คณิตศาสตร์ภายใต้การเปลี่ยนแปลงของโลก
MATHEMATICS IN A CHANGING WORLD

CONFERENCE PROCEEDINGS

AMM
28TH ANNUAL MEETING IN MATHEMATICS 2024

การประชุมวิชาการทางคณิตศาสตร์ ครั้งที่ 28

29-31 พฤษภาคม 2567

จัดโดย สมาคมคณิตศาสตร์แห่งประเทศไทย ในพระบรมราชูปถัมภ์
ร่วมกับ คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี

AMM2024

สารจากอธิการบดี มหาวิทยาลัยอุบลราชธานี

มหาวิทยาลัยอุบลราชธานี มีวิสัยทัศน์ในการเป็นมหาวิทยาลัยชั้นนำในอาเซียน ที่ยกระดับคุณภาพชีวิตให้แก่สังคม โดยมีพันธกิจ 4 ด้านประกอบไปด้วย ด้านที่1 สร้างบัณฑิตที่มีสมรรถนะสูง มีทักษะการเป็นผู้ประกอบการ สามารถปฏิบัติงานได้จริงเพื่อตอบโจทย์อุตสาหกรรม ด้านที่2 สร้างองค์ความรู้และนวัตกรรมเพื่อพัฒนาคุณภาพชีวิตและสร้างมูลค่าเพิ่มให้กับเศรษฐกิจและสังคม ด้านที่3 บริการวิชาการ เพื่อถ่ายทอดองค์ความรู้ เทคโนโลยีและนวัตกรรมที่ตอบสนองความต้องการของสังคมและภาคอุตสาหกรรม และด้านที่4 ส่งเสริมวัฒนธรรมและภูมิปัญญาอีสานใต้อย่างสร้างสรรค์เพื่อสร้างมูลค่าเพิ่มทางเศรษฐกิจ ซึ่งในการจัดการประชุมวิชาการทางคณิตศาสตร์ ครั้งที่ 28 ประจำปี 2567 The 28th Annual Meeting in Mathematics 2024 (AMM2024) ในหัวข้อ Mathematics in a Changing World (คณิตศาสตร์ภายใต้การเปลี่ยนแปลงของโลก) ซึ่งมหาวิทยาลัยอุบลราชธานีโดยภาควิชาคณิตศาสตร์ สถิติและคอมพิวเตอร์ คณะวิทยาศาสตร์ได้เป็นเจ้าภาพ เพื่อนำเสนอผลงานวิจัยทางด้านกลุ่มคณิตศาสตร์และคณิตศาสตร์ประยุกต์ สถิติ สถิติประยุกต์ วิทยาการข้อมูล คณิตศาสตรศึกษา และกลุ่มวิจัยอื่น ๆ ที่เกี่ยวข้อง ซึ่งเป็นเวทีสำหรับนักวิจัยทั้งระดับชาติและนานาชาติได้เผยแพร่งานวิจัยและแลกเปลี่ยน องค์ความรู้ เพื่อจะนำไปสู่คณิตศาสตร์ภายใต้การเปลี่ยนแปลงโลกต่อไป

ในการจัดงานครั้งนี้เป็นกิจกรรมทางวิชาการที่มีความสำคัญที่จะช่วยส่งเสริมสนับสนุนการพัฒนาคุณภาพการศึกษาและงานวิจัยตลอดจนพัฒนาองค์ความรู้จากการวิจัยที่มีคุณภาพไปสู่การพัฒนาและประยุกต์ใช้เพื่อเป็นประโยชน์ทั้งต่อองค์กร สังคมและประเทศชาติ อีกทั้งยังสอดรับกับวิสัยทัศน์และพันธกิจของมหาวิทยาลัยอุบลราชธานีด้วย การประชุมวิชาการครั้งนี้ได้เปิดโอกาสให้ นักศึกษา คณาจารย์และนักวิชาการได้นำเสนอผลงานวิจัยต่อที่ประชุม เพื่อเผยแพร่ผลงานวิจัยสู่สาธารณะชน อันจะทำให้เกิดการแลกเปลี่ยนความคิดเห็นระหว่างนักวิจัยในสาขาวิชาต่าง ๆ ที่เกี่ยวข้อง ทั้งในสถาบันการศึกษาเดียวกันและระหว่างสถาบันการศึกษาอันจะนำไปสู่การพัฒนาคุณภาพงานวิจัยต่อไป

ดิฉันหวังเป็นอย่างยิ่งว่าการประชุมวิชาการในครั้งนี้จะเป็นอีกก้าวหนึ่งที่เปิดโอกาสให้กับอาจารย์ นักวิจัย นิสิตนักศึกษา ของมหาวิทยาลัยต่างๆ ตลอดจนผู้สนใจทุกท่านแลกเปลี่ยนเรียนรู้ร่วมกัน เพื่อเป็นเครือข่ายการสร้างสรรค์งานวิจัย และสามารถนำองค์ความรู้ที่ได้จากงานวิจัยไปประยุกต์ใช้ให้เกิดประสิทธิภาพและประสิทธิผลอย่างแท้จริง กับสังคมและประเทศชาติในอนาคต

(รองศาสตราจารย์ ดร.ชุตินันท์ ประสิทธิ์ภูริปรีชา)

อธิการบดีมหาวิทยาลัยอุบลราชธานี

สารจากนายกสมาคมคณิตศาสตร์แห่งประเทศไทย ในพระบรมราชูปถัมภ์

ในนามของสมาคมคณิตศาสตร์แห่งประเทศไทย ในพระบรมราชูปถัมภ์ ดิฉันขอแสดงความยินดีเป็นอย่างยิ่ง ที่การจัดประชุมวิชาการคณิตศาสตร์ครั้งที่ 28 ประจำปี 2567 (AMM 2024) ภายใต้หัวข้อ “Mathematics in a Changing World หรือ คณิตศาสตร์ภายใต้การเปลี่ยนแปลงของโลก” สำเร็จไปได้ด้วยดี โดยความร่วมมือร่วมแรงร่วมใจของบุคลากรภาควิชาคณิตศาสตร์ สถิติ และคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี เป็นสำคัญ

การจัดการประชุมวิชาการประจำปี หรือ AMM ของชาวคณิตศาสตร์จากทั่วประเทศที่ศูนย์ส่งเสริมการวิจัยคณิตศาสตร์แห่งประเทศไทย (CEPMART) ภายใต้สมาคมคณิตศาสตร์แห่งประเทศไทย ในพระบรมราชูปถัมภ์ และมหาวิทยาลัยต่าง ๆ ทั่วประเทศ มีส่วนร่วมด้วยนั้น มีความมุ่งหมายหลักเพื่อให้เป็นเวทีในการนำเสนอผลงานวิจัยใหม่ ๆ ทางคณิตศาสตร์ คณิตศาสตร์ประยุกต์ สถิติ สถิติประยุกต์ วิทยาการคอมพิวเตอร์ วิทยาการข้อมูล และคณิตศาสตร์ศึกษา อีกทั้งยังเป็นเวทีในการแลกเปลี่ยนเรียนรู้ความก้าวหน้าทางวิชาการ ด้านคณิตศาสตร์แขนงต่าง ๆ โดยมีกำหนดจัดประชุมเป็นประจำทุกปีในการประชุมครั้งนี้ มีการบรรยายพิเศษ และการเสวนาทางวิชาการโดยวิทยากรผู้ทรงคุณวุฒิจากทั่วทุกมุมโลกเหมือนเช่นเคย สมาคมคณิตศาสตร์แห่งประเทศไทย ในพระบรมราชูปถัมภ์ หวังเป็นอย่างยิ่งว่าครู อาจารย์ นิสิต นักศึกษา และผู้สนใจคณิตศาสตร์ตลอดจนศาสตร์ที่เกี่ยวข้อง ที่ได้เข้าร่วมประชุมวิชาการในครั้งนี้ จะได้รับประโยชน์จากการนำประสบการณ์ที่ได้รับไปพัฒนา และต่อยอดองค์ความรู้ อีกทั้งนำไปถ่ายทอดให้แพร่หลายต่อไปด้วย

สมาคมคณิตศาสตร์แห่งประเทศไทย ในพระบรมราชูปถัมภ์ ขอขอบคุณภาควิชาคณิตศาสตร์ สถิติ และคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี ที่ให้เกียรติเป็นเจ้าภาพการจัดการประชุม AMM ครั้งนี้ และขอขอบคุณทุกท่านที่มีส่วนร่วมในการจัดประชุมด้วยความวิริยะอุตสาหะยิ่งสุดท้ายนี้ ขอให้การประชุมวิชาการครั้งนี้ เป็นดั่งสะพานเชื่อมไปสู่ความร่วมมือกันระหว่างนักคณิตศาสตร์ทุกแขนงจากสถาบันและองค์กรต่าง ๆ ทั่วประเทศและทั่วโลก และนำไปสู่การประยุกต์ใช้ความรู้ ความสามารถไปพัฒนานวัตกรรมด้านวิทยาศาสตร์ คณิตศาสตร์ และเทคโนโลยี เพื่อความเจริญรุ่งเรืองที่ยั่งยืนของประเทศชาติอันเป็นที่รักของพวกเราสืบไป

(ศาสตราจารย์กิตติคุณ ดร. พัฒนี อุดมกะวานิช)

นายกสมาคมคณิตศาสตร์แห่งประเทศไทย ในพระบรมราชูปถัมภ์

สารจากผู้อำนวยการศูนย์ส่งเสริมการวิจัยคณิตศาสตร์แห่งประเทศไทย

การจัดประชุมวิชาการทางคณิตศาสตร์ประจำปี (Annual Meeting in Mathematics) เป็นการประชุมที่สำคัญของประชาคมชาวคณิตศาสตร์ในประเทศไทยที่จะมาพบกันเพื่อฟังการบรรยายจากผู้เชี่ยวชาญเฉพาะด้าน ในสาขาต่าง ๆ ทางคณิตศาสตร์และสาขาที่เกี่ยวข้องกับคณิตศาสตร์ เพื่อให้เกิดการตระหนักรู้ของ ความก้าวหน้าและวิทยาการใหม่ ๆ รวมไปถึงการนำคณิตศาสตร์ไปใช้ในด้านต่าง ๆ และยังเป็นเวทีให้นักวิจัย ทั้งรุ่นเก่าและรุ่นใหม่ได้นำเสนอผลงาน เพื่อแลกเปลี่ยนเรียนรู้กับผู้ที่สนใจที่อยู่ต่างสถาบันกัน ซึ่งอาจจะนำไปสู่ ความร่วมมือทางด้านงานวิจัยต่อไปในอนาคต

สถาบันที่มีหลักสูตรคณิตศาสตร์ในประเทศไทยได้ร่วมมือหมุนเวียนกันเป็นเจ้าภาพร่วมจนถึงครั้งนี้ โดยภาควิชาคณิตศาสตร์ สถิติและคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี ได้รับเป็นเจ้าภาพจัดการประชุมวิชาการทางคณิตศาสตร์ ประจำปี 2567 โดยเป็นการประชุมวิชาการระดับชาติ ครั้งที่ 28 ระหว่างวันที่ 29 - 31 พฤษภาคม 2567 ในหัวข้อ “Mathematics in a Changing World คณิตศาสตร์ภายใต้การเปลี่ยนแปลงของโลก” ทั้งนี้ การจัดประชุมวิชาการทางคณิตศาสตร์ จะเป็นเวทีสำหรับนักวิจัยทั้งระดับชาติและระดับนานาชาติได้เผยแพร่งานวิจัย และแลกเปลี่ยนองค์ความรู้ เพื่อจะนำไปสู่คณิตศาสตร์ภายใต้การเปลี่ยนแปลงของโลก ต่อไป

ในนามของผู้อำนวยการศูนย์ส่งเสริมการวิจัยคณิตศาสตร์แห่งประเทศไทย (CEPMART) ขอขอบคุณคณะทำงาน ภาควิชาคณิตศาสตร์ สถิติและคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี เป็นอย่างยิ่ง ที่ให้ความกรุณารับเป็นเจ้าภาพการจัดการประชุมทางคณิตศาสตร์ประจำปี ในครั้งนี้ และขออวยพรให้การจัดงานครั้งนี้สำเร็จราบรื่นด้วยดีทุกประการ และขอให้คณะทำงาน ทุกท่านมีสุขภาพแข็งแรงทั้งกายและใจ มีกำลังใจในการทำงาน ก้าวผ่านปัญหาและอุปสรรคต่าง ๆ อย่างราบรื่นด้วยดี

(รองศาสตราจารย์ ดร.ศจี เพียรสกุล)
ผู้อำนวยการศูนย์ส่งเสริมการวิจัยคณิตศาสตร์แห่งประเทศไทย

สารจากคณบดีคณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี

คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี มีวิสัยทัศน์ในการเป็นสถาบันชั้นนำด้านวิจัยวิทยาศาสตร์ระดับประเทศ โดยมีพันธกิจ 3 ด้านประกอบไปด้วย ด้านที่ 1 ผลิตบัณฑิตที่พึงประสงค์ด้านวิทยาศาสตร์และเทคโนโลยีที่มีความโดดเด่นทางด้านทักษะดิจิทัล (Digital Literacy and Accessibility) ด้านที่ 2 ผลิตงานวิจัยที่เป็นที่ยอมรับในระดับสากลและสร้างนวัตกรรมเพื่อตอบโจทย์ความต้องการของประเทศ และสร้างความยั่งยืนให้ชุมชน และด้านที่ 3 บริการวิชาการตอบโจทย์ความต้องการของผู้รับบริการ สร้างคุณค่าร่วมกับสังคมเพื่อการพัฒนาที่ยั่งยืนซึ่งในการจัดการประชุมวิชาการทางคณิตศาสตร์ ครั้งที่ 28 ประจำปี 2567 The 28th Annual Meeting in Mathematics 2024 (AMM2024) ในหัวข้อ Mathematics in a Changing World (คณิตศาสตร์ภายใต้การเปลี่ยนแปลงของโลก) โดยภาควิชาคณิตศาสตร์ สถิติและคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานีมหาวิทยาลัยได้เป็นเจ้าภาพ เพื่อนำเสนอผลงานวิจัยทางด้านกลุ่มคณิตศาสตร์และคณิตศาสตร์ประยุกต์ สถิติ สถิติประยุกต์ วิทยาการข้อมูล คณิตศาสตรศึกษา และกลุ่มวิจัยอื่น ๆ ที่เกี่ยวข้อง ซึ่งเป็นเวทีสำหรับนักวิจัยทั้งระดับชาติและ นานาชาติได้เผยแพร่งานวิจัยและแลกเปลี่ยน องค์ความรู้ เพื่อจะนำไปสู่คณิตศาสตร์ภายใต้การเปลี่ยนแปลงโลกต่อไป

ในการจัดงานครั้งนี้เป็นกิจกรรมทางวิชาการที่มีความสำคัญที่จะช่วยส่งเสริมสนับสนุนการพัฒนาคุณภาพการศึกษาและงานวิจัยตลอดจนพัฒนาองค์ความรู้จากการวิจัยที่มีคุณภาพไปสู่การพัฒนาและประยุกต์ใช้เพื่อเป็นประโยชน์ทั้งต่อองค์กร สังคมและประเทศชาติ อีกทั้งยังสอดรับกับวิสัยทัศน์และพันธกิจของคณะวิทยาศาสตร์ด้วย การประชุมวิชาการครั้งนี้ได้เปิดโอกาสให้ นักศึกษา คณาจารย์และนักวิชาการได้นำเสนอผลงานวิจัยต่อที่ประชุม เพื่อเผยแพร่ผลงานวิจัยสู่สาธารณะชน อันจะทำให้เกิดการแลกเปลี่ยนความคิดเห็นระหว่างนักวิจัยในสาขาวิชาต่าง ๆ ที่เกี่ยวข้อง ทั้งในสถาบันการศึกษาเดียวกันและระหว่างสถาบันการศึกษาอันจะนำไปสู่การพัฒนาคุณภาพงานวิจัยต่อไป

ดิฉันหวังเป็นอย่างยิ่งว่าการประชุมวิชาการในครั้งนี้จะเป็นอีกก้าวหนึ่งที่เปิดโอกาสให้กับอาจารย์ นักวิจัย นิสิตนักศึกษา ของมหาวิทยาลัยต่างๆ ตลอดจนผู้สนใจทุกท่านแลกเปลี่ยนเรียนรู้ร่วมกัน เพื่อเป็นเครือข่ายการสร้างสรรค์งานวิจัย และสามารถนำองค์ความรู้ที่ได้จากงานวิจัยไปประยุกต์ใช้ให้เกิดประสิทธิภาพและประสิทธิผลอย่างแท้จริง กับสังคมและประเทศชาติในอนาคต

(ศาสตราจารย์ ดร.ศิริพร จึงสุทธิวงษ์)

คณบดีคณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี

คำนำ

สมาคมคณิตศาสตร์แห่งประเทศไทยในพระบรมราชูปถัมภ์ โดยศูนย์ส่งเสริมการวิจัยคณิตศาสตร์แห่งประเทศไทย (Center for Promotion of Mathematical Research of Thailand) ได้เริ่มจัดประชุมวิชาการทางคณิตศาสตร์ระดับประเทศ ตั้งแต่ปีพุทธศักราช 2538 และจัดต่อเนื่องเป็นประจำทุกปี โดยมีภาควิชา คณิตศาสตร์ของมหาวิทยาลัยต่าง ๆ หมุนเวียนกันเป็นเจ้าภาพ

ในปีพุทธศักราช 2567 สาขาวิชาคณิตศาสตร์ ภาควิชาคณิตศาสตร์ สถิติ และคอมพิวเตอร์ คณะวิทยาศาสตร์มหาวิทยาลัยอุบลราชธานี ได้รับมอบหมายจากศูนย์ส่งเสริมการวิจัยคณิตศาสตร์แห่งประเทศไทยให้เป็นเจ้าภาพจัดการประชุมวิชาการทางคณิตศาสตร์ระดับชาติ ครั้งที่ 28 ประจำปีพุทธศักราช 2567 ในหัวข้อ “Mathematics in a changing world คณิตศาสตร์ในโลกที่กำลังเปลี่ยนแปลง” ระหว่างวันที่ 29 – 31 พฤษภาคม 2567 โดยกลุ่มงานวิจัยที่สามารถนำเสนอได้ในการประชุมวิชาการครั้งนี้ได้แก่ กลุ่มคณิตศาสตร์และคณิตศาสตร์ประยุกต์ กลุ่มสถิติ สถิติประยุกต์ และวิทยาการข้อมูล กลุ่มคณิตศาสตรศึกษา และกลุ่มวิจัยอื่น ๆ ที่เกี่ยวข้อง ทั้งนี้มีจุดมุ่งหมายเพื่อที่จะให้ครู อาจารย์ นักวิชาการ นักวิจัย นิสิตและนักศึกษา รวมทั้งผู้ที่สนใจและทำงานที่เกี่ยวข้องกับคณิตศาสตร์ในสาขาต่าง ๆ ได้มาพบปะ แลกเปลี่ยนความรู้และประสบการณ์ทางด้านการทำวิจัย การเรียนการสอนทางคณิตศาสตร์ ซึ่งทำให้เกิดความร่วมมือในการทำงานทางคณิตศาสตร์ระหว่างสถาบัน และเสริมสร้างความเข้มแข็งทางวิชาการด้านคณิตศาสตร์

การประชุมครั้งนี้ได้รับเกียรติจากผู้ทรงคุณวุฒิมาเป็นวิทยากรบรรยายพิเศษจำนวน 7 ท่าน ได้แก่ Professor Dr. Malgorzata Peszynska Mr. Alain Jean Alherbe รองศาสตราจารย์ ดร.ธีระเดช เจียรสุขสกุล รองศาสตราจารย์ ดร.ชัชวาล ปานรักษา รองศาสตราจารย์ ดร.สายันต์ แก่นนาคำ ผู้ช่วยศาสตราจารย์ ดร.วีระชัย สาระคร และ ดร.วุฒิศักดิ์ ตรงศิริวัฒน์ นอกจากนี้ยังมีเสวนาวิชาการในหัวข้อคณิตศาสตร์ภายใต้การเปลี่ยนแปลงของโลก โดยมีผู้ร่วมเสวนาคือ รองศาสตราจารย์ ดร.กิตติกร นาคประสิทธิ์ รองศาสตราจารย์ ดร.นพรัตน์ โพธิ์ชัย รองศาสตราจารย์ ดร.รตินันท์ บุญเคลือบ และว่าที่ ร.อ. ดร.ภณัฐ ก้วยเจริญพานิชก์ เป็นพิธีกรดำเนินรายการ รวมทั้งมีการนำเสนอผลงานภาคบรรยายทางคณิตศาสตร์ในกลุ่มคณิตศาสตร์และคณิตศาสตร์ประยุกต์ กลุ่มสถิติ สถิติประยุกต์ และวิทยาการข้อมูล กลุ่มคณิตศาสตรศึกษาและกลุ่มวิจัยอื่น ๆ ที่เกี่ยวข้อง จำนวน 82 ผลงาน และมีผู้เข้าร่วมประชุมประมาณ 210 คน โดยบทคัดย่อของผลงานนำเสนอทั้งหมดจะถูกตีพิมพ์ในหนังสือรวบรวมบทคัดย่อ (Book of Abstracts) และบางผลงานวิจัยฉบับเต็ม (Full Papers) จะถูกนำไปตีพิมพ์ในเอกสารสืบเนื่องการประชุม (Proceedings) วารสารวิทยาศาสตร์และวิทยาศาสตร์ศึกษา วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยอุบลราชธานี Thai Journal of Mathematics (Special Issue: AMM 2024)

ภาควิชาคณิตศาสตร์ สถิติ และคอมพิวเตอร์ ขอขอบคุณสมาคมคณิตศาสตร์แห่งประเทศไทยในพระบรมราชูปถัมภ์ ศูนย์ส่งเสริมการวิจัยคณิตศาสตร์แห่งประเทศไทย และคณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี ที่เป็นส่วนหนึ่งในการสนับสนุนงบประมาณในการจัดประชุมครั้งนี้ นอกจากนี้ ยังขอขอบคุณแขกรับเชิญ วิทยากรบรรยายพิเศษ ผู้เข้าร่วมเสวนา ผู้ทรงคุณวุฒิพิจารณาบทความ ประธานนำเสนอในแต่ละห้อง ผู้นำเสนอผลงาน และผู้เข้าร่วมประชุมทุกท่าน สุดท้ายนี้ขอขอบคุณกรรมการดำเนินงานทุกท่านที่อุทิศเวลาแรงกาย แรงใจ อย่างสุดความสามารถจนทำให้การประชุมครั้งนี้ประสบความสำเร็จดังวัตถุประสงค์ ทั้งนี้หากเอกสารฉบับนี้มีข้อบกพร่อง รวมทั้งการจัดประชุมมีข้อบกพร่องประการใด ทางภาควิชา ฯ ขออภัยท่านไว้ ณ ที่นี้ และขอน้อมรับคำติชมจากทุกท่านเพื่อนำไปปรับปรุงในการจัดประชุมในโอกาสต่อไป

(รองศาสตราจารย์ศราวุธ แสนการุณ)

ประธานดำเนินงาน

สารบัญ

หน้า

AMM
28th ANNUAL MEETING IN MATHEMATICS 2024

SCHEDULE
กำหนดการ

การประชุมวิชาการทางคณิตศาสตร์ ครั้งที่ 28 ประจำปี 2567

The 28th Annual Meeting in Mathematics (AMM 2024)

Mathematics in a Changing World (คณิตศาสตร์ภายใต้การเปลี่ยนแปลงของโลก)

วันที่ 29 – 31 พฤษภาคม พ.ศ. 2567

ณ อาคารเฉลิมพระเกียรติ 7 รอบพระชนมพรรษา มหาวิทยาลัยอุบลราชธานี

วันพุธที่ 29 พฤษภาคม พ.ศ. 2567	
8.00 น.	รถบัสรับส่งผู้เข้าร่วมประชุมออกเดินทางจากโรงแรมบ้านสวนคุณตา กอล์ฟ แอนด์ รีสอร์ท โรงแรมอยู่ด้วยกัน การ์เด้น โฮม และโรงแรมแหวนเพชรเพลส ไปยังอาคารเฉลิมพระเกียรติ 7 รอบพระชนมพรรษา มหาวิทยาลัยอุบลราชธานี
8.00 – 9.00 น.	ลงทะเบียนเข้าร่วมงาน
9.00 – 9.30 น.	**กล่าวรายงาน** โดย รองศาสตราจารย์ ดร.ศราวุธ แสนการุณ ประธานจัดการประชุมฯ **พิธีเปิด** โดยอธิการบดี มหาวิทยาลัยอุบลราชธานี **กล่าวต้อนรับ** โดย - ผู้อำนวยการศูนย์ส่งเสริมการวิจัยคณิตศาสตร์แห่งประเทศไทย (CEPMART) - คณบดีคณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี
9.30 – 9.45 น.	ถ่ายรูปหมู่ร่วมกัน
9.45 – 10.00 น.	พักรับประทานอาหารว่าง
10.00 – 11.00 น.	การบรรยายพิเศษแบบ online เรื่อง Underground Computational Mathematics: Models and Analyses of an Evolving Subsurface of Planet Earth โดย Professor Dr. Malgorzata Peszynska, Oregon State University, USA
11.00 – 12.00 น.	การบรรยายพิเศษ เรื่อง Safeguarding Data Privacy: Exploring Full Homomorphic Encryption โดย Mr. Alain Jean Alherbe มหาวิทยาลัยอุบลราชธานี
12.00 – 13.00 น.	พักรับประทานอาหารกลางวัน
13.00 – 13.50 น.	การบรรยายพิเศษเรื่อง Arithmetic Dynamics: Bridging Order and Chaos โดย รองศาสตราจารย์ ดร.ชัชวาล ปานรักษา มหาวิทยาลัยมหิดล
	การบรรยายพิเศษเรื่อง Unleashing the Potential of Applied Mathematics in AI and Machine Learning for Modern Industry โดย รองศาสตราจารย์ ดร.สายันต์ แก่นนาคำ มหาวิทยาลัยเทคโนโลยีสุรนารี
13.50 – 14.50 น.	การนำเสนอผลงานกลุ่มย่อย
14.50 – 15.10 น.	พักรับประทานอาหารว่าง
15.10 – 16.30 น.	การนำเสนอผลงานกลุ่มย่อย

16.30 – 18.00 น.	เยี่ยมชม เฮือนกำนันคาเฟ่ หนองอีเจม
14.00 – 17.00 น.	ประชุมคณะกรรมการศูนย์ส่งเสริมการวิจัยคณิตศาสตร์แห่งประเทศไทย (CEPMART)
18.00 – 20.00 น.	งานเลี้ยงรับรอง
20.00 น.	รถบัสรับส่งผู้เข้าร่วมประชุมออกเดินทางจากอาคารเฉลิมพระเกียรติ 7 รอบพระชนมพรรษา มหาวิทยาลัยอุบลราชธานี ไปยังโรงแรมบ้านสวนคุณตา กอล์ฟ แอนด์ รีสอร์ท โรงแรมอยู่ด้วยกัน การ์เด้น โฮม และโรงแรมแหวนเพชรเพลส
วันพฤหัสบดีที่ 30 พฤษภาคม พ.ศ. 2567	
8.00 น.	รถบัสรับส่งผู้เข้าร่วมประชุมออกเดินทางจากโรงแรมบ้านสวนคุณตา กอล์ฟ แอนด์ รีสอร์ท โรงแรมอยู่ด้วยกัน การ์เด้น โฮม และโรงแรมแหวนเพชรเพลส ไปยังอาคารเฉลิมพระเกียรติ 7 รอบพระชนมพรรษา มหาวิทยาลัยอุบลราชธานี
9.00 – 10.00 น.	การบรรยายพิเศษเรื่อง การพัฒนาสมรรถนะด้านคณิตศาสตร์ของ PISA ให้กับครูและนักเรียนในยุคดิจิทัล โดย รองศาสตราจารย์ ดร.ธีระเดช เจียรสุขสกุล ผู้อำนวยการสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสวท.)
10.00 – 10.20 น.	พักรับประทานอาหารว่าง
10.20 – 12.00 น.	การนำเสนอผลงานกลุ่มย่อย
12.00 – 13.00 น.	พักรับประทานอาหารกลางวัน
13.00 – 13.50 น.	การบรรยายพิเศษเรื่อง KKU Smart Mathematics Learning Platform for Secondary Schools โดย ผู้ช่วยศาสตราจารย์ ดร.วีระชัย สาระคร มหาวิทยาลัยขอนแก่น
	การบรรยายพิเศษเรื่อง Decoding Modern Banking: A Mathematician's Guide โดย ดร.วุฒิศักดิ์ ตรงศิริวัฒน์ รองผู้อำนวยการฝ่าย Data Innovation ธนาคารกรุงไทย
13.50 – 14.50 น.	การนำเสนอผลงานกลุ่มย่อย
14.50 – 15.10 น.	พักรับประทานอาหารว่าง
15.10 – 16.30 น.	การนำเสนอผลงานกลุ่มย่อย
16.30 น.	รถบัสรับส่งผู้เข้าร่วมประชุมออกเดินทางจากอาคารเฉลิมพระเกียรติ 7 รอบพระชนมพรรษา มหาวิทยาลัยอุบลราชธานี ไปยังโรงแรมบ้านสวนคุณตา กอล์ฟ แอนด์ รีสอร์ท โรงแรมอยู่ด้วยกัน การ์เด้น โฮม และโรงแรมแหวนเพชรเพลส
วันศุกร์ที่ 31 พฤษภาคม พ.ศ. 2567	
8.00 น.	รถบัสรับส่งผู้เข้าร่วมประชุมออกเดินทางจากโรงแรมบ้านสวนคุณตา กอล์ฟ แอนด์ รีสอร์ท โรงแรมอยู่ด้วยกัน การ์เด้น โฮม และโรงแรมแหวนเพชรเพลส ไปยังอาคารเฉลิมพระเกียรติ 7 รอบพระชนมพรรษา มหาวิทยาลัยอุบลราชธานี
9.00 – 10.00 น.	การนำเสนอผลงานกลุ่มย่อย
10.00 – 10.10 น.	พักรับประทานอาหารว่าง

10.10 – 11.30 น.	เสวนาวิชาการ: Mathematics in a Changing World (คณิตศาสตร์ภายใต้การเปลี่ยนแปลงของโลก) รองศาสตราจารย์ ดร.กิตติกร นาคประสิทธิ์ มหาวิทยาลัยขอนแก่น รองศาสตราจารย์ ดร.นพรัตน์ โพธิ์ชัย สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง รองศาสตราจารย์ ดร.รตินันท์ บุญเคลือบ จุฬาลงกรณ์มหาวิทยาลัยและเลขาธิการสมาคมคณิตศาสตร์แห่งประเทศไทย ในพระบรมราชูปถัมภ์ ว่าที่ ร.อ. ดร.ภณัฐ ก้วยเจริญพานิชก์ สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสวท.) พิธีกรดำเนินรายการ
11.30 – 12.00 น.	**พิธีปิด** โดยนายกสมาคมคณิตศาสตร์แห่งประเทศไทย ในพระบรมราชูปถัมภ์ **พิธีมอบธง** - มอบธงจากมหาวิทยาลัยอุบลราชธานีโดยคณบดีคณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานีสู่สมาคมคณิตศาสตร์แห่งประเทศไทย ในพระบรมราชูปถัมภ์ รับโดยนายกสมาคมคณิตศาสตร์แห่งประเทศไทย ในพระบรมราชูปถัมภ์ - มอบธงจากสมาคมคณิตศาสตร์แห่งประเทศไทยโดยนายกสมาคมคณิตศาสตร์แห่งประเทศไทย ในพระบรมราชูปถัมภ์สู่มหาวิทยาลัยศรีนครินทรวิโรฒ
12.00 – 13.00 น.	พักรับประทานอาหารกลางวัน
12.00 น. และ 13.00 น.	รถบัสรับส่งผู้เข้าร่วมประชุมออกเดินทางจากอาคารเฉลิมพระเกียรติ 7 รอบพระชนมพรรษา มหาวิทยาลัยอุบลราชธานีไปยังสนามบิน (แวะซื้อของฝาก) และเดินทางกลับโดยสวัสดิภาพ

หมายเหตุ 1. **ไม่มี**รถรับส่งจากสนามบินมายังโรงแรม หรือสถานที่ประชุม (อาจใช้บริการรถแท็กซี่ตรงหน้าทางออกอาคารสนามบิน)

2. มีรถรับส่งจากสถานที่ประชุมไปยังสนามบิน ในวันศุกร์ที่ 31 พฤษภาคม 2567 เวลา 12.00 น. และ 13.00 น.

3. ตารางอาจมีการเปลี่ยนแปลงตามความเหมาะสม

AMM 28th ANNUAL MEETING IN MATHEMATICS 2024

กำหนดการ
นำเสนอผลงาน

กำหนดการนำเสนอผลงาน (รหัสบทความ)

การประชุมวิชาการทางคณิตศาสตร์ ครั้งที่ 28 ประจำปี 2567

วันที่ 29 – 31 พฤษภาคม พ.ศ. 2567

ณ อาคารเฉลิมพระเกียรติ 7 รอบพระชนมพรรษา มหาวิทยาลัยอุบลราชธานี

วันพุธที่ 29 พฤษภาคม พ.ศ. 2567

เวลา	รายการ			
08.00 – 09.00 น.	**ลงทะเบียน**			
09.00 – 09.30 น.	**พิธีเปิด**			
09.30 – 09.45 น.	**ถ่ายรูปหมู่ร่วมกัน**			
09.45 – 10.00 น.	พักรับประทานอาหารว่าง			
10.00 – 11.00 น.	**Keynote Speaker I (KNS-01)** ห้องกันเกรา			
11.00 – 12.00 น.	**Keynote Speaker II (KNS-02)** ห้องกันเกรา			
12.00 – 13.00 น.	พักรับประทานอาหารกลางวัน			
	ห้องกันเกรา		**ห้องพวงพะยอม**	
13.00 – 13.50 น.	Invited Speaker I (IVS-01)		Invited Speaker II (IVS-02)	
	ห้องพวงพะยอม	**ห้องประดู่ 1**	**ห้องประดู่ 2**	**ห้องประดู่ 3**
13.50 – 14.10 น.	PTS-01	ANA-01	DNM-01	NUT-01
14.10 – 14.30 น.	PTS-02	ANA-02	DNM -02	NUT-02
14.30 – 14.50 น.	PTS-03	ANA-03	DNM -03	NUT-03
14.50 – 15.10 น.	พักรับประทานอาหารว่าง			
15.10 – 15.30 น.	CGT-01	ANA-04	DNM-04	ALG-01
15.30 – 15.50 น.	CGT-02	ANA-05	DNM-05	ALG-02
15.50 – 16.10 น.	CGT-03	ANA-06	MMF-01	ALG-03
16.10 – 16.30 น.	CGT-04			ALG-04
16.30 – 18.00 น.	**เยี่ยมชม เฮือนกำนันคาเฟ่ หนองอีเจม**			
14.00 – 17.00 น.	**ประชุมคณะกรรมการ CEPMART**			
18.00 – 20.00 น.	**งานเลี้ยงรับรอง**			

วันพฤหัสบดีที่ 30 พฤษภาคม พ.ศ. 2567

เวลา	รายการ			
09.00 – 10.00 น.	**Keynote Speaker III (KNS-03)** ห้องกันเกรา			
10.00 – 10.20 น.	พักรับประทานอาหารว่าง			
	ห้องพวงพะยอม	**ห้องประดู่ 1**	**ห้องประดู่ 2**	**ห้องประดู่ 3**
10.20 – 10.40 น.	NUT-04	MED-01	DCS-01	ALG-05
10.40 – 11.00 น.	NUT-05	MED-02	DCS-02	ALG-06
11.00 – 11.20 น.	NUT-06	MED-03	DCS-03	ALG-07
11.20 – 11.40 น.	NUT-07	MED-04	DCS-04	ALG-08
11.40 – 12.00 น.	NUT-08	MED-05	DCS-05	ALG-09
12.00 – 13.00 น.	พักรับประทานอาหารกลางวัน			
	ห้องกันเกรา		**ห้องพวงพะยอม**	
13.00 – 13.50 น.	Invited Speaker III (IVS-03)		Invited Speaker IV (IVS-04)	
	ห้องพวงพะยอม	**ห้องประดู่ 1**	**ห้องประดู่ 2**	**ห้องประดู่ 3**
13.50 – 14.10 น.	PTS-04	ANA-07	MMF-02	ALG-10
14.10 – 14.30 น.	PTS-05	ANA-08	MMF-03	CGT-05
14.30 – 14.50 น.	PTS-06	ANA-09	MMF-08	CGT-06
14.50 – 15.10 น.	พักรับประทานอาหารว่าง			
15.10 – 15.30 น.	NUT-09	CGT-07	MMF-05	ALG-11
15.30 – 15.50 น.	NUT-10	CGT-08	ORT-02	ALG-12
15.50 – 16.10 น.	NUT-11	CGT-09	ORT-03	ALG-13
16.10 – 16.30 น.				ALG-14

วันศุกร์ที่ 31 พฤษภาคม พ.ศ. 2567

เวลา	รายการ			
	ห้องพวงพะยอม	**ห้องประดู่ 1**	**ห้องประดู่ 2**	**ห้องประดู่ 3**
09.00 – 09.20 น.	PTS -07	ANA-10	MMF-06	CGT-10
09.20 – 09.40 น.	PTS -08	~~ANA-11~~	MMF-07	CGT-11
09.40 – 10.00 น.	PTS -09	ORT-01	MMF-04	CGT-12
10.00 – 10.10 น.	พักรับประทานอาหารว่าง			
10.10 – 11.30 น.	**เสวนาวิชาการ: Mathematics in a Changing World** **(คณิตศาสตร์ภายใต้การเปลี่ยนแปลงของโลก)** ห้องกันเกรา			
11.30 – 12.00 น.	**พิธีปิด**			
12.00 – 13.00 น.	พักรับประทานอาหารกลางวัน			

คำอธิบายอักษรย่อ

KNS	Keynote Speakers
IVS	Invited Speakers
ALG	Algebra
ANA	Analysis, Fixed Point Theory and Applications, Topology and Geometry
CGT	Combinatorics and Graph Theory
DCS	Data Science and Computer Science
DNM	Differential Equations and Numerical Mathematics
MMF	Mathematical Modeling and Mathematical Finance
MED	Mathematics Education
NUT	Number Theory
ORT	Other Related Topics in Mathematics
PTS	Probability Theory and Statistics

รายการรหัสบทความนำเสนอผลงาน

1. Keynote Speakers (KNS)

KNS-01	Underground Computational Mathematics: Models and Analyses of an Evolving Subsurface of Planet Earth ***Professor Dr. Malgorzata Peszynska****, Oregon State University, USA*
KNS-02	Safeguarding Data Privacy: Exploring Full Homomorphic Encryption ***Mr. Alain Jean Alherbe*** *มหาวิทยาลัยอุบลราชธานี*
KNS-03	การพัฒนาสมรรถนะด้านคณิตศาสตร์ของ PISA ให้กับครูและนักเรียนในยุคดิจิทัล ***รองศาสตราจารย์ ดร.ธีระเดช เจียรสุขสกุล*** *ผู้อำนวยการสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสวท.)*

2. Invited Speakers (IVS)

IVS-01	Arithmetic Dynamics: Bridging Order and Chaos ***รองศาสตราจารย์ ดร.ชัชวาล ปานรักษา*** *มหาวิทยาลัยมหิดล*
IVS-02	Unleashing the Potential of Applied Mathematics in AI and Machine Learning for Modern Industry ***รองศาสตราจารย์ ดร.สายันต์ แก่นนาคำ*** *มหาวิทยาลัยเทคโนโลยีสุรนารี*
IVS-03	KKU Smart Mathematics Learning Platform for Secondary Schools ***ผู้ช่วยศาสตราจารย์ ดร.วีระชัย สาระคร*** *มหาวิทยาลัยขอนแก่น*
IVS-04	Decoding Modern Banking: A Mathematician's Guide ***ดร.วุฒิศักดิ์ ตรงศิริวัฒน์*** *รองผู้อำนวยการฝ่าย Data Innovation ธนาคารกรุงไทย*

3. Algebra (ALG)

ALG-01	A New Approach to Ordered Semigroup Theory: Soft Union Ordered Semigroups *Panuwat Luangchaisri and* ***Thawhat Changphas***
ALG-02	Magnifiers in some Subsemigroups of the Full Transformation Semigroups ***Pongsan Prakitsri***
ALG-03	Posets of Ideals in Certain Semigroups of Partial Transformations with Invariant Sets ***Jitsupa Srisawat*** *and Yanisa Chaiya*
ALG-04	Some Algebraic Properties of Translations on n-Ary Semihypergroups ***Anak Nongmanee*** *and Sorasak Leeratanavalee*
ALG-05	Transformation Semigroups Which Are Disjoint Union of General Linear Groups ***Utsithon Chaichompoo*** *and Kritsada Sangkhanan*
ALG-06	Soft Semigroups in Terms of Rough Approximations ***Rukchart Prasertpong****, Nares Sawatraksa, and Sasisophit Buada*
ALG-07	The Pre-period of a Finite Cyclic Group ***Pongsaphat Prachumdang*** *and Udom Chotwattakawanit*
ALG-08	The Isomorphism Theorems for LU13-algebras *Jidapa Wongthipparat and* ***Lee Sassanapitax***
ALG-09	Farey Graphs and Continued Fractions over Certain Finite Fields ***Arlisa Janjing****, Teeraphong Phongpattanacharoen, and Tuangrat Chaichana*

ALG-10	The Diameter and Girth of Subspace Inclusion Graphs Modulo Prime Powers ***Juthamas Sangwisat*** *and Siripong Sirisuk*
ALG-11	Solutions of Systems of PDEs and Representations of A_2 ***Sarawut Saenkarun***
ALG-12	Upper Bounds for the Length of SEL Egyptian Fraction Expansions for Rational Elements of Certain Discrete-Valued Non-Archimedean Fields *Narakorn Rompurk Kanasri and* ***Mayurachat Janthawee***
ALG-13	Some Shallow Elements of Coxeter Groups of Type B ***Kittitat Iamthong****, Sittinon Jirattikansakul, and Korkeat Korkeatikhun*
ALG-14	Functional Graphs of Non-Monic Linear Polynomials on Finite Field Extensions ***Suphawich Sengpanich*** *and Nithi Rungtanapirom*

4. Analysis, Fixed Point Theory and Applications, Topology and Geometry (ANA)

ANA-01	A Fast Forward-Backward Algorithm Using Linesearch and Inertial Techniques for Convex Bilevel Optimization Problems with Applications in Data Classification of Some Noncommunicable Diseases ***Piti Thongsri*** *and Suthep Suantai*
ANA-02	A Novel Double Inertial Viscosity Algorithm for Convex Bilevel Optimization Problems with Application to Image Restoration Problems ***Kobkoon Janngam****, Rattanakorn Wattanataweekul, and Suthep Suantai*
ANA-03	Convergence and Stability of a New Hybrid Iteration Scheme for a Contraction Operator in Banach Spaces with Applications ***Chonjaroen Chairasiripong****, Damrongsak Yambangwai, Papinwich Paimsang, and Tanakit Thianwan*
ANA-04	Convergence Analysis and Polynomiographic Visualization of Picard-SP Hybrid Iterative Methods ***Kaiwich Baewnoi****, Damrongsak Yambangwai, Papinwich Paimsang, and Tanakit Thianwan*
ANA-05	Approximation Theorems for G-nonexpansive Mappings in Hyperbolic Spaces by Using Two-step Iterations ***Tanakit Thianwan****, Maliha Rashid, Amna Kalsoom, and Sana Jabeen*
ANA-06	Accelerated Common Fixed Point Algorithm for Convex Minimization Problems and Applications ***Jirayut Butwang*** *and Suthep Suantai*
ANA-07	Fixed Point Theory for α-G-Contraction Types on Uniform Spaces with a Graph G ***Sittichoke Songsa-ard***
ANA-08	Endpoint Theorems of Diametrically Regular Mappings in Uniformly Convex Hyperbolic Spaces ***Thanomsak Laokul***
ANA-09	Some Characterizations of a Closed Geodesic Polygon and a Closed Spherical Curve in a CAT(k) Space ***Areeyuth Sama-Ae****, Aniruth Phon-on, Nifatamah Makaje, Areena Hazanee, and Pakwan Riyapan*
ANA-10	An Explicit Formula for Quasi-Arithmetic Mean Sequences ***Thanatkrit Kaewtem***
ANA-11	อัตราส่วนของผลรวมพื้นที่รูปสี่เหลี่ยมขนาบข้างอันดับ 2 ต่อผลรวมของพื้นที่รูปสี่เหลี่ยมขนาบข้างอันดับ 1 ของรูปสามเหลี่ยมทั่วไป ***เกวลิน เกิดประวัติ*** *อรรณพ แก้วขาว และ สมคิด อินเทพ* **(ยกเลิกการนำเสนอ)**

5. Combinatorics and Graph Theory (CGT)

CGT-01	Proper Magic Sigma Coloring of Special Graphs ***Panuvit Chuaephon*** *and Kittikorn Nakprasit*
CGT-02	The (3, 3)-Colorability of Planar Graphs with Specific Cycles *Pongpat Sittitrai and* ***Wannapol Pimpasalee***
CGT-03	Solving a 4-Colored 5-Cube Puzzle by Graph Theory ***Pichaya Kankonsue****, Sayan Panma, and Piyashat Sripratak*
CGT-04	ปัญหาการพับแถบแสตมป์ n ดวง เมื่อ n = 2, 3, 4, 5, 6 ***ศิริญญา โปร่งจิตร์*** *ประกายแสง โคตรมิตร ทศพร สายเสมา และ วัชราภรณ์ อดทน*
CGT-05	Secret Sharing from Combinatorial Designs ***Nada Somswasdi*** *and Wutichai Chongchitmate*
CGT-06	Ternary LDPC Codes Based on Projective Plane ***Chanya Lawong*** *and Penying Rochanakul*
CGT-07	The Extreme Case of 3-PGDD's with Block Size 4 and 2 Groups ***Apiwat Peereeyaphat****, Dinesh G. Sarvate, and Chariya Uiyyasathian*
CGT-08	Perfect Matchings in Latin Square Graphs after Vertex Deletions ***Thammanoon Puirod***
CGT-09	Solvability Conditions for $\left(n^2-1\right)$-puzzle with 1 or 2 Fixed Cells ***Waitin Sinthu-urai*** *and Piyashat Sripratak*
CGT-10	Girths and Diameters of a Graph, its δ-Complement, and its δ'-Complement ***Supakorn Srisawat*** *and Panupong Vichitkunakorn*
CGT-11	Local Antimagic Chromatic Number of the Cartesian Product of Graphs *Teeradej Kittipassorn and* ***Kiattiyot Phibul***
CGT-12	List Coloring and List Edge Coloring on King's Graphs ***Papon Tantiwanichanon*** *and Kittikorn Nakprasit*

6. Data Science and Computer Science (DCS)

DCS-01	Deep Learning and Quantum Image Processing in Optometry *Monchita Toopsuwan and* ***Umaporn Nuntaplook***
DCS-02	Graph Convolutional Network for Multiple Traveling Salesman Problem ***Chanoknun Phunnasorn****, Wasakorn Laesanklang, and Tipaluck Krityakierne*
DCS-03	Artificial Intelligence for Forecasting Rice Yields in Thailand ***Thoedsak Saengthong****, Thanathat Khottiam, Chakhrit Utamapokai*, and *Wanyok Atisattapong*
DCS-04	Detection of Parvovirus Infection in Shrimps with VGG16 ***Tharyar Aung****, Pallop Huabsomboon, Kittisak Chayantrakom, Somkid Amornsamankul, and Rapeepun Vanichviriyakit*

DCS-05	การเปรียบเทียบประสิทธิภาพของแบบจำลองพยากรณ์จำนวนผู้เสียชีวิตจากการเกิดอุบัติเหตุจราจรบนโครงข่ายถนนของกระทรวงคมนาคม ***สุภาพร ครองยุทธ*** *และ ปรียานุช เชื้อสุข*

7. Differential Equations and Numerical Mathematics (DNM)

DNM-01	A Non-dimensional Mathematical Model for Predicting Coastlines with a Double-Groin Structure Using the Forward Time-Centered Space Finite Difference Scheme ***Surasak Manilam*** *and Nopparat Pochai*
DNM-02	วิธีการสปริทเบรกแมนสำหรับกำจัดสัญญาณรบกวนแบบการคูณออกจากภาพดิจิทัล ***โสภิดา สุขญาณกิจ*** *และ* ***ศิริวรรณ จันทร์แก่น***
DNM-03	อัลกอริทึมผสมใหม่สำหรับการหาผลเฉลยของสมการไม่เชิงเส้นโดยใช้วิธีของนิวตันและวิธีแก้ตำแหน่งผิด ***ลลิตภัทร สาโรจน์*** *และ อภิชาติ เนียมวงษ์*
DNM-04	Applying the Residual Power Series Method to a Time Fractional Black Scholes European Option Pricing with Two Assets ***Pitsinee Winyarat*** *and Panumart Sawangtong*
DNM-05	An Approximate Analytical Solution of the Time-Fractional Navier-Stokes Equations by the Generalized Shehu Residual Power Series Method ***P. Dunnimit****, W. Sawangtong, and P. Sawangtong*

8. Mathematical Modeling and Mathematical Finance (MMF)

MMF-01	Estimating the Value at Risk of Buy-and-Sell Strategy Using the RSI Indicator on the EUR/USD Exchange Market ***Rattaporn Supama*** *and Watcharin Klongdee*
MMF-02	Mechanistic Modeling of Financial Bubble Driven by Herding Behavior and Safe-Haven Asset ***Sorathan Juanjenkit*** *and Klot Patanarapeelert*
MMF-03	Mathematical Model for the Dynamic of COVID-19 Spread and Impacts of Vaccination, Quarantine, and Hospitalization among the 5th Wave of COVID-19 in Thailand ***Jiraporn Lamwong*** *and Puntani Pongsumpun*
MMF-04	Modified NEH Algorithms for Flowshop Scheduling Problem ***Rungrot Pholyiam****, Pannarat Guayjarernpanishk, and Tawun Remsungnen*
MMF-05	ตัวแบบเชิงคณิตศาสตร์ SI_aI_sR การแพร่ระบาดของโรคโควิด-19 ที่มีผลจากระยะเวลาในการเข้ารับการรักษา ***อภิชญา เกลี้ยงสง*** *กันยากร อ่อนรักษ์ กรกนก ตันติษัณสกุล เกตุกนก หนูดี อัญชุลี ณ ตะกั่วทุ่ง และ ศุภชัย ดำคำ*
MMF-06	A Mathematical Simulation of Airborne Infection Risk Evaluation for Bus Passengers ***Jenjira Sooknum*** *and Nopparat Pochai*
MMF-07	2-D Magnetotelluric Modeling Using Back-Propagation Multilayer Perceptron Approach: Preliminary Results ***Phongphan Mukwachi****, Samak Boonpan, and Weerachai Sarakorn*
MMF-08	Encapsulation of Endofullerene Fe@C_{20} into Single-Walled Carbon Nanotube ***Tana Sunpatanon*** *and Prangsai Tiangtrong*

9. Mathematics Education (MED)

MED-01	การพัฒนาทักษะการแก้ปัญหาทางคณิตศาสตร์และการทำงานเป็นทีมของนักเรียนระดับประกาศนียบัตรวิชาชีพชั้นปีที่ 1 เรื่อง พื้นที่ผิวและปริมาตร โดยใช้การจัดการเรียนรู้แบบปัญหาเป็นฐาน ***ธวัชชัย อินทโฉม*** *และ ธีระพล สลีวงศ์*
MED-02	บทเรียนออนไลน์ เรื่อง สถิติ บน Platform DBAC Style ส่งเสริมทักษะการสื่อสารทางคณิตศาสตร์ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 2 ***พัชรินทร์ เศรษฐีชัยชนะ***
MED-03	การใช้กิจกรรมการเรียนรู้ร่วมมือเทคนิค TGT ร่วมกับสื่อประสมเพื่อพัฒนาทักษะการเรียนรู้และผลสัมฤทธิ์ทางการเรียนเรื่องวงกลม ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ***ยุทธศาสตร์ กองพวง*** *สมฤทัย เย็นใจ และ กุณฑลีรัฐ พิมพิลา*
MED-04	ผลของการใช้ชุดการสอนเกมมิฟิเคชันที่มีต่อผลสัมฤทธิ์ทางการเรียนเรื่องตัวแปรสุ่มและการแจกแจงความน่าจะเป็นของนักเรียนชั้นมัธยมศึกษาปีที่ 4 ***สิทธิโชค โสมอ่ำ***
MED-05	การจัดการเรียนการสอนแบบ Active Learning ในรายวิชาสถิติสำหรับนักวิทยาศาสตร์ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 4 โรงเรียนมหิดลวิทยานุสรณ์ ***เดี่ยว ใจบุญ***

10. Number Theory (NUT)

NUT-01	Equations Related to the Sum and Product of the Fibonacci Numbers *Aram Tangboonduangjit and* ***Shayathorn Wanasawat***
NUT-02	Relation Between the Digit Sum of Numbers: From 1 to $10^n - 1$ and 10^{n-1} to $10^n - 1$ ***Perawit Boonsomchua***
NUT-03	Divisibility Algorithm of Even Number ***Itsara Saenjaroen*** *and Apisit Pakapongpun*
NUT-04	More on the Quadratic Exponential Diophantine Equation $\left(p^k - 1\right)^x + \left(p^k\right)^y = z^2$ *Phornpassorn Boonchu, Janyarak Tongsomporn, and* ***Saeree Wananiyakul***
NUT-05	สมการไดโอแฟนไทน์ $n^x + p^y = z^2$ เมื่อ p เป็นจำนวนเฉพาะ และ $n \equiv 2(mod\ 3p)$ ***อนุสรา ประสิทธิ์นอก*** *และ วีรยุทธ นิลสระคู*
NUT-06	All the Positive Solutions of $p^x - p^y = z^p$ in the Fibonacci and Lucas Numbers when $p = 2$ and $p = 3$ ***Phitthayathon Phetnun***
NUT-07	Integral Representations of the Pell and Pell-Lucas Numbers ***Achariya Nilsrakoo***
NUT-08	Some Properties of k–Narayana Quaternions ***Chansouk Sikhammountri*** *and Narawadee Phudolsitthiphat*
NUT-09	Some Quadratic and Quartic Diophantine Equations with Solutions Involving Fibonacci and Lucas Numbers *Shayathorn Wanasawat,* ***Panida Krongkaew, Orrawan Prathumwan,*** *and* ***Onanong Wimolrat***

NUT-10	Sums of Iterated Partial Sums of the k-Fibonnacci Sequence ***Supamit Pimsri**, Somthawin Khunkhet, and Boonyen Thongkam*
NUT-11	สมบัติบางประการสำหรับลำดับ k-โอเรสเมในรูปแบบเชิงซ้อน ***ชนนิกานต์ คนเพียร** และ บุญยงค์ ศรีพลแผ้ว*

11. Other Related Topics in Mathematics (ORT)

ORT-01	A Generalization of Decomposition Theorem in D-minimal Expansions of the Real Field ***Thanathip Phokhaw** and Athipat Thamrongthanyalak*
ORT-02	System of Stochastic Grey Differential Equations with Singular Spectrum Analysis for Precious Metal Prices Forecasting ***Rammarat Panadsako** and Raywat Tanadkithirun*
ORT-03	อิทธิพลของปัจจัยทางอุตุนิยมวิทยาที่ส่งผลต่อผลผลิตทุเรียนรายปีในจังหวัดสุราษฎร์ธานี ***อินทฤทธิ์ หอมหวล** อรวรรณ สืบเสน และ ปุริณชญาน์ วิสุทธิ์สิริ*

12. Probability Theory and Statistics (PTS)

PTS-01	Local Limit Theorems without Assuming Finite Third Moment ***Punyapat Kammoo**, Kritsana Neammanee, and Kittipong Laipaporn*
PTS-02	Some Properties of Two-Dimensional Trinomial Random Walks Conditioned on End Points ***Yuparat Hommai**, Monchai Kooakachai, and Wasamon Jantai*
PTS-03	Non-uniform Bound on Translated Poisson Approximation for Poisson Binomial Random Variables via Exchangeable Pair Coupling ***Kamonrat Kamjornkittikoon** and Suporn Jongpreechaharn*
PTS-04	Stochastic Models for Breaking Large Bills and Coins *Nakharin Kabbun, Wasamon Jantai, Duong Than, and **Monchai Kooakachai***
PTS-05	การแจกแจงความน่าจะเป็นของความเร็วลมในพื้นที่ที่มีศักยภาพในการตั้งฟาร์มลม: ความเร็วลม ***วนิดา พงษ์ศักดิ์ชาติ** และ พรหมพร ธรรมสาร*
PTS-06	การศึกษาความแกร่งของสถิติทดสอบความแตกต่างของค่าเฉลี่ยประชากรสองกลุ่มอิสระกัน เมื่อข้อมูลมีการแจกแจงปรกติแบบผสมและการแจกแจงแกมมาแบบผสม ***ภัทราภรณ์ กิจผลเจริญ** สุวิมล ชูเปรม และ บำรุงศักดิ์ เผื่อนอารีย์*
PTS-07	Modelling Volleyball Match Outcomes by Using Modified Estimators for the Binomial Parameter ***Jeeraphat Monnoi**, Sutimon Jamrat, and Monchai Kooakachai*
PTS-08	Hidden Population Size Estimator of Poisson Lognormal Distribution for Capture-Recapture Data ***Orasa Nunkaw** and Jutamas Boonradsamee*
PTS -09	ความรู้ความเข้าใจและพฤติกรรมการป้องกันโรคโควิด-19 หลังการระบาดใหญ่ของประชาชนในจังหวัดสุราษฎร์ธานี *อัญชุลี ณ ตะกั่วทุ่ง **ศุภชัย ดำคำ** เกตุกนก หนูดี และ กันยากร อ่อนรักษ์*

AMM 28th ANNUAL MEETING IN MATHEMATICS 2024

1. KEYNOTE SPEAKERS

Professor
Malgorzata Peszynska
AAAS Honorary Fellow,
SIAM GS Career Award,
Oregon State University

Alain Jean Alherbe
Former Microsoft Network Consultant,
Ubon Ratchathani University

Associate Professor
Thiradet Jiarasuksakun
Director of the Institute for the Promotion of Teaching Science and Technology (IPST)

The 28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Underground Computational Mathematics: Models and Analyses of an Evolving Subsurface of Planet Earth

Malgorzata Peszynska[1,†]

[1]Joel Davis Faculty Scholar and Professor (Dr. hab.)
Department of Mathematics, Oregon State University, USA
Corvallis, OR 97331 - 4605

Abstract

In the talk we discuss mathematical models of complex phenomena in the subsurface of the Earth such as flow, transport, and heat conduction, as well as mechanical deformation. The models are coupled systems of nonlinear partial differential equations which typically have solutions of low regularity; they also require a lot of data, frequently given at disparate multiple scales. To use the models for prediction, we run simulations based on our computational algorithms constructed based on rigorous analyses. However, the simulations are only useful if the data for the models are also reasonably accurate. We show how one can construct such data from first principles starting from xray micro-CT tomography at the millimeter scale up to the Darcy scale of meters and further to the kilometer scale of the Arctic landscape. We illustrate with simulation examples and present current work including the challenges going forward.

[†]Keynote Speaker.
Email: Malgo.Peszynska@oregonstate.edu

The 28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Safeguarding Data Privacy: Exploring Full Homomorphic Encryption

Alain Jean Alherbe[1,†]

[1]Department of Mathematics Statistics and Computer, Faculty of Science
Ubon Ratchathani University, Ubon Ratchathani 34190, Thailand

Abstract

Encryption is the process of securing the confidentiality of stored or transmitted data. It involves encoding the information in such a way that only authorized parties can access it. There are several cryptography architectures designed to ensure secure data transmission and storage. For example, Advanced Encryption Standard (AES) and Secure Hash Algorithm (SHA). When data is transmitted over the internet with those architectures, there is a risk of interception by unauthorized parties and sensitive information can be compromised, leading to security and privacy breaches. Full Homomorphic Encryption (FHE) is an innovative encryption technique that enables computations to be performed on encrypted data without the need for decryption. This means that sensitive information remains private while computations are carried out on the encrypted data, ensuring that the output is also encrypted. TenSEAL is a software library developed by Google. It is specifically designed for building homomorphic applications requiring secure computations on sensitive data. This library enables the implementation of secure computations while maintaining the confidentiality of the underlying data. We provide an overview of FHE, examine the benefits and limitations of using TenSEAL, and demonstrate the procedure of using the library to perform basic computations on encrypted data.

[†]Keynote Speaker.
Email: alain.j@ubu.ac.th

การพัฒนาสมรรถนะด้านคณิตศาสตร์ของ PISA ให้กับครูและนักเรียนในยุคดิจิทัล

รองศาสตราจารย์ ดร.ธีระเดช เจียรสุขสกุล[1,†] และนางสุชาดา ปัทมวิภาต[1]

[1]สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสวท.) 475 อาคารสิริภิญโญ ชั้น 9 เขตราชเทวี กรุงเทพมหานคร 10400 (สำนักงานชั่วคราว)

บทคัดย่อ

ในศตวรรษที่ 21 เทคโนโลยีเข้ามามีบทบาทมากขึ้นในชีวิตประจำวัน ข้อมูลที่หลากหลายและมีความซับซ้อนส่วนใหญ่จึงอยู่ในรูปดิจิทัล ซึ่งข้อมูลเหล่านี้สามารถนำมาใช้ในการตัดสินใจทั้งในเรื่องส่วนตัว ไปจนถึงเรื่องที่มีผลต่อสังคมส่วนรวมได้ สิ่งเหล่านี้ทำให้การใช้การดำเนินการทางคณิตศาสตร์เพียงอย่างเดียวนั้นไม่เพียงพอ แต่จำเป็นต้องมีการคิดอย่างเป็นเหตุเป็นผลและสามารถอธิบายเหตุผลได้ ด้วยเหตุนี้ ใน PISA 2022 ซึ่งเป็นรอบการประเมินล่าสุดที่เน้นด้านคณิตศาสตร์ จึงได้ปรับกรอบการประเมินคณิตศาสตร์ให้สอดคล้องกับการเปลี่ยนแปลงดังกล่าว โดยได้เพิ่มการให้เหตุผลทางคณิตศาสตร์เข้ามา เป็นส่วนหนึ่งของการประเมินร่วมกับกระบวนการแก้ปัญหาทางคณิตศาสตร์ ดังนั้น นักเรียนจึงควรได้รับการส่งเสริมให้มีการแสดงเหตุผลร่วมกับการใช้หลักการพื้นฐานทางคณิตศาสตร์ รวมถึงการส่งเสริมให้แก้ปัญหาทางคณิตศาสตร์ผ่านกิจกรรมและแบบฝึกที่ส่งเสริมและกระตุ้นให้ฝึกคิดและฝึกแก้ปัญหาอย่างเป็นระบบและให้เหตุผลทางคณิตศาสตร์อย่างสมเหตุสมผลตามหลักการ เพื่อนำไปสู่ความฉลาดรู้ด้านคณิตศาสตร์สำหรับการใช้ชีวิตในโลกศตวรรษที่ 21 ต่อไป

คำสำคัญ: PISA, ความฉลาดรู้ด้านคณิตศาสตร์

---

†Keynote Speaker

อีเมล: thiradet@ipst.ac.th (ธีระเดช เจียรสุขสกุล), sthai@ipst.ac.th (สุชาดา ปัทมวิภาต).

2. INVITED SPEAKERS

Associate Professor
Chatchawan Panraksa
Mahidol University
International College

Associate Professor
Sayan Kaennakham
Suranaree University of Technology

Assistant Professor
Weerachai Sarakorn
Khon Kaen University

Wuttisak Trongsiriwat
Vice President, Data Innovation,
Krungthai Bank

The 28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Arithmetic Dynamics: Bridging Order and Chaos

Chatchawan Panraksa[1,†]

[1]Mahidol University International College, Nakhon Pathom, Thailand 73170

Abstract

Arithmetic Dynamics stands at the crossroads of number theory and dynamical systems, exploring how numerical patterns evolve over time. This talk introduces its core principles—focusing on the iteration of functions over fields, the significance of periodic and preperiodic points, and the interplay between arithmetic properties and dynamical behavior. We will then highlight current research frontiers, including advances in the distribution of periodic points, applications of height functions, and emerging conjectures that promise to redefine our understanding of the field. This presentation aims to provide a clear and thorough overview of Arithmetic Dynamics, illustrating its role in addressing complex mathematical problems and highlighting opportunities for future research.

[†]Invited Speaker.
Email: chatchawan.pan@mahidol.edu

The 28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Unleashing the Potential of Applied Mathematics in AI and Machine Learning for Modern Industry

Sayan Kaennakham[1,2,†]

[1]School of Mathematics and Geoinformatics, Institute of Science
[2]The Multidisciplinary Innovation Research Centre for Digital Transformation towards Smart Healthcare and Modern Industry (MIDTHaI)

Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand

Abstract

This talk explores the indispensable role of applied mathematics in driving innovations in artificial intelligence (AI) and machine learning (ML). Aimed at applied mathematics undergraduates, we journey from the core mathematical theories underpinning AI/ML to their practical applications in various industries. By interweaving personal experiences with insights into foundational concepts and emerging trends, we highlight the transformative potential of applied mathematics. Attendees will learn about the mathematical backbone of AI technologies, the transition from theoretical models to practical solutions in modern industry, and the exciting research opportunities that await in fields. Through this session, we aim to inspire students to apply their mathematical skills towards pioneering solutions in AI and ML, paving the way for a future where their contributions lead to significant technological advancements.

[†]Invited Speaker.
Email: sayan_kk@g.sut.ac.th

The 28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

KKU Smart Mathematics Learning Platform for Secondary Schools

Weerachai Sarakorn[1,†], Thotsaphon Thongjunthug[1], Warisa Nakpim[1], Somnuek Worawiset[1], and Watcharin Klongdee[1]

[1]Department of Mathematics, Faculty of Science
Khon Kaen University, Khon Kaen 40002, Thailand

Abstract

This study focuses on a digital platform to enhance secondary school students' mathematical interactive learning experience in grades 7-9 (M.1-3). The platform comprises six courses aligned with Thailand's core learning standards and the Programme for International Student Assessment (PISA). It adapts previous smart mathematical learning innovations with carefully selected digital tools for each learning activity. Then, the platform trial testing at networked secondary schools in Northeast Thailand and the primary learning outcome data were collected and analyzed. The results demonstrate that the platform has promising outcomes in promoting student engagement and learning in mathematics.

[†]Invited Speaker.
Email: wsarakorn@kku.ac.th

The 28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Decoding Modern Banking: A Mathematician's Guide

Wuttisak Trongsiriwat[1,†]

[1]Vice President–Data Innovation, Krungthai Bank

Abstract

Banking is a cornerstone of modern economies. Its operations are deeply intertwined with mathematical principles. This talk will delve into the fundamentals of banking operations, emphasizing the critical role of mathematics. We will examine how the rising trend of artificial intelligence presents both opportunities and challenges for the mathematically inclined within the banking sector. In addition, this talk will highlight the enduring importance of a strong mathematical foundation for those seeking to navigate the evolving landscape of banking.

[†]Invited Speaker.
Email: wuttisak.tr@gmail.com

CONTRIBUTED PAPERS
บทความฉบับเต็ม

3. Algebra

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Soft Semigroups in Terms of Rough Approximations

Rukchart Prasertpong[1,†,‡], Nares Sawatraksa[1], and Sasisophit Buada[1]

[1]Division of Mathematics and Statistics, Faculty of Science and Technology,
Nakhon Sawan Rajabhat University, Nakhon Sawan 60000, Thailand

Abstract

In this work, we introduce the lower and upper rough approximations for uni-soft (resp., int-soft) semigroups, uni-soft (resp., int-soft) left ideals, uni-soft (resp., int-soft) right ideals, and uni-soft (resp., int-soft) quasi-ideals via congruence relations on semigroups. Then, we verify the relationship between these concepts and the classical concept of uni-soft (resp., int-soft) ideal theory in semigroups.

Keywords: rough set, soft set, uni-soft ideal, int-soft ideal, semigroup.

2020 MSC: 08A72, 03E20, 06F99.

1 Introduction and Earlier Works

Another issue discussed in connection with the concept of a set or a notion is vagueness. Mathematics requires that all mathematical notions must be exact. The concept of a set is not only fundamental for the whole of mathematics but it also plays an important role in natural language. We often speak about sets (collections) of various objects of interest such as collections of food, tourism locations, and people. Almost all concepts we are using in natural language are vague. In classical set theory, a set is uniquely determined by its elements. In other words, every element must be uniquely classified as belonging to the set or not. It follows that the notion of a set is a crisp (precise) one. Then, common sense reasoning based on natural language must be based on vague concepts and not on classical logic. Observe that beauty is not a precise but a vague concept, because we cannot classify all interesting images uniquely into two classes: beautiful and not beautiful. That is the doubtful area that exists for some interesting images based on beauty. With this point of view, rough set theory can be seen as a new mathematical approach to vagueness. In the proposed approach, assume that any vague concept is replaced by a pair of precise concepts called the lower and the upper approximation of the vague concept. The lower approximation consists of all objects which surely belong to the concept and the upper approximation contains all objects which possibly belong to the concept. At this point, approximations are two basic operations in rough set theory classified by the basic building

†Speaker. ‡Corresponding author.
Email: rukchart.p@nsru.ac.th (R. Prasertpong), nares.sa@nsru.ac.th (N. Sawatraksa), sasisophit.b@nsru.ac.th (S. Buada).

blocks (equivalence classes) so-called the crisp lower and the crisp upper approximation. The difference between the lower and the upper approximation constitutes the boundary region of the vague concept. That is the boundary region of a set consists of all elements that cannot be classified uniquely to the set or its complement. If the boundary region of a set is empty, then it means that the set is definable (or exact). In the opposite case, the set is rough (or inexact). Observe that vagueness is usually associated with the boundary region approach. Rough set theory was proposed by Pawlak [1] in 1982. The rough set approach seems to be of fundamental importance in artificial intelligence. Especially, it is a powerful tool in research areas such as machine learning and decision analysis. This literature is contained in the rudiments of rough sets [2], and the review article in a survey on rough set theory and its applications [3].

In 1999, Molodtsov [4] initiated the notion of soft set theory as a tuple that is associated with a set of parameters and a function from a parameter set to the collection of subsets of a universal set. At this point, a parameter is attributes, characteristics or statements. The major advantage of soft set theory is that it does need not to bother with any additional information about the data such as probability in statistics or possibility value in fuzzy set theory. Especially, the research of the theory for combining the soft set with other mathematical theories has been developed by many authors. This literature is contained in the review on soft set-based parameter reduction and decision-making [5].

Throughout this paper, S and U are non-empty sets. In addition, we denote the collection of subsets of U by $\mathcal{C}(U)$.

Definition 1.1. [4] (F, S) is called a soft set over U with respect to S if F is a function from S to $\mathcal{C}(U)$.

Throughout this paper, We generally denote by $\mathcal{C}(U \sim S)$ a collection of soft sets over U with respect to S.

Definition 1.2. [4] If $(F, S) \in \mathcal{C}(U \sim S)$ is defined by

$$F(a) = U \text{ (resp., } F(a) = \emptyset)$$

for all $a \in S$, then it is called a relative whole (resp., null) soft set over U with respect to S.

Throughout this work, we write $\mathfrak{W}_{U_S} := (W_{U_S}, S)$ (resp., $\mathfrak{N}_{\emptyset_S} := (N_{\emptyset_S}, S)$) instead of a relative whole (resp., null) soft set over U with respect to S.

Definition 1.3. [4] Let $\mathfrak{F} := (F, S), \mathfrak{G} := (G, S) \in \mathcal{C}(U \sim S)$. Recall that $\mathfrak{F}$ is a soft subset of $\mathfrak{G}$ if $F(a) \subseteq G(a)$ for all $a \in S$. At this point, we write $\mathfrak{F} \Subset \mathfrak{G}$. The statement $\mathfrak{F} \Supset \mathfrak{G}$ means $\mathfrak{G} \Subset \mathfrak{F}$.

Definition 1.4. [4] Let $\mathfrak{F} := (F, S), \mathfrak{G} := (G, S) \in \mathcal{C}(U \sim S)$. A soft union of $\mathfrak{F}$ and $\mathfrak{G}$ is defined to be the soft set $\mathfrak{F} \Cup \mathfrak{G} := (F \Cup G, S) \in \mathcal{C}(U \sim S)$ in which $F \Cup G$ is defined by

$$(F \Cup G)(a) = F(a) \cup G(a)$$

for all $a \in S$.

Definition 1.5. [4] Let $\mathfrak{F} := (F, S), \mathfrak{G} := (G, S) \in \mathcal{C}(U \sim S)$. A soft intersection of $\mathfrak{F}$ and $\mathfrak{G}$ is defined to be the soft set $\mathfrak{F} \Cap \mathfrak{G} := (F \Cap G, S) \in \mathcal{C}(U \sim S)$ in which $F \Cap G$ is defined by

$$(F \Cap G)(a) = F(a) \cap G(a)$$

for all $a \in S$.

In 2013, Kim et al. [6] introduced the concept of uni-soft ideals of semigroups based on soft set theory. Then, they proved some properties via the concept of uni-soft products. As mentioned above, we review this concept as follows.

Definition 1.6. [7] Let $*$ be a given binary operation on S. Recall that a semigroup is denoted by an algebraic system $(S, *)$ in which $*$ is associative. For simplicity, we shall write S instead of $(S, *)$. An element a of a semigroup S is said to be a regular element if there exists $b \in S$ such that $a = aba$. A semigroup S is called a regular semigroup if all elements of S are regular.

In the following, if $(S, *)$ is a semigroup, then $a * b$ is denoted by ab for all $a, b \in S$. Given two non-empty subsets A and B of a semigroup S, the product $A * B$ (simply AB) is defined by

$$AB = \{ab : a \in A \text{ and } b \in B\}.$$

Furthermore, for an element a of a semigroup S, we put $\mathcal{R}_a := \{(b, c) \in S \times S : a = bc\}$.

Definition 1.7. [6] Let S be a semigroup, and let $\mathfrak{F} := (F, S), \mathfrak{G} := (G, S) \in \mathcal{C}(U \sim S)$. A uni-soft product of $\mathfrak{F}$ and $\mathfrak{G}$ is defined to be the soft set $\mathfrak{F} \bigtriangledown \mathfrak{G} := (F \bigtriangledown G, S) \in \mathcal{C}(U \sim S)$ in which $F \bigtriangledown G$ is defined by

$$(F \bigtriangledown G)(a) = \begin{cases} \bigcap_{(b,c)\in\mathcal{R}_a} (F(b) \cup G(c)) & \text{if } \mathcal{R}_a \neq \emptyset, \\ U & \text{otherwise} \end{cases}$$

for all $a \in S$.

Remark 1.8. [6] Based on Definition 1.7, $\bigtriangledown$ is associative on $\mathcal{C}(U \sim S)$.

Definition 1.9. [6] Let S be a semigroup, and let $\mathfrak{F} := (F, S) \in \mathcal{C}(U \sim S)$.

(1) $\mathfrak{F}$ is called a uni-soft semigroup if $F(ab) \subseteq F(a) \cup F(b)$ for all $a, b \in S$.

(2) $\mathfrak{F}$ is called a uni-soft left ideal if $F(ab) \subseteq F(b)$ for all $a, b \in S$.

(3) $\mathfrak{F}$ is called a uni-soft right ideal if $F(ab) \subseteq F(a)$ for all $a, b \in S$.

(4) $\mathfrak{F}$ is called a uni-soft quasi-ideal if $(\mathfrak{F} \bigtriangledown \mathfrak{W}_{U_S}) \Cup (\mathfrak{W}_{U_S} \bigtriangledown \mathfrak{F}) \Supseteq \mathfrak{F}$.

Theorem 1.10. *[6] Let S be a semigroup, and let $\mathfrak{F} := (F, S) \in \mathcal{C}(U \sim S)$. Then, the following statements hold.*

(1) *$\mathfrak{F}$ is a uni-soft-soft semigroup if and only if $\mathfrak{F} \bigtriangledown \mathfrak{F} \Supseteq \mathfrak{F}$.*

(2) *$\mathfrak{F}$ is a uni-soft-soft left ideal if and only if $\mathfrak{W}_{U_S} \bigtriangledown \mathfrak{F} \Supseteq \mathfrak{F}$.*

(3) *$\mathfrak{F}$ is a uni-soft-soft right ideal if and only if $\mathfrak{F} \bigtriangledown \mathfrak{W}_{U_S} \Supseteq \mathfrak{F}$.*

(4) *S is a regular semigroup if and only if $\mathfrak{F} \bigtriangledown \mathfrak{W}_{U_S} \bigtriangledown \mathfrak{F} = \mathfrak{F}$ for every uni-soft quasi-ideal $\mathfrak{F}$.*

In 2014, Song et al. [8] proposed the notion of int-soft ideals of semigroups based on soft set theory. Then, they proved some properties via the concept of int-soft products. We review this concept as the following.

Definition 1.11. [8] Let S be a semigroup, and let $\mathfrak{F} := (F, S), \mathfrak{G} := (G, S) \in \mathcal{C}(U \sim S)$. An int-soft product of $\mathfrak{F}$ and $\mathfrak{G}$ is defined to be the soft set $\mathfrak{F} \bigtriangleup \mathfrak{G} := (F \bigtriangleup G, S) \in \mathcal{C}(U \sim S)$ in which $F \bigtriangleup G$ is defined by

$$(F \bigtriangleup G)(a) = \begin{cases} \bigcup_{(b,c)\in\mathcal{R}_a} (F(b) \cap G(c)) & \text{if } \mathcal{R}_a \neq \emptyset, \\ \emptyset & \text{otherwise} \end{cases}$$

for all $a \in S$.

Remark 1.12. [8] Based on Definition 1.11, $\bigtriangleup$ is associative on $\mathcal{C}(U \sim S)$.

Definition 1.13. [8] Let S be a semigroup, and let $\mathfrak{F} := (F, S) \in \mathcal{C}(U \sim S)$.

(1) $\mathfrak{F}$ is called an int-soft semigroup if $F(ab) \supseteq F(a) \cap F(b)$ for all $a, b \in S$.

(2) $\mathfrak{F}$ is called an int-soft left ideal if $F(ab) \supseteq F(b)$ for all $a, b \in S$.

(3) $\mathfrak{F}$ is called an int-soft right ideal if $F(ab) \supseteq F(a)$ for all $a, b \in S$.

(4) $\mathfrak{F}$ is called an int-soft quasi-ideal if $(\mathfrak{F} \triangle \mathfrak{W}_{U_S}) \Cap (\mathfrak{W}_{U_S} \triangle \mathfrak{F}) \Subset \mathfrak{F}$.

Theorem 1.14. *[8] Let S be a semigroup, and let $\mathfrak{F} := (F, S) \in \mathcal{C}(U \sim S)$. Then, the following statements hold.*

(1) *$\mathfrak{F}$ is an int-soft semigroup if and only if $\mathfrak{F} \triangle \mathfrak{F} \Subset \mathfrak{F}$.*

(2) *$\mathfrak{F}$ is an int-soft left ideal if and only if $\mathfrak{W}_{U_S} \triangle \mathfrak{F} \Subset \mathfrak{F}$.*

(3) *$\mathfrak{F}$ is an int-soft right ideal if and only if $\mathfrak{F} \triangle \mathfrak{W}_{U_S} \Subset \mathfrak{F}$.*

(4) *S is a regular semigroup if and only if $\mathfrak{F} \triangle \mathfrak{W}_{U_S} \triangle \mathfrak{F} = \mathfrak{F}$ for every int-soft quasi-ideal $\mathfrak{F}$.*

To support solving the roughness for uni-soft ideal theory and int-soft ideal theory in semigroups, this paper first constructs lower and upper approximation operations to novel soft sets together with a corresponding example. Then, investigate the sufficient conditions for the lower and upper rough approximations of uni-soft (resp., int-soft) semigroups, uni-soft (resp., int-soft) left ideals, uni-soft (resp., int-soft) right ideals, and uni-soft (resp., int-soft) quasi-ideals via congruence relations on semigroups. At this point, the concept of congruence relations on semigroups is presented as follows.

Definition 1.15. [7] Let R be an equivalence relation on a semigroup S. R is called a congruence relation if for all $x, a, b \in S, (a, b) \in R$ implies $(xa, xb) \in R$ and $(ax, bx) \in R$.

Definition 1.16. [7] Let R be a congruence relation on a semigroup S and $a \in S$. The set

$$[a]_R := \{b \in S : (a, b) \in R\}$$

is called a congruence class of R (briefly, R-congruence class) containing a.

Remark 1.17. [9] If R is a congruence relation on a semigroup S, then

$$[a]_R[b]_R \subseteq [ab]_R$$

for all $a, b \in S$.

Definition 1.18. [9] Let R be a congruence relation on a semigroup S. R is a complete congruence relation if

$$[a]_R[b]_R = [ab]_R$$

for all $a, b \in S$.

2 Main Results

In the following, S instead of a semigroup. We now construct lower and upper approximation operations of a soft set in S below.

Definition 2.1. Let R be a congruence relation on S and $\mathfrak{F} := (F, S) \in \mathcal{C}(U \sim S)$. A lower rough approximation of $\mathfrak{F}$ is defined to be the soft set $\llcorner\mathfrak{F}\lrcorner_R := (\llcorner F\lrcorner_R, S) \in \mathcal{C}(U \sim S)$ in which $\llcorner F\lrcorner_R$ is defined by

$$\llcorner F\lrcorner_R(a) = \bigcap_{b\in[a]_R} F(b)$$

for all $a \in S$. An upper rough approximation of $\mathfrak{F}$ is defined to be the soft set $\ulcorner\mathfrak{F}\urcorner_R := (\ulcorner F\urcorner_R, S) \in \mathcal{C}(U \sim S)$ in which $\ulcorner F\urcorner_R$ is defined by

$$\ulcorner F\urcorner_R(a) = \bigcup_{b\in[a]_R} F(b)$$

for all $a \in S$. Based on this point, we say that $\mathfrak{F}$ is a definable soft set if $\ulcorner\mathfrak{F}\urcorner_R = \llcorner\mathfrak{F}\lrcorner_R$; otherwise, $\mathfrak{F}$ is a rough soft set.

We consider the following example.

Example 2.2. Let $S := \{a, b, c, d\}$ be a semigroup with multiplication rules defined by Table 1. Let R be a congruence relation on S such that the R-congruence classes of S are the subsets

Table 1: The Cayley table of a semigroup S

$*$	a	b	c	d
a	a	b	c	d
b	b	b	b	b
c	c	c	c	c
d	d	c	b	a

$\{a\}, \{b, c\}$, and $\{d\}$. Let τ_1, τ_2, τ_3, and τ_4 be subsets of U such that $\tau_1 \subset \tau_2 \subset \tau_3 \subset \tau_4$, and let $\mathfrak{F} := (F, S) \in \mathcal{C}(U \sim S)$ be a soft set over U with respect to S in which F is defined by

$$F(\alpha) = \begin{cases} \tau_1 & \text{if } \alpha = a, \\ \tau_2 & \text{if } \alpha = b, \\ \tau_3 & \text{if } \alpha = c, \\ \tau_4 & \text{if } \alpha = d. \end{cases}$$

Then

$$\llcorner F\lrcorner_R(\alpha) = \begin{cases} \tau_1 & \text{if } \alpha = a, \\ \tau_2 & \text{if } \alpha = b, \\ \tau_2 & \text{if } \alpha = c, \\ \tau_4 & \text{if } \alpha = d \end{cases}$$

and

$$\ulcorner F\urcorner_R(\alpha) = \begin{cases} \tau_1 & \text{if } \alpha = a, \\ \tau_3 & \text{if } \alpha = b, \\ \tau_3 & \text{if } \alpha = c, \\ \tau_4 & \text{if } \alpha = d. \end{cases}$$

Therefore, it is easy to see that $\mathfrak{F}$ is a rough soft set.

Remark 2.3. Based on Example 2.2, observe that $\llcorner F\lrcorner_R(b) = \llcorner F\lrcorner_R(c)$ and $\ulcorner F\urcorner_R(b) = \ulcorner F\urcorner_R(c)$ whenever $(b, c) \in R$. In generality, the statement is also true. Indeed, we assume α and β are

elements of S and $(\alpha, \beta) \in R$. Then $[\alpha]_R = [\beta]_R$. Suppose $u \in \llcorner F \lrcorner_R(\alpha)$. Then $u \in \bigcap_{\gamma \in [\alpha]_R} F(\gamma)$. Hence $u \in \bigcap_{\gamma \in [\beta]_R} F(\gamma)$. Thus $u \in \llcorner F \lrcorner_R(\beta)$. Whence $\llcorner F \lrcorner_R(\alpha) \subseteq \llcorner F \lrcorner_R(\beta)$. Similarly, we can show that $\llcorner F \lrcorner_R(\beta) \subseteq \llcorner F \lrcorner_R(\alpha)$. It follows that $\llcorner F \lrcorner_R(\alpha) = \llcorner F \lrcorner_R(\beta)$. In the same way, it is true that $\ulcorner F \urcorner_R(\alpha) = \ulcorner F \urcorner_R(\beta)$.

Lemma 2.4. *Let R be a congruence relation on S, and let $\mathfrak{F} := (F, S), \mathfrak{G} := (G, S) \in \mathcal{C}(U \sim S)$. Then, the following statements hold.*

(1) $\llcorner \mathfrak{F} \lrcorner_R \Subset \mathfrak{F}$.

(2) $\mathfrak{F} \Subset \ulcorner \mathfrak{F} \urcorner_R$.

(3) *If* $\mathfrak{F} = \mathfrak{N}_{\emptyset_S}$, *then* $\llcorner \mathfrak{F} \lrcorner_R = \mathfrak{N}_{\emptyset_S}$.

(4) *If* $\mathfrak{F} = \mathfrak{N}_{\emptyset_S}$, *then* $\ulcorner \mathfrak{F} \urcorner_R = \mathfrak{N}_{\emptyset_S}$.

(5) *If* $\mathfrak{F} = \mathfrak{W}_{U_S}$, *then* $\llcorner \mathfrak{F} \lrcorner_R = \mathfrak{W}_{U_S}$.

(6) *If* $\mathfrak{F} = \mathfrak{W}_{U_S}$, *then* $\ulcorner \mathfrak{F} \urcorner_R = \mathfrak{W}_{U_S}$.

(7) $\llcorner \mathfrak{F} \Cap \mathfrak{G} \lrcorner_R = \llcorner \mathfrak{F} \lrcorner_R \Cap \llcorner \mathfrak{G} \lrcorner_R$.

(8) $\ulcorner \mathfrak{F} \Cup \mathfrak{G} \urcorner_R = \ulcorner \mathfrak{F} \urcorner_R \Cup \ulcorner \mathfrak{G} \urcorner_R$.

(9) $\llcorner \mathfrak{F} \Cup \mathfrak{G} \lrcorner_R \Supseteq \llcorner \mathfrak{F} \lrcorner_R \Cup \llcorner \mathfrak{G} \lrcorner_R$.

(10) $\ulcorner \mathfrak{F} \Cap \mathfrak{G} \urcorner_R \Subset \ulcorner \mathfrak{F} \urcorner_R \Cap \ulcorner \mathfrak{G} \urcorner_R$.

(11) *If* $\mathfrak{F} \Subset \mathfrak{G}$, *then* $\llcorner \mathfrak{F} \lrcorner_R \Subset \llcorner \mathfrak{G} \lrcorner_R$.

(12) *If* $\mathfrak{F} \Subset \mathfrak{G}$, *then* $\ulcorner \mathfrak{F} \urcorner_R \Subset \ulcorner \mathfrak{G} \urcorner_R$.

Proof. We consider the following proofs.

(1) We have $\llcorner \mathfrak{F} \lrcorner_R \Subset \mathfrak{F}$. In fact,

$$\llcorner F \lrcorner_R(a) = \bigcap_{b \in [a]_R} F(b) \subseteq F(a)$$

for all $a \in S$.

(2) We have $\mathfrak{F} \Subset \ulcorner \mathfrak{F} \urcorner_R$. In fact,

$$F(a) \subseteq \bigcup_{b \in [a]_R} F(b) = \ulcorner F \urcorner_R(a)$$

for all $a \in S$.

(3) Suppose that $\mathfrak{F} = \mathfrak{N}_{\emptyset_S}$. Then $F(a) = \emptyset$ for all $a \in S$. Assume that there exists $a \in S$ such that $\llcorner F \lrcorner_R(a) \neq \emptyset$. Let $u \in \llcorner F \lrcorner_R(a)$. Then $u \in \bigcap_{b \in [a]_R} F(b)$. Thus $u \in F(\alpha)$ for all $\alpha \in [a]_R$. This is a contradiction with $\mathfrak{F} = \mathfrak{N}_{\emptyset_S}$. Then, it is true that

$$\llcorner F \lrcorner_R(\beta) = \emptyset = N_{\emptyset_S}(\beta)$$

for all $\beta \in S$. Whence $\llcorner \mathfrak{F} \lrcorner_R = \mathfrak{N}_{\emptyset_S}$.

(4) Suppose that $\mathfrak{F} = \mathfrak{N}_{\emptyset_S}$. Then $F(a) = \emptyset$ for all $a \in S$. Assume that there exists $a \in S$ such that $\ulcorner F \urcorner_R(a) \neq \emptyset$. Let $u \in \ulcorner F \urcorner_R(a)$. Then $u \in \bigcup_{b \in [a]_R} F(b)$. Thus, there exists $\alpha \in [a]_R$ such that $u \in F(\alpha)$. This is a contradiction with $\mathfrak{F} = \mathfrak{N}_{\emptyset_S}$. Thus, it follows that

$$\ulcorner F \urcorner_R(\beta) = \emptyset = N_{\emptyset_S}(\beta)$$

for all $\beta \in S$. Therefore $\ulcorner \mathfrak{F} \urcorner_R = \mathfrak{N}_{\emptyset_S}$.

(5) Suppose that $\mathfrak{F} = \mathfrak{W}_{U_S}$. Then $F(a) = U$ for all $a \in S$. Assume that there exists $a \in S$ such that $U \setminus \llcorner F \lrcorner_R(a) \neq \emptyset$. Let $u \in U \setminus \llcorner F \lrcorner_R(a)$. Then $u \in U$ and $u \notin \llcorner F \lrcorner_R(a)$. Hence $u \notin \bigcap_{b \in [a]_R} F(b)$. Thus, there exists $\alpha \in [a]_R$ such that $u \notin F(\alpha)$. This is a contradiction with $\mathfrak{F} = \mathfrak{W}_{U_S}$. It follows that

$$W_{U_S}(\beta) \setminus \llcorner F \lrcorner_R(\beta) = U \setminus \llcorner F \lrcorner_R(\beta) = \emptyset$$

for all $\beta \in S$. Consequently, $\llcorner \mathfrak{F} \lrcorner_R = \mathfrak{W}_{U_S}$.

(6) Suppose that $\mathfrak{F} = \mathfrak{W}_{U_S}$. Then $F(a) = U$ for all $a \in S$. Assume that there exists $a \in S$ such that $U \setminus \ulcorner F \urcorner_R(a) \neq \emptyset$. Let $u \in U \setminus \ulcorner F \urcorner_R(a)$. Then $u \in U$ and $u \notin \ulcorner F \urcorner_R(a)$. Thus $u \notin \bigcup_{b \in [a]_R} F(b)$. Hence $u \notin F(\alpha)$ for all $\alpha \in [a]_R$. This is a contradiction with $\mathfrak{F} = \mathfrak{W}_{U_S}$. This means that

$$W_{U_S}(\beta) \setminus \ulcorner F \urcorner_R(\beta) = U \setminus \ulcorner F \urcorner_R(\beta) = \emptyset$$

for all $\beta \in S$, which yields $\ulcorner \mathfrak{F} \urcorner_R = \mathfrak{W}_{U_S}$.

(7) Let $a \in S$ be given. Then

$$\begin{aligned}
\llcorner F \Cap G \lrcorner_R(a) &= \bigcap_{b \in [a]_R} (F \Cap G)(b) \\
&= \bigcap_{b \in [a]_R} (F(b) \cap G(b)) \\
&= \left(\bigcap_{b \in [a]_R} F(b) \right) \cap \left(\bigcap_{b \in [a]_R} G(b) \right) \\
&= \llcorner F \lrcorner_R(a) \cap \llcorner G \lrcorner_R(a) \\
&= (\llcorner F \lrcorner_R \Cap \llcorner G \lrcorner_R)(a).
\end{aligned}$$

This means that $\llcorner \mathfrak{F} \Cap \mathfrak{G} \lrcorner_R = \llcorner \mathfrak{F} \lrcorner_R \Cap \llcorner \mathfrak{G} \lrcorner_R$.

(8) Let $a \in S$ be given. Then

$$\begin{aligned}
\ulcorner F \Cup G \urcorner_R(a) &= \bigcup_{b \in [a]_R} (F \Cup G)(b) \\
&= \bigcup_{b \in [a]_R} (F(b) \cup G(b)) \\
&= \left(\bigcup_{b \in [a]_R} F(b) \right) \cup \left(\bigcup_{b \in [a]_R} G(b) \right) \\
&= \ulcorner F \urcorner_R(a) \cup \ulcorner G \urcorner_R(a) \\
&= (\ulcorner F \urcorner_R \Cup \ulcorner G \urcorner_R)(a).
\end{aligned}$$

This implies that $\ulcorner \mathfrak{F} \Cup \mathfrak{G} \urcorner_R = \ulcorner \mathfrak{F} \urcorner_R \Cup \ulcorner \mathfrak{G} \urcorner_R$.

(9) Let $a \in S$ be given. Then

$$\begin{aligned}
\llcorner F \Cup G \lrcorner_R(a) &= \bigcap_{b \in [a]_R} (F \Cup G)(b) \\
&= \bigcap_{b \in [a]_R} (F(b) \cup G(b)) \\
&\supseteq \bigcap_{b \in [a]_R} F(b) \\
&= \llcorner F \lrcorner_R(a).
\end{aligned}$$

In the same way, we get that $\llcorner F \Cup G \lrcorner_R(a) \supseteq \llcorner G \lrcorner_R(a)$. Observe that

$$\llcorner F \Cup G \lrcorner_R(a) \supseteq \llcorner F \lrcorner_R(a) \cup \llcorner G \lrcorner_R(a) = (\llcorner F \lrcorner_R \Cup \llcorner G \lrcorner_R)(a).$$

This means that $\llcorner \mathfrak{F} \Cup \mathfrak{G} \lrcorner_R \Supseteq \llcorner \mathfrak{F} \lrcorner_R \Cup \llcorner \mathfrak{G} \lrcorner_R$.

(10) Let $a \in S$ be given. Then

$$\begin{aligned} \ulcorner F \Cap G \urcorner_R(a) &= \bigcup_{b \in [a]_R} (F \Cap G)(b) \\ &= \bigcup_{b \in [a]_R} (F(b) \cap G(b)) \\ &\subseteq \bigcup_{b \in [a]_R} F(b) \\ &= \ulcorner F \urcorner_R(a). \end{aligned}$$

Similarly, we can prove that $\ulcorner F \Cap G \urcorner_R(a) \subseteq \ulcorner G \urcorner_R(a)$. Then, it is true that

$$\ulcorner F \Cap G \urcorner_R(a) \subseteq \ulcorner F \urcorner_R(a) \cap \ulcorner G \urcorner_R(a) = (\ulcorner F \urcorner_R \Cap \ulcorner G \urcorner_R)(a).$$

This implies that $\ulcorner \mathfrak{F} \Cap \mathfrak{G} \urcorner_R \Subseteq \ulcorner \mathfrak{F} \urcorner_R \Cap \ulcorner \mathfrak{G} \urcorner_R$.

(11) Suppose that $\mathfrak{F} \Subseteq \mathfrak{G}$. Then $\llcorner \mathfrak{F} \lrcorner_R \Subseteq \llcorner \mathfrak{G} \lrcorner_R$. Indeed,

$$\llcorner F \lrcorner_R(a) = \bigcap_{b \in [a]_R} F(b) \subseteq \bigcap_{b \in [a]_R} G(b) = \llcorner G \lrcorner_R(a)$$

for all $a \in S$.

(12) Suppose that $\mathfrak{F} \Subseteq \mathfrak{G}$. Then $\ulcorner \mathfrak{F} \urcorner_R \Subseteq \ulcorner \mathfrak{G} \urcorner_R$. In fact,

$$\ulcorner F \urcorner_R(a) = \bigcup_{b \in [a]_R} F(b) \subseteq \bigcup_{b \in [a]_R} G(b) = \ulcorner G \urcorner_R(a)$$

for all $a \in S$. □

Lemma 2.5. *Let R be a complete congruence relation on S, and let $\mathfrak{F} := (F, S), \mathfrak{G} := (G, S) \in \mathcal{C}(U \sim S)$. Then, the following statements hold.*

(1) $\llcorner \mathfrak{F} \lrcorner_R \bigtriangledown \llcorner \mathfrak{G} \lrcorner_R \Supseteq \llcorner \mathfrak{F} \bigtriangledown \mathfrak{G} \lrcorner_R$.

(2) $\llcorner \mathfrak{F} \lrcorner_R \bigtriangleup \llcorner \mathfrak{G} \lrcorner_R \Subseteq \llcorner \mathfrak{F} \bigtriangleup \mathfrak{G} \lrcorner_R$.

Proof. We consider the following proofs.

(1) Let $a \in S$ be given. Then, we consider the following two cases.

Case 1. Suppose $\mathcal{R}_a \neq \emptyset$. Then

$$\begin{aligned}
(\llcorner F \lrcorner_R \bigtriangledown \llcorner G \lrcorner_R)(a) &= \bigcap_{a=bc} (\llcorner F \lrcorner_R(b) \cup \llcorner G \lrcorner_R(c)) \\
&= \bigcap_{a=bc} \left(\left(\bigcup_{x \in [b]_R} F(x) \right) \cup \left(\bigcup_{y \in [c]_R} G(y) \right) \right) \\
&= \bigcap_{a=bc} \left(\bigcup_{x \in [b]_R, y \in [c]_R} (F(x) \cup G(y)) \right) \\
&\supseteq \bigcap_{a=bc} \left(\bigcup_{x \in [b]_R, y \in [c]_R} \bigcap_{xy=\alpha\beta} (F(\alpha) \cup G(\beta)) \right), \text{where } \alpha, \beta \in S \\
&= \bigcap_{a=bc} \left(\bigcup_{x \in [b]_R, y \in [c]_R} (F \bigtriangledown G)(xy) \right) \\
&= \bigcap_{a=bc} \left(\bigcup_{xy \in [bc]_R} (F \bigtriangledown G)(xy) \right) \\
&= \bigcap_{a=bc} (\llcorner F \bigtriangledown G \lrcorner_R)(bc) \\
&= (\llcorner F \bigtriangledown G \lrcorner_R)(a).
\end{aligned}$$

Case 2. Suppose $\mathcal{R}_a = \emptyset$. Then

$$(\llcorner F \lrcorner_R \bigtriangledown \llcorner G \lrcorner_R)(a) = U \supseteq (\llcorner F \bigtriangledown G \lrcorner_R)(a).$$

Thus $\llcorner \mathfrak{F} \lrcorner_R \bigtriangledown \llcorner \mathfrak{G} \lrcorner_R \ni \llcorner \mathfrak{F} \bigtriangledown \mathfrak{G} \lrcorner_R$.

(2) Let $a \in S$ be given. Then, we consider the following two cases.

Case 1. Suppose $\mathcal{R}_a \neq \emptyset$. Then

$$\begin{aligned}
(\llcorner F \lrcorner_R \bigtriangleup \llcorner G \lrcorner_R)(a) &= \bigcup_{a=bc} (\llcorner F \lrcorner_R(b) \cap \llcorner G \lrcorner_R(c)) \\
&= \bigcup_{a=bc} \left(\left(\bigcap_{x \in [b]_R} F(x) \right) \cap \left(\bigcap_{y \in [c]_R} G(y) \right) \right) \\
&= \bigcup_{a=bc} \left(\bigcap_{x \in [b]_R, y \in [c]_R} (F(x) \cap G(y)) \right) \\
&\subseteq \bigcup_{a=bc} \left(\bigcap_{x \in [b]_R, y \in [c]_R} \bigcup_{xy=\alpha\beta} (F(\alpha) \cap G(\beta)) \right), \text{where } \alpha, \beta \in S \\
&= \bigcup_{a=bc} \left(\bigcap_{x \in [b]_R, y \in [c]_R} (F \bigtriangleup G)(xy) \right) \\
&= \bigcup_{a=bc} \left(\bigcap_{xy \in [bc]_R} (F \bigtriangleup G)(xy) \right) \\
&= \bigcup_{a=bc} (\llcorner F \bigtriangleup G \lrcorner_R)(bc) \\
&= (\llcorner F \bigtriangleup G \lrcorner_R)(a).
\end{aligned}$$

Case 2. Suppose $\mathcal{R}_a = \emptyset$. Then

$$(\llcorner F \lrcorner_R \triangle \llcorner G \lrcorner_R)(a) = \emptyset \subseteq (\llcorner F \triangle G \lrcorner_R)(a).$$

Therefore $\llcorner \mathfrak{F} \lrcorner_R \triangle \llcorner \mathfrak{G} \lrcorner_R \Subset \llcorner \mathfrak{F} \triangle \mathfrak{G} \lrcorner_R$. □

Lemma 2.6. *Let R be a congruence relation on S, and let $\mathfrak{F} := (F, S), \mathfrak{G} := (G, S) \in \mathcal{C}(U \sim S)$. Then, the following statements hold.*

(1) $\ulcorner \mathfrak{F} \urcorner_R \bigtriangledown \ulcorner \mathfrak{G} \urcorner_R \Supset \ulcorner \mathfrak{F} \bigtriangledown \mathfrak{G} \urcorner_R$.

(2) $\ulcorner \mathfrak{F} \urcorner_R \triangle \ulcorner \mathfrak{G} \urcorner_R \Subset \ulcorner \mathfrak{F} \triangle \mathfrak{G} \urcorner_R$.

Proof. We consider the following proofs.

(1) Let $a \in S$ be given. Then, we consider the following two cases.

Case 1. Suppose $\mathcal{R}_a \neq \emptyset$. Then, by Remark 1.17, it follows that

$$\begin{aligned}
(\ulcorner F \urcorner_R \bigtriangledown \ulcorner G \urcorner_R)(a) &= \bigcap_{a=bc} (\ulcorner F \urcorner_R(b) \cup \ulcorner G \urcorner_R(c)) \\
&= \bigcap_{a=bc} \left(\left(\bigcap_{x \in [b]_R} F(x) \right) \cup \left(\bigcap_{y \in [c]_R} G(y) \right) \right) \\
&= \bigcap_{a=bc} \left(\bigcap_{x \in [b]_R, y \in [c]_R} (F(x) \cup G(y)) \right) \\
&\supseteq \bigcap_{a=bc} \left(\bigcap_{xy \in [bc]_R} (F(x) \cup G(y)) \right) \\
&= \bigcap_{xy \in [a]_R} (F(x) \cup G(y)) \\
&= \bigcap_{z \in [a]_R, z = xy} (F(x) \cup G(y)) \\
&= \bigcap_{z \in [a]_R} \bigcap_{z = xy} (F(x) \cup G(y)) \\
&= \bigcap_{z \in [a]_R} (F \bigtriangledown G)(z) \\
&= (\ulcorner F \bigtriangledown G \urcorner_R)(a).
\end{aligned}$$

Case 2. Suppose $\mathcal{R}_a = \emptyset$. Then

$$(\ulcorner F \urcorner_R \bigtriangledown \ulcorner G \urcorner_R)(a) = U \supseteq (\ulcorner F \bigtriangledown G \urcorner_R)(a).$$

Hence $\ulcorner \mathfrak{F} \urcorner_R \bigtriangledown \ulcorner \mathfrak{G} \urcorner_R \Supset \ulcorner \mathfrak{F} \bigtriangledown \mathfrak{G} \urcorner_R$.

(2) Let $a \in S$ be given. Then, we consider the following two cases.

Case 1. Suppose $\mathcal{R}_a \neq \emptyset$. Then, by Remark 1.17, we obtain

$$\begin{aligned}
(\ulcorner F \urcorner_R \triangle \ulcorner G \urcorner_R)(a) &= \bigcup_{a=bc} (\ulcorner F \urcorner_R(b) \cap \ulcorner G \urcorner_R(c)) \\
&= \bigcup_{a=bc} \left(\left(\bigcup_{x \in [b]_R} F(x) \right) \cap \left(\bigcup_{y \in [c]_R} G(y) \right) \right) \\
&= \bigcup_{a=bc} \left(\bigcup_{x \in [b]_R, y \in [c]_R} (F(x) \cap G(y)) \right) \\
&\subseteq \bigcup_{a=bc} \left(\bigcup_{xy \in [bc]_R} (F(x) \cap G(y)) \right) \\
&= \bigcup_{xy \in [a]_R} (F(x) \cap G(y)) \\
&= \bigcup_{z \in [a]_R, z=xy} (F(x) \cap G(y)) \\
&= \bigcup_{z \in [a]_R} \bigcup_{z=xy} (F(x) \cap G(y)) \\
&= \bigcup_{z \in [a]_R} (F \triangle G)(z) \\
&= (\ulcorner F \triangle G \urcorner_R)(a).
\end{aligned}$$

Case 2. Suppose $\mathcal{R}_a = \emptyset$. Then

$$(\ulcorner F \urcorner_R \triangle \ulcorner G \urcorner_R)(a) = \emptyset \subseteq (\ulcorner F \triangle G \urcorner_R)(a).$$

Therefore $\ulcorner \mathfrak{F} \urcorner_R \triangle \ulcorner \mathfrak{G} \urcorner_R \Subset \ulcorner \mathfrak{F} \triangle \mathfrak{G} \urcorner_R$. □

Theorem 2.7. *Let R be a complete congruence relation on S and $\mathfrak{F} := (F, S) \in \mathcal{C}(U \sim S)$. Then, the following statements hold.*

(1) *If $\mathfrak{F}$ is a uni-soft (resp., an int-soft) semigroup, then $\llcorner \mathfrak{F} \lrcorner_R$ is a uni-soft (resp., an int-soft) semigroup.*

(2) *If $\mathfrak{F}$ is a uni-soft (resp., an int-soft) left ideal, then $\llcorner \mathfrak{F} \lrcorner_R$ is a uni-soft (resp., an int-soft) left ideal.*

(3) *If $\mathfrak{F}$ is a uni-soft (resp., an int-soft) right ideal, then $\llcorner \mathfrak{F} \lrcorner_R$ is a uni-soft (resp., an int-soft) right ideal.*

(4) *If S is a regular semigroup and $\mathfrak{F}$ is a uni-soft quasi-ideal, then $\llcorner \mathfrak{F} \lrcorner_R$ is a uni-soft quasi-ideal.*

(5) *If $\mathfrak{F}$ is an int-soft quasi-ideal, then $\llcorner \mathfrak{F} \lrcorner_R$ is an int-soft quasi-ideal.*

Proof. We consider the following proofs.

(1) Assume that $\mathfrak{F}$ is a uni-soft semigroup. Then $\mathfrak{F} \bigtriangledown \mathfrak{F} \Supseteq \mathfrak{F}$ due to Theorem 1.10 (1). Using Lemma 2.4 (11) and Lemma 2.5 (1), we deduce that

$$\llcorner \mathfrak{F} \lrcorner_R \bigtriangledown \llcorner \mathfrak{F} \lrcorner_R \Supseteq \llcorner \mathfrak{F} \bigtriangledown \mathfrak{F} \lrcorner_R \Supseteq \llcorner \mathfrak{F} \lrcorner_R.$$

Therefore $\llcorner\mathfrak{F}\lrcorner_R$ is a uni-soft semigroup due to Theorem 1.10 (1). We now assume $\mathfrak{F}$ is an int-soft semigroup. Then $\mathfrak{F} \triangle \mathfrak{F} \Subset \mathfrak{F}$ due to Theorem 1.14 (1). Thus, it follows from Lemma 2.4 (11) and Lemma 2.5 (2) that

$$\llcorner\mathfrak{F}\lrcorner_R \triangle \llcorner\mathfrak{F}\lrcorner_R \Subset \llcorner\mathfrak{F} \triangle \mathfrak{F}\lrcorner_R \Subset \llcorner\mathfrak{F}\lrcorner_R.$$

This means that $\llcorner\mathfrak{F}\lrcorner_R$ is an int-soft semigroup due to Theorem 1.14 (1).

(2) Assume that $\mathfrak{F}$ is a uni-soft left ideal. Then, it is shown in Theorem 1.10 (2) that $\mathfrak{W}_{U_S} \triangledown \mathfrak{F} \Supseteq \mathfrak{F}$. Therefore

$$\begin{aligned} \mathfrak{W}_{U_S} \triangledown \llcorner\mathfrak{F}\lrcorner_R &= \llcorner\mathfrak{W}_{U_S}\lrcorner_R \triangledown \llcorner\mathfrak{F}\lrcorner_R \\ &\Supseteq \llcorner\mathfrak{W}_{U_S} \triangledown \mathfrak{F}\lrcorner_R \\ &\Supseteq \llcorner\mathfrak{F}\lrcorner_R \end{aligned}$$

due to the arguments (5) and (11) of Lemma 2.4 and Lemma 2.5 (1). As a consequence, $\llcorner\mathfrak{F}\lrcorner_R$ is a uni-soft left ideal due to Theorem 1.10 (2). We now assume $\mathfrak{F}$ is an int-soft left ideal. Then, by Theorem 1.14 (2), we get $\mathfrak{W}_{U_S} \triangle \mathfrak{F} \Subset \mathfrak{F}$. Using the arguments (5) and (11) of Lemma 2.4 and Lemma 2.5 (2), it follows that

$$\begin{aligned} \mathfrak{W}_{U_S} \triangle \llcorner\mathfrak{F}\lrcorner_R &= \llcorner\mathfrak{W}_{U_S}\lrcorner_R \triangle \llcorner\mathfrak{F}\lrcorner_R \\ &\Subset \llcorner\mathfrak{W}_{U_S} \triangle \mathfrak{F}\lrcorner_R \\ &\Subset \llcorner\mathfrak{F}\lrcorner_R. \end{aligned}$$

Whence $\llcorner\mathfrak{F}\lrcorner_R$ is an int-soft left ideal due to Theorem 1.14 (2).

(3) Suppose that $\mathfrak{F}$ is a uni-soft right ideal. Then, we get that $\mathfrak{F} \triangledown \mathfrak{W}_{U_S} \Supseteq \mathfrak{F}$ due to Theorem 1.10 (3). From the arguments (5) and (11) of Lemma 2.4 and Lemma 2.5 (1), it is true that

$$\begin{aligned} \llcorner\mathfrak{F}\lrcorner_R \triangledown \mathfrak{W}_{U_S} &= \llcorner\mathfrak{F}\lrcorner_R \triangledown \llcorner\mathfrak{W}_{U_S}\lrcorner_R \\ &\Supseteq \llcorner\mathfrak{F} \triangledown \mathfrak{W}_{U_S}\lrcorner_R \\ &\Supseteq \llcorner\mathfrak{F}\lrcorner_R. \end{aligned}$$

Therefore $\llcorner\mathfrak{F}\lrcorner_R$ is a uni-soft right ideal due to Theorem 1.10 (3). We now assume $\mathfrak{F}$ is an int-soft right ideal. Then $\mathfrak{F} \triangle \mathfrak{W}_{U_S} \Subset \mathfrak{F}$ due to Theorem 1.14 (3). Using the arguments (5) and (11) of Lemma 2.4 and Lemma 2.5 (2), it follows that

$$\begin{aligned} \llcorner\mathfrak{F}\lrcorner_R \triangle \mathfrak{W}_{U_S} &= \llcorner\mathfrak{F}\lrcorner_R \triangle \llcorner\mathfrak{W}_{U_S}\lrcorner_R \\ &\Subset \llcorner\mathfrak{F} \triangle \mathfrak{W}_{U_S}\lrcorner_R \\ &\Subset \llcorner\mathfrak{F}\lrcorner_R. \end{aligned}$$

This implies that $\llcorner\mathfrak{F}\lrcorner_R$ is an int-soft right ideal due to Theorem 1.14 (3).

(4) Suppose that S is regular. Then, for every uni-soft right ideal $\mathfrak{G} := (G, S) \in \mathcal{C}(U \sim S)$,

$$\mathfrak{G} \Cup \mathfrak{F} \Supseteq \mathfrak{G} \triangledown \mathfrak{F}.$$

In fact, let $\mathfrak{G} := (G, S) \in \mathcal{C}(U \sim S)$ be a uni-soft right ideal. Let $a \in S$ be given. Then, there exists $b \in S$ such that $a = aba$. Now, observe that $\mathcal{R}_a \neq \emptyset$. Thus, it is true that

$$\begin{aligned} (G \triangledown F)(a) &= \bigcap_{(b,c)\in\mathcal{R}_a} (G(b) \cup F(c)) \\ &\subseteq G(ab) \cup F(a) \\ &\subseteq G(a) \cup F(a) \\ &= (G \Cup F)(a), \end{aligned}$$

as required. Note that $\mathfrak{W}_{U_S}$ is a uni-soft semigroup, which yields $\mathfrak{W}_{U_S} \bigtriangledown \mathfrak{W}_{U_S} \Supseteq \mathfrak{W}_{U_S}$ due to Theorem 1.10 (1). Furthermore, we see that $\mathfrak{F} \bigtriangledown \mathfrak{W}_{U_S}$ is a uni-soft right ideal induced by Theorem 1.10 (3). Indeed,

$$(\mathfrak{F} \bigtriangledown \mathfrak{W}_{U_S}) \bigtriangledown \mathfrak{W}_{U_S} = \mathfrak{F} \bigtriangledown (\mathfrak{W}_{U_S} \bigtriangledown \mathfrak{W}_{U_S}) \Supseteq \mathfrak{F} \bigtriangledown \mathfrak{W}_{U_S}$$

due to Remark 1.8. Thus, it is shown in the argument (3) that $\llcorner \mathfrak{F} \bigtriangledown \mathfrak{W}_{U_S} \lrcorner_R$ is a uni-soft right ideal. We now suppose $\mathfrak{F}$ is a uni-soft quasi-ideal. Then, by Theorem 1.10 (4), the arguments (5) and (11) of Lemma 2.4, and Lemma 2.5 (1), it follows that

$$\begin{aligned}
(\llcorner \mathfrak{F} \lrcorner_R \bigtriangledown \mathfrak{W}_{U_S}) \Cup (\mathfrak{W}_{U_S} \bigtriangledown \llcorner \mathfrak{F} \lrcorner_R) &= (\llcorner \mathfrak{F} \lrcorner_R \bigtriangledown \llcorner \mathfrak{W}_{U_S} \lrcorner_R) \Cup (\llcorner \mathfrak{W}_{U_S} \lrcorner_R \bigtriangledown \llcorner \mathfrak{F} \lrcorner_R) \\
&\Supseteq (\llcorner \mathfrak{F} \bigtriangledown \mathfrak{W}_{U_S} \lrcorner_R) \Cup (\llcorner \mathfrak{W}_{U_S} \bigtriangledown \mathfrak{F} \lrcorner_R) \\
&\Supseteq (\llcorner \mathfrak{F} \bigtriangledown \mathfrak{W}_{U_S} \lrcorner_R) \bigtriangledown (\llcorner \mathfrak{W}_{U_S} \bigtriangledown \mathfrak{F} \lrcorner_R) \\
&\Supseteq \llcorner (\mathfrak{F} \bigtriangledown \mathfrak{W}_{U_S}) \bigtriangledown (\mathfrak{W}_{U_S} \bigtriangledown \mathfrak{F}) \lrcorner_R \\
&= \llcorner \mathfrak{F} \bigtriangledown (\mathfrak{W}_{U_S} \bigtriangledown \mathfrak{W}_{U_S}) \bigtriangledown \mathfrak{F} \lrcorner_R \\
&\Supseteq \llcorner \mathfrak{F} \bigtriangledown \mathfrak{W}_{U_S} \bigtriangledown \mathfrak{F} \lrcorner_R \\
&= \llcorner \mathfrak{F} \lrcorner_R.
\end{aligned}$$

As a consequence, $\llcorner \mathfrak{F} \lrcorner_R$ is a uni-soft quasi-ideal.

(5) Assume that $\mathfrak{F}$ is an int-soft quasi-ideal. Then $(\mathfrak{F} \bigtriangleup \mathfrak{W}_{U_S}) \Cap (\mathfrak{W}_{U_S} \bigtriangleup \mathfrak{F}) \Subseteq \mathfrak{F}$. Using the arguments (5), (7), and (11) of Lemma 2.4 and Lemma 2.5 (2), it follows that

$$\begin{aligned}
(\llcorner \mathfrak{F} \lrcorner_R \bigtriangleup \mathfrak{W}_{U_S}) \Cap (\mathfrak{W}_{U_S} \bigtriangleup \llcorner \mathfrak{F} \lrcorner_R) &= (\llcorner \mathfrak{F} \lrcorner_R \bigtriangleup \llcorner \mathfrak{W}_{U_S} \lrcorner_R) \Cap (\llcorner \mathfrak{W}_{U_S} \lrcorner_R \bigtriangleup \llcorner \mathfrak{F} \lrcorner_R) \\
&\Subseteq (\llcorner \mathfrak{F} \bigtriangleup \mathfrak{W}_{U_S} \lrcorner_R) \Cap (\llcorner \mathfrak{W}_{U_S} \bigtriangleup \mathfrak{F} \lrcorner_R) \\
&= \llcorner (\mathfrak{F} \bigtriangleup \mathfrak{W}_{U_S}) \Cap (\mathfrak{W}_{U_S} \bigtriangleup \mathfrak{F}) \lrcorner_R \\
&\Subseteq \llcorner \mathfrak{F} \lrcorner_R.
\end{aligned}$$

This implies that $\llcorner \mathfrak{F} \lrcorner_R$ is an int-soft quasi-ideal. □

It is not difficult to see that the converse of Theorem 2.7 does not hold in general. We consider the context of uni-soft semigroups in Example 2.8 below.

Example 2.8. Let $S := \{a, b, c, d\}$ be a semigroup with multiplication rules of the binary operation $*$ on S defined by Table 2. Let R be a complete congruence relation on S in which

Table 2: The Cayley table of a semigroup S

$*$	a	b	c	d
a	a	a	a	d
b	a	b	a	d
c	a	a	c	d
d	d	d	d	d

the R-congruence classes of S are the subsets $\{a, b, c\}$ and $\{d\}$. Let τ_1, τ_2, τ_3, and τ_4 be subsets of U such that $\tau_1 \supset \tau_2 \supset \tau_3 \supset \tau_4$, and let $\mathfrak{F} := (F, S) \in \mathcal{C}(U \sim S)$ be a soft set over U with respect to S in which F is defined by

$$F(\alpha) = \begin{cases} \tau_1 & \text{if } \alpha = a, \\ \tau_2 & \text{if } \alpha = b, \\ \tau_3 & \text{if } \alpha = c, \\ \tau_4 & \text{if } \alpha = d. \end{cases}$$

Observe that $F(bc) \not\subseteq F(b) \cup F(c)$. Thus $\mathfrak{F}$ is not a uni-soft semigroup. But, we know $\llcorner\mathfrak{F}\lrcorner_R$ is a uni-soft semigroup. Indeed,

$$\llcorner F\lrcorner_R(\alpha) = \begin{cases} \tau_3 & \text{if } \alpha = a, \\ \tau_3 & \text{if } \alpha = b, \\ \tau_3 & \text{if } \alpha = c, \\ \tau_4 & \text{if } \alpha = d, \end{cases}$$

which yields $\llcorner F\lrcorner_R(\beta\gamma) \subseteq \llcorner F\lrcorner_R(\beta) \cup \llcorner F\lrcorner_R(\gamma)$ for all $\beta, \gamma \in S$.

Theorem 2.9. *Let R be a congruence relation on S and $\mathfrak{F} := (F, S) \in \mathcal{C}(U \sim S)$. Then, the following statements hold.*

(1) *If $\mathfrak{F}$ is a uni-soft (resp., an int-soft) semigroup, then $\ulcorner\mathfrak{F}\urcorner_R$ is a uni-soft (resp., an int-soft) semigroup.*

(2) *If $\mathfrak{F}$ is a uni-soft (resp., an int-soft) left ideal, then $\ulcorner\mathfrak{F}\urcorner_R$ is a uni-soft (resp., an int-soft) left ideal.*

(3) *If $\mathfrak{F}$ is a uni-soft (resp., an int-soft) right ideal, then $\ulcorner\mathfrak{F}\urcorner_R$ is a uni-soft (resp., an int-soft) right ideal.*

(4) *If $\mathfrak{F}$ is a uni-soft quasi-ideal, then $\ulcorner\mathfrak{F}\urcorner_R$ is a uni-soft quasi-ideal.*

(5) *If S is a regular semigroup and $\mathfrak{F}$ is an int-soft quasi-ideal, then $\ulcorner\mathfrak{F}\urcorner_R$ is an int-soft quasi-ideal.*

Proof. We consider the following proofs.

(1) Assume that $\mathfrak{F}$ is a uni-soft semigroup. Then $\mathfrak{F} \bigtriangledown \mathfrak{F} \Supseteq \mathfrak{F}$ due to Theorem 1.10 (1). By Lemma 2.4 (12) and Lemma 2.6 (1), it follows that

$$\ulcorner\mathfrak{F}\urcorner_R \bigtriangledown \ulcorner\mathfrak{F}\urcorner_R \Supseteq \ulcorner\mathfrak{F} \bigtriangledown \mathfrak{F}\urcorner_R \Supseteq \ulcorner\mathfrak{F}\urcorner_R.$$

Therefore $\ulcorner\mathfrak{F}\urcorner_R$ is a uni-soft semigroup due to Theorem 1.10 (1). We now suppose $\mathfrak{F}$ is an int-soft semigroup. Then $\mathfrak{F} \bigtriangleup \mathfrak{F} \Subseteq \mathfrak{F}$ due to Theorem 1.14 (1). It follows from Lemma 2.4 (12) and Lemma 2.6 (2) that

$$\ulcorner\mathfrak{F}\urcorner_R \bigtriangleup \ulcorner\mathfrak{F}\urcorner_R \Subseteq \ulcorner\mathfrak{F} \bigtriangleup \mathfrak{F}\urcorner_R \Subseteq \ulcorner\mathfrak{F}\urcorner_R.$$

Hence $\ulcorner\mathfrak{F}\urcorner_R$ is an int-soft semigroup due to Theorem 1.14 (1).

(2) Suppose $\mathfrak{F}$ is a uni-soft left ideal. Then, it is shown in Theorem 1.10 (2) that $\mathfrak{W}_{U_S} \bigtriangledown \mathfrak{F} \Supseteq \mathfrak{F}$. It follows that

$$\begin{aligned} \mathfrak{W}_{U_S} \bigtriangledown \ulcorner\mathfrak{F}\urcorner_R &= \ulcorner\mathfrak{W}_{U_S}\urcorner_R \bigtriangledown \ulcorner\mathfrak{F}\urcorner_R \\ &\Supseteq \ulcorner\mathfrak{W}_{U_S} \bigtriangledown \mathfrak{F}\urcorner_R \\ &\Supseteq \ulcorner\mathfrak{F}\urcorner_R \end{aligned}$$

due to the arguments (6) and (12) of Lemma 2.4 and Lemma 2.6 (1). Consequently, $\ulcorner\mathfrak{F}\urcorner_R$ is a uni-soft left ideal due to Theorem 1.10 (2). We now assume $\mathfrak{F}$ is an int-soft left ideal. Then, by Theorem 1.14 (2), we have $\mathfrak{W}_{U_S} \bigtriangleup \mathfrak{F} \Subseteq \mathfrak{F}$. Using the arguments (6) and (12) of Lemma 2.4 and Lemma 2.6 (2), we deduce that

$$\begin{aligned} \mathfrak{W}_{U_S} \bigtriangleup \ulcorner\mathfrak{F}\urcorner_R &= \ulcorner\mathfrak{W}_{U_S}\urcorner_R \bigtriangleup \ulcorner\mathfrak{F}\urcorner_R \\ &\Subseteq \ulcorner\mathfrak{W}_{U_S} \bigtriangleup \mathfrak{F}\urcorner_R \\ &\Subseteq \ulcorner\mathfrak{F}\urcorner_R. \end{aligned}$$

Thus $\ulcorner\mathfrak{F}\urcorner_R$ is an int-soft left ideal due to Theorem 1.14 (2).

(3) Assume that $\mathfrak{F}$ is a uni-soft right ideal. Then, we get that $\mathfrak{F} \bigtriangledown \mathfrak{W}_{U_S} \Supseteq \mathfrak{F}$ due to Theorem 1.10 (3). From the arguments (6) and (12) of Lemma 2.4 and Lemma 2.6 (1), it follows that

$$\begin{aligned} \ulcorner \mathfrak{F} \urcorner_R \bigtriangledown \mathfrak{W}_{U_S} &= \ulcorner \mathfrak{F} \urcorner_R \bigtriangledown \ulcorner \mathfrak{W}_{U_S} \urcorner_R \\ &\Supseteq \ulcorner \mathfrak{F} \bigtriangledown \mathfrak{W}_{U_S} \urcorner_R \\ &\Supseteq \ulcorner \mathfrak{F} \urcorner_R. \end{aligned}$$

This implies that $\ulcorner \mathfrak{F} \urcorner_R$ is a uni-soft right ideal due to Theorem 1.10 (3). We now assume $\mathfrak{F}$ is an int-soft right ideal. Then, we have $\mathfrak{F} \bigtriangleup \mathfrak{W}_{U_S} \Subseteq \mathfrak{F}$ due to Theorem 1.14 (3). By the arguments (6) and (12) of Lemma 2.4 and Lemma 2.6 (2), we obtain

$$\begin{aligned} \ulcorner \mathfrak{F} \urcorner_R \bigtriangleup \mathfrak{W}_{U_S} &= \ulcorner \mathfrak{F} \urcorner_R \bigtriangleup \ulcorner \mathfrak{W}_{U_S} \urcorner_R \\ &\Subseteq \ulcorner \mathfrak{F} \bigtriangleup \mathfrak{W}_{U_S} \urcorner_R \\ &\Subseteq \ulcorner \mathfrak{F} \urcorner_R. \end{aligned}$$

Whence $\ulcorner \mathfrak{F} \urcorner_R$ is an int-soft right ideal due to Theorem 1.14 (3).

(4) Assume that $\mathfrak{F}$ is a uni-soft quasi-ideal. Then $(\mathfrak{F} \bigtriangledown \mathfrak{W}_{U_S}) \Cup (\mathfrak{W}_{U_S} \bigtriangledown \mathfrak{F}) \Supseteq \mathfrak{F}$. Using the arguments (6), (8), and (12) of Lemma 2.4 and Lemma 2.6 (1), it follows that

$$\begin{aligned} (\ulcorner \mathfrak{F} \urcorner_R \bigtriangledown \mathfrak{W}_{U_S}) \Cup (\mathfrak{W}_{U_S} \bigtriangledown \ulcorner \mathfrak{F} \urcorner_R) &= (\ulcorner \mathfrak{F} \urcorner_R \bigtriangledown \ulcorner \mathfrak{W}_{U_S} \urcorner_R) \Cup (\ulcorner \mathfrak{W}_{U_S} \urcorner_R \bigtriangledown \ulcorner \mathfrak{F} \urcorner_R) \\ &\Supseteq (\ulcorner \mathfrak{F} \bigtriangledown \mathfrak{W}_{U_S} \urcorner_R) \Cup (\ulcorner \mathfrak{W}_{U_S} \bigtriangledown \mathfrak{F} \urcorner_R) \\ &= \ulcorner (\mathfrak{F} \bigtriangledown \mathfrak{W}_{U_S}) \Cup (\mathfrak{W}_{U_S} \bigtriangledown \mathfrak{F}) \urcorner_R \\ &\Supseteq \ulcorner \mathfrak{F} \urcorner_R. \end{aligned}$$

Consequently, $\ulcorner \mathfrak{F} \urcorner_R$ is a uni-soft quasi-ideal.

(5) Suppose that S is regular. Then, for every int-soft left ideal $\mathfrak{G} := (G, S) \in \mathcal{C}(U \sim S)$,

$$\mathfrak{F} \Cap \mathfrak{G} \Subseteq \mathfrak{F} \bigtriangleup \mathfrak{G}.$$

Indeed, let $\mathfrak{G} := (G, S) \in \mathcal{C}(U \sim S)$ be an int-soft left ideal. Let $a \in S$ be given. Then, there exists $b \in S$ such that $a = aba$. Now, observe that $\mathcal{R}_a \neq \emptyset$. Thus, it is true that

$$\begin{aligned} (F \bigtriangleup G)(a) &= \bigcup_{(b,c) \in \mathcal{R}_a} (F(b) \cap G(c)) \\ &\supseteq F(a) \cap G(ba) \\ &\supseteq F(a) \cap G(a) \\ &= (F \Cap G)(a), \end{aligned}$$

as required. It is not difficult to see that $\mathfrak{W}_{U_S}$ is an int-soft semigroup. Thus, we obtain that $\mathfrak{W}_{U_S} \bigtriangleup \mathfrak{W}_{U_S} \Subseteq \mathfrak{W}_{U_S}$ due to Theorem 1.14 (1). In addition, observe that $\mathfrak{W}_{U_S} \bigtriangleup \mathfrak{F}$ is an int-soft left ideal induced by Theorem 1.14 (2). In fact,

$$\mathfrak{W}_{U_S} \bigtriangleup (\mathfrak{W}_{U_S} \bigtriangleup \mathfrak{F}) = (\mathfrak{W}_{U_S} \bigtriangleup \mathfrak{W}_{U_S}) \bigtriangleup \mathfrak{F} \Subseteq \mathfrak{W}_{U_S} \bigtriangleup \mathfrak{F}$$

due to Remark 1.12. Thus, it is shown in the argument (2) that $\ulcorner \mathfrak{W}_{U_S} \bigtriangleup \mathfrak{F} \urcorner_R$ is an int-soft left ideal. We now assume $\mathfrak{F}$ is an int-soft quasi-ideal. Then, by Theorem 1.14 (4), the

arguments (6) and (12) of Lemma 2.4, and Lemma 2.6 (2), it follows that

$$\begin{aligned}(\ulcorner\mathfrak{F}\urcorner_R \triangle \mathfrak{W}_{U_S}) \Cap (\mathfrak{W}_{U_S} \triangle \ulcorner\mathfrak{F}\urcorner_R) &= (\ulcorner\mathfrak{F}\urcorner_R \triangle \ulcorner\mathfrak{W}_{U_S}\urcorner_R) \Cap (\ulcorner\mathfrak{W}_{U_S}\urcorner_R \triangle \ulcorner\mathfrak{F}\urcorner_R)\\ &\Subset (\ulcorner\mathfrak{F} \triangle \mathfrak{W}_{U_S}\urcorner_R) \Cap (\ulcorner\mathfrak{W}_{U_S} \triangle \mathfrak{F}\urcorner_R)\\ &\Subset (\ulcorner\mathfrak{F} \triangle \mathfrak{W}_{U_S}\urcorner_R) \triangle (\ulcorner\mathfrak{W}_{U_S} \triangle \mathfrak{F}\urcorner_R)\\ &\Subset \ulcorner(\mathfrak{F} \triangle \mathfrak{W}_{U_S}) \triangle (\mathfrak{W}_{U_S} \triangle \mathfrak{F})\urcorner_R\\ &= \ulcorner\mathfrak{F} \triangle (\mathfrak{W}_{U_S} \triangle \mathfrak{W}_{U_S}) \triangle \mathfrak{F}\urcorner_R\\ &\Subset \ulcorner\mathfrak{F} \triangle \mathfrak{W}_{U_S} \triangle \mathfrak{F}\urcorner_R\\ &= \ulcorner\mathfrak{F}\urcorner_R.\end{aligned}$$

This means that $\ulcorner\mathfrak{F}\urcorner_R$ is an int-soft quasi-ideal. □

It is not difficult to see that the converse of Theorem 2.9 does not hold in general. We consider the context of int-soft semigroups in Example 2.10 below.

Example 2.10. Let $S := \{a, b, c, d\}$ be a semigroup with multiplication rules of the binary operation $*$ on S defined by Table 3. Let R be a given congruence relation on S such that the

Table 3: The Cayley table of a semigroup S

$*$	a	b	c	d
a	a	b	b	d
b	b	b	b	d
c	b	b	b	d
d	d	d	d	d

R-congruence classes of S are the subsets $\{a\}, \{b, d\}$, and $\{c\}$. Let τ_1, τ_2, τ_3, and τ_4 be subsets of U such that $\tau_1 \subset \tau_2 \subset \tau_3 \subset \tau_4$, and let $\mathfrak{F} := (F, S) \in \mathcal{C}(U \sim S)$ be a soft set over U with respect to S in which F is defined by

$$F(\alpha) = \begin{cases} \tau_1 & \text{if } \alpha = a,\\ \tau_2 & \text{if } \alpha = b,\\ \tau_3 & \text{if } \alpha = c,\\ \tau_4 & \text{if } \alpha = d.\end{cases}$$

We see that $F(cc) \not\supseteq F(c) \cap F(c)$. Thus $\mathfrak{F}$ is not an int-soft semigroup. But, we know $\ulcorner\mathfrak{F}\urcorner_R$ is an int-soft semigroup. In fact,

$$\ulcorner F\urcorner_R(\alpha) = \begin{cases} \tau_1 & \text{if } \alpha = a,\\ \tau_4 & \text{if } \alpha = b,\\ \tau_3 & \text{if } \alpha = c,\\ \tau_4 & \text{if } \alpha = d,\end{cases}$$

and so $\ulcorner F\urcorner_R(\beta\gamma) \supseteq \ulcorner F\urcorner_R(\beta) \cap \ulcorner F\urcorner_R(\gamma)$ for all $\beta, \gamma \in S$.

3 Summarized Frameworks

In Section 2, we proved that the lower (resp., upper) rough approximation of uni-soft semigroups, uni-soft left ideals, uni-soft right ideals, and uni-soft quasi-ideals is uni-soft semigroups, uni-soft left ideals, uni-soft right ideals, and uni-soft quasi-ideals, respectively. Furthermore, we found

that the lower (resp., upper) rough approximation of int-soft semigroups, int-soft left ideals, int-soft right ideals, and int-soft quasi-ideals is int-soft semigroups, int-soft left ideals, int-soft right ideals, and int-soft quasi-ideals, respectively.

Acknowledgment. We would like to thank the expert reviewers for their qualitative suggestions. We would like to thank supporter organizations: Division of Mathematics and Statistics, Faculty of Science and Technology, Nakhon Sawan Rajabhat University, Thailand.

References

[1] Z. Pawlak, *Rough sets*, Int. J. Comput. Inf. Sci. **11** (1982), 341–356. doi: 10.1007/BF01001956.

[2] Z. Pawlak and A. Skowron, *Rudiments of rough sets*, Inform. Sci. **177** (2007), 3–27. doi: 10.1016/j.ins.2006.06.003.

[3] Q. Zhang, Q. Xie, and G. Wang, *A survey on rough set theory and its applications*, CAAI T. Intell. Techno. **1** (2016), 323–333. doi: 10.1016/j.trit.2016.11.001.

[4] D. Molodtsov, *Soft set theory-first results*, Comput. Math. Appl. **37** (1999), 19–31. doi: 10.1016/S0898-1221(99)00056-5.

[5] S. Danjuma, T. Herawan, M. A. Ismail, H. Chiroma, A. I. Abubakar, and A. M. Zeki, *A review on soft set-based parameter reduction and decision making*, IEEE Access **5** (2017), 4671–4689. doi: 10.1109/ACCESS.2017.2682231.

[6] C. S. Kim, J. G. Kang, and J. S. Kim, *Uni-soft (quasi) ideals of semigroups*, Applied Mathematical Sciences **7** (2013), 2455–2468. doi: 10.12988/ams.2013.13222.

[7] J. M. Howie, *Fundamentals of semigroup theory*, United States: Oxford University Press, New York, 1995.

[8] S. Z. Song, H. S. Kim, and Y. B. Jun, *Ideal theory in semigroups based on intersectional soft sets*, Sci. World J. (2014), 1–7. doi: 10.1155/2014/136424.

[9] N. Kuroki, *Rough ideals in semigroups*, Inform. Sci. **100** (1997), 139–163. doi: 10.1016/S0020-0255(96)00274-5.

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Farey Graphs and Continued Fractions over Certain Finite Fields

Arlisa Janjing[1,†,‡], Teeraphong Phongpattanacharoen[1], and Tuangrat Chaichana[1]

[1]Department of Mathematics and Computer Science, Faculty of Science
Chulalongkorn University, Bangkok 10300, Thailand

Abstract

In this work, we construct Farey graphs in the fields of rational functions over certain finite fields. We explore theirs properties and establish some relationships between these graphs and regular continued fractions.

Keywords: continued fraction, Farey graph, rational function.

2020 MSC: Primary 11A55; Secondary 05C90.

1 Introduction

In 1991, Jones, Singerman, and Wicks [1] studied structures of certain graphs $\mathcal{F}_{u,N}$, where $u, N \in \mathbb{N}$ and $\gcd(u, N) = 1$, defined as follows: the vertex set of the graph $\mathcal{F}_{u,N}$ is given by

$$\chi_N = \left\{ \frac{p}{q} : p, q \in \mathbb{Z}, q > 0, (p, q) = 1 \text{ and } N \mid q \right\} \cup \{\infty\}, \tag{1.1}$$

and there is an edge joining $\frac{p}{q}$ and $\frac{r}{s}$ if and only if $rq - sp = N$ with $p \equiv ur \pmod{N}$ or $rq - sp = -N$ with $p \equiv -ur \pmod{N}$. The graph $\mathcal{F}_{1,1}$ is called the *Farey graph*. A large body of research showed the close connection between these graphs and continued fractions. In 2015, Sarma, Kushwaha, and Krishnan [6] introduced a specific kind of semi-regular continued fractions which is referred to as an $\mathcal{F}_{1,2}$-continued fraction as follows: a finite continued fraction of the form

$$\frac{1}{0+}\frac{2}{b+}\frac{\epsilon_1}{a_1+}\frac{\epsilon_2}{a_2}\cdots\frac{\epsilon_n}{a_n} \qquad (n \geq 0)$$

or an infinite continued fraction of the form

$$\frac{1}{0+}\frac{2}{b+}\frac{\epsilon_1}{a_1+}\frac{\epsilon_2}{a_2+}\cdots\frac{\epsilon_n}{a_n+}\cdots$$

*The first author was financially supported by the Development and Promotion of Science and Technology Talents Project (DPST).

†Speaker. ‡Corresponding author.
Email: arlisa.janjing@gmail.com (A. Janjing), teeraphong.p@chula.ac.th (T. Phongpattanacharoen), tuangrat.c@chula.ac.th (T. Chaichana).

where b is an odd integer, $a_1, a_2, \ldots$ are even positive integers, and $\epsilon_1, \epsilon_2, \cdots \in \{\pm 1\}$, is called an $\mathcal{F}_{1,2}$-*continued fraction*. They established that each finite $\mathcal{F}_{1,2}$-continued fraction corresponds to a path from ∞ to its value. In 2018, similar results for a graph $\mathcal{F}_{1,3}$ were also studied by Kushwaha and Sarma [3]. Recently, in 2022, Kushwaha and Sarma [4] relaxed the conditions of two adjacent vertices in the graph $\mathcal{F}_{u,N}$ and got a new family of graphs $\mathcal{F}_N$ defined as follows: the set of vertices is χ_N (as in Equation (1.1)) and two vertices $\frac{p}{q}$ and $\frac{r}{s}$ are connected by an edge if and only if $rq - sp = \pm N$. Similarly, they constructed $\mathcal{F}_N$-continued fraction and established the parallel results of their earlier works.

Motivated by the idea of Kushwaha and Sarma, we are interested in some relationships between graphs and continued fractions but in fields of rational functions over finite fields instead. For the continued fractions part, we focus on the regular continued fractions that have already been constructed and well-known, see e.g. [2]. In this study, our objective is to construct Farey graphs analogous to the one defined in [1] within the field of rational functions over finite fields. We aim to explore their properties and establish some relationships between these graphs and regular continued fractions.

2 Continued Fractions in Fields of Rational Functions

Let $\mathbb{F}_p$ be a finite fields of p elements (p not necessary a prime), $\mathbb{F}_p(x)$ the field of rational functions over $\mathbb{F}_p$ and $\mathbb{F}_p((x^{-1}))$ the field of formal series over $\mathbb{F}_p$ complete with respect to the degree valuation $|\cdot|$. Recall that for each nonzero element

$$\alpha = c_m x^m + \ldots + c_1 x + c_0 + \frac{c_{-1}}{x} + \frac{c_{-2}}{x^2} + \ldots \in \mathbb{F}_p((x^{-1}))$$

where $m \in \mathbb{Z}, c_i \in \mathbb{F}_p$ $(i \leq m)$ with $c_m \neq 0$, the degree valuation is defined by $|\alpha| = p^m$ and $|0| = 0$. The *integral part* of α, denoted by $[\alpha]$ is defined to be $[\alpha] = c_m x^m + \ldots + c_1 x + c_0$. We summarize the definitions and basic results of the regular continued fractions over $\mathbb{F}_p$ in [2] as follows: every element $\alpha \in \mathbb{F}_p$ can be uniquely represented as a finite or infinite expression of the form

$$\alpha = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots}} = [a_0, a_1, a_2, \ldots]$$

where $a_0 \in \mathbb{F}_p[x]$ is the integral part $[\alpha]$ and a_n's are in $\mathbb{F}_p[x]\backslash\mathbb{F}_p$ $(n \geq 1)$. The polynomials a_n are called the *partial quotients* of α and $\alpha_n = [a_n, a_{n+1}, \ldots]$ are called the nth *complete quotient* of α. In order to establish convergence to α, we define two sequences $\{A_n\}$ and $\{B_n\}$ in the following way

$$\begin{array}{llll} A_{-1} = 1, & A_0 = a_0, & A_{n+1} = a_{n+1}A_n + A_{n-1} & (n \geq 0), \\ B_{-1} = 0, & B_0 = 1, & B_{n+1} = a_{n+1}B_n + B_{n-1} & (n \geq 0). \end{array}$$

The two sequences then satisfies the properties below.

Lemma 2.1. *For any $n \geq 0, \beta \in \mathbb{F}_p\backslash\{0\}$, we have*

1. $\dfrac{\beta A_n + A_{n-1}}{\beta B_n + B_{n-1}} = [a_0, a_1, a_2, \ldots, a_n, \beta]$,
2. $A_n B_{n-1} - A_{n-1} B_n = (-1)^{n-1}$,
3. $|B_n| > |B_{n-1}| > 0$,
4. $\left|\alpha - \dfrac{A_n}{B_n}\right| = \dfrac{1}{|a_{n+1}||B_n|^2}$ $(n \geq 1)$.

From Lemma 2.1 (1) and (2), we have $\frac{A_n}{B_n}$ are reduced fractions and satisfy

$$\frac{A_n}{B_n} = [a_0, a_1, a_2, \ldots, a_n].$$

Moreover, by Lemma 2.1 (3) and (4), we have

$$\left|\alpha - \frac{A_n}{B_n}\right| = \frac{1}{|a_{n+1}||B_n|^2} \to 0, \ (n \to \infty).$$

Then we call $\frac{A_n}{B_n}$ the nth *convergent* of the regular continued fraction of α where A_n and B_n are called the nth *partial numerator* and nth *partial denominator*, respectively. A characterization of rationality was also provided in [2], namely, α is rational if and only if the continued fraction of α is finite.

3 Farey Graphs

We now introduce the Farey graph over $\mathbb{F}_p$ as follows: the vertex set is

$$\chi_p = \left\{ \frac{p(x)}{q(x)} : p(x), q(x) \in \mathbb{F}_p[x] \text{ with } q(x) \neq 0 \text{ and } (p(x), q(x)) = 1 \right\}$$

and $\frac{p(x)}{q(x)}$ and $\frac{r(x)}{s(x)}$ are adjacent, denoted by $\frac{p(x)}{q(x)} \sim \frac{r(x)}{s(x)}$, if and only if

$$r(x)q(x) - s(x)p(x) \in \mathbb{F}_p \setminus \{0\}.$$

We consider the Farey graph over $\mathbb{F}_p$ as a simple and undirected graph. When there is no ambiguity, we use χ_p, or briefly χ, to stand for the Farey graph over $\mathbb{F}_p$.

It is clear to see that, for any c in $\mathbb{F}_p$ and for any $q(x)$ in $\mathbb{F}_p[x]\backslash\mathbb{F}_p$, two vertices c and $c+\frac{1}{q(x)}$ $\left(= \frac{cq(x)+1}{q(x)}\right)$ in χ are adjacent. In addition, the path from c to $c+\frac{1}{q(x)}$ defines the regular continued fraction of $c + \frac{1}{q(x)}$.

Note that in the next section we usually use long arrows to indicate and emphasize the direction of a path between two vertices, as it helps us visualize the relation between paths and its associated continued fraction.

Example 3.1. In the Farey graph χ_3, some examples of paths starting from the vertex x are shown below.

1. $x \longrightarrow \frac{x^2+1}{x} \longrightarrow \frac{x^3-x}{x^2+1} \longrightarrow \frac{x^4+1}{x^3-x} \longrightarrow \frac{x^5+x^3}{x^4+1}$
2. $x \longrightarrow \frac{x^2-1}{x} \longrightarrow \frac{x^3}{x^2+1} \longrightarrow \frac{x^4+x^2-1}{x^3-x} \longrightarrow \frac{-x^5+x}{-x^4-x^2+1}$
3. $x \longrightarrow \frac{x^2+1}{x} \longrightarrow \frac{x^3}{x^2-1} \longrightarrow \frac{x^4-x^2-1}{x^3+x} \longrightarrow \frac{-x^5-x^4+x^2+x+1}{-x^4-x^3+x^2-x+1}$
4. $x \longrightarrow \frac{x^2-1}{x} \longrightarrow \frac{x^3+x}{x^2-1} \longrightarrow \frac{-x^4+x^2+1}{-x^3} \longrightarrow \frac{x^6-x^4-x^3-x^2-x}{x^5-x^2+1}$

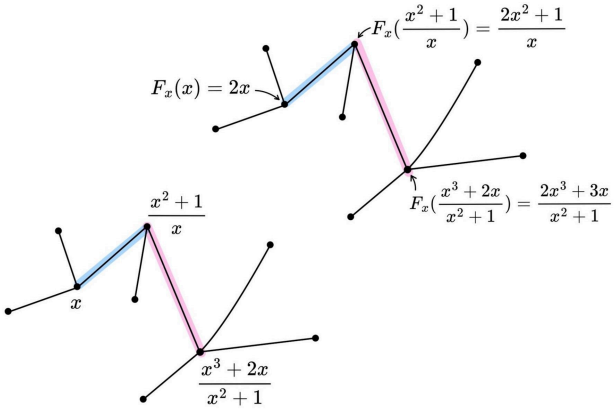
Now, for a fixed polynomial $T(x)$, we define F_T to be the bijection on χ that sends $\frac{p(x)}{q(x)}$ to $T(x) + \frac{p(x)}{q(x)}$. Then F_T is an automorphism as it also preserves the adjacency and non-adjacency. The map can be considered as

$$\frac{p(x)}{q(x)} \mapsto \begin{bmatrix} 1 & T(x) \\ 0 & 1 \end{bmatrix} \cdot \frac{p(x)}{q(x)}$$

where $\begin{bmatrix} a(x) & b(x) \\ c(x) & d(x) \end{bmatrix} \cdot \frac{p(x)}{q(x)} = \frac{a(x)p(x)+b(x)q(x)}{c(x)p(x)+d(x)q(x)}$. It is straightforward to see that $(F_T)^{-1} = F_{-T}$. Therefore, for any given polynomials $u(x)$ and $v(x)$, we have

$$F_{v-u}\left(\frac{u(x)}{1}\right) = \begin{bmatrix} 1 & v(x)-u(x) \\ 0 & 1 \end{bmatrix} \cdot \frac{u(x)}{1} = \frac{v(x)}{1} \tag{3.1}$$

The figure below shows how F_x preserves the adjacency of the graph χ_5.



Moreover, we have another class of automorphisms on χ that affects the denominator of the vertex. For any polynomial $T(x)$, we define G_T to be the bijection on χ defined by

$$\frac{p(x)}{q(x)} \mapsto \begin{bmatrix} 1 & 0 \\ T(x) & 1 \end{bmatrix} \cdot \frac{p(x)}{q(x)}$$

where the action $\cdot$ is as discussed above. That is,

$$G_T\left(\frac{p(x)}{q(x)}\right) = \frac{p(x)}{p(x)T(x)+q(x)}.$$

Then, G_T is an automorphism. Also, we have $G_T^{-1} = G_{-T}$ and

$$G_{v-u}\left(\frac{1}{u(x)}\right) = \begin{bmatrix} 1 & 0 \\ v(x)-u(x) & 1 \end{bmatrix} \cdot \frac{1}{u(x)} = \frac{1}{v(x)} \tag{3.2}$$

for any polynomials $u(x)$ and $v(x)$.

Now, we consider a very specific case when $T(x) = 1$. One can see that, for any polynomial $h(x)$, the vertices $h(x)$ and $F_T(h(x)) = F_1(h(x))\,(= h(x)+1)$ are adjacent. In addition, $F_1^k(h(x)) = h(x)+k$, and if our field of consideration has characteristic m, then the graph have a cycle of length m, namely, $(h(x), h(x)+1, h(x)+2, \ldots, h(x)+m-1)$. This implies that χ contains infinitely many (disjoint) cycles of length m.

Theorem 3.2. *The graph χ_p contains infinitely many cycles of length* $\mathrm{char}(\mathbb{F}_p)$*, the characteristic of $\mathbb{F}_p$.*

In the rest of this section we provide some useful local information of vertices in χ; more details can be found in [7]. The definitions given below introduce a notion of neighbors of a vertex in our Farey graphs.

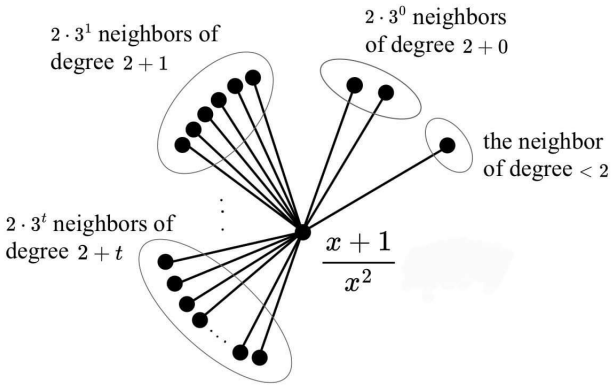
Definition 3.3. For any distinct vertices $\frac{p(x)}{q(x)}, \frac{h(x)}{k(x)}$ in χ with $q(x)$ and $k(x)$ monic, we say that $\frac{h(x)}{k(x)}$ is a deg $k(x)$-neighbor of $\frac{p(x)}{q(x)}$ if $\left|\frac{p(x)}{q(x)} - \frac{h(x)}{k(x)}\right| \leq \left|\frac{p(x)}{q(x)} - \frac{h'(x)}{k'(x)}\right|$ for every $\frac{h'(x)}{k'(x)}$ with deg $k'(x) =$ deg $k(x)$.

Definition 3.4. The fraction $\frac{p(x)}{q(x)}, \frac{h(x)}{k(x)}$ in χ are neighbor if $\frac{p(x)}{q(x)}$ is a $\deg q(x)$-neighbor of $\frac{h(x)}{k(x)}$ and $\frac{h(x)}{k(x)}$ is a $\deg k(x)$-neighbor of $\frac{p(x)}{q(x)}$.

Theorem 3.5. [7, Theorem 1] *Any distinct vertices $\frac{p(x)}{q(x)}, \frac{h(x)}{k(x)}$ in χ, with $q(x)$ and $k(x)$ monic, are neighbors if and only if* $\deg\left(p(x)k(x) - h(x)q(x)\right) = 0$.

Theorem 3.6. [7, Theorem 2] *Let $\frac{p(x)}{q(x)}$ be a vertex in χ_p, with $q(x)$ monic, and $q = \deg q(x)$. Then $\frac{p(x)}{q(x)}$ has exactly $(p-1)p^t$ neighbors of degree $q+t$ for any $t \geq 0$, and has only one neighbor of degree less than q.*

The figure below demonstrates local information at vertex $\frac{x+1}{x^2}$ in χ_3.



4 Some Relationships between Farey Graphs and Continued Fractions

In this section, some relationships between Farey graphs and continued fractions are provided. We first show how continued fractions define their associated paths in our Farey graphs.

Theorem 4.1. *The value of every finite regular continued fraction belongs to χ and every finite continued fraction defines a path from its integral part to its value with the convergents as the vertices.*

Proof. The first part is clear. Let $[a_0, a_1, a_2, \ldots, a_n]$ where $a_0 \in \mathbb{F}_p[x], a_i \in \mathbb{F}_p[x] \backslash \mathbb{F}_p$ $(1 \leq i \leq n)$ be a regular continued fraction with the ith convergent $\frac{A_i}{B_i}$ for $0 \leq i \leq n$. For each $1 \leq i \leq n$, by Lemma 2.1 (2), we have $A_iB_{i-1} - A_{i-1}B_i = (-1)^{i-1}$. Therefore, $\frac{A_{i-1}}{B_{i-1}}$ and $\frac{A_i}{B_i}$ are adjacent and the given regular continued fraction defines the path

$$a_0 \longrightarrow \frac{A_1}{B_1} \longrightarrow \frac{A_2}{B_2} \longrightarrow \cdots \longrightarrow \frac{A_n}{B_n}$$

from a_0 to $\frac{A_n}{B_n}$ as required.

□

Throughout, let $\mathbb{F} := \mathbb{F}_p$ where $p = 2, 3$. We use the following notation. Let $n \in \mathbb{N}$, $p_{-1} = 1 = q_0$, $q_{-1} = 0$ and $p_0 \in \mathbb{F}[x]$. For all $i \in \{1, 2, ..., n\}$, let $p_i, q_i \in \mathbb{F}[x]$ with $(p_i, q_i) = 1$ and $\deg(q_{i-1}) < \deg(q_i)$. Note that, in the remainder, our results may have no difference when considered in $\mathbb{F}_2(x)$ as $1 \equiv -1 \pmod 2$. However, they become varied and interesting over $\mathbb{F}_3$.

Theorem 4.2. *If $q_{i-1} \mid (q_i - q_{i-2})$ and $p_i q_{i-1} - p_{i-1} q_i = (-1)^{i-1}$ $(1 \leq i \leq n)$, then the path*

$$a_0 := \frac{p_0}{q_0} \longrightarrow \frac{p_1}{q_1} \longrightarrow \frac{p_2}{q_2} \longrightarrow \frac{p_3}{q_3} \longrightarrow \cdots \longrightarrow \frac{p_n}{q_n}$$

from a_0 to $\frac{p_n}{q_n}$ defines the finite regular continued fraction of $\frac{p_n}{q_n}$ where each vertex $\frac{p_i}{q_i}$ defines its ith convergent. In particular, the ith partial numerator and partial denominator are p_i and q_i, respectively, with the partial quotient $a_i = \frac{q_i - q_{i-2}}{q_{i-1}}$ $(1 \leq i \leq n)$.

Proof. With those assumptions we will prove the statement by induction. Consider the path

$$\frac{p_0}{q_0} \longrightarrow \frac{p_1}{q_1}.$$

By the assumption, we have

$$p_1 q_0 - p_0 q_1 = 1$$

which implies that $p_1 = a_0 q_1 + 1$. Moreover, since $\deg(q_1) > \deg(q_0) = 0$, we have $q_1 \in \mathbb{F}[x] \backslash \mathbb{F}$. Then

$$\frac{p_1}{q_1} = \frac{a_0 q_1 + 1}{q_1} = a_0 + \frac{1}{q_1} = a_0 + \frac{1}{a_1}$$

where the latter expression is the regular continued fraction of $\frac{p_1}{q_1}$ with the partial quotient

$$a_1 = q_1 = \frac{q_1 - q_{-1}}{q_0}.$$

Note that $A_0 = p_0, B_0 = q_0$ and the first partial denominator $B_1 = a_1 B_0 + B_{-1} = a_1 = q_1$. Since $\frac{p_1}{q_1} = \frac{A_1}{B_1}$, we have $A_1 = p_1$.

Suppose that the statement is true for k. Given a path

$$\frac{p_0}{q_0} \longrightarrow \frac{p_1}{q_1} \longrightarrow \frac{p_2}{q_2} \longrightarrow \frac{p_3}{q_3} \longrightarrow \cdots \longrightarrow \frac{p_k}{q_k} \longrightarrow \frac{p_{k+1}}{q_{k+1}}.$$

Then the inductive hypothesis implies that the path

$$\frac{p_0}{q_0} \longrightarrow \frac{p_1}{q_1} \longrightarrow \frac{p_2}{q_2} \longrightarrow \frac{p_3}{q_3} \longrightarrow \cdots \longrightarrow \frac{p_k}{q_k}$$

from $\frac{p_0}{q_0}$ to $\frac{p_k}{q_k}$ defines the finite regular continued fraction of $\frac{p_k}{q_k}$, say

$$\frac{p_k}{q_k} = [a_0, a_1, \ldots, a_k]$$

where $a_i = \frac{q_i - q_{i-2}}{q_{i-1}} \in \mathbb{F}[x] \backslash \mathbb{F}$ $(1 \leq i \leq k)$ and each vertex $\frac{p_i}{q_i}$ $(1 \leq i \leq k)$ defines its ith partial numerator and denominator, respectively. The assumption $p_{k+1} q_k - p_k q_{k+1} = (-1)^k$ implies that $p_{k+1} = \frac{p_k q_{k+1} + (-1)^k}{q_k}$. Moreover, since $q_k \mid (q_{k+1} - q_{k-1})$ and $\deg\left(q_{k+1} - q_{k-1}\right) = \deg\left(q_{k+1}\right) > \deg\left(q_k\right)$, we have $\frac{q_{k+1} - q_{k-1}}{q_k} \in \mathbb{F}[x] \backslash \mathbb{F}$. Now, one can see that

$$\begin{aligned}
\frac{p_{k+1}}{q_{k+1}} = \frac{\left(\frac{p_k q_{k+1} + (-1)^k}{q_k}\right)}{q_{k+1}} &= \frac{\left(\frac{q_{k+1} - q_{k-1}}{q_k}\right) p_k + p_{k-1}}{\left(\frac{q_{k+1} - q_{k-1}}{q_k}\right) q_k + q_{k-1}} \\
&= \frac{\left(\frac{q_{k+1} - q_{k-1}}{q_k}\right) A_k + A_{k-1}}{\left(\frac{q_{k+1} - q_{k-1}}{q_k}\right) B_k + B_{k-1}} \\
&= [a_0, a_1, \ldots, a_{k+1}],
\end{aligned}$$

where $a_{k+1} = \frac{q_{k+1} - q_{k-1}}{q_k} \in \mathbb{F}[x] \backslash \mathbb{F}$, by Lemma 2.1(1). We also have, $B_{k+1} = a_{k+1} B_k + B_{k-1} = q_{k+1}$. This implies that $A_{k+1} = p_{k+1}$ as required. □

Theorem 4.3. *If $q_{i-1} \mid (q_i - q_{i-2})$ and $p_i q_{i-1} - p_{i-1} q_i = (-1)^i$ $(1 \leq i \leq n)$, then the path*

$$a_0 := \frac{p_0}{q_0} \longrightarrow \frac{p_1}{q_1} \longrightarrow \frac{p_2}{q_2} \longrightarrow \frac{p_3}{q_3} \longrightarrow \cdots \longrightarrow \frac{p_n}{q_n}$$

from a_0 to $\frac{p_n}{q_n}$ defines the finite regular continued fraction of $\frac{p_n}{q_n}$ where each vertex $\frac{p_i}{q_i}$ defines its ith convergent. In particular, the ith partial numerator and partial denominator are $(-1)^i p_i$ and $(-1)^i q_i$, respectively, with the partial quotient $a_i = -\left(\frac{q_i - q_{i-2}}{q_{i-1}}\right)$ $(1 \leq i \leq n)$.

Proof. Consider the path

$$\frac{p_0}{q_0} \longrightarrow \frac{p_1}{q_1}.$$

By the assumption, we have $p_1 = a_0 q_1 - 1$. Moreover, since $\deg(q_1) > \deg(q_0) = 0$, we have $q_1 \in \mathbb{F}[x] \backslash \mathbb{F}$. We then have the regular continued fraction of $\frac{p_1}{q_1}$ as

$$\frac{p_1}{q_1} = \frac{a_0 q_1 - 1}{q_1} = a_0 + \frac{1}{-q_1} = a_0 + \frac{1}{a_1}$$

with the partial quotient

$$a_1 = -q_1 = -\left(\frac{q_1 - q_{-1}}{q_0}\right).$$

Note that $A_0 = p_0, B_0 = q_0$ the first partial denominator $B_1 = a_1 B_0 + B_{-1} = a_1 = (-1)^1 q_1$. Since $\frac{p_1}{q_1} = \frac{A_1}{B_1}$ and they are reduced fractions, $A_1 = (-1)^1 p_1$.

Suppose that the statement is true for k. Given a path

$$\frac{p_0}{q_0} \longrightarrow \frac{p_1}{q_1} \longrightarrow \frac{p_2}{q_2} \longrightarrow \frac{p_3}{q_3} \longrightarrow \cdots \longrightarrow \frac{p_k}{q_k} \longrightarrow \frac{p_{k+1}}{q_{k+1}}.$$

Then the path

$$\frac{p_0}{q_0} \longrightarrow \frac{p_1}{q_1} \longrightarrow \frac{p_2}{q_2} \longrightarrow \frac{p_3}{q_3} \longrightarrow \cdots \longrightarrow \frac{p_k}{q_k}$$

from $\frac{p_0}{q_0}$ to $\frac{p_k}{q_k}$ defines the finite regular continued fraction of $\frac{p_k}{q_k}$, say

$$\frac{p_k}{q_k} = [a_0, a_1, \ldots, a_k]$$

where $a_i = -\left(\frac{q_i - q_{i-2}}{q_{i-1}}\right) \in \mathbb{F}[x] \backslash \mathbb{F}$ $(1 \leq i \leq k)$ and for each $1 \leq i \leq k$, its ith partial numerator and denominator are $(-1)^i p_i$ and $(-1)^i q_i$, respectively. The assumption $p_{k+1} q_k - p_k q_{k+1} = (-1)^{k+1}$ implies that $p_{k+1} = \frac{p_k q_{k+1} + (-1)^{k+1}}{q_k}$. Moreover, since $q_k \mid (q_{k+1} - q_{k-1})$ and $\deg\left(q_{k+1} - q_{k-1}\right) = \deg\left(q_{k+1}\right) > \deg\left(q_k\right)$, we have $\frac{q_{k+1} - q_{k-1}}{q_k} \in \mathbb{F}[x] \backslash \mathbb{F}$. Now, one can see that

$$\begin{aligned}
\frac{p_{k+1}}{q_{k+1}} = \frac{\left(\frac{p_k q_{k+1} + (-1)^{k+1}}{q_k}\right)}{q_{k+1}} &= \frac{\left(\frac{q_{k+1} - q_{k-1}}{q_k}\right) p_k + p_{k-1}}{\left(\frac{q_{k+1} - q_{k-1}}{q_k}\right) q_k + q_{k-1}} \\
&= \frac{-\left(\frac{q_{k+1} - q_{k-1}}{q_k}\right)(-1)^k p_k + (-1)^{k-1} p_{k-1}}{-\left(\frac{q_{k+1} - q_{k-1}}{q_k}\right)(-1)^k q_k + (-1)^{k-1} q_{k-1}} \\
&= \frac{-\left(\frac{q_{k+1} - q_{k-1}}{q_k}\right) A_k + A_{k-1}}{-\left(\frac{q_{k+1} - q_{k-1}}{q_k}\right) B_k + B_{k-1}} \\
&= [a_0, a_1, \ldots, a_{k+1}],
\end{aligned}$$

where $a_{k+1} = -\left(\frac{q_{k+1}-q_{k-1}}{q_k}\right) \in \mathbb{F}[x]\backslash\mathbb{F}$, by Lemma 2.1(1). Therefore,

$$B_{k+1} = a_{k+1}B_k + B_{k-1} = a_{k+1}(-1)^k q_k + (-1)^{k-1}q_{k-1} = (-1)^{k+1}(-a_{k+1}q_k + q_{k-1}) = (-1)^{k+1}q_{k+1}.$$

This implies that $A_{k+1} = (-1)^{k+1}p_{k+1}$ as required. □

Theorem 4.4. *If $q_{i-1} \mid (q_i + q_{i-2})$ and $p_i q_{i-1} - p_{i-1}q_i = 1$ $(1 \le i \le n)$, then the path*

$$a_0 := \frac{p_0}{q_0} \longrightarrow \frac{p_1}{q_1} \longrightarrow \frac{p_2}{q_2} \longrightarrow \frac{p_3}{q_3} \longrightarrow \cdots \longrightarrow \frac{p_n}{q_n}$$

from a_0 to $\frac{p_n}{q_n}$ defines the finite regular continued fraction of $\frac{p_n}{q_n}$ where each vertex $\frac{p_i}{q_i}$ defines its ith convergent. In particular, the ith partial numerator and partial denominator are

$$A_i = \begin{cases} p_i, & \text{if } i \equiv 0,1 \pmod 4 \\ -p_i, & \text{if } i \equiv 2,3 \pmod 4 \end{cases}, B_i = \begin{cases} q_i, & \text{if } i \equiv 0,1 \pmod 4 \\ -q_i, & \text{if } i \equiv 2,3 \pmod 4 \end{cases},$$

respectively, with the partial quotient $a_i = (-1)^{i+1}\left(\frac{q_i+q_{i-2}}{q_{i-1}}\right)$ $(1 \le i \le n)$.

Proof. Note that, for any $i \in \{1, 2, \ldots, n-1\}$, we have

$$\frac{p_{i+1}}{q_{i+1}} = \frac{\left(\frac{p_i q_{i+1}+1}{q_i}\right)}{q_{i+1}} = \frac{\left(\frac{q_{i+1}+q_{i-1}}{q_i}\right)p_i - p_{i-1}}{\left(\frac{q_{i+1}+q_{i-1}}{q_i}\right)q_i - q_{i-1}}. \tag{4.1}$$

Consider the path $\frac{p_0}{q_0} \longrightarrow \frac{p_1}{q_1}$. We have

$$\frac{p_1}{q_1} = \frac{a_0 q_1 + 1}{q_1} = a_0 + \frac{1}{q_1} = [a_0, a_1]$$

where $a_1 = q_1 = (-1)^2\left(\frac{q_1+q_{-1}}{q_0}\right)$. Again, we have $B_0 = q_0$ and $A_0 = p_0$. Moreover, $B_1 = a_1B_0 + B_{-1} = a_1 = q_1$ and so $A_1 = p_1$. By equation (4.1), we have

$$\frac{p_2}{q_2} = \frac{\left(\frac{q_2+q_0}{q_1}\right)p_1 - p_0}{\left(\frac{q_2+q_0}{q_1}\right)q_1 - q_0} = \frac{-\left(\frac{q_2+q_0}{q_1}\right)p_1 + p_0}{-\left(\frac{q_2+q_0}{q_1}\right)q_1 + q_0} = \frac{a_2A_1 + A_0}{a_2B_1 + B_0} = [a_0, a_1, a_2].$$

Here, $a_2 = (-1)^3\left(\frac{q_2+q_0}{q_1}\right) \in \mathbb{F}[x] \setminus \mathbb{F}$. It is easy to see that $B_2 = -q_2$ and so $A_2 = -p_2$. Again by equation (4.1), we have

$$\frac{p_3}{q_3} = \frac{\left(\frac{q_3+q_1}{q_2}\right)p_2 - p_1}{\left(\frac{q_3+q_1}{q_2}\right)q_2 - q_1} = \frac{\left(\frac{q_3+q_1}{q_2}\right)(-p_2) + p_1}{\left(\frac{q_3+q_1}{q_2}\right)(-q_2) + q_1} = \frac{a_3A_2 + A_1}{a_3B_2 + B_1} = [a_0, a_1, a_2, a_3]$$

where $a_3 = (-1)^4\left(\frac{q_3+q_1}{q_2}\right) \in \mathbb{F}[x] \setminus \mathbb{F}$. Similarly, we get $B_3 = a_3B_2 + B_1 = -a_3q_2 + q_1 = -q_3$ and so $A_3 = -p_3$. Continuing in the same manner, we have

$$\frac{p_4}{q_4} = \frac{\left(\frac{q_4+q_2}{q_3}\right)p_3 - p_2}{\left(\frac{q_4+q_2}{q_3}\right)q_3 - q_2} = \frac{a_4A_3 + A_2}{a_4B_3 + B_2} = [a_0, a_1, a_2, a_3, a_4]$$

where $a_4 = (-1)^5\left(\frac{q_4+q_2}{q_3}\right) \in \mathbb{F}[x] \setminus \mathbb{F}$, $B_4 = q_4$ and $A_4 = p_4$.

Assume that the statement is true for k. Given a path

$$\frac{p_0}{q_0} \longrightarrow \frac{p_1}{q_1} \longrightarrow \frac{p_2}{q_2} \longrightarrow \frac{p_3}{q_3} \longrightarrow \cdots \longrightarrow \frac{p_k}{q_k} \longrightarrow \frac{p_{k+1}}{q_{k+1}}.$$

Then the path

$$\frac{p_0}{q_0} \longrightarrow \frac{p_1}{q_1} \longrightarrow \frac{p_2}{q_2} \longrightarrow \frac{p_3}{q_3} \longrightarrow \cdots \longrightarrow \frac{p_k}{q_k}$$

from $\frac{p_0}{q_0}$ to $\frac{p_k}{q_k}$ defines the finite regular continued fraction of $\frac{p_k}{q_k}$, say

$$\frac{p_k}{q_k} = [a_0, a_1, a_2, \ldots, a_k]$$

where, for each $1 \leq i \leq k$, $a_i = (-1)^{i+1}\left(\frac{q_i + q_{i-2}}{q_{i-1}}\right) \in \mathbb{F}[x] \setminus \mathbb{F}$ and its ith partial numerator and denominator are

$$A_i = \begin{cases} p_i, & \text{if } i \equiv 0,1 \pmod 4 \\ -p_i, & \text{if } i \equiv 2,3 \pmod 4 \end{cases}, \quad B_i = \begin{cases} q_i, & \text{if } i \equiv 0,1 \pmod 4 \\ -q_i, & \text{if } i \equiv 2,3 \pmod 4 \end{cases},$$

respectively. We divide the proof into 4 cases as follows:

Case 1: $k \equiv 0 \pmod 4$. Then $k-1 \equiv 3 \pmod 4$ and we therefore have $A_k = p_k, B_k = q_k, A_{k-1} = -p_{k-1}$ and $B_{k-1} = -q_{k-1}$. By equation (4.1), we have

$$\frac{p_{k+1}}{q_{k+1}} = \frac{\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right)p_k - p_{k-1}}{\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right)q_k - q_{k-1}} = \frac{a_{k+1}A_k + A_{k-1}}{a_{k+1}B_k + B_{k-1}} = [a_0, a_1, \ldots, a_{k+1}]$$

where $a_{k+1} = (-1)^{k+2}\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right) \in \mathbb{F}[x] \setminus \mathbb{F}$. We also have, $B_{k+1} = a_{k+1}B_k + B_{k-1} = a_{k+1}q_k - q_{k-1} = q_{k+1}$. Therefore, $A_{k+1} = p_{k+1}$.

Case 2: $k \equiv 1 \pmod 4$. Then $k-1 \equiv 0 \pmod 4$ and $k+2 \equiv 3 \pmod 4$. We then have $A_k = p_k, B_k = q_k, A_{k-1} = p_{k-1}$ and $B_{k-1} = q_{k-1}$. By equation (4.1), we have

$$\begin{aligned}\frac{p_{k+1}}{q_{k+1}} = \frac{\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right)p_k - p_{k-1}}{\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right)q_k - q_{k-1}} &= \frac{(-1)^{k+2}\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right)p_k + p_{k-1}}{(-1)^{k+2}\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right)q_k + q_{k-1}} \\ &= \frac{a_{k+1}A_k + A_{k-1}}{a_{k+1}B_k + B_{k-1}} = [a_0, a_1, \ldots, a_{k+1}]\end{aligned}$$

where $a_{k+1} = (-1)^{k+2}\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right) \in \mathbb{F}[x] \setminus \mathbb{F}$. We also have, $B_{k+1} = a_{k+1}B_k + B_{k-1} = a_{k+1}q_k + q_{k-1} = -q_{k+1}$. Therefore, $A_{k+1} = -p_{k+1}$.

Case 3: $k \equiv 2 \pmod 4$. Then $k-1 \equiv 1 \pmod 4$ and $k+2 \equiv 0 \pmod 4$. We then have $A_k = -p_k, B_k = -q_k, A_{k-1} = p_{k-1}$ and $B_{k-1} = q_{k-1}$. By the equation (4.1), we have

$$\begin{aligned}\frac{p_{k+1}}{q_{k+1}} = \frac{\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right)p_k - p_{k-1}}{\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right)q_k - q_{k-1}} &= \frac{(-1)^{k+2}\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right)(-p_k) + p_{k-1}}{(-1)^{k+2}\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right)(-q_k) + q_{k-1}} \\ &= \frac{a_{k+1}A_k + A_{k-1}}{a_{k+1}B_k + B_{k-1}} = [a_0, a_1, \ldots, a_{k+1}]\end{aligned}$$

where $a_{k+1} = (-1)^{k+2}\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right) \in \mathbb{F}[x] \setminus \mathbb{F}$. We also have, $B_{k+1} = a_{k+1}B_k + B_{k-1} = a_{k+1}(-q_k) + q_{k-1} = -q_{k+1}$. Therefore, $A_{k+1} = -p_{k+1}$.

Case 4: $k \equiv 3 \pmod 4$. Then $k-1 \equiv 2 \pmod 4$ and $k+2 \equiv 0 \pmod 4$. We then have $A_k = -p_k, B_k = -q_k, A_{k-1} = -p_{k-1}$ and $B_{k-1} = -q_{k-1}$. By equation (4.1), we have

$$\frac{p_{k+1}}{q_{k+1}} = \frac{\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right)p_k - p_{k-1}}{\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right)q_k - q_{k-1}} = \frac{(-1)^{k+2}\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right)(-p_k) - p_{k-1}}{(-1)^{k+2}\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right)(-q_k) - q_{k-1}}$$
$$= \frac{a_{k+1}A_k + A_{k-1}}{a_{k+1}B_k + B_{k-1}} = [a_0, a_1, \ldots, a_{k+1}]$$

where $a_{k+1} = (-1)^{k+2}\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right) \in \mathbb{F}[x] \setminus \mathbb{F}$. We also have, $B_{k+1} = a_{k+1}B_k + B_{k-1} = a_{k+1}(-q_k) + (-q_{k-1}) = q_{k+1}$. So $A_{k+1} = p_{k+1}$. □

Theorem 4.5. *If $q_{i-1} \mid (q_i + q_{i-2})$ and $p_iq_{i-1} - p_{i-1}q_i = -1$ $(1 \le i \le n)$, then the path*

$$a_0 := \frac{p_0}{q_0} \longrightarrow \frac{p_1}{q_1} \longrightarrow \frac{p_2}{q_2} \longrightarrow \frac{p_3}{q_3} \longrightarrow \cdots \longrightarrow \frac{p_n}{q_n}$$

from a_0 to $\frac{p_n}{q_n}$ defines the finite regular continued fraction of $\frac{p_n}{q_n}$ where each vertex $\frac{p_i}{q_i}$ defines its ith convergent. In particular, the ith partial numerator and partial denominator are

$$A_i = \begin{cases} p_i, & \text{if } i \equiv 0,3 \pmod 4 \\ -p_i, & \text{if } i \equiv 1,2 \pmod 4 \end{cases}, B_i = \begin{cases} q_i, & \text{if } i \equiv 0,3 \pmod 4 \\ -q_i, & \text{if } i \equiv 1,2 \pmod 4 \end{cases},$$

respectively, with the partial quotient $a_i = (-1)^i\left(\frac{q_i+q_{i-2}}{q_{i-1}}\right)$ $(1 \le i \le n)$.

Proof. Note that, for each $i \in \{1, 2, \ldots, n-1\}$, the equation (4.1) is also true. That is,

$$\frac{p_{i+1}}{q_{i+1}} = \frac{\left(\frac{q_{i+1}+q_{i-1}}{q_i}\right)p_i - p_{i-1}}{\left(\frac{q_{i+1}+q_{i-1}}{q_i}\right)q_i - q_{i-1}}.$$

Consider the path $\frac{p_0}{q_0} \longrightarrow \frac{p_1}{q_1}$. We have

$$\frac{p_1}{q_1} = \frac{a_0q_1 - 1}{q_1} = a_0 + \frac{1}{-q_1} = [a_0, a_1]$$

where $a_1 = -q_1 = (-1)^1\left(\frac{q_1+q_{-1}}{q_0}\right)$. Again, we have $B_0 = q_0$ and $A_0 = p_0$. Moreover, $B_1 = a_1B_0 + B_{-1} = a_1 = -q_1$ and so $A_1 = -p_1$. By equation (4.1), we have

$$\frac{p_2}{q_2} = \frac{\left(\frac{q_2+q_0}{q_1}\right)p_1 - p_0}{\left(\frac{q_2+q_0}{q_1}\right)q_1 - q_0} = \frac{\left(\frac{q_2+q_0}{q_1}\right)(-p_1) + p_0}{\left(\frac{q_2+q_0}{q_1}\right)(-q_1) + q_0} = \frac{a_2A_1 + A_0}{a_2B_1 + B_0} = [a_0, a_1, a_2].$$

Here, $a_2 = (-1)^2\left(\frac{q_2+q_0}{q_1}\right) \in \mathbb{F}[x] \setminus \mathbb{F}$. It is easy to see that $B_2 = -q_2$ and so $A_2 = -p_2$. Again by equation (4.1), we have

$$\frac{p_3}{q_3} = \frac{\left(\frac{q_3+q_1}{q_2}\right)p_2 - p_1}{\left(\frac{q_3+q_1}{q_2}\right)q_2 - q_1} = \frac{-\left(\frac{q_3+q_1}{q_2}\right)(-p_2) - p_1}{-\left(\frac{q_3+q_1}{q_2}\right)(-q_2) - q_1} = \frac{a_3A_2 + A_1}{a_3B_2 + B_1} = [a_0, a_1, a_2, a_3]$$

where $a_3 = (-1)^3\left(\frac{q_3+q_1}{q_2}\right) \in \mathbb{F}[x] \setminus \mathbb{F}$. Similarly, we get $B_3 = q_3$ and $A_3 = p_3$. Continuing in the same manner, we have

$$\frac{p_4}{q_4} = \frac{\left(\frac{q_4+q_2}{q_3}\right)p_3 - p_2}{\left(\frac{q_4+q_2}{q_3}\right)q_3 - q_2} = \frac{a_4A_3 + A_2}{a_4B_3 + B_2} = [a_0, a_1, a_2, a_3, a_4]$$

where $a_4 = (-1)^4\left(\frac{q_4+q_2}{q_3}\right) \in \mathbb{F}[x] \setminus \mathbb{F}$, $B_4 = q_4$ and $A_4 = p_4$.

Assume that the statement is true for k. Given a path

$$\frac{p_0}{q_0} \longrightarrow \frac{p_1}{q_1} \longrightarrow \frac{p_2}{q_2} \longrightarrow \frac{p_3}{q_3} \longrightarrow \cdots \longrightarrow \frac{p_k}{q_k} \longrightarrow \frac{p_{k+1}}{q_{k+1}}.$$

Then the path

$$\frac{p_0}{q_0} \longrightarrow \frac{p_1}{q_1} \longrightarrow \frac{p_2}{q_2} \longrightarrow \frac{p_3}{q_3} \longrightarrow \cdots \longrightarrow \frac{p_k}{q_k}$$

from $\frac{p_0}{q_0}$ to $\frac{p_k}{q_k}$ defines the finite regular continued fraction of $\frac{p_k}{q_k}$, say

$$\frac{p_k}{q_k} = [a_0, a_1, a_2, \ldots, a_k]$$

where, for each $1 \leq i \leq k$, $a_i = (-1)^{i+1}\left(\frac{q_i + q_{i-2}}{q_{i-1}}\right) \in \mathbb{F}[x] \setminus \mathbb{F}$ and its ith partial numerator and denominator are

$$A_i = \begin{cases} p_i, & \text{if } i \equiv 0,3 \pmod 4 \\ -p_i, & \text{if } i \equiv 1,2 \pmod 4 \end{cases}, B_i = \begin{cases} q_i, & \text{if } i \equiv 0,3 \pmod 4 \\ -q_i, & \text{if } i \equiv 1,2 \pmod 4 \end{cases},$$

respectively. We divide the proof into 4 cases as follows:

Case 1: $k \equiv 0 \pmod 4$. Then $k-1 \equiv 3 \pmod 4$ and we therefore have $A_k = p_k, B_k = q_k, A_{k-1} = p_{k-1}$ and $B_{k-1} = q_{k-1}$. By equation (4.1), we have

$$\frac{p_{k+1}}{q_{k+1}} = \frac{(-1)^{k+1}\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right)p_k + p_{k-1}}{(-1)^{k+1}\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right)q_k + q_{k-1}} = \frac{a_{k+1}A_k + A_{k-1}}{a_{k+1}B_k + B_{k-1}} = [a_0, a_1, \ldots, a_{k+1}]$$

where $a_{k+1} = (-1)^{k+1}\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right) \in \mathbb{F}[x] \setminus \mathbb{F}$. We also have $B_{k+1} = q_{k+1}$ and $A_{k+1} = p_{k+1}$.

Case 2: $k \equiv 1 \pmod 4$. Then $k-1 \equiv 0 \pmod 4$. We then have $A_k = -p_k, B_k = -q_k, A_{k-1} = p_{k-1}$ and $B_{k-1} = q_{k-1}$. By equation (4.1), we have

$$\frac{p_{k+1}}{q_{k+1}} = \frac{(-1)^{k+1}\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right)(-p_k) + p_{k-1}}{(-1)^{k+1}\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right)(-q_k) + q_{k-1}} = \frac{a_{k+1}A_k + A_{k-1}}{a_{k+1}B_k + B_{k-1}} = [a_0, a_1, \ldots, a_{k+1}]$$

where $a_{k+1} = (-1)^{k+1}\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right) \in \mathbb{F}[x] \setminus \mathbb{F}$. We also have $B_{k+1} = q_{k+1}$ and $A_{k+1} = p_{k+1}$.

Case 3: $k \equiv 2 \pmod 4$. Then $k-1 \equiv 1 \pmod 4$. We then have $A_k = -p_k, B_k = -q_k, A_{k-1} = -p_{k-1}$ and $B_{k-1} = -q_{k-1}$. By equation (4.1), we have

$$\frac{p_{k+1}}{q_{k+1}} = \frac{(-1)^{k+1}\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right)(-p_k) - p_{k-1}}{(-1)^{k+1}\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right)(-q_k) - q_{k-1}} = \frac{a_{k+1}A_k + A_{k-1}}{a_{k+1}B_k + B_{k-1}} = [a_0, a_1, \ldots, a_{k+1}]$$

where $a_{k+1} = (-1)^{k+1}\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right) \in \mathbb{F}[x] \setminus \mathbb{F}$. We also have $B_{k+1} = q_{k+1}$ and $A_{k+1} = p_{k+1}$.
Case 4: $k \equiv 3 \pmod 4$. Then $k-1 \equiv 2 \pmod 4$. We then have $A_k = p_k, B_k = q_k, A_{k-1} = -p_{k-1}$ and $B_{k-1} = -q_{k-1}$. By equation (4.1), we have

$$\frac{p_{k+1}}{q_{k+1}} = \frac{(-1)^{k+1}\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right)p_k - p_{k-1}}{(-1)^{k+1}\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right)q_k - q_{k-1}} = \frac{a_{k+1}A_k + A_{k-1}}{a_{k+1}B_k + B_{k-1}} = [a_0, a_1, \ldots, a_{k+1}]$$

where $a_{k+1} = (-1)^{k+1}\left(\frac{q_{k+1}+q_{k-1}}{q_k}\right) \in \mathbb{F}[x] \setminus \mathbb{F}$. We also have $B_{k+1} = q_{k+1}$ and $A_{k+1} = p_{k+1}$. □

Example 4.6. For convenience, we set the notation as follows: any two vertices $p(x)/q(x)$ and $r(x)/s(x)$ in χ_3,

$$\begin{array}{lllll} \frac{p(x)}{q(x)} \xrightarrow{+} \frac{r(x)}{s(x)} & \text{means} & \frac{p(x)}{q(x)} \sim \frac{r(x)}{s(x)} & \text{with} & r(x)q(x) - s(x)p(x) = 1 \quad \text{and} \\ \frac{p(x)}{q(x)} \xrightarrow{-} \frac{r(x)}{s(x)} & \text{means} & \frac{p(x)}{q(x)} \sim \frac{r(x)}{s(x)} & \text{with} & r(x)q(x) - s(x)p(x) = -1. \end{array}$$

From Example 3.1, we obtain

1. by Theorem 4.2 that the path $x \xrightarrow{+} \frac{x^2+1}{x} \xrightarrow{-} \frac{x^3-x}{x^2+1} \xrightarrow{+} \frac{x^4+1}{x^3-x} \xrightarrow{-} \frac{x^5+x^3}{x^4+1}$ defines the regular continued fraction of
$$\frac{x^5+x^3}{x^4+1} = [x, x, x, x, x],$$
2. by Theorem 4.3 that the path $x \xrightarrow{-} \frac{x^2-1}{x} \xrightarrow{+} \frac{x^3}{x^2+1} \xrightarrow{-} \frac{x^4+x^2-1}{x^3-x} \xrightarrow{+} \frac{-x^5+x}{-x^4-x^2+1}$ defines the regular continued fraction of
$$\frac{-x^5+x}{-x^4-x^2+1} = [x, -x, -x, -x, x],$$
3. by Theorem 4.4 that the path $x \xrightarrow{+} \frac{x^2+1}{x} \xrightarrow{+} \frac{x^3}{x^2-1} \xrightarrow{+} \frac{x^4-x^2-1}{x^3+x} \xrightarrow{+} \frac{-x^5-x^4+x^2+x+1}{-x^4-x^3+x^2-x+1}$ defines the regular continued fraction of
$$\frac{-x^5-x^4+x^2+x+1}{-x^4-x^3+x^2-x+1} = [x, x, -x, x, x+1],$$
4. by Theorem 4.5 that the path $x \xrightarrow{-} \frac{x^2-1}{x} \xrightarrow{-} \frac{x^3+x}{x^2-1} \xrightarrow{-} \frac{-x^4+x^2+1}{-x^3} \xrightarrow{-} \frac{x^6-x^4-x^3-x^2-x}{x^5-x^2+1}$ defines the regular continued fraction of
$$\frac{x^6-x^4-x^3-x^2-x}{x^5-x^2+1} = [x, -x, x, x, -x^2].$$

5 Conclusion

In this article, we introduce Farey graphs over certain finite fields analogous to the classical case of rational numbers. We establish some relationships between these graphs and regular continued fractions. The study confirms the relationships between Farey graphs and continued fractions in such a way that continued fractions define paths whose vertices are convergents of continued fractions. On the other hand, we also provide explicit formulae of partial quotients and nth convergents of associated continued fractions for a given path.

References

[1] J.G. Jones and D. Singerman, *The modular group and generalized Farey graphs*, LMS Lect. Note Ser, **160** (1991), 316–338.

[2] T. Chaichana, V. Laohakosol and A. Harnchoowong, *Linear Independence of continued fractions in the field of formal series over a finite field*, Thai J. Math **4** (2006), 163–177.

[3] S. Kushwaha and R. Sarma, *Continued fractions arising from* $\mathcal{F}_{1,3}$, Ramanujan J. **46**(1) (2018), 605–631.

[4] S. Kushwaha and R. Sarma, *Farey-subgraphs and Continued Fractions*, Studia Sci. Math. Hungar. **59**(2) (2022), 164–182.

[5] C.D. Olds, *Continued Fractions*, The L.W. Singer Company, New York, 1963.

[6] R. Sarma, S. Kushwaha and K. Wicks, *Continued fractions arising from* $\mathcal{F}_{1,2}$, J. Number Theory **154** (2015), 179–200.

[7] W.A. Webb, *The Farey Series of polynomials over a finite field*, El. Math. **41** (1986), 6–11.

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

The Diameter and Girth of Subspace Inclusion Graphs Modulo Prime Powers

Siripong Sirisuk[1], and Juthamas Sangwisat[1,†,‡]

[1]Department of Mathematics and Statistics, Faculty of Science and Technology
Thammasat University, Pathum Thani 12120, Thailand

Abstract

Let $\mathbb{Z}_{p^s}$ denote the ring of integers modulo p^s, where p is a prime number and s is a positive integer. In this talk, we introduce the subspace inclusion graph of $\mathbb{Z}_{p^s}$, which is a graph whose vertices are the non-trivial proper subspaces of $\mathbb{Z}_{p^s}^n$ (for $n \geq 2$) and two distinct vertices are adjacent if and only if one includes the other. We determine some properties of the graph, including its order, vertex degrees, diameter and girth.

Keywords: subspace, diameter, girth.

2020 MSC: Primary 05C25; Secondary 05C50, 05C69, 15A03.

1 Introduction

Graphs associated with algebraic structures play a significant role in mathematics, as well as in other areas. Various types of graphs can be linked to many algebraic structures. Fields are commonly used algebraic objects to define and explore different types of algebraic graphs. Many studies have looked into graphs connected to subspaces of vector spaces. (cf. [1, 4, 20]). Das [3] introduced the subspace inclusion graph over a field. Its vertex set is the collection of non-trivial proper subspaces of a finite-dimensional vector space. Two vertices are adjacent if one is contained in the other. Various fundamental properties have been explored. Furthermore, many studies and applications have emerged since this research (cf. [2, 8, 18]).

Graphs over finite commutative rings have received significant attention, as evidenced by previous studies (cf. [7, 16]). Many studies have built upon the ideas of graphs initially used for finite fields, as shown by the works of [9, 17, 19]. Notable examples of graphs defined on rings of integers modulo prime powers include bilinear forms graphs [12, 13] and Grassmann graphs [10, 11]. These rings of integers modulo prime powers hold significant potential for applications in mathematics, coding theory and information theory.

This paper aims to extend Das's concept [3] of subspace inclusion graphs from fields to rings of integers modulo prime powers and explore their properties. The paper is organized as follows: we revisit properties related to subspaces and review key definitions and concepts in

†Speaker. ‡Corresponding author.
Email: siripong@mathstat.sci.tu.ac.th (S. Sirisuk), juthamas.sang@dome.tu.ac.th (J. Sangwisat).

graph theory. In section 3, we introduce the concept of subspace inclusion graphs over rings of integers modulo prime powers, providing fundamental results on their order, vertex degrees, diameter and girth.

2 Preliminaries

Throughout this paper, our rings are commutative and always contain the identity $1 \neq 0$.

Let $\mathbb{Z}_{p^s}$ denote the ring of integers modulo p^s, where p is a prime number and s is a positive integer. It is well-known that when $s = 1$, $\mathbb{Z}_p$ forms a finite field. $\mathbb{Z}_{p^s}$ is a significant algebraic structure with multiple properties. It is a Galois ring, a finite chain ring, a principal ideal ring and a commutative local ring (cf. [5, 14, 15, 21]). The ideals of $\mathbb{Z}_{p^s}$ form a chain as follows:

$$\{0\} = p^s\mathbb{Z}_{p^s} \subsetneq p^{s-1}\mathbb{Z}_{p^s} \subsetneq p^{s-2}\mathbb{Z}_{p^s} \subsetneq \cdots \subsetneq p^2\mathbb{Z}p^s \subsetneq p\mathbb{Z}_{p^s} \subsetneq \mathbb{Z}_{p^s}.$$

Hence, the principal ideal $p\mathbb{Z}_{p^s}$ serves as the unique maximal ideal of $\mathbb{Z}_{p^s}$. This implies that an element u is a unit in $\mathbb{Z}_{p^s}$ if and only if u does not belong to $p\mathbb{Z}_{p^s}$. It is also known that any element $a \in \mathbb{Z}_{p^s}$ can be expressed as $a = up^t$, where u is a unit and $0 \leq t \leq s$. Furthermore, the cardinality of $p^i\mathbb{Z}_{p^s}$ is given by $|p^i\mathbb{Z}_{p^s}| = p^{s-i}$ for all $i = 0, 1, \ldots, s$ and the order of $\mathbb{Z}^*_{p^s}$, the set of units, is $|\mathbb{Z}^*_{p^s}| = (p-1)p^{s-1}$.

Let n be a positive integer. Consider the $\mathbb{Z}_{p^s}$-module $\mathbb{Z}^n_{p^s}$. A set $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m\}$ of vectors in $\mathbb{Z}^n_{p^s}$ is said to be *linearly independent* if for any $a_1, a_2, \ldots, a_m$ in $\mathbb{Z}_{p^s}$, $a_1\vec{x}_1 + a_2\vec{x}_2 + \cdots + a_m\vec{x}_m = \vec{0}$ implies $a_1 = a_2 = \cdots = a_m = 0$. The *dimension* of a submodule X of $\mathbb{Z}^n_{p^s}$ is denoted by $\dim(X)$ and is defined to be the number of vectors in the largest linearly independent subset of X. Note that a linearly independent set in $\mathbb{Z}^n_{p^s}$ is equivalent to a unimodular set (see [10, 11]).

Next, if $X = \langle \vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m \rangle$ is the submodule of $\mathbb{Z}^n_{p^s}$ generated by a linearly independent set $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m\}$, then X is called an *m-subspace* or simply a *subspace* of $\mathbb{Z}^n_{p^s}$ and the set $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m\}$ is called a *basis* of X. It is worth noting that a subspace of $\mathbb{Z}^n_{p^s}$ is also known as a *free submodule*. Naturally, the subspace $\{\vec{0}\}$ is a trivial subspace with dimension 0 and an empty basis. Furthermore, $\mathbb{Z}^n_{p^s}$ possesses the *standard basis* $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ where for each $i = 1, 2, \ldots, n$, $\vec{e}_i = (e_{i1}, e_{i2}, \ldots, e_{in})$ with $e_{ii} = 1$ and $e_{ij} = 0$ for all $i \neq j$. Consequently, $\dim(\mathbb{Z}^n_{p^s}) = n$. In general, if X is an m-subspace of $\mathbb{Z}^n_{p^s}$, then $\dim(X) = m$.

It is well-known that a submodule may not necessarily be a subspace, even though it has a dimension. However, if X is a submodule of $\mathbb{Z}^n_{p^s}$ with $\dim(X) = m$, then X contains an m-subspace of $\mathbb{Z}^n_{p^s}$. It is important to note that every basis of a subspace of $\mathbb{Z}^n_{p^s}$ can be extended to a basis of $\mathbb{Z}^n_{p^s}$ (cf. [6]). Additionally, if X is a subspace of $\mathbb{Z}^n_{p^s}$ with dimension m and a basis $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m\}$, it can be shown that any element $\vec{x} \in X$ can be uniquely expressed as $\vec{x} = a_1\vec{x}_1 + a_2\vec{x}_2 + \cdots + a_m\vec{x}_m$, resulting in $|X| = p^{sm}$. Furthermore, for any subspaces X and Y of $\mathbb{Z}^n_{p^s}$, if $X \subseteq Y$, then $\dim(X) \leq \dim(Y)$ and if $\dim(X) = \dim(Y)$, then $X = Y$.

Next, let X and Y be subspaces of $\mathbb{Z}^n_{p^s}$. A *join* of X and Y is defined to be a subspace of $\mathbb{Z}^n_{p^s}$ containing both X and Y. A join of X and Y with the minimum dimension is called a *minimum join*. Denoted by $X \vee Y$, the set of minimum joins of X and Y. We write $\dim(X \vee Y)$ for the dimension of a minimum join of X and Y. Note that $X \cap Y$ may not be a subspace of $\mathbb{Z}^n_{p^s}$; however, $\dim(X \cap Y)$ always exists.

Lemma 2.1. *[11, Theorem 3.3] Let X and Y be subspaces of $\mathbb{Z}^n_{p^s}$. Then*

$$\dim(X \vee Y) = \dim(X) + \dim(Y) - \dim(X \cap Y).$$

Let m, n, q be non-negative integers with $q \geq 2$. The *Gaussian binomial coefficient* is given by

$$\begin{bmatrix} n \\ m \end{bmatrix}_q = \prod_{i=1}^{m} \frac{q^{n+1-i} - 1}{q^i - 1},$$

where $\begin{bmatrix} n \\ 0 \end{bmatrix}_q := 1$ and $\begin{bmatrix} 0 \\ m \end{bmatrix}_q := 0$ if $m > 0$. Note that $\begin{bmatrix} n \\ m \end{bmatrix}_q = \begin{bmatrix} n \\ n-m \end{bmatrix}_q$ and $\begin{bmatrix} n \\ m \end{bmatrix}_q = \frac{q^n-1}{q^m-1}\begin{bmatrix} n-1 \\ m-1 \end{bmatrix}_q$. In fact, $\begin{bmatrix} n \\ m \end{bmatrix}_q$ is the number of m-dimensional subspaces in an n-dimensional vector space over the finite field of q elements. The number of m-subspaces of $\mathbb{Z}_{p^s}^n$ is then studied in [11].

Lemma 2.2. *[11, Theorem 3.5] Let $1 \leq k \leq m \leq n$. Then:*

1. *The number of m-subspaces of $\mathbb{Z}_{p^s}^n$ is $p^{(s-1)m(n-m)}\begin{bmatrix} n \\ m \end{bmatrix}_p$.*
2. *In $\mathbb{Z}_{p^s}^n$, the number of k-subspaces in a given m-subspace is $p^{(s-1)k(m-k)}\begin{bmatrix} m \\ k \end{bmatrix}_p$.*
3. *In $\mathbb{Z}_{p^s}^n$, the number of m-subspaces containing a given k-subspace is $p^{(s-1)(m-k)(n-m)}\begin{bmatrix} n-k \\ m-k \end{bmatrix}_p$.*

Finally, we review some basic background from graph theory. A (*simple*) *graph* $G = (V, E)$ consists of a non-empty set $V := V(G)$ of vertices and an edge set $E := E(G)$ of unordered pairs of two distinct elements in V. The cardinalities of V and E are called the *order* and *size* of G, respectively. A graph with no edges is called an *edgeless graph*. If there is an edge $\{u, v\} \in E$, we say that u is *adjacent* to v denoted by $u \sim v$. A *subgraph* H of G is a graph in which $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. If a vertex set V of a graph G can be partitioned into k disjoint partite sets $V_1, V_2, \ldots, V_k$ such that no two vertices in the same partite set are adjacent, then G is called a *k-partite graph*. A *bipartite graph* is a 2-partite graph. The *degree* of a vertex u in a graph G is the number of vertices in G adjacent to u and is denoted by $\deg(u)$. If every vertex in a graph G has the same degree k, we say that G is *regular* and k is called the *valency* of G. A *complete graph* is a graph in which every two distinct vertices are adjacent. A complete graph of order n is denoted by K_n.

For any two vertices u and v of a graph G, a *u-v path* of length k is a sequence of $k+1$ distinct vertices $u = w_0, w_1, w_2, \ldots, w_{k-1}, w_k = v$ in G such that $w_i \sim w_{i+1}$ for all $i = 0, 1, \ldots, k-1$. A graph G is said to be *connected* if there is a u-v path in G for any two vertices u and v in G. The *distance* between two vertices u and v in a connected graph G, denoted by $\mathrm{d}_G(u, v)$ or $\mathrm{d}(u, v)$, is the length of the shortest u-v path in G. The *diameter* of a connected graph G, denoted by $\mathrm{diam}(G)$, is the largest distance between pairs of vertices in G. A *cycle* of length $k \geq 2$ is a sequence of vertices $u_0, u_1, \ldots, u_k, u_0$ where $u_0, u_1, \ldots, u_k$ form a u_0-u_k path and $u_k \sim u_0$. A cycle of length 3 is called a *triangle*. The *girth* of G, denoted by $\mathrm{g}(G)$, is the length of the shortest cycle if it exists; otherwise, it is defined as $\mathrm{g}(G) = \infty$.

3 Main Results

From now on, we assume that p is a prime number and s and n are positive integers such that $n \geq 2$, unless otherwise specified. We define the *subspace inclusion graph* of $\mathbb{Z}_{p^s}^n$, denoted by $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$, to be the graph whose vertices are the non-trivial proper subspaces of $\mathbb{Z}_{p^s}^n$ and any two distinct vertices X and Y are adjacent if and only if $X \subseteq Y$ or $Y \subseteq X$.

In this paper, we present some fundamental results concerning the subspace inclusion graphs. To begin, the order of the graph $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$ can be readily determined through Lemma 2.2 (1), as follows:

Theorem 3.1. *The subspace inclusion graph of $\mathbb{Z}_{p^s}^n$ is of order* $\displaystyle\sum_{m=1}^{n-1} p^{(s-1)m(n-m)}\begin{bmatrix} n \\ m \end{bmatrix}_p$.

Theorem 3.2. *Let X be an m-subspace of $\mathbb{Z}_{p^s}^n$ where $1 \leq m \leq n-1$. Then the degree of X in $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$ is given by*

$$\deg(X) = \sum_{k=1}^{m-1} p^{(s-1)k(m-k)}\begin{bmatrix} m \\ k \end{bmatrix}_p + \sum_{k=1}^{n-m-1} p^{(s-1)k(n-m-k)}\begin{bmatrix} n-m \\ k \end{bmatrix}_p.$$

Here, the empty sums (sums with no summands) are defined to be equal to 0.

Proof. Note that if Y is a vertex in $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$ and is adjacent to X, then $Y \subseteq X$ or $X \subseteq Y$. By Lemma 2.2, the number of subspaces Y contained in X with $1 \leq \dim(Y) \leq m-1$ is given by $\sum_{k=1}^{m-1} p^{(s-1)k(m-k)} \begin{bmatrix} m \\ k \end{bmatrix}_p$. On the other hand, the number of subspaces Y containing X with $m+1 \leq \dim(Y) \leq n-1$ is $\sum_{k=m+1}^{n-1} p^{(s-1)(k-m)(n-k)} \begin{bmatrix} n-m \\ k-m \end{bmatrix}_p = \sum_{k=1}^{n-m-1} p^{(s-1)k(n-m-k)} \begin{bmatrix} n-m \\ k \end{bmatrix}_p$. Hence, the degree of X in $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$ is $\sum_{k=1}^{m-1} p^{(s-1)k(m-k)} \begin{bmatrix} m \\ k \end{bmatrix}_p + \sum_{k=1}^{n-m-1} p^{(s-1)k(n-m-k)} \begin{bmatrix} n-m \\ k \end{bmatrix}_p$ as required. □

Corollary 3.3. *If X and Y are two subspaces of $\mathbb{Z}_{p^s}^n$ of dimension m and $n-m$, respectively, where $1 \leq m \leq n-1$, then $\deg(X) = \deg(Y)$.*

Proof. By Theorem 3.2, we obtain that

$$\deg(X) = \sum_{k=1}^{m-1} p^{(s-1)k(m-k)} \begin{bmatrix} m \\ k \end{bmatrix}_p + \sum_{k=1}^{n-m-1} p^{(s-1)k(n-m-k)} \begin{bmatrix} n-m \\ k \end{bmatrix}_p$$

and

$$\deg(Y) = \sum_{k=1}^{n-m-1} p^{(s-1)k(n-m-k)} \begin{bmatrix} n-m \\ k \end{bmatrix}_p + \sum_{k=1}^{m-1} p^{(s-1)k(m-k)} \begin{bmatrix} m \\ k \end{bmatrix}_p.$$

Therefore, $\deg(X) = \deg(Y)$. □

Next, we present special properties of $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$ when $n = 3$.

Proposition 3.4. *If $n = 3$, then $\mathcal{I}n(\mathbb{Z}_{p^s}^3)$ is of order $2p^{2(s-1)}(p^2+p+1)$, is regular with valency $p^{(s-1)}(p+1)$ and is of size $p^{3(s-1)}(p+1)(p^2+p+1)$.*

Proof. By Theorem 3.1, the order of $\mathcal{I}n(\mathbb{Z}_{p^s}^3)$ is

$$\sum_{m=1}^{2} p^{(s-1)m(3-m)} \begin{bmatrix} 3 \\ m \end{bmatrix}_p = p^{2(s-1)} \begin{bmatrix} 3 \\ 1 \end{bmatrix}_p + p^{2(s-1)} \begin{bmatrix} 3 \\ 2 \end{bmatrix}_p = 2p^{2(s-1)}(p^2+p+1).$$

Since all vertices of $\mathcal{I}n(\mathbb{Z}_{p^s}^3)$ are of dimensions 1 or 2, Corollary 3.3 implies that their degrees are equal. Hence, $\mathcal{I}n(\mathbb{Z}_{p^s}^3)$ is regular of degree $p^{(s-1)} \begin{bmatrix} 2 \\ 1 \end{bmatrix}_p = p^{(s-1)}(p+1)$. Therefore, the size of $\mathcal{I}n(\mathbb{Z}_{p^s}^3)$ is $p^{3(s-1)}(p+1)(p^2+p+1)$. □

Lemma 3.5. *If X and Y are two distinct vertices of $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$ of the same dimension, then $X \nsim Y$ in $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$.*

Proof. Suppose $X \sim Y$. Then $X \subseteq Y$ or $Y \subseteq X$. Since $\dim(X) = \dim(Y)$, it implies that $X = Y$, a contradiction. □

By applying the preceding lemma, we can establish numerous properties.

Corollary 3.6. *The subspace inclusion graph of $\mathbb{Z}_{p^s}^n$ is not complete.*

Proof. Since $n \geq 2$, we consider the standard basis vectors $\vec{e}_1$ and $\vec{e}_2$. It is clear that $\langle \vec{e}_1 \rangle$ and $\langle \vec{e}_2 \rangle$ are two distinct 1-subspaces of $\mathbb{Z}_{p^s}^n$. By Lemma 3.5, $\langle \vec{e}_1 \rangle \nsim \langle \vec{e}_2 \rangle$. It thus implies that $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$ is not complete. □

Corollary 3.7. *The subspace inclusion graph of $\mathbb{Z}_{p^s}^n$ is an $(n-1)$-partite graph.*

Proof. Let V_i be the set of i-subspaces of $\mathbb{Z}_{p^s}^n$ for $i = 1, 2, \ldots, n-1$. Then $V_1, V_2, \ldots$ and V_{n-1} partition the vertices of $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$. Moreover, Lemma 3.5 implies that no two vertices in V_i are adjacent. Thus, $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$ is an $(n-1)$-partite graph. □

Corollary 3.8. *The graph $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$ is an edgeless graph if and only if $n = 2$.*

Proof. Assume that $n = 2$. Note that the vertices of the graph $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$ are 1-subspaces of $\mathbb{Z}_{p^s}^2$. From Lemma 3.5, it follows that $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$ is a graph without edges.

On the other hand, let $n \geq 3$. Consider the subspaces $\langle \vec{e}_1 \rangle$ and $\langle \vec{e}_1, \vec{e}_2 \rangle$ of $\mathbb{Z}_{p^s}^n$ where $\vec{e}_1$ and $\vec{e}_2$ are standard basis vectors in $\mathbb{Z}_{p^s}^n$. Since $n \geq 3$, they are adjacent vertices in $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$. Thus, $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$ is not an edgeless graph. □

Next, we explore the diameter and girth of the subspace inclusion graph, along with related findings. To begin, we establish the graph's connectivity.

Lemma 3.9. *If $n \geq 3$, then $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$ is a connected graph and* $\mathrm{diam}(\mathcal{I}n(\mathbb{Z}_{p^s}^n)) \leq 3$.

Proof. Let X and Y be two vertices in $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$. If $X \sim Y$, then $\mathrm{d}(X, Y) = 1$. Assume $X \nsim Y$, i.e., $\mathrm{d}(X, Y) \neq 1$ and so $X \nsubseteq Y$ and $Y \nsubseteq X$. If $X \cap Y = X$, then $X \subseteq Y$, a contradiction. Thus, $X \cap Y$ is a proper submodule of X, so that $\dim(X \cap Y) < \dim(X)$. Similarly, $\dim(X \cap Y) < \dim(Y)$. We divide the cases by the dimensions of X and Y as follows:

Case 1: $\dim(X) = \dim(Y) = 1$. Then $\dim(X \cap Y) = 0$. By Lemma 2.1, $\dim(X \vee Y) = \dim(X) + \dim(Y) - \dim(X \cap Y) = 1 + 1 - 0 = 2$. Hence, there exists a subspace $Z \in X \vee Y$, i.e., Z is a subspace of $\mathbb{Z}_{p^s}^n$ containing X and Y with $\dim(Z) = \dim(X \vee Y) = 2$. Thus, Z is a vertex in $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$ such that $X \sim Z \sim Y$. Hence, $\mathrm{d}(X, Y) = 2$.

Case 2: $\dim(X) = 1$ and $\dim(Y) > 1$. Then there exists a 1-subspace Z of $\mathbb{Z}_{p^s}^n$ contained in Y, it implies that $Z \sim Y$. Note that $X \cap Z \subseteq X \cap Y$. Then $\dim(X \cap Z) \leq \dim(X \cap Y) < \dim(X) = 1$ in X. Thus, $\dim(X \cap Z) = 0$. Since $\dim(X \vee Z) = \dim(X) + \dim(Z) - \dim(X \cap Z) = 1 + 1 - 0 = 2$, there exists a subspace $W \in X \vee Z$, i.e., W is a subspace of $\mathbb{Z}_{p^s}^n$ containing X and Z with $\dim(W) = \dim(X \vee Z) = 2$. Thus, W is a vertex in $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$ such that $X \sim W \sim Z \sim Y$. Therefore, $\mathrm{d}(X, Y) \leq 3$.

Case 3: $\dim(X) > 1$ and $\dim(Y) = 1$. It is similar to Case 2.

Case 4: $\dim(X) > 1$ and $\dim(Y) > 1$.

Case 4.1: $\dim(X \vee Y) < n$. Then there exists $Z \in X \vee Y$, i.e., Z is a subspace of $\mathbb{Z}_{p^s}^n$ containing X and Y with $\dim(Z) = \dim(X \vee Y) < n$. Since $X \subseteq Z$, $\dim(Z) > 1$. Hence, Z is a vertex in $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$ such that $X \sim Z \sim Y$. Hence, $\mathrm{d}(X, Y) = 2$.

Case 4.2: $\dim(X \cap Y) \geq 1$. Then $X \cap Y$ contains a subspace Z of $\mathbb{Z}_{p^s}^n$ such that $\dim(Z) = \dim(X \cap Y) \geq 1$. Note that $\dim(Z) \leq \dim(X) < n$. Hence, Z is a vertex in $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$ such that $X \sim Z \sim Y$. Thus, $\mathrm{d}(X, Y) = 2$.

Case 4.3: $\dim(X \vee Y) = n$ and $\dim(X \cap Y) = 0$. Since $\dim(Y) > 1$, there exists a 1-subspace Z of $\mathbb{Z}_{p^s}^n$ contained in Y. Hence, $Z \sim Y$. Note that $X \cap Z \subseteq X \cap Y$. Then $\dim(X \cap Z) \leq \dim(X \cap Y) = 0$, i.e., $\dim(X \cap Z) = 0$. Thus, $\dim(X \vee Z) = \dim(X) + \dim(Z) - \dim(X \cap Z) = \dim(X) + \dim(Z) < \dim(X) + \dim(Y) = \dim(X) + \dim(Y) - \dim(X \cap Y) = \dim(X \vee Y) = n$. Hence, there exists $W \in X \vee Z$, i.e., W is a subspace of $\mathbb{Z}_{p^s}^n$ containing X and Z with $\dim(W) = \dim(X \vee Z) < n$. Thus, W is a vertex in $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$ such that $X \sim W \sim Z \sim Y$. Hence, $\mathrm{d}(X, Y) \leq 3$.

From all cases, it follows that the graph $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$ is complete and $\mathrm{d}(X, Y) \leq 3$. □

Theorem 3.10. *If $n \geq 3$, then* $\mathrm{diam}(\mathcal{I}n(\mathbb{Z}_{p^s}^n)) = 3$.

Proof. Assume $n \geq 3$. Let $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ be the standard basis of $\mathbb{Z}_{p^s}^n$. Then $X = \langle \vec{e}_1 \rangle$ and $Y = \langle \vec{e}_2, \vec{e}_3, \ldots, \vec{e}_n \rangle$ are subspaces of $\mathbb{Z}_{p^s}^n$ such that $\dim(X) = 1$ and $\dim(Y) = n - 1 \geq 2$. It is easy to see that $X \cap Y = \{\vec{0}\}$. Clearly, $Y \nsubseteq X$ and $X \nsubseteq Y$. Thus, $X \nsim Y$. Hence, $\mathrm{d}(X, Y) \neq 1$.

Assume $\mathrm{d}(X,Y) = 2$. Then there exists a vertex Z in $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$ such that $X \sim Z \sim Y$. Since $X \sim Z$, we have $X \subseteq Z$ or $Z \subseteq X$. Suppose $Z \subseteq X$. Then $Z \cap Y = \{\vec{0}\}$. As well, $Z \subseteq Y$ or $Y \subseteq Z$ since $Z \sim Y$. As a result, $Z = \{\vec{0}\}$ or $Y = \{\vec{0}\}$, a contradiction. Thus, $X \subseteq Z$. Since $\mathrm{d}(X,Y) = 2$, we also have $Y \subseteq Z$. Hence, Z is a join of X and Y. Note that $\dim(X \vee Y) = \dim(X) + \dim(Y) - \dim(X \cap Y) = 1 + (n-1) - 0 = n$. Then a minimum join of X and Y is $\mathbb{Z}_{p^s}^n$. Thus, $Z = \mathbb{Z}_{p^s}^n$ which is a contradiction. Therefore, $\mathrm{d}(X,Y) \neq 2$.

Lemma 3.9 shows that $\mathrm{diam}(\mathcal{I}n(\mathbb{Z}_{p^s}^n)) \leq 3$. Since $\mathrm{d}(X,Y) \neq 1$ and 2, it implies that $\mathrm{d}(X,Y) = 3$. Therefore, $\mathrm{diam}(\mathcal{I}n(\mathbb{Z}_{p^s}^n)) = 3$. □

Finally, we determine the girth of $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$ in the following theorem.

Theorem 3.11. *The girth of the subspace inclusion graph of $\mathbb{Z}_{p^s}^n$ is*

$$\mathrm{g}\left(\mathcal{I}n(\mathbb{Z}_{p^s}^n)\right) = \begin{cases} \infty & \text{if } n = 2, \\ 6 & \text{if } n = 3 \text{ and } s = 1, \\ 4 & \text{if } n = 3 \text{ and } s \geq 2, \\ 3 & \text{if } n \geq 4. \end{cases}$$

Proof. Note that the length of any cycle is at least 3. Hence, $\mathrm{g}(\mathcal{I}n(\mathbb{Z}_{p^s}^n) \geq 3$ or $\mathrm{g}(\mathcal{I}n(\mathbb{Z}_{p^s}^n)) = \infty$. *Case 1:* $n = 2$. $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$ is an edgeless graph by Corollary 3.8. Hence, $\mathrm{g}\left(\mathcal{I}n(\mathbb{Z}_{p^s}^n)\right) = \infty$.

Case 2: $n = 3$ and $s = 1$. By Theorem 4.2 in [3], we have $\mathrm{g}(\mathcal{I}n(\mathbb{Z}_p^n)) = 6$.

Case 3: $n = 3$ and $s \geq 2$. We show that $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$ has no cycle of length 3. Let $X \sim Y \sim Z \sim X$ be a cycle of length 3 in $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$. Then $\dim(X), \dim(Y)$ and $\dim(Z)$ are either 1 or 2 because $n = 3$. By Lemma 3.5, $\dim(X), \dim(Y)$ and $\dim(Z)$ must be different. This is impossible. Hence, $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$ does not contain any cycle of length 3. Thus, $\mathrm{g}(\mathcal{I}n(\mathbb{Z}_{p^s}^n)) \geq 4$.

Let $X_1 = \langle(1,0,0)\rangle, X_2 = \langle(1,0,0),(0,1,0)\rangle, X_3 = \langle(1,p^{s-1},0)\rangle$ and $X_4 = \langle(1,0,0),(0,1,p)\rangle$. It is easy to see that $\{(1,0,0)\}$, $\{(1,p^{s-1},0)\}$, $\{(1,0,0),(0,1,0)\}$ and $\{(1,0,0),(0,1,p)\}$ are linearly independent sets. Then X_1, X_2, X_3 and X_4 are non-trivial proper distinct subspaces of $\mathbb{Z}_{p^s}^n$. Clearly, $X_1 \subseteq X_2$ and $X_1 \subseteq X_4$. Since $(1,p^{s-1},0) = (1,0,0) + p^{s-1}(0,1,0)$, we have $X_3 \subseteq X_2$. As well, $(1,p^{s-1},0) = (1,0,0) + p^{s-1}(0,1,p)$ implies that $X_3 \subseteq X_4$. Therefore, $X_1 \sim X_2 \sim X_3 \sim X_4 \sim X_1$ is a cycle of length 4 in $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$. Thus, $\mathrm{g}(\mathcal{I}n(\mathbb{Z}_{p^s}^n)) = 4$.

Case 4: $n \geq 4$. Let $X_1 = \langle\vec{e}_1\rangle$, $X_2 = \langle\vec{e}_1,\vec{e}_2\rangle$ and $X_3 = \langle\vec{e}_1,\vec{e}_2,\vec{e}_3\rangle$ where $\vec{e}_1$, $\vec{e}_2$ and $\vec{e}_3$ are standard basis vectors. Then X_1, X_2 and X_3 are non-trivial proper subspaces of $\mathbb{Z}_{p^s}^n$ and $X_1 \subseteq X_2 \subseteq X_3$. Therefore, $X_1 \sim X_2 \sim X_3 \sim X_1$ is a cycle of length 3 in $\mathcal{I}n(\mathbb{Z}_{p^s}^n)$. Thus, $\mathrm{g}(\mathcal{I}n(\mathbb{Z}_{p^s}^n)) = 3$. □

We can see from the previous theorem that the girths of the graphs over the field $\mathbb{Z}_p$ and the ring $\mathbb{Z}_{p^s}$ (with $s \geq 2$) differ. These are some properties of the subspace inclusion graphs over $\mathbb{Z}_{p^s}$. Many more properties of these graphs could be explored in the future.

References

[1] A. Das, *Nonzero component graph of a finite dimensional vector space*, Commun. Algebra. **44**(9) (2016), 3918–3926.

[2] A. Das, *On subspace inclusion graph of a vector space*, Linear Multilinear Algebra. **66**(3) (2018), 554–564.

[3] A. Das, *Subspace inclusion graph of a vector space*, Commun. Algebra. **44**(11) (2016), 4724–4731.

[4] A. E. Brouwer, A. M. Cohen and A. Neumaier, *Distance-Regular Graphs*, Springer Verlag, New York, 1989.

[5] B. R. McDonald, *Finite Rings with Identity*, Marcel Dekker, New York, 1974.

[6] B. R. McDonald, *Geometric Algebra over Local Rings*, Marcel Dekker, Inc., New York, 1976.

[7] D. F. Anderson and P. S. Livingston, *The zero-divisor graph of a commutative ring*, J. Algebra. **217** (1999), 434–-447.

[8] D. Wong, X. Wang and C. Xia, *On two conjectures on the subspace inclusion 6 graph of a vector space*, J. Algebra Appl. **17**(10) (2018), 1850189.

[9] F. Li, K. Wang and J. Guo, *Symplectic graphs modulo pq*, Discrete Math. **313**(5) (2013), 650–655.

[10] L. P. Huang, B. Lv and K. Wang, *Automorphisms of Grassmann graphs over a residue class ring*, Discrete Math. **343**(4) (2020), 111693.

[11] L. P. Huang, B. Lv and K. Wang, *Erdős-Ko-Rado theorem, Grassmann graphs and p^s-Kneser graphs for vector spaces over a residue class ring*, J. Comb. Theory Ser. A. **164** (2019), 125–158.

[12] L. P. Huang, *Generalized bilinear forms graphs and MRD codes over a residue class ring*, Finite Fields Appl. **51** (2018), 306–324.

[13] L. P. Huang, H. Su, G. Tang and J. B. Wang, *Bilinear forms graphs over residue class rings*, Linear Algebra Appl. **523** (2017), 13–32.

[14] N. H. McCoy, *Rings and Ideals*, American Mathematical Soc., 1948.

[15] P. M. Cohn, *Free Rings and Their Relations*, 2nd ed., Academic Press, London, 1985.

[16] S. Akbari, M. Habibi, A. Majidinya and R, Manaviyat, *The inclusion ideal graph of rings*, Commun. Algebra. **43**(6) (2015), 2457–2465.

[17] S. Sirisuk and Y. Meemark, *Generalized symplectic graphs and generalized orthogonal graphs over finite commutative rings*, Linear Multilinear Algebra. **67**(12) (2019), 2427–2450.

[18] X. Ma and D. Wang, *Independence number of subspace inclusion graph and subspace sum graph of a vector space*, Linear Multilinear Algebra. **66**(12) (2018), 2430–2437.

[19] Y. Meemark and T. Prinyasart, *On symplectic graphs modulo p^n*, Discrete math. **311**(17) (2011), 1874–1878.

[20] Z. Tang and Z. X. Wan, *Symplectic graphs and their automorphisms*, Eur. J. Comb. **27**(1) (2006), 38–50.

[21] Z. X. Wan, *Lectures on Finite Fields and Galois Rings*, World Scientific Publishing Company, 2003.

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Functional Graphs of Non-Monic Linear Polynomials on Finite Field Extensions*

Suphawich Sengpanich† and Nithi Rungtanapirom‡

Department of Mathematics and Computer Science, Faculty of Science
Chulalongkorn University, Bangkok 10330, Thailand

Abstract

Functional graphs are introduced to study iteration behaviors of functions via their graph structures. Many papers consider monomial functions over commutative ring. In this work, we are interested in graph-theoretic properties of functional graphs of linear polynomials on finite field extensions; for example, the indegrees of vertices and the structure of components – the number of components and order of symmetry. Furthermore, the quotient digraph of functional graph by a suitable group is introduced to observe the similarity of merged vertices and components. The main ingredients of this work are the Möbius Inversion Formula and the Galois theory of finite field extensions.

Keywords: functional graphs, finite fields, group actions on graphs, arithmetic functions.

2020 MSC: Primary 37P25; Secondary 05C25, 05E18, 12E20, 11A25.

1 Introduction

This article deals with the functional graph. In the general context, let X be a nonempty set and f be a map on X. The functional graph is defined as follows:

Definition 1.1. The *functional graph* of f on X is a digraph $\Gamma(X, f)$ whose vertex set is X and directed edges are $(x, f(x))$ for all $x \in X$.

Note that the functional graph is used to study the behavior of iterations of a function because some properties of a function correspond to those of its functional graph.

The following propositions are well-known facts about the structure of *connected components*, maximal connected undirected subgraphs, of the functional graph. The proofs of these can be found in [8].

*This research was financially supported by Faculty of Science, Chulalongkorn University.
†Speaker. ‡Corresponding author.
Email: s.sengpanich@gmail.com (S. Sengpanich), Nithi.R@chula.ac.th (N. Rungtanapirom)

Proposition 1.2. *Every connected component of $\Gamma(X, f)$ contains one directed cycle and no infinite forward path, or contains no directed cycle and an infinite forward path, and any two such paths have a common point. In particular, if X is finite, then every component is a directed cycle attached by directed trees with single root on cycle.*

Proposition 1.3. *Assume that f is a bijective map on X. Then every connected component of $\Gamma(X, f)$ is exactly a directed cycle or an infinite two-way path. In particular, if X is finite, then every component is a directed cycle.*

To obtain the structure of the functional graph, the main related parameters are the indegree of each vertex, the number of connected components, the lengths of directed cycles, and the length of the longest directed path.

Further interesting properties of digraph are the (semi-)regularity and the symmetry.

Definition 1.4. A digraph is *regular* if every vertex has the same indegree and outdegree. Consequently, a functional graph is regular if every vertex has the same indegree.

Definition 1.5. A digraph is *symmetric of order* $M \geq 2$ if its set of components can be partitioned into subsets of M isomorphic components.

In Algebra, many researches are concerned with the case when X is a commutative ring and f is a polynomial function over X. Many graph-theoretic properties of $\Gamma(X, f)$ are investigated in the case that f is a monomial; for example, the formula for indegree and structure of connected components. Most of these are studied on the rings of integers modulo [7] and the quotient rings of polynomials over finite fields [6, 9]. In Section 2.2, we study graph-theoretic properties and structure of $\Gamma(\mathbb{F}, f)$ when $\mathbb{F}$ is a field and f is a non-monic linear polynomial over $\mathbb{F}$.

Furthermore, we are interested in actions of certain groups on functional graphs of which the result are not known to us so far. In what follows, let Γ be a digraph with vertex set $V(\Gamma)$ and directed edge set $E(\Gamma)$, and let G be a group.

Definition 1.6. A *group action on a digraph* Γ by G is a group action on $V(\Gamma)$ by G such that for each $\sigma \in G$, if $(v, w) \in E(\Gamma)$, then $(\sigma(v), \sigma(w)) \in E(\Gamma)$.

Definition 1.7. For a group action on the digraph Γ by G, a *quotient digraph* Γ/G is a digraph whose vertices are the orbits $Gv = \{\sigma(v) : \sigma \in G\}$ for $v \in V(\Gamma)$ and directed edges are (Gv, Gw) for $v, w \in V(\Gamma)$ such that there exist $v' \in Gv$ and $w' \in Gw$ with $(v', w') \in E(\Gamma)$.

Example 1.8. Consider the finite field extension $\mathbb{F}_3(\alpha)/\mathbb{F}_3$ where $\alpha^2 + 1 = 0$. Note that the Galois group $\mathrm{Gal}(\mathbb{F}_3(\alpha)/\mathbb{F}_3)$ consists of the identity map and the map $\sigma : \alpha \mapsto -\alpha$. So, the orbits of this group are $\{0\}, \{1\}, \{-1\}, \{\pm\alpha\}, \{\pm\alpha + 1\}, \{\pm\alpha - 1\}$.

Let $f(x) = -x + 1 \in \mathbb{F}_3[x]$. Then the functional graph $\Gamma(\mathbb{F}_3(\alpha), f)$ and its quotient digraph by $\mathrm{Gal}(\mathbb{F}_3(\alpha)/\mathbb{F}_3)$ are as follows.

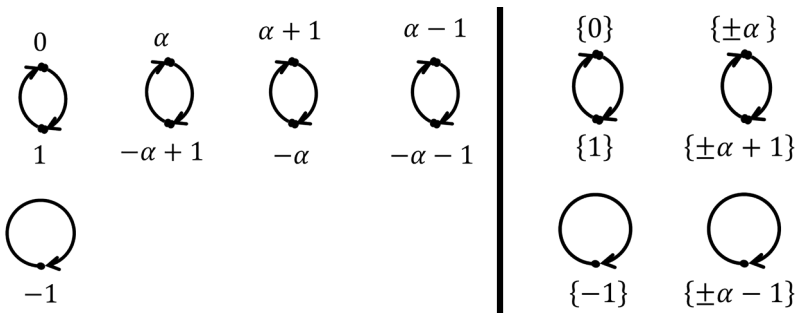


Figure 1: Functional graph (left) and its quotient digraph (right)

Based on this notion, we study the quotient digraph of $\Gamma(\mathbb{F}, f)$ by some subgroups of $\mathrm{Aut}(\mathbb{F})$ that preserve coefficients of f. For example, for a prime power q and positive integer d, the canonical action of the Galois group $\mathrm{Gal}(\mathbb{F}_{q^d}/\mathbb{F}_q)$ on $\mathbb{F}_{q^d}$ yields a group action on the functional graph $\Gamma(\mathbb{F}_{q^d}, f)$, where $\mathbb{F}_{q^d}/\mathbb{F}_q$ is a finite field extension and f is a polynomial over $\mathbb{F}_q$. In fact, each $\sigma \in \mathrm{Gal}(\mathbb{F}_{q^d}/\mathbb{F}_q)$ preserves every coefficient of f, so that σ gives rise to an automorphism of the functional graph $\Gamma(\mathbb{F}_{q^d}, f)$. In Section 3.1, we would like to figure out how connected components of the functional graphs $\Gamma(\mathbb{F}_{q^d}, f)$ collapse into the quotient digraph, for instance, when a directed cycle becomes a shorter cycle or when several directed cycles become a single cycle. By the Galois theory of finite field extensions, the structure of quotient digraphs is transformed to number-theoretic conditions that can be investigated via arithmetic-function properties recapitulated in Section 2.1. The number of components and the order of symmetry of quotient digraphs are discussed in Section 3.2. The final two sections give examples of quotient digraphs demonstrating the results in the previous section.

Notation For a set S, we denote by $\#S$ the cardinality of S and if $S \subseteq \mathbb{N}$, we denote by $\gcd S$ the greatest common divisor of elements in S.

2 Preliminaries

2.1 Arithmetic Functions

In this section, we summarize some well-known definitions and properties in Number Theory [2, 4, 5] that are used in this article. We begin with the following basic definitions.

- An *arithmetic function* is a function from $\mathbb{N}$ to $\mathbb{C}$.
- A *multiplicative function* is an arithmetic function A such that $A(M_1M_2) = A(M_1)A(M_2)$ for all relatively prime $M_1, M_2 \in \mathbb{N}$. Note that, for each $M \in \mathbb{N}$ with prime factorization $M = p_1^{\alpha_1} \dots p_r^{\alpha_r}$, we have
$$A(M) = A(p_1^{\alpha_1}) \dots A(p_r^{\alpha_r}).$$
- The Möbius function is an arithmetic function μ defined by, for $M \in \mathbb{N}$,
$$\mu(M) = \begin{cases} 1 & \text{if } M = 1, \\ (-1)^r & \text{if } M \text{ is square-free}, \\ 0 & \text{otherwise.} \end{cases}$$
- The Euler's phi function is an arithmetic function φ defined by, for $M \in \mathbb{N}$,
$$\varphi(M) = \#\{1 \le x \le M : \gcd(x, M) = 1\}.$$
It is known that, for each $M \in \mathbb{N}$,
$$\varphi(M) = M \cdot \prod_{p|M} \left(1 - \frac{1}{p}\right).$$
- For a prime number p, the p-adic valuation is an arithmetic function v_p defined by, for $M \in \mathbb{N}$, $v_p(M)$ is the highest exponent $\alpha \ge 0$ such that $p^\alpha \mid M$.
- The radical $\mathrm{rad}(M)$ of $M \in \mathbb{N}$ is the product of the distinct prime divisors of M.

Proposition 2.1. *Let $f, F : \mathbb{N} \to \mathbb{C}$ be such that, for $M \in \mathbb{N}$,*

$$F(M) = \sum_{x|M} f(x).$$

If f is multiplicative, then so is F.

Note that both μ and φ are multiplicative. Furthermore, μ satisfies the following property.

Proposition 2.2. *For $M \in \mathbb{N}$,*

$$\sum_{x|M} \mu(x) = \sum_{x|M} \mu\left(\frac{M}{x}\right) = \begin{cases} 1 & \text{if } M = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 2.3 (Möbius Inversion Formula for one variable)**.** *Let $f, F : \mathbb{N} \to \mathbb{C}$. The following statements are equivalent:*

(1) For each $M \in \mathbb{N}$,

$$\sum_{x|M} f(x) = F(M).$$

(2) For each $M \in \mathbb{N}$,

$$\sum_{x|M} \mu\left(\frac{M}{x}\right) F(x) = f(M).$$

Proposition 2.4 (Möbius Inversion Formula for two variables)**.** *Let $f, F : \mathbb{N} \times \mathbb{N} \to \mathbb{C}$. The following statements are equivalent:*

(1) For each $M, N \in \mathbb{N}$,

$$\sum_{x|M} \sum_{y|N} f(x, y) = F(M, N).$$

(2) For each $M, N \in \mathbb{N}$,

$$\sum_{x|M} \sum_{y|N} \mu\left(\frac{M}{x}\right) \mu\left(\frac{N}{y}\right) F(x, y) = f(M, N).$$

Lemma 2.5. *Let $n \in \mathbb{N}$. Define the divisor-checking function $f(*; n)$ by, for $M \in \mathbb{N}$, $f(M; n) = 1$ if $M \mid n$, and $f(M; n) = 0$ otherwise.*

Then $f(*; n)$ is multiplicative and, for each $M \in \mathbb{N}$,

$$\gcd(M, n) = \sum_{x|M} \varphi(x) f(x; n). \tag{2.1}$$

Proof. First, to show that $f(*; n)$ is multiplicative, let $M_1, M_2 \in \mathbb{N}$ be relatively prime. Observe that $M_1 \mid n$ and $M_2 \mid n$ if and only if $M_1 M_2 \mid n$. If $M_1 M_2 \mid n$, then $f(M_1 M_2; n) = 1 = f(M_1; n) f(M_2; n)$. Otherwise, $M_1 \nmid n$ or $M_2 \nmid n$, so $f(M_1 M_2; n) = 0 = f(M_1; n) f(M_2; n)$.

Next, to show that $\gcd(*, n)$ is multiplicative, let $M_1, M_2 \in \mathbb{N}$ be relatively prime. Observe that every divisor of $M_1 M_2$ can be written as a product of divisors of M_1 and M_2, respectively. This implies that $\gcd(M_1 M_2, n) \mid \gcd(M_1, n) \cdot \gcd(M_2, n)$. On the other hand, it is clear that $\gcd(M_1, n) \cdot \gcd(M_2, n) \mid M_1 M_2$. Furthermore, since M_1 and M_2 are relatively prime, so are $\gcd(M_1, n)$ and $\gcd(M_2, n)$. It follows that $\gcd(M_1, n) \cdot \gcd(M_2, n) \mid n$, thus $\gcd(M_1, n) \cdot \gcd(M_2, n) \mid \gcd(M_1 M_2, n)$. Therefore, $\gcd(M_1 M_2, n) = \gcd(M_1, n) \cdot \gcd(M_2, n)$.

Finally, by Proposition 2.1 and the multiplicity of φ and $f(*; n)$, the function given by right-hand side of (2.1) is multiplicative. Since $\gcd(*, n)$ is also multiplicative, it suffices to show that (2.1) holds for prime powers. Let p be a prime number and $\alpha \in \mathbb{N}$. We distinguish three cases.

1. If $p \nmid n$, then $f(p^i; n) = 0$ for all $i \geq 1$, so
$$\gcd(p^\alpha, n) = 1 = 1 + 0 = \varphi(1)f(1; n) + \sum_{i=1}^{\alpha} \varphi(p^i)f(p^i; n).$$

2. If $p \mid n$ and $\alpha \leq v_p(n)$, then $f(p^i; n) = 1$ for all $0 \leq i \leq \alpha$, so
$$\gcd(p^\alpha, n) = p^\alpha = 1 + \sum_{i=1}^{\alpha}(p^i - p^{i-1}) = \varphi(1)f(1; n) + \sum_{i=1}^{\alpha} \varphi(p^i)f(p^i; n).$$

3. If $p \mid n$ and $\alpha > v_p(n)$, then $f(p^i; n) = 1$ for all $0 \leq i \leq v_p(n)$ and $f(p^i; n) = 0$ for all $v_p(n) + 1 \leq i \leq \alpha$, so
$$\begin{aligned}\gcd(p^\alpha, n) = p^{v_p(n)} &= 1 + \sum_{i=1}^{v_p(n)}(p^i - p^{i-1}) + 0 \\ &= \varphi(1)f(1; n) + \sum_{i=1}^{v_p(n)} \varphi(p^i)f(p^i; n) + \sum_{i=v_p(n)+1}^{\alpha} \varphi(p^i)f(p^i; n).\end{aligned}$$ □

Möbius Inversion Formula for one variable (Proposition 2.3) implies the following lemma.

Lemma 2.6. *Let $n \in \mathbb{N}$. Then, for each $M \in \mathbb{N}$,*
$$\sum_{x|M} \mu\left(\frac{M}{x}\right) \gcd(x, n) = \begin{cases} \varphi(M) & \text{if } M \mid n, \\ 0 & \text{otherwise.} \end{cases}$$

2.2 Functional Graphs of Linear Polynomials on Fields

In this section, an *isomorphism of digraphs* is introduced to discuss the structures of $\Gamma(\mathbb{F}, ax+b)$ for some specific $a, b \in \mathbb{F}$. For a digraph Γ, denote vertex set and directed edge set of Γ by $V(\Gamma)$ and $E(\Gamma)$, respectively.

Definition 2.7. Let Γ_1 and Γ_2 be digraphs. A map $\Phi : V(\Gamma_1) \to V(\Gamma_2)$ is called an *isomorphism of digraphs* from Γ_1 to Γ_2 if Φ is bijective and $(v, w) \in E(\Gamma_1)$ if and only if $(\Phi(v), \Phi(w)) \in E(\Gamma_2)$.

If there is an isomorphism from Γ_1 onto Γ_2, we say that Γ_1 is *isomorphic* to Γ_2 and denote this $\Gamma_1 \cong \Gamma_2$.

Theorem 2.8. *Let $a, b_1, b_2 \in \mathbb{F}$. If $a \neq 1$, then $\Gamma(\mathbb{F}, ax + b_1) \cong \Gamma(\mathbb{F}, ax + b_2)$.*

Moreover, for each $b \in \mathbb{F}^\times$, $\Gamma(\mathbb{F}, ax + b) \cong \Gamma(\mathbb{F}, ax)$ if and only if $a \neq 1$.

Proof. Assume that $a \neq 1$. Write $c = (1 - a)^{-1} \cdot (b_2 - b_1)$. Define $\Phi : \mathbb{F} \to \mathbb{F}$ by $\Phi(x) = x + c$ for $x \in \mathbb{F}$. Clearly, Φ is bijective. Note that $\Phi(ax + b_1) = (ax + b_1) + c = a(x + c) + b_2 = a\Phi(x) + b_2$ for all $x \in \mathbb{F}$. Therefore, $\Gamma(\mathbb{F}, ax + b_1) \cong \Gamma(\mathbb{F}, ax + b_2)$.

Let $b \in \mathbb{F}^\times$ be such that $\Gamma(\mathbb{F}, ax) \cong \Gamma(\mathbb{F}, ax + b)$ with an isomorphism Φ. Then $\Phi(ax) = a\Phi(x)+b$ for all $x \in \mathbb{F}$. Therefore, $\Phi(0) = a\Phi(0)+b$, so $a \neq 1$; otherwise, $b = (1-a)\Phi(0) = 0$. □

Next, the structure of $\Gamma(\mathbb{F}, ax)$ for $a \in \mathbb{F}^\times$ is as follows:

Theorem 2.9. *Let $a \in \mathbb{F}^\times$. The components of $\Gamma(\mathbb{F}, ax)$ are $[x]_a = \{a^t x : t \in \mathbb{Z}\}$ for $x \in \mathbb{F}$. The **trivial** component of $\Gamma(\mathbb{F}, ax)$ is the component $[0]_a = \{0\}$. In particular, if $n = \mathrm{ord}(a, \mathbb{F}^\times)$ is finite, every **nontrivial** component of $\Gamma(\mathbb{F}, ax)$ is a directed cycle of length n. Otherwise, it is an infinite two-way path.*

Therefore, for each $a_1, a_2 \in \mathbb{F}^\times$, $\Gamma(\mathbb{F}, a_1x) \cong \Gamma(\mathbb{F}, a_2x)$ if and only if $\mathrm{ord}(a_1, \mathbb{F}^\times) = \mathrm{ord}(a_2, \mathbb{F}^\times)$.

Furthermore, if $\mathbb{F}$ is finite, then the number of nontrivial components is $\#\mathbb{F}^\times/n$.

To conclude, for each $a \in \mathbb{F}^\times$ and $b \in \mathbb{F}$, we have $\Gamma(\mathbb{F}, ax+b)$ is regular with indegree 1 and if $a \neq 1$ and $[1]_a \neq \mathbb{F}^\times$, we have $\Gamma(\mathbb{F}, ax+b) - [0]_a$ is symmetric.

To end this section, we state a sufficient condition related to a group action on a digraph and a homomorphism of quotient digraphs.

Definition 2.10. Let X and Y be sets with group actions by a group G. A function ψ from X to Y is said to be *G-equivariant* if $\sigma \circ \psi = \psi \circ \sigma$ for every $\sigma \in G$.

By definitions of a group action of digraph and quotient digraph, we obtain the following propositions.

Proposition 2.11. *Assume that a function f on X is G-equivariant. Then a group action on X by G induces a group action on $\Gamma(X, f)$ by G.*

Proposition 2.12. *Let Γ_1 and Γ_2 be digraphs with group actions on Γ_1 and Γ_2 by G. If $\Gamma_1 \cong \Gamma_2$ with a G-equivariant isomorphism Φ, then $\Gamma_1/G \cong \Gamma_2/G$ with an isomorphism $\mathbf{\Phi} : Gx \mapsto G\Phi(x)$.*

3 Main Results

3.1 On Finite Field Extension

In this section, let q be a prime power and d be a positive integer. Consider a finite field extension $\mathbb{F}_{q^d}/\mathbb{F}_q$ with a cyclic group $\mathbb{F}_{q^d}^\times$ of order $q^d - 1$.

Let $a \in \mathbb{F}_q \setminus \{0, 1\}$. Let $n = \mathrm{ord}(a, \mathbb{F}_q^\times)$ be the multiplicative order of a. Note that $n \mid q-1$.

It is known [3] that the Galois group $G := \mathrm{Gal}(\mathbb{F}_{q^d}/\mathbb{F}_q)$ is a cyclic group of order d generated by the Frobenius automorphism $\mathrm{Frob}_q : x \mapsto x^q$. For $x \in \mathbb{F}_{q^d}^\times$, denote the orbit of x in G by $Gx = \left\{\mathrm{Frob}_q^e(x) = x^{q^e} : 0 \leq e < d\right\}$.

For every $a, b \in \mathbb{F}_q$ with $a \neq 0, 1$, Theorem 2.8 implies that $\Gamma(\mathbb{F}_{q^d}, ax+b) \cong \Gamma(\mathbb{F}_{q^d}, ax)$ with an isomorphism $x \mapsto x - (1-a)^{-1}b$. Furthermore, G acts on both functional graphs canonically by Proposition 2.11 and $\Gamma(\mathbb{F}_{q^d}, ax+b)/G \cong \Gamma(\mathbb{F}_{q^d}, ax)/G$ by Proposition 2.12.

Therefore, it suffices to study the structure of the quotient digraph Γ/G of a functional graph $\Gamma := \Gamma(\mathbb{F}_{q^d}, ax)$ where $a \in \mathbb{F}_q \setminus \{0, 1\}$. We already know the structure of Γ by Theorem 2.9.

As the trivial component $[0] = \{0\}$ in Γ must be collapsed to the **trivial** component $[G0] = \{0\}$ in Γ/G, we consider **nontrivial** components $[Gx]$ in Γ/G for $x \in \mathbb{F}_{q^d}^\times$.

For $x \in \mathbb{F}_{q^d}^\times$, we define the following parameters

- D is the least positive integer such that $x^{q^D} = x$, i.e. $D = \#Gx$;
- k is the least positive integer such that $x^{q^k} \in [x]$ where $[x]$ is the component of x in Γ.

Note that D is the order of q modulo $\mathrm{ord}(x, \mathbb{F}_{q^d}^\times)$ and k is the order of q modulo $\mathrm{ord}(x^n, \mathbb{F}_{q^d}^\times)$, so $k \mid D$ and $D \mid nk$.

Lemma 3.1. *Let $x, x' \in \mathbb{F}_{q^d}^\times$ be in the same component of Γ. Let D', k' be the corresponding parameters of x'. Then $D' = D$ and $k' = k$.*

Proof. Write $x' = a^t x$ for some $t \in \mathbb{Z}$. As $a \in \mathbb{F}_q$, we have $a^q = a$, so $(x')^{q^e} = (a^t x)^{q^e} = (a^{q^e})^t x^{q^e} = a^t x^{q^e}$ for all $0 \leq e < d$. Hence parameters of x and x' coincide by minimality. □

Lemma 3.2. *Let $x, x' \in \mathbb{F}_{q^d}^\times$ be in the same orbit over G. Let D', k' be the corresponding parameters of x'. Then $D' = D$ and $k' = k$.*

Proof. Clearly, $D' = D$. Write $x' = x^{q^e}$ for some $0 \leq e < d$. As $a \in \mathbb{F}_q$, we have $a^q = a$. Note that if $x^{q^k} = a^t x$ for some $t \in \mathbb{Z}$, then $(x')^{q^k} = (x^{q^k})^{q^e} = (a^{q^e})^t x^{q^e} = a^t x'$, so $k' \leq k$. Similarly, $k \leq k'$. □

Observe that D is the number of vertices in Γ that are collapsed into a vertex Gx in Γ/G and k is the number of components, namely $[x], [x^q], [x^{q^2}], \ldots, [x^{q^{k-1}}]$, in Γ that are collapsed into the component $[Gx]$ in Γ/G. Since the component $[x]$ in Γ is a directed cycle of length n, it follows that the component $[Gx]$ in Γ/G is a directed cycle of length $\ell := nk/D$ because there are $\ell D = nk$ vertices in Γ collapsed to $[Gx]$.

Therefore, to count the number of components in Γ/G with given length ℓ, it suffices by Lemmas 3.1, 3.2, and the above observation to count the number of vertices in Γ with parameters D, k such that $nk = \ell D$.

For $D, k \in \mathbb{N}$, let $\gamma(D, k; n)$ be the number of elements in $\mathbb{F}_{q^d}^\times$ with parameters D, k with respect to Γ if $D \mid d$, $k \mid D$, and $D \mid nk$; otherwise, we set $\gamma(D, k; n) = 0$. Then the number of components in Γ/G with parameters D, k is $\dfrac{1}{nk}\gamma(D, k; n)$.

Lemma 3.3. *Let Z be a cyclic group and let $\delta, \kappa, n \in \mathbb{N}$ be such that $\delta \mid \#Z$ and $\kappa \mid \#Z$. Then*

$$\#\{x \in Z : \operatorname{ord}(x, Z) \mid \delta \text{ and } \operatorname{ord}(x^n, Z) \mid \kappa\} = \gcd(\delta, n\kappa).$$

Proof. Note that, since Z is a cyclic group, we have for each $x \in Z$,

$$\operatorname{ord}(x^n, Z) \mid \kappa \quad \text{if and only if} \quad \operatorname{ord}(x, Z) \mid n\kappa.$$

As $\gcd(\delta, n\kappa) \mid \#Z$, it follows that

$$\begin{aligned}\#\{x \in Z : \operatorname{ord}(x, Z) \mid \delta \text{ and } \operatorname{ord}(x^n, Z) \mid \kappa\} &= \#\{x \in Z : \operatorname{ord}(x, Z) \mid \delta \text{ and } \operatorname{ord}(x^n, Z) \mid n\kappa\} \\ &= \#\{x \in Z : \operatorname{ord}(x, Z) \mid \gcd(\delta, n\kappa)\} \\ &= \gcd(\delta, n\kappa).\end{aligned}$$

□

To calculate $\gamma(D, k; n)$, we use the relation between parameters and orders of q in specific modulos as shown in the following theorem.

Theorem 3.4. *Let $D, k \in \mathbb{N}$. Then*

$$\gamma(D, k; n) = \sum_{h \mid D} \sum_{e \mid k} \mu\left(\frac{D}{h}\right) \mu\left(\frac{k}{e}\right) \left(q^{\gcd(h,e)} - 1\right) \gcd\left(\frac{h}{\gcd(h, e)}, n\right).$$

Proof. Note that, for each $D, k \in \mathbb{N}$ and $x \in \mathbb{F}_{q^d}^\times$, if $\operatorname{ord}(x, \mathbb{F}_{q^d}^\times) \mid q^D - 1$ and $\operatorname{ord}(x^n, \mathbb{F}_{q^d}^\times) \mid q^k - 1$, then $\operatorname{ord}(x, \mathbb{F}_{q^d}^\times) \mid q^h - 1$ and $\operatorname{ord}(x^n, \mathbb{F}_{q^d}^\times) \mid q^e - 1$ for the least $h \mid D$ and $e \mid k$, so h is the order of q modulo $\operatorname{ord}(x, \mathbb{F}_{q^d}^\times)$ and k is the order of q modulo $\operatorname{ord}(x^n, \mathbb{F}_{q^d}^\times)$.

By Lemma 3.3 applied to $\mathbb{F}_{q^d}^\times$ and definition of $\gamma(*, *; n)$, we have, for each $D, k \in \mathbb{N}$,

$$\sum_{h \mid D} \sum_{e \mid k} \gamma(h, e; n) = \gcd\left(q^D - 1, n(q^k - 1)\right).$$

As $n \mid q - 1$, it follows from an elementary calculation that for each $D, k \in \mathbb{N}$,

$$\sum_{h \mid D} \sum_{e \mid k} \gamma(h, e; n) = \left(q^{\gcd(D,k)} - 1\right) \gcd\left(\frac{D}{\gcd(D, k)}, n\right).$$

Hence, the result follows from Möbius Inversion Formula for two variables (Proposition 2.4). □

Now, the number of components in Γ/G with given length can be calculated. More precisely, if $\ell \in \mathbb{N}$ with $\ell \mid n$ and $n \mid \ell d$, let $\mathcal{C}(\ell; n)$ be the number of nontrivial components in Γ/G with length ℓ. We have the following formula

$$\mathcal{C}(\ell; n) = \sum_{k \mid \frac{\ell d}{n}} \frac{1}{nk} \gamma\left(\frac{nk}{\ell}, k; n\right).$$

Since we know the number of components with given length, we can calculate the order of symmetry of $\Gamma/G - [G0]$ based on the fact that the set of nontrivial components is partitioned into subsets of $M_\Gamma := \gcd\{\mathcal{C}(\ell; n) : \ell \mid n, n \mid \ell d\}$ isomorphic component.

3.2 General Results

In this section, we simplify the formula of $\gamma(D, k; n)$ in Theorem 3.4.

Observe first that, for $D, k \in \mathbb{N}$ with $D \mid d$, $k \mid D$, and $D \mid nk$,

$$\begin{aligned}\gamma(D,k;n) &= \sum_{h|D}\sum_{e|k}\mu\left(\frac{D}{h}\right)\mu\left(\frac{k}{e}\right)\left(q^{\gcd(h,e)}-1\right)\gcd\left(\frac{h}{\gcd(h,e)},n\right)\\ &= \sum_{g|\gcd(D,k)}(q^g-1)\left[\sum_{\substack{h|D,\ e|k\\ \gcd(h,e)=g}}\mu\left(\frac{D}{h}\right)\mu\left(\frac{k}{e}\right)\gcd\left(\frac{h}{g},n\right)\right]\\ &= \sum_{g|k}(q^g-1)\left[\sum_{\substack{h|\frac{D}{g},\ e|\frac{k}{g}\\ \gcd(h,e)=1}}\mu\left(\frac{D/g}{h}\right)\mu\left(\frac{k/g}{e}\right)\gcd(h,n)\right].\end{aligned}$$

For a further simplification, we define functions $F(*,*;n), \widehat{F}(*,*) : \mathbb{N}\times\mathbb{N}\to\mathbb{C}$ by

$$F(M,N;n) = \sum_{\substack{x|M,\ y|N\\ \gcd(x,y)=1}}\mu\left(\frac{M}{x}\right)\mu\left(\frac{N}{y}\right)\gcd(x,n)$$

and

$$\widehat{F}(M,N) = \sum_{\substack{x|M,\ y|N\\ \gcd(x,y)=1}}\mu\left(\frac{M}{x}\right)\mu\left(\frac{N}{y}\right).$$

for each $M, N \in \mathbb{N}$. It follows that

$$\gamma(D,k;n) = \sum_{g|k}(q^g-1)\cdot F(D/g, k/g; n).$$

Furthermore, if $\gcd(n,d) = 1$, then

$$\gamma(D,k;n) = \sum_{g|k}(q^g-1)\cdot \widehat{F}(D/g, k/g).$$

Now, we apply Möbius Inversion Formulas on some suitable functions to calculate $\widehat{F}$, F. To this end, we define functions $H, \Delta : \mathbb{N}\times\mathbb{N}\to\mathbb{C}$ by

$$H(M,N) = \begin{cases}1 & \text{if } \gcd(M,N)=1,\\ 0 & \text{otherwise,}\end{cases}$$

and

$$\Delta(M,N) = \begin{cases}1 & \text{if } M=N,\\ 0 & \text{otherwise}\end{cases}$$

for each $M, N \in \mathbb{N}$. By Proposition 2.2,

$$H(M,N) = \sum_{z|\gcd(M,N)}\mu(z) = \sum_{\substack{x|M,\ y|N\\ x=y}}\mu(y) = \sum_{x|M}\sum_{y|N}\mu(y)\Delta(x,y).$$

By Möbius Inversion Formula for two variables (Proposition 2.4),

$$\widehat{F}(M,N) = \sum_{\substack{x|M,\, y|N \\ \gcd(x,y)=1}} \mu\left(\frac{M}{x}\right)\mu\left(\frac{N}{y}\right) = \sum_{x|M}\sum_{y|N} \mu\left(\frac{M}{x}\right)\mu\left(\frac{N}{y}\right) H(x,y)$$
$$= \mu(N)\Delta(M,N) = \begin{cases} \mu(N) & \text{if } M = N, \\ 0 & \text{otherwise.} \end{cases}$$

For $M, N \in \mathbb{N}$, define

$$\widetilde{H}(M,N) = \sum_{y|N} \mu\left(\frac{N}{y}\right) H(M,y) = \sum_{\substack{y|N \\ \gcd(M,y)=1}} \mu\left(\frac{N}{y}\right).$$

Then

$$\widehat{F}(M,N) = \sum_{x|M} \mu\left(\frac{M}{x}\right)\left[\sum_{\substack{y|N \\ \gcd(x,y)=1}} \mu\left(\frac{N}{y}\right)\right] = \sum_{x|M} \mu\left(\frac{M}{x}\right)\widetilde{H}(x,N).$$

By Möbius Inversion Formula for one variable (Proposition 2.3),

$$\widetilde{H}(M,N) = \sum_{x|M} \widehat{F}(x,N) = \sum_{\substack{x|M \\ x=N}} \mu(N) = \begin{cases} \mu(N) & \text{if } N \mid M, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore,

$$F(M,N;n) = \sum_{\substack{x|M,\, y|N \\ \gcd(x,y)=1}} \mu\left(\frac{M}{x}\right)\mu\left(\frac{N}{y}\right)\gcd(x,n)$$
$$= \sum_{x|M} \mu\left(\frac{M}{x}\right)\gcd(x,n)\left[\sum_{\substack{y|N \\ \gcd(x,y)=1}} \mu\left(\frac{N}{y}\right)\right]$$
$$= \sum_{x|M} \mu\left(\frac{M}{x}\right)\gcd(x,n)\widetilde{H}(x,N)$$
$$= \sum_{\substack{x|M \\ N|x}} \mu\left(\frac{M}{x}\right)\gcd(x,n)\mu(N).$$

It follows that $F(M,N;n) = 0$ if $N \nmid M$, and if $N \mid M$, then

$$F(M,N;n) = \mu(N)\sum_{x|\frac{M}{N}} \mu\left(\frac{M/N}{x}\right)\gcd(Nx,n)$$
$$= \mu(N)\sum_{x|\frac{M}{N}} \mu\left(\frac{M/N}{x}\right)\gcd(\gcd(N,n)x,n)$$
$$= \mu(N)\gcd(N,n)\sum_{x|\frac{M}{N}} \mu\left(\frac{M/N}{x}\right)\gcd\left(x,\frac{n}{\gcd(N,n)}\right).$$

By Lemma 2.6, we have

$$F(M,N;n)=\mu(N)\gcd(N,n)\varphi\left(\frac{M}{N}\right)$$

if $N \mid M$ and $\gcd(N,n)\left|\frac{nN}{M}\right.$; otherwise, $F(M,N;n)=0$.

Hence, for each $D \mid d$, $k \mid D$ with $D \mid nk$,

$$\begin{aligned}\gamma(D,k;n)&=\sum_{g|k}(q^g-1)\cdot F\left(\frac{D}{g},\frac{k}{g};n\right)\\&=\varphi\left(\frac{D}{k}\right)\sum_{\substack{g|k\\ \gcd\left(\frac{k}{g},n\right)\mid\frac{nk}{D}}}\mu\left(\frac{k}{g}\right)\gcd\left(\frac{k}{g},n\right)\cdot(q^g-1)\\&=\varphi\left(\frac{D}{k}\right)\sum_{\substack{g|k\\ \gcd(g,n)\mid\frac{nk}{D}}}\mu(g)\gcd(g,n)\cdot\left(q^{k/g}-1\right).\end{aligned}$$

So, for each $\ell \mid n$ with $n \mid \ell d$, we have

$$\begin{aligned}\mathcal{C}(\ell;n)&=\sum_{k|\frac{\ell d}{n}}\frac{1}{nk}\gamma\left(\frac{nk}{\ell},k;n\right)\\&=\varphi\left(\frac{n}{\ell}\right)\sum_{k|\frac{\ell d}{n}}\sum_{\substack{g|k\\ \gcd(g,n)|\ell}}\mu(g)\frac{\gcd(g,n)}{g}\cdot\frac{q^{k/g}-1}{n\cdot k/g}\\&=\varphi\left(\frac{n}{\ell}\right)\sum_{\substack{g|\frac{\ell d}{n}\\ \gcd(g,n)|\ell}}\sum_{k|\frac{\ell d}{n}\atop g|k}\mu(g)\frac{\gcd(g,n)}{g}\cdot\frac{q^{k/g}-1}{n\cdot k/g}\\&=\varphi\left(\frac{n}{\ell}\right)\sum_{\substack{g|\frac{\ell d}{n}\\ \gcd(g,n)|\ell}}\sum_{m|\frac{\ell d}{ng}}\mu(g)\frac{\gcd(g,n)}{g}\cdot\frac{q^m-1}{n\cdot m}\\&=\varphi\left(\frac{n}{\ell}\right)\sum_{m|\frac{\ell d}{n}}\left[\sum_{\substack{g|\frac{\ell d}{nm}\\ \gcd(g,n)|\ell}}\mu(g)\frac{\gcd(g,n)}{g}\right]\cdot\frac{q^m-1}{n\cdot m}.\end{aligned}$$

For $M \in \mathbb{N}$, define

$$A(M)=\sum_{\substack{x|M\\ \gcd(x,n)|\ell}}\mu(x)\frac{\gcd(x,n)}{x}=\sum_{x|M}\mu(x)\frac{\gcd(x,n)}{x}\cdot f(\gcd(x,n);\ell),$$

where $f(*;\ell)$ is the divisor-checking function as in Lemma 2.5. Since $f(*;\ell)$ and $\gcd(*,n)$ are multiplicative and $\gcd(\gcd(M_1,n),\gcd(M_2,n))=1$ for all M_1, M_2 with $\gcd(M_1,M_2)=1$, it follows that $f(\gcd(*,n);\ell)$ is multiplicative, whence also A. It suffices to calculate the value of

A for prime powers. Let p be a prime number and $\alpha \in \mathbb{N}$. Then

$$\begin{aligned} A(p^\alpha) &= \sum_{x|p^\alpha} \mu(x) \frac{\gcd(x,n)}{x} \cdot f(\gcd(x,n);\ell) \\ &= \mu(1)\frac{\gcd(1,n)}{1} \cdot f(\gcd(1,n);\ell) + \mu(p)\frac{\gcd(p,n)}{p} \cdot f(\gcd(p,n);\ell) \\ &= \begin{cases} 1 - \dfrac{1}{p} & \text{if } p \nmid n, \\ 1 & \text{if } p \mid n \text{ and } p \nmid \ell, \\ 0 & \text{if } p \mid n \text{ and } p \mid \ell. \end{cases} \end{aligned}$$

Since A is multiplicative, for each $M \in \mathbb{N}$, we have

$$A(M) = \prod_{\substack{p|M \\ p\nmid n}} A\left(p^{v_p(M)}\right) \cdot \prod_{\substack{p|M \\ p|n,\ p\nmid \ell}} A\left(p^{v_p(M)}\right) \cdot \prod_{\substack{p|M \\ p|n,\ p|\ell}} A\left(p^{v_p(M)}\right).$$

The second term is always 1. The last term is 0 if there is a prime number p such that $p \mid M$, $p \mid n$, and $p \mid \ell$, i.e. $\gcd(M,n,\ell) > 1$; otherwise, the last term is 1. Therefore

$$\begin{aligned} A(M) &= \prod_{\substack{p|M \\ p\nmid n}} \left(1 - \frac{1}{p}\right) \\ &= \left[\prod_{p|M} \left(1 - \frac{1}{p}\right)\right] \cdot \left[\prod_{\substack{p|M \\ p|n}} \left(1 - \frac{1}{p}\right)\right]^{-1} = \frac{\varphi(M)}{M} \cdot \frac{\gcd(M,n)}{\varphi(\gcd(M,n))} \end{aligned}$$

if $\gcd(M,n,\ell) = 1$; otherwise, $A(M) = 0$.

Therefore, for $\ell \mid n$ with $n \mid \ell d$,

$$\begin{aligned} \mathcal{C}(\ell;n) &= \varphi\left(\frac{n}{\ell}\right) \sum_{m|\frac{\ell d}{n}} A\left(\frac{\ell d}{nm}\right) \cdot \frac{q^m - 1}{n \cdot m} \\ &= \varphi\left(\frac{n}{\ell}\right) \sum_{\substack{m|\frac{\ell d}{n} \\ \gcd\left(\frac{\ell d}{nm},\ell\right)=1}} \left[\prod_{\substack{p|\frac{\ell d}{nm} \\ p\nmid n}} \left(1 - \frac{1}{p}\right)\right] \cdot \frac{q^m - 1}{n \cdot m}. \end{aligned}$$

We are interested in the following three special cases, namely when $\gcd(n,d) = 1$, $d \mid n$, and d is a prime power.

First, assume that $\gcd(n,d) = 1$. Then the length of nontrivial component in Γ/G is $\ell = n$ (since $\frac{n}{\ell}\Big|n$ and $\frac{n}{\ell}\Big|d$) and $\frac{\ell d}{n} = d$, so $\gcd\left(\frac{d}{m},\ell\right) = \gcd\left(\frac{d}{m},n\right) = 1$ for all $m \mid d$ and

$$\mathcal{C}(\ell;n) = \mathcal{C}(n;n) = \varphi(1) \sum_{m|d} \left[\prod_{p|\frac{d}{m}} \left(1 - \frac{1}{p}\right)\right] \cdot \frac{q^m - 1}{n \cdot m} = \sum_{m|d} \frac{m}{d} \varphi\left(\frac{d}{m}\right) \cdot \frac{q^m - 1}{n \cdot m}.$$

Therefore, $M_\Gamma = \mathcal{C}(n;n) \geq \dfrac{q^d - 1}{n \cdot d} \geq 2$, so $\Gamma/G - [G0]$ is symmetric.

Other cases are discussed in the next sections.

3.3 Special Case: $d \mid n$

In this section, we assume that $d \mid n$. Then the length of nontrivial component in Γ/G is of the form $\ell = \frac{n}{d} \cdot s$, where $s \mid d$. Note that $\ell d/n = s$ and $s \mid \ell$ because $\frac{n}{d} \in \mathbb{N}$. For each $m \mid s$, observe that $\gcd\left(\frac{s}{m}, \ell\right) = 1$ if and only if $m = s$. Then

$$\mathcal{C}(\ell; n) = \varphi\left(\frac{n}{\ell}\right) \cdot \frac{q^{\frac{\ell d}{n}} - 1}{n \cdot \frac{\ell d}{n}} = \varphi\left(\frac{d}{s}\right) \cdot \frac{q^s - 1}{n \cdot s}.$$

Next, we calculate the order of symmetry M_Γ using the following proposition and lemma.

Proposition 3.5 (Lifting The Exponent lemma [1])**.** *Let π be a prime number, $x, y \in \mathbb{Z}$ be such that $\pi \nmid x$, $\pi \nmid y$, and $\pi \mid x - y$, and let $n \in \mathbb{N}$. Then*

- *if π is odd, then $v_\pi(x^n - y^n) = v_\pi(x - y) + v_\pi(n)$;*
- *if $\pi = 2$ and n is even, then $v_2(x^n - y^n) = v_2(x - y) + v_2(x + y) + v_2(n) - 1$;*
- *if $\pi = 2$ and n is odd, then $v_2(x^n - y^n) = v_2(x - y)$,*

Lemma 3.6. *Let $\pi \nmid q$ be a prime number and η be the order of q modulo π. Then, for $s \in \mathbb{N}$ with $\operatorname{rad}(s) \mid q - 1$,*

$$v_\pi\left(\frac{q^s - 1}{(q-1) \cdot s}\right) = \begin{cases} 0 & \text{if } \eta \nmid s, \\ v_\pi(q^\eta - 1) - v_\pi(q - 1) & \text{if } \pi \text{ is odd and } \eta \mid s, \\ v_2(q + 1) - 1 & \text{if } \pi = 2 \text{ and } s \text{ is even}, \\ 0 & \text{if } \pi = 2 \text{ and } s \text{ is odd}. \end{cases}$$

Proof. Note that $\eta \leq \pi - 1 < \pi$, so $\eta = 1$ if $\pi = 2$. By LTE (Proposition 3.5), we distinguish four cases.

1. If $\eta \nmid s$, then $\gcd(\eta, s) < s$, so $\pi \nmid q^{\gcd(\eta,s)} - 1$ by minimality of η. Since $\gcd(q^s - 1, q^\eta - 1) = q^{\gcd(\eta,s)} - 1$ and $\pi \mid q^\eta - 1$, we have $\pi \nmid q^s - 1$, i.e. $v_\pi(q^s - 1) = 0$. Observe that $\pi \nmid q - 1$; for otherwise we would have $\eta = 1$, which is absurd as $\eta \nmid s$. So $\pi \nmid s$. Therefore, $v_\pi(q - 1) = 0 = v_\pi(s)$.
2. If π is odd and $\eta \mid s$, then $v_\pi(q^s - 1) = v_\pi(q^\eta - 1) + v_\pi(s) - v_\pi(\eta) = v_\pi(q^\eta - 1) + v_\pi(s)$.
3. If $\pi = 2$ and s is even, then $v_2(q^s - 1) = v_2(q - 1) + v_2(q + 1) + v_2(s) - 1$.
4. If $\pi = 2$ and s is odd, then $v_2(q^s - 1) = v_2(q - 1)$ and $v_2(s) = 0$.

From all the four cases, we obtain the result by subtracting $v_\pi(q - 1) + v_\pi(s)$. □

Observe that if $\pi \mid q$, then $\pi \nmid s$ as $s \mid q - 1$, so $v_\pi(q^s - 1) = v_\pi(q - 1) = v_\pi(s) = 0$. For each $s \mid q - 1$, it follows that $\frac{q^s - 1}{(q-1) \cdot s}$ has nonnegative π-adic valuation for all prime number π, so it is an integer. Note that Lemma 3.6 and this result do not use the assumption $d \mid n$.

Back to the calculation of M_Γ. As $M_\Gamma = \gcd\left\{\varphi\left(\frac{d}{s}\right) \cdot \frac{q-1}{n} \cdot \frac{q^s - 1}{(q-1) \cdot s} : s \mid d\right\}$, we have $\frac{q-1}{n} \Big| M_\Gamma$, so $M_\Gamma > 1$ if $n < q - 1$.

Assume $n = q - 1$. We claim that $M_\Gamma = 1$ if (d is odd or $v_2(q + 1) = 1 = v_2(d)$) and $2 \mid M_\Gamma$ otherwise. To prove this, let π be a prime number. We compute $v_\pi\left(\varphi\left(\frac{d}{s}\right) \cdot \frac{q^s - 1}{(q-1) \cdot s}\right)$ for some suitable $s \mid d$ to yield a value of $v_\pi(M_\Gamma)$. We distinguish three cases via Lemma 3.6.

1. If $\pi \mid q$, then $v_\pi\left(\dfrac{q^d-1}{(q-1)\cdot d}\right) = 0$ by the above observation, so $v_\pi(M_\Gamma) = 0$.

2. If $\pi \nmid q$ and π is odd, then $v_\pi(M_\Gamma) = v_\pi\left(\dfrac{q^d-1}{(q-1)\cdot d}\right) = 0$ if $\eta \nmid d$ or $\eta = 1$. Assume that $\eta \mid d$ and $\eta \neq 1$, i.e. $\pi \nmid q-1$, so $\pi \nmid d$. Pick

$$s = \prod_{\substack{p \mid d \\ \pi \mid p-1}} p^{v_p(d)}.$$

Clearly, $s \mid d$. Suppose to the contrary that $\eta \mid s$. Then there exists a prime number $p \mid d$ such that $\pi \mid p-1$ and $p \mid \eta$, so $\pi < \eta$ which is absurd. So, $\eta \nmid s$. Because $\pi \nmid d$, we have

$$v_\pi(M_\Gamma) = v_\pi\left(\varphi\left(\frac{d}{s}\right)\cdot\frac{q^s-1}{(q-1)\cdot s}\right) = \sum_{\substack{p \mid d \\ \pi \nmid p-1}} v_\pi(\varphi(p^{v_p(d)})) + 0 = 0.$$

3. If $\pi \nmid q$ and $\pi = 2$, then $v_2(M_\Gamma) = v_2\left(\dfrac{q^d-1}{(q-1)\cdot d}\right) = 0$ if d is odd or $v_2(q+1) = 1$. Assume that d is even and $v_2(q+1) > 1$. If $v_2(d) = 1$, then pick $s = \dfrac{d}{2}$, which is odd, so $v_2(M_\Gamma) = v_2\left(\varphi\left(\dfrac{d}{s}\right)\cdot\dfrac{q^s-1}{(q-1)\cdot s}\right) = 0$. Assume that $v_2(d) > 1$. Let $s \mid d$.

   - If s is even, then $v_2\left(\dfrac{q^s-1}{(q-1)\cdot s}\right) = v_2(q+1) - 1 > 0$.
   - If s is odd, then $v_2\left(\dfrac{d}{s}\right) = v_2(d) > 1$, so $v_2(M_\Gamma) = v_2\left(\varphi\left(\dfrac{d}{s}\right)\right) \geq v_2\left(\dfrac{d}{s}\right) - 1 > 0$.

Thus, $\Gamma/G - [G0]$ is not symmetric unless $n < q-1$ or both $v_2(q+1)$ and $v_2(d)$ are greater than 1. If $n < q-1$, then $\Gamma/G - [G0]$ is symmetric of order $\dfrac{q-1}{n}$. If $v_2(q+1) > 1$ and $v_2(d) > 1$, then $\Gamma/G - [G0]$ is symmetric of order 2.

3.4 Special Case: Prime-power degree of extension

In this section, we assume that $d > 1$ is a power of a prime number p. Then the length of nontrivial component in Γ/G is of the form $\ell = n/p^\lambda$, where $0 \leq \lambda \leq \min(v_p(d), v_p(n))$. Note that $\ell d/n = d/p^\lambda$.

1. If $v_p(n) = 0$, then $\gcd(n,d) = 1$, so

$$\mathcal{C}(\ell;n) = \mathcal{C}(n;n) = \frac{q^d-1}{n\cdot d} + \sum_{\substack{m \mid d \\ m<d}} \left(1-\frac{1}{p}\right)\cdot\frac{q^m-1}{n\cdot m}.$$

2. If $v_p(d) \leq v_p(n)$, then $d \mid n$, so

$$\mathcal{C}(\ell;n) = \varphi(p^\lambda)\cdot\frac{q^{d/p^\lambda}-1}{n\cdot d/p^\lambda}.$$

3. If $0 < v_p(n) < v_p(d)$, then $\gcd\left(\frac{d}{p^\lambda m}, \ell\right) = 1$ if and only if ($\lambda < v_p(n)$ and $m = \frac{d}{p^\lambda}$) or $\lambda = v_p(n)$ for all $m \Big| \frac{d}{p^\lambda}$, so

$$\mathcal{C}(\ell; n) = \begin{cases} \varphi(p^\lambda) \cdot \dfrac{q^{d/p^\lambda} - 1}{n \cdot d/p^\lambda} & \text{if } 0 \le \lambda < v_p(n), \\ \varphi(p^\lambda) \cdot \displaystyle\sum_{m \mid \frac{d}{p^\lambda}} \frac{q^m - 1}{n \cdot m} & \text{if } \lambda = v_p(n). \end{cases}$$

Finally, we calculate the order of symmetry M_Γ using the following lemma.

Lemma 3.7. *Assume that $d > 1$ is a power of a prime number p. If $p \mid n$, then*

$$\gcd(q^{d-1} + \cdots + q + 1, p - 1) = 1.$$

Proof. Assume $p \mid n$. Since $n \mid q - 1$, write $q = mp + 1$ for some $m \in \mathbb{N}$.
Assume $p > 2$. Let r be a prime divisor of $p - 1$. Then $p \equiv 1 \bmod r$, so $q \equiv m + 1 \bmod r$.

1. If $r \mid m$, then $q \equiv 1 \bmod r$, so $q^{d-1} + \cdots + q + 1 \equiv d \bmod r$. Since $r < p$ and d is a power of p, we have $\gcd(d, r) = 1$. It follows that $r \nmid q^{d-1} + \cdots + q + 1$.

2. If $r \nmid m$, then $m+1 \not\equiv 1 \bmod r$ and $q^{d-1}+\cdots+q+1 \equiv \dfrac{(m+1)^d - 1}{(m+1) - 1} \bmod r$. Since $r-1 < p$ and d is a power of p, we have $\gcd(d, r-1) = 1$, so the order of $m+1$ modulo r does not divide d. It follows that $(m+1)^d \not\equiv 1 \bmod r$, i.e. $r \nmid q^{d-1} + \cdots + q + 1$.

From the both cases, we conclude that $\gcd(q^{d-1} + \cdots + q + 1, p - 1) = 1$. □

Back to the calculation of M_Γ.

1. If $v_p(n) = 0$, then $M_\Gamma = \mathcal{C}(n; n) \ge 2$, so $\Gamma/G - [G0]$ is symmetric.

2. If $v_p(d) \le v_p(n)$, then $M_\Gamma = 1$ unless $n < q-1$ or both $v_2(q+1)$ and $v_2(d)$ are greater than 1. In these exceptional cases, $\Gamma/G - [G0]$ is symmetric of order $\dfrac{q-1}{n}$ and 2, respectively.

3. If $0 < v_p(n) < v_p(d)$, then

$$M_\Gamma = \gcd\left\{ \varphi(p^{v_p(n)}) \cdot \sum_{k \mid d/p^{v_p(n)}} \frac{q-1}{n} \cdot \frac{q^k - 1}{(q-1) \cdot k},\ \varphi(p^\lambda) \cdot \frac{q-1}{n} \cdot \frac{q^{d/p^\lambda} - 1}{n \cdot d/p^\lambda} : 0 \le \lambda < v_p(n) \right\}.$$

Since $v_p(n) \ge 1$, we have $p \mid q-1$, so $\dfrac{q^{d/p^\lambda} - 1}{(q-1) \cdot d/p^\lambda} \in \mathbb{N}$ for every $0 \le \lambda \le v_p(n)$ by Lemma 3.6. Then $\Gamma/G - [G0]$ is symmetric of order $\dfrac{q-1}{n}$ if $n < q - 1$ as $\dfrac{q-1}{n} \Big| M_\Gamma$.

Assume that $n = q - 1$. Note that $\dfrac{q^{p^\alpha} - 1}{(q-1) \cdot p^\alpha} \Big| \dfrac{q^{p^\beta} - 1}{(q-1) \cdot p^\beta}$ for all $0 \le \alpha \le \beta$. By Lemmas 3.6 and 3.7,

$$M_\Gamma = \gcd\left\{ p^{v_p(n)-1} \cdot \sum_{k \mid d/p^{v_p(n)}} \frac{q^k - 1}{(q-1) \cdot k},\ \frac{q^{d/p^{v_p(n)-1}} - 1}{(q-1) \cdot d/p^{v_p(n)-1}} \right\}.$$

If $p = 2$ and $q \equiv 3 \bmod 4$ and $v_2(n) \geq 2$, then $2^{\min(v_2(n)-1, v_2(q+1)-1)} \mid M_\Gamma$, i.e. $\Gamma/G - [G0]$ is symmetric of order 2. In other cases, we have

$$M_\Gamma = \gcd\left\{ \sum_{k \mid d/p^{v_p(n)}} \frac{q^k - 1}{(q-1)\cdot k},\ \frac{q^{d/p^{v_p(n)-1}} - 1}{(q-1)\cdot d/p^{v_p(n)-1}} \right\}.$$

It is still hard to compute M_Γ.

In order to illustrate our results, we provide examples of the quotient digraph of a functional graph $\Gamma = \Gamma(\mathbb{F}_{7^d}, 3x)/G$ by $G = \mathrm{Gal}(\mathbb{F}_{7^d}/\mathbb{F}_7)$, where $d = 3, 4, 5, 6, 10$. Note that $q = 7$ and $n = \mathrm{ord}(3, \mathbb{F}_7^\times) = 6 = q - 1$.

Table 1: $\mathcal{C}(\ell; n)$ for $\ell \mid n$ with $n \mid \ell d$

d	condition	ℓ	$\mathcal{C}(\ell; n)$
3	$v_3(n) = 1 = v_3(d)$	$6 = \frac{6}{1}$	$\varphi(1) \cdot \frac{7^{3/1} - 1}{6 \cdot 3/1} = 29$
		$2 = \frac{6}{3}$	$\varphi(3) \cdot \frac{7^{3/3} - 1}{6 \cdot 3/3} = 2$
4	$v_2(n) = 1 < 2 = v_2(d)$	$6 = \frac{6}{1}$	$\varphi(1) \cdot \frac{7^{4/1} - 1}{6 \cdot 4/1} = 100$
		$3 = \frac{6}{2}$	$\varphi(2) \cdot \left[\frac{7^{4/2} - 1}{6 \cdot 4/2} + \frac{7^{4/4} - 1}{6 \cdot 4/4}\right] = 5$
5	$v_5(n) = 0$	6	$\frac{7^5 - 1}{6 \cdot 5} + \left(1 - \frac{1}{5}\right) \cdot \frac{7^1 - 1}{6 \cdot 1} = 561$
6	$d \mid n$	6	$\varphi\left(\frac{6}{6}\right) \cdot \frac{7^6 - 1}{6 \cdot 6} = 3268$
		3	$\varphi\left(\frac{6}{3}\right) \cdot \frac{7^3 - 1}{6 \cdot 3} = 29$
		2	$\varphi\left(\frac{6}{2}\right) \cdot \frac{7^2 - 1}{6 \cdot 2} = 8$
		1	$\varphi\left(\frac{6}{1}\right) \cdot \frac{7^1 - 1}{6 \cdot 1} = 2$
10	General	6	$\varphi\left(\frac{6}{6}\right) \cdot \left[\left(1 - \frac{1}{5}\right) \cdot \frac{7^2 - 1}{6 \cdot 2} + 1 \cdot \frac{7^{10} - 1}{6 \cdot 10}\right] = 4707924$
		3	$\varphi\left(\frac{6}{3}\right) \cdot \left[\left(1 - \frac{1}{5}\right) \cdot \frac{7^1 - 1}{6 \cdot 1} + 1 \cdot \frac{7^5 - 1}{6 \cdot 5}\right] = 561$

Therefore, $\Gamma/G - [G0]$ is not symmetric when $d = 3, 6$, and is symmetric of order 5, 561, and 3 when $d = 4, 5, 10$ respectively.

Acknowledgment. The authors are grateful to the referees for their careful reading of the manuscript and their useful comments.

References

[1] A.H. Parvardi, *Lifting The Exponent Lemma (LTE)*, 2011, https://s3.amazonaws.com/aops-cdn.artofproblemsolving.com/resources/articles/lifting-the-exponent.pdf.

[2] D.M. Burton, *Elementary Number Theory*, 6th ed., McGraw-Hill, New York, 2007.

[3] D.S. Dummit and R.M. Foote, *Abstract Algebra*, 3rd ed., Wiley, New York, 2003.

[4] I. Niven, H.S. Zuckerman and H.L. Montgomery, *An Introduction to the Theory of Numbers*, 5th ed., Wiley, New York, 1991.

[5] R.P. Stanley, *Enumerative Combinatorics, Volume 1*, 2nd ed., Cambridge University Press, New York, 2012.

[6] A. Sawkmie and M.M. Singh, *Digraphs associated with the kth power map on the quotient ring of polynomials over finite fields*, J. Algebra Appl., **18**(11) (2019), 1950218.

[7] L. Somer and M. Křížek, *The structure of digraphs associated with the congruence* $x^k \equiv y \pmod{n}$, Czechoslovak Math. J., **61**(136) (2011), 337–358.

[8] T. Chen and E. Scheinerman, *Finding a Compositional Square Root of Sine*, Amer. Math. Monthly, **129**(9) (2022), 816–830.

[9] Y. Meemark and N. Wiroonsri, *The digraph of the kth power mapping of the quotient ring of polynomials over finite fields*, Finite Fields Appl., **18** (2012), 179–191.

4.
Analysis, Fixed Point Theory and Applications, Topology and Geometry

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Fixed Point Theory for α-G-Contraction Types on Uniform Spaces with a Graph G

Sittichoke Songsa-ard[1,†]

[1]Mathematics Program, Faculty of Science and Technology
Suratthani Rajabhat University, Suratthani 84100, Thailand

Abstract

In this work, we investigate the properties of α-G-contraction selfmaps in the context of uniform spaces with an associated graph. We provide sufficient conditions that ensure the existence of fixed points and demonstrate our findings through a collection of illustrative examples.

Keywords: α-G-contraction, uniform spaces with a graph, Picard operator.

2020 MSC: Primary 47H10; Secondary 47H09.

1 Introduction

In recent years, fixed point theory has become a significant tool in various areas of mathematics, especially in the study of functional equations, differential equations, and optimization. A fundamental result in fixed point theory is the Banach Contraction Principle. Since its introduction, many researchers have aimed to generalize and extend the Banach Contraction Principle to different contexts, such as metric spaces with a graph or partially ordered sets.

The study of maps on complete metric spaces endowed with a partial ordering has been a subject of interest since the work of A.C.M. Ran and M.C.B. Reurings [8]. Later, J.J. Nieto and R. Rodríguez-López [7] further generalized these results and introduced Picard operators in the context of partially ordered sets. In 2008, Jacek Jachymski [5] introduced the concept of generalizations of contractions on metric spaces with partially ordered sets to metric spaces endowed with a graph.

Parallel to these developments, progress has been made in examining contractions on uniform spaces. In 1987, Angelov [1] put forth the concept of Φ-contractions on Hausdorff uniform spaces, which concurrently generalizes the Banach contractions on metric spaces and γ-contractions [6] on locally convex spaces, and demonstrated the existence of fixed points under various circumstances. Later, in 1991 [2], he expanded the notion of Φ-contractions to encompass j-nonexpansive maps and established conditions to ensure the existence of their fixed points.

†Speaker. ‡Corresponding author.
Email: sittichoke.son@sru.ac.th (S. Songsa-ard)

Expanding upon this research, we delved into functionally lipschitzian (FL) and functionally uniformly lipschitzian (FUL) maps in locally convex spaces [4]. We focused on the weak topology in normed spaces and provided criteria for FL and FUL maps, demonstrating that FL maps are weakly continuous. Furthermore, we investigated fixed points in uniform spaces [3], examining sufficient conditions for the existence of fixed points of J-contractions in uniform spaces equipped with a collection of pseudometrics. We also presented examples of ordinary differential equations (ODEs) that employ the main theorem to ensure the existence of solutions.

Inspired by these previous works and the developments in the field, we focus our investigation on α-G-contractions on uniform spaces generated by a collection of pseudometrics. We aim to provide criteria for these contractions by drawing ideas from our work on FL and FUL maps and to demonstrate their applicability through examples similar to the integral equations presented in [3].

This paper is organized into three chapters, beginning with an introduction that provides motivation, background, and essential definitions. In Chapter 2, we introduce the novel concept of α-G-contraction maps on uniform spaces, present criteria for analyzing maps on ℓ_p equipped with the weak topology, which is a uniform space induced by a collection of pseudometrics resulting from seminorms, and explore the connection between weakly connected and Cauchy equivalent in uniform spaces. Chapter 3 presents two main theorems outlining sufficient conditions for maps to be Picard operators, along with criteria to ensure their satisfaction.

2 Preliminaries

In this part, we provide some background and preliminary concepts in graph theory and uniform spaces, necessary for the understanding of our main results. A directed graph (or digraph) is a pair consisting of a nonempty set of vertices and a set of ordered pairs of distinct vertices, called edges or arcs. The graph of interest has an infinite number of vertices, is directed, has weights on the connecting lines as distances between two vertices along that line, and every vertex has a loop.

We also introduce the concepts of conversion and undirected graphs. A conversion graph of a graph G is obtained by reversing the direction of the edges of G, denoted by G^{-1}. The conversion graph has the same vertex set as G, $V(G^{-1}) = V(G)$, and the edge set is defined as $E(G^{-1}) = \{(x, y) \in X \times X : (y, x) \in E(G)\}$.

An undirected graph of G, denoted by $\tilde{G}$, is defined such that $V(\tilde{G}) = V(G)$ and $E(\tilde{G}) = E(G) \cup E(G^{-1})$. In other words, if graph G has an edge connecting vertices x and y, the undirected graph of G has edges connecting x to y and y to x.

A path in a graph G from vertex x to vertex y with length N ($N \in \mathbb{N} \cup 0$) is a sequence $(x_i)_{i=0}^{N}$ of $N + 1$ vertices such that $x_0 = x$, $x_N = y$, and $(x_{i-1}, x_i) \in E(G)$ for $i = 1, ..., N$. A graph G is connected if, for every $x, y \in V(G)$, there exists a path in G from x to y and from y to x. Finally, a graph G is weakly connected if its undirected graph $\tilde{G}$ is connected. The definition of G-contraction on metric spaces endowed with a graph G, as presented in Jacek's work, is as follows:

Definition 2.1. Let $f : X \to X$. f is called a *G-contraction* if f preserves edges of G, that is,

$$\forall x, y \in X[(x, y) \in E(G) \Rightarrow (fx, fy) \in E(G)]$$

and f decreases the weight of the edges of G, that is,

$$\exists c \in (0, 1) \forall x, y \in X[(x, y) \in E(G) \Rightarrow d(fx, fy) \leq cd(x, y)].$$

Next, we lay the foundation for our main results by introducing essential concepts and proving several key lemmas. Firstly, we present the definition of an α-G-contraction on a uniform space generated by a collection of pseudometrics.

Next, we establish a lemma stating that if T is an α-G-contraction, then T is also an α-G^{-1}-contraction and an α-$\tilde{G}$-contraction.

Following this, we introduce a significant lemma that is essential for bounding the tail of a series, similar to the Cauchy sense. This lemma will be instrumental in our upcoming theorems.

Lastly, we present the concepts of Cauchy equivalence and weakly connectedness in uniform spaces generated by a collection of pseudometrics. We then prove that two specific types of iteration sequences starting from distinct points are Cauchy equivalent if and only if the graph G is weakly connected. This result offers additional understanding of the relationship between graph theory and uniform spaces.

Definition 2.2. Let $(E, \mathcal{A})$ be a uniform space generated by a collection of pseudometrics indexed by a set A, X be a nonempty subset of E, and $T : X \to X$.
T is said to be *α-G-contraction* if for each $\alpha \in A$, there is a graph G such that

1. for any $(x, y) \in E(G)$, $(Tx, Ty) \in E(G)$, and
2. there exists $c_\alpha \in (0, 1)$ such that for any $(x, y) \in E(G)$, then

$$p_\alpha(Tx, Ty) \leq c_\alpha p_\alpha(x, y),$$

where p_α is a pseudometric indexed by α.

We also provide a criterion for checking whether a map $T : \ell_p \to \ell_p$ is an α-G-contraction on ℓ_p with respect to the weak topology where $1 < p < \infty$, as follows in the next theorem:

Theorem 2.3. *Let $1 < p < \infty$, $f_1, f_2, ..., f_N$ be Lipschitz functions with Lipschitz constants $k_1, k_2, ..., k_N$, $c \in (0, 1)$, and $T : \ell_p \to \ell_p$ defined by*

$$T((x_m)) = (f_1 x_1, f_2 x_2, ..., f_N x_N, c x_{N+1}, c x_{N+2}, ...).$$

Let G be a graph with $V(G) = \ell_p$ and

$$E(G) = \{((x_m), (y_m)) : x_i = y_i \ \forall \ i = 1, \dots, N\}.$$

Then T is an α-G-contraction on ℓ_p with respect to the weak topology.

Proof. We will show that T preserves the edges of G. Let $((x_m), (y_m)) \in E(G)$, i.e., $x_i = y_i$ for all $i = 1, \dots, N$. Then, $(T((x_m)), T((y_m))) \in E(G)$, so T preserves the edges of G.

Now, we will show that T decreases the weight of the edges of G. Let $(\alpha_m) \in \ell_q$ with $\frac{1}{q} + \frac{1}{p} = 1$ and $((x_m), (y_m)) \in E(G)$.
Consider $T(x_m) - T(y_m) = (f_1 x_1 - f_1 y_1, \dots, f_N x_N - f_N y_N, c x_{N+1} - c y_{N+1}, \dots)$, which satisfies

$$\left| \sum_{m=1}^{\infty} \alpha_m \cdot e_m^*(T(x_m) - T(y_m)) \right| \leq \sum_{i=1}^{N} k_i |\alpha_i| \, |(x_i - y_i)| + c \left| \sum_{i=N+1}^{\infty} \alpha_i \cdot (x_i - y_i) \right|.$$

Since $x_i = y_i$ for every $i = 1, \dots, N$, it follows that

$$\left| \sum_{m=1}^{\infty} \alpha_m \cdot e_m^*(T(x_m) - T(y_m)) \right| \leq c \left| \sum_{i=1}^{\infty} \alpha_i \cdot (x_i - y_i) \right|.$$

Thus, T is an α-G-contraction on ℓ_p with respect to the weak topology. □

To further illustrate the concept of α-G-contractions and their properties, we present the following example. This example demonstrates that every γ-contraction [6] can be considered as an α-G-contraction when associated with a complete graph.

Example 2.4. Let $(E, \mathcal{A})$ be a uniform space generated by a collection of pseudometrics indexed by a set A, X be a nonempty subset of E, and $T : X \to X$.
If T is a γ-contraction for each $\gamma \in A$, then T is also an α-G-contraction when associated with a complete graph.

Proof. Recall the definition of γ-contraction in [6], for each $\gamma \in A$, there is a constant $c \in (0, 1)$ such that $p_\gamma(Tx, Ty) \leq cp_\gamma(x, y)$ for any $x, y \in X$. It is easy to see that T is also an α-G-contraction when associated with a complete graph i.e.,

$$V(G) = X \text{ and } E(G) = X \times X.$$

□

Example 2.5. Let $1 < p < \infty$ and $T : \ell_p \to \ell_p$ with

$$T((x_m)) = c(x_1, x_2, \dots, x_N, \dots) + (z_n) \quad \text{where } (z_n) \in \ell_p.$$

If $c \in (0, 1)$, then T is an α-G contraction, where G is a complete graph, i.e.,

$$V(G) = \ell_p \text{ and } E(G) = \ell_p \times \ell_p.$$

Next, we introduce some additional graph theory concepts before exploring various properties and characteristics of α-G-contractions in the context of uniform spaces and graphs. The *equivalence class of* x on a graph G, denoted by $[x]_G$, consists of vertices related through a relation R on $V(G)$. Here, yRz holds if and only if there exists a path in G from y to z and from z to y. Furthermore, the *component of* G *containing vertex* x denoted by G_x refers to the largest connected subgraph of G that includes x as a vertex, with its vertices making up the equivalence class $[x]_G$.

Lemma 2.6. *Let $(E, \mathcal{A})$ be a uniform space generated by a collection of pseudometrics indexed by a set A, X be a nonempty subset of E, and $T : X \to X$.*
If T is α-G-contraction, then T is α-G^{-1}-contraction and α-$\tilde{G}$-contraction.

Proof. Let $(x, y) \in E(G^{-1})$. Then $(y, x) \in E(G)$, so $(Ty, Tx) \in E(G)$ and hence $(Tx, Ty) \in E(G^{-1})$. Since T is α-G-contraction, there exists $c_\alpha \in (0, 1)$ such that for any $(x, y) \in E(G)$, then $p_\alpha(Tx, Ty) \leq c_\alpha p_\alpha(x, y)$. Let $(x, y) \in E(G^{-1})$. Then $(y, x) \in E(G)$, so $p_\alpha(Tx, Ty) = p_\alpha(Ty, Tx) \leq c_\alpha p_\alpha(y, x) = c_\alpha p_\alpha(x, y)$. Then T is α-G^{-1}-contraction and T is α-$\tilde{G}$-contraction.
□

Lemma 2.7. *Let X be nonempty subset of a uniform space generated by a collection of pseudometrics indexed by a set A, and $T : X \to X$ be an α-G-contraction with a constant c_α. Then for any $x \in X$, and $y \in [x]_{\tilde{G}}$, there exists $r_\alpha(x, y) \geq 0$ such that*

$$p_\alpha(T^n x, T^n y) \leq c_\alpha^n r_\alpha(x, y) \quad \textit{for any } \alpha \in A, \textit{ and } n \in \mathbb{N}.$$

Proof. Let $x \in X$ and $y \in [x]_{\tilde{G}}$. Then there is a path $(x_i)_{i=0}^N$ in $\tilde{G}$ from x to y for some $N \in \mathbb{N}$ such that $x_0 = x$ and $x_N = y$ and $(x_{i-1}, x_i) \in E(\tilde{G})$ for each $i = 1, 2, \dots, N$. Let $\alpha \in A$. Then there exists $c_\alpha \in (0, 1)$ such that

$$p_\alpha(T^n x_{i-1}, T^n x_i) \leq c_\alpha^n p_\alpha(x_{i-1}, x_i) \quad \text{for any } i = 1, 2, \dots, N.$$

By letting $r_\alpha(x, y) = \sum_{i=1}^N p_\alpha(x_{i-1}, x_i)$, then

$$\begin{aligned} p_\alpha(T^n x, T^n y) &\leq \sum_{i=1}^N p_\alpha(T^n x_{i-1}, T^n x_i) \\ &\leq c_\alpha^n \sum_{i=1}^N p_\alpha(x_{i-1}, x_i) = c_\alpha^n r_\alpha(x, y). \end{aligned}$$

□

And next, we discuss the concepts of Cauchy equivalence and weakly connected graphs in the context of uniform spaces generated by a collection of pseudometrics. We focus on the relationship between these properties and the uniform structure induced by the pseudometrics. We begin by introducing the concept of Cauchy equivalence:

Definition 2.8. A sequence (x_n) and (y_n) are said to be *Cauchy equivalent* in a uniform space $(X, \mathcal{A})$ generated by a collection of pseudometrics $\mathcal{A}$ with A as the index set, if for all sequences (x_n) and (y_n) that are Cauchy sequences, $p_\alpha(x_n, y_n) \to 0$ as $n \to \infty$ for every $\alpha \in A$.

Throughout this part, we will explore the implications of Cauchy equivalence in the context of weakly connected graphs and the properties of uniform spaces induced by pseudometrics.

Theorem 2.9. *G is weakly connected if and only if for any α-G-contraction $T : X \to X$ and for any $x, y \in X$, $(T^n x)$ and $(T^n y)$ are Cauchy equivalent.*

Proof. ($\Rightarrow$) Let $x, y \in X$. Since G is weakly connected, $[x]_{\tilde{G}} = X$, so $Tx \in [x]_{\tilde{G}}$. To show that $T^n x$ is a Cauchy sequence, let $\alpha \in A$ and $\epsilon > 0$. By lemma 2.7 and definition of α-G-contraction, there exist $r_\alpha(x, Tx) > 0$ such that

$$p_\alpha(T^n x, T^{n+1} x) \leq c_\alpha^n r_\alpha(x, Tx).$$

Since $c_\alpha \in (0, 1)$, there exists $N \in \mathbb{N}$ such that for any $m, n \geq N$, $\sum_{i=m}^{n} c_\alpha^i < \epsilon$. Let $m, n \geq N$ such that $m \leq n$. Then

$$\begin{aligned} p_\alpha(T^m x, T^n x) &\leq \sum_{i=m}^{n-1} p_\alpha(T^i x, T^{i+1} x) \\ &\leq \sum_{i=m}^{n-1} c_\alpha^i r_\alpha(x, Tx) \\ &= r_\alpha(x, Tx) \sum_{i=m}^{n-1} c_\alpha^i. \end{aligned}$$

Since $\sum_{i=m}^{n} c_\alpha^i < \epsilon$, $(T^n x)$ is a Cauchy sequence. Since $y \in [x]_{\tilde{G}}$, by lemma 2.7, for all $\alpha \in A$, there exists $r_\alpha(x.y) \geq 0$ such that

$$p_\alpha(T^n x, T^n y) \leq c_\alpha^n r_\alpha(x, y) \quad \text{for any } n \in \mathbb{N}.$$

Because $\lim_{n\to\infty} c_\alpha^n = 0$, $(T^n x)$ and $(T^n y)$ are Cauchy equivalent.
($\Leftarrow$) Suppose that G is not weakly connected, then there are two point x_0 and y_0 in X such that $y_0 \notin [x_0]_{\tilde{G}}$. Let $T : X \to X$ be defined by

$$Tx = \begin{cases} x_0 & \text{if } x \in [x_0]_{\tilde{G}} \\ y_0 & \text{if } x \notin [x_0]_{\tilde{G}}. \end{cases}$$

It is clear that $(T^n x_0) = (x_0)$ and $(T^n y_0) = (y_0)$ are constant sequences, however, $x_0 \neq y_0$, so $(T^n x_0)$ and $(T^n y_0)$ are not Cauchy equivalent. It remains to show that T is α-G-contraction. Let $\alpha \in A$.

1. Let $(x, y) \in E(G)$. Then $y \in [x]_{\tilde{G}}$, so $(Tx, Ty) = (x_0, x_0) \in E(G)$ or $(Tx, Ty) = (y_0, y_0) \in E(G)$. Hence T preserves edges.

2. Let $(x, y) \in E(G)$. Then
$$p_\alpha(Tx, Ty) = 0 \leq c_\alpha p_\alpha(x, y).$$

Then T is an α-G-contraction, so we are done. □

In the following theorem, we explore a specific property that arises when restricting an α-G-contraction to a certain component of a graph $\tilde{G}$. The result demonstrates that the equivalence class $[x_0]_{\tilde{G}}$ is T-invariant, and the restricted mapping $T|_{[x_0]_{\tilde{G}}}$ is an α-$\tilde{G}_{x_0}$-contraction. This finding provides valuable insights into the behavior of the contraction mapping in relation to the graph-theoretic properties of the component.

Theorem 2.10. *Let X be nonempty subset of a uniform space generated by a collection of pseudo metrics indexed by a set A, and $T : X \to X$ be an α-G-contraction. Assume that there is a point $x_0 \in X$ such that $Tx_0 \in [x_0]_{\tilde{G}}$. Let $\tilde{G}_{x_0}$ be the component of $\tilde{G}$ containing x_0. Then $[x_0]_{\tilde{G}}$ is T-invariant and is $T|_{[x_0]_{\tilde{G}}}$ is an α-$\tilde{G}_{x_0}$-contraction.*

Proof. Let $x \in [x_0]_{\tilde{G}}$. Then there is a path $(x_i)_{i=0}^N$ from x_0 to x such that $x_N = x$ and $(x_{i-1}, x_i) \in E(\tilde{G})$ for each $i = 1, 2, \ldots, N$. Since T preserves edges of $\tilde{G}$, $(Tx_{i-1}, Tx_i) \in E(\tilde{G})$ for each $i = 1, 2, \ldots, N$. Then $(Tx_i)_{i=0}^N$ is a path in $\tilde{G}$ from Tx_0 to Tx, so $Tx \in [Tx_0]_{\tilde{G}} = [x_0]_{\tilde{G}}$ and hence $T|_{[x_0]_{\tilde{G}}}$ is a selfmap on $[x_0]_{\tilde{G}}$.
To show that $T|_{[x_0]_{\tilde{G}}}$ is an α-G-contraction, let $\alpha \in A$.
(1) : Let $(x, y) \in E(\tilde{G}_{x_0})$. There is a path $(x_i)_{i=0}^N$ from x_0 to y such that $x_{N-1} = x$, $x_N = y$, and $(x_{i-1}, x_i) \in E(\tilde{G}_{x_0})$. Since $(x_{N-1}, x_N) = (x, y) \in E(\tilde{G}_{x_0}) \subseteq E(\tilde{G})$, $(Tx_{N-1}, Tx_N) = (Tx, Ty) \in E(\tilde{G})$. Since $Tx_0 \in [x_0]_{\tilde{G}}$, there is a path $(y_i)_{i=0}^M$ from x_0 to Tx_0 such that $y_0 = x_0$, $y_M = Tx_0$ and $(y_{i-1}, y_i) \in E(\tilde{G}_{x_0})$. Since T is an α-$\tilde{G}$-contraction, $(Tx_i)_{i=0}^N$ is a path in $\tilde{G}$ from Tx_0 to Ty. Then $(y_0 = x_0, y_1, \ldots, y_M = Tx_0, Tx_0, Tx_1, \ldots, Tx_{N-1} = Tx, Tx_N = Ty)$ is a path in $\tilde{G}$ from x_0 to Ty, by the definition of $\tilde{G}_{x_0}$, $(Tx, Ty) = (Tx_{N-1}, Tx_N) \in E(\tilde{G}_{x_0})$. Hence $T|_{[x_0]_{\tilde{G}}}$ preserves edges of $\tilde{G}_{x_0}$.
(2) : Clearly, by T is an α-$\tilde{G}$-contraction, there exists $c_\alpha \in (0, 1)$ such that for any $(x, y) \in E(\tilde{G}_{x_0})$, then

$$p_\alpha(T|_{[x_0]_{\tilde{G}}}x, T|_{[x_0]_{\tilde{G}}}y) \leq c_\alpha p_\alpha(x, y).$$

□

3 Main Results

In this section, we present two main theorems that focus on the conditions under which an α-G-contraction is a Picard Operator (P.O.) in the context of uniform spaces generated by a collection of pseudometrics. We begin with Main Theorem I, which establishes conditions for a Picard Operator in the presence of a specific property $(*)$ that involves convergent sequences in X and the existence of subsequences with edges in $E(\tilde{G})$. Under these conditions and with T being an α-G-contraction, we demonstrate that $T|_{[x]_{\tilde{G}}}$ is a Picard Operator for $x \in X_T$, where X_T is a set of vertices in X with edges in $E(\tilde{G})$.

Moving on, we present Main Theorem II, which explores a different set of conditions for an α-G-contraction to be a Picard Operator. In this case, we consider an orbitally G-continuous α-G-contraction T and again define X_T as the set of vertices with edges in $E(G)$. If $x \in X_T$, we show that $T|_{[x]_{\tilde{G}}}$ is a Picard Operator.

These two main theorems contribute to a deeper understanding of the properties and behavior of α-G-contractions under different conditions in the context of uniform spaces and graphs. As we progress through this section, we will elaborate on these results and their implications for the study of α-G-contractions.

Theorem 3.1 (The main theorem I)**.** *Let X be nonempty sequentially complete subset of a uniform space generated by a collection of pseudo metrics indexed by a set A, and $(X, \mathcal{A}, G)$ have the following property $(*)$:*

for each sequence (x_n) in X, if $x_n \to x$ as $n \to \infty$ and $(x_n, x_{n+1}) \in E(G)$ for any $n \in \mathbb{N}$, then there exists a subsequence (x_{n_k}) of (x_n) with $(x_{n_k}, x) \in E(\tilde{G})$ for any $k \in \mathbb{N}$.

Let $T : X \to X$ be an α-G-contraction, and $X_T = \{x \in X : (x, Tx) \in E(\tilde{G})\}$. If $x \in X_T$, then $T|_{[x]_{\tilde{G}}}$ is a Picard operator.

Proof. Let $x \in X_T$. Then $(x, Tx) \in E(G)$, by theorem 2.10, $[x]_{\tilde{G}}$ is T-invariant and $T|_{[x]_{\tilde{G}}}$ is an α-$\tilde{G}_x$-contraction. By lemma 2.7, since $\tilde{G}_x$ is weakly connceted, for any $y \in \tilde{G}_x$, $(T^n x)$ is a Cauchy sequence. Since X is sequentially complete, there exists $x_0 \in X$ such that $T^n x \to x_0$ as $n \to \infty$. Since $(x, Tx) \in E(G)$, so is $(T^n x, T^{n+1}x)$ for any $n \in \mathbb{N}$. By property $(*)$, there is a subsequence $(T^{n_k}x)$ of $(T^n x)$ with $(T^{n_k}x, x_0) \in E(G)$ for any $k \in \mathbb{N}$. Then $(T^{n_k+1}x, Tx_0) \in E(G)$, so $(x, Tx, T^2x, \ldots, T^{n_1}x, x_0)$ is a path in $\tilde{G}$ from x to x_0. Hence $x_0 \in [x]_{\tilde{G}}$ and for each $\alpha \in A$

$$p_\alpha(T^{n_k+1}x, Tx_0) \leq c_\alpha p_\alpha(T^{n_k}x, x_0) \quad \text{for any } k \in \mathbb{N}.$$

Since T^n converges to x_0, so are (T^{n_k+1}) and (T^{n_k}). By continuity of p_α for any $\alpha \in A$, $p_\alpha(x_0, Tx_0) = 0$, then we have $Tx_0 = x_0$. □

In the proof of Theorem 3.1, we showed that under the given conditions, $T|_{[x]_{\tilde{G}}}$ is a Picard operator. Now, we can extend this result further by considering the case when the graph G is weakly connected, which leads us to the following corollary.

Corollary 3.2. *Let $T : X \to X$ be an α-G-contraction, and G be weakly connected. Then T is a Picard operator.*

Proof. Since G is weakly connected, $X_T = X$, so for all $x \in X$, $[x]_{\tilde{G}} = X$. By the proof of Theorem 3.1, we are done. □

In Theorem 3.1, we establish a set of conditions under which an α-G-contraction T on a nonempty sequentially complete subset X of a uniform space generated by a collection of pseudometrics becomes a Picard operator. One crucial condition is property $(*)$, which relates to the behavior of sequences in X and their connection to the graph G.

Theorem 3.3 provides a concrete example of a situation where condition $(*)$ is satisfied, by considering the graph G defined on the space ℓ_p. By constructing the edges of G in a specific way, we show that (ℓ_p, ℓ_p^*, G) indeed satisfies condition $(*)$.

Theorem 3.3. *Let G be a graph defined by $V(G) = \ell_p$. The edges of graph G are defined as*

$$E(G) = \{((x_n), (y_n)) : x_i = y_i \text{ for all } i = 1, \ldots, N\}.$$

It follows that (ℓ_p, ℓ_p^*, G) satisfies condition $(*)$ in theorem 3.1.

Proof. Let $(x_n^k)_{k\in\mathbb{N}}$ be a sequence in ℓ_p. Assume that (x_n^k) converges weakly to (x_n^0) as $k \to \infty$ and $((x_n^k), (x_n^{k+1})) \in E(G)$ for all $k \in \mathbb{N}$. It follows that for every $(\alpha_n) \in \ell_q$,

$$\sum_{n=1}^{\infty} \alpha_n x_n^k \to \sum_{n=1}^{\infty} \alpha_n x_n^0 \text{ as } k \to \infty.$$

Since $((x_n^k), (x_n^{k+1})) \in E(G)$ for all $k \in \mathbb{N}$, it holds that for $i = 1, 2, \ldots, N$, $e_i^*(x_n^k) = e_i^*(x_n^{k+1})$. Hence, $x_i^k = x_i^{k+1}$ for all $k \in \mathbb{N}$ and $i = 1, 2, \ldots, N$. As $e_i \in \ell_q$ for all $i = 1, 2, \ldots, N$, we have

$$\sum_{n=1}^{\infty} (e_n^*(e_i) \cdot x_n^k) \to \sum_{n=1}^{\infty} (e_n^*(e_i) \cdot x_n^0) \text{ as } k \to \infty.$$

This implies that $x_i^k \to x_i^0$ as $k \to \infty$. Thus, $x_i^k = x_i^0$. Therefore, $((x_n^k), (x_n^0)) \in E(G)$. Consequently, (ℓ_p, ℓ_p^*, G) satisfies condition $(*)$. □

Before presenting Lemma 3.4, it is essential to ensure that the set $(\ell_p)_T$ is nonempty, as this has implications for the existence of fixed points for the Lipschitz functions $f_1, f_2, \ldots, f_N$. In the context of the α-G-contraction T on ℓ_p and the graph G defined as in Theorem 3.3, Lemma 3.4 provides a characterization of the existence of fixed points for these Lipschitz functions. Specifically, we show that $f_1, f_2, \ldots, f_N$ have fixed points if and only if $(\ell_p)_T \neq \emptyset$. This result further supports the applicability of Theorem 3.1 in the case of ℓ_p and the graph G defined in Theorem 3.3.

Lemma 3.4. *Let $1 < p < \infty$ and $f_1, f_2, \ldots, f_N$ be Lipschitz functions with Lipschitz constants $k_1, k_2, \ldots, k_N$ respectively, $c \in (0,1)$, and let $T : \ell_p \to \ell_p$ be defined by*

$$T((x_m)) = (f_1x_1, f_2x_2, \ldots, f_Nx_N, cx_{N+1}, cx_{N+2}, \ldots).$$

Let G be a graph defined by $V(G) = \ell_p$ and

$$E(G) = \{((x_m), (y_m)) \ : \ x_i = y_i \ \text{ for all } i = 1, \ldots, N\}$$

and

$$(\ell_p)_T = \{(x_m) \in \ell_p : ((x_m), T(x_m)) \in E(G)\}.$$

Then $f_1, f_2, \ldots, f_n$ have fixed points if and only if $(\ell_p)_T \neq \emptyset$.

Proof. ($\Rightarrow$) Suppose $f_1, f_2, \ldots, f_N$ have fixed points. Then there exists x_i such that $f_ix_i = x_i$, where $i = 1, 2, \ldots, N$.

$$\begin{aligned} T((x_m)) &= (f_1x_1, f_2x_2, \ldots, f_Nx_N, cx_{N+1}, cx_{N+2}, \ldots) \\ &= (x_1, x_2, \ldots, x_N, cx_{N+1}, cx_{N+2}, \ldots). \end{aligned}$$

Therefore, $((x_m), T(x_m)) \in E(G)$.
($\Leftarrow$) Suppose $(\ell_p)_T \neq \emptyset$. Then there exists $(x_m) \in (\ell_p)_T$ such that $((x_m), T(x_m)) \in E(G)$. This means $f_ix_i = x_i$ for all $i = 1, 2, \ldots, N$. Hence, x_i is a fixed point of f_i for all $i = 1, 2, \ldots, N$. Therefore, $f_1, f_2, \ldots, f_N$ have fixed points.

□

Theorem 3.5. *Let $f_1, f_2, \ldots, f_N$ be Lipschitz functions with constants $k_1, k_2, \ldots, k_N$ respectively, $f_1, f_2, \ldots, f_N$ have some fixed points, and let $c \in (0,1)$. Define the operator $T : \ell_p \to \ell_p$ by*

$$T((x_n)) = (f_1x_1, f_2x_2, \ldots, f_Nx_N, cx_{N+1}, cx_{N+2}, \ldots).$$

Consider a graph G defined by $V(G) = \ell_p$. The edges of the graph G are defined as

$$E(G) = \{((x_n), (y_n)) : x_i = y_i \text{ for all } i = 1, \ldots, N\}.$$

Let $(\ell_p)_T = \{(x_n) \in \ell_p : ((x_n), T(x_n)) \in E(G)\}$. Then, there exists $(w_n) \in (\ell_p)_T$ such that $T|_{[(w_n)]_{\tilde{G}}}$ is a Picard operator.

Proof. We aim to show that there exists $(w_m) \in (\ell_p)_T$ such that $T|_{[(w_m)]_{\tilde{G}}}$ is a Picard operator, using the main theorem 3.1.

Considering ℓ_p with the weak topology as a uniform space, generated by a collection $\mathcal{A}$ of pseudomatrices induced by seminorms, with ℓ_p^* serving as the index set. Thus $\mathcal{A}$ is saturated since the weak topology is Hausdroff and we have:

1. From Lemma 3.4: $(\ell_p)_T \neq \emptyset$, indicating the existence of $(w_m) \in (\ell_p)_T$.
2. From Theorem 2.3: T is a α-G-contraction on ℓ_p with the weak topology.
3. From Theorem 3.3: (ℓ_p, ℓ_p^*, G) satisfies condition $(*)$.

Hence, $T|_{[(w_m)]_{\tilde{G}}}$ is a Picard operator. □

Having established the conditions under which an α-G-contraction is a Picard operator in Theorem 3.1, we now present a second main theorem that considers a slightly different scenario, specifically examining the role of orbital G-continuity. We say that a function $f : X \to X$ is *orbitally G-continuous* if, for every $x, y \in X$ and every sequence $(k_n)_{n\in\mathbb{N}}$ of positive integers such that

$$f^{k_n}x \to y \text{ and } (f^{k_n}x, f^{k_{n+1}}x) \in E(G) \text{ for } n \in \mathbb{N},$$

it follows that $f(f^{k_n}x) \to fy$. In this case, we examine the implications of orbital G-continuity for the Picard operator property of $T|_{[x]_{\tilde{G}}}$ for some $x \in X$.

Theorem 3.6 (The main theorem II)**.** *Let X be nonempty sequentially complete subset of a uniform space generated by a collection of pseudometrics indexed by a set A. Let $T : X \to X$ be an α-G-contraction and orbitally G-continuous, and $X_T = \{x \in X : (x, Tx) \in E(G)\}$. If $x \in X_T$ and $[x]_{\tilde{G}}$ is closed, then $T|_{[x]_{\tilde{G}}}$ is a Picard operator.*

Proof. Let $x \in X_T$ and $y \in [x]_{\tilde{G}}$. Then $(x, Tx) \in E(G)$. By Theorem 2.10, $T|_{[x]_{\tilde{G}}}$ is α-G-contraction. Since Theorem 2.9, $(T^n x)$ and $(T^n y)$ are Cauchy equivalent. By the sequential completeness, $(T^n x)$ and $(T^n y)$ converge to the same point $x_0 \in [x]_{\tilde{G}}$. It remains to show that x_0 is a fixed point of $T|_{[x]_{\tilde{G}}}$. Since $(x, Tx) \in E(G)$, $(T^n x, T^{n+1}x) \in E(G)$. By orbital G-continuity of T, $T(T^n x) \to Tx_0$ and hence $(T^{n+1}x) \to Tx_0$. Since $\mathcal{A}$ is saturated and $[x]_{\tilde{G}}$ is closed, $Tx_0 = x_0 \in [x]_{\tilde{G}}$. □

Having established the conditions for an α-G-contraction to be a Picard operator in the context of orbital G-continuity in Theorem 3.6, we now provide some criteria for maps, specifically Lipschitz functions and the mapping $T : \ell_p \to \ell_p$, that satisfy these requirements in Theorem 3.7. This demonstrates the implications of orbital G-continuity.

Theorem 3.7. *Let $1 < p < \infty$, and let $f_1, f_2, \ldots, f_N$ be Lipschitz functions with $k_1, k_2, \ldots, k_N$ as their respective Lipschitz constants. Let $c \in (0, 1)$ and define a mapping $T : \ell_p \to \ell_p$ by*

$$T((x_m)) = (f_1x_1, f_2x_2, \ldots, f_Nx_N, cx_{N+1}, cx_{N+2}, \ldots).$$

Let G be a graph defined by $V(G) = \ell_p$ and

$$E(G) = \{((x_m), (y_m)) \ : \ x_i = y_i \ \text{ for all } i = 1, \ldots, N\}.$$

Then, T is orbitally G-continuous.

Proof. Let $(x_m), (y_m) \in \ell_p$ and a sequence (k_a) of positive integers where $a \in \mathbb{N}$.
Suppose $T^{k_a}((x_m)) \xrightarrow{w} (y_m)$ as $a \to \infty$ and $(T^{k_a}((x_m)), T^{k_a+1}((x_m))) \in E(G)$ for all $a \in \mathbb{N}$. Since $T^{k_a}((x_m)) \xrightarrow{w} (y_m)$ as $a \to \infty$, for every $(\alpha_m) \in \ell_q$ where $\frac{1}{q} + \frac{1}{p} = 1$, we have

$$\sum_{m=1}^{\infty} \alpha_m \cdot e_m^*(T^{k_a}(x_m)) \to \sum_{m=1}^{\infty} \alpha_m \cdot y_m.$$

Thus,

$$\sum_{m=1}^{\infty} \alpha_m \cdot e_m^*(T^{k_a}(x_m)) = \sum_{i=1}^{N} \alpha_i(f_i^{k_a}x_i) + \sum_{i=N+1}^{\infty} \alpha_i(c^{k_a}x_i).$$

Let $(\beta_m) \in \ell_q$. Thus, we have

$$\sum_{m=1}^{\infty} \beta_m \cdot e_m^*(T(T^{k_a}(x_m))) = \sum_{i=1}^{N} \beta_i(f_i^{k_a+1}x_i) + \sum_{i=N+1}^{\infty} \beta_i(c^{k_a+1}x_i),$$

and

$$\sum_{m=1}^{\infty} \beta_m \cdot e_m^*(T(y_m)) = \sum_{i=1}^{N} \beta_i(f_i y_i) + \sum_{i=N+1}^{\infty} \beta_i(cy_i).$$

Let $(\alpha_m) = (\beta_1, \beta_2, ..., \beta_N, c\beta_{N+1}, c\beta_{N+2}...) \in \ell_q$.
Since $(T^{k_a}((x_m)), T^{k_a+1}((x_m))) \in E(G)$, we have $f_i^{k_a} x_i = f_i^{k_a+1} x_i$ for $i = 1, 2, ..., N$.
Therefore,

$$\sum_{m=1}^{\infty} \alpha_m \cdot e_m^*(T^{k_a}(x_m)) = \sum_{m=1}^{\infty} \beta_m \cdot e_m^*(T(T^{k_a}(x_m))).$$

Since f_i is a continuous function for all $i = 1, 2, ..., N$ and $f_i^{k_a} x_i \to y_i$, we have

$$\alpha_i f_i(f_i^{k_a} x_i) \to \alpha_i f_i(y_i).$$

$$\begin{aligned}\sum_{m=1}^{\infty} \beta_m \cdot e_m^*(T(T^{k_a}(x_m))) &= \sum_{i=1}^{N} \beta_i \cdot e_i^*(T(T^{k_a}(x_m))) + \sum_{i=N+1}^{\infty} \beta_i c(c^{k_a} x_i) \\ &= \sum_{i=1}^{N} \alpha_i f_i(f_i^{k_a} x_i) + \sum_{i=N+1}^{\infty} \alpha_i (c^{k_a} x_i)\end{aligned}$$

and

$$\sum_{m=1}^{\infty} \beta_m \cdot e_m^*(T(y_m)) = \sum_{i=1}^{N} \beta_i \cdot e_i^*(T(y_m)) + \sum_{i=N+1}^{\infty} c\beta_i y_i = \sum_{i=1}^{N} \alpha_i f_i(y_i) + \sum_{i=N+1}^{\infty} \alpha_i y_i.$$

Since

$$\sum_{i=1}^{N} \alpha_i f_i(f_i^{k_a} x_i) + \sum_{i=N+1}^{\infty} \alpha_i (c^{k_a} x_i) \to \sum_{i=1}^{N} \alpha_i f_i(y_i) + \sum_{i=N+1}^{\infty} \alpha_i y_i,$$

we have

$$\sum_{m=1}^{\infty} \beta_m \cdot e_m^*(T(T^{k_a}(x_m))) \to \sum_{m=1}^{\infty} \beta_m \cdot e_m^*(T(y_m)).$$

Therefore, $T(T^{k_a}(x_m)) \xrightarrow{w} (T(y_m))$ as $a \to \infty$. □

In order to satisfy the condition in Theorem 3.6 that $[x]_{\tilde{G}}$ must be a closed set, we present Lemma 3.8 which guarantees that $[(x_m)]_{\tilde{G}}$ is indeed a closed set for the mapping $T : \ell_p \to \ell_p$ in Lemma 3.4.

Lemma 3.8. *Let $1 < p < \infty$ and $f_1, f_2, \ldots, f_N$ be Lipschitz functions with Lipschitz constants $k_1, k_2, \ldots, k_N$ respectively, $c \in (0, 1)$, and let $T : \ell_p \to \ell_p$ be defined by*

$$T((x_m)) = (f_1 x_1, f_2 x_2, \ldots, f_N x_N, cx_{N+1}, cx_{N+2}, \ldots).$$

Let G be a graph defined by $V(G) = \ell_p$ and

$$E(G) = \{((x_m), (y_m)) \ : \ x_i = y_i \ \text{ for all } i = 1, \ldots, N\}.$$

Then $[(x_m)]_{\tilde{G}}$ is a closed set.

Proof. To show that $[(x_m)]_{\tilde{G}}$ is a closed set, let (w_m^α) be a net in $[(x_m)]_{\tilde{G}}$, where $\alpha \in \Lambda$, and $(y_m) \in \ell_p$ with $1 < p < \infty$. Suppose $(w_m^\alpha) \xrightarrow{w} (y_m)$ as $\alpha \to \infty$. Since $(w_m^\alpha) \in [(x_m)]_{\tilde{G}}$, we have $w_i^\alpha = x_i$ for all $i = 1, 2, \ldots, N$. Since $e_i^* \in \ell_p^*$ for all $i = 1, 2, \ldots, N$, and $(w_m^\alpha) \xrightarrow{w} (y_m)$ as $\alpha \to \infty$, we have $e_i^*(w_m^\alpha) \to e_i^*(y_m)$. That is, $w_i^\alpha \to y_i$ for all $i = 1, 2, \ldots, N$. Thus, $(w_i^\alpha) = (x_i)$, where $\alpha \in \Lambda$ and (x_i) is a constant net. That is, $(x_i) \to y_i$ and $(x_i) \to x_i$. Hence, $y_i = x_i$ for all $i = 1, 2, \ldots, N$. Therefore, $(y_m) \in [(x_m)]_{\tilde{G}}$. Consequently, $[(x_m)]_{\tilde{G}}$ is a closed set. □

Now, we present Theorem 3.9 which provides a complete criteria for verifying if a map in ℓ_p satisfies all the conditions outlined in Theorem 3.6.

Theorem 3.9. *Let $1 < p < \infty$, and let $f_1, f_2, \ldots, f_N$ be Lipschitz functions with $k_1, k_2, \ldots, k_N$ as their Lipschitz constants, respectively and $f_1, f_2, \ldots, f_N$ have fixed points. Let $c \in (0,1)$ and $T : \ell_p \to \ell_p$ be defined by*

$$T((x_m)) = (f_1x_1, f_2x_2, \ldots, f_Nx_N, cx_{N+1}, cx_{N+2}, \ldots).$$

Let G be a graph defined by $V(G) = \ell_p$ and

$$E(G) = \{((x_m), (y_m)) \ : \ x_i = y_i \ \text{ for all } i = 1, \ldots, N\}.$$

Let

$$(\ell_p)_T = \{(x_m) \in \ell_p : ((x_m), T(x_m)) \in E(G)\}.$$

Then, there exists $(x_m) \in (\ell_p)_T$ such that $T|_{[(x_m)]_{\tilde{G}}}$ is a Picard operator.

Proof. We aim to show that there exists $(x_m) \in (\ell_p)_T$ such that $T|_{[(x_m)]_{\tilde{G}}}$ is a Picard operator, using the main theorem 3.6.

Considering ℓ_p with the weak topology as a uniform space, generated by a collection $\mathcal{A}$ of pseudomatrices induced by seminorms, with ℓ_p^* serving as the index set. Thus $\mathcal{A}$ is saturated since the weak topology is Hausdroff and we have:

1. From Lemma 3.4: $(\ell_p)_T \neq \emptyset$, indicating the existence of $(x_m) \in (\ell_p)_T$.
2. From Theorem 2.3: T is a α-G-contraction on ℓ_p with the weak topology.
3. From Theorem 3.7: T is orbitally-G-continuous.
4. From Lemma 3.8: $[(x_m)]_{\tilde{G}}$ is a closed set.

Hence, $T|_{[(x_m)]_{\tilde{G}}}$ is a Picard operator. □

Acknowledgment. The authors are grateful to the referees for their careful reading of the manuscript and their useful comments.

References

[1] Vasil G. Angelov, *Fixed point theorems in uniform spaces and applications*, Czechoslovak Mathematical Journal, **37** (1987), 19–33.

[2] Vasil G. Angelov, *J–nonexpansive mappings in uniform spaces and applications*, Bulletin of the Australian Mathematical Society, **43** (1991), 331–339.

[3] P. Chaoha and S. Songsa–ard, *Fixed points in uniform spaces*, Fixed Point Theory and Applications, **2014**(1), 2014.

[4] P. Chaoha and S. Songsa–ard, *Fixed points of functionally lipschitzian maps*, Journal of Nonlinear and Convex Analysis, **15**(4) (2014), 665–679.

[5] Jacek Jachymski, *The contraction principle for mappings on a metric space with a graph*, Proc. Amer. Math. Soc., **136**(4) (2008), 1359–1373.

[6] Jr. Nashed, G. L. and Nashed MZ, *Fixed points and stability for a sum of two operators in locally convex spaces*, Pacific Journal of Mathematics, **39**(3), 1971.

[7] J.J. Nieto and R. Rodríguez–López, *Existence and uniqueness of fixed point in partially ordered sets and applications to ordinary differential equations*, Acta Math. Sin. (Engl. Ser.), **23**(12) (2007), 2205–2212.

[8] A.C.M. Ran and M.C.B. Reurings, *A fixed point theorem in partially ordered sets and some applications to matrix equations*, Proc. Amer. Math. Soc., **132**(5) (2004), 1435–1443.

5. Combinatorics and Graph Theory

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Solving a 4-Colored 5-Cube Puzzle by Graph Theory

Pichaya Kankonsue[1,†], Sayan Panma[2], and Piyashat Sripratak[2,‡]

[1]Graduate Master Degree Program in Applied Mathematics, Department of Mathematics, Faculty of Science
Chiang Mai University, Chiang Mai 50200, Thailand

[2]Department of Mathematics, Faculty of Science
Chiang Mai University, Chiang Mai 50200, Thailand

Abstract

Instant Insanity is a puzzle consisting of four cubes where each face is colored with one of the four colors. The goal of Instant Insanity is to arrange the cubes in a stack so that each color appears exactly once on each of their four long sides (front, back, left, right). In this study, we propose a new puzzle consisting of five cubes where the first four cubes are the original cubes from Instant Insanity, and the last cube is a copy of one of those cubes, called a 4-colored 5-cube puzzle. This puzzle aims to stack the original four cubes, and then attach the last cube to a face of one of the four cubes, creating a structure known as a tower, so that each color appears exactly once on the vertical line and the horizontal line of each side (front, back, left, and right). To solve the puzzle, we apply graph theory to construct graphs that arrange a tower. We show all ways of arranging the cubes to solve the puzzle.

Keywords: Instant Insanity, cube puzzle, directed graph.

2020 MSC: Primary 91A46; Secondary 05C20, 05C30.

1 Introduction

Instant Insanity is a 4-colored 4-cube puzzle introduced around 1900. Parker Brothers named the puzzle as "Instant Insanity" and the puzzle became popular in the 1960's. The puzzle has appeared under the variety of names, for examples, "The Great Tantalizer", "Groceries" and "Katzenjammer", see [5]. Instant Insanity consists of four cubes where each face of each cube is colored yellow (Y), green (G), white (W) or cyan (C). To solve the puzzle, one needs to arrange the cubes in the stack such that each of the four long sides have different colors ([2], [6], [7] and [10]). A cube can be unfolded so that each side is adjacent to at least one of its neighboring faces. An unfolded cube is called a *net*. One possible net of the cube is shown in Figure 1. The nets of cube numbers 1, 2, 3 and 4 for Instant Insanity are given in Figure 2.

†Speaker. ‡Corresponding author.
Email: pichaya_kankonsue@cmu.ac.th (P. Kankonsue), sayan.panma@cmu.ac.th (S. Panma),
piyashat.sripratak@cmu.ac.th (P. Sripratak).

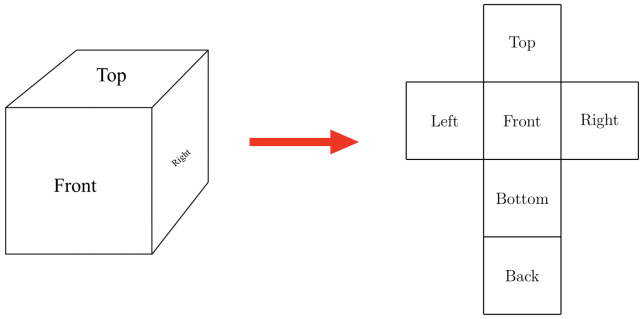


Figure 1: The net of the cube

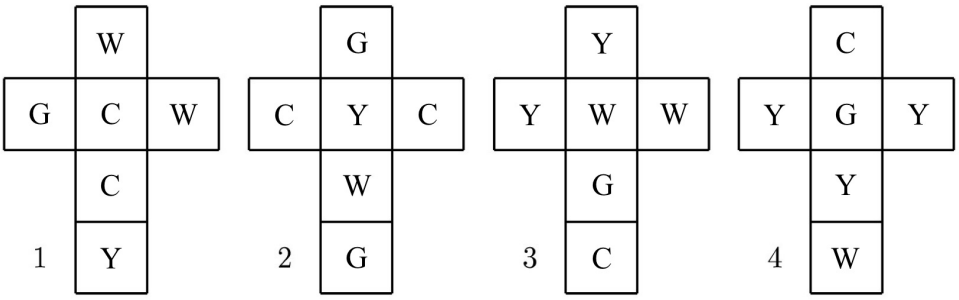


Figure 2: The nets of cubes for Instant Insanity

After Instant Insanity became popular in 1967, Brown [2] showed that Instant Insanity has 82,944 different cases to consider by solving an associated problem in number theory, released in 1968. Schwartz [10] improved Brown's approach. Their methods reduce the possible number of cases from 82,944 to 81. Subsequently, Grecos and Gibberd [6] enhanced Brown's and Schwartz's works and replaced the problem by a simple graphical problem. This method can be used to solve the problems proposed by Brown [2]. In 1969, Deventer [11] introduced the methods of graph theory to solve the puzzle. The colors of the puzzle can be represented by vertices and the pairs of opposite sides can be represented by edges. A solution of the game can be represented by two cycles.

Many researchers have studied the problem in various solids. In 2002, Jebasingh and Simoson [8] extended Instant Insanity puzzle to the five Platonic solids which are cube, octahedron, dodecahedron, icosahedron and tetrahedron. They studied the number of distinct ways of stacking the solids. In 2013, Demaine et al. [4] studied variants of the Instant Insanity puzzle, exploring the relationship between the complexity and the shapes of the pieces. Moreover, they analyzed different types of triangular prism puzzles and rectangular prism puzzles.

Recently, Roldán [9] studied a mathematical model of Instant Insanity by analyzing all possible ways of coloring cubes to create a similar puzzle. They have done this analysis for n cubes and n colors for $n = 4, 5$ and 6, released in 2016. Alsardary et al. [1] introduced a new technique to solve Instant Insanity using the Perl programming language, released in 2016.

The solution for Instant Insanity, obtained from [11], is presented in Table 1.

Table 1: The solution for Instant Insanity from [11]

cube	**front**	**back**	**left**	**right**
1	white	cyan	cyan	yellow
2	green	white	yellow	green
3	yellow	green	white	cyan
4	cyan	yellow	green	white

The following theorem shows that the solution given in Table 1 is unique, see [11].

Theorem 1.1. *[11] There is a unique solution to the Instant Insanity puzzle, up to rotations, flips and permutations of the cubes.*

Our game is a puzzle of five cubes where each face of each cube is colored with one of the four colors (yellow, green, white and cyan). The faces of four from five cubes are colored with four colors as in Figure 2. The last cube is a copy of one of the four cubes. The objective is to stack the four cubes and attach the last cube to one of the four cubes, so that each side (front, back, left and right) of the structure displays each color exactly once, both vertically and horizontally. This structure is called a tower. Two examples of tower are shown in Figure 3.

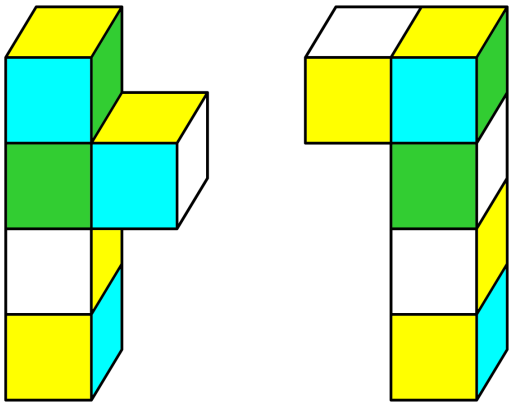


Figure 3: Two examples of tower

2 Preliminaries

In this section, we introduce notations for cubes in the puzzle and provide definitions for graphs. Additionally, we will convert the nets of cubes into graphs to help us obtain the solution and present a theorem for Instant Insanity.

First, we give the notation of the cubes. Let i represent the number of the cube with $i \in \{1, 2, 3, 4\}$. Cube i' is the cube with the same color pattern as cube i. Let $j \in \{1, 2, 3, 4\}$ represent the number of the cube such that cube i' is attached to. The face of cube j such that cube i' is attached to is defined by k, where $k \in \{F, B, L, R\}$ represents the front, the back, the left and the right faces, respectively. Let Y, G, W and C be the colors of the faces on the cubes correspond to yellow, green, white and cyan, respectively.

Then, we provide definitions for graphs. A *directed graph* is a triple $G = (V(G), E(G), p_G)$ with the *vertex set* $V(G)$, the *edge set* $E(G)$ and the *incidence mapping* $p_G : E(G) \to V(G)^2$ defined as $p_G(e) := (o_G(e), t_G(e))$ where $o_G(e)$ represents the *origin* and $t_G(e)$ represents the *tail* of edge e.

Let G be a graph. For $e \in E(G)$ where $o_G(e) \in V(G)$ and $t_G(e) \in V(G)$, let $\bar{e}$ be an edge with $o_G(\bar{e}) = t_G(e)$ and $t_G(\bar{e}) = o_G(e)$.

For $v \in V(G)$, the *out-neighborhood* of v is denoted as $N_G^+(v) := \{t_G(e) | e \in E(G) \text{ and } o_G(e) = v\}$ and the *in-neighborhood* of v is denoted as $N_G^-(v) := \{o_G(e) | e \in E(G) \text{ and } t_G(e) = v\}$.

For definition of the graph of cube i, let $G_i = (V(G_i), E(G_i), p_{G_i})$ be the graph of cube i with $V(G_i) = \{Y, G, W, C\}$ and $E(G_i) = \{e_1, e_2, e_3\}$. Let $\mathbb{G}_i(V(\mathbb{G}_i), E(\mathbb{G}_i), p_{\mathbb{G}_i})$ be the graph of the copy of cube i with $V(\mathbb{G}_i) = V(G_i)$, $E(\mathbb{G}_i) = E(G_i) \cup \{\bar{e} | e \in E(G_i)\} = \{e_1, e_2, e_3, \bar{e}_1 = e_4, \bar{e}_2 = e_5, \bar{e}_3 = e_6\}$ and $p_{\mathbb{G}_i} := E(\mathbb{G}_i) \to V(\mathbb{G}_i)^2$, where

$$p_{\mathbb{G}_i}(e_\alpha) = \begin{cases} p_{G_i}(e_\alpha), & \text{if } \alpha = 1, 2, 3, \\ (t_{G_i}(e_{\alpha-3}), o_{G_i}(e_{\alpha-3})), & \text{if } \alpha = 4, 5, 6. \end{cases}$$

Next, we transform the nets of cubes into graphs. A graph can be constructed from the nets of four cubes (see Figure 3) as shown in Figure 4. The vertices represent the four colors of the puzzle, and each edge represents a pair of opposite faces. Moreover, each cube has three pairs of opposite faces which are front-back, left-right and top-bottom. Hence, there are three edges for each cube ([3], [11] and [12]).

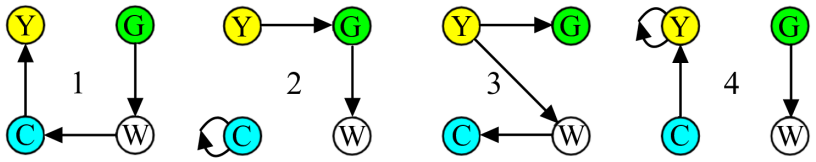


Figure 4: The graphs constructed from the cubes

The four graphs can be combined into a graph whose edges are labeled by that cube number shown in Figure 5. This graph is called the graph for Instant Insanity.

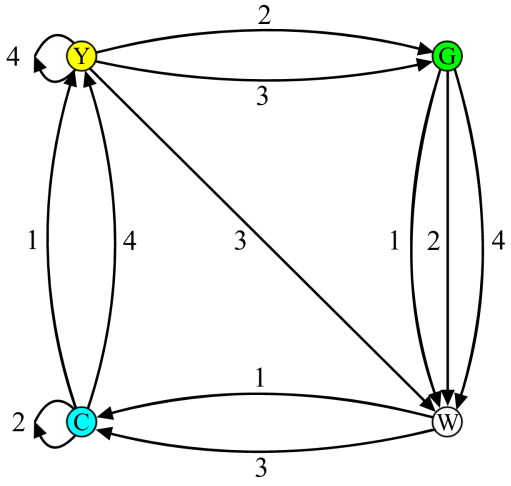


Figure 5: The graph for Instant Insanity

Our goal is to find two directed cycles within the graph in Figure 5. The first cycle will identify the front and the back faces of the stack, referred to as C^F. The second cycle will identify the left and the right faces of the stack, referred to as C^L. Each edge in C^F and C^L is obtained from each cube exactly once, and there are no edges shared between C^F and C^L. Thus, C^F and C^L must be *Hamiltonian cycles*, meaning each cycle visits each vertex exactly once. C^F and C^L are given in Figure 6.

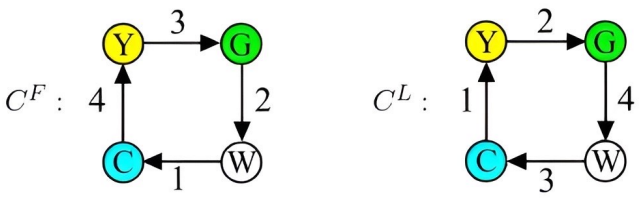


Figure 6: C^F and C^L

There is an edge e with label 1 in C^F such that $p_{C^F}(e) = (W, C)$, and there is an edge e' with label 1 in C^L such that $p_{C^L}(e') = (C, Y)$. Thus, the color of the front face is white, the color of the back face is cyan, the color of the left face is cyan and the color of the right face is yellow. Additionally, there is an edge e with label 2 in C^F such that $p_{C^F}(e) = (G, W)$, and there is an edge e' with label 2 in C^L such that $p_{C^L}(e') = (Y, G)$. Thus, the color of the front face is green, the color of the back face is white, the color of the left face is yellow and the color of the right face is green. Moreover, there is an edge e with label 3 in C^F such that $p_{C^F}(e) = (Y, G)$, and there is an edge e' with label 3 in C^L such that $p_{C^L}(e') = (W, C)$. Thus, the color of the front face is yellow, the color of the back face is green, the color of the left face is white and the color of the right face is green. Finally, there is an edge e with label 4 in C^F such that $p_{C^F}(e) = (C, Y)$, and there is an edge e' with label 4 in C^L such that $p_{C^L}(e') = (G, W)$. Thus, the color of the front face is cyan, the color of the back face is yellow, the color of the left face is green and the color of the right face is white.

The cubes can be arranged according to C^F and C^L. Hence, we can solve the puzzle.

3 Main Results

In this section, we introduce an algorithm for constructing solutions to a 4-colored 5-cube puzzle and provide some relevant theorems to the solutions of the puzzle. Moreover, we give an example of how to find the solutions and show all of them.

Recall that C^F and C^L are graphs that construct the solution to the arrangement of the cube numbers 1, 2, 3 and 4 as in Figure 6.

For every edge, we define a function ϕ on $\{1, 2, 3, 4, 5, 6\}$ to pair each edge e with its opposite-direction edge.

A function ϕ is defined as follows:

$$\phi(\alpha) = \begin{cases} \alpha + 3; & \alpha = 1, 2, 3, \\ \alpha - 3; & \alpha = 4, 5, 6. \end{cases}$$

Let $k \in \{F, B, L, R\}$ and $j \in \{1, 2, 3, 4\}$. The color of face k of cube j is denoted by $k(j)$.

Next, we determine an algorithm that outputs the solutions for our puzzle.

Algorithm for Solving a 4-Colored 5-Cube Puzzle

1. If cube i' is attached to face $k \in \{F, B\}$ of cube j.

   1.1. If $k = F$, then choose $e^F \in N_F = \left\{e \in E(\mathbb{G}_i) | o_{\mathbb{G}_i}(e) = F(j) \text{ and } t_{\mathbb{G}_i}(e) \in N^+_{\mathbb{G}_i}(F(j))\right\}$.
   If $k = B$, then choose $e^F \in N_B = \left\{e \in E(\mathbb{G}_i) | t_{\mathbb{G}_i}(e) = B(j) \text{ and } o_{\mathbb{G}_i}(e) \in N^-_{\mathbb{G}_i}(B(j))\right\}$.

   1.2. Let α be the index such that $e^F = e_\alpha$.
   If $E(\mathbb{G}_i) - \left(\left\{e_\alpha, e_{\phi(\alpha)}\right\} \cup \{e \in E(\mathbb{G}_i) | o_{\mathbb{G}_i}(e) = L(j) \text{ or } t_{\mathbb{G}_i}(e) = R(j)\}\right) \neq \varnothing$,
   then choose $e^L \in E(\mathbb{G}_i) - \left(\left\{e_\alpha, e_{\phi(\alpha)}\right\} \cup \{e \in E(\mathbb{G}_i) | o_{\mathbb{G}_i}(e) = L(j) \text{ or } t_{\mathbb{G}_i}(e) = R(j)\}\right)$.
   Otherwise, e^L does not exist.

1.3. Construct $G^F := C^F + e^F$ and $G^L := C^L + e^L$.
1.4. Construct the solutions to the puzzle from G^F and G^L using "**Generate Solution**".

2. If cube i' is attached to face $k \in \{L, R\}$ of cube j.
2.1. If $k = L$, then choose $e^L \in N_L = \Big\{e \in E(\mathbb{G}_i) | o_{\mathbb{G}_i}(e) = L(j) \text{ and } t_{\mathbb{G}_i}(e) \in N^+_{\mathbb{G}_i}(L(j))\Big\}$.
If $k = R$, then choose $e^L \in N_R = \Big\{e \in E(\mathbb{G}_i) | t_{\mathbb{G}_i}(e) = R(j) \text{ and } o_{\mathbb{G}_i}(e) \in N^-_{\mathbb{G}_i}(R(j))\Big\}$.
2.2. Let α be the index such that $e^L = e_\alpha$.
If $E(\mathbb{G}_i) - \big(\{e_\alpha, e_{\phi(\alpha)}\} \cup \{e \in E(\mathbb{G}_i) | o_{\mathbb{G}_i}(e) = F(j) \text{ or } t_{\mathbb{G}_i}(e) = B(j)\}\big) \neq \varnothing$,
then choose $e^F \in E(\mathbb{G}_i) - \big(\{e_\alpha, e_{\phi(\alpha)}\} \cup \{e \in E(\mathbb{G}_i) | o_{\mathbb{G}_i}(e) = F(j) \text{ or } t_{\mathbb{G}_i}(e) = B(j)\}\big)$.
Otherwise, e^F does not exist.
2.3. Construct $G^F := C^F + e^F$ and $G^L := C^L + e^L$.
2.4. Construct the solutions to the puzzle from G^F and G^L using "**Generate Solution**".

Generate Solution

From graphs $G^F := C^F + e^F$ and $G^L := C^L + e^L$, we can construct the solution of a 4-colored 5-cube Puzzle. This process consists of two steps.

In the first step, we arrange cubes numbered 1, 2, 3 and 4 using C^F and C^L. The solution to the arrangement of cubes numbered 1, 2, 3 and 4 is given in Table 1, see [11].

In the second step, we attach cube i' to face k of cube j in the stack of cubes numbered 1, 2, 3 and 4.

On cube 5, the color on the front face corresponds to the color of vertex $o_{\mathbb{G}_i}(e^F)$, the color on the back face corresponds to the color of vertex $t_{\mathbb{G}_i}(e^F)$, the color on the left face corresponds to the color of vertex $o_{\mathbb{G}_i}(e^L)$, and the color on the right face corresponds to the color of vertex $t_{\mathbb{G}_i}(e^L)$.

Now, we prove the next theorem to show that the graph obtained from the algorithm can be used to generate solutions to the puzzle.

Theorem 3.1. *Any output from the algorithm is a solution to the puzzle.*

Proof. The solution to the arrangement of the cubes number 1, 2, 3 and 4 can be constructed from C^F and C^L. The front (back, left and right) side of each cube has all four different colors as shown in Table 1.

Let the copy of cube i be the cube i', and let $\mathbb{G}_i$ be the graph of cube i'. Suppose that cube i' attaches to face $k \in \{F, B, L, R\}$ of cube j.

Case 1 Consider the graph constructed by attaching cube i' to face $k \in \{F, B\}$ of cube j. If $k = F$, we have $o_{\mathbb{G}_i}(e^F) = F(j)$. Therefore, face F (front face) of cube i' and face F of cube j having the same color implies that the front side of the puzzle have all different colors. Similarly, if $k = B$, we have $t_{\mathbb{G}_i}(e^F) = B(j)$. Therefore, face B (back face) of cube i' and face B of cube j having the same color implies that the back side of the puzzle have all different colors. Let $e^L \in E(\mathbb{G}_i) - \big(\{e_\alpha, e_{\phi(\alpha)}\} \cup \{e \in E(\mathbb{G}_i) | o_{\mathbb{G}_i}(e) = L(j) \text{ or } t_{\mathbb{G}_i}(e) = R(j)\}\big)$. Then, $e^L \notin \{e_\alpha, e_{\phi(\alpha)}\}$. Therefore, face L (left face) and face R (right face) of cube i' are not face F or face B of cube i'. Since $e^L \notin \{e \in E(\mathbb{G}_i) | o_{\mathbb{G}_i}(e) = L(j) \text{ or } t_{\mathbb{G}_i}(e) = R(j)\}$, we have $o_{\mathbb{G}_i}(e^L) \neq L(j)$ and $t_{\mathbb{G}_i}(e^L) \neq R(j)$. Hence, face L of cube i' and face L of cube j have different colors. Furthermore, the colors on face R of cube i' and face R of cube j are different.

Case 2 Consider the graph constructed by attaching cube i' to face $k \in \{L, R\}$ of cube j. If $k = L$, we have $o_{\mathbb{G}_i}(e^L) = L(j)$. Therefore, face L of cube i' and face L of cube j having the same color implies that the left side of the puzzle have all different colors. Similarly, if $k = R$, we have $t_{\mathbb{G}_i}(e^L) = R(j)$. Therefore, face R of cube i' and face R of cube j having the same color implies that the right side of the puzzle have all different colors. Let $e^F \in E(\mathbb{G}_i) - \big(\{e_\alpha, e_{\phi(\alpha)}\} \cup \{e \in E(\mathbb{G}_i) | o_{\mathbb{G}_i}(e) = F(j) \text{ or } t_{\mathbb{G}_i}(e) = B(j)\}\big)$. Then, $e^F \notin$

$\{e_\alpha, e_{\phi(\alpha)}\}$. Therefore, face F and face B of cube i' are not face L or face R of cube i'. Since $e^F \notin \{e \in E(\mathbb{G}_i) | o_{\mathbb{G}_i}(e) = F(j) \text{ or } t_{\mathbb{G}_i}(e) = B(j)\}$, we have $o_{\mathbb{G}_i}(e^F) \neq F(j)$ and $t_{\mathbb{G}_i}(e^F) \neq B(j)$. Hence, face F of cube i' and face F of cube j have different colors. Furthermore, the colors on face B of cube i' and face B of cube j are different.

Therefore, $G^F := C^F + e^F$ and $G^L := C^L + e^L$ can be used to construct a solution to the puzzle as in Table 2. □

Table 2: The solution to a 4-colored 5-cube Puzzle

cube	**front**	**back**	**left**	**right**
1	white	cyan	cyan	yellow
2	green	white	yellow	green
3	yellow	green	white	cyan
4	cyan	yellow	green	white
5	$o_{\mathbb{G}_i}(e^F)$	$t_{\mathbb{G}_i}(e^F)$	$o_{\mathbb{G}_i}(e^L)$	$t_{\mathbb{G}_i}(e^L)$

The following theorem shows that we can generate all solutions of the puzzle using the algorithm.

Theorem 3.2. *If a solution to a 4-colored 5-cube puzzle exists, then it can be constructed using the algorithm.*

Proof. Consider the case when cube i' is attached to sides F or B. If there is a solution to the puzzle, there exist graphs H_1 and H_2 corresponding to the solution. Since graphs C^F and C^L construct the unique solution to Instant Insanity, we get $H_1 := C^F + e_q$ for some $e_q \in E(\mathbb{G}_i)$ and $H_2 := C^L + e_r$ for some $e_r \in E(\mathbb{G}_i)$, where e_q represents the pair of faces F and B of cube i' and e_r represents the pair of faces L and R of cube i'.

If we attach cube i' to face F of cube j and the colors on all faces of side F in the tower must be different, then face F of cube i' and face F of cube j must have the same color. We get $o_{\mathbb{G}_i}(e_q) = F(j)$ and $t_{\mathbb{G}_i}(e_q) \in N^+_{\mathbb{G}_i}(F(j))$. Similarly, if cube i' attached to face B of cube j and all faces on side B of the tower must have different colors, then face B of cube i' and face B of cube j must have the same color. We get $t_{\mathbb{G}_i}(e_q) = B(j)$ and $o_{\mathbb{G}_i}(e_q) \in N^-_{\mathbb{G}_i}(B(j))$. Therefore, edge e_q can be chosen as e^F in step 1.1 of the algorithm.

Moreover, the pair of opposite faces represented by edge e_q is defined as face F and face B of cube i'. Hence, it is not possible to use this pair of faces as face L and face R. We obtain $e_r \neq e_q$ and $e_r \neq e_{\phi(q)}$. In addition, face L of cube i' must not have the same color as face L of cube j, and face R of cube i' must not have the same colors as face R of cube j. We get $o_{\mathbb{G}_i}(e_r) = L(j)$ and $t_{\mathbb{G}_i}(e_r) = R(j)$. Therefore, edge e_r can be chosen as e^L in step 1.2 of the algorithm.

We obtain $H_1 = G^F = C^F + e_q$ and $H_2 = G^L = C^L + e_r$ using the algorithm, and these graphs correspond to the given solution to the puzzle.

Consider the case when cube i' is attached to sides L or R. If there is a solution to the puzzle, there exist graphs T_1 and T_2 corresponding to the solution. Since graphs C^L and C^F construct the unique solution to Instant Insanity, we get $T_1 := C^L + e_r$ for some $e_r \in E(\mathbb{G}_i)$ and $T_2 := C^F + e_q$ for some $e_q \in E(\mathbb{G}_i)$, where e_q represents the pair of faces F and B of cube i' and e_r represents the pair of faces L and R of cube i'.

In the same way, if we attach cube i' to face L of cube j and the colors on all faces of side L in the tower must be different, then face L of cube i' and face L of cube j must have the same color. We get $o_{\mathbb{G}_i}(e_r) = L(j)$ and $t_{\mathbb{G}_i}(e_r) \in N^+_{\mathbb{G}_i}(L(j))$. If cube i' attached to face R of cube j and all faces on side R of the tower must have different colors, then face R of cube i' and face R

of cube j must have the same color. We get $t_{\mathbb{G}_i}(e_r) = R(j)$ and $o_{\mathbb{G}_i}(e_r) \in N^-_{\mathbb{G}_i}(R(j))$. Therefore, edge e_r can be chosen as e^L in step 2.1 of the algorithm.

Moreover, the pair of opposite faces represented by edge e_r is defined as face L and face R of cube i'. Hence, it is not possible to use this pair of faces as face F and face B. We obtain $e_q \neq e_r$ and $e_q \neq e_{\phi(r)}$. In addition, face F of cube i' must not have the same color as face F of cube j, and face B of cube i' must not have the same colors as face B of cube j. We get $o_{\mathbb{G}_i}(e_q) = F(j)$ and $t_{\mathbb{G}_i}(e_q) = B(j)$. Therefore, edge e_q can be chosen as e^F in step 2.2 of the algorithm.

We obtain $T_1 = G^L = C^L + e_r$ and $T_2 = G^F = C^F + e_q$ using the algorithm, and these graphs correspond to the given solution to the puzzle. □

Next, we demonstrate an example of the solutions in the case that cube i' is attached to face k of cube j.

Denote $P(i, j, k)$ 4-colored 5-cube puzzle which cube i' is attached to face k of cube j.

For $P(1, 1, F)$, attach cube $1'$ to face F of cube 1.

Let $E(\mathbb{G}_1) = \{e_1, e_2, e_3, \bar{e}_1, \bar{e}_2, \bar{e}_3\}$, $p_{\mathbb{G}_1}(e_1) = (C, Y), p_{\mathbb{G}_1}(e_2) = (W, C), p_{\mathbb{G}_1}(e_3) = (G, W)$, $p_{\mathbb{G}_1}(\bar{e}_1) = (Y, C), p_{\mathbb{G}_1}(\bar{e}_2) = (C, W)$ and $p_{\mathbb{G}_1}(\bar{e}_3) = (W, G)$.

Since $F(1) = W$, we have $N^+_{\mathbb{G}_1}(F(1)) = \{C, G\}$. Then,

$$N_F = \left\{e \in E(\mathbb{G}_1) | o_{\mathbb{G}_1}(e) = F(1) \text{ and } t_{\mathbb{G}_1}(e) \in N^+_{\mathbb{G}_1}(F(1))\right\} = \{e_2, \bar{e}_3\},$$

that is $|N_F| = 2$.

Therefore, we can choose an edge e^F in 2 ways, namely, $e^F \in \{e_2, \bar{e}_3\}$.

Case 1 $e^F = e_2$: Since $L(1) = C$ and $R(1) = Y$, we have

$$E(\mathbb{G}_1) - (\{e_2, \bar{e}_2\} \cup \{e \in E(\mathbb{G}_1) | o_{\mathbb{G}_1}(e) = C \text{ or } t_{\mathbb{G}_1}(e) = Y\}) = \{\bar{e}_1, e_3, \bar{e}_3\}.$$

Therefore, we can choose an edge e^L in 3 ways, namely, $e^L \in \{\bar{e}_1, e_3, \bar{e}_3\}$.

Case 2 $e^F = \bar{e}_3$: Since $L(1) = C$ and $R(1) = Y$, we have

$$E(\mathbb{G}_1) - (\{\bar{e}_3, e_3\} \cup \{e \in E(\mathbb{G}_1) | o_{\mathbb{G}_1}(e) = C \text{ or } t_{\mathbb{G}_1}(e) = Y\}) = \{\bar{e}_1, e_2\}.$$

Therefore, we can choose also an edge e^L in 2 ways, namely, $e^L \in \{\bar{e}_1, e_2\}$.

Hence, there are five pairs of G^F and G^L for $P(1, 1, F)$:

1. $G^F = C^F + e_2$ and $G^L = C^L + \bar{e}_1$
2. $G^F = C^F + e_2$ and $G^L = C^L + e_3$
3. $G^F = C^F + e_2$ and $G^L = C^L + \bar{e}_3$
4. $G^F = C^F + \bar{e}_3$ and $G^L = C^L + \bar{e}_1$
5. $G^F = C^F + \bar{e}_3$ and $G^L = C^L + e_2$.

In the case $P(1, j, k)$, we obtain all possible edges e^F and e^L from the algorithm. When $k = F$ or $k = B$, the first edge obtained is e^F and the following edge is e^L. However, if $k = L$ or $k = R$, the first edge obtained is e^L an the following edge is e^F. Thus, we get graphs corresponding to the solution of $P(1, j, k)$ which are $C^F + e^F$ and $C^L + e^L$.

Table 3: Edges e^F and e^L for $P(1,j,k)$ obtained from the algorithm

$P(i,j,k)$	**Edge** e^F	**Edge** e^L
$P(1,1,F)$	e_2	$\bar{e}_1, e_3, \bar{e}_3$
	$\bar{e}_3$	$\bar{e}_1, e_2$
$P(1,2,F)$	e_3	$e_1, e_2, \bar{e}_2$
$P(1,3,F)$	$\bar{e}_1$	$\bar{e}_2, e_3$
$P(1,4,F)$	e_1	$e_2, \bar{e}_3$
	$\bar{e}_2$	$e_1, \bar{e}_1, \bar{e}_3$
$P(1,1,B)$	$\bar{e}_1$	$e_2, e_3, \bar{e}_3$
	e_2	$\bar{e}_1, e_3, \bar{e}_3$
$P(1,2,B)$	$\bar{e}_2$	e_1, e_3
	e_3	$e_1, e_2, \bar{e}_2$
$P(1,3,B)$	$\bar{e}_3$	$e_1, \bar{e}_2$
$P(1,4,B)$	e_1	$e_2, \bar{e}_3$

$P(i,j,k)$	**Edge** e^L	**Edge** e^F
$P(1,1,L)$	e_1	$\bar{e}_2, e_3$
	$\bar{e}_2$	e_1, e_3
$P(1,2,L)$	$\bar{e}_1$	$e_2, \bar{e}_3$
$P(1,3,L)$	e_2	e_1, e_3
	$\bar{e}_3$	$e_1, e_2, \bar{e}_2$
$P(1,4,L)$	e_3	$\bar{e}_1, e_2$
$P(1,1,R)$	e_1	$\bar{e}_2, e_3$
$P(1,2,R)$	$\bar{e}_3$	$e_1, \bar{e}_1, e_2$
$P(1,3,R)$	$\bar{e}_1$	$e_2, \bar{e}_2, e_3$
	e_2	e_1, e_3
$P(1,4,R)$	$\bar{e}_2$	$\bar{e}_1, e_3, \bar{e}_3$
	e_3	$\bar{e}_1, e_2$

For example, in the case $P(1,1,F)$, the edges e^F can be e_2 and $\bar{e}_3$. If $e^F = e_2$, the edges e^L can be $\bar{e}_1, e_3$ and $\bar{e}_3$. Likewise, if $e^F = \bar{e}_3$, the edges e^L can be $\bar{e}_1$ and e_2.

Hence, graphs corresponding to the solutions of $P(1,1,F)$ are (i) $C^F + e_2$ and $C^L + \bar{e}_1$, (ii) $C^F + e_2$ and $C^L + e_3$, (iii) $C^F + e_2$ and $C^L + \bar{e}_3$, (iv) $C^F + \bar{e}_3$ and $C^L + \bar{e}_1$ and (v) $C^F + \bar{e}_3$ and $C^L + e_2$.

In the same way, we obtain pairs of e^F and e^L for $P(2,j,k)$, $P(3,j,k)$, and $P(4,j,k)$.

For $P(2,j,k)$, let $E(\mathbb{G}_2) = \{e_1, e_2, e_3, \bar{e}_1, \bar{e}_2, \bar{e}_3\}$, $p_{\mathbb{G}_2}(e_1) = (C,C), p_{\mathbb{G}_2}(e_2) = (Y,G)$, $p_{\mathbb{G}_2}(e_3) = (G,W), p_{\mathbb{G}_2}(\bar{e}_1) = (C,C), p_{\mathbb{G}_2}(\bar{e}_2) = (G,Y)$ and $p_{\mathbb{G}_2}(\bar{e}_3) = (W,G)$.

Table 4: Edges e^F and e^L for $P(2,j,k)$ obtained from the algorithm

$P(i,j,k)$	**Edge** e^F	**Edge** e^L
$P(2,1,F)$	$\bar{e}_3$	e_2
$P(2,2,F)$	$\bar{e}_2$	$e_1, \bar{e}_1, e_3$
	e_3	$e_1, \bar{e}_1, \bar{e}_2$
$P(2,3,F)$	e_2	e_3
$P(2,4,F)$	e_1	$e_2, \bar{e}_3$
	$\bar{e}_1$	$e_2, \bar{e}_3$
$P(2,1,B)$	e_1	$e_2, e_3, \bar{e}_3$
	$\bar{e}_1$	$e_2, e_3, \bar{e}_3$
$P(2,2,B)$	e_3	$e_1, \bar{e}_1, \bar{e}_2$
$P(2,3,B)$	e_2	e_3
	$\bar{e}_3$	$e_2, \bar{e}_2$
$P(2,4,B)$	$\bar{e}_2$	$e_1, \bar{e}_1, \bar{e}_3$

$P(i,j,k)$	**Edge** e^L	**Edge** e^F
$P(2,1,L)$	e_1	$e_2, \bar{e}_2, e_3$
	$\bar{e}_1$	$e_2, \bar{e}_2, e_3$
$P(2,2,L)$	e_2	$e_1, \bar{e}_1, \bar{e}_3$
$P(2,3,L)$	$\bar{e}_3$	$e_1, \bar{e}_1, \bar{e}_2$
$P(2,4,L)$	$\bar{e}_2$	$e_3, \bar{e}_3$
	e_3	e_2
$P(2,1,R)$	$\bar{e}_2$	e_3
$P(2,2,R)$	e_2	$e_1, \bar{e}_1, \bar{e}_3$
	$\bar{e}_3$	$e_1, \bar{e}_1, e_2$
$P(2,3,R)$	e_1	$\bar{e}_2, e_3$
	$\bar{e}_1$	$\bar{e}_2, e_3$
$P(2,4,R)$	e_3	e_2

For $P(3,j,k)$, let $E(\mathbb{G}_3)=\{e_1,e_2,e_3,\bar{e}_1,\bar{e}_2,\bar{e}_3\}$, $p_{\mathbb{G}_3}(e_1)=(Y,G)$, $p_{\mathbb{G}_3}(e_2)=(Y,W)$, $p_{\mathbb{G}_3}(e_3)=(W,C)$, $p_{\mathbb{G}_3}(\bar{e}_1)=(G,Y)$, $p_{\mathbb{G}_3}(\bar{e}_2)=(W,Y)$ and $p_{\mathbb{G}_3}(\bar{e}_3)=(C,W)$.

Table 5: Edges e^F and e^L for $P(3,j,k)$ obtained from the algorithm

$P(i,j,k)$	**Edge** e^F	**Edge** e^L
$P(3,1,F)$	$\bar{e}_2$	e_1,e_3
	e_3	e_1,e_2
$P(3,2,F)$	$\bar{e}_1$	$\bar{e}_2,e_3,\bar{e}_3$
$P(3,3,F)$	e_1	$e_2,\bar{e}_3$
	e_2	$e_1,\bar{e}_1,\bar{e}_3$
$P(3,4,F)$	$\bar{e}_3$	$e_1,\bar{e}_2$
$P(3,1,B)$	e_3	e_1,e_2
$P(3,2,B)$	e_2	$\bar{e}_1,e_3,\bar{e}_3$
	$\bar{e}_3$	$\bar{e}_1,\bar{e}_2$
$P(3,3,B)$	e_1	$e_2,\bar{e}_3$
$P(3,4,B)$	$\bar{e}_1$	$\bar{e}_2,e_3$
	$\bar{e}_2$	e_1,e_3

$P(i,j,k)$	**Edge** e^L	**Edge** e^F
$P(3,1,L)$	$\bar{e}_3$	$e_1,\bar{e}_1,e_2$
$P(3,2,L)$	e_1	$\bar{e}_2,e_3$
	e_2	e_1,e_3
$P(3,3,L)$	$\bar{e}_2$	$\bar{e}_1,e_3,\bar{e}_3$
	e_3	$\bar{e}_1,\bar{e}_2$
$P(3,4,L)$	$\bar{e}_1$	e_2,e_3
$P(3,1,R)$	$\bar{e}_1$	$e_2,\bar{e}_3$
	$\bar{e}_2$	$e_1,\bar{e}_1,\bar{e}_3$
$P(3,2,R)$	e_1	$\bar{e}_2,e_3$
$P(3,3,R)$	e_3	$\bar{e}_1,\bar{e}_2$
$P(3,4,R)$	e_2	e_1,e_3
	$\bar{e}_3$	e_1,e_2

For $P(4,j,k)$, let $E(\mathbb{G}_4)=\{e_1,e_2,e_3,\bar{e}_1,\bar{e}_2,\bar{e}_3\}$, $p_{\mathbb{G}_4}(e_1)=(Y,Y)$, $p_{\mathbb{G}_4}(e_2)=(C,Y)$, $p_{\mathbb{G}_4}(e_3)=(G,W)$, $p_{\mathbb{G}_4}(\bar{e}_1)=(Y,Y)$, $p_{\mathbb{G}_4}(\bar{e}_2)=(Y,C)$ and $p_{\mathbb{G}_4}(\bar{e}_3)=(W,G)$.

Table 6: Edges e^F and e^L for $P(4,j,k)$ obtained from the algorithm

$P(i,j,k)$	**Edge** e^F	**Edge** e^L
$P(4,1,F)$	$\bar{e}_3$	$\bar{e}_2$
$P(4,2,F)$	e_3	e_2
$P(4,3,F)$	e_1	e_2,e_3
	$\bar{e}_1$	e_2,e_3
	$\bar{e}_2$	$e_1,\bar{e}_1,e_3$
$P(4,4,F)$	e_2	$e_1,\bar{e}_1,\bar{e}_3$
$P(4,1,B)$	$\bar{e}_2$	$e_3,\bar{e}_3$
$P(4,2,B)$	e_3	e_2
$P(4,3,B)$	$\bar{e}_3$	$e_1,\bar{e}_1,e_2$
$P(4,4,B)$	e_1	$e_2,\bar{e}_2,\bar{e}_3$
	$\bar{e}_1$	$e_2,\bar{e}_2,\bar{e}_3$
	e_2	$e_1,\bar{e}_1,\bar{e}_3$

$P(i,j,k)$	**Edge** e^L	**Edge** e^F
$P(4,1,L)$	e_2	$e_1,\bar{e}_1,e_3$
$P(4,2,L)$	e_1	$e_2,\bar{e}_2,\bar{e}_3$
	$\bar{e}_1$	$e_2,\bar{e}_2,\bar{e}_3$
	$\bar{e}_2$	$e_1,\bar{e}_1,\bar{e}_3$
$P(4,3,L)$	$\bar{e}_3$	e_2
$P(4,4,L)$	e_3	$\bar{e}_2$
$P(4,1,R)$	e_1	e_2,e_3
	$\bar{e}_1$	e_2,e_3
	e_2	$e_1,\bar{e}_1,e_3$
$P(4,2,R)$	$\bar{e}_3$	$e_1,\bar{e}_1,e_2,\bar{e}_2$
$P(4,3,R)$	$\bar{e}_2$	e_3
$P(4,4,R)$	e_3	$\bar{e}_2$

In summary, there are 220 pairs of G^F and G^L for $P(i,j,k)$. Thus, the 4-colored 5-cube puzzle has 220 solutions.

It is also interesting to consider the puzzle that attaches $n\in\{1,2,3,4\}$ copies of the original cubes to the stack from Instant Insanity, called a 4-colored $(4+n)$-cube puzzle. We can use a similar idea to develop an algorithm for the puzzle and find the number of the solutions.

References

[1] S. Y. Alsardary, H. J. Kim, and J. George, *A new technique to solve the Instant Insanity problem*, Journal of Advances in Mathematics, **11**(10) (2016), 5766–5773.

[2] T. A. Brown, *A note on "Instant Insanity"*, Mathematics Magazine, **41**(4) (1968), 167–169.

[3] G. Chartrand, L. Lesniak, and P. Zhang, *Graphs & digraphs*, 5th ed., CRC press, New York, 2010.

[4] E. D. Demaine, M. L. Demaine, S. Eisenstat, T. D. Morgan, and R. Uehara, *Variations on instant insanity*, ch. Space-Efficient Data Structures, Streams, and Algorithms, pp. 33–47, Springer-Verlag, Berlin, Heidelberg, 2013.

[5] E. D. Demaine, M. L. Demaine, and T. Rodgers (Eds.), *A Lifetime of Puzzles*. CRC Press, Wellesley, 2008.

[6] A. P. Grecos, and R. W. Gibberd, *A diagrammatic solution to "Instant Insanity" problem*, Mathematics Magazine, **44**(3) (1971), 119–124.

[7] F. Harary, *On "The Tantalizer" and "Instant Insanity"*, Historia Mathematica, **4** (1977), 205–206.

[8] A. Jebasingh, and A. Simoson, *Platonic solid insanity*, Congressus Numerantium, **154** (2002), 101–112.

[9] É. B. R. Roa, *The Mutando of Insanity*, G4G12 Exchange Book Vol. 2, pp. 135 – 144, 2016.

[10] B. L. Schwartz, *An improved solution to "Instant Insanity"*, Mathematics Magazine, **43**(1) (1970), 20–23.

[11] J. V. Deventer, *Graph theory and "Instant Insanity"*, Proceedings of the Conference held at Western Michigan University, Kalamazoo/MI., October 31–November 2, 1968, pp. 283–286.

[12] D. B. West, *Introduction to Graph Theory*, 2nd ed., Prentice hall, Upper Saddle River, 2001.

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

ปัญหาการพับแถบแสตมป์ n ดวง เมื่อ $n = 2, 3, 4, 5, 6$

ศิริญญา โปร่งจิตร์[1,†,‡] ประกายแสง โคตรมิตร[1] ทศพร สายเสมา[1] และ วัชราภรณ์ อดทน[1]

[1]ภาควิชาวิทยาศาสตร์ทั่วไป คณะวิทยาศาสตร์และวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์
วิทยาเขตเฉลิมพระเกียรติ จังหวัดสกลนคร 47000

บทคัดย่อ

ในงานวิจัยนี้ เราได้แสดงคำตอบของข้อปัญหาในการพับแถบแสตมป์ n ดวง เมื่อ $n = 2, 3, 4, 5, 6$ คำตอบของข้อปัญหาซึ่งเราเรียกว่า *n-ทบ* อยู่ในรูปของการเรียงสับเปลี่ยนของสมาชิกในเซต $\{1, 2, \ldots, n\}$ เราได้ศึกษาความสัมพันธ์ระหว่าง *n-ทบ* กับ *การเรียงสับเปลี่ยนรอยพับภูเขาหุบเขา* ได้นิยาม *แพทเทิร์นของ n-ทบ* เมื่อ n เป็นจำนวนนับที่มากกว่า 1 และได้เสนอทฤษฎีบทซึ่งเป็นผลลัพธ์จากการศึกษาความสัมพันธ์ดังกล่าว

คำสำคัญ: การเรียงสับเปลี่ยน, ปัญหาการพับแสตมป์, โอริกามิ

2020 MSC: ปฐมภูมิ 05 ทุติยภูมิ 05A05

1 บทนำ

โอริกามิเป็นศิลปะรูปแบบหนึ่งซึ่งมีรากฐานอยู่ในทวีปเอเชียมานานกว่า 1000 ปี ในภาษาญี่ปุ่นคำว่า *โอริกามิ* มีความหมายตามตัวอักษรว่า *พับ, กระดาษ* ความสนใจในคณิตศาสตร์ที่ซ่อนในโอริกามิเพิ่งเกิดขึ้นในช่วงศตวรรษที่ผ่านมาเท่านั้น มีการเผยแพร่บทความวิจัยทางคณิตศาสตร์ที่เกี่ยวข้องกับโอริกามิหลายบทความ เช่น บทความเรื่อง Folding a Strip of Stamps [3] เมื่อ ปี ค.ศ. 1968 เรื่อง On the Mathematics of Flat Origamis [1] เมื่อ ปี ค.ศ. 1994 เรื่อง A method for Designing Crease Patterns for Flat-Foldable Origami with Numerical Optimization [4] เมื่อ ปี ค.ศ. 2011 และเรื่อง An Application of A Theorem of Alternative to Origami เมื่อปี ค.ศ. 2017 [2] บทความวิจัยเรื่องนี้ เริ่มต้นมาจากข้อปัญหาในหนังสือ How to Fold It [5] ซึ่งท้าทายให้หาจำนวนวิธีในการพับแถบแสตมป์ 4 ดวง ที่ตราหมายเลข $1, 2, 3, 4$ ไว้บนแสตมป์โดยเรียงลำดับจากแสตมป์ดวงซ้ายไปดวงขวา และตราหมายเลขไว้เพียงด้านเดียว กติกามีว่า เมื่อพับแถบแสตมป์

†ผู้นำเสนอ ‡ผู้แต่งหลัก
อีเมล: sirinya.pr@ku.th (ศิริญญา โปร่งจิตร์), prakaisang.k@ku.th (ประกายแสง โคตรมิตร),
todsaporn.sa@ku.th (ทศพร สายเสมา), watcharaporn.od@ku.th (วัชราภรณ์ อดทน).

ตามแนวรอยปรุแล้วต้องวางทบแสตมป์โดยให้หมายเลข 1 หงายขึ้น แล้วอ่านหมายเลขจากทบแสตมป์โดยเรียงลำดับจากชั้นบนลงมาชั้นล่าง แน่นอนว่าลำดับหมายเลขที่อ่านได้นี้ต้องเป็นการเรียงสับเปลี่ยนอันใดอันหนึ่งในเซต $\{1,2,3,4\}$ แต่จริงหรือไม่ที่ทุก ๆ การเรียงสับเปลี่ยนในเซต $\{1,2,3,4\}$ จะทำให้เราสามารถพับเพื่อให้ได้ทบแสตมป์ที่สอดคล้องกับการเรียงสับเปลี่ยนนั้นได้ ?

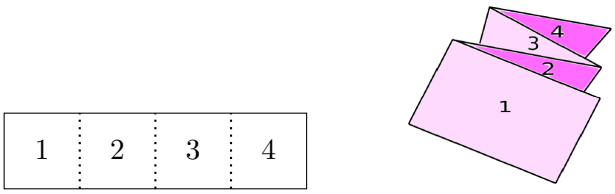
ในบทความนี้ เราได้แสดงคำตอบของข้อปัญหาดังกล่าว รวมถึงได้แสดงคำตอบจากการขยายข้อปัญหาไปสู่กรณีแถบแสตมป์ n ดวง เมื่อ $n = 2,3,4,5,6$ จากความพยายามที่จะหาคำตอบของข้อปัญหาเมื่อ n เป็นจำนวนนับใด ๆ ได้นำไปสู่การนิยาม *การเรียงสับเปลี่ยนรอยพับภูเขาหุบเขา* เราได้ศึกษาความสัมพันธ์ระหว่าง *การเรียงสับเปลี่ยนที่เป็นคำตอบของข้อปัญหาดังกล่าว* ซึ่งต่อไปเราจะเรียกว่า n-ทบ กับ *การเรียงสับเปลี่ยนรอยพับภูเขาหุบเขา* เราได้เสนอทฤษฎีบทที่เกิดจากการอนุมานผลการศึกษาความสัมพันธ์ดังกล่าว และเพื่อพิสูจน์ทฤษฎีบทเราได้นิยาม *แบบจำลองสองมิติสำหรับ n-ทบ แพทเทิร์นของ n-ทบ* และ *ลำดับเรียงเส้นของ n-ทบ* ด้วย

2 ปัญหาการพับแสตมป์

เริ่มต้นด้วยข้อปัญหาในหนังสือ How to Fold It [5] ดังนี้

ข้อปัญหา 2.1. ถ้าคุณมีแถบแสตมป์ 4 ดวง ซึ่งมีหมายเลข $1,2,3$ และ 4 ตราไว้บนแสตมป์โดยเรียงลำดับจากซ้ายไปขวา (ภาพที่ 1 ซ้าย) เมื่อเราพับแถบแสตมป์ตามรอยปรุซ้อนกันเป็นทบแสตมป์แล้ว เรากำหนดทิศทางของทบแสตมป์ โดยให้แสตมป์หมายเลข 1 หงายขึ้น (ไม่ว่าแสตมป์หมายเลข 1 จะอยู่ที่ชั้นใด) จากนั้นให้อ่านหมายเลขที่ตราไว้บนแสตมป์โดยเรียงลำดับจากชั้นบนลงมาชั้นล่าง เช่น ภาพที่ 1 ขวา ได้แสดงทบแสตมป์ที่สัมพันธ์กับการเรียงสับเปลี่ยน 1234

คำถามคือ จากการเรียงสับเปลี่ยนของสมาชิกในเซต $\{1,2,3,4\}$ ซึ่งมีทั้งหมด $4! = 24$ แบบนั้น จริงหรือไม่ที่ทุก ๆ การเรียงสับเปลี่ยนสามารถเกิดทบแสตมป์ได้ ?



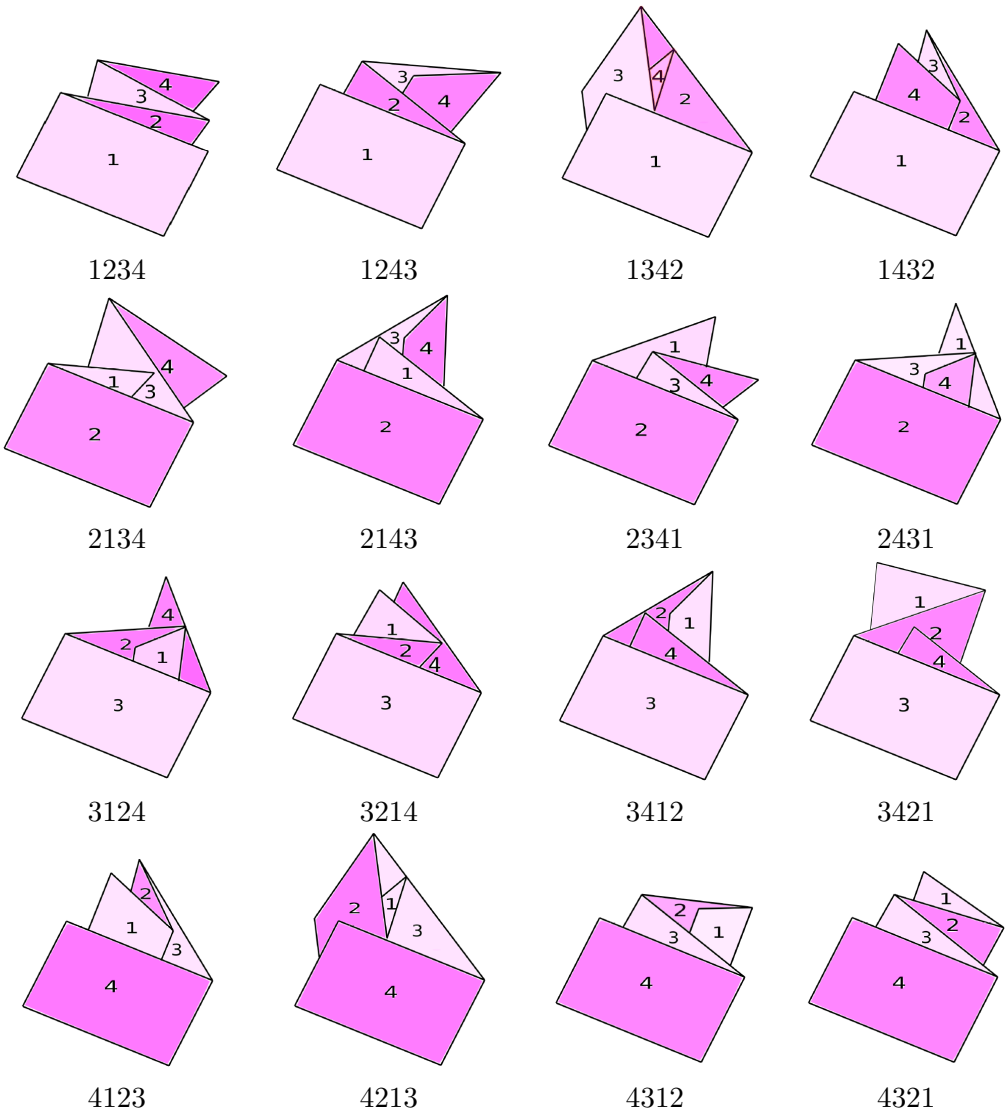
ภาพที่ 1: ทบแสตมป์สี่ดวงกับการเรียงสับเปลี่ยน 1234

หมายเหตุ 2.2. ในภาพที่ 1 ขวา ซึ่งเป็นตัวอย่างของทบแสตมป์ที่พับได้ เราถือว่าแสตมป์ชั้นบนสุดของทบแสตมป์คือรูปสี่เหลี่ยมด้านหน้า ส่วนที่ระบายสีอ่อนคือด้านหน้าของแสตมป์ซึ่งตราหมายเลข ส่วนที่ระบายสีเข้มคือด้านหลังของแสตมป์ซึ่งไม่ตราหมายเลข แต่ด้วยข้อจำกัดของรูปเราจำเป็นต้องใส่หมายเลขไว้ตรงส่วนที่เป็นสีเข้มด้วย

2.1 4-ทบ

เราได้ทดลองพับแถบแสตมป์และพบว่า มีการเรียงสับเปลี่ยนของสมาชิกในเซต $\{1,2,3,4\}$ เพียง 16 แบบจากการเรียงสับเปลี่ยนทั้งหมด 24 แบบ ที่สอดคล้องกับทบแสตมป์ทั้งหมดที่พับได้ และได้แสดงคำตอบทั้งหมดนี้ในภาพที่ 2

เราเรียกแต่ละการเรียงสับเปลี่ยนที่เกิดทบแสตมป์ในรูปที่ 2 ว่า *4-ทบ* เช่น 2134 เป็น *4-ทบ* แต่ 2314 ไม่เป็น *4-ทบ* และถ้า n เป็นจำนวนนับที่มากกว่า 1 แล้ว *n-ทบ* หมายถึง การเรียงสับเปลี่ยน π บนเซต $\{1,2,\ldots,n\}$ ที่สามารถพับแถบแสตมป์ n ดวงให้สอดคล้องกับ π ได้



ภาพที่ 2: 4-*ทบ* ทั้งหมด 16 แบบ

2.2 n-ทบ เมื่อ $n = 2, 3, 4, 5, 6$

เราได้ขยายข้อปัญหา 2.1 จากแถบแสตมป์ 4 ดวง เป็นแถบแสตมป์ n ดวง เมื่อ $n = 2, 3, 4, 5, 6$ และพบว่าจำนวนของการเรียงสับเปลี่ยนของสมาชิกในเซต $\{1, 2, \ldots, n\}$ ที่เป็น n-ทบ คือ $2, 6, 16, 50, 144$ เมื่อ $n = 2, 3, 4, 5, 6$ ตามลำดับ (จะกล่าวโดยละเอียดในหัวข้อที่ 3) ตรงนี้จะสังเกตได้ว่า เมื่อแถบแสตมป์ไม่เกิน 3 ดวง ทุก ๆ การเรียงสับเปลี่ยนสามารถเกิดทบแสตมป์ได้ อย่างไรก็ตาม ตัวเลข $2, 6, 16, 50, 144$ ที่เราค้นพบจากการทดลองพับจริงตรงกันกับข้อมูลในบทความวิจัยของ Koehler [3] หรืออาจกล่าวได้ว่า จากบทความวิจัยของ Koehler ทำให้เราสามารถสรุปคำตอบของข้อปัญหาดังกล่าวได้โดยอาศัยการพับด้วยมือเท่านั้น

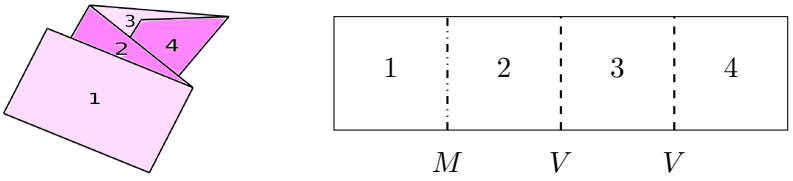
การพยายามที่จะหาคำตอบของข้อปัญหาเมื่อ n เป็นจำนวนนับใด ๆ ได้นำไปสู่การนิยาม *การเรียงสับเปลี่ยนรอยพับภูเขาหุบเขา* และศึกษาความสัมพันธ์ระหว่าง n-ทบ กับ การเรียงสับเปลี่ยนรอยพับภูเขาหุบเขา ดังหัวข้อถัดไป

3 การเรียงสับเปลี่ยนรอยพับภูเขาหุบเขา

พิจารณา 4-ทบ 1243 ดังภาพที่ 3 ซ้าย ถ้าเราคลี่ทบแสตมป์ออกโดยหงายด้านที่ตราหมายเลขขึ้น เราเรียกรอยพับระหว่างแสตมป์ดวงที่ 1 กับดวงที่ 2 ว่า *รอยพับภูเขา* เรียกรอยพับระหว่างแสตมป์ดวงที่ 2 กับดวงที่ 3 และรอยพับระหว่างแสตมป์ดวงที่ 3 กับดวงที่ 4 ว่า *รอยพับหุบเขา*

นิยาม 3.1. กำหนดให้ M แทนรอยพับภูเขา และ V แทนรอยพับหุบเขา สำหรับจำนวนนับ n ใด ๆ *การเรียงสับเปลี่ยนรอยพับภูเขาหุบเขา* คือการเรียงสับเปลี่ยน n สิ่งซ้ำได้ในเซต $\{M, V\}$

จาก 4-ทบ 1243 ซึ่งคลี่แล้ว ถ้าเราลากเส้นตรงตามแนวรอยปรุของแถบแสตมป์บนด้านที่ตราหมายเลข โดยให้เส้นประจุดแทนรอยพับภูเขา (M) และเส้นประแทนรอยพับหุบเขา (V) จะได้ภาพที่ 3 ขวา ในกรณีนี้เรากล่าวว่า 1243 *สัมพันธ์กับการเรียงสับเปลี่ยนรอยพับภูเขาหุบเขา* MVV หรือกล่าวโดยย่อว่า 1243 *สัมพันธ์กับ* MVV



ภาพที่ 3: 1243 สัมพันธ์กับ MVV

เราศึกษาความสัมพันธ์ระหว่าง *การเรียงสับเปลี่ยนที่เป็น n-ทบ* กับ *การเรียงสับเปลี่ยนรอยพับภูเขาหุบเขา* เมื่อ $n = 2, 3, 4, 5, 6$ ได้ผลดังตารางที่ 1 และตารางที่ 2 ต่อไปนี้

ตารางที่ 1: n-*ทบ* กับการเรียงสับเปลี่ยนรอยพับภูเขาหุบเขา เมื่อ $n = 2, 3, 4, 5$

n	n-ทบ	การเรียงสับเปลี่ยนรอยพับภูเขาหุบเขา
2	12	M
	21	V
3	132, 312	MM
	231, 213	VV
	123	MV
	321	VM
4	3124, 3412, 1342	MMM
	4213, 2143, 2431	VVV
	4312, 1432	MMV
	2134, 2341	VVM
	1234	MVM
	4321	VMV
	3214, 3421	VMM
	1243, 4123	MVV
5	53124, 35412, 31254, 13542	$MMMM$
	42135, 21453, 45213, 24531	$VVVV$
	34512, 34152, 13452, 31245	$MMMV$
	21543, 25143, 25431, 54213	$VVVM$
	54312, 51432, 15432	$MMVM$
	21345, 23415, 23451	$VVMV$
	12354, 51234, 15234	$MVMM$
	45321, 43215, 43251	$VMVV$
	53214, 35421, 32154, 32514	$VMMM$
	41235, 12453, 45123, 41523	$MVVV$
	43125, 45312, 14532, 14325	$MMVV$
	52134, 21354, 23541, 52341	$VVMM$
	12345	$MVMV$
	54321	$VMVM$
	54123, 12543	$MVVM$
	32145, 34521	$VMMV$

ตารางที่ 2: 6-*ทบ* กับการเรียงสับเปลี่ยนรอยพับภูเขาหุบเขา

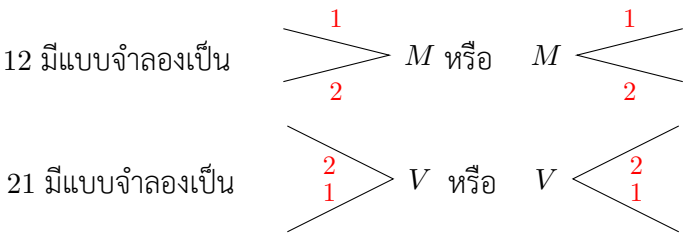
6-ทบ	การเรียงสับเปลี่ยน รอยพับภูเขาหุบเขา
135642, 312564, 356412, 531246, 563124 246531, 465213, 214653, 642135, 421365	$MMMMM$ $VVVVV$
365412, 136542, 653124, 361254, 312654 214563, 245631, 421356, 452163, 456213	$MMMMV$ $VVVVM$
345612, 341562, 134562, 345126, 312456 216543, 265143, 265431, 621543, 654213	$MMMVM$ $VVVMV$
564312, 543612, 561432, 156432, 154362, 543126, 514326 213465, 216345, 234165, 234651, 263451, 621345, 623415	$MMVMM$ $VVMVV$
561234, 156234, 512634, 123564, 512346 432165, 432651, 436215, 465312, 643215	$MVMMM$ $VMVVV$
321564, 325614, 356412, 532146, 563214 465123, 416523, 124653, 641235, 412365	$VMMMM$ $MVVVV$
634512, 163452, 341652, 134562, 631245, 312465 215436, 254361, 256143, 256431, 542136, 564213	$MMMVV$ $VVVMM$
456312, 145632, 431256, 143256 213654, 236541, 652134, 652341	$MMVVM$ $VVMMV$
564123, 125643, 541236, 125436 321465, 346521, 632145, 634521	$MVVMM$ $VMMVV$
213564, 235641, 521346, 521634, 523416, 562134, 562341 465312, 146532, 643125, 436125, 614325, 432165, 143265	$VVMMM$ $MMVVV$
654312, 651432, 165432 213456, 234156, 234561	$MMVMV$ $VVMVM$
651234, 165234, 123654 432156, 432561, 456321	$MVMMV$ $VMVVM$
321654, 326514, 362154, 365421, 653214 456123, 415623, 451263, 124563, 412356	$VMMMV$ $MVVVM$
321465, 346521, 632145, 634521 564123, 125643, 541236, 125436	$VMMVM$ $MVVMV$
215643, 543216, 543621, 564321 346512, 612345, 126345, 123465	$VMVMM$ $MVMMV$
123456 654321	$MVMVM$ $VMVMV$

4 แพทเทิร์นของ n-ทบ เมื่อ $n \in \mathbb{N}$ โดยที่ $n \geq 2$

เพื่อวิเคราะห์ความสัมพันธ์ระหว่าง n-ทบ กับ การเรียงสับเปลี่ยนรอยพับภูเขาหุบเขา เราได้สร้างแบบจำลอง 2 มิติสำหรับแต่ละ n-ทบ ขึ้นมา แล้วปรับแบบจำลองดังกล่าวให้เป็นแพทเทิร์นของ n-ทบ

4.1 แบบจำลองสองมิติ

เริ่มต้นที่ 2-ทบ ซึ่งมีเพียง 12 กับ 21 จากทบแสตมป์ใน 3 มิติ ที่วางโดยให้หมายเลข 1 หงายขึ้น เราจำลองเป็นลายเส้นใน 2 มิติได้ดังภาพที่ 4 จุดที่เส้นสองเส้นชนกันเป็นมุมแหลมคือรอยพับ เส้นแต่ละเส้นคือแสตมป์แต่ละดวงซึ่งหมายเลขที่กำกับคือหมายเลขที่ตราบนแสตมป์ การกำกับหมายเลขมีทิศทางซึ่งสอดคล้องกับทบแสตมป์ใน 3 มิติ กล่าวคือ ถ้าเป็นรอยพับภูเขาเรากำกับหมายเลขไว้นอกมุม แต่ถ้าเป็นรอยพับหุบเขาเรากำกับหมายเลขไว้ในมุม หรืออีกนัยหนึ่ง เส้นตรงแต่ละเส้นมีสองด้าน ด้านที่ติดหมายเลขแทนด้านหน้าของแสตมป์ (ด้านที่ตราหมายเลข) ส่วนด้านที่ไม่ติดหมายเลขแทนด้านหลังของแสตมป์ เนื่องจากต้องวางทบแสตมป์โดยให้แสตมป์หมายเลข 1 หงายขึ้นเสมอ ดังนั้นในแบบจำลองนี้หมายเลข 1 ต้องอยู่เหนือเส้นของมันเสมอ ส่วนหมายเลขอื่น ๆ อาจจะอยู่เหนือเส้นหรืออยู่ใต้เส้นของมันก็ได้ ในภาพเรากำกับอักษร M และ V ไว้สำหรับรอยพับภูเขาและรอยพับหุบเขาตามลำดับ

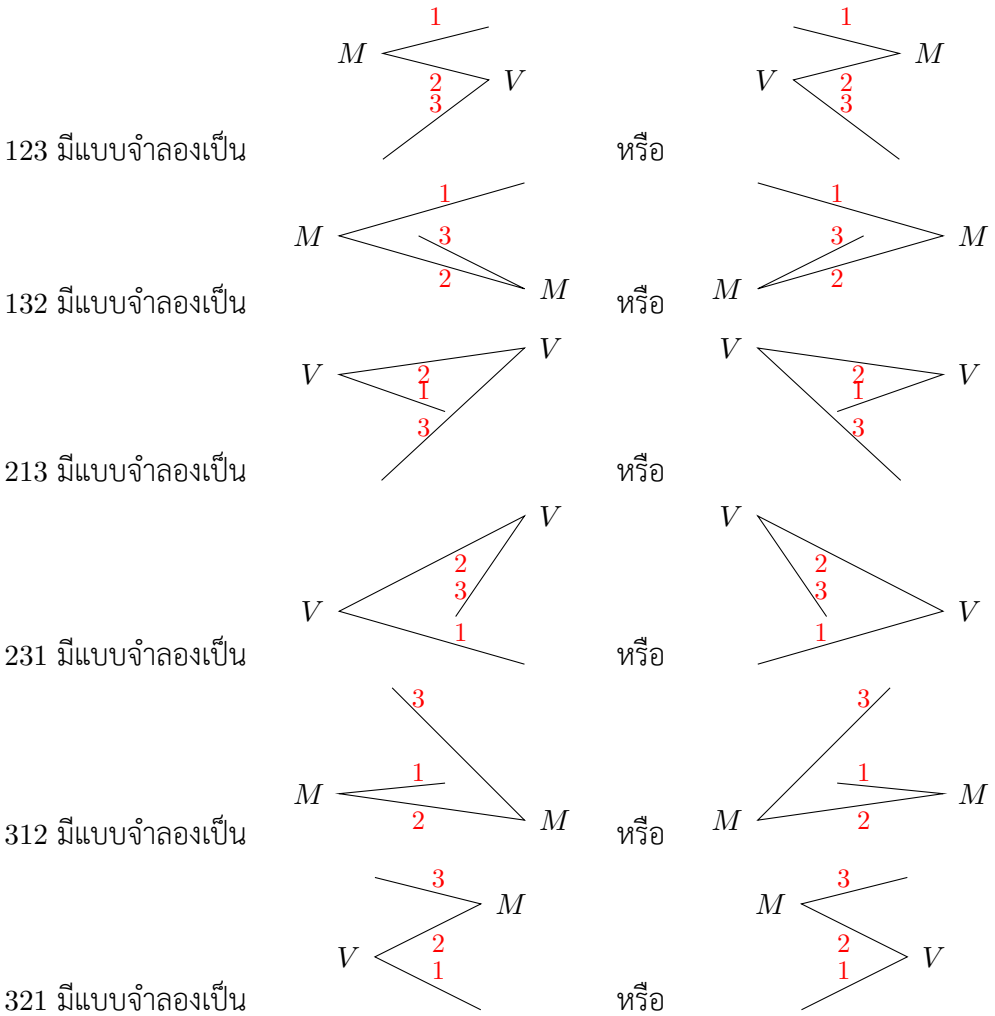


ภาพที่ 4: แบบจำลอง 2 มิติ สำหรับ 2-ทบ

สำหรับ 3-ทบ ซึ่งมีทั้งหมด 6 แบบ เราได้แบบจำลอง 2 มิติ ดังภาพที่ 5

การวาดแบบจำลองสองมิติสำหรับ n-ทบ

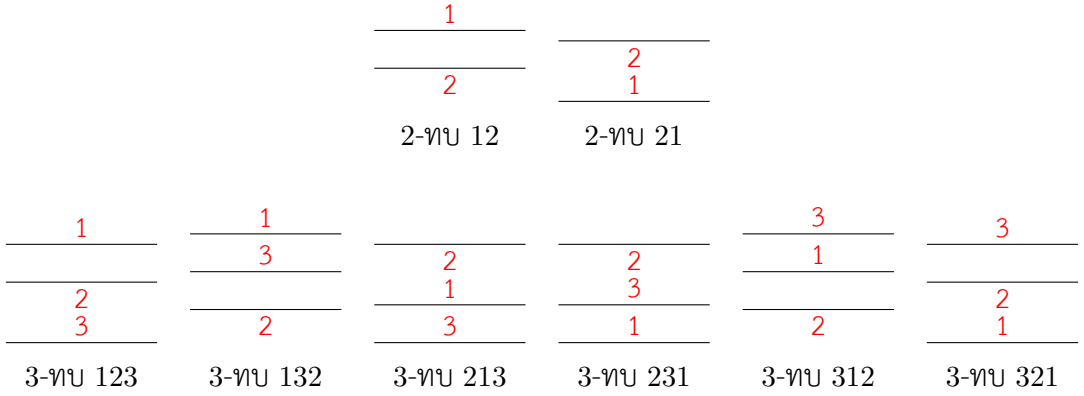
เราสามารถวาดแบบจำลองของ n-ทบ ดังนี้ นำ n-ทบ มาย้อมสีตรงเส้นขอบของดวงแสตมป์โดยย้อมต่อเนื่องไม่ยกมือ จากขอบแสตมป์ดวงที่ 1 ไปขอบแสตมป์ดวงที่ 2 จนถึงขอบแสตมป์ดวงที่ n แสตมป์แต่ละดวงถูกย้อมเพียงเส้นขอบเดียว จากนั้นตั้งทบแสตมป์โดยให้หมายเลข 1 หงายขึ้นตามกติกา แล้วประทับรอยสีย้อมบนกระดาษโดยให้ส่วนที่เป็นมุมอยู่ทางขวาหรือทางซ้ายของกระดาษ เราจะได้ลายเส้นของแบบจำลองขึ้นมา แต่ละมุมที่เส้นชนกันเราเขียนอักษร M และ V กำกับไว้ สุดท้ายใส่หมายเลขโดยเริ่มที่เลข 1 ต้องอยู่เหนือเส้นของมันตามด้วยหมายเลข $2, 3, \ldots, n$ ซึ่งจะอยู่เหนือเส้นหรืออยู่ใต้เส้นของมันขึ้นอยู่กับว่าเป็นรอยพับภูเขาหรือรอยพับหุบเขา การย้อมสีดังกล่าวทำได้ 2 แบบ เราจะย้อมขอบบนหรือขอบล่างของแสตมป์ก็ได้ซึ่งส่งผลให้แต่ละ n-ทบ มีแบบจำลองได้ 2 แบบ ดังภาพที่ 4 และภาพที่ 5



ภาพที่ 5: แบบจำลอง 2 มิติ สำหรับ 3-ทบ

4.2 แพทเทิร์นของ n-ทบ

จากแบบจำลอง 2 มิติ สำหรับ 2-ทบ และ 3-ทบ ดังภาพที่ 4 และภาพที่ 5 จะพบว่า ถ้าเราตัดมุมที่เส้นชนกันออกแล้วจัดเส้นให้เป็นแนวขนานกันโดยคงหมายเลขไว้ เราจะได้แพทเทิร์นของ 2-ทบ และ 3-ทบ ดังภาพที่ 6



ภาพที่ 6: แพทเทิร์นของ 2-ทบ และ 3-ทบ

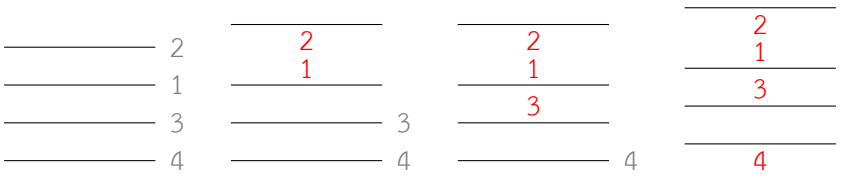
ดังนั้น แพทเทิร์นของ n-ทบ คือการจัดเรียงเส้นทั้งหมด n เส้นให้เป็นแนวขนานกันโดยมีหมายเลขกำกับ เริ่มจากหมายเลข 1 แทนแสตมป์ดวงที่ 1 ต้องเขียนไว้เหนือเส้นของมันเสมอ ถัดมาคือหมายเลข 2 จะอยู่เหนือเส้นหรืออยู่ใต้เส้นของมันขึ้นอยู่กับรอยพับระหว่างแสตมป์ดวงที่ 1 กับแสตมป์ดวงที่ 2 ว่าเป็นรอยพับภูเขาหรือเป็นรอยพับหุบเขา จนกระทั่งหมายเลข n ที่ต้องเทียบกับเส้นที่ $n-1$ ว่ารอยพับระหว่างแสตมป์ดวงที่ n กับ $n-1$ เป็นแบบใด เราได้แจกแจงกรณีทั้งหมดที่เป็นไปได้ในแพทเทิร์นของ n-ทบ เมื่อตัดเฉพาะส่วนระหว่างแสตมป์ดวงที่ k กับ ดวงที่ $k+1$ ($k=2,3,\ldots,n-1$) มาพิจารณาดังภาพที่ 7

(a) รอยพับภูเขา (b) รอยพับภูเขา (c) รอยพับหุบเขา (d) รอยพับหุบเขา

ภาพที่ 7: แสตมป์ดวงที่ k กับ $k+1$ ในแพทเทิร์น

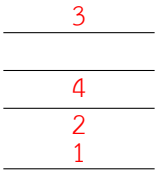
ตัวอย่าง 4.1. เราจะแสดงการวาดแพทเทิร์นของ 4-ทบ 2134 ซึ่งสัมพันธ์กับการเรียงสับเปลี่ยน VVM ดังนี้

1. จัดเรียงเส้นทั้งหมด 4 เส้นให้เป็นแนวขนานกันดังภาพที่ 8 (a) หมายเลขทางขวาคือลำดับของดวงแสตมป์ในแถบแสตมป์
2. เขียนหมายเลข 1 ไว้เหนือเส้นของมัน เนื่องจากรอยพับระหว่างแสตมป์ดวงที่ 1 กับดวงที่ 2 เป็นรอยพับหุบเขา ดังนั้นหมายเลข 2 ต้องอยู่ใต้เส้นของมันดังภาพที่ 8 (b)
3. เนื่องจากรอยพับระหว่างแสตมป์ดวงที่ 2 กับดวงที่ 3 เป็นรอยพับหุบเขา ดังนั้นหมายเลข 3 ต้องอยู่เหนือเส้นของมันดังภาพที่ 8 (c)
4. เนื่องจากรอยพับระหว่างแสตมป์ดวงที่ 3 กับดวงที่ 4 เป็นรอยพับภูเขา ดังนั้นหมายเลข 4 ต้องอยู่ใต้เส้นของมันดังภาพที่ 8 (d) เป็นแพทเทิร์นของ 4-ทบ 2134 ตามต้องการ



ภาพที่ 8: แพทเทิร์นของ 4-ทบ 2134

ตัวอย่าง 4.2. ให้ π เป็น 4-ทบ ที่มีแพทเทิร์นดังภาพที่ 9 จะได้ว่า π คือ 3421 เพื่อหาการเรียงสับเปลี่ยนรอยพับภูเขาหุบเขาของ π เราเริ่มที่เส้นหมายเลข 1 กับหมายเลข 2 ซึ่งจะพบว่าเป็นกรณีดังภาพที่ 7 (c) ดังนั้นรอยพับระหว่างแสตมป์ดวงที่ 1 กับดวงที่ 2 เป็นรอยพับหุบเขา (V) ต่อไปดูเส้นหมายเลข 2 กับหมายเลข 3 จะพบว่าเป็นกรณีดังภาพที่ 7 (a) ดังนั้นรอยพับระหว่างแสตมป์ดวงที่ 2 กับดวงที่ 3 เป็นรอยพับภูเขา (M) สำหรับเส้นหมายเลข 3 กับหมายเลข 4 จะพบว่าเป็นกรณีดังภาพที่ 7 (b) ดังนั้นรอยพับระหว่างแสตมป์ดวงที่ 3 กับดวงที่ 4 เป็นรอยพับภูเขา (M) ดังนั้น π สัมพันธ์กับการเรียงสับเปลี่ยน VMM



ภาพที่ 9: แพทเทิร์นของ 4-ทบ π

นิยาม 4.3. ให้ π เป็น n-ทบ *ลำดับเรียงเส้น* ในแพทเทิร์นของ π คือลำดับ $\langle 1x_2^2x_3^3\dots x_n^n\rangle$ เมื่อ $x^k \in \{N_k, S_k\}$ สำหรับ $k = 2, 3, \dots, n$ โดย N_k หมายถึง เส้นหมายเลข k อยู่เหนือเส้นหมายเลข $k-1$ และ S_k หมายถึง เส้นหมายเลข k อยู่ใต้เส้นหมายเลข $k-1$

ตัวอย่าง 4.4. จากภาพที่ 9 จะได้ลำดับเรียงเส้นของ 4-ทบ 3421 คือ $\langle 1N_2N_3S_4\rangle$

5 ทฤษฎีบท

จากตารางที่ 1 และตารางที่ 2 จะพบว่า นอกจากกรณี $n = 2$ แล้ว แต่ละการเรียงสับเปลี่ยนรอยพับภูเขาหุบเขามีความสัมพันธ์กับ n-ทบ มากกว่าหนึ่งแบบ ยกเว้นการเรียงสับเปลี่ยนที่ M กับ V อยู่ติดกันและมีการเรียงสลับกันไป นั่นคือการเรียงสับเปลี่ยน MV, VM การเรียงสับเปลี่ยน MVM, VMV การเรียงสับเปลี่ยน $MVMV, VMVM$ และการเรียงสับเปลี่ยน $MVMVM, VMVMV$ สำหรับกรณี $n = 3, 4, 5, 6$ ตามลำดับ เราจึงอนุมานว่าในกรณีที่ $n \geq 2$ ความสัมพันธ์ดังกล่าวควรจะเป็นไปในทำนองเดียวกัน และได้เสนอบทตั้งและทฤษฎีบทดังต่อไปนี้

บทตั้ง 5.1. *ให้ $n \in \mathbb{N}$ โดยที่ $n \geq 2$ และให้ π เป็น n-ทบใด ๆ จะได้ว่า*

(i) ถ้า π สัมพันธ์กับการเรียงสับเปลี่ยน $MVM\dots MV$ หรือ $MVM\dots VM$ แล้ว แพทเทิร์นของ π มีลำดับเรียงเส้นเป็น $\langle 1S_2S_3\dots S_n\rangle$

(ii) ถ้า π สัมพันธ์กับการเรียงสับเปลี่ยน $VMV\dots VM$ หรือ $VMV\dots MV$ แล้ว แพทเทิร์นของ π มีลำดับเรียงเส้นเป็น $\langle 1N_2N_3\dots N_n\rangle$

พิสูจน์. เราจะพิสูจน์โดยใช้หลักการอุปนัยเชิงคณิตศาสตร์แบบเข้มดังนี้ จากภาพที่ 6 เราได้ว่าบทตั้งเป็นจริงสำหรับ $n = 2, 3$ สมมติว่าบทตั้งเป็นจริงสำหรับทุก ๆ จำนวนนับ $k \geq 3$ ให้ π เป็น $(k+1)$-ทบ

(i) เราแบ่งการพิสูจน์เป็น 2 กรณี ดังนี้

กรณีที่ 1 k เป็นจำนวนคู่

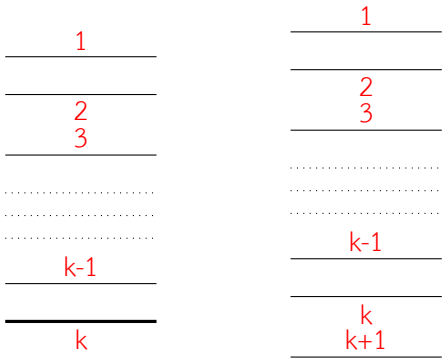
จะได้ว่า π สัมพันธ์กับการเรียงสับเปลี่ยน $MVM\dots MV$ ถ้าเราติดกาวแสตมป์ดวงที่ k กับ $k+1$ ให้เป็นดวงเดียว เราจะได้ k- ทบ ที่สัมพันธ์กับการเรียงสับเปลี่ยน $MVM\dots VM$ โดยสมมติฐานการอุปนัยจะได้ว่าแพทเทิร์นของ k-ทบ นี้มีลำดับเรียงเส้นเป็น $\langle 1S_2S_3\dots S_k\rangle$ ดังนั้น k-ทบ มีแพทเทิร์นดังภาพที่ 10 ซ้าย เนื่องจากรอยพับระหว่างแสตมป์ดวงที่ k กับ $k+1$ เป็นรอยพับหุบเขา ดังนั้นเมื่อลอกกาวออกเพื่อกลับคืนเป็น $(k+1)$-ทบ π ดังเดิม จะได้แพทเทิร์นของ π ดังภาพที่ 10 ขวา ซึ่งมีลำดับเรียงเส้นเป็น $\langle 1S_2S_3\dots S_{k+1}\rangle$

กรณีที่ 2 k เป็นจำนวนคี่

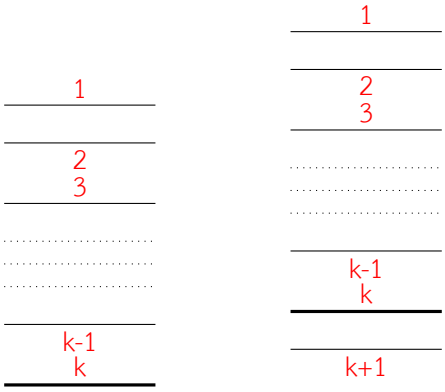
จะได้ว่า π สัมพันธ์กับการเรียงสับเปลี่ยน $MVM\dots VM$ ถ้าเราติดกาวแสตมป์ดวงที่ k กับ $k+1$ ให้เป็นดวงเดียว เราจะได้ k- ทบ ที่สัมพันธ์กับการเรียงสับเปลี่ยน $MVM\dots MV$ ในทำนองเดียวกันกับกรณี k เป็นจำนวนคู่ เราได้โดยสมมติฐานการอุปนัยว่า แพทเทิร์นของ k-ทบ นี้มีลำดับเรียงเส้นเป็น $\langle 1S_2S_3\dots S_k\rangle$ และมีแพทเทิร์นดังภาพที่ 11 ซ้าย เนื่องจากรอยพับระหว่างแสตมป์ดวงที่ k กับ $k+1$ เป็นรอยพับภูเขา ดังนั้นเมื่อลอกกาวออกเพื่อกลับคืนเป็น $(k+1)$-ทบ π จะได้แพทเทิร์นของ π ดังภาพที่ 11 ขวา ซึ่งมีลำดับเรียงเส้นเป็น

$\langle 1S_2S_3\dots S_{k+1}\rangle$ ตามต้องการ

(ii) สามารถพิสูจน์ได้ในทำนองเดียวกันกับ (i)



ภาพที่ 10: แพทเทิร์นของ k-ทบ และ $(k+1)$-ทบ กรณี k เป็นจำนวนคู่



ภาพที่ 11: แพทเทิร์นของ k-ทบ และ $(k+1)$-ทบ กรณี k เป็นจำนวนคี่

□

ทฤษฎีบท 5.2. *ให้ $n\in\mathbb{N}$ โดยที่ $n\geq 2$ และให้ π เป็น n-ทบ ใด ๆ จะได้ว่า*

(i) ถ้า π สัมพันธ์กับการเรียงสับเปลี่ยน $MVM\dots VM$ หรือ $MVM\dots MV$ แล้ว π คือ n-ทบ $12\dots n$

(ii) ถ้า π สัมพันธ์กับการเรียงสับเปลี่ยน $VMV\dots MV$ หรือ $VMV\dots VM$ แล้ว π คือ n-ทบ $n(n-1)\dots 1$

พิสูจน์. (i) สมมติว่า π เป็น n-ทบ ที่สัมพันธ์กับการเรียงสับเปลี่ยน $MVM\dots VM$ หรือ $MVM\dots MV$ โดยบทตั้ง 5.1 (i) จะได้ว่า แพทเทิร์นของ π มีลำดับเรียงเส้นเป็น $\langle 1S_2S_3\dots S_n\rangle$ เมื่ออ่านหมายเลขในแพทเทิร์นจากด้านบนลงมาด้านล่าง จะได้ว่า π คือ n-ทบ $12\dots n$

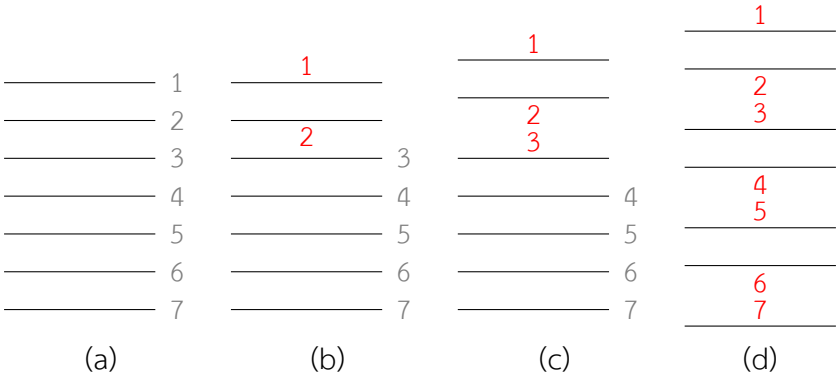
(ii) สมมติว่า π เป็น n-ทบ ที่สัมพันธ์กับการเรียงสับเปลี่ยน $VMV\dots MV$ หรือ $VMV\dots VM$ โดยบทตั้ง 5.1 (ii) จะได้ว่า แพทเทิร์นของ π มีลำดับเรียงเส้นเป็น $\langle 1N_2N_3\dots N_n\rangle$ เมื่ออ่านหมายเลขในแพทเทิร์นจากด้านบนลงมาด้านล่าง จะได้ว่า π คือ n-ทบ $n(n-1)\dots 1$ □

ทฤษฎีบท 5.2 ได้พิสูจน์ว่าเป็นจริงเฉพาะเงื่อนไขที่เพียงพอสำหรับการเป็น n-ทบ $12\dots n$ กับ $n(n-1)\dots 1$ ในส่วนของเงื่อนไขที่จำเป็นหรือบทกลับของทฤษฎีบทนั้นผู้วิจัยอนุมานได้ว่าเป็นจริง และยังอยู่ระหว่างการศึกษา ตัวอย่าง 5.3 และตัวอย่าง 5.4 ต่อไปนี้ เราแสดงเพื่อสนับสนุนการอนุมานดังกล่าว ยิ่งไปกว่านั้นเรายังสนใจการวิเคราะห์ความสัมพันธ์ระหว่างการเรียงสับเปลี่ยนรอยพับภูเขาหุบเขา กับ n-ทบ อื่น ๆ ต่อไปด้วย

ตัวอย่าง 5.3. เราจะแสดงว่า 7-ทบ 1234567 สัมพันธ์กับการเรียงสับเปลี่ยน $MVMVMV$ โดยการวาดแพทเทิร์นดังนี้

1. จัดเรียงเส้นทั้งหมด 7 เส้นให้เป็นแนวขนานกันดังภาพที่ 12 (a) หมายเลขทางขวาคือลำดับของดวงแสตมป์ในแถบแสตมป์
2. เขียนหมายเลข 1 ไว้เหนือเส้นของมัน สำหรับหมายเลข 2 เราต้องเทียบกับหมายเลข 1 จากภาพที่ 6 จะพบว่าเราต้องเขียนหมายเลข 2 ไว้ใต้เส้นของมันดังภาพที่ 12 (b)
3. สำหรับหมายเลข 3 ซึ่งต้องเทียบกับหมายเลข 2 จะพบว่าเป็นกรณีของภาพที่ 7 (d) ดังนั้นเราต้องเขียนหมายเลข 3 ไว้เหนือเส้นของมันดังภาพที่ 12 (c)
4. ทำซ้ำตามขั้นตอนที่ 3 กับหมายเลข $4, 5, 6$ และ 7 จะได้แพทเทิร์นของ 7-ทบ 1234567 ดังภาพที่ 12 (d)

ดังนั้น 7-ทบ 1234567 สัมพันธ์กับการเรียงสับเปลี่ยน $MVMVMV$

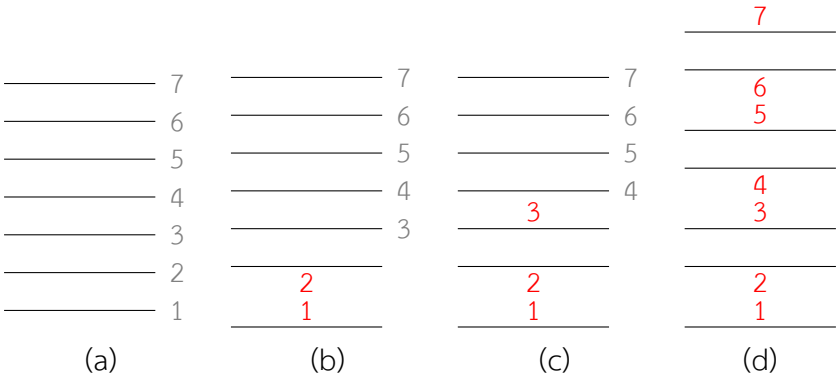


ภาพที่ 12: แพทเทิร์นของ 7-ทบ 1234567

ตัวอย่าง 5.4. เราจะแสดงว่า 7-ทบ 7654321 สัมพันธ์กับการเรียงสับเปลี่ยน $VMVMVM$ โดยการวาดแพทเทิร์นดังนี้

1. จัดเรียงเส้นทั้งหมด 7 เส้นให้เป็นแนวขนานกันดังภาพที่ 13 (a) หมายเลขทางขวาคือลำดับของดวงแสตมป์ในแถบแสตมป์
2. เขียนหมายเลข 1 ไว้เหนือเส้นของมัน สำหรับหมายเลข 2 เราต้องเทียบกับหมายเลข 1 จากภาพที่ 6 จะพบว่าเราต้องเขียนหมายเลข 2 ไว้ใต้เส้นของมันดังภาพที่ 13 (b)
3. สำหรับหมายเลข 3 ซึ่งต้องเทียบกับหมายเลข 2 จะพบว่าเป็นกรณีของภาพที่ 7 (a) ดังนั้นเราต้องเขียนหมายเลข 3 ไว้เหนือเส้นของมันดังภาพที่ 13 (c)
4. ทำซ้ำตามขั้นตอนที่ 3 กับหมายเลข $4, 5, 6$ และ 7 จะได้แพทเทิร์นของ 7-ทบ 7654321 ดังภาพที่ 13 (d)

ดังนั้น 7-ทบ 7654321 สัมพันธ์กับการเรียงสับเปลี่ยน $VMVMVM$



ภาพที่ 13: แพทเทิร์นของ 7-ทบ 7654321

กิตติกรรมประกาศ ผู้วิจัยขอขอบคุณผู้ทรงคุณวุฒิทุกท่านที่ได้ให้ข้อคิดเห็นและข้อเสนอแนะต่าง ๆ เพื่อปรับปรุงบทความวิจัยนี้

เอกสารอ้างอิง

[1] T. Hull, *On the Mathematics of Flat Origamis*, Congressus Numerantium Vol.100, Utilitas Mathematica Publ., 1994, pp. 215–224.

[2] H. Kawasaki, *An Application of A Theorem of Alternatives to Origami*, J. Oper. Res. Soc. Jpn. **60**(3) (2017), 393–399.

[3] J. E. Koehler, *Folding a Strip of Stamps*, J. Comb. Theory **5** (1968), 135–152.

[4] J. Mitani, *A Method for Designing Crease Patterns for Flat-Foldable Origami with Numerical Optimization*, Journal for Geometry and Graphics **15**(2) (2011), 195–201.

[5] J. O'Rourke, *How to Fold It*, Cambridge University Press, New York, 2011, pp. 56–71.

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Secret Sharing from Combinatorial Designs

Nada Somswasdi[1,†] and Wutichai Chongchitmate[1,‡]

[1]Department of Mathematics and Computer Science, Faculty of Science
Chulalongkorn University, Bangkok 10330, Thailand

Abstract

A secret sharing scheme is a process where a secret is divided into shares and distributed to each of the n parties, where a group of these parties can reconstruct the secret only when they satisfy some conditions. In the threshold secret sharing schemes, only a group of t parties or more can recover the secret and a group of less than t parties have no information about the secret. By associating parties and shares to blocks and treatments of a combinatorial design, we can construct a threshold secret sharing scheme with the property that if a party loses their share, they can recover it using a secure protocol with the help of some parties involved in the scheme. The protocol used to recover the lost share is called the repairability protocol, and a threshold secret sharing scheme with such protocol is called a repairable threshold scheme or RTS. In this study, we constructed four new RTS's using four different designs with more flexible parameters and efficiency than the existing schemes.

Keywords: secret sharing, combinatorial design, share repairability.

2020 MSC: Primary 94A62; Secondary 05B15, 05B99.

1 Introduction

A secret sharing scheme is a process where a dealer chooses a secret and distributes shares to each of the n parties, where a group of these parties can reconstruct the secret only when some conditions are met. One of the most well-known secret sharing schemes is the *threshold secret sharing schemes*. In this type of secret sharing scheme, only a group of t parties or more can recover the secret and a group of less than t parties have no information about the secret. An example of the threshold secret sharing schemes is the *Shamir's secret sharing scheme* where we view the secret as the constant term of a polynomial of degree $t-1$ over a field $\mathbb{F}$, where the coefficients of other terms besides the constant term are randomly chosen from elements of $\mathbb{F}$. The share distribution of the scheme can be done by substituting i to the same polynomial and then giving the result to the i^{th} party, where $i = 1, 2, ..., n$. A group of t parties can recover the secret through the Lagrange interpolation formula and a group of less than t parties do not have any information about the secret since they can not reconstruct the polynomial [1]. The threshold secret sharing schemes can be generalized into the ramp schemes where instead of

†Speaker. ‡Corresponding author.
Email: 6470174123@student.chula.ac.th (N. Somswasdi), wutichai.ch@chula.ac.th (W. Chongchitmate).

having one threshold, a ramp scheme has two thresholds, denoted in this study by k_1 and k_2, such that any group of more than or equal to k_2 parties can reconstruct the secret and any group of less than or equal to k_1 have no information about the secret. A threshold secret sharing scheme with threshold t is a ramp scheme with $k_2 = t$ and $k_1 = t - 1$ [5].

There are many studies regarding secret sharing schemes, however, this study focuses on the study of share repairability of threshold secret sharing schemes, as conducted by Stinson and Wei [5]. The purpose of the studies of share repairability is to find a way to recover the share of a party in case they lose it, without any help from the dealer. The protocol we execute to repair the lost share is called the *repairability protocol*. We want these repairability protocols to be not only able to repair the lost share but also want them to be *secure*, that is, any group of parties that do not satisfy the conditions to reconstruct the secret still do not know the secret with the information they have and all the information they sent and receive during the repairability protocol, regardless of how many time the protocol is executed. We call a threshold secret sharing scheme with a secure repairability protocol a *repairable threshold Scheme (RTS)* [5].

There are some parameters we need to be concerned with when considering an RTS other than the threshold of the scheme, including the repairing degree, information rate, and communication complexity defined by Stinson and Wei in [5] to determine the efficiency of an RTS. There is also the repairability index which defined by Liang and Stinson in [6] to further determine the properties of an RTS.

In [5], Stinson and Wei found an RTS that has the optimal information rate and repairability index, but also has bad communication complexity which means that this scheme is inefficient to execute. They then discovered a way to obtain new RTS's with less communication complexity by linking parties and shares to blocks and treatments of *combinatorial designs*.

A combinatorial design is a way of selecting subsets called *blocks* from a finite set of *treatments* in such a way that satisfies some conditions [7]. Stinson and Wei defined the distribution designs that have the conditions they need to obtain new RTS's with better communication complexity. An RTS with repairability protocol that uses a distribution design to repair the lost share is called the *combinatorial RTS* by Kacsmar and Stinson in [2]. Some known designs, including the balanced incomplete block designs and $\tau-$designs, are considered as distribution designs in the work of Stinson and Wei [5] and a work of Kacsmar and Stinson [2]. However, some parameters in the combinatorial RTS constructed by Stinson and Wei are not flexible, meaning those schemes only work in very few cases. And while Kacsmar and Stinson obtained new schemes with more flexible parameters, they do not have very good efficiency compared to schemes in [5]. So, in this study, we use different designs as distribution designs to construct RTS's with more flexible parameters and more efficiency than the existing ones.

In Section 2, we give more details about the secret sharing and combinatorial designs as well as the process of share repairability using the designs. We will also give some more details about the previous works of Stinson and Wei and Kacsmar and Stinson mentioned above. Section 4 contains our main results and the comparison between our results and the previous works.

2 Preliminaries

In this section, we provide proper definitions and some facts about combinatorial designs, secret sharing schemes, and share repairability mentioned in the introduction. We also talk about some existing schemes and their parameters as well as introduce the designs used to construct new schemes.

2.1 Notations

Let S be a set. We write $|S|$ to denote the number of elements in S. The notation $S \to a$ means that we uniformly sample an element from S. Additionally, if B and C are sets, a *randomized*

algorithm $A : B \to C$ is an algorithm on an input $x \in B$ that calculates a function $f(x, r)$ mapping from $B \times R$ to C, where $r \in R$ is obtained by A uniformly sampling an element from R. We write $A(x) \to y$ to denote that we give A an input x and receive y as a result.

2.2 Combinatorial Designs

First, we cover the definitions of the combinatorial design as well as the definitions of designs that are used to construct the new schemes in this study and some later theorems [4] [7].

Definition 2.1. A **combinatorial design** (on set S) is a way of selecting subsets from a finite set S in such a way that some specified conditions are satisfied. S is called the **support set** of the design, the members of S are called **treatments**, the chosen subsets of S in the design are called **blocks**, and the collection (i.e., multiset) of all blocks of a design is called the **block set**.

Definition 2.2. Let D be a design on support set S.

1. D is **incomplete** if at least one block does not contain all treatments from S.
2. D is **regular** if every block of D has the same size and each treatment occurs equally often in the design.
3. D is called a **balanced design** if any two treatments occur together in precisely λ blocks, where λ is a constant.

Definition 2.3. A **balanced incomplete block design (BIBD)** is a design that is incomplete, regular, and balanced.

We refer to a BIBD by using five parameters (m, b, r, k, λ), each parameter is defined as follows;

- m is the size of the support set,
- b is the number of blocks in the design,
- r is the number of blocks that contain each treatment of the design,
- k is the size of blocks in the design and
- λ is the number of blocks that contain each pair of different treatments.

We denote a BIBD with those parameters by (m, b, r, k, λ)- BIBD. Note that we can define parameters m and b similarly in any general design, and r also exists in a regular design that is not a BIBD. However, not all designs have all blocks in the same size and we can find a constant λ only in the balanced designs.

Using the defined conditions of regular designs and BIBDs, one can obtain the following properties regarding parameters m, b, r, k, and λ.

Theorem 2.4 (Theorem 1.1 of [7])**.** *In a regular design,*

$$bk = mr$$

Theorem 2.5 (Theorem 1.2 of [7])**.** *In an* $(m, b, r, k, \lambda)-$ *BIBD,*

$$r(k-1) = \lambda(m-1)$$

As we can always find other parameters from m, k and λ, we sometimes denote an (m, b, r, k, λ)-BIBD by (m, k, λ)- BIBD.

The following are some special types and properties of BIBDs that will be mentioned again afterward [7].

Definition 2.6. BIBDs with $k = 3$ and $\lambda = 1$ are called **Steiner triple systems (STS)**. We denote a Steiner triple system with the support set of size m by $STS(m)$.

Definition 2.7. A **symmetric balanced incomplete block design (SBIBD)** is an (m, b, r, k, λ)-BIBD with $m = b$ and $r = k$, denoted by (m, k, λ)-SBIBD.

Theorem 2.8 (Corollary 2.10.2 of [7]). *The intersection of two distinct blocks of an (m, k, λ)-SBIBD always contains λ treatments.*

Definition 2.9. A design is called **resolvable** if one can divide its blocks into disjoint partitions, where each treatment of the design appears in each partition exactly once. Such partition is called a **parallel class**.

For example, an $STS(9)$ with the block set $\beta = \{\{1,2,3\},\{4,5,6\},\{7,8,9\},\{1,4,7\},\{2,5,8\},\{3,6,9\},\{1,5,9\},\{2,6,7\},\{3,4,8\},\{1,6,8\},\{2,4,9\},\{3,5,7\}\}$ is resolvable since its blocks can be divided into the following parallel classes:

- $\{1,2,3\},\{4,5,6\},\{7,8,9\}$
- $\{1,4,7\},\{2,5,8\},\{3,6,9\}$
- $\{1,5,9\},\{2,6,7\},\{3,4,8\}$
- $\{1,6,8\},\{2,4,9\},\{3,5,7\}$

Note that a design does not need to be a BIBD to be resolvable, but in this study, we will consider only resolvable BIBDs.

The next definition is for the design used in [2] and a theorem used to calculate the parameters of some schemes from the same article.

Definition 2.10. A **τ-(m, k, λ)-design** is a design where:

1. The support set is of size m.
2. Each block contains exactly k treatments.
3. Every set of τ treatments from the support set occurs in exactly λ blocks.

Theorem 2.11 (Theorem 1.10 of [2]). *The ith replication number, denoted r_i, of a $\tau-(m, k, \lambda)$-design is the number of blocks containing any given set of i treatments. It is known that*

$$r_i = \frac{\lambda\binom{m-i}{\tau-i}}{\binom{k-i}{\tau-i}},$$

for $1 \leq i \leq \tau$

Lastly, we will cover some of the objects that are initially not a design but can be considered as or used to construct one [7].

Definition 2.12. An **affine plane** consists of a set P of objects called **points** and a set L of nonempty subsets of P called lines such that:

1. Given any two distinct points P and Q, there is exactly one line that contains them both.
2. There is a set of four points, not three of which belong to one common line.
3. Given any point P and given any line q, that does not contain P, there is exactly one line that contains P and contains no point of q.

A **finite affine plane** is an affine plane with finite P.

Theorem 2.13 (Lemma 3.2 of [7]). *In a finite affine plane, there is a parameter n such that every line contains n points and every point lies on $n+1$ lines. We denote a finite affine plane with parameter n by $AG(2, n)$.*

According to [7], we can always construct an $AG(2,n)$ from a finite field of order n, so if such a field exists, then $AG(2,n)$ exists. Thus, we can always find an $AG(2,n)$ when n is a prime power.

An affine plane can be considered as a resolvable BIBD as stated in Theorem 2.14 and it has some similar properties to a resolvable design as stated in Theorem 2.15.

Theorem 2.14 (Theorem 3.3 of [7]). *If "points" are identified with "treatments" and "lines" are identified with "blocks", then a finite affine plane with parameter n is precisely a resolvable BIBD with parameters $(n^2, n^2+n, n+1, n, 1)$.*

Theorem 2.15 (Corollary 3.12.2 of [7]). *The lines of $AG(2,n)$ can be partitioned into $n+1$ subsets of size n, called **parallel classes**, such that two lines meet if and only if they are in different parallel classes.*

Theorem 2.16 (Theorem 3.13 of [7]). *There exists an $AG(2,n)$ if and only if there exists an $(n^2+n+1, n+1, 1)-$ SBIBD.*

Since an $AG(2,n)$ exists when n is a prime power, by Theorem 2.16, we get that an $(n^2+n+1, n+1, 1)-$ SBIBD exists when n is a prime power.

Another important object is the Room square defined as the following [7].

Definition 2.17. Let S be a set of $r+1$ elements. A **Room square of side** r (or of order $r+1$) is an $r \times r$ array such that:

1. Each cell is either empty or contains an unordered pair of symbols chosen from S.
2. Each row and each column contains each element of S precisely once.
3. Each of the $\binom{r+1}{2}$ possible distinct pairs of symbols occurs exactly once in a cell of the square.

It follows that r must be odd, that is, $r = 2n-1$ for some $n \in \mathbb{N}$, and each row and each column of a room square of side $r = 2n-1$ has n non-empty cells and $n-1$ empty cells. We often use $S = \{1, 2, ..., 2n-1\} \cup \{\infty\}$ [7].

Theorem 2.18 (Chapter 15 of [7]). *A Room square of side $r = 2n-1$ exists when $n \geq 4$.*

2.3 Secret Sharing

There are various types of secret sharing schemes, but the ones we focus on are *threshold secret sharing schemes* and *ramp secret sharing schemes*. According to [3], the threshold secret sharing Scheme (sometimes the name was shortened to *secret sharing schemes*) can be formally defined as the following. Note that, if $V = (v_1, v_2, ..., v_a)$ is an a-tuple and $I = \{i_1, i_2, ..., i_b\}$ is an index set, $V|_I$ denotes the projection of V onto its $i_1^{th}, i_2^{th}, ..., i_b^{th}$ position. For example, if $V = (v_1, v_2, v_3, v_4, v_5)$ and $I = \{1, 3, 4\}$, then $V|_I = (v_1, v_3, v_4)$.

Definition 2.19. Let D be the domain of secrets and D_1 be the domain of shares. Let $Shr : D \to D_1^n$ be a randomized sharing algorithm, and $Rec : D_1^k \to D$ be a reconstruction algorithm.

Let t and n be positive integers such that $n \geq t$. A (t,n)**- (threshold) secret sharing scheme** is a pair of algorithms (Shr, Rec) that satisfies these two properties:

- Reconstruction. For all $s \in D$, if $Shr(s) \to (s_1, s_2, , s_n)$ then

$$Rec(s_{i_1}, s_{i_2}, , s_{i_k}) = s,$$

for all $\{i_1, i_2, , i_k\} \subseteq \{1, 2, , n\}$ where $k \geq t$.

- Secrecy. For any two secrets $a, b \in D$, any index set $I = \{i_1, i_2, , i_k\} \subseteq \{1, 2, , n\}$ and any possible vector of shares $v = (v_1, v_2, , v_k) \in D_1^k$, such that $k < t$,

$$Pr[v = Shr(a)|_I] = Pr[v = Shr(b)|_I],$$

where Pr denotes the probability on randomness of the sharing algorithm.

This concept of the threshold secret sharing schemes can be generalized in the form of *ramp schemes* as follows [3] [5].

Definition 2.20. Let D be the domain of secrets and D_1 be the domain of shares. Let $Shr : D \to D_1^n$ be a randomized sharing algorithm, and $Rec : D_1^k \to D$ be a reconstruction algorithm.

Let k_1 and k_2 be positive integers such that $k_2 > k_1$. A (k_1, k_2, n)**-ramp scheme** is a pair of algorithms (Shr, Rec) that satisfies these two properties:

- Reconstruction. For all $s \in D$, if $Shr(s) \to (s_1, s_2, , s_n)$ then

$$Rec(s_{i_1}, s_{i_2}, , s_{i_k}) = s,$$

for all $\{i_1, i_2, , i_k\} \subseteq \{1, 2, , n\}$ where $k \geq k_2$.

- Secrecy. For any two secrets $a, b \in D$, any index set $I = \{i_1, i_2, , i_k\} \subseteq \{1, 2, , n\}$ and any possible vector of shares $v = (v_1, v_2, , v_k) \in D_1^k$, such that $k \leq k_1$,

$$Pr[v = Shr(a)|_I] = Pr[v = Shr(b)|_I],$$

where Pr denotes the probability on randomness of the sharing algorithm.

Note that a (t, n)- threshold secret sharing scheme is a $(t-1, t, n)$- ramp scheme.

Now, let l_1, l_2 and m be positive integers such that $m \geq l_2 > l_1$. We can always construct a (l_1, l_2, m)-ramp scheme using the following construction, where the construction takes place over a finite field $\mathbb{F}_Q$ of order $Q \geq m + 1$ [5]:

1. In the **Initialization Phase**, the dealer, denoted by P, chooses n distinct, non-zero elements of $\mathbb{F}_Q$, denoted x_i, where $1 \leq i \leq m$. The values x_i are public for $1 \leq i \leq m$.

2. Let $l = l_2 - l_1$. In the **Share Distribution phase**, P chooses a secret

$$s = (a_0, a_1, ..., a_{l-1}) \in \mathbb{F}_Q^l.$$

Then define $Shr : \mathbb{F}_Q^l \to \mathbb{F}_Q^n$ as $Shr(s) \to (a(x_1), a(x_2), ..., a(x_m))$, where $a(x) = \sum_{j=0}^{l_2-1} a_j x^j$ and $\mathbb{F}_Q^{l_1} \to (a_l, a_{l+1}, ..., a_{l_2-1})$. In other words, P secretly chooses (independently and uniformly at random) $a_l, a_{l+1}..., a_{l_2-1} \in \mathbb{F}_Q$, then for $1 \leq i \leq n$, P computes $y_i = a(x_i)$, where $a(x) = \sum_{j=0}^{l_2-1} a_j x^j$, and gives it to P_i, where P_i is the i^{th} party that receive the i^{th} share of the ramp scheme.

Reconstruction is easily accomplished using the Lagrange interpolation formula. That is, $Rec : \mathbb{F}_Q^k \to \mathbb{F}_Q^l$ where $Rec(y_{i_1}, y_{i_2}, ..., y_{i_k}) = (c_0, c_1, ..., c_{l-1})$, c_0 is the constant and $c_1, c_2, ..., c_{l-1}$ are the coefficients of $x, x^2, ..., x^{l-1}$ of the polynomial
$\frac{(x-x_{i_2})(x-x_{i_3})...(x-x_{i_{l_2}})}{(x_{i_1}-x_{i_2})(x_{i_1}-x_{i_3})...(x_{i_1}-x_{i_{l_2}})} y_{i_1} + \frac{(x-x_{i_1})(x-x_{i_3})...(x-x_{i_{l_2}})}{(x_{i_2}-x_{i_1})(x_{i_2}-x_{i_3})...(x_{i_2}-x_{i_{l_2}})} y_{i_2} + ...$
$+ \frac{(x-x_{i_1})(x-x_{i_2})...(x-x_{i_{l_2-1}})}{(x_{i_{l_2}}-x_{i_1})(x_{i_{l_2}}-x_{i_2})...(x_{i_{l_2}}-x_{i_{l_2-1}})} y_{i_{l_2}}$, as $k \geq l_2$.

In this study, the **size** of an object (such as secrets, shares, etc.) is considered as the number of the object's $\mathbb{F}_Q$ components. For example, the size of each share in the ramp scheme is 1 and the size of the secret in the ramp scheme is $k = l_2 - l_1$, this makes the size of the secret in the threshold scheme 1. We can convert the size of an object as defined above into bit length by multiplying $\lceil \log_2 Q \rceil$ to the size.

3 Combinatorial Repairability for Threshold Scheme

The problem of share repairability is that a certain party P_l in a secret sharing scheme loses its share. The goal is to find a secure protocol involving P_l and a subset of the other parties that allows the missing share x_l to be reconstructed [5].

We consider protocols that operate in two phases:

1. In the **message exchange phase**, a d-subset of parties other than P_l exchange messages among themselves.

2. In the **repairing phase**, these same d parties each send a message to P_l. The messages received by P_l allow P_l's share to be reconstructed. We consider protocols that appoint the same number of parties to help reconstruct each party's share, that is, d is constant for any P_l. We call this constant d the **repairing degree** of the protocol.

The protocol above is called the **repairability protocol**. This study only focuses on the repairability protocols for threshold secret sharing schemes.

In a (t, n)- threshold scheme, a repairability protocol is said to be **secure** if any coalition of $t-1$ parties cannot reconstruct the secret with all information they have, including their shares and all messages they sent or received during the repairability protocol, regardless of how many time it is executed [5]. We note that $d \geq t$ is a necessary condition for a secure repairability protocol with repairing degree to exist for a (t, n)- threshold scheme. Otherwise, if $d \leq t-1$, then a group of $t-1$ parties have enough information to reconstruct another share, making this $t-1$ parties hold t shares and can obtain the secret which is a contradiction to the condition of threshold schemes.

If a (t, n)- threshold scheme has a repairability protocol with d repairing degree that satisfies the above security requirement, then we say that it is a (t, n, d)**- repairable threshold scheme ((t, n, d)-RTS)** [5].

We say that a (t, n, d)-RTS has **universal repairability** if any subset of d parties can repair P_l's share.

Some parameters can be used to determine the efficiency of an RTS, first is the **information rate**. It is the ratio of the size of the secret to the maximum size of a share [5]. The high information rate means that we can share a large-size secret with small-size shares, and the small-size shares make it easier for the dealer to send them to parties or for a party to send them to another party. Hence, the higher the information rate of a secret sharing scheme, the better. According to [5], the information rate of a threshold scheme is always less than or equal to 1.

The next parameter is the **communication complexity**. The communication complexity of a repairability protocol is the sum of the sizes of all messages transmitted during the protocol divided by the size of the secret [5]. High communication complexity means that there are a lot of shares (or shares of large size) transmitted during the protocol, making the execution of the protocol inefficient. So the lower the communication complexity of a repairability protocol, the better.

The last parameter is the **repairability index**. The repairability index, denoted by κ, of a (t, n, d)-RTS is the ratio of the number of d-subsets of $n-1$ parties excluding P_l that can repair P_l's share to the number of all possible d-subsets of $n-1$ parties besides P_l, considering

conditions of the repairability protocol. If the repairability index is high, it means that a party can rely on many groups of d parties to repair their share, so the higher the reliability index of an RTS, the better.

The repairability index was first defined in [6] by Laing and Stinson with a slightly different name. There, it is called the *repairability* and also denoted by κ. However, we decided to change the name of the parameter to repairability index to avoid confusion with the share repairability which is what we call the initial problem.

In [5], Stinson and Wei constructed a (t, n, t)- RTS with universal repairability with the information rate 1 using Shamir's secret sharing scheme. But the communication complexity is t^2 which is quite high. They showed in [5] that there are RTS's with lower communication complexity, although we have to trade off the universal repairability and high information rate. These schemes can be constructed with the help of combinatorial designs, with the following definition as a key.

Definition 3.1. Let l_1 and l_2 be positive integers such that $l_2 - l_1 \geq 1$. A (t, l_1, l_2)**- distribution design** is a design that satisfies the following two properties [5]:

1. The union of any t blocks contains at least l_2 elements.
2. The union of any $t-1$ blocks contains at most l_1 elements.

Let t, d, n, l_1, l_2 and m be positive integers such that $d \leq t \leq n$, $m \geq l_2 > l_1$ and $m \geq n$. With a (t, l_1, l_2)- distribution design and an (l_1, l_2, m)- ramp scheme, we can construct a (t, n, d)- RTS using the following process [5]:

1. Consider a (t, l_1, l_2)- distribution design with each block of size d and the support set of size m such that there are n blocks of the design that contain all of the treatments, with each treatment appears in at least two of these blocks. We start with an (l_1, l_2, m)- ramp scheme called the "base scheme", the shares of this scheme are called "subshares".

2. We label each of the m subshares with a treatment of the (t, l_1, l_2)- distribution design so that two distinct subshares are labeled as two different treatments of the design. Then assign d subshares to each of the n parties we want in the result RTS in a way that each party holds subshares that are labeled by d treatments that contain in the same block in the (t, l_1, l_2)- distribution design. If a party P_i holds subshares labeled by treatments from block B_j, we say that P_i represents block B_j. We must assign subshares to the n parties in such a way that two distinct parties represent two different blocks and the n blocks of the distribution design that the n parties represent must have each treatment appear in at least two of these blocks. The design used in this construction is public information, that is, every party knows what design was used to distribute subshares and which block each party represents but each party only knows the subshares it holds itself.

3. Suppose that we want to repair the share of P_l that represent block B_l of the (t, l_1, l_2)- distribution design. For each treatment $x \in B_l$, since each treatment occurs in at least 2 blocks out of the n blocks represented by the n parties, we can find another party that represents a block containing x. That party can send the subshare labeled x to P_l whose share is being repaired. Since each party holds d subshares, we can assign d parties to help reconstruct P_l's share. So the result scheme has the repairing degree d. Since each party represents a block in the (t, l_1, l_2)- distribution design, we know that t parties hold at least l_2 subshares in total and $t-1$ parties hold at most l_1 subshares in total. Using the same secret as in the (l_1, l_2, m)- ramp scheme, since each subshare is a share of a (l_1, l_2, m)- ramp scheme and a group of t or more parties hold at least l_2 subshares in total, we know that any group of more than or equal to t parties can reconstruct the secret. Furthermore, since a group of less than or equal to $t-1$ parties holds at most l_1 subshares, any group of less than or equal to $t-1$ parties has no information about the secret. Thus, the result

is a (t, n, d)- RTS where each party has a share that contains d subshares and the shared secret is the same secret we share in the base scheme.

Note that the d subshares sent to P_l do not need to be from d different parties as a party may hold more than one common subshare with P_l. In this case, it is possible to have that party send two subshares and add another party that does not have to send its subshare so that the group that helps reconstruct P_l's share is still a group of d parties. However, in our result, we will have each party send exactly one subshare to P_l since we want to distribute the work equally between all the parties. Whether we can do this depends on the design used as a distribution design in the above construction. We will see in Section 4 that we can always choose d blocks in our results that can repair a share this way. It is obvious that the distribution designs that Stinson and Wei used in their work in [5] and distribution designs that Kacsmar and Stinson used in [2] also give results RTS's that allow P_l's share to be reconstructed by having each party send exactly one subshare to P_l. So, in this study, when we count the d-sets of parties that can repair the share of a party when computing the repairability index of RTS's constructed by this construction, we count only the set of d parties that can repair the share by having each of them send exactly one subshare.

The property that every treatment occurs in at least 2 out of n blocks in the (t, l_1, l_2)- distribution design used for the reconstruction of shares is a necessary and sufficient condition for this kind of repairability to be possible. Therefore, if this property is satisfied, we say that the distribution design is **repairable**. An RTS with the repairability protocol constructed by the previous construction is called a **combinatorial repairable threshold scheme (combinatorial RTS)** by Kacsmar and Stinson in [2].

Since we can always construct a (l_1, l_2, m)- ramp scheme over $\mathbb{F}_Q$ using the construction in Subsection 2.3 when $Q \geq m + 1$, if we can find a repairable (t, l_1, l_2)- distribution design, then we can construct a (t, n, d)- RTS using the construction above as stated in the following theorem by Stinson and Wei [5].

Theorem 3.2 (Theorem 4.1 of [5]). *Suppose that there exists a repairable (t, l_1, l_2)- distribution design on m treatments, having n blocks of size d, and suppose that $Q \geq m + 1$. Then, there is a (t, n, d)- RTS having information rate $\frac{l_2 - l_1}{d}$ and communication complexity $\frac{d}{l_2 - l_1}$, where every share is in $\mathbb{F}_Q^d$ as $\mathbb{F}_Q$ is a finite field of order Q.*

The following theorems show us some combinatorial designs that can be considered as a distribution design, and so can be used to construct an RTS according to [5].

Theorem 3.3 (Theorem 5.1 of [5]). *Suppose that $m \equiv 3 \pmod 6$, Q is a prime power such that $Q \geq m+1$ and $\frac{2m}{3} \leq n \leq \frac{m(m-1)}{6}$. Then there exists a $(2, n, 3)$- RTS with restricted repairability, with shares from $\mathbb{F}_Q^3$, having information rate $\frac{2}{3}$ and communication complexity $\frac{3}{2}$.*

Theorem 3.4 (Theorem 5.2 of [5]). *Suppose that $m \equiv 4 \pmod{12}$, Q is a prime power such that $Q \geq m+1$ and $\frac{m}{2} \leq n \leq \frac{m(m-1)}{12}$. Then there exists a $(2, n, 4)$- RTS with restricted repairability, with shares from $\mathbb{F}_Q^4$, having information rate $\frac{3}{4}$ and communication complexity $\frac{4}{3}$.*

Theorem 3.5 (Theorem 5.3 of [5]). *Suppose that $m \equiv 5 \pmod{20}$ and there exists a resolvable $(m, 5, 1)$-BIBD. Let Q be a prime power such that $Q \geq m + 1$ and $\frac{2m}{5} \leq n \leq \frac{m(m-1)}{20}$. Then, the following RTS exists:*

1. *A $(2, n, 5)$- RTS with restricted repairability, with shares from $\mathbb{F}_Q^5$, having information rate $\frac{4}{5}$ and communication complexity $\frac{5}{4}$.*
2. *A $(3, n, 5)$- RTS with restricted repairability, with shares from $\mathbb{F}_Q^5$, having information rate $\frac{2}{5}$ and communication complexity $\frac{5}{2}$.*

Theorem 3.6 (Theorem 5.4 of [5]). *Suppose that $m \equiv 8 \pmod{56}$ and there exists a resolvable $(m, 8, 1)$-BIBD. Let Q be a prime power such that $Q \geq m+1$ and $\frac{m}{4} \leq n \leq \frac{m(m-1)}{56}$. Then, the following RTS exists:*

1. *A $(2, n, 8)$- RTS with restricted repairability, with shares from $\mathbb{F}_Q^8$, having information rate $\frac{7}{8}$ and communication complexity $\frac{8}{7}$.*
2. *A $(3, n, 8)$- RTS with restricted repairability, with shares from $\mathbb{F}_Q^8$, having information rate $\frac{5}{8}$ and communication complexity $\frac{8}{5}$.*
3. *A $(4, n, 8)$- RTS with restricted repairability, with shares from $\mathbb{F}_Q^8$, having information rate $\frac{1}{4}$ and communication complexity 4.*

Aside from these schemes, we also get other schemes from $3-(m, 4, 1)$-designs as follows.

Theorem 3.7 (Theorem 1.9 & 3.1 of [2]). *Suppose that $m \equiv 2, 4 \pmod{6}$ and Q is a prime power such that $q \geq m+1$. Then we can construct a $(2, n, 4)$- RTS with restricted repairability, with shares from $\mathbb{F}_Q^4$ from a $3-(m, 4, 1)-$ design.*

This theorem can be generalized using $\tau-(m, d, 1)-$ designs as in the next theorem.

Theorem 3.8 (Theorem 3.3 of [2]). *Suppose that $\tau-(m, d, 1)-$ designs exist and Q is a prime power such that $q \geq m+1$. Then there exists a (t, n, d)- RTS with restricted repairability, with shares from $\mathbb{F}_Q^d$, where $t, d, \tau \in \mathbb{N}$ $t \geq 2$, $\tau \geq 3$, $n = \frac{\binom{m}{\tau}}{\binom{d}{\tau}}$ and $d \geq \binom{t}{2}(\tau-1)+1$.*

Unlike the schemes in [5], Kacsmar and Stinson use a threshold scheme as the base scheme for the construction of the previous two schemes, so the size of the secret is only 1. Thus, the information rate and communication complexity of the scheme in Theorem 3.7 is $\frac{1}{4}$ and 4 respectively. Similarly, the information rate and communication complexity of the scheme in Theorem 3.8 are $\frac{1}{d}$ and d respectively.

Returning to the repairability index of combinatorial repairable threshold schemes, the definition Laing and Stinson gave us in [6] concerns only the case when the number of parties is equal to the number of all blocks in the distribution designs. For example, in theorem 1.3.2 where $\frac{2m}{3} \leq n \leq \frac{m(m-1)}{6}$, we will calculate the repairability index of the scheme in the case that $n = \frac{m(m-1)}{6}$. We have the following theorem from [6] that helps us calculate the repairability index of a combinatorial RTS.

Theorem 3.9 (Theorem 4.4 of [6]). *A randomly chosen subset of d parties in a (t, n, d)- RTS, constructed using an underlying $(m, d, 1)$- BIBD with $n = b$ parties, has probability*

$$\kappa = \frac{(r-1)^d}{\binom{n-1}{d}}$$

of successfully repairing the share of a party P_l, where r is the replication number of the BIBD as defined in Subsection 2.2.

Table 1 shows us the parameters of all the schemes in this section. Note that, since a pair of treatments can occur together in various blocks, the number of d-subset that can repair a party's share when using a $\tau-(m, d, 1)$-design as distribution design is up to the blocks in the design and thus, is very complicated to find in general cases. So the columns "repairability index" of Scheme (6) and (7) are marked with a hyphen.

Table 1: Parameters of the existing schemes

Design/Method used	**Result**	**Information rate**	**Communication complexity**	**Rapairability index**
(1) $STS(m)$ (Thm.3.3) (Stinson and Wei [5])	$(2,n,3)$-RTS with shares from $\mathbb{F}_Q^3$	$\frac{2}{3}$	$\frac{3}{2}$	$\frac{(m-3)^3}{8\left(\frac{m(m-1)}{6}_3-1\right)}$
(2) $(m,4,1)$-BIBD (Thm.3.4) (Stinson & Wei [5])	$(2,n,4)$-RTS with shares from $\mathbb{F}_Q^4$	$\frac{3}{4}$	$\frac{4}{3}$	$\frac{(m-4)^4}{81\left(\frac{m(m-1)}{12}_4-1\right)}$
(3) $(m,5,1)$-BIBD (Thm.3.5) (Stinson & Wei [5])	$(2,n,5)$-RTS with shares from $\mathbb{F}_Q^5$	$\frac{4}{5}$	$\frac{5}{4}$	$\frac{(m-5)^5}{4^5\left(\frac{m(m-1)}{20}_5-1\right)}$
	$(3,n,5)$-RTS with shares from $\mathbb{F}_Q^5$	$\frac{2}{5}$	$\frac{5}{2}$	
(4) $(m,8,1)$-BIBD (Thm.3.6) (Stinson & Wei [5])	$(2,n,8)$-RTS with shares from $\mathbb{F}_Q^8$	$\frac{7}{8}$	$\frac{8}{7}$	$\frac{(m-8)^8}{7^8\left(\frac{m(m-1)}{56}_8-1\right)}$
	$(3,n,8)$-RTS with shares from $\mathbb{F}_Q^8$	$\frac{5}{8}$	$\frac{8}{5}$	
	$(4,n,8)$-RTS with shares from $\mathbb{F}_Q^8$	$\frac{1}{4}$	4	
(5) Shamir's secret sharing scheme (Stinson & Wei [5])	(t,n,t)-RTS with shares from $\mathbb{F}_Q$ having universal repairability	1	d^2	1
(6) $3-(m,4,1)$- design (Thm.3.7) (Kacsmar & Stinson [2])	$(2,n,4)$-RTS with shares from $\mathbb{F}_Q^4$	$\frac{1}{4}$	4	-
(7) $\tau-(m,d,1)$- design (Thm (3.8) (Kacsmar & Stinson [2])	(t,n,d)-RTS with shares from $\mathbb{F}_Q^d$	$\frac{1}{d}$	d	-

4 Main Results

This section contains the structure, conditions, and construction of new schemes obtained during this study. The designs used for the construction and parameters of those schemes are gathered in Table 2, we will cover the details of each scheme, including the conditions and constructions, in Subsection 4.1 - 4.4. There are two main parts in the construction of these schemes, first, we need to verify that the design we want to use is a distribution design, then we use Theorem 3.2 and the process in Subsection 3 to construct the scheme and find its parameters. We then compare the scheme with those in Table 1 at the end of each subsection.

The parameter l in Table 2 is an integer greater than or equal to 4. Unlike in Scheme (b), l is not the repairing degree of the result RTS, instead, in the final result, we obtain $d = 2l - 1$ and $n = d + 1 = 2l$.

Note that, though it is not shown in Table 2, the parameter n of Scheme (a) depends on the parameter d as will be explained in Subsection 4.1. Additionally, the parameter n of Scheme (d) also depends on d as well as the parameter λ. We will discuss the relation between these parameters of Scheme (d) in Subsection 4.4.

Table 2: Parameters of newly constructed schemes

Design used	**Result**	**Information rate**	**Communication complexity**	**Repairability index**
(a) Affine plane ($AG(2,d)$)	(t,n,d)-RTS with shares from $\mathbb{F}_Q^d$	$1-\frac{t(t-1)}{2d}$	$\frac{2d}{2d-t(t-1)}$	$\frac{d^d}{\binom{d^2+d-1}{d}}$
(b) Room square of side $r=2d-1$ (Ver.1)	$(t,4d-2,d)$-RTS with shares from $\mathbb{F}_Q^d$	$1-\frac{1}{d}\lceil\frac{t}{2}\rceil\lfloor\frac{t}{2}\rfloor$	$\frac{d}{d-\lceil\frac{t}{2}\rceil\lfloor\frac{t}{2}\rfloor}$	$\frac{1}{\binom{4d-3}{d}}$
(c) Room square of side $r=2l-1$ (Ver.2)	$(t,d+1,d)$-RTS with shares from $\mathbb{F}_Q^{n-1}$	$\frac{d+1}{2d}$	$\frac{2d}{d+1}$	1
(d) (n,d,λ)- SBIBD	(d1) $(2,n,d)$-RTS with shares from $\mathbb{F}_Q^d$ (d2) $(3,n,d)$-RTS with shares from $\mathbb{F}_Q^d$	(d1) $1-\frac{\lambda}{d}$ (d2) $1-\frac{2\lambda}{d}$	(d1) $\frac{d}{d-\lambda}$ (d2) $\frac{d}{d-2\lambda}$	$\frac{(d-1)^d}{\binom{n-1}{d}}$, when $\lambda=1$

4.1 Scheme (a), Constructed by an Affine Plane

Scheme (a) in Table 2 is constructed by an affine plane of order d. As stated in Subsection 2.2, if d is a prime power, then an $AG(2,d)$ exists. However, there may be some affine plane of order d when d is not a prime power. The construction in this subsection works with those affine planes as well.

First, we need to verify $AG(2,d)$ as a distribution design. This will be done in the following lemma.

Lemma 4.1. *Suppose that d is a prime power and t is a positive integer such that $t(t-1)<2d$. Let D be a design where the support set is the set of points of an $AG(2,d)$ and the block set is the set of lines in the same $AG(2,d)$, then D is a $(t,d(t-1),dt-\binom{t}{2})$ - distribution design over the support set of size d^2*

Proof. Since d is a prime power, we know that an affine plane of order d exists. Denote t blocks of D by $L_1,L_2,...,L_t$.

We know by Theorem 2.14 that there are d^2 points and d^2+d lines in $AG(2,d)$. So D has the support set of size d^2. Furthermore, since each line of an $AG(2,d)$ contains exactly d points and lines in the same parallel class do not intersect each other (thus contain the most points), we have $\left|\bigcup_{i=1}^{t-1} L_i\right| \leq d(t-1)$.

Note that $\left|\bigcup_{i=1}^{t} L_i\right|$ is the smallest when there are the most intersecting points. So $\left|\bigcup_{i=1}^{t} L_i\right|$ is the smallest when L_i intersects L_j for all $i,j=1,2,...,t$ such that $i\neq j$.

Since 1 line of an $AG(2,d)$ contains d points, any 2 lines intersect at most 1 point and there are $\binom{t}{2}$ ways to pair any 2 of t lines, we have $\left|\bigcup_{i=1}^{t} L_i\right| \geq dt-\binom{t}{2}$.

Since $t(t-1)<2d$, we get that $d(t-1)<dt-\binom{t}{2}$.

Thus, D is a $(t,d(t-1),dt-\binom{t}{2})$ - distribution design over the support set of size d^2. □

Now, using Lemma 4.1, we can obtain Scheme (a) as in Theorem 4.2.

Theorem 4.2. *Suppose that d is a prime power, Q is a prime power such that $Q\geq d^2+1$, and t is a positive integer such that $t(t-1)<2d$, then there exists an (t,n,d) - RTS with shares from $\mathbb{F}_Q^d$ that has restricted repairability, where n is an integer such that $2d\leq n\leq d^2+d$.*

Proof. Let D be the same design as in Lemma 4.1. Then we know that D is a $(t, d(t-1), dt-\binom{t}{2})$ - distribution design over the support set of size d^2.

Since lines of $AG(2,d)$ can be partitioned into parallel classes of size d (each line contains exactly d points) where lines in each class do not intersect each other and there are exactly d^2 points in an $AG(2,d)$, each parallel class contains all the points of $AG(2,d)$.

Thus, each treatment occurs in exactly 2 blocks of $2d$ blocks from 2 parallel classes of D. So D is repairable and we can accommodate any number of n parties such that $2d \leq n \leq d^2+d$. By Theorem 3.2, since D is a $(t, d(t-1), dt-\binom{t}{2})$ - distribution design and is repairable with blocks of size d, we can use D and a $(d(t-1), dt-\binom{t}{2}, d^2)$ - ramp scheme defined over $\mathbb{F}_Q$ to construct a (t,n,d) - RTS with shares from $\mathbb{F}_Q^d$ having restricted repairability, information rate $1-\frac{t(t-1)}{2d}$ and communication complexity $\frac{2d}{2d-t(t-1)}$, where $2d \leq n \leq d^2+d$.

Furthermore, by Theorem 3.9 and Theorem 2.14, we obtain the repairability index of this scheme which is equal to $\frac{d^d}{\binom{d^2+d-1}{d}}$. □

In Table 3, we will compare Scheme (a) to schemes from Table 1 in terms of information rate, communication complexity, and repairability index for each possible t and d for Scheme (a), and considering n in the same range or value of each comparing scheme.

From now on, in this subsection and all the following subsections, if we write (i) $<$ (j), it means that the parameter of that cell in the Scheme (i) is less than the same parameter in the Scheme (j), same goes for (i) $=$ (j), (i) $>$ (j), (i) $\leq$ (j), etc. Also, if a cell is colored green, it means that our scheme gives better results than the comparing scheme from Table 1. If a cell is colored red, it means that the scheme from Table 1 gives better results. And if a cell is white, it means that the parameter of our scheme is equal to the parameter of the scheme from Table 1. Note that, since Scheme (6) and (7) have no comparable repairability index, the cells of tables that compare the repairability index of schemes from Table 2 with Scheme (6) and (7) are marked with a hyphen just like in Table 1.

As we observe from Table 3, Scheme (a)'s parameters, including information rate, communication complexity, and repairability index, are mostly better than or equal to parameters of schemes from Table 1, except Scheme (7)'s information rate and communication complexity which are better than Scheme (a) in some cases and Scheme (5)'s information rate and repairability index which has the greatest possible value. However, the communication complexity of Scheme (a) is smaller compared to Scheme (5) just as we hoped. Scheme (a) is also more flexible than Scheme (1) - (4) in terms of t and d, but Scheme (1) - (4) are more flexible in terms of n. Note that there is always an m for each of Scheme (1) - (4) that makes Scheme (a)'s information rate, communication complexity, and repairability index equal to the parameters of that scheme.

Table 3: Comparison of Scheme (a) to schemes in Table 1

Threshold and Repairing degree	Scheme (a)'s parameters	Comparing scheme	Parameters of the comparing scheme corresponding to (a)	Information rate comparison	Communication complexity comparison	κ comparison
$t=2$, $d=2$	$4 \le n \le 6$	(5)	$4 \le n \le 6$	(5) > (a)	(5) > (a)	(5) ≥ (a)
$t=2$, $d=3$	$6 \le n \le 12$	(1)	$m=9$, $(6 \le n \le 12)$	(1) = (a)	(1) = (a)	(1) = (a)
			$m=15$, $(10 \le n \le 35)$			(1) < (a)
		(7)	$\tau=3$, $m=5$, $n=10$	(7) < (a)	(7) > (a)	-
$t=2$, $d=4$	$8 \le n \le 20$	(2)	$m=16$, $(8 \le n \le 20)$	(2) = (a)	(2) = (a)	(2) = (a)
			$m=28$, $(14 \le n \le 63)$			(2) < (a)
			$m=40$, $(20 \le n \le 130)$			(2) < (a)
		(6)	$m=8$, $n=14$	(6) < (a)	(6) > (a)	-
$t=2$ or 3, $d=5$	$10 \le n \le 30$	(3)	$m=25$, $(10 \le n \le 30)$	(3) = (a)	(3) = (a)	(3) = (a)
			$m=45$, $(18 \le n \le 99)$			(3) < (a)
			$m=65$, $(26 \le n \le 208)$			(3) < (a)
		(7)	$\tau=5$, $m=7$, $n=21$	(7) < (a)	(7) > (a)	-
$t=2$, 3 or 4, $d=8$	$16 \le n \le 72$	(4)	$m=64$, $(16 \le n \le 72)$	(4) = (a)	(4) = (a)	(4) = (a)
			$m=120$, $(30 \le n \le 255)$			(4) < (a)
			$m=176$, $(44 \le n \le 550)$			(4) < (a)
			$m=232$, $(58 \le n \le 957)$			(4) < (a)
			$m=288$, $(72 \le n \le 1476)$			(4) < (a)
		(7)	$\tau=8$, $m=10$, $n=45$	(7) < (a)	(7) > (a)	-
Other t,d satisfying (a)'s conditions	$2d \le n \le d^2+d$	(7)	$2d \le n = \frac{\binom{m}{\tau}}{\binom{d}{\tau}} \le d^2+d$, as m, τ satisfying conditions of Scheme (7)	(7) < (a) when $2(d-1) > t(t-1)$	(7) > (a) when $2d - t(t-1) > 2$	-
				(7) = (a) when $2(d-1) = t(t-1)$	(7) = (a) when $2d - t(t-1) = 2$	
				(7) > (a) when $2(d-1) < t(t-1)$	(7) < (a) when $2d - t(t-1) < 2$	

4.2 Scheme (b), Constructed by a Room Square of Order r (Version 1)

We use a Room square to construct both Scheme (b) and (c), but the parameters of the Room square in this subsection and in Subsection 4.3 are slightly different. In this subsection, when we say a Room square of order r, we define $r = 2d - 1$ as d is the repairable degree of the final scheme. Furthermore, the designs used in this subsection and Subsection 4.3, although both used a Room square to construct, are different. These are the reasons we label the two constructions "Version 1" and "Version 2" with the construction in this subsection being Version 1. We define the design used in this subsection, as well as the proof that it is a distribution design, in Lemma 4.3.

Lemma 4.3. *Let t and d be positive integers where $d \geq 4$ and $\lfloor \frac{t}{2} \rfloor \lceil \frac{t}{2} \rceil < d$, R be a Room square of side $r = 2d - 1$ and D be a design over the support set S and the block set β, where S is the set of all possible unordered pairs of elements from $\{1, 2, ..., 2d-1\} \cup \{\infty\}$ and $\beta = A \cup B$ where $A = \{\{xy : xy \in S \text{ and } xy \text{ contains in the } i^{th} \text{ row of } R\} : 1 \leq i \leq 2d-1\}$, and $B = \{\{xy : xy \in S \text{ and } xy \text{ contains in the } j^{th} \text{ column of } R\} : 1 \leq j \leq 2d-1\}$.*

Then D is a $(t, d(t-1), dt - \lfloor \frac{t}{2} \rfloor \lceil \frac{t}{2} \rceil)$ - distribution design.

Proof. Since $d \geq 4$, By Theorem 2.18 we know that a Room square of side $r = 2d - 1$ exists.

Since there are d nonempty cells in each row and each column of a Room square, we know that the union of any $t - 1$ blocks of D contains at most $d(t - 1)$ treatments.

Since blocks from A do not intersect one another and blocks from B do not intersect one another, we know that the union of t blocks from D contains the least treatments when some blocks are from A and some are from B. Note that a block from A and a block from B have at most 1 common treatment.

So, if a blocks are from A and b blocks are from B, where $a + b = t$, we know that there are at most ab treatments that belong to 2 blocks.

Let $k \in \mathbb{N}$, if t is even, then $\lfloor \frac{t}{2} \rfloor \lceil \frac{t}{2} \rceil = \left(\frac{t}{2}\right)\left(\frac{t}{2}\right) = \frac{t^2}{4}$ and $\left(\lfloor \frac{t}{2} \rfloor + k\right)\left(\lceil \frac{t}{2} \rceil - k\right) = \left(\frac{t}{2} + k\right)\left(\frac{t}{2} - k\right) = \frac{t^2}{4} - k^2$. Thus, ab is the largest when $a = \lfloor \frac{t}{2} \rfloor$ and $b = \lceil \frac{t}{2} \rceil$ (or vice versa).

Similarly, if t is odd, then we have $\lfloor \frac{t}{2} \rfloor \lceil \frac{t}{2} \rceil = \left(\frac{t-1}{2}\right)\left(\frac{t+1}{2}\right) = \frac{t^2-1}{4}$, $\left(\lfloor \frac{t}{2} \rfloor - k\right)\left(\lceil \frac{t}{2} \rceil + k\right) = \frac{t^2-1}{4} - k - k^2$ and $\left(\lfloor \frac{t}{2} \rfloor + k\right)\left(\lceil \frac{t}{2} \rceil - k\right) = \frac{t^2-1}{4} - k - k^2$. So ab is the largest when $a = \lfloor \frac{t}{2} \rfloor$ and $b = \lceil \frac{t}{2} \rceil$ (or vice versa).

Thus, the union of t blocks contains at least $dt - \lfloor \frac{t}{2} \rfloor \lceil \frac{t}{2} \rceil$ treatments. So we can conclude that D is a $(t, d(t-1), dt - \lfloor \frac{t}{2} \rfloor \lceil \frac{t}{2} \rceil)$ - distribution design. □

With the distribution design from Lemma 4.3, we obtain an RTS as in the following theorem.

Theorem 4.4. *Let t and d be positive integers where $d \geq 4$ and $\lfloor \frac{t}{2} \rfloor \lceil \frac{t}{2} \rceil < d$, Q is a prime power such that $Q \geq d^2 - d + 1$. Then there exists an $(t, 4d-2, d)$ - RTS with shares from $\mathbb{F}_Q^d$ that has restricted repairability.*

Proof. Let D be the same design as in Lemma 4.3. Then we know that D is a $(t, d(t-1), dt - \lfloor \frac{t}{2} \rfloor \lceil \frac{t}{2} \rceil)$ - distribution design.

Note that there are $r + r = 4d - 2$ blocks in β, each block of size d. Note also that the size of the support set S of D is $\binom{2d}{2} = 2d^2 - 1$.

By the definition of Room squares, we know that each treatment in S occurs in exactly 2 blocks, one from A and one from B. Thus, D is repairable.

By Theorem 3.2, since D is a $(t, d(t-1), dt - \lfloor \frac{t}{2} \rfloor \lceil \frac{t}{2} \rceil)$ - distribution design and is repairable with $4d-2$ blocks of size d, we can use D and a $(d(t-1), dt - \lfloor \frac{t}{2} \rfloor \lceil \frac{t}{2} \rceil, 2d^2 - 2)$ - ramp scheme to construct a $(t, 4d-2, d)$ - RTS with shares from $\mathbb{F}_Q^d$ having restricted repairability, information rate $\frac{d - \lfloor \frac{t}{2} \rfloor \lceil \frac{t}{2} \rceil}{d}$ and communication complexity $\frac{d}{d - \lfloor \frac{t}{2} \rfloor \lceil \frac{t}{2} \rceil}$.

We also obtain that the repairability index of this scheme is $\frac{1}{\binom{4d-3}{d}}$. □

Just like in the previous subsection, we now compare Scheme (b) with schemes from Table 1 for each possible t and d for Scheme (b), then consider n that has the same value for both schemes, where the color code has the same meaning as in the previous section. Table 4 is the said comparison.

Table 4: Comparison of Scheme (b) to schemes in Table 1

Threshold and Repairing degree	Scheme (b)'s parameters	Comparing scheme	Parameters of the comparing scheme corresponding to (b)	Information rate comparison	Communication complexity comparison	κ comparison
$t=2$, $d=4$	$n=14$	(2)	$m=16$, $(8\le n\le 20)$ or $m=28$, $(14\le n\le 63)$	(2) = (b)	(2) = (b)	(2) > (b)
		(6)	$m=8$, $n=14$	(6) < (b)	(6) > (b)	-
$t=2$ or 3, $d=5$	$n=18$	(3)	$m=25$, $(10\le n\le 30)$ or $m=45$, $(18\le n\le 99)$	(3) = (b) when $t=2$	(3) = (b) when $t=2$	(3) > (b)
				(3) < (b) when $t=3$	(3) > (b) when $t=3$	
$t=2$,3 or 4, $d=8$	$n=30$	(4)	$m=64$, $(16\le n\le 72)$ or $m=120$, $(30\le n\le 255)$	(4) = (b) when $t=2$	(4) = (b) when $t=2$	(3) > (b)
				(4) < (b) when $t=3,4$	(4) > (b) when $t=3,4$	
Other t,d satisfying Scheme (b)'s conditions	$n=4d-2$	(7)	$n=\frac{\binom{m}{\tau}}{\binom{d}{\tau}}=4d-2$, as m,τ satisfying conditions of Scheme (7)	(7) < (b) when $\frac{1}{d}\left(1+\lceil\frac{t}{2}\rceil\lfloor\frac{t}{2}\rfloor\right)<1$	(7) > (b) when $d-\lceil\frac{t}{2}\rceil\lfloor\frac{t}{2}\rfloor>1$	-
				(7) = (b) when $\frac{1}{d}\left(1+\lceil\frac{t}{2}\rceil\lfloor\frac{t}{2}\rfloor\right)=1$	(7) = (b) when $d-\lceil\frac{t}{2}\rceil\lfloor\frac{t}{2}\rfloor=1$	
				(7) > (b) when $\frac{1}{d}\left(1+\lceil\frac{t}{2}\rceil\lfloor\frac{t}{2}\rfloor\right)>1$	(7) < (b) when $d-\lceil\frac{t}{2}\rceil\lfloor\frac{t}{2}\rfloor<1$	

Similar to Scheme (a), Scheme (b)'s information rate and communication complexity are mostly better than or equal to schemes from Table 1 except Scheme (7) in some cases. It is noteworthy that those two parameters of Scheme (b) are better than most of those from Table 1 when $t>2$. Scheme (b)'s repairability index, however, is less than all of the schemes since we need very specific d parties to reconstruct a lost share.

4.3 Scheme (c), Constructed by a Room Square of Order r (Version.2)

As mentioned in Section 4.2, we also use a Room square to construct Scheme (c), but in a different way than Scheme (b). For starter, we define the order of a Room square r to be $2l-1$ in this section, where l is an integer that is greater or equal to 4 while the repairing degree of this scheme is $n-1$ where $n=2l$. Furthermore, the design we use in this section is different than the one in Section 4.2, the design is defined as follows.

Let $t,l\in\mathbb{N}$ where $l\ge 4$ and $2\le t\le 2l-1$ and let Q be a prime power such that $Q\ge 2l^2-l+1$. Since $l\ge 4$, we know by Theorem 2.18 that the Room square of side $r=2l-1$ exists.

Let R be a Room square of side $r=2l-1$ over the set $\{1,2,...,2l-1\}\cup\{\infty\}$.

Let D be a design over the support set S and the block set β, where S is the set of positions of non-empty cells in R and β is the set of subsets of S that contains i, for each $i\in\{1,2,...,2l-1\}\cup\{\infty\}$.

Example : Consider a Room square of $r=7$ as follows

$\infty 1$				36	27	45
46	$\infty 2$	17			35	
25		$\infty 3$	16	47		
37	15		$\infty 4$			26
	67		23	$\infty 5$	14	
		24	57		$\infty 6$	13
	34	56		12		$\infty 7$

We have $S = \{(1,1),(1,5),(1,6),(1,7)...,(7,2),(7,3),(7,5),(7,7)\}$. And $\beta = \{\{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6),(7,7)\},\{(1,1),(2,3),(3,4),(4,2),(5,6),(6,7),(7,5)\}$, $...,\{(1,6),(2,3),(3,5),(4,1),(5,2),(6,4),(7,7)\}\}$ as $(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)$ and $(7,7)$ are positions of cells that contain ∞, $(1,1),(2,3),(3,4),(4,2),(5,6),(6,7)$ and $(7,5)$ are positions of cells that contain 1, and so on.

Note that $|S| = lr = 2l^2 - l$, $|\beta| = 2l$ and each block is of size $r = 2l - 1$. We verify D as a distribution design and use it to construct an RTS in Theorem 4.5.

Theorem 4.5. *There exist an $(t, n, n-1)-$RTS with shares from $\mathbb{F}_Q^{n-1}$ that has restricted repairability, where $n = 2l$.*

Proof. Since R contains every unordered pair of elements of $\{1,2,...,2l-1\} \cup \{\infty\}$ exactly once, we know that every pair of blocks in β has exactly one common treatment. So the union of t blocks contains at least $tl - \binom{t}{2}$ treatments and the union of $t-1$ blocks contains at most $(t-1)l - \binom{t}{2}$ treatments.

That is D is a $(t, (t-1)l - \binom{t}{2}, tl - \binom{t}{2})-$distribution design.

Since each non-empty cell of a Room square contains 2 elements of $\{1,2,...,2l-1\} \cup \{\infty\}$, we know that each treatment occurs in exactly 2 blocks of D. Thus, D is repairable.

Let the base scheme be a $((t-1)l - \binom{t}{2}, tl - \binom{t}{2}, 2l^2 - l)-$ramp scheme and $n = 2l$, then, by Theorem 3.2 and since D is repairable with $2l$ blocks of size $2l - 1 = n - 1$, we can use D to construct a $(t, n, n-1)-$RTS with information rate $\frac{l}{2l-1} = \frac{n}{2(n-1)}$ and communication complexity $\frac{2l-1}{l} = \frac{2(n-1)}{n}$.

Note that, since this scheme uses all the other parties to repair a lost share, it trivially has universal repairability and so κ of this scheme is 1. □

As we mentioned in Section 3, in each scheme and for each party P_l that lost a share, we want d parties that repair the share to be able to do so by sending exactly one share per party to P_l. We claimed in Section 3 that our results can do so and now that we know the construction of Scheme (a), (b), and (c), it is time we prove that claim. Note that the only case where this way of repairing a share cannot be done is when there exist subshares x_1 and x_2 of P_l such that x_1 and x_2 are held by another party P_k and no other party besides P_l and P_k has x_1 or x_2. However, this case cannot happen in Scheme (a), (b), and (c) since, in the distribution design used for these schemes, each pair of treatments can be together in only one block, i.e. x_1 and x_2 can be together only in P_l so they have to be held by two different parties, not including P_l. Thus, in Scheme (a), (b), and (c), there are d parties that can repair the share of P_l in the way we want.

Just as before, we now compare Scheme (c) with schemes in Table 1 for each possible t and d for Scheme (c). Note that, since $l \geq 4$ and $n = 2l$ in Scheme (c), we get that $d = n - 1$ must be an odd number and $d \geq 7$.

Table 5: Comparison of Scheme (c) to schemes in Table 1

Threshold and Repairing degree	Comparing scheme	Parameters of the comparing scheme corresponding to (c)	Information rate comparison	Communication complexity comparison	κ comparison
d is an odd number such that $d \geq 7$ and $t = d$	(5)	$d+1 \in \mathbb{N}$	(5) > (c)	(5) > (c)	(5) = (c)
d is an odd number such that $d \geq 7$ and $t < d$	(7)	$n = d+1 = \frac{\binom{m}{\tau}}{\binom{d}{\tau}}$, as m, τ satisfying conditions of Scheme (7)	(7) < (c)	(7) > (c)	(7) ≤ (c)

We can see from Table 5 that Scheme (c)'s parameters, including information rate, communication complexity, and repairability index, are better than the parameters of Scheme (5) and (7), except Scheme (5)'s information rate which is more than (c), again, this is not very surprising since the information rate of Scheme (5) has the highest possible value.

4.4 Scheme (d), Constructed by a Symmetric Balanced Incomplete Block Design (SBIBD)

We obtain Scheme (d) from an (n, d, λ)-SBIBD with Theorem 2.8 playing an important part in the proof. Scheme (d) consists of Scheme (d1) and (d2), as an SBIBD can be both a $(2, d, 2d-\lambda)-$distribution design and, with an additional condition, can also be a $(3, 2d-\lambda, 3d-3\lambda)-$distribution design as shown in Lemma 4.6.

Lemma 4.6. *Let $n, d, \lambda \in \mathbb{N}$ where $d \geq 2$, $0 < \lambda < d$. Suppose that an $(n, d, \lambda)-$SBIBD exists, then we obtain the following facts :*

1. *An $(n, d, \lambda)-$SBIBD is a $(2, d, 2d-\lambda)-$distribution design.*
2. *If $d > 2\lambda$, then an $(n, d, \lambda)-$SBIBD is a $(3, 2d-\lambda, 3d-3\lambda)-$distribution design.*

Proof. 1. Since the intersection of any 2 distinct blocks of an $(n, d, \lambda)-$SBIBD contains exactly λ treatments by Theorem 2.8, we know that the union of 2 blocks contains at least $2d-\lambda$ treatments while 1 block contains at most d treatments. Since $d > \lambda$, we have $2d-\lambda > d$. So an $(n, d, \lambda)-$SBIBD is a $(2, d, 2d-\lambda)-$distribution design.

2. Suppose that $d > 2\lambda$. Consider the union of 3 arbitrary blocks of an $(n, d, \lambda)-$SBIBD. suppose that the intersection of those 3 blocks contains x treatments, where $x \leq \lambda$. Then, since the intersection between any 2 blocks contains exactly λ treatments, we know that the union of those 3 blocks contains $3d-3\lambda+x$ treatments. Thus, the union of any 3 blocks of D contains at least $3d-3\lambda$ treatments (which is when $x = 0$).

Furthermore, by Theorem 2.8, we know that the union of any 2 blocks contains exactly $2d-\lambda$ treatments.

Since $d < 2\lambda$, we have $3d-3\lambda > 2d-\lambda$. So an $(n, d, \lambda)-$SBIBD is a $(3, 2d-\lambda, 3d-3\lambda)-$distribution design. □

We still cannot construct RTS for $t > 3$ using an SBIBD since Theorem 2.8 only tells us about the intersection between two blocks. With what we have, we can construct an RTS with $t = 2$ and $t = 3$ as in Theorem 4.7.

Theorem 4.7. *Suppose there exist an $(n, d, \lambda)-$SBIBD over the support set S and the block set β, where $n, d, \lambda \in \mathbb{N}$, $d \geq 2$ and $0 < \lambda < d$. Let D be the said SBIBD and Q be a prime power such that $Q \geq n + 1$. Then*

1. There exist a $(2, n, d)-$RTS with shares from $\mathbb{F}_Q^d$ having restricted repairability.

2. If $d > 2\lambda$, then there exist a $(3, n, d)-$RTS with shares from $\mathbb{F}_Q^d$ having restricted repairability.

Proof. Since D is an $(n, d, \lambda)-$SBIBD over S with the block set β, we know that there are n treatments in S, there are n total blocks in β and each block in β contains d treatments. Furthermore, since the repetition number d of D is greater or equal to 2, we know that each treatment of S is contained in at least 2 blocks of D. Thus, D is repairable.

1. By Lemma 4.6, we know that D is a $(2, d, 2d - \lambda)-$distribution design with n blocks of size d. So, by Theorem 3.2, we can use D and a $(d, 2d - \lambda, n)-$ramp scheme defined over $\mathbb{F}_Q$ to construct a $(2, n, d)-$RTS having restricted repairability with information rate $\frac{d-\lambda}{d} = 1 - \frac{\lambda}{d}$ and communication complexity $\frac{d}{d-\lambda}$.

Furthermore, by Theorem 3.9, we get that the repairability index of this scheme when $\lambda = 1$ is $\frac{(d-1)^d}{\binom{n-1}{d}}$. We denote this scheme we obtained from 1. by (d1).

2. Suppose that $d > 2\lambda$, then we know by Lemma 4.6 that D is a Let the base scheme be a a $(3, 2d - \lambda, 3d - 3\lambda)-$distribution design with n blocks of size d. So, by Theorem 3.2, we can use D and a $(2d - \lambda, 3d - 3\lambda, n)-$ramp scheme defined over $\mathbb{F}_Q$ to construct a $(3, n, d)-$RTS having restricted repairability with information rate $\frac{d-2\lambda}{d} = 1 - \frac{2\lambda}{d}$ and communication complexity $\frac{d}{d-2\lambda}$.

Moreover, just like in 1., we know that the repairability index of this scheme when $\lambda = 1$ is $\frac{(d-1)^d}{\binom{n-1}{d}}$. We denote this scheme in 2. by (d2). □

We know by Theorem 2.5 that if an $(n, d, \lambda)-$ SBIBD exists then $d(d-1) = \lambda(n-1)$, conversely, if $d(d-1) \neq \lambda(n-1)$, then we know that the $(n, d, \lambda)-$ SBIBD does not exist and so we need not concern ourselves with that case of n, d, λ. Note that just because a case of n, d, and λ satisfies the equation above does not mean the $(n, d, \lambda)-$ SBIBD exists, just that it may exist. However, when $\lambda = 1$ and $n = d(d-1)+1$, we obtain by substituting n in Theorem 2.16 with $d-1$ and using the fact that an $AG(2, d-1)$ exists when $d-1$ is a prime power that an $(d(d-1)+1, d, 1)-$ SBIBD exists when $d-1$ is a prime power. Thus, by Theorem 4.7, we can always find a $(2, n, d)-$RTS when $d-1$ is a prime power and $n = d(d-1)+1$. Furthermore, if $d \geq 3$, we can find a $(3, n, d)-$RTS when $d-1$ is a prime power and $n = d(d-1)+1$. So we obtain the following corollary.

Corollary 4.8. *Let $d \in \mathbb{N}$ such that $d-1$ is a prime power and Q be a prime power such that $Q \geq d^2 - d + 2$. Then*

1. There exist a $(2, d^2 - d + 1, d)-$RTS with shares from $\mathbb{F}_Q^d$ having restricted repairability.

2. If $d \geq 3$, then there exist a $(3, d^2 - d + 1, d)-$RTS with shares from $\mathbb{F}_Q^d$ having restricted repairability.

Now, as we have done for Scheme (a), (b), and (c), we have to check whether Scheme (d) can repair the share of P_l the way we want, that is, whether there are d parties that can repair P_l's share by sending exactly one share per party to P_l. Recall that there would be a problem only when there exist subshares x_1 and x_2 of P_l such that x_1 and x_2 are held by another party P_k and no other party besides P_l and P_k has x_1 or x_2. In Scheme (d), this case may happen when $d = 2$ since d is also the repetition number of the SBIBD used in the scheme, but since $\lambda < d$ in Scheme (d), we get that, when d = 2, λ can only be 1, meaning that no other party besides P_l can have both x_1 and x_2.

Next, we compare Scheme (d1) and (d2) with schemes from Table 1 for each possible t and d for Scheme (d1) and (d2), and consider n that has the same value in both schemes just like

in the previous sections. Since the information rate and communication complexity depend on different parameters than the repairability index, we will separate the comparing table of repairability index from the other two parameters. The comparison of the information rate and communication complexity of Scheme (d1) and (d2) with schemes from Table 1 is shown in Table 6 and Table 7 respectively.

We can see from Table 6 and Table 7 that the common n's between Scheme (d1) and (d2) are mostly pair with $\lambda = 1$. We obtain from Table 6 that the information rate and communication complexity of Scheme (d1) is equal to Scheme (1) - (4). On the other hand, when there are common n's that pair with $\lambda = 2$, the parameters of Scheme (d1) are worse than (1) - (4). However, (d1) still has better communication complexity than Scheme (5) and better information rate and communication complexity than Scheme (7) just like our previous schemes.

Table 6: Comparison of information rate and communication complexity of Scheme (d1) to schemes in Table 1

Threshold and Repairing degree	Scheme (d1)'s parameters	Comparing scheme	Parameters of the comparing scheme corresponding to (d1)	Information rate comparison	Communication complexity comparison
$t=2$, $d=2$	$n=3$, $\lambda=1$	(5)	$n=3$	(5) > (d1)	(5) > (d1)
$t=2$, $d=3$	$n=7$ when $\lambda=1$, $n=3$ when $\lambda=2$	(1)	$m=9$, $n=7$	(1) = (d1) when $\lambda=1$	(1) = (d1) when $\lambda=1$
$t=2$, $d=4$	$n=13$ when $\lambda=1$, $n=7$ when $\lambda=2$, $n=5$ when $\lambda=3$	(2)	$m=16$, $n=13$	(2) = (d1) when $\lambda=1$	(2) = (d1) when $\lambda=1$
$t=2$, $d=5$	$n=21$ when $\lambda=1$, $n=11$ when $\lambda=2$, $n=6$ when $\lambda=4$	(3)	$m=25$, $n=11,21$	(3) = (d1) when $\lambda=1$	(3) = (d1) when $\lambda=1$
				(3) > (d1) when $\lambda=2$	(3) < (d1) when $\lambda=2$
		(7)	$\tau=5$, $m=6$, $n=6$ and $\tau=5$, $m=7$, $n=21$	(7) < (d1)	(7) > (d1)
$t=2$, $d=8$	$n=57$ when $\lambda=1$, $n=29$ when $\lambda=2$, $n=15$ when $\lambda=4$, $n=9$ when $\lambda=7$	(4)	$m=64$, $n=29,57$	(4) = (d1) when $\lambda=1$	(4) = (d1) when $\lambda=1$
				(4) > (d1) when $\lambda=2$	(4) < (d1) when $\lambda=2$
		(7)	$\tau=7$, $m=10$, $n=15$	(7) < (d1)	(7) > (d1)
Other cases of t, d	$n\in\mathbb{N}$ such that (n,d,λ)−SBIBD exists where $0<\lambda<d$	(7)	$n=\binom{m}{\tau}/\binom{d}{\tau}$, as m,τ satisfying conditions of Scheme (7)	(7) < (d1) when $\lambda<d-1$	(7) > (d1) when $\lambda<d-1$
				(7) = (d1) when $\lambda=d-1$	(7) = (d1) when $\lambda=d-1$

Table 7: Comparison of information rate and communication complexity of Scheme (d2) to schemes in Table 1

Threshold and Repairing degree	Scheme (d2)'s parameters	Comparing scheme	Parameters of the comparing scheme corresponding to (d2)	Information rate comparison	Communication complexity comparison
$t=3$, $d=3$	$n=7$, $\lambda=1$	(5)	$n=7$	(5) > (d2)	(5) > (d2)
$t=3$, $d=5$	$n=21$ when $\lambda=1$, $n=11$ when $\lambda=2$	(3)	$m=25$, $n=11,21$	(3) < (d2) when $\lambda=1$	(3) > (d2) when $\lambda=1$
				(3) > (d2) when $\lambda=2$	(3) < (d2) when $\lambda=2$
$t=3$, $d=8$	$n=57$ when $\lambda=1$, $n=29$ when $\lambda=2$	(4)	$m=64$, $n=29,57$	(4) < (d2) when $\lambda=1$	(4) > (d2) when $\lambda=1$
				(4) > (d2) when $\lambda=2$	(4) < (d2) when $\lambda=2$
Other cases of t, d	$n\in\mathbb{N}$ such that (n,d,λ)−SBIBD exists where $0<\lambda<\frac{d}{2}$	(7)	$n=\binom{m}{\tau}/\binom{d}{\tau}$, as m,τ satisfying conditions of Scheme (7)	(7) < (d2) when $2\lambda<d-1$	(7) > (d2) when $2\lambda<d-1$
				(7) = (d2) when $2\lambda=d-1$	(7) = (d2) when $2\lambda=d-1$

In Table 8, we compare the repairability index of Scheme (d1) and (d2) to schemes in Table 1. Since (d1) and (d2) have the same repairability index, in Table 8, we will write (d) to refer to both schemes. In this table, we only use the case of n where $\lambda=1$ to calculate κ since the formula we have in Table 2 only applies for $\lambda=1$. This works well for us since we know for sure by the proof of Corollary 4.8 that a $(d(d-1)+1,d,1)-$ SBIBD exists when $d-1$ is a prime power. Note that we also know that the $(3,2,1)-$ SBIBD for the case $t=d=2$ exists despite $d-1$ not being a prime power because if we let the blocks set be the set of all possible pairing of 3 treatments, it satisfies all the conditions of an SBIBD.

From Table 8, we learn that the repairability index of Scheme (d) is greater than Scheme (1) - (4) for all possible n's that satisfy the condition of Corollary 4.8. However, similar to Scheme (a), we have (5) > (d) when $t=d=3$. We get (5) = (d) when $t=d=2$ since the SBIBD used in this case has very few blocks that trivially achieve universal repairability.

Table 8: Comparison of Scheme (d)'s repairability index to schemes in Table 1

Threshold and Repairing degree	Scheme (d)'s parameters	Comparing scheme	Parameters of the comparing scheme corresponding to (d)	Repairability index (κ) comparison
$t = 2$, $d = 2$	$n = 3$	(5)	$n = 3$	(5) = (d)
$t = 3$, $d = 3$	$n = 7$	(5)	$n = 7$	(5) > (d)
$t = 2$, $d = 3$	$n = 7$	(1)	$m = 9$, $n = 7$	(1) < (d)
$t = 2$, $d = 4$	$n = 13$	(2)	$m = 16$, $n = 13$	(2) < (d)
$t = 2$ or 3, $d = 5$	$n = 21$	(3)	$m = 25$, $n = 21$	(3) < (d)
$t = 2$ or 3, $d = 8$	$n = 57$	(4)	$m = 64$, $n = 57$	(3) < (d)

5 Comparison Between the Main Results

Now that we cover all the details of schemes in Table 2, we then compare their parameters between themselves. Recall that, in Section 4, we compare schemes from Table 1 and Table 2 for each t and d when considering n in the same values or intervals. However, we cannot do the same comparing the four schemes in Table 2 since we cannot find any mutual n for all four schemes when considering d and t in the same values. So, in this section, we compare the four schemes when the parameters d and t are the same while n varies between the four schemes, depending on the conditions of each scheme. We choose to consider d and t instead of n because the information rate, communication complexity, and repairability index of the schemes in Table 2 depend on t and d more than n.

In Table 9, we calculate and compare their information rate in some cases of t, d, and λ. Since the communication complexity of schemes from Table 2 is the reciprocal of the information rate, we can obtain the comparison of communication complexity from Table 9 as well. The cells that are colored green in each row mean that the scheme in the cell's column has the best result for that row's case.

We can obtain directly from Table 2 that the information rate of Scheme (d1) and (d2) will only decrease as λ gets bigger, and if $\lambda = 1$, the information rate of Scheme (d1) is equal to Scheme (a) and (b) (as in row 3) and the information rate of Scheme (d2) is equal to Scheme (b) (as in row 5). Furthermore, when $\lambda = 1$, we obtain that Scheme (a)'s information rate is less than Scheme (d2) which means that, at the same d and t, (a) < (b) in term of information rate. In fact, (a) $\leq$ (b) for all t, d with the equality holds only when $t = 2$. When $\lambda > 1$, we also obtain that the information rate of Scheme (d1) and (d2) is less than Scheme (a) and (b). So we can conclude that (b) $\geq$ (a) $\geq$ (d1) for any t, d, λ in terms of information rate (and the opposite for communication complexity). Furthermore, in terms of information rate, (b) $\geq$ (a) > (d2) for all d when $1 < \lambda < \frac{d}{2}$ and (b) = (d2) > (a) for all d when $\lambda = 1$.

Scheme (c) is a little complicated to compare with the other schemes, in the third and fourth

rows of the table, the information rate of (c) is the lowest but then in rows 5 and 6, (c)'s information rate becomes the greatest.

From the expressions in Table 2, we get that, in terms of information rate, (c) > (a) when $t(t-1) > n-2$ (as in row 7 to 9), (c) > (b) when $t^2 > 2(n-2)$ (as in row 7), (c) > (d1) when $\lambda < \frac{n-2}{2}$ (as in row 1) and (c) > (d2) when $\lambda < \frac{n-2}{4}$ (as in row 2). So we need to consider the comparison of Scheme (c) case by case, however, from all the previous inequality, we can practically say that (c) has better results in terms of information rate and communication complexity compared to other schemes from Table 2 if t's value is close enough to n (which is equal to $d+1$ for Scheme (c)).

Table 9: Comparison of information rate in some cases of t, d and λ between schemes from Table 2

Cases \ Scheme	(a)	(b)	(c)	(d1)	(d2)
$t=2,\ d=9,$ $\lambda=6$	$\frac{8}{9}$	$\frac{8}{9}$	$\frac{5}{9}$	$\frac{3}{9}$	-
$t=3,\ d=9,$ $\lambda=3$	$\frac{2}{3}$	$\frac{7}{9}$	$\frac{5}{9}$	-	$\frac{1}{3}$
$t=2,\ d=11,$ $\lambda=1$	$\frac{10}{11}$	$\frac{10}{11}$	$\frac{6}{11}$	$\frac{10}{11}$	-
$t=2,\ d=11,$ $\lambda=2$	$\frac{10}{11}$	$\frac{10}{11}$	$\frac{6}{11}$	$\frac{9}{11}$	-
$t=3,\ d=11,$ $\lambda=1$	$\frac{8}{11}$	$\frac{9}{11}$	$\frac{6}{11}$	-	$\frac{9}{11}$
$t=3,\ d=11,$ $\lambda=2$	$\frac{8}{11}$	$\frac{9}{11}$	$\frac{6}{11}$	-	$\frac{7}{11}$
$t=4,\ d=7$	$\frac{1}{7}$	$\frac{3}{7}$	$\frac{4}{7}$	-	-
$t=4,\ d=9$	$\frac{1}{3}$	$\frac{5}{9}$	$\frac{5}{9}$	-	-
$t=4,\ d=11$	$\frac{5}{11}$	$\frac{7}{11}$	$\frac{6}{11}$	-	-

Next, we will compare the repairability index of schemes from Table 2 in some cases of d. Since the repairability index of Scheme (c) is 1 which is the highest possible value, we know that the repairability index of Scheme (c) is always the highest among schemes from Table 2. Thus, in Table 10, we only compare κ of the Scheme (a), (b), and (d) (as (d1) and (d2) both have the same κ, we will refer to both schemes as Scheme (d)). Again, the scheme in the column that is colored green is the best in each case.

Just like in Subsection 4.4, we only use $\lambda = 1$ and n that satisfies the conditions of Corollary 4.8. Furthermore, since we only confirmed the existence of $(n, d, 1)-$ SBIBD when $d-1$ is a prime power, we will consider only the case of d such that $d-1$ is a prime power in Table 10. Additionally, d must also be a prime power that is greater or equal to 4 so that it satisfies the conditions of Scheme (a) and (b) as well. We still obtain κ of Scheme (a) and (b) in the regular way which is by substituting d.

Table 10: Comparison of repairability index in some cases of d between schemes from Table 2

Cases \ Scheme	(a)	(b)	(d)
$d = 4$	≈ 0.066	≈ 0.0014	≈ 0.1636
$d = 5$	≈ 0.02631	≈ 0.00016	≈ 0.066
$d = 8$	≈ 0.00158	$\approx 2.3\times10^{-7}$	≈ 0.00406
$d = 9$	≈ 0.00061	$\approx 2.6\times10^{-8}$	≈ 0.00158

In these cases of d, we can see that Scheme (d) gives the best result, with Scheme (a) as the second runner-up. Since we obtained Corollary 4.8, that is, the existence of a $(d(d-1)+1, d, 1)-$ SBIBD from $AG(2, d-1)$, we get that the number of blocks n of an $AG(2, d-1)$ and the number of blocks of $(d(d-1)+1, d, 1)-$ SBIBD are equal. So κ of Scheme (d) in the case of Corollary 4.8 when calculating at d is equal to κ of Scheme (a) at $d-1$ as we can see in Table 10.

In conclusion, Scheme (b) is the best for optimizing the information rate and communication complexity, while Scheme (c) has the best repairability index since it has universal repairability. However, Scheme (c)'s property of universal repairability is trivial since, when a party P_l loses its share, it requires all other parties to repair the lost share. Scheme (d), on the other hand, does not involve all other parties besides P_l in the reconstruction of P_l's share, and the results in Table 10 lead us to think that it has second-best repairability index next to Scheme (c), though more study is required to confirm this fact.

Notice that, of all schemes in Table 2, Scheme (a) is the only one with various possible n's for each t and d while the other schemes have fixed n. So even though Scheme (a) does not have optimal information rate, communication complexity, or repairability index, the flexibility of its parameter makes it useful when we need an RTS for some specific t, d, and n that does not satisfy the conditions of Scheme (b), (c), or (d).

References

[1] A. Shamir, *How to share a secret*, Communications of the ACM. **22**(11) (1979), 612–613.

[2] B. Kacsmar and D. R. Stinson, *A network reliability approach to the analysis of combinatorial repairable threshold schemes*, Advances in Mathematics of Communications. **13**(4)(2019), 601–612.

[3] D. Evans, V. Kolesnikov and M. Rosulek, *A pragmatic introduction to secure multi-party computation*, NOW Publishers, 2018.

[4] D. R. Stinson, *Combinatorial designs: constructions and analysis*, Springer, New York, 2004.

[5] D. R. Stinson and R. Wei, *Combinatorial repairability for threshold schemes*, Designs, Codes and Cryptography,**86**(1) (2017), 195–-210.

[6] T. M. Liang and D. R. Stinson, *A survey and refinement of repairable threshold schemes*, Journal of Mathematical Cryptology. **12**(1) (2018), 57–81.

[7] W. D. Wallis, *Introduction to combinatorial designs*, 2nd ed., Chapman & Hall/CRC, 2007.

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Ternary LDPC Codes Based on Projective Plane

Chanya Lawong[1,†] and Penying Rochanakul[2,‡]

[1]Graduate Master Degree Program in Applied Mathematics, Department of Mathematics, Faculty of Science
Chiang Mai University, Chiang Mai 50200, Thailand

[2]Department of Mathematics, Faculty of Science
Chiang Mai University, Chiang Mai 50200, Thailand

Abstract

Since the late 1990s, low-density parity-check (LDPC) codes have emerged as highly efficient error-correcting codes and extensively utilized in communication systems. Tanner graphs are considered one of the powerful LDPC codes representations. In this work, we consider tanner graphs for ternary LDPC codes over finite fields and integer residue rings. In particular, we expand the method for constructing tanner graphs of binary LDPC codes proposed by Polak and Zhupa into a ternary Tanner graph over a finite field and integer residue rings. This extension introduces a novel approach within the field, aiming to explore the potential benefits and applications of ternary LDPC codes.

Keywords: low-density parity-check (LDPC) codes, tanner graphs, binary LDPC codes, ternary LDPC codes.

2020 MSC: Primary 94B05; Secondary 05C50, 05C75.

1 Introduction

Low-density parity-check (LDPC) codes are a class of error-correcting codes widely used in modern digital communication systems. These codes were initially proposed by Robert G. Gallager [3, 4] in 1960 but gained practical significance in the early 2000s due to the discovery of efficient decoding algorithms. LDPC codes are characterized by their sparse parity check matrices, which enable efficient decoding. They offer excellent error correction performance, approaching the theoretical limits defined by Shannon's channel coding theorem [1]. LDPC codes are employed in various communication standards and applications, including wireless communication, digital video broadcasting, satellite communication, and storage systems. Decoding LDPC codes involves iteratively exchanging messages between variable nodes and check nodes, gradually improving the estimates of the transmitted codeword until convergence. Overall, LDPC codes play a crucial role in enabling reliable data transmission over noisy channels in modern communication systems.

†Speaker. ‡Corresponding author.
Email: Chanya_l@cmu.ac.th (C. Lawong), Penying.Rochanakul@cmu.ac.th (P. Rochanakul)

Parity check is a technique used in digital communication systems to verify whether transmitted data has been corrupted during transmission. It involves adding extra bits to the transmitted data to ensure accuracy. By employing parity check, the system can detect errors in the transmitted data and, in some cases, correct them. The Tanner graph, introduced by Michael Tanner in 1981 [6], provides a graphical representation of the relationships between the bits and parity check equations in a code. This bipartite graph consists of variable nodes representing bits and check nodes representing parity check equations, with edges denoting the connections between them. Understanding the structure of LDPC codes through Tanner graphs is crucial for developing efficient decoding algorithms and optimizing code performance in various communication systems.

Research on ternary LDPC codes has explored their design, performance analysis, decoding algorithms, and applications. Methods include protograph-based designs, progressive edge growth algorithms, and density evolution techniques. Performance evaluation involves simulations and theoretical analysis under different channel conditions. Decoding algorithms like belief propagation and sum-product algorithm are proposed to efficiently decode ternary LDPC codes. Potential applications include high-speed communication over noisy channels, optical communication systems, and storage systems employing multi-level flash memory.

2 Preliminaries

In this section, we will review the article's fundamental content, which is divided into two parts: linear codes and Tanner graphs.

2.1 Linear Codes

Let n be a fixed positive integer and let the input and output symbols of the channel belong to $\mathbb{F}_q$ the finite field with q elements. The set of q-ary vectors is denoted by $\mathbb{F}_q^n$.

A distance between two vectors x and y in $\mathbb{F}_q^n$ is the number of coordinates, where x and y differ.

Definition 2.1. The *Hamming distance* between $x = x_1x_2\dots x_n$ and $y = y_1y_2\dots y_n$ in $\mathbb{F}_q^n$, denoted by $d(x, y)$, is the number of positions in which x and y are different. That is

$$d(x, y) = |\{i \in \{1, 2, \dots, n\} : x_i \neq y_i\}|$$

We often abbreviate 'Hamming distance' to 'distance'.

Example 2.2. $d(000, 011) = 2$, $d(10101, 11110) = 3$.

A q-ary *block code* C of length n is any nonempty subset of $\mathbb{F}_q^n$. The elements of C are called *codewords*. If $|C| = 1$, the code is called *trivial*. The *minimum distance* d of a non-trivial code C is given by $d = \min\{d(x, y) : x \in C, y \in C, x \neq y\}$. One often refers to a 'block code' as a 'code'. A linear subspace C of $\mathbb{F}_q^n$ is called a *linear code*.

If C has dimension k and minimum distance d, one says that C is an $[n, k, d]$ code. The parameter d in the notation $[n, k, d]$ is sometimes omitted.

A code over the code alphabet $\mathbb{F}_3 = \{0, 1, 2\}$ is called a *ternary code*, while the term *quaternary code* is sometimes used for a code over the code alphabet $\mathbb{F}_4$. However, a code over the code alphabet $\mathbb{Z}_4 = \{0, 1, 2, 3\}$ is also sometimes referred to as a quaternary code.

There are two standard ways of describing a k-dimensional linear subspace: one by means of k independent basis vectors; the other uses $n - k$ linearly independent equations.

Definition 2.3. A *generator matrix* G of an $[n, k, d]$ code C is a $k \times n$ matrix, of which the k rows form a basis of C.

Example 2.4. The matrix $G = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$ is a *generator matrix* of the linear binary codes $C = \{0000, 1001, 0110, 1111\}$.

Definition 2.5. A *parity-check matrix* H of an $[n, k, d]$ code C is an $(n-k) \times n$ matrix, satisfying

$$c \in C \Leftrightarrow cH^t = \underline{0},$$

where H^t denotes the transpose of H and $\underline{0}$ is the all-zeros word of length $n - k$.

In other words C is the null space (solution space) of the $n-k$ linearly independent equations $cH^t = \underline{0}$.

Example 2.6. Show that $H = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$ is a parity check matrix of the linear binary code $C = \{0000, 1001, 0110, 1111\}$.
Solution. Consider all possible binary words of length 4. We have

$(0000)H^t = 00$,	$(0100)H^t = 10$,	$(1000)H^t = 01$,	$(1100)H^t = 11$,
$(0001)H^t = 01$,	$(0101)H^t = 11$,	$(1001)H^t = 00$,	$(1101)H^t = 10$,
$(0010)H^t = 10$,	$(0110)H^t = 00$,	$(1010)H^t = 11$,	$(1110)H^t = 01$,
$(0011)H^t = 11$,	$(0111)H^t = 01$,	$(1011)H^t = 10$,	$(1111)H^t = 00$.

Definition 2.7. Let $G(V_1 \cup V_2, E)$ be a bipartite graph.
• If each vertex in both parts has degrees s and r, respectively, then if $s = r$, G is called a *regular s-graph*, denoted by s-graph.
• If each vertex in both parts has degrees s and r, respectively, and $s \neq r$, then G is called a *bi-regular graph* (s, r), denoted by bi-regularity - (s, r), and (s, r) is referred to as the *bi-degree*.

Example 2.8. The following graph is an example of a regular graph and a bi-regular graph.

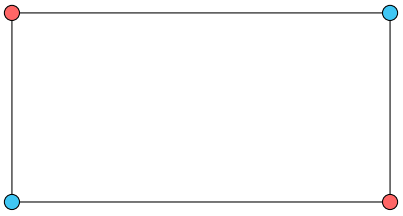


Figure 1: 2 - regular

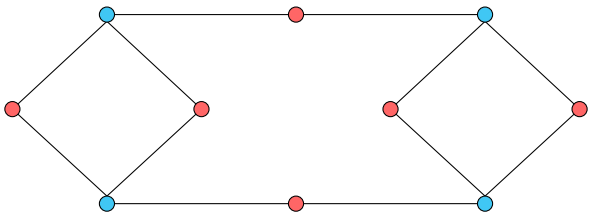


Figure 2: bi-regular - $(2, 3)$

Theorem 2.9 ([7], Theorem 4.5.6)**.** *Let C be a linear code and let H be a parity-check matrix for C. Then $d(C)$ equals to minimum number of columns that give zero linear combination.*

Example 2.10. The minimum distance $d(C)$ of linear binary code C with parity check matrix $H = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$ is 2.

Theorem 2.11 ([5], Definition 2.1.2)**.** *A code with distance d can detect up to $d-1$ errors and correct up to $\lfloor \frac{d-1}{2} \rfloor$ errors.*

Example 2.12. Let $H = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$ is a parity check matrix of the linear binary code $C = \{0000, 1001, 0110, 1111\}$.
Solution. Suppose $u = 1001$ was sent and $v = 1111$ was recieved.
We have $uH^t = 0$ and $vH^t = 0$.
Thus, errors cannot be detected (and cannot be corrected).

Definition 2.13. A *low-density parity-check (LDPC) codes* is a linear code for which the parity-check matrix H has a low density of 1's.

Definition 2.14. A *regular* (n, k) *LDPC codes* is a linear code whose parity-check matrix H contains constant number of 1's in each column, denoted by W_c and contains constant number of 1's in each row, denoted by $W_r = W_c(n/n-k)$ 1's per row, where $W_c << n-k$.

Example 2.15. H is a parity check matrix for a regular $(8, 4)$ LDPC codes where $W_r = 2$ and $W_c = 1$.

$$H = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

2.2 Tanner Graph

In 1981, Tanner [6] introduced a highly effective graphical representation for LDPC codes known as the Tanner graph. This graphical model [9] is bipartite, meaning it consists of two distinct sets of vertices: V_1, representing the codeword bits, and V_2, representing the parity checks. A vertex from V_1 is connected to a vertex from V_2 if and only if the bit corresponding to the V_1 vertex is involved in the parity check corresponding to the V_2 vertex.

Let $V(G)$ be the set of vertices and $E(G)$ be the set of edges of a graph G.

Definition 2.16. Vertices u, v are *adjacent* in G if $\{u, v\} \in E(G)$.

Definition 2.17. An edge $e \in E(G)$ is *incident* to a vertex $v \in V(G)$ if v is an endpoint of e.

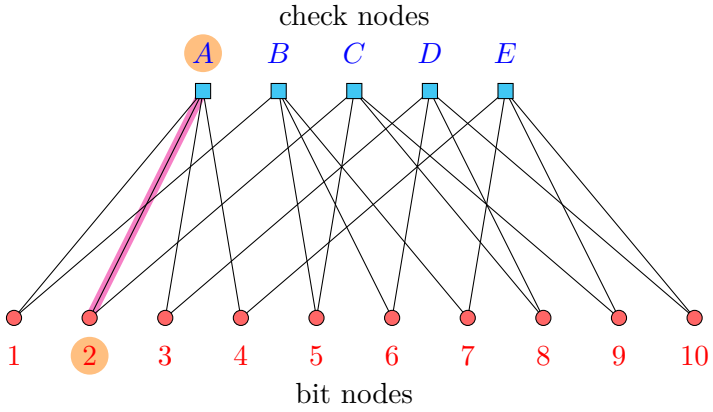
Definition 2.18. The number of edges incident to a vertex v in a graph is called the *degree* of vertex v, denoted by $deg(v)$.

Definition 2.19. A *bipartite graph* is a graph (nodes or vertices connected by undirected edges) whose nodes may be separated into two classes, and where edges may only connect two nodes not residing in the same class.

Definition 2.20. *Tanner graph* of a code is a bipartite graph containing the two classes of nodes, the n *variable nodes* (or *bit* nodes) and the $n-k$ *check nodes* (or *function nodes*). The graph drawn according to the following rule: check node j is connected to variable node i whenever element h_{ji} in H is 1.

Example 2.21. The Tanner graph represents the parity-check matrix H, where the presence of 1 in the matrix corresponds to connections between check nodes and bit nodes. Highlighted in magenta, the graph illustrates the connection between check node A and bit node 2.

$$H = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{matrix} \\ \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \end{array} \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix}$$



$n = 10$
$m = 5$
$k = 6$
$W_r = 4$
$W_c = 2$

Figure 3: Tanner graph represented parity-check matrix H

The Tanner graph serves as a visual depiction of the parity check equations within a code. There exists a conventional method for constructing error-correcting codes based on the adjacency matrix of a bipartite, bi-regular graph. The parity check matrix H is extracted from the adjacency matrix A of the graph, possessing specified properties essential for code generation:

$$A = \begin{pmatrix} 0 & H \\ H^t & 0 \end{pmatrix}$$

The establishment of a matrix H defines the code configuration. Nonetheless, the parity check matrix is not singular. Rearranging columns does not alter the code characteristics, providing an equivalent code.

A code represented by a *sparse matrix* or a *sparse Tanner graph* is known as an LDPC code [2]. A matrix is considered sparse when the number of ones it contains is significantly smaller than the number of zeros. LDPC codes are characterized by having a very sparse parity check matrix. A sparse graph exhibits a low ratio of edges to vertices. A straightforward expression describing the graph density $G(V, E)$ is as follows:

$$D = \frac{2|E|}{|V|(|V|-1)} \tag{2.1}$$

where $|E|$ is the number of edges and $|V|$ the number of vertices of graph G.

Definition 2.22. A *cycle of length* ℓ in a Tanner graph is a path comprising ℓ edges which closes back on itself.

Example 2.23. The Tanner graph represents the parity-check matrix H, where the presence of 1 indicates connections between check nodes and bit nodes. Highlighted in magenta, the graph illustrates a cycle pattern of length 6, indicating interconnections between nodes in both check and bit nodes.

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

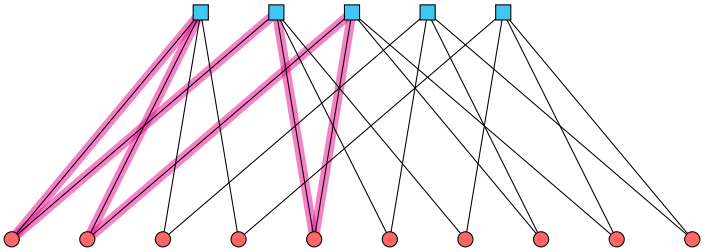


Figure 4: A cycle of length six as seen in both the Tanner graph and the parity check matrix

Definition 2.24. The *girth* of a Tanner graph is the minimum cycle length of the graph.

Remark. [8] The girth of Tanner graph must be more than 4.

3 Main Results

In this section, we will discuss the steps involved in constructing a bipartite graph that can be used as a Tanner graph.

Assume that q is a *prime power*. The quadratic extension of $\mathbb{F}_q$ is $\mathbb{F}_{q^2}$. Ustimenko and Woldar [9] introduced the family of $F = F(\mathbb{F}_q, \mathbb{F}_{q^2})$. Those graphs are bipartite with a set of vertices $V = V_1 \cup V_2$, where $V_1 \cap V_2 = \phi$. They have girths of at least 8 and very different bi-regularities (q, q^2). Due to geometric construction, one partition set, $V_1 = P$, is traditionally referred to as the set of points, and one, $V_2 = L$, as the set of lines:

$$P = \{(a, b, c) : a \in \mathbb{F}_q, b \in \mathbb{F}_{q^2}, c \in \mathbb{F}_q\}$$
$$L = \{[x, y, z] : x \in \mathbb{F}_{q^2}, y \in \mathbb{F}_{q^2}, z \in \mathbb{F}_q\}$$

Two types of brackets are used to distinguish points and lines. We say point (p) is incident to line $[l]$ in graph $F(\mathbb{F}_q, \mathbb{F}_{q^2})$, and we define incidence relation I (between (p) and $[l]$) as: $(a, b, c)I[x, y, z]$ iff

$$\begin{cases} y - b = ax \\ z - c = ay + ay^q \end{cases} \tag{3.1}$$

The set of vertices is $V(F) = P \cup L$, and the set of edges consists of all pairs $((p), [l])$, for which $(p)I[l]$. Because $a \in \mathbb{F}_q, b \in \mathbb{F}_{q^2}, c \in \mathbb{F}_q, x \in \mathbb{F}_{q^2}, y \in \mathbb{F}_{q^2}, z \in \mathbb{F}_q$, we have $|P| = q^4$, $|L| = q^5$, and $|V(F)| = q^5 + q^4 = q^4(q + 1)$.

Instead of using elements of fields $\mathbb{F}_{q^2}$ and $\mathbb{F}_q$ as coordinates, Polak and Zhupa [2] propose to use two rings $\mathbb{Z}_{n^2}, \mathbb{Z}_n$ and modulo operations. In this case, the graph $F(\mathbb{Z}_n, \mathbb{Z}_{n^2})$ has sets P and L are the following:

$$P = \{(a, b, c) : a \in \mathbb{Z}_n, b \in \mathbb{Z}_{n^2}, c \in \mathbb{Z}_n\}$$
$$L = \{[x, y, z] : x \in \mathbb{Z}_{n^2}, y \in \mathbb{Z}_{n^2}, z \in \mathbb{Z}_n\}$$

They define the incidence relation I (between (p) and $[l]$) as: $(a, b, c)I[x, y, z]$ iff

$$\begin{cases} (y - b) \equiv (ax) \pmod{n^2} \\ (z - c) \equiv (ay + ay^n) \pmod{n} \end{cases} \tag{3.2}$$

Graphs with coordinates specified in terms of finite rings are bipartite, bi-regularity (n, n^2), and girth at least 6 (possibly 8, but not tested). In this case, they are not affine parts of generalized quadrangles. There are n elements in the set L, and $|P| = n^4$. There are $|V| = n^4(n + 1)$ elements in the set of vertices and n^6 elements in the set of edges.

Building upon the foundation laid by Polak and Zhupa, who introduced equations 3.1 and 3.2 elucidating the relationship between a set of points and a set of lines, we have formulated a novel equation aimed at further refining the categorization of connections between these entities. In accordance with this equation:

$$(a + c) \equiv (x + z) \pmod{2} \tag{3.3}$$

If the computations follow the equations given above, the edges between the set of points and the set of lines denoted by dashed lines. Otherwise, they will be marked with normal lines.

Example 3.1. Constructing the code of the graph $F(\mathbb{Z}_2, \mathbb{Z}_4)$:

- $\mathbb{Z}_2 = \{0, 1\}$ is a ring with addition and multiplication identities represented by 0 and 1, respectively. Addition and multiplication operations are carried out under the modulus of 2.
- $\mathbb{Z}_4 = \{0, 1, 2, 3\}$ is a ring with 4 elements, where addition and multiplication operations are performed under the modulus of 4.

$+$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

$\cdot$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

Each point (a, b, c) where $a \in \mathbb{Z}_2$, $b \in \mathbb{Z}_4$, and $c \in \mathbb{Z}_2$ is adjacent to the line $[x, y, z]$ where $x \in \mathbb{Z}_4$, $y \in \mathbb{Z}_4$, and $z \in \mathbb{Z}_2$, satisfying the conditions: $y - b \equiv ax \,(\mathrm{mod}\, n^2)$, and $z - c \equiv ay + ay^2 \,(\mathrm{mod}\, n)$, where the components of matrix is 2 when $a + c \equiv x + z \,(\mathrm{mod}\, 2)$.

$$H = \begin{bmatrix}
2&1&2&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&2&1&2&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&2&1&2&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&2&1&2&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&2&1&2&1&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&2&1&2&1&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&2&1&2&1&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&2&1\\
1&0&0&0&0&0&0&0&0&2&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&2&0&0\\
0&0&0&0&1&0&0&0&0&0&0&0&0&2&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0\\
0&0&0&2&0&0&0&0&1&0&0&0&0&0&0&0&0&2&0&0&0&0&0&0&0&0&1&0&0&0\\
0&0&0&0&0&0&0&2&0&0&0&0&1&0&0&0&0&0&0&0&0&0&2&0&0&0&0&0&0&0\\
0&0&1&0&0&0&0&0&0&0&0&2&0&0&0&0&1&0&0&0&0&0&0&0&0&2&0&0&0&0\\
0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&2&0&0&0&0&1&0&0&0&0&0&0&0&2&0\\
0&2&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&2&0&0&0&0&1&0&0&0&0\\
0&0&0&0&0&2&0&0&0&0&0&0&0&0&1&0&0&0&0&0&0&0&0&0&2&0&0&0&1&0
\end{bmatrix}$$

Figure 5: The low-density parity-check matrix H of the graph $F(\mathbb{Z}_2, \mathbb{Z}_4)$ (where 0 indicates no edge between vertex-point pairs, 1 indicates the presence of an edge between vertex-point pairs, and 2 indicates an edge between vertex-point pairs satisfying the condition $(a + c) \equiv (x + z) \,(\mathrm{mod}\, 2)$

3.2. Code construction. In graphs $F(\mathbb{Z}_n, \mathbb{Z}_{n^2})$: We denote with $[N, K]$ the code with code words length N and K information bits. In such a code there are $R = N - K$ parity check equations. The ratio K/N is called *code rate* and is denoted by R_C.

A set of lines is larger than a set of points. Lines correspond to code word bits, while points correspond to parity checks. We decide whether to include one or zero in the parity check matrix by determining whether relations 3.2 between the corresponding points and lines hold. Every bit of the codeword is checked for parity using q checks. In this case, the parity check matrix H is part of the graph's adjacency matrix. Since $|L| = n^5$ and $|P| = n^4$, the size of H is $n^4 \times n^5$ and

$$R_C = \frac{n^5 - n^4}{n^5} = \frac{n^4(n-1)}{n^5} = \frac{n-1}{n}.$$

4 Conclusion

Our work extends the method for constructing Tanner graphs of binary LDPC codes proposed by Polak and Zhupa into a ternary Tanner graph over a finite field and integer residue rings. Ternary LDPC codes are constructed based on the modified family of graphs. The resulting Tanner graphs have girth at least 6 and the resulting ternary LDPC codes have rate at least $(n-1)/n$.

Acknowledgment. The authors wish to express their sincere gratitude to the Applied Mathematics Program, Department of Mathematics, Chiang Mai University, for their invaluable support and guidance throughout this paper. I am deeply thankful to all the faculty members who have contributed to this work with their expertise and insights. Thank you all for your unwavering support and encouragement.

References

[1] C.E. Shannon, *A mathematical theory of communication*, Bell Syst. Tech. J., pp. 372–423, 623–656, 1948.

[2] M. Polak and E. Zhupa, *Graph based linear error correcting codes*, Albanian Journal of Mathematics, Volume 10, Number 1, Pages 37–45 ISSN: 1930-1235; (2016).

[3] R. G. Gallager, *Low-density parity-check codes*. IRE Transactions on information theory, 8(1), 21-28, (1962).

[4] R. G. Gallager, *Low-Density Parity Check Codes*. Cambridge, MA: MIT Press, 1963.

[5] R. Hill, *A first course in coding theory*. Oxford University Press, 1986.

[6] R. Tanner, *A recursive approach to low complexity codes*. IEEE Transactions on information theory, 27(5), 533-547, (1981).

[7] S. Ling and C. Xing, *Coding theory: a first course*. Cambridge University Press, 2004.

[8] T. R. Halford, A. J. Grant, and K. M. Chugg, *Which Codes Have* 4*-Cycle-Free Tanner Graphs?*. IEEE transactions on information theory, 52(9) (2006), 4219-4223.

[9] V. A. Ustimenko and A. J. Woldar, *Extremal properties of regular and affine generalized m-gons as tactical configurations*, European J. Combin. 24, no. 1, 99–111, (2003).

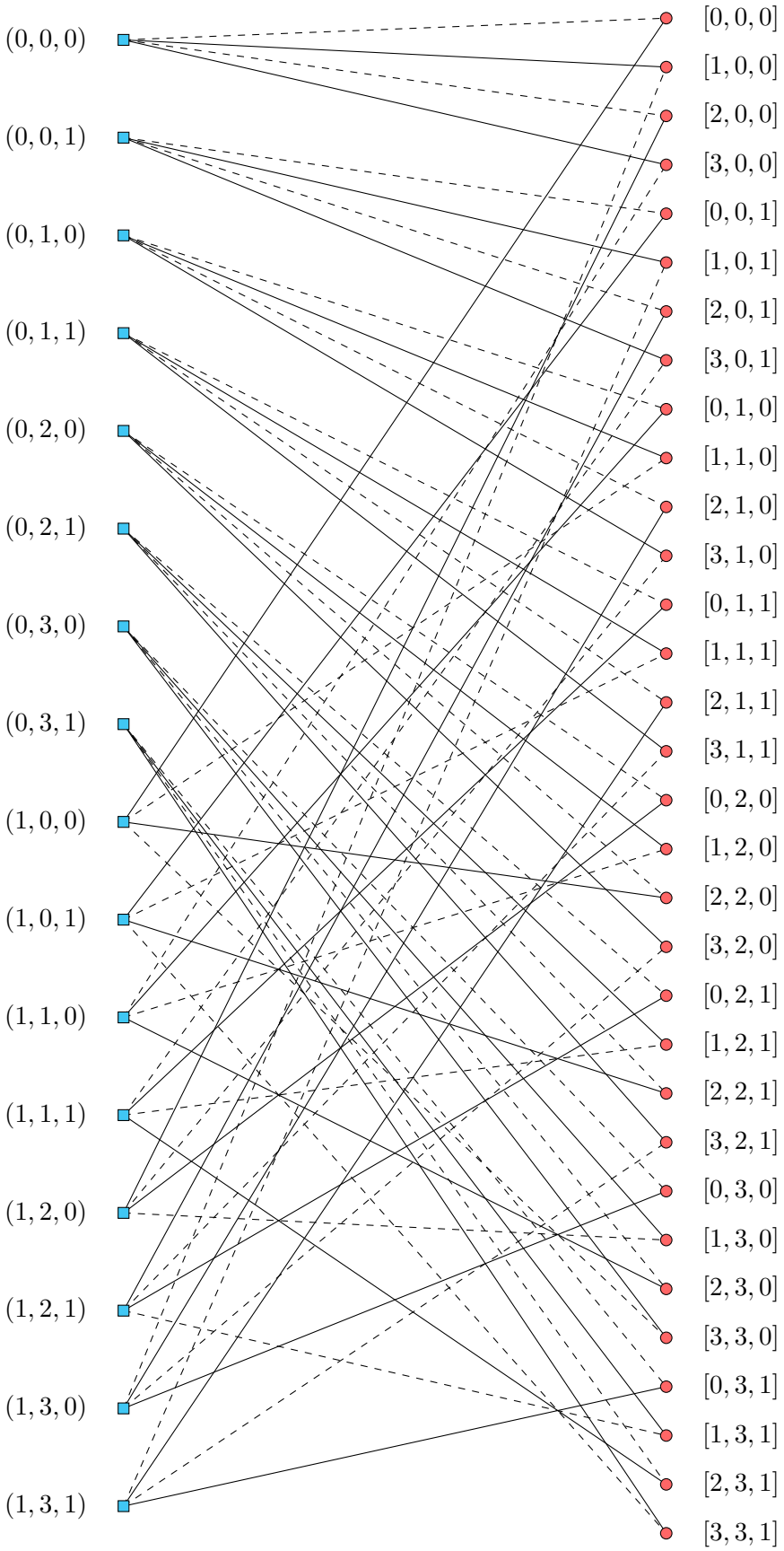


Figure 6: From the graph representing the connectivity from the matrix H in Figure 5, it can be concluded that to depict the edges between vertex-vertex pairs (on the left-hand side: $P = (a, b, c)$) and vertex-edge pairs (on the right-hand side: $L = [x, y, z]$) in the graph $F(\mathbb{Z}_2, \mathbb{Z}_4)$, normal and dashed lines represent adherence to the equation $(a + c) \equiv (x + z) \pmod 2$, respectively

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Solvability Conditions for (n^2-1)-puzzle with 1 or 2 Fixed Cells

Waitin Sinthu-urai[1,†] and Piyashat Sripratak[2,‡]

[1]Graduate Master Degree Program in Applied Mathematics, Department of Mathematics, Faculty of Science
Chiang Mai University, Chiang Mai 50200, Thailand

[2]Department of Mathematics, Faculty of Science
Chiang Mai University, Chiang Mai 50200, Thailand

Abstract

(n^2-1)-puzzle is a puzzle within square board with $n \times n$ unit square cells where $n \geq 3$, labelled as cell $c \in \{1, 2, 3, \ldots, n^2\}$, in order from left to right, and then from the upper row to the lower row. Each of the first n^2-1 cells contains a unit square tile labelled by number $t \in \{1, 2, 3, \ldots, n^2-1\}$. The other cell at the bottom-right corner contains a single hole. Beginning with an initial configuration of the board, a player has to make moves by switching the hole and a tile next to the hole, so that we can transform the board to the configuration that all tiles are arranged in order from 1 to n^2-1 with the hole in the bottom-right corner cell. The more challenging puzzle is when a board consists of some fixed cells. The tile located at a fixed cell cannot be moved. This research focuses on solvability conditions of an initial configuration of a board with a single fixed cell and a board with two fixed cells. We conclude that for an $n \times n$ board with a fixed cell, any even configuration is solvable if and only if the fixed cell is not in $\{2, n-1, n+1, 2n, n^2-2n+1, n^2-n, n^2-n+2, n^2-1\}$. As for a board with two fixed cells, we give conditions on the positions of the fixed cells where not all even configuration are solvable. Moreover, some sufficient conditions that make all even configurations solvable are provided.

Keywords: (n^2-1)-puzzle, solvability, permutation.

2020 MSC: Primary 91A46; Secondary 05A05, 20B05.

1 Introduction

15-puzzle is a puzzle within square board with 4×4 unit square cells. These cells are called *cell* c, where $c = 1, 2, 3, \ldots, 16$, located in order from left to right, and then from the upper row to the lower row. In each of the first fifteen cells, there is a unit square tile with label t where $t = 1, 2, 3, \ldots, 15$. We shortly name the tile labelled by number t as *tile* t. The other cell contains a single hole. The game starts with an initial configuration of the board, where a

†Speaker. ‡Corresponding author.
Email: waitin.sint@gmail.com (W. Sinthu-urai), psripratak@gmail.com (P. Sripratak).

player makes a *move* which is to switch the hole, located at cell 16, and a tile next to the hole. The goal of the game is to transform the board to the target configuration. In this research, the target configuration, called the *standard configuration*, is the board that tile t is located in cell t for all $t = 1, 2, 3, \ldots, 15$ and cell 16 in the bottom-right-corner of the board is left for the hole. The standard configuration is shown in Figure 1.

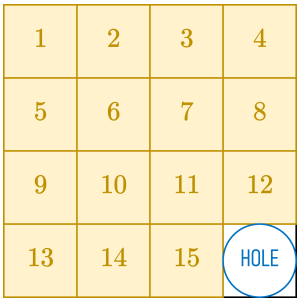


Figure 1: standard configuration of 15-puzzle

An initial configuration of the board is said to be *solvable* if there is a sequence of moves that transforms the initial configuration to the target configuration which is the standard configuration. Otherwise, it is said to be *unsolvable*. In the 1870s, Sam Loyd proposed a dramatic problem of 15-puzzle throughout the world. He came up with a 15-puzzle whose initial configuration was set as Figure 2; the tiles were arranged in order except tile 14 and tile 15 being switched. This famous problem was named after him as Sam Loyd's puzzle or 14-15 puzzle. His problem has inspired lots of mathematicians and computational scientists since it was discovered to be unsolvable [8]. Hence, at the beginning, many of the researchers were interested in solvability of the puzzle [2, 6].

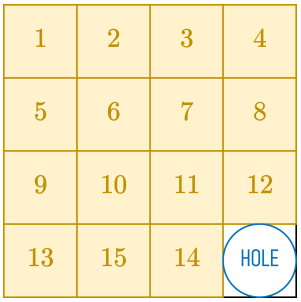


Figure 2: initial configuration of Sam Loyd's puzzle

Afterwards, 15-puzzle has been generalized to $(n^2 - 1)$-puzzle with $n^2 - 1$ unit square tiles labelled by number $1, 2, 3, \ldots, n^2 - 1$ for the positive integer $n \geq 3$. These tiles and a single hole are located within the square board with $n \times n$ unit square cells. The results for the original 15-puzzle can be easily extended to the $(n^2 - 1)$-puzzle. Various appearances of the board have been examined for solvability as well. Johnson and Story [8] and Muralidharan [10] provided necessary and sufficient conditions for solvability of the $m \times n$ (rectangular) board. Davies [4] examined solvability of the $m \times n$ (rectangular) board where the numbers on the tiles are printed diagonally. Liebeck [9] considered solvability of the rotated board. Recently, Hamersma [5] analyzed solvability of the board with hexagonal cells.

Thenceforth, several related problems have been considered. Berenbom et al. [3], Archer [1] and Yang [13] studied the generalized 15-puzzle with graph and vertices instead of board and tiles. One of those vertices is called a blank vertex. The problem is determining whether an initial configuration can be transformed into a target configuration by swapping the blank vertex with its adjacent vertex through the incident line. Berenbom et al. [3] and Archer [1] applied algebra

to analyze the problem whereas Yang [13] did it in graph theoretical way. Besides, Schwartz [12] considered the 4×4 board with colored tiles instead of numbered tiles and modified the objective of the game: no two tiles of the same color are in the same line.

The more challenging case of the game is solving the puzzle in a board that consists of a fixed cell. The tile located at the fixed cell cannot be moved, that is a player cannot switch the position of the hole and the tile in the fixed cell. This generalization of the puzzle is found in a mobile application, *15 Puzzle Polygon*. This application, released in 2020, contains several styles of board structures and shapes of cells. Boards with a fixed cell have been considered in this application as well. This research focuses on solvability conditions of an initial configuration of the board with a single fixed cell and the board with two fixed cells.

2 Preliminaries

Some algebra definitions and theorems are applied in this research. Let S_n be a set of bijiections from $\{1, 2, \ldots, n\}$ to $\{1, 2, \ldots, n\}$. The set S_n is a group under function composition $\circ$ with the identity ι, $\iota(i) = i$ for each $i \in \{1, 2, \ldots, n\}$, and the inverse of an element τ is denoted by τ^{-1}. S_n is called a *symmetric group* whose elements are called *permutations*. We denote $(a_1, a_2, \ldots, a_k)$ as the permutation such that $a_1 \mapsto a_2$, $a_2 \mapsto a_3, \ldots, a_{k-1} \mapsto a_k$ and $a_k \mapsto a_1$ while $a_i \mapsto a_i$ for all $i \notin \{a_1, a_2, \ldots, a_k\}$ where $a_1, a_2, \ldots, a_k$ are distinct elements in $\{1, 2, \ldots, n\}$ and $k \leq n$. Such permutation is said to be a *cycle* of length k or a *k-cycle*. In particular, a 2-cycle is said to be a *transposition*. We usually denote a product of permutations $\tau \circ \sigma$ as $\tau\sigma$. It was proved that every permutation can be written as a product of transpositions [7]. A permutation is said to be *even* (*odd*) if it can be written as a product of an even (odd) number of transpositions. Next, we provide an interesting property of the parity of the number of transpositions.

Theorem 2.1. *[11] Let ω be a permutation that can be expressed as products of transpositions $\omega = \tau_k\tau_{k-1}\ldots\tau_2\tau_1$ and $\omega = \sigma_m\sigma_{m-1}\ldots\sigma_2\sigma_1$ for some transpositions τ_i and σ_j for $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, m$, and some distinct positive integers k and m. If k is even (odd), then m is even (odd).*

According to Theorem 2.1, we also have the following result.

Theorem 2.2. *[7] For $n \geq 2$, a permutation in S_n cannot be both even and odd.*

We can conclude from Theorem 2.2 that the parity of any permutation is unique. Moreover, the next theorem gives a nice property of even permutations.

Theorem 2.3. *[11] For $n \geq 3$, an even permutation in S_n can be written as a product of 3-cycles.*

In a board, a permutation $(a_1, a_2, \ldots, a_k)$ transfers the tile in cell a_i to replace the tile in cell a_{i+1}, for $i = 1, 2, \ldots, k-1$, and transfers the tile in cell a_k to replace the tile in cell a_1.

An initial configuration which is constructed by taking even (odd) permutation from the target configuration is called an even (odd) initial configuration. Theorem 2.3 leads to the following important results for solvability of 15-puzzle.

Theorem 2.4. *[8] For 15-puzzle, every even initial configuration is solvable.*

Theorem 2.5. *[8] For 15-puzzle, every odd initial configuration is unsolvable.*

3 Main Results

As the original problem, 15-puzzle, is proved to be solvable if and only if the initial configuration is even, we consider the effect of fixing some cells in the board on the solvability of even initial

configuration. We offer conditions on the positions of the fixed cells that determine whether all even configurations are solvable or not. The first and the second sections are devoted for the results for a board with one fixed cell and two fixed cells, respectively. The last section provides the construction of the sequences of moves that play important roles in our solvability proofs.

3.1 Characteristic of Solvable Boards with 1 Fixed Cell

We identify a sufficient condition on the position of the fixed cell that allows unsolvable even configurations. For solvable configurations, we apply abstract algebra to provide sufficient conditions on the position of the fixed cell and the parity of initial configuration.

Theorem 2.5, for the original 15-puzzle, implies that a solvable initial configuration is even. We offer similar result for (n^2-1)-puzzle where $n \geq 3$.

Theorem 3.1. *For an $n \times n$ board where $n \geq 3$, if an initial configuration is solvable, it is an even initial configuration.*

Proof. We follow the proof of the (4×4)-board case. Let ω be a solvable initial configuration. Then $\omega = \tau_m \tau_{m-1} \dots \tau_2 \tau_1$ for some moves τ_i where $i = 1, 2, \dots, m$. Both in initial configuration and target configuration, the hole has to be located in the same cell. Thus, the number of moves shifting the hole to the left equals to the number of moves shifting the hole to the right, and the number of moves shifting the hole up equals to the number of moves shifting the hole down. Hence, the total number of moves has to be even. Then m is even. Therefore, ω is an even initial configuration. □

Equivalently, if an initial configuration is odd, then it is unsolvable. Due to this reason, we consider solvability conditions for even initial configurations only.

However, there are some cells in the board that cannot be fixed, otherwise there exists an unsolvable even initial configuration. We call them *forbidden cells*. The set of forbidden cells for an $n \times n$ board with a fixed cell is denoted by F_1^n. Note that by the definition, given the board with a fixed cell f, if all even initial configurations are solvable, then $f \notin F_1^n$.

Let $E_1^n = \{2, n-1, n+1, 2n, n^2-2n+1, n^2-n, n^2-n+2, n^2-1\}$. We will show that this is the set of all forbidden cells for an $n \times n$ board with a fixed cell, that is $E_1^n = F_1^n$, starting from the following proposition.

Proposition 3.2. *For an $n \times n$ board with a fixed cell, $E_1^n \subseteq F_1^n$ where $n \geq 4$.*

Proof. We claim that the cells in E_1^n are forbidden cells. Consider the case of cell 2. Assume that cell 2 is fixed. Then there is an even initial configuration where the tile located in cell 1 is not tile 1, named tile $t \neq 1$. To solve such puzzle, we have to take tile t away from cell 1 and take tile 1 to cell 1 instead. First, we have to transfer the hole to cell $n+1$ as Figure 3. Then move the hole up to swap the hole and tile t. Now, the hole is locked in cell 1. After that, we cannot make any moves in the board without the hole. Hence, we have to take the hole out of cell 1, and the only way is moving it down to cell $n+1$. That makes tile t get back to cell 1 and cannot be transferred to cell t. Thus, the configuration is unable to solve. Therefore, cell 2 is a forbidden cell. The cases of cells in $\{n-1, n+1, 2n, n^2-2n+1, n^2-n+2\}$ can be proved in a similar way due to the symmetry of the board.

Consider the case of cell n^2-n. Assume that cell n^2-n is fixed. There exists an even initial configuration where the tile located in cell n^2-1 is not cell n^2-1, named tile $t' \neq n^2-1$. To solve such puzzle, we have to take tile t' away from cell n^2-1 and take tile n^2-1 to cell n^2-1 instead. First, we move the hole to the left to swap the hole and tile t'. Now, tile t' is locked in cell n^2. Whatever moves we make, we finally have to transfer the hole back to cell n^2-1 so that we can return the hole to cell n^2. This only way makes tile t' back to cell n^2-1 and cannot

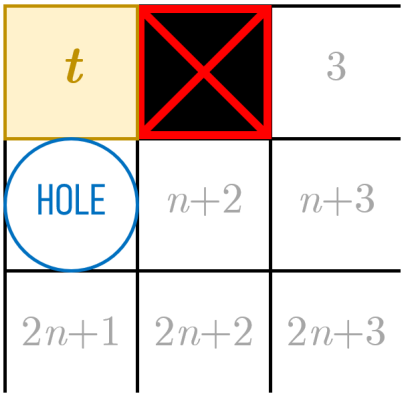


Figure 3: an $n \times n$ board with cell 2 being fixed

be transferred to cell t'. Thus, the configuration is unable to solve. Therefore, cell $n^2 - n$ is a forbidden cell. The case of cell $n^2 - 1$ can be proved in a similar way due to the symmetry of the board. Hence, $E_1^n \subseteq F_1^n$. □

Note that Theorem 2.3 is only available for the board without fixed cells. We will provide an important theorem, which is similar to Theorem 2.3, for the board with fixed cells.

Theorem 3.3. *The even permutation representing an initial configuration of a board with fixed cells can be written as a product of 3-cycles where each 3-cycle does not contain any fixed cells.*

Proof. Let ω be the permutation representing a given even initial configuration of a board B with k fixed cells. It is known that ω can be written as a product of disjoint cycles. Then $\omega = \gamma_m\gamma_{m-1}\dots\gamma_2\gamma_1$ for some disjoint cycles γ_i's of length at least 2 where $i = 1, 2, \dots, m$. Since the tiles in fixed cells cannot be transferred, γ_i in this product cannot contain fixed cells. Note that each $\gamma_i = (a_1, a_2, \dots, a_p)$ can be written as a product of $p-1$ transpositions $(a_1, a_p)(a_1, a_{p-1})\cdots(a_1, a_3)(a_1, a_2)$. Then we can transform ω to a product of transpositions without fixed cells. Since ω is the even permutation, from Theorem 2.1, it is guaranteed that the product we obtain also consists of even number of transpositions. Next, we consider transpositions in such product in pairs. We will transform each pair of transpositions to 3-cycles. For a pair of transpositions that contains one identical cell, we observe that $(a, b)(a, c) = (a, c, b)$. For a pair of transpositions that contains distinct cells, we observe that $(a, b)(c, d) = (c, b, a)(a, c, d)$. Then we obtain ω as the product of 3-cycles that does not contain any fixed cells. □

Next, we intend to prove that $F_1^n \subseteq E_1^n$ by contrapositive.

We need to deal with an even initial configuration containing no fixed cells in E_1^n. According to Theorem 3.3, we can write the permutation representing the configuration as a product of 3-cycles without any appearances of the fixed cells. Then we offer a process to transform any 3-cycle to a product of transpositions where each transposition represents a move in the game, which is switching the hole and a tile next to the hole. Hence, the product of 3-cycles denotes the sequence of moves leading the board to the standard configuration. Therefore, we can conclude that the initial configuration is solvable.

The solvability of most of the even initial configurations with such condition can be proved by applying the process that we provide in Theorem 3.4. Nevertheless, there are some positions of the fixed cell that still counteract the process. Hence, we have to modify some steps to deal with the specific case in Theorem 3.5. Therefore, to prove the hypothesis, we separate it into Theorem 3.4 and Theorem 3.5.

Theorem 3.4. *For an $n \times n$ board B with a fixed cell, if the fixed cell is not in $E_1^n \cup \{n^2 - 2n - 1, n^2 - n - 1, n^2 - 2\}$, then all even initial configurations of the board B are solvable.*

Proof. Let ω be an even initial configuration of the board B. By Theorem 3.3, ω can be written as a product of 3-cycles with no fixed cells appearing. We will prove that any 3-cycle (i, j, k) with no fixed cells where $i, j, k \notin \{n^2 - n - 1, n^2 - n\}$ can be obtained via the puzzle's moves by modifying a routine construction from the (4×4)-board case. Then we prove that any 3-cycle (i, j, k) with no fixed cells where at least one of i, j, k is in $\{n^2 - n - 1, n^2 - n\}$ can be obtained via the puzzle's moves as well.

Case 1: $i, j, k \notin \{n^2 - n - 1, n^2 - n\}$**.**

Let i be a cell in the board that is neither cell $n^2 - n - 1$ nor cell $n^2 - n$. We claim that $\left(n^2 - n - 1, n^2 - n, i\right)$ can be obtained by the puzzle's moves. For $i = n^2 - 1$, we can construct $\sigma = \left(n^2 - n - 1, n^2 - n, n^2 - 1\right)$ by

$$\sigma = \left(n^2, n^2 - 1\right)\left(n^2 - 1, n^2 - n - 1\right)\left(n^2 - n - 1, n^2 - n\right)\left(n^2 - n, n^2\right).$$

For $i = n^2$, we can construct $\alpha = \left(n^2 - n - 1, n^2 - n, n^2\right)$ by

$$\alpha = \left(n^2 - n - 1, n^2 - n\right)\left(n^2 - n, n^2\right).$$

Then we will construct a sequence of moves in the game that represents the permutation $\left(n^2 - n - 1, n^2 - n, i\right)$ where cell i is not a cell in $\{n^2 - n - 1, n^2 - n, n^2 - 1, n^2\}$. Let β be a permutation transferring a tile in cell i to cell $n^2 - 1$ without passing the fixed cell, cell $n^2 - n$, and cell n^2, and the hole is in cell $n^2 - n - 1$ at the beginning and goes back to the same cell at the end. Claim that we can construct β for all configurations of the board B, which is shown in the appendix. If we get the claim, we obtain a permutation $\alpha^{-1}\beta^{-1}\alpha\sigma\alpha^{-1}\beta\alpha$ which is a sequence of moves in the game.

By such permutation, the tiles in some cells are transferred. The tile in cell $n^2 - n - 1$ goes to cell $n^2 - n$ and back to cell $n^2 - n - 1$ by α and α^{-1} respectively. Next, it goes to cell $n^2 - n$ by σ. After that, it goes to cell n^2 and back to cell $n^2 - n$ by α and α^{-1} respectively. The tile in cell $n^2 - n$ goes to cell n^2 and back to cell $n^2 - n$ by α and α^{-1} respectively. Next, it goes to cell $n^2 - 1$ by σ and to cell i by β^{-1} at the end. The tile in cell i goes to cell $n^2 - 1$ by β. Next, it goes to cell $n^2 - n - 1$ by σ. After that, it goes to cell $n^2 - n$ and back to cell $n^2 - n - 1$ by α and α^{-1} respectively. After applying $\beta\alpha$, the tiles in all cells that are transferred by β except cell i go to other cells that are not cells in $\{n^2 - n - 1, n^2 - n, n^2 - 1, n^2\}$ and they rest there throughout the permutation $\alpha\sigma\alpha^{-1}$. Then they go back to their initial cells by β^{-1} and are not affected by α^{-1}. The hole in cell n^2 goes to cell $n^2 - n - 1$ and back to cell n^2 by α and α^{-1}, respectively. It is not relocated by σ. After that, it again goes to cell $n^2 - n - 1$ and back to cell n^2 by α and α^{-1}, respectively. The remaining tiles end up at their initial cells. That implies

$$\left(n^2 - n - 1, n^2 - n, i\right) = \alpha^{-1}\beta^{-1}\alpha\sigma\alpha^{-1}\beta\alpha.$$

Thus, we can construct $\left(n^2 - n - 1, n^2 - n, i\right)$ by a sequence of moves in the puzzle where $i \notin \{n^2 - n - 1, n^2 - n\}$. Note that

$$\left(n^2 - n - 1, j\right)\left(n^2 - n, k\right) = \left(n^2 - n - 1, n^2 - n, j\right)\left(n^2 - n - 1, n^2 - n, k\right).$$

Hence, for $i, j, k \notin \{n^2 - n - 1, n^2 - n\}$,

$$\begin{aligned}(i, j, k) &= \left(n^2 - n - 1, j\right)\left(n^2 - n, k\right)\left(n^2 - n - 1, n^2 - n, i\right)\left(n^2 - n - 1, j\right)\left(n^2 - n, k\right) \\ &= \left(n^2 - n - 1, n^2 - n, j\right)\left(n^2 - n - 1, n^2 - n, k\right)\left(n^2 - n - 1, n^2 - n, i\right) \\ &\quad \left(n^2 - n - 1, n^2 - n, j\right)\left(n^2 - n - 1, n^2 - n, k\right).\end{aligned}$$

Case 2: At least one of i, j, k **is in** $\{n^2 - n - 1, n^2 - n\}$**.**

It is enough to show that we can obtain 3-cycles which are in the forms $\left(n^2 - n, n^2 - n - 1, i\right)$, $\left(n^2 - n - 1, j, k\right)$ and $\left(n^2 - n, j, k\right)$ where $i, j, k \notin \{n^2 - n - 1, n^2 - n\}$ via the puzzle's moves as well.

Note that $\left(n^2-n, n^2-n-1, i\right)$ is the inverse of $\left(n^2-n-1, n^2-n, i\right)$. Hence,

$$\begin{aligned}\left(n^2-n, n^2-n-1, i\right) &= \left(n^2-n-1, n^2-n, i\right)^{-1} \\ &= (\alpha^{-1}\beta^{-1}\alpha\sigma\alpha^{-1}\beta\alpha)^{-1} \\ &= \alpha^{-1}\beta^{-1}\alpha\sigma^{-1}\alpha^{-1}\beta\alpha.\end{aligned}$$

Thus, the 3-cycle $\left(n^2-n, n^2-n-1, i\right)$ can be constructed by a sequence of moves in the puzzle where $i \notin \{n^2-n-1, n^2-n\}$.

Furthermore, we can construct $\left(n^2-n-1, j, k\right)$ and $\left(n^2-n, j, k\right)$ by applying 3-cycles $\left(n^2-n-1, n^2-n, i\right)$ and $\left(n^2-n, n^2-n-1, i\right)$. Then

$$\begin{aligned}\left(n^2-n-1, j, k\right) &= \left(n^2-n, n^2-n-1, k\right)\left(n^2-n-1, j\right)\left(n^2-n, k\right) \\ &= \left(n^2-n, n^2-n-1, k\right)\left(n^2-n-1, n^2-n, j\right)\left(n^2-n-1, n^2-n, k\right)\end{aligned}$$

and

$$\begin{aligned}\left(n^2-n, j, k\right) &= \left(n^2-n-1, n^2-n, k\right)\left(n^2-n, j\right)\left(n^2-n-1, k\right) \\ &= \left(n^2-n-1, n^2-n, k\right)\left(n^2-n, n^2-n-1, j\right)\left(n^2-n, n^2-n-1, k\right)\end{aligned}$$

where $j, k \notin \{n^2-n-1, n^2-n\}$.

Thus, the 3-cycles in the forms $\left(n^2-n-1, j, k\right)$ and $\left(n^2-n, j, k\right)$ can be constructed by a sequence of moves in the puzzle where $j, k \notin \{n^2-n-1, n^2-n\}$.

Therefore, any 3-cycle can be obtained via the puzzle's moves. Thus, ω can be obtained via the puzzle's moves as well. Then ω is solvable. □

It remains to show that any even initial configuration of the board with a fixed cell in $\{n^2-2n-1, n^2-n-1, n^2-2\}$ is solvable by modifying the previous process.

Theorem 3.5. *For an $n \times n$ board B with a fixed cell in $\{n^2-2n-1, n^2-n-1, n^2-2\}$, all even initial configurations of the board B are solvable.*

Proof. Let ω be an even initial configuration of the board B, and δ be a permutation which transfers the hole from cell n^2 to cell n throughout the rightmost column by repeatedly switching the hole with the tile in the above cell $n-1$ times, and then transfers the hole from cell n to cell 1 throughout the uppermost row by repeatedly switching the hole with the tile in the left cell $n-1$ times, that is

$$\delta = (1,2)(2,3)\cdots(n-1,n)(n,2n)(2n,3n)\cdots\left(n^2-2n, n^2-n\right)\left(n^2-n, n^2\right).$$

Since δ is the product of $2(n-1)$ transpositions, δ is an even permutation. We define a permutation ω' which is a configuration after transferring the hole from cell n^2 to cell 1 by δ, that is $\omega' = \delta\omega$. Since ω is an even configuration, and δ makes even puzzle's moves, ω' is an even configuration as well. Claim that ω' can be obtained via the puzzle's moves. If we get the claim, then $\omega = \delta^{-1}\omega'$ can be obtained via the puzzle's moves as well. We prove such claim by following the idea of Theorem 3.4's proof.

Note that ω' is an even configuration containing a fixed cell in $\{n^2-2n-1, n^2-n-1, n^2-2\}$ with the hole at cell 1. By Theorem 3.3, ω' can be written as a product of 3-cycles. We will prove that any 3-cycle (i,j,k) with no fixed cells where $i, j, k \notin \{n+1, n+2\}$ can be obtained via the puzzle's moves by modifying a routine construction from the previous case. Then we prove that any 3-cycle (i,j,k) with no fixed cells where at least one of i, j, k is in $\{n+1, n+2\}$ can be obtained via the puzzle's moves as well.

Case 1: $i, j, k \notin \{n+1, n+2\}$**.**

Let i be a cell in the board that is neither cell $n+1$ nor cell $n+2$. We claim that $(n+2, n+1, i)$ can be obtained by the puzzle's moves. For $i = 1$, we can construct $\alpha' = (n+2, n+1, 1)$ by $\alpha' = (n+2, n+1)\,(n+1, 1)$. For $i = 2$, we can construct $\sigma' = (n+2, n+1, 2)$ by $\sigma' = (1, 2)\,(2, n+2)\,(n+2, n+1)\,(n+1, 1)$.

Then we will construct a sequence of moves in the game that represents the permutation $(n+2, n+1, i)$ where cell i is not a cell in $\{1, 2, n+1, n+2\}$. Let β' be a permutation transferring a tile in cell i to cell 2 without passing cell 1, cell $n+1$ and the fixed cell, and the hole is in cell $n+2$ at the beginning and goes back to the same cell at the end. Claim that we can construct β' for all boards with a fixed cell in $\{n^2-2n-1, n^2-n-1, n^2-2\}$, which is shown in the appendix. If we get the claim, we obtain a permutation $(\alpha')^{-1}(\beta')^{-1}\alpha'\sigma'(\alpha')^{-1}\beta'\alpha'$ which is a sequence of moves in the game.

By such permutation, it implies that $(n+2, n+1, i) = (\alpha')^{-1}(\beta')^{-1}\alpha'\sigma'(\alpha')^{-1}\beta'\alpha'$.

Thus, we can construct $(n+2, n+1, i)$ by a sequence of moves in the puzzle where $i \notin \{n+1, n+2\}$. Note that $(n+2, j)\,(n+1, k) = (n+2, n+1, j)\,(n+2, n+1, k)$. Hence, for $i, j, k \notin \{n+1, n+2\}$,

$$\begin{aligned}(i, j, k) &= (n+2, j)\,(n+1, k)\,(n+2, n+1, i)\,(n+2, j)\,(n+1, k)\\ &= (n+2, n+1, j)\,(n+2, n+1, k)\,(n+2, n+1, i)\,(n+2, n+1, j)\,(n+2, n+1, k)\,.\end{aligned}$$

Case 2: At least one of i, j, k **is in** $\{n+1, n+2\}$**.**

It is enough to show that we can obtain 3-cycles in the forms $(n+1, n+2, i)$, $(n+1, j, k)$ and $(n+2, j, k)$ where $i, j, k \neq n+1$ and $i, j, k \neq n+2$ via the puzzle's moves as well.

Note that $(n+1, n+2, i)$ is the inverse of $(n+2, n+1, i)$. Hence,

$$\begin{aligned}(n+1, n+2, i) &= (n+2, n+1, i)^{-1}\\ &= ((\alpha')^{-1}(\beta')^{-1}\alpha'\sigma'(\alpha')^{-1}\beta'\alpha')^{-1}\\ &= (\alpha')^{-1}(\beta')^{-1}\alpha'(\sigma')^{-1}(\alpha')^{-1}\beta'\alpha'.\end{aligned}$$

Thus, the 3-cycle $(n+1, n+2, i)$ can be constructed by a sequence of moves in the puzzle where $i \notin \{n+1, n+2\}$.

Furthermore, we can construct $(n+1, j, k)$ and $(n+2, j, k)$ by applying 3-cycles $(n+2, n+1, i)$ and $(n+1, n+2, i)$. Then

$$\begin{aligned}(n+1, j, k) &= (n+2, n+1, k)\,(n+1, j)\,(n+2, k)\\ &= (n+2, n+1, k)\,(n+1, n+2, j)\,(n+1, n+2, k)\,,\end{aligned}$$

and

$$\begin{aligned}(n+2, j, k) &= (n+1, n+2, k)\,(n+2, j)\,(n+1, k)\\ &= (n+1, n+2, k)\,(n+2, n+1, j)\,(n+2, n+1, k)\end{aligned}$$

where $j, k \notin \{n+1, n+2\}$.

Thus, the 3-cycles in the forms $(n+1, j, k)$ and $(n+2, j, k)$ can be constructed by a sequence of moves in the puzzle where $j, k \notin \{n+1, n+2\}$.

Therefore, any 3-cycle with no fixed cells can be obtained via the puzzle's moves. Thus, ω' can be obtained via the puzzle's moves as well. Hence, ω can be obtained via the puzzle's moves, and is solvable. □

By Proposition 3.2, Theorem 3.4 and Theorem 3.5, we obtain the following theorem.

Theorem 3.6. *For an $n \times n$ board with a fixed cell f, all even configurations are solvable if and only if $f \notin F_1^n = \{2, n-1, n+1, 2n, n^2-2n+1, n^2-n, n^2-n+2, n^2-1\}$ where $n \geq 4$.*

For example, in case of a 4×4 board with a fixed cell, $F_1^n = \{2, 3, 5, 8, 9, 12, 14, 15\}$.

3.2 Solvability Conditions of Boards with 2 Fixed Cells

In this section, we expand similar theorems in the case of 2 fixed cells. Firstly, we provide some necessary definitions and notations.

From this section on, the positions of cells in the board are sometimes rewritten in ordered pairs. The position of the cell c in the row c_h (from the top) and the column c_v (from the left) is denoted by $\langle c_h, c_v\rangle$ where $c_h, c_v \in \{1, 2, 3, \ldots, n\}$. Note that $c_h = \lceil \frac{c}{n} \rceil$ and $c_v = c - n\lfloor \frac{c}{n} \rfloor$. Moreover, for convenience, we name some cells specifically. A cell $\mathscr{B}_m$ is defined by

$$\mathscr{B}_m = \begin{cases} \langle n-m-1, n-1\rangle & ; m = 0, 1 \\ \langle n-m, n\rangle & ; 2 \le m \le n-1 \\ \langle 1, 2n-m-1\rangle & ; n \le m \le 2n-2 \\ \langle m-2n+3, 1\rangle & ; 2n-1 \le m \le 3n-3 \\ \langle n, m-3n+4\rangle & ; 3n-2 \le m \le 4n-5 \\ \mathscr{B}_0 & ; m = 4n-4 \end{cases}$$

where $m = 0, 1, 2, \ldots, 4n-4$.

The *distance* between cells c and d in the board is defined by

$$\mathcal{D}(c, d) = \max\{|c_h - d_h|, |c_v - d_v|\}.$$

In the previous section, we defined the forbidden cells for the board with 1 fixed cell. However, for the board with 2 fixed cells, there are some sets of two cells such that their elements cannot be fixed simultaneously; otherwise, there exists an unsolvable even initial configuration. For an $n \times n$ board with 2 fixed cells, a 2-element set whose elements cannot be fixed simultaneously is called a *forbidden 2-set*. The set of all forbidden 2-sets for an $n \times n$ board with 2 fixed cells is denoted by F_2^n.

Proposition 3.7. *Let c and d be the only 2 fixed cells in an $n \times n$ board. If $\{c, d\}$ satisfies at least one of these conditions:*

(i) $c \in F_1^n$ *or* $d \in F_1^n$

(ii) $\{c, d\} \in \{\{3, 2n+1\}, \{n-2, 3n\}, \{n^2-3n+1, n^2-n+3\}\}$

(iii) $\{c, d\} = \{n^2-2n, n^2-2\}$

(iv) $\mathcal{D}(c, d) = 2$ *where* $c_h = d_h \in \{1, n\}$ *or* $c_v = d_v \in \{1, n\}$

(v) $|c_h - d_h| = |c_v - d_v| = 1$ *where* $\{c, d\} \cap \{\mathscr{B}_2, \mathscr{B}_3, \ldots, \mathscr{B}_{4n-5}\} \neq \phi$,

then $\{c, d\} \in F_2^n$ *where* $n \ge 4$.

Proof. (i) Suppose that $c \in F_1^n$ or $d \in F_1^n$. The proof of Proposition 3.2 is still available. There exists a configuration that is unable to solve. Hence, $\{c, d\}$ is a forbidden 2-set and then is in F_2^n.

(ii) Claim that the pairs of cells $\{3, 2n+1\}, \{n-2, 3n\}$ and $\{n^2-3n+1, n^2-n+3\}$ are forbidden 2-sets. Consider the case of $\{3, 2n+1\}$. Assume that cell 3 and cell $2n+1$ are fixed and the tiles located in cell 1 and cell 2 are tile 2 and tile 1, respectively. To solve such puzzle, the tiles in those cells have to be swapped. Firstly, we have to transfer the hole to cell $n+2$. From this position, there are two ways for the hole to move, going up and going to the left. Without loss of generality, we move the hole up, swapping the hole and tile 1 to move tile 1 out of cell 2. To avoid the repetition, the only way for the second move is moving to the left. Now the hole is in cell 1 and tile 2 is in cell 2. Similar to the

previous step, we move the hole down to cell $n+1$. Then, to transfer tile 1 to cell 1, we move the hole to the right, swapping with tile 1. This process changes the tiles in cell 1, cell 2 and cell $n+1$ from tile 2, tile 1 and tile $n+1$ to tile $n+1$, tile 2 and tile 1, respectively. If we repeat the process, the tiles in cell 1, cell 2 and cell $n+1$ are changed into tile 1, tile $n+1$ and tile 2, respectively. If we again repeat the process, the tiles in those cells are transferred back to their initial cells. Notice that we cannot transfer tile 1 and tile 2 to cell 1 and cell 2, respectively. Thus, the configuration is unable to solve. Therefore, $\{3, 2n+1\}$ is a forbidden 2-set. The other cases can be proved in a similar way due to the symmetry of the board. Hence, $\{3, 2n+1\}, \{n-2, 3n\}, \{n^2-3n+1, n^2-n+3\} \in F_2^n$.

(iii) Assume that cell n^2-2n and cell n^2-2 are fixed and the tiles located in cell n^2-n-1 and cell n^2-1 are tile n^2-1 and tile n^2-n-1, respectively. To solve such puzzle, the tiles in those cells have to be swapped. From the initial position of the hole, there are two ways for the hole to move, going up and going to the left. Without loss of generality, we move the hole to the left, swapping the hole and tile n^2-n-1 to move tile n^2-n-1 out of cell n^2-1. To avoid the repetition, the only way for the second move is moving up. Now the hole is in cell n^2-n-1 and tile n^2-1 is in cell n^2-1. Similar to the previous step, we move the hole to the right to cell n^2-n. Then, to transfer tile n^2-n-1 to cell n^2-n-1, we move the hole down, swapping with tile n^2-n-1. This process changes the tiles in cell n^2-n-1, cell n^2-n and cell n^2-1 from tile n^2-1, tile n^2-n and tile n^2-n-1 to tile n^2-n, tile n^2-n-1 and tile n^2-1, respectively. If we repeat the process, the tiles in cell n^2-n-1, cell n^2-n and cell n^2-1 are changed into tile n^2-n-1, tile n^2-1 and tile n^2-n, respectively. If we again repeat the process, the tiles in those cells are transferred back to their initial cells. Notice that we cannot transfer tile n^2-n-1 and tile n^2-1 to cell n^2-n-1 and cell n^2-1, respectively. Thus, the configuration is unable to solve. Therefore, $\{n^2-2n, n^2-2\}$ is a forbidden 2-set. Hence, $\{n^2-2n, n^2-2\} \in F_2^n$.

(iv) Consider the case that $c_h = d_h \in \{1, n\}$. Since $\mathcal{D}(c,d) = 2$, without loss of generality, let $c = \langle 1, c_v\rangle$ and $d = \langle 1, c_v+2\rangle = c+2$ where $c_v \notin \{2, n-3\}$. Assume that cell c and d are fixed and the tile located in cell $c+1$ is not tile $c+1$, named tile $t \neq c+1$. To solve such puzzle, we have to take tile t away from cell $c+1$ and take tile $c+1$ to cell $c+1$ instead. First, we have to transfer the hole to cell $n+c+1$. Then move the hole up to swap the hole and tile t. Now, the hole is locked in cell $c+1$. After that, we cannot make any moves in the board without the hole. Hence, we have to take the hole out of cell $c+1$, and the only way is moving it down to cell $n+c+1$. That makes tile t get back to cell $c+1$ and cannot be transferred to cell t. Thus, the configuration is unable to solve. Therefore, $\{c, d\}$ is a forbidden 2-set. The other cases can be proved similarly. Hence, $\{c, d\} \in F_2^n$.

(v) Without loss of generality, let $c \in \{\mathscr{B}_2, \mathscr{B}_3, \ldots, \mathscr{B}_{4n-5}\}$. We first consider the case when $c = \langle 1, c_v\rangle$ and $d = \langle 2, c_v+1\rangle$ where $c_v \neq n$. Moreover, $c_v \neq n-1$ since the case $c_v = n-1$ is considered in (i). Assume that cell c and d are fixed and the tile located in cell $c+1$ is not tile $c+1$, named tile $t \neq c+1$. To solve such puzzle, we have to take tile t away from cell $c+1$ and take tile $c+1$ to cell $c+1$ instead. First, we have to transfer the hole to cell $c+2$. Then move the hole to the left to swap the hole and tile t. Now, the hole is locked in cell $c+1$. After that, we cannot make any moves in the board without the hole. Hence, we have to take the hole out of cell $c+1$, and the only way is moving it to the right to cell $c+2$. That makes tile t get back to cell $c+1$ and cannot be transferred to cell t. Thus, the configuration is unable to solve. Therefore, $\{c, d\}$ is a forbidden 2-set. The other cases can be proved similarly. Hence, $\{c, d\} \in F_2^n$.

□

In an $n \times n$ board, let E_2^n be the set of pair of cells $\{c, d\}$ which satisfies at least one of these

conditions:

(i) $c \in F_1^n$ or $d \in F_1^n$

(ii) $\{c, d\} \in \{\{3, 2n+1\}, \{n-2, 3n\}, \{n^2-3n+1, n^2-n+3\}\}$

(iii) $\{c, d\} = \{n^2-2n, n^2-2\}$

(iv) $\mathcal{D}(c, d) = 2$ where $c_h = d_h \in \{1, n\}$ or $c_v = d_v \in \{1, n\}$

(v) $|c_h - d_h| = |c_v - d_v| = 1$ where $\{c, d\} \cap \{\mathscr{B}_2, \mathscr{B}_3, \ldots, \mathscr{B}_{4n-5}\} \neq \phi$

According to Proposition 3.7, $E_2^n \subseteq F_2^n$. We provide sufficient conditions for the positions of the two fixed cells that make all even initial configurations solvable in Theorem 3.8 and Theorem 3.9.

Theorem 3.8. *For an $n \times n$ board B with 2 fixed cells c and d, if $\{c, d\} \notin E_2^n$ and $c, d \notin \{n^2-2n-1, n^2-n-1, n^2-2\}$, then all even initial configurations of the board B are solvable.*

Proof. Let B be an $n \times n$ board with 2 fixed cells c and d where $\{c, d\} \notin E_2^n$ and $c, d \notin \{n^2-2n-1, n^2-n-1, n^2-2\}$, and ω be an even initial configuration of B. The proof can be demonstrated by following the proof of Theorem 3.4 with the claim that we can construct β for all configurations of B, which is shown in the appendix. Note that cell n^2-n-1 is not fixed because we cannot have fixed cells in $\{n^2-2n-1, n^2-n-1, n^2-2\}$. Cell n^2-n and cell n^2-1 cannot be fixed since both cells are in F_1^n; otherwise, a pair of cells that contains cell n^2-n or cell n^2-1 is in E_2^n. Cell n^2 is the position of the hole. Therefore, $\sigma = (n^2-n-1, n^2-n, n^2-1)$ and $\alpha = (n^2-n-1, n^2-n, n^2)$ are still available for B. Finally, it results that ω can be obtained via the puzzle's moves. Therefore, ω is solvable. □

Theorem 3.9. *For an $n \times n$ board B with 2 fixed cells c and d in which $\{c, d\} \notin E_2^n$, if $c, d \in \{n^2-2n-1, n^2-n-1, n^2-2\}$, then all even initial configurations of the board B are solvable.*

Proof. Let B be an $n \times n$ board with 2 fixed cells c and d where $\{c, d\} \notin E_2^n$ and $c, d \in \{n^2-2n-1, n^2-n-1, n^2-2\}$, and ω be an even initial configuration of B. The proof can be demonstrated by following the proof of Theorem 3.5 with the claim that we can construct β' for all configurations of B, which is shown in the appendix. Note that cells in $\{1, 2, n+1, n+2\}$ are not fixed since $c, d \in \{n^2-2n-1, n^2-n-1, n^2-2\}$. Therefore, $\sigma' = (n+2, n+1, 2)$ and $\alpha' = (n+2, n+1, 1)$ are still available for B. Finally, it results that ω can be obtained via the puzzle's moves. Therefore, ω is solvable. □

Acknowledgment. The authors are grateful to the Institute for the Promotion of Teaching Science and Technology (IPST) for giving us a scholarship under the Development and Promotion of Science and Technology Talents (DPST) Project.

References

[1] A. F. Archer, *A Modern Treatment of the 15 Puzzle*, American Math Monthly **106** (1999) 793–799.

[2] A. K. Austin, *The 14-15 puzzle*, The Mathematical Gazette **63** (1979) 45–46.

[3] J. Berenbom, J. Fendel, G. T. Gilbert and R. L. Hatcher, *Sliding piece puzzles with oriented tiles*, Discrete Mathematics **175** (1997) 23–33.

[4] A. L. Davies, *Rotating the Fifteen Puzzle*, The Mathematical Gazette **54** (1970) 237–240.

[5] J. Hamersma, *An examination of the solvability of a sliding puzzle on a hexagonal grid*, Master Thesis, Utrecht University (2018).

[6] J. T. Hollist, *The Fifteen Puzzle*, The Mathematics Teacher **72** (1979) 603–607.

[7] T. W. Hungerford, *Algebra*, Springer-Verlag New York, Inc., 1974.

[8] W. W. Johnson and W. E. Story, *Notes on the "15" puzzle*, American Journal of Mathematics **2** (1879) 397–404.

[9] H. Liebeck, *Some Generalizations of the 14-15 Puzzle*, Mathematics Magazine **44** (1971) 185–189.

[10] S. Muralidharan, *The Fifteen Puzzle - A New Approach*, Mathematics Magazine **90** (2017) 48–57.

[11] W. K. Nicholson, *Introduction to abstact algebra*, John Wiley & Sons, Inc., 2012.

[12] B. L. Schwartz, *A New Sliding Block Puzzle*, The Mathematics Teacher **66** (1973) 277–280.

[13] C. Yang, *Sliding puzzles and rotating puzzles on graphs*, Discrete Mathematics **311** (2011) 1290–1294.

4 Appendix

In this part, we construct β and β'. We show the algorithms to construct β and β' for an $n \times n$ board with 1 or 2 fixed cells. Now, we provide some more definitions and notations.

Let ϵ be the number of fixed cells in an $n \times n$ board and $\mathscr{F} = \{f^e : 1 \leq e \leq \epsilon\}$ be the set of all fixed cells in the board. The position of the fixed cell f^e is denoted by $\langle f_h^e, f_v^e \rangle$ and the position of cell i is denoted by $\langle i_h, i_v \rangle$.

The set of cells located around cell c is called the *neighborhood* of cell c, which is defined as $\mathscr{N}_v = \{x : \mathcal{D}(x, c) = 1\}$. The elements in $\mathscr{N}_v$ are called the *neighbors* of cell c. Furthermore, the set of (non-fixed) neighbors of all fixed cells in $\mathscr{F}$ is called the neighborhood of $\mathscr{F}$, defined by $\mathscr{N}_{\mathscr{F}} = \bigcup_{f^e \in \mathscr{F}} \mathscr{N}_{f^e} \backslash \mathscr{F}$.

4.1 Construction of β

The main idea refers to a sequence of cells, called the boundary route. In case of β, the *boundary route* is a sequence of cells $\mathscr{B}_0, \mathscr{B}_1, \mathscr{B}_2, \ldots, \mathscr{B}_{4n-4}$ as defined in Section 3.2. We denote $\mathscr{B} = \{\mathscr{B}_m : m = 0, 1, 2, \ldots, 4n-4\}$.

To construct β, we have to establish a closed route of hole containing cell i (position $\langle i_h, i_v \rangle$), cell $n^2 - n - 1$ (position $\langle n-1, n-1 \rangle$), and cell $n^2 - 1$ (position $\langle n, n-1 \rangle$). Through this closed route, β transfers the tile in $\langle i_h, i_v \rangle$ to $\langle n, n-1 \rangle$ by shifting the hole from $\langle n-1, n-1 \rangle$ to the next cell along this closed route consecutively until the tile in $\langle i_h, i_v \rangle$ arrives at $\langle n, n-1 \rangle$, and the hole gets back to $\langle n-1, n-1 \rangle$. Here comes an algorithm that constructs β.

Our route begins at $\langle n-1, n-1 \rangle$, we keep stepping the hole to the following cell in the boundary route as long as the following cell is not fixed. If the fixed cells are not in the boundary route, the method can be done repeatedly until we reach $\langle n-1, n-1 \rangle$ again, and then we obtain the closed route. If there are fixed cells in the boundary route, we do "AvoidFixedCell" to go through the neighbors of fixed cells instead, and then get back to the boundary route and go forward as before.

If $\langle i_h, i_v \rangle$ is in the boundary route, such boundary route is actually the closed route that contains $\langle i_h, i_v \rangle$ as desire. If $\langle i_h, i_v \rangle$ is not in the boundary route, we step along the boundary route until we are in the same row or column as $\langle i_h, i_v \rangle$. Note that there are at least 4 cells in the boundary route that is in the same row or column as $\langle i_h, i_v \rangle$. The board we considered contains at most 2 fixed cells. Hence, it is guaranteed that there are cell a and cell b in the boundary

route such that we can take path from cell a along row i_h or column i_v to reach $\langle i_h, i_v\rangle$, and then take path from $\langle i_h, i_v\rangle$ along row i_h or column i_v to cell b in the boundary route, without any appearances of fixed cells along the path. After that, we continue along the boundary route to complete the closed route.

The pseudocode representing the whole method are shown below.

Procedure: Main
Set $k := 0$, $j := 1$, $\langle x_h, x_v\rangle := \mathscr{B}_0$, $\beta_0 := \mathscr{B}_0$. Input $\mathscr{F}$.
while $k \leq 4n-5$ **do**
if $\langle i_h, i_v\rangle \notin \mathscr{B}$ and $x_h = i_h$
if $i_v = n-1$ and $\exists f^1, f^2 \in \mathscr{F}$ such that $f^1_v = i_v$, $f^1_h < i_h$ and $f^2_h = i_h$
repeat Set $\langle x_h, x_v\rangle := \mathscr{B}_{k+1}$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_h = i_h$
Set $\langle x_h, x_v\rangle := \langle i_h, i_v\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$.
repeat Set $\langle x_h, x_v\rangle := \langle x_h+1, x_v\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_h = n-3$
Set $\langle x_h, x_v\rangle := \langle x_h, x_v-1\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$.
repeat Set $\langle x_h, x_v\rangle := \langle x_h+1, x_v\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_h = n$
end if
if $\exists f^e \in \mathscr{F}$ such that $f^e_h = x_h$ and $i_v < f^e_v < x_v$
repeat Set $\langle x_h, x_v\rangle := \mathscr{B}_{k+1}$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_v = i_v$
if $\exists f^e \in \mathscr{F}$ such that $f^e_v = x_v$ and $x_h < f^e_h < i_h$
repeat Set $\langle x_h, x_v\rangle := \mathscr{B}_{k+1}$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_h = i_h$
repeat Set $\langle x_h, x_v\rangle := \langle x_h, x_v+1\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_v = i_v$
repeat Set $\langle x_h, x_v\rangle := \langle x_h+1, x_v\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_h = n$
else
repeat Set $\langle x_h, x_v\rangle := \langle x_h+1, x_v\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_h = i_h$
if $\exists f^e \in \mathscr{F}$ such that $f^e_h = x_h$ and $f^e_v < x_v$
repeat Set $\langle x_h, x_v\rangle := \langle x_h+1, x_v\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_h = n$
else
repeat Set $\langle x_h, x_v\rangle := \langle x_h, x_v-1\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_v = 1$
else
repeat Set $\langle x_h, x_v\rangle := \langle x_h, x_v-1\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_v = i_v$
if $\exists f^e \in \mathscr{F}$ such that $f^e_v = x_v$ and $f^e_h < x_h$
if $\exists f^e \in \mathscr{F}$ such that $f^e_h = x_h$ and $f^e_v < x_v$
repeat Set $\langle x_h, x_v\rangle := \langle x_h+1, x_v\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_h = n$
else
repeat Set $\langle x_h, x_v\rangle := \langle x_h, x_v-1\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_v = 1$
else
repeat Set $\langle x_h, x_v\rangle := \langle x_h-1, x_v\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_h = 1$
Let k be the index such that $\mathscr{B}_k = \langle x_h, x_v\rangle$.
end if
if $\mathscr{B}_{k+1} = \langle f^e_h, f^e_v\rangle$ for some $f^e \in \mathscr{F}$
Perform "**AvoidFixedCell**($\langle x_h, x_v\rangle$, q)".
Let k be the index such that $\mathscr{B}_k = \langle x_h, x_v\rangle$. Set $j := j+q+1$.
end if
Set $\langle x_h, x_v\rangle := \mathscr{B}_{k+1}$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$.
end while
"**Construct** β"

If the following cell in the boundary route is fixed, we need to choose another route by applying "AvoidFixedCell". At the end of "AvoidFixedCell", the hole is back in the boundary route again to continue the main method.

Procedure: AvoidFixedCell

If $\mathscr{B}_{k+1} = f^e$ for some $f^e \in \mathscr{F}$, then $\mathscr{B}_k \in \mathscr{N}_{f^e} \subseteq \mathscr{N}_{\mathscr{F}}$. Note that there is a non-fixed cell $\mathscr{B}_l \neq \mathscr{B}_k$ in $\mathscr{B} \cap \mathscr{N}_{\mathscr{F}}$. Starting from $\mathscr{B}_k$, we then step to the other cell in $\mathscr{N}_{\mathscr{F}}$ that is next to the current cell in clockwise direction around the fixed cells. We keep doing this until we reach $\mathscr{B}_l$. Assume that it takes p steps from $\mathscr{B}_0$ to $\mathscr{B}_k$ and we need q steps from $\mathscr{B}_k$ to $\mathscr{B}_l$ via this process. Let $\beta_p = \mathscr{B}_k$ and β_{p+q} be the cell that the hole is in after q steps in this process. Note that $\beta_{p+q} = \mathscr{B}_l$. At the end of the process, it outputs the final position $\langle x_h, x_v\rangle$ of the cell that the hole is in and the number of steps q.

Procedure: Construct β

According to the procedure, we obtain the sequence of cells $\beta_0, \beta_1, \beta_2, \ldots, \beta_s$ for some positive integer s. Then we construct β by letting

$$\mathcal{B} = (\beta_0, \beta_s)\,(\beta_s, \beta_{s-1}) \cdots (\beta_2, \beta_1)\,(\beta_1, \beta_0)\,.$$

After applying $\mathcal{B}$, the hole goes along the route, which contains cell i, from cell n^2-n-1 and get back to cell n^2-n-1 without passing any fixed cells, the tile located at cell β_j is moved to cell β_{j-1} for $j = 2, 3, 4, \ldots, s$ and the tile at cell β_1 is moved to cell β_s. Hence, to obtain β, we apply $\mathcal{B}$ repeatedly until tile i is in cell n^2-1 and the hole is in cell n^2-n-1. We then obtain β as desire.

4.2 Construction of β'

In case of β', the *boundary route* is a sequence of cells $\mathscr{B}'_0, \mathscr{B}'_1, \mathscr{B}'_2, \ldots, \mathscr{B}'_{4n-4}$ where $\mathscr{B}'_m$ is defined by

$$\mathscr{B}'_m = \begin{cases} \langle m+2, 2\rangle & ; m = 0, 1 \\ \langle m+1, 1\rangle & ; 2 \leq m \leq n-1 \\ \langle n, m-n+2\rangle & ; n \leq m \leq 2n-2 \\ \langle 3n-m-2, n\rangle & ; 2n-1 \leq m \leq 3n-3 \\ \langle 1, 4n-m-3\rangle & ; 3n-2 \leq m \leq 4n-5 \\ \mathscr{B}'_0 & ; m = 4n-4. \end{cases}$$

We denote $\mathscr{B}' = \{\mathscr{B}'_m : m = 0, 1, 2, \ldots, 4n-4\}$. To construct β', similar to the previous case, we have to establish a closed route of hole containing cell i (position $\langle i_h, i_v\rangle$), cell 2 (position $\langle 1, 2\rangle$), and cell $n+2$ (position $\langle 2, 2\rangle$). Through this closed route, β' transfers the tile in $\langle i_h, i_v\rangle$ to $\langle 1, 2\rangle$ by shifting the hole from $\langle 2, 2\rangle$ to the next cell along this closed route consecutively until the tile in $\langle i_h, i_v\rangle$ arrives at $\langle 1, 2\rangle$ and the hole gets back to $\langle 2, 2\rangle$. An algorithm constructing β' follows the idea of the algorithm for β by changing the initial cell from $\langle n-1, n-1\rangle$ to $\langle 2, 2\rangle$ and considering $\mathscr{B}'$ and β'_j instead of $\mathscr{B}$ and β_j, respectively.

The pseudocode representing the whole method are shown below.

Procedure: Main

Set $k := 0$, $j := 1$, $\langle x_h, x_v\rangle := \mathscr{B}'_0$, $\beta_0 := \mathscr{B}'_0$. Input $\mathscr{F}$.
while $k \leq 4n-5$ **do**
  if $\langle i_h, i_v\rangle \notin \mathscr{B}'$ and $x_h = i_h$
    if $i_v = 2$ and $\exists f^1, f^2 \in \mathscr{F}$ such that $f^1_v = i_v$, $f^1_h > i_h$ and $f^2_h = i_h$
      repeat Set $\langle x_h, x_v\rangle := \mathscr{B}'_{k+1}$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_h = i_h$
      Set $\langle x_h, x_v\rangle := \langle i_h, i_v\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$.
      repeat Set $\langle x_h, x_v\rangle := \langle x_h - 1, x_v\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_h = 4$
      Set $\langle x_h, x_v\rangle := \langle x_h, x_v + 1\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$.
      repeat Set $\langle x_h, x_v\rangle := \langle x_h - 1, x_v\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_h = 1$

end if
if $\exists f^e \in \mathscr{F}$ such that $f_h^e = x_h$ and $x_v < f_v^e < i_v$
repeat Set $\langle x_h, x_v\rangle := \mathscr{B}'_{k+1}$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_v = i_v$
if $\exists f^e \in \mathscr{F}$ such that $f_v^e = x_v$ and $i_h < f_h^e < x_h$
repeat Set $\langle x_h, x_v\rangle := \mathscr{B}'_{k+1}$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_h = i_h$
repeat Set $\langle x_h, x_v\rangle := \langle x_h, x_v - 1\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_v = i_v$
repeat Set $\langle x_h, x_v\rangle := \langle x_h - 1, x_v\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_h = 1$
else
repeat Set $\langle x_h, x_v\rangle := \langle x_h - 1, x_v\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_h = i_h$
if $\exists f^e \in \mathscr{F}$ such that $f_h^e = x_h$ and $f_v^e > x_v$
repeat Set $\langle x_h, x_v\rangle := \langle x_h - 1, x_v\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_h = 1$
else
repeat Set $\langle x_h, x_v\rangle := \langle x_h, x_v + 1\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_v = n$
else
repeat Set $\langle x_h, x_v\rangle := \langle x_h, x_v + 1\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_v = i_v$
if $\exists f^e \in \mathscr{F}$ such that $f_v^e = x_v$ and $f_h^e > x_h$
if $\exists f^e \in \mathscr{F}$ such that $f_h^e = x_h$ and $f_v^e > x_v$
repeat Set $\langle x_h, x_v\rangle := \langle x_h - 1, x_v\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_h = 1$
else
repeat Set $\langle x_h, x_v\rangle := \langle x_h, x_v + 1\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_v = n$
else
repeat Set $\langle x_h, x_v\rangle := \langle x_h + 1, x_v\rangle$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$. **until** $x_h = n$
Let k be the index such that $\mathscr{B}'_k = \langle x_h, x_v\rangle$.
end if
if $\mathscr{B}'_{k+1} = \langle f_h^e, f_v^e\rangle$ for some $f^e \in \mathscr{F}$
Perform "**AvoidFixedCell**($\langle x_h, x_v\rangle$, q')".
Let k be the index such that $\mathscr{B}'_k = \langle x_h, x_v\rangle$. Set $j := j + q' + 1$.
end if
Set $\langle x_h, x_v\rangle := \mathscr{B}'_{k+1}$, $\beta_j := \langle x_h, x_v\rangle$, $j := j+1$.
end while
"**Construct** β'"

If the following cell in the boundary route is fixed, we need to choose another route by applying "AvoidFixedCell". At the end of "AvoidFixedCell", the hole is back in the boundary route again to continue the main method.

Procedure: AvoidFixedCell

If $\mathscr{B}'_{k+1} = \langle f_h^e, f_v^e\rangle$ for some $f^e \in \mathscr{F}$, then $\mathscr{B}'_k \in \mathscr{N}_{f^e} \subseteq \mathscr{N}_{\mathscr{F}}$. Note that there is a non-fixed cell $\mathscr{B}'_l \neq \mathscr{B}'_k$ in $\mathscr{B}' \cap \mathscr{N}_{\mathscr{F}}$. Starting from $\mathscr{B}'_k$, we then step to the other cell in $\mathscr{N}_{\mathscr{F}}$ that is next to the current cell in clockwise direction around the fixed cells. We keep doing this until we reach $\mathscr{B}'_l$. Assume that it takes p' steps from $\mathscr{B}'_0$ to $\mathscr{B}'_k$ and we need q' steps from $\mathscr{B}'_k$ to $\mathscr{B}'_l$ via this process. Let $\beta'_{p'} = \mathscr{B}'_k$ and $\beta'_{p'+q'}$ be the cell that the hole is in after q' steps in this process. Note that $\beta'_{p'+q'} = \mathscr{B}'_l$. At the end of the process, it outputs the final position $\langle x_h, x_v\rangle$ of the cell that the hole is in and the number of steps q'.

Procedure: Construct β'

According to the procedure, we obtain the sequence of cells $\beta'_0, \beta'_1, \beta'_2, \ldots, \beta'_{s'}$ for some positive integer s'. Then we construct β' by letting

$$\mathcal{B}' = \left(\beta'_0, \beta'_{s'}\right)\left(\beta'_{s'}, \beta'_{s'-1}\right)\cdots\left(\beta'_2, \beta'_1\right)\left(\beta'_1, \beta'_0\right).$$

After applying $\mathcal{B}'$, the hole goes along the route, which contains cell i, from cell $n+2$ and get back to cell $n+2$ without passing any fixed cells, the tile located at cell β'_j is moved to cell β'_{j-1}

for $j = 2, 3, 4, \ldots, s'$ and the tile at cell β'_1 is moved to cell $\beta'_{s'}$. Hence, to obtain β', we apply $\mathcal{B}'$ repeatedly until tile i is in cell 2 and the hole is in cell $n + 2$. We then obtain β' as desire.

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Girths and Diameters of a Graph, its δ-Complement, and its δ'-Complement*

Supakorn Srisawat[1,†] and Panupong Vichitkunakorn[1,‡]

[1]Division of Computational Science, Faculty of Science
Prince of Songkla University, Hat Yai, Songkhla 90110, Thailand

Abstract

The δ-complement G_δ and the δ'-complement $G_{\delta'}$ of a graph G, introduced in 2022 by Pai et al., are two variants of the graph complement. Two vertices are adjacent in G_δ if and only if they are of the same degree but not adjacent in G or they are of different degrees but adjacent in G. On the other hand, two vertices are adjacent in G'_δ if and only if they are not adjacent in G_δ, i.e., G'_δ is the complement of G_δ. We provide the Nordhaus-Gaddum-type bounds, applied from Nordhaus and Gaddum (1956), over the girths and the diameters of a graph and its δ-complement. We also provide the Nordhaus-Gaddum-type bounds over the girths and the diameters of a graph and its δ'-complement.

Keywords: Nordhaus-Gaddum-type relations, δ-complement graph, δ'-complement graph, girth.

2020 MSC: 05C35, 05C38, 05C69, 05C76.

1 Introduction

In 1956, Nordhaus and Gaddum [8] showed the following relations between the chromatic numbers of a graph G and $\overline{G}$.

$$2\sqrt{n} \leq \chi(G) + \chi(\overline{G}) \leq n + 1,$$

and

$$n \leq \chi(G) \cdot \chi(\overline{G}) \leq \left(\frac{n+1}{2}\right)^2.$$

Some times after, researchers studied the relations between the same invariant of a graph and its complement in a similar manner to Nordhaus and Gaddum. This kind of relation was then called *Nordhaus-Gaddum-type relations*. The parameters they were concerning are, for instance, minimum degrees [2], maximum degrees [12], diameters [11], girths [11], circumferences [11], and domination numbers [7]. Modern mathematicians even collected those results in a survey [3].

*S. Srisawat was supported by a Graduate Fellowship (Research Assistant),
Faculty of Science, Prince of Songkla University, Contract no. 1-2565-02-028.
†Speaker. ‡Corresponding author.
Email: supakorn.swt@gmail.com (S. Srisawat), panupong.v@psu.ac.th (P. Vichitkunakorn).

In 2022, Pai, et al. [9] defined the δ-complement and the δ'-complement of a graph. They are defined similarly to the usual graph complement but regarding the degree of the vertices further than just the edges. In 2023, Vichitkunakorn, et al. [10] then studied the Nordhaus-Gaddum-type relations between G and G_δ on the chromatic numbers by applying the original theorem from [8].

In this paper, we provide the Nordhaus-Gaddum-type relations over the girths, radii, and diameters of a graph and its δ-complement. Then, we give a similar result on the δ'-complement. The paper is organized as follows. In Section 2, we review the Nordhaus-Gaddum-type relations on the chromatic numbers, girths, radii, an diameters of a graph and its complement. Then we recall the definition of the δ-complement and the δ'-complement of a graph. After that, we revisit the Nordhaus-Gaddum-type relation over the chromatic numbers of a graph and its δ-complement. In Section 3, we show the Nordhaus-Gaddum-type relation on the girths and the diameters of a graph and its δ-complement. Moreover, we show the same relation over the girths and the diameters of a graph and its δ'-complement. The sharpness of the bounds is then discussed. Finally, the conclusion and some discussions on this study are in Section 4.

2 Preliminaries

The complement of a simple graph $G = (V, E)$, denoted by $\overline{G} = (V, \overline{E})$, is a graph such that $uv \in \overline{E}$ if and only if $uv \notin E$. The chromatic number $\chi(G)$ of a graph G is the least amount of colours required to label each vertex in G so that no two adjacent vertices share the same colour.

In 1956, Nordhaus and Gaddum were concerned with finding the relations between the chromatic number $\chi(G)$ of a graph G and the chromatic number $\chi(\overline{G})$ of the complement $\overline{G}$. They found the upper bound and lower bound of the sum and the product of $\chi(G)$ and $\chi(\overline{G})$ which they have provided in [8].

Theorem 2.1 ([8]). *Let G be a graph of n vertices. Then,*

$$2\sqrt{n} \leq \chi(G) + \chi(\overline{G}) \leq n + 1,$$

and

$$n \leq \chi(G) \cdot \chi(\overline{G}) \leq \left(\frac{n+1}{2}\right)^2.$$

Moreover, the bounds are sharp for all n.

The girth $\text{girth}(G)$ of a graph G is the length of the smallest cycle contained in a graph. If G does not contain any cycles, then $\text{girth}(G) = \infty$.

In 1990, Xu [11] provided the upper bound and the lower bound of $\text{girth}(G) + \text{girth}(\overline{G})$ and $\text{girth}(G) \cdot \text{girth}(\overline{G})$ as follows.

Theorem 2.2 ([11]). *For $n \geq 6$, let G be a graph of n vertices such that G and its complement $\overline{G}$ are containing cycles. Then,*

$$6 \leq \text{girth}(G) + \text{girth}(\overline{G}) \leq n + 3,$$

and

$$9 \leq \text{girth}(G) \cdot \text{girth}(\overline{G}) \leq 3n.$$

Moreover, the bounds are sharp for all $n \geq 6$.

The distance between any two vertices $u, v \in V(G)$, denoted by $d_G(u, v)$, is the length of the shortest path between u and v in G. If there are no such paths between them, then $d_G(u, v) = \infty$.

The diameter of G, denoted by $\text{diam}(G)$, is the maximum distance between any two vertices in G. If G is not connected, then $\text{diam}(G) = \infty$.

For the diameter of a graph and its complement, the following theorem was proven separately by [1], [4], [5], and [11].

Theorem 2.3 ([1], [4], [5], [11])**.** *For $n \geq 6$, let G be a graph of n vertices such that G and its complement $\overline{G}$ are connected. Then,*

$$4 \leq \text{diam}(G) + \text{diam}(\overline{G}) \leq n + 1,$$

and

$$4 \leq \text{diam}(G) \cdot \text{diam}(\overline{G}) \leq 2n - 2.$$

Moreover, the bounds are sharp for all $n \geq 6$.

In 2022, Pai et al. [9] have defined the δ-complement and the δ'-complement of a graph as follows.

Definition 2.4. Consider a graph $G = (V, E)$. A graph $G_\delta = (V, E_\delta)$ such that for any $u, v \in V$, $uv \in E_\delta$ if and only if $\deg(u) = \deg(v)$ and $uv \notin E$, or $\deg(u) \neq \deg(v)$ and $uv \in E$. Then G_δ is called **a δ-complement of** G.

Definition 2.5. Consider a graph $G = (V, E)$. A graph $G_{\delta'} = (V, E_{\delta'})$ such that for any $u, v \in V$, $uv \in E_{\delta'}$ if and only if $\deg(u) = \deg(v)$ and $uv \in E$, or $\deg(u) \neq \deg(v)$ and $uv \notin E$. Then $G_{\delta'}$ is called **a δ'-complement of** G.

We notice from the definitions that $G_{\delta'} = \overline{G_\delta}$.

After that, in the year after, Vichitkunakorn et al. [10] considered a δ-complement variant of the Nordhaus-Gaddum-type relation as follows.

Theorem 2.6 ([10])**.** *For $n \geq 4$, let G be a graph of n vertices. Let $d_1, d_2, \ldots, d_m$ be the distinct degrees of vertices in G, and V_{d_i} be the set of vertices of degree d_i. Then*

$$2 \cdot \sqrt{\max_{1 \leq i \leq m} \{|V_{d_i}|\}} \leq \chi(G) + \chi(G_\delta) \leq m + n,$$

and

$$\max_{1 \leq i \leq m} \{|V_{d_i}|\} \leq \chi(G) \cdot \chi(G_\delta) \leq \left(\frac{m+n}{2}\right)^2.$$

To the best of the authors' knowledge, other works on Nordhaus-Gaddum-type relations over other invariants of a graph and its δ-complement (or its δ'-complement) are yet to be found.

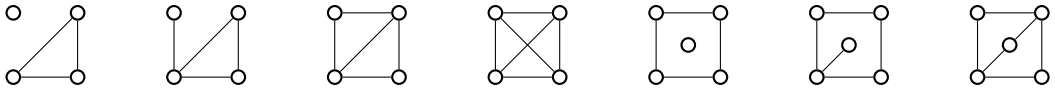
3 Main Results

Throughout this section, we denote K_n the complete graph of order n, C_n the cycle of order n, and P_n the path of order n. For two graphs G and H, we denote $G + H$ and $G \vee H$ the disjoint union and the join of G and H, respectively. If H is a subgraph of G, the graph $G - H$ is the resulting graph after deleting all edges of H from G.

3.1 Girths

We recall that the girth of a graph G, denoted by $\text{girth}(G)$, is the length of the smallest cycle contained in G. If that graph does not contain cycles, i.e., it is acyclic, then $\text{girth}(G) = \infty$.

Before we get to the theorem, there is a necessary fact to give the result.



(I) (II) (III) (IV) (V) (VI) (VII)

Figure 1: All non-isomorphic graphs of order $n = 4$ or 5 with girth $n - 1$

Lemma 3.1. *For $n \geq 6$, let G be a graph of n vertices. Then* $\text{girth}(G) = n - 1$ *if and only if G is a cycle of order $n - 1$ with a pendant vertex adjacent to at most one vertex in the cycle.*

Proof. Assume $n \geq 6$, and $\text{girth}(G) = n - 1$. Then the smallest cycle contained in G is C_{n-1}. No two non-consecutive vertices in the cycle can be adjacent.

Let v be the vertex outside the cycle. Suppose v is adjacent to at least two vertices in the cycle. Since the length of the cycle is at least 5, this creates a smaller cycle, which is a contradiction. So v is adjacent to at most one vertex in the cycle. So G is $C_{n-1} + K_1$ or a cycle with a pendant vertex attached.

The converse obviously holds. □

Lemma 3.2. *For $n = 4$ or 5, a graph G of n vertices has* $\text{girth}(G) = n - 1$ *if it is one of the seven graphs in Figure. 1.*

Proof. Assume $n = 4$. For a graph with 4 vertices whose girth is $n - 1 = 3$, it implies that there exists a 3-cycle contained in the graph. The vertex outside the cycle can be adjacent to one, two, all, or none of the three vertices of the cycle. So we obtain the four graphs from Figure. 1, namely, (I), (II), (III), and (IV). Also, it is obvious that the girth of each graph is three.

Now assume $n = 5$. A graph with 5 vertices whose girth is $n - 1 = 4$ implies that there exists a 4-cycle contained in the graph. It is clear that there are no other edges between two vertices of the cycle; otherwise, there will be a 3-cycle. The vertex outside the cycle cannot be adjacent to two adjacent vertices in the cycle; otherwise, it creates a 3-cycle. So v is adjacent to at most two vertices in the cycle. Hence, we obtain the result. □

Theorem 3.3. *For $n \geq 4$, let G be a graph of n vertices such that G and G_δ contain cycles. Then*

$$6 \leq \text{girth}(G) + \text{girth}(G_\delta) \leq \begin{cases} 2n - 2 & \text{if } n = 4 \text{ or } n \geq 6, \\ 2n & \text{if } n = 5, \end{cases}$$

and

$$9 \leq \text{girth}(G) \cdot \text{girth}(G_\delta) \leq \begin{cases} (n-1)^2 & \text{if } n = 4 \text{ or } n \geq 6, \\ n^2 & \text{if } n = 5. \end{cases}$$

In addition, the lower bounds are sharp for all $n \geq 4$.

Proof. The lower bounds are obvious since both G and G_δ must contain cycles. An example of the sharp bounds for all $n \geq 4$ is when $G = K_1 \vee P_{n-1}$. When $n = 4$, we can easily check that $G_\delta \cong G$. So, $\text{girth}(G) = \text{girth}(G_\delta) = 3$. When $n > 4$, the vertex joining the path is of degree $n - 1$ in G. An endpoint of the path is of degree 2 in G, while its unique neighbour in the path is of degree 3 in G. These three vertices are of different degrees and form a 3-cycle in G. We get $\text{girth}(G) = 3$. In addition, This 3-cycle remains in G_δ. Therefore, $\text{girth}(G_\delta) = 3$.

We now prove the upper bounds. For $n = 4$, Lemma 3.2 implies that G must be either C_4 or one of the graphs (I)–(IV) in Figure 1. It is easy to verify that the upper bounds hold. For $n = 5$, the upper bounds are obvious. For $n \geq 6$, it suffices to show that $\text{girth}(G) + \text{girth}(G_\delta) \leq 2n - 2$.

Suppose to the contrary that $\text{girth}(G) + \text{girth}(G_\delta)$ is greater than $2n - 2$. Then there are three possibilities:

Case 1: $\text{girth}(G) = n$ and $\text{girth}(G_\delta) = n$.

Case 2: $\text{girth}(G) = n$ and $\text{girth}(G_\delta) = n - 1$.

Case 3: $\text{girth}(G) = n - 1$ and $\text{girth}(G_\delta) = n$.

If $\text{girth}(G) = n$, then $G = C_n$. Since cycle graphs are 2-regular, $G_\delta = \overline{G}$. Consider the Ramsey number $R(3,3) = 6$. Since $n \geq 6$ but G does not contain a triangle, we have G_δ contains triangles. Hence, $\text{girth}(G_\delta) = 3$, but $3 < n - 1$ for all $n \geq 6$. This is a contradiction. So the first two cases cannot happen.

If $\text{girth}(G) = n - 1$ and $\text{girth}(G_\delta) = n$, then by Lemma 3.1, G is either $C_{n-1} + K_1$ or a cycle of order $n - 1$ with one pendant vertex, and $G_\delta = C_n$. Both cases are impossible for any n. Hence, the third case cannot happen as well.

Therefore, $\text{girth}(G) + \text{girth}(G_\delta) \leq 2n - 2$.

The multiplicative upper bound follows from the additive upper bound using the AM-GM inequality as follows.

$$\begin{aligned}\text{girth}(G) \cdot \text{girth}(G_\delta) &\leq \left(\frac{\text{girth}(G) + \text{girth}(G_\delta)}{2}\right)^2 \\ &\leq \left(\frac{2n-2}{2}\right)^2 \\ &= (n-1)^2.\end{aligned}$$

□

Remark 3.4. It is still unknown when both upper bounds of Theorem 3.3 are sharp. For $4 \leq n \leq 10$, there are only 3 graphs that achieve the bounds. They are $K_1 \vee P_3$ (the graph (III) in Figure 1), C_5, and $C_5 + K_1$, for $n = 4, 5$, and 6, respectively.

For large n, we expect that either G or G_δ will be likely to contain C_3. Hence, either $\text{girth}(G)$ or $\text{girth}(G_\delta)$ is 3. So we conjecture, by using the same manner as Theorem 2.2, that C_n is the graph with maximum values of $\text{girth}(G) + \text{girth}(G_\delta)$ and $\text{girth}(G) \cdot \text{girth}(G_\delta)$ for infinitely many values of n.

Conjecture 3.5. *There are infinitely many values of n such that*

$$\text{girth}(G) + \text{girth}(G_\delta) \leq n + 3, \quad \textit{and} \quad \text{girth}(G) \cdot \text{girth}(G_\delta) \leq 3n$$

for each graph G of n vertices such that both G and G_δ contain cycles.

In the case of the sum and the product between $\text{girth}(G)$ and $\text{girth}(G_{\delta'})$, the obvious bounds are sharp for all $n \geq 3$.

Theorem 3.6. *For $n \geq 3$, let G be a graph of n vertices such that G and $G_{\delta'}$ contain cycles. Then*

$$6 \leq \text{girth}(G) + \text{girth}(G_{\delta'}) \leq 2n,$$

and

$$9 \leq \text{girth}(G) \cdot \text{girth}(G_{\delta'}) \leq n^2.$$

Moreover, all bounds are sharp for all $n \geq 3$.

Proof. All bounds are obvious. An example of the sharp lower bounds is when $G = K_n$ since $G_{\delta'} = K_n$ and $\text{girth}(K_n) = 3$ for any $n \geq 3$. The upper bounds are sharp if and only if $\text{girth}(G) = n$ and $\text{girth}(G_{\delta'}) = n$. This only happens when $G = C_n$ for $n \geq 3$. □

3.2 Diameters

We recall that the diameter of a graph G is the maximum distance between any two vertices in G. If G is not connected, then $\text{diam}(G) = \infty$.

Before we get to the Nordhaus-Gaddum relation between the diameter of G and G_δ, there are some facts to give the result.

Lemma 3.7. *Let $G = P_n$ be a path graph of n vertices. Then*

$$\text{diam}(G_\delta) = \begin{cases} 0 & \text{if } n = 1, \\ \infty & \text{if } n = 2 \text{ or } n = 5, \\ 1 & \text{if } n = 3, \\ 3 & \text{if } n = 4 \text{ or } n \geq 6. \end{cases}$$

Proof. Assume $G = P_n$. If $n = 1$, then $G_\delta = K_1$. So, the diameter is zero. If $n = 2$, then $G_\delta = 2K_1$. Since G_δ is not connected, we have $\text{diam}(G_\delta) = \infty$. If $n = 3$, then $G_\delta = K_3$. So, $\text{diam}(G_\delta) = 1$. If $n = 4$, then $G_\delta = P_4$, so $\text{diam}(G_\delta) = 3$. If $n = 5$, then $G_\delta = C_4 + K_1$ which is not connected, so $\text{diam}(G_\delta) = \infty$.

Consider $n \geq 6$. Let $v_1, v_2, \ldots, v_n$ be the vertices in the path where v_i and v_{i+1} are adjacent for $1 \leq i \leq n-1$. The edges of G_δ are v_1v_n, v_1v_2, $v_{n-1}v_n$, and v_iv_j where $2 \leq i < j \leq n-1$ and $j \neq i+1$.

Since $N_{G_\delta}(v_1) = \{v_2, v_n\}$, $N_{G_\delta}(v_n) = \{v_1, v_{n-1}\}$, $N_{G_\delta}(v_2) = V(G) \setminus \{v_3, v_n\}$, and $N_{G_\delta}(v_{n-1})$ $= V(G) \setminus \{v_1, v_{n-2}\}$, the distance between v_1 and all other vertices in G_δ and the distance between v_n and all other vertices in G_δ are at most 3.

Since $N_{G_\delta}(v_2) = V(G) \setminus \{v_3, v_n\}$, $N_{G_\delta}(v_3) = V(G) \setminus \{v_1, v_2, v_4, v_n\}$, $N_{G_\delta}(v_{n-2}) = V(G) \setminus \{v_1, v_{n-3}, v_{n-1}, v_n\}$, and $N_{G_\delta}(v_{n-1}) = V(G) \setminus \{v_1, v_{n-2}\}$, the distance between v_2 and all other vertices in G_δ and the distance between v_{n-1} and all other vertices in G_δ are at most 2.

We also notice that $d_{G_\delta}(v_i, v_{i+1}) = 3$ for any $3 \leq i \leq n-3$ as v_i and $v_i + 1$ have no common neighbours. Hence, the distance between any two vertices in G_δ is at most 3. Therefore, if $G = P_n$ such that $n \geq 6$, $\text{diam}(G_\delta) = 3$. □

Lemma 3.8. *Let G be a graph of $n \geq 3$ vertices. Then* $\text{diam}(G) = n - 2$ *if and only if G can be formed by P_{n-1} and a vertex v satisfies one of the following conditions:*

1. *v is adjacent to one vertex which is not an endpoint of the path.*
2. *v is adjacent to two vertices in the path which are either adjacent or have a distance of two between them.*
3. *v is adjacent to three consecutive vertices in the path.*

Proof. Assume $\text{diam}(G) = n-2$. Then G must contain a path of length $n-2$ and a pair of vertices of distance $n-2$. Let $v_1, v_2, \ldots, v_{n-1}$ be the vertices in the path where $d_G(v_1, v_{n-1}) = n-2$. We see that two nonconsecutive vertices in the path cannot be adjacent. Let v be the vertex outside the path. The vertex v cannot be adjacent to two vertices v_i and v_j such that $|j-i| > 2$; otherwise $d_G(v_1, v_{n-1}) < n-2$. This gives the conditions 2 and 3. Furthermore, if v is adjacent to only one vertex which is the endpoint v_1 or v_{n-1}, then $G = P_n$ and $\text{diam}(G) = n-1$, which is a contradiction. This gives the condition 1.

The converse is easy to verify as $d_G(v_1, v_{n-1}) = n-2$. □

Theorem 3.9. *For $n \geq 6$, let G be a graph of n vertices such that G and G_δ are connected. Then*

$$3 \leq \text{diam}(G) + \text{diam}(G_\delta) \leq 2n - 4,$$

and

$$2 \leq \operatorname{diam}(G) \cdot \operatorname{diam}(G_\delta) \leq (n-2)^2.$$

Moreover, the lower bounds are sharp for all $n \geq 6$.

Proof. Suppose $\operatorname{diam}(G) + \operatorname{diam}(G_\delta) < 3$ or $\operatorname{diam}(G) \cdot \operatorname{diam}(G_\delta) < 2$. Then $\operatorname{diam}(G) = 1$ and $\operatorname{diam}(G_\delta) = 1$. Then $G = G_\delta = K_n$, which is a contradiction. An example of the sharp bounds is when $G = K_{1,n-1}$ where $\operatorname{diam}(G) = 2$, and $G_\delta = K_n$ where $\operatorname{diam}(G_\delta) = 1$.

For the additive upper bound, we suppose to the contrary that $\operatorname{diam}(G)+\operatorname{diam}(G_\delta) > 2n-4$.

Clearly, the diameter of any graph is at most $n-1$. So there are three possibilities:

Case 1. $\operatorname{diam}(G) = n-1$ and $\operatorname{diam}(G_\delta) = n-1$.

Case 2. $\operatorname{diam}(G) = n-1$ and $\operatorname{diam}(G_\delta) = n-2$.

Case 3. $\operatorname{diam}(G) = n-2$ and $\operatorname{diam}(G_\delta) = n-1$.

If $\operatorname{diam}(G) = n-1$, then $G = P_n$. By Lemma 3.7, we have $\operatorname{diam}(G_\delta) = 3$ when $n \geq 6$. But $n-1 \neq 3$ and $n-2 \neq 3$ for all $n \geq 6$, the first two possibilities are not satisfied. For the third possibility, from $\operatorname{diam}(G) = n-2$, Lemma 3.8 gives three cases for G. In all cases, we can show that $G_\delta \neq P_n$. Hence, $\operatorname{diam}(G_\delta) \neq n-1$. Therefore, $\operatorname{diam}(G) + \operatorname{diam}(G_\delta) \leq 2(n-2)$.

The multiplicative upper bound follows from the additive upper bound using the AM-GM inequality as follows.

$$\begin{aligned}
\operatorname{diam}(G) \cdot \operatorname{diam}(G_\delta) &\leq \left(\frac{\operatorname{diam}(G) + \operatorname{diam}(G_\delta)}{2}\right)^2 \\
&\leq \left(\frac{2n-4}{2}\right)^2 \\
&= (n-2)^2.
\end{aligned}$$

□

Remark 3.10. The lower bounds from Theorem 3.9 are also sharp for $n = 3, 4$, and 5, with the same example of $G = K_{1,n-1}$.

In the case of the sum and the product between $\operatorname{diam}(G)$ and $\operatorname{diam}(G_{\delta'})$, we are going to use the following lemma.

Lemma 3.11. *Let* $G = P_n$ *be a path graph of* n *vertices. Then*

$$\operatorname{diam}(G_{\delta'}) = \begin{cases} 0 & \text{if } n = 1, \\ 1 & \text{if } n = 2, \\ \infty & \text{if } n = 3, \\ 2 & \text{if } n = 5, \\ 3 & \text{if } n = 4 \text{ or } n \geq 6. \end{cases}$$

Proof. Assume $G = P_n$. If $n = 1$, then $G_{\delta'} = K_1$, so the diameter is zero. If $n = 2$, then $G_{\delta'} = K_2$, so $\operatorname{diam}(G_{\delta'}) = 1$. If $n = 3$, then $G_{\delta'} = 3K_1$. Since $G_{\delta'}$ is not connected, we have $\operatorname{diam}(G_{\delta'}) = \infty$. If $n = 4$, then $G_{\delta'} = P_4$, so $\operatorname{diam}(G_{\delta'}) = 3$. If $n = 5$, $G_{\delta'}$ is a bow-shaped graph as shown in Figure 2 , so $\operatorname{diam}(G_{\delta'}) = 2$.

Consider $n \geq 6$. Let $v_1, v_2, \ldots, v_{n-1}, v_n$ be the vertices in the path where v_i and v_{i+1} are adjacent for $1 \leq i \leq n-1$. The edges of $G_{\delta'}$ are v_1v_i such that $i \neq 2$ and $i \neq n$, v_nv_j such that $j \neq 1$ and $j \neq n-1$, and $v_{i'}v_{j'}$ such that $2 \leq i' \leq n-2$ and $j' = i' + 1$.

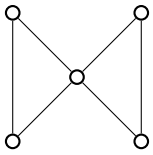


Figure 2: The bow-shaped graph whose diameter is 2

Since $N_{G_{\delta'}}(v_1) = V(G) \setminus \{v_2, v_n\}$, $N_{G_{\delta'}}(v_n) = V(G) \setminus \{v_1, v_{n-1}\}$, $N_{G_{\delta'}}(v_2) = \{v_3, v_n\}$, and $N_{G_{\delta'}}(v_{n-1}) = \{v_1, v_{n-2}\}$, the distance between v_1 and all other vertices in $G_{\delta'}$ and the distance between v_n and all other vertices in $G_{\delta'}$ are at most 2.

Since $N_{G_{\delta'}}(v_2) = \{v_3, v_n\}$, $N_{G_{\delta'}}(v_3) = \{v_1, v_2, v_4, v_n\}$, $N_{G_{\delta'}}(v_{n-2}) = \{v_1, v_{n-3}, v_{n-1}, v_n\}$, and $N_{G_{\delta'}}(v_{n-1}) = \{v_1, v_{n-2}\}$, the distance between v_2 and all other vertices in $G_{\delta'}$ and the distance between v_{n-1} and all other vertices in $G_{\delta'}$ are at most 3.

We also notice that $d_{G'_\delta}(v_i, v_{i+1}) = 1$ for any $3 \leq i \leq n-3$ as v_i and $v_i + 1$ are adjacent. Hence, the distance between any two vertices in G'_δ is at most 3. Therefore, if $G = P_n$ such that $n \geq 6$, $\operatorname{diam}(G_{\delta'}) = 3$. □

Theorem 3.12. *For $n \geq 2$, let G be a graph of n vertices such that G and $G_{\delta'}$ are connected. Then*

$$2 \leq \operatorname{diam}(G) + \operatorname{diam}(G_{\delta'}),$$

and

$$1 \leq \operatorname{diam}(G) \cdot \operatorname{diam}(G_{\delta'}).$$

Moreover, both bounds are sharp for all $n \geq 2$.

Proof. Both bounds are obvious. They are sharp if and only if $\operatorname{diam}(G) = 1$ and $\operatorname{diam}(G_{\delta'}) = 1$. This is when $G = G_{\delta'} = K_n$. □

For the upper bounds, we conjecture in the same manner as Theorem 2.3 that P_n is the graph with the maximum values of $\operatorname{diam}(G) \cdot \operatorname{diam}(G_{\delta'})$ for infinitely many values of n. Using the fact from Lemma 3.11, we get the following conjecture.

Conjecture 3.13. *There are infinitely many values of n such that*

$$\operatorname{diam}(G) + \operatorname{diam}(G_{\delta'}) \leq n + 2, \quad \text{and} \quad \operatorname{diam}(G) \cdot \operatorname{diam}(G_{\delta'}) \leq 3(n-1)$$

for each graph G of n vertices such that both G and $G_{\delta'}$ are connected.

4 Conclusion and Discussion

We provide the Nordhaus-Gaddum-type relations on the girths and the diameters of a graph and its δ-complement. The lower bounds for the girths are found sharp for all $n \geq 4$, and for the diameters, they are found sharp for all $n \geq 6$. The sharpness of the upper bounds for both invariants is still unknown. However, we conjecture that such upper bounds are not sharp for all n. On the girths, it is also conjectured that there are upper bounds, for infinitely many values of n.

We also give the Nordhaus-Gaddum-type relations on the girths and the diameters of a graph and its δ'-complement. All lower bounds are found obvious and sharp for all n. Both upper bounds for girths are found obvious and sharp for all n while the upper bounds for diameters are yet to be found sharp. However, we conjecture that such bounds are not sharp for all n.

For further research, results on the Nordhaus-Gaddum-type relation on other graph invariants will be interesting. In addition to the Nordhaus-Gaddum-type relation that considers

one invariant, the relations between two (or more) different invariants of a graph and its δ-complement (or δ'-complement) are also interesting to study. See [6] and the references therein for more examples.

Acknowledgements

Supakorn Srisawat was supported by Graduate Fellowship (Research Assistant), Faculty of Science, Prince of Songkla University, Contract no. 1-2565-02-028.

References

[1] N. Achuthan, N. R. Achuthan, and L. Caccetta, *On the Nordhaus-Gaddum class problems*, Australasian Journal of Combinatorics **2** (1990), 5–27.

[2] Y. Alavi and J. Mitchem, *The connectivity and line connectivity of complementary graphs*, Lecture Notes in Mathematics (1971), 1–3.

[3] M. Aouchiche and P. Hansen, *A survey of Nordhaus–Gaddum type relations*, Discrete Applied Mathematics **161** (2013), 466–546.

[4] J. A. Bondy, *A note on the diameter of a graph*, Canadian Mathematical Bulletin **11** (1968), 499–501.

[5] J. Bosák, A. Rosa, and S. Znám, *On decompositions of complete graphs into factors with given diameters*, Theory of Graphs (Proc. Colloq., Tihany, 1966) (1968), 37–56.

[6] R. Brigham and R. Dutton, *A compilation of relations between graph invariants*, Networks **15** (200610), 73 –107.

[7] F. Jaeger and C. Payan, *Relations du type Nordhaus–Gaddum pour le nombre d'absorption d'un graphe simple*, C. R. Acad. Sci. Paris Sér. A **274** (1972), 728–730.

[8] E. A. Nordhaus and J. W. Gaddum, *On complementary graphs*, The American Mathematical Monthly **63** (1956), no. 3, 175.

[9] A. Pai, H. A. Rao, S. D'Souza, P. G. Bhat, and S. Upadhyay, *δ-complement of a graph*, Mathematics **10** (2022), no. 8, 1203.

[10] P. Vichitkunakorn, R. Maungchang, and W. Tangjai, *On Nordhaus-Gaddum type relations of δ-complement graphs*, Heliyon **9** (2023), no. 6, e16630.

[11] S. J. Xu, *Some parameters of graph and its complement*, Discrete Mathematics **65** (1987), no. 2, 197 –207.

[12] S. J. Xu, *Relations between parameters of a graph*, Discrete Mathematics **89** (1991), 65–88.

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Local Antimagic Chromatic Number of the Cartesian Product of Graphs

Teeradej Kittipassorn[1] and Kiattiyot Phibul[1,†,‡]

[1]Department of Mathematics and Computer Science, Faculty of Science
Chulalongkorn University, Bangkok 10330, Thailand

Abstract

A *local antimagic labeling* of a graph $G = (V, E)$ is a bijection from the set of edges to the set of integers $\{1, 2, 3, \ldots, |E|\}$ such that $w(u) \neq w(v)$ for any adjacent vertices u and $v \in V(G)$ where the weight $w(u) = \sum_{e \in E(u)} f(e)$ and $E(u)$ is the set of edges incident to u. The *local antimagic chromatic number* $\chi_{la}(G)$ is the minimum number of colors taken over all colorings of G induced by local antimagic labelings of G. In this article, we determine some bounds for the local antimagic chromatic numbers of the grid graphs $P_2 \times P_n$, the prism graphs $P_2 \times C_n$ and the toroidal grid graphs $C_m \times C_n$.

Keywords: local antimagic labeling, local antimagic chromatic number, cartesian product, paths, cycles.

2020 MSC: 05C15, 05C78, 05C76, 05C38.

1 Introduction

In 1990, Hartsfield and Ringel studied a variant of the magic labeling, called an *antimagic labeling* [7]. A graph with q edges is called *antimagic* if it has an edge labeling with $1, 2, 3, \ldots, q$ without repetition such that the sums of the labels of all the edges incident with a vertex are distinct. They conjectured that all connected graphs except K_2 are antimagic. Recently, Bensmail, Senhaji and Lyngsie proved that this conjecture is true [4].

The concept of local antimagic graphs was presented in 2017 [2]. A local antimagic labeling of a graph $G = (V, E)$ is a bijection from the set of edges to the integers $\{1, 2, 3, \ldots, |E|\}$ such that $w(u) \neq w(v)$ for any adjacent vertices u and $v \in V(G)$ where the weight $w(u) = \sum_{e \in E(u)} f(e)$ and $E(u)$ is the set of edges incident to u. A graph G is called *local antimagic* if G has an local antimagic labeling. Clearly, if G is antimagic, then G is local antimagic. An *induced color* of u under f is the vertex label $w(u)$. The number of distinct induced colors under f is denoted by $c(f)$, and is called the *color number* of f. The *local antimagic chromatic number* of G, denoted

*This research was financially supported by the Development and Promotion of Science and Technology Talents Project (DPST).

†Speaker. ‡Corresponding author.
Email: teeradej.k@chula.ac.th (T. Kittipassorn), skeattiyos@gmail.com (K. Phibul).

by $\chi_{la}(G)$, is $\min\{c(f) : f \text{ is a local antimagic labeling of } G\}$. Obviously, the local antimagic chromatic number for any graph is at least the chromatic number for its, that is, $\chi_{la}(G) \geq \chi(G)$.

In the intervening years, the local antimagic chromatic numbers for several graphs have been determined, such as paths [2], cycles [2], friendship [2], complete bipartite [2, 11], wheel [2, 11], kite [2], and copies of graph [3]. In addition, the local chromatic numbers of some products of graphs, including join product [10, 12, 21], lexicographic product [15] and corona product [1, 9] were investigated by various researchers. Consequently, an open problem in their papers asks for the local antimagic chromatic number of the cartesian product of graphs.

Let G and H be two graphs. The *cartesian product* of graphs G and H, written $G \times H$, is the graph with vertex set $V(G) \times V(H)$ specified by putting (u, v) adjacent to (u', v') if and only if (1) $u = u'$ and $vv' \in E(H)$, or (2) $v = v'$ and $uu' \in E(G)$. A *path* P_n is a graph with n vertices $u_1, u_2, u_3, \ldots, u_n$ and $n-1$ edges $u_1u_2, u_2u_3, u_3u_4, \ldots, u_{n-1}u_n$. A *cycle* C_n is a graph with n vertices $u_1, u_2, u_3, \ldots, u_n$ and n edges $u_1u_2, u_2u_3, u_3u_4, \ldots, u_{n-1}u_n, u_nu_1$.

In 2004, Wang [18] showed that the *toroidal grids* which are the cartesian products of two cycles are antimagic. A year later, Cheng [5] proved that *grid graphs* and *prism graphs* which are the cartesian products of two paths and of a cycle and a path, respectively are antimagic. It follows that these graphs are local antimagic graphs. Furthermore, Lau and Shiu [14] were the first to study and determine the local antimagic chromatic numbers of $P_2 \times C_3$ and $P_2 \times C_4$.

Theorem 1.1. *$\chi_{la}(C_3 \times P_2) = 3$ and $\chi_{la}(C_4 \times P_2) = 4$.*

Apart from this, the local antimagic chromatic number of no other cartesian product of graphs is known. Motivated by this, in this article, our main results determine some bounds for the local antimagic chromatic numbers of the grid graphs $P_2 \times P_n$, the prism graphs $P_2 \times C_n$ and the toroidal grids graph $C_m \times C_n$ in the following theorems.

Theorem 1.2. *(i) For $n \geq 3$, $3 \leq \chi_{la}(P_2 \times P_n) \leq 6$.*

(ii) $\chi_{la}(P_2 \times P_2) = 3$ and $\chi_{la}(P_2 \times P_3) = 4$.

Theorem 1.3. *For $n \geq 5$,*

$$3 \leq \chi_{la}(P_2 \times C_n) \leq \begin{cases} 5 & \text{if } n \text{ is odd,} \\ 6 & \text{if } n \text{ is even.} \end{cases}$$

Theorem 1.4. *(i) For even $m, n \geq 3$, $\chi_{la}(C_m \times C_n) \leq 5$ unless $m = n = 4$.*

(ii) For odd $m \geq 3$, even $n \geq 3, m < n$, $\chi_{la}(C_m \times C_n) \leq n + 2$.

(iii) For even $m \geq 3$, odd $n \geq 3, m < n$, $\chi_{la}(C_m \times C_n) \leq m + 4$.

(iv) For odd $m, n \geq 3$, $m \neq n$, $\chi_{la}(C_m \times C_n) \leq m + n + 1$.

(v) For $m, n \geq 3$, $\chi_{la}(C_m \times C_n) \geq 3$.

The rest of this paper is organized as follows. In Section 2 is divided into three subsections to study three cartesian product graphs including the grid graphs $P_2 \times P_n$, the prism graphs $P_2 \times C_n$ and the toroidal grids $C_m \times C_n$. In Subsection 2.1, we prove Theorem 1.2. The proof of Theorem 1.3 is given in Subsection 2.2. Subsection 2.3 is devoted to the proof of Theorem 1.4. Finally, we conclude the paper in Section 3 with some open problems.

2 Proofs of Theorems

The following lemma of Lau, Shiu and Ng [12] which gives a sufficient condition for a bipartite graph G to have $\chi_{la}(G) \geq 3$ is needed to prove the lower bounds for the local antimagic chromatic numbers of our graphs.

Lemma 2.1. *Let G be a graph of size q. Suppose there is a local antimagic labeling of G inducing a 2-coloring of G with colors x and y, where $x < y$. Let X and Y be the sets of vertices of colored x and y, respectively. Then G is a bipartite graph with bipartition (X, Y) and $|X| > |Y|$, and*

$$x|X| = y|Y| = \frac{q(q+1)}{2}. \tag{2.1}$$

Proof. Since the vertices of X(or Y) are not adjacent by the definition of the local antimagic labeling, G is bipartite with bipartition (X, Y). Thus, the sum of the labels of all edges is $x|X| = \frac{q(q+1)}{2} = y|Y|$. Since $x < y$, we obtain that $|X| > |Y|$. □

Corollary 2.2. *Suppose G is a connected bipartite graph of q edges with bipartition (V_1, V_2). If $\chi_{la}(G) = 2$, then $|V_1| \neq |V_2|$ and $\binom{q+1}{2}$ is divisible by both $|V_1|$ and $|V_2|$.*

Proof. Suppose $\chi_{la}(G) = 2$. By Lemma 2.1, we obtain that $|X| \neq |Y|$ and $\frac{q(q+1)}{2}$ is divisible by both $|X|$ and $|Y|$. Since bipartition of G is unique by connectedness, $\{X, Y\} = \{V_1, V_2\}$. Hence, $|V_1| \neq |V_2|$ and $\frac{q(q+1)}{2}$ is divisible by both $|V_1|$ and $|V_2|$. □

2.1 Local Antimagic Chromatic Number of $P_2 \times P_n$

Let $n \geq 2$. Since $P_2 \times P_n$ is a bipartite graph, the vertices can be divided into two disjoint sets $\{u_i : 1 \leq i \leq n\}$ and $\{v_i : 1 \leq i \leq n\}$ with edge set $\{u_i v_{i+1} : 1 \leq i < n\} \cup \{v_i u_{i+1} : 1 \leq i < n\} \cup \{u_i v_i : 1 \leq i \leq n\}$. Clearly, $|E(P_2 \times P_n)| = 3n - 2$. See the graph $P_2 \times P_n$ for odd and even n in Figures 1 and 2, respectively.

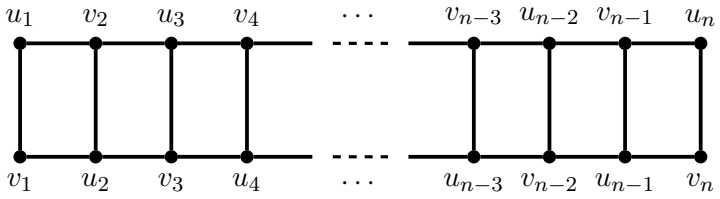


Figure 1: The graph $P_2 \times P_n$ for odd n

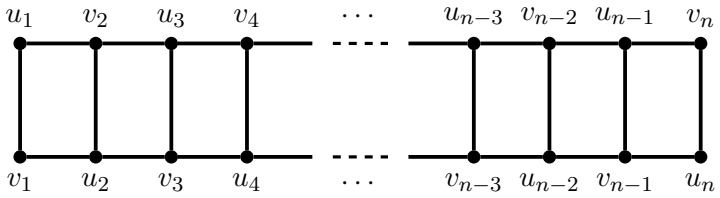


Figure 2: The graph $P_2 \times P_n$ for even n

Proof of Theorem 1.2. (i) For the lower bound, since $P_2 \times P_n$ is a bipartite graph where partite sets are of the same size, we have $\chi_{la}(P_2 \times P_n) > 2$ by Corollary 2.2. For the upper bound, it suffices to define a local antimagic labeling $f : E(P_2 \times P_n) \to \{1, 2, 3, \ldots, 3n-2\}$ that induces 6 distinct vertex colors. Let us separate into two cases by the parity of n.

Case 1. n is odd.

We will show an algorithm of this labeling in the following five steps.

Step 1 We divide the set $\{1, 2, 3, \ldots, 3n-2\}$ into six subsets as follows:

$\{2, 4, 6, \ldots, n-1\}$, $\{1, 3, 5, \ldots, n-2\}$, $\left\{n, n+1, , n+2 \ldots, \frac{3n-1}{2}\right\}$, $\left\{\frac{3n+1}{2}\right\}$, $\left\{\frac{3n+1}{2}+1, \frac{3n+1}{2}+2, \frac{3n+1}{2}+3, \ldots, 2n-1\right\}$ and $\{2n, 2n+1, 2n+2, \ldots, 3n-2\}$.

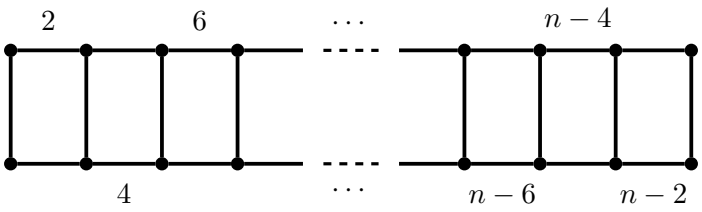


Figure 3: The edge labeling in Step 2 of $P_2 \times P_n$ for odd n

Step 2 We label the edge u_iv_{i+1} for all $1 \leq i < n$ with the numbers $2, 4, 6, \ldots, n-1, 1, 3, 5, \ldots, n-2$ in that order from left to right (see Figure 3).

Step 3 We label the edge v_iu_{i+1} for all $1 \leq i \leq \frac{n-1}{2}$ with the numbers $n, n+1, n+2, \ldots, \frac{3n-1}{2}$ in that order in backward direction from $i = \frac{n-1}{2}$ to $i = 1$ (see Figures 4 and 5).

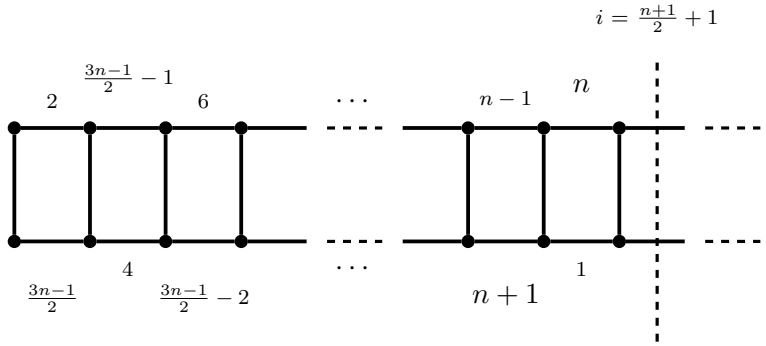


Figure 4: The edge labeling in Step 3 of $P_2 \times P_n$ for $n \equiv 3 \mod 4$

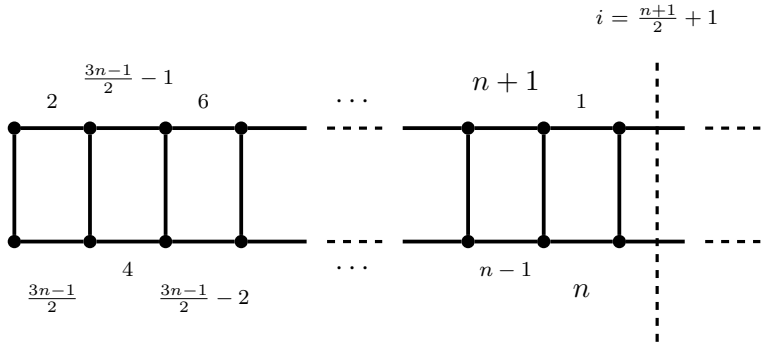


Figure 5: The edge labeling in Step 3 of $P_2 \times P_n$ for $n \equiv 1 \mod 4$

Step 4 We label the edge v_iu_{i+1} for all $\frac{n-1}{2} < i < n$ with the numbers $\frac{3n+1}{2} + 1, \frac{3n+1}{2} + 2, \frac{3n+1}{2} + 3, \ldots, 2n-1$ in that order in backward direction from $i = n-1$ to $i = \frac{n+1}{2}$ (see Figures 6 and 7).

Step 5 We label the edge u_iv_i for all $1 \leq i \leq n$ with the numbers $2n, 2n+1, 2n+2, \ldots, 3n-3, 3n-2, \frac{3n+1}{2}$ in that order from right to left (see Figure 8).

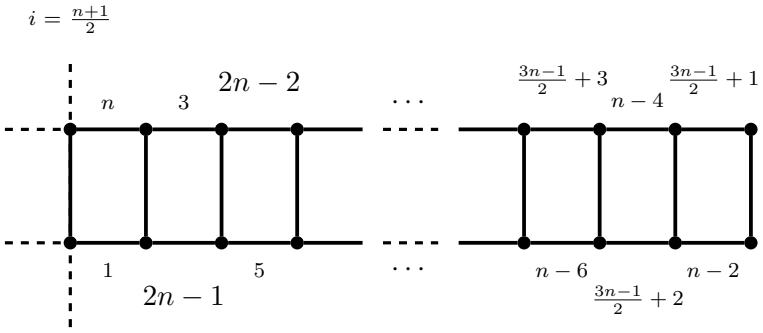


Figure 6: The edge labeling of $P_2 \times P_n$ for $n \equiv 3 \mod 4$

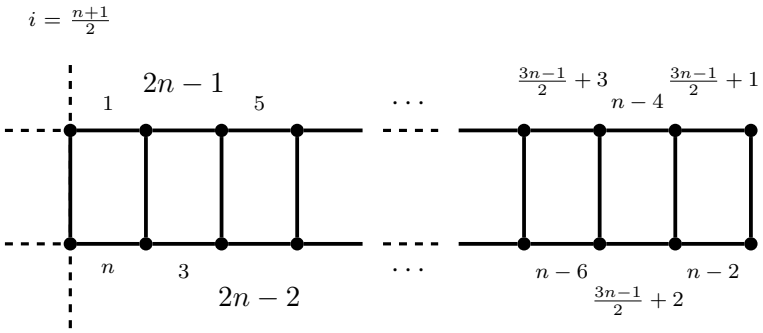


Figure 7: The edge labeling of $P_2 \times P_n$ for $n \equiv 1 \mod 4$

We can observe that this labeling can be summarized as follows:

$$f(u_iv_{i+1}) = \begin{cases} 2i & ;1 \le i \le \lfloor \frac{n}{2} \rfloor \\ 2i-n & ;\lfloor \frac{n}{2} \rfloor < i < n \end{cases}$$

$$f(v_iu_{i+1}) = \begin{cases} 2n - \lfloor \frac{n}{2} \rfloor - i & ;1 \le i \le \lceil \frac{n}{2} \rceil \\ 3n - \lfloor \frac{n}{2} \rfloor - i & ;\lceil \frac{n}{2} \rceil < i < n \end{cases}$$

$$f(u_1v_1) = \frac{3n+1}{2}$$

$$f(u_iv_i) = 3n - i \text{ for all } 2 \le i \le n.$$

Thus, we obtain that $w(u_1) = f(u_1v_1)+f(u_1v_2) = \frac{1}{2}(3n+5)$, $w(u_n) = f(u_nv_n)+f(v_{n-1}u_n) = \frac{1}{2}(7n+3)$ and $w(u_i) = f(v_{i-1}u_i) + f(u_iv_i) + f(u_iv_{i+1}) = \frac{1}{2}(9n+3)$ for $1 < i < \lceil \frac{n}{2} \rceil$ and $\lceil \frac{n}{2} \rceil + 1 < i < n$.

In the case $i = \lceil \frac{n}{2} \rceil, \lceil \frac{n}{2} \rceil + 1$, $w(u_i) = f(v_{i-1}u_i) + f(u_iv_i) + f(u_iv_{i+1}) = \frac{1}{2}(7n+3)$. Likewise, $w(v_1) = f(u_1v_1)+f(v_1u_2) = 3n$, $w(v_n) = f(u_nv_n)+f(u_{n-1}v_n) = 3n-2$ and $w(v_i) = f(u_{i-1}v_i)+f(u_iv_i) + f(v_iu_{i+1}) = \frac{1}{2}(9n-3)$ for $1 < i < n$.

Therefore, f is a local antimagic labeling that induces 6 distinct vertex colors, including

(1) $w(u_1) = \frac{1}{2}(3n+5)$,

(2) $w(u_i) = \frac{1}{2}(9n+3)$ when $i \in \{2,3,\ldots,n-1\} \setminus \{\lceil \frac{n}{2} \rceil, \lceil \frac{n}{2} \rceil + 1\}$,

(3) $w\left(u_{\lceil \frac{n}{2} \rceil}\right) = w\left(u_{\lceil \frac{n}{2} \rceil + 1}\right) = w(u_n) = \frac{1}{2}(7n+3)$,

(4) $w(v_1) = 3n$,

(5) $w(v_i) = \frac{1}{2}(9n-3)$ for $1 < i < n$,

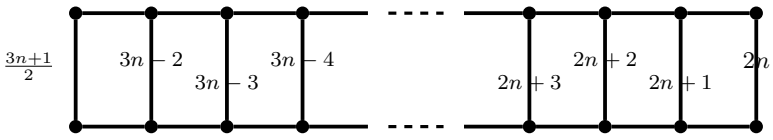


Figure 8: The edge labeling in Step 5 of $P_2 \times P_n$ for odd n

(6) $w(v_n) = 3n - 2$.

Case 2. n is even.

Similarly, we can show an algorithm of this labeling in the following five steps.

Step 1 We divide the set $\{1, 2, 3, \ldots, 3n-2\}$ into five subsets as follows:
$\{2, 4, 6, \ldots, n-2\}$, $\{1, 3, 5, \ldots, n-1\}$, $\left\{n, n+1, n+2, \ldots, \frac{3n-1}{2}\right\}$, $\left\{\frac{3n+1}{2}, \frac{3n+1}{2}+1, \frac{3n+1}{2}+2, \ldots, 2n-2\right\}$ and $\{2n-1, 2n, 2n+1, \ldots, 3n-2\}$.

Step 2 We label the edge u_iv_{i+1} for all $1 \leq i < n$ with the numbers $2, 4, 6, \ldots, n-2$, $1, 3, 5, \ldots, n-1$ in that order from left to right.

Step 3 We label the edge v_iu_{i+1} for all $1 \leq i \leq \frac{n}{2}$ with the numbers $n, n+1, n+2, \ldots, \frac{3n-1}{2}$ in that order in backward direction from $i = \frac{n}{2}$ to $i = 1$.

Step 4 We label the edge v_iu_{i+1} for all $\frac{n}{2} < i < n$ with the numbers $\frac{3n+1}{2}+1, \frac{3n+1}{2}+2, , \frac{3n+1}{2}+3, \ldots, 2n-1$ in that order in backward direction from $i = n-1$ to $i = \frac{n}{2}+1$.

Step 5 We label the edge u_iv_i for all $1 \leq i \leq n$ with the numbers $2n-1, 2n, 2n+1, , \ldots, 3n-3, 3n-2$ in that order from right to left.

We can observe that this labeling can be summarized as follows:

$$f(u_iv_{i+1}) = \begin{cases} 2i & ; 1 \leq i \leq \left\lfloor \frac{n}{2} \right\rfloor \\ 2i - n + 1 & ; \left\lfloor \frac{n}{2} \right\rfloor < i < n \end{cases}$$

$$f(v_iu_{i+1}) = \begin{cases} 2n - \left\lfloor \frac{n}{2} \right\rfloor - i & ; 1 \leq i \leq \left\lceil \frac{n}{2} \right\rceil \\ 3n - \left\lfloor \frac{n}{2} \right\rfloor - i - 1 & ; \left\lceil \frac{n}{2} \right\rceil < i < n \end{cases}$$

$$f(u_iv_i) = 3n - i - 1 \text{ for all } 1 \leq i \leq n.$$

Thus, we obtain that $w(u_1) = f(u_1v_1) + f(u_1v_2) = 3n$, $w(u_n) = f(u_nv_n) + f(v_{n-1}u_n) = \frac{1}{2}(7n-2)$ and $w(u_i) = f(v_{i-1}u_i) + f(u_iv_i) + f(u_iv_{i+1}) = \frac{9n}{2}$ for $1 < i < \left\lceil \frac{n}{2} \right\rceil$, $\left\lceil \frac{n}{2} \right\rceil + 1 < i < n$.

In case $i = \left\lceil \frac{n}{2} \right\rceil, \left\lceil \frac{n}{2} \right\rceil + 1$, $w(u_i) = f(v_{i-1}u_i) + f(u_iv_i) + f(u_iv_{i+1}) = \frac{1}{2}(7n+2)$. Likewise, $w(v_i) = f(u_{i-1}v_i) + f(u_iv_i) + f(v_iu_{i+1}) = \frac{1}{2}(9n-6)$ for $1 \leq i < n$ and $w(v_n) = f(u_nv_n) + f(u_{n-1}v_n) = 3n-2$.

Therefore, f is a local antimagic labeling that induces 6 distinct vertex colors, including

(1) $w(u_1) = 3n$,

(2) $w(u_i) = \frac{9n}{2}$ when $i \in \{2, 3, \ldots, n-1\} \setminus \left\{\left\lceil \frac{n}{2} \right\rceil, \left\lceil \frac{n}{2} \right\rceil + 1\right\}$,

(3) $w\left(u_{\left\lceil \frac{n}{2} \right\rceil}\right) = w\left(u_{\left\lceil \frac{n}{2} \right\rceil + 1}\right) = \frac{1}{2}(7n+2)$,

(4) $w(u_n) = \frac{1}{2}(7n-2)$,

(5) $w(v_i) = \frac{1}{2}(9n-6)$ for $1 \leq i < n$,

(6) $w(v_n) = 3n-2$.

Hence, $\chi_{la}(P_2 \times P_n) \leq 6$.

(ii) Since $P_2 \times P_2$ is the cycle C_4, we have $\chi_{la}(P_2 \times P_2) = \chi_{la}(C_4) = 3$. In Figure 9, we obtain a local antimagic labeling of $P_2 \times P_3$ with $c(f) = 4$. Thus, $\chi_{la}(P_2 \times P_3) \leq 4$. By the

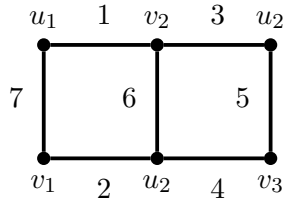


Figure 9: An edge labeling of $P_2 \times P_3$ with induced vertex colors in $\{8, 9, 10, 12\}$

upper bound in Theorem 1.2 *(i)*, it remains to show that $\chi_{la}(P_2 \times P_3) \neq 3$. We suppose for contradiction that there is a local antimagic labeling $f : E(P_2 \times P_3) \to \{1, 2, 3, \ldots, 7\}$ such that $c(f) = 3$. Then we consider a possible vertex coloring of $P_2 \times P_3$ with three colors x, y, z which can be divided into two cases.

Case 1. There are at least three vertices with the same color.

Obviously, there is exactly three vertices with the same color such that $w(u_1) = w(u_2) = w(u_3)$ or $w(v_1) = w(v_2) = w(v_3)$. Without loss of generality, we let $w(u_1) = w(u_2) = w(u_3) = x$. Observe that $3x = w(u_1) + w(u_2) + w(u_3) = \sum_{e \in E(P_2 \times P_3)} f(e) = \sum_{k=1}^{7} k = 28$. It implies that $x = \frac{28}{3} \notin \mathbb{Z}$ which contradicts with $x \in \mathbb{Z}$.

Case 2. Each colors appears on at most two vertices.

Obvious that each colors label two vertices. Without loss of generality, let $w(u_2) = x$. Then the other vertex with color x is u_1 or u_3. Without loss of generality, suppose $w(u_1) = x$ and $w(u_3) = y$. Then it implies that $w(v_2) = w(v_3) = z$ and $w(v_1) = y$. Consider $2x + 2y + 2z = \sum_{u \in V(P_2 \times P_3)} w(u) = 2\sum_{e \in E(P_2 \times P_3)} f(e)$. Thus, $x + y + z = \sum_{e \in E(P_2 \times P_3)} f(e)$. On the other hand, we have $x + x + y = w(u_1) + w(u_2) + w(u_3) = \sum_{e \in E(P_2 \times P_3)} f(e)$. Thus, we get that $x + y + z = x + x + y$ implying $x = z$ which is a contradiction. Hence, we conclude that $\chi_{la}(P_2 \times P_3) \geq 4$. □

2.2 Local Antimagic Chromatic Number of $P_2 \times C_n$

Proof of Theorem 1.3. For the lower bound, if n is even, we observe that $P_2 \times C_n$ is the bipartite graph and both partite sets are the same size. Thus, $\chi_{la}(P_2 \times P_n) > 2$ by Corollary 2.2. If n is odd, we consider $\chi_{la}(P_2 \times C_n) \geq \chi(P_2 \times C_n) = \chi(C_n) = 3$ since $\chi(G \times H) = \max\{\chi(G), \chi(H)\}$ for any graph G and H.

For the upper bound, let $P_2 \times C_n$ be the graph for any $n \geq 5$. Let $V(P_2 \times C_n) = \{u_i : 1 \leq i \leq n\} \cup \{v_i : 1 \leq i \leq n\}$ be the vertex set of $P_2 \times C_n$ and $E(P_2 \times C_n) = \{u_i v_{i+1} : 1 \leq i \leq n\} \cup \{v_i u_{i+1} : 1 \leq i \leq n\}$. Note that all additions within the index is performed in modulo n. See the graph $P_2 \times C_n$ for odd and even n in Figures 10 and 11, respectively. Clearly, $|E(P_2 \times C_n)| = 3n$.

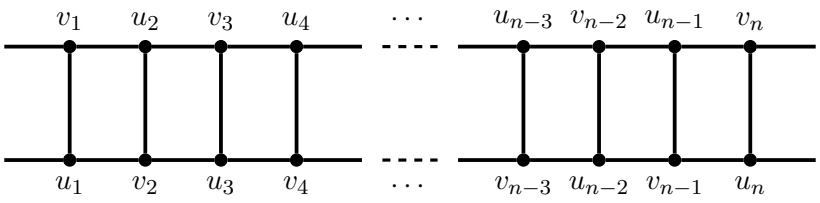


Figure 10: The graph $P_2 \times C_n$ for odd n

It suffices to define a local antimagic labeling $f : E(P_2 \times C_n) \to \{1, 2, 3, \ldots, 3n\}$ that induces 5 distinct vertex colors if n is odd and induces 6 distinct vertex colors if n is even. Let us separate into two cases by the parity of n.

Case 1. n is odd.

We will show an algorithm of this labeling in the following five steps.

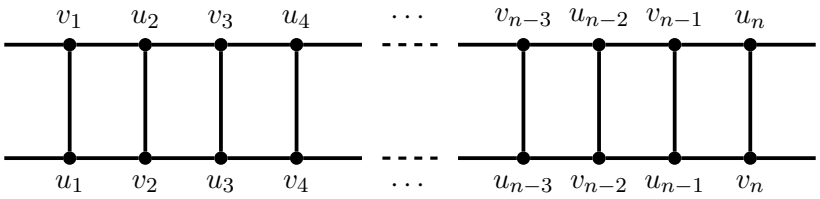


Figure 11: The graph $P_2 \times C_n$ for even n

Step 1 We divide the set $\{1, 2, 3, \ldots, 3n\}$ into five subsets as follows:

$\{2, 4, 6, \ldots, n-1\}$, $\{1, 3, 5, \ldots, n\}$, $\left\{n+1, n+2, n+3, \ldots, \frac{3n+1}{2}\right\}$, $\left\{\frac{3n+1}{2}+1, \frac{3n+1}{2}+2, \frac{3n+1}{2}+3, \ldots, 2n\right\}$ and $\{2n+1, 2n+2, 2n+3, \ldots, 3n\}$.

Step 2 We label the edge $v_n v_1$ with the number 2 and label the edge $u_i v_{i+1}$ for all $1 \leq i < n$ with the numbers $4, 6, \ldots, n-1, 1, 3, 5, \ldots, n$ in that order from left to right (see Figure 12).

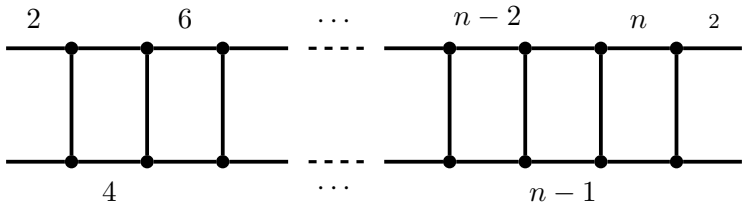


Figure 12: The edge labeling in Step 2 of $P_2 \times C_n$ for odd n

Step 3 We label the edge $u_n u_1$ with the number $\frac{3n+1}{2}$ and label the edge $v_i u_{i+1}$ for all $1 \leq i \leq \frac{n-1}{2}$ with the numbers $n+1, n+2, n+3, \ldots, \frac{3n+1}{2}-1$ in that order in backward direction from $i = \frac{n-1}{2}$ to $i = 1$ (see Figures 13 and 14).

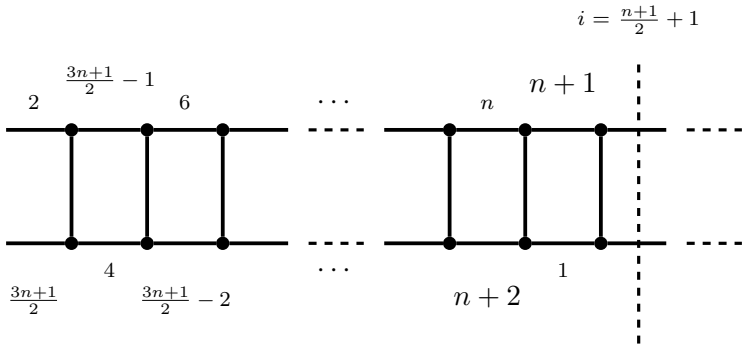


Figure 13: The edge labeling in Step 3 of $P_2 \times C_n$ for $n \equiv 3 \mod 4$

Step 4 We label the edge $v_i u_{i+1}$ for all $\frac{n-1}{2} < i < n$ with the numbers $\frac{3n+1}{2}+1, \frac{3n+1}{2}+2, \frac{3n+1}{2}+3, \ldots, 2n$ in that order in backward direction from $i = n-1$ to $i = \frac{n+1}{2}$ (see Figures 15 and 16).

Step 5 We label the edge $u_i v_i$ for all $1 \leq i \leq n$ with the numbers $2n+1, 2n+2, 2n+3, \ldots, 3n$ in that order from right to left (see Figure 17).

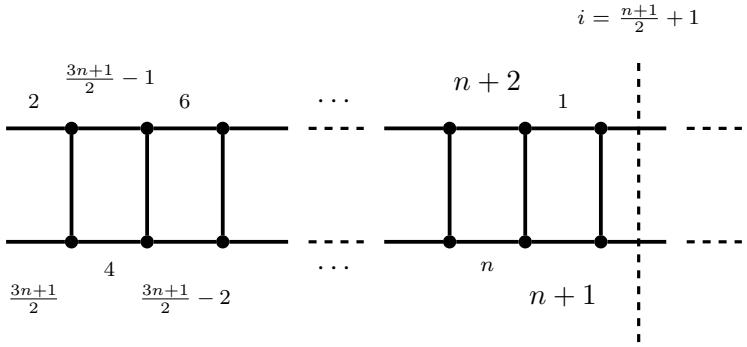


Figure 14: The edge labeling in Step 3 of $P_2 \times C_n$ for $n \equiv 1 \mod 4$

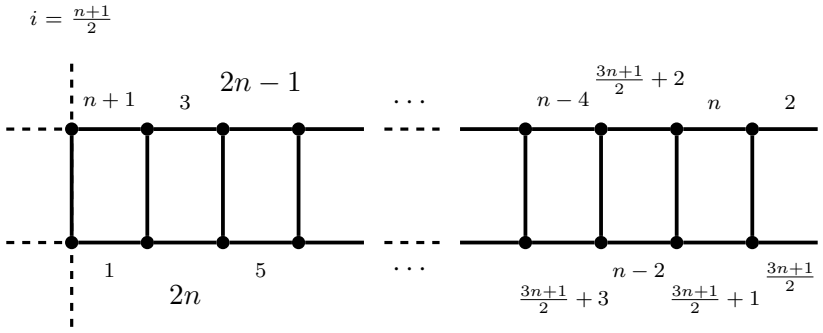


Figure 15: The edge labeling in Step 4 of $P_2 \times C_n$ for $n \equiv 3 \mod 4$

We can observe that this labeling can be summarized as follows:

$$f(u_i v_{i+1}) = \begin{cases} 2(i+1) & ; 1 \le i \le \lfloor \frac{n}{2} \rfloor \\ 2(i+1) - n & ; \lfloor \frac{n}{2} \rfloor < i < n \end{cases}$$

$$f(v_i u_{i+1}) = \begin{cases} 2n - \lfloor \frac{n}{2} \rfloor - i & ; 1 \le i \le \lceil \frac{n}{2} \rceil \\ 3n - \lfloor \frac{n}{2} \rfloor - i & ; \lceil \frac{n}{2} \rceil < i < n \end{cases}$$

$$f(v_n v_1) = 2$$

$$f(u_n u_1) = \frac{3n+1}{2}$$

$$f(u_i v_i) = 3n - i + 1 \text{ for all } 1 \le i \le n.$$

Thus, we obtain that $w(v_1) = f(v_n v_1) + f(u_1 v_1) + f(v_1 u_2) = \frac{1}{2}(9n+3)$, $w(u_1) = f(u_n u_1) + f(u_1 v_1) + f(u_1 v_2) = \frac{1}{2}(9n+9)$, $w(u_i) = f(v_{i-1} u_i) + f(u_i v_i) + f(u_i v_{i+1}) = \frac{1}{2}(9n+9)$ for $1 < i < \lceil \frac{n}{2} \rceil$, $\lceil \frac{n}{2} \rceil + 1 < i < n$.

In case $i = \lceil \frac{n}{2} \rceil, \lceil \frac{n}{2} \rceil + 1$, $w(u_i) = f(v_{i-1} u_i) + f(u_i v_i) + f(u_i v_{i+1}) = \frac{1}{2}(7n+9)$. Likewise, $w(v_n) = f(v_n v_1) + f(u_n v_n) + f(u_{n-1} v_n) = 3n+3$, $w(u_n) = f(u_n u_1) + f(u_n v_n) + f(v_{n-1} u_n) = 5n+3$, $w(v_i) = f(u_{i-1} v_i) + f(u_i v_i) + f(v_i u_{i+1}) = \frac{1}{2}(9n+3)$ for $1 < i < n$.

Therefore, f is a local antimagic labeling that induces 5 distinct vertex colors, including

(1) $w(u_n) = 5n+3$,

(2) $w(u_i) = \frac{1}{2}(9n+9)$ when $i \in \{1, 2, 3, \ldots, n-1\} \setminus \{\lceil \frac{n}{2} \rceil, \lceil \frac{n}{2} \rceil + 1\}$,

(3) $w\left(u_{\lceil \frac{n}{2} \rceil}\right) = w\left(u_{\lceil \frac{n}{2} \rceil + 1}\right) = \frac{1}{2}(7n+9)$,

(4) $w(v_i) = \frac{1}{2}(9n+3)$ for $1 \le i < n$,

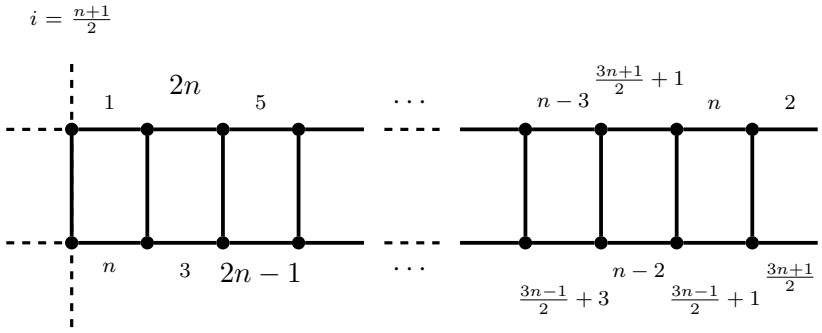


Figure 16: The edge labeling in Step 4 of $P_2 \times C_n$ for $n \equiv 1 \mod 4$

Figure 17: The edge labeling in Step 5 of $P_2 \times C_n$ for odd n

(5) $w(v_n) = 3n + 3$.

Case 2. n is even.

Similarly, we can show an algorithm of this labeling in the following five steps.

Step 1 We divide the set $\{1, 2, 3, \ldots, 3n\}$ into six subsets as follows: $\{2, 4, \ldots, n\}$, $\{1, 3, \ldots, n-1\}$, $\{n+1, n+2, n+3, \ldots, \frac{3n}{2}\}$, $\{\frac{3n+2}{2}\}$, $\{\frac{3n+2}{2}+1, \frac{3n+2}{2}+2, \frac{3n+2}{2}+3, \ldots, 2n+1\}$ and $\{2n+2, 2n+3, 2n+4, \ldots, 3n\}$.

Step 2 We label the edge $u_n v_1$ with the numbers 2 and label the edge $u_i v_{i+1}$ for all $1 \leq i < n$ with the numbers $4, 6, \ldots, n, 1, 3, 5, \ldots, n-1$ in that order from left to right.

Step 3 We label the edge $v_n u_1$ with the numbers $\frac{3n+2}{2}+1$ and label the edge $v_i u_{i+1}$ for all $1 \leq i \leq \frac{n-1}{2}$ with the numbers $n+1, n+2, n+3, \ldots, \frac{3n}{2}$ in that order in backward direction from $i = \frac{n-1}{2}$ to $i = 1$.

Step 4 We label the edge $v_i u_{i+1}$ for all $\frac{n}{2} < i < n$ with the numbers $\frac{3n+2}{2}+2, \frac{3n+2}{2}+3, \frac{3n+2}{2}+4, \ldots, 2n+1$ in that order in backward direction from $i = n-1$ to $i = \frac{n}{2}+1$.

Step 5 We label the edge $u_i v_i$ for all $1 \leq i \leq n$ with the numbers $2n+2, 2n+3, 2n+4, \ldots, 3n-1, 3n, \frac{3n+2}{2}$ in that order from right to left (see Figure 18).

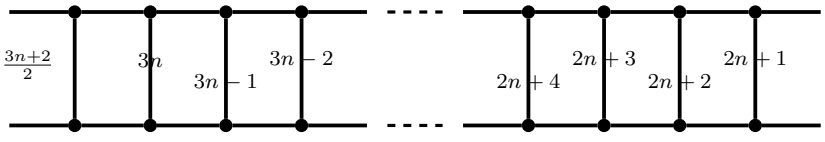


Figure 18: The edge labeling in Step 5 of $P_2 \times C_n$ for even n

We can observe that this labeling can be summarize in the following functions.

$$f(u_i v_{i+1}) = \begin{cases} 2(i+1) & ; 1 \leq i \leq \lfloor \frac{n}{2} \rfloor \\ 2(i+1) - n - 1 & ; \lfloor \frac{n}{2} \rfloor < i < n \end{cases}$$

$$f(v_i u_{i+1}) = \begin{cases} 2n - \lfloor \frac{n}{2} \rfloor - i + 1 & ; 1 \leq i \leq \lceil \frac{n}{2} \rceil \\ 3n - \lfloor \frac{n}{2} \rfloor - i + 2 & ; \lceil \frac{n}{2} \rceil < i < n \end{cases}$$

$$f(u_n v_1) = 2$$

$$f(v_nu_1) = 2n - \left\lfloor \frac{n}{2} \right\rfloor + 2$$
$$f(u_1v_1) = \frac{3n+2}{2}$$
$$f(u_iv_i) = 3n - i + 2 \text{ for all } 1 < i \leq n.$$

Thus, we obtain that $w(v_1) = f(v_nv_1) + f(u_1v_1) + f(v_1u_2) = 3n + 3$, $w(u_1) = f(u_nu_1) + f(u_1v_1) + f(u_1v_2) = 3n + 7$, $w(u_i) = f(v_{i-1}u_i) + f(u_iv_i) + f(u_iv_{i+1}) = \frac{9n}{2} + 6$ for $1 < i < \left\lceil \frac{n}{2} \right\rceil$, $\left\lceil \frac{n}{2} \right\rceil + 1 < i < n$.

In case $i = \left\lceil \frac{n}{2} \right\rceil, \left\lceil \frac{n}{2} \right\rceil + 1$, $w(u_i) = f(v_{i-1}u_i) + f(u_iv_i) + f(u_iv_{i+1}) = \frac{7n}{2} + 5$. Likewise, $w(v_n) = f(v_nv_1) + f(u_nv_n) + f(u_{n-1}v_n) = \frac{9n}{2} + 3$, $w(u_n) = f(u_nu_1) + f(u_nv_n) + f(v_{n-1}u_n) = \frac{7n}{2} + 7$, $w(v_i) = f(u_{i-1}v_i) + f(u_iv_i) + f(v_iu_{i+1}) = \frac{9n}{2} + 3$ for $1 < i < n$.

Therefore, f is a local antimagic labeling that induces 6 distinct vertex colors, including

(1) $w(u_1) = 3n + 7$,

(2) $w(u_i) = \frac{9n}{2} + 6$ when $i \in \{2, 3, \ldots, n-1\} \setminus \left\{ \left\lceil \frac{n}{2} \right\rceil, \left\lceil \frac{n}{2} \right\rceil + 1 \right\}$,

(3) $w\left(u_{\left\lceil \frac{n}{2} \right\rceil}\right) = w\left(u_{\left\lceil \frac{n}{2} \right\rceil + 1}\right) = \frac{7n}{2} + 5$,

(4) $w(u_n) = \frac{7n}{2} + 7$,

(5) $w(v_1) = 3n + 3$,

(6) $w(v_i) = \frac{9n}{2} + 3$ for $1 < i \leq n$.

Hence, we conclude that

$$\chi_{la}(P_2 \times C_n) \leq \begin{cases} 5 & \text{if } n \text{ is odd} \\ 6 & \text{if } n \text{ is even.} \end{cases}$$

□

2.3 Local Antimagic Chromatic Number of $C_m \times C_n$

In this subsection, we consider the graph $C_m \times C_n$ for any $m, n \geq 3$ with $V(C_m \times C_n) = \{u_{i,j} : 1 \leq i \leq m, 1 \leq j \leq n\}$ and $E(C_m \times C_n) = \{u_{i,j}u_{i+1,j} : 1 \leq i \leq m, 1 \leq j \leq n\} \cup \{u_{i,j}u_{i,j+1} : 1 \leq i \leq m, 1 \leq j \leq n\}$ (see Figure 19). Note that all addition within the first and second indices is performed in modulo m and n, respectively. Clearly, $|E(C_m \times C_n)| = 2mn$.

We separate into four cases by the parities of m and n, namely (1) m, n are even, (2) m is odd and n is even where $m < n$, (3) m is even and n is odd where $m < n$, (4) m, n are odd. To obtain an upper bound for $\chi_{la}(C_m \times C_n)$ in each case, we will use the edge labeling of $C_m \times C_n$ given by the following algorithm.

Step 1 We divide the set $\{1, 2, 3, \ldots, 2mn\}$ into four subsets as follows: $\{1, 2, 3, \ldots, x\}$, $\{x+1, x+2, x+3, \ldots, mn\}$, $\{mn+1, mn+2, mn+3 \ldots, mn+y\}$ and $\{mn+y+1, mn+y+2, mn+y+3, \ldots, 2mn\}$ where $x = \left\lceil \frac{m}{2} \right\rceil \left\lfloor \frac{n}{2} \right\rfloor + \left\lceil \frac{n}{2} \right\rceil \left\lfloor \frac{m}{2} \right\rfloor$ and $y = \left\lceil \frac{m}{2} \right\rceil \left\lceil \frac{n}{2} \right\rceil + \left\lfloor \frac{m}{2} \right\rfloor \left\lfloor \frac{n}{2} \right\rfloor$.

Step 2 We label the edge $u_{i,j}u_{i+1,j}$ for all i and j of different parity with the numbers $1, 2, 3, \ldots, x-1, x$ starting from the edges with $i = 1$ from left to right followed by the edges with $i = 2$ from left to right, and so on (see an example in Figure 20).

Step 3 We label the edge $u_{i,j}u_{i,j+1}$ for all i and j of the same parity with the numbers $x+1, x+2, x+3, \ldots, mn-1, mn$ in the opposite direction of Step 2. It means that we start from the edges with $i = m$ from right to left followed by the edges with $i = m-1$ from right to left, and so on (see an example in Figure 21).

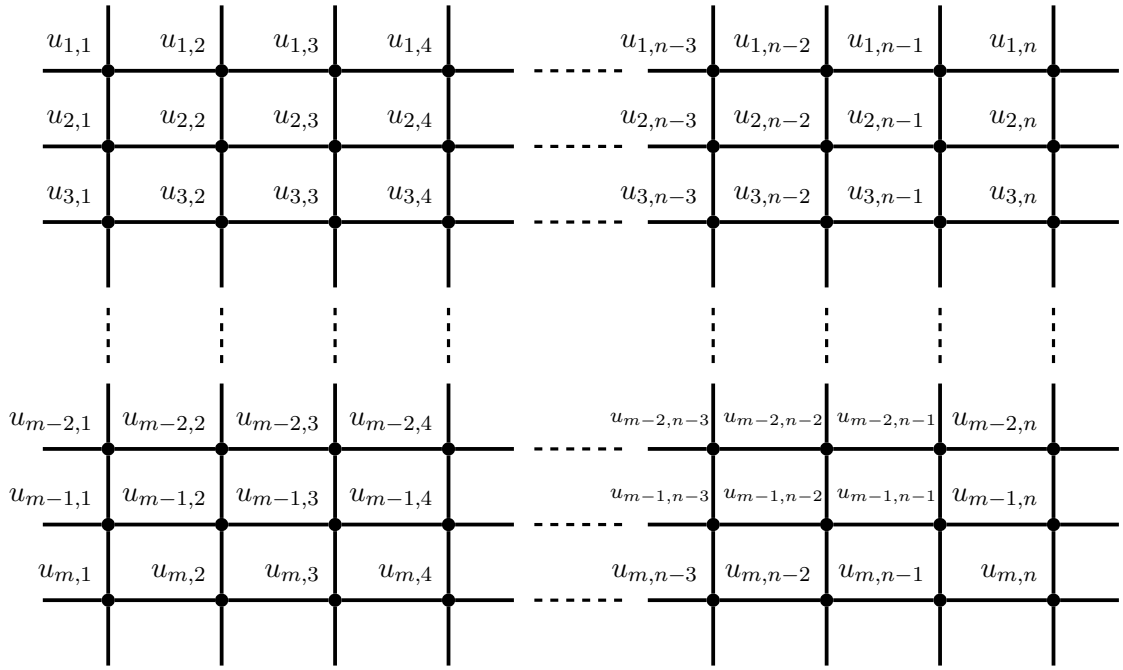


Figure 19: The graph $C_m \times C_n$

Step 4 We label the edge $u_{i,j}u_{i,j+1}$ for all i and j of different parity if m is even, and for all i and j of the same parity if m is odd with the numbers $mn+1, mn+2, mn+3, \ldots, mn+y-1, mn+y$ starting from the edges with $j = 1$ from bottom to top followed by the edges with $j = 2$ from bottom to top, and so on (see an example in Figure 22).

Step 5 We label the edge $u_{i,j}u_{i,j+1}$ for all i and j of the same parity if m is even or i and j of different parity if m is odd with the numbers $mn+y+1, mn+y+2, mn+y+3, \ldots, 2mn-1, 2mn$ in the opposite direction of step 4. It means that we start from the edges with $j = n$ from top to bottom followed by the edges with $j = n-1$ from top to bottom, and so on (see an example in Figure 23).

Example 2.3. *The algorithm of labeling on the graph $C_5 \times C_8$.*

Step 1 We divide the set $\{1, 2, 3, \ldots, 80\}$ into four subsets as follows:
$\{1, 2, 3, \ldots, 20\}$, $\{21, 22, 23, \ldots, 40\}$, $\{41, 42, 43 \ldots, 60\}$ and $\{61, 62, 63, \ldots, 80\}$.

Step 2 We label the edge $u_{i,j}u_{i+1,j}$ for all i and j of different parity with the numbers $1, 2, 3, \ldots, 19, 20$ starting from the edges with $i = 1$ from left to right followed by the edges with $i = 2$ from left to right, and so on (see Figure 20).

Step 3 We label the edge $u_{i,j}u_{i,j+1}$ for all i and j of the same parity with the numbers $21, 22, 23, \ldots, 39, 40$ in the opposite direction of step 2. It means that we start from the edges with $i = 5$ from right to left followed by the edges with $i = 4$ from right to left, and so on (see Figure 21).

Step 4 We label the edge $u_{i,j}u_{i,j+1}$ for all i and j of the same parity with the numbers $41, 42, 43, \ldots, 59, 60$ starting from the edges with $j = 1$ from bottom to top followed by the edges with $j = 2$ from bottom to top, and so on (see Figure 22).

Step 5 We label the edge $u_{i,j}u_{i,j+1}$ for all i and j of different parity with the numbers $61, 62, 63, \ldots, 79, 80$ in the opposite direction of step 4. It means that we start from the edges with $j = 8$ from top to bottom followed by the edges with $j = 7$ from top to bottom, and so on (see Figure 23).

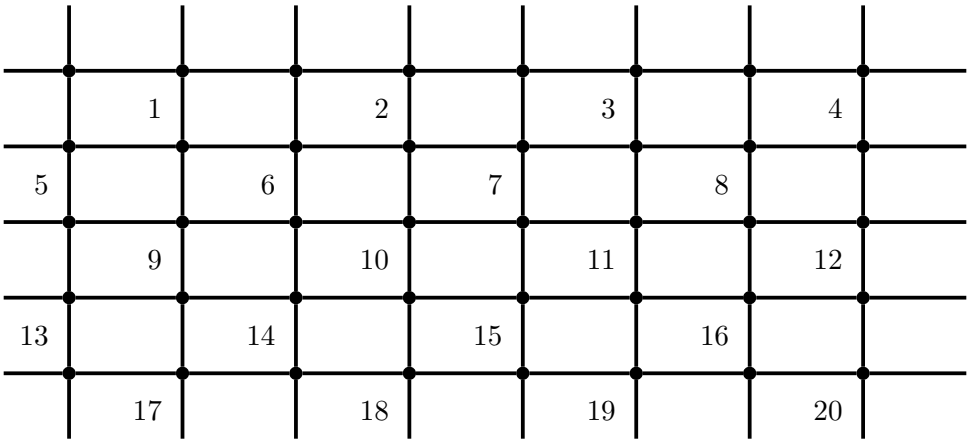


Figure 20: The edge labeling in Step 2 of $C_5 \times C_8$

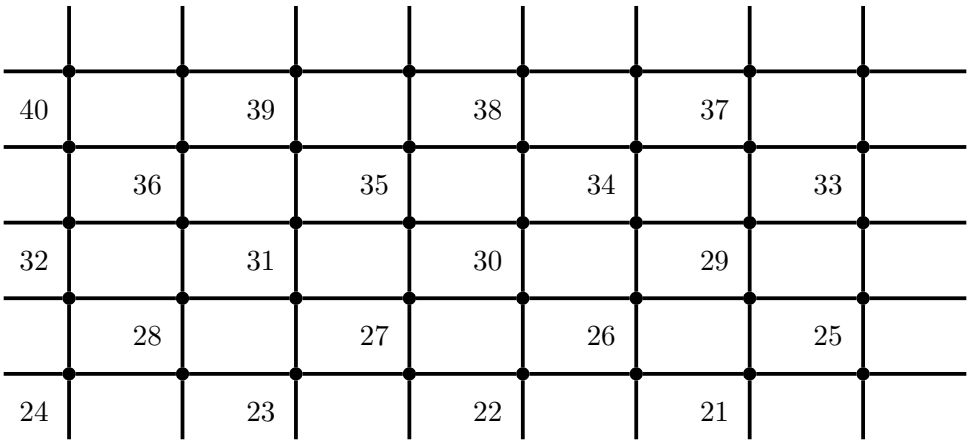


Figure 21: The edge labeling in Step 3 of $C_5 \times C_8$

Next, we prove Theorem 1.4 *(i)-(iv)* corresponding with each case of m and n.

Proof of Theorem 1.4. (i) Let $f : E(C_m \times C_n) \to \{1, 2, 3, \ldots, 2mn\}$ be the edge labeling obtained from the algorithm. It suffices to check that f is a local antimagic labeling that induces 5 distinct vertex colors.

Step 2

$$f(u_{i,j}u_{i+1,j}) = \begin{cases} \left(\frac{i-1}{2}\right) n + \frac{j}{2} & \text{if } i \text{ is odd, } j \text{ is even} \\ \left(\frac{i-1}{2}\right) n + \frac{j+1}{2} & \text{if } i \text{ is even, } j \text{ is odd.} \end{cases}$$

Step 3

$$f(u_{i,j}u_{i+1,j}) = \begin{cases} mn - \left(\frac{i-1}{2}\right) n - \frac{j-1}{2} & \text{if } i, j \text{ are odd} \\ mn - \left(\frac{i-1}{2}\right) n - \frac{j-2}{2} & \text{if } i, j \text{ are even.} \end{cases}$$

Step 4

$$f(u_{i,j}u_{i,j+1}) = \begin{cases} mn + \left(\frac{j}{2}\right) m - \frac{i-1}{2} & \text{if } i \text{ is odd, } j \text{ is even} \\ mn + \left(\frac{j}{2}\right) m - \frac{i-2}{2} & \text{if } i \text{ is even, } j \text{ is odd.} \end{cases}$$

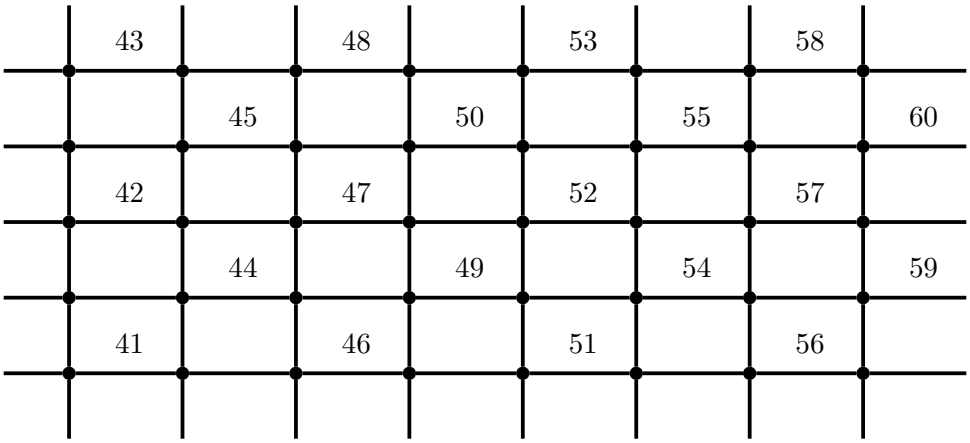


Figure 22: The edge labeling in Step 4 of $C_5 \times C_8$

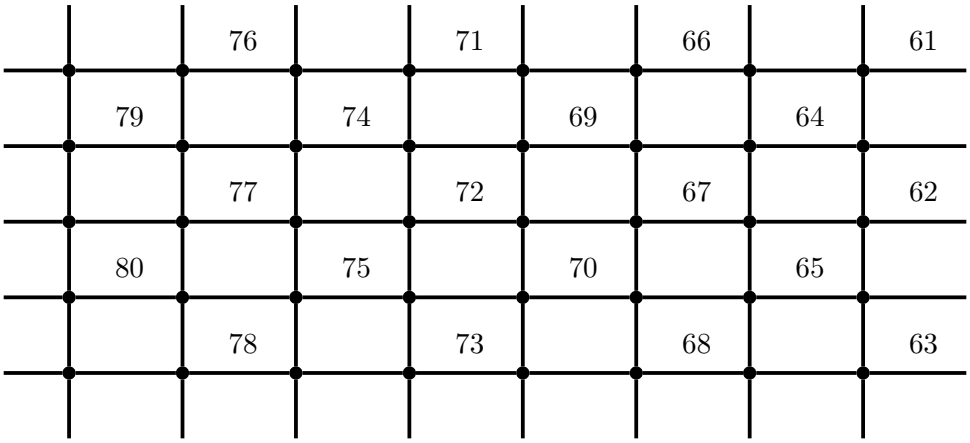


Figure 23: The edge labeling in Step 5 of $C_5 \times C_8$

Step 5

$$f(u_{i,j}u_{i,j+1}) = \begin{cases} 2mn - \left(\frac{j}{2}\right)m + \frac{i+1}{2} & \text{if } i,j \text{ are odd} \\ 2mn - \left(\frac{j}{2}\right)m + \frac{i}{2} & \text{if } i,j \text{ are even.} \end{cases}$$

It is clear that f is a bijection. Thus, we obtain that $w(u_{i,j}) = f(u_{i-1,j}u_{i,j}) + f(u_{i,j}u_{i+1,j}) + f(u_{i,j+1}u_{i,j+2}) + f(u_{i,j}u_{i,j+1})$ for all $2 \leq i \leq m$ and $2 \leq j \leq n$ such that

(1) $w(u_{i,j}) = \frac{1}{2}(8mn - m - n + 4)$ if $i \equiv j \mod 2$

(2) $w(u_{i,j}) = \frac{1}{2}(8mn + m + n + 4)$ if $i \not\equiv j \mod 2$.

Moreover,

$$\begin{aligned} w(u_{1,1}) &= f(u_{m,1}u_{1,1}) + f(u_{1,1}u_{2,1}) + f(u_{1,1}u_{1,2}) + f(u_{1,n}u_{1,1}) \\ &= \frac{1}{2}(10mn - m - n + 4), \\ w(u_{1,j}) &= f(u_{m,j}u_{1,j}) + f(u_{1,j}u_{2,j}) + f(u_{1,j-1}u_{1,j}) + f(u_{1,j}u_{1,j+1}) \\ &= \begin{cases} \frac{1}{2}(7mn + m + n + 4) & \text{if } j \text{ is even} \\ \frac{1}{2}(9mn - m - n + 4) & \text{if } j \text{ is odd} \end{cases} \end{aligned}$$

and

$$\begin{aligned} w(u_{i,1}) &= f(u_{i-1,1}u_{i,1}) + f(u_{i,1}u_{i+1,1}) + f(u_{i,n}u_{i,1}) + f(u_{i,1}u_{i,2}) \\ &= \begin{cases} \frac{1}{2}(7mn+m+n+4) & \text{if } i \text{ is even} \\ \frac{1}{2}(9mn-m-n+4) & \text{if } i \text{ is odd.} \end{cases} \end{aligned}$$

It is clear that $w(u_{1j}) = w(u_{i1})$ if i and j have the same parity. Therefore, f is a local antimagic labeling that induces 5 distinct vertex colors, including

(1) $A_1 = \frac{1}{2}(8mn-m-n+4) = \frac{1}{2}(7mn-m-n+4) + \frac{1}{2}(mn)$,

(2) $A_2 = \frac{1}{2}(8mn+m+n+4) = \frac{1}{2}(7mn-m-n+4) + \frac{1}{2}(mn+2m+2n)$,

(3) $A_3 = \frac{1}{2}(10mn-m-n+4) = \frac{1}{2}(7mn-m-n+4) + \frac{1}{2}(3mn)$,

(4) $A_4 = \frac{1}{2}(7mn+m+n+4) = \frac{1}{2}(7mn-m-n+4) + \frac{1}{2}(2m+2n)$,

(5) $A_5 = \frac{1}{2}(7mn+m+n+4) = \frac{1}{2}(7mn-m-n+4) + \frac{1}{2}(2mn)$.

Clearly, for a given $k = 1, 2, 3, 4, 5$, the vertices of color A_k are not adjacent. To check that A_1, A_2, A_3, A_4, A_5 are distinct, we claim that $A_4 < A_1 < A_2 < A_5 < A_3$. It is obvious that $A_1 < A_2$ and $A_5 < A_3$. Since $(m,n) \neq (4,4)$, we have $m > 4$ or $n > 4$. Then $mn = \frac{1}{2}mn + \frac{1}{2}mn > \frac{4}{2}m + \frac{4}{2}n = 2m + 2n$. implying $A_4 < A_1$. Since $2mn = mn + mn > mn + 2m + 2n$, we have that $A_2 < A_5$.

(ii) Let $f : E(C_m \times C_n) \to \{1, 2, 3, \ldots, 2mn\}$ be the edge labeling obtained from the algorithm. It suffices to check that f is a local antimagic labeling that induces $n+2$ distinct vertex colors.

Step 2

$$f(u_{i,j}u_{i+1,j}) = \begin{cases} \left(\frac{i-1}{2}\right)n + \frac{j}{2} & \text{if } i \text{ is odd, } j \text{ is even} \\ \left(\frac{i-1}{2}\right)n + \frac{j+1}{2} & \text{if } i \text{ is even, } j \text{ is odd.} \end{cases}$$

Step 3

$$f(u_{i,j}u_{i+1,j}) = \begin{cases} mn - \left(\frac{i}{2}\right)n + \frac{n-j+1}{2} & \text{if } i,j \text{ are odd} \\ mn - \left(\frac{i}{2}\right)n + \frac{n-j+2}{2} & \text{if } i,j \text{ are even.} \end{cases}$$

Step 4

$$f(u_{i,j}u_{i,j+1}) = mn + \left(\frac{j-1}{2}\right)m + \frac{m-i+2}{2}.$$

Step 5

$$f(u_{i,j}u_{i,j+1}) = 2mn - \left(\frac{j}{2}\right)m + \frac{i+1}{2}.$$

It is clear that f is a bijection. Thus, we obtain that $w(u_{i,j}) = f(u_{i-1,j}u_{i,j}) + f(u_{i,j}u_{i+1,j}) + f(u_{i,j+1}u_{i,j+2}) + f(u_{i,j}u_{i,j+1})$ for all $2 \leq i \leq m$ and $2 \leq j \leq n$ such that

(1) $w(u_{i,j}) = \frac{1}{2}(8mn+m-n+5)$ if $i \equiv j \mod 2$

(2) $w(u_{i,j}) = \frac{1}{2}(8mn-m+n+5)$ if $i \not\equiv j \mod 2$.

Moreover,

$$\begin{aligned} w(u_{1,1}) &= f(u_{m,1}u_{1,1}) + f(u_{1,1}u_{2,1}) + f(u_{1,1}u_{1,2}) + f(u_{1,n}u_{1,1}) \\ &= \frac{1}{2}(8mn + m + n + 3), \\ w(u_{1,j}) &= f(u_{m,j}u_{1,j}) + f(u_{1,j}u_{2,j}) + f(u_{1,j-1}u_{1,j}) + f(u_{1,j}u_{1,j+1}) \\ &= \begin{cases} \frac{1}{2}(7mn - m - n + 3 + 2j) & \text{if } j \text{ is even} \\ \frac{1}{2}(9mn + m + n + 5 - 2j) & \text{if } j \text{ is odd} \end{cases} \end{aligned}$$

and

$$\begin{aligned} w(u_{i,1}) &= f(u_{i-1,1}u_{i,1}) + f(u_{i,1}u_{i+1,1}) + f(u_{i,n}u_{i,1}) + f(u_{i,1}u_{i,2}) \\ &= \begin{cases} \frac{1}{2}(9mn - m + n + 5) & \text{if } i \text{ is even} \\ \frac{1}{2}(7mn + m - n + 5) & \text{if } i \text{ is odd.} \end{cases} \end{aligned}$$

Therefore, f is a local antimagic labeling that induces $n + 2$ distinct vertex colors, including

(1) $A_1 = \frac{1}{2}(8mn + m - n + 5) = \frac{1}{2}(7mn - m - n + 3) + \frac{1}{2}(mn + 2m + 2)$,

(2) $A_2 = \frac{1}{2}(8mn - m + n + 5) = \frac{1}{2}(7mn - m - n + 3) + \frac{1}{2}(mn + 2n + 2)$,

(3) $A_3 = \frac{1}{2}(8mn + m + n + 3) = \frac{1}{2}(7mn - m - n + 3) + \frac{1}{2}(mn + 2m + 2n)$,

(4) $A_4 = \frac{1}{2}(9mn - m + n + 5) = \frac{1}{2}(7mn - m - n + 3) + \frac{1}{2}(2mn + 2n + 2)$,

(5) $A_5 = \frac{1}{2}(7mn + m - n + 5) = \frac{1}{2}(7mn - m - n + 3) + \frac{1}{2}(2m + 2)$,

(6) $A_6^{(j)} = \frac{1}{2}(7mn - m - n + 3 + 2j) = \frac{1}{2}(7mn - m - n + 3) + \frac{1}{2}(2j)$ for even $2 \le j \le n$,

(7) $A_7^{(j)} = \frac{1}{2}(9mn + m + n + 5 - 2j) = \frac{1}{2}(7mn - m - n + 3) + \frac{1}{2}(2mn + 2m + 2n + 2 - 2j)$ for odd $2 \le j \le n$.

Clearly, for a given $k = 1, 2, 3, 4, 5$, the vertices of color A_k are not adjacent and for a given $k = 6, 7$, the vertices of color $A_k^{(j)}$ are not adjacent. Then we claim that $A_1, A_2, A_3, A_4, A_5, A_6^{(j)}$ and $A_7^{(j)}$ are distinct except $A_6^{(m+1)} = A_5$ and $A_7^{(m)} = A_4$. Note that the vertex of color $A_6^{(m+1)}$ is not adjcent to the vertices of color A_5 and the vertex of color $A_7^{(m)}$ is not adjcent to the vertices of color A_4. Obviously, each $A_6^{(j)}$ is different and each $A_7^{(j)}$ is different. It is enough to show that $A_6^{(j)} < A_1 < A_2 < A_3 < A_7^{(j)}$. It is obvious that $A_1 < A_2 < A_3$. Then $A_6^{(j)} \le \frac{1}{2}(7mn - m - n + 3) + \frac{1}{2}(2n) < \frac{1}{2}(7mn - m - n + 3) + \frac{1}{2}(mn + 2m + 2) = A_1$ and $A_7^{(j)} \ge \frac{1}{2}(7mn - m - n + 3) + \frac{1}{2}(2mn + 2m + 2) > \frac{1}{2}(7mn - m - n + 3) + \frac{1}{2}(mn + 2m + 2n) = A_3$.

(iii) Let $f : E(C_m \times C_n) \to \{1, 2, 3, \ldots, 2mn\}$ be the edge labeling obtained from the algorithm. It suffices to check that f is a local antimagic labeling that induces $m + 4$ distinct vertex colors.

Step 2

$$f(u_{i,j}u_{i+1,j}) = \left(\frac{i-1}{2}\right)n + \frac{j}{2}.$$

Step 3

$$f(u_{i,j}u_{i+1,j}) = mn - \left(\frac{i}{2}\right)n + \frac{n - j + 1}{2}.$$

Step 4

$$f(u_{i,j}u_{i,j+1}) = \begin{cases} mn + \left(\frac{j}{2}\right)m - \frac{i-1}{2} & \text{if } i \text{ is odd, } j \text{ is even} \\ mn + \left(\frac{j}{2}\right)m - \frac{i-2}{2} & \text{if } i \text{ is even, } j \text{ is odd.} \end{cases}$$

Step 5

$$f(u_{i,j}u_{i,j+1}) = \begin{cases} 2mn - \left(\frac{j}{2}\right)m + \frac{i+1}{2} & \text{if } i,j \text{ are odd} \\ 2mn - \left(\frac{j}{2}\right)m + \frac{i}{2} & \text{if } i,j \text{ are even.} \end{cases}$$

It is clear that f is a bijection. Thus, we obtain that $w(u_{i,j}) = f(u_{i-1,j}u_{i,j}) + f(u_{i,j}u_{i+1,j}) + f(u_{i,j+1}u_{i,j+2}) + f(u_{i,j}u_{i,j+1})$ for all $2 \le i \le m$ and $2 \le j \le n$ such that

(1) $w(u_{i,j}) = \frac{1}{2}(8mn - m - n + 3)$ if $i \equiv j \mod 2$,

(2) $w(u_{i,j}) = \frac{1}{2}(8mn + m + n + 3)$ if $i \not\equiv j \mod 2$.

Moreover,

$$\begin{aligned} w(u_{1,1}) &= f(u_{m,1}u_{1,1}) + f(u_{1,1}u_{2,1}) + f(u_{1,1}u_{1,2}) + f(u_{1,n}u_{1,1}) \\ &= \frac{1}{2}(10mn - m - n + 5), \\ w(u_{1,j}) &= f(u_{m,j}u_{1,j}) + f(u_{1,j}u_{2,j}) + f(u_{1,j-1}u_{1,j}) + f(u_{1,j}u_{1,j+1}) \\ &= \begin{cases} \frac{1}{2}(7mn + m + n + 3) & \text{if } j \text{ is even} \\ \frac{1}{2}(9mn - m - n + 3) & \text{if } j \text{ is odd} \end{cases} \end{aligned}$$

and

$$\begin{aligned} w(u_{i,1}) &= f(u_{i-1,1}u_{i,1}) + f(u_{i,1}u_{i+1,1}) + f(u_{i,n}u_{i,1}) + f(u_{i,1}u_{i,2}) \\ &= \begin{cases} \frac{1}{2}(7mn + m + n + 5 - 2i) & \text{if } i \text{ is even} \\ \frac{1}{2}(9mn - m - n + 3 + 2i) & \text{if } i \text{ is odd.} \end{cases} \end{aligned}$$

Therefore, f is a local antimagic labeling that induces $m + 4$ distinct vertex colors, including

(1) $A_1 = \frac{1}{2}(8mn - m - n + 3) = \frac{1}{2}(7mn - m - n + 3) + \frac{1}{2}(mn)$,

(2) $A_2 = \frac{1}{2}(8mn + m + n + 3) = \frac{1}{2}(7mn - m - n + 3) + \frac{1}{2}(mn + 2m + 2n)$,

(3) $A_3 = \frac{1}{2}(10mn - m - n + 5) = \frac{1}{2}(7mn - m - n + 3) + \frac{1}{2}(3mn + 2)$,

(4) $A_4 = \frac{1}{2}(7mn + m + n + 3) = \frac{1}{2}(7mn - m - n + 3) + \frac{1}{2}(2m + 2n)$,

(5) $A_5 = \frac{1}{2}(9mn - m - n + 3) = \frac{1}{2}(7mn - m - n + 3) + \frac{1}{2}(2mn)$,

(6) $A_6^{(i)} = \frac{1}{2}(7mn + m + n + 5 - 2i) = \frac{1}{2}(7mn - m - n + 3) + \frac{1}{2}(2m + 2n + 2 - 2i)$ for even $2 \le i \le m$,

(7) $A_7^{(i)} = \frac{1}{2}(9mn - m - n + 3 + 2i) = \frac{1}{2}(7mn - m - n + 3) + \frac{1}{2}(2mn + 2i)$ for odd $2 \le i \le m$,.

Clearly, for a given $k = 1, 2, 3, 4, 5$, the vertices of color A_k are not adjacent and for a given $k = 6, 7$, the vertices of color $A_k^{(i)}$ are not adjacent. Obviously, each $A_6^{(i)}$ is different and each $A_7^{(i)}$ is different. To check that $A_1, A_2, A_3, A_4, A_5, A_6^{(i)}$ and $A_7^{(i)}$ are distinct, it is enough to show that $A_6^{(i)} < A_4 < A_1 < A_2 < A_5 < A_7^{(i)} < A_3$. Obviously, $A_1 < A_2 < A_5 < A_7^{(i)}$. Since $mn \ge 4m > 2m + 2n$, we have $A_4 < A_1$. We can see that $A_6^{(i)} \le \frac{1}{2}(7mn - m - n + 3) + \frac{1}{2}(2m + 2n - 2) < A_4$ for even $2 \le i \le m$. Moreover, $A_7^{(i)} < \frac{1}{2}(7mn - m - n + 3) + \frac{1}{2}(2mn + 2m) <$

$\frac{1}{2}(7mn - m - n + 3) + \frac{1}{2}(3mn + 2) = A_3$ for odd $2 \leq i \leq m$.

(iv) Let $f : E(C_m \times C_n) \to \{1, 2, 3, \ldots, 2mn\}$ be the edge labeling obtained from the algorithm. It suffices to check that f is a local antimagic labeling that induces $m + n + 1$ distinct vertex colors.

Step 2

$$f(u_{i,j}u_{i+1,j}) = \left(\frac{i-1}{2}\right)n + \frac{j}{2}.$$

Step 3

$$f(u_{i,j}u_{i+1,j}) = mn - \left(\frac{i}{2}\right)n + \frac{n-j+1}{2}.$$

Step 4

$$f(u_{i,j}u_{i,j+1}) = mn + \left(\frac{j-1}{2}\right)m + \frac{m-i+2}{2}.$$

Step 5

$$f(u_{i,j}u_{i,j+1}) = 2mn - \left(\frac{j}{2}\right)m + \frac{i+1}{2}.$$

It is clear that f is a bijection. Thus, we obtain that $w(u_{i,j}) = f(u_{i-1,j}u_{i,j}) + f(u_{i,j}u_{i+1,j}) + f(u_{i,j+1}u_{i,j+2}) + f(u_{i,j}u_{i,j+1})$ for all $2 \leq i \leq m$ and $2 \leq j \leq n$ such that

(1) $w(u_{i,j}) = \frac{1}{2}(8mn + m - n + 4)$ if $i \equiv j \mod 2$,

(2) $w(u_{i,j}) = \frac{1}{2}(8mn - m + n + 4)$ if $i \not\equiv j \mod 2$.

Moreover,

$$\begin{aligned} w(u_{1,1}) &= f(u_{m,1}u_{1,1}) + f(u_{1,1}u_{2,1}) + f(u_{1,1}u_{1,2}) + f(u_{1,n}u_{1,1}) \\ &= \frac{1}{2}(8mn + m + n + 2), \\ w(u_{1,j}) &= f(u_{m,j}u_{1,j}) + f(u_{1,j}u_{2,j}) + f(u_{1,j-1}u_{1,j}) + f(u_{1,j}u_{1,j+1}) \\ &= \begin{cases} \frac{1}{2}(7mn - m - n + 3 + 2j) & \text{if } j \text{ is even} \\ \frac{1}{2}(9mn + m + n + 5 - 2j) & \text{if } j \text{ is odd} \end{cases} \end{aligned}$$

and

$$\begin{aligned} w(u_{i,1}) &= f(u_{i-1,1}u_{i,1}) + f(u_{i,1}u_{i+1,1}) + f(u_{i,n}u_{i,1}) + f(u_{i,1}u_{i,2}) \\ &= \begin{cases} \frac{1}{2}(9mn - m + n + 3 + 2i) & \text{if } i \text{ is even} \\ \frac{1}{2}(7mn + m - n + 5 - 2i) & \text{if } i \text{ is odd.} \end{cases} \end{aligned}$$

Therefore, f is a local antimagic labeling that induces $m+n+1$ distinct vertex colors, including

(1) $A_1 = \frac{1}{2}(8mn + m - n + 4) = \frac{1}{2}(7mn - m - n + 2) + \frac{1}{2}(mn + 2m + 2)$,

(2) $A_2 = \frac{1}{2}(8mn - m + n + 4) = \frac{1}{2}(7mn - m - n + 2) + \frac{1}{2}(mn + 2n + 2)$,

(3) $A_3 = \frac{1}{2}(8mn + m + n + 2) = \frac{1}{2}(7mn - m - n + 2) + \frac{1}{2}(mn + 2m + 2n)$,

(4) $A_4^{(j)} = \frac{1}{2}(7mn - m - n + 3 + 2j) = \frac{1}{2}(7mn - m - n + 2) + \frac{1}{2}(1 + 2j)$ for even $2 \leq j \leq n$,

(5) $A_5^{(j)} = \frac{1}{2}(9mn+m+n+5-2j) = \frac{1}{2}(7mn-m-n+2) + \frac{1}{2}(2mn+2m+2n+3-2j)$ for odd $2 \leq j \leq n$,

(6) $A_6^{(i)} = \frac{1}{2}(9mn-m+n+3+2i) = \frac{1}{2}(7mn-m-n+2) + \frac{1}{2}(2mn+2n+1+2i)$ for even $2 \leq i \leq m$,

(7) $A_7^{(i)} = \frac{1}{2}(7mn+m-n+5-2i) = \frac{1}{2}(7mn-m-n+2) + \frac{1}{2}(2m+2-2i)$ for odd $2 \leq i \leq m$.

Clearly, for a given $k = 1, 2, 3$, the vertices of color A_k are not adjacent, for a given $k = 4, 5$, the vertices of color $A_k^{(j)}$ are not adjacent and for a given $k = 6, 7$, the vertices of color $A_k^{(i)}$ are not adjacent. Obviously, each $A_4^{(j)}$ is different, each $A_5^{(j)}$ is different, each $A_6^{(i)}$ is different and each $A_7^{(i)}$ is different. To check that $A_1, A_2, A_3, A_4^{(j)}, A_5^{(j)}, A_6^{(i)}$ and $A_7^{(i)}$ are distinct, we claim that $\max\{A_4^{(j)}, A_7^{(i)}\} < A_1 < A_2 < A_3 < \min\{A_5^{(j)}, A_6^{(i)}\}$, $A_4^{(j)} \neq A_7^{(i)}$ and $A_5^{(j)} \neq A_6^{(i)}$. It is obvious that $A_1 < A_2 < A_3$. We can see that $A_4^{(j)} \leq \frac{1}{2}(7mn-m-n+3) + \frac{1}{2}(2n-1) < A_1$ and $A_7^{(i)} < \frac{1}{2}(7mn-m-n+3)+\frac{1}{2}(2m-4) < A_1$. Moreover, $A_6^{(i)} \geq \frac{1}{2}(7mn-m-n+3)+\frac{1}{2}(2mn+2n+5) > A_3$ and $A_5^{(j)} \geq \frac{1}{2}(7mn-m-n+3) + \frac{1}{2}(2mn+2m+3) > A_3$. Since $1+2j$ and $2m+2-2i$ have different parities, we have $A_4^{(j)} \neq A_7^{(i)}$. Next, we will show $A_5^{(j)} \neq A_6^{(i)}$. It is enough to check that $2mn+2m+2n+3-2j \neq 2mn+2n+1+2i$. We suppose $2m+3-2j = 1+2i$. Then $m+1 = i+j$ for all odd i, even j which is a contradiction. Thus, $A_5^{(j)} \neq A_6^{(i)}$.

(v) To find a lower bound for the local antimagic number of $C_m \times C_n$, if both m and n are even, then we observe that $C_m \times C_n$ is a bipartite graph whose partite sets have the same size. Thus, $\chi_{la}(C_m \times C_n) > 2$ by Corollary 2.2. Otherwise, we have $\chi_{la}(C_m \times C_n) \geq \chi(C_m \times C_n) = \max\{\chi(C_m), \chi(C_n)\} = 3$. □

3 Concluding Remarks

In this research, we obtained some bounds for the local antimagic chromatic number of cartesian product of some graphs, namely $P_2 \times P_n$, $P_2 \times C_n$ and $C_m \times C_n$. Moreover, we proved the exact local antimagic chromatic number of $P_2 \times P_3$. Lau and Shiu [14] found the exact values of $\chi_{la}(P_2 \times C_3)$ and $\chi_{la}(P_2 \times C_4)$. It would be interesting to determine the exact values of $\chi_{la}(P_2 \times P_n)$, $\chi_{la}(P_2 \times C_n)$ and $\chi_{la}(C_m \times C_n)$.

Problem 3.1. Determine the exact local antimagic chromatic number of $P_2 \times P_n$ for $n \geq 4$.

Problem 3.2. Determine the exact local antimagic chromatic number of $P_2 \times C_n$ for $n \geq 5$.

Problem 3.3. Determine the exact local antimagic chromatic number of $C_m \times C_n$ for $m, n \geq 3$.

In addition, the idea in the proof the upper bound for $\chi_{la}(C_m \times C_n)$ in Theorem 1.4 can be used to give upper bounds for $\chi_{la}(P_m \times P_n)$ and $\chi_{la}(P_m \times C_n)$. However, we are not able to determine the exact values.

Problem 3.4. Determine the exact local antimagic chromatic number of $P_m \times P_n$ and $P_m \times C_n$.

Acknowledgment. In the completion of this article, the second author would like to thank the Development and Promotion of Science and Technology Talents Project (DPST) scholarship for their financial support.

References

[1] S. Arumugam, Y. C. Lee, K. Premalatha, T. M. Wang, *On local antimagic vertex coloring for corona products of graphs*, Available from: https://arxiv.org/pdf/1808.04956v1.pdf. (2022, Jun 1).

[2] S. Arumugam, K. Premalatha, M. Bača, A. Semaničová-Fenovčíková, *Local antimagic vertex coloring of a graph*, Graphs Combi. **33** (2017), 275–285.

[3] M. Bača, A. Semaničová-Fenovčíková, T. M. Wang, *Local antimagic chromatic number for copies of graphs*, Mathematics **1230**(9) (2021), 1–12.

[4] J. Bensmail, M. Senhaji, K. S. Lyngsie, *On a combination of the 1-2-3 Conjecture and the Antimagic Labelling Conjecture*, Discret. Math. Theor. Comput. Sci. **19** (2017), 1–17.

[5] Y. Cheng, *Lattice Grids and Prisms are Antimagic*, Theor. Comput. Sci. **374** (2006), 66–73.

[6] Dafik, I. H. Agustin, Slamin, R. Adawiyah, E. Y. Kurniawati, *On the study of local antimagic vertex coloring of graphs and their operations*, J. Phys. Conf. Ser. **1836** (2021), 1–8.

[7] N. Hartsfield, G. Ringel, *Pearls in graph theory*, Academic Press, Inc., Boston, 1994.

[8] J. Haslegrave, *Proof of a local antimagic conjecture*, Discrete Math. Theor. Comput. Sci. **20**(1) (2018), 1–14.

[9] G. C. Lau, M. Nalliah, *On local antimagic chromatic number of a corona product graph*, Bull. Inst. Comb. Appl. **98** (2023), 111–121.

[10] G. C. Lau, K. Premalatha, S. Arumugam, W. C. Shiu, *On local antimagic chromatic number of cycle-related join graphs II*, Available from: https://arxiv.org/pdf/2112.04142v1.pdf. (2022, Jun 2).

[11] G. C. Lau, W. C. Shiu, H. K. Ng, *Affirmative Solutions On Local Antimagic Chromatic Number*, Available from: https://arxiv.org/pdf/1805.02886.pdf. (2022, Jun 1).

[12] G. C. Lau, W. C. Shiu, H. K. Ng, *On local antimagic chromatic number of cycle-related join graphs*, Available from: https://arxiv.org/pdf/1805.04888v1.pdf. (2022, Jun 1).

[13] G. C. Lau, W. C. Shiu, H. K. Ng, *On local antimagic chromatic number of graphs with cut-vertices*, Iran. J. Math. Sci. Inform, Available from: https://arxiv.org/pdf/1805.04801v8.pdf. (2022, Jun 2).

[14] G. C. Lau, W. C. Shiu, *On join product and local antimagic chromatic number of regular graphs*, Available from: https://arxiv.org/pdf/2203.06594v1.pdf. (2022, Jun 2).

[15] G. C. Lau, W. C. Shiu, *On local antimagic chromatic number of lexicographic product graphs*, Available from: https://arxiv.org/pdf/2203.16359.pdf. (2022, Jun 1).

[16] R. Shankar, M. Nalliah, *Local vertex antimagic chromatic number of some wheel related graphs*, Proyecciones **41**(1) (2022), 319–334.

[17] J. Sedláček, *Problem 27, in Theory of Graphs and its Applications*, Proc. Symposium Smolenice (2022), 163–167.

[18] T. M. Wang, *Toroidal Grids Are Anti-magic*, Computing and Combinatorics **3595** (2005), 671–679.

[19] T. M. Wang, C. C. Hsiao, *On anti-magic labeling for graph products*, Discrete Math. **308** (2008), 3624–3633.

[20] D. B. West, *Introduction to graph theory*, Pearson Education, Inc., India, 2001.

[21] X. Yang, H. Bian, H. Yu, D. Liu, *The Local Antimagic Chromatic Numbers of Some Join Graphs*, Math. Comput. Appl. **26**(80) (2021), 1–13.

6. Data Science and Computer Science

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Graph Convolutional Network for Multiple Traveling Salesman Problem

Chanoknun Phunnasorn[1,†], Wasakorn Laesanklang[1], and Tipaluck Krityakierne[1,‡]

[1]Department of Mathematics, Faculty of Science
Mahidol University, Bangkok 10400, Thailand

Abstract

This study investigates the application of Graph Convolutional Network (GCN) coupled with beam search to solve the Multiple Traveling Salesman Problem (MTSP). The GCN is trained to model and understand the structure of the problem, including features of various locations, their interconnections, and the number of salesmen. Subsequently, beam search is employed to extract the final optimal routes. Our findings confirm the applicability of this approach, yielding solutions with small optimality gaps and highlighting its efficiency in addressing the complexities of the MTSP.

Keywords: graph convolutional network, multiple traveling salesmen, beam search.

2020 MSC: Primary 68T07; Secondary 90B06, 90-08.

1 Introduction

The traveling salesman problem (TSP) involves finding the shortest route to visit all assigned nodes. It is a well-known problem within the realms of computer science and operations research, with practical applications in logistics and delivery businesses. Solving TSP has the potential to address real-life business operations, including shipping time, cost, and scheduling. The Multiple Traveling Salesman Problem (MTSP) is an extension of the single salesman problem, where multiple routes corresponding to the number of salesmen are built, with the condition that each node must be visited by one salesman, and the total distance of all routes must be minimized.

TSP and MTSP are typically solved by heuristic algorithms, such as genetic algorithms [1], tabu search [5], ant colony optimization [13], as well as LP solvers like Concorde [4]. Moreover, machine learning techniques have also been employed. Examples include accelerated augmented

*This research was partially supported by Office of the Permanent Secretary, Ministry of Higher Education, Science, Research and Innovation, Thailand, through the Grant No. RGNS 64-151.

†Speaker. ‡Corresponding author.
Email: chanoknun.phn@student.mahidol.edu (C. Phunnasorn), wasakorn.lae@mahidol.ac.th (W. Laesanklang), tipaluck.kri@mahidol.edu (T. Krityakierne).

Lagrangian Hopfield neural network algorithms [6] and decentralized attention-based neural networks [2].

Since TSP problems are represented in the form of graphs with connected paths between nodes, Graph Convolutional Network (GCN) can be naturally employed to analyze the connections between data nodes in the problem. GCN facilitates the sharing of information among nodes, revealing relationships that influence the determination of the optimal route by progressively combining node information from nearby locations. The idea of using a GCN has been explored in [12] to extract features from the graph. In this approach, each node in the graph is represented by a feature vector, and it merges information from its neighboring nodes. Graph Neural Network (GNN) local search for the traveling salesman problem, as considered by [7], is a hybrid data-driven approach for solving the TSP based on graph neural networks and guided local search. The work of [11] reviews Hopfield neural networks, graph neural networks, and neural networks with reinforcement learning for solving the TSP. The work of [8] uses a graph as input and outputs probabilities of edges occurring on a TSP tour from the GCN model. Deep policy dynamic programming for vehicle routing problems, proposed by [9], combines the strengths of learned neural heuristics with dynamic programming algorithms.

Objective: The objective of this research is to investigate the applicability of GCN in solving MTSP. Specifically, in the numerical results, we train the model for 20-node MTSP instances with 1, 2, or 3 salesmen using datasets containing coordinates and their corresponding tour solutions. The results are subsequently evaluated by comparing the efficiency of decoding methods used to extract the final optimal MTSP routes.

2 Methodology

The technique used in this study was adapted from that of [8], where the GCN was initially employed to solve a single-TSP problem. Building upon this foundation, we have tailored the methodology to tackle the MTSP.

An overview of the main procedure is now given.

Algorithm 1 Procedure for solving MTSP problems with n nodes and m salesmen

Input: Training and validation datasets (Section 2.1)

1: For each input data, duplicate $m-1$ number of artificial origin nodes $n+1,\ldots,n+m-1$, all having the same coordinate locations as that of the origin node 1.
2: Apply GCN for TSP (Section 2.2).
3: Decode a probabilistic heat-map into an optimal route (Section 2.3).
4: Convert an optimal TSP route back into optimal MTSP tours by reverting the procedure in Step 1.

The procedure begins by transforming the MTSP input data into a single-TSP, achieved by adding $m-1$ artificial origin nodes. For example, in Figure 1, the original 2-TSP is transformed into a single-TSP by the addition of node 11.

Next, the GCN is applied to obtain outputs in the form of a prediction heat map, representing the adjacency matrix of edges (with details provided in Section 2.2). Subsequently, the prediction heat map serves as input for the next step, which determines an optimal route (as discussed in Section 2.3). Finally, after the optimal route solution has been determined, we convert a single-TSP tour back into optimal tours for MTSP by collapsing all artificial nodes back to the origin node 1.

We will now provide the necessary details and explanations for each step of the procedure starting with the dataset structure and generation.

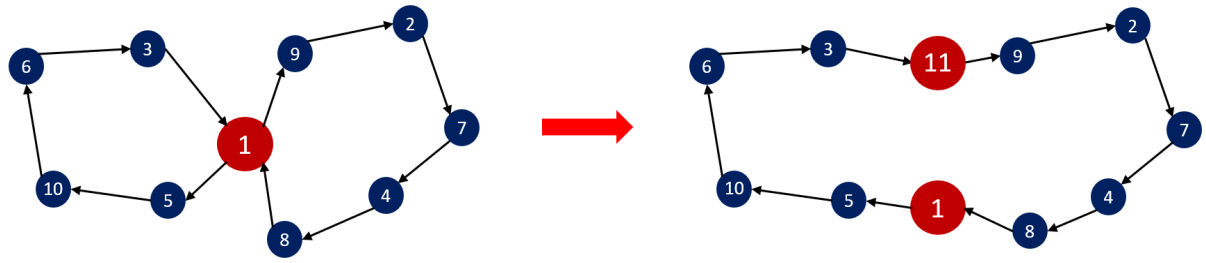


Figure 1: The 2-TSP transformed into a single-TSP by adding an artificial node 11

2.1 MTSP Dataset Structure and Generation

2.1.1 Dataset Structure

The input data is a collection of two-dimensional coordinates (x, y) based on the number of nodes of interest, followed by the optimal tours, where node 1 is always the origin node. For example, for a problem with 10 nodes and 2 salesmen, the input data will take the form:

(0.5222797461609513,0.8095372968646599),
(0.117167965988632, 0.7118656499868851),
(0.5755570629747176, 0.4779300930503926),
(0.17851080649823037, 0.4215499396635761),
(0.6930025843040617, 0.6361822518619119),
(0.8158842838579384, 0.24317747043216498),
(0.0018452807967549445, 0.5670706020799816),
(0.3103900468905868, 0.6491108438482176),
(0.10556622600366072, 0.8425337796202614),
(0.8159933439685663, 0.38616766472072805)
salesman 1: 1 5 10 6 3 1;
salesman 2: 1 9 2 7 4 8 1

2.1.2 Dataset Generation via Linear Programming for MTSP

The datasets for MTSP for training and testing consist of coordinates (x, y), and the corresponding optimal tour solution, which is obtained through the Miller-Tucker-Zemlin formulation. The objective and constraint equations are as follows:

$$\min \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} c_{ij} x_{ij} \tag{2.1}$$

$$\sum_{i=2}^{n} x_{i,1} = m \tag{2.2}$$

$$\sum_{j=2}^{n} x_{1,j} = m \tag{2.3}$$

$$\sum_{i=2, i\neq j}^{n} x_{ij} = 1, \quad \forall j = 1, ..., n; \tag{2.4}$$

$$\sum_{j=2, i\neq j}^{n} x_{ij} = 1, \quad \forall i = 1, ..., n; \tag{2.5}$$

$$u_i - u_j + 1 \leq (n-1)(1 - x_{ij}), \quad \forall i, j,\ 2 \leq i \neq j \leq n; \tag{2.6}$$

$$0 \leq u_i \leq n, \quad \forall i = 2, \ldots, n; \tag{2.7}$$

$$x_{ij} \in \{0, 1\}, \quad \forall i, j = 1, \ldots, n; \tag{2.8}$$

$$u_i \in Z, \quad \forall i = 2, \ldots, n, \tag{2.9}$$

where the node indices are represented by numbers $1, \ldots, n$. The variable x_{ij} equals 1 when the path goes from node i to node j and 0 otherwise. Here, m is the number of salesmen, u_i is an auxiliary variable, c_{ij} represents the distance from node i to node j.

The objective function (2.1) aims to minimize the distance of each salesman's tour. The first two constraints, (2.2) and (2.3), ensure that there are m salesmen who return to and depart from the origin node. Following are the next two constraints, (2.4) and (2.5), which guarantee that each node is reached from exactly one other node and that from each node there is an exit to exactly one other node. The constraints on the auxiliary variables (2.6) ensure that no salesman passes through the same node twice.

2.2 Application of GCN for TSP

Graphs consist of nodes connected by edges, representing relationships and interactions between entities. To apply GNN, the input graph needs to be transformed into a matrix (adjacency matrix or node attributes matrix), filled with 0s and 1s to denote connections between nodes. GNN then learns embeddings or representations for each node in the graph. These representations encode information about the node's features and its relationships with neighboring nodes. Subsequently, GNN uses information from a node's neighbors to update its own information as shown in Figure 2. This updating of node information is commonly referred to as message passing as shown in Figure 3.

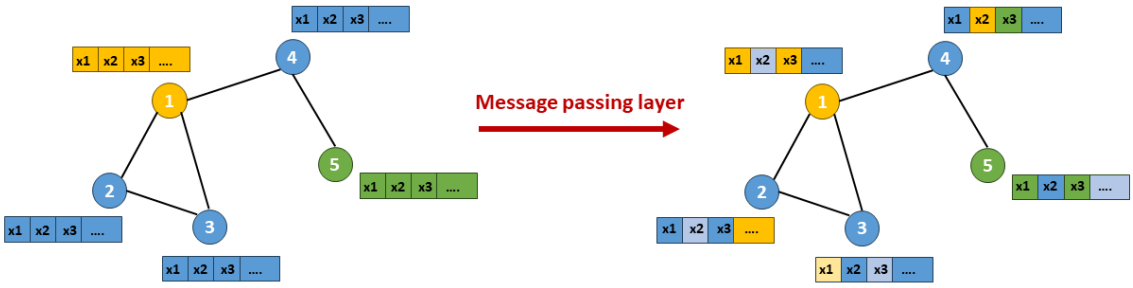


Figure 2: After the message passing layer, each node in the graph possesses information about its neighbors

GCN is a specific type of GNN that employs convolution-like operations on graphs. It adapts the concept of convolution from image processing to capture local structures within graphs. Features are learned by examining neighboring nodes. However, GCN differs from Convolutional Neural Networks (CNNs) in that CNNs are designed to operate on data with regular structures, while GCN is a generalized version of CNN where the number of node connections varies, and the nodes are unordered. See Figure 4.

In this work, we apply GCN to learn and solve MTSP in a similar manner to that of [8] for TSP. The details are as follow:

1. **Input layer**
   For the input node feature, we are given the two-dimensional coordinates $x_i \in [0, 1]^2$ which are embedded into h-dimensional features. Here, h depends on the batch size, node, and

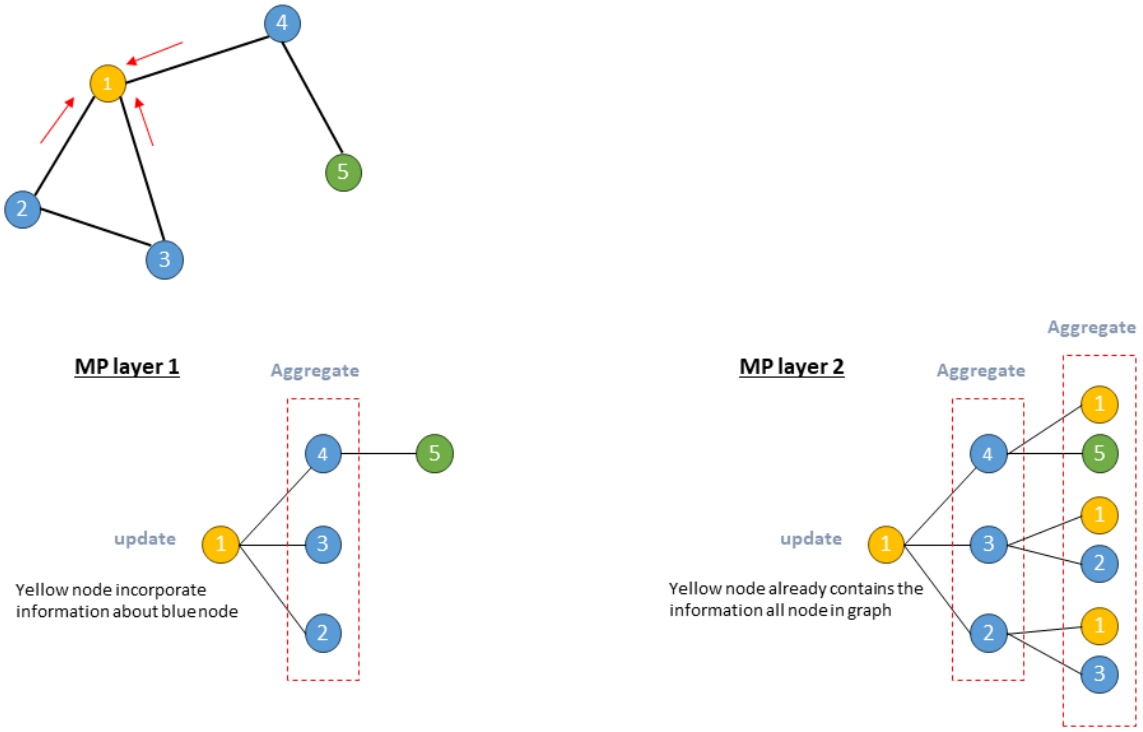


Figure 3: Example illustrating how each message passing layer operates when concentrating on updating information in the yellow node

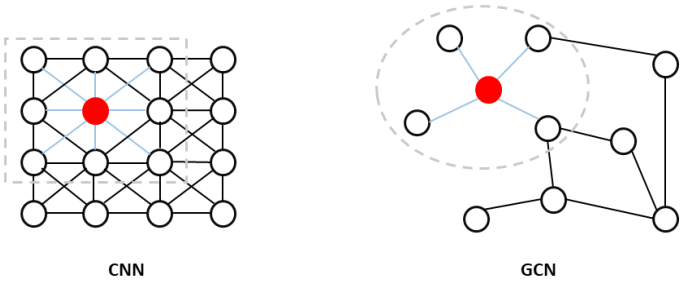


Figure 4: Structuring data for CNN and GCN

hidden dimension. The edge distance between nodes i and j is represented as a $h/2$-dimensional feature vector. An indicator function of a TSP edge is denoted by δ_{ij}^{k-NN}, with a value of one if node j is one of the k-nearest neighbors of node i, a value of two for self-connections, and a value zero otherwise.

2. **Graph convolution layer**
The graph convolution layer is used to calculate representations for nodes and edges within a graph. Both node and edge information are used to compute these representations, and multiple layers of graph convolution are applied to iteratively extract increasingly complex features from the input graph. More precisely, let x_i^l and e_i^l denote the node feature vector and edge feature vector at layer l associated with node i and edge ij, respectively. The notation $j \sim i$ denotes the set of neighboring node centered at node i. At the input layer, $x_i^{l=0} = \alpha_i$ and $e_{ij}^{l=0} = \beta_{ij}$. We define the node feature and edge feature at the next layer as:

$$x_i^{l+1} = x_i^l + ReLU(BN(W_1^l x_i^l + \sum_{j \sim i} n_{ij}^l W_2^l x_j^l))$$

$$e_{ij}^{l+1} = e_{ij}^{l} + ReLU(BN(W_3^l e_{ij}^l + W_4^l x_i^l + W_5^l x_i^l)),$$

with

$$n_{ij}^{l} = \frac{\sigma(e_{ij}^l)}{\sum_{j' \sim i} \sigma(e_{ij}^l) + \varepsilon},$$

where $W_i \in \mathbb{R}^{h \times h}$ is the matrix of size $h \times h$ in PyTorch's "nn.Linear" modules [10]. These weights are automatically initialized by PyTorch and are updated during the training of a neural network. Here, σ represents the sigmoid function, ε denotes a small value, $ReLU$ is the rectified linear unit, and BN refers to batch normalization. In this work, batch normalization normalizes node and edge features independently, helping stabilize and accelerate the training of neural networks for graph-related tasks. Figure 5 illustrates the graph convolution layer.

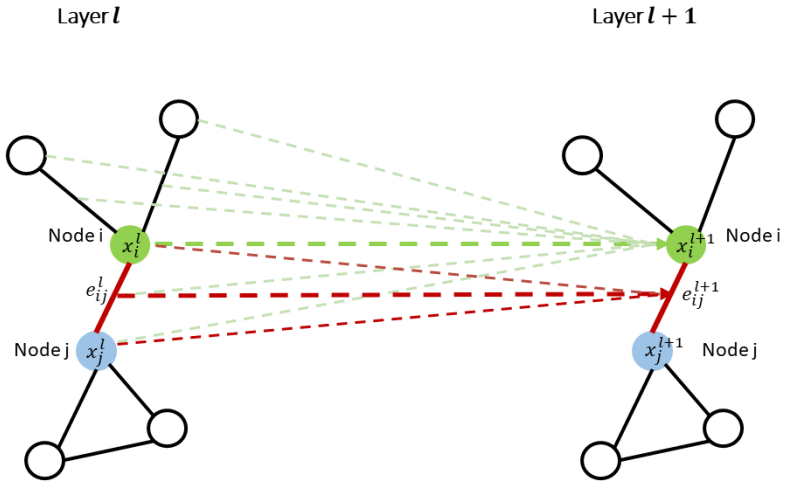


Figure 5: The h-dimensional representations x_i for node i, and e_{ij} for the edge connecting node i and j in the graph, are computed by the graph convolution layer. The information for computing in the next layer is indicated by green and red arrows.

3. **Multi-layer perceptron (MLP) classifier**
Once the GCN layer is completed, the algorithm acquires edges with more complex features (e_{ij}^l). The edge embedding of the last layer is then used to compute the probability of edge connection in the tour of the graph. This probability can be seen as computing a probabilistic heat map over the adjacency matrix of tour connections

$$p_{ij}^{TSP} = MLP(e_{ij}^L),$$

where $p_{ij}^{TSP} \in [0,1]^2$ and L is the layer of the MLP.

We provide an example of the outputs from the MLP layer with 10 nodes: [[[0.3176, 0.3986], [-0.0411, 0.0290], [-0.0586, -0.1276], [0.1145, 0.1033], [-0.0287, 0.1249], [0.6307, 0.8379], [-0.0568, -0.2325], [-0.0545, -0.1104], [-0.0247, -0.1831], [0.0276, 0.4686]], [[-0.0411, 0.0290], [0.3932, 0.4884], [0.1436, 0.2179], [0.0385, 0.1141], [-0.2165, -0.0985], [0.6611, 0.7397], [0.0333, 0.0074], [0.0780, 0.1625], [0.0711, -0.0255], [0.0328, 0.2008]], ...], where each component is a range of values and can be transformed into a prediction heat map as shown in Figure 6.

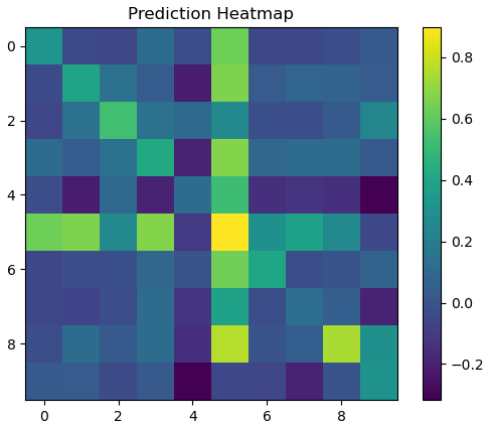


Figure 6: Heatmap illustrating the predicted output from an example within the MLP layer

4. **Loss function**
   Each element of the adjacency matrix, transformed from the ground-truth TSP tour, indicates whether there is an edge between nodes i and j in the TSP tour. To optimize this process, a weighted binary cross-entropy loss is minimized over mini-batches. As the size of the problem grows, there is an issue of class imbalance in the classification task, where the negative class dominates. To address this imbalance, appropriate class weights are needed to balance the effect. The balance class weight of classes 0 and 1 are computed by [8]

$$w_0 = \frac{n^2}{(n^2 - 2n) * c}$$

and

$$w_1 = \frac{n^2}{(2n) * c},$$

where $c = 2$ denotes the number of classes (0 and 1), and n is the number of nodes in each instance.

The input for computing the loss value includes predictions for edges, targets for edges, and weights for edge loss. The algorithm utilizes the edge predictions with "F.log_softmax" (a function call from the PyTorch library) and calculates the logarithm of the softmax function applied element-wise to the input to obtain the log probability tensor as input for the loss function. Next, it computes negative log-likelihood loss using the "nn.NLLLoss" function. This loss is employed to measure the dissimilarity between the predicted probability distribution for edges and the actual class labels, while also considering the specified class weights. This loss is computed between the log probabilities and the target values. The class weights are used during the loss calculation. PyTorch's negative log-likelihood loss, "nn.NLLLoss" is defined as [3]:

$$l(x, y) = L = \sum_{n=1}^{N} \frac{l_n}{\sum_{n=1}^{N} w_{y_n}},$$

$$l_n = -w_{y_n} x_{n,y_n},$$

where x computes the logarithm of the softmax function along the edge predictions, y is the target for edges, w is the class weight, and N is the batch size.

2.3 Decoding Optimal Routes

The output of the previous step is a probabilistic heat map over the adjacency matrix of tour connections. To extract this probabilistic edge heat map into a valid permutation of nodes, opti-

mization search strategies are required. Although alternative search strategies can be employed for this task, in our numerical experiment, we will primarily use beam search. Therefore, we will now provide a brief refresher on beam search.

Beam search: Beam search is a technique used in natural language processing and generative models to find a set of high-probability sequences. It involves exploring the possibilities by expanding the most likely connections among neighboring nodes, starting with the first node and progressively expanding the top candidates at each stage. The process continues until all nodes have been visited, and a valid tour is constructed using a masking strategy. The mask tensor is updated to exclude nodes that have already been added to the beam during the search. The final prediction is the tour with the highest probability among a specified number of complete tours, known as the beam width.

In this work, beam search is used to extract the final optimal TSP tour based on a probabilistic heat map. The steps are as follows:

1. Calculate probabilities: Probabilities for edge prediction are calculated using either softmax or log softmax.

2. Initialize Beam search: An instance of the beam search is created with specified parameters such as beam size, batch size, number of nodes, data types, probability type, etc.

3. Perform beam search: A loop is employed to advance the beam search step by step. At each step, probabilities are calculated, and the beam is advanced accordingly.

4. Extract the TSP tour: After the beam search is complete, the TSP tour with the highest probability among the beam candidates is obtained.

Heuristic beam search: As in [8], another variant of beam search will also be implemented for comparison. In this variant, instead of selecting the tour with the highest probability at the end of beam search, we select the shortest tour among the set of b complete tours as the final solution. We shall refer to this variant as the heuristic beam search.

3 Numerical Experiments

3.1 Experimental Setup

The hyperparameters and setting used for training are referenced from [8]. In particular, each model consists of $l_{conv} = 30$ graph convolutional layers and $l_{mlp} = 3$ layers in the MLP, with hidden dimension $h = 300$ for each layer. We use a fixed beam width $b = 1{,}280$ and fix $k = 20$ nearest neighbors for each node in the adjacency matrix.

For each training epoch, a random 10,000 problem instances are selected from a 100,000 training set. The Adam optimizer is used to minimize the cross-entropy loss for each mini-batch.

The model's performance is evaluated on a separate validation set consisting of 10,000 instances. If the validation loss does not decrease by at least 1% compared to the previous validation loss, the learning rate of the optimizer is reduced. This small learning rate decay strategy helps the models learn more efficiently and converge to better local minima during training.

3.2 Performance Metrics

We use the following two metrics for comparing algorithm performance.

1. Average tour length: The average predicted MTSP tour length is computed by

$$\frac{1}{T}\sum_{i=1}^{T} l_i,$$

where T is the number of test instances and l_i is the predicted MTSP tour length.

2. Optimality gap: The average percentage ratio of the predicted tour length relative to the optimal solution is computed as

$$\frac{1}{T}\sum_{i=1}^{T}\left(\frac{l_i}{l_i^*}-1\right),$$

where l_i^* is the optimal MTSP tour length.

3.3 Numerical Results

Three variants of the algorithm are considered, depending on the decoding method used to extract the final optimal TSP routes: **I. greedy search (GS)**, **II. beam search (BS)**, and **III. heuristic beam search (BS*)**. The explanation of BS and BS* can be found in Section 2.3. Regarding GS, the method begins at the first node and greedily selects the next node from its neighbors based on the highest probability of an edge's presence. The search terminates once all nodes have been visited.

Due to the three-week timeframe required to generate MTSP training dataset containing optimal tours of 100,000 problem instances (on a computer with an Intel Core i7 @5.4GHz CPU), we were only able to demonstrate the method for problems involving 20 nodes and 1, 2, or 3 salesmen. This process alone consumed over 1.5 months solely for data generation.

The exact solution, serving as a baseline, was obtained from GUROBI 10.0.3. Table 1 presents the average tour length over 10,000 instances (Avg. Len.), optimality gap (Opt. gap.), and total computation time (Time) taken for 10,000 test instances.

Table 1: Numerical results from the GCN model with greedy search (GS), beam search (BS), or heuristic beam search (BS*) and Gurobi solver

#Salesmen	Measurement	Gurobi	GCN+GS	GCN+BS	GCN+BS*
1	Avg. Len.	3.831	3.935	3.854	3.831
	Opt. gap. (%)		2.715	0.600	0
	Time (sec)		402.07	511.87	868.4
2	Avg. Len.	4.009	4.189	4.065	4.011
	Opt. gap. (%)		4.489	1.397	0.049
	Time (sec)		461.69	564.98	1394.86
3	Avg. Len.	4.311	4.545	4.388	4.320
	Opt. gap. (%)		5.428	1.786	0.209
	Time (sec)		496.82	615.17	1946.35

From the table, it is evident that while GCN+GS requires relatively short computational time, it does not provide a satisfactory solution for route optimization. On the other hand, GCN+BS* appears to offer the best solution, albeit at the expense of longer computational time especially for the problem having more than one salesman. GCN+BS represents a middle ground between accuracy and computational efficiency when compared to the other two methods. Figure 7 illustrates the comparison of optimality gap across the three methods with varying numbers of salesmen, thereby confirming the results observed in the table.

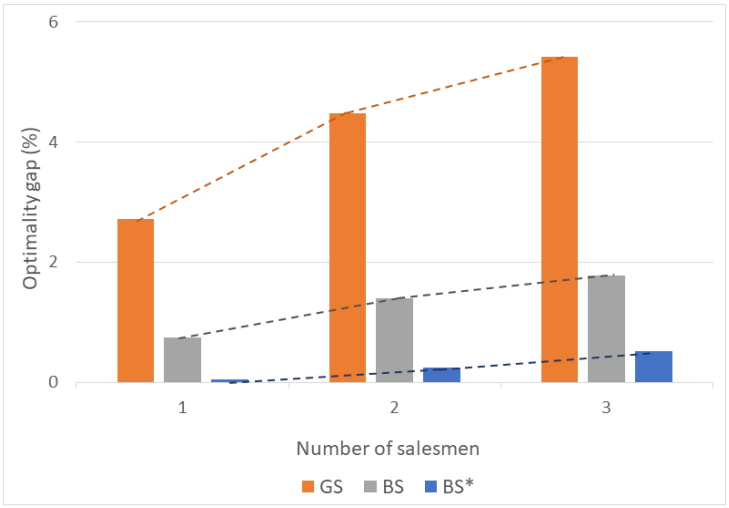


Figure 7: The optimality gap for the 20-node problem instances with 1, 2, and 3 salesmen

4 Conclusions

In this study, we addressed the MTSP using a graph convolutional network. Our numerical results demonstrated the effectiveness of GCN in solving MTSP and emphasized the crucial role of the optimization search method in extracting optimal routes efficiently. Future research directions could include exploring alternative search methods, incorporating training inputs with varying numbers of nodes, and focusing on algorithm enhancements and scalability testing using larger datasets or real-world MTSP scenarios.

References

[1] Maha Ata Al-Furhud and Zakir Hussain Ahmed, *Genetic algorithms for the multiple travelling salesman problem*, International Journal of Advanced Computer Science and Applications, **11** (2020), no. 7, 553–560.

[2] Yuhong Cao, Zhanhong Sun, and Guillaume Sartoretti, *Dan: Decentralized attention-based neural network to solve the minmax multiple traveling salesman problem*, arXiv preprint arXiv:2109.04205 (2021).

[3] PyTorch Contributors, *NLLLOSS*, https://pytorch.org/docs/stable/generated/torch.nn.NLLLoss.html, 2023.

[4] William J Cook, David L Applegate, Robert E Bixby, and Vasek Chvatal, *The traveling salesman problem: a computational study*, Princeton university press, 2011.

[5] C-N Fiechter, *A parallel tabu search algorithm for large traveling salesman problems*, Discrete Applied Mathematics, **51** (1994), no. 3, 243–267.

[6] Yun Hu and Qianqian Duan, *Solving the tsp by the AALHNN algorithm*, Mathematical Biosciences and Engineering, **19** (2022), 3427–3488.

[7] Benjamin Hudson, Qingbiao Li, Matthew Malencia, and Amanda Prorok, *Graph neural network guided local search for the traveling salesperson problem*, arXiv preprint arXiv:2110.05291 (2021).

[8] Chaitanya K Joshi, Thomas Laurent, and Xavier Bresson, *An efficient graph convolutional network technique for the travelling salesman problem*, arXiv preprint arXiv:1906.01227 (2019).

[9] Wouter Kool, Herke van Hoof, Joaquim Gromicho, and Max Welling, *Deep policy dynamic programming for vehicle routing problems*, International conference on integration of constraint programming, artificial intelligence, and operations research, Springer, June 20–23, 2022, pp. 190–213.

[10] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala, *Pytorch: An imperative style, high-performance deep learning library*, Advances in Neural Information Processing Systems 32 (H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, eds.), Curran Associates, Inc., 2019, pp. 8024–8035.

[11] Yong Shi and Yuanying Zhang, *The neural network methods for solving traveling salesman problem*, Procedia Computer Science, **199** (2022), 681–686.

[12] Zhihao Xing, Shikui Tu, and Lei Xu, *Solve traveling salesman problem by monte carlo tree search and deep neural network*, arXiv preprint arXiv:2005.06879 (2020).

[13] Jinhui Yang, Xiaohu Shi, Maurizio Marchese, and Yanchun Liang, *An ant colony optimization method for generalized tsp problem*, Progress in Natural Science, **18** (2008), no. 11, 1417–1422.

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Artificial Intelligence for Forecasting Rice Yields in Thailand

Thoedsak Saengthong[1,†], Thanathat Khottiam[1], Chakhrit Utamapokai[1], and Wanyok Atisattapong[1,‡]

[1]Department of Mathematics and Statistics, Faculty of Science and Technology
Thammasat University, Pathum Thani 12120, Thailand

Abstract

The use of artificial intelligence in developing a rice production forecasting model for Thailand was investigated in this work. The planting area, rice varieties, irrigation area, harvesting area, amount of fertilizer applied, selling price, average rainfall, temperature, and humidity were all taken into consideration during the cultivation process. The rice yields were estimated using the following four models: Artificial Neural Network (ANN), Decision Tree Regressor (DTR), Extreme Gradient Boosting (XGBoost), and Multiple Linear Regression (MLR). The results indicate that XGBoost performed better than the other three models in terms of prediction accuracy. Therefore, this technique was used to predict Thailand's rice production. In addition, we separated the anticipated scenario for the years 2023–2025 into three categories: typical occurrences, flood situation, and drought situation.

Keywords: rice yield prediction, artificial intelligence, multiple linear regression, decision tree regressor, XGB regressor, artificial neural network.

2020 MSC: Primary 68T01; Secondary 68T20.

1 Introduction

Nowadays, the agriculture sector accounts for the majority of Thailand's economy. The National Statistical Office [7] reports that as of the fourth quarter of 2022, 12.22% million Thais, or 17.47% of the labor force, were employed in agriculture. This demonstrates the importance of employment and household income to the country. According to projections, between 2040 and 2049, greenhouse gas emissions will cost Thailand's agribusiness between $24 billion and $94 billion [1]. Climate change is making agricultural output more unpredictable, thus the agriculture sector is depending more and more on yield forecasts. If the forecast turns out to be accurate, agencies will have the knowledge required to create appropriate policies that will assist farmers in better planning their agricultural operations and preparing for any potential

†Speaker: Thoedsak saengthong. ‡Corresponding author: Wanyok Atisattapong.
Email: thoedsak.sae@dome.tu.ac.th (T. saengthong), thanathat9394@gmail.com (T. Khottiam), chakildball@gmailcom (C. utamapokai), wanyok@mathstat.sci.tu.ac.th (W. Atisattapong)

changes in the scheduling or management of different resources, including labor, capital, water, and land.

Agricultural yield forecasting develops a model to predict future crop production based on historical crop yield data and independent variables impacting agricultural yield. Crop yield forecasting creates an agricultural yield projection at the end of the growing season by importing all data once, which is static data [8]. Nevertheless, this kind of forecasting ignores data collected during the agricultural season such as temperature, precipitation, humidity. As a result, the Agricultural Information Center [11] forecasts yields divided into four quarter, with the first quarter occurring in March, the second in June, the third in September, and the fourth in December. Each forecast is followed by an adjustment to improve the accuracy of the subsequent forecast based on the outcomes. At every stage of agricultural production, the forecasting model incorporates data that influences and correlates many aspects of agricultural use.

In this study, we create models utilizing four different techniques: (1) multiple linear regression (MLR), (2) decision tree regression (DTR), (3) extreme gradient boosting (XGBoost), and (4) artificial neural network (ANN) to forecast Thailand's rice yield in the future. When predicting rice yield, three phases will be considered: preparation, planting, and harvesting the crop. It is expected that the forecast's outcomes would facilitate goal-setting and the drafting of appropriate policies by public and private organizations involved with Thailand's agricultural sector.

The rest of the paper is organized as follows. In Section 2, the related works are discussed. In Section 3, the models are proposed. The results and simulation are reported in Section 4, and our conclusions are discussed in Section 5.

2 Literature Review

Byoung-Hoon Lee, et al. [2] proposed regression models to forecast county wheat yield and wheat quality using meteorological data. Precipitation and temperature are included in the models as explanatory factors for the various stages of wheat development. In addition, the models include a spatial lag effect, crop year random effects, and county fixed effects. Weather factors have a significant impact on both yield and quality level; precipitation and yield have a positive, nonlinear relationship, whereas average monthly temperatures have a negative link. The forecasting ability of the models is enhanced by adding the spatial lag effect, and out-of-sample tests confirm the usefulness of the models in predicting wheat yield and quality.

P. S. Maya Gopal, et al. [9] proposed a hybrid MLR-ANN model for crop yield prediction in agriculture. The MLR intercept and coefficients are utilized to initialize the input layer bias and weights of the ANN. Based on performance criteria, the hybrid model outperforms the traditional MLR, ANN, support vector machine (SVR), K-nearest neighbor (KNN), and random forest (RF) models in terms of prediction accuracy. Crop yields are predicted by Pallavi Shankarrao Mahore, et al. [10] using a variety of machine learning algorithms, including RF, SVM, and KNN. These methods provide better performance outcomes for specific meteorological conditions, and the suggested system uses data mining techniques to process all the data and estimate harvest output. This can assist farmers in making well-informed decisions about which crops to plant at different times of the year to maximize profits.

A greenhouse drip-irrigated tomato crop evapotranspiration (ET) prediction model (XGBR-ET) based on XGBoost regression was developed by Jiankun Ge, et al. [5] and demonstrated good modeling accuracy for daily ET for greenhouse tomatoes. Additionally, the XGBR-ET model performed better in terms of prediction accuracy when compared to seven other regression models. The effectiveness of XGBR-ET in modeling daily ET for greenhouse tomatoes was proved by its statistical indicators, including mean square error, root mean square error, mean absolute error, mean absolute percentage error, and coefficient of determination.

Ervin Gubin Moung, et al. [3] proposed the XGBoost regression model in 2022 as a helpful tool for forecasting crop yield in Malaysia, and it has a strong R-squared value of 0.98. The most important independent variables in predicting crop output were found to be the quantity of pesticides applied, the average rainfall, and the average temperature. Lastly, to develop a global regression model that can predict crop output across national borders, future research endeavors to integrate yield prediction data from diverse country sources.

The cherry coffee yield forecast for 2022 by Yotsaphat Kittichotsatsawat, et al. [13] includes four stages from plantation to harvest. The area, rainfall, temperature, and relative humidity (RH) datasets are the inputs, and the crop yield of cherry coffee is the output. The productivity of cherry coffee crop yield was estimated using the MLR analysis. With an RMSE of 0.0784 tons and an R^2 value of 0.9235, the MLR model was found to be a good predictor of crop production. The MLR model maintained the linear relationship between crop yield and input factors.

3 Methods

The operational framework of the research methods used in this study as shown in Figure 1.

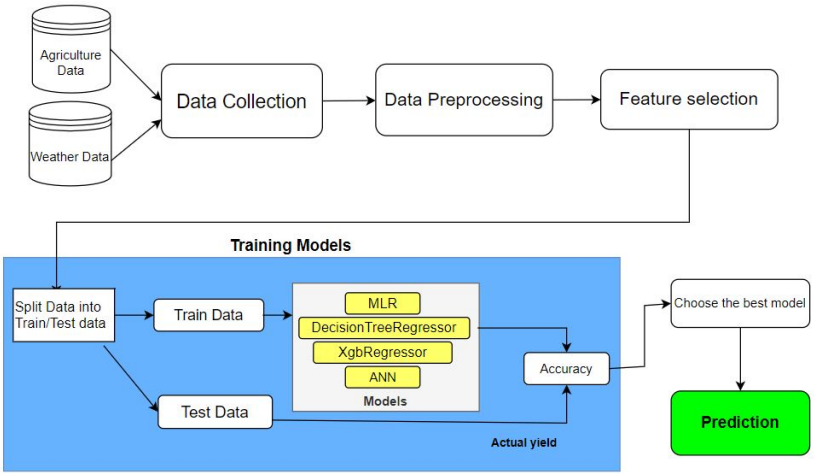


Figure 1: Operational framework

3.1 Data Collection

The primary tasks carried out throughout the cultivation process are associated with the vegetative, reproductive, and ripening phases of rice production [12]. The Ministry of Agriculture's Office of Agricultural Economics of Cooperatives provided the input factors, which included the amount of rice harvested, the area under cultivation, the area both inside and outside the irrigation zone, the amount of fertilizer applied, the type of rice farmed, and the selling price, that affected the value of forecasting rice output at each step. The Meteorological Department and other websites supplied the average temperature, humidity, wind speed, and rainfall, among other meteorological data.

The nine years between 2012 and 2020 were used for collecting all of the data. The twelve variables (eleven inputs and one output) and 693 data points that were obtained through the data collection process are shown in Table 1.

3.2 Data Preparation

The data was enhanced, updated, and values from analytical statistics were calculated. Next, to increase the model's accuracy, prepare the data before training and import it into the analysis

Table 1: Dataset list

Variable	Definition	Vegetative	Reproduction	Ripening
X_1	Plantation Area (Rai)	✓	✓	
X_2**	Rice varieties	✓	✓	✓
X_3	Selling price (Bath)	✓		
X_4	Average temperature (Celsius)	✓	✓	✓
X_5	Average humidity (Percent)	✓	✓	✓
X_6	Average wind force (Knot)	✓	✓	✓
X_7	Average rainfall(milli meters) (Knot)	✓	✓	✓
X_8	Fertilizer quantity (Ton)		✓	
X_9	Irrigated area (Rai*)		✓	
X_{10}	Out Irrigated area (Rai*)		✓	
X_{11}	Harvest area (Rai*)			✓
Y	Rice yield (Ton)	✓	✓	✓

*1 Rai = 1,600 m^2, $X_1, X_2, X_8, X_9, X_{10}$, and X_{11} are retrieved from https://www.oae.go.th/view/1/ /TH-TH, X_3 is from https://www.oae.go.th/view/1/ /TH-TH, X_4, X_5, X_6, and X_7 are obtained from https://en.tutiempo.net/climate/01-2023/ws-484540.html
**X_2 (Rice varieties): KD6 rice, Thai jasmine rice 105, Native rice, Non-photoperiod sensitivity Rice

model. The following is an explanation of the preparation steps.

Table 2: Descriptive statistics of independent and dependent variables

Phases	Variable	Mean	S.D.	Min	Max
Vegetative	X_1	988,357.9	1,053,044.3	63	4,314,831
	X_3	13,525.7	1,941.4	10,189	15,582
	X_4	29.5	1.2	26.5	36.5
	X_5	75.1	4.9	60.5	86.9
	X_6	2.2	1.4	0	7.2
	X_7	186.1	126.3	2.8	816.1
Reproduction	X_1	988,357.9	1,053,044.3	63	4,314,831
	X_4	28.2	3.1	25.8	93.3
	X_5	80.2	3.7	64.3	88.5
	X_6	2	1.3	0	6.2
	X_7	226.4	132.5	3.7	857.1
	X_8	27,523.7	29,897.5	2	112,045
	X_9	201,481.1	199,312	0	1,026,722
	X_{10}	778,916.4	955,957.3	0	3,954,591
Ripening	X_4	26.1	3.1	20.3	77.5
	X_5	70.4	6.4	53.8	87.9
	X_6	2.2	1.5	0	6.8
	X_7	53.2	100.2	0	4,163,693
	X_{11}	919,287.3	964,638.4	63	4,163,693
	Y	398,476.2	381,446.2	31	1,432,101

3.2.1 Data Cleaning

To ensure that the data was as accurate and useful as feasible, it was updated, reviewed, and replaced with inaccurate data. The dataset was also filtered to remove information that was not accurate. The dataset had 12 variables and 472 data points after data cleaning.

3.2.2 Data Transformation

The data used to build the prediction model has a range of values and units, thus to make the data suitable for training the model, normalization and modifying the data units within the same unit are required. The standard scalar approach is used in this study, as shown in Eq. (3.1).

$$X_{i,scale} = \frac{X_i - \mu_{X_i}}{\sigma_{X_i}}, \quad \text{for } i = 1, 2, \ldots, 11, \quad \text{and} \quad Y_{scale} = \frac{Y - \mu_Y}{\sigma_Y}, \tag{3.1}$$

where there are 472 data points for each X_i for $i = 1, 2, \ldots, 11$. Each variable's mean is denoted by μ_{X_i} or μ_y, and its variance is denoted by σ_{X_i} or σ_y. After data transformation, scaled variables have a mean of zero and a variation of one.

3.3 Feature Selection

The process of selecting features from a dataset that will increase the prediction model's performance and accuracy while reducing overfitting is known as feature selection. The goal behind feature selection is to employ the best outcomes as input variables in the prediction model, ranking each aspect based on relevance or the most important relationship. First, a MLR model containing all independent variables will be constructed as in Eq. (3.2).

$$Y = \hat{\beta} X \tag{3.2}$$

where

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}, \text{ and } \hat{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}.$$

Here p is the number of factors ($p = 11$), n is the number of data points ($n = 472$), Y is a vector of rice yields, X is a matrix of input factors with its first column equal to one, and $\hat{\beta}$ is the coefficient vector of input factors.

Then, to find the correlation coefficient vector, Eq. (3.2) can be solved by.

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{3.3}$$

An inverse relationship exists when the correlation coefficient is negative, which indicates that as the value of the input element rises, the rice production will fall. A positive correlation coefficient indicates that an increase in the input factor's value will likewise increase the output value. The variable has little to no relate at all if the correlation coefficient is near zero.

After the model was created, the variables with the highest p-value were removed one at a time using the backward elimination technique. After the feature selection process, the input factors for each planting stage are shown in Table 3.

3.4 Building a Prediction Model

In this work, the Python programming language was used to create prediction models for the total rice production. The training set comprised 80% of the data, while the test set was created using the remaining 20%. There were four models created using supervised learning.

Table 3: Prediction factors after the feature selection process of each phase

Phase	Input factors
Vegetative	X_1, X_2, X_4, X_5, X_6
Reproduction	$X_1, X_2, X_5, X_6, X_7, X_8, X_9, X_{10}$
Ripening	$X_2, X_5, X_6, X_7, X_{11}$

3.4.1 Multiple Linear Regression (MLR)

Multiple regression analysis is used to examine the relationship between several input components $(X_1, X_2, \ldots . X_n)$ and a single predicted value of rice production (Y). The expected outcome depends on the values of the input elements, as shown by the linear relationship between the variables as shown by Eq. (3.2). The following equations demonstrate how we can obtain the predicted value of rice yield for each phase of cultivation by changing the coefficients for each input element in the equation.

Phase	**MLR**
Vegetative	$Y = 0.977X_1 + 0.160X_2 + 0.031X_4 - 0.039X_5 - 0.098X_6$
Reproduction	$Y = -0.365X_1 + 0.045X_2 + 0.036X_5 - 0.052X_6 - 0.022X_7$ $+0.315X_8 + 0.371X_9 + 0.848X_{10}$
Ripening	$Y = 0.148X_2 + 0.040X_5 - 0.104X_6 - 0.041X_7 + 1.011X_{11}$

Note that in the three equations above, the value of β_0 is almost zero. all variables were statistically significant $(p \leq 0.5)$ by OLS.

3.4.2 Decision Tree Regressor (DTR)

The decision tree regressor is one algorithm used in ensemble learning, a technique for building machine learning models. It predicts variable values by constructing decision trees based on the bagging technique, as shown in Figure 2.

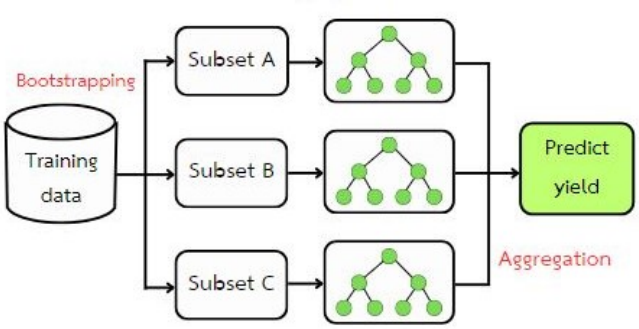


Figure 2: Bagging technique

Several decision tree regressors are built by bagging, and each is trained using a different subset of the training set. Usually, these subsets are sampled with replacement. The predictions of each individual decision tree in the ensemble are averaged once all the decision tree regressors have been trained. Better generalization performance on unobserved data results from this averaging's ability to lower the model's variance and overfitting. In this study, 50 batches of samples are sampled with replacement, yielding 10 subsets. Next, an average method is used to combine the prediction results for each subset.

3.4.3 Extreme Gradient Boosting (XGB Regressor)

One ensemble learning approach that is comparable to the decision tree regressor is the XGB Regressor. The prediction principle, however, is different. Boosting is used by the XGB Regressor, shown in Figure 3.

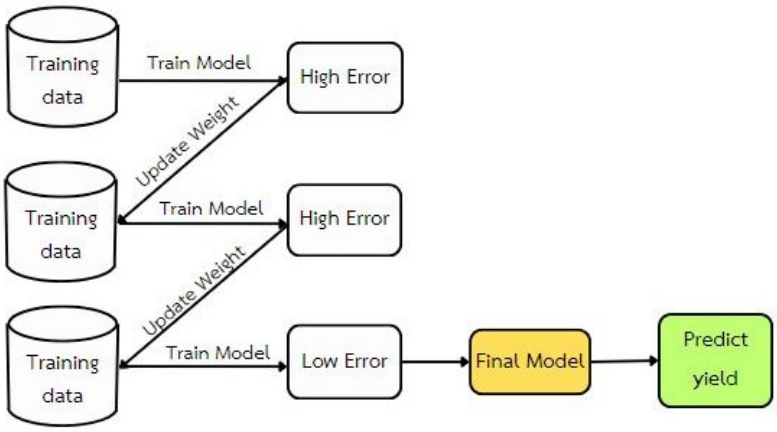


Figure 3: Boosting technique

In this study, a weighted starting dataset with a value of one, a learning rate of 0.1, and a total of 100 trials per run are used. The weights are then modified after the program forecasts the amount of rice yield. To increase the efficiency of the model, this modification is made by increasing the error value for that iteration by the learning rate. This process is repeated recursively for 100 cycles.

3.4.4 Artificial Neural Network (ANN)

Artificial Neural Network (ANN) or artificial neural network is a mathematical model that mimics the workings of the human neural network. It can solve complex problems and identify relationships between data by adjusting weight values (Weight) and bias values (Bias) in the learning process. The structure of ANN has three main parts: the input layer, the hidden layer, and the output layer. Each layer is composed of a different number of nodes. The ANN workflow is shown in Figure 4.

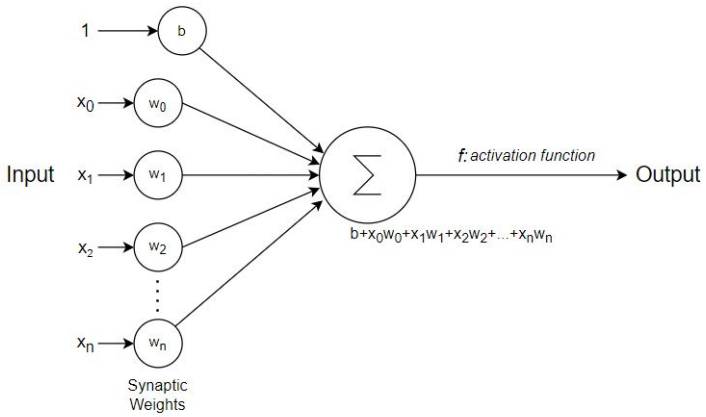


Figure 4: Working structure of ANN [6]

The input data in the first layer is composed of factors that affect the estimated value. The

data was then sent to the activation function of the hidden layer for further processing after being multiplied by the weight. The Sigmoid function, which adjusts a variable's value to fall between 0 and 1, is used in this study, as shown by Eq. (3.4).

$$f(x) = \frac{1}{1 + e^{-x}} \tag{3.4}$$

where $f(x)$ is the activation function and x is the input data from Table 1. Eq. (3.5) states that the expected value is present in the output layer.

$$y = f\left(\sum_{i=1}^{n} x_i w_i\right) \tag{3.5}$$

where y is the rice yield prediction, x_i is the input data and w_i is the weight of i^{th} factor

The number of factors in each period will determine how many nodes are in the input layer of the ANN model. There will be five nodes during the vegetative phase, eight nodes during the reproductive phase, and five nodes during the ripening phase. Next, it is calculated that there are seven hidden layers for each period, each having the same number of nodes as the square of the period's input data. Finally, there will only be one node in each output layer, as shown in Table 4.

Table 4: Parameter settings for ANN model

Parameter	Vegetative	Reproduction	Ripening
Number of nodes in input layer	5	8	5
Number of hidden layers	7	7	7
Number of nodes in hidden layers	25	64	25
Number of nodes in output layers	1	1	1
Learning rate	0.01	0.01	0.01
Number of running models	10	10	10

3.5 Model Performance

The expected values of rice output for each agricultural period for each of the four models will be compared to identify which model generates the best predictions. Moving forward, the most accurate model will be used. The following three criteria are used in this study to quantify performance.

Table 5: Regression model performance evaluation

Method	Formula
Mean Absolute Error (MAE)	$MAE = \frac{\sum_{i=1}^{n} \lvert y_i - \hat{y}_i \rvert}{n}$
Root Mean Square Error:(RMSE)	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$
R-Squared (R^2)	$R^2 = 1 - \frac{SSE}{SST}$

Note that y_i is the raw data of rice yield, $\hat{y}_i$ is the rice yield prediction, and n is the number of data points ($n = 472$). The estimate of rice output has little to no error when the MAE and

RMSE values are minimal, close to zero, or equal to zero. If a prediction model's R^2 value is close to or equal to 1, it is considered suitable for the data.

4 Results

Finding the best-performing model with the highest projected accuracy is the aim of this research. The development outcomes of all four models are therefore compared in Table 6.

Table 6: Compare the efficiency of rice yield prediction models

Cultivation phases	Method	MAE	RMSE	RMSE(ton)	R^2
Vegetative	MLR	0.213	0.298	112,879	0.913
	DecisionTreeRegressor	0.105	0.188	72,181	0.960
	XGBRegressor	0.092	0.158	60,280	0.974
	ANN	0.122	0.166	62,795	0.973
Reproduction	MLR	0.149	0.207	78,769	0.956
	DecisionTreeRegressor	0.095	0.163	61,523	0.974
	XGBRegressor	0.065	0.120	45,523	0.985
	ANN	0.096	0.133	50,280	0.982
Ripening	MLR	0.187	0.247	93,941	0.938
	DecisionTreeRegressor	0.079	0.137	52,045	0.981
	XGBRegressor	0.067	0.110	42,184	0.987
	ANN	0.092	0.132	50,223	0.983

Comparing the XGB Regressor model to the other models, it is evident from Table 6 that it performs the best, providing the most precise forecasts at every stage of cultivation. Its MAE is 0.092, its RMSE is 0.158 (or 60,280 tons when transformed back to actual numbers), and its R^2 is 0.974 at the vegetative stage. Its MAE in the reproduction stage is 0.065, its RMSE is 0.120 (45,523 tons), and its R^2 is 0.985. Finally, it has an R^2 of 0.987, an MAE of 0.067, and an RMSE of 0.110 (or 42,184 tons) during the ripening stage. As a result, the XGB Regressor model was selected by the researchers to forecast Thailand's rice harvest in the future.

The XGB Regressor model, which had the best predictive accuracy during the model-development process, was used to forecast rice yield. The Thailand Water Situation Report's meteorological information and rainfall totals for the year 2022 [4] were then examined to classify the predicted scenarios. As shown in Table 7, the prediction possibilities are separated into three scenarios: normal occurrences, flooding events, and drought events.

Table 7: Thailand's water situation by region

Region\Year	2012	2013	2014	2015	2016	2017	2018	2019	2020
North	normal	normal	drought	drought	normal	flooding	normal	drought	drought
Northeast	drought	normal	normal	drought	normal	flooding	normal	drought	drought
Central	flooding	normal	drought	drought	normal	flooding	normal	drought	drought
South east side	drought	drought	drought	drought	drought	flooding	drought	drought	normal
South west side	flooding	normal	normal	drought	flooding	flooding	drought	drought	normal

The rice yield for the years 2023 to 2025 will be forecasted for each scenario. Table 8 shows the expected rice yield results for each scenario.

It is evident from the results of rice yield prediction that the reproductive period, on average, yields the highest predicted yield, followed by the vegetative phase and the ripening phase. This

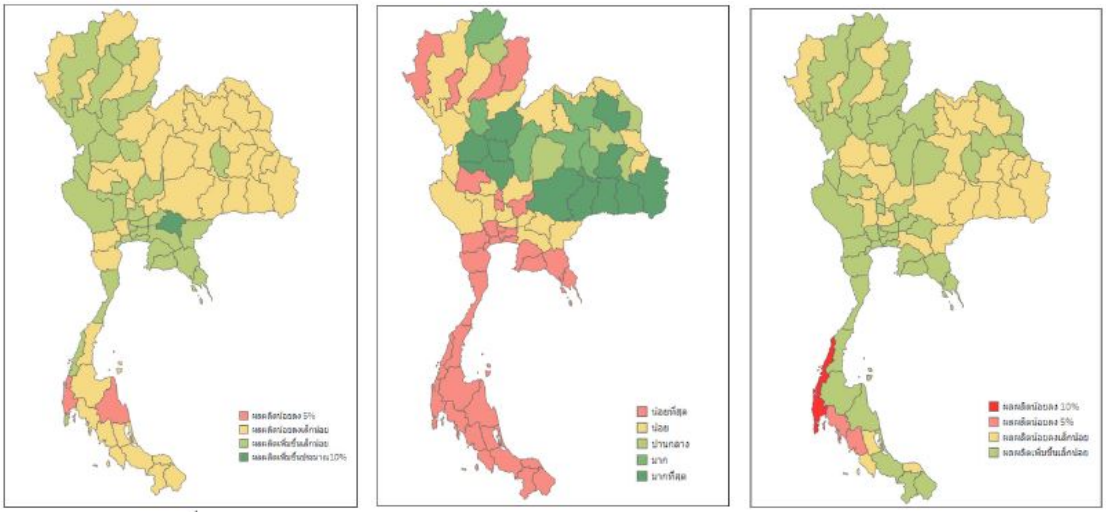
Table 8: Rice yield prediction in each scenario.

Cultivation phases	Scenario/Year	2023	2024	2025
Vegetative	drought	26.292	25.602	25.845
	normal	26.020	25.339	25.339
	flooding	26.053	25.366	25.620
Reproduction	drought	26.119	25.369	25.859
	normal	26.064	25.317	25.814
	flooding	25.952	25.204	25.689
Ripening	drought	24.350	23.739	23.907
	normal	24.300	23.717	23.863
	flooding	24.339	23.749	23.894

Unit:million tons

is due to the fact that the amount of rice planted is initially determined by the area under cultivation. But as farming advances, more variables are taken into account, such the amount of fertilizer used and the area inside and outside of irrigation zones, which raises rice yield. This suggests that these elements are required to increase output. However, the anticipated yield typically begins to progressively decrease during the ripening phase. This is probably because there is a chance of damaging agricultural areas and because the expected rice production does not vary all that much from year to year and situation to situation.

The normal scenario produces the least amount of rice, whereas flooding produces the most, according to predictions made during drought situations. This is due to the fact that weather has minimal influence on rice output forecasts and that parameters pertaining to the cultivated area vary somewhat every year. In general, the year 2023 (BE 2566) is expected to have the highest rice output, followed by the year 2025 (BE 2568), and the year 2024 (BE 2567) is predicted to have the lowest yield, indicating a downward tendency, when taking into account the overall rice yield for the nation in each year.



(a) yield in dought scenario (b) yield in norm scenario (c) yield in flooding scenario

Figure 5: Results of yield prediction according to the cultivation phase in 2025

As can be shown from the prediction results, the province with the greatest expected rice yield in Thailand is Ubon Ratchathani Province. The reason for this is that its cultivated area value is the highest, meaning that it has the biggest influence on the forecast value. The

outcomes of rice yield predictions will be comparable for every year and circumstance. This is due to the fact that the farmed area has barely altered and the prediction is not greatly impacted by weather factors. Taking into account the nation's overall rice production for each year, it was determined that 2023 will yield the most, with 2025 coming in second and 2024 coming in last, with a tendency toward decline.

5 Conclusions

In this study, a model to estimate rice output was developed using input data from agricultural data collected in Thailand between 2012 and 2020. The three stages of the prediction—vegetation, reproduction, and ripening—were selected to align with the main activities associated with rice cultivation. The most effective model, according to the results, was the XGB Regressor model, which produced the most accurate prediction values throughout all cultural phases. Furthermore, this model was the most appropriate for the data in comparison to the other three.

Next, the XGB Regressor model was applied to forecast Thailand's rice output. Three scenarios—a normal scenario, a flood scenario, and a drought scenario—were used to anticipate various outcomes. According to the prediction's findings, Thailand produces the most rice under the drought scenario overall, followed by the normal and flood scenarios. This is due to the fact that it may benefit rice farming during periods of moderate drought. For instance, managing water during a drought requires caution to avoid wasting it, which increases the effectiveness of water use. It may be able to prevent insects and diseases from proliferating across rice fields. The rice forecast's results also show that there is minimal variance in the amount produced. This can be explained by the weak relationship between the estimated value of rice output and the weather component.The amount of production is likewise not significantly different when the weather values in each event are not substantially diverse. We may examine the variables influencing the quantity of rice yield produced during each agricultural cycle by breaking down the prediction into planting periods. This enables us to think about the optimal times to put agricultural development programs into action and the regions that should be developed to maximize productivity.

To enhance the model's prediction ability, researchers can investigate incorporating more important features or modifying the model's parameters to better fit the data.

References

[1] W. Attavanich., *The effect of climate change on Thailand's agriculture.*, Proceedings International Institute of Social and Economic Sciences ,2013 , Prague Czech Republic, September 1–4, pp 23–40.

[2] L. Byoung-Hoon, K. Phil Kenkel, and B. W. Brorsen., *Pre-harvest forecasting of county wheat yield and wheat quality using weather information.*, Agricultural and Forest Meteorology, **168** (2013), pp. 26–35

[3] D. A. -L. Mariadass, E. G. Moung, M. M. Sufian and A. Farzamnia, *Extreme gradient boosting (XGBoost) regressor and shapley additive explanation for crop yield prediction in agriculture.*, Proceedings of the 12th International Conference on Computer and Knowledge Engineering (ICCKE), Mashhad, Iran, November 17-18, 2022, pp. 219–224,

[4] Hydro-informatics institute (Public organization), *Report on the water situation in Thailand A.D.2022*, (2022).

[5] J. Ge, L. Zhao, Z. Yu, H. Liu, L. Zhang, X. Gong, and H. Sun, *Prediction of greenhouse tomato crop evapotranspiration Using XGBoost machine learning model*, Plants, **11** (2022), pp.1-7.

[6] K. Matsumura, C. F. Gaitan, K. Sugimoto, A. J. Cannon, W. W. Hsieh, *Maize yield forecasting by linear regression and artificial neural networks in Jilin, China.*, The Journal of Agricultural Science, **153**(3) (2015), pp.399-410

[7] National Statistical Office Thailand, *Survey of the working conditions of the population*, (2023).

[8] L. Norawat and K. Nantachai, *Agricultural yields forecasting by time series methods*, Thai Industrial Engineering Network Journal, 1(1) (2015), pp 7-13.

[9] P. S. Maya Gopal and R. Bhargavi., *A novel approach for efficient crop yield prediction*, Computers and Electronics in Agriculture **165** 2019, pp. 104968.

[10] P. S. Mahore and A. A. Bardekar, *Crop yield prediction using different machine learning techniques*, Proceedings of the International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 2021, May-June, 2021, pp. 561–569.

[11] P. Seesawang., *Agricultural products with the development of agricultural information work*, Agricultural Information, Center Office of Agricultural Economics (2018).

[12] T. Sriwongchai and S. Rungmekarat, *Rice cultivation*, 4th ed., Department of Agronomy, Faculty of Agriculture, Kasetsart University, Bangkok, 2016.

[13] Y. Kittichotsatsawat, N. Tippayawong and K. Y. Tippayawong, *Prediction of arabica coffee production using artificial neural network and multiple linear regression techniques.* Scientific Reports **12** 2022, pp.1-14.

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Detection of Parvovirus Infection in Shrimps with VGG16

Tharyar Aung[1,†], Pallop Huabsomboon[1,2,‡], Kittisak Chayantrakom[1,2], Somkid Amornsamankul[1,2], and Rapeepun Vanichviriyakit[3,4]

[1]Department of Mathematics, Faculty of Science, Mahidol University, Bangkok, 10400, Thailand

[2]Centre of Excellence in Mathematics, CHE Bangkok 10400, Thailand

[3]Department of Anatomy, Faculty of Science, Mahidol University, Bangkok 10400, Thailand

[4]Centre of Excellence for Shrimp Molecular Biology and Biotechnology (Centex Shrimp), Faculty of Science, Mahidol University, Rama 6 Road, Bangkok 10400, Thailand

Abstract

Given that we all coexist within an ecosystem and depend on one another, it is imperative to prioritize the well-being of all entities rather than solely focusing on human beings. The major aim of this paper is to identify the parvovirus infection in shrimps, a dangerous and harmful infection that specifically targets the hepatopancreas which is the internal organ responsible for the intake and absorption of nutrients, essential for the growth of shrimps. Implementing measures to prevent shrimps from contracting that infection could have both environmental and economic advantages. However, it is a formidable and arduous undertaking to develop a high-quality software or program capable of detecting prawn infections. This research will utilize the VGG16 model, which is well renowned for its exceptional popularity in image classification, to identify parvovirus infection in the hepatopancreas region of a given picture file. The VGG16 model is customized in this study by implementing alterations to its conventional configuration. The near-perfect accuracy rates the model generates at times implies that it is highly convincing in generating prediction results.

Keywords: CNN, deep learning, transfer learning, VGG16, image classification.

1 Introduction

Human beings are an integral element of the Earth's ecosystem, which also includes plants, animals, and several other organisms. All components within the ecosystem are interdependent in order to sustain a healthy ecosystem. It is essential to prioritize the health of all species, not

[†]Speaker. ‡Corresponding author.
E-mail address: tharyar.au1@student.mahidol.edu (T. Aung), pallop.hua@mahidol.ac.th (P. Huabsomboon), kittisak.cha@mahidol.ac.th (K. Chayantrakom), somkid.amo@mahidol.ac.th (S. Amornsamankul), rapeepun.van@mahidol.ac.th (R. Vanichviriyakit)

just humans. This paper aims to identify one very serious and deadly infection that is caused by parvovirus that affects the internal portion of shrimps, specifically the hepatopancreas.

With parvovirus infection, shrimps encounter their growth process without reaching to the actual stage but terminate at a very early phase or even lead to mortality in severe cases. Scientists have been conducting relentless research on image analysis processes to analyze inputted photos and extract valuable information with high precision [1].

Conventional image processing systems require time for feature extraction and matching, which can lead to inefficiency and reduced accuracy due to the complexity of the process and the generated outputs. Researchers and scientists have been striving to achieve improved categorization results by combining deep learning and machine learning techniques [2–4].

Neural networks are widely used in conjunction with image processing systems [5]. One of the most common methods of deep learning is called convolutional neural networks, and it involves the processing of images through a number of layers in order to analyze and extract information, which ultimately results in the production of an output. It can be difficult for traditional image processing systems to produce correct results due to the fact that it takes a significant amount of time to independently perform feature extraction and matching procedures.

In this paper, the system concentrates on gaining access to a convolutional neural network in order to perform an analysis on an image dataset consisting of many image files of hepatopancreatic regions of shrimps. Once this is done, it will yield results that are both important and informative regarding the existence of parvovirus infection in shrimps. It is known that the secretion coming out from the infected ones can really affect the healthy ones as they are sharing the same habitat and resources that exist within it. The sooner the infected ones are known, the better as this can motivate some to remove the infected ones from the habitat shared by other which are non-infected ones.

2 Deep Learning Neural Networks

The structure and operation of the human brain served as the inspiration for the development of deep learning neural networks, which are a subset of machine learning algorithms. Deep learning neural networks are hierarchical, consisting of layers of linked nodes, also known as neurons or units. It is the duty of each layer to digest the information and then transmit it to the subsequent layer for further tasks. The main layers contained in the neural networks are: the input layer, the hidden layers and the output layer. The first layer is responsible for receiving the raw data that are entered. Consequently, in the second place, the hidden layers exist and there are numerous neurons in each hidden layer, and these neurons are responsible for doing calculations on the inputted data. The final layer is responsible for outputting the predictions based on the calculations that were carried out by the layers that came before it.

Even though there are many different kinds of popular neural networks, Convolutional Neural Networks (CNN) are the most astonishingly popular networks especially in terms of dealing with generating predictions based on the input images. Moreover, among all the existing deep learning architectures, convolution neural networks (CNNs) are the ones where they need multi layers fabricate in between the inputs and the outputs where the incoming input layers going forward and passing through many hidden layers located in between to reach the finalized output layers. As its name goes, CNNs are not really simple and easy to understand architectures as the word "convolution-al" included is technically providing how they operate to generate the prediction results in a complex way as the way neural nerves work in our brain consisting of a multitude of layers. However, they are popular especially in the world of scientists as they could always learn the data and their connected patterns efficiently and can output the highly accurate results and outcomes. So, regardless of their complicated working flow, they are still accepted and used in many researches conducted by data scientists. The sample structure of the neural networks is provided below [6]:

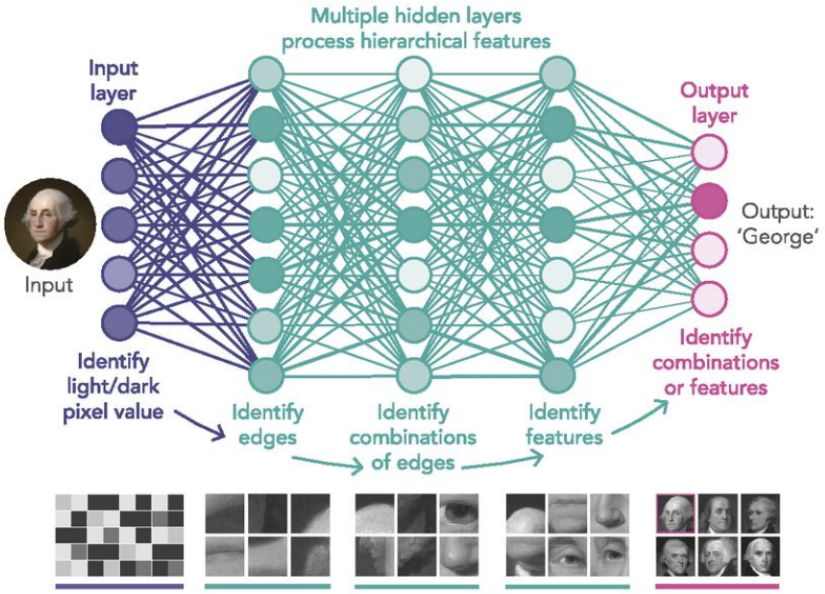


Figure 1: Architecture of neural networks

In CNNs, there are normally two types of propagations [7] including forward propagation and backward propagation. Basically, the former process is to propagate the data from the input layer while the latter process is to carry the data with the reverse order to be able to make necessary adjustments to network fabricated.

2.1 Forward Propagation

In forward propagation, to be able for it to run as smoothly as it can, neuron nodes existing among the hidden layers which reside between the input and output layers play a vital role and their major function is to get an input from its predecessor nodes and operate their tasks with the use of linear function given by

$$y = w_i x + b_i \tag{2.1}$$

where w_i is the weight value coming out from the i^{th} node and b_i is the bias value coming out from the i^{th} node.

After a particular neuron completes its task, it puts its results in an acceptance interval for the upcoming node by using an activation function $f(x)$. This can be illustrated as follows:

$$\theta^{([L]\{i\})} = f(X^{[L]\{i\}}) \tag{2.2}$$

where $\theta^{([L]\{i\})}$ is the output from the activation function of the i^{th} node in the L^{th} layer and $X^{[L]\{i\}}$ is the output from the linear function of the i^{th} node in the L^{th} layer.

2.2 Backward Propagation

Back propagation starts off right at the point where the forward propagation finishes its entire process meaning that the output of the forward propagation is the input of the backward propagation. During the phase of back propagation, normally and in most cases a cross-entropy loss function is used together with the gradient descent algorithm in order to generate accurate results from the neural network. The major purpose of the combination of cross-entropy loss function and gradient descent algorithm is to reduce the percentage of deriving loose results so that the whole model setup can produce better results which could be accessed in many real-world situations and scenarios coming from different environments.

3 Methodology

3.1 Data Set

The data is collected from the Ministry of Fisheries, Thailand. There are 960 images of shrimps in total belonging to 2 classes: HPV (Hepatopancreatic Parvovirus or Penaeus monodon Denso virus) infected shrimps and shrimps with healthy hepatopancreases. The resolution of all the images is fixed to 224×224 for the processing in order not to bump with any sort of turbulations in training and testing processes.

Table 1: The number of images in the two classes according to the shrimp data set

Class	Disease Type	Number of training images	Number of testing images	**Total**
A	HPV	96	384	480
B	Normal	96	384	480
	Total	192	768	960

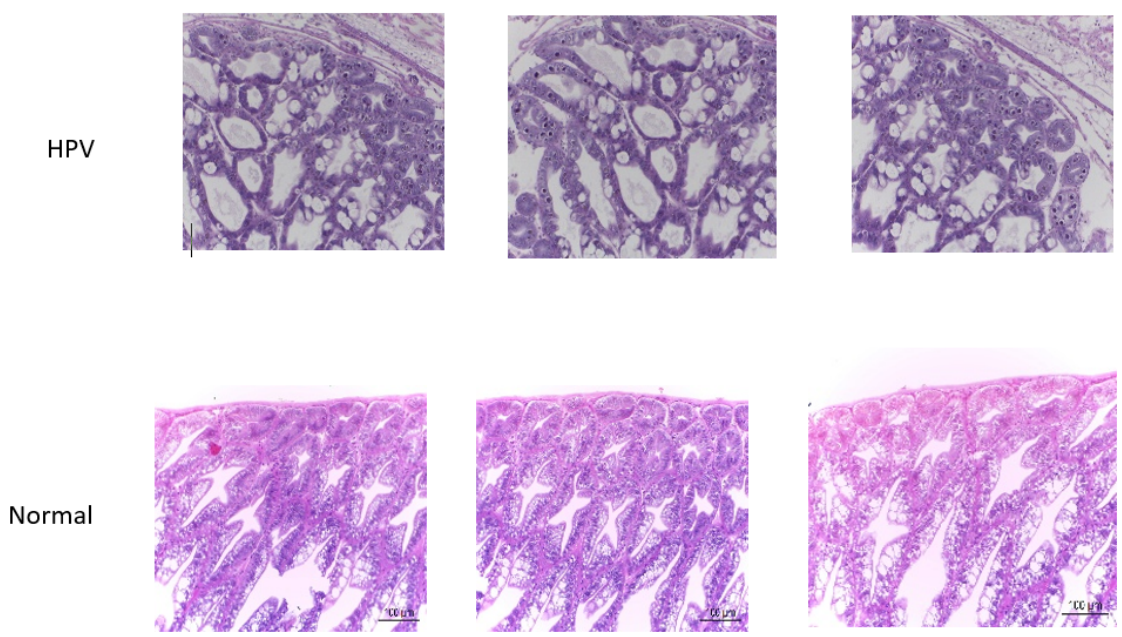


Figure 2: Sample images of hepatopancreatic regions of both cases obtained from the shrimp data set

3.2 VGG16

The VGG16 model is a convolutional neural network design astonishingly popular for its simplicity at certain parts and efficiency in applications related to image classification. Its architecture is composed of 16 layers, predominantly consisting of convolutional layers, followed by max-pooling layers, and building up to fully linked layers.

3.2.1 Architecture of VGG16

The VGG16 model has two main components embedded within its architecture. The first component is a convolutional base which is packed with 13 convolutional layers. Each layer is

closed set up with an activation function which is famously known as a Rectified Linear Unit (ReLU) function. Additionally, there are some max-pooling layers inserted in between those layers which perform to combine all the results coming out from activation layer into a single output. Within the convolution layers of traditional and classical VGG16 architecture setup, the model uses small (3×3) receptive fields with 1 stride and equal padding in order to preserve the spatial resolution of the input. The max-pooling layers, with a size of 2×2 and a stride of 2, decrease the size of the spatial dimensions while simultaneously increasing the number of feature maps making required connections for the model to understand the distinctive features of classes. The second component consists of 3 dense layers which are located at the end of the model along the way of process. Their main responsibility is to make the predictions which are precise and convincing enough and if they happen to loosely structured, there is a humongous possibility that the model will output the results with so many flaws and discrepancies. The architecture of the VGG16 model is shown below [8]:

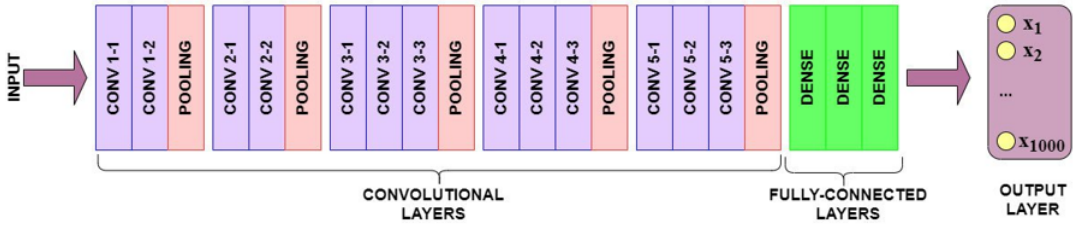


Figure 3: The architecture of the VGG16 model

3.2.2 Actions of Each Layer Contained in the VGG16 Model

In the VGG16 model, the combination of 13 convolution layers and 3 fully connected dense layers are not only important layers for the model to make good classification results. There are three more layers contained in the model that support the operations of 16 layers making the model has 5 different kinds of important layers exist inside. For the 13 convolution layers, there are two layers working along side by side with them.

Firstly, convolutional layers perform complex operations on the input data and are commonly used in computer vision tasks to extract compelling features from images. Within those layers, there are trainable filters to extract features from the inputted images. They perceive and identify boundaries, surface qualities, arrangements, and various characteristics at varying degrees of conceptualization. The activation function layers are also with them and utilized after each convolutional layer intentionally to introduce non-linearity, which enables the network to get a deeper understanding of the data by learning more detailed relationships. Then, Max-Pooling Layers decrease the spatial dimensions of the feature maps while preserving the crucial information obtained from the convolutional layers. In 3 fully connected layers, each neuron is connected to every neuron in the previous and next layers. The layers in the later stages of the network carry out categorization by using the retrieved characteristics and associating them with specific output classes and then its outputs are taken as the input to soft-max activation layer which is mainly responsible for making the most appropriate choice out of all the available options to pick one final result to produce through the output layer. The important layers of the VGG16 model are shown in the figure below [8]:

3.2.3 Changes to be Made in the VGG16 Model

Due to the fact that the main focus of the system of this paper is only on two categories such as HPV infected shrimps and shrimps with healthy hepatopancreases, there are some changes to be made in the structure of Figure 3. It is not necessary to have so many output labels coming out.

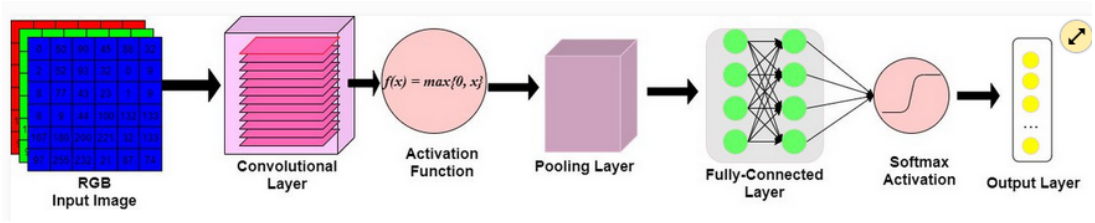


Figure 4: The Important Layers contained in the VGG16 model

Here, we can remove a couple of sessions such as fully connected layers zone and output layer which is also known as the later or top portion of the model as shown in the figure below [8]:

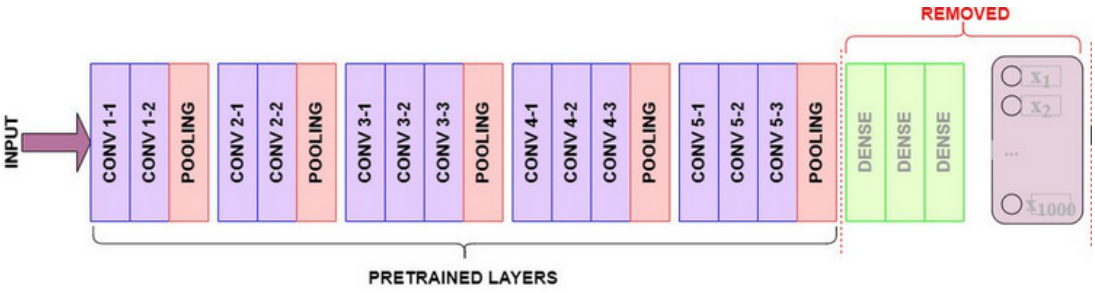


Figure 5: The removal of portions of the VGG16 model note that the removed portions could generate 1000 output labels

In the place of the chopped off fully connected dense layers and the output layer, customized layers are interpolated in order to be fir with the number of classes that are expected to see through the output layer as shown in the figure below [8]:

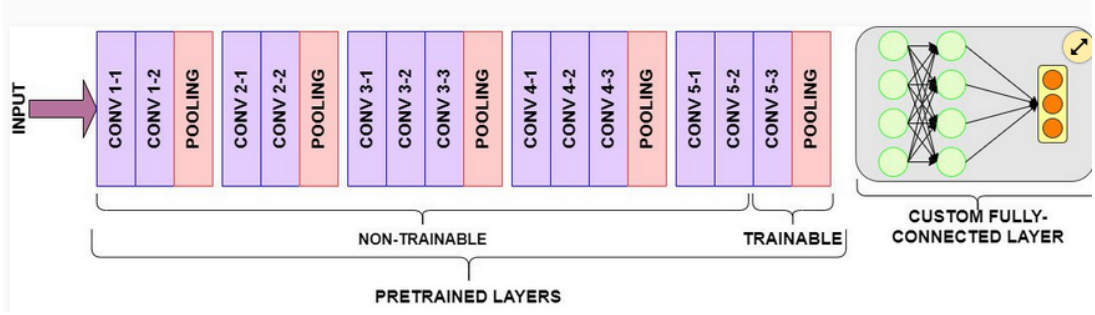


Figure 6: The inserted customized fully connected layer with the output layer together

Employment of a strategy called fine tuning will pack in the system. It transforms to be the one with customized fully connected layer in order to allow a portion of the pre-trained layers to retrain and also to increase the accuracy rate of the results derived by the model. In the process of fine tuning a pre-trained model, there are three main steps involving which are bootstrapping which is to customize fully connected layers and the output layer, freezing some pre-trained convolutional layer out of the 13 layers and unfreezing the last few pre-trained layers for training while some are frozen. The frozen convolutional layers convolve visual features as usual whereas the non-frozen convolutional layers are trained with the enough amount of data set which are a multitude of hepatopancreatic region images from both non-infected shrimps and infected shrimps.

3.2.4 Impacts of Learning Rate and the Number of Epochs

When the VGG16 model comes inside the frame, there are two major parameters that tag along with such as the learning rate and number of epochs which significantly affect the training of the VGG16 model as they control the learning process.

The step size used by the model to update its weights during training is determined by the learning rate. Increasing the learning rate can could accomplish the convergence of the model sooner, but it may result in overshooting the ideal solution or causing instability. A decreased learning rate may result in a slower convergence rate, but it can enhance the model's ability to discover a more precise solution.

An epoch signifies a whole iteration through the entire training dataset. Increasing the number of epochs in the training process enables the model to observe the data several times, which has the potential to enhance its performance. Excessive training for numerous epochs might result in overfitting, a situation where the model becomes excessively proficient in learning the training data but performs inadequately on new data.

In order to achieve the optimal outcomes or results, it is necessary to optimize certain hyper-parameters, such as identifying an optimal learning rate and considering the optimal number of epochs. This can lead to achieve high convergence and avoid problems like overfitting or poor convergence during training.

3.3 Performance Metrics

Getting the model with the best performance is dependent upon the demands of a classification task and an adequate amount of input data for the model to make itself familiar with all the cases contained in the task. Performance evaluation metrics such as precision, recall (sensitivity) and F1 score (F measure) are essential for the performance evaluation of classification models. They provide valuable perspectives of model performance, optimize model performance of the tasks assigned to a certain extent, and are being able to make decision-making in model deployment and enhancement serene. Additionally, evaluating the performance metrics aforementioned is always the mandatory phase to go through for one particular classification process in order to obtain some sorts of input data and then to generate the prediction results just to make certain the target audience knows how much the results are convincing especially when they can offer knowledgeable and useful insights into many elements of a classification model's performance.

To evaluate precision, recall and F1 measures, there are four instances required including TP (true positive) which is a cluster of instances and predicted positive results are within and all of them are actual positive cases, TN (true negative) which is a cluster of instances and predicted negative results are within and all of them are actual negative cases, FP (false positive) which is a group of instances and predicted positive results are within but they are negative ones actually and FN (false negative) which is a group of instances and predicted negative results are within but they are positive ones actually. If the four kinds of instances aforementioned are in hand, the following three performance metrics can be calculated [10–12].

Precision is one of the measures that calculates the accuracy percentage based on the positive predictions produced by the model. The value of the measure can be found by

$$Precision = \frac{TP}{(TP + FP)}. \tag{3.1}$$

Recall, known as the sensitivity of the model, computes the performance of the model to identify all appropriate instances based on the inputted dataset. The value of the metric can be calculated by recall or sensitivity equation. The only difference that exists with the formulas of precision measure and sensitivity measure is that in the denominator region of forms, which are the sums, the precision uses FP but the sensitivity uses FN.

$$Recall(Sensitivity) = \frac{TP}{(TP + FP)}. \tag{3.2}$$

F1 score, famously known as F measure, is calculated as the linear average of precision and recall, offering a harmonious equilibrium between these two measurements. This technique proves to be particularly advantageous in cases when there exists an unequal distribution of classes within the dataset. Since, F1 score is based on the precision and recall and due to the requirement of the values of both, if some sorts of discrepancies are involved in the two, F1 score always get impacted and might not be able to deduce the correct value. The F1 score can also be calculated by

$$F1score = 2 \times \frac{Precision \times Recall}{Precision \times Recall}. \tag{3.3}$$

4 Experimental Results

4.1 Accuracy Vs Loss Graphs

The primary factor to be taken into account is the level of efficiency exhibited by the model. Even though the VGG16 model was trained with a different number of epochs (25, 26, 27, 31 and 50), it can be seen that the validation accuracy value is always beyond the validation loss value meaning that the outcomes from the model are precise and convincing enough. Figures 7 and 8 both display with or without fine tuning, the optimal validation loss and validation accuracy can always be achieved during the model's training across a span of 50 epochs. The evaluation is conducted using a data set including almost 1000 shrimp images. Typically, and technically, convolutional neural network models that have a greater number of layers possess the capability to acquire more intricate characteristics from the images in the training dataset.

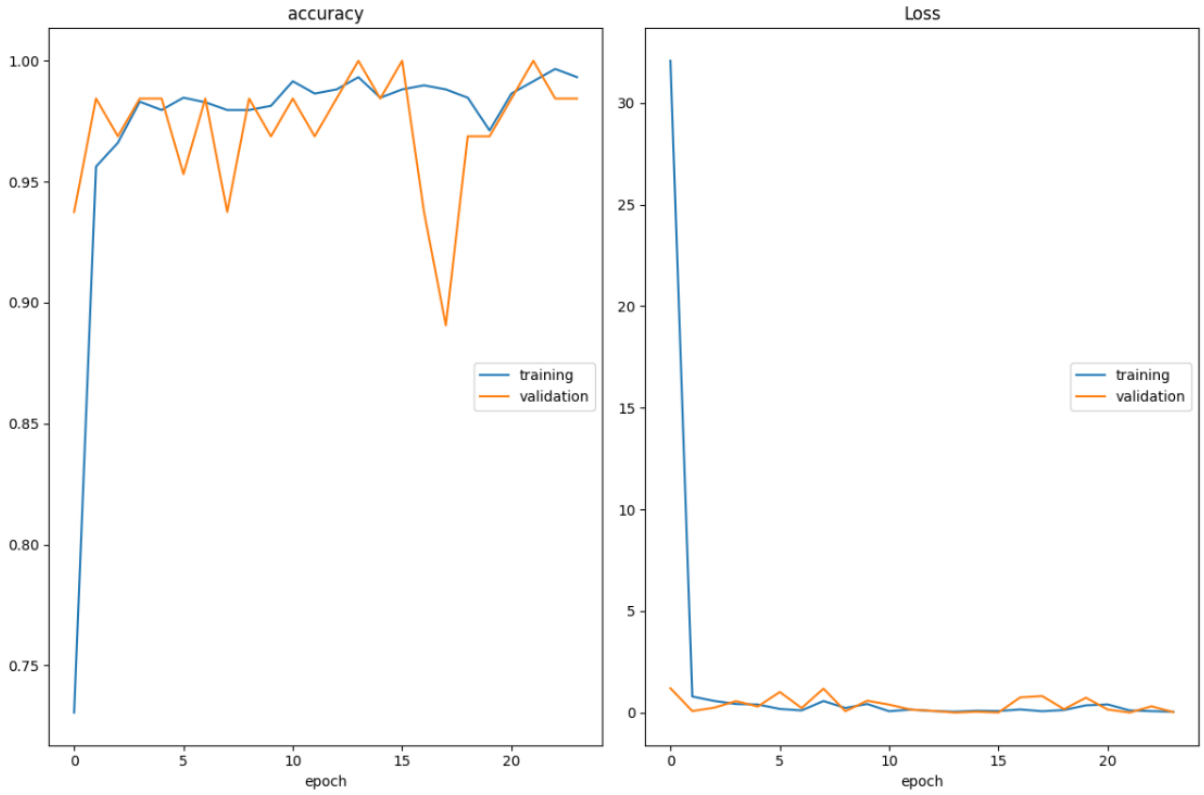


Figure 7: Training and Testing curve BEFORE FINE TUNING: orange line from accuracy means test accuracy, orange line from loss means loss accuracy, blue line from accuracy means train accuracy, blue line from loss means train loss

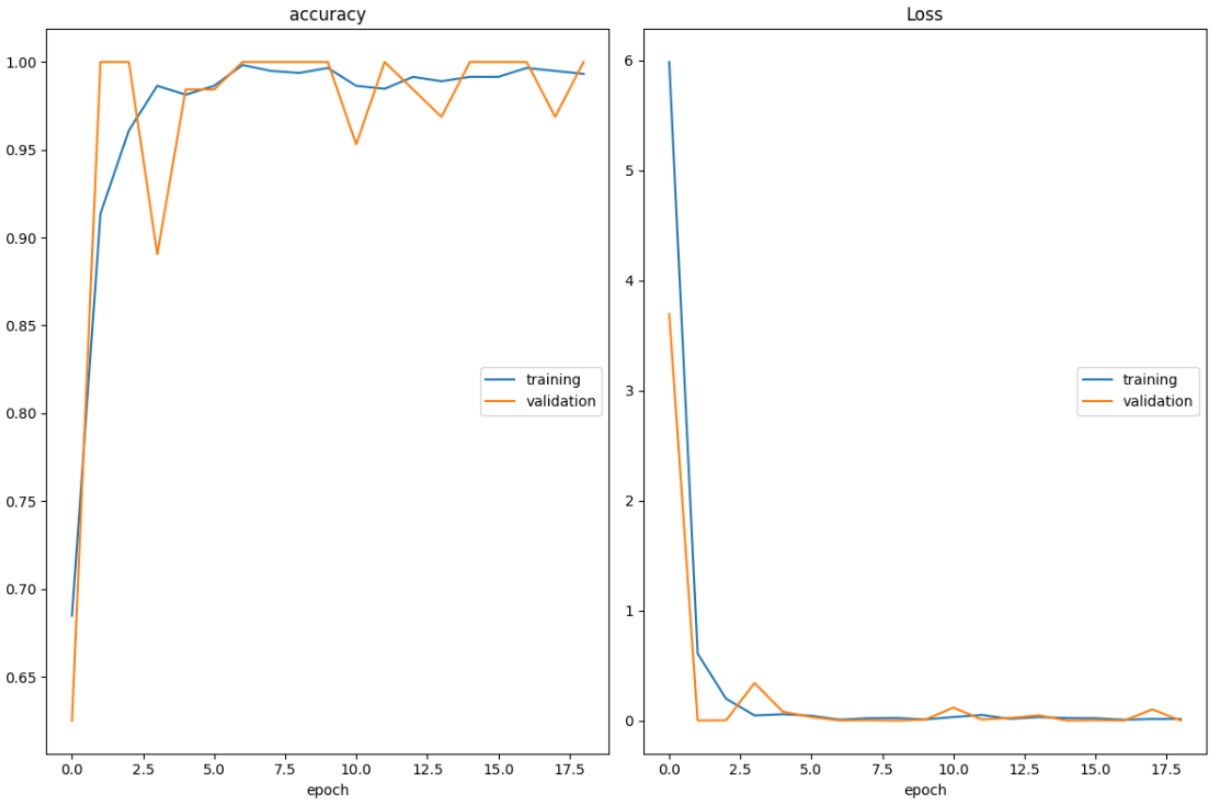


Figure 8: Training and Testing curve AFTER FINE TUNING: orange line from accuracy means test accuracy, orange line from loss means loss accuracy, blue line from accuracy means train accuracy, blue line from loss means train loss

4.2 Input and Output

After training the model for a required amount of time in order to increase the index of familiarity with both categories of the system such as HPV infected shrimps and healthy shrimps with or without fine tuning, it is observed the fact that the accuracy is always around 98% for whatever datasets: train and test. This indicates the sign of the model being able to produce the prediction results such as HPV for the hepatopancreas image input which is infected with Penaeus monodon Denso virus and NORMAL for the hepatopancreas image input which is the one without having any sorts of infections and healthy. The model is able to receive image input which has all the requirements of being able to get inside the system as the one as with the code shown in below:

```
upload = files.upload()
```

Figure 9: Python code to accept the input

With the code line shown above, the model can take one image file in and for instance one unhealthy hepatopancreas which has some spores showing that viruses consume good parts of it and make it dysfunctional. Then, shrimps encounter with growth retardation which is a significant cause of infection. The model deduces the around 97% to 100% accurate result stating the input is suffering HPV as shown in the figure below.

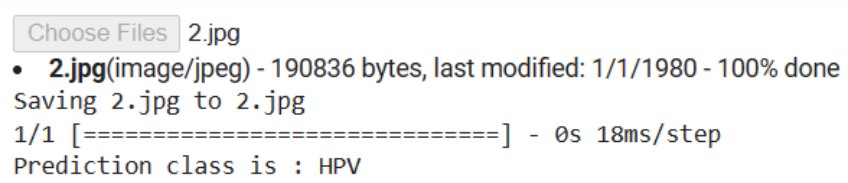


Figure 10: Python code showing the output for the input image file numbered 2

4.3 Results from Performance Metrics

The model exhibits 98% for precision measure, 97% each for both recall and F1 measure. The precision of the model is 98%, the recall is 97%, and the F1 score is also 97%. 98% of precision is a sign that showcases that the model precisely predicts positive cases and that is a really high score and meaning that the model is doing great significantly. 97% for recall measure indicates that the model can correctly identify 97% of all positive cases for the data set which is a collection of many images which captured various hepatopancreatic regions of shrimps. This indicates that the model is proficient in identifying the majority of two cases which are the major goals of the system such as infected shrimps and non-infected shrimps. One final performance measure F1 score has 97% which is the same percentage as recall states a favorable equilibrium between precision and recall since the percentage evaluated is almost about to hit 100% and only 3 points away from being able to achieve it. According to many surveys and researches conducted by many scientists, all the models which have a high level of precision while maintaining a high recall rate at the same time are highly convincing in numerous classification applications. In summary, the percentages of not just one but all the measures indicate that the model exhibits exceptional performance demonstrating a harmonious balance according to the f-1 score.

5 Conclusion

This paper showcases a route of applying one of the popular CNN models, VGG16 for diagnosing parvovirus infection occurring or not in the hepatopancreas region of shrimps from many images collected around the hepatopancreas regions of both infected and non-infected shrimps. VGG16 could really manage to extract some distinct architectures from each input by making itself familiar with all of them such as all the implicit features of both good and harmed hepatopancreas with a decent number of data set, 960 images in total by taking a decent amount of time with or without fine tuning. Most importantly, the model demonstrates a high level of accuracy rate which is measured in terms of percentage and that welcomes other hazardous diseases to come in the frame and spot not just parvo virus infected one and non-infected one.

References

[1] M. P. Safeena, P. Rai, and I. Karunasagar, *Molecular Biology and Epidemiology of Hepatopancreatic parvovirus of Penaeid Shrimp*, Indian J Virol. **23** (2012), 191–202.

[2] M. H. Al-Adhaileh, E. M. Senan, F. Alsaade, T. Aldhyani, N. Alsharif, A. A. Alqarni, M. Uddin, M. Alzahrani, E. Alzain, and M. Jadhav, *Deep learning algorithms for detection and classification of gastrointestinal diseases*, Complexity **2021** (2021), 1–12.

[3] N. Tajbakhsh, J. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, *Convolutional neural networks for medical image analysis: Full training or fine tuning?*, IEEE Transactions on Medical Imaging **35** (2016), 1299–1312.

[4] F. Chollet, *Deep Learning with Python*, 2nd ed., Manning Publications, Shelter Island, New York, 2021.

[5] J. Torres, *Learning process of a deep neural network*, Available online: https://towardsdatascience.com/learning-process-of-a-deep-neural-network-5a9768d7a651 (Accessed: 28 January 2024).

[6] J.Tanaka, and V. Philibert, *The Expertise of Perception*, [Edition unavailable]. Cambridge University Press. Retrieved from https://www.perlego.com/book/4230077/the-expertise-of-perception-how-experience-changes-the-way-we-see-the-world-pdf (Accessed: 14 December 2023), 2022.

[7] H. Li, R. Zhao, and X. Wang, *Highly efficient forward and backward propagation of CNNs for pixelwise classification*, arXiv preprint arXiv:1412.4526 (2014).

[8] Keras, Available online: https://www.learndatasci.com/tutorials/hands-on-transfer-learning-keras/ (Accessed: 28 February 2024).

[9] C. Kittinaradorn, *Neural network algorithm*, Available online: https://guopai.github.io/ml-blog14.html (Accessed: 11 January 2024)

[10] W.Yu, J. Chang, C. Yang, L. Zhang, H. Shen, Y. Xia, and J. Sha, *Automatic classification of leukocytes using deep neural network*, Proceedings of the International Conference on ASIC, 2017, Guiyang, China, 25–28 October 2017, pp. 1041–1044.

[11] C. Marzahl, M. Aubreville, J. Voigt, and A. Maier, *Classification of leukemic B-Lymphoblast cells from blood smear microscopic images with an attention-based deep learning method and advanced augmentation techniques*, In Lecture Notes in Bio-engineering. Springer: Singapore, 2019, pp. 13–22.

[12] S. H. Kassani, and P. H. Kassani, *A comparative study of deep learning architectures on melanoma detection*, Tissue and Cell **58** (2019), 76–83.

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

การเปรียบเทียบประสิทธิภาพของแบบจำลองพยากรณ์จำนวนผู้เสียชีวิตจากการเกิดอุบัติเหตุจราจรบนโครงข่ายถนนของกระทรวงคมนาคม

สุภาพร ครองยุทธ [1,†] และ ปรียานุช เชื้อสุข[1,‡]

[1]ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา 20131

บทคัดย่อ

งานวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ การศึกษาและเปรียบเทียบประสิทธิภาพของอัลกอริทึมการเรียนรู้ของเครื่องในการพยากรณ์จำนวนผู้เสียชีวิตจากการเกิดอุบัติเหตุบนโครงข่ายถนน ของกระทรวงคมนาคม ตั้งแต่เดือนมกราคม พ.ศ. 2562 ถึงเดือนมกราคม พ.ศ. 2566 ของจังหวัดนครราชสีมา โดยประยุกต์ใช้วิธีการถดถอยเชิงเส้นโครงข่ายประสาทเทียมแบบเพอร์เซฟตรอนหลายชั้น และซัพพอร์ตเวกเตอร์รีเกรสชันที่อาศัยการเรียนรู้ของเครื่องสำหรับการพัฒนาแบบจำลองในการเรียนรู้แบบผู้สอนในการพยากรณ์จำนวนผู้เสียชีวิต และวัดประสิทธิภาพการพยากรณ์ของแบบจำลองด้วยค่าคลาดเคลื่อนกำลังสองเฉลี่ย และค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย ผลการวิจัยพบว่า การพัฒนาแบบจำลองโดยใช้วิธีซัพพอร์ตเวกเตอร์รีเกรสชันมีประสิทธิภาพในการสร้างแบบจำลองการพยากรณ์จำนวนผู้เสียชีวิตจากการเกิดอุบัติเหตุบนโครงข่ายถนน ของกระทรวงคมนาคมมากที่สุดและมีค่าคลาดเคลื่อนกำลังสองเฉลี่ย ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ยต่ำสุด เมื่อเปรียบเทียบกับแบบจำลองที่สร้างจากวิธีการถดถอยเชิงเส้น โครงข่ายประสาทเทียมแบบเพอร์เซฟตรอนหลายชั้น

คำสำคัญ: การเรียนรู้ของเครื่อง, โครงข่ายประสาทเทียม, การถดถอยเชิงเส้น, ซัพพอร์ตเวกเตอร์รีเกรสชัน

2020 MSC: 68T07, 62J05

*งานวิจัยเรื่องนี้ได้รับทุนสนับสนุนจากภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา
†ผู้นำเสนอ ‡ผู้แต่งหลัก
อีเมล: 63030512@go.buu.ac.th, preeyanuch.ch@buu.ac.th.

1 บทนำ

ปัญหาอุบัติเหตุโครงข่ายถนนของกระทรวงคมนาคมเป็นปัญหาที่สำคัญอันดับต้น ๆ ของประเทศไทยที่จำเป็นต้องป้องกันและแก้ไขอย่างเป็นระบบและเร่งด่วน เนื่องจากมีแนวโน้มความรุนแรงของการเกิดอุบัติเหตุจราจรบนโครงข่ายถนนของกระทรวงคมนาคมเพิ่มมากขึ้น อุบัติเหตุบนท้องถนนเป็นหนึ่งในสาเหตุหลักที่นำไปสู่การเสียชีวิตและบาดเจ็บสาหัสในหลายประเทศทั่วโลก อุบัติเหตุเหล่านี้มักเกิดจากหลายปัจจัยรวมกัน เช่น ความผิดพลาดของผู้ขับขี่ สภาพถนนที่ไม่ดี สภาพอากาศ การขาดความตระหนักรู้เกี่ยวกับความปลอดภัยในการขับขี่ และการใช้แอลกอฮอล์หรือสารเสพติดขณะขับขี่ ผลกระทบของอุบัติเหตุทางถนนไม่เพียงแต่ส่งผลต่อผู้ที่ได้รับบาดเจ็บหรือเสียชีวิตเท่านั้น แต่ยังรวมถึงครอบครัว เพื่อนฝูง และชุมชน นอกจากนี้ยังมีผลกระทบทางเศรษฐกิจ เช่น ค่าใช้จ่ายทางการแพทย์ การสูญเสียแรงงาน

จากข้อมูลของศูนย์เทคโนโลยีสารสนเทศและการสื่อสาร สำนักงานปลัดกระทรวงคมนาคมตั้งแต่ปี พ.ศ. 2560 ถึง พ.ศ. 2566 พบว่า การเกิดอุบัติเหตุบนถนนมีจำนวนเพิ่มขึ้น จากข้อมูลดังกล่าวมีการเกิดอุบัติเหตุบนถนนจาก 18,951 ครั้งในปี 2560 เป็น 23,057 ครั้งในปี 2566 และในขณะที่จำนวนผู้บาดเจ็บและเสียชีวิตยังคงสูงอย่างต่อเนื่อง การวิเคราะห์สาเหตุของอุบัติเหตุเหล่านี้ระบุว่าการขับขี่เร็วเกินกำหนดเป็นสาเหตุหลัก โดยส่วนใหญ่เกิดกับรถปิคอัพบรรทุก 4 ล้อ และสถานที่เกิดเหตุส่วนใหญ่อยู่บนถนนทางตรงที่ไม่มีความลาดชัน สำหรับจังหวัดนครราชสีมาที่เปรียบเสมือนประตูสู่ภาคตะวันออกเฉียงเหนือ เพราะมีเส้นทางเชื่อมไปยังหลายจังหวัด ทำให้ทุกปีมีสถิติการเกิดอุบัติเหตุที่มากเป็นอันดับต้น ๆ ของประเทศ ในช่วงปี 2560 ถึง 2566 พบว่าพื้นที่นี้มีการเกิดอุบัติเหตุรวมทั้งสิ้น 6,386 ครั้ง ทำให้มีผู้บาดเจ็บ 6,657 ราย และผู้เสียชีวิตถึง 879 ราย จากสถิติเหล่านี้ทำให้สะท้อนถึงความเร่งด่วนในการดำเนินการอย่างจริงจังเพื่อลดอุบัติเหตุบนถนน และปรับปรุงนโยบายการจราจรเพื่อลดความเสี่ยง และปกป้องความสูญเสียของประชาชนจากการเกิดอุบัติเหตุทางถนนในอนาคต

การลดอุบัติเหตุทางถนนยังคงเป็นความท้าทายที่สำคัญของประเทศไทยและหลายประเทศทั่วโลก รัฐบาลไทยพยายามตอบสนองปัญหานี้ด้วยการดำเนินนโยบายและกลยุทธ์มากมายเพื่อลดอุบัติเหตุบนท้องถนน ซึ่งรวมถึงการกำหนดมาตรการป้องกันอุบัติเหตุทางถนน การพัฒนาโครงสร้างพื้นฐานทางถนน และการสร้างความตระหนักเรื่องความปลอดภัยให้แก่ประชาชน อย่างไรก็ดี การทำให้มาตรการเหล่านี้บรรลุผลตามเป้าหมายยังคงเป็นความท้าทายอย่างยิ่ง ดังนั้น การพยากรณ์อุบัติเหตุทางถนนที่แม่นยำจึงเป็นเครื่องมือสำคัญที่ช่วยให้หน่วยงานที่เกี่ยวข้องสามารถดำเนินการได้อย่างมีประสิทธิภาพ และมองเห็นแนวโน้มของอุบัติเหตุทางถนนซึ่งจะช่วยในการกำหนดและจัดสรรทรัพยากรได้อย่างเหมาะสม โดยเฉพาะการพยากรณ์จำนวนผู้เสียชีวิตจากอุบัติเหตุ เนื่องจากช่วยให้หน่วยงานที่เกี่ยวข้องสามารถจัดสรรทรัพยากรและกำหนดมาตรการที่เหมาะสมได้อย่างตรงจุด การมีข้อมูลที่แม่นยำเกี่ยวกับจำนวนผู้เสียชีวิตจากการเกิดอุบัติเหตุที่อาจเกิดขึ้น จะสามารถช่วยให้รัฐบาลและหน่วยงานความปลอดภัยทางถนนวางแผนได้ดีขึ้นในการป้องกันและลดความรุนแรงของอุบัติเหตุ

นักวิจัยหลายท่านได้ทำการวิเคราะห์ปัญหาอุบัติเหตุบนโครงข่ายถนนด้วยเทคนิคต่าง ๆ ยกตัวอย่างเช่น ในปี พ.ศ. 2561 ปทิตญา บุญรักษา และจารี ทองคำ [1] ได้ทำการศึกษาเปรียบเทียบประสิทธิภาพของแบบจำลองต่าง ๆ ในการพยากรณ์จำนวนผู้ประสบอุบัติเหตุบนท้องถนนในจังหวัดขอนแก่น โดยใช้วิธี 5 วิธี คือ การถดถอยเชิงเส้น โครงข่ายประสาทเทียม การถดถอยเวกเตอร์ (SMOreg: Sequential Minimal Optimization Regression) ซัพพอร์ตเวกเตอร์รีเกรสชัน และกระบวนการเกาส์เซียน พวกเขาใช้หลักการหน้าต่างบานเลื่อน (sliding window) ในการแบ่งข้อมูลเป็นชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบ โดยวัดประสิทธิภาพการพยากรณ์ด้วยค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย และรากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ย ผลการวิจัยพบว่าวิธี ซัพพอร์ตเวกเตอร์รีเกรสชัน มีประสิทธิภาพสูงสุดในการสร้างแบบจำลองที่มีความแม่นยำสูงที่สุด เมื่อเปรียบเทียบกับเทคนิคอื่น ๆ ในปีเดียวกันนั้น Dali Wu และ Sanming Wang [14] ได้ศึกษาการพยากรณ์การเกิดอุบัติเหตุทางถนนในประเทศจีน โดยใช้การวิเคราะห์องค์ประกอบหลักเพื่อลดมิติข้อมูลสถิติการเกิดอุบัติเหตุจราจรทางถนน และ

เปรียบเทียบวิธีการวิเคราะห์ระหว่างซัพพอร์ตเวกเตอร์รีเกรสชัน และโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ ผลการศึกษาพบว่า ซัพพอร์ตเวกเตอร์รีเกรสชันมีความแม่นยำสูงกว่า โครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ และสามารถตอบโจทย์ความต้องการของการพยากรณ์การเกิดอุบัติเหตุจราจรทางถนนได้เป็นอย่างแม่นยำ

การวิจัยนี้ มุ่งเน้นไปที่แบบจำลองการพยากรณ์จำนวนผู้เสียชีวิตจากการเกิดอุบัติเหตุจราจรบนโครงข่ายถนนของกระทรวงคมนาคม ในจังหวัดนครราชสีมา โดยเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์ที่ต่างกัน คือ การถดถอยเชิงเส้น โครงข่ายประสาทเทียมแบบเพอร์เซฟตรอนหลายชั้น และซัพพอร์ตเวกเตอร์รีเกรสชัน โดยมีวัตถุประสงค์ในการหาแบบจำลองที่มีความแม่นยำและประสิทธิภาพสูงสุดในการพยากรณ์ เพื่อเป็นเครื่องมือสนับสนุนการตัดสินใจของผู้กำหนดนโยบาย และการวางแผนของหน่วยงานที่เกี่ยวข้องในการป้องกันและลดอุบัติเหตุบนถนน และจัดทำมาตรการป้องกันได้อย่างมีประสิทธิภาพ นอกจากนี้ทางผู้วิจัยได้นำกระบวนการวิเคราะห์ข้อมูลด้วย Cross-Industry Standard Process for Data Mining (CRISP-DM) มาประยุกต์ใช้กับข้อมูลการเกิดอุบัติเหตุบนโครงข่ายถนนของกระทรวงคมนาคม

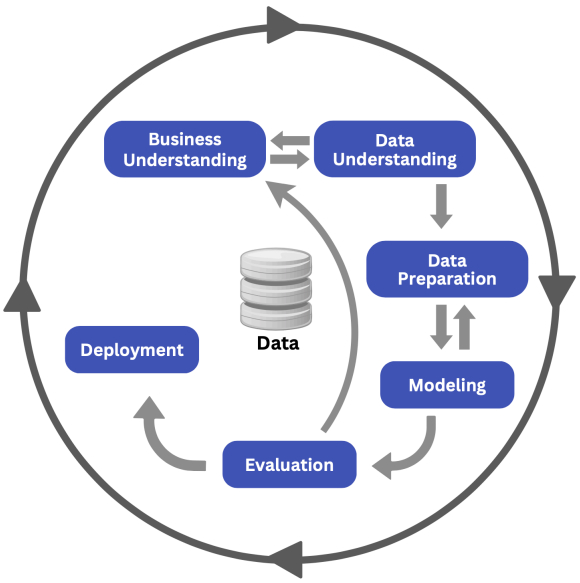
2 ความรู้พื้นฐาน

ในหัวข้อนี้ จะกล่าวถึงกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM หลักการของแต่ละเทคนิคในการสร้างแบบจำลอง และการวิเคราะห์ความหมายของค่าสถิติ

2.1 กระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM

CRISP-DM ย่อมาจาก “Cross-Industry Standard Process for Data Mining” [16] เป็นกระบวนการมาตรฐานที่ใช้สำหรับการทำเหมืองข้อมูล (Data Mining) โดยกระบวนการ CRISP-DM จะประกอบด้วย 6 ขั้นตอน ซึ่งแต่ละขั้นตอนจะเป็นขั้นตอนที่ต่อเนื่องกันถูกแสดงด้วยลูกศรที่เชื่อมระหว่างขั้นตอนแต่ละขั้นตอน ดังภาพที่ 1 ขั้นตอนในกระบวนการ CRISP-DM มีดังนี้

1. การทำความเข้าใจธุรกิจ (Business Understanding)
   ขั้นตอนแรกจะมุ่งไปที่การทำความเข้าใจในจุดประสงค์ทางธุรกิจ การระบุปัญหา และการกำหนดวัตถุประสงค์ เพื่อที่จะแปลงปัญหาเหล่านั้นเป็นโจทย์ในการวิเคราะห์ข้อมูล และวางแผนในการนำข้อมูลไปใช้ต่อไป



ภาพที่ 1: ขั้นตอนของกระบวนการ CRISP-DM

2. การทำความเข้าใจข้อมูล (Data Understanding)
   ขั้นตอนนี้ คือ การทำความเข้าใจกับข้อมูล โดยเริ่มต้นจากการเก็บรวบรวมข้อมูลที่เกี่ยวข้องและเชื่อถือได้ ขั้นตอนที่ 1 และ 2 สามารถทำกลับไปมาได้ เนื่องจากการทำความเข้าใจธุรกิจช่วยให้เราได้รับความรู้เกี่ยวกับข้อมูลมากขึ้น ในทางตรงกันข้ามการได้ทำความเข้าใจข้อมูลอย่างลึกซึ้งจะช่วยเพิ่มความเข้าใจในแง่มุมธุรกิจให้กว้างขึ้นเช่นเดียวกัน

3. การเตรียมข้อมูล (Data Preparation)
   ขั้นตอนนี้ คือ การเตรียมข้อมูลดิบ (raw material) ที่ถูกรวบรวมมา ให้สามารถนำไปวิเคราะห์ในขั้นตอนที่ 4 ได้ โดยจะประกอบด้วยขั้นตอนย่อยหลัก ๆ คือ การทำความสะอาดข้อมูล การเลือกและสร้างชุดข้อมูลที่จะใช้ในแบบจำลอง การแปลงข้อมูลให้เหมาะสมกับวิธีการวิเคราะห์ที่จะใช้ รวมถึงการจัดการข้อมูลที่หายไปหรือผิดพลาด

4. การสร้างแบบจำลองวิเคราะห์ข้อมูล (Modeling)
   ในขั้นตอนนี้ เป็นการนำข้อมูลจากขั้นตอนที่ 3 มาทดลองสร้างแบบจำลองจากวิธีหลาย ๆ วิธี ที่น่าจะสามารถแก้ไขปัญหาที่ต้องการได้ และทำการปรับเปลี่ยนค่าพารามิเตอร์ต่าง ๆ เพื่อหาแบบจำลองที่ดีที่สุด

5. การประเมินผลลัพธ์ (Evaluation)
   ขั้นตอนนี้ คือ ขั้นตอนของการตรวจสอบและประเมินผลแบบจำลองที่ได้จากขั้นตอนที่ 4 เพื่อวัดว่าแบบจำลองมีประสิทธิภาพเพียงพอต่อการนำไปใช้งานแล้วหรือไม่

6. การนำไปใช้งานจริง (Deployment)
   ขั้นตอนนี้ เป็นแสดงผลที่ได้มาจาก ขั้นตอนที่ 5 และนำผลลัพธ์ที่ได้จากแบบจำลองไปใช้งานจริง เพื่อวิเคราะห์และแก้ปัญหาที่ต้องการ

2.2 การถดถอยเชิงเส้น (Linear Regression: LR)

การวิเคราะห์การถดถอย (Regression Analysis) [5] เป็นการศึกษาความสัมพันธ์ระหว่างแปร ตั้งแต่ 2 ตัวแปรขึ้นไป โดยมีวัตถุประสงค์ที่ต้องการประมาณหรือพยากรณ์ค่าของตัวแปรตามจากตัวแปรอื่น ๆ ที่เกี่ยวข้อง การวิเคราะห์ความถดถอยแบ่งออกได้ 2 ประเภท

1. การวิเคราะห์ความถดถอยอย่างง่าย (Simple Regression Analysis) เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปร 2 ตัว ซึ่งจะประกอบด้วยตัวแปรตาม y จำนวน 1 ตัวแปร และมีตัวแปรอิสระ 1 ตัวแปร โดยที่มีความสัมพันธ์อยู่ในรูปเชิงเส้น สามารถเขียนเป็นสมการได้ดังนี้

$$y = \beta_0 + \beta_1 x + \varepsilon$$

   โดยที่ y คือ ตัวแปรตาม x คือ ตัวแปรอิสระ
   β_0 คือ ระยะตัดแกน Y หรือค่าของ y เมื่อ x มีค่าเป็นศูนย์
   β_1 คือ สัมประสิทธิ์การถดถอย (Regression Coefficient) เป็นความชันของเส้นสมการถดถอย และ
   ε คือ ค่าความคลาดเคลื่อน

2. การวิเคราะห์การถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression) เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรตาม y จำนวน 1 ตัวแปร และตัวแปรอิสระ จำนวน 2 ตัวแปรขึ้นไป โดยที่มีความสัมพันธ์อยู่ในรูปเชิงเส้น ซึ่งสามารถเขียนเป็นความสัมพันธ์ได้ดังนี้

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n + \varepsilon$$

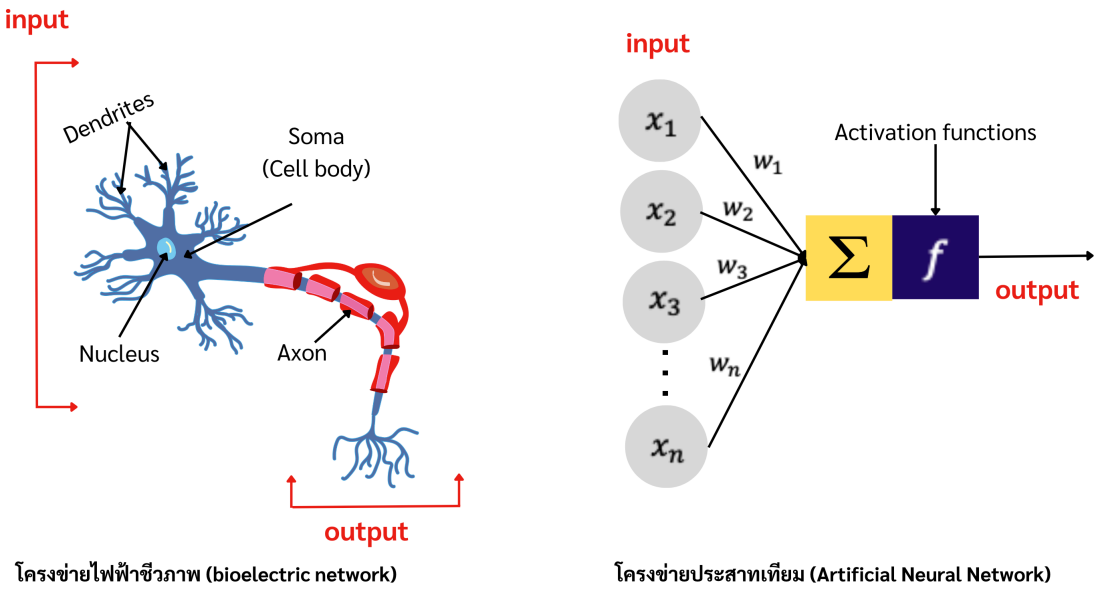
โดยที่ y คือ ค่าของตัวแปรตาม x_i คือ ค่าของตัวแปรอิสระที่ i
β_0 คือ เป็นระยะตัดแกน Y หรือค่าเริ่มต้นของเส้นสมการถดถอย
$\beta_1, ..., \beta_n$ คือ ค่าสัมประสิทธิ์การถดถอย ของตัวแปรอิสระ x_i และ
ε คือ ค่าความคลาดเคลื่อน

2.3 โครงข่ายประสาทเทียม (Artificial Neural Networks: ANN)

โครงข่ายประสาทเทียม (Artificial Neural Network) [6] คือ การสร้างโปรแกรมคอมพิวเตอร์ที่จำลองวิธีการทำงานของสมองมนุษย์ หรือเป็นการทำให้คอมพิวเตอร์รู้จักการคิดและการจดจำ แนวคิดเริ่มต้นของเทคนิคนี้ได้มาจากการศึกษาโครงข่ายไฟฟ้าชีวภาพ (Bioelectric Network) ในสมอง ซึ่งประกอบด้วย เซลล์ประสาท (Neurons) และ จุดประสานประสาท (Synapses) ซึ่งโครงสร้างหลักของเซลล์ประสาท 1 เซลล์ จะประกอบด้วย 3 ส่วน คือ ตัวเซลล์ (Soma) ทำหน้าที่ประมวลผลสัญญาณ เดนไดรต์ (Dendrite) ทำหน้าที่ รับสัญญาณเข้า และแอกซอน (Axon) ทำหน้าที่ ถ่ายโอนสัญญาณออกไปยังเซลล์สมองอื่น ในส่วนของโครงข่ายประสาทเทียมจะประกอบด้วย หน่วยประมวลผลเล็ก ๆ ที่เรียกว่า “โหนด” (Node) ซึ่งแต่ละโหนดนั้นจะทำงานคล้ายกับเซลล์ประสาทในสมองของมนุษย์ โดยมีการส่งข้อมูลระหว่างกันผ่านทาง “ค่าน้ำหนัก” (Weight) ที่เทียบเท่ากับจุดประสานประสาท ในโครงข่ายไฟฟ้าชีวภาพ ซึ่งโหนดจะมีการรวมตัวกันเป็นชั้น โหนดชั้นอินพุตของโครงข่ายประสาทเทียมจะรับสัญญาณเข้า โหนดชั้นซ่อนจะคำนวณสัญญาณเข้าเหล่านี้ และโหนดชั้นเอาต์พุตจะคำนวณผลลัพธ์สุดท้ายโดยใช้ ฟังก์ชันกระตุ้น (Activation Functions) ดังภาพที่ 2

โครงข่ายประสาทเทียมแบบเพอร์เซฟตรอนหลายชั้น (Multi-Layer Perceptron : MLP) เป็นโครงข่ายประสาทเทียมที่มีความซับซ้อนโดยมีการเชื่อมต่อกันของหลายชั้น แบบจำลองทางคณิตศาสตร์ของโครงข่ายประสาทเทียมเป็นการจำลองความสัมพันธ์ระหว่างข้อมูลและผลลัพธ์ที่ต้องการ มีสมการดังนี้

$$y = f(\sum_{i=1}^{n} w_i x_i + b)$$



ภาพที่ 2: โครงข่ายไฟฟ้าชีวภาพในสมองและโครงข่ายประสาทเทียม

โดยที่ w_i คือ ค่าน้ำหนัก ของตัวแปรนำเข้า x_i คือ ตัวแปรนำเข้า f คือ ฟังก์ชันกระตุ้น และ b คือ ค่าความเอนเอียง (Bias) การปรับค่าน้ำหนักให้เหมาะสมกับข้อมูลที่ให้ อาจใช้เทคนิคการแพร่ย้อนกลับ (Backpropagation) มีสมการดังนี้

$$\Delta w_{ji} = \eta \delta_j o_i$$

โดยที่ w_{ji} คือ การเปลี่ยนแปลงของน้ำหนักในการเชื่อมต่อระหว่างโหนด η คือ อัตราการเรียนรู้ o_i คือ ค่าเอาต์พุตของโหนด i ในชั้นปัจจุบัน และ δ_j คือ ค่าคลาดเคลื่อนของโหนด j คำนวณจาก

$$\delta_j = \begin{cases} x_j(1-x_j)(t_j - x_j) & \text{เมื่อ } j \text{ เป็นยูนิตที่อยู่ในชั้นส่งออก} \\ x_j(1-x_j)\sum_k \delta_k w_{kj} & \text{เมื่อ } j \text{ เป็นยูนิตที่อยู่ในชั้นซ่อน} \end{cases}$$

2.4 ซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression: SVR)

ซัพพอร์ตเวกเตอร์รีเกรสชัน (Support Vector Regression) [3] เป็นการประยุกต์ใช้หลักการของซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) ที่เป็นวิธีที่มีประสิทธิภาพสูงในกลุ่มของการเรียนรู้โดยมีผู้สอน (Supervised Machine Learning) ที่ใช้สำหรับการทำนายเชิงตัวเลข การจำแนกประเภท และการจดจำรูปแบบในข้อมูลที่ซับซ้อน วิธี SVR เป็นการขยายขอบเขตการใช้งานของ SVM จากการจำแนกประเภท (Classification) เป็นการทำนายค่าตัวเลข (Regression) วิธีการนี้เป็นประโยชน์ในการพยากรณ์ค่าของตัวแปรต่อเนื่อง และสามารถนำมาพยากรณ์ข้อมูลอนุกรมเวลา (Time Series Data) ได้ โดยมีเป้าหมายเพื่อค้นหาฟังก์ชันถดถอยที่สามารถทำนายค่าเอาต์พุต $(y \in \mathbb{R})$ จากอินพุตเวกเตอร์ $(x \in \mathbb{R}^n)$ ได้อย่างแม่นยำที่สุด ฟังก์ชันนี้จะถูกแสดงในรูปของสมการเชิงเส้น

$$f(x) = w^T x + b$$

โดยที่ w คือ เวกเตอร์น้ำหนัก และ b คือ ค่าเอนเอียง (Bias Term) โดยที่วิธี SVR จะพยายามหาค่าของ w และ b ที่ทำให้ฟังก์ชัน $f(x)$ สามารถทำนายค่าเอาต์พุตที่ได้ดีที่สุด โดยพิจารณาถึงการปรับความแม่นยำในการทำนายเพื่อลดความเสี่ยงของการทำนายที่ผิดพลาด การหาค่าของ w และ b ทำได้ด้วยวิธีการหาค่าต่ำสุดของสมการที่ (2.1)

$$R = \frac{1}{2}\|w\|^2 + \frac{c}{l}\sum_{i=1}^{l} |y_i - f(x_i)|_\varepsilon \tag{2.1}$$

การใช้วิธี SVR ในการทำนายค่าเอาท์พุตจากอินพุตเวกเตอร์ เป็นการสร้างแบบจำลองที่มีการใช้ท่อเอปซิลอน (Epsilon Tube) เพื่อกำหนดขอบเขตในการพยากรณ์ โดยใช้ฟังก์ชันสูญเสีย (Loss Function) ในการปรับแบบจำลองเพื่อให้ค่าพยากรณ์ใกล้เคียงกับค่าจริงในช่วงที่กำหนดได้เยอะที่สุด ดังสมการที่ (2.2)

$$|y_i - f(x_i)|_\varepsilon = \begin{cases} 0, & \text{if } |y_i - f(x_i)|_\varepsilon = \varepsilon \\ |y_i - f(x_i)|_\varepsilon - \varepsilon, & \text{otherwise} \end{cases} \tag{2.2}$$

การแก้ปัญหาของสมการที่ (2.1) ที่มีเงื่อนไขตามสมการที่ (2.2) สามารถปรับให้อยู่ในรูปแบบการแก้ปัญหาแบบคู่ (Dual problem) ด้วยการใช้ตัวคูณลากรานจ์ (Lagrange multipliers) ดังสมการที่ (2.3) และ (2.4) [7]

$$Maximize L_p(\alpha_i, \alpha_i^*) = -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)x_i^T x_j - \varepsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)y_i \tag{2.3}$$

$$subject\ to \begin{cases} \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 & \\ 0 \leq \alpha_i \leq C, & i = 1, ..., l \\ 0 \leq \alpha_i^* \leq C, & i = 1, ..., l \end{cases} \tag{2.4}$$

เมื่อเรามี α_i และ α_i^* เป็นตัวคูณลากรานจ์ และ C เป็นจำนวนเต็มที่เป็นค่าคงที่ เมื่อเกิดข้อผิดพลาด (Error) ขนาด ε และ ε คือความกว้างของท่อเอปซิลอน (Epsilon Tube) หรือความคลาดเคลื่อนของชุดข้อมูลฝึกฝน และ l คือจำนวนของซัพพอร์ตเวกเตอร์ (Support Vectors) ซึ่งส่วนอินพุตเวกเตอร์ที่เป็นซัพพอร์ตเวกเตอร์ จะมี $\alpha_i, \alpha_i^* > 0$ ส่วนอินพุตเวกเตอร์ที่ไม่ใช่ซัพพอร์ตเวกเตอร์ จะมี $\alpha_i, \alpha_i^* = 0$ และหลังจากที่คำนวณค่า α_i และ α_i^* จากชุดข้อมูลฝึกฝน เราจะสามารถสร้างสมการ SVR เพื่อใช้ทำนายค่าเอาท์พุตจากอินพุตเวกเตอร์ ได้ดังสมการที่ (2.5)

$$f(x) = w_0^T x + b = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)x_i^T x + b \tag{2.5}$$

โดยที่ เวกเตอร์ถ่วงน้ำหนัก w_0 เป็นดังสมการที่ (2.6)

$$w_0 = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)x_i \tag{2.6}$$

สมการที่ (2.5) มีการใช้การถดถอยเชิงเส้น เพื่อให้มีการปรับค่าในการเปลี่ยนแปลงของตัวแปรต้นและมีค่าเอนเอียง ที่บวกอยู่ด้วยเพื่อทำให้การถดถอยนั้นมีความแม่นยำมากขึ้น ส่วนในกรณีที่เราต้องการการถดถอยไม่เชิงเส้น เราสามารถใช้ฟังก์ชันเคอร์เนล (Kernel Function) มาช่วยในการแปลงข้อมูลให้อยู่ในมิติที่สูงขึ้น เพื่อให้การถดถอยสามารถประมาณค่าได้ดีขึ้น ซึ่งฟังก์ชันเคอร์เนลที่นิยมใช้ใน SVR อยู่ 3 แบบ [7] คือ

1. Linear Kernel ใช้สำหรับ การถดถอยเชิงเส้น ซึ่งมีสมการเป็น

$$k(x_i, x) = x_i^T x$$

2. Polynomial Kernel ใช้สำหรับ การถดถอยไม่เชิงเส้น ซึ่งมีสมการเป็น

$$k(x_i, x) = (1 + x_i \cdot x_j)^d$$

โดยที่ d คือ Polynomial degree

3. Gaussian Kernel ใช้สำหรับ การถดถอยไม่เชิงเส้น ซึ่งมีสมการเป็น

$$k(x_i, x) = exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\|)$$

ดังนั้น สมการที่ (2.5) สามารถเขียนใหม่ในรูปแบบการถดถอยไม่เชิงเส้นโดยใช้เคอร์เนลฟังก์ชัน ได้ดังสมการที่ (2.7)

$$f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)k(x_i - x) \tag{2.7}$$

วิธีการถดถอยเชิงเส้น ถึงแม้จะได้รับความนิยมอย่างกว้างขวางสำหรับการประยุกต์ใช้สร้างแบบจำลองการพยากรณ์ที่ใช้ความสัมพันธ์แบบเชิงเส้น อย่างไรก็ตาม วิธี ANN และ SVR กลายเป็นที่นิยมมากขึ้นในงานวิจัย เนื่องจากความแม่นยำสูง และความสามารถในการจัดการข้อมูลที่มีความสัมพันธ์แบบไม่เชิงเส้น นอกจากนี้วิธี SVR เป็นวิธีการที่ประมวลผลได้รวดเร็วและเหมาะสมกับชุดข้อมูลที่มีขนาดเล็ก [3]

2.5 วิเคราะห์ความหมายของค่าสถิติ

การทำข้อมูลให้เป็นปกติ (Data Normalization) [12] เป็นวิธีการหนึ่งของการแปลงข้อมูล (Data Transformation) โดยงานวิจัยชิ้นนี้ เราจะแปลงข้อมูลโดย เทคนิค Min-Max Scaling เป็นการปรับค่าคุณลักษณะของข้อมูลให้อยู่ในช่วงค่าน้อย และค่ามากที่กำหนด ซึ่งนิยมใช้ค่าน้อยเป็น 0 และค่ามากเป็น 1 ซึ่งวิธีการแปลงค่าจะคำนวณได้จากสมการดังนี้

$$x_i^{scaled} = \frac{x_i - \min(X)}{\max X - \min X}$$

โดยที่ x_i^{scaled} คือ ค่าใหม่ของคุณลักษณะตัวที่ i
x_i คือ ค่าของคุณลักษณะตัวที่ i
$\min(X)$ คือ ค่าที่น้อยที่สุดของของคุณลักษณะนั้น และ
$\max(X)$ คือ ค่าที่มากที่สุดของคุณลักษณะนั้น

ค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error) ใช้ในการวัดความแตกต่างระหว่างค่าที่คำนวณได้จากแบบจำลองหรือการพยากรณ์กับค่าจริง โดยใช้ค่าเฉลี่ยของความแตกต่างระหว่างทุกจุดข้อมูลที่มีในชุดข้อมูลที่กำหนด คำนวณด้วยการยกกำลังสองของความแตกต่างแต่ละจุดข้อมูล และหาค่าเฉลี่ยของความแตกต่างทั้งหมดนั้น หากค่าคลาดเคลื่อนกำลังสองเฉลี่ยน้อยแสดงถึงค่าพยากรณ์สามารถประมาณค่าได้ใกล้เคียงกับค่าจริง มีสมการดังนี้

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

โดยที่ n คือ จำนวนข้อมูลทั้งหมด
y_i คือ ค่าจริงในชุดข้อมูลที่ i และ
$\hat{y}_i$ คือ ค่าพยากรณ์ในชุดข้อมูลที่ i

ค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Error) ใช้ในการวัดความแตกต่างระหว่างค่าที่คำนวณได้จากแบบจำลองหรือการพยากรณ์กับค่าจริง โดยใช้วิธีการหาค่าเฉลี่ยของความแตกต่างสมบูรณ์ระหว่างค่าพยากรณ์และค่าจริง หากค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ยน้อยแสดงถึงค่าพยากรณ์สามารถประมาณค่าได้ใกล้เคียงกับค่าจริง มีสมการดังนี้

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

โดยที่ n คือ จำนวนข้อมูลทั้งหมด
y_i คือ ค่าจริงในชุดข้อมูลที่ i และ
$\hat{y}_i$ คือ ค่าพยากรณ์ในชุดข้อมูลที่ i

MAE จะมีความอ่อนไหวต่อข้อมูลที่มีค่าผิดปกติ น้อยกว่าเมื่อเทียบกับ MSE การใช้ MAE จึงเหมาะสมกว่าในสถานการณ์ที่ข้อมูลมีค่าผิดปกติอยู่เป็นจำนวนมาก

ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) [9] เป็นค่าที่บ่งชี้ถึงความสัมพันธ์ระหว่างตัวแปรสองตัว จะมีค่าอยู่ระหว่าง -1 ถึง 1 โดยหากพบว่าค่าสัมประสิทธิ์สหสัมพันธ์ เข้าใกล้ -1 หมายความว่าตัวแปรทั้งสองตัวมีความสัมพันธ์กันในเชิงตรงกันข้าม แต่หากค่าสัมประสิทธิ์สหสัมพันธ์ มีค่าเข้าใกล้ 1 หมายความว่าตัวแปรทั้งสองมีความสัมพันธ์ไปในทิศทางเดียวกัน สามารถคำนวณได้ดังนี้

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

โดยที่ r คือ ค่าสัมประสิทธิ์สหสัมพันธ์

n คือ จำนวนข้อมูลทั้งหมด
x_i คือ ค่าตัวแปร x ของชุดข้อมูลที่ i
$\bar{x}$ คือ ค่าเฉลี่ยของตัวแปร x
y_i คือ ค่าตัวแปร y ของชุดข้อมูลที่ i และ
$\bar{y}$ คือ ค่าเฉลี่ยของตัวแปร y

3 งานวิจัยที่เกี่ยวข้อง

จากงานวิจัยของ Zeng Qing-Wei และคณะ [11] กล่าวไว้ว่า การใช้ซัพพอร์ตเวกเตอร์รีเกรสชัน รวมกับวิธีหาค่าเหมาะสมที่สุดแบบกลุ่มอนุภาค (Particle Swarm Optimization : PSO) เพื่อพยากรณ์การเกิดอุบัติเหตุทางจราจร การวิเคราะห์เวลาในรูปแบบชุดข้อมูลเป็นทิศทางสำคัญในการทำนายอุบัติเหตุทางจราจร ผลการวิจัยพบว่า เทคนิควิธี PSO-SVR มีประสิทธิภาพสูงกว่าโครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ ในการพยากรณ์อุบัติเหตุทางจราจร

จากงานวิจัยของ Wei-wei Wu และคณะ [15] ใช้วิธีซัพพอร์ตเวกเตอร์รีเกรสชัน เพื่อการพยากรณ์ระยะเวลาของเหตุการณ์จราจร โดยใช้ข้อมูลเหตุการณ์จราจรจากทางด่วนในเนเธอร์แลนด์ ผลการวิจัยพบว่า แบบจำลองจากวิธี ซัพพอร์ตเวกเตอร์รีเกรสชัน สามารถพยากรณ์ระยะเวลาของเหตุการณ์จราจรได้อย่างแม่นยำ โดยสามารถนำไปใช้ในระบบตรวจจับและแก้ไขปัญหาเหตุการณ์จราจรได้อย่างมีประสิทธิภาพ

จากงานวิจัยของ Chunjiao Dong และคณะ [8] ใช้การรวมวิธีระหว่าง ซัพพอร์ตเวกเตอร์รีเกรสชัน และ State-Space Model (SSM) พยากรณ์การเกิดอุบัติเหตุทางถนน จากการพยากรณ์พบว่าผลลัพธ์มีประสิทธิภาพที่ดีและมีค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยที่ต่ำเมื่อเปรียบเทียบกับวิธีการอื่น ๆ

จากงานวิจัย Junyou Zhang และคณะ [17] เน้นการศึกษาเกี่ยวกับ การตัดสินใจในการขับขี่ของรถยนต์อัตโนมัติซึ่งเป็นปัจจัยสำคัญในการให้ความปลอดภัยในการขับขี่ โดยนำเสนอวิธี ซัพพอร์ตเวกเตอร์รีเกรสชัน ในการวิเคราะห์ข้อมูลถนนและการออกแบบกลไกการตัดสินใจในการขับขี่ของรถยนต์อัตโนมัติ ผลการวิจัยแสดงให้เห็นถึงความสามารถในการปรับปรุงการตัดสินใจในการขับขี่โดยพิจารณาถึงเงื่อนไขของถนน แบบจำลองจากวิธีซัพพอร์ตเวกเตอร์รีเกรสชันที่ถูกปรับปรุงมีประสิทธิภาพดีกว่าโมเดลอื่น ๆ และสภาพถนนมีผลกระทบมากที่สุดต่อการตัดสินใจในการขับขี่ในสภาพจราจรที่หนาแน่นต่ำ ซึ่งมีความสำคัญในการพัฒนาระบบขับขี่อัตโนมัติในสภาพทางเมืองที่ซับซ้อน

จากงานวิจัยของ Nidhi Nidhi และ DK Lobiyal [10] ศึกษาการพยากรณ์การไหลของการจราจรในพื้นที่มหาวิทยาลัย Jawaharlal Nehru (JNU) ที่ตั้งอยู่ใน New Delhi ประเทศอินเดีย และได้สร้างแบบจำลองการพยากรณ์โดยใช้วิธีซัพพอร์ตเวกเตอร์รีเกรสชัน เพื่อช่วยในการจัดการการจราจรและลดการแออัดในพื้นที่มหาวิทยาลัย โดยใช้ข้อมูลการจราจรเรียลไทม์จากประตูทางเหนือของวิทยาเขต ผลการวิจัยพบว่า การใช้วิธีซัพพอร์ตเวกเตอร์รีเกรสชัน ทำให้แบบจำลองมีผลลัพธ์ของการพยากรณ์ที่ดี และให้ค่ารากที่สองของค่าคลาดเคลื่อนกำลังสองเฉลี่ย และค่าคลาดเคลื่อนค่าสัมบูรณ์เฉลี่ยต่ำสุดในชุดข้อมูลฝึกฝน นอกจากนี้แบบจำลองได้ถูกนำมาใช้เพื่อทดสอบความแม่นยำของการพยากรณ์การไหลสำหรับการจราจรทั้งเข้า และออกของวิทยาเขตนี้ในแต่ละวันของเดือนมกราคม พ.ศ. 2556

4 ผลการศึกษา

ในหัวข้อนี้ผู้วิจัยได้นำกระบวนการ CRISP-DM มาประยุกต์ใช้กับข้อมูลอุบัติเหตุบนโครงข่ายถนนของกระทรวงคมนาคมในประเทศไทย เดือนมกราคม พ.ศ. 2562 ถึง เดือนมกราคม พ.ศ. 2566 ของจังหวัดนครราชสีมา และเปรียบเทียบประสิทธิภาพการพยากรณ์จำนวนผู้เสียชีวิตจากการเกิดอุบัติเหตุบนโครงข่ายถนนของกระทรวงคมนาคมของแบบจำลองที่ใช้เทคนิควิธี การถดถอยเชิงเส้น โครงข่ายประสาทเทียมแบบเพอร์เซฟตรอนหลายชั้น และซัพพอร์ตเวกเตอร์รีเกรสชัน

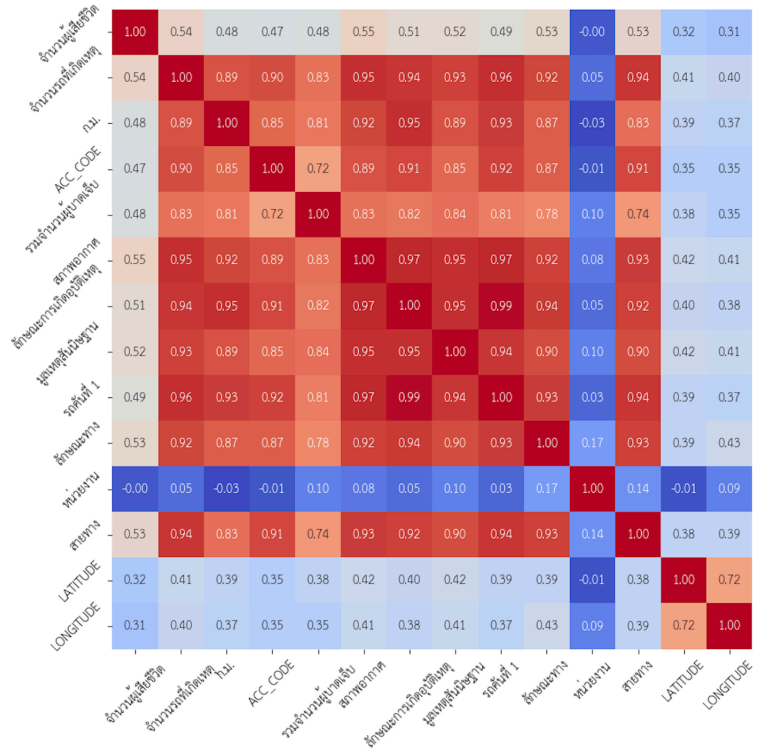
กระบวนการ CRISP-DM สำหรับข้อมูลอุบัติเหตุบนโครงข่ายถนนของกระทรวงคมนาคม ประกอบด้วย 6 ขั้นตอน

1. การทำความเข้าใจธุรกิจ (Business Understanding)
   การศึกษานี้ เป็นการเปรียบเทียบประสิทธิภาพของแบบจำลอง เพื่อประเมินและพยากรณ์จำนวนผู้ที่สูญเสียชีวิตจากการเกิดอุบัติเหตุบนเส้นทางการจราจร ภายใต้การดูแลของกระทรวงคมนาคมในพื้นที่จังหวัดนครราชสีมา ผลลัพธ์จากการพยากรณ์นี้จะทำให้เราเข้าใจถึงแนวโน้มและจำนวนของการเกิดอุบัติเหตุที่อาจเกิดขึ้นในอนาคต ซึ่งจะเป็นข้อมูลสำคัญที่จะช่วยให้หน่วยงานที่เกี่ยวข้องสามารถวางแผน และดำเนินการตามมาตรการป้องกันอย่างมีประสิทธิภาพ

2. การทำความเข้าใจข้อมูล (Data Understanding)
   การศึกษานี้เป็นการนำข้อมูลของการเกิดอุบัติเหตุบนโครงข่ายถนนของกระทรวงคมนาคมในประเทศไทย จากแหล่งข้อมูลสาธารณะ https://datagov.mot.go.th/dataset/roadaccident ของศูนย์เทคโนโลยีสารสนเทศและการสื่อสาร สำนักงานปลัดกระทรวงคมนาคม ตั้งแต่ วันที่ 5 เดือน มกราคม พ.ศ. 2562 ถึง วันที่ 31 เดือน มกราคม พ.ศ. 2566 ทางผู้วิจัยได้นำข้อมูลของจังหวัดนครราชสีมา มาศึกษาและวิเคราะห์ตัวแปรและสาเหตุการเสียชีวิตจากการเกิดอุบัติเหตุ เพื่อสร้างแบบจำลองที่มีประสิทธิภาพในการพยากรณ์จำนวนผู้เสียชีวิตจากการเกิดอุบัติเหตุ ข้อมูลมีจำนวนทั้งหมด 3,330 แถว 20 คอลัมน์ คือ ปีที่เกิดอุบัติเหตุ วันที่เกิดอุบัติเหตุ เวลาที่เกิดอุบัติเหตุ วันที่รายงานผลอุบัติเหตุ เวลาที่รายงานผลอุบัติเหตุ เลขที่อ้างอิง พื้นที่สังกัดหน่วยงาน เส้นที่กิโลเมตรที่ จังหวัดที่เกิดเหตุ รถคันที่ 1 บริเวณที่เกิดอุบัติเหตุ สาเหตุมาจาก ลักษณะการเกิดอุบัติเหตุ จำนวนรถที่เกิดอุบัติเหตุ จำนวนผู้เสียชีวิต รวมจำนวนได้รับผู้บาดเจ็บ สภาพอากาศสถานที่เกิดเหตุ สายทาง LATITUDE LONGITUDE

3. การเตรียมข้อมูล (Data Preparation)
   จากชุดข้อมูลในขั้นตอนที่ 2 เราจะการเตรียมข้อมูลด้วยขั้นตอนต่อไปนี้
   3.1. การทำความสะอาดข้อมูล จากชุดข้อมูล มีข้อมูลที่ขาดหายไป (Missing Value) 245 ข้อมูล เราจะจัดการกับข้อมูลที่ขาดหายไป โดยจะลบแถวที่มีข้อมูลที่ขาดหายไป ซึ่งจะทำให้เหลือแถวที่นำมาใช้ในการวิเคราะห์ข้อมูลเพื่อพัฒนาแบบจำลองทั้งหมด 3065 แถว และทางผู้วิจัยได้ลบคอลัมน์ ปีที่เกิดเหตุ วันที่เกิดเหตุ เวลา วันที่รายงาน เวลาที่รายงาน และจังหวัด ออก เหลือ 14 คอลัมน์

   3.2. การวิเคราะห์ระดับความสัมพันธ์ของตัวแปร ผ่านเทคนิคการหาค่าสัมประสิทธิ์สหสัมพันธ์ จากการสำรวจข้อมูลจะมีคอลัมน์ที่เป็นข้อมูลชนิดสตริง (string) ประเภทหมวดหมู่ (categorical data) ได้แก่ หน่วยงาน สายทาง รถคันที่ 1 บริเวณที่เกิดเหตุ/ลักษณะทาง มูลเหตุสันนิษฐาน ลักษณะการเกิดอุบัติเหตุ สภาพอากาศ ผู้วิจัยได้การทำแปลงข้อมูลจากคอลัมน์เหล่านี้โดยใช้ Ordinal Encoder ในเครื่องมือ ColumnTransformer ในไลบรารี scikit-learn ของภาษา Python ก่อนนำไปหาค่าสัมประสิทธิ์สหสัมพันธ์ ผลการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ แสดงดังภาพที่ 3



ภาพที่ 3: ค่าความสัมพันธ์ของข้อมูล

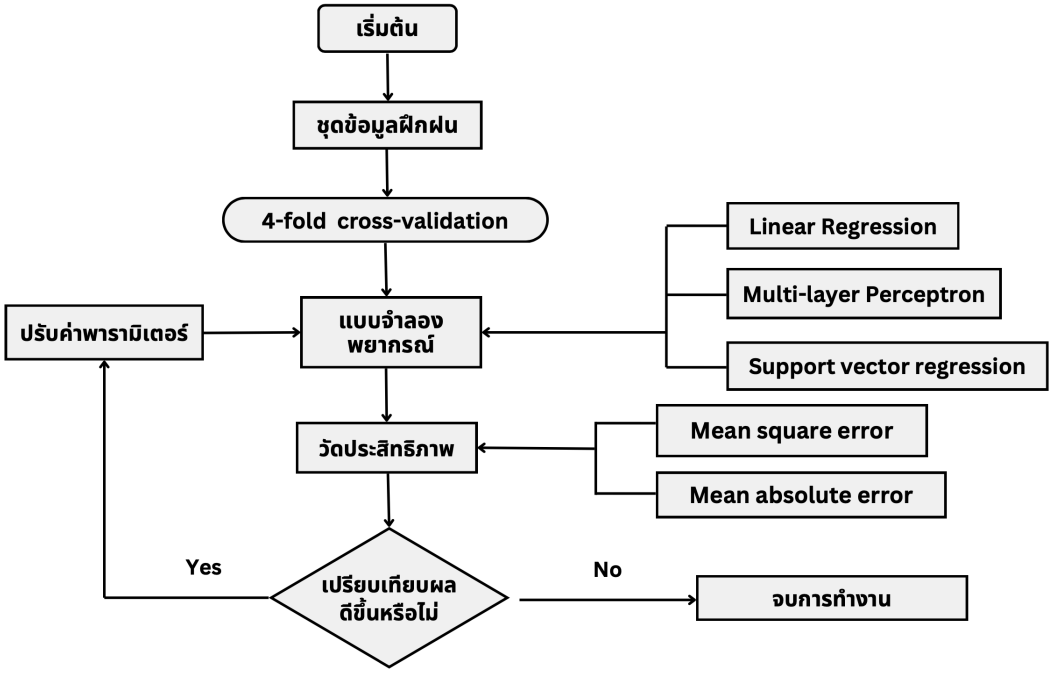
จากผลที่แสดงในดังภาพที่ 3 เราจะทำงานลบคอลัมน์ ”หน่วยงาน” ออก เนื่องจากมีค่าสัมประสิทธิ์สหสัมพันธ์กับจำนวนผู้เสียชีวิตอยู่ในระดับน้อย

3.3. กำหนดตัวแปรต้นและตัวแปรตาม
ตัวแปรตาม คือ จำนวนผู้เสียชีวิต
ตัวแปรต้น คือ จำนวนรถที่เกิดเหตุ เลขที่อ้างอิง รวมจำนวนผู้บาดเจ็บ สภาพอากาศ ลักษณะการเกิดอุบัติเหตุ สาเหตุมาจาก รถคันที่ 1 บริเวณที่เกิดอุบัติเหตุ เส้นทางกิโลที่ สายทาง LATITUDE LONGITUDE

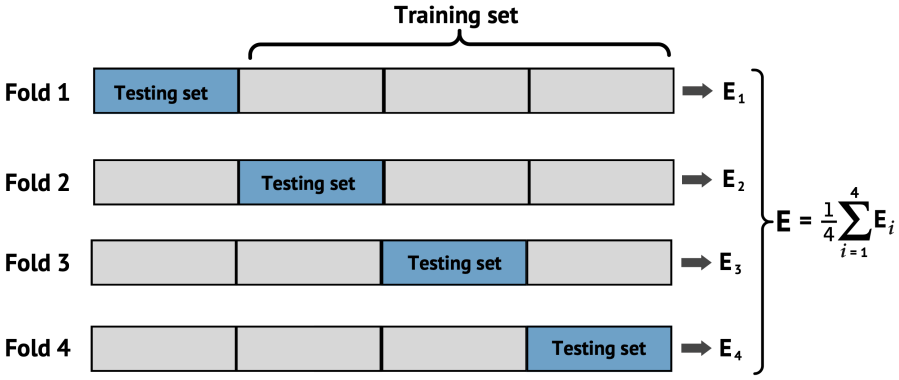
3.4. การแบ่งชุดข้อมูล โดยแบ่งชุดข้อมูลออกเป็น 2 ส่วน (80:20)
- ชุดข้อมูลฝึกฝน (Training Data Set) ใช้ข้อมูล 5 ม.ค. 62 - 30 เม.ย 65 ทั้งหมด 2,452 ข้อมูล
- ชุดข้อมูลทดสอบ (Testing Data Set) ใช้ข้อมูล 1 พ.ค. 65 – 31 ม.ค 66 ทั้งหมด 613 ข้อมูล

3.5. การแปลงข้อมูล (Data transformation)
จากชุดข้อมูล ทางผู้วิจัยมีการปรับค่าของข้อมูลในชุดข้อมูลฝึกฝนให้มีสเกลเดียวกัน ก่อนนำเข้าสู่แบบจำลองโดยใช้วิธีการ Min-Max Scaling

4. การสร้างแบบจำลองวิเคราะห์ข้อมูล (Modeling)
จากข้อมูลที่ผ่านการเตรียมข้อมูล (Data Preparation) และถูกแบ่งออกเป็น 2 ส่วน ได้แก่ ชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบ จากขั้นตอนที่ 3 ในขั้นตอนนี้ผู้วิจัยสร้างแบบจำลองโดยใช้วิธี การถดถอยเชิงเส้น (LR) โครงข่ายประสาทเทียมแบบเพอร์เซฟตรอนหลายชั้น (MLP) และซัพพอร์ตเวกเตอร์รีเกรสชัน (SVR) โดยจะมีการปรับค่าพารามิเตอร์เพื่อให้ได้แบบจำลองที่มีประสิทธิภาพที่ดีที่สุด



ภาพที่ 4: ขั้นตอนการออกแบบและพัฒนาแบบจำลองพยากรณ์



ภาพที่ 5: 4-fold Cross-Validation

จากภาพที่ 4 เป็นการแสดงขั้นตอนการออกแบบและพัฒนาแบบจำลองพยากรณ์ โดยเริ่มจากการแบ่งข้อมูลชุดฝึกฝนด้วยวิธี Cross-validation ตามภาพที่ 5 นำข้อมูลมาสร้างการเรียนรู้ให้กับการเรียนรู้ของเครื่อง โดยใช้วิธี การถดถอยเชิงเส้น โครงข่ายประสาทเทียมแบบเพอร์เซฟตรอนหลายชั้น และซัพพอร์ตเวกเตอร์รีเกรสชัน สร้างตัวแบบจำลองพยากรณ์ แล้ววัดประสิทธิภาพของแบบจำลองพยากรณ์ด้วยค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (MEAN-MSE) และค่าเฉลี่ยของค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MEAN-MAE) นำผลไปวัดผลประสิทธิภาพเพื่อใช้เปรียบเทียบกับตัวแบบถัดไป และจากการปรับค่าพารามิเตอร์ เราได้ค่าพารามิเตอร์ของวิธีทั้ง 3 วิธี ดังนี้

4.1. การถดถอยเชิงเส้น (LR) ไม่มีการปรับค่าพารามิเตอร์

4.2. โครงข่ายประสาทเทียมแบบเพอร์เซฟตรอนหลายชั้น (MLP) กำหนดให้ มีจำนวนชั้นซ่อน คือ 1000 แต่ละชั้นมีจำนวนโหนด คือ 1 และให้มีการหมุนวนซ้ำ 1000 รอบ

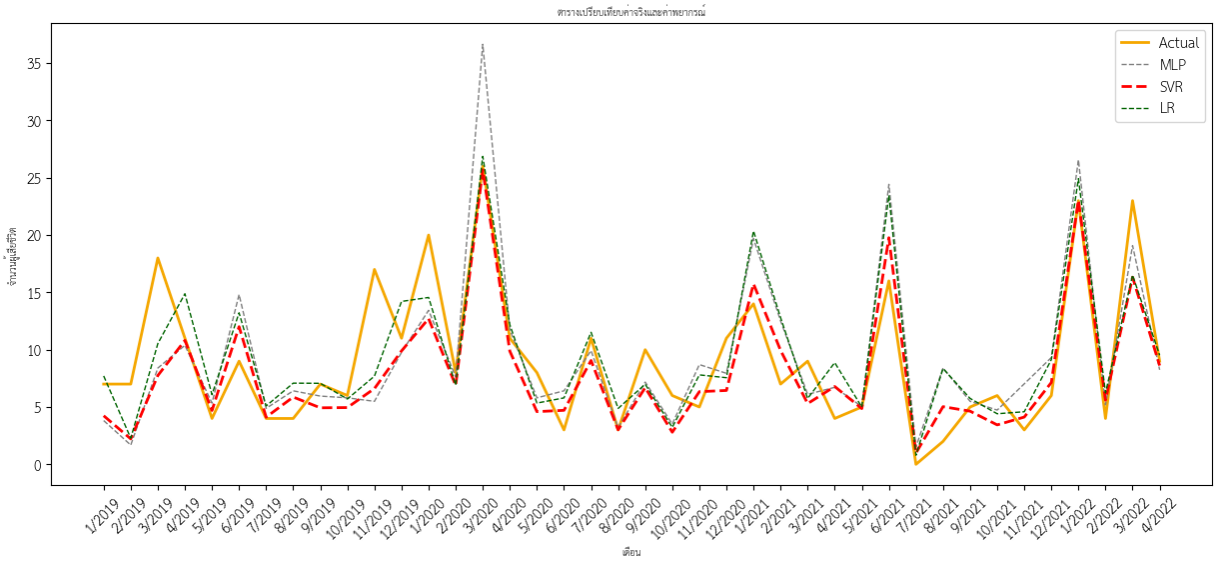
4.3. ซัพพอร์ตเวกเตอร์รีเกรสชัน (SVR) กำหนดให้ เคอร์เนลฟังก์ชัน เป็น เกาส์เซียนเคอร์เนล (Gaussian Kernel) ที่มี stopping criterion = 0.001 C=2 และ epsilon = 0.15

ตารางที่ 1: การเปรียบเทียบประสิทธิภาพของแบบจำลอง

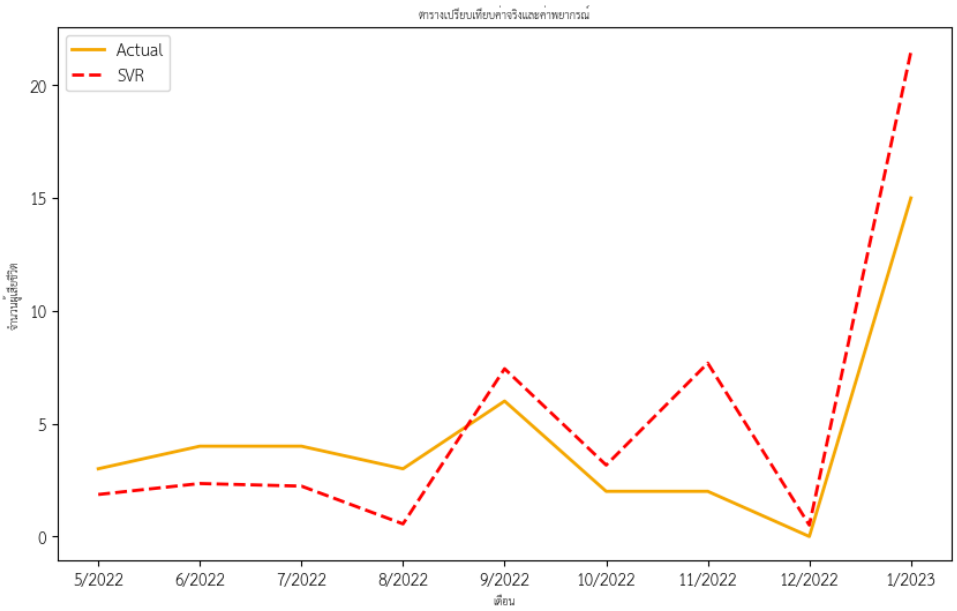
Model	MEAN-MSE	MEAN-MAE
Support Vector Regression	0.178138	0.175904
Linear Regression	0.187041	0.238318
Multi-Layer Perceptron	0.229159	0.253268

5. การประเมินผลลัพธ์ (Evaluation) การประเมินประสิทธิภาพของแบบจำลองจากชุดข้อมูลฝึกฝนที่ทำการแบ่งข้อมูลเป็นชุดฝึกฝนและชุดตรวจสอบหรือทดสอบ(validation) ในรูปแบบ 4-fold และนำค่าที่ได้มาทำการคิดค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย และค่าเฉลี่ยของค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย จะได้ผลลัพธ์ตามตารางที่ 1 จากผลลัพธ์จะเห็นได้ว่าแบบจำลองที่ใช้วิธี ซัพพอร์ตเวกเตอร์รีเกรสชัน ให้หาค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย และค่าเฉลี่ยของค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ยน้อยที่สุด คือ 0.178138 และ 0.175904 ตามลำดับ เมื่อเปรียบเทียบกับแบบจำลองที่ใช้วิธี การถดถอยเชิงเส้น และโครงข่ายประสาทเทียมแบบเพอร์เซฟตรอนหลายชั้น และจากผลลัพธ์ที่ได้สามารถนำมาแสดงเป็นกราฟเปรียบเทียบประสิทธิภาพของแบบจำลอง ตามภาพที่ 6 จากกราฟจะเห็นได้ว่า แบบจำลองที่ใช้เทคนิควิธีซัพพอร์ตเวกเตอร์รีเกรสชัน สามารถพยากรณ์จำนวนผู้เสียชีวิตได้ใกล้เคียงกับค่าจริงได้ดีกว่าแบบจำลองที่ใช้วิธี การถดถอยเชิงเส้น และโครงข่ายประสาทเทียมแบบเพอร์เซฟตรอนหลายชั้น

6. การอธิบายผลและการนำไปใช้งานจริง (Deployment)
จากการประเมินประสิทธิภาพจะเห็นได้ว่าแบบจำลองจากวิธีซัพพอร์ตเวกเตอร์รีเกรสชัน เป็นแบบจำลองที่ให้ประสิทธิภาพที่ดีที่สุดเมื่อเปรียบเทียบกับแบบจำลองอื่น เราจึงได้นำแบบจำลองมาใช้กับชุดข้อมูลทดสอบ แสดงกราฟเปรียบเทียบค่าจริงและค่าพยากรณ์ของแบบจำลอง ดังภาพที่ 7 และมีค่าคลาดเคลื่อนกำลังสองเฉลี่ย คือ 0.09177 จากแบบจำลองที่พัฒนาขึ้นมา สามารถนำไปใช้พยากรณ์จำนวนผู้เสียชีวิตจากข้อมูลที่ได้รับเกี่ยวกับการเกิดอุบัติเหตุในอนาคต และใช้ผลลัพธ์จากการพยากรณ์เพื่อวางแผนดำเนินการป้องกันและลดการเกิดอุบัติเหตุบนถนน



ภาพที่ 6: กราฟเปรียบเทียบประสิทธิภาพของแบบจำลองโดยใช้วิธี การถดถอยเชิงเส้น (LR) โครงข่ายประสาทเทียมแบบเพอร์เซฟตรอนหลายชั้น (MLP) และซัพพอร์ตเวกเตอร์รีเกรสชัน (SVR) ของชุดข้อมูลฝึกฝน



ภาพที่ 7: กราฟเปรียบเทียบค่าจริงและค่าพยากรณ์ของแบบจำลองของชุดข้อมูลทดสอบ

5 การอภิปรายผลและการเสนอแนะ

จากการศึกษา และวิเคราะห์เปรียบเทียบประสิทธิภาพของอัลกอริทึมการเรียนรู้ของเครื่องในการพยากรณ์จำนวนผู้เสียชีวิตจากการเกิดอุบัติเหตุบนโครงข่ายถนนของกระทรวงคมนาคม ซึ่งเป็นข้อมูลของศูนย์เทคโนโลยีสารสนเทศ และการสื่อสาร สำนักงานปลัดกระทรวงคมนาคม จากแหล่งข้อมูลสาธารณะ https://datagov.mot.go.th/dataset/roadaccident ตั้งแต่เดือนมกราคม พ.ศ.2562 ถึงเดือนมกราคม พ.ศ. 2566 ของจังหวัดนครราชสีมา โดยสร้างแบบจำลอง ด้วยวิธีที่แตกต่างกัน 3 วิธี ได้แก่ การถดถอยเชิงเส้น (LR) โครงข่ายประสาทเทียมแบบเพอร์เซฟตรอนหลายชั้น (MLP) และซัพพอร์ตเวกเตอร์รีเกรสชัน (SVR) แบ่งข้อมูลออกเป็นชุดฝึกฝนและชุดทดสอบ สร้างแบบจำลองด้วยชุดข้อมูลฝึกฝน และวัดผลการเปรียบเทียบประสิทธิภาพการพยากรณ์การเกิดอุบัติเหตุโดยใช้ ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย และค่าเฉลี่ยของค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย พบว่า วิธี LR มีค่าของการเสียชีวิตเฉลี่ย 0.187041 และ 0.238318 วิธี MLP มีค่าของการเสียชีวิตเฉลี่ยอยู่ที่ 0.229159 และ 0.253268 วิธี SVR มีค่าของการเสียชีวิตเฉลี่ยอยู่ที่ 0.178138 และ 0.175904 การสรุปผลการวิเคราะห์พบว่า วิธี SVR เป็นวิธีที่มีค่าความคลาดเคลื่อนต่ำที่สุด ต่อมาผู้วิจัยได้นำชุดข้อมูลทดสอบวัดผลพยากรณ์จำนวนผู้เสียชีวิตจากชุดข้อมูลทดสอบได้ค่าคลาดเคลื่อนกำลังสองเฉลี่ย 0.0917 จะเห็นได้ว่า วิธี SVR มีความเหมาะสมในการพัฒนาแบบจำลองเพื่อพยากรณ์จำนวน ผู้เสียชีวิตและใช้ผลลัพธ์จากการพยากรณ์เพื่อวางแผนดำเนินการป้องกันและลดการเกิดอุบัติเหตุบนถนน ดังนั้น ข้อเสนอแนะเพื่อให้ผลการพยากรณ์มีความถูกต้องแม่นยำมากขึ้น ผู้วิจัยควรพิจารณาปัจจัยแวดล้อมอื่น ๆ ที่มี ผลต่อความเสี่ยงต่อการเสียชีวิตของการเกิดอุบัติเหตุ เช่น การดื่มแอลกอฮอล์ พักผ่อนไม่เพียงพอ บริเวณที่เกิดเหตุ ทางโค้งอันตราย เป็นต้น และจากการศึกษาวิธีการพยากรณ์ จำนวน 3 วิธี ในงานวิจัยนี้ หากมีการศึกษาวิธีอื่นเข้ามาเปรียบเทียบเพิ่มเติม จะทำให้ได้ค่าการพยากรณ์ที่ต่างกันในแต่ละวิธี และจะสามารถหาแบบจำลองที่มีการพยากรณ์แม่นยำมากยิ่งขึ้น

กิตติกรรมประกาศ คณะผู้วิจัยขอขอบคุณผู้ทรงคุณวุฒิทุกท่านที่ได้ให้ข้อคิดเห็นและข้อเสนอแนะต่าง ๆ เพื่อปรับปรุงบทความวิจัย และขอขอบคุณ ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา สำหรับทุนสนับสนุนการทำวิจัยในครั้งนี้

เอกสารอ้างอิง

[1] ปทิตญา บุญรักษา และจารี ทองคำ. (2017). การเปรียบเทียบประสิทธิภาพของแบบจำลองการเกิดอุบัติเหตุทางถนน โดยใช้ เทคนิคอนุกรมเวลา. *วารสารวิชาการการจัดการเทคโนโลยี มหาวิทยาลัยราชภัฏมหาสารคาม*, *4*(2), 39-46

[2] กลุ่มสถิติสารสนเทศ, *รายงานประจำปี 2565 อุบัติเหตุจราจรบนทางหลวงแผ่นดิน*, สำนักอำนวยความปลอดภัยกรมทางหลวง ปี 2566

[3] รณชัย ชื่นธวัช กิตติศักดิ์ เกิดประสพ และนิตยา เกิดประสพ. (2560). การพยากรณ์ความต้องการใช้งานหน่วยจำหน่ายไฟฟ้าด้วยซัพพอร์ตเวกเตอร์รีเกรสชันแบบตรวจสอบสลับ 3 ส่วน. *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยอุบลราชธานี*, 19(1), 216-232.

[4] ณัฐพล กวีกิจกำพล, ธนัตถ์ เรือนน้อย ภาณุ ฐิติมโนรักษ์ และพรทิพย์ เดชพิชัย. (2562). การเปรียบเทียบตัวแบบการพยากรณ์ปริมาณการใช้ไฟฟ้าของกรุงเทพมหานคร. *วารสารวิทยาศาสตร์รำไพพรรณี*, *1*(1), 26-33.

[5] สุกิจ อินทร์เจริญ. (2556). ปัจจัยการปฏิบัติงานการบริหารงานก่อสร้าง ของสำนักการช่างเทศบาลนครปากเกร็ด การศึกษาปัญหาการบริหารงานก่อสร้าง กรณีศึกษาเทศบาลนครปากเกร็ด จังหวัดนนทบุรี. *สารนิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาวิชาการบริหารงานก่อสร้าง คณะสถาปัตยกรรมศาสตร์ มหาวิทยาลัยศรีปทุม*

[6] Artificial Neural Networks and its Applications สืบค้นจาก https://www.geeksforgeeks.org/artificial-neural-networks-and-its-applications/

[7] Bagheripour, P., Gholami, A., Asoodeh, M., and and Asadi, M.V. (2015). Support vector regression based determination of shear wave velocity, *Journal of Petroleum Science and Engineering*, 2015(25), 95-99.

[8] Dong, C., Xie, K., Sun, X., Lyu, M., and Yue, H. (2019). Roadway traffc crash prediction using a state-space model based support vector regression approach. *PLOS ONE*, 14(4). https://doi.org/10.1371/journal.pone.0214866

[9] Forthofer, Ronald N., Lee, Eun S. and Hermandez, M.(2007). Biostatistics: A Guide to Design, Analysis, and Discovery. 2^{nd}*.ed California: Elsevier Academic Press*.

[10] Nidhi, N., Lobiyal, D. K. (2022). Traffc fow prediction using support vector regression. *International Journal of Information Technology*, 14(5), 619–626.

[11] Qing-wei, Z., Ai-Ying, F., and Zhi-Hai, X. (2009). Application of support vector regression and particle swarm optimization in traffc accident forecasting. *International Conference on Information Management, Innovation Management and Industrial Engineering, Xi'an, China*, (pp. 188-191). Doi: 10.1109/ICIII.2009.506

[12] Shalabi, L.A., Zyad, S. and Basel, K. (2006). Data mining: A preprocessing engine. *Journal of Computer Science*, 2(9), 735.

[13] Shearer, C. (2000). The CRISP-DM model: the new blueprint for data mining. *J. Data Warehousing*, 5(4) 13–22.

[14] Wu, D., and Wang, S. (2020). Comparison of road traffc accident prediction effects based on SVR and BP neural network. *IEEE International Conference on Information Technology, Big Data and Artifcial Intelligence (ICIBA), Chongqing, China* (pp. 1150-1154). Doi: 10.1109/ICIBA50161.2020.9277150.

[15] Wu, W., Chen, S., and Zheng, C. (2011). Traffc incident durationprediction based on support vector regression. *In 11th International Conference of Chinese Transportation Professionals (ICCTP), Nanjing, China* (pp. 2412-2421). https://doi.org/10.1061/41186(421)241

[16] Wirth, R. and Hipp, J. (2000) CRISP-DM: Towards a standard process model for data mining, *Citeseer in Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, (pp. 29-39)

[17] Zhang, J., Liao, Y., Wang, S., Han, J. (2017). Study on driving decision-making mechanism of autonomous vehicle based on an optimized support vector machine regression. *Applied Sciences*, *8*(1), 12-22. https://doi.org/10.3390/app8010013

7. Differential Equations and Numerical Mathematics

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

วิธีการสปริทเบรกแมนสำหรับกำจัดสัญญาณรบกวนแบบการคูณออกจากภาพดิจิทัล

โสภิดา สุขญาณกิจ[1,†] และ ศิริวรรณ จันทร์แก่น[2,†,‡]

[1]สาขาวิชาคณิตศาสตร์ประยุกต์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏพระนครศรีอยุธยา 13000

[2]สาขาวิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏกาญจนบุรี 71190

บทคัดย่อ

เนื่องจากการที่ภาพถ่ายปรากฏสัญญาณรบกวนนั้นเป็นสิ่งที่หลีกเลี่ยงไม่ได้ กระบวนการซ่อมแซมภาพจึงเข้ามามีบทบาทที่สำคัญ ในงานวิจัยนี้ ผู้วิจัยได้นำเสนอตัวแบบเชิงการแปรผันจำนวน 2 ตัวแบบ คือ ตัวแบบ JYTL และตัวแบบ JYBH สำหรับกำจัดสัญญาณรบกวนแบบการคูณออกจากภาพ ซึ่งใช้ข้อดีของตัวแบบ JY และอนุพันธ์อันดับสูง เพื่อลดปรากฏการณ์ขั้นบันได พร้อมทั้งวิธีการสปริทเบรกแมนซึ่งเป็นวิธีการเชิงตัวเลขที่มีประสิทธิภาพในการแก้ปัญหา ผลการทดลองเชิงตัวเลขแสดงให้เห็นว่าตัวแบบที่ได้นำเสนอพร้อมด้วยวิธีการสปริทเบรกแมนดังกล่าวสามารถกำจัดสัญญาณรบกวนออกจากภาพอย่างแม่นยำ และให้คุณภาพของภาพผลลัพธ์ที่ดีขึ้น โดยตัวแบบที่ได้นำเสนอคือตัวแบบ JYBH ให้ความแม่นยำสูงกว่าตัวแบบ JY และตัวแบบ JYTL ในทุกกรณี นอกจากนี้ผู้วิจัยได้ทำการทดสอบความมีประสิทธิภาพของตัวแบบที่ได้นำเสนอพร้อมด้วยวิธีการสปริทเบรกแมนในการกำจัดสัญญาณออกจากภาพถ่ายทางการแพทย์ ผลการทดสอบพบว่าตัวแบบที่นำเสนอสามารถกำจัดสัญญาณรบกวนออกจากภาพได้อย่างมีประสิทธิภาพ

คำสำคัญ: การซ่อมแซมภาพ, การกำจัดสัญญาณรบกวนออกจากภาพ, ตัวแบบเชิงการแปรผัน, วิธีการสปริทเบรกแมน, สัญญาณรบกวนแบบการคูณ

2020 MSC: ปฐมภูมิ 65N22 ทุติยภูมิ 68U10

[†]ผู้นำเสนอ [‡]ผู้แต่งหลัก

อีเมล: sopida.jew@aru.ac.th (โสภิดา สุขญาณกิจ), siriwan.c@kru.ac.th (ศิริวรรณ จันทร์แก่น).

1 บทนำ

ปัญหาการกำจัดสัญญาณรบกวนออกจากภาพ (Image Denoising Problems) เป็นปัญหาการประมวลผลภาพ (Image Processing Problems) ที่ได้รับความสนใจเป็นอย่างมาก เนื่องจากการที่ภาพถ่ายปรากฏสัญญาณรบกวนนั้นเป็นสิ่งที่หลีกเลี่ยงไม่ได้ โดยสัญญาณรบกวนเหล่านี้อาจเกิดจากกระบวนการสร้างภาพ การบันทึกภาพ หรือการรับ-ส่งภาพ เป็นต้น การประยุกต์การซ่อมแซมภาพมีความจำเป็นในหลายสาขา เช่น การประยุกต์ทางด้านศิลปะ ฟิสิกส์ดาราศาสตร์ ชีววิทยา เคมี ฟิสิกส์ ธรณีฟิสิกส์ อาชญาวิทยา และศาสตร์แขนงอื่น ๆ ที่เกี่ยวข้องกับการใช้และสร้างภาพถ่าย นอกจากนี้ปัญหาการประมวลผลภาพยังได้รับความสนใจอย่างมากทางการแพทย์ เนื่องจากภาพถ่ายคลื่นเสียงความถี่สูง (Ultrasound Images) เป็นภาพถ่ายที่ได้รับความนิยมนำมาตรวจวินิจฉัยโรคอย่างแพร่หลาย แต่ภาพดังกล่าวมักปรากฏสัญญาณรบกวน ซึ่งส่งผลให้การแปลความหมายจากภาพถ่ายคลื่นเสียงความถี่สูงมีความคลาดเคลื่อน ดังนั้นการกำจัดสัญญาณรบกวนที่เกิดขึ้นในภาพดังกล่าวจึงมีความจำเป็นอย่างยิ่งในการประยุกต์ทางการแพทย์

โดยทั่วไปตัวแบบสัญญาณรบกวนสามารถแบ่งเป็น 2 ประเภท คือ ตัวแบบสัญญาณรบกวนแบบการบวก (Additive Noise Model) และ ตัวแบบสัญญาณรบกวนแบบการคูณ (Multiplicative Noise Model) [1] ตัวแบบสัญญาณรบกวนแบบการบวกมักปรากฏในภาพถ่ายซึ่งเกิดจากกระบวนการบันทึกสัญญาณภาพด้วยเครื่องมือดิจิทัล ตัวแบบสัญญาณรบกวนแบบการคูณหรือตัวแบบสัญญาณรบกวนแบบสเป็กเคิล (Speckle Noise Model) มักถูกพบในภาพถ่ายคลื่นเสียงความถี่สูง ภาพเลเซอร์ และภาพจากระบบเรดาร์ที่ติดตั้งบนเครื่องบินหรือดาวเทียม ซึ่งเป็นภาพที่ได้จากระบบการสร้างภาพแบบโคฮีเลนท์ (Coherent Imaging System)

ในการกำจัดสัญญาณรบกวนออกจากภาพได้มีการศึกษาและนำเสนอเทคนิควิธีการต่าง ๆ โดยสามารถแบ่งออกได้เป็น 2 กลุ่ม คือ กลุ่มงานวิจัยด้านการกำจัดสัญญาณรบกวนภาพแบบใช้อัลกอริทึม และกลุ่มงานวิจัยด้านการกำจัดสัญญาณรบกวนของภาพแบบใช้ชุดข้อมูลมาฝึกสอนให้กับโมเดล [2] กลุ่มงานวิจัยด้านการกำจัดสัญญาณรบกวนของภาพแบบใช้ชุดข้อมูลมาฝึกสอนให้กับโมเดล ได้แก่ วิธีการเรียนรู้แบบอัตโนมัติด้วยการเลียนแบบการทำงานของโครงข่ายประสาทของมนุษย์ วิธีการเวฟเลต วิธีการสโตรแคสติก วิธีการที่ใช้วิธีการวิเคราะห์องค์ประกอบหลัก และกลุ่มงานวิจัยด้านการกำจัดสัญญาณรบกวนภาพแบบใช้อัลกอริทึม ได้แก่ วิธีการเชิงการแปรผัน [3] จากการศึกษาพบว่าวิธีหนึ่งที่เป็นเทคนิควิธีการทางคณิตศาสตร์ที่น่าเชื่อถือและมีความแม่นยำสูงมาก คือ วิธีการเชิงการแปรผัน (Variational Method) โดยแนวคิดในการหาคำตอบเริ่มต้นจากการพิจารณาภาพเป็นฟังก์ชัน สร้างตัวแบบเชิงการแปรผันสำหรับกำจัดสัญญาณรบกวนออกจากภาพ จากนั้นใช้แคลคูลัสของการแปรผัน (Calculus of Variations) ในการสร้างสมการออยเลอร์-ลากรานจ์ที่สมนัยกับตัวแบบดังกล่าว และใช้เทคนิควิธีการเชิงตัวเลขสำหรับแก้สมการออยเลอร์-ลากรานจ์อย่างมีประสิทธิภาพ โดยทั่วไปสมการออยเลอร์-ลากรานจ์ที่ได้จากตัวแบบเชิงการแปรผันมักเป็นสมการเชิงอนุพันธ์ย่อยไม่เป็นเชิงเส้น จากแนวคิดในการหาคำตอบข้างต้นพบว่าแนวทางที่ใช้ในการศึกษาแก้ปัญหาการกำจัดสัญญาณรบกวนออกจากภาพสามารถแบ่งออกได้เป็น 2 แนวทาง คือ การสร้างตัวแบบทางคณิตศาสตร์สำหรับกำจัดสัญญาณรบกวนออกจากภาพที่มีความน่าเชื่อถือและแม่นยำ ซึ่งได้ถูกนำเสนอครั้งแรกโดยคณะวิจัยของ Rudin ในปี ค.ศ. 1992 [4] และการพัฒนาวิธีการเชิงตัวเลขที่มีประสิทธิภาพสูงและรวดเร็วสำหรับแก้สมการออยเลอร์-ลากรานจ์ที่สมนัยกับตัวแบบ

Goldstein และ Osher [5] ได้นำเสนอแนวคิดในการแก้ปัญหาเชิงการแปรผันโดยวิธีการสปริทเบรกแมน (Split Bregmam (SB) Method) แนวทางในการหาคำตอบของวิธีการ SB คือ การแนะนำตัวแปรเสริมเพื่อแปลงปัญหาที่ซับซ้อนเป็นปัญหาย่อย และใช้กระบวนการทำซ้ำแบบสลับเพื่อหาคำตอบ ซึ่งเป็นวิธีการที่มีประสิทธิภาพสูงในการแก้ปัญหาเชิงการแปรผันและยังแสดงให้เห็นว่าวิธีการนี้ลู่เข้าสู่คำตอบได้อย่างรวดเร็วอีกด้วย

ในงานวิจัยนี้ผู้วิจัยทำการศึกษาและพัฒนาตัวแบบเชิงการแปรผันสำหรับกำจัดสัญญาณรบกวนแบบการคูณออกจากภาพถ่ายคลื่นเสียงความถี่สูง และนำเสนอวิธีการ SB ซึ่งเป็นเทคนิควิธีการเชิงตัวเลขที่มีประสิทธิภาพและรวดเร็วในการกำจัดสัญญาณรบกวนออกจากภาพ

2 ความรู้พื้นฐาน

ในขั้นตอนการสร้างตัวแบบเชิงการแปรผัน จะพิจารณาภาพเป็นฟังก์ชัน $I:\Omega \subset \mathbb{R}^2 \to V \subset [0,\infty)$ โดยที่โดเมนภาพ (Image Domain) Ω มีรูปร่างเป็นรูปสี่เหลี่ยม โดยกำหนดให้เรนจ์ของภาพ $R(I) \subset [0,\infty)$ เพื่อระบุว่า I เป็นภาพที่มีความเข้มของภาพ (Image Intensity) อยู่ในอัตราส่วนความเข้มของภาพในโทนสีเทา (Grayscale) [6, 7] กล่าวคือ ภาพ I เกี่ยวข้องกับแต่ละสมาชิก $x\epsilon\Omega$ ด้วยค่าความเข้มโทนสีเทา $I(x)\epsilon V$ ซึ่งในที่นี้สามารถสมมติได้โดยไม่เสียหลักการสำคัญว่า $\Omega = [1,M] \times [1,N] \subset \mathbb{R}^2$ และ $V = [0,255]$ เมื่อ M และ N เป็นจำนวนเต็มบวก

ในกระบวนการบันทึกสัญญาณภาพด้วยเครื่องมือดิจิทัลมักปรากฏสัญญาณรบกวนแบบการบวก [4] ในตัวแบบสัญญาณรบกวนแบบการบวก เรามีเป้าหมายเพื่อกู้คืนหรือซ่อมแซมภาพต้นฉบับ (ไม่ทราบ) $u:\Omega \to \mathrm{V}$ จากภาพที่มีสัญญาณรบกวน (ทราบ) $z:\Omega \to \mathrm{V}$ ซึ่งเจือปนด้วยสัญญาณรบกวนแบบการบวก ดังนี้

$$z = u + \eta \tag{1}$$

โดยทั่วไป η แทนสัญญาณรบกวนแบบเกาส์เซียนซึ่งมีค่าเฉลี่ยศูนย์

ในระบบการสร้างภาพแบบโคฮีเลนท์ที่สร้างภาพถ่ายคลื่นเสียงความถี่สูง และภาพเลเซอร์ มักพบสัญญาณรบกวนแบบการคูณหรือสัญญาณรบกวนแบบสเป็กเคิล ในตัวแบบสัญญาณรบกวนแบบการคูณ [8, 9] ภาพต้นฉบับถูกเจือปนด้วยสัญญาณรบกวนแบบการคูณ η ซึ่งถูกกำหนดโดย

$$z = u\eta \tag{2}$$

เราสามารถสมมติได้โดยไม่เสียหลักการสำคัญว่า $u,\eta > 0$ ในการกำจัดสัญญาณรบกวนชนิดนี้ออกจากภาพทำได้ค่อนข้างยากกว่าการกำจัดสัญญาณรบกวนแบบการบวก ทั้งนี้เนื่องจากการคูณระหว่างสัญญาณรบกวนและภาพต้นฉบับส่งผลให้มีความถี่สูง รวมทั้งการแจกแจงของสัญญาณรบกวนชนิดนี้ไม่เป็นแบบเกาส์เซียน โดยทั่วไป การแจกแจงของสัญญาณรบกวนชนิดนี้เป็นการแจกแจงแบบแกมม่า (Gamma) หรือแบบเรย์ลี (Rayleigh) สำหรับการศึกษาเกี่ยวกับการกำจัดสัญญาณรบกวนแบบการบวกและการคูณที่ได้รับความนิยมมีดังต่อไปนี้

Rudin Osher และ Fatemi [4] เป็นนักวิจัยกลุ่มแรกที่นำเสนอตัวแบบในการกำจัดสัญญาณแบบการบวก ซึ่งเป็นตัวแบบที่มีชื่อเสียงในการให้ผลลัพธ์ที่มีความคมชัดดี โดยตัวแบบ ROF กำหนดดังนี้

$$\min_{u\epsilon U}\{\mathcal{J}^{TV}(u) = \alpha D_{AN}(u) + R^{TV}(u)\} \tag{3}$$

เมื่อ $R^{TV}(u) = \int_\Omega |\nabla u| \ d\Omega$ แทนเร็กกิวลาร์ไรซ์เซชันแบบการแปรผันรวม $D_{AN}(u) = \frac{1}{2}\int_\Omega (u-z)^2 \ d\Omega$ แทนเทอมวัดความผิดปกติของข้อมูล แต่เนื่องจากเร็กกิวลาร์ไรซ์เซชันแบบการแปรผันรวมทำให้เกิดปรากฏการณ์ขั้นบันได คือ แปลงสัญญาณที่เรียบให้เป็นขั้นบันได เพื่อแก้ปัญหานี้ You และ Kaveh [10] ได้นำเสนอเร็กกิวลาร์ไรซ์เซชันแบบลาปลาซ

$$R^{TL}(u) = \int_\Omega |\Delta u| \ d\Omega$$

เมื่อ Δu แทนการดำเนินการลาปลาซ และนำเสนอตัวแบบ TL ดังนี้

$$\min_{u\epsilon U}\{\mathcal{J}^{TL}(u) = \alpha D_{AN}(u) + R^{TL}(u)\} \tag{4}$$

และนอกจากนี้ยังมีคณะวิจัยของ Scherzer [11] ใช้ Bounded Hessian เร็กกิวลาร์ไรซ์เซชัน (BH regularization)

$$R^{BH}(u) = \int_\Omega |\nabla^2 u| \ d\Omega$$

เมื่อ $\nabla^2 u = \begin{pmatrix} u_{xx} & u_{yx} \\ u_{xy} & u_{yy} \end{pmatrix}$ เป็นเมทริกซ์เฮสเซียนของ u และ $|\nabla^2 u| = \sqrt{u_{xx}^2 + u_{yx}^2 + u_{xy}^2 + u_{yy}^2}$ และนำเสนอตัวแบบ BH ดังนี้

$$\min_{u\epsilon U}\{\mathcal{J}^{BH}(u) = \alpha D_{AN}(u) + R^{BH}(u)\} \tag{5}$$

ในงานวิจัยที่ศึกษาตัวแบบสัญญาณรบกวนแบบการคูณ โดยปกติจะใช้เทอมของเร็กกิวลาร์ไรซ์เซชันเป็นเร็กกิวลาร์ไรซ์เซชันแบบการแปรผันรวม และพัฒนาเทอมวัดความผิดปกติของข้อมูล ดังต่อไปนี้

งานวิจัยของ Rudin และคณะ [12] ได้นำเสนอตัวแบบสำหรับกำจัดสัญญาณรบกวนแบบการคูณที่เรียกว่าตัวแบบ RLO ดังนี้

$$\min_{u\epsilon U}\left\{\mathcal{J}^{RLO}(u) = \gamma_1 \int_\Omega \frac{z}{u} \ d\Omega + \gamma_2 \int_\Omega \left(\frac{z}{u} - 1\right)^2 \ d\Omega + \int_\Omega |\nabla u| \ d\Omega\right\}$$

เมื่อเทอมที่ 1 และเทอมที่ 2 แทนเทอมวัดความผิดปกติของข้อมูล และ γ_1 และ γ_2 เป็นพารามิเตอร์ถ่วงน้ำหนัก

จากนั้น Aubert และ Aujol [8] ได้ใช้สมมติฐานที่ว่า สัญญาณรบกวนแบบการคูณมีการแจกแจงแบบแกมม่าที่มีค่าเฉลี่ยเป็น 1 และนำเสนอตัวแบบ AA ดังต่อไปนี้

$$\min_{u\epsilon U}\left\{\mathcal{J}^{AA}(u) = \gamma \int_\Omega \left(\log u + \frac{z}{u}\right) \ d\Omega + \int_\Omega |\nabla u| \ d\Omega\right\}$$

เนื่องจากเทอมวัดความผิดปกติของข้อมูลของตัวแบบ AA ไม่เป็นคอนเวกซ์ ส่งผลให้การหาคำตอบทำได้ยากและช้า เพื่อแก้ปัญหาของตัวแบบ AA ที่ไม่เป็นคอนเวกซ์ Shi และ Osher [13] ได้แปลง $\bar{u} = \log u$ และนำเสนอตัวแบบ SO ดังต่อไปนี้

$$\min_{\bar{u}\epsilon U}\left\{\mathcal{J}^{SO}(\bar{u}) = \gamma \int_\Omega \left(aze^{-\bar{u}} + \frac{b}{2}z^2 e^{-2\bar{u}} + (a+b)\bar{u}\right) \ d\Omega + \int_\Omega |\nabla \bar{u}| \ d\Omega\right\}$$

เมื่อ a, b เป็นค่าคงที่ที่มากกว่าศูนย์ และ $\bar{u} = e^{\bar{u}}$ สังเกตว่าตัวแบบ SO เป็นคอนเวกซ์ แต่ในการหาคำตอบยังคงช้า เพื่อแก้ปัญหาในการคำนวณช้า คณะวิจัยของ Huang [14] ได้นำเสนอการแปลงเทอม $\log u + \frac{z}{u}$ ในตัวแบบ AA เป็น $u + ze^{-u}$ ภายใต้การแปลง $u \to e^u$ และนำเสนอตัวแบบ HNW ดังนี้

$$\min_{u,w\epsilon U}\left\{\mathcal{J}^{HNW}(u) = \int_\Omega (u + ze^{-u}) \ d\Omega + \gamma_1 \int_\Omega |\mathrm{u} - \mathrm{w}|^2 + \gamma_2 \int_\Omega |\nabla \mathrm{w}| \ d\Omega\right\}$$

จากนั้น Jin และ Yang [9] ได้ปรับปรุงตัวแบบ AA และ ตัวแบบ HNW และนำเสนอตัวแบบที่มีชื่อเสียงในการกำจัดสัญญาณรบกวนแบบการคูณ ที่เรียกว่าตัวแบบ JY ดังต่อไปนี้

$$\min_{u\epsilon U}\{\mathcal{J}^{JY}(u) = \gamma \int_\Omega (u + ze^{-u}) \ d\Omega + \int_\Omega |\nabla u| \ d\Omega\} \tag{6}$$

ซึ่งผลงานวิจัยพบว่าตัวแบบ JY ให้ภาพผลลัพธ์ที่ดีกว่าและใช้เวลาในการหาคำตอบน้อยกว่าตัวแบบ AA และตัวแบบ HNW

จากงานวิจัยที่ผ่านมาพบว่า สำหรับตัวแบบสัญญาณรบกวนแบบการบวกส่วนใหญ่พัฒนาเทอมเร็กกิวลาร์ไรซ์เซชันเพื่อแก้ปัญหาปรากฏการณ์ขั้นบันได และสำหรับตัวแบบสัญญาณรบกวนแบบการคูณมุ่งเน้นในการพัฒนาเทอมการวัดความผิดปกติของข้อมูลเพื่อพัฒนาตัวแบบการสร้างภาพ และทั้งในการพัฒนาตัวแบบสัญญาณรบกวนแบบการบวกและการคูณยังคงต้องพัฒนาวิธีการตัวเลขแบบเร็วสำหรับกำจัดสัญญาณรบกวน

ในงานวิจัยนี้ผู้วิจัยมุ่งเน้นพัฒนาตัวแบบกำจัดสัญญาณรบกวนออกจากภาพถ่ายคลื่นเสียงความถี่สูงซึ่งถูกเจือปนด้วยสัญญาณรบกวนแบบการคูณ โดยใช้ข้อดีของตัวแบบ JY และเพื่อแก้ปัญหาปรากฏการณ์ขั้นบันได ผู้วิจัยได้ปรับปรุงและเปรียบเทียบตัวแบบดังกล่าวโดยใช้ $R^{TL}(u)$ และ $R^{BH}(u)$ พร้อมทั้งพัฒนาวิธีการเชิงตัวเลขแบบเร็วสำหรับกำจัดสัญญาณรบกวนออกจากภาพ โดยในการหาคำตอบของปัญหาเชิงการแปรผันผู้วิจัยได้นำเสนอวิธีการ SB

2.1 ตัวแบบการกำจัดสัญญาณรบกวนที่นำเสนอ

อย่างที่กล่าวมาตัวแบบ JY ใน (6) เป็นตัวแบบสำหรับกำจัดสัญญาณรบกวนออกจากภาพที่มีชื่อเสียงในการให้ผลลัพธ์ที่มีความคมชัด ผู้วิจัยจึงได้พัฒนาตัวแบบซึ่งใช้ข้อดีในการให้ผลลัพธ์ที่ดีจากตัวแบบดังกล่าว สังเกตว่าเทอมเร็กกิวลาร์ไรซ์เซชันของตัวแบบ JY ใน (6) เป็นเร็กกิวลาร์ไรซ์เซชันแบบการแปรผันรวม $R^{TV}(u)$ แม้ว่าประสิทธิภาพในการรักษาขอบของภาพในการกำจัดสัญญาณรบกวนของเร็กกิวลาร์ไรซ์เซชันแบบการแปรผันรวมสามารถทำได้อย่างดี แต่มักพบการแปลงสัญญาณให้มีความเรียบเป็นขั้นบันไดโดยไม่จำเป็น ผู้วิจัยจึงได้ปรับปรุงตัวแบบสัญญาณรบกวนออกจากภาพโดยใช้ข้อดีของอนุพันธ์อันดับสูงของลาปลาซจาก [10] และอนุพันธ์อันดับสูงของเมทริกซ์เฮสเซียนจาก [11] โดยตัวแบบเชิงการแปรผันสำหรับกำจัดสัญญาณรบกวนที่ปรับปรุง กำหนดโดย

$$\min_{u}\{\mathcal{J}^{JYTL} = \alpha D_{MN}(u) + R^{TL}(u)\} \tag{7}$$

และ

$$\min_{u}\{\mathcal{J}^{JYBH} = \alpha D_{MN}(u) + R^{BH}(u)\} \tag{8}$$

เมื่อ $D_{MN}(u) = \int_{\Omega}(u + ze^{-u})\ d\Omega$ ในที่นี้เราเรียกตัวแบบใน (7) และ (8) ว่าตัวแบบ JYTL และตัวแบบ JYBH ตามลำดับ

2.2 วิธีการผลต่างอันตะ

ในการแก้ปัญหาค่าขอบโดยวิธีการผลต่างอันตะ (Finite Difference Method) เราเริ่มจากการดิสครีตไทซ์โดเมนภาพ Ω เป็นเมชแบบคงรูปในแต่ละทิศทาง ซึ่งจะได้โดเมนภาพแบบดิสครีต

$$\Omega_h = \{(x,y) \in \Omega | (x,y) = (x_i, y_j), x_i = i, y_j = j, 1 \le i \le M, 1 \le j \le N\}$$

เพื่อความสะดวก สำหรับแต่ละจุดกริด $(x_i, y_j) \in \Omega_h$ เราเขียนแทนด้วย (i,j) โดยที่พิกัด x และ y จะวางแนวตามคอลัมน์และแถว ตามลำดับ สำหรับการประมาณแบบผลต่างอันตะของอนุพันธ์ย่อยอันดับหนึ่งกำหนดโดย

$$\partial_x^+(u)_{i,j} = \begin{cases} (u)_{i,j+1} - (u)_{i,j}, & 1 \le i \le M, 1 \le j \le N \\ (u)_{i,1} - (u)_{i,j}, & 1 \le i \le M, j = N \end{cases}$$

$$\partial_y^+(u)_{i,j} = \begin{cases} (u)_{i+1,j} - (u)_{i,j}, & 1 \le i \le M, 1 \le j \le N \\ (u)_{i,1} - (u)_{i,j}, & i = M, 1 \le j \le N \end{cases}$$
$$\partial_x^-(u)_{i,j} = \begin{cases} (u)_{i,j} - (u)_{i,j-1}, & 1 \le i \le M, 1 \le j \le N \\ (u)_{i,1} - (u)_{i,N}, & 1 \le i \le M, j = 1 \end{cases}$$
$$\partial_y^-(u)_{i,j} = \begin{cases} (u)_{i,j} - (u)_{i,j-1}, & 1 \le i \le M, 1 \le j \le N \\ (u)_{i,1} - (u)_{M,j}, & i = 1, 1 \le j \le N \end{cases}$$

การประมาณค่าแบบผลต่างอันตะของอนุพันธ์ย่อยอันดับสองของ u ที่แต่ละจุดกริด (i,j) สามารถกำหนดโดย

$$\partial_x^+\partial_x^-(u)_{i,j} = \partial_x^-\partial_x^+(u)_{i,j} = \begin{cases} (u)_{i,N} - 2(u)_{i,j} + (u)_{i,j+1}, & 1 \le i \le M, j = 1 \\ (u)_{i,j-1} - 2(u)_{i,j} + (u)_{i,j+1}, & 1 \le i \le M, 1 < j < N \\ (u)_{i,j-1} - 2(u)_{i,j} + (u)_{i,1}, & 1 \le i \le M, j = N \end{cases}$$
$$\partial_y^+\partial_y^-(u)_{i,j} = \partial_y^-\partial_y^+(u)_{i,j} = \begin{cases} (u)_{M,j} - 2(u)_{i,j} + (u)_{i+1,j}, & i = 1, 1 \le j \le N \\ (u)_{i-1,j} - 2(u)_{i,j} + (u)_{i,+1j}, & 1 < i < M, 1 \le j \le N \\ (u)_{i-1,j} - 2(u)_{i,j} + (u)_{1,j}, & i = M, 1 \le j \le N \end{cases}$$
$$\partial_x^+\partial_y^-(u)_{i,j} = \begin{cases} (u)_{i,j+1} - (u)_{i,j} - (u)_{M,j+1} + (u)_{M,j}, & i = 1, 1 \le j < N \\ (u)_{i,1} - (u)_{i,j} - (u)_{M,1} + (u)_{M,j}, & i = 1, j = N \\ (u)_{i,j+1} - (u)_{i,j} - (u)_{i-1,j+1} + (u)_{i-1,j}, & 1 < i \le M, 1 \le j < N \\ (u)_{i,1} - (u)_{i,j} - (u)_{i-1,1} + (u)_{i-1,j}, & 1 < i \le M, j = N \end{cases}$$
$$\partial_y^+\partial_x^-(u)_{i,j} = \begin{cases} (u)_{i+1,j} - (u)_{i,j} - (u)_{i+1,N} + (u)_{i,N}, & 1 \le j < M, j = 1 \\ (u)_{1,j} - (u)_{i,j} - (u)_{1,N} + (u)_{i,N}, & i = M, j = 1 \\ (u)_{i+1,j} - (u)_{i,j} - (u)_{i+1,j-1} + (u)_{i,j-1}, & 1 \le i < M, 1 < j \le N \\ (u)_{1,j} - (u)_{i,j} - (u)_{1,j-1} + (u)_{i,j-1}, & i = M, 1 < j \le N \end{cases}$$

2.3 วิธีการสปริทเบรกแมน (วิธีการ SB)

ในหัวข้อนี้จะกล่าวถึงวิธีการเชิงตัวเลขที่มีประสิทธิภาพซึ่งในที่นี้คือวิธีการ SB สำหรับแก้ปัญหาตัวแบบเชิงการแปรผันทั้ง 3 ตัวแบบ ได้แก่ ตัวแบบ JY ตัวแบบ JYTL และตัวแบบ JYBH

2.3.1 วิธีการ SB สำหรับตัวแบบ JY

ในการแก้ปัญหาเชิงการแปรผันสำหรับตัวแบบ JY ใน (6) ด้วยวิธีการ SB จะเริ่มต้นจากการแนะนำเวกเตอร์เสริม $w = (w_1, w_2)^T$ พารามิเตอร์การทำซ้ำเบรกแมน (Bregman iterative parameter) $b = (b_1, b_2)^T$ และพารามิเตอร์ตัวโทษ (penalty parameter) $\theta > 0$ เพื่อแปลงเป็นปัญหาเชิงการแปรผันซึ่งกำหนดโดย

$$\min_u \left\{ \mathcal{J}^{JY}(w;b) = \int_\Omega |w| d\Omega + \frac{\theta}{2}\int_\Omega (w - \nabla u - b)^2 d\Omega + \alpha \int_\Omega (u + ze^{-u}) d\Omega \right\} \quad (9)$$

สังเกตว่าเป็นการยากที่จะแก้ไขตัวแปร u และ w ไปพร้อมกัน จึงแบ่งปัญหาออกเป็นปัญหาการหาค่าต่ำสุดสองปัญหาย่อยดังนี้

$$u^{[new]} = \underset{u}{\operatorname{argmin}} \left\{ \frac{\theta}{2}\int_\Omega \left(w^{[old]} - \nabla u - b^{[old]}\right)^2 d\Omega + \alpha \int_\Omega (u + ze^{-u}) d\Omega \right\} \quad (10)$$
$$w^{[new]} = \underset{w}{\operatorname{argmin}} \left\{ \int_\Omega |w| d\Omega + \frac{\theta}{2}\int_\Omega \left(w - \nabla u^{[new]} - b^{[old]}\right)^2 d\Omega \right\} \quad (11)$$

เพื่อแก้ปัญหาดังกล่าวเราใช้เทคนิคการทำซ้ำแบบสลับ จากนั้นทำการปรับปรุงพารามิเตอร์เบรกแมน b

$$b^{[new]} = b^{[old]} + \nabla u^{[new]} - w^{[new]} \quad (12)$$

ในกระบวนการนี้ เราจะดำเนินการทำซ้ำแบบสลับจนกระทั่งลำดับของ u สอดคล้องกับเกณฑ์การหยุด

$$\frac{\|u^{[new]} - u^{[old]}\|_{l_2}^2}{\|u^{[new]}\|_{l_2}^2} < \varepsilon_1^{SB} \quad \text{หรือ} \quad m \ge \varepsilon_2^{SB} \quad (13)$$

เมื่อ $u^{[new]}$ และ $u^{[old]}$ แทนเวกเตอร์ของ u ที่ได้จากการทำซ้ำรอบปัจจุบันและการทำซ้ำรอบก่อนหน้า ตามลำดับ $\varepsilon_1^{SB} > 0$ แทนค่าความแม่นยำ และ ε_2^{SB} แทนจำนวนรอบการทำซ้ำสูงสุดของวิธีการ SB โดย m แทนรอบการทำซ้ำของวิธีการ SB ในงานวิจัยนี้กำหนด $\varepsilon_1^{SB} = 10^{-4}$ และ $\varepsilon_2^{SB} = 200$ สองปัญหาย่อยข้างต้นสามารถแก้ได้ดังนี้

ปัญหาย่อย u เมื่อตรึงตัวแปร $(w; b)$ ใน (10) แล้วใช้แคลคูลัสของการแปรผันเพื่อแก้ปัญหาค่าต่ำที่สุดจะได้สมการออยเลอร์-ลากรานจ์ดังนี้

$$-\theta\Delta u = \bar{G}(u) \tag{14}$$

เมื่อ $\bar{G}(u) = -\alpha(1 - ze^{-u}) - \theta div(w - b)$ เพื่อแก้สมการเชิงอนุพันธ์ย่อยไม่เป็นเชิงเส้นใน (14) ถูกทำให้เป็นเชิงเส้นโดยวิธีการทำซ้ำแบบจุดตรึง ซึ่งกำหนดโดย

$$\gamma u^{[\nu+1]} - \theta\Delta u^{[\nu+1]} = G\left(u^{[\nu]}\right)$$

เมื่อ $G\left(u^{[\nu]}\right) = \bar{G}\left(u^{[\nu]}\right) + \gamma u^{[\nu]}$ และ $\gamma > 0$ แทนพารามิเตอร์จุดตรึงที่ช่วยในการคำนวณให้มีความเสถียร จากนั้นทำการดิสครีตไทซ์โดยวิธีการผลต่างอันตะ จะได้ว่า

$$\gamma\left(u^{[\nu+1]}\right)_{i,j} - \theta\left(\partial_x^-\partial_x^+\left(u^{[\nu+1]}\right)_{i,j} + \partial_x^-\partial_x^+\left(u^{[\nu+1]}\right)_{i,j}\right) = G\left(u^{[\nu]}\right)_{i,j} \tag{15}$$

โดยที่ $G\left(u^{[\nu]}\right)_{i,j} = -\alpha\left(1 - (z)_{i,j}e^{-\left(u^{[\nu]}\right)_{i,j}}\right) - \theta\left(\partial_x^-\left((w_1)_{i,j} - (b_1)_{i,j}\right) + \partial_y^-\left((w_2)_{i,j} - (b_2)_{i,j}\right)\right) + \gamma\left(u^{[\nu]}\right)_{i,j}$ สำหรับการทำซ้ำภายนอก เริ่มต้นจากการกำหนดคำตอบเริ่มต้น $u^{[0]}$ (ในกรณีเฉพาะ $u^{[0]} = z$) จากนั้นใช้การแปลงฟูเรียร์แบบดิสครีต F กับ (15) จะได้

$$F\left(\gamma\left(u^{[\nu+1]}\right)_{i,j} - \theta\left(\partial_x^-\partial_x^+\left(u^{[\nu+1]}\right)_{i,j} + \partial_x^-\partial_x^+\left(u^{[\nu+1]}\right)_{i,j}\right)\right) = F\left(G\left(u^{[\nu]}\right)_{i,j}\right)$$

หรือ $$\zeta F\left(\left(u^{[\nu+1]}\right)_{i,j}\right) = F\left(G\left(u^{[\nu]}\right)_{i,j}\right)$$

เมื่อ $\zeta = \gamma - 2\theta\left(\cos\left(\frac{2\pi s}{N}\right) + \cos\left(\frac{2\pi r}{M}\right) - 2\right)$, $i \in [1, M]$ และ $j \in [1, N]$ แทนดัชนีในโดเมนเวลา $r \in [0, M)$ และ $s \in [0, N)$ แทนดัชนีในโดเมนความถี่ ในขั้นตอนสุดท้ายเราได้รูปแบบปิดของคำตอบของ $u^{[\nu+1]}$ ที่จุดกริด (i, j)

$$\left(u^{[\nu+1]}\right)_{i,j} = Re\left(F^{-1}\left(\frac{F\left(G\left(u^{[\nu]}\right)_{i,j}\right)}{\zeta}\right)\right)$$

ในที่นี้ F^{-1} แทนการแปลงฟูเรียร์ผกผันแบบดิสครีต และ Re เป็นส่วนจริงของจำนวนเชิงซ้อน

ปัญหาย่อย w เมื่อตรึงตัวแปร $(u; b)$ ใน (11) แล้วใช้แคลคูลัสของการแปรผันจะได้สมการออยเลอร์-ลากรานจ์ที่เกี่ยวข้องกับตัวแปร w ดังนี้

$$\frac{w}{|w|} + \theta(w - \nabla u - b) = 0$$

ซึ่งคำตอบสามารถใช้สูตรรูปแบบปิด [15]

$$(w)_{i,j} = \max\left(\left|\nabla(u)_{i,j} - (b)_{i,j}\right| - \frac{1}{\theta}, 0\right)\frac{\nabla(u)_{i,j} - (b)_{i,j}}{\left|\nabla(u)_{i,j} - (b)_{i,j}\right|}$$

และขั้นตอนสุดท้ายคือการปรับปรุงพารามิเตอร์การทำซ้ำเบรกแมน โดยกำหนดให้ $b \leftarrow b + \nabla u - w$

2.3.2 วิธีการ SB สำหรับตัวแบบ JYTL

เพื่อที่จะแก้ปัญหาเชิงการแปรผันสำหรับตัวแบบ JYTL ใน (7) ด้วยวิธีการ SB เราเริ่มต้นจากการแนะนำเวกเตอร์เสริม $w = (w_1, w_2)^T$ พารามิเตอร์การทำซ้ำเบรกแมน (Bregman iterative parameter) $b = (b_1, b_2)^T$ และพารามิเตอร์ตัวโทษ (penalty parameter) $\theta > 0$ เพื่อแปลงเป็นปัญหาเชิงการแปรผันซึ่งกำหนดโดย

$$\min_{u}\left\{\mathcal{J}^{JYTL}(w;b)=\int_{\Omega}|w|d\Omega+\frac{\theta}{2}\int_{\Omega}(w-\Delta u-b)^2d\Omega+\alpha\int_{\Omega}(u+ze^{-u})d\Omega\right\} \quad (16)$$

ในทำนองเดียวกัน การแก้ปัญหาสามารถทำได้โดยใช้เทคนิคการทำซ้ำแบบสลับ โดยปัญหาการหาค่าต่ำสุดสองปัญหาย่อยสำหรับ u และ w และการปรับปรุงพารามิเตอร์ b กำหนดโดย

$$u^{[new]}=\underset{u}{\operatorname{argmin}}\left\{\frac{\theta}{2}\int_{\Omega}\left(w^{[old]}-\Delta u-b^{[old]}\right)^2d\Omega+\alpha\int_{\Omega}(u+ze^{-u})d\Omega\right\} \quad (17)$$

$$w^{[new]}=\underset{w}{\operatorname{argmin}}\left\{\int_{\Omega}|w|d\Omega+\frac{\theta}{2}\int_{\Omega}\left(w-\Delta u^{[new]}-b^{[old]}\right)^2d\Omega\right\} \quad (18)$$

$$b^{[new]}=b^{[old]}+\Delta u^{[new]}-w^{[new]} \quad (19)$$

โดยเราจะดำเนินการทำซ้ำจนกระทั่งสอดคล้องเกณฑ์การหยุด (13) โดยปัญหาย่อยทั้งสองสามารถแก้ได้ดังนี้

ปัญหาย่อย u เมื่อตรึงตัวแปร $(w;b)$ ใน (17) แล้วใช้แคลคูลัสของการแปรผันจะได้สมการออยเลอร์-ลากรานจ์ดังนี้

$$-\theta\Delta(\Delta u)=\bar{G}(u) \quad (20)$$

เมื่อ $\bar{G}(u)=-\alpha(1-ze^{-u})-\theta\Delta(w-b)$ เพื่อแก้ปัญหาสมการเชิงอนุพันธ์ย่อยไม่เป็นเชิงเส้นใน (20) เราใช้วิธีการทำซ้ำแบบจุดตรึง กำหนดโดย

$$\gamma u^{[\nu+1]}-\theta\Delta\left(\Delta u^{[\nu+1]}\right)=G\left(u^{[\nu]}\right)$$

เมื่อ $G\left(u^{[\nu]}\right)=\bar{G}\left(u^{[\nu]}\right)+\gamma u^{[\nu]}$ และ $\gamma>0$ แทนพารามิเตอร์จุดตรึง จากนั้นทำการดิสครีตไทซ์โดยวิธีการผลต่างอันตะจะได้ว่า

$$\gamma\left(u^{[\nu+1]}\right)_{i,j}-\theta\begin{pmatrix}\partial_x^-\partial_x^+\partial_x^-\partial_x^+\left(u^{[\nu+1]}\right)_{i,j}+\partial_x^-\partial_x^+\partial_y^-\partial_y^+\left(u^{[\nu+1]}\right)_{i,j}\\+\partial_y^-\partial_y^+\partial_x^-\partial_x^+\left(u^{[\nu+1]}\right)_{i,j}+\partial_y^-\partial_y^+\partial_y^-\partial_y^+\left(u^{[\nu+1]}\right)_{i,j}\end{pmatrix}=G\left(u^{[\nu]}\right)_{i,j} \quad (21)$$

โดยที่ $G\left(u^{[\nu]}\right)_{i,j}=-\alpha\left(1-(z)_{i,j}e^{-\left(u^{[\nu]}\right)_{i,j}}\right)-\theta\left(\partial_x^-\partial_x^+\left((w)_{i,j}-(b)_{i,j}\right)+\partial_y^-\partial_y^+\left((w)_{i,j}-(b)_{i,j}\right)\right)+\gamma\left(u^{[\nu]}\right)_{i,j}$ สำหรับการทำซ้ำภายนอก เริ่มต้นจากการกำหนดคำตอบเริ่มต้น $u^{[0]}$ (ในกรณีเฉพาะ $u^{[0]}=z$) จากนั้นใช้การแปลงฟูเรียร์แบบดิสครีต F กับ (21) จะได้

$$F\left(\gamma\left(u^{[\nu+1]}\right)_{i,j}-\theta\begin{pmatrix}\partial_x^-\partial_x^+\partial_x^-\partial_x^+\left(u^{[\nu+1]}\right)_{i,j}+\partial_x^-\partial_x^+\partial_y^-\partial_y^+\left(u^{[\nu+1]}\right)_{i,j}\\+\partial_y^-\partial_y^+\partial_x^-\partial_x^+\left(u^{[\nu+1]}\right)_{i,j}+\partial_y^-\partial_y^+\partial_y^-\partial_y^+\left(u^{[\nu+1]}\right)_{i,j}\end{pmatrix}\right)=F\left(G\left(u^{[\nu]}\right)_{i,j}\right)$$

หรือ $$\zeta F\left(\left(u^{[\nu+1]}\right)_{i,j}\right)=F\left(G\left(u^{[\nu]}\right)_{i,j}\right)$$

เมื่อ $\zeta=\gamma+4\theta\left(\cos\left(\frac{2\pi s}{N}\right)+\cos\left(\frac{2\pi r}{M}\right)-2\right)^2$, $i\in[1,M]$ และ $j\in[1,N]$ แทนดัชนีในโดเมนเวลา $r\in[0,M)$ และ $s\in[0,N)$ แทนดัชนีในโดเมนความถี่ ดังนั้นผลเฉลยของ (20) ถูกกำหนดโดย

$$\left(u^{[\nu+1]}\right)_{i,j}=Re\left(F^{-1}\left(\frac{F\left(G\left(u^{[\nu]}\right)_{i,j}\right)}{\zeta}\right)\right)$$

ในที่นี้ F^{-1} แทนการแปลงฟูเรียร์ผกผันแบบดิสครีต และ Re เป็นส่วนจริงของจำนวนเชิงซ้อน

ปัญหาย่อย w สำหรับการคำนวณเวกเตอร์ w สามารถคำนวณได้จาก [15]

$$\frac{w}{|w|}+\theta(w-\Delta u-b)=0$$

ซึ่งมีผลเฉลยแม่นตรง (exact solution) คือ

$$(w)_{i,j}=\max\left(\left|\Delta(u)_{i,j}-(b)_{i,j}\right|-\frac{1}{\theta},0\right)sign\left(\Delta(u)_{i,j}-(b)_{i,j}\right)$$

และขั้นตอนสุดท้ายคือการปรับปรุงพารามิเตอร์การทำซ้ำเบรกแมน โดยกำหนดให้ $b\leftarrow b+\Delta u-w$

2.3.3 วิธีการ SB สำหรับตัวแบบ JYBH

เพื่อแก้ปัญหาเชิงการแปรผันสำหรับตัวแบบ JYBH ใน (8) เราจะแปลงปัญหาดังกล่าวเป็นปัญหาเชิงการแปรผันที่มีหลายตัวแปรซึ่งกำหนดโดย

$$\min_{u\in U}\left\{\mathcal{J}^{JYBH}(w;b)=\int_\Omega|w|d\Omega+\frac{\theta}{2}\int_\Omega(w-\nabla^2u-b)^2d\Omega+\alpha\int_\Omega(u+ze^{-u})d\Omega\right\} \quad (22)$$

ขั้นตอนถัดไป จะทำการแก้ปัญหาย่อยสองปัญหาสำหรับ u และ w และปรับปรุงพารามิเตอร์ b ซึ่งกำหนดโดย

$$u^{[new]}=\operatorname*{argmin}_u\left\{\frac{\theta}{2}\int_\Omega\left(w^{[old]}-\nabla^2u-b^{[old]}\right)^2d\Omega+\alpha\int_\Omega(u+ze^{-u})d\Omega\right\} \quad (23)$$

$$w^{[new]}=\operatorname*{argmin}_w\left\{\int_\Omega|w|d\Omega+\frac{\theta}{2}\int_\Omega\left(w-\nabla^2u^{[new]}-b^{[old]}\right)^2d\Omega\right\} \quad (24)$$

$$b^{[new]}=b^{[old]}+\nabla^2u^{[new]}-w^{[new]} \quad (25)$$

ในกระบวนการนี้ เราจะดำเนินการทำซ้ำแบบสลับจนกระทั่งลำดับของ u สอดคล้องกับเกณฑ์การหยุด (13)

ปัญหาย่อย u เมื่อตรึงตัวแปร $(w;b)$ ใน (23) แล้วใช้แคลคูลัสของการแปรผันเพื่อแก้ปัญหาค่าต่ำที่สุดจะได้สมการออยเลอร์-ลากรานจ์ที่เกี่ยวข้องดังนี้

$$-\theta div^2(\nabla^2u)=\bar{G}(u) \quad (26)$$

เมื่อ $\bar{G}(u)=-\alpha(1-ze^{-u})-\theta div^2(w-b)$ ใช้วิธีการทำซ้ำแบบจุดตรึงเพื่อแก้สมการเชิงอนุพันธ์ย่อยไม่เป็นเชิงเส้นใน (26) โดยที่ v แทนดัชนีสำหรับขั้นตอนการทำซ้ำภายนอกที่กำหนดโดย

$$\gamma u^{[v+1]}-\theta div^2\left(\nabla^2u^{[v+1]}\right)=G\left(u^{[v]}\right)$$

เมื่อ $G\left(u^{[v]}\right)=\bar{G}\left(u^{[v]}\right)+\gamma u^{[v]}$ และ $\gamma>0$ แทนพารามิเตอร์จุดตรึง จากนั้นทำการดิสครีตไทซ์โดยวิธีการผลต่างอันตะจะได้ว่า

$$\gamma\left(u^{[v+1]}\right)_{i,j}-\theta\begin{pmatrix}\partial_x^-\partial_x^+\partial_x^-\partial_x^+\left(u^{[v+1]}\right)_{i,j}+\partial_y^-\partial_x^+\partial_y^-\partial_x^+\left(u^{[v+1]}\right)_{i,j}\\+\partial_x^-\partial_y^+\partial_x^-\partial_y^+\left(u^{[v+1]}\right)_{i,j}+\partial_y^-\partial_y^+\partial_y^-\partial_y^+\left(u^{[v+1]}\right)_{i,j}\end{pmatrix}=G\left(u^{[v]}\right)_{i,j} \quad (27)$$

โดยที่ $G\left(u^{[v]}\right)_{i,j}=-\alpha\left(1-(z)_{i,j}e^{-\left(u^{[v]}\right)_{i,j}}\right)-\theta\left(\partial_x^-\partial_x^+\left((w_1)_{i,j}-(b_1)_{i,j}\right)+\partial_y^-\partial_x^+\left((w_2)_{i,j}-(b_2)_{i,j}\right)+\partial_x^-\partial_y^+\left((w_3)_{i,j}-(b_3)_{i,j}\right)+\partial_y^-\partial_x^+\left((w_4)_{i,j}-(b_4)_{i,j}\right)\right)+\gamma\left(u^{[v]}\right)_{i,j}$

จากนั้นใช้การแปลงฟูเรียร์แบบดิสครีต F กับ (27) จะได้

$$F\left(\gamma\left(u^{[v+1]}\right)_{i,j}+\theta\begin{pmatrix}\partial_x^-\partial_x^+\partial_x^-\partial_x^+\left(u^{[v+1]}\right)_{i,j}+\partial_x^-\partial_x^+\partial_y^-\partial_y^+\left(u^{[v+1]}\right)_{i,j}\\+\partial_y^-\partial_y^+\partial_x^-\partial_x^+\left(u^{[v+1]}\right)_{i,j}+\partial_y^-\partial_y^+\partial_y^-\partial_y^+\left(u^{[v+1]}\right)_{i,j}\end{pmatrix}\right)=F\left(G\left(u^{[v]}\right)_{i,j}\right)$$

หรือ $$\zeta F\left(\left(u^{[v+1]}\right)_{i,j}\right)=F\left(G\left(u^{[v]}\right)_{i,j}\right)$$

เมื่อ $\zeta=\gamma+4\theta\left(\cos\left(\frac{2\pi s}{N}\right)+\cos\left(\frac{2\pi r}{M}\right)-2\right)^2$, $i\in[1,M]$ และ $j\in[1,N]$ แทนดัชนีในโดเมนเวลา $r\in[0,M)$ และ $s\in[0,N)$ แทนดัชนีในโดเมนความถี่ ดังนั้นผลเฉลยของ (26) ถูกกำหนดโดย

$$\left(u^{[v+1]}\right)_{i,j}=Re\left(F^{-1}\left(\frac{F\left(G\left(u^{[v]}\right)_{i,j}\right)}{\zeta}\right)\right)$$

ปัญหาย่อย w เมื่อตรึง $(u;b)$ การคำนวณเวกเตอร์ w สามารถคำนวณได้จาก [15]

$$\frac{w}{|w|}+\theta(w-\nabla^2u-b)=0$$

ซึ่งมีผลเฉลยแม่นตรง (exact solution) คือ

$$(w)_{i,j}=\max\left(\left|\nabla^2(u)_{i,j}-(b)_{i,j}\right|-\frac{1}{\theta},0\right)\frac{\nabla^2(u)_{i,j}-(b)_{i,j}}{\left|\nabla^2(u)_{i,j}-(b)_{i,j}\right|}$$

จากนั้นปรับปรุงพารามิเตอร์การทำซ้ำเบรกแมน โดยกำหนดให้ $b \leftarrow b + \nabla^2 u - w$

3 ผลการศึกษา

เพื่อทดสอบประสิทธิภาพของวิธีการเชิงตัวเลขที่ได้นำเสนอ ผู้วิจัยได้ทำการทดลองเชิงตัวเลขกับภาพถ่ายดิจิทัลทั้งภาพสังเคราะห์และภาพจริง โดยภาพต้นฉบับแสดงดังรูปที่ 1 ในการทดลองเชิงตัวเลขผู้วิจัยได้ประยุกต์ใช้ค่าอัตราส่วนของสัญญาณรบกวนสูงสุด (Peak Signal to noise ratio : PSNR) (หน่วยเป็นเดซิเบล) ระหว่างเวกเตอร์ของภาพต้นฉบับที่ไม่มีสัญญาณรบกวน u^* และเวกเตอร์ของภาพผลลัพธ์ u ที่ได้จากวิธีการที่นำเสนอ เพื่อทำการประเมินคุณภาพของตัวแบบที่ได้นำเสนอ และประยุกต์ใช้ค่าความคลาดเคลื่อนกำลังสองโดยเฉลี่ย (Mean Square Error : MSE) เพื่อตรวจสอบประสิทธิภาพของวิธีการที่ได้นำเสนอ โดยค่าอัตราส่วนของสัญญาณรบกวนสูงสุดถูกนิยามโดย

$$PSNR = 10 \log_{10} \frac{255^2}{MSE}$$

และค่าคลาดเคลื่อนกำลังสองโดยเฉลี่ยถูกนิยามโดย

$$MSE = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \left((u^*)_{i,j} - (u)_{i,j} \right)^2$$

สำหรับการประเมินประสิทธิภาพของการกำจัดสัญญาณรบกวนที่ได้นำเสนอทุกการทดลองเชิงตัวเลขใช้ภาพที่มีความคมชัดขนาด 512 x 512 และในการประเมินประสิทธิภาพของขั้นตอนวิธีเชิงตัวเลขที่ได้พัฒนาขึ้นจะทำซ้ำจนกระทั่งเวกเตอร์ผลเฉลยลู่เข้าด้วยเกณฑ์การหยุดค่าคลาดเคลื่อนสัมพัทธ์ $\frac{\|u^{[new]} - u^{[old]}\|_{l_2}^2}{\|u^{[new]}\|_{l_2}^2} < 10^{-4}$ หรือจำนวนการทำซ้ำสูงสุด 200 รอบ เมื่อ $u^{[new]}$ และ $u^{[old]}$ แทนเวกเตอร์ของ u ที่ได้จากการทำซ้ำรอบปัจจุบันและการทำซ้ำรอบก่อนหน้า ตามลำดับ สำหรับแต่ละการทดลองเชิงตัวเลข ผู้วิจัยเลือกใช้พารามิเตอร์ α, γ และ θ ที่ให้ผลลัพธ์ที่ดีที่สุดภายใต้จำนวนรอบการทำซ้ำแบบจุดตรึงเท่ากับ 1

รูปที่ 1 : ภาพดิจิทัลของภาพต้นฉบับขนาด 512 x 512 พิกเซล โดยภาพสังเคราะห์ (คอลัมน์แรก) และภาพจริง (คอลัมน์ที่ 2 และ 3)

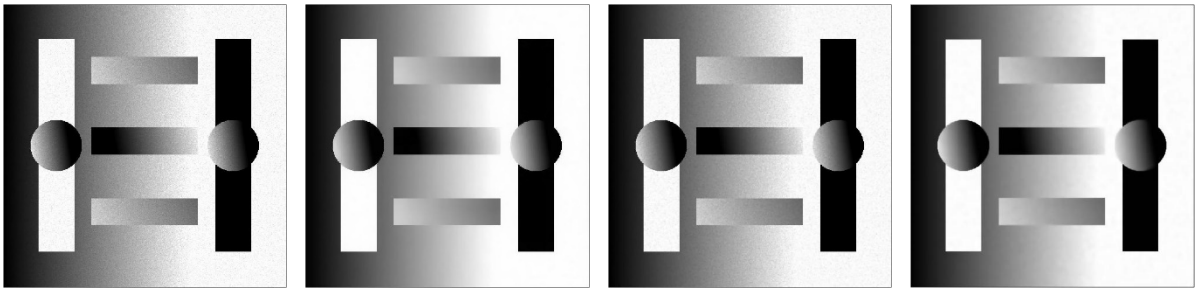
เพื่อแสดงให้เห็นประสิทธิภาพของตัวแบบที่ได้นำเสนอ พร้อมด้วยวิธีการ SB ผู้วิจัยจึงได้นำเสนอผลการวิเคราะห์จากการกำจัดสัญญาณรบกวนออกจากภาพที่มีสัญญาณรบกวนแบบการคูณที่มีค่าความหนาแน่น 0.03, 0.04 และ 0.05 โดยพิจารณาจากค่า PSNR และค่า MSE ในการทดลองบนภาพสังเคราะห์ (ภาพ Smooth) และภาพจริง

(ภาพ Boat และภาพ Airplane) จากตารางที่ 1 แสดงให้เห็นว่าตัวแบบ JYBH ให้ค่า PSNR ที่สูงกว่าตัวแบบ JY และตัวแบบ JYTL และเมื่อพิจารณาค่า MSE จะเห็นว่าตัวแบบ JYBH ให้ค่า MSE ที่ต่ำกว่าตัวแบบ JY และตัวแบบ JYTL ในทุกกรณี ซึ่งแสดงให้เห็นว่าตัวแบบ JYBH พร้อมด้วยวิธีการ SB สามารถกำจัดสัญญาณรบกวนออกจากภาพอย่างแม่นยำ และให้คุณภาพของภาพผลลัพธ์ที่ดีขึ้น โดยตัวแบบ JYBH ให้ความแม่นยำสูงกว่าตัวแบบ JY และตัวแบบ JYTL ในทุกกรณี ซึ่งผลการทดลองเชิงตัวเลขแสดงดังตารางที่ 1

ภาพ	k	$PSNR^0$	$PSNR$			MSE		
			ตัวแบบ JY	ตัวแบบ JYTL	ตัวแบบ JYBH	ตัวแบบ JY	ตัวแบบ JYTL	ตัวแบบ JYBH
Smooth	0.03	31.7418	33.0841	33.171	**33.5271**	31.9639	31.3305	**28.8637**
	0.04	30.4258	32.6883	32.4976	**32.9989**	35.0137	36.5858	**32.5968**
	0.05	29.1508	32.2177	31.5879	**32.3728**	39.0209	45.1109	**37.6519**
Boat	0.03	35.8861	40.2188	41.0886	**42.0915**	6.1821	5.0598	**4.0163**
	0.04	33.3958	38.0455	39.1573	**40.6474**	10.1973	7.8939	**5.6009**
	0.05	31.4541	37.2948	37.4415	**39.3605**	12.1218	11.719	**7.5331**
Airplane	0.03	33.2338	37.0639	37.0694	**37.8076**	12.7837	12.7676	**10.7717**
	0.04	30.7226	34.6961	34.974	**35.7918**	22.0522	20.6854	**17.1345**
	0.05	28.7981	34.5056	33.2382	**34.7883**	23.0409	30.8492	**21.5889**

ตารางที่ 1 : ผลการทดลองเชิงตัวเลขด้วยตัวแบบเชิงการแปรผันกับภาพสังเคราะห์ (ภาพ Smooth) และภาพจริง (ภาพ Boat และภาพ Airplane)

นอกจากนี้ เพื่อเป็นการแสดงให้เห็นประสิทธิภาพของตัวแบบที่ได้นำเสนอพร้อมด้วยวิธีการ SB ดังกล่าว ผู้วิจัยได้แสดงภาพผลลัพธ์จากการกำจัดสัญญาณรบกวนออกจากภาพในกรณีสัญญาณรบกวน k = 0.05 ดังแสดงในดังรูปที่ 2



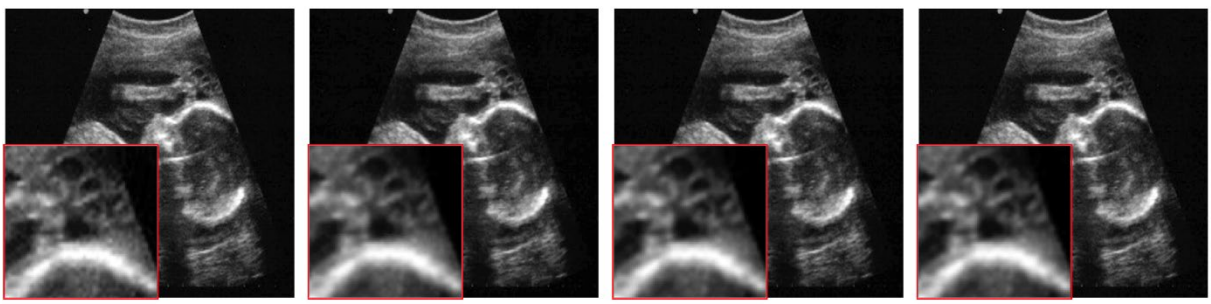
$PSNR^0$ = 29.1508 $PSNR$ = 32.2177 $PSNR$ = 31.5879 $PSNR$ = 32.3728

$PSNR^0$ = 31.4541 $PSNR$ = 37.2948 $PSNR$ = 37.4415 $PSNR$ = 39.3605

$PSNR^0$ = 28.7981 $PSNR$ = 34.5056 $PSNR$ = 33.2382 $PSNR$ = 34.7883

รูปที่ 2 : คอลัมน์แรกแสดงภาพที่มีสัญญาณรบกวน คอลัมน์ที่ 2 แสดงภาพผลลัพธ์จากการกำจัดสัญญาณรบกวนด้วยตัวแบบ JY คอลัมน์ที่ 3 และ 4 แสดงภาพผลลัพธ์จากการกำจัดสัญญาณรบกวนด้วยตัวแบบที่ได้นำเสนอ (ตัวแบบ JYTL และตัวแบบ JYBH ตามลำดับ)

ในลำดับถัดไป ผู้วิจัยได้ทำการทดสอบประสิทธิภาพของตัวแบบเชิงการแปรผันพร้อมด้วยวิธีการที่ได้นำเสนอในการกำจัดสัญญาณรบกวนออกจากภาพถ่ายทางการแพทย์ซึ่งไม่ทราบค่าของสัญญาณรบกวน ดังแสดงในรูปที่ 3 โดยคอลัมน์แรกแสดงภาพถ่ายทางการแพทย์ คอลัมน์ที่ 2 - 4 แสดงภาพผลลัพธ์จากการกำจัดสัญญาณรบกวนด้วยตัวแบบ JY ตัวแบบ JYTL และตัวแบบ JYBH ตามลำดับ จากรูปที่ 3 พบว่าตัวแบบ JY ตัวแบบ JYTL และตัวแบบ JYBH สามารถกำจัดสัญญาณรบกวนออกจากภาพได้ และตัวแบบ JYTL และตัวแบบ JYBH ให้ภาพผลลัพธ์ที่คมชัดกว่าตัวแบบ JY แม้ว่าตัวแบบทั้งสามจะสามารถกำจัดสัญญาณรบกวนออกจากภาพถ่ายทางการแพทย์ได้ แต่ภาพผลลัพธ์อาจเกิดความเบลอจากการกำจัดสัญญาณรบกวนในบางจุดมากเกินความจำเป็น



รูปที่ 3 : คอลัมน์แรกแสดงภาพถ่ายทางการแพทย์ คอลัมน์ที่ 2 - 4 แสดงภาพผลลัพธ์จากการกำจัดสัญญาณรบกวนด้วยตัวแบบ JY ตัวแบบ JYTL และตัวแบบ JYBH ตามลำดับ

4 สรุปผล

ในงานวิจัยนี้ ผู้วิจัยได้นำเสนอตัวแบบเชิงการแปรผันจำนวน 2 ตัวแบบ คือ ตัวแบบ JYTL และตัวแบบ JYBH สำหรับกำจัดสัญญาณรบกวนแบบการคูณออกจากภาพดิจิทัลซึ่งใช้ข้อดีของตัวแบบ JY และการแก้ปัญหาปรากฏการณ์ขั้นบันไดโดยใช้อนุพันธ์อันดับสูง ซึ่งผู้วิจัยได้ทำการทดลองกำจัดสัญญาณรบกวนออกจากภาพทั้งภาพสังเคราะห์และภาพจริง เพื่อแก้ปัญหาเชิงการแปรผันที่เกี่ยวข้องผู้วิจัยได้นำเสนอวิธีการ SB ซึ่งเป็นวิธีการที่มีประสิทธิภาพในการแก้ปัญหา ผลการทดลองเชิงตัวเลขแสดงให้เห็นว่าตัวแบบที่ได้นำเสนอพร้อมด้วยวิธีการ SB สามารถกำจัดสัญญาณรบกวนออกจากภาพอย่างแม่นยำ และให้คุณภาพของภาพผลลัพธ์ที่ดีขึ้น โดยตัวแบบที่ได้นำเสนอคือตัวแบบ JYBH ให้ความแม่นยำสูงกว่าตัวแบบ JY และตัวแบบ JYTL ในทุกกรณี นอกจากนี้ผู้วิจัยได้ทำการทดสอบความมีประสิทธิภาพของตัวแบบที่ได้นำเสนอพร้อมด้วยวิธีการ SB ในการกำจัดสัญญาณรบกวนออกจากภาพถ่ายทางการแพทย์ ผลการทดสอบพบว่าตัวแบบที่นำเสนอสามารถกำจัดสัญญาณรบกวนออกจากภาพได้อย่างมีประสิทธิภาพ แต่ภาพผลลัพธ์อาจเกิดความเบลอจากการกำจัดสัญญาณรบกวนในบางจุดมากเกินความจำเป็น เพื่อเพิ่มประสิทธิภาพในการกำจัดสัญญาณรบกวนออกจากภาพถ่ายทางการแพทย์อาจมีการพัฒนาตัวแบบหรือวิธีการเชิงตัวเลขเพื่อลดปัญหาดังกล่าว

กิตติกรรมประกาศ ผู้แต่งขอขอบคุณผู้ทรงคุณวุฒิทุกท่านที่ได้ให้ข้อคิดเห็นและข้อเสนอแนะต่าง ๆ เพื่อปรับปรุงบทความวิจัยนี้

เอกสารอ้างอิง

[1] C. Zhao, J. Liu, and J. Zhang, *A Dual Model for restoring image corrupted by mixture of additive and multiplicative noise.* IEEE Access. **9** (2021), 168869-168888.

[2] C. Supakorn, *Image denoising for Gaussian noise using deep learning and edge feature,* Master Thesis of Chulalongkorn University. 2018.

[3] A. Ullah, W. Chen, M. A. Khan, and H. Sun, *An efficient variational method for restoring images with combined additive and multiplicative noise.* Int. J. Appl. Comput. Math. **3**(3) (2017), 1999-2019.

[4] L. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms.* Physica D. **60** (1992), 259–268.

[5] T. Goldstein and S. Osher, *The split bregman method for l1-regularized problems.* SIAM Journal on Imaging Sciences. **2**(2) (2009), 323–343.

[6] G. Aubert and P. Kornprobst, *Mathematical problems in image processing: partial differential equations and the calculus of variations*, 2nd ed., Springer, New York, 2006.

[7] N. Chumchob, K. Chen, and C. B. Loeza, A new variational model for removal of combined additive and multiplicative noise and a fast algorithm for its numerical approximation. Int. J. Comput. Math. **90**(1), (2013), 140-161.

[8] G. Aubert and J.-F. Aujol, *A variational approach to removing multiplicative noise.* SIAM J. Appl. Math. **68**(4) (2008), 925–946.

[9] Z. Jin and X. Yang, *Analysis of a new variational model for multiplicative noise removal*. J. Math. Anal. Appl. **362** (2010), 415–426.

[10] Y.L. You and M. Kaveh, *Fourth-order partial differential equations for noise removal*. IEEE Transactions on Image Processing. **9**(10) (2000), 1723–1730.

[11] O. Scherzer, *Denoising with higher order derivatives of bounded variation and an application to parameter estimation*. Computing, **60**(1) (1998), 1–27.

[12] L. I. Rudin, P. L. Lions, and S. Osher, *Multiplicative denoising and deblurring: theory and algorithms, in Geometric Level Set Methods in Imaging, Vision, and Graphics,* S. Osher and N.Paragios, Eds., pp. 103-120, Springer, Berlin, Germany, 2003.

[13] J. Shi and S. Osher, *A nonlinear inverse scale space method for a convex multiplicative noise model.* SIAM J. Imaging Sci. **1**(3) (2008), 294–321.

[14] Y. Huang, M. Ng, and Y. Wen, *A new total variation method for multiplicative noise removal*. SIAM J. Imaging Sci. **2**(1) (2009), 20–40.

[15] W. Lu, J. Duan, Z. Qiu, Z. Pan, R.W. Lid, and L. Bai, *Implementation of high-order variational models made easy for image processing.* Mathematical Methods in the Applied Sciences. **39** (2016), 4208–4233.

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

อัลกอริทึมผสมใหม่สำหรับการหาผลเฉลยของสมการไม่เชิงเส้นโดยใช้วิธีของนิวตันและวิธีแก้ตำแหน่งผิด*

ลลิตภัทร สาโรจน์[1,†] และ อภิชาติ เนียมวงษ์[2,‡]

[1]สาขาวิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา 20131

[2]ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา 20131

บทคัดย่อ

งานวิจัยนี้นำเสนออัลกอริทึมผสมใหม่สำหรับการหาผลเฉลยเชิงตัวเลขของสมการไม่เชิงเส้นที่อยู่ในรูปแบบ $f(x) = 0$ อัลกอริทึมดังกล่าวใช้วิธีของนิวตัน (ซึ่งเป็นวิธีแบบเปิด) ร่วมกับวิธีแก้ตำแหน่งผิด (ซึ่งเป็นวิธีกำหนดค่าขอบ) โดยนำข้อดีของวิธีแบบเปิดคือสามารถลู่เข้าสู่ผลเฉลยได้อย่างรวดเร็ว และข้อดีของวิธีกำหนดค่าขอบคือสามารถลู่เข้าสู่ผลเฉลยได้อย่างแน่นอน ผลการเปรียบเทียบการหาผลเฉลยเชิงตัวเลขของสมการไม่เชิงเส้นจำนวน 6 สมการ พบว่าอัลกอริทึมผสมใหม่มีจำนวนรอบในการทำซ้ำน้อยกว่าวิธีแก้ตำแหน่งผิด วิธีของนิวตัน และอัลกอริทึมผสม CJ แต่วิธีของนิวตันใช้เวลาในการคำนวณหาผลเฉลยทั้ง 6 สมการน้อยที่สุด

คำสำคัญ: ผลเฉลยของสมการไม่เชิงเส้น, วิธีแก้ตำแหน่งผิด, วิธีของนิวตัน

2020 MSC: ปฐมภูมิ 65H04 ทุติยภูมิ 65H05

1 บทนำ

ในปัจจุบันการหาผลเฉลยของสมการไม่เชิงเส้น (Nonlinear equations) นอกจากจะถูกใช้ในสาขาคณิตศาสตร์แล้ว ยังถูกใช้งานในสาขาอื่น ๆ อีกด้วย เช่น วิทยาศาสตร์ วิศวกรรมศาสตร์ เป็นต้น ซึ่งสมการดังกล่าวสามารถหาผลเฉลยของสมการได้ทั้งวิธีวิเคราะห์ (Analytic method) และวิธีเชิงตัวเลข (Numerical method) แต่ในบางสมการไม่สามารถหาผลเฉลยโดยวิธีวิเคราะห์ได้ (หรือหาได้ยาก) ดังนั้นสมการเหล่านี้จำเป็นจะต้องใช้วิธีเชิง

*งานวิจัยเรื่องนี้ได้รับทุนสนับสนุนจากภาควิชาคณิตศาสตร์ และคณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา
†ผู้นำเสนอ ‡ผู้แต่งหลัก
อีเมล: 63030242@go.buu.ac.th (ลลิตภัทร สาโรจน์), apichat@buu.ac.th (อภิชาติ เนียมวงษ์).

ตัวเลขโดยวิธีทำซ้ำ (Iterative method) เป็นทางเลือกหนึ่งที่ใช้ในการหาผลเฉลยเชิงตัวเลข (Numerical solutions) ของสมการไม่เชิงเส้น นั่นคือใช้อัลกอริทึม (Algorithm) เพื่อหาค่าของผลเฉลยเชิงตัวเลข x ที่ทำให้สมการ $f(x) = 0$ เป็นจริง และหนึ่งในวิธีทำซ้ำคือ วิธีแบ่งครึ่งช่วง (Bisection method) เป็นวิธีกำหนดค่าขอบ (Bracketing method) ซึ่งวิธีในกลุ่มนี้จะลู่เข้าสู่ผลเฉลยอย่างแน่นอน โดยมีแนวคิดในการหาผลเฉลยโดยใช้ทฤษฎีบทค่าระหว่างกลางเพื่อสร้างช่วงปิดย่อยที่มีผลเฉลยในช่วงปิดที่กำหนด $[a, b]$ จะได้ $c = \frac{a+b}{2}$ แล้วพิจารณาเงื่อนไขของการมีผลเฉลยในช่วงปิดย่อย $[a, c]$ หรือ $[c, b]$ เป็นช่วงเริ่มต้นในรอบถัดไป และ [4] ได้กล่าวว่าวิธีนี้มีอันดับการลู่เข้าเท่ากับหนึ่ง อีกวิธีหนึ่งคือวิธีแก้ตำแหน่งผิด (Regula-falsi method) เป็นวิธีกำหนดค่าขอบ ซึ่งจะต้องกำหนดขอบเขตเริ่มต้นคือช่วงปิด $[a, b]$ และมีแนวคิดในการหาผลเฉลยโดยสร้างเส้นตรงเชื่อมจุด $(a, f(a))$ และ $(b, f(b))$ แล้วใช้จุดตัดของเส้นตรงดังกล่าวกับแกน x นั่นคือ $c = \frac{af(b)-bf(a)}{f(b)-f(a)}$ เป็นค่าประมาณของผลเฉลย โดยวิธีนี้มีจำนวนรอบทำซ้ำน้อยกว่าวิธีแบ่งครึ่งช่วง นอกจากนี้ยังมีวิธีแบบเปิด (Open method) ซึ่งวิธีในกลุ่มนี้จะลู่เข้าสู่ผลเฉลยอย่างรวดเร็ว และหนึ่งในวิธีที่ใช้กันอย่างแพร่หลาย คือวิธีของนิวตัน (Newton's method) ซึ่งจะต้องกำหนดค่าเริ่มต้นคือ x_0 ที่ใกล้กับผลเฉลย และสร้างเส้นตรงสัมผัสเส้นโค้งที่จุด $(x_0, f(x_0))$ แล้วใช้จุดตัดของเส้นสัมผัสดังกล่าวกับแกน x นั่นคือ $x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$ เป็นจุดในการประมาณค่าในรอบถัดไป วิธีนี้หาผลเฉลยได้อย่างรวดเร็วและมีอันดับการลู่เข้าเท่ากับสอง

ปัจจุบันมีนักวิจัยเป็นจำนวนมากทำการปรับปรุงวิธีของนิวตันเพื่อให้มีอันดับการลู่เข้าที่มากขึ้น หนึ่งในนั้นคือ [5] นำเสนอการปรับปรุงวิธีของนิวตันจนทำให้มีอันดับการลู่เข้าเป็นสาม ต่อมา [2] ได้ทำการปรับปรุงวิธีของนิวตันจนทำให้มีอันดับการลู่เข้าเป็นสี่ อย่างไรก็ตามวิธีของนิวตันอาจเกิดการลู่ออก นั่นคือไม่สามารถหาผลเฉลยได้ ขึ้นอยู่กับการเลือกจุดเริ่มต้นที่เหมาะสม ดังนั้น [1], [3], [6] และ [7] จึงได้นำเสนออัลกอริทึมผสม (Hybrid algorithm) โดยการใช้สองวิธีร่วมกัน คือ วิธีแบ่งครึ่งช่วง และวิธีของนิวตัน โดยนำข้อดีของวิธีแบบเปิดคือสามารถลู่เข้าสู่ผลเฉลยได้อย่างรวดเร็ว และข้อดีของวิธีกำหนดค่าขอบคือสามารถลู่เข้าสู่ผลเฉลยได้อย่างแน่นอน อีกทั้งยังนำเสนอผลการเปรียบเทียบจำนวนรอบการทำซ้ำ (Number of iterations) เพื่อทดสอบประสิทธิภาพของอัลกอริทึมดังกล่าวอีกด้วย

อย่างไรก็ตามอัลกอริทึมผสมดังกล่าวนั้นยังไม่ครอบคลุมความเป็นไปได้ให้ครบทุกกรณี ดังนั้นในงานวิจัยนี้ผู้วิจัยได้นำเสนอ**อัลกอริทึมผสมใหม่ (New hybrid algorithm)** สำหรับหาผลเฉลยของสมการไม่เชิงเส้น โดยใช้สองวิธี คือ วิธีของนิวตันร่วมกับวิธีแก้ตำแหน่งผิด โดยเปรียบเทียบจำนวนรอบในการทำซ้ำ และเวลาที่ในการประมวลผล (CPU time) ของอัลกอริทึมผสมใหม่กับวิธีอื่น ๆ

2 ความรู้พื้นฐาน

2.1 งานวิจัยที่เกี่ยวข้อง

ในงานวิจัยนี้ได้นำเสนออัลกอริทึมเพื่อหาผลเฉลยของสมการไม่เชิงเส้นที่อยู่ในรูปแบบ

$$f(x) = 0 \tag{2.1}$$

โดยที่ f เป็นฟังก์ชันต่อเนื่องและหาอนุพันธ์ได้บนช่วงปิด $[a, b]$ และสมการที่ (2.1) มีผลเฉลยอย่างน้อยหนึ่งผลเฉลยในช่วงดังกล่าว โดยที่มีเงื่อนไขคือ $f(a)f(b) < 0$

ผู้วิจัยได้ศึกษาอัลกอริทึมผสมของ [6] ขั้นแรกหาค่าของ B โดยวิธีแบ่งครึ่งช่วง และนำค่า B ที่ได้ไปหาค่า N โดยวิธีของนิวตัน ขั้นต่อไปถ้า $N \in [a, b]$ นั่นคือ N ลู่เข้าสู่ผลเฉลย แล้วจะได้ผลเฉลยคือ N แต่ถ้า $N \notin [a, b]$ แล้วจะใช้ B เป็นผลเฉลย โดยอัลกอริทึมนี้เรียกว่า **อัลกอริทึมผสม KNOP** และมี 7 ขั้นตอนดังนี้

ขั้นตอน 1 กำหนด $f(x), f'(x)$, ช่วงปิด$[a, b], i = 1$ และ Tolerance error (ϵ)
ขั้นตอน 2 หาค่า $B = \frac{a+b}{2}$ และ $N = B - \frac{f(B)}{f'(B)}$ จากสูตรของวิธีแบ่งครึ่งช่วงและวิธีของนิวตัน

ขั้นตอน 3 ถ้า $N \in [a,b]$ แล้ว $x_i = N$ แต่ถ้าไม่ $x_i = B$
ขั้นตอน 4 ถ้า $|f(x_i)| < \epsilon$ และได้ผลเฉลยโดยประมาณคือ x_i แล้วไปขั้นตอนที่ 7
ขั้นตอน 5 ถ้า $f(a)f(x_i) < 0$ แล้วให้ $b = x_i$ แต่ถ้าไม่ แล้วให้ $a = x_i$
ขั้นตอน 6 กำหนดช่วงในรอบถัดไปคือ $[a,b]$ และ $i = i+1$ แล้วกลับไปขั้นตอนที่ 2
ขั้นตอน 7 หยุดการทำซ้ำ

ต่อมา [3] ได้ปรับปรุงอัลกอริทึมผสมของ [6] โดยขั้นแรกใช้วิธีของนิวตันที่ขอบช่วง $[a,b]$ เพื่อหาค่า a_1, b_1 ตามลำดับ ขั้นต่อไปถ้า a_1 หรือ $b_1 \notin [a,b]$ แล้วจะใช้ผลเฉลยของวิธีแบ่งครึ่งช่วง $[a,b]$ (ไม่ใช้ช่วง $[a_1,b_1]$ เนื่องจาก a_1 หรือ b_1 อยู่ห่างจากผลเฉลยมากกว่า a หรือ b) แต่ถ้า a_1 และ $b_1 \in [a,b]$ แล้วจะใช้ผลเฉลยของวิธีแบ่งครึ่งช่วง $[a_1,b_1]$ (เนื่องจาก a_1 และ b_1 ที่ได้จากวิธีของนิวตันอยู่ใกล้ผลเฉลยมากกว่า a และ b) โดยอัลกอริทึมนี้เรียกว่า **อัลกอริทึมผสม CJ** และมี 8 ขั้นตอนดังนี้

ขั้นตอน 1 กำหนด $f(x), f'(x),$ ช่วงปิด$[a,b], i = 1$ และ Tolerance error (ϵ)
ขั้นตอน 2 หาค่า $a_1 = a - \frac{f(a)}{f'(a)}$, $b_1 = b - \frac{f(b)}{f'(b)}$, $c_1 = \frac{a_1+b_1}{2}$ และ $c = \frac{a+b}{2}$
ขั้นตอน 3 ถ้า a_1 หรือ $b_1 \notin [a,b]$ แล้ว
$a^* = a$ และ $b^* = c$ เมื่อ $f(a)f(c) < 0$ หรือ
$a^* = c$ และ $b^* = b$ เมื่อ $f(a)f(c) > 0$
ผลเฉลยคือ $x_i = c$ และไปขั้นตอนที่ 6
ขั้นตอน 4 ถ้า $f(a_1)f(b_1) < 0$ นั่นคือผลเฉลยอยู่ในช่วง $[a_1,b_1]$ แล้ว
$a^* = a_1$ และ $b^* = c_1$ เมื่อ $f(a_1)f(c_1) < 0$ หรือ
$a^* = c_1$ และ $b^* = b_1$ เมื่อ $f(a_1)f(c_1) > 0$
ผลเฉลยคือ $x_i = c_1$ และไปขั้นตอนที่ 6
ขั้นตอน 5 ถ้า $f(a_1)f(b_1) > 0$ นั่นคือผลเฉลยไม่อยู่ในช่วง $[a_1,b_1]$ แล้ว
(พิจารณาจากการเข้าใกล้ผลเฉลย โดยการเปรียบเทียบค่าของ $|f(a_1)|$ และ $|f(b_1)|$)
ผลเฉลยคือ $x_i = a_1$ เมื่อ $|f(a_1)| < |f(b_1)|$ หรือ
ผลเฉลยคือ $x_i = b_1$ เมื่อ $|f(a_1)| > |f(b_1)|$
ให้ $a^* = a_1$, $b^* = b_1$ และไปขั้นตอนที่ 6
ขั้นตอน 6 ถ้า $|f(x_i)| < \epsilon$ และได้ x_i เป็นผลเฉลยโดยประมาณ แล้วไปขั้นตอนที่ 8
ขั้นตอน 7 กำหนดช่วงในรอบถัดไปคือ $[a,b] = [a^*,b^*]$ และ $i = i+1$ แล้วกลับไปขั้นตอนที่ 2
ขั้นตอน 8 หยุดการทำซ้ำ

ข้อสังเกตุ ในขั้นตอนที่ 4 ของอัลกอริทึมผสม CJ มีความเป็นไปได้ที่ค่าของ a^* อาจจะมากกว่า b^* ซึ่งทำให้การกำหนดช่วง $[a,b]$ ในรอบถัดไปไม่ถูกต้อง (โดยทั่วไปวิธีกำหนดขอบค่าของ $a < b$ เสมอ) และในขั้นตอนที่ 5 การเปรียบเทียบค่าของ $|f(a_1)|$ และ $|f(b_1)|$ ไม่สามารถระบุได้ว่าค่าของ a_1 และ b_1 ค่าใดมีค่าเข้าใกล้ผลเฉลยมากกว่ากัน

3 อัลกอริทึมผสมใหม่ HA

ผู้วิจัยได้ปรับปรุงอัลกอริทึมผสมของ [3] โดยเปลี่ยนจากวิธีแบ่งครึ่งช่วงเป็นวิธีแก้ตำแหน่งผิด และปรับกระบวนการของ [3] ให้มีเงื่อนไขครอบคลุมทุกกรณี (ดังภาพที่ 1-11) คือ กรณี 1: a_1 และ $b_1 \notin [a,b]$, กรณี 2: $a_1 \in [a,b]$ แต่ $b_1 \notin [a,b]$, กรณี 3: $b_1 \in [a,b]$ แต่ $a_1 \notin [a,b]$ และ กรณี 4: a_1 และ $b_1 \in [a,b]$ เรียกว่า **อัลกอริทึมผสมใหม่ HA (New Hybrid Algorithm)** โดยมี 10 ขั้นตอนดังนี้

ขั้นตอน 1 กำหนด $f(x), f'(x)$, ช่วงปิด$[a,b], i=1$ และ Tolerance error (ϵ)

ขั้นตอน 2 หาค่า $a_1 = a - \frac{f(a)}{f'(a)}$, $b_1 = b - \frac{f(b)}{f'(b)}$,

$RF_1 = \frac{af(b)-bf(a)}{f(b)-f(a)}$, $RF_2 = \frac{af(a_1)-a_1f(a)}{f(a_1)-f(a)}$, $RF_3 = \frac{a_1f(b)-bf(a_1)}{f(b)-f(a_1)}$,

$RF_4 = \frac{af(b_1)-b_1f(a)}{f(b_1)-f(a)}$, $RF_5 = \frac{b_1f(b)-bf(b_1)}{f(b)-f(b_1)}$ และ , $RF_6 = \frac{a_1f(b_1)-b_1f(a_1)}{f(b_1)-f(a_1)}$

ขั้นตอน 3 (กรณี 1) ถ้า a_1 และ $b_1 \notin [a,b]$ (ดังภาพที่ 1) แล้วหา RF_1 บนช่วงปิด $[a,b]$

$a^* = a$ และ $b^* = RF_1$ เมื่อ $f(a)f(RF_1) < 0$ หรือ

$a^* = RF_1$ และ $b^* = b$ เมื่อ $f(a)f(RF_1) > 0$

ผลเฉลยคือ $x_i = RF_1$ และไปขั้นตอนที่ 8

ขั้นตอน 4 (กรณี 2) ถ้า $a_1 \in [a,b]$ แต่ $b_1 \notin [a,b]$ และ

(กรณี 2A) ถ้า $f(a)f(a_1) < 0$ (ดังภาพที่ 2) แล้วหา RF_2 บน $[a,a_1]$

$a^* = a$ และ $b^* = RF_2$ เมื่อ $f(a)f(RF_2) < 0$ หรือ

$a^* = RF_2$ และ $b^* = b$ เมื่อ $f(a)f(RF_2) > 0$

ผลเฉลยคือ $x_i = RF_2$ และไปขั้นตอนที่ 8

(กรณี 2B) แต่ถ้า $f(a)f(a_1) > 0$ (ดังภาพที่ 3) แล้วหา RF_3 บน $[a_1,b]$

$a^* = a_1$ และ $b^* = RF_3$ เมื่อ $f(a_1)f(RF_3) < 0$ หรือ

$a^* = RF_3$ และ $b^* = b$ เมื่อ $f(a_1)f(RF_3) > 0$

ผลเฉลยคือ $x_i = RF_3$ และไปขั้นตอนที่ 8

ขั้นตอน 5 (กรณี 3) ถ้า $b_1 \in [a,b]$ แต่ $a_1 \notin [a,b]$ และ

(กรณี 3A) ถ้า $f(a)f(b_1) < 0$ (ดังภาพที่ 4) แล้วหา RF_4 บน $[a,b_1]$

$a^* = a$ และ $b^* = RF_4$ เมื่อ $f(a)f(RF_4) < 0$ หรือ

$a^* = RF_4$ และ $b^* = b_1$ เมื่อ $f(a)f(RF_4) > 0$

ผลเฉลยคือ $x_i = RF_4$ และไปขั้นตอนที่ 8

(กรณี 3B) แต่ถ้า $f(a)f(b_1) > 0$ (ดังภาพที่ 5) แล้วหา RF_5 บน $[b_1,b]$

$a^* = b_1$ และ $b^* = RF_5$ เมื่อ $f(b_1)f(RF_5) < 0$ หรือ

$a^* = RF_5$ และ $b^* = b$ เมื่อ $f(b_1)f(RF_5) > 0$

ผลเฉลยคือ $x_i = RF_5$ และไปขั้นตอนที่ 8

ขั้นตอน 6 (กรณี 4.1) ถ้า a_1 และ $b_1 \in [a,b]$ & $f(a_1)f(b_1) < 0$

(กรณี 4.1A) ถ้า $f(a)f(a_1) > 0$ (ดังภาพที่ 6) แล้วหา RF_6 บน $[a_1,b_1]$

$a^* = a_1$ และ $b^* = RF_6$ เมื่อ $f(a_1)f(RF_6) < 0$ หรือ

$a^* = RF_6$ และ $b^* = b_1$ เมื่อ $f(a_1)f(RF_6) > 0$

ผลเฉลยคือ $x_i = RF_6$ และไปขั้นตอนที่ 8

(กรณี 4.1B) ถ้า $f(a)f(a_1) < 0$ (ดังภาพที่ 7) แล้วหา RF_6 บน $[b_1,a_1]$

$a^* = b_1$ และ $b^* = RF_6$ เมื่อ $f(b_1)f(RF_6) < 0$ หรือ

$a^* = RF_6$ และ $b^* = a_1$ เมื่อ $f(b_1)f(RF_6) > 0$

ผลเฉลยคือ $x_i = RF_6$ และไปขั้นตอนที่ 8

ขั้นตอน 7 (กรณี 4.2) ถ้า a_1 และ $b_1 \in [a,b]$ & $f(a_1)f(b_1) > 0$

(กรณี 4.2A) ถ้า $f(a)f(a_1) < 0$ & $a_1 < b_1$ (ดังภาพที่ 8) แล้วหา RF_2 บน $[a,a_1]$

$a^* = a$ และ $b^* = RF_2$ เมื่อ $f(a)f(RF_2) < 0$ หรือ

$a^* = RF_2$ และ $b^* = a_1$ เมื่อ $f(a)f(RF_2) > 0$

ผลเฉลยคือ $x_i = RF_2$ และไปขั้นตอนที่ 8

(กรณี 4.2B) ถ้า $f(a)f(a_1) < 0$ & $b_1 < a_1$ (ดังภาพที่ 9) แล้วหา RF_4 บน $[a,b_1]$

$a^* = a$ และ $b^* = RF_4$ เมื่อ $f(a)f(RF_4) < 0$ หรือ

$a^* = RF_4$ และ $b^* = b_1$ เมื่อ $f(a)f(RF_4) > 0$

ผลเฉลยคือ $x_i = RF_4$ และไปขั้นตอนที่ 8

(กรณี 4.2C) ถ้า $f(a)f(a_1) > 0$ & $b_1 > a_1$ (ดังภาพที่ 10) แล้วหา RF_5 บน $[b_1, b]$

$a^* = b_1$ และ $b^* = RF_5$ เมื่อ $f(b_1)f(RF_5) < 0$ หรือ

$a^* = RF_5$ และ $b^* = b$ เมื่อ $f(b_1)f(RF_5) > 0$

ผลเฉลยคือ $x_i = RF_5$ และไปขั้นตอนที่ 8

(กรณี 4.2D) ถ้า $f(a)f(a_1) > 0$ & $a_1 > b_1$ (ดังภาพที่ 11) แล้วหา RF_3 บน $[a_1, b]$

$a^* = a_1$ และ $b^* = RF_3$ เมื่อ $f(a_1)f(RF_3) < 0$ หรือ

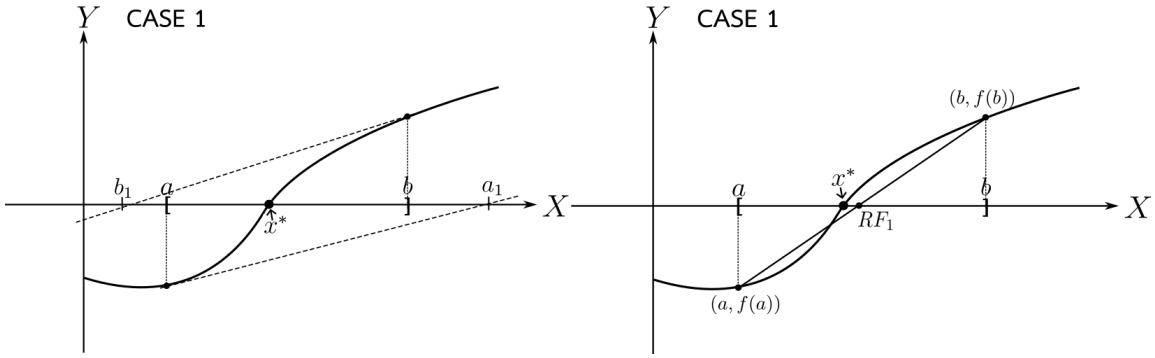
$a^* = RF_3$ และ $b^* = b$ เมื่อ $f(a_1)f(RF_3) > 0$

ผลเฉลยคือ $x_i = RF_3$ และไปขั้นตอนที่ 8

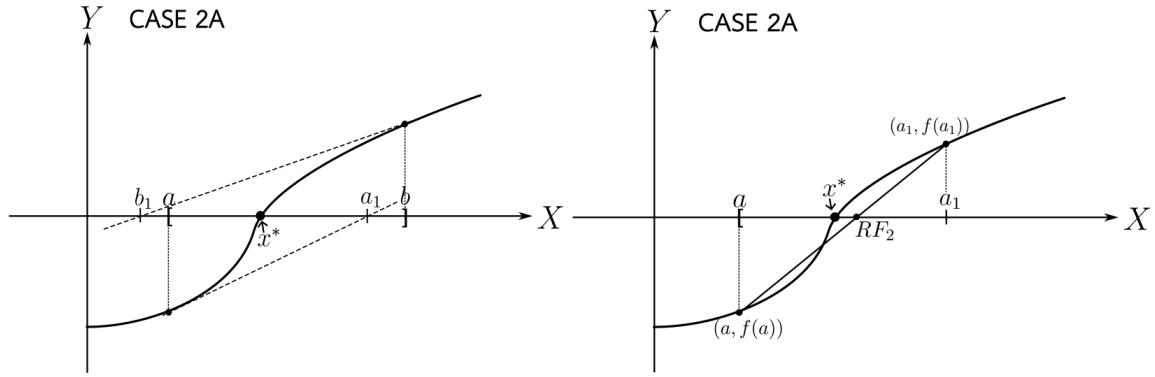
ขั้นตอน 8 ถ้าค่าคลาดเคลื่อนสัมพัทธ์ $\frac{|f(x_i)-f(x_{i-1})|}{|f(x_i)|} < \epsilon$ และได้ x_i เป็นผลเฉลยโดยประมาณ แล้วไปขั้นตอนที่ 10

ขั้นตอน 9 กำหนดช่วงในรอบถัดไปคือ $[a, b] = [a^*, b^*]$ และ $i = i + 1$ แล้วกลับไปขั้นตอนที่ 2

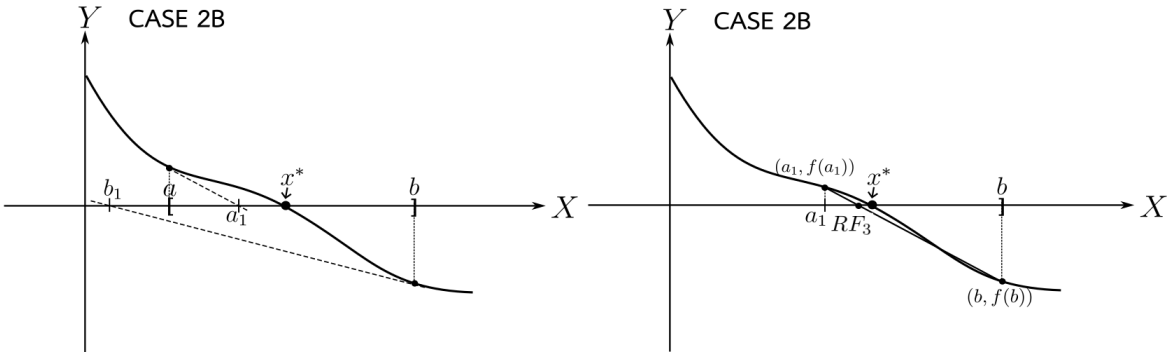
ขั้นตอน 10 หยุดการทำซ้ำ



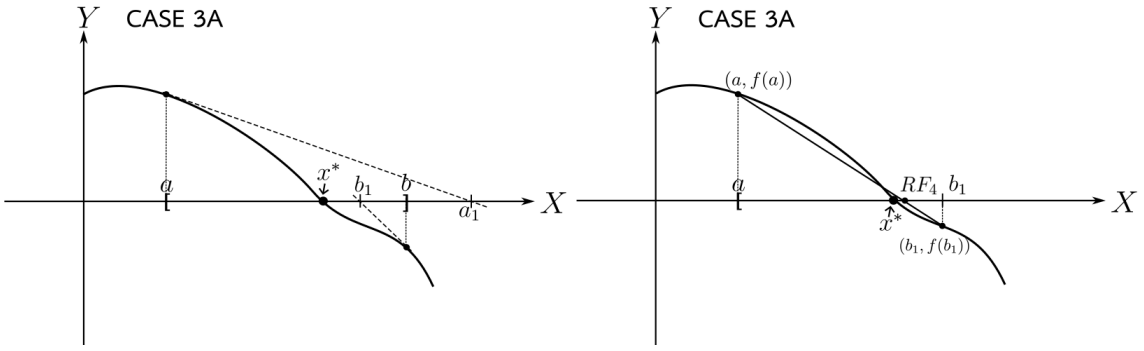
ภาพที่ 1: กราฟ (ซ้าย) แสดง กรณี 1: a_1 และ $b_1 \notin [a, b]$ และ (ขวา) แสดงจุด RF_1 ที่ได้จากวิธีแก้ตำแหน่งผิดบนช่วงปิด $[a, b]$ เมื่อ x^* คือผลเฉลยแท้จริง



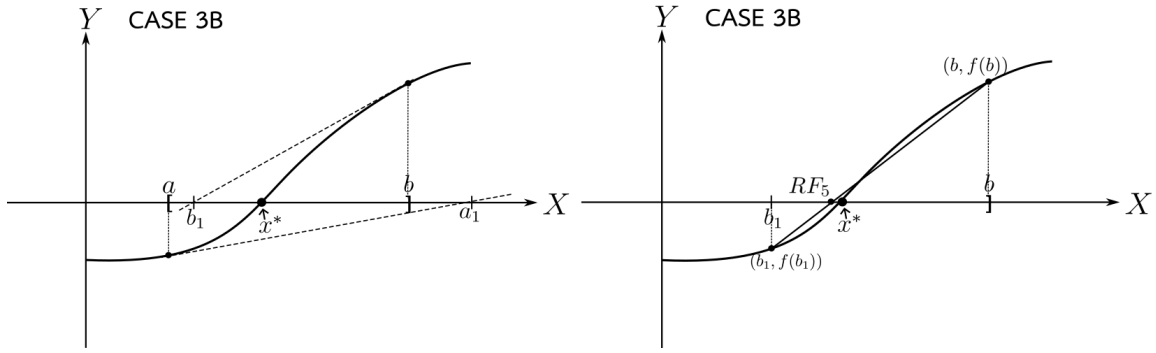
ภาพที่ 2: กราฟ (ซ้าย) แสดง กรณี 2A: $a_1 \in [a, b]$ แต่ $b_1 \notin [a, b]$ & $f(a)f(a_1) < 0$ และ (ขวา) แสดงจุด RF_2 ที่ได้จากวิธีแก้ตำแหน่งผิดบนช่วงปิด $[a, a_1]$ เมื่อ x^* คือผลเฉลยแท้จริง



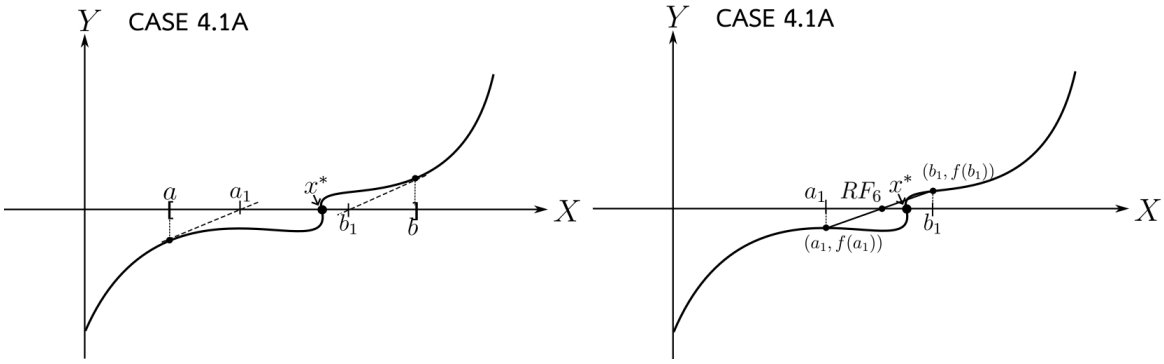
ภาพที่ 3: กราฟ (ซ้าย) แสดง กรณี 2B: $a_1 \in [a,b]$ แต่ $b_1 \notin [a,b]$ & $f(a)f(a_1) > 0$ และ (ขวา) แสดงจุด RF_3 ที่ได้จากวิธีแก้ตำแหน่งผิดบนช่วงปิด $[a_1, b]$ เมื่อ x^* คือผลเฉลยแท้จริง



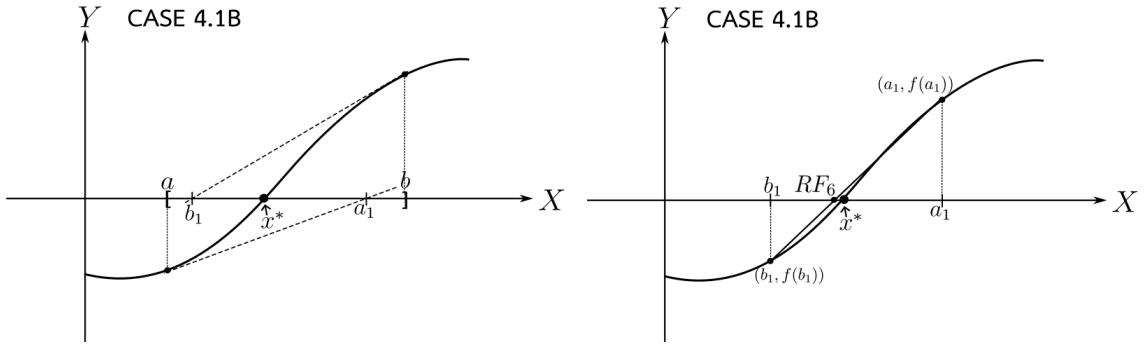
ภาพที่ 4: กราฟ (ซ้าย) แสดง กรณี 3A: $b_1 \in [a,b]$ แต่ $a_1 \notin [a,b]$ & $f(a)f(b_1) < 0$ และ (ขวา) แสดงจุด RF_4 ที่ได้จากวิธีแก้ตำแหน่งผิดบนช่วงปิด $[a, b_1]$ เมื่อ x^* คือผลเฉลยแท้จริง



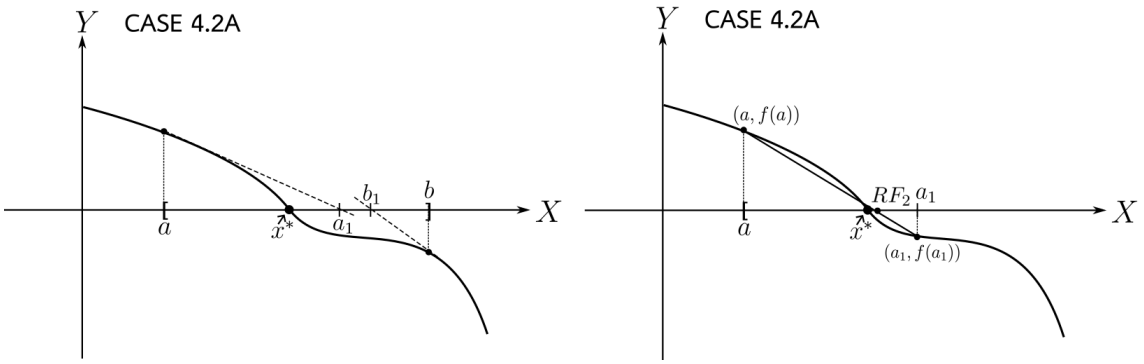
ภาพที่ 5: กราฟ (ซ้าย) แสดง กรณี 3B: $b_1 \in [a,b]$ แต่ $a_1 \notin [a,b]$ & $f(a)f(b_1) > 0$ และ (ขวา) แสดงจุด RF_5 ที่ได้จากวิธีแก้ตำแหน่งผิดบนช่วงปิด $[b_1, b]$ เมื่อ x^* คือผลเฉลยแท้จริง



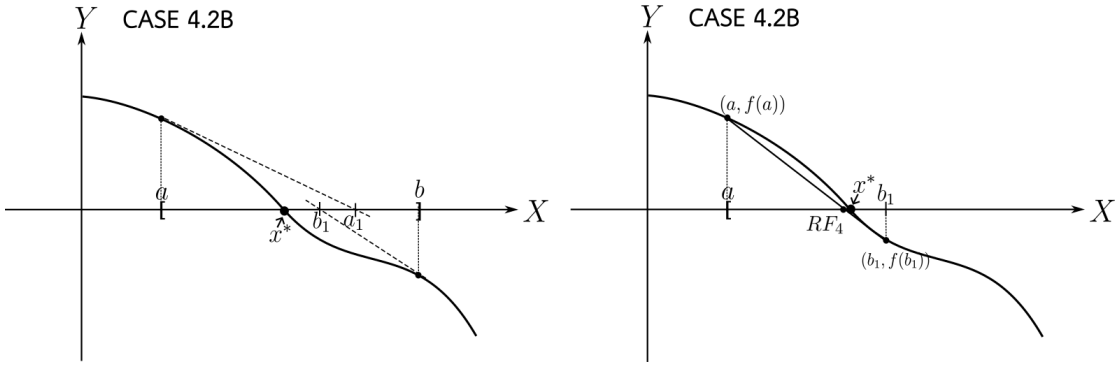
ภาพที่ 6: กราฟ (ซ้าย) แสดง กรณี 4.1A: a_1 และ $b_1 \in [a, b]$ & $f(a_1)f(b_1) < 0$ & $f(a)f(a_1) > 0$ และ (ขวา) แสดงจุด RF_6 ที่ได้จากวิธีแก้ตำแหน่งผิดบนช่วงปิด $[a_1, b_1]$ เมื่อ x^* คือผลเฉลยแท้จริง



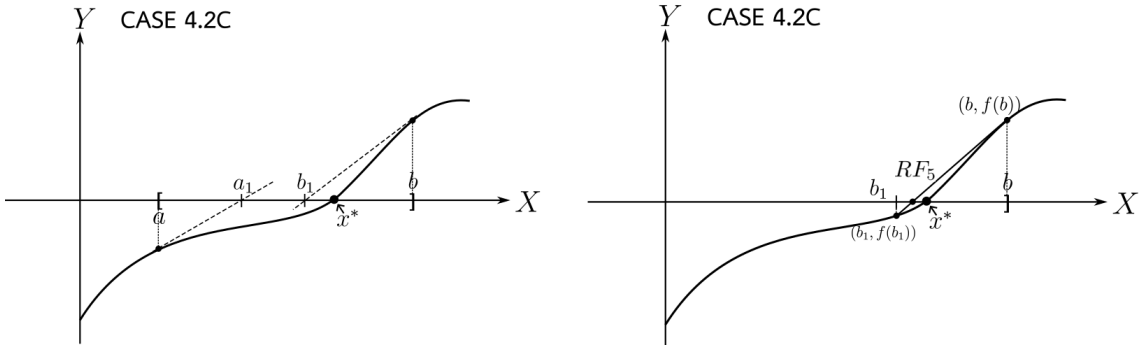
ภาพที่ 7: กราฟ (ซ้าย) แสดง กรณี 4.1B: a_1 และ $b_1 \in [a, b]$ & $f(a_1)f(b_1) < 0$ & $f(a)f(a_1) < 0$ และ (ขวา) แสดงจุด RF_6 ที่ได้จากวิธีแก้ตำแหน่งผิดบนช่วงปิด $[b_1, a_1]$ เมื่อ x^* คือผลเฉลยแท้จริง



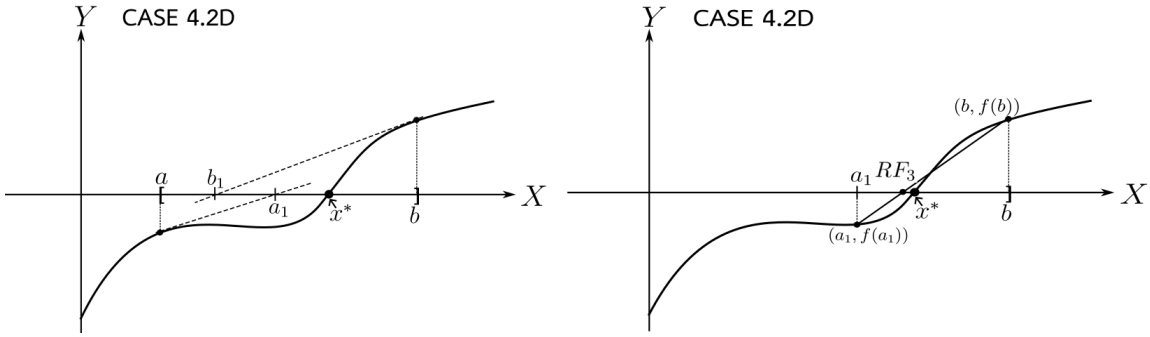
ภาพที่ 8: กราฟ (ซ้าย) แสดง กรณี 4.2A: a_1 และ $b_1 \in [a, b]$ & $f(a_1)f(b_1) > 0$ & $f(a)f(a_1) < 0$ & $a_1 < b_1$ และ (ขวา) แสดงจุด RF_2 ที่ได้จากวิธีแก้ตำแหน่งผิดบนช่วงปิด $[a, a_1]$ เมื่อ x^* คือผลเฉลยแท้จริง



ภาพที่ 9: กราฟ (ซ้าย) แสดง กรณี 4.2B: a_1 และ $b_1 \in [a, b]$ & $f(a_1)f(b_1) > 0$ & $f(a)f(a_1) < 0$ & $b_1 < a_1$ และ (ขวา) แสดงจุด RF_4 ที่ได้จากวิธีแก้ตำแหน่งผิดบนช่วงปิด $[a, b_1]$ เมื่อ x^* คือผลเฉลยแท้จริง



ภาพที่ 10: กราฟ (ซ้าย) แสดง กรณี 4.2C: a_1 และ $b_1 \in [a, b]$ & $f(a_1)f(b_1) > 0$ & $f(a)f(a_1) > 0$ & $b_1 > a_1$ และ (ขวา) แสดงจุด RF_5 ที่ได้จากวิธีแก้ตำแหน่งผิดบนช่วงปิด $[b_1, b]$ เมื่อ x^* คือผลเฉลยแท้จริง



ภาพที่ 11: กราฟ (ซ้าย) แสดง กรณี 4.2D: a_1 และ $b_1 \in [a, b]$ & $f(a_1)f(b_1) > 0$ & $f(a)f(a_1) > 0$ & $a_1 > b_1$ และ (ขวา) แสดงจุด RF_3 ที่ได้จากวิธีแก้ตำแหน่งผิดบนช่วงปิด $[a_1, b]$ เมื่อ x^* คือผลเฉลยแท้จริง

4 ผลการศึกษา

ในตารางที่ 1 แสดงผลการเปรียบเทียบจำนวนรอบการทำซ้ำ (รอบ) เวลาที่ใช้ในการประมวลผล (วินาที) ผลเฉลยโดยประมาณ และค่าคลาดเคลื่อนสัมพัทธ์ ของวิธีแก้ตำแหน่งผิด (RF), วิธีของนิวตัน (NR), อัลกอริทึมผสม CJ และอัลกอริทึมผสมใหม่ HA ในการหาผลเฉลยของสมการไม่เชิงเส้นที่อยู่ในรูปแบบสมการที่ (2.1) จำนวน 6 สมการ และมีผลเฉลยอย่างน้อยหนึ่งผลเฉลยบนช่วงปิด $[a,b]$ (ให้ a เป็นจุดเริ่มต้นสำหรับวิธีของนิวตัน) โดยกำหนดไว้คือ $f_1(x) : \sin(x^3) = 0$ บนช่วง $[1.4, 1.6]$, $f_2(x) : e^{\sin(x)} - 2x = 0$ บนช่วง $[0.1, 4.0]$, $f_3(x) : -x^5 + 3x - 2 = 0$ บนช่วง $[-2.0, -1.0]$, $f_4(x) : x^3 - 8x - 4 = 0$ บนช่วง $[2.0, 4.0]$, $f_5(x) : \cos(x) - x^3 = 0$ บนช่วง $[0.1, 1.0]$ และ $f_6(x) : 10xe^{-x^2} - 1 = 0$ บนช่วง $[1.0, 2.0]$ โดยใช้ค่าคลาดเคลื่อนสัมพัทธ์ $\frac{|x_i - x_{i-1}|}{|x_i|}$ และ Tolerance error, $\epsilon = 1 \times 10^{-10}$

และในตารางที่ 2 เป็นการเปรียบเทียบเวลาที่ใช้ในการประมวลผลและค่าคลาดเคลื่อนสัมพัทธ์ โดยการกำหนดจำนวนรอบการทำซ้ำที่เท่ากันคือ 3 รอบ

ตารางที่ 1: แสดงจำนวนรอบการทำซ้ำ (รอบ) เวลาที่ใช้ในการประมวลผล (วินาที) ผลเฉลยโดยประมาณ และค่าคลาดเคลื่อนสัมพัทธ์ของวิธีแก้ตำแหน่งผิด (RF), วิธีของนิวตัน (NR), อัลกอริทึมผสม CJ และอัลกอริทึมผสมใหม่ HA โดยใช้ $\epsilon = 1 \times 10^{-10}$

ฟังก์ชัน $f(x) = 0$	วิธี	จำนวนรอบ	เวลา (sec.)	ผลเฉลยโดยประมาณ	ค่าคลาดเคลื่อน
$f_1(x) : \sin(x^3) = 0$	RF	5	0.000591	1.46459188756152	1.7586588e-14
	NR	5	0.000523	1.46459188756152	1.5160852e-16
	CJ	7	0.001377	1.46459188756152	4.3194631e-11
	HA	3	0.001542	1.46459188756152	1.0712961e-11
$f_2(x) : e^{\sin(x)} - 2x = 0$	RF	38	0.026655	1.31594217966101	5.6688825e-11
	NR	6	0.000416	1.31594217974302	7.5420865e-12
	CJ	8	0.001287	1.31594217974302	∼0.0000000
	HA	5	0.001683	1.31594217974302	∼0.0000000
$f_3(x) : -x^5 + 3x - 2 = 0$	RF	39	0.001655	-1.44685724776065	8.1818542e-11
	NR	7	0.000619	-1.44685724791387	1.5346684e-16
	CJ	13	0.001382	-1.44685724791387	9.0545434e-15
	HA	5	0.001979	-1.44685724791387	3.7138975e-14
$f_4(x) : x^3 - 8x - 4 = 0$	RF	21	0.001068	3.05137424163987	6.2137358e-11
	NR	8	0.000580	3.05137424173104	∼0.0000000
	CJ	10	0.001513	3.05137424173104	4.4869194e-13
	HA	4	0.001851	3.05137424173104	8.3080050e-11
$f_5(x) : \cos(x) - x^3 = 0$	RF	13	0.000978	0.86547403310019	1.2135339e-11
	NR	11	0.000804	0.86547403310161	1.1332183e-12
	CJ	9	0.001434	0.86547403310161	∼0.0000000
	HA	4	0.001744	0.86547403310161	1.2827918e-16
$f_6(x) : 10xe^{-x^2} - 1 = 0$	RF	19	0.001179	1.67963061048606	8.0497634e-11
	NR	5	0.000573	1.67963061042845	1.8677001e-11
	CJ	8	0.001393	1.67963061042845	1.4647591e-13
	HA	4	0.001699	1.67963061042845	1.3219848e-16

ตารางที่ 2: แสดงจำนวนรอบการทำซ้ำ (รอบ) เวลาที่ใช้ในการประมวลผล (วินาที) ผลเฉลยโดยประมาณ และค่าคลาดเคลื่อนสัมพัทธ์ของวิธีแก้ตำแหน่งผิด (RF), วิธีของนิวตัน (NR), อัลกอริทึมผสม CJ และอัลกอริทึมผสมใหม่ HA โดยกำหนดให้จำนวนรอบการทำซ้ำ คือ 3 รอบ

ฟังก์ชัน $f(x)=0$	วิธี	จำนวนรอบ	เวลา (sec.)	ผลเฉลยโดยประมาณ	ค่าคลาดเคลื่อน
$f_1(x):\ \sin(x^3)=0$	RF	3	0.000437	1.46459188514296	1.0555032e-05
	NR	3	0.000372	1.46459188806748	1.8591463e-05
	CJ	3	0.000782	1.46459188806748	1.8591463e-05
	HA	3	0.001542	1.46459188756152	1.0712961e-11
$f_2(x):\ e^{\sin(x)}-2x=0$	RF	3	0.000432	0.96338604616086	1.9337086e-01
	NR	3	0.000324	1.31788083269659	3.4873415e-02
	CJ	3	0.000762	1.31788083269659	2.0363591e-01
	HA	3	0.001429	1.31594217514529	1.3641307e-03
$f_3(x):\ -x^5+3x-2=0$	RF	3	0.000607	-1.32671339825160	5.6786598e-02
	NR	3	0.000442	-1.45284370164767	4.0375926e-02
	CJ	3	0.000825	-1.45284370164767	4.0375926e-02
	HA	3	0.001762	-1.44682970720008	1.4930916e-02
$f_4(x):\ x^3-8x-4=0$	RF	3	0.000529	2.99583866628475	3.6289087e-02
	NR	3	0.000413	3.21704489807048	1.7842458e-01
	CJ	3	0.001103	3.05163835253246	7.8380862e-03
	HA	3	0.001732	3.05137424147753	3.4831530e-04
$f_5(x):\ \cos(x)-x^3=0$	RF	3	0.000473	0.86309198477770	2.0046884e-02
	NR	3	0.000402	3.44334400191801	5.0591468e-01
	CJ	3	0.000892	0.86547407595298	2.4274236e-04
	HA	3	0.001568	0.86547403310161	4.5812050e-06
$f_6(x):\ 10xe^{-x^2}-1=0$	RF	3	0.000434	1.69357702223456	1.7942062e-02
	NR	3	0.000369	1.67962488207810	1.4571204e-03
	CJ	3	0.000710	1.67962488207810	1.4571204e-03
	HA	3	0.001507	1.67963061042845	2.0359108e-07

5 สรุปผลและข้อเสนอแนะ

ผู้วิจัยได้นำเสนออัลกอริทึมผสมใหม่เป็นวิธีทำซ้ำเพื่อหาผลเฉลยเชิงตัวเลขของสมการไม่เชิงเส้นจำนวน 6 สมการ มีผลการเปรียบเทียบดังตารางที่ 1 และผู้วิจัยพบว่าอัลกอริทึมผสมใหม่ HA มีจำนวนรอบของการทำซ้ำคือ 3, 5, 5, 4, 4 และ 4 รอบ ตามลำดับ ซึ่งน้อยกว่าวิธีแก้ตำแหน่งผิด วิธีของนิวตัน และอัลกอริทึมผสม CJ [3] โดยทั้ง 4 วิธีให้ผลเฉลยโดยประมาณใกล้เคียงกัน แต่วิธีของนิวตันใช้เวลาในการคำนวณหาผลเฉลยทั้ง 6 สมการน้อยที่สุด คือ 0.000523, 0.000416, 0.000619, 0.000580, 0.000804 และ 0.000573 วินาที ตามลำดับ และเมื่อกำหนดให้ทุกวิธีมีจำนวนรอบการทำซ้ำ 3 รอบเท่ากันในตารางที่ 2 พบว่าอัลกอริทึมผสมใหม่ HA มีค่าคลาดเคลื่อนสัมพัทธ์น้อยที่สุด คือ 1.0712961e-11, 1.3641307e-03, 1.4930916e-02, 3.4831530e-04, 4.5812050e-06 และ 2.0359108e-07 ตามลำดับ อย่างไรก็ตามผู้วิจัยมีข้อสังเกตคืออัลกอริทึมผสมใหม่ HA มีขั้นตอนที่ซับซ้อนเมื่อเทียบกับทั้ง 3 วิธีที่นำมาเปรียบเทียบ เพราะว่าจะต้องกำหนดเงื่อนไขให้ครอบคลุมทุกกรณีที่เป็นไปได้ ดังนั้นผู้วิจัยคิดว่าอัลกอริทึมผสมใหม่ HA นี้ควรจะถูกปรับปรุงให้มีความกระชับมากกว่านี้ในอนาคต อีกทั้งควรเพิ่มผลการวิเคราะห์เรื่องอัตราและอันดับของการลู่เข้า (Rate and order of convergence) ของอัลกอริทึมผสมใหม่ HA ด้วย

กิตติกรรมประกาศ ผู้วิจัยขอขอบคุณผู้ทรงคุณวุฒิทุกท่านที่ได้ให้ข้อคิดเห็นและข้อเสนอแนะต่าง ๆ เพื่อปรับปรุงบทความวิจัยนี้ และขอบคุณภาควิชาคณิตศาสตร์ และคณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา ที่ให้ทุนวิจัยและนำเสนอผลงานนี้

เอกสารอ้างอิง

[1] A. Altaee, K. Hoomod and K. Hussein, *A new approach to find roots of nonlinear equations by hybrid algorithm to bisection and Newton-Raphson algorithms*, Iraq. J. Inform. Tech. **1**(7) (2015), 75–82.

[2] C. Chun and B. Neta, *Some modification of Newton's method by the method of undetermined coefficients*, Comput. and Math. Appl. **56** (2008), 2528–2538.

[3] C. Comemuang and P. Janngam, *Hybrid algorithm to Newton Raphson method and bisection method*, J. Math. Sci. **11**(6) (2021), 7082–7088.

[4] J. C. Ehiwario and S. O. Aghamie, *Comparative study of bisection, Newton-Raphson and secant methods of root-finding problems*, IOSR J. of Eng. **4**(4) (2014), 1–7.

[5] H. Homeier, *On Newton-type methods with cubic convergence*, J. Comput. and Appl. Math. **11**(1) (2005), 2528–2538.

[6] J. Kim, T. Noh, W. Oh, S. Park and N. Hahm, *An improved hybrid algorithm to bisection method and Newton-Raphson method*, Appl. Math. Sci. **11**(56) (2017), 2789–2797.

[7] S. Tanakan, *A new algorithm of modified bisection method for nonlinear equation*, Appl. Math. Sci. **7**(123) (2013), 6107–6114.

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Applying the Residual Power Series Method to a Time Fractional Black Scholes European Option Pricing with Two Assets

Pitsinee Winyarat[1,†] and Panumart Sawangtong[1,2,‡]

[1]Department of Mathematics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand

[2]Research group for fractional calculus theory and applications, Science and Technology Research Institute, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand

Abstract

The Black-Scholes pricing model is a significant tool for the financial market to predict the current value of the European call option. In this paper, the Black-Scholes model with two assets is modified in the form of time fractional derivative. The fractional derivative used here is the Caputo derivative. An approximate analytical solution of the fractional Black-Scholes European option pricing with two assets is investigated by utilizing the residual power series method (RPSM). An analytical solution for such a fractional problem is in the form of a special function, the Mittag-Leffler function. The RPSM technique is to assume the solution of differential equations as a fractional power series and then solve for the coefficients of the series iteratively under certain requirements. Another primary outcome demonstrates that the RPSM approach is not more complicated but is more effective in solving both differential equations and fractional differential equations.

Keywords: Residual power series method, fractional Black-Scholes equation, approximate analytical solution, fractional power series.

2020 MSC: Primary 26A33; Secondary 30B10, 32A05, 33E12.

1 Introduction

Options trading can be an effective way for investors to speculate on the future direction of the overall stock market or individual securities, like stocks or bonds. Options contracts give traders the flexibility to buy or sell an underlying asset at a specified price (also known as the strike price) by a specified date without actually having to buy the asset. This means that traders can potentially profit from changes in the price of an asset without having to invest large amounts of capital upfront. Mainly, options can be isolated into two types of options: put and call [1]. The terms "put option" and "call option" refer to contract options that give the holder the right,

†Speaker. ‡Corresponding author.
Email: s6304021820082@email.kmutnb.ac.th (P.Winyarat), panumart.s@sci.kmutnb.ac.th (P.Sawangtong).

but not the obligation, to buy or sell an underlying asset at a predetermined price (strike price) within a certain time period.

Put Option: a put option gives the holder the right to sell the underlying asset at the strike price. The put option holder benefits when the market price of the underlying asset is lower than the strike price.

Call Option: a call option gives the holder the right to buy the underlying asset at the strike price. The call option holder benefits when the market price of the underlying asset is higher than the strike price.

Put and call options allow investors to manage risk and implement investment strategies based on their expectations of the future price movements of the underlying asset.

This study will focus on the European call option, a contract that allows the buyer or seller to execute the option only at its expiration date, following the assumptions of the Black-Scholes model (BSM) [2].

The BSM is a widely used options pricing model that was developed by Fisher Black, Myron Scholes, and Robert Merton in 1973, referred to as [3, 4]. The Black-Scholes model is used to calculate the premium value of a call or put option based on current stock prices, expected dividends, the option's strike price, anticipated interest rates, expiry date, and volatility. By using the BSM, investors can calculate the fair price of options and make informed tool for investors looking to manage risk and maximize returns in today's complex financial markets.

In the past, there have been many researchers studying the BSM by various analytical methods [5–10] such as the generalized differential transform technique, the Homotopy perturbation scheme, the Adomian decomposition scheme, the variation iteration method, etc. The residual power series (RPS) approach is a powerful numeric-analytic technique that has been developed to solve a wide range of ordinary, partial, fuzzy differential equations, integral-differential equations, and integral-differential equations of fractional order. This approach is particularly effective because it provides closed-form solutions in terms of known functions, making it an attractive optimization technique for solving complex problems. For example, it has been successfully applied to solve time-fractional Fokker-Planck models, Newell-Whitehead-Segel equations of fractional order, and fractional integral equations, among others. This versatility makes it a valuable tool in many areas of science and engineering [11]. Another advantage of the RPS approach is its ability to provide accurate and efficient solutions to non-linear problems. By coupling analytical approaches with the Laplace transform operator, the RPS approach can increase the accuracy of solutions and reduce the time required to solve complex problems. This is particularly important in fields of natural science where accurate solutions are crucial for understanding complex phenomena.

In the following, we provide a brief overview of the history of fractional calculus and the reasons why authors use fractional derivatives to develop the BS model in fractional-order form.

Fractional calculus was first introduced in 1695, when L'Hopital's wrote a letter to Leibniz [12] asking what would be the result of $\frac{\partial^{\frac{1}{2}} x}{\partial x^{\frac{1}{2}}}$. Since then, numerous mathematicians have started to be interested in this field. Led to many famous fractional derivatives that were developed [13], for example, Caputo fractional derivatives, Reiman-Liouville fractional derivatives, Atangana-Baleanu fractional derivatives, etc. Fractional-order ordinary and partial differential equations have been applied to many fields, such as finance, engineering, and science. Fractional-order derivatives and integrals have indeed proven to be very useful in describing the memory and hereditary properties of various real-world processes [14]. This is because these derivatives and integrals can capture the behavior of systems that exhibit non-local, long-range, or multi-scale dependencies. In other words, they can account for the way in which past events influence the behavior of the system at any given moment in time, even if those events occurred a long time ago [15].

In this article, we will find the approximate analytical solution of a two-dimensional fractional Black-Scholes European option pricing equation based on the Caputo derivative by applying the

RPS method.

The Black-Scholes European option pricing equation with two assets is in the following form [16]:

$$\frac{\partial u}{\partial t} + \frac{1}{2}\sigma_1^2 S_1^2 \frac{\partial^2 u}{\partial S_1^2} + \frac{1}{2}\sigma_2^2 S_2^2 \frac{\partial^2 u}{\partial S_2^2} + wS_1S_2\sigma_1\sigma_2 \frac{\partial^2 u}{\partial S_1 S_2} + l(S_1 \frac{\partial u}{\partial S_1} + S_2 \frac{\partial u}{\partial S_2}) - lu = 0 \tag{1.1}$$

for $S_1, S_2 \in (0, \infty)$, and $t \in [0, T]$. The terminal condition is given by:

$$u(S_1, S_2, T) = \max(\beta_1 S_1 + \beta_2 S_2 - K, 0), \tag{1.2}$$

where the meaning of each variable is described in Table 1.

Table 1: Parameters Identification of The two-dimensional fractional Black-Scholes European option pricing equation

Symbol	Identification
u	the value of the call option
S_1	price of underlying asset 1
S_2	price of underlying asset 2
T	expiring date
σ_1	volatility of underlying asset 1
σ_1	volatility of underlying asset 2
w	correlation coefficient between price of underlying asset 1 and asset 2
l	risk free rate of interest
β_1	proportion of investment on asset 1
β_2	proportion of investment on asset 2
K	$\max(k_1, k_2)$
k_1	strike price of asset 1
k_2	strike price of asset 2

Next, we simplify the BS Model (1.1) and its terminal condition (1.2) via the following transformation:

$$x = \ln(S_1) - (l - \frac{1}{2}\sigma_1^2)\tau, \qquad y = \ln(S_2) - (l - \frac{1}{2}\sigma_2^2)\tau, \qquad t = T - \tau,$$

and

$$u = (x, y, \tau) = e^{-r(T-\tau)} v(x, y, t).$$

Note that the procedure for obtaining the simplified version of the Black-Scholes model may be found in references [17–19]. The reduction process provided a rewritten Black-Scholes partial differential equation for a European call option with two assets:

$$v_t = \frac{1}{2}\sigma_1^2 \frac{\partial^2 v}{\partial x^2} + \frac{1}{2}\sigma_2^2 \frac{\partial^2 v}{\partial y^2} + w\sigma_1\sigma_2 \frac{\partial^2 v}{\partial x \partial y}, \text{ for } (x, y, t) \in \mathbb{R} \times \mathbb{R} \times [0, T], \tag{1.3}$$

and the initial condition (IC):

$$v(x, y, 0; \alpha) = \max(c_1 e^x + c_2 e^y - K, 0), \tag{1.4}$$

where c_1 and c_2 are constants determined by $c_1 = \beta_1 e^{(l - \frac{1}{2}\sigma_1^2)T}$ and $c_2 = \beta_2 e^{(l - \frac{1}{2}\sigma_2^2)T}$.

The Black-Scholes European option pricing equation with two assets will be modified in this article by substituting the integer-order time derivative in equation (1.3) with the time-fractional derivative in the Caputo sense with order $\alpha \in (0, 1]$ as follows:

$$D_t^\alpha v = \frac{1}{2}\sigma_1^2 \frac{\partial^2 v}{\partial x^2} + \frac{1}{2}\sigma_2^2 \frac{\partial^2 v}{\partial y^2} + w\sigma_1\sigma_2 \frac{\partial^2 v}{\partial x \partial y} \text{ for } (x, y, t) \in \mathbb{R} \times \mathbb{R} \times [0, T], \tag{1.5}$$

with the IC:

$$v(x, y, 0; \alpha) = \max(c_1 e^x + c_2 e^y - K, 0), \tag{1.6}$$

where c_1 and c_2 are constants defined by (1.4) and the symbol D_t^α represents the Caputo derivative with order α as defined in Definition 2.5.

The aim of this study is to use the RPS approach to provide an approximate analytical solution for the fractional Black-Scholes European option pricing problem with two assets (1.5)-(1.6), based on the Caputo derivative.

2 Preliminaries

This section provides definitions of special functions, fractional integrals, and derivatives [20], which will subsequently be used throughout the work. Let us start with the first special function, the Gamma function.

Definition 2.1. The Gamma function, denoted by Γ, is defined as:

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt \text{ for } z > 0.$$

The next lemma deals with some properties of Gamma function.

Lemma 2.2. *(Some properties of Gamma function)*

a) $\Gamma(z + 1) = z\Gamma(z)$ *for any* $z > 0$.

b) $\Gamma(z + 1) = z!$ *if* z *is positive integer.*

Definition 2.3. The Mittag-Leffler (ML) function with order $\alpha > 0$, denoted by E_α, is given by:

$$E_\alpha(z) = \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(\alpha n + 1)} \text{ for } z \text{ is any real number.}$$

It follows from Definition 2.3 that $E_1(z) = \sum_{n=0}^{\infty} \frac{z^n}{\Gamma(n+1)} = e^z$.

Next, we will introduce the definition of fractional integral and derivative, used in this work, together with their respective properties [13].

Definition 2.4. The Reimann-Liouville fractional integral operator of order $\alpha > 0$ for the function g is determined by:

$$J_t^\alpha g(t) = \int_0^t (t - \tau)^{\alpha-1} g(\tau) d\tau, \text{ for } t > 0,$$

if the integral exists.

Definition 2.5. The Caputo fractional derivative with order $0 < \alpha \leq 1$ for the function g is defined by:

$$D_t^\alpha g(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t (t-\tau)^{-\alpha} g(\tau) d\tau, \text{ for } t > 0,$$

if the integral exists.

The following give some properties of the fractional integral and derivative used throughout this paper.

Lemma 2.6. *(Some properties of the fractional integral and derivative) Let β and γ be any constant.*

a) $J_t^0 g(t) = g(t)$ *for* $t > 0$.

b) $D_t^\beta t^\gamma = 0$ *if* $\gamma < \beta$.

c) $D_t^\beta t^\gamma = \frac{\Gamma(\gamma+1)}{\Gamma(\gamma+1-\beta)} t^{\gamma-\beta}$ *if* $\gamma \geq \beta$ *and* $\gamma + 1 - \beta$ *is not equal to zero.*

We next introduce the fractional-order power series with the parameter α [21].

Definition 2.7. (The fractional-order power series with the parameter α) The fractional-order power series with the parameter α (in three variables) around 0 with respect to $t \geq 0$ is an infinite series of the form:

$$\sum_{n=0}^{\infty} g_n(x,y) \frac{t^{n\alpha}}{\Gamma(1+n\alpha)} \text{ for any } (x,y,t) \in \mathbb{R} \times \mathbb{R} \times [0,T],$$

where g_n represents the coefficient of the nth term for any $n \in \mathbb{N} \cup \{0\}$.

Note that the fractional-order power series with the parameter α can be reduced to the classical power series when $\alpha = 1$.

3 RPSM Methodology

Let us begin by considering the generalized fractional nonlinear differential equation:

$$D_t^\alpha h(x,y,t) + K[x,y]h(x,y,t) + Q[x,y]h(x,y,t) = G(x,y,t), \ (x,y,t) \in \mathbb{R} \times \mathbb{R} \times [0,T], \quad (3.1)$$

and satisfies initial condition:

$$h(x,y,0) = g(x,y), \ (x,y) \in \mathbb{R} \times \mathbb{R}, \quad (3.2)$$

where $K[x,y]$ and $Q[x,y]$ are the linear and non-linear operator in x and y, $G(x,y,t)$ is the continuous function, and D_t^α denotes the Caputo fractional derivative with order α.

The RPSM process, modified from [10], consists of the following steps.

Step 1. We assume that the solution h of the generalized fractional differential equation (3.1) and (3.2) is in the form of the fractional power series about the point $t = 0$:

$$h(x,y,t) = \sum_{n=0}^{\infty} g_n(x,y) \frac{t^{n\alpha}}{\Gamma(1+n\alpha)}, \quad (3.3)$$

where the function $g_n(x, y)$ is determined by the process below.
It is clear that $h(x, y, 0) = g(x, y) = g_0(x, y)$.
We next define the k^{th} truncated series $h_k(x, y, t)$ for h by:

$$h_k(x, y, t) = \sum_{n=0}^{k} g_n(x, y) \frac{t^{n\alpha}}{\Gamma(1 + n\alpha)} \text{ for any } k = 1, 2, 3, \ldots,$$

or

$$h_k(x, y, t) = g(x, y) + \sum_{n=1}^{k} g_n(x, y) \frac{t^{n\alpha}}{\Gamma(1 + n\alpha)} \text{ for any } k = 1, 2, 3, \ldots. \tag{3.4}$$

Step 2. Before finding the value of coefficient $g_n(x, y)$ for $n = 1, 2, 3, \ldots$ in equation (3.4), we define the residual function Res_h for the fractional problem (3.1) and (3.2) by:

$$Res_h(x, y, t) = D_t^{\alpha} h(x, y, t) + K[x, y]h(x, y, t) + Q[x, y]h(x, y, t) - G(x, y, t), \tag{3.5}$$

and the k^{th} residual function $Res_{h,k}$ for the fractional problem (3.1) and (3.2) by: for any $k = 1, 2, 3, \ldots,$

$$Res_{h,k}(x, y, t) = D_t^{\alpha} h_k(x, y, t) + K[x, y]h_k(x, y, t) + Q[x, y]h_k(x, y, t) - G(x, y, t), \tag{3.6}$$

It is obvious that $Res_h(x, y, t) = 0$ and $\lim_{k\to\infty} Res_{h,k}(x, y, t) = Res_h(x, y, t)$ for any $(x, y, t) \in \mathbb{R} \times \mathbb{R} \times [0, T]$.

In order to specific the value of coefficient g_n, we must to use the following requirements:

$$D_t^{(k-1)\alpha} Res_{h,k}(x, y, 0) = 0, \text{ for any } k = 1, 2, 3, \ldots. \tag{3.7}$$

Step 3. We substitute all values of g_n in equation (3.3). In the end, we get the desired solution for the generalized fractional differential equations (3.1) and (3.2).

4 Main Results

In this section, we solve the fractional Black-Scholes European option pricing equation with two assets defined by (1.5) and (1.6) by the RPSM technique. Let us start:

Step 1. We assume that the solution v of the fractional Black-Scholes European option pricing equation with two assets defined by (1.5) and (1.6) is in the form of the fractional power series about the point $t = 0$:

$$v(x, y, t) = \sum_{n=0}^{\infty} g_n(x, y) \frac{t^{n\alpha}}{\Gamma(1 + n\alpha)}, \text{ for } (x, y, t) \in \mathbb{R} \times \mathbb{R} \times [0, T]. \tag{4.1}$$

It is clear that

$$v(x, y, 0) = g_0(x, y) = \max(c_1 e^x + c_2 e^y - K, 0).$$

We next define the k^{th} truncated series $v_k(x, y, t)$ for v by:

$$v_k(x, y, t) = g_0(x, y) + \sum_{n=1}^{k} g_n(x, y) \frac{t^{n\alpha}}{\Gamma(1 + n\alpha)} \text{ for } , k = 1, 2, 3, \ldots. \tag{4.2}$$

Step 2. The residual function Res_v for the fractional Black-Scholes European option pricing equation with two assets (1.5) and (1.6) is constructed by:

$$Res_v(x,y,t) = D_t^{\alpha} v - \frac{1}{2}\sigma_1^2 x^2 \frac{\partial^2 v}{\partial x^2} - \frac{1}{2}\sigma_2^2 y^2 \frac{\partial^2 v}{\partial y^2} - w\sigma_1\sigma_2 \frac{\partial^2 v}{\partial x \partial y},$$

and the k^{th} residual function Res_v of the fractional problem (1.5) and (1.6) is given by:

$$Res_{v,k}(x,y,t) = D_t^{\alpha} v_k - \frac{1}{2}\sigma_1^2 x^2 \frac{\partial^2}{\partial x^2} v_k - \frac{1}{2}\sigma_2^2 y^2 \frac{\partial^2}{\partial y^2} v_k - w\sigma_1\sigma_2 \frac{\partial^2}{\partial x \partial y} v_k. \tag{4.3}$$

To find the first unknown coefficient g_1, we let $k = 1$ in equation (4.2) and (4.3) and then, we have:

$$Res_{v,1}(x,y,t) = D_t^{\alpha} v_1 - \frac{1}{2}\sigma_1^2 x^2 \frac{\partial^2}{\partial x^2} v_1 - \frac{1}{2}\sigma_2^2 y^2 \frac{\partial^2}{\partial y^2} v_1 - w\sigma_1\sigma_2 \frac{\partial^2}{\partial x \partial y} v_1, \tag{4.4}$$

and

$$v_1(x,y,t) = g_0(x,y) + g_1(x,y)\frac{t^{\alpha}}{\Gamma(1+\alpha)}. \tag{4.5}$$

By substituting equation (4.5) into equation (4.4), we obtain:

$$\begin{aligned} Res_{v,1}(x,y,t) &= g_1(x,y) - \frac{1}{2}\sigma_1^2\left[g_{0xx}(x,y) + g_{1xx}(x,y)\frac{t^{\alpha}}{\Gamma(1+\alpha)}\right] + \frac{1}{2}\sigma_2^2\left[g_{0yy}(x,y)\right. \\ &+ \left. g_{1yy}(x,y)\frac{t^{\alpha}}{\Gamma(1+\alpha)}\right] - w\sigma_1\sigma_2\left[g_{0xy}(x,y) + g_{1xy}(x,y)\frac{t^{\alpha}}{\Gamma(1+\alpha)}\right]. \end{aligned}$$

Then, the residual function at $t = 0$ is:

$$Res_{v,1}(x,y,0) = g_1(x,y) - \frac{1}{2}\sigma_1^2 g_{0xx}(x,y) - \frac{1}{2}\sigma_2^2 g_{0yy}(x,y) - w\sigma_1\sigma_2 g_{0xy}(x,y).$$

By the condition (3.7), $Res_{v,1}(x,y,0) = 0$, that we obtain that:

$$g_1(x,y) = \frac{1}{2}\sigma_1^2 g_{0xx}(x,y) + \frac{1}{2}\sigma_2^2 g_{0yy}(x,y) + w\sigma_1\sigma_2 g_{0xy}(x,y),$$

or

$$g_1(x,y) = \frac{1}{2}\sigma_1^2 \max(c_1 e^x, 0) + \frac{1}{2}\sigma_2^2 \max(c_2 e^y, 0).$$

Hence, the first RPS approximate analytical solution of the fractional problem (1.5) and (1.6) is:

$$v_1(x,y,t) = \max(c_1 e^x + c_2 e^y - k, 0) + \frac{1}{2}\sigma_1^2 \max(c_1 e^x, 0) + \frac{1}{2}\sigma_2^2 \max(c_2 e^y, 0)\frac{t^{\alpha}}{\Gamma(1+\alpha)}.$$

Now putting $k = 2$ in equations (4.2) and (4.3) to determine the second unknown coefficient $g_2(y)$, we have:

$$Res_{v,2}(x,y,t) = D_t^{\alpha} v_2 - \frac{1}{2}\sigma_1^2 x^2 \frac{\partial^2}{\partial x^2} v_2 - \frac{1}{2}\sigma_2^2 y^2 \frac{\partial^2}{\partial y^2} v_2 - w\sigma_1\sigma_2 \frac{\partial^2}{\partial x \partial y} v_2, \tag{4.6}$$

and

$$v_2(x,y,t) = g(x,y) + g_1(x,y)\frac{t^{\alpha}}{\Gamma(1+\alpha)} + g_2(x,y)\frac{t^{2\alpha}}{\Gamma(1+2\alpha)}. \tag{4.7}$$

From equations (4.6) and (4.7), we get:

$$\begin{aligned} Res_{v,2}(x,y,t) &= g_1(x,y) + g_2(x,y)\frac{t^\alpha}{\Gamma(1+\alpha)} \\ &- \frac{1}{2}\sigma_1^2\left[g_{0xx}(x,y) + g_{1xx}(x,y)\frac{t^\alpha}{\Gamma(1+\alpha)} + g_{2xx}(x,y)\frac{t^{2\alpha}}{\Gamma(1+2\alpha)}\right] \\ &- \frac{1}{2}\sigma_2^2\left[g_{0yy}(x,y) + g_{1yy}(x,y)\frac{t^\alpha}{\Gamma(1+\alpha)} + g_{2yy}(x,y)\frac{t^{2\alpha}}{\Gamma(1+2\alpha)}\right] \\ &- w\sigma_1\sigma_2\left[g_{0xy}(x,y) + g_{1xy}(x,y)\frac{t^\alpha}{\Gamma(1+\alpha)} + g_{2xy}(x,y)\frac{t^{2\alpha}}{\Gamma(1+2\alpha)}\right]. \end{aligned}$$

Let us consider:

$$\begin{aligned} D_t^\alpha Res_{v,2}(x,y,t) &= g_2(x,y) \\ &- \frac{1}{2}\sigma_1^2\left[g_{1xx}(x,y) + g_{2xx}(x,y)\frac{t^\alpha}{\Gamma(1+\alpha)}\right] \\ &- \frac{1}{2}\sigma_2^2\left[g_{1yy}(x,y) + g_{2yy}(x,y)\frac{t^\alpha}{\Gamma(1+\alpha)}\right] \\ &- w\sigma_1\sigma_2\left[g_{1xy}(x,y) + g_{2xy}(x,y)\frac{t^\alpha}{\Gamma(1+\alpha)}\right]. \end{aligned}$$

It follows by the condition (3.7), $D_t^\alpha Res_{v,2}(x,y,0) = 0$, that we have:

$$g_2(x,y) = \frac{1}{2}\sigma_1^2 g_{1xx}(x,y) + \frac{1}{2}\sigma_2^2 g_{1yy}(x,y) + w\sigma_1\sigma_2 g_{1xy}(x,y),$$

or

$$g_2(x,y) = \left(\frac{1}{2}\sigma_1^2\right)^2 \max(c_1e^x, 0) + \left(\frac{1}{2}\sigma_2^2\right)^2 \max(c_2e^y, 0).$$

Hence the second RPS approximate solution of the fractional problem (1.5) and (1.6) is:

$$\begin{aligned} v_2(x,y,t) &= \max(c_1e^x + c_2e^y - k, 0) + \left[\frac{1}{2}\sigma_1^2\max(c_1e^x,0) + \frac{1}{2}\sigma_2^2\max(c_2e^y,0)\right]\frac{t^\alpha}{\Gamma(1+\alpha)} \\ &+ \left[\left(\frac{1}{2}\sigma_1^2\right)^2\max(c_1e^x,0) + \left(\frac{1}{2}\sigma_2^2\right)^2\max(c_2e^y,0)\right]\frac{t^{2\alpha}}{\Gamma(1+2\alpha)}. \end{aligned}$$

To determine the third unknown coefficient g_3, we let $k = 3$ in equations (4.2) and (4.3), and then we get:

$$Res_{v,3}(x,y,t) = D_t^\alpha v_3 - \frac{1}{2}\sigma_1^2x^2\frac{\partial^2}{\partial x^2}v_3 - \frac{1}{2}\sigma_2^2y^2\frac{\partial^2}{\partial y^2}v_3 - w\sigma_1\sigma_2\frac{\partial^2}{\partial x\partial y}v_3, \tag{4.8}$$

and

$$v_3(x,y,t) = g(x,y) + g_1(x,y)\frac{t^\alpha}{\Gamma(1+\alpha)} + g_2(x,y)\frac{t^{2\alpha}}{\Gamma(1+2\alpha)} + g_3(x,y)\frac{t^{3\alpha}}{\Gamma(1+3\alpha)}. \tag{4.9}$$

In following, we substitute equation (4.9) into equation (4.8) and then we use the condition (3.7), $D_t^{2\alpha} Res_{v3}(x,y,0) = 0$. Finally, we obtain:

$$g_3(x,y) = \left(\frac{1}{2}\sigma_1^2\right)^3 \max(c_1e^x, 0) + \left(\frac{1}{2}\sigma_2^2\right)^3 \max(c_2e^y, 0),$$

and the third RPS approximate solution is:

$$\begin{aligned} v_3(x,y,t) &= \max(c_1e^x + c_2e^y - k, 0) \\ &+ \left[\left(\frac{1}{2}\sigma_1\right)^2 \max(c_1e^x, 0) + \left(\frac{1}{2}\sigma_2\right)^2 \max(c_2e^y, 0)\right]\frac{t^\alpha}{\Gamma(1+\alpha)} \\ &+ \left[\left(\frac{1}{2}\sigma_1^2\right)^2 \max(c_1e^x, 0) + \left(\frac{1}{2}\sigma_2^2\right)^2 \max(c_2e^y, 0)\right]\frac{t^{2\alpha}}{\Gamma(1+2\alpha)} \\ &+ \left[\left(\frac{1}{2}\sigma_1^2\right)^3 \max(c_1e^x, 0) + \left(\frac{1}{2}\sigma_2^2\right)^3 \max(c_2e^y, 0)\right]\frac{t^{3\alpha}}{\Gamma(1+3\alpha)}. \end{aligned}$$

Likewise, the coefficients of g_n and the n-th RPS approximate solution v_n for any $n \geq 4$ can be evaluated as previously discussed and

$$g_n(x,y) = \left(\frac{1}{2}\sigma_1^2\right)^n \max(c_1e^x, 0) + \left(\frac{1}{2}\sigma_2^2\right)^n \max(c_2e^y, 0),$$

and

$$\begin{aligned} v_n(x,y,t) &= \max(c_1e^x + c_2e^y - k, 0) \\ &+ \left[\left(\frac{1}{2}\sigma_1\right)^2 \max(c_1e^x, 0) + \left(\frac{1}{2}\sigma_2\right)^2 \max(c_2e^y, 0)\right]\frac{t^\alpha}{\Gamma(1+\alpha)} \\ &+ \left[\left(\frac{1}{2}\sigma_1^2\right)^2 \max(c_1e^x, 0) + \left(\frac{1}{2}\sigma_2^2\right)^2 \max(c_2e^y, 0)\right]\frac{t^{2\alpha}}{\Gamma(1+2\alpha)} \\ &+ \left[\left(\frac{1}{2}\sigma_1^2\right)^3 \max(c_1e^x, 0) + \left(\frac{1}{2}\sigma_2^2\right)^3 \max(c_2e^y, 0)\right]\frac{t^{3\alpha}}{\Gamma(1+3\alpha)} \\ &\vdots \\ &+ \left[\left(\frac{1}{2}\sigma_1^2\right)^n \max(c_1e^x, 0) + \left(\frac{1}{2}\sigma_2^2\right)^n \max(c_2e^y, 0)\right]\frac{t^{n\alpha}}{\Gamma(1+n\alpha)}, \end{aligned} \tag{4.10}$$

respectively.

Step 3. By equation (4.10), the solution of the fractional Black-Scholes European option pricing equation with two assets defined by (1.5) and (1.6) is in the following form:

$$\begin{aligned} v(x,y,t) &= \sum_{n=0}^{\infty} g_n(x,y)\frac{t^{n\alpha}}{\Gamma(1+n\alpha)} \\ &= \max(c_1e^x + c_2e^y - k, 0) \\ &+ \sum_{n=1}^{\infty}\left[\left(\frac{1}{2}\sigma_1^2\right)^n \max(c_1e^x, 0) + \left(\frac{1}{2}\sigma_2^2\right)^n \max(c_2e^y, 0)\right]\frac{t^{n\alpha}}{\Gamma(1+n\alpha)} \\ &= \max(c_1e^x + c_2e^y - k, 0) \\ &+ \max(c_1e^x, 0)\sum_{n=1}^{\infty}\left(\frac{1}{2}\sigma_1^2\right)^n \frac{t^{n\alpha}}{\Gamma(1+n\alpha)} + \max(c_2e^y, 0)\sum_{n=1}^{\infty}\left(\frac{1}{2}\sigma_2^2\right)^n \frac{t^{n\alpha}}{\Gamma(1+n\alpha)} \\ &= \max(c_1e^x + c_2e^y - k, 0) \\ &+ \max(c_1e^x, 0)\left[\sum_{n=0}^{\infty}\left(\frac{\sigma_1^2}{2}\right)^n \frac{t^{n\alpha}}{\Gamma(1+n\alpha)} - 1\right] + \max(c_2e^y, 0)\left[\sum_{n=0}^{\infty}\left(\frac{\sigma_2^2}{2}\right)^n \frac{t^{n\alpha}}{\Gamma(1+n\alpha)} - 1\right] \\ &= \max(c_1e^x + c_2e^y - k, 0) \\ &+ \max(c_1e^x, 0)\left[\sum_{n=0}^{\infty}\frac{\left(\frac{\sigma_1^2t^\alpha}{2}\right)^n}{\Gamma(1+n\alpha)} - 1\right] + \max(c_2e^y, 0)\left[\sum_{n=0}^{\infty}\frac{\left(\frac{\sigma_2^2t^\alpha}{2}\right)^n}{\Gamma(1+n\alpha)} - 1\right], \end{aligned}$$

or

$$v(x,y,t) = \max(c_1e^x + c_2e^y - k, 0) + \max(c_1e^x, 0)\left[E_\alpha\left(\frac{\sigma_1^2}{2}t^\alpha\right) - 1\right] + \max(c_2e^y, 0)\left[E_\alpha\left(\frac{\sigma_2^2}{2}t^\alpha\right) - 1\right],$$

where E_α is the ML function with order α.

The following theorem is the main result.

Theorem 4.1. *The analytical solution for the fractional Black-Scholes European option pricing equation with two assets defined by (1.5) with the IC (1.6) is in the following form:*

$$v(x,y,t) = max(c_1e^x + c_2e^y - K, 0) + max(c_1e^x, 0)\left[E_\alpha\left(\frac{\sigma_1^2}{2}t^\alpha\right) - 1\right] + max(c_2e^y, 0)\left[E_\alpha\left(\frac{\sigma_2^2}{2}t^\alpha\right) - 1\right] \quad (4.11)$$

where E_α is the ML function with order α. Furthermore, by the RPSM approach, the n^{th} truncated series of the approximate analytical solution for (1.5) with the IC (1.6) is given by:

$$\begin{aligned} v_n(x,y,t) &= max(c_1e^x + c_2e^y - K, 0) \\ &+ \left[\left(\frac{1}{2}\sigma_1\right)^2 max(c_1e^x, 0) + \left(\frac{1}{2}\sigma_2\right)^2 max(c_2e^y, 0)\right]\frac{t^\alpha}{\Gamma(1+\alpha)} \\ &+ \left[\left(\frac{1}{2}\sigma_1^2\right)^2 max(c_1e^x, 0) + \left(\frac{1}{2}\sigma_2^2\right)^2 max(c_2e^y, 0)\right]\frac{t^{2\alpha}}{\Gamma(1+2\alpha)} \\ &+ \left[\left(\frac{1}{2}\sigma_1^2\right)^3 max(c_1e^x, 0) + \left(\frac{1}{2}\sigma_2^2\right)^3 max(c_2e^y, 0)\right]\frac{t^{3\alpha}}{\Gamma(1+3\alpha)} \\ &\vdots \\ &+ \left[\left(\frac{1}{2}\sigma_1^2\right)^n max(c_1e^x, 0) + \left(\frac{1}{2}\sigma_2^2\right)^n max(c_2e^y, 0)\right]\frac{t^{n\alpha}}{\Gamma(1+n\alpha)}. \end{aligned} \quad (4.12)$$

The following is consequence from Theorem 4.1.

Corollary 4.2. *The analytical solution for the Black-Scholes European option pricing equation with two assets is in the following form:*

$$v(x,y,t) = max(c_1e^x + c_2e^y - K, 0) + max(c_1e^x, 0)\left[e^{\frac{\sigma_1^2 t}{2}} - 1\right] + max(c_2e^y, 0)\left[e^{\frac{\sigma_2^2 t}{2}} - 1\right].$$

Proof. To get the proof of this corollary, we may analyze Theorem 4.1 with the value of α set to 1. □

5 Numerical Results and Discussions

In this part, we use Python program to plot graph of the analytical solution of the European call option for two assets in the time fractional Black Scholes model, based on the Caputo derivative. Each value of model parameters is shown in Table 2.

Graphs of the analytical solution u, given by (4.11), for the fractional Black-Scholes equation (1.5) with the IC (1.6) with $\alpha = 1$, $\alpha = 0.8$, $\alpha = 0.5$, and $\alpha = 0.25$, are illustrated in Figure 1. Obviously, the analytic solution of the two assets time fractional Black Scholes European call Option in Caputo derivative sense of $\alpha = 1$ and other alphas have an agree tendency as shown in Figure 1. The graphs of option prices in Equation (4.11) corresponding to the assets x and y in Figure 2-3 illustrate the outcome in the identical direction.

The European call option prices have lower values when $\alpha < 1$ compared to when $\alpha = 1$, as seen in Table 3 and Figure 2-3. This implies that there is a direct proportionality between α and v. As the quantity of alpha decreases, the value of European call options also lowers.

Table 2: The values of model parameters

Model parameters	Value
strike price(dollars): K	70
risk free rate of interest: l;	0.05
expiration date(year): t	2
volatility of the underlying stock 1: σ_1	0.1
volatility of the underlying stock 2: σ_2	0.2
proportion of investment on asset 1: β_1	2
proportion of investment on asset 2: β_2	1

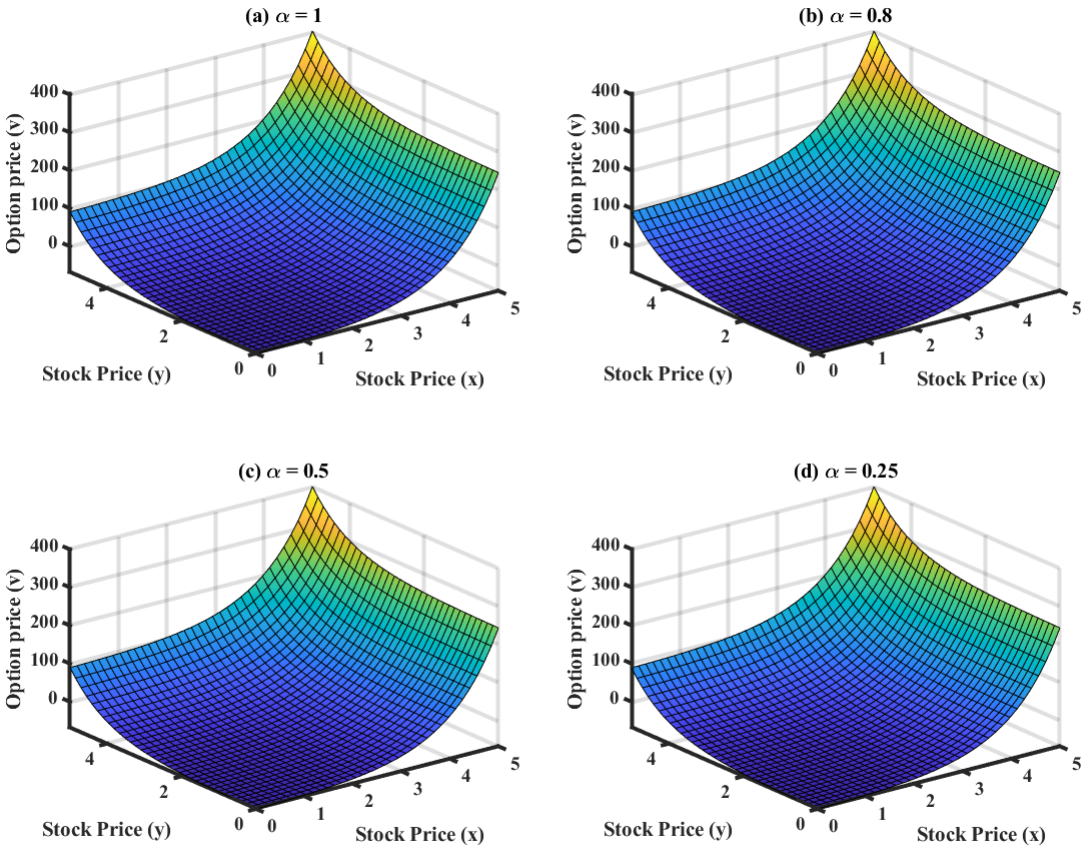


Figure 1: European call option prices for $\alpha = 1$(a),$\alpha = 0.8$(b),$\alpha = 0.5$(c),$\alpha = 0.25$(d)

Table 3: Values of European call option as $\alpha = 1$,$\alpha = 0.8$,$\alpha = 0.5$,$\alpha = 0.25$

alpha	$\alpha = 1$	$\alpha = 0.8$	$\alpha = 0.5$	$\alpha = 0.25$
x	2.9057	2.9057	2.9057	2.9057
y	3.8520	3.8520	3.8520	3.8520
t	2	2	2	2
v	19.1274	18.9782	18.6581	18.3201

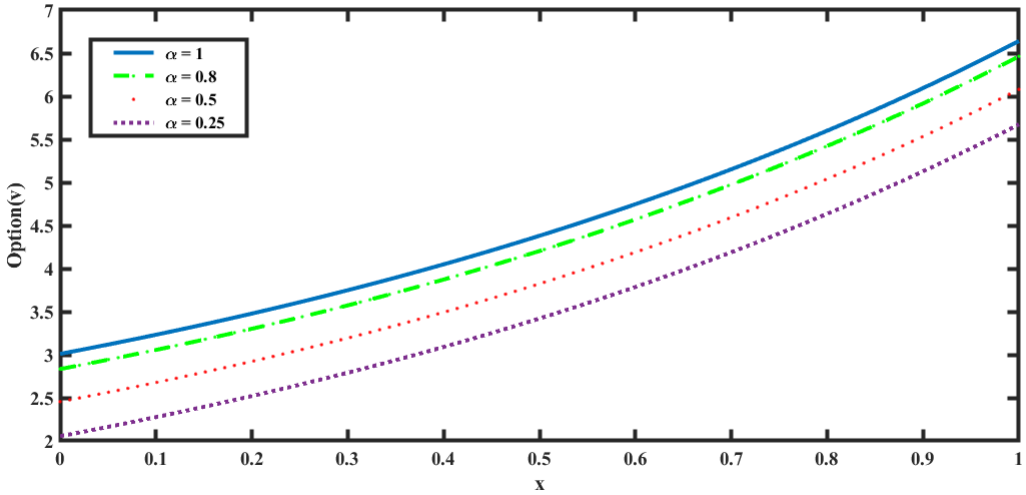


Figure 2: European call option prices for $\alpha = 1$,$\alpha = 0.8$, $\alpha = 0.5$,$\alpha = 0.25$,$t = 2$ years with $y = 4.19118$

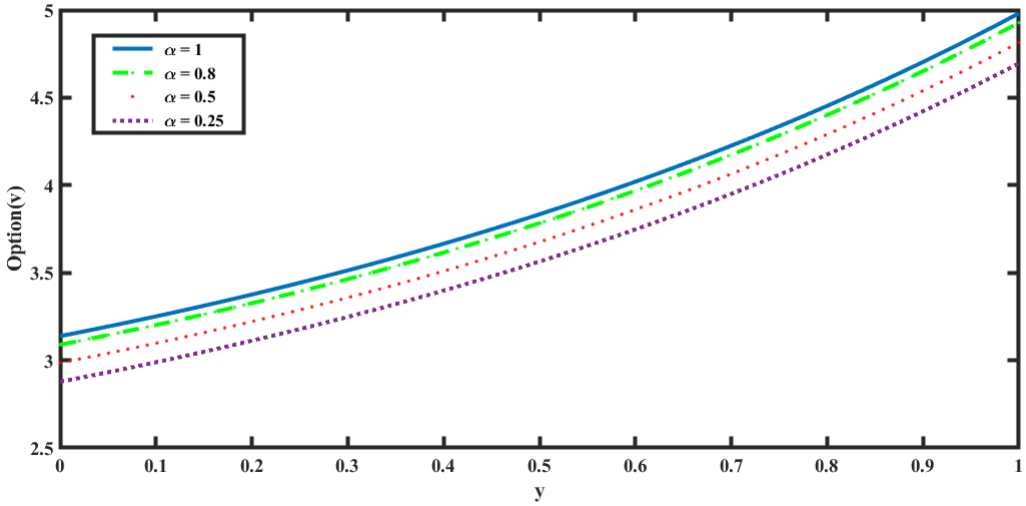


Figure 3: European call option prices for $\alpha = 1$,$\alpha = 0.8$, $\alpha = 0.5$,$\alpha = 0.25$,$t = 2$ years with $x = 3.52941$

6 Conclusion

The application of the residual power series method (RPSM) allows us to derive an analytical solution, as given by equation (4.11), and an approximate analytical solution, as determined by equation (4.12), for a two-dimensional fractional Black-Scholes pricing model in the Caputo sense, as described in equation (1.5), with the IC specified in equation (1.6), as presented in Theorem 4.1. As with the main result, the residual power series method only needs a few iterations to get a good answer, as shown in Theorem 4.1. According to the numerical result in Figures 1-3, option pricing for $\alpha = 1$ and $\alpha < 1$ displays an agreed tendency, which shows that the solutions of the time-fractional Black-Scholes equation in the sense of Caputo fractional derivative and the classical Black-Scholes equation with two assets are concurred. However, we observe that the order α of the fractional derivative in the Caputo sense has an effect on the price of the European option. The smaller the number of alphas, the lower the price of the European option, according to Table 3 and Figure 4.

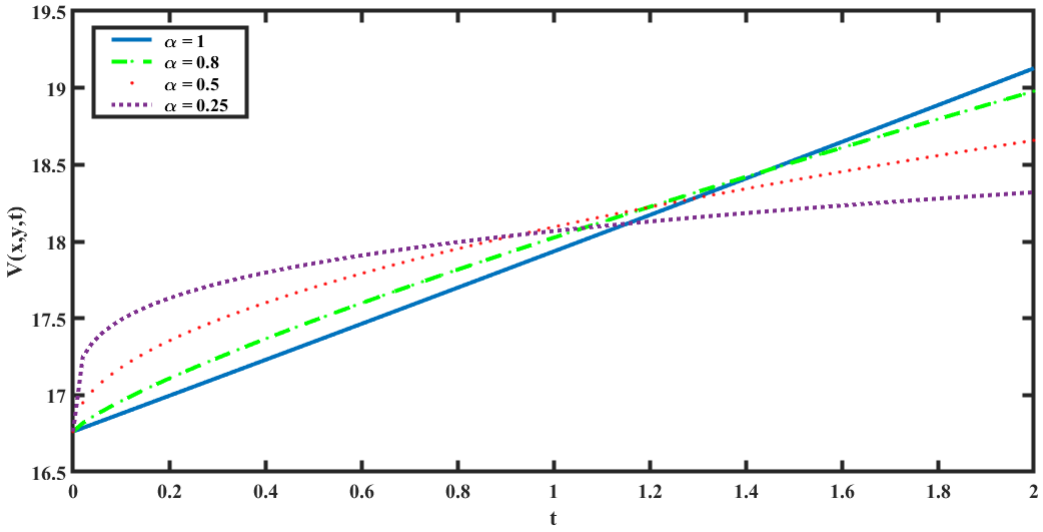


Figure 4: Value of the options for $\alpha = 1$(a),$\alpha = 0.8$(b),$\alpha = 0.5$(c),$\alpha = 0.25$(d)

References

[1] A. L. Jackson, *Option trading- A beginner's guide on how to trade options*, Forbes Advisor.,23 Mar 2023,www.forbes.com/advisor/in/investing/options-trading/.

[2] C. Team, *Option trading- A beginner's guide on how to trade options*, Corp. Finan. Insti.,23 Feb 2023,www.corporatefinanceinstitute.com/resources/derivatives/european-option/.

[3] F. Black and M. Scholes, *The pricing of options and corporate liabilities*, J. Polit. Econ. **81**(3) (1973), 637–654.

[4] A. Hayes, *Black-Scholes Model: What It Is, How It Works, Options Formula*, Investopedia,31 Oct 2023,www.investopedia.com/terms/b/blackscholes.asp.

[5] R.M. Jena and S. Chakraverty, *A new iterative method based solution for fractinal Black-Scholes option pricing equations(BSOPE)*, SN. Appl. Sci. **1**(1) (2019).

[6] S. Ampun and P. Sawangtong, *The approximate analytic solution of the time-fractional Black-Scholes equation with a European option based on the Katugampola fractional derivative*, Math. **9**(3) (2021), 214.

[7] S. Ampun, P. Sawangtong and W. Sawangtong, *An analysis of the fractional-order option pricing problem for two assets by the generalized laplace variational iteration approach*, Fractal Fract. **6**(11) (2022), 667.

[8] S. Rawat, *A Black-scholes Options Pricing Model (BSOPM)*, Anal. Step.,22 Aug 2021,www.analyticssteps.com/blogs/black-scholes-options-pricing-model-bsopm.

[9] S. Thanompolkrang, W. Sawangtong and P. Sawangtong, *Application of the generalized Laplace homotopy perturbation method to the time-fractional Black-Scholes equations based on the Kaatugampola fractional derivative in Caputo type*, Computation **9**(3) (2021), 33.

[10] V. P. Dubey, R. Kumar and D. Kumar, *A reliable treatment of residual power seires method for time-fractional Black-Scholes European option pricing equations*, Physica A: Stat. Mech. Appl. **533**(1) (2019), 122040.

[11] M. Alaround, *Application of Laplace residual power seires method for approximate solutions of fractional IVP's*, Alex. Engi. J. **61**(2) (2021), 1585–1595.

[12] R. Khalil, M.A. Horani and A. Yousef, *A new definition of fractional derivative*, J. Comp. Appl.Math **264**(1) (2014), 65–70.

[13] C. Li, D. Qian and Y. Chen, *On Riemann-Liouville and Caputo derivatives*, Dis. Dyn. Natu. Soci.**2011**(562494) (2011),15.

[14] M.D. Ortigueira and J.A. Tenreiro Machado, *What is a fractional derivative?*, J. Comp. Phy.**293**(1) (2015),4–13.

[15] J.M. Kimeu, *Fractional calculus: definitions and applications*, West. Ken. Uni. Top. Scho.**4**(1) (2009).

[16] K. Zakaria and S. Hafeez *Options pricing for two stocks by Black-Scholes time fractional order non-linear partial differential equation*, Proceedings of the 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies, 2020, Sukkkur, January 29–30, 2020, pp. 1–13.

[17] M. Contreras, A. Llanquihuén and M. Villena, *On the Solution of the Multi-Asset Black-Scholes Model: Correlations, Eigenvalues and Geometry*, Dis. Dyn. Natu. Soci.**6**(4) (2016),562–579.

[18] P. Sawangtong, K. Trachoo, W. Sawangtong and B. Wiwattanapataphee, *he Analytical Solution for the Black-Scholes Equation with Two Assets in the Liouville-Caputo Fractional Derivative Sense*, Math.**6**(8) (2018),129.

[19] Y.H. Ro and N. Wan, *A Method of Reducing Dimension of Space Variables in Multi-Dimensional Black-Scholes Equations*, Math.**30**(2) (2014),145–158.

[20] J.F. Gomez-Aguilar and A. Atangana, *Applications of fractional calculus to modeling in dynamics and chaos*, CRC Press. (2022).

[21] P. Dunnimit, W. Sawangtong and P. Sawangtong, *An approximate analytical solution of the time-fractional Navier–Stokes equations by the generalized Laplace residual power series method*, Part. Diff. Equa. Appl. Math. **9**(1) (2024), 100629.

8. Mathematical Modeling and Mathematical Finance

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Estimating the Value at Risk of Buy-and-Sell Strategy Using the RSI Indicator on the EUR/USD Exchange Market

Rattaporn Supama[1,†] and Watcharin Klongdee[1,‡]

[1]Department of Mathematics, Faculty of Science, Khon Kaen University, Khon Kaen 40002, Thailand

Abstract

This article estimates the value-at-risk of the buy-and-sell strategy by using the relative strength index (RSI) indicator on the EUR/USD exchange rate to assess market risk in financial asset portfolios. It focusing on potential declines in market value due to fluctuations in interest rates, foreign exchange rates, equity prices, or commodity prices. The historical sample covers January 4, 2021, to December 29, 2023 (780 days). We simulate the buy-and-sell strategy in 10,000 scenarios, using the relative strength index (RSI) indicator for the EUR/USD exchange rate. For each scenario, we generate the sequence of daily rate-of-return of the EUR/USD exchange rate over 260 days to approximate the probability of loss. Then, we use quadratic polynomial regression to determine the value-at-risk. The simulation measures investment risk at 95% and 99% confidence levels, indicating the probability that portfolio losses are smaller than estimated by the risk measure. The simulation result is that the maximum loss will not exceed 9.48% with 95% confidence and 12.27% with 99% confidence.

Keywords: value at risk, relative strength index, forex, quadratic polynomial regression.

2020 MSC: Primary 90-10.

1 Introduction

The value-at-risk model assesses market risk by gauging the potential decline in a portfolio's value within a specified timeframe and probability attributable to fluctuations in market prices or rates. Value-at-risk measurements typically represent percentiles aligned with the chosen confidence level. In practical applications, these estimates are computed across

*This research was financially supported by the research capability enhancement program through graduate student scholarship, Faculty of Science, Khon Kaen University, is gratefully acknowledged.

† Speaker. ‡ Corresponding author.
E-mail address: rattapornsupama@kkumail.com (R. Supama), kwatch@kku.ac.th (W. Klongdee).

percentiles ranging from the 90th to the 99.9th, with the 95th to 99th percentile range being the most frequently employed [5].

For calculating the value at risk of an investment in one year, we need to know the number of trading days in a year to calculate an investment's risk value. As previously noted, the forex market operates around the clock, 24 hours each day, but exclusively for six days a week, excluding Saturdays. Even on Sundays, the market has limited hours, commencing at 5 p.m. EST (Eastern Standard Time) with minimal volatility until early Monday morning, when market liquidity is high. The value-at-risk calculation considers only five trading days per week, disregarding Saturdays and Sundays [4]. Consequently, there are a total of $5\times52 = 260$ days annually.

Despite over 50 currencies being traded regularly, the US dollar (USD) reigns supreme, with the euro (EUR) and Japanese yen (JPY) trailing behind. The US dollar (USD) holds a dominant position as a vehicle currency, involved in nearly 90% of global foreign exchange transactions. Vehicle currency is used as a unit of account, medium of exchange, and store of value not only for transactions within the country but also for international public and private transactions. Illustrating the global role of the US dollar as the main vehicle currency, the top three most traded currency pairs, with EUR/USD leading at 23%, followed by USD/JPY at 14%, and GBP/USD at 10%. So, EUR/USD remains the largest currency pair [3].

Welles Wilder created the relative strength index (RSI) indicator and published it in 1978 [7]. In several trading systems, the Relative Strength Index (RSI) is a popular oscillator [6]. Anson et al. provide three situations: the buy signal appears when the RSI crosses 30 from above, and the sell signal appears when the RSI crosses 70 from below. Second, when the RSI goes back to 30 from below, it signals a buy, and when it goes back to 70 from above, it signals a sell. Third, a more complicated analysis of the RSI is made to obtain buying and selling signals [2].

The relative strength index (RSI) is a technical analysis indicator used to assess overbought or oversold conditions in an asset's price. It is a bounded oscillator ranging from 0 to 100. Classically, an RSI above 70 suggests overbought conditions, signaling a potential price correction, while an RSI below 30 indicates oversold conditions, suggesting a potential price rebound. The formula for calculating RSI is

$$RSI = 100 - \frac{100}{1 + \frac{\text{Average of } n \text{ days' up closes}}{\text{Average of } n \text{ days' down closes}}},$$

where n represents the number of periods that the trader chooses to analyze. The RSI (30,70) is widely used; for example, Anderson, B., and Li, S. (2015) explore the Relative Strength Index (RSI) trading profitability using daily data for the Swiss franc/US dollar exchange rate. The standard RSI thresholds of ≤ 30 and ≥ 70 for buy or sell signals have shown no trading profit but a slight loss over the past decade. However, modifying the threshold parameters

reveals that deviating from the commonly used combination can result in profitable trading signals using RSI [1].

Marek and Sediva [6] compare four trading strategies: RSI with standard parameters, daily optimized parameters, a simple buy-and-hold strategy, and a Kelly gambling-based strategy. Simulations were conducted using randomized time intervals spanning from February 15, 2007, to February 14, 2017. Sixteen companies from the S&P 500, which ranked among the top 10 largest companies from 2006 to 2009, were selected for simulation. The results revealed that the best strategy is buy-and-hold strategy [6].

In this article, we generate the sequence of the daily rate of return of EUR/USD exchange rates over 260 days from the average and standard deviation of the historical sample covering January 4, 2021, to December 29, 2023 (780 days). Next, we simulate buy-and-sell strategy in 10,000 scenarios, using the Relative Strength Index (RSI) indicator for the EUR/USD exchange rate to approximate the probability of loss. Then, we use quadratic polynomial regression to determine the value at risk. The simulation measures investment risk at 95% and 99% confidence levels.

2 Data and Simulation Methodology

2.1 The RSIBS (30,70) Strategy

In this section, we shall introduce the RSIBS (30,70) strategy, which is considered the signal for buying and selling with an RSI equal to 30 and 70, respectively.
The procedure of the RSIBS (30,70) strategy is as follows:

1. *We start with the initial capital of* $I_0 = 10000$ *USD. In the first order, investors wait until the RSI of the exchange rate EUR/USD reaches 30 to open a buy or sell order.*
2. *In the* $k-th$ *order,*
   2.1 If the $(k-1)-th$ *order is a buy order, we shall wait until* $RSI \geq 70$ *and close it. After that, we shall open a sell order immediately.*
   2.2 If the $(k-1)-th$ *order is a sell order, we shall wait until* $RSI \leq 30$ *and close it. After that, we shall open a buy order immediately.*
   2.3 If the last order wasn't closed within the specified timeframe, it would not be considered.

However, most investors focus on returns without calculating the risks involved. This research presents another form of risk measurement called value at risk using the RSIBS $(30,70)$ strategy.

Let T_k be the time of investment in k-th order and P_n be the price of EUR/USD at time n. The activity of the strategy is described as follows:
The first order is considered in two situations:

Case 1: The 1-st order is a buy order, the investor has $\frac{I_0}{P_{T_1}}$ lots. At the time T_2, the investor close the buy order and open the sell order immediately. The profit of the portfolio at the time T_2 is $\frac{I_0 \times P_{T_2}}{P_{T_1}} - I_0 = I_0(\frac{P_{T_2}-P_{T_1}}{P_{T_1}})$ and the value of the portfolio at the time T_2 is

$$I_1 = I_0 + I_0\left(\frac{P_{T_2}-P_{T_1}}{P_{T_1}}\right) = I_0\left(1 + \frac{P_{T_2}-P_{T_1}}{P_{T_1}}\right).$$

Case 2: The 1-st order is a sell order, the investor has $\frac{I_0}{P_{T_1}}$ lots. At the time T_2, the investor close the sell order and open the buy order immediately. The profit of the portfolio at the time T_2 is $I_0 - \frac{I_0 \times P_{T_2}}{P_{T_1}} = I_0(\frac{P_{T_1}-P_{T_2}}{P_{T_1}})$ and the value of the portfolio at the time T_2 is

$$I_1 = I_0 + I_0(\frac{P_{T_1}-P_{T_2}}{P_{T_1}}) = I_0\left(1 - \frac{P_{T_2}-P_{T_1}}{P_{T_1}}\right).$$

Similarly, the value of the portfolio in the $k-th$ order at the time T_k is

$$I_k = \begin{cases} I_{k-1}\left(1 + \dfrac{P_{T_k} - P_{T_{k-1}}}{P_{T_{k-1}}}\right) & \text{, the } (k-1)-\text{th order is a buy order,} \\ I_{k-1}\left(1 - \dfrac{P_{T_k} - P_{T_{k-1}}}{P_{T_{k-1}}}\right) & \text{, the } (k-1)-\text{th order is a sell order.} \end{cases}$$

For the convenient, we denote

$$\pi(k-1) = \begin{cases} 0 & \text{, the } (k-1)-\text{th order is a buy order,} \\ 1 & \text{, the } (k-1)-\text{th order is a sell order.} \end{cases}$$

Therefore, we have

$$\begin{aligned} I_k &= I_{k-1}\left(1 + (-1)^{\pi(k-1)}\frac{P_{T_k} - P_{T_{k-1}}}{P_{T_{k-1}}}\right) \\ &= I_0 \prod_{m=1}^{k-1}\left(1 + (-1)^{\pi(m)}\frac{P_{T_{m+1}} - P_{T_m}}{P_{T_m}}\right). \end{aligned}$$

So that, the investor has the percent of loss at the time T_k as

$$\begin{aligned} Loss(T_k) &= -\frac{I_k - I_0}{I_0} \\ &= -\left(\prod_{m=1}^{k-1}\left(1 + (-1)^{\pi(m)}\frac{P_{T_{m+1}} - P_{T_m}}{P_{T_m}}\right) - 1\right) \\ &= \left(1 - \prod_{m=1}^{k-1}\left(1 + (-1)^{\pi(m)}\frac{P_{T_{m+1}} - P_{T_m}}{P_{T_m}}\right)\right). \end{aligned}$$

Finally, the value at risk $(VaR_\alpha(T))$ of the portfolio at the confidence level α is the minimum percent of loss that could occur to the portfolio over the last order at the time T defined by

$$VaR_\alpha(T) = \min\{x \in R : Pr(Loss(T) \le x) \ge \alpha\}$$

where $\alpha \in (0,1)$.

2.2 Estimating the Value at Risk

This article analyzes value-at-risk approaches. We generate simulation approaches over 260 days. The historical sample data covers January 4, 2021, to December 29, 2023 (780 days). The data consists of daily exchange rates (close prices collected by Metra trader 5) against the U.S. dollar for the Euro. For example, the EUR/USD exchange rate is 1.07. It takes 1.07 US Dollars (USD) to buy one Euro (EUR). The simulation methodology consists of two steps:

1. (Parameter estimation) We use the Kolmogorov-Smirnov test to test the rate of return. The results show that there is a statistical value equal to 0.04092. At a significance level of 0.05, there will be a critical value equal to 0.04866, making it accepted that there is a normal distribution $N(\mu, \sigma^2)$ with a significance level of 0.05. We consider EUR/USD exchange rate simulation via daily rate of return, which has a normal distribution $N(\mu, \sigma^2)$. Using maximum likelihood estimation, we have

$$\mu = \frac{1}{779}\sum_{i=1}^{779} r_i \quad \text{and} \quad \sigma^2 = \frac{1}{779}\sum_{i=1}^{779}(r_i - \mu)^2$$

where $r_i = \frac{p_i - p_{i-1}}{p_{i-1}}$, $i = 1,2,3,\ldots,779$, and $p_0, p_1, \ldots, p_{779}$ are historical EUR/USD exchange rate (780 days). We obtain that $\mu = -0.00012$ and $\sigma = 0.00496$.

2. (Approximate probability of loss) We generate the sequence of daily rate of return of EUR/USD exchange rates over 260 days to approximate the probability of loss using 10,000 scenarios. For each percent of loss $x = 0.01,\ 0.02, \ldots, 0.20$.

   2.1 Set $count = 0$ and $j = 1$.

   2.2 For the *j-th* scenarios, the price at time m is calculated by

$$P_m^{(j)} = P_0^{(j)} \prod_{i=1}^{m} \left(1 + R_i^{(j)}\right)$$

   where $R_i^{(j)} \sim N(-0.00012, 0.00496), m = 1,2,3,\ldots,259$ and $P_0^{(j)} = 1.2247$ is the fixed EUR/USD exchange rate on January 4, 2021, for all j.

   2.3 Use RSIBS (30,70) strategy on $P_0^{(j)}, P_1^{(j)}, \ldots, P_{259}^{(j)}$. We have the percent of loss is $Loss(j)$.

$$count = \begin{cases} count + 1, & \text{If } Loss(j) \le x \\ count + 0, & \text{If } Loss(j) > x \end{cases}$$

   2.4 Let $j = j + 1$ and repeat 2.2 and 2.3 until $j > 10{,}000$.

   2.5 Probability of loss is approximated by $\frac{count}{10000}$.

Finally, we use quadratic polynomial regression to determine the value at risk. The simulation measures investment risk at 95% and 99% confidence levels.

3 Results

Estimating parameters of the rate of return with a normal distribution using maximum likelihood estimation, the average and standard deviation are -0.00012 and 0.00496, respectively. So, we create a rate of return from the parameters obtained, and then, we simulate the situation by creating a closing price of EUR/USD (260 days) using the rate of return value. We repeat this process 10,000 times and get the results in table 1. Table 1 shows the probability of loss of the portfolio at loss from 0.01 to 0.20.

Table 1: Probability of loss obtained from simulation by setting value at risk

Percent of loss	**Probability of loss**	**Percent of loss**	**Probability of loss**
0.01	0.6753	0.11	0.9726
0.02	0.7144	0.12	0.9796
0.03	0.7707	0.13	0.9883
0.04	0.8056	0.14	0.9919
0.05	0.8544	0.15	0.9943
0.06	0.8771	0.16	0.9968
0.07	0.9095	0.17	0.9977
0.08	0.9273	0.18	0.9988
0.09	0.9488	0.19	0.9991
0.10	0.9614	0.20	0.9996

We want to estimate the value at risk at 95% and 99% confidence levels. So, we took the loss and the probability of loss and plotted the graph in Figure 1. We found that the quadratic polynomial graph of cumulative distribution function of loss was the closest to the point. The quadratic polynomial regression with an R-square value of 98.96% is $y = -14.759x^2 + 4.6428x + 0.6425$. So, we solve the quadratic polynomial regression to find the value at risk at the probability of loss from 0.80 to 0.99. Table 2 shows the value at risk of the portfolio at 95% and 99% confidence levels, which are 0.0948 and 0.1227, respectively. This means that with 95% confidence level, the maximum loss will not exceed 9.48%. Similarly, with 99% confidence level, the maximum loss will not exceed 12.27%.

Table 2: The value at risk obtained from solving the quadratic polynomial regression

α	**Value at risk**	α	**Value at risk**
0.80	0.0387	0.90	0.0719
0.81	0.0416	0.91	0.0760
0.82	0.0445	0.92	0.0802
0.83	0.0476	0.93	0.0848
0.84	0.0507	0.94	0.0896
0.85	0.0539	0.95	0.0948
0.86	0.0573	0.96	0.1005
0.87	0.0607	0.97	0.1068
0.88	0.0643	0.98	0.1140
0.89	0.0680	0.99	0.1227

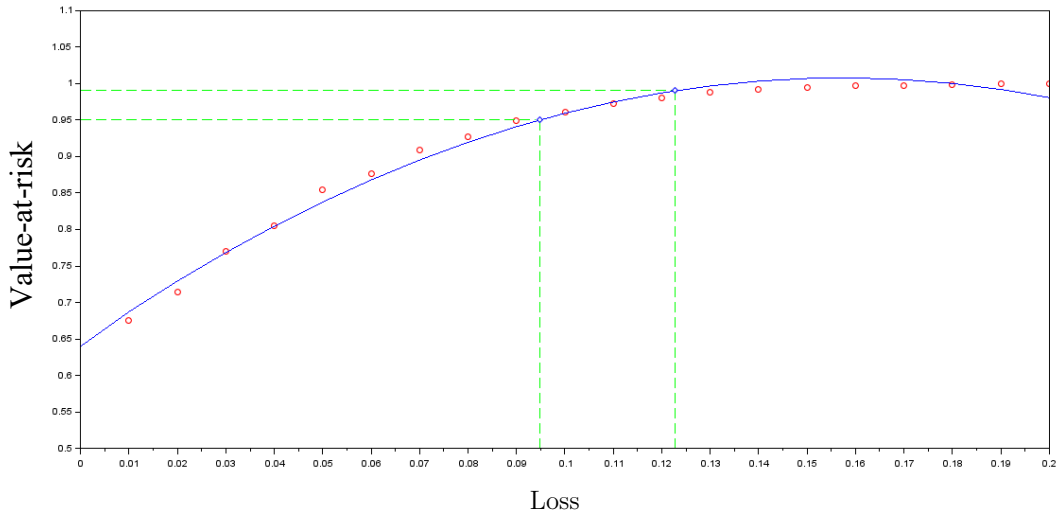


Figure 1: The quadratic polynomial graph of loss and the value at risk. The red circle is the loss. The blue line is the quadratic polynomial regression. The blue diamond is the value-at-risk for the portfolio at 95% and 99% confidence levels

4 Conclusions

From the historical sample, the daily EUR/USD exchange rates data covers January 4, 2021, to December 29, 2023 (780 days). We generate a rate of return over 260 days by average and standard deviation (-0.00012,0.00496) of this historical sample. We use the daily close price of EUR/USD on January 4, 2021, as the starting point for the calculation. Then, we generate a daily close price of EUR/USD by using the rate of return that was generated. Next, we simulate the buy-and-sell strategy in the EUR/USD exchange rate in 10,000 scenarios, using the Relative Strength Index (RSI) indicator to approximate the probability of loss. Then, we use quadratic polynomial regression with an R-square value of 98.96%, $y = -14.759x^2 + 4.6428x + 0.6425$, to estimate the value at risk. The resulting value-at-risk for the portfolio at 95% and 99% confidence levels were estimated to be 0.0948 and 0.1227, respectively. Thus, the maximum loss will not exceed 9.48% at a 95% confidence level. The maximum loss will also not exceed 12.27%, with 99% confidence. Thus, there is less risk. Most investors focus on returns without calculating the risks involved. This study recommends estimating the value-at-risk of an asset before investing. In addition, the value obtained should be acceptable to investors.

Further studies should investigate the value-at-risk of another buy/sell threshold parameter. For example, Is the RSI(20,80) strategy more or less risky than the RSI(30,70) strategy on the EUR/USD exchange rate? In addition, further studies should investigate and compare the value-at-risk of other foreign exchange rates.

Acknowledgment. The research capability enhancement program through graduate student scholarship, Faculty of Science, Khon Kaen University, is gratefully acknowledged.

References

[1] Anderson, B., and Li, S., *An investigation of the relative strength index*. Banks & bank systems, 10(1) (2015), 92-96.

[2] Anson, M. J. P., Chambers, D. R., Black, K. H., and Kazemi, H., CAIA Level I: An Introduction to Core Topics in Alternative Investments. 2nd Edition. Hoboken, New Jersey: John Wiley & Sons, 2012.

[3] Bank for International Settlements, *Triennial Central Bank Survey of Foreign Exchange and Over-the-counter (OTC) Derivatives Markets in 2022,* October, 2022.

[4] Hakim, H., *Forex Trading and Investment*. Diss., Worcester Polytechnic Institute, 2012.

[5] Hendricks, D., *Evaluation of value-at-risk models using historical data*. Economic policy review, 2(1) (1996).

[6] Marek, P., and Sediva, B., *Optimization and Testing of RSI*. In 11th International Scientific Conference on Financial Management of Firms and Financial Institutions, 2017.

[7] Wilder, J. W., New concepts in technical trading systems. Greensboro, NC, 1978.

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Mechanistic Modeling of Financial Bubble Driven by Herding Behavior and Safe-Haven Asset

Sorathan Juanjenkit[1,†] and Klot Patanarapeelert[1,‡]

[1]Department of Mathematics, Faculty of Science
Silpakorn University, Nakhon Pathom, 73000, Thailand

Abstract

Safe-haven strategy usually used to reduce the risk among the market turbulence. It is hypothesized that inclusion of safe-haven asset may reduce the market volatility during the bubble. In this study, we propose the new model of financial bubble that generalizes the previous models by adding the safe-haven asset that interacts with the behavioral change of investors. The stability condition is derived to confine the parameter space avoiding the stable fixed point. The numerical results are used to calculate the amplitude and duration of bubbles. The effect of involved parameters are analyzed. This result indicates that information from a safe-haven asset model based on mean reversion helps reduce the severity of financial bubbles resulting from herd behavior of profit seekers in the market. Additionally, it suggests that if these profit seekers consistently use data from safe-haven assets in the market, the severity of financial bubbles would decrease significantly compared to when profit seekers are interested in safe-haven assets only during crisis events.

Keywords: financial bubbles, safe-haven asset, price dynamic, herding behavior.

2020 MSC: Primary 91B55; Secondary 34A34, 37-XX, 82-XX.

1 Introduction

Financial bubbles are economic phenomena that have occurred multiple times in history. The definition or description of financial bubbles and the process of their bursting continue to vary and have diverse interpretations. For instance, a definition related to financial bubbles by Didier Sornette suggests that if the price of an asset experiences rapid growth beyond exponential, there is a possibility that the asset may become a financial bubble [10]. Another definition highlights that financial bubbles and the bursting of financial bubbles are temporary events where asset prices deviate and fluctuate around their fundamental value temporarily [8]. One prominent example of a financial bubble event is the Subprime Crisis of 2008. According to 'Review of economic bubble (2016)' [5], the crisis was initiated by a continuous increase in real estate accompanied by loose monetary policies of central banks and governments, which reduced interest rates to encourage more people to own real estate. Additionally, the softening

†Speaker. ‡Corresponding author.
Email: juanjenkit.sorathan@gmail.com (S. Juanjenkit), klotpat@gmail.com (K. Patanarapeelert).

of lending standards brought subprime borrowers into the market. All these factors compounded the growth of real estate, leading people to speculate and invest more, resulting in skyrocketing real estate prices. While everyone was enjoying the prosperity of life, some events were unfolding in the background. 'Inflation' has started creeping in gradually. The low-interest rates, combined with subprime borrowers, led to people defaulting on their loans, and debts began to pile up rapidly. Many homes were foreclosed by banks and released into the market simultaneously with decreased consumer spending. People panicked and wanted to minimize their losses as much as possible, but it was too late.

The research about the financial bubble has been conducted and explored from various perspectives in recent years [11], [7], [10] and [4]. Questions such as where financial bubbles originate, how the mechanics of financial bubbles work, when financial bubbles form and burst, or what factors are related to the occurrence or size of financial bubbles are central to current research. These questions were addressed through various disciplines. For instance, [2] suggested that risky monetary policies by governments and central banks are factors in the emergence of profit-seeking bubbles in the market. Thomas Lux states that financial bubbles arise from the collective behavior and sequential actions of investors in buying or selling until an imbalance occurs between buying and selling demand [8]. What supports the readiness in the behavior of investors to follow each other is fundamental economic variables such as actual returns. While the previous study highlighted the possible influence of herding behavior and the feedback of price during the bubble event, hedging strategy that helps to minimize and offset risks within the portfolio of investors was neglected. Financial hedging is more common amongst short-term noised traders, as market volatility tends to increase. However, research analyzing the impact of other assets on the financial bubble of another asset is relatively limited. Which asset is most important to people in the market?

According to Baur and Lucey (2010), A safe-haven is defined as an asset that is uncorrelated of negatively correlated with another asset or portfolio in times of market stress or turmoil. As an example, gold or land, which are well-known and have long lasting value over time may be considered as safe-haven assets for stock trading and other risky asset. A safe-haven asset must therefore be some asset that holds its value in 'stormy weather' or adverse market conditions" [3]. For some profit-seeking investors, using safe haven assets to hedge against risk is one of the investment and risk mitigation strategies [1]. In some cases, it is not necessary to include the safe-haven asset into their portfolio but use safe-haven asset data, price volatility of assets, or market returns to make trading decisions in other profit-generating assets. If this is a case during the financial bubble event, how the bubble pattern changes or conditioned might be a crucial issue. In this study, we aim to address these questions and will provide insight into what happens when profit-seeking investors in the market employ strategies and track the price movements of profit-generating assets. How will the financial bubble of those assets develop, shrink, or expand, and what impact will it have on the severity of the economic aftermath?

Due to the wide variety of valuable researches on financial bubbles, all accompanied by various definitions of bubbles, we must choose just one that we deem suitable as a solid foundation to begin addressing the questions. This study extends Thomas Lux's model that emphasized on the impact of Herding behavior on Bubble, and Crash. Herding behavior was explained as events stemming from the collective behavior and sequential actions of investors in buying or selling, leading to an imbalance between buying and selling demand. What reinforces the readiness in investors' behavior to follow each other is fundamental economic variables such as actual returns. This type of model will be integrated with the model of safe-haven asset that we will present in the next section. The model results will be subsequently used to investigate and compare the effect of crucial parameters.

2 Models

2.1 Review of Lux's Model

First, we will present for mutual understanding the characteristics of the market under consideration and the definition of the financial bubble based on the previous study. A key feature of the market is that profit-seekers in the market exhibit a behavior known as herd behavior, wherein profit-seekers tend to follow the direction of the crowd in one direction. We presume a market population consisting of a total of $2N$ market participants. Within this population, individuals are divided into two ideological groups: those who perceive the market negatively, denoted as n_-, representing individuals predisposed to selling assets, and those who view the market positively, denoted as n_+, representing individuals inclined to purchase assets. Additionally, investors are assumed to make buy or sell decisions based on the contagion process, where each individual is immediately prepared to switch from their current group to the larger or predominant group. We introduce the superiority of each group's population with $x = (n_+ - n_-)/2N$, where x is within the range $[-1, 1]$. In cases where $x > 0$, it indicates a prevailing demand for buying assets in the market; conversely, $x < 0$ denotes a predominance of selling. When $x = 0$, it signifies market equilibrium, while $x = 1$ and $x = -1$ represent extreme cases where all market participants converge on the same perspective.

Next, our focus shifts to the properties of market participants, specifically herd behavior or the contagion process within the market. In the market under consideration, we assume that individuals' decisions depend on others within the market, meaning each market participant's decision to buy or sell assets depends on the prevailing sentiment or noise in the market. We further assume that at any given moment, individuals in the market have a probability of switching from being buyers to sellers or vice versa, denoted as p_{-+} and p_{+-}, respectively. Conversely, in the opposite direction, we have p_{+-} and p_{-+}, which, combined with the contagion process, are determined by the collective sentiment of market participants x. Thus, we define $p_{-+} = p_{-+}(x)$ and $p_{+-} = p_{+-}(x)$ based on the overall market sentiment x.

Since there are the probabilities of the transition between optimistic one and pessimistic one, such that we are starting to consider the change of average disposition x. Consequently, we expect fraction $n_-\mathrm{p}_{+-}$ to switch from the n_- to the n_+ group which means those who are pessimistic traders turn to an optimistic attitude with probability p_{+-}, and vice versa. From this it follows that the change in time of the number of optimistic and pessimistic traders is : $dn_+/dt = n_-\mathrm{p}_{+-} - n_+\mathrm{p}_{-+}$ and $dn_-/dt = n_+\mathrm{p}_{-+} - n_-\mathrm{p}_{+-}$. Including with n and x that we defined:

$$\begin{aligned} dx/dt &= [(N-n)\mathrm{p}_{+-}(x) - (N+n)\mathrm{p}_{-+}(x)]/N \\ &= (1-x)\mathrm{p}_{+-}(x) - (1+x)\mathrm{p}_{-+}(x). \end{aligned} \tag{2.1}$$

We note that the original arguments serve the stochastic model. However, the derivation of this equation was carried out via the Master equation approach which calculated the dynamic of expectation of the population variables.

The transition probabilities will be specified in order to perceive how (2.1) potentially-describes. Note that the requirements for p_{+-} and p_{-+} is, (1) all transition probabilities have to be positive, (2) if the prevailing disposition of the population is already optimistic then $\mathrm{p}_{-+} > \mathrm{p}_{+-}$. Moreover, it seems reasonable to assume that $\mathrm{dp}_{+-}/\mathrm{p}_{+-} = a\ dx$, that is the relative changes in probability to switch from pessimism to optimism increases linearly with-changes in x, and vice versa $\mathrm{dp}_{-+}/\mathrm{p}_{-+} = -a\ dx$. These assumptions may suggest the following functional form commonly chosen in the related literature:

$$\mathrm{p}_{+-}(x) = ve^{ax}, \quad \mathrm{p}_{-+}(x) = ve^{-ax}. \tag{2.2}$$

Here, a gives a measure for the strength of herd behavior, v is a variable for the speed of change ($x = 0$, balanced disposition we have $\mathrm{p}_{+-} = \mathrm{p}_{-+} = v > 0$). This means that a little change from equilibrium point is the starting point of herd behavior.

Follow by properties of the hyperbolic sine and cosine and this specification of transition rates the time development of the mean value of the index x becomes:

$$\begin{aligned} dx/dt &= (1-x)ve^{ax} - (1+x)ve^{-ax} = 2v[\sinh(ax) - x\cosh(ax)] \\ &= 2v[\tanh(ax) - x]\cosh(ax). \end{aligned} \tag{2.3}$$

The equation (2.3) represents changes in the majority Sentiment of the market. As the price of focusing securities changes according to the excess demand, the further assumption relies on the direct proportionality of the excess demand on the market sentiment and the deviation of price from the fundamental value. These two factors used the different proportionality constants that distinguishes between the trading volume of speculative investors and of fundamentalists. The corresponding dynamics are given by

$$\begin{aligned} \frac{dx}{dt} &= 2v[\tanh(a_1\dot{p}/v + a_2x) - x]\cosh(a_1\dot{p}/v + a_2x), \\ \frac{dp}{dt} &= \beta[xT_N + T_F(p_f - p)], \end{aligned} \tag{2.4}$$

where dp/dt, representing the rate of change in the price of the underlying asset. Fundamental traders, who trade based on the perceived discrepancy between current prices-and fundamental values, and Noise traders, who follow others' actions. The excess demand of Fundamental traders is denoted by $T_F(p_f - p)$, where T_F is the trading volume of Fundamental-traders, and p_f is the Fundamental price of the underlying asset. On the other hand, Noise-traders' excess demand is represented by xT_N, with T_N being the trading volume of Noise-traders. a_1 is weight factor describing how much information investors try to draw from price and a_2 is weight factor describing how much information investors drawn from the behavior of others.

Furthermore, the contagion process and price dynamics have different mean time lags, denoted by $1/v$ and $1/\beta$, respectively. Assuming instantaneous market clearing, the equation implies that $p = p_f + (T_N/T_F)x$ and $\dot{p} = (T_N/T_F)\dot{x}$, where the expected returns influence the readiness of profit-seekers to follow suit in the market. This readiness is influenced by the cumulative difference between the true returns of the underlying asset and the expected returns in the market.

$$\begin{aligned} \frac{dx}{dt} &= 2v[\tanh(a_0 + a_2x) - x]\cosh(a_0 + a_2x), \\ \frac{da_0}{dt} &= \tau\left[\frac{r + \tau^{-1}(T_N/T_F)\dot{x}}{p_f + (T_N/T_F)x} - R\right], \end{aligned} \tag{2.5}$$

Here, r is the nominal dividend payment and defines $R = r/p_f$ as the expected return, with τ interpreted as an adjustment coefficient. Finally, it is noted that when the accumulated market return a_0 becomes less than 0, it indicates the occurrence of a financial bubble burst.

2.2 The Model with Safe-Haven Asset

Suppose that the safe-haven has an impact on the decision of all traders with some weight. In this section we include the price dynamic of the safe-haven in Equation (2.5). We adopt the assumption that the change in return of a safe-haven asset denoted by s follows a mean-reverting process, as discussed in [9]. To adhere to the definition of a safe-haven asset as outlined in [3], it is an asset that attracts profit-seekers when the underlying asset market enters a downturn, and diminishes in attractiveness when the market returns to a profitable state. Building on the research by [2], which emphasizes the significance of economic indicators on financial bubbles, we introduce $-Ex$ as a factor in our safe-haven asset model, where $E > 0$ represents a basic economic factor influencing investors (akin to a weight factor). Hence, the safe-haven asset model is given by:

$$\frac{ds}{dt} = \alpha\left[T_S(s_f - s) - Ex\right] \tag{2.6}$$

In this context, α signifies the speed of change of the safe-haven asset, T_S represents the trading volume for the safe-haven asset, and s_f denotes the fundamental price of the safe-haven asset. In this equation, the return of safe-haven tends to decrease as the current financial market booms. Furthermore, considering the economic indicators' involvement with financial bubbles, it becomes imperative to contemplate a new model for the underlying asset. Therefore, we derive the following equation:

$$\frac{dp}{dt} = \beta[xT_N + T_F(p_f - p) + E] \tag{2.7}$$

It is clear that Equation (2.6)-(2.7) describes that for both asset price are driven by the market sentiment. As a result, the dynamic of accumulative return is adjusted as

$$\frac{da_0}{dt} = \tau \left[\frac{r + \tau^{-1}(T_N/T_F)\dot{x}}{p_f + (T_N/T_F)x + E/T_F} - R \right] \tag{2.8}$$

Here, $R = r/(p_f + E/T_F)$. To complete the model modification, we extend the transition probability by assuming that additional information is also drawn from the safe-haven return with directly but negatively proportional to the change in return of the safe-haven asset. By this assumption, dynamic of safe-haven becomes negatively associated with the price of the focusing asset. However, this assumption can be considered as two possibilities that is the relationship between two assets can be discrete and continuous. Thus, it is reasonable to separate the model into two sub-models as follows.

2.2.1 Model 1: Continuous Relationship

In light of the definition of safe-haven assets as cited in [3], it is reasonable to include $a_3 ds/dt$ as a factor influencing the readiness of profit-seekers to follow the crowd in the market. Consequently, the system of equations for the financial bubble model we are considering is represented by the following equation:

$$\begin{aligned} \frac{dx}{dt} &= 2v[\tanh(a_0 + a_2 x + a_3 \dot{s}) - x]\cosh(a_0 + a_2 x + a_3 \dot{s}), \\ \frac{da_0}{dt} &= \tau \left[\frac{r + \tau^{-1}(T_N/T_F)\dot{x}}{p_f + (T_N/T_F)x + E/T_F} - R \right], \\ \frac{ds}{dt} &= \alpha \left[T_S(s_f - s) - Ex \right], \end{aligned} \tag{2.9}$$

where a_3 is an adjustment coefficient that express strength of safe-haven asset to herd behavior.

2.2.2 Model 2: Discrete Relationship

In order to align more closely with the safe-haven asset's definition we have discussed. We now define function $A(a_0)$,

$$A(a_0) = \begin{cases} 0, & \text{if } a_0 \geq 0 \\ 1, & \text{if } a_0 < 0 \end{cases}$$

Incorporating the term $A(a_0)$ to refine and adjust equation (2.9) would enhance the system to adhere more closely to the defined definition. The influence of the returns of safe-haven assets on market participants' decision-making would come into play only when the value of a_0, or the accumulated actual return of the market is negative. Now we have

$$\frac{dx}{dt} = 2v[\tanh(a_0 + a_2 x + a_3 \dot{s} A(a_0)) - x]\cosh(a_0 + a_2 x + a_3 \dot{s} A(a_0)). \tag{2.10}$$

where a_3 is defined as the same as the previous model. Therefore, the present models are (2.10)

2.3 Stability Analysis

In this section, we aim to determine the (local) stability condition for the equilibrium point of model (2.9). This is because when considering the definition of a bubble as a transient situation where prices oscillate around the fundamental price, analyzing the stability of the system becomes an important aspect. As the bubble may occur when the system undergoes the unstable equilibrium state, the derived condition can be used to confine the parameter space for further investigation.

To determine the equilibrium point of the system (2.9), we first put $dx/dt = 0$, $da_0/dt = 0$, and $ds/dt = 0$, respectively. As a result, it is obvious that $da_0/dt = 0$ is always true. So, we consider only remained two equations. We also observe that $x = 0$ is only solution for the first equation. Hence, $s = s_f$ is a result. Therefore, we can conclude that our system inherently possesses a unique equilibrium $E(x, a_0, s) = E(0, 0, s_f)$, representing a scenario where the majority of dispositions are balanced, actual returns are zero, and the price of the safe-haven asset equals its fundamental price. For assessing system stability, we rely on the Routh-Hurwitz stability criterion [6]

After computing the coefficients of the characteristic equation of our differential equation system and constructing the Routh-Hurwitz array, we identified the stability conditions as follows: the equilibrium is stable if and only if either $a < 0$ and $b < 0$. The values of a and b are determined as follows:

$$\begin{aligned} a &= -\alpha T_S + 2vC - 2a_3\alpha Ev + \frac{2T_N Rv}{rT_F}, \\ b &= 2v\left(\frac{RT_N(\alpha T_S - \tau R)}{rT_F} + \alpha T_S\left(C + \frac{\tau R^2 T_N}{\alpha r T_F(T_S + 2a_3 Ev) - 2v(CrT_F + RT_N)}\right)\right), \end{aligned} \tag{2.11}$$

where $R = r/(p_f + E/T_F)$ and $C = a_2 - 1$.

Before proceeding to the next section, it's important to acknowledge the scope of our stability analysis. While we have successfully identified conditions under which our system exhibits instability, it's essential to note that our focus has been primarily on understanding fluctuations around the fundamental price. However, it's worth mentioning that determining conditions for periodic events remains an ongoing challenge. Despite this limitation, our analysis provides valuable insights into the behavior of our system within the context of instability.

3 Results

In the context of financial bubble phenomena, two factors can indicate its severity. First is its size, which refers to the magnitude of its price fluctuations around the fundamental value, represented by the height from crest to trough. The second factor is its duration, which represents the time it takes for the price fluctuations to complete one cycle, indicated by the length from crest to crest.

We have omitted the analysis of events in the early stages of the mechanism concerning size and duration in both (2.9) and (2.10) due to their non-periodic nature. Instead, we focus on the analysis of events in the second stage when the system exhibits periodic solutions, as depicted in Figure 1.

In this section, we calculate the two indicators from the numerical solutions of the models using the parameter values in Table 1. To verify whether the results align with our hypothesis, which posits that the inclusion of information from safe-haven assets reduces the severity of financial bubbles which are temporary events where asset prices deviate and follow with the fluctuation around their fundamental value, we consider the stability conditions outlined in the previous section. Given the unique equilibrium point of the system, it is sufficient to select parameters that induce instability in system (2.9) for this analysis.

Table 1: Parameter values used in numerical calculations

Parameter	Description	Value(unit)
a_2	Strength of herd behavior	1.125
a_3	Strength of safe-haven asset	1.25
r	Constant nominal dividend payment	1.0
T_N	Trading volume of speculative investor	21/160
T_F	Trading volume of fundamental investor	3/4
p_F	Fundamental price of underlying asset	7/10
T_S	Trading volume of safe-haven asset	1.0
s_F	Fundamental price of safe-haven asset	1.3
α	Speed of change on safe-haven asset	1.0
β	Speed of change on underlying asset	1.0
E	Economic factor	0.02
v	Speed of change on probability	0.5
τ	Adjustment coefficient	1.0

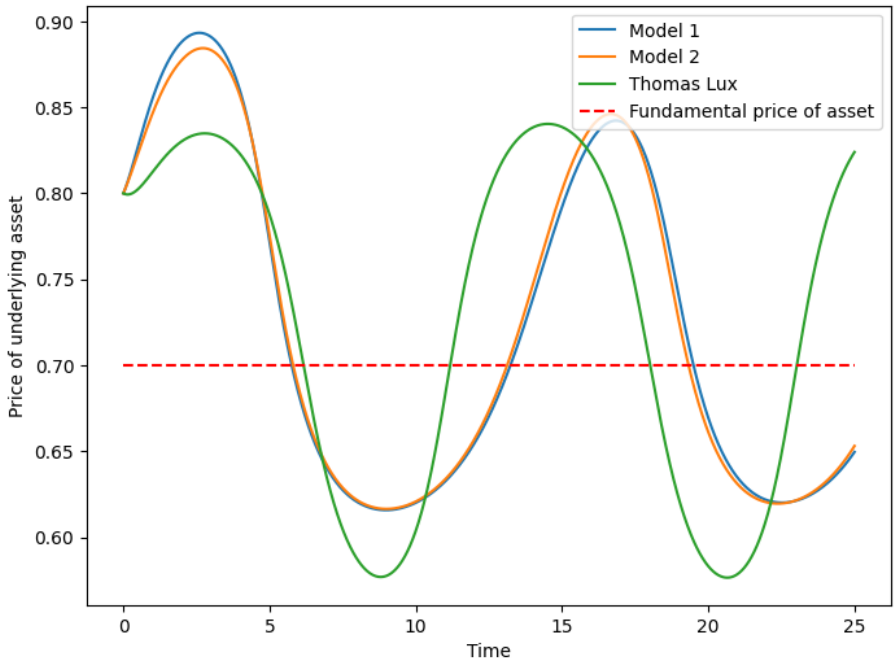


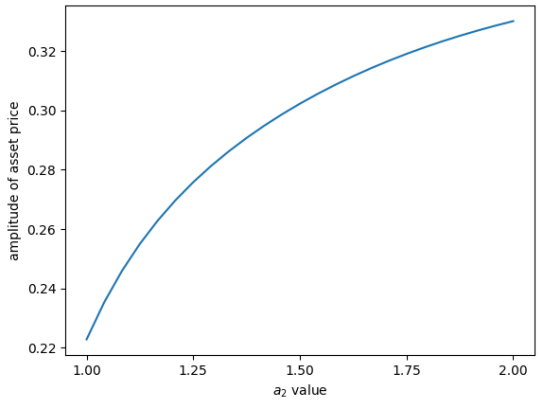
Figure 1: Sample of price dynamics/movements of the underlying asset for each model with a set of initial conditions $p = 0.8$, $x = 0.5$, $a_0 = 1$ and $s = 1$

3.1 Model 1's Result

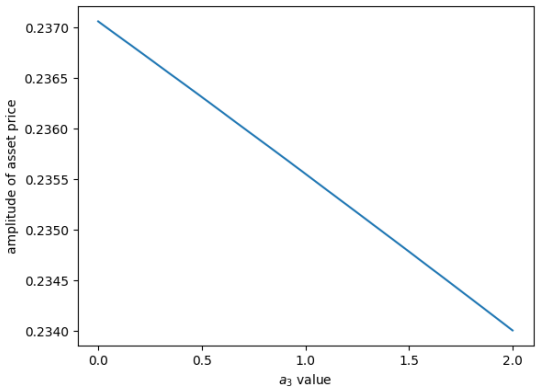
According to model (2.9), it demonstrates how safe-haven assets play a role in investors' decision-making at all times. When considering the weight factor variable a_3, which represents the weight that profit-seekers give to information about safe-haven assets, from Figure 2b, it can be observed that as a_3 increases, the height of the bubble decreases. In this scenario, we might argue that when profit-seekers who exhibit herding behavior take a moment to observe information from safe-haven assets before considering buying/selling the underlying asset they are interested in, in cases where these profit-seekers make mistakes in their decision-making, it may help reduce the resulting losses.

As for the weight factor variable a_2, which represents the weight that profit-seekers give to

the noise of the crowd before considering buying/selling the underlying asset they are interested in, from Figure 2a, it can be observed that as a_2 increases, the height of the bubble also increases. It is evident that when people are ready to make decisions to buy/sell the underlying asset solely because others are doing so, it is not surprising that the price of this asset may soar to the sky or plummet underground.

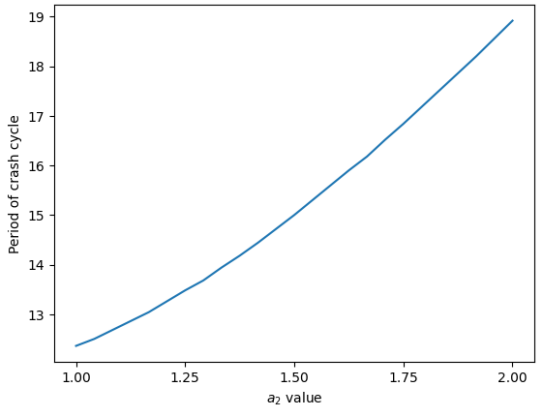


(a) Impact of a_2 on asset's amplitude as a_3 is 1

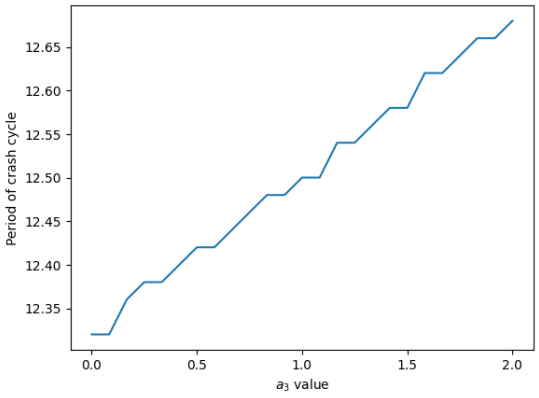


(b) Impact of a_3 on asset's amplitude as a_2 is 1

Figure 2: Impacts of a_2 and a_3 on underlying asset's amplitude of model 1 as a_3 is 1 and a_2 is 1 respectively



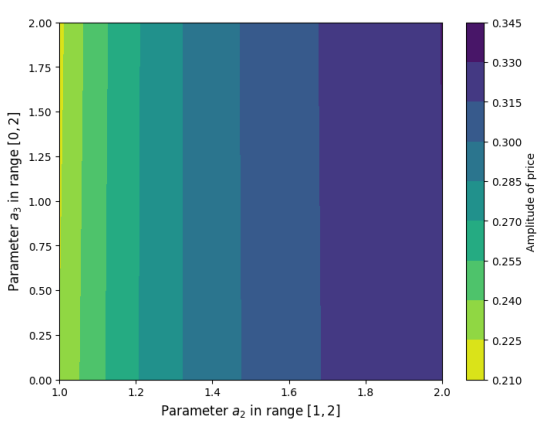
(a) Impact of a_2 on asset's period as a_3 is 1



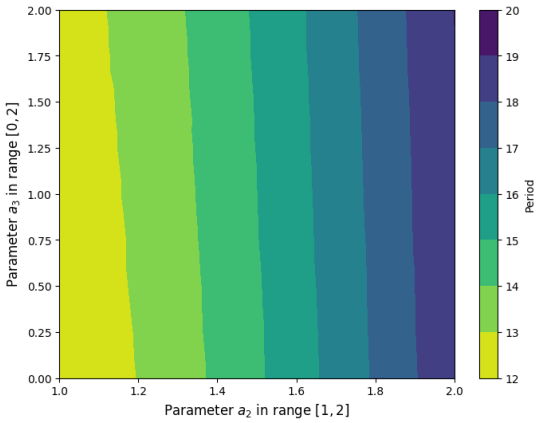
(b) Impact of a_3 on asset's period as a_2 is 1

Figure 3: Impacts of a_2 and a_3 on underlying asset's period of model 1 as a_3 is 1 and a_2 is 1 respectively

Upon examining financial bubble in term of the duration in Figures 3a and 3b, both variables a_2 and a_3 yield similar results. That is, as these variables increase, the duration of price fluctuations around the fundamental value for one cycle also increases. This may be beneficial as it suggests a decrease in market volatility.



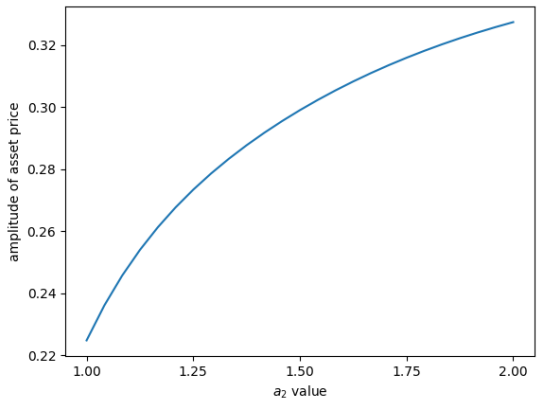
(a) Impacts of a_2 and a_3 on underlying asset's amplitude



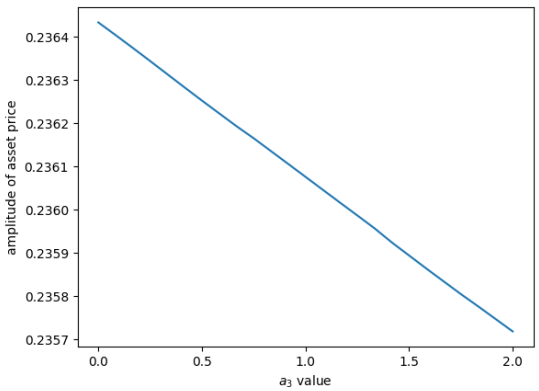
(b) Impacts of a_2 and a_3 on underlying asset's period

Figure 4: Impacts of a_2 and a_3 on underlying asset's price of model 1 in contour plot

Figures 4a and 4b represent contour plots illustrating the impacts of a_2 and a_3 on the underlying asset's price in terms of amplitude and period in Model 1. We have seen that the change in combination of two parameters does not make significant change of the amplitudes and periods from the pattern when fixing one parameter. The results are more relatively sensitive to the change of a2 than a3. The contour plot shows that the safe-haven strategy and the herding behavior are uncorrelated.



(a) Impact of a_2 on asset's amplitude as a_3 is 1

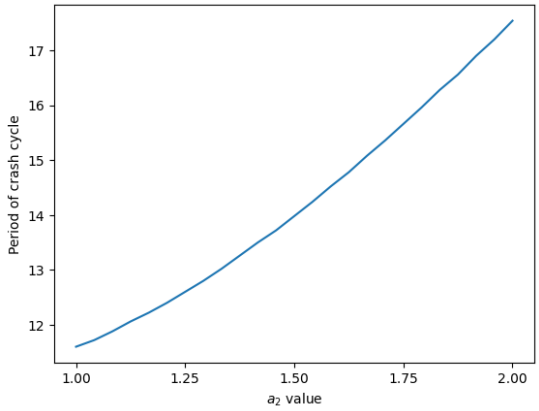


(b) Impact of a_3 on asset's amplitude as a_2 is 1

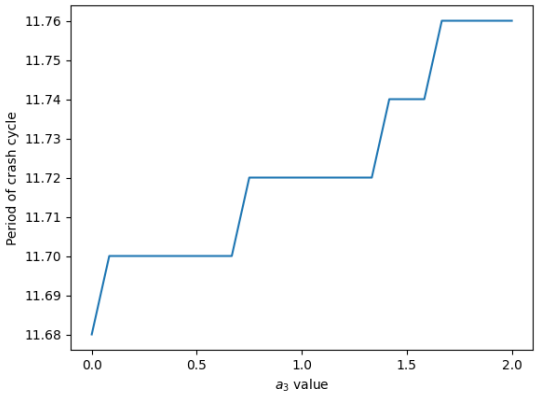
Figure 5: Impacts of a_2 and a_3 on underlying asset's amplitude of model 2 as a_3 is 1 and a_2 is 1 respectively

3.2 Model 2's Result

For the results of model (2.10), where we stated that safe-haven assets play a role in investors' decision-making only when the market enters a crisis or downturn, as indicated by the actual return a_0 being less than 0, the outcomes, whether in terms of amplitude as shown in figures 5a and 5b, or in terms of period as shown in figures 6a and 6b, yield similar Model 1(a continuous relationship) both numerical result and interpretations. However, when comparing the outcomes of both models from both the amplitude and period perspectives by the effect from a_3, it is evident that Model 1 provides better results in both aspects, as depicted in figures 8a and 8b. Therefore, we can conclude that investors' continuous interest in safe-haven assets at all times leads to less market volatility compared to when they only pay attention to them during crisis periods.

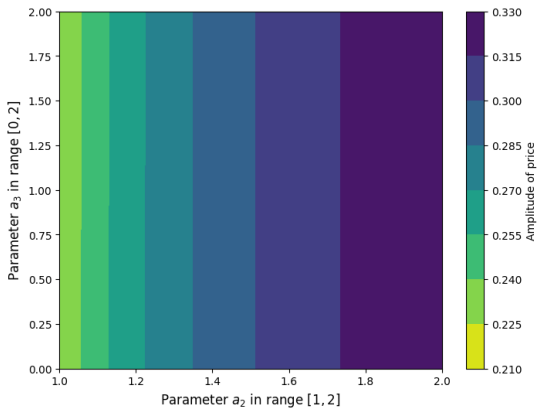


(a) Impact of a_2 on asset's period as a_3 is 1.

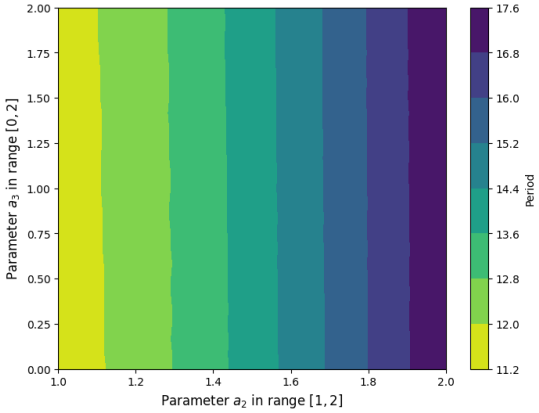


(b) Impact of a_3 on asset's period as a_2 is 1

Figure 6: Impacts of a_2 and a_3 on underlying asset's period of model 2 as a_3 is 1 and a_2 is 1 respectively

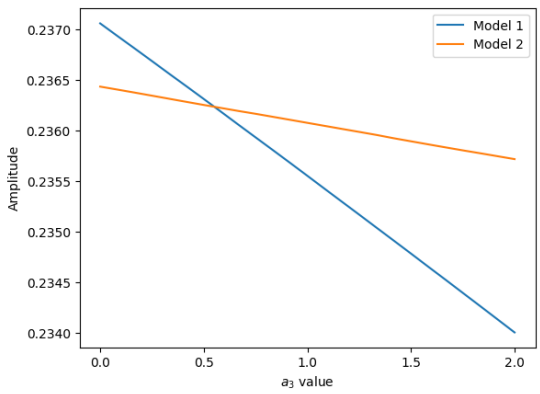


(a) Impacts of a_2 and a_3 on underlying asset's amplitude

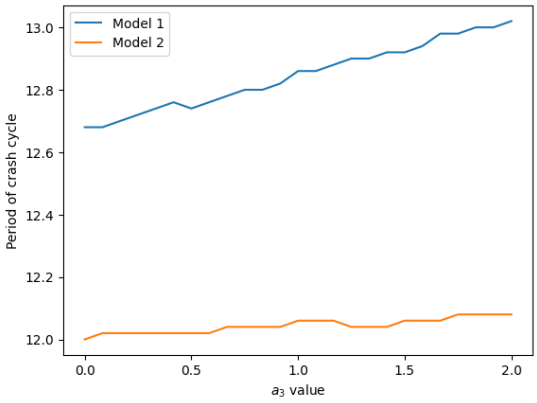


(b) Impacts of a_2 and a_3 on underlying asset's period

Figure 7: Impacts of a_2 and a_3 on underlying asset's price of model 2 in contour plot



(a) Comparison between model 1&2 on amplitude from impact of a_3



(b) Comparison between model 1&2 on period from impact of a_3

Figure 8: Comparison between model 1&2 on amplitude and period from impact of a_3

Figures 7a and 7b represent contour plots illustrating the impacts of a_2 and a_3 on the underlying asset's price in terms of amplitude and height in Model 2.

4 Conclusion

Our findings from the safe-haven asset model, as illustrated by the Mean Reversion model, confirm our underlying assumption in both respects. This indicates that when investors base their decisions to buy or sell the underlying asset on information beyond mere consensus, it can reduce market volatility. In this context, volatility refers to the intensity of financial bubbles. Or in other words, wisdom prevails over emotions.

We have proposed the mechanistic models extended from the previous work. The model composes of the dynamics of disposition variables, the accumulative difference of returns and safe-haven asset. The present model always has only one equilibrium point. The stability conditions are more complicate than of the previous work since the number of parameters of safe-haven asset are added. However, the common necessary conditions are that $a_2 < 1$. This implies that the onset of financial bubble requires the strong influence of herding behavior.

Understanding the existence of financial bubbles and being able to explain them in another form, as we have proposed, would be beneficial for analyzing whether the current situation warrants diversification of our investment risks or not. In addition to that, safe-haven assets are likely to be another option for hedging or portfolio allocation. Since the parameters used in our experiments are not specified, it may be possible to consider the proportions of holding safe-haven assets for hedging or portfolio allocation.

This research, while explaining the influence of safe-haven assets on financial bubbles in a deterministic form, also paves the way for exploring stochastic models. This extension could encompass various aspects, including price prediction models or financial bubble models, sentiment analysis of profit seekers in the market, or expressing it in other forms. There are numerous avenues to explore. Another potential direction is to include other assets beyond safe-haven assets to observe the behavior of profit seekers, price movements, and sentiment, which could be beneficial for hedging or portfolio allocation. Undoubtedly, there is much more to investigate.

As mentioned earlier, this paper is an extension of Thomas Lux's work on "Herd behavior, bubbles and crashes". In this regard, it raises the question of what would happen if other assets were involved with the underlying asset, and we chose it as the safe-haven asset. While our proposed safe-haven asset model may not fully capture the characteristics indicative of a safe-haven asset and could prompt questions about its efficacy, this could serve as a starting point for further development of Thomas Lux's model from another interesting perspective.

References

[1] M. Akhtaruzzaman, S. Boubaker, B. M. Lucey, and A. Sensoy, *Is gold a hedge or a safe-haven asset in the covid–19 crisis?*, Economic Modelling **102** (2021), 105588.

[2] F. Allen and D. Gale, *Bubbles and crises*, The Economic Journal **110** (2000), no. 460, 236–255.

[3] D. G. Baur and T. K.J. McDermott, *Is gold a safe haven? international evidence*, Journal of Banking & Finance **34** (2010), no. 8, 1886–1898.

[4] D. G. Baur and T.K.J. McDermott, *Safe Haven Assets and Investor Behaviour Under Uncertainty*, The Institute for International Integration Studies Discussion Paper Series (2011), no. iiisdp392.

[5] V. Chang, R. Newman, R. J. Walters, and G. B. Wills, *Review of economic bubbles*, International Journal of Information Management **36** (2016), no. 4, 497–506.

[6] G. Franklin, J.D. Powell, and Abbas Emami-Naeini, *Feedback control of dynamic systems*, 1994.

[7] I. Giardina and J. Bouchaud, *Bubbles, crashes and intermittency in agent based market models*, The European Physical Journal B **31** (2003), 421–437.

[8] T. Lux, *Herd behaviour, bubbles and crashes*, The Economic Journal **105** (1995), no. 431, 881–896.

[9] E. S. Schwartz, *The stochastic behavior of commodity prices: Implications for valuation and hedging*, The Journal of Finance **52** (1997), no. 3, 923–973.

[10] D. Sornette and P. Cauwels, *Financial bubbles: Mechanisms and diagnostics*, Review of Behavioral Economics **2** (2015), no. 3, 279–305.

[11] I. Wöckl, *Bubble detection in financial markets - a survey of theoretical bubble models and empirical bubble detection tests*, Working Paper (August 2019).

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Mathematical Model for the Dynamic of COVID-19 Spread and Impacts of Vaccination, Quarantine, and Hospitalization among the 5th Wave of COVID-19 in Thailand

Jiraporn Lamwong[1] and Puntani Pongsumpun[2,‡]
[1]Department of Mathematics, School of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand
[2]Department of Mathematics, School of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand

Abstract

The novel Coronavirus or COVID-19 pandemic is a massive outbreak that has affected almost every country in the world. Many methods have been sought to stop its spreading. A mathematical model is an effective instrument that helps analyze the pandemic situation. In this research, a new model of transmission in Thailand consisting of vaccination, quarantine, and hospitalization is presented, aiming at seeking factors affecting the pandemic and guidelines for reducing the spread of this disease. Equilibrium points and basic reproduction numbers were analyzed and stability was tested. Model fitting was performed to obtain parameter values suitable for the pandemic. Besides, numerical results revealed that infection rates and the efficiency of vaccines played a significant role in reducing the number of patients and controlling the pandemic situation.

Keywords: COVID-19, standard dynamical modeling, model fitting, sensitivity analysis, globally.

2020 MSC: 92-10; 93D20.

1 Introduction

Recently, the world faced the fifth wave of the spread of severe acute respiratory syndrome Coronavirus (SARS-COV-2), widely known as COVID-19 [1]. It was indicated as the most infectious wave since the pandemic was reported. It had a huge effect on those

*This research was financially supported by School of Science, King Mongkut's Institute of Technology Ladkrabang, grant number RA/TA-2565-D-001.
‡ Corresponding author. Puntani Pongsumpun
E-mail address: 65056018@kmitl.ac.th (J. Lamwong), puntani.po@kmitl.ac.th (P. Pongsumpun)

having underlying diseases since the pandemic occurred rapidly. The infection can be transmitted by small respiratory droplets, such as sneezing or coughing or exposure to secretions on surfaces [2]. Therefore, social distancing and wearing surgical masks are one of the various methods that can help prevent the spread of the disease. After getting the infection, the incubation period for the coronavirus is between 2 and 14 days [3-4]. Next, if the human body loses immunity, symptoms among infected people range from muscle pain, body aches, sore throat, dry throat, and high fever to severe symptoms that can destroy the respiratory system [3, 5].

According to the global situation report on 2 November 2023, there were 771,679,618 confirmed cases and 6,977,023 deaths. On 23 October 2023, a total of 13,534,457,273 vaccine doses were reported. As for the situation in Thailand, there were 4,758,125 confirmed cases and 34,487 deaths and on 31 August 2023, a total of 139,343,323 vaccine doses were reported [6]. Based on the current situation, prevention by vaccination is a strategy that the government is focusing on helping to control the disease spread [2,7-8]. Many companies develop their vaccines to be efficient to meet people's needs promptly. Vaccines that are accepted and widely used by the Thai government and private sector are AstraZeneca which is suitable for people aged 18 years and above with 2 doses of the vaccine, 10-12 weeks apart, CoronaVac or Sinovac COVID-19 vaccine is an inactivated vaccine suitable for people aged 18 – 59 years with 2 doses, 2-4 weeks apart, Pfizer is a messenger RNA (mRNA) vaccine suitable for those aged 16 years and above with doses, 21-28 days or 3-4 weeks apart, and other vaccines [9-10]. Preventing vaccination is one of the strategies. Many other strategies will help control the situation like social distancing, wearing surgical masks, and quarantine to help reduce the spread of the virus.

Mathematical modeling plays a vital role in assessing the situation, control efficiency, and preparedness to cope with a future outbreak [11-12]. A lot of researchers are interested in developing a model to keep pace with the current situation. Yang [13] proposed an epidemic model by considering the quarantine population in the pre-incubation phase including the home isolation and hospital isolation. It was found that early isolation could help to control the spread of disease effectively. Ibrahim et al. [14] designed $SVEI_sI_aI_mR$ model to keep up-to-date with the situation by considering vaccination factors, asymptomatic infection, symptomatic infection, and Omicron infection to finely isolate people. The study revealed that vaccination alone was not enough to fight against the spread of COVID-19. There should be other measures to help stop the spread of co-infection. Lamwong et al. [15] designed a standard dynamic model by considering vaccinated people, asymptomatic people, symptomatic people, and hospitalized people and determining the most suitable strategy to control the spread of the disease. Strategies used for the control were vaccination measures and people who received immunity from vaccination. It was found that disease control could be implemented by setting other measures to help control the situation, such as wearing face masks and social distancing, making the disease control more effective.

In this article, importance is given to vaccination, quarantine, and hospitalization. Topics are arranged as follows: Part 2 designs and describes the dynamic of the disease in the model. Equilibrium points and basic reproduction number are found and the stability of DEF and EE is tested. Part 3 presents a numerical model by analyzing actual data of the spread in Thailand in conjunction with the model. Meanwhile, the sensitivity index is analyzed to

examine input parameters affecting basic reproduction number. The final part prepares the conclusion as shown in Part 4.

2 Materials and Methods

2.1 Model Formulation

Mathematical modeling is a method that helps analyze the spread situation, designing control measures and finding strategies to help prevent the spread of COVID-19. In this research, the model was designed by dividing people into 7 groups, i.e. susceptible group (S), vaccinated group (V), exposed group (E), infected group (I), quarantine group (Q), hospitalized group (H), and recovered group (R). The basis is from the SEIQR model and importance is given to vaccinated people, quarantine people, and hospitalized people as shown in Figure 1.

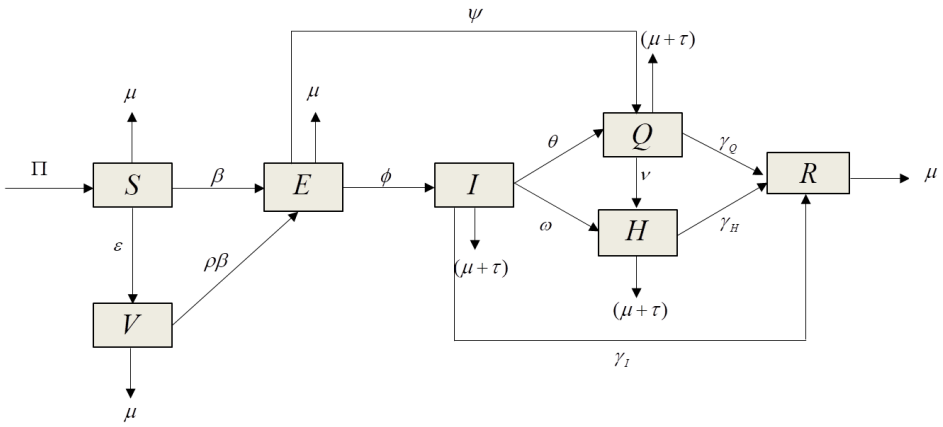


Figure 1. Diagram showing the relationship of the 5th wave of COVID-19 spread

Table 1. Definitions of variables and parameters

Variables/Parameters	Description	Units
S	The number of susceptible group	Person
V	The number of vaccinated group	Person
E	The number of exposed group	Person
I	The number of infected group	Person
Q	The number of quarantine group	Person
H	The number of hospitalized group	Person
R	The number of recovered group	Person
Π	Initial population.	day^{-1}
β	Infection rate.	Per person · $days^{-1}$
ε	Vaccination rate.	day^{-1}
ρ	Vaccination prevention efficacy.	N/A
ϕ	Incubation rate.	$days^{-1}$
ψ	Transition rate from incubation group to quarantine group.	$days^{-1}$

θ	Transition rate from infected group to quarantine group.	days^{-1}
ω	Transition rate from infected group to hospitalized group.	days^{-1}
ν	Transition rate from quarantine group to hospitalized group.	days^{-1}
γ_I	Recovery rate from infected group.	days^{-1}
γ_Q	Recovery rate from quarantine group.	days^{-1}
γ_H	Recovery rate from hospitalized group.	days^{-1}
μ	Natural mortality rate.	days^{-1}
τ	Mortality rate from COVID-19.	days^{-1}

From Figure 1, the relationship of the spread can be described as follow: The initial population Π is at risk of getting infected with COVID-19 from the group of susceptible population and the group of vaccinated population at a rate of β and $\rho\beta$ respectively. When the population is infected, the virus incubates in the body at a rate of ϕ. Some people get vaccinated to prevent the spread of disease at a rate of ε. When getting vaccinated, some individuals improve their immunity while some people have low immunity and they can get infected. Once they get infected with COVID-19, they are required to stay in quarantine at a rate of θ as their symptoms are not much severe. However, during staying in quarantine, they express severe symptoms, they need to be transferred to a hospital at a rate of ω . During the infection period, infected group, quarantine group, and hospitalized group, patients may die from the disease at a rate of τ. After they completely undergo treatments, they enter into recovered group at a rate of γ_I, γ_Q and γ_H respectively. The differential equation can be written in the following form.

$$\begin{cases} S'(t) = \Pi - \beta S(t)I(t) - (\varepsilon + \mu)S(t), \\ V'(t) = \varepsilon S(t) - \rho\beta V(t)I(t) - \mu V(t), \\ E'(t) = \beta S(t)I(t) + \rho\beta V(t)I(t) - (\phi + \psi + \mu)E(t), \\ I'(t) = \phi E(t) - (\theta + \omega + \gamma_I + \mu + \tau)I(t), \\ Q'(t) = \theta I(t) + \psi E(t) - (\nu + \gamma_Q + \mu + \tau)Q(t), \\ H'(t) = \omega I(t) + \nu Q(t) - (\gamma_H + \mu + \tau)H(t), \\ R'(t) = \gamma_I I(t) + \gamma_Q Q(t) + \gamma_H H(t) - \mu R(t). \end{cases} \tag{2.1}$$

Where $$N(t) = S(t) + V(t) + E(t) + I(t) + Q(t) + H(t) + R(t). \tag{2.2}$$

With initial conditions as follow:

$$S(0) > 0, V(0) > 0, E(0) > 0, I(0) > 0, Q(0) > 0, H(0) > 0, R(0) > 0. \tag{2.3}$$

Since the initial conditions are all positive (2.3), all time t > 0, the biologically feasible region will be considered:

$$\Omega = \left\{ (S, V, E, I, Q, H, R) \in \mathbb{R}_+^7 : N \le \frac{\Pi}{\mu} \right\}. \tag{2.4}$$

2.2 Stability Analysis

In this subpart, standard dynamical modeling is performed to analyze an equilibrium point, basic reproduction number and stability of the model, which can be seen as follows.

2.2.1 Equilibrium Point and Basic Reproduction Number

To find the equilibrium point of the system, simply set all the ordinary differential equations in the system (2.1) equal to zero as follows: $S'(t)=0, V'(t)=0, E'(t)=0, I'(t)=0, Q'(t)=0, H'(t)=0, R'(t)=0$. Two equilibrium points are obtained, i.e. disease-free equilibrium point

$$K_0^* = (S_0^*, V_0^*, E_0^*, I_0^*, Q_0^*, H_0^*, R_0^*) = \left(\frac{\Pi}{\varepsilon+\mu}, \frac{\varepsilon\Pi}{\mu(\varepsilon+\mu)}, 0, 0, 0, 0, 0 \right) \tag{2.5}$$

where $R_0 < 1$

and the endemic equilibrium point $K_1^* = (S_1^*, V_1^*, E_1^*, I_1^*, Q_1^*, H_1^*, R_1^*)$, (2.6)

where $S_1^* = \frac{\Pi}{\beta I_1^* + \varepsilon + \mu}$, $V_1^* = \frac{\varepsilon\Pi}{(\beta I_1^* + \varepsilon + \mu)(\rho\beta I_1^* + \mu)}$, $E_1^* = \frac{\Pi\beta I_1^*(\mu + (\beta I_1^* + \varepsilon)\rho)}{(\beta I_1^* + \varepsilon + \mu)(\rho\beta I_1^* + \mu)(\phi + \psi + \mu)}$,

$I_1^* = \frac{\phi E_1^*}{\theta + \omega + \gamma_I + \mu + \tau}$, $Q_1^* = \frac{\theta I_1^* + \psi E_1^*}{\nu + \gamma_Q + \mu + \tau}$, $H_1^* = \frac{\nu Q_1^* + \omega I_1^*}{\gamma_H + \mu + \tau}$, $R_1^* = \frac{\gamma_I I_1^* + \gamma_Q Q_1^* + \gamma_H H_1^*}{\mu}$,

where $R_0 > 1$.

Where R_0 is basic reproduction number. Basic reproduction number is calculated by using next-generation method. In this study $E(t), I(t), Q(t)$ and $H(t)$ expressions are taken into consideration for calculating basic reproduction number from next-generation method [16-17]. The non-linear differential equation is arranged in the following form: $\frac{dx}{dt} = F(x) - V(x)$, where $F(x)$ is the matrix of new infection and $V(x)$ is the matrix of transfer as follows:

$$F = \begin{bmatrix} \beta SI + \rho\beta VI \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad V = \begin{bmatrix} (\phi + \psi + \mu)E \\ -\phi E + (\theta + \omega + \gamma_I + \mu + \tau)I \\ -\theta I - \psi E + (\nu + \gamma_Q + \mu + \tau)Q \\ -\omega I - \nu Q + (\gamma_H + \mu + \tau)H \end{bmatrix}.$$

Thus, the Jacobian matrix can be obtained at the disease-free equilibrium point (2.5) as follow:

$$F = \begin{bmatrix} 0 & \beta S + \rho\beta V & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, V = \begin{bmatrix} (\phi + \psi + \mu) & 0 & 0 & 0 \\ -\phi & (\theta + \omega + \gamma_I + \mu + \tau) & 0 & 0 \\ -\psi & -\theta & (\nu + \gamma_Q + \mu + \tau) & 0 \\ 0 & -\omega & -\nu & (\gamma_H + \mu + \tau) \end{bmatrix}.$$

The basic reproduction number (R_0) can be calculated from the spectral radius of $\rho(FV^{-1})$ by considering eigenvalues which can be obtained from the following:

$$R_0 = \rho(FV^{-1}) = \frac{\Pi\beta\phi(\mu + \varepsilon\rho)}{\mu(\varepsilon + \mu)(\phi + \psi + \mu)(\theta + \omega + \gamma_I + \mu + \tau)}. \tag{2.7}$$

2.2.2 Global Stability Analysis

Theorem 2.1. If $R_0 < 1$ and

$$\beta = \frac{\mu + \tau}{(S_0^* + \rho V)}, \tag{2.8}$$

then the disease-free equilibrium point K_0^* is globally asymptotically stable in its feasible region.

Proof. To reveal the result, Lyapunov function is considered as follows:

$$X(t) = (S - S_0^* - S_0^* \ln\frac{S}{S_0^*}) + E + I + Q + H + R.$$

The derivative of $X(t)$ will be:

$$\begin{aligned}
X'(t) &= S'\left(1-\frac{S_0^*}{S}\right) + E' + I' + Q' + H' + R' \\
&= (\Pi - \beta SI - (\varepsilon+\mu)S)\left(1-\frac{S_0^*}{S}\right) + (\beta SI + \rho\beta VI - (\phi+\psi+\mu)E) + (\phi E - (\theta+\omega+\gamma_I+\mu+\tau)I) \\
&\quad + (\theta I + \psi E - (\nu+\gamma_Q+\mu+\tau)Q) + (\omega I + \nu Q - (\gamma_H+\mu+\tau)H) + (\gamma_I I + \gamma_Q Q + \gamma_H H - \mu R) \\
&= \Pi\left(1-\frac{S_0^*}{S}\right) - (\varepsilon+\mu)S\left(1-\frac{S_0^*}{S}\right) + \beta S_0^* I + \rho\beta VI - \mu E - (\mu+\tau)I - (\mu+\tau)Q - (\mu+\tau)H - \mu R \\
&= \Pi\left(1-\frac{S_0^*}{S}\right) + (\varepsilon+\mu)S_0^*\left(1-\frac{S}{S_0^*}\right) + \left(\beta(S_0^* + \rho V) - (\mu+\tau)\right)I - \mu E - (\mu+\tau)Q - (\mu+\tau)H - \mu R.
\end{aligned}$$

From the hypothesis (2.8), the following equation is obtained.

$$X'(t) = \Pi\left(1-\frac{S_0^*}{S}\right) + (\varepsilon+\mu)S_0^*\left(1-\frac{S}{S_0^*}\right) - \mu E - (\mu+\tau)Q - (\mu+\tau)H - \mu R.$$

Replace the equilibrium point $K_0^* = (S_0^*, V_0^*, E_0^*, I_0^*, Q_0^*, H_0^*, R_0^*) = \left(\frac{\Pi}{\varepsilon+\mu}, \frac{\varepsilon\Pi}{\mu(\varepsilon+\mu)}, 0, 0, 0, 0, 0\right)$, shall be obtained.

$$\begin{aligned}
X'(t) &= \Pi\left(1-\frac{S_0^*}{S}\right) + \Pi\left(1-\frac{S}{S_0^*}\right) - \mu E - (\mu+\tau)Q - (\mu+\tau)H - \mu R \\
&= \Pi\left(2-\frac{S_0^*}{S}-\frac{S}{S_0^*}\right) - \mu E - (\mu+\tau)Q - (\mu+\tau)H - \mu R \\
X'(t) &= -\left[\Pi\left(\frac{(S_0^* - S)^2}{S_0^* S}\right) + \mu E + (\mu+\tau)Q + (\mu+\tau)H + \mu R\right] \le 0
\end{aligned} \tag{2.9}$$

Since all parameters have positive value, $X'(t) \le 0$. $X'(t) = 0$, if $S_0^* = S$, $E = 0, Q = 0, H = 0$ and $R = 0$. Therefore, it is compliant with the LaSalle's Invariance Principle. It means that the model (2.1) is globally asymptotically stable in Ω. □

Theorem 2.2. If $R_0 > 1$, then the endemic equilibrium point K_1^* is globally asymptotically stable in its feasible region.

Proof. Lyapunov function is determined as follow [14]:

$$\begin{aligned}
Y(t) &= (S - S_1^* - S_1^* \ln\frac{S}{S_1^*}) + (V - V_1^* - V_1^* \ln\frac{V}{V_1^*}) + (E - E_1^* - E_1^* \ln\frac{E}{E_1^*}) + (I - I_1^* - I_1^* \ln\frac{I}{I_1^*}) \\
&\quad + (Q - Q_1^* - Q_1^* \ln\frac{Q}{Q_1^*}) + (H - H_1^* - H_1^* \ln\frac{H}{H_1^*}) + (R - R_1^* - R_1^* \ln\frac{R}{R_1^*}).
\end{aligned}$$

The derivative of Lyapunov function is considered, is obtained.

$$Y'(t) = S'\left(1-\frac{S_1^*}{S}\right) + V'\left(1-\frac{V_1^*}{V}\right) + E'\left(1-\frac{E_1^*}{E}\right) + I'\left(1-\frac{I_1^*}{I}\right) + Q'\left(1-\frac{Q_1^*}{Q}\right) + H'\left(1-\frac{H_1^*}{H}\right) + R'\left(1-\frac{R_1^*}{R}\right).$$

The derivative from the system (2.1) is replaced,

$$Y'(t)=\{\Pi-\beta SI-(\varepsilon+\mu)S\}\left(1-\frac{S_1^*}{S}\right)+\{\varepsilon S-\rho\beta VI-\mu V\}\left(1-\frac{V_1^*}{V}\right)+\{\beta SI+\rho\beta VI-(\phi+\psi+\mu)E\}\left(1-\frac{E_1^*}{E}\right)$$
$$+\{\phi E-(\theta+\omega+\gamma_I+\mu+\tau)I\}\left(1-\frac{I_1^*}{I}\right)+\{\theta I+\psi E-(\nu+\gamma_Q+\mu+\tau)Q\}\left(1-\frac{Q_1^*}{Q}\right)$$
$$+\{\omega I+\nu Q-(\gamma_H+\mu+\tau)H\}\left(1-\frac{H_1^*}{H}\right)+\{\gamma_I I+\gamma_Q Q+\gamma_H H-\mu R\}\left(1-\frac{R_1^*}{R}\right).$$

Putting $S=S-S_1^*, V=V-V_1^*, E=E-E_1^*, I=I-I_1^*, Q=Q-Q_1^*, H=H-H_1^*$ and $R=R-R_1^*$ is obtained.

$$Y'(t)=\left\{\Pi-\beta I\left(S-S_1^*\right)-(\varepsilon+\mu)\left(S-S_1^*\right)\right\}\left(\frac{S-S_1^*}{S}\right)+\left\{\varepsilon S-\rho\beta I\left(V-V_1^*\right)-\mu\left(V-V_1^*\right)\right\}\left(\frac{V-V_1^*}{V}\right)$$
$$+\left\{\beta SI+\rho\beta VI-(\phi+\psi+\mu)\left(E-E_1^*\right)\right\}\left(\frac{E-E_1^*}{E}\right)+\left\{\phi E-(\theta+\omega+\gamma_I+\mu+\tau)\left(I-I_1^*\right)\right\}\left(\frac{I-I_1^*}{I}\right)$$
$$+\left\{\theta I+\psi E-(\nu+\gamma_Q+\mu+\tau)\left(Q-Q^*\right)\right\}\left(\frac{Q-Q_1^*}{Q}\right)+\left\{\omega I+\nu Q-(\gamma_H+\mu+\tau)\left(H-H_1^*\right)\right\}\left(\frac{H-H_1^*}{H}\right)$$
$$+\left\{\gamma_I I+\gamma_Q Q+\gamma_H H-\mu\left(R-R_1^*\right)\right\}\left(\frac{R-R_1^*}{R}\right)$$
$$=\Pi-\Pi\left(\frac{S_1^*}{S}\right)-\beta I\frac{\left(S-S_1^*\right)^2}{S}-(\varepsilon+\mu)\frac{\left(S-S_1^*\right)^2}{S}+\varepsilon S-\varepsilon S\left(\frac{V_1^*}{V}\right)-\rho\beta I\frac{\left(V-V_1^*\right)^2}{V}-\mu\frac{\left(V-V_1^*\right)^2}{V}$$
$$+\beta SI-\beta SI\left(\frac{E_1^*}{E}\right)+\rho\beta VI-\rho\beta VI\left(\frac{E_1^*}{E}\right)-(\phi+\psi+\mu)\frac{\left(E-E_1^*\right)^2}{E}+\phi E-\phi EI_1^*-(\theta+\omega+\gamma_I+\mu+\tau)\frac{\left(I-I_1^*\right)^2}{I}$$
$$+\theta I-\theta I\left(\frac{Q_1^*}{Q}\right)+\psi E-\psi E\left(\frac{Q_1^*}{Q}\right)-(\nu+\gamma_Q+\mu+\tau)\frac{\left(Q-Q_1^*\right)^2}{Q}+\omega I-\omega I\left(\frac{H_1^*}{H}\right)+\nu Q-\nu Q\left(\frac{H_1^*}{H}\right)$$
$$-(\gamma_H+\mu+\tau)\frac{\left(H-H_1^*\right)^2}{H}+\gamma_I I-\gamma_I I\left(\frac{R_1^*}{R}\right)+\gamma_Q Q-\gamma_Q Q\left(\frac{R_1^*}{R}\right)+\gamma_Q Q-\gamma_Q Q\left(\frac{R_1^*}{R}\right)-\mu\frac{\left(R-R_1^*\right)^2}{R}.$$

A new equation is arranged as

$$Y'(t)=A-B$$

where

$$A=\Pi+\varepsilon S+\beta SI+\rho\beta VI+\phi E+\theta I+\psi E+\omega I+\nu Q+\gamma_I I+\gamma_Q Q+\gamma_Q Q,$$
$$B=\Pi\left(\frac{S_1^*}{S}\right)+\beta I\frac{\left(S-S_1^*\right)^2}{S}+(\varepsilon+\mu)\frac{\left(S-S_1^*\right)^2}{S}+\varepsilon S\left(\frac{V_1^*}{V}\right)+\rho\beta I\frac{\left(V-V_1^*\right)^2}{V}+\mu\frac{\left(V-V_1^*\right)^2}{V}$$
$$+\beta SI\left(\frac{E_1^*}{E}\right)+\rho\beta VI\left(\frac{E_1^*}{E}\right)+(\phi+\psi+\mu)\frac{\left(E-E_1^*\right)^2}{E}+\phi E\left(\frac{I_1^*}{I}\right)+(\theta+\omega+\gamma_I+\mu+\tau)\frac{\left(I-I_1^*\right)^2}{I}$$
$$+\theta I\left(\frac{Q_1^*}{Q}\right)+\psi E\left(\frac{Q_1^*}{Q}\right)+(\nu+\gamma_Q+\mu+\tau)\frac{\left(Q-Q_1^*\right)^2}{Q}+\omega I\left(\frac{H_1^*}{H}\right)+\nu Q\left(\frac{H_1^*}{H}\right)+(\gamma_H+\mu+\tau)\frac{\left(H-H_1^*\right)^2}{H}$$
$$+\gamma_I I\left(\frac{R_1^*}{R}\right)+\gamma_Q Q\left(\frac{R_1^*}{R}\right)+\gamma_Q Q\left(\frac{R_1^*}{R}\right)+\mu\frac{\left(R-R_1^*\right)^2}{R}.$$

It can be seen that, $Y'(t)<0$, when $A<B$ for $R_0>1$ and $Y'(t)=0$ when $S=S_1^*, V=V_1^*, E=E_1^*$, $I=I_1^*, Q=Q_1^*$, $H=H_1^*$ and $R=R_1^*$. Since all parameters have positive values, it is compliant with LaSalle's invariance principle. The endemic equilibrium point K_1^* is global asymptotically stable in its feasible region, if $A<B$. □

3 Numerical Results

3.1 Model Fitting

In this part, the fitting of the parameters of the model (2.1) is performed. As some parameters are difficult to predict. We obtain parameters that are precise and suitable for the model, fmincon algorithm in MATLAB is employed to analyze parameters suitable for the actual data of the disease spread in Thailand. The parameters performed fitting are shown in Table 2. The remaining parameters are obtained from the observation of disease behavior and demographic factors connected to the disease. In this study, the data analyzed referred to the actual data of the disease spread in Thailand, in which daily infection data from 11 January 2022 to 1 May 2022 were considered (since the spread of Omicron was reported), concerning the data collection of Ministry of Public Health [18]. The black circle displays the data on daily infection in Thailand while the opaque line displays the numerical analysis of the model (2.1) with $R^2 = 0.9544$ as seen in Figure 2.

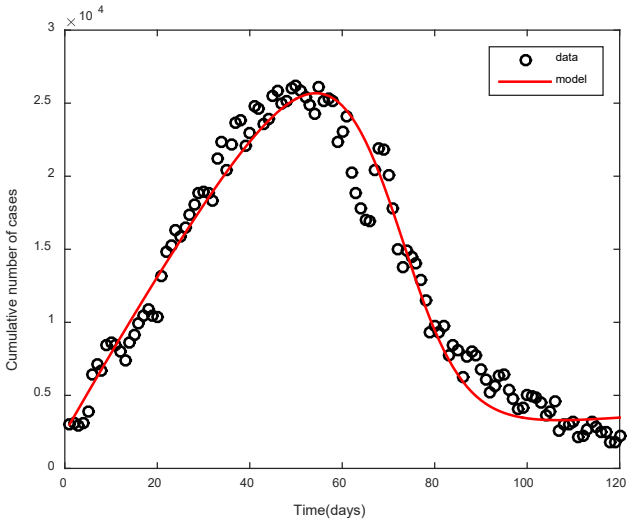


Figure 2. Fitting model with the data of daily infection in Thailand

3.2 Numerical Analysis Result

In this subpart, numerical simulation of the model (2.1) was presented by considering the stability of the endemic equilibrium point. The parameters used in this simulation are shown in Table 2. In the simulation, initial population values were determined as follow: $S(0) = 1000, V(0) = 1000, E(0) = 100, I(0) = 100, Q(0) = 1053, H(0) = 1002$ and $R(0) = 13456000$ show the stability of the endemic equilibrium point. It can be seen that the time is passed, the results were convergent to the equilibrium point at K_1^* Figure 4 – Figure 5 show numerical results in 2D and 3D trajectories. In the simulation, $\beta = 0.00000009$ was used to display the trajectory of convergence to the equilibrium more clearly. A comparison between infection rate (β) and vaccination prevention efficacy (ρ) was made and presented in Figure 6 – Figure 7. From Figure 6, when the infection rate reduced from $\beta = 0.0000009, 0.0000008, 0.0000007, 0.0000006, 0.0000005,$ it can be clearly seen that the population number increased, indicating that a high infection rate results in a faster control period than a lower infection rate. Figure 7 shows an increase in the vaccination prevention

efficacy. From $\rho = 0.4, 0.5, 0.6, 0.7, 0.8$, it can be noticeable that when the vaccine efficacy is higher, the control of disease spread is better. It is evident that vaccination strategy is a method to control the spread of COVID-19 in an efficient manner.

Table 2. Shows the parameters used in the numerical analysis

Parameters	Description	Value	Source
Π	Initial population.	560	Fitted
β	Infection rate.	0.000009	Fitted
ε	Vaccination rate.	0.4	Fitted
ρ	Vaccination prevention efficacy.	0.5	Fitted
ϕ	Incubation rate.	1/6	Fitted
ψ	Transition rate from incubation group to quarantine group.	0.08	Fitted
θ	Transition rate from infected group to quarantine group.	0.02	Fitted
ω	Transition rate from infected group to hospitalized group.	0.005	[19]
ν	Transition rate from quarantine group to hospitalized group.	0.03	Fitted
γ_I	Recovery rate from infected group.	0.001	[13]
γ_Q	Recovery rate from quarantine group.	0.03	[13]
γ_H	Recovery rate from hospitalized group.	1/14	[19]
μ	Natural mortality rate.	0.000036529	[14]
τ	Mortality rate from COVID-19.	0.00286	[15]

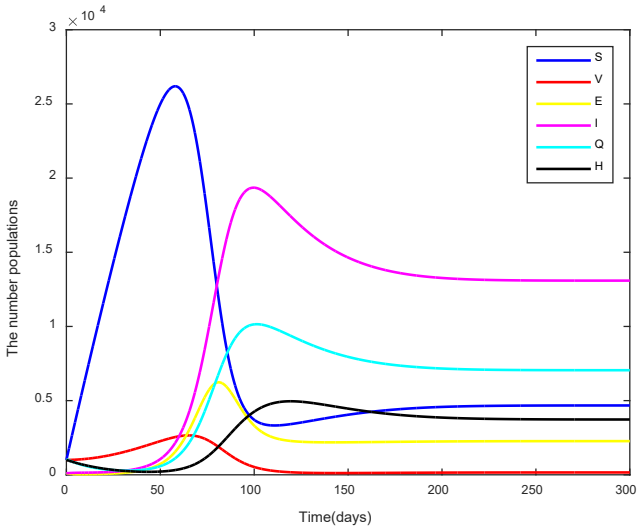


Figure 3. Graph showing the numerical results of the model (2.1) for $R_0 > 1$ and the parameters used in this simulation are shown in Table 2

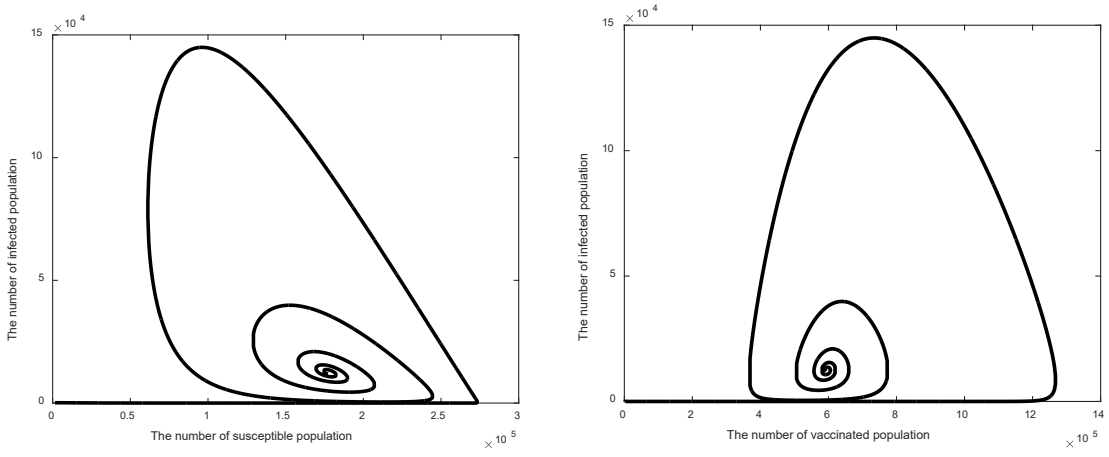


Figure 4. Graph showing 2D trajectory of the results (2.1) on the plane (S_1^*, I_1^*) and (V_1^*, I_1^*) for $R_0 > 1$ and the parameters used in this simulation are shown in Table 2

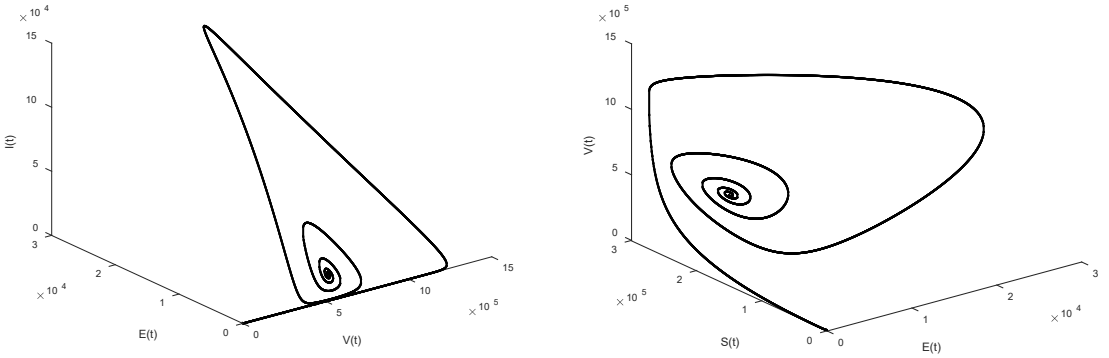


Figure 5. Graph showing 3D trajectory of the results (2.1) on the plane for $R_0 > 1$ and the parameters used in this simulation are shown in Table 2

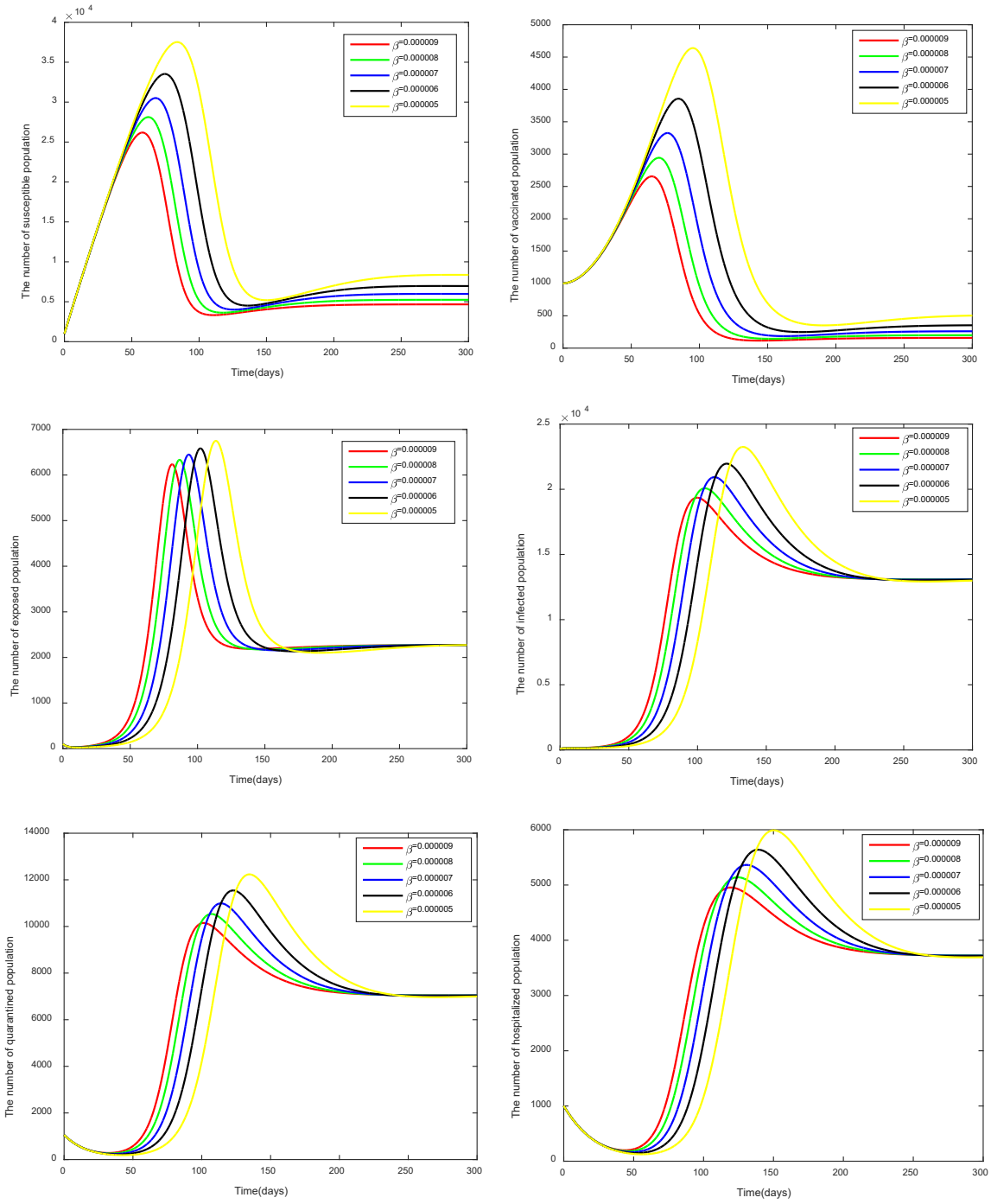


Figure 6. Graph showing the numerical results of the model (2.1) by comparing the infection rate (β) for $R_0 > 1$ and the parameters used in this simulation are shown in Table 2

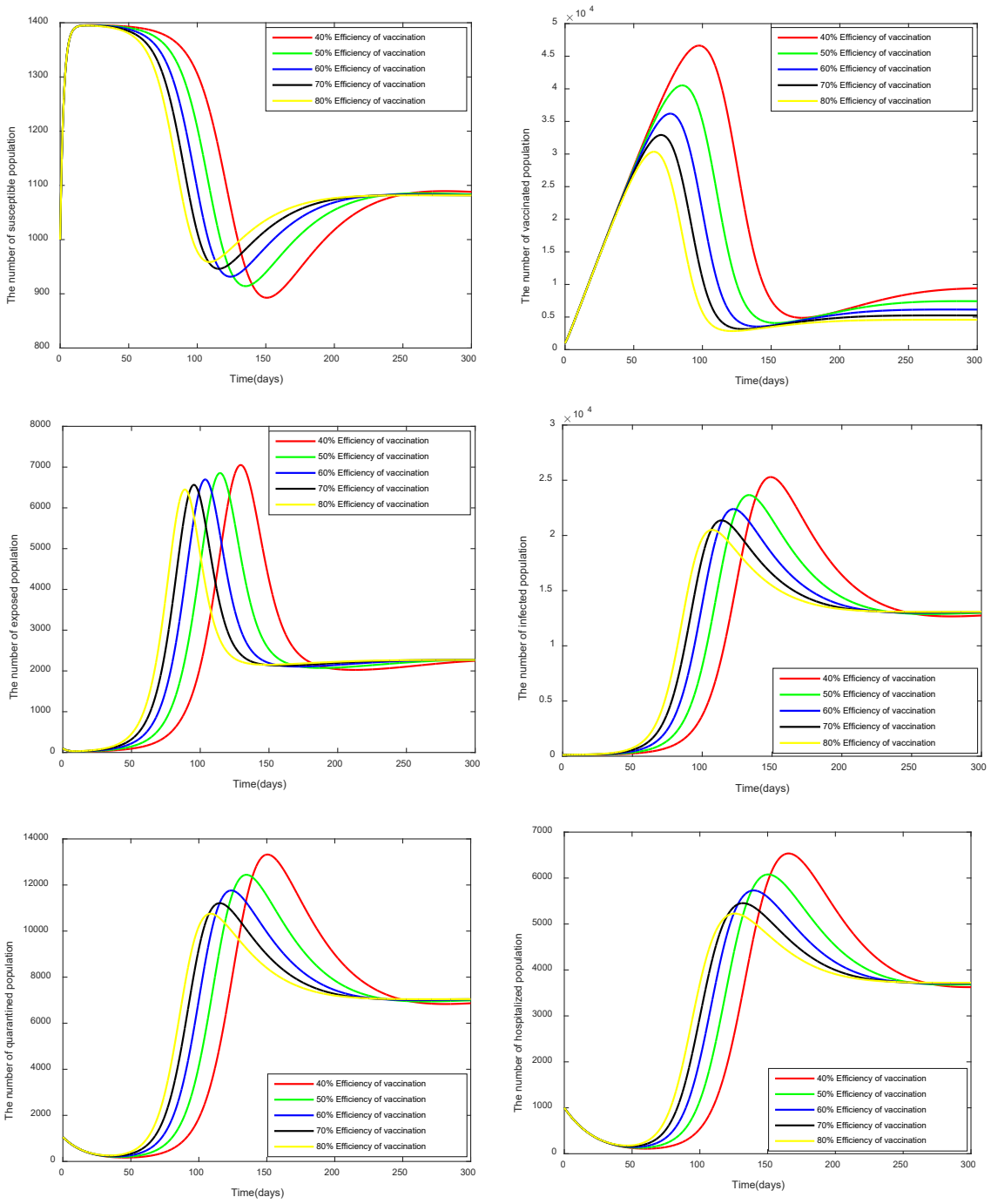


Figure 7. Graph showing the numerical results of the model (2.1) by comparing the vaccination prevention efficacy (ρ) for $R_0 > 1$ and the parameters used in this simulation are shown in Table 2

3.3 Sensitivity Analysis

In this subpart, sensitivity analysis of the basic reproduction number (R_0) was performed since the basic reproduction number indicates the status of the disease spread. Sensitivity analysis was performed to verify parameters significantly affecting the disease spread. The normalized forward sensitivity index can be calculated as follows:

$$\Upsilon_{\sigma}^{R_0} = \frac{\partial R_0}{\partial \sigma} \times \frac{\sigma}{R_0}. \tag{3.1}$$

Where σ are parameters of the disease spread and R_0 is the basic reproduction number. The parameters used in the sensitivity index are shown in Table 2 and the analysis results are shown in Table 3.

Table 3. Basic reproduction number sensitivity index

Parameters	Sensitivity
Π	1.000000
β	1.00000
ε	-0.017304
ρ	0.964758
ϕ	0.827651
ψ	-0.827274
θ	-0.692125
ω	-0.173031
γ_I	-0.034606
μ	-0.984337
τ	-0.098973

From Table 3, the sensitivity index results showed that the parameters that most likely affected the disease spread were initial population (Π) and infection rate (β), which can be described as follow: The initial population and infection rate is equal to 1, meaning that an increase or decrease of the initial population and the infection rate of 10% shall result in an increase or decrease of the basic reproduction number by 10%. Consequently, to achieve efficient disease control, it is necessary to have tight control, avoid meeting people at risk of getting infected, maintain social distancing, and wear a face mask to reduce the infection rate.

4 Conclusions

In conclusion, to describe the dynamic of COVID-19 spread, a new model for the Omicron variant in Thailand was introduced. The model gives importance to the vaccinated population, quarantine population, and hospitalized population. Equilibrium points, basic reproduction number, and model stability were analyzed. The findings from the study revealed that at the disease-free equilibrium point, the model was stable when $R_0 < 1$ and at the endemic equilibrium point, the model was stable when $R_0 > 1$. According to the numerical result analysis, we describe the dynamic of the disease spread and to ensure the results obtained are close to actual data of the spread in Thailand. Model fitting was performed to obtain parameter values suitable for the model and the disease spread in Thailand. Meanwhile, the basic reproduction number sensitivity index was analyzed. The basic reproduction number was defined in the form of R_0 and it was given by

$$R_0 = \frac{\Pi\beta\phi(\mu + \varepsilon\rho)}{\mu(\varepsilon + \mu)(\phi + \psi + \mu)(\theta + \omega + \gamma_I + \mu + \tau)}.$$

According to the basic reproductive number analysis of the parameters, the top 3 positive parameters that affected the basic reproduction number are the initial population (Π), infection rate (β) and vaccination prevention efficacy (ρ), respectively and the top 3 negative parameters that affected the basic reproduction number are natural mortality rate (μ), transition rate from incubation group to quarantine group (ψ) and transition rate from infected group to quarantine group (θ), respectively. The analysis of parameters affecting the basic reproduction number indicated that an increase in infection rate results in faster control of the disease spread and an increase in vaccination efficacy results in significant reduction of the infection. It can be noticeable that prevention of the disease by vaccination is a strategy that helps control the disease spread. However, a combination of measures can be imposed for the prevention and reduction of COVID-19 to reduce uncertainty about the spread of this disease in the future.

Acknowledgment. Jiraporn Lamwong is the recipient of the Graduate Study Fellowship of the School of Science, King Mongkut's Institute of Technology Ladkrabang, Thailand. This research was funded by the RA-TA graduate scholarship from the School of Science, King Mongkut's Institute of Technology Ladkrabang, grant number RA/TA-2565-D-001.

References

[1] J. Tian, J. Wu, Y. Bao, X. Weng, L. Shi, B. Liu, X. Yu, L. Qi and Z. Liu, *Modeling analysis of COVID-19 based on morbidity data in Anhui, China,* Math. Biosci. Eng. **17**(4) (2020), 2842–2852.

[2] T. Theparod, P. Kreabkhontho and W. Teparos, *Booster Dose Vaccination and Dynamics of COVID-19 Pandemic in the Fifth Wave: An Efficient and Simple Mathematical Model for Disease Progression,* Vaccines. **11**(3) (2023), 1-17.

[3] Z. S. Kifle and L.L Obsu, *Mathematical modeling for COVID-19 transmission dynamics: A case study in Ethiopia,* Results Phys. **34**(1) (2022), 1-13.

[4] N.I. Akinwande, T. T. Ashezua, R. I. Gweryina, S. A. Somma, F. A. Oguntolu, A. Usman, O. N. Abdurrahman, F. S. Kaduna, T.P. Adajime, F.A. Kuta, S. Abdulrahman, R. O. Olayiwola, A. I. Enagi, G. A. Bolarin and M. D. Shehu, *Mathematical model of COVID-19 transmission dynamics incorporating booster vaccine program and environmental contamination,* Heliyon. **8**(8) (2022), 1-14.

[5] P. Kumari, S. Singh and H. P. Singh, *Dynamical Analysis of COVID-19 Model Incorporating Environmental Factors,* Iran J. Sci. Technol. Trans. Sci. **46**(6) (2022), 1651–1666.

[6] World Health Organization. WHO Coronavirus (COVID-19) Dashboard [online]. Available from: https://covid19.who.int/ (December 16, 2023).

[7] S. Saha and A. K. Saha, *Modeling the Dynamics of COVID-19 in the Presence of Delta and Omicron Variants with Vaccination and Non-Pharmaceutical Interventions,* Heliyon, **9**(7) (2023), 1-26.

[8] A. Rajput, Tanvi, R. Aggarwal, A. Sharma, S. K. Sahdev, M. Kumar and Jaimala, *Fractional Order on Modeling the Transmission of Devastative COVID-19Infection:*

Efficacy of Vaccination, Applications and Applied Mathematics: An International Journal (AAM). **18**(1) (2023), 1-24.

[9] Department of Disease Control. Vaccine covid-19 of Thailand. [online]. Available from: https://ddc.moph.go.th/vaccine-covid19/ (December 3, 2023).

[10] Department of Disease Control. Vaccine covid-19 of Thailand. [online]. Available from: https://ddc.moph.go.th/vaccine-covid19/guidelines (December 3, 2023).

[11] M. Yavuz, F.Ö. Coşar, F. Günay and F.N. Özdemir, *A New Mathematical Modeling of the COVID-19 Pandemic Including the Vaccination Campaign*, Open Journal of Modelling and Simulation. **9**(3) (2021), 299-321.

[12] S. H. A. Khoshnaw, R. H. Salih and S. Sulaimany, *Mathematical Modelling for Coronavirus Disease (COVID-19) in Predicting Future Behaviours and Sensitivity Analysis,* Math. Model. Nat. Phenom. **15**(33) (2020), 1-13.

[13] W.Yang, *Modeling COVID-19 Pandemic with Hierarchical Quarantine and Time Delay*, Dyn. Games. Appl. **11**(4) (2021), 892–914.

[14] A. Ibrahim, U. W. Humphries, P. S. Ngiamsunthorn, I. Baba, S. Qureshi and A. Khan, *Modeling the dynamics of COVID-19 with real data from Thailand,* Sci Rep. **13**(1) (2023), 1-26.

[15] J. Lamwong, P. Pongsumpun, I.-M. Tang and N. Wongvanich, *Vaccination's Role in Combating the Omicron Variant Outbreak in Thailand: An Optimal Control Approach,* Mathematics. **10**(20) (2022), 1-29.

[16] Y.Kim, S. Lee, C. Chu, S. Choe, S. Hong and Y. Shin, *The Characteristics of Middle Eastern Respiratory Syndrome Coronavirus Transmission Dynamics in South Korea*, Osong Public Health Res Perspect. **7**(1) (2016), 49-55.

[17] K. Sarkar, S. Khajanchi and J. J. Nieto, *Modeling and forecasting the COVID-19 pandemic in India*,Chaos Solitons Fractals. **139** (2020), 1-16.

[18] World Health Organization. COVID-19 - WHO Thailand Situation Reports. [online] Available:https://www.who.int/thailand/emergencies/novel-coronavirus-2019/situation-reports, (December 15, 2023).

[19] I.U. Haq, N. Ullah, N. Ali and K.S. Nisar. *A New Mathematical Model of COVID-19 with Quarantine and Vaccination,* Mathematics. **42**(11) (2023), 1-21.

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Modified NEH Algorithms for Flowshop Scheduling Problem

Rungrot Pholyiam[1,†], Pannarat Guayjarernpanishk[1] and Tawun Remsungnen[1,‡]

[1]Department of Technology and Engineering, Faculty of Interdisciplinary Sciences
Khon Kaen University, Nongkhai Campus, Nongkhai 43000, Thailand

Abstract

The Flowshop Scheduling Problem (FSP) is a powerful optimization technique used to maximize resource utilization and operational efficiency across various industries and applications. This includes production planning in manufacturing, logistics scheduling, service appointment optimization, and even task scheduling in computer science and software engineering. The NEH (Nawaz, Enscore, and Ham) algorithm is a well-established construction heuristic method for FSP. However, its effectiveness depends on the initial job sequence selection. This research proposes an enhanced NEH algorithm that leverages a combination of diverse data shapes and a robust tie-breaking rule to improve decision-making capabilities. Numerical experiments conducted with standard benchmarks demonstrate that the proposed approach, NEHDL, reduces the relative percentage deviation (RPD) compared to the classic NEH algorithm, emerging as the preferred method for minimizing completion time. Additionally, NEHDL offers simplicity compared to E-NEH, NEH3TF, and NEH4TF methods, making it straightforward to apply.

Keywords: flowshop scheduling problem, construction heuristic method,

NEH algorithm, resource utilization, optimization.

2020 MSC: Primary 90B35; Secondary 90C35, 90C59, 68M20.

1 Introduction

The flowshop scheduling problem (FSP) has garnered significant research attention due to its applicability across diverse sectors [1]. This optimization technique finds utility in production lines (e.g., automotive, electronics), logistics (e.g., delivery routes, flight scheduling), healthcare surgery scheduling, service appointment management, and even task scheduling within computer science and software engineering (e.g., parallel processing). In essence, FSP serves as a valuable tool for optimizing resource utilization and enhancing operational efficiency across a wide spectrum of industries and applications.

† Speaker. ‡ Corresponding author.
E-mail address: rungrot_p@kkumail.com (R. Pholyiam), panngu@kku.ac.th (P. Guayjarernpanishk), rtawun@kku.ac.th (T. Remsungnen).

The selection of the objective function within FSP is contingent upon the specific priorities and constraints of the given instance. For example, minimizing tardiness might be prioritized if timely delivery is paramount. Conversely, minimizing completion time (makespan) or idle time might be more relevant if cost minimization is a primary objective. The core elements of FSP are:

(i) Fixed Machines and Jobs: This entails a predefined set of M machines arranged in a specific order, along with N jobs slated for processing on those machines in a particular sequence.

(ii) Single Processing: Under this condition, each job is restricted to being processed on only one machine at any given time, while each machine can handle only one job simultaneously.

(iii) Once an operation begins, the sequence of jobs cannot be interrupted, and the processing times of each job are unaffected by the order in which the N jobs are arranged. The solution to the Flowshop Scheduling Problem (FSP) involves determining a sequence of N jobs that minimizes operation times or other relevant objective functions.

The makespan (Cmax) represents the total time elapsed between when the first job begins processing on the first machine and when the last job finishes on the last machine. It essentially reflects the total completion time for all jobs in the schedule.

While an analytical solution exists for the FSP with just two machines [2, 3], the problem becomes computationally intractable for more machines. It is proven to be NP-complete [4], signifying that finding the optimal solution becomes exponentially more time-consuming as the problem size increases. Given the immense number of possible job sequences ($N!$), various algorithms have been developed to efficiently identify "good" (near-optimal) solutions in a reasonable amount of time.

Nawaz-Enscore-Ham [5] proposed NEH algorithmm, is known for its effectiveness in FSP. However, its performance is highly dependent on the initial sequence selection, which can be challenging to optimize. Subsequently, multiple enhancements to the NEH algorithms have been suggested [6-8]. Framinan et al. [9] tackled this issue by examining the effects of different initial sequencing rules on the NEH heuristic while aiming to minimize three objectives: makespan, idle time, and flow time. Through their comprehensive investigation, they identified specific initial sequencing rules that consistently surpassed the standard NEH approach across all three objectives. This underscores the continuous endeavors to refine and optimize NEH methodologies [10-12]. Through systematic experimentation and analysis, this research seeks to provide valuable insights into optimizing the NEH algorithm for solving the FSP, by identifying effective initial sequencing rules. The study aims to contribute to advancing heuristic methods in optimization and addressing practical challenges in industrial scheduling.

2 Preliminaries

Dong et al. [8] proposed an improved method for generating the initial sequence. They sorted jobs based on the sum of the average and standard deviation of their processing times. This approach aims to consider both central tendency and processing time variability when selecting initial jobs. Moonsan and Remsungnen [13] explored a broader range of data characteristics for job selection. They defined several sorting criteria using combinations of factors like average processing time, standard deviation, skewness, and kurtosis. Additionally, they incorporated the sum of total flowtime and total operation time as a secondary criterion during the partial sequence selection step. Ito et al. [14] introduced a new method called E-

NEH for flowshop scheduling. This method dynamically adapts to different problem scenarios by employing various priority rules at each step of the algorithm. While E-NEH outperforms other NEH-based methods in terms of solution quality, it requires more computational time. Our proposed approach will also be evaluated alongside these existing works for comparison.

2.1 Completion Time (C_{max})

Let $P = j_1 j_2 ... j_N$ be a member of possible solutions in $N!$ permutation space, and let p(i,j) be a given processing time of job j on machine i, then a completion time of job j on machine i, $C_{i,j}$ denoted as,

$$C_{i,j} = \max(C_{i,j-1}, C_{i-1,j}) + p(i,j), \tag{2.1}$$

where $i = 1, 2, 3, ..., M$; $j = 1, 2, 3, ..., N$; $C_{0,j} = 0$ and $C_{i,0} = 0$ the $C_{max}(P) = C_{M,N}$.

Let we have a task of 5 machines (M) and 5 jobs (N), their processing times and data shapes which obtained from equation (2.2)-(2.5) are shown in Table 1.

Table 1. The processing time of job j on machine i, p(i,j) with data shapes of each job, i.e., total (T_j), average (AVGj), standard deviation (STD_j), skewness ($SKEW_j$) and kurtosis ($KURT_j$). Their corresponding C(i,j) tables for $J_3J_1J_2J_4J_5$ and $J_2J_1J_4J_5J_3$ orders which result C_{max} values of 62 and 65, respectively

p(i,j)	J_1	J_2	J_3	J_4	J_5
M1	6	8	7	3	8
M2	8	4	9	7	6
M3	7	6	4	5	7
M4	6	3	7	2	7
M5	5	4	8	4	6
T_j	32	25	35	21	34
AVG_j	6.4	5.0	7.0	4.2	6.8
STD_j	1.02	1.79	1.67	1.72	0.75
$SKEW_j$	0.405	0.94	-1.15	0.59	0.51
$KURT_j$	-1.04	-1.05	-0.5	-1.01	-1.15

$C_{i,j}$	J3	J1	J2	J4	J5
M1	7	7+6=13	13+8=21	21+3=24	24+8=32
M2	7+9=16	16+8=24	24+4=28	28+7=35	35+6=41
M3	16+4=20	24+7=31	31+6=37	37+5=42	42+7=49
M4	20+7=27	31+6=37	37+3=40	42+2=44	49+7=56
M5	27+8=35	37+5=42	42+4=46	46+4=48	56+6=62

$C_{i,j}$	J2	J1	J4	J5	J3
M1	8	8+6 =14	14+3 = 17	17+8 = 25	25+7 = 32
M2	8+4 = 12	14+8 = 22	22+7 = 29	29+6 = 35	35+9 = 46
M3	12+6 = 18	22+7=29	29+5 = 34	35+7 = 41	46+4 = 50
M4	18+3 = 21	29+6 = 35	35+2 = 37	41+7 = 48	50+7 = 57
M5	21+4 = 25	35+5 = 40	40+4 = 44	48+6 = 54	57+8 = 65

In Table 1 illustrates the processing times, p(i,j), for each job j on each machine i, along with their corresponding completion times, $C_{i,j}$ for two job scheduling sequences: $J_3J_1J_2J_4J_5$ and $J_2J_1J_4J_5J_3$. Notably, the order in which jobs are processed can significantly impact the C_{max}.

As demonstrated here, the first sequence achieves a C_{max} of 62, while the second sequence results in a C_{max} of 65. This highlights the importance of optimizing the job scheduling sequence to minimize completion time in FSP.

2.2 NEH Algorithm

Classical NEH algorithm is show step by step as follows;

Step 1: Obtain the total processing time of each job, T_j, which is the sum of its processing times across all machines. Then sort jobs in decreasing order of T_j.

Step 2: Take the first two jobs and schedule them so as to minimize the partial makespan.

Step 3: For k = 3 to N, insert the k^{th} job into the partial schedule, in the k possible position which minimizes the makespan.

Since the classical NEH algorithm relies on T_j to order jobs initially. This work proposes a more informative approach for step one, incorporating not only T_j but also the data shape of processing times to create a richer initial sequence. This data shape is captured by a combination of statistical measures, including:

- Average Processing Time (AVG_j): Represents the central tendency of processing times.

$$AVG_j = \frac{\sum_{i=1}^{M} p(i,j)}{M} \tag{2.2}$$

- Standard Deviation (STD_j): Indicates the variability of processing times around the average.

$$STD_j = \frac{\sum_{i=1}^{M} (p(i,j) - AVG_j)^2}{M} \tag{2.3}$$

- Skewness ($SKEW_j$): Measures the asymmetry in the distribution of processing times (positive or negative skew).

$$SKEW_j = \frac{M}{(M-1)(M-2)} \sum_{i=1}^{M} \left(\frac{p(i,j) - AVG_j}{STD_j} \right)^3 \tag{2.4}$$

- Kurtosis ($KURT_j$): Captures the "peakedness" of the processing time distribution compared to a normal distribution.

$$KURT_j = \frac{\sum_{i=1}^{M} (p(i,j) - AVG_j)^4}{M \cdot STD_j^4} \tag{2.5}$$

By incorporating these data shape elements alongside T_j, we aim to create a more robust and informative initial job sequence for subsequent NEH steps. Additionally, we address the potential for encountering multiple partial sequences with the same C_{max} in step two with simple tie-breaking strategies first or last. The combinations are shown in Table 2.

It is important to note that the core complexity of the NEH algorithm, which is $O(MN^3)$ remains unchanged. This complexity is primarily determined by the sequence insertion process and the calculation of C_{max} throughout the algorithm. The data shape-based approach for step one does not significantly impact this complexity.

Table 2. The combinations of data shapes and tie-breaking strategies

Name	Data shape combination	First-Last tie-breaking
NEHCF	T_j	First
NEHCL	T_j	Last
NEHDF	AVG_j + STD_j	First
NEHDL	AVG_j + STD_j	Last
NEH3TF	AVG_j + STD_j + $SKEW_j$	First
NEH3TL	AVG_j + $STDj$ + $SKEW_j$	Last
NEH4TF	AVG_j + STD_j + $SKEW_j$ + $KURT_j$	First
NEH4TL	AVG_j + STD_j + $SKEW_j$ + $KURT_j$	Last

3 Main Results

To assess the effectiveness of our enhanced NEH approach, we conducted extensive numerical experiments. The well-known Taillard's benchmark and widely employed as standard for FSP, had been utilized. This benchmark provides a diverse set of test instances with varying complexities, allowing for a comprehensive evaluation.

The performance of proposed method will be compared against the classical NEH algorithm, NEHC and some recently published NEH variants. The results are presented in Table 3, displaying the performance metrics known as relative percentage deviation (RPD), as outlined in equation (3.1). This comparison will allow us to quantify the improvements achieved by incorporating the data-shape analysis and tie-breaking strategies within our enhanced NEH approach.

$$RPD_p = 100\frac{(H_p - UB_p)}{UB_p} \tag{3.1}$$

Where H_p, Solution value (e.g., C_{max}) obtained by a heuristic algorithm for problem instance p, and UB_p, Upper bound (known best possible solution value) for problem instance p provided by Taillard's benchmark, which consists of 10 instances for each of the 12 problem sizes. The average RPD metric serves as the benchmark for solution quality, with lower RPD signifying solutions closer to the optimal value and hence, superior performance.

The proposed NEH variants (NEHDL, NEH3TF, and potentially others) consistently achieve lower average RPD values compared to the classical NEH (NEHC, NEHD) and other existing methods (NEHSKE) across various problem sizes. This observation underscores the effectiveness of incorporating data shape analysis and strategic tie-breaking strategies in improving solution quality.

While E-NEH [14] demonstrates the best overall average RPD (2.75), it requires significantly more computational steps as it selects the best solution after applying at least four initial sequences from different rules. This underscores a pivotal strength of our approach: it yields highly competitive RPD values (NEHDL: 2.85, NEH3TF: 2.86) while demonstrably improving computational efficiency. Moreover, in some instances, our approach even outperforms E-NEH, showcasing its superior solution quality in addition to its efficiency gains.

Table 3. Average RPD values of different NEH variants on Taillard's Benchmark: Average RPD values for each problem size and overall average

p	Size	NEHC*	NEHCF	NEHCL	NEHD*	NEHDF	NEHDL	NEHSKE*	E-NEH*	NEH3TF	NEH3TL	NEH4TF	NEH4TL
1	20 x 5	3.30	3.07	2.77	2.70	2.91	2.56	2.71	2.16	3.12	3.04	2.63	2.60
2	20 x 10	4.60	5.02	4.55	4.08	4.10	3.90	3.68	3.68	4.09	4.02	4.07	4.75
3	20 x 20	3.73	3.66	3.60	3.82	3.69	3.24	2.91	3.06	3.11	3.42	3.11	3.47
4	50 x 5	0.73	0.76	0.88	0.89	1.04	1.03	0.88	0.64	1.07	1.00	0.86	0.84
5	50 x 10	5.07	4.53	4.72	4.90	4.61	4.31	4.48	4.25	4.23	5.20	4.52	4.29
6	50 x 20	6.65	6.05	5.43	6.12	6.05	5.50	6.42	6.15	5.55	5.72	5.89	6.04
7	100 x 5	0.53	0.52	0.40	0.41	0.42	0.51	0.54	0.36	0.33	0.50	0.47	0.34
8	100 x 10	2.21	2.20	2.36	2.16	2.07	2.40	2.24	1.72	2.17	2.08	2.10	2.34
9	100 x 20	5.34	4.43	4.46	5.65	4.43	4.42	4.99	4.81	4.36	4.17	4.61	4.84
10	200 x 10	1.26	1.24	1.10	1.24	1.16	1.29	1.24	0.89	1.25	1.26	1.47	1.21
11	200 x 20	4.41	3.31	3.41	4.57	3.22	3.33	4.14	3.65	3.29	3.12	3.35	3.20
12	500 x20	2.07	1.76	1.90	2.13	1.73	1.68	2.12	1.62	1.78	1.71	1.70	1.65
	Overall AVG	3.32	3.05	2.97	3.22	2.95	2.85	3.06	2.75	2.86	2.94	2.90	2.96

* Take values from [14].

Since the NEH method relies on arranging jobs from largest to smallest from the first position to the last. By utilizing the NEHDL rule, which prioritizes placing the current job in the last feasible slot while ensuring priority positions for subsequent smaller jobs, one can effectively reorder and optimize job sequences.

However, while NEH variants like NEHCF and NEHDF, which prioritize the initial job encounter order during tie-breaking, yield lower RPD compared to NEHCL and NEHDL, but NEH3TF and NEH4TF result in higher RPD compared to NEH3TL and NEH4TL. These findings suggest that a tie-breaking rule based on encounter order may not consistently enhance solution quality.

Moreover, the impact of different methods on RPD seems to vary depending on the problem size (number of jobs and machines). For instance, the proposed NEHDL performs well across most sizes, while some methods like NEHSKE show more size-dependent performance. This warrants further investigation into the influence of problem characteristics on optimal NEH variant selection.

In conclusion, the proposed NEH enhancements, particularly the NEHDL variant, offer a significant contribution to the domain of flowshop scheduling algorithms. NEHDL stands out for its simplicity and effectiveness in minimizing completion time while maintaining competitive solution quality (low RPD) and improved computational efficiency compared to E-NEH and other variants. However, further research may be needed to fully validate its effectiveness across various scenarios. Future work could explore problem size dependence and conduct statistical comparisons between the proposed methods and existing ones for a more robust evaluation.

Acknowledgment. The authors are grateful to the referees for their careful reading of the manuscript and their useful comments.

References

[1] J. M. Framinan, J. N. Gupta, and R. Leisten, *A review and classification of heuristics for permutation flow-shop scheduling with makespan objective*, Journal of the Operational Research Society, 55, (2004), 1243–1255.

[2] S. M. Johnson, *Optimal two-and three-stage production schedules with setup times included*. Naval research logistics quarterly, 1(1), (1954), 61-68.

[3] M. R. Garey, D. S. Johnson, and R. Sethi, *The complexity of flowshop and jobshop scheduling*. Mathematics of operations research, 1(2), (1976), 117–129.

[4] J. N. Gupta, and E. F. Stafford Jr, *Flowshop scheduling research after five decades*. European Journal of Operational Research, 169(3), (2006), 699–711.

[5] M. Nawaz, E. E. Enscore Jr, and I. Ham, *A heuristic algorithm for the m-machine, n-job flow-shop sequencing problem.* Omega, 11(1), (1983), 91–95.

[6] C. Koulamas, *A new constructive heuristic for the flowshop scheduling problem.* European Journal of Operational Research, 105(1), (1998), 66–71.

[7] P. J. Kalczynski, and J. Kamburowski, *On the NEH heuristic for minimizing the makespan in permutation flow shops*. Omega, 35(1), (2007), 53–60.

[8] X. Dong, H. Huang, and P. Chen, *An improved NEH-based heuristic for the permutation flowshop problem*. Computers & Operations Research, 35(12), (2008), 3962–3968.

[9] J. M. Framinan, R. Leisten, and C. Rajendran, *Different initial sequences for the heuristic of Nawaz, Enscore and Ham to minimize makespan, idle time or flowtime in the static permutation flowshop sequencing problem.* International Journal of Production Research, 41(1), (2003), 121–148.

[10] C. Sauvey, and N. Sauer, *Two NEH heuristic improvements for flowshop scheduling problem with makespan criterion.* Algorithms, 13(5), (2020), 112.

[11] F. Jin, S, Song, and C. Wu, *An improved version of the NEH algorithm and its application to large-scale flow-shop scheduling problems.* IIE Transactions, 39(2), (2007), 229–234.

[12] W. Phaphan, *A New Hybrid Heuristic for Minimizing Total Flow Time in Permutation Flow Shop.* In Proceedings of the 3rd International Conference on Industrial and Business Engineering, (2017), 81–86.

[13] T. Moonsan, *Combintion of Construction Heuristics and Metaheuristic Methods for Flowshop Scheduling Problem.* Doctor of Philosophy Thesis in Applied Mathematics, Graduate School, Khon Kaen University. 2014.

[14] S. Ito, K., Kanahara, T. Oda and K. Katayama, *An Extended NEH based Method for Permutation Flowshop Scheduling Problem.* In Proceedings of the 10th International Conference on Computer and Communications Management, (2022, July), 252–256.

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Encapsulation of Endofullerene $Fe@C_{20}$ into Single-Walled Carbon Nanotube

Tana Sunpatanon[1,†] and Prangsai Tiangtrong[1,‡]
[1]Department of Mathematics, Faculty of Science
Ramkhamhaeng University, Bangkok 10240, Thailand

Abstract

Encapsulating endofullerene in a carbon nanotube results in the development of innovative nanomaterials with distinct characteristics and applications. Encapsulating an iron atom in the middle of a C_{20} fullerene, forming $Fe@C_{20}$, within a carbon nanotube provides benefits such as increased stability, higher electrical conductivity, and adjustable magnetic characteristics. Additionally, the $Fe@C_{20}$ encapsulated within the carbon nanotube shows promise for several applications including biomedical imaging, medication administration, and energy storage. This study employs a continuum approach to examine the encapsulation behavior of the van der Waals interaction between an endofullerene $Fe@C_{20}$ and a single-walled carbon nanotube. The Lennard-Jones potential is used to calculate the acceptance energy and suction energy. The results indicate that the force of interaction between endofullerene enclosed in the carbon nanotube becomes apparent at nanotube radii of 4.728 Å, 4.977 Å, and 5.250 Å. When the radius of the tube is greater than or equal to 4.728 Å, the endofullerene will be accepted into the carbon nanotube because of the non-negative acceptance energy. The endofullerene will reach its peak suction energy when moving through a nanotube with a radius of 5.250 Å. This paper demonstrates a method to determine the encapsulation procedure for endofullerene $Fe@C_{20}$ within a carbon nanotube, enabling the creation of a more intricate system for investigating its further features.

Keywords: carbon nanotube, encapsulation, Lennard-Jones potential, endofullerene.

2020 MSC: Primary 00A71.

1 Introduction

Iijima made the discovery of carbon nanotubes (CNTs) in 1991 [1]. CNTs, with their unique properties of high electrical conductivity and thermal, chemical, and mechanical stability [2], find application in various fields such as nanoelectronics [3], biosensors [4], chemical sensors [5], chemical and biological separation [6], purification, and catalysis [7]. Also, the discovery of C_{60} fullerene in 2000 [8] got a lot of attention because of its unique mechanical properties caused by the van der Waals force and its electronic properties because

† Speaker. ‡ Corresponding author.
E-mail address: tanainaot@gmail.com (T. Sunpatanon), Prangsai@ru.ac.th (P. Tiangtrong).

it has a high surface-to-volume ratio [9, 10]. There is much research on the mechanism of C_{60} fullerene inside the carbon nanotube called "nanopeapods" acting as superconducting nanowires [11]. Many scientists tried to formulate and create many possible members of fullerenes like C_{28} and C_{36} to further investigate their properties. However, in 2000, scientists successfully synthesized C_{20} fullerene, the smallest member of the fullerenes, as predicted [12]. Because it has so many unique qualities, C_{20} fullerene creates magnetization discontinuities when an external magnetic field is applied [13]. It also shows that its shape stays mostly the same, with no obvious distortion, even at temperatures as high as 1,500 Kelvin [14]. These astonishing properties have a possible application in nanotechnology, material science, or drug transportation. However, many scientists tried to create fullerene, or the closed-cage carbon molecules containing additional atoms or a molecule inside the cage, which is called "endohedral fullerenes" or "endofullerenes" [15] because of the discovery of lanthanum (La) atoms trapped inside a C_{60} fullerene in 1985 [16]. Endofullerenes, which encapsulate individual atoms of different elements, have been the subject of extensive research and experimentation, with potential applications in molecular chemistry, physics, biomedicine, electronics, optics, and nanotechnology. A lot of people are interested in endofullerenes that are surrounded by ferromagnetic material, like an iron atom [17, 18]. This is because they have special structural and physicochemical properties [19] that could be used in molecular electronics [20], magnetic resonance imaging [21], and nuclear magnetic resonance (NMR) analysis [22, 23]. Poklonski et al. used a semi-empirical approach to study the activated nanorelay caused by the bending of nanotubes by a magnetic force [24]. They looked at the properties of an (8,8) carbon nanotube with a single $Fe@C_{20}$ inside it.

The point of this study is to use the continuum approach to look into how the van der Waals forces act on single-walled carbon nanotubes and the endofullerene $Fe@C_{20}$. We also studied the acceptance and suction energies using the potential energy of the Lennard-Jones potential function. Another goal is to determine the minimum radius of the nanotube that accepts the endofullerene $Fe@C_{20}$ and the optimal radius that provides the maximum suction energy. We can organize this research in the following ways: Section 2 shows the concept of mathematical modeling of the system consisting of sub-two systems: the interaction of the C_{20} fullerene and the tube, and the interaction of an iron atom and the tube. Then we formulate the mathematical relation to calculate the interaction energies, acceptance energies, and suction energies for each subsystem. Section 3 is the numerical results of total interaction energies, acceptance energies, and suction energies from all sub-two systems. Section 4 summarizes the conclusions.

2 Mathematical Modelling

In this mathematical mode, the following assumptions are required: The model considers only the van der Waals interaction; the electrostatic effect is neglected; and the fullerene molecule is a perfect sphere [25]. Due to nonbonded interaction, the Lennard-Jones potential function is used to calculate the potential energy, which is given by

$$\Phi(\rho) = -\frac{A}{\rho^6} + \frac{B}{\rho^{12}} = 4\epsilon\left[-\left(\frac{\sigma}{\rho}\right)^6 + \left(\frac{\sigma}{\rho}\right)^{12}\right], \tag{1}$$

where $\Phi(\rho)$ is the potential function such that ρ is a distance between two atoms, A and B are the attractive and repulsive constants, respectively, ϵ is the energy well depth, and σ is the van der Waals diameter.

Based on a continuous approach [26], there is a uniform distribution of atoms over the surface of each molecule. We can calculate the molecular interatomic energy by using the integrals over the surface or the volume of each molecule, as indicated by the following equation:

$$E = \eta_1 \eta_2 \int_{S_1} \int_{S_2} \left(-\frac{A}{\rho^6} + \frac{B}{\rho^{12}} \right) dS_2 dS_1, \tag{2}$$

where η_1 and η_2 are the mean surface densities or the mean volume densities of atoms on each molecule, whereas dS_1 and dS_2 are typical surface elements on each molecule. However, the equation (2) can be reduced to

$$E = \eta_1 \eta_2 (-AI_3 + BI_6), \tag{3}$$

where the integrals I_n $(n = 3,6)$ can be defined by

$$I_n = \int_{S_1} \int_{S_2} \rho^{-2n} \, dS_2 dS_1. \tag{4}$$

In this research, we construct a mathematical model that explains the mechanism of the encapsulation between an endofullerene $\mathrm{Fe@C_{20}}$ and the carbon nanotube with any radius to calculate the interaction force, the acceptance energy, and the suction energy. In this case, the Cartesian coordinate system (x, y, z) will be used as a reference to construct the system in which the carbon nanotube is assumed to be a perfect and well-defined cylinder, and the endofullerene is spherical. An endofullerene $\mathrm{Fe@C_{20}}$ consists of a $\mathrm{C_{20}}$ fullerene in which an iron atom (Fe) is located at the center of the fullerene. Each carbon atom (C) of the fullerene is located at 0.205 nm from its center of the sphere. In the ferrocene $\mathrm{C_{10}H_{10}Fe}$, an iron atom is at the same distance from the carbon atoms [27], thus it may be placed within the $\mathrm{C_{20}}$ fullerene to symmetrically place it at the center.

The interaction energy between the endofullerene $\mathrm{Fe@C_{20}}$ and the carbon nanotube is the sum of an iron atom and a $\mathrm{C_{20}}$ fullerene's interactions with the carbon nanotube. The total interaction energy between the endofullerene and the carbon nanotube E_{tot} is given by

$$E_{tot} = E_1 + E_2, \tag{5}$$

where E_1 and E_2 are interaction energies between the carbon nanotube and the $\mathrm{C_{20}}$ fullerene and the iron atom, respectively.

The van der Waals force (F_{vdw}) means the attractive and repulsive forces between molecules, also known as the intermolecular force. The van der Waals interaction force between two typical atoms on two nonbonded molecules is given by [28],

$$F_{vdw} = -\nabla E, \tag{6}$$

where E is the interaction energy between the atoms and the operator ∇ is the vector gradient.

The gradient in Cartesian coordinates is given by

$$\nabla E(x,y,z) = \frac{\partial E}{\partial x}\hat{\mathbb{i}} + \frac{\partial E}{\partial y}\hat{\mathbb{j}} + \frac{\partial E}{\partial z}\hat{\mathbb{k}}. \tag{7}$$

The resulting axial force along the Z-axis can be derived from the differentiation of the integrated interaction energy with respect to Z, which represents the distance between the centers of two molecules. The van der Waal force can be expressed in the form of the resulting axial force as [28],

$$F_Z = -\frac{\partial E}{\partial Z}. \tag{8}$$

Acceptance energy and suction energy are the two main characteristics that were first introduced by Cox et al. [29]. These energies are useful to study the suction and acceptance behaviors of the encapsulation mechanism for some applications like drug transportation [29].

The acceptance energy (W_a) can be defined as the total work performed by the van der Waals interactions on the particle entering the nanotube, up until the point Z_0 that the van der Waals force becomes attractive [29]. The total acceptance energy along the Z-axis can determine whether the molecule will be sucked into the carbon nanotube or not, which can be expressed as

$$W_a = \int_{-\infty}^{Z_0} F_Z dZ. \tag{9}$$

Meanwhile, the suction energy (W_s) is a criterion for the total increase in the kinetic energy experienced by the inner tube [30], which can be expressed as the total work done by the van der Waals interaction force to move the molecule from $Z = -\infty$ to $Z = \infty$. It can be written as

$$W_s = \int_{-\infty}^{\infty} F_Z dZ. \tag{10}$$

The suction energy is also an indicator of the total increase in the kinetic energy experienced by the inner carbon nanotube [30].

2.1 Interaction Between C_{20} Fullerene and Carbon Nanotube

According to Figure 1, the system consists of a C_{20} fullerene with a radius of b and a carbon nanotube with a radius of a. Based on the continuum approach and Lennard-Jones potential, the carbon nanotube is assumed to be a perfect cylinder. An atom on the tube has a coordinate of $(acos\theta, asin\theta, Z)$, where $-\pi \leq \theta \leq \pi$ and $-\infty < Z < \infty$. The coordinate of a point on the fullerene is $(0,0,Z^*)$, which is far Z^* units from the open-end carbon nanotube in the direction of the Z-axis. The distance between the center of the fullerene and an atom on the nanotube can be denoted by ρ, which is given by $\rho^2 = a^2 + (Z^* - Z)^2$. The interaction force between the spherical fullerene and the cylindrical carbon nanotube is determined by using the potential energy. According to Cox et al. [29], the potential energy $E_1(\rho)$ for all atoms of the fullerene of radius b interacting with an atom on the carbon nanotube of radius a can be expressed as

$$E_1(\rho) = -C_6(\rho) + C_{12}(\rho), \tag{11}$$

where $C_n(\rho)$ is defined by

$$C_n(\rho) = F_n \eta_{C_{20}} \int_0^{2\pi} \int_0^{\pi} \frac{b^2 \sin\phi}{r^n} d\phi d\theta\,, \tag{12}$$

r denotes the distance between atoms at the points on the fullerene and the carbon nanotube, respectively,

$\eta_{C_{20}}$ means the atomic surface density of the C_{20} fullerene,

ϕ denotes for the polar angle of the fullerene, and

θ denotes for the azimuth angle of the fullerene.

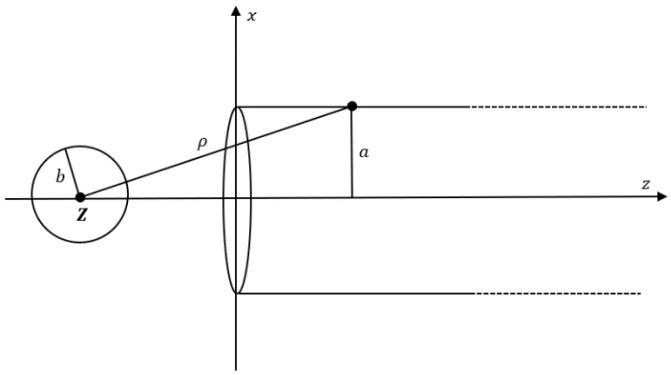


Figure 1: A system consists of a C_{20} fullerene and an open-end carbon nanotube

Since
$$r^2 = \rho^2 + b^2 - 2b\rho\cos\phi, \tag{13}$$
we have

$$C_n(\rho) = F_n \eta_{C_{20}} \int_0^{2\pi} \int_0^{\pi} \frac{b^2 \sin\phi}{(\rho^2 + b^2 - 2b\rho\cos\phi)^{\frac{n}{2}}} d\phi d\theta\,, \tag{14}$$

$$C_n(\rho) = \frac{2F_n \eta_{C_{20}} \pi b}{\rho(2-n)} \left[(\rho + b)^{(2-n)} - (\rho - b)^{(2-n)}\right]. \tag{15}$$

Thus

$$C_6(\rho) = \frac{F_6 \eta_{C_{20}} \pi b}{\rho(-2)} [(\rho + b)^{-4} - (\rho - b)^{-4}]. \tag{16}$$

and

$$C_{12}(\rho) = \frac{F_{12} \eta_{C_{20}} \pi b}{\rho(-5)} [(\rho + b)^{-10} - (\rho - b)^{-10}]. \tag{17}$$

If we let $F_6 = A$ and $F_{12} = B$, then we obtain the following potential energy as

$$E_1(\rho) = \frac{A\eta_{C_{20}} \pi b}{2\rho} [(\rho + b)^{-4} - (\rho - b)^{-4}] - \frac{B\eta_{C_{20}} \pi b}{5\rho} [(\rho + b)^{-10} - (\rho - b)^{-10}]. \tag{18}$$

The resulting axial force can be expressed as

$$F_{Z_1} = -2\pi\eta_{CNT}a\int_0^{\infty}\frac{(Z^*-Z)}{\rho}\frac{dE_1(\rho)}{d\rho}dZ = 2\pi\eta_{CNT}a\int_{\sqrt{a^2+Z^2}}^{\infty}\frac{dE_1(\rho)}{d\rho}d\rho, \tag{19}$$

$$F_{Z_1} = -2\pi^2\eta_{CNT}\eta_{C_{20}}ab\left[\frac{A}{2\rho}[(\rho+b)^{-4}-(\rho-b)^{-4}]-\frac{B}{5\rho}[(\rho+b)^{-10}-(\rho-b)^{-10}]\right]. \tag{20}$$

We can transform the following terms into simplified terms;

$$\frac{A}{2\rho}[(\rho+b)^{-4}-(\rho-b)^{-4}] = (-4A)\left[1+\frac{2b^2}{(\rho^2-b^2)^4}\right], \tag{21}$$

and

$$\frac{B}{5\rho}[(\rho+b)^{-10}-(\rho-b)^{-10}] = \frac{5}{(\rho^2-b^2)^6}+\frac{80b^2}{(\rho^2-b^2)^7}+\frac{336b^4}{(\rho^2-b^2)^8} + \frac{512b^6}{(\rho^2-b^2)^9}+\frac{256b^8}{(\rho^2-b^2)^{10}}. \tag{22}$$

Substitute (21) and (22) into (20), then we get

$$F_{Z_1} = \frac{8\pi^2\eta_{C_{20}}\eta_{CNT}a}{\tau^3b^4}\left[A_{g-C_{20}}\left(1+\frac{2}{\tau}\right)-\frac{B_{g-C_{20}}}{5\tau^3b^6}\left(5+\frac{80}{\tau}+\frac{336}{\tau^2}+\frac{512}{\tau^3}+\frac{256}{\tau^4}\right)\right], \tag{23}$$

where $$\tau = \left(a^2-b^2+Z^2\right)/b^2.$$

To calculate the Lennard-Jones potential constants A and B where $A = 4\epsilon\sigma^6$ and $B = 4\epsilon\sigma^{12}$, we need to use other parameters which are the well depth energy ϵ and the van der Waals diameter σ. The related parameters in this research are given by $\epsilon_{C_{20}} = 4.2038 \times 10^{-3}$ eV, and $\sigma_{C_{20}} = 0.337$ nm [31]. Moreover, the parameters of non-bonded C_{20}-C_{20} and C_{20}-graphene are obtained by using the Lorentz-Berthelot mixing rules $\sigma_{12} = \frac{(\sigma_1+\sigma_2)}{2}$ and $\epsilon_{12} = \sqrt{\epsilon_1\epsilon_2}$. They are used to calculate the Lennard-Jones constants in a system with different atomic species [32]. The Lennard-Jones constants for each interaction can be calculated in accordance with Table 1, as well as other related constant parameters.

Based on the assumption that the C_{20} fullerene is initially at rest, the acceptance energy is used to determine the condition in which the C_{20} fullerene will be accepted into the carbon nanotube. The mathematical expression of the acceptance energy W_{a_1}, in case of interaction between the fullerene and the tube, can be obtained as

$$W_{a_1} = \int_{-\infty}^{Z_0} F_{Z_1}dZ, \tag{24}$$

$$W_{a_1} = \frac{8\pi^2\eta_{C_{20}}\eta_{CNT}a}{\tau^3b^4}\int_{-\infty}^{Z_0}\left[A_{g-C_{20}}\left(1+\frac{2}{\tau}\right)-\frac{B_{g-C_{20}}}{5\tau^3b^6}\left(5+\frac{80}{\tau}+\frac{336}{\tau^2}+\frac{512}{\tau^3}+\frac{256}{\tau^4}\right)\right]dZ, \tag{25}$$

Table 1. Values of parameters used in the model. [31, 32]

Radius of C_{20}	$b = 2.040$ Å
$C - C$ bond length	$\sigma = 1.421$ Å
Mean Surface density for C_{20}	$\eta_{C_{20}} = 0.3824$ Å^{-2}
Mean Surface density for Carbon Nanotube (CNT)	$\eta_{CNT} = 0.3812$ Å^{-2}
Mass of a single C atom	$m_C = 19.92 \times 10^{-27}$ kg
Mass of a single fullerene C_{20}	$M_{C_{20}} = 398.4 \times 10^{-27}$ kg
Attractive constant for $C - C$ interaction	$A_{C-C} = 15.41$ eV Å^6
Repulsive constant for $C - C$ interaction	$B_{C-C} = 22534.75$ eV Å^{12}
Attractive constant for graphene - graphene	$A_{g-g} = 15.2$ eV Å^6
Repulsive constant for graphene - graphene	$B_{g-g} = 24.1 \times 10^3$ eV Å^{12}
Attractive constant for $C_{20} - C_{20}$	$A_{C_{20}-C_{20}} = 24.63$ eV Å^6
Repulsive constant for $C_{20} - C_{20}$	$B_{C_{20}-C_{20}} = 36.1 \times 10^3$ eV Å^{12}
Attractive constant for graphene – C_{20}	$A_{g-C_{20}} = 19.35$ eV Å^6
Repulsive constant for graphene – C_{20}	$B_{g-C_{20}} = 29.49 \times 10^3$ eV Å^{12}

Let $Z = \sqrt{a^2 - b^2}\tan\emptyset$ such that $dZ = \sqrt{a^2 - b^2}\sec^2\emptyset d\emptyset$. We have

$$W_{a_1} = \frac{8\pi^2 \eta_{C_{20}} \eta_{CNT} a}{\tau^3 b^4} \int_{-\frac{\pi}{2}}^{\emptyset_0} \left[A_{g-C_{20}} \left(1 + \frac{2}{\tau}\right) - \frac{B_{g-C_{20}}}{5\tau^3 b^6} \left(5 + \frac{80}{\tau} + \frac{336}{\tau^2} + \frac{512}{\tau^3} + \frac{256}{\tau^4}\right) \right] \sqrt{a^2 - b^2}\sec^2\emptyset d\emptyset, \quad (26)$$

where $\emptyset_0 = \tan^{-1}\left(\frac{Z_0}{\sqrt{a^2-b^2}}\right)$,

and Z_0 is the real roots of the equation (23).

By using the relationship

$$\tau^n = \frac{(a^2 - b^2)^n \sec^{2n}\emptyset}{b^{2n}},$$

we can rearrange the acceptance energy W_{a_1} into the following;

$$W_{a_1} = \frac{8\pi^2\eta_{C_{20}}\eta_{CNT}a}{b^4\sqrt{a^2-b^2}} \int_{-\frac{\pi}{2}}^{\emptyset_0} \left[\frac{A_{g-C_{20}}b^6}{(a^2-b^2)^2\sec^4\emptyset} + \frac{2A_{g-C_{20}}b^8}{(a^2-b^2)^3\sec^6\emptyset}\right.$$

$$-\frac{B_{g-C_{20}}}{5b^6}\left(\frac{5b^{12}}{(a^2-b^2)^5\sec^{10}\emptyset} + \frac{80b^{14}}{(a^2-b^2)^6\sec^{12}\emptyset} + \frac{336b^{16}}{(a^2-b^2)^7\sec^{14}\emptyset}\right.$$

$$\left.\left. + \frac{512b^{18}}{(a^2-b^2)^8\sec^{16}\emptyset} + \frac{256b^{20}}{(a^2-b^2)^9\sec^{18}\emptyset}\right)\right] d\emptyset, \tag{27}$$

$$W_{a_1} = \frac{8\pi^2\eta_{C_{20}}\eta_{CNT}a}{b^2\sqrt{a^2-b^2}} \int_{-\frac{\pi}{2}}^{\emptyset_0} [A_{g-C_{20}}b^4(a^2-b^2)^{-2}\cos^4\emptyset + 2A_{g-C_{20}}b^6(a^2-b^2)^{-3}\cos^6\emptyset$$

$$-\frac{B_{g-C_{20}}}{5b^6}(5b^{10}(a^2-b^2)^{-5}\cos^{10}\emptyset + 80b^{12}(a^2-b^2)^{-6}\cos^{12}\emptyset$$

$$+336b^{14}(a^2-b^2)^{-7}\cos^{14}\emptyset + 512b^{16}(a^2-b^2)^{-8}\cos^{16}\emptyset$$

$$+256b^{18}(a^2-b^2)^{-9}\cos^{18}\emptyset)]d\emptyset, \tag{28}$$

$$W_{a_1} = \frac{8\pi^2\eta_{C_{20}}\eta_{CNT}a}{b^2\sqrt{a^2-b^2}}\left[\left(A_{g-C_{20}}\int_{-\frac{\pi}{2}}^{\emptyset_0} b^4(a^2-b^2)^{-2}\cos^4\emptyset d\emptyset\right.\right.$$

$$\left. + A_{g-C_{20}}\int_{-\frac{\pi}{2}}^{\emptyset_0} b^6(a^2-b^2)^{-3}\cos^6\emptyset d\emptyset\right) - \frac{B_{g-C_{20}}}{5b^6}\left(5\int_{-\frac{\pi}{2}}^{\emptyset_0} b^{10}(a^2-b^2)^{-5}\cos^{10}\emptyset d\emptyset\right.$$

$$+80\int_{-\frac{\pi}{2}}^{\emptyset_0} b^{12}(a^2-b^2)^{-6}\cos^{12}\emptyset d\emptyset + 336\int_{-\frac{\pi}{2}}^{\emptyset_0} b^{14}(a^2-b^2)^{-7}\cos^{14}\emptyset d\emptyset$$

$$\left.\left. +512\int_{-\frac{\pi}{2}}^{\emptyset_0} b^{16}(a^2-b^2)^{-8}\cos^{16}\emptyset d\emptyset + 256\int_{-\frac{\pi}{2}}^{\emptyset_0} b^{18}(a^2-b^2)^{-9}\cos^{18}\emptyset d\emptyset\right)\right]. \tag{29}$$

Thus we can rearrange (29) into the following simplified expressions

$$W_{a_1} = \frac{8\pi^2\eta_{C_{20}}\eta_{CNT}a}{b^2\sqrt{a^2-b^2}}\left[A_{g-C_{20}}(I_2+2I_3) - \frac{B_{g-C_{20}}}{5b^6}(5I_5+80I_6+336I_7+512I_8+256I_9)\right], \tag{30}$$

where

$$I_n = \int_{-\frac{\pi}{2}}^{\emptyset_0} b^{2n}(a^2-b^2)^{-n}\cos^{2n}\emptyset d\emptyset.$$

The suction energy is defined as the total work done by the van der Waals interaction between a C_{20} fullerene molecule moving inside a carbon nanotube [29]. It can be obtained as

$$W_{s_1} = \int_{-\infty}^{\infty} F_{Z_1} dZ, \tag{31}$$

Substitute (23) and into (31), then we get

$$W_{s_1} = \frac{8\pi^2\eta_{C_{20}}\eta_{CNT}a}{\tau^3 b^4}\int_{-\infty}^{\infty}\left[A_{g-C_{20}}\left(1+\frac{2}{\tau}\right)-\frac{B_{g-C_{20}}}{5\tau^3 b^6}\left(5+\frac{80}{\tau}+\frac{336}{\tau^2}+\frac{512}{\tau^3}+\frac{256}{\tau^4}\right)\right]dZ. \tag{32}$$

Let $Z=\sqrt{a^2-b^2}\tan\emptyset$ such that $dZ=\sqrt{a^2-b^2}\sec^2\emptyset d\emptyset$. Thus we will get the following relation

$$W_{s_1} = \frac{8\pi^2\eta_{C_{20}}\eta_{CNT}a}{\tau^3 b^4}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\left[A_{g-C_{20}}\left(1+\frac{2}{\lambda}\right)\right.$$

$$\left.-\frac{B_{g-C_{20}}}{5\tau^3 b^6}\left(5+\frac{80}{\tau}+\frac{336}{\tau^2}+\frac{512}{\tau^3}+\frac{256}{\tau^4}\right)\right]\sqrt{a^2-b^2}\sec^2\emptyset d\emptyset, \tag{33}$$

$$W_{s_1} = \frac{8\pi^2\eta_{C_{20}}\eta_{CNT}a}{b^2\sqrt{a^2-b^2}}\left[\left(A_{g-C_{20}}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} b^4(a^2-b^2)^{-2}\cos^4\emptyset d\emptyset\right.\right.$$

$$\left.+2A_{g-C_{20}}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} b^6(a^2-b^2)^{-3}\cos^6\emptyset d\emptyset\right)-\frac{B_{g-C_{20}}}{5b^6}\left(5\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} b^{10}(a^2-b^2)^{-5}\cos^{10}\emptyset d\emptyset\right.$$

$$+80\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} b^{12}(a^2-b^2)^{-6}\cos^{12}\emptyset d\emptyset+336\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} b^{14}(a^2-b^2)^{-7}\cos^{14}\emptyset d\emptyset$$

$$\left.\left.+512\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} b^{16}(a^2-b^2)^{-8}\cos^{16}\emptyset d\emptyset+256\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} b^{18}(a^2-b^2)^{-9}\cos^{18}\emptyset d\emptyset\right)\right]. \tag{34}$$

Because

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^{2n}\emptyset d\emptyset = \frac{(2n-1)!!}{(2n)!!}\pi,$$

thus

$$W_{s_1} = \frac{8\pi^2\eta_{C_{20}}\eta_{CNT}a}{b^2\sqrt{a^2-b^2}}\left[A_{g-C_{20}}b^4(a^2-b^2)^{-2}\left(\frac{3\pi}{8}\right) + 2A_{g-C_{20}}b^6(a^2-b^2)^{-3}\left(\frac{15\pi}{48}\right)\right.$$
$$-\frac{5B_{g-C_{20}}}{5b^6}b^{10}(a^2-b^2)^{-5}\left(\frac{945\pi}{3840}\right) - \frac{80B_{g-C_{20}}}{5b^6}b^{12}(a^2-b^2)^{-6}\left(\frac{10395\pi}{46080}\right)$$
$$-\frac{336B_{g-C_{20}}}{5b^6}b^{14}(a^2-b^2)^{-7}\left(\frac{135135\pi}{645120}\right) - \frac{512B_{g-C_{20}}}{5b^6}b^{16}(a^2-b^2)^{-8}\left(\frac{2027025\pi}{10321920}\right)$$
$$\left.-\frac{256B_{g-C_{20}}}{5b^6}b^{18}(a^2-b^2)^{-9}\left(\frac{34459425\pi}{195794560}\right)\right]. \quad (35)$$

We can rearrange the suction energy W_{s_1} into the following;

$$W_{s_1} = \frac{\pi^3\eta_{C_{20}}\eta_{CNT}ab^2}{(a^2-b^2)^{\frac{5}{2}}}\times$$
$$\left[3A_{g-C_{20}} + 5A_{g-C_{20}}\kappa - \frac{B_{g-C_{20}}(315+4620\kappa+18018\kappa^2+25740\kappa^3+12155\kappa^4)}{160(a^2-b^2)^3}\right], \quad (36)$$

where $$\kappa = \frac{b^2}{(a^2-b^2)}.$$

2.2 Interaction between Iron Atom (Fe) and Carbon Nanotube

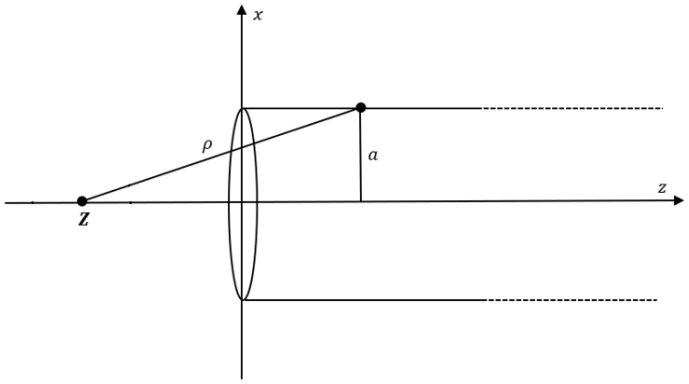


Figure 2: A system consists of an iron atom and an open-end carbon nanotube

According to Figure 2, the system consists of an iron atom (Fe) and a carbon nanotube with radius a. The Fe atom is placed along the Z-axis and its coordinate is $(0,0,Z^*)$, which is far Z^* units from a semi-infinite carbon nanotube that has a coordinate of $(acos\theta, asin\theta, Z)$,

where $-\pi \le \theta \le \pi$ and $-\infty < Z < \infty$. The distance between the iron atom and an atom on the nanotube can be denoted by ρ, which is given by $\rho^2 = a^2 + (Z^* - Z)^2$. The potential energy $E_2(\rho)$ for an iron atom (Fe) interacting with an atom on the carbon nanotube of radius a can be expressed as

$$E_2(\rho) = \eta_{CNT} a \int_0^{2\pi} \int_0^{\infty} \Phi(\rho)\, dZ d\theta, \tag{37}$$

where

$$\Phi(\rho) = -\frac{A_{\mathrm{c-Fe}}}{\rho^6} + \frac{B_{\mathrm{c-Fe}}}{\rho^{12}}.$$

The interaction force between the iron atom and the carbon nanotube is obtained by [29]

$$F_{Z_2} = -2\pi\eta_{CNT} a \int_0^{\infty} \frac{d\Phi}{d\rho}\frac{(Z^* - Z)}{\rho} dZ = 2\pi\eta_{CNT} a \int_{\sqrt{a^2+Z^2}}^{\infty} \frac{d\Phi(\rho)}{d\rho} d\rho, \tag{38}$$

Thus

$$F_{Z_2} = 2\pi\eta_{CNT} a \int_{\sqrt{a^2+Z^2}}^{\infty} \frac{d}{d\rho}\left(-\frac{A_{\mathrm{c-Fe}}}{\rho^6} + \frac{B_{\mathrm{c-Fe}}}{\rho^{12}}\right) d\rho, \tag{39}$$

$$F_{Z_2} = 2\pi\eta_{CNT} a \left[\frac{A_{\mathrm{c-Fe}}}{\left(a^2 + Z^{*2}\right)^3} - \frac{B_{\mathrm{c-Fe}}}{\left(a^2 + Z^{*2}\right)^6}\right]. \tag{40}$$

The values of van der Waals diameter σ and well-depths ϵ for the Fe atom and other related parameters are from Rappe et al. [33] to calculate the attractive and repulsive constants as appeared in Table 2.

Table 2. Values of parameters used in the model. [31 - 34]

C – C bond length	$\sigma = 1.421$ Å
Mass of a single C atom	$m_{\mathrm{C}} = 19.92 \times 10^{-27}$ kg
Mass of a single Fe atom	$m_{\mathrm{Fe}} = 19.92 \times 10^{-27}$ kg
Attractive constant for C – Fe interaction	$A_{\mathrm{C-Fe}} = 10.005$ eV Å^6
Repulsive constant for C – Fe interaction	$B_{\mathrm{C-Fe}} = 15{,}620.156$ eV Å^{12}

From Figure 2, when the Fe atom appeared as a point, located in front of the carbon nanotube, and it is assumed initially to be at rest, we can determine the acceptance energy W_{a_2} between Fe and the nanotube as

$$W_{a_2} = \int_{-\infty}^{Z_0} F_{Z_2} dZ, \tag{41}$$

$$W_{a_2} = \int_{-\infty}^{Z_0} \left[2\pi\eta_{CNT} a \left[\frac{A_{\mathrm{c-Fe}}}{\left(a^2 + Z^{*2}\right)^3} - \frac{B_{\mathrm{c-Fe}}}{\left(a^2 + Z^{*2}\right)^6} \right] \right] dZ, \tag{42}$$

$$W_{a_2} = 2\pi\eta_{CNT} a^2 \int_{-\frac{\pi}{2}}^{Z_0} \left[\frac{A_{\mathrm{c-Fe}}}{\left(a^2 + Z^{*2}\right)^3} - \frac{B_{\mathrm{c-Fe}}}{\left(a^2 + Z^{*2}\right)^6} \right] \sec^2 \emptyset d\emptyset. \tag{43}$$

Let $Z = a\tan\emptyset$ such that $dZ = a\sec^2\emptyset d\emptyset$,

$$W_{a_2} = \frac{2\pi\eta_{CNT}}{a^4} \int_{-\frac{\pi}{2}}^{\emptyset_0} \left[A_{\mathrm{c-Fe}} \cos^4\emptyset - \frac{B_{\mathrm{c-Fe}}}{a^6} \cos^{10}\emptyset \right] d\emptyset, \tag{44}$$

where

$$\emptyset_0 = \tan^{-1}\left(\frac{Z_0}{a}\right),$$

and Z_0 is the real roots of the equation (40).

By using the relationship

$$I = \int \cos^m x dx = \frac{\sin x \cos^{m-1} x dx}{m} + \frac{m-1}{m} \int \cos^{m-2} x dx,$$

we can find the following

$$\int_{-\frac{\pi}{2}}^{\emptyset_0} A_{\mathrm{c-Fe}} \cos^4\emptyset \, d\emptyset = \frac{A_{\mathrm{c-Fe}}}{8} \left[\sin\emptyset_0 (2\cos^3\emptyset_0 + 3\cos\emptyset_0) + 3\left(\emptyset_0 + \frac{\pi}{2}\right) \right], \tag{45}$$

and

$$\int_{-\frac{\pi}{2}}^{\emptyset_0} \frac{B_{\mathrm{c-Fe}}}{a^6} \cos^{10}\emptyset \, d\emptyset = \frac{B_{\mathrm{c-Fe}}}{a^6} \left[\frac{1}{5} \sin\emptyset_0 \left(\frac{1}{2}\cos^9\emptyset_0 + \frac{9}{16}\cos^7\emptyset_0 + \frac{63}{96}\cos^5\emptyset_0 + \frac{630}{768}\cos^3\emptyset_0 + \frac{945}{768}\cos\emptyset_0 \right) + \frac{945}{3840}\left(\emptyset_0 + \frac{\pi}{2}\right) \right]. \tag{46}$$

Substitute (45) and (46) into (44), we can rearrange the acceptance energy W_{a_2} as

$$\int_{-\frac{\pi}{2}}^{\emptyset_0}\left[A_{\mathrm{c-Fe}}\cos^4\emptyset-\frac{B_{\mathrm{c-Fe}}}{a^6}\cos^{10}\emptyset\right]d\emptyset=\left(\frac{32A_{\mathrm{c-Fe}}}{256}\right)\left[\sin\emptyset_0(2\cos^3\emptyset_0+3\cos\emptyset_0)+3\left(\emptyset_0+\frac{\pi}{2}\right)\right]$$

$$-\left(\frac{B_{\mathrm{c-Fe}}}{256a^6}\right)\left[\frac{1}{5}\sin\emptyset_0(128\cos^9\emptyset_0+144\cos^7\emptyset_0\right.$$

$$+168\cos^5\emptyset_0+210\cos^3\emptyset_0+315\cos\emptyset_0)$$

$$\left.+63\left(\emptyset_0+\frac{\pi}{2}\right)\right]. \tag{47}$$

Thus

$$W_{a_2}=\frac{\pi\eta_{CNT}}{128a^4}\left\{(32A_{\mathrm{c-Fe}})\left[\sin\omega_0(2\cos^3\omega_0+3\cos\omega_0)+3\left(\omega_0+\frac{\pi}{2}\right)\right]\right.$$

$$-\left(\frac{B_{\mathrm{c-Fe}}}{a^6}\right)\left[\frac{1}{5}\sin\omega_0(128\cos^9\omega_0+144\cos^7\omega_0+168\cos^5\omega_0+210\cos^3\omega_0\right.$$

$$\left.\left.+315\cos\omega_0)+63\left(\omega_0+\frac{\pi}{2}\right)\right]\right\}, \tag{48}$$

where

$$\omega_0=\tan^{-1}\left\{\left[\frac{B_{\mathrm{c-Fe}}}{(A_{\mathrm{c-Fe}}a^6)}\right]^{\frac{1}{3}}-1\right\}^{\frac{1}{2}}.$$

Consider the suction energy that is acquired by the Fe atom and the carbon nanotube. In this scenario, we can calculate the energy as

$$W_{s_2}=\int_{-\infty}^{\infty}F_{Z_2}dZ, \tag{49}$$

$$W_{s_2}=\int_{-\infty}^{\infty}\left[2\pi\eta_{CNT}a\left[\frac{A_{\mathrm{c-Fe}}}{\left(a^2+Z^{*2}\right)^3}-\frac{B_{\mathrm{c-Fe}}}{\left(a^2+Z^{*2}\right)^6}\right]\right]dZ. \tag{50}$$

Let $Z=a\tan\emptyset$ such that $dZ=a\sec^2\emptyset d\emptyset$. Then

$$W_{s_2}=2\pi\eta_{CNT}a\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}\left[\frac{A_{\mathrm{c-Fe}}}{a^6(1+\tan^2\emptyset)^3}-\frac{B_{\mathrm{c-Fe}}}{a^{12}(1+\tan^2\emptyset)^6}\right](a\sec^2\emptyset d\emptyset), \tag{51}$$

$$W_{s_2}=2\pi\eta_{CNT}a\left[\frac{1}{a^5}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}A_{c-Fe}\cos^4\emptyset d\emptyset-\frac{1}{a^{11}}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}B_{c-Fe}\cos^{10}\emptyset d\emptyset\right]. \tag{52}$$

We note that

$$\frac{1}{a^5}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} A_{\mathrm{c-Fe}}\cos^4\emptyset d\emptyset = \frac{A_{\mathrm{c-Fe}}}{a^5}\left[\frac{3\pi}{8}\right], \tag{53}$$

and

$$\frac{1}{a^{11}}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} B_{\mathrm{c-Fe}}\cos^{10}\emptyset d\emptyset = \frac{B_{\mathrm{c-Fe}}}{a^{11}}\left[\frac{945\pi}{3840}\right], \tag{54}$$

$$\frac{1}{a^5}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} A_{\mathrm{c-Fe}}\cos^4\emptyset d\emptyset - \frac{1}{a^{11}}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} B_{\mathrm{c-Fe}}\cos^{10}\emptyset d\emptyset = 32A_{\mathrm{c-Fe}} - \frac{21B_{\mathrm{c-Fe}}}{a^6}. \tag{55}$$

Substitute (53) and (54) into (52), we can rearrange the suction energy W_{s_2} as

$$W_{s_2} = \frac{3\pi^2\eta_{CNT}}{128}\left[32A_{\mathrm{c-Fe}} - \frac{21B_{\mathrm{c-Fe}}}{a^6}\right]. \tag{56}$$

3 Main Results

In this section, we sum up all sub-interactions: the interactions between the carbon nanotube and the $\mathrm{C_{20}}$ fullerene and the Fe atom, respectively, to calculate their interaction force, the acceptance energy, and the suction energy. First, we study the numerical results for the interaction forces between the endofullerene $\mathrm{Fe@C_{20}}$ and the carbon nanotube with different radii which are 4.728 Å, 4.977 Å, 5.250 Å and 5.500 Å, respectively. The graphs of the interaction forces versus the axial position, as shown in Figure 3, illustrate the results. These graphs are similar to those of Cox et al. (2007) [29]. When we consider a carbon nanotube with a radius less than 4.728 Å, the endofullerene will be rejected from the nanotube due to the negative acceptance energy. However, at the other radii of the nanotubes, which are 4.977 Å, 5.250 Å, and 5.500 Å, the nanotubes will accept the endofullerene. This result shows that the encapsulation process is strongly dependent on the size of the radius of the nanotube [28].

According to the acceptance energies, which are shown in Figure 4, the results show a relationship between the acceptance energies and the radii of the nanotubes, ranging from 4.5 Å to 5.5 Å. If the acceptance energies are positive, the endofullerene will be admitted into the carbon nanotube. Conversely, if the energies are negative, an extra energy is needed for the endofullerene to get inside the carbon nanotube. From the results, the acceptance energy is greater than zero when the radius is greater than 4.728 Å. If the radius is below this critical value, the carbon nanotube will not accept the endofullerene, and an additional energy is required. We also note that the nanotube with a radius greater than 5.250 Å will accept the endofullerene.

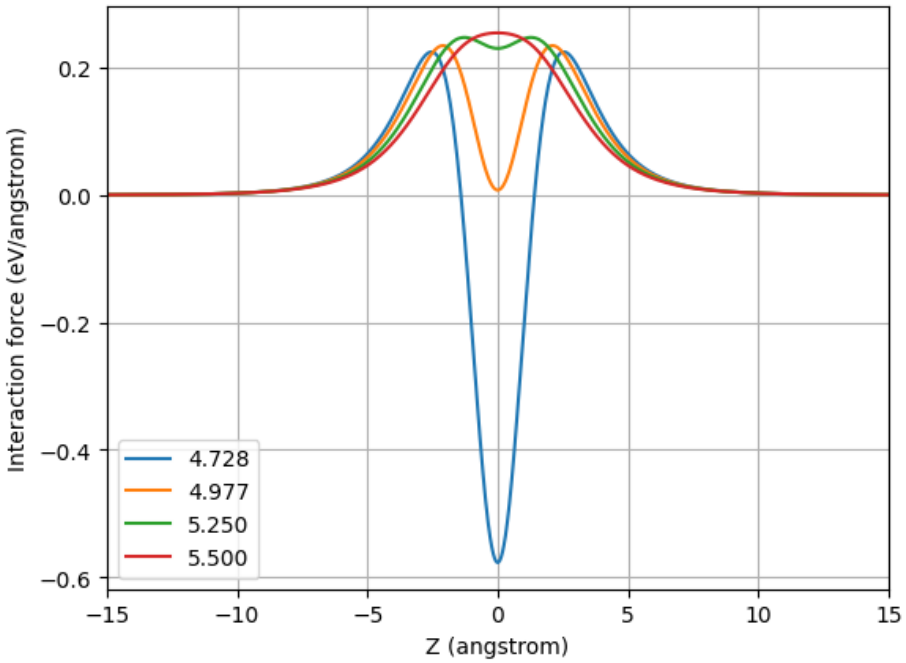


Figure 3: Interaction forces between an endofullerene $Fe@C_{20}$ and the carbon nanotubes with radii 4.728, 4.977, 5.250 and 5.500 Å

Figure 5 displays the suction energies for the endofullerene entering the carbon nanotubes at various radii. The radii of the nanotubes range from 4 Å to 10 Å. Based on the assumption that the endofullerene is initially at rest, the interaction energies are positive when the radius of the nanotube is greater than or equal to 4.75 Å. It provides the maximum kinetic energy when the nanotube radius is 5.25 Å. This result agrees with Zhou et al. [31].

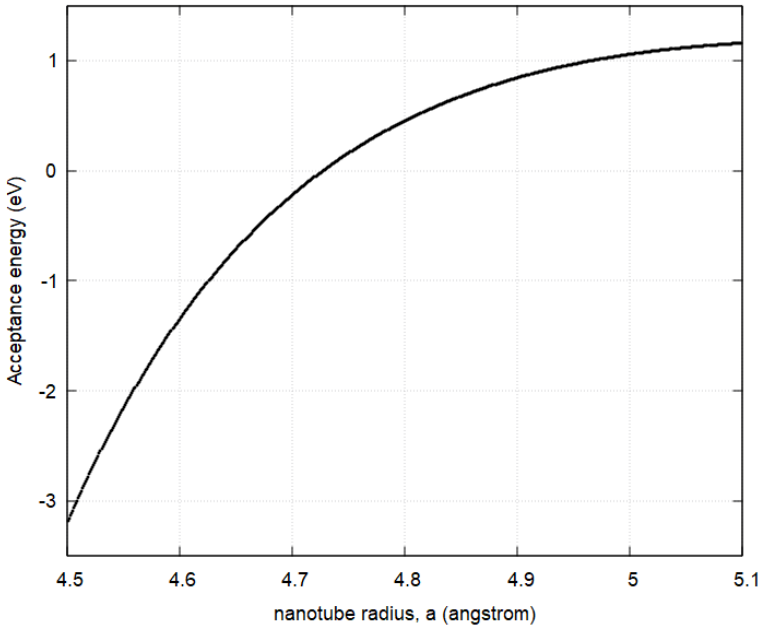


Figure 4: Acceptance energies of carbon nanotubes with different radii interacting with the endofullerene $Fe@C_{20}$

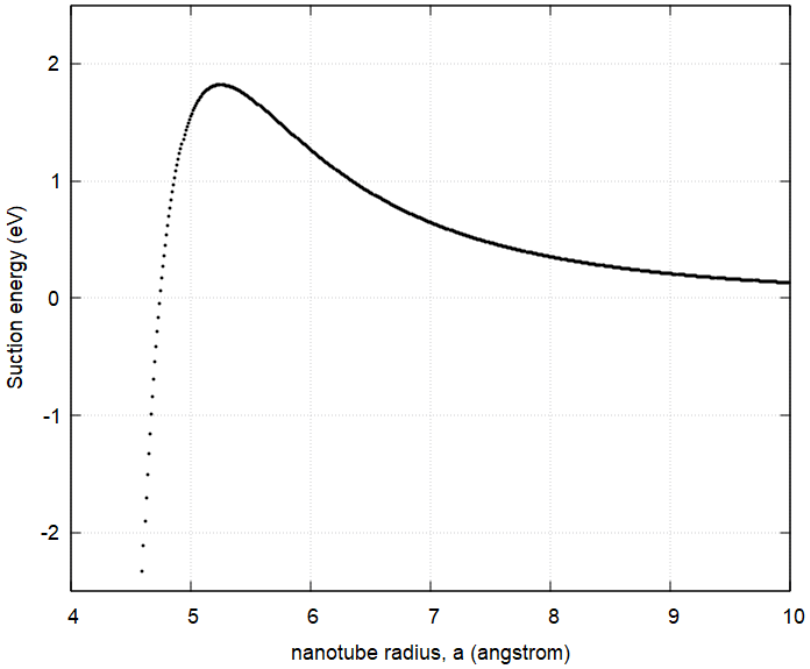


Figure 5: Suction energies of the carbon nanotubes with different radii interacting with the endofullerene $Fe@C_{20}$

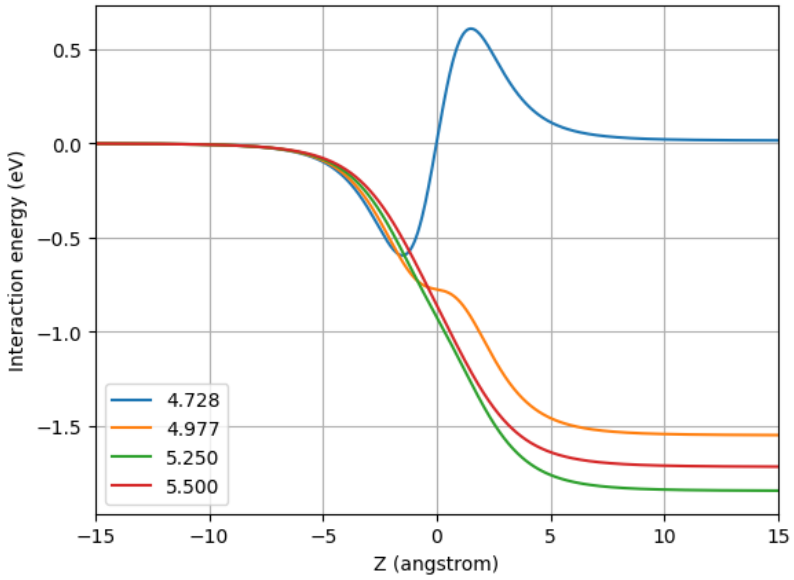


Figure 6: Total Interaction Energies of the carbon nanotubes with different radii interacting with the endofullerene $Fe@C_{20}$

According to Figure 6, the total interaction energies between the endofullerene and the carbon nanotubes with different radii, the endofullerene will be accepted into the nanotubes with radii greater than 4.977 Å because the interaction energies inside the nanotubes are less than the energies outside. There is no appearance of the barrier energy in radii of 4.977 Å, 5.250 Å and 5.500 Å, respectively. This means that the endofullerene is accepted into the carbon nanotubes with radii of 4.977 Å, 5.250 Å and 5.500 Å, respectively.

4 Conclusion

In this paper, we use the continuum approach with the van der Waals interaction and the Lennard-Jones potential function to determine the interaction energies between an endofullerene $Fe@C_{20}$ interacting with semi-infinite length carbon nanotubes with different radii. According to the results, we can consider such a system in two parts: we first studied the interaction between the C_{20} fullerene and the carbon nanotubes to find the optimal radii of nanotubes that accept the fullerene. The second part is the study of the interaction between the iron atom and the carbon nanotubes based on the assumption that the atom is located at the center of the C_{20} fullerene. In this paper, we can calculate the total interaction energies, the acceptance energies, and the suction energies between the endofullerene and the carbon nanotubes. The results show that the endofullerene $Fe@C_{20}$ is encapsulated into the carbon nanotubes with radii of 4.728 Å, 4.977 Å, 5.250 Å, and 5.500 Å. The encapsulation of the endofullerene into the carbon nanotubes depends strongly on the radii of the carbon nanotubes. In addition, the results of the acceptance energies show that when the radii of the carbon nanotubes are greater than or equal to 4.75 Å, the acceptance energies will be positive, meaning that the endofullerene will be accepted into the carbon nanotubes. The results are similar to those of Cox et al. (2007) [29], and the suction energy provides the maximum kinetic energy when the radius of the nanotube is 5.25 Å which is in accordance with Zhou et al. (2006) [31]. They state that a (8,8) single-walled carbon nanotube is the most stable configuration for the encapsulated C_{20} fullerene. However, this study only presents theoretical conclusions for the encapsulation of endofullerene inside carbon nanotubes of semi-infinite length. The outcomes of an experiment are also required in the future.

Acknowledgment. The authors express their gratitude to the Department of Mathematics, Faculty of Science, Ramkhamhaeng University, as well as the free and open-source Python and Gnuplot software.

References

[1] Iijima, S. *Helical microtubules of graphitic carbon.* Nature 354 (1991), 56–58.

[2] Sabu Thomas, Nandakumar Kalarikkal and Ann Rose Abraham. *Fundamentals and Properties of Multifunctional Nanomaterials.* (1st Edition). Elsevier, 2021.

[3] Che Y, Chen H, Gui H, et al. *Review of carbon nanotube nanoelectronics and macroelectronics.* Semicond Sci Technol. 29 (2014), 073001.

[4] Popov VN. *Carbon nanotubes: properties and application.* Mater Sci Eng R Rep. **43**(3) (2004), 61–102.

[5] Froudakis GE. *Hydrogen storage in nanotubes and nanostructures.* Matter Today. **14** (7–8) (2011), 324–328.

[6] Wang J. *Carbon-nanotube based electrochemical biosensors: a review.* Electroanalysis. **17**(1) (2005), 7-14.

[7] Kauffman DR, Star A. *Carbon nanotube gas and vapor sensors*. Angew Chem, **47**(35) (2008), 6550–6570.

[8] Kroto, H., Heath, J., O'Brien, S. et al. *C60: Buckminsterfullerene*. Nature. 318 (1985), 162–163.

[9] M. S. Dresselhaus, G. Dresselhaus, and P. C. Eklund. *Science of Fullerenes and Carbon Nanotubes*. (1st Edition). California: Academic, 1995

[10] P. J. F. Harris. *Carbon Nanotubes and Related Structures*. (1st Edition). Cambridge: Cambridge University Press, 2003

[11] J. Vavro, M. C. Llaguno, B. C. Satishkumar, D. E. Luzzi, and J. E. Fischer. *Electrical and thermal properties of C60-filled single-wall carbon nanotubes*. Appl. Phys. Lett. 80 (2002), 1450–1452.

[12] Prinzbach, H., Weiler, A., Landenberger, P. et al. *Gas-phase production and photoelectron spectroscopy of the smallest fullerene, C20*. Nature. 407 (2000), 60–63.

[13] Konstantinidis, N. P. *Unconventional magnetic properties of the icosahedral symmetry antiferromagnetic Heisenberg model*. Phys. Rev. B. 76 (2007), 104434.

[14] Yupeng Shen, Fancy Qian Wang, Jie Liu, Yaguang Guo, Xiaoyin Li, Guangzhao Qin, Ming Hu and Qian Wang. *A C_{20} fullerene-based sheet with ultrahigh thermal conductivity*. Nanoscale. 10 (2018), 6099-6104.

[15] Koichi Komatsu, Michihisa Murata and Yasujiro Murata. *Encapsulation of Molecular Hydrogen in Fullerene C_{60} by Organic Synthesis*. Science 307 (2005), 238-240.

[16] J.R. Heath et al. *Lanthanum complexes of spheroidal carbon shells*. J. Am. Chem. Soc. 107, 25 (1985), 7779–7780.

[17] T. Pradeep, G. U. Kulkarni, K. R. Kannan, T. N. G. Row, and C. N. R. Rao, *A novel iron fullerene (FeC_{60}) adduct in the solid state,* J. Am. Chem. Soc. **114**(6) (1992), 2272–2273.

[18] G. N. Churilov, O. A. Bayukov, E. A. Petrakovskaya, A. Y. Korets, V. G. Isakova, and Y. N. Titarenko, *Synthesis and study of iron-containing fullerene complexes*, Tech. Phys. **42**(9) (1997), 1111–1113.

[19] T. Lebedev et al. *Endometallofullerenes and their derivatives: Synthesis, physicochemical properties, and perspective application in biomedicine*. Colloids and Surfaces B: Biointerfaces. 222 (2023), 113–133.

[20] S. Kobayashi et al. *Conductivity and Field Effect Transistor of $La_2@C_{80}$ Metallofullerene*. J. Am. Chem. Soc. 125, 27 (2003), 8116-8117.

[21] Komatsu, K., Murata, M., & Murata, Y. *Encapsulation of Molecular Hydrogen in Fullerene C_{60} by Organic Synthesis*. *Science*, ***307***(5707) (2005), 238–240.

[22] Martin Saunders et al. *Noble Gas Atoms Inside Fullerenes*. *Science* **271** (1996), 1693-1697.

[23] Saunders, M., Jiménez-Vázquez, H., Cross, R. et al. (1994). *Probing the interior endohedral 3He@C_{60} and 3He@C_{70}*. Nature 367 (1994), 256-258.

[24] Nikolai A. Poklonski et al. *Magnetically operated nanorelay based on two single-walled carbon nanotubes filled with endofullerenes Fe@C20*. Journal of Nanophotonics, Vol. 4, Issue 1 (2010), 041675.

[25] Duangkamon Baowan and Ngamta Thamwattana. *Modelling encapsulation of gold and silver nanoparticles inside lipid nanotubes*. Physica A. 396 (2014), 149-154.

[26] Thamwattana N, Hill JM. *Continuum modelling for carbon and boron nitride nanostructures*. J. Phys. Condens. Matter. **19**(40) (2007), 406209.

[27] Hans Peter Lüthi, John Ammeter, Jan Almlöf and Knut Korsell. *The metal to ring distance of ferrocene as determined by ab initio calculations*. Chem. Phys. Lett. **69**(3) (1980), 540-542.

[28] Duangkamon Baowan, Barry J.Cox, Tamsyn A. Hilder, James M. Hill and Ngamta Thamwattana. *Modelling and Mechanics of Carbon-Based Nanostructured Materials*. Elsevier, 2017.

[29] Cox Barry J, Thamwattana Ngamta and Hill James M. *Mechanics of atoms and fullerenes in single-walled carbon nanotube. I.Acceptance and suction energies*. Proc. R. Soc. A. 463 (2007), 461-477.

[30] Sadeghi F, Ansari R, Darvizeh M. *Mechanics of metallic nanoparticles inside lipid nanotubes: Suction and acceptance energies*. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science. **231**(13) (2017), 2540-2553.

[31] L Zhou, Z Y Pan, Y X Wang, J Zhu, T J Liu and X M Jiang. *Stable configurations of C20 and C28 encapsulated in single wall carbon nanotubes*. Nanotechnology. 17 (2006), 1891.

[32] Thamwattana N, Hill JM. *Continuum modelling for carbon and boron nitride nanostructures*. J. Phys. Condens. Matter. **19**(40) (2007), 406209.

[33] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard III, and W. M. Skiff. UFF, *a full periodic table force field for molecular mechanics and molecular dynamics simulations*. Journal of the American Chemical Society **114**(25) (1992), 10024-10035.

[34] Alshehri MH. *Modeling Interactions of Iron Atoms Encapsulated in Nanotubes*. Crystals **11**(8) (2021), 845.

9. Mathematics education

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

การพัฒนาทักษะการแก้ปัญหาทางคณิตศาสตร์และการทำงานเป็นทีม ของนักเรียนระดับประกาศนียบัตรวิชาชีพชั้นปีที่ 1 เรื่อง พื้นที่ผิวและปริมาตร โดยใช้การจัดการเรียนรู้แบบปัญหาเป็นฐาน

ธวัชชัย อินทโฉม[1,†] และ ธีระพล สลีวงศ์[2]

[1]สาขาวิชาคณิตศาสตร์ศึกษา คณะวิทยาศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี 10140
[2]ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี 10140

บทคัดย่อ

งานวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) เปรียบเทียบทักษะการแก้ปัญหาทางคณิตศาสตร์ เรื่อง พื้นที่ผิวและปริมาตร ก่อนและหลังการจัดการเรียนรู้โดยใช้ปัญหาเป็นฐาน 2) เปรียบเทียบทักษะการแก้ปัญหาทางคณิตศาสตร์ เรื่อง พื้นที่ผิวและปริมาตร ที่ได้รับการจัดการเรียนรู้โดยใช้ปัญหาเป็นฐานกับเกณฑ์ร้อยละ 70 3) ศึกษาการทำงานเป็นทีมของนักเรียนระดับชั้นประกาศนียบัตรวิชาชีพชั้นปีที่ 1 กลุ่มตัวอย่างที่ใช้ในการวิจัยในครั้งนี้ ได้แก่ นักเรียนระดับชั้นประกาศนียบัตรวิชาชีพชั้นปีที่ 1 สาขาอาหารและโภชนาการ โรงเรียนจิตรลดาวิชาชีพ ภาคเรียนที่ 2 ปีการศึกษา 2566 จำนวน 29 คน ที่ได้มาจากการเลือกแบบเจาะจง เครื่องมือวิจัยประกอบด้วย 1) แผนการจัดการเรียนรู้โดยใช้ปัญหาเป็นฐาน เรื่อง พื้นที่ผิวและปริมาตร จำนวน 5 แผน 2) แบบทดสอบวัดทักษะการแก้ปัญหาทางคณิตศาสตร์ก่อนและหลังการจัดการเรียนรู้ จำนวน 5 ข้อ 3) แบบประเมินการทำงานเป็นทีม 4) แบบประเมินทักษะการแก้ปัญหาทางคณิตศาสตร์ ผลการวิจัยพบว่า คะแนนทักษะการแก้ปัญหาทางคณิตศาสตร์หลังการจัดการเรียนรู้โดยใช้ปัญหาเป็นฐานสูงกว่าก่อนการจัดการเรียนรู้อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อเปรียบเทียบกับเกณฑ์ร้อยละ 70 พบว่าสูงกว่าเกณฑ์ร้อยละ 70 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และการทำงานเป็นทีมพบว่าในภาพรวมอยู่ในเกณฑ์ดีถึงดีมาก และจากวิเคราะห์แยกเป็นรายด้านในภาพรวม พบว่า นักเรียนสามารถยอมรับความคิดเห็นซึ่งกันและกันได้

คำสำคัญ: การจัดการเรียนรู้แบบปัญหาเป็นฐาน, การทำงานเป็นทีม, ทักษะการแก้ปัญหาทางคณิตศาสตร์
2020 MSC: 97D50

*งานวิจัยเรื่องนี้ได้รับทุนสนับสนุนจากคณะวิทยาศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
†ผู้นำเสนอ ธวัชชัย อินทโฉม
อีเมล: thawatchai.int@cdti.ac.th (ธวัชชัย อินทโฉม), teerapol.sal@kmutt.ac.th (ธีระพล สลีวงศ์)

1 บทนำ

สมรรถนะสำคัญของผู้เรียนในศตวรรษที่ 21 ที่ทุกคนจะต้องเรียนรู้ตลอดชีวิต คือ การเรียนรู้ 3R และ 7C ซึ่ง 3R คือ อ่านออก เขียนได้ คิดเลขเป็น และ 7C ได้แก่ ทักษะด้านการคิดอย่างมีวิจารณญาณและทักษะในการแก้ปัญหา ทักษะด้านการสร้างสรรค์และนวัตกรรม ทักษะด้านความเข้าใจความต่างวัฒนธรรมต่างกระบวนทัศน์ ทักษะด้านความร่วมมือ การทำงานเป็นทีม และภาวะผู้นำ ทักษะด้านการสื่อสารสารสนเทศ และรู้เท่าทันสื่อ ทักษะด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศและการสื่อสาร ทักษะอาชีพและทักษะการเรียนรู้ กล่าวคือให้ผู้เรียนมีความสามารถในการคิดวิเคราะห์ การคิดสังเคราะห์ การคิดอย่างสร้างสรรค์ การคิดอย่างมีวิจารณญาณ และการคิดอย่างเป็นกระบวนการเพื่อนำไปสู่การสร้างองค์ความรู้ของตนเองและสังคมได้อย่างเหมาะสมส่งเสริมความสามารถในการแก้ปัญหาได้อย่างถูกต้อง (สำนักบริหารงานการมัธยมศึกษาตอนปลาย สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน และกระทรวงศึกษาธิการ, 2558)

ทักษะเพื่อการดำรงชีวิตในศตวรรษที่ 21 ว่า สาระวิชาหลัก จะนำมาสู่การกำหนดเป็นกรอบแนวคิดและยุทธศาสตร์สำคัญต่อการจัดการเรียนรู้ในเนื้อหาเชิงสหวิทยาการ หรือหัวข้อสำหรับศตวรรษที่ 21 โดยการส่งเสริมความเข้าใจในเนื้อหา วิชาแกนหลัก และสอดแทรกทักษะแห่งศตวรรษที่ 21 เข้าไปในทุกวิชาแกนหลัก (วิจารณ์ พานิช, 2555) ทักษะศตวรรษที่ 21 มีความสำคัญและจำเป็นอย่างยิ่งต่อการดำรงชีวิตในศตวรรษใหม่จะช่วยเตรียมความพร้อมให้คนรู้จักคิด เรียนรู้ ทำงาน แก้ปัญหา สื่อสารและร่วมมือทำงานได้อย่างมีประสิทธิผลไปตลอดชีวิต (บันเย็น เพ็งกระจ่าง, 2561)

สภาครูคณิตศาสตร์แห่งสหรัฐอเมริกา ซึ่งเป็นสภาที่มีบทบาทในการกำหนดทิศทางในการจัดการเรียนการสอนคณิตศาสตร์ในสหรัฐอเมริกาในปัจจุบัน จุดประสงค์ของการเรียนการสอนคณิตศาสตร์ในศตวรรษที่ 21 ที่สหรัฐอเมริกามุ่งเน้นและกำหนดเป็นจุดประสงค์กว้าง ๆ ได้แก่ เพื่อให้ผู้เรียนได้ตระหนักถึงคุณค่าของคณิตศาสตร์ เพื่อให้ผู้เรียนเป็นหนักแก้ปัญหา สื่อสารคณิตศาสตร์ได้ ให้เหตุผลทางคณิตศาสตร์ได้ (NCTM, 1989) ซึ่งสอดคล้องกับกระทรวงศึกษาธิการ กล่าวว่าการจัดการศึกษาตามแนวทางหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน ซึ่งได้กำหนดทักษะและกระบวนการทางคณิตศาสตร์ (สถาบันส่งเสริมวิทยาศาสตร์และเทคโนโลยี กระทรวงศึกษาธิการ, 2551)

สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน, 2566) การทดสอบทางการศึกษาระดับชาติด้านอาชีวศึกษา (V- NET) ด้วยระบบดิจิทัล มีวัตถุประสงค์เพื่อทดสอบความรู้และประเมินความพร้อมในการเข้าสู่โลกอาชีพทักษะในศตวรรษที่ 21 และการพัฒนาตนเองอย่างต่อเนื่อง และจากรายงานผลการทดสอบทางการศึกษาระดับชาติด้านอาชีวศึกษา (V-NET) ในปีการศึกษา 2565 พบว่าคะแนนเฉลี่ยสมรรถนะที่จำเป็นในการเข้าสู่อาชีพระดับประเทศของสาขาอาหารและโภชนาการ เท่ากับ 52.11 คะแนน โรงเรียนจิตรลดาวิชาชีพได้คะแนนเฉลี่ยของสาขาอาหารและโภชนาการ เท่ากับ 56.98 คะแนน

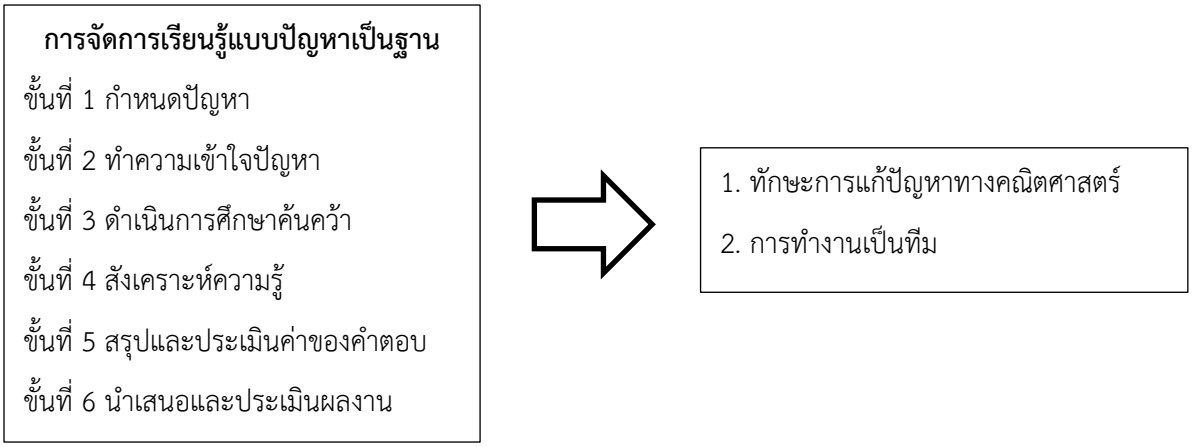
ปัญหาการจัดการเรียนรู้วิชาคณิตศาสตร์ของผู้วิจัยในชั้นเรียนของนักเรียนชั้นระดับประกาศนียบัตรวิชาชีพชั้นปีที่ 1 สถาบันเทคโนโลยีจิตรลดา สังกัดโรงเรียนจิตรลดาวิชาชีพ ภาคเรียนที่ 1 ปีการศึกษา 2566 พบว่า นักเรียน

สามารถคิดคำนวณหาค่าจากสูตรพื้นฐานได้อย่างถูกต้อง ตามหลักการทางคณิตศาสตร์ แต่นักเรียนไม่สามารถนำความรู้และกระบวนการต่าง ๆ ทางคณิตศาสตร์ที่เรียนรู้มาแล้วมาวางแผนการแก้ปัญหาเพื่อจะนำไปสู่การดำเนินการแก้โจทย์ปัญหาได้ถูกต้อง และจากการแบ่งให้นักเรียนทำงานเป็นทีมพบว่า นักเรียนไม่แบ่งหน้าที่กันในการทำงาน ทุกคนในทีมไม่ช่วยกันทำงาน มีการปรึกษาหารือกันที่ไม่มากพอ เมื่อเกิดปัญหาไม่ช่วยกันแก้ปัญหาที่เกิดขึ้น และขาดการวางแผนในการทำงาน ทางผู้วิจัยจึงหาแนวทางในการจัดกิจกรรมการเรียนรู้ โดยสนใจที่จะนำการจัดกิจกรรมการเรียนรู้โดยใช้ปัญหาเป็นฐานมาใช้เพราะเห็นว่าเป็นกระบวนการเรียนรู้ที่เหมาะสมกับนักเรียนในการสร้างองค์ความรู้ด้วยตนเองจากปัญหาหรือสถานการณ์ที่ครูกำหนดเพื่อกระตุ้นให้นักเรียนเกิดความสนใจ

การจัดการเรียนรู้โดยใช้ปัญหาเป็นฐาน (Problem-Based Learning) เป็นกระบวนการเรียนรู้โดยใช้ปัญหาเป็นตัวกระตุ้นให้ผู้เรียนตั้งสมมติฐาน หาเหตุและกลไกของการเกิดปัญหานั้น รวมถึงการค้นคว้าความรู้พื้นฐานที่เกี่ยวข้องกับปัญหา เพื่อนำไปสู่การแก้ปัญหาต่อไป นอกจากนี้ยังมุ่งให้ผู้เรียนใฝ่หาความรู้เพื่อแก้ไขปัญหา ได้คิดเป็นทำเป็น มีการตัดสินใจที่ดี และสามารถเรียนรู้การทำงานเป็นทีม โดยเน้นให้ผู้เรียนได้เกิดการเรียนรู้ด้วยตนเอง และสามารถนำทักษะจากการเรียนมาช่วยแก้ปัญหาในชีวิต (สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, 2551)

การทำงานเป็นทีมเป็นการที่สมาชิกเสียสละความเป็นส่วนตัวในการทำงานร่วมกันเพื่อให้บรรลุวัตถุประสงค์ (Nolan, 1989) และเป็นพฤติกรรมของสมาชิกในกลุ่มที่แบ่งปันข้อมูลและร่วมมือกันทำงาน (Dickinson & McIntyre, 1997) อีกทั้งยังมีการทำงานร่วมกับผู้อื่น มีการช่วยเหลือสนับสนุน และให้กำลังใจกันและกัน ร่วมกันแก้ปัญหาความขัดแย้ง อีกทั้งให้คำแนะนำผู้อื่นด้วย (Wang, et al., 2009) ซึ่งองค์ประกอบของการทำงานเป็นทีมให้ประสบผลสำเร็จได้แก่ การพึ่งพาซึ่งกันและกัน มีความรับผิดชอบต่อคนอื่น ๆ ความสัมพันธ์ระหว่างบุคคล ยอมรับความแตกต่างของสมาชิกในทีม มีการสื่อสารกัน มีบทบาทหน้าที่ สมาชิกในทีมมีความร่วมมือ (Tarricone & Luca, 2002)

จากที่มาและความสำคัญทั้งหมดจึงทำให้เกิดการวิจัยในหัวข้อ การพัฒนาทักษะการแก้ปัญหาทางคณิตศาสตร์และการทำงานเป็นทีมของนักเรียนประกาศนียบัตรวิชาชีพชั้นปีที่ 1 เรื่อง พื้นที่ผิวและปริมาตร โดยใช้การจัดการเรียนรู้แบบปัญหาเป็นฐาน โดยผู้วิจัยได้กำหนดกรอบแนวคิดในการวิจัยครั้งนี้ ดังแสดงในภาพที่ 1



ภาพที่ 1 กรอบแนวคิดในการวิจัย

2 วัตถุประสงค์ของการวิจัย

1. เพื่อเปรียบเทียบทักษะการแก้ปัญหาทางคณิตศาสตร์ เรื่อง พื้นที่ผิวและปริมาตร ของนักเรียนระดับประกาศนียบัตรวิชาชีพชั้นปีที่ 1 ก่อนและหลังการจัดการเรียนรู้โดยใช้ปัญหาเป็นฐาน

2. เพื่อเปรียบเทียบทักษะการแก้ปัญหาทางคณิตศาสตร์ เรื่อง พื้นที่ผิวและปริมาตร ของนักเรียนระดับประกาศนียบัตรวิชาชีพชั้นปีที่ 1 ที่ได้รับการจัดการเรียนรู้โดยใช้ปัญหาเป็นฐานกับเกณฑ์ร้อยละ 70

3. เพื่อศึกษาการทำงานเป็นทีมของนักเรียนระดับประกาศนียบัตรวิชาชีพชั้นปีที่ 1 โดยการจัดการเรียนรู้โดยใช้ปัญหาเป็นฐาน

3 สมมติฐานของการวิจัย

1. นักเรียนระดับประกาศนียบัตรวิชาชีพชั้นปีที่ 1 ที่ได้รับการจัดการเรียนรู้โดยใช้ปัญหาเป็นฐานมีทักษะการแก้ปัญหาทางคณิตศาสตร์ เรื่อง พื้นที่ผิวและปริมาตร หลังเรียนสูงกว่าก่อนเรียน

2. นักเรียนระดับประกาศนียบัตรวิชาชีพชั้นปีที่ 1 ที่ได้รับการจัดการเรียนรู้โดยใช้ปัญหาเป็นฐานมีทักษะการแก้ปัญหาทางคณิตศาสตร์ เรื่อง พื้นที่ผิวและปริมาตร สูงกว่าเกณฑ์ร้อยละ 70

3. นักเรียนระดับประกาศนียบัตรวิชาชีพชั้นปีที่ 1 โดยการจัดการเรียนรู้โดยใช้ปัญหาเป็นฐานมีการทำงานเป็นทีมในระดับที่ดีถึงดีมาก

4 นิยามและศัพท์เฉพาะ

4.1 ประชากรกลุ่มตัวอย่าง

1. ประชากร คือ นักเรียนระดับประกาศนียบัตรวิชาชีพชั้นปีที่ 1 โรงเรียนจิตรลดาวิชาชีพ เขตดุสิต จังหวัดกรุงเทพมหานคร ภาคเรียนที่ 2 ปีการศึกษา 2566 ในสังกัดกระทรวงอุดมศึกษา วิทยาศาสตร์ วิจัยและนวัตกรรม ซึ่งมีการจัดห้องเรียนแบบแยกระดับความสามารถของนักเรียนโดยแบ่งเป็นสาขาวิชาต่าง ๆ จำนวน 11 ห้อง ซึ่งมีจำนวน 142 คน

2. กลุ่มตัวอย่าง คือ นักเรียนระดับประกาศนียบัตรวิชาชีพชั้นปีที่ 1 สาขาอาหารและโภชนาการ โรงเรียนจิตรลดาวิชาชีพ เขตดุสิต จังหวัดกรุงเทพมหานคร ภาคเรียนที่ 2 ปีการศึกษา 2566 จำนวน 29 คน ในสังกัดกระทรวงอุดมศึกษา วิทยาศาสตร์ วิจัยและนวัตกรรม ที่ได้มาจากการเลือกแบบเจาะจง

4.2 การจัดการเรียนรู้โดยใช้ปัญหาเป็นฐาน หมายถึง กระบวนการเรียนรู้ที่มีการจัดกิจกรรมการเรียนที่จัดให้นักเรียนเป็นทีมโดยคละความสามารถ โดยใช้เกณฑ์จากผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ในภาคเรียนที่ 1 ปีการศึกษา 2566 มีนักเรียนกลุ่มเก่งจำนวน 10 คน นักเรียนกลุ่มกลางจำนวน 10 คน นักเรียนกลุ่มอ่อนจำนวน 9 คน โดยภาพรวมจะมีทีมละ 3 คน โดยแบ่งเป็นนักเรียนเก่ง นักเรียนกลาง และนักเรียนอ่อนทีมละ 1 คน โดยใช้

ประเด็นปัญหาเหตุการณ์กระตุ้นให้นักเรียนวิเคราะห์ ค้นหา สืบค้น สำรวจ ค้นคว้าหาแนวทางแก้ไขปัญหา เพื่อนำไปสู่การอภิปรายและสรุปองค์ความรู้ที่เป็นคำตอบของปัญหานั้นร่วมกัน ซึ่งประกอบด้วย 6 ขั้นตอน ดังนี้

ขั้นที่ 1 กำหนดปัญหา เป็นขั้นที่ครูจัดสถานการณ์ต่าง ๆ กระตุ้นให้ผู้เรียนเกิดความสนใจ และมองเห็นปัญหาสามารถกำหนดสิ่งที่เป็นปัญหาที่นักเรียนอยากรู้อยากเรียนได้และเกิดความสนใจที่จะค้นหาคำตอบ

ขั้นที่ 2 ทำความเข้าใจปัญหา นักเรียนจะต้องทำความเข้าใจปัญหาที่ต้องการเรียนรู้ ซึ่งนักเรียนจะต้องอธิบายถึงสิ่งต่าง ๆ ที่เกี่ยวข้องกับปัญหาได้

ขั้นที่ 3 ดำเนินการศึกษาค้นคว้า นักเรียนกำหนดสิ่งที่ต้องเรียนและดำเนินการศึกษาค้นคว้า ด้วยวิธีที่หลากหลาย

ขั้นที่ 4 สังเคราะห์ความรู้ เป็นขั้นที่นักเรียนนำความรู้ที่ได้ค้นคว้ามาแลกเปลี่ยนเรียนรู้ ร่วมมือกัน อภิปรายผลและสังเคราะห์ความรู้ที่ได้มาว่ามีความเหมาะสมหรือไม่เพียงใด เพียงพอกับการตรวจสอบสมมติฐานที่ตั้งไว้หรือไม่ แล้วนำข้อมูลที่ได้ไปตรวจสอบสมมติฐานและแก้ปัญหา ถ้าไม่เพียงพอ สมาชิกภายในทีมจะต้องกำหนดสิ่งที่ต้องเรียนเพิ่มเติม แผนการเรียนรู้ และแหล่งข้อมูลแล้วดำเนินการศึกษาอีกครั้งหนึ่งเพื่อให้ได้ข้อมูลที่สมบูรณ์ก่อน

ขั้นที่ 5 สรุปและประเมินค่าของคำตอบ นักเรียนแต่และทีมสรุปผลงานของทีมตนเองและประเมินผลงานว่าข้อมูลที่ศึกษาค้นคว้ามีความเหมาะสมหรือไม่เพียงใด โดยพยายามตรวจสอบแนวคิดภายในทีมของตนเองอย่างอิสระทุกทีมช่วยกันสรุปองค์ความรู้ในภาพรวมของปัญหาอีกครั้ง

ขั้นที่ 6 นำเสนอและประเมินผลงาน นักเรียนนำข้อมูลที่ได้มาจัดระบบองค์ความรู้นำเสนอเป็นผลงานในรูปแบบที่หลากหลาย นักเรียนทุกทีมรวมทั้งนักเรียนที่เกี่ยวข้องกับปัญหาร่วมกันประเมินผล

4.3 ทักษะการแก้ปัญหาทางคณิตศาสตร์ หมายถึง ความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ของนักเรียน โดยใช้กระบวนการแก้ปัญหาของโพลยา (Polya, 1985) ซึ่งประกอบด้วย 4 ขั้นตอน ดังนี้

ขั้นการทำความเข้าใจปัญหา เป็นขั้นที่นักเรียนวิเคราะห์ปัญหา โดยจะต้องระบุถึงสิ่งที่ปัญหากำหนดและสิ่งที่ปัญหาต้องการทราบ

ขั้นการเลือกกลยุทธวิธีในการแก้ปัญหา เป็นขั้นที่นักเรียนต้องพิจารณาหลักการหรือวิธีการทางคณิตศาสตร์มากำหนดเป็นแนวทางในการแก้ปัญหา

ขั้นการใช้วิธีการแก้ปัญหา เป็นขั้นที่นักเรียนดำเนินการตามแนวทางที่วางไว้ โดยใช้การดำเนินการทางคณิตศาสตร์อย่างเป็นระบบ เพื่อหาคำตอบของปัญหา

ขั้นการสรุปคำตอบ เป็นขั้นที่นักเรียนสรุปผลที่ได้จากการดำเนินการแก้ปัญหา ให้สอดคล้องกับสิ่งที่ปัญหาต้องการทราบ

ซึ่งวัดได้จากเกณฑ์การประเมินทักษะการแก้ปัญหาทางคณิตศาสตร์ดังแสดงในตารางที่ 1

4.4 การทำงานเป็นทีม หมายถึง การที่นักเรียน 2 - 3 คน มารวมตัวกันเพื่อปฏิบัติหน้าที่ที่ได้รับมอบหมาย แต่ละคนมีบทบาทหน้าที่ในการทำงานของทีม มีส่วนรวมในวางแผนและการดำเนินการทำงาน การยอมรับฟังความคิดเห็น

ของผู้อื่นและการแสดงความคิดเห็น มีส่วนร่วมกัน มีการติดต่อสื่อสารกันกับสมาชิกภายในทีม ซึ่งวัดได้จากแบบประเมินการทำงานเป็นทีมที่ผู้วิจัยสร้างขึ้นมาซึ่งประกอบด้วย 5 องค์ประกอบ คือ ความรับผิดชอบ การติดต่อสื่อสาร การวางแผนการทำงานร่วมกัน ความร่วมมือ และการยอมรับความคิดเห็นซึ่งกันและกัน ซึ่งมีเกณฑ์การประเมินดังแสดงในตารางที่ 2

4.5 เกณฑ์ หมายถึง คะแนนเฉลี่ยขั้นต่ำที่จะยอมรับได้ว่านักเรียนมีทักษะการแก้ปัญหาทางคณิตศาสตร์ ซึ่งผู้วิจัยใช้เกณฑ์ร้อยละ 70 ของคะแนนรวม ซึ่งอยู่ในระดับดี ตามกระทรวงศึกษาธิการ (สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, 2551)

5 เครื่องมือที่ใช้ในการวิจัย

5.1 เครื่องมือที่ใช้ในงานวิจัย

1. แผนการจัดการเรียนรู้โดยใช้ปัญหาเป็นฐาน จำนวน 5 แผน ประกอบด้วย

พื้นที่ผิวและปริมาตรของปริซึม	จำนวน 2 ชั่วโมง
พื้นที่ผิวและปริมาตรของพีระมิด	จำนวน 2 ชั่วโมง
พื้นที่ผิวและปริมาตรของทรงกระบอก	จำนวน 2 ชั่วโมง
พื้นที่ผิวและปริมาตรของกรวย	จำนวน 2 ชั่วโมง
พื้นที่ผิวและปริมาตรของทรงกลม	จำนวน 2 ชั่วโมง

ในแต่ละแผนจะมีขั้น 6 ขั้นตอน ประกอบด้วย ขั้นกำหนดปัญหา ขั้นทำความเข้าใจปัญหา ขั้นดำเนินการศึกษา ขั้นสังเคราะห์ความรู้ ขั้นสรุปและประเมินค่าของคำตอบ และขั้นนำเสนอและประเมินผลงาน

2. แบบทดสอบวัดทักษะการแก้ปัญหาทางคณิตศาสตร์ก่อนและหลังการจัดการเรียนรู้แบบปัญหาเป็นฐาน จำนวน 5 ข้อ

3. แบบประเมินการทำงานเป็นทีมที่ผู้วิจัยสร้างขึ้นดังตารางที่ 2

4. แบบประเมินทักษะการแก้ปัญหาทางคณิตศาสตร์ ซึ่งใช้เกณฑ์การให้คะแนนการแก้ปัญหาทางคณิตศาสตร์ของสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี แสดงในตารางที่ 1

5. แบบบันทึกหลังสอน

5.2 วิธีการสร้างเครื่องมือวิจัย

1. ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องเกี่ยวกับปัญหาเป็นฐาน ทักษะการแก้ปัญหาทางคณิตศาสตร์ และการทำงานเป็นทีม

2. สร้างแผนการจัดการเรียนรู้แบบปัญหาเป็นฐาน เรื่อง พื้นที่ผิวและปริมาตร จำนวน 5 แผน ซึ่งมีขั้นตอนการสร้างแผนการจัดการเรียนรู้ ดังนี้

2.1 ศึกษาหลักสูตรระดับประกาศนียบัตรวิชาชีพชั้นปีที่ 1 วิชาคณิตศาสตร์เพื่อการออกแบบ พุทธศักราช 2556 สำนักงานคณะกรรมการอาชีวศึกษา กระทรวงศึกษาธิการ

2.2 สำรวจเกี่ยวกับการนำเรื่องพื้นที่ผิวและปริมาตรไปใช้ในวิชาอื่น ๆ ที่เรียนในสาขากับนักเรียนระดับ

ประกาศนียบัตรวิชาชีพชั้นปีที่ 3 ที่เคยเรียนเรื่องพื้นที่ผิวและปริมาตรมาก่อน เพื่อนำไปใช้ในการกำหนดปัญหาในขั้นที่ 1 ของปัญหาเป็นฐาน

2.3 กำหนดเนื้อหาและจุดประสงค์การเรียนรู้วิชาคณิตศาสตร์เพื่อการออกแบบ ตลอดจนเขียนแผนการจัดการเรียนรู้ โดยประกอบด้วย สาระการเรียนรู้ ผลการเรียนรู้ จุดประสงค์การเรียนรู้ สาระสำคัญ กิจกรรมการเรียนรู้ สื่อและแหล่งการเรียนรู้ การวัดประเมินผลการเรียนรู้

2.4 นำแผนการจัดการเรียนรู้ที่สร้างขึ้นเรียบร้อยแล้ว เสนอต่อผู้เชี่ยวชาญจำนวน 5 คน เพื่อหาค่าดัชนีความเที่ยงตรงเชิงเนื้อหา (CVI) ความสอดคล้องระหว่างจุดประสงค์การเรียนรู้กับกิจกรรมการเรียนรู้ และความถูกต้องของภาษาที่ใช้ จากนั้นนำแผนการเรียนรู้มาปรับปรุงแก้ไขตามข้อเสนอแนะของผู้เชี่ยวชาญ และนำเสนอต่ออาจารย์ที่ปรึกษาอีกครั้ง เพื่อตรวจสอบความถูกต้อง ก่อนนำไปใช้เป็นเครื่องมือในการจัดกิจกรรมการเรียนรู้ของการวิจัยต่อไป

ค่าดัชนีความตรงเชิงเนื้อหา (Index of content validity (CVI)) หมายถึง สัดส่วนของข้อความที่ผู้เชี่ยวชาญให้คะแนน 3 หรือ 4 ซึ่งมีสูตรในการคำนวณและมีเกณฑ์การให้คะแนนความคิดเห็น ดังนี้

$$CVI = \frac{\sum n_3 \ or \ n_4}{N}$$

เมื่อ $\sum n_3 \ or \ n_4$ ผลรวมของความคิดเห็นของผู้เชี่ยวชาญที่ให้คะแนน 3 หรือ 4

โดยมีความคิดเห็นเป็น 4 ระดับ ดังนี้

1 หมายถึง ไม่เกี่ยวข้อง

2 หมายถึง เกี่ยวข้องบ้าง

3 หมายถึง ค่อนข้างเกี่ยวข้อง

4 หมายถึง เกี่ยวข้องมาก

$\sum n_3$ คือ จำนวนผู้เชี่ยวชาญที่เห็นว่าข้อคำถามค่อนข้างเกี่ยวข้องกับสิ่งที่ต้องการวัด

$\sum n_4$ คือ จำนวนผู้เชี่ยวชาญที่เห็นว่าข้อคำถามเกี่ยวข้องมากกับสิ่งที่ต้องการวัด

N คือ จำนวนผู้เชี่ยวชาญทั้งหมด

ค่า Item-CVI คำนวณจากสัดส่วนของผู้เชี่ยวชาญที่มีความเห็นตรงกันว่าข้อความนั้น ๆ เกี่ยวข้องกับสิ่งที่วัด ถ้าค่า Item-CVI ที่มีค่ามากกว่า .80 สามารถนำไปใช้ได้ (วีระยุทธ, 2565)

3. สร้างแบบบันทึกหลังสอนสำหรับครูผู้สอน เพื่อบันทึกทุกครั้งหลังการเรียนการสอนที่เน้นปัญหาเป็นฐาน

4. สร้างแบบทดสอบวัดทักษะการแก้ปัญหาทางคณิตศาสตร์ เรื่อง พื้นที่ผิวและปริมาตร จำนวน 15 ข้อ ซึ่งแบ่งเป็นเรื่องพื้นที่ผิวและปริมาตรปริซึม พีระมิด ทรงกระบอก กรวย และทรงกลมอย่างละ 3 ข้อ แล้วเสนอต่อผู้เชี่ยวชาญจำนวน 5 คน เพื่อหาค่าดัชนีความเที่ยงตรงเชิงเนื้อหา (CVI) ความสอดคล้องระหว่างจุดประสงค์การเรียนรู้กับข้อคำถามของแบบทดสอบ และความถูกต้องของภาษาที่ใช้ ซึ่งพบว่าค่า Item-CVI ตั้งแต่ 0.80 ขึ้นไป มีจำนวน 13 ข้อ จากนั้นนำแบบทดสอบวัดทักษะการแก้ปัญหาทางคณิตศาสตร์ ดำเนินทดสอบกับนักเรียนระดับประกาศนียบัตรวิชาชีพชั้นปีที่ 2 สาขาอาหารและโภชนาการ จำนวน 13 คน เพื่อหาค่าความยากและค่าอำนาจจำแนกของแบบทดสอบ ซึ่งพบว่าแบบทดสอบจำนวน 13 ข้อ สามารถนำมาใช้วัดทักษะการแก้ปัญหาทางคณิตศาสตร์ได้ ผู้วิจัยได้

เลือกแบบทดสอบเรื่องพื้นที่ผิวและปริมาตรปริซึม พีระมิด ทรงกระบอก กรวย และทรงกลมอย่างละ 1 ข้อ แล้วนำมาใช้เป็นแบบทดสอบวัดทักษะการแก้ปัญหาทางคณิตศาสตร์ก่อนเรียนและหลังเรียน เรื่อง พื้นที่ผิวและปริมาตร จำนวน 5 ข้อ ซึ่งแบบทดสอบก่อนเรียนและหลังเรียนเป็นแบบสอบชุดเดียวกัน โดยมีเกณฑ์การให้คะแนนสำหรับวัดทักษะการแก้ปัญหาทางคณิตศาสตร์ในแต่ละข้อ (สถาบันส่งเสริมวิทยาศาสตร์และเทคโนโลยี, 2551) ดังแสดงในตารางที่ 1 และจะใช้ผลรวมในทุกรายการประเมิน สรุปเป็นระดับทักษะการแก้ปัญหาทางคณิตศาสตร์ โดยมีเกณฑ์ประเมินคุณภาพ ดังต่อไปนี้

1 - 4 คะแนน หมายถึง อยู่ในระดับต้องปรับปรุง

5 - 8 คะแนน หมายถึง อยู่ในระดับพอใช้

9 - 12 คะแนน หมายถึง อยู่ในระดับดี

ตารางที่ 1 เกณฑ์การประเมินให้คะแนนสำหรับแบบทดสอบวัดทักษะการแก้ปัญหาทางคณิตศาสตร์

รายการประเมิน	ระดับคุณภาพ	เกณฑ์การประเมิน
1. ความเข้าใจปัญหา	3 (ดี) 2 (พอใช้) 1 (ต้องปรับปรุง)	เข้าใจปัญหาได้อย่างถูกต้อง เข้าใจปัญหาบางส่วนไม่ถูกต้อง เข้าใจปัญหาน้อยมากหรือไม่เข้าใจปัญหา
2. การเลือกกลยุทธวิธีในการแก้ปัญหา	3 (ดี) 2 (พอใช้) 1 (ต้องปรับปรุง)	เลือกวิธีการแก้ปัญหาได้เหมาะสมและเขียนประโยคสัญลักษณ์คณิตศาสตร์ได้ถูกต้อง เลือกวิธีการแก้ปัญหา ซึ่งอาจจะนำไปสู่คำตอบที่ถูก แต่ยังมีส่วนผิดโดยอาจเขียนประโยคสัญลักษณ์คณิตศาสตร์ไม่ถูกต้อง เลือกวิธีการแก้ปัญหาส่วนใหญ่ไม่ถูกต้อง
3. การใช้วิธีการแก้ปัญหา	3 (ดี) 2 (พอใช้) 1 (ต้องปรับปรุง)	นำวิธีการแก้ปัญหาไปใช้ได้อย่างถูกต้อง นำวิธีการแก้ปัญหาไปใช้ได้อย่างถูกต้องเป็นบางส่วน นำวิธีการแก้ปัญหาไปใช้ได้ไม่ถูกต้อง
4. การสรุปคำตอบ	3 (ดี) 2 (พอใช้) 1 (ต้องปรับปรุง)	สรุปคำตอบได้อย่างถูกต้อง สมบูรณ์ สรุปคำตอบที่ไม่สมบูรณ์หรือใช้สัญลักษณ์ไม่ถูกต้อง ไม่มีการสรุปคำตอบ

5. สร้างแบบประเมินการทำงานเป็นทีม โดยดำเนินการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องเกี่ยวกับการทำงานเป็นทีม จากนั้นสรุปเป็นองค์ประกอบของการทำงานเป็นทีมได้ 5 องค์ประกอบ คือ ความรับผิดชอบ การติดต่อสื่อสาร การวางแผนการทำงานร่วมกัน ความร่วมมือ การยอมรับความคิดเห็นซึ่งกันและกัน จากนั้นกำหนดเกณฑ์การให้คะแนนเป็นระดับคะแนน 3 (ดีมาก), 2 (ดี), 1 (พอใช้) และ 0 (ปรับปรุง) เสนอต่อผู้เชี่ยวชาญจำนวน 5 คน เพื่อหาค่าดัชนีความเที่ยงตรงเชิงเนื้อหา (CVI) ความสอดคล้องระหว่างจุดประสงค์การเรียนรู้กับเกณฑ์การประเมิน

การทำงานเป็นทีม และความถูกต้องของภาษาที่ใช้ แล้วนำมาปรับปรุงตามข้อเสนอแนะของผู้เชี่ยวชาญ โดยมีเกณฑ์การให้คะแนนสำหรับแบบการทำงานเป็นทีม ดังแสดงในตารางที่ 2 และจะใช้ผลรวมในทุกรายการประเมิน สรุปเป็นระดับการทำงานเป็นทีม โดยมีเกณฑ์ประเมินคุณภาพ ดังต่อไปนี้

0 - 3 คะแนน หมายถึง อยู่ในระดับปรับปรุง

4 - 7 คะแนน หมายถึง อยู่ในระดับพอใช้

8 - 11 คะแนน หมายถึง อยู่ในระดับดี

12 – 15 คะแนน หมายถึง อยู่ในระดับดีมาก

ตารางที่ 2 เกณฑ์การประเมินให้คะแนนการทำงานเป็นทีม

รายการประเมิน	**เกณฑ์การประเมิน**			
	ดีมาก	**ดี**	**พอใช้**	**ปรับปรุง**
	3	**2**	**1**	**0**
ความรับผิดชอบ	ทุกคนมีหน้าที่และความรับผิดชอบต่อหน้าที่ของตนเอง	มีอย่างน้อยร้อยละ 30 ไม่มีหน้าที่และไม่รับผิดชอบ	มีอย่างน้อยร้อยละ 60 ไม่มีหน้าที่และไม่รับผิดชอบ	ไม่มีการแบ่งหน้าที่รับผิดชอบกันภายในทีม
การติดต่อสื่อสาร	ทุกคนมีการติดติดต่อสื่อสารกัน	มีอย่างน้อยร้อยละ 30 ที่ไม่มีการติดต่อสื่อสารกัน	มีอย่างน้อยร้อยละ 60 ที่ไม่มีการติดต่อสื่อสารกัน	ไม่มีการติดต่อสื่อสารกันภายในทีม
การวางแผนการทำงานร่วมกัน	1. ปรึกษาหารือ 2. เตรียมข้อมูลได้เหมาะสม 3. วางแผนการทำงาน 4. ปฏิบัติตามแผนและพัฒนาผลงาน	ขาด 1 ขั้นตอน จาก 4 ขั้นตอน	ขาด 2 - 3 ขั้นตอน จาก 4 ขั้นตอน	ไม่มีการวางแผนกันภายในทีม
ความร่วมมือ	ทุกคนมีส่วนร่วมให้ความร่วมมืออย่างเต็มที่	มีผู้ไม่ให้ความร่วมมืออย่างน้อยร้อยละ 30	มีผู้ไม่ให้ความร่วมมืออย่างน้อยร้อยละ 60	ไม่มีผู้ให้ความร่วมมือเลย
การยอมรับความคิดเห็นซึ่งกันและกัน	ทุกคนยอมรับฟัง ความคิดเห็นของผู้อื่นและมีการแสดงความคิดเห็น	อย่างน้อยร้อยละ 60 ของทีม ยอมรับฟังความคิดเห็นของผู้อื่น และแสดงความ คิดเห็น	อย่างน้อยร้อยละ 30 ของทีม ยอมรับฟังความคิดเห็นของผู้อื่น และแสดงความคิดเห็น	ไม่ยอมรับฟังความคิดเห็นของผู้อื่นและไม่แสดงความคิดเห็นกันภายในทีม

6 การเก็บรวบรวมข้อมูลและการวิเคราะห์ข้อมูล

6.1 การเก็บรวบรวมข้อมูล

1. การเก็บรวบรวมข้อมูลก่อนเรียน ผู้วิจัยได้เก็บรวบรวมข้อมูลจากนักเรียนกลุ่มตัวอย่างโดยใช้แบบวัดทักษะทักษะการแก้ปัญหาทางคณิตศาสตร์ก่อนเรียน จำนวน 5 ข้อ ที่ผู้วิจัยได้สร้างขึ้นเพื่อตรวจสอบทักษะการแก้ปัญหาทางคณิตศาสตร์ของนักเรียนก่อนการจัดการเรียนรู้

2. การเก็บรวบรวมข้อมูลหลังเรียน ผู้วิจัยได้เก็บรวบรวมข้อมูลจากนักเรียนกลุ่มตัวอย่าง เป็นการวัดทักษะการแก้ปัญหาทางคณิตศาสตร์หลังเรียน โดยใช้แบบทดสอบวัดทักษะทางคณิตศาสตร์หลังเรียน จำนวน 5 ข้อ โดยแบบทดสอบเป็นชุดเดียวกันกับแบบทดสอบวัดทักษะการแก้ปัญหาทางคณิตศาสตร์ก่อนเรียน

6.2 การวิเคราะห์ข้อมูล

1. การวิเคราะห์ข้อมูลเชิงปริมาณ พิจารณาจากแบบทดสอบวัดทักษะการแก้ปัญหาทางคณิตศาสตร์ เรื่อง พื้นที่ผิวและปริมาตร แบบประเมินการทำงานเป็นทีม โดยใช้สถิติวิเคราะห์ ได้แก่ ร้อยละ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และการทดสอบที (t-test)

2. เกณฑ์การประเมินทักษะการแก้ปัญหาทางคณิตศาสตร์ โดยคิดเป็นร้อยละจากคะแนนรวมที่ได้ในแบบทดสอบวัดทักษะการแก้ปัญหาทางคณิตศาสตร์ ซึ่งมีคะแนนเต็ม 60 คะแนน โดยใช้เกณฑ์ในการประเมินดังนี้

ร้อยละ 70 ขึ้นไป	อยู่ในระดับดี
ร้อยละ 50 - 69	อยู่ในระดับพอใช้
ต่ำกว่าร้อยละ 50	อยู่ในระดับปรับปรุง

3. เกณฑ์การประเมินการทำงานเป็นทีม โดยคิดเป็นร้อยละจากคะแนนรวมที่ได้จากแบบประเมินการทำงานเป็นทีมทั้งหมด 5 สัปดาห์ ซึ่งมีคะแนนเต็ม 75 คะแนน โดยใช้เกณฑ์ในการประเมินดังนี้

ร้อยละ 80 ขึ้นไป	อยู่ในระดับดีมาก
ร้อยละ 60 - 79	อยู่ในระดับดี
ร้อยละ 41 – 59	อยู่ในระดับพอใช้
ต่ำกว่าร้อยละ 40	อยู่ในระดับปรับปรุง

6.3 ตัวอย่างแผนการจัดการเรียนรู้

ในงานวิจัยจะขอนำเสนอแผนการจัดการเรียนรู้ เรื่อง พื้นที่ผิวและปริมาตรของปริซึม ซึ่งมีขั้นตอนดังหัวข้อ 5.1 ซึ่งมีกิจกรรมการเรียนรู้ ดังนี้

ขั้นที่ 1 กำหนดปัญหา ครูจัดสถานการณ์ กระตุ้นให้นักเรียนเรียนเกิดความสนใจ และมองเห็นปัญหาสามารถกำหนดสิ่งที่เป็นปัญหาที่ผู้เรียนอยากรู้อยากเรียนได้และเกิดความสนใจที่จะค้นหาคำตอบ ดังนี้

สถานการณ์
นักเรียนต้องการออกแบบบรรจุภัณฑ์รูปทรงปริซึมเพื่อนำไปบรรจุช็อกโกแลตจำนวน 6 ชิ้น โดยรูปทรงของช็อกโกแลตจะประกอบไปด้วย ปริซึมฐานสามเหลี่ยมด้านเท่า 2 ชิ้น ปริซึมฐานสี่เหลี่ยมจัตุรัส 2 ชิ้น และปริซึมฐานหกเหลี่ยมด้านเท่า 2 ชิ้น โดยแต่ละด้านจะมีความยาวด้านเท่ากับ 2 เซนติเมตร จากนั้นต้องนำบรรจุภัณฑ์ที่ได้มาห่อกระดาษของขวัญ นักเรียนจะต้องได้กระดาษของขวัญอย่างน้อยกี่ตารางเซนติเมตรจึงจะเพียงพอต่อการห่อของขวัญ

ขั้นที่ 2 ทำความเข้าใจปัญหา นักเรียนภายในทีมช่วยวิเคราะห์ว่าโจทย์กำหนดอะไรมาให้บ้าง และสิ่งที่โจทย์ต้องการสิ่งใด และต้องใช้อุปกรณ์ใดบ้างในการทำ

ขั้นที่ 3 ดำเนินการศึกษาค้นคว้า นักเรียนสืบค้นข้อมูลจากใบความรู้และจากแหล่งอื่น ๆ เพื่อนำมาใช้ในการแก้ปัญหาที่กำหนด

ขั้นที่ 4 สังเคราะห์ความรู้ นักเรียนภายในทีมนำข้อมูลที่ได้มาแลกเปลี่ยนกันว่าข้อมูลที่หามาได้เพียงพอต่อการแก้ปัญหาในสถานการณ์ดังกล่าวหรือไม่ หากไม่เพียงพอให้ดำเนินการสืบค้นเพิ่มเติม และดำเนินการออกแบบบรรจุภัณฑ์ที่กำหนด

ขั้นที่ 5 สรุปและประเมินค่าของคำตอบ นักเรียนช่วยกันประเมินผลงานภายในทีมของตนเองว่ามีความเหมาะสมหรือไม่ และตรงกับสถานการณ์ที่กำหนดให้หรือไม่ จากนั้นสรุปในใบกิจกรรมการเรียนรู้

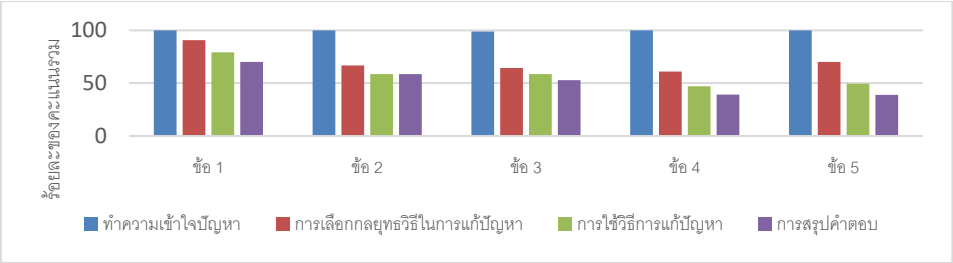
ขั้นที่ 6 นำเสนอและประเมินผลงาน นักเรียนแต่ละทีมนำเสนอผลงานที่ทีมตนเองสร้างขึ้นมาจากนั้นเพื่อนในชั้นเรียนช่วยกันประเมินผลงานของทีมที่ออกมานำเสนอ

7 ผลการวิจัย

7.1 ผลการวิเคราะห์ทักษะการแก้ปัญหาทางคณิตศาสตร์ เรื่อง พื้นที่ผิวและปริมาตร โดยใช้การจัดการเรียนรู้แบบปัญหาเป็นฐาน

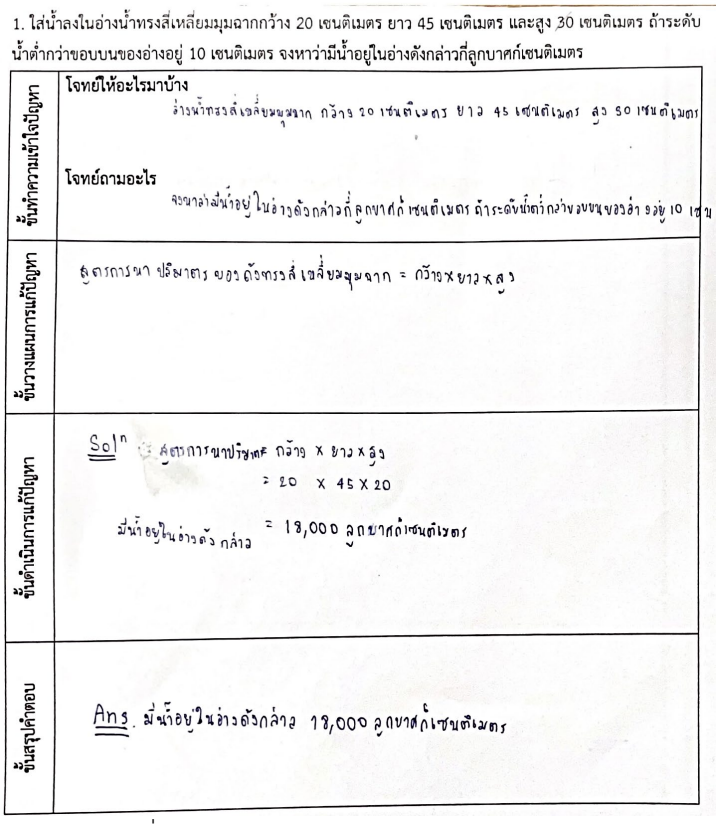
ผลการวิเคราะห์ทักษะการแก้ปัญหาทางคณิตศาสตร์ เรื่อง พื้นที่ผิวและปริมาตร จากแบบทดสอบวัดทักษะการแก้ปัญหาทางคณิตศาสตร์ก่อนและหลังการจัดการเรียนรู้แบบปัญหาเป็นฐาน พบว่า ได้คะแนนเฉลี่ยทักษะการแก้ปัญหาทางคณิตศาสตร์ก่อนการจัดการเรียนรู้ เท่ากับ 27.96 คะแนน และหลังการจัดการเรียนรู้ เท่ากับ 42.13 คะแนน ซึ่งจากการทดสอบสมมติฐานพบว่าหลังการจัดการเรียนรู้สูงกว่าก่อนการจัดการเรียนรู้โดยใช้การจัดการเรียนรู้แบบปัญหาเป็นฐานอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ผลการวิเคราะห์แบบวัดทักษะการแก้ปัญหาทางคณิตศาสตร์ เรื่อง พื้นที่ผิวและปริมาตร หลังการจัดการเรียนรู้แบบปัญหาเป็นฐานเปรียบเทียบกับเกณฑ์ร้อยละ 70 เมื่อเปรียบเทียบกับเกณฑ์ร้อยละ 70 พบว่าคะแนนเฉลี่ยทักษะการแก้ปัญหาทางคณิตศาสตร์หลังการจัดการเรียนรู้สูงกว่าเกณฑ์ร้อยละ 70 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และผ่านเกณฑ์ร้อยละ 70 จำนวน 15 คน คิดเป็นร้อยละ 51.72 และจากวิเคราะห์เป็นรายด้านของร้อยละคะแนนทักษะการแก้ปัญหาทางคณิตศาสตร์หลังการจัดการเรียนรู้โดยใช้ปัญหาเป็นฐานแสดงดังภาพที่ 2



ภาพที่ 2 วิเคราะห์เป็นรายด้านของทักษะการแก้ปัญหาทางคณิตศาสตร์หลังการจัดการเรียนรู้โดยใช้ปัญหาเป็นฐาน

จากภาพที่ 2 พบว่า ในแต่ละข้อนักเรียนสามารถทำความเข้าใจปัญหาได้ จากภาพจะพบว่านักเรียนส่วนใหญ่ยังไม่สามารถเลือกกลยุทธวิธีในการแก้ปัญหาและใช้วิธีการแก้ปัญหาได้อย่างครบถ้วนจึงทำให้ไม่สามารถสรุปคำตอบได้อย่างถูกต้อง


1. ใส่น้ำลงในอ่างน้ำทรงสี่เหลี่ยมมุมฉากกว้าง 20 เซนติเมตร ยาว 45 เซนติเมตร และสูง 30 เซนติเมตร ถ้าระดับน้ำต่ำกว่าขอบบนของอ่างอยู่ 10 เซนติเมตร จงหาว่ามีน้ำอยู่ในอ่างดังกล่าวกี่ลูกบาศก์เซนติเมตร

ขั้นทำความเข้าใจปัญหา
โจทย์ให้อะไรมาบ้าง
อ่างน้ำทรงสี่เหลี่ยมมุมฉาก กว้าง 20 เซนติเมตร ยาว 45 เซนติเมตร สูง 30 เซนติเมตร
โจทย์ถามอะไร
จงหาว่ามีน้ำอยู่ในอ่างดังกล่าวกี่ลูกบาศก์เซนติเมตร ถ้าระดับน้ำต่ำกว่าขอบบนของอ่าง อยู่ 10 เซน

ขั้นวางแผนการแก้ปัญหา
สูตรการหา ปริมาตร ของทรงสี่เหลี่ยมมุมฉาก = กว้าง x ยาว x สูง

ขั้นดำเนินการแก้ปัญหา
Soln สูตรการหาปริมาตร กว้าง x ยาว x สูง
= 20 x 45 x 20
มีน้ำอยู่ในอ่างดังกล่าว = 18,000 ลูกบาศก์เซนติเมตร

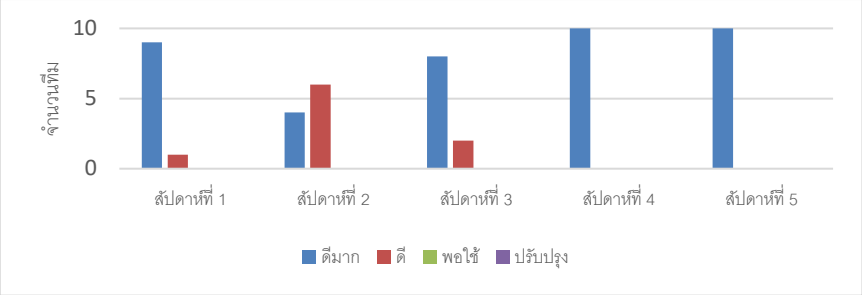
ขั้นสรุปคำตอบ
Ans. มีน้ำอยู่ในอ่างดังกล่าว 18,000 ลูกบาศก์เซนติเมตร

ภาพที่ 3 ตัวอย่างนักเรียนที่ได้คะแนนจากแบบวัดทักษะการแก้ปัญหาทางคณิตศาสตร์หลังเรียน เรื่อง พื้นที่ผิวและปริมาตร อยู่ในระดับดี

จากภาพที่ 3 ตัวอย่างนักเรียนที่ได้คะแนนจากแบบวัดทักษะการแก้ปัญหาทางคณิตศาสตร์หลังเรียน เรื่อง พื้นที่ผิวและปริมาตร อยู่ในระดับคุณภาพดีทุกรายการประเมิน

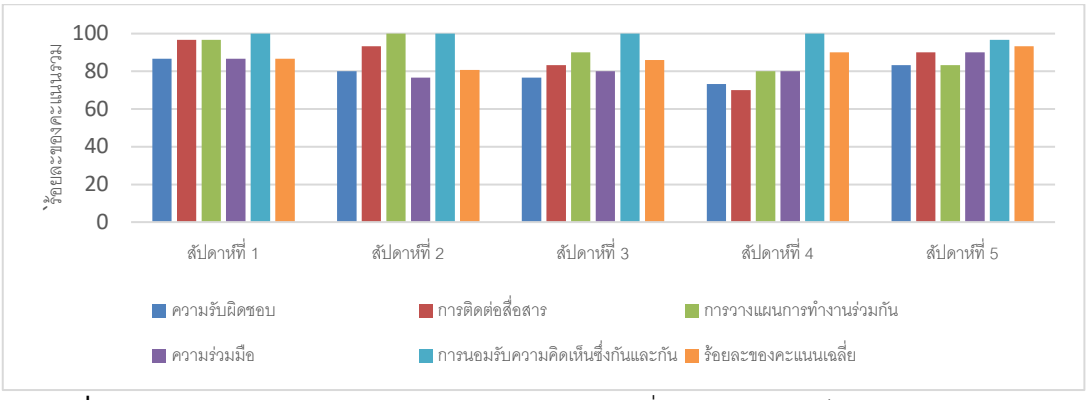
7.2 ผลการวิเคราะห์การทำงานเป็นทีม

ผลการวิเคราะห์เพื่อศึกษาการทำงานเป็นทีมของนักเรียนระดับประกาศนียบัตรวิชาชีพชั้นปีที่ 1 โดยใช้การจัดการเรียนรู้แบบปัญหาเป็นฐาน ผลการวิจัยพบว่า การจัดการเรียนรู้โดยใช้ปัญหาเป็นฐานสามารถทำให้นักเรียนทำงานเป็นทีมอยู่ในเกณฑ์ที่ดีและดีมาก ซึ่งแสดงในภาพที่ 4



ภาพที่ 4 ผลการวิเคราะห์การทำงานเป็นทีมเป็นรายสัปดาห์

ผลจากการวิเคราะห์เป็นรายองค์ประกอบทั้งหมด 5 องค์ประกอบ คือ ด้านความรับผิดชอบ ด้านการติดต่อสื่อสาร ด้านการวางแผนการทำงานร่วมกัน ด้านความร่วมมือ และด้านการยอมรับความคิดเห็นซึ่งกันและกัน พบว่า นักเรียนส่วนใหญ่มีการยอมรับยอมรับความคิดเห็นซึ่งกันและกันมากที่สุด ดังแสดงในภาพที่ 5



ภาพที่ 5 ร้อยละของคะแนนรายองค์ประกอบและคะแนนเฉลี่ยในการทำงานเป็นทีมของแต่ละสัปดาห์

จากภาพที่ 5 จะเห็นว่าคะแนนเฉลี่ยการทำงานเป็นทีมในสัปดาห์ที่ 1 ถึง สัปดาห์ที่ 5 อยู่ในเกณฑ์ดีมาก แต่ในสัปดาห์ที่ 2 และสัปดาห์ที่ 3 พบว่าคะแนนเฉลี่ยการทำงานเป็นทีมได้คะแนนลดลง เนื่องจากโจทย์ปัญหาในสัปดาห์ที่ 2 คือ “ให้นักเรียนสร้างพีระมิดที่มีปริมาตรอย่างน้อย 50 ถึง 70 ลูกบาศก์เซนติเมตร และมีพื้นที่ผิวทั้งหมดอย่างน้อย 120 ถึง 150 ตารางเซนติเมตร” ซึ่งส่วนใหญ่สร้างได้ไม่ตรงกับโจทย์ที่กำหนดและโจทย์ข้อนี้มีขั้นตอนการแก้ปัญหาที่ค่อนข้างซับซ้อนจึงส่งผลให้นักเรียนมีส่วนร่วมในการทำงานเป็นทีมน้อย จึงทำให้มีคะแนนเฉลี่ยของการทำงานเป็นทีมน้อยที่สุด จากตัวอย่างโจทย์ปัญหาดังกล่าวข้างต้นการที่นักเรียนในหลายทีมไม่สามารถแก้ปัญหาจากปัญหาเป็นฐานในเรื่อง พื้นที่ผิวและปริมาตรของพีระมิด ได้ เช่น นักเรียนไม่เข้าใจปัญหาและไม่สามารถสร้างรูปทรงพีระมิดที่มีปริมาตรและพื้นที่ผิวตามที่กำหนดได้ ดังภาพที่ 6

ภาพที่ 6 ตัวอย่างผลงานนักเรียนจากปัญหาเป็นฐานในเรื่อง พื้นที่ผิวและปริมาตรของพีระมิด

8 สรุปผลการวิจัย

จากการวิจัยเรื่อง การพัฒนาทักษะการแก้ปัญหาทางคณิตศาสตร์และการทำงานเป็นทีมของนักเรียนระดับประกาศนียบัตรวิชาชีพชั้นปีที่ 1 เรื่อง พื้นที่ผิวและปริมาตร โดยใช้การจัดการเรียนรู้แบบปัญหาเป็นฐาน สรุปผลการวิจัยได้ดังนี้

1. นักเรียนที่ได้การจัดการเรียนรู้โดยใช้ปัญหาเป็นฐานมีทักษะการแก้ปัญหาทางคณิตศาสตร์ เรื่อง พื้นที่ผิวและปริมาตร หลังเรียนสูงกว่าก่อนการจัดการเรียนรู้อย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ .05

2. นักเรียนที่ได้การจัดการเรียนรู้โดยใช้ปัญหาเป็นฐานมีทักษะการแก้ปัญหาทางคณิตศาสตร์ เรื่อง พื้นที่ผิวและปริมาตร สูงกว่าเกณฑ์ร้อยละ 70 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

3. นักเรียนที่ได้การจัดการเรียนรู้โดยใช้ปัญหาเป็นฐานมีการทำงานเป็นทีมในภาพรวมอยู่ในเกณฑ์ดีถึงดีมาก

9 อภิปรายผลการวิจัย

จากผลการทดสอบก่อนและหลังการจัดการเรียนรู้โดยใช้ปัญหาเป็นฐาน พบว่า มีทักษะการแก้ปัญหาทางคณิตศาสตร์ เรื่อง พื้นที่ผิวและปริมาตร หลังการจัดจัดการเรียนรู้โดยใช้ปัญหาเป็นฐานสูงกว่าก่อนการจัดการเรียนรู้โดยใช้ปัญหาเป็นฐาน และผ่านเกณฑ์ร้อยละ 70 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อพิจารณาคะแนนนักเรียนที่มาจากในแต่ละขั้นตอนของการแก้ปัญหา จะเห็นได้ว่านักเรียนทำคะแนนในขั้นทำความเข้าใจปัญหาได้มากที่สุด และทำคะแนนในขั้นวางแผนการแก้ปัญหาได้น้อยที่สุด อาจเป็นเพราะการวางแผนการแก้ปัญหาวางแผนไม่ครบถ้วนและถ้าในขั้นดำเนินการแก้ปัญหานักเรียนคำนวณผิดพลาดซึ่งเกิดจากความเข้าใจที่คลาดเคลื่อนของนักเรียน อาจส่งผลให้ขั้นสรุปคำตอบผิดพลาดด้วยเช่นกัน และถึงแม้นักเรียนจะมีคะแนนในขั้นวางแผนการแก้ปัญหาน้อยที่สุด แต่ผลการทดสอบก็ยังผ่านเกณฑ์ร้อยละ 70 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นั่นคือนักเรียนส่วนใหญ่มีทักษะการแก้ปัญหาทางคณิตศาสตร์อยู่ในระดับดี แสดงว่าการจัดการเรียนรู้โดยใช้ปัญหาเป็นฐาน ในเรื่องพื้นที่ผิวและปริมาตร สามารถทำให้นักเรียนมีทักษะการแก้ปัญหาทางคณิตศาสตร์ได้ ผู้วิจัยคิดว่าอาจเป็นเพราะการจัดการเรียนรู้โดยใช้ปัญหาเป็นฐานเป็นรูปแบบการจัดการเรียนรู้ที่เป็นขั้นตอนที่ทำให้นักเรียนได้ร่วมกันได้วิเคราะห์ปัญหาจากสถานการณ์ที่กำหนด ช่วยกันวางแผนการแก้ปัญหา ดำเนินการศึกษาค้นคว้าเพิ่มเติมเพื่อที่จะนำไปใช้ในการแก้ปัญหา อีกทั้งยังได้ร่วมกันสรุปคำตอบจากการแก้ปัญหาอีกด้วย นอกจากนี้การฝึกให้นักเรียนมีทักษะในการแก้ปัญหาอย่างเป็นระบบและเปิดโอกาสให้นักเรียนรู้จักแก้ปัญหาด้วยตนเองให้มากที่สุด จะทำให้นักเรียนมีความเชื่อมั่นในการแก้ปัญหา สอดคล้องกับงานวิจัยของพรทิพา เมืองโคตร และคณะ (2559) ได้ศึกษาความสามารถในการแก้ปัญหาทางคณิตศาสตร์โดยการจัดการเรียนรู้และโดยใช้ปัญหาเป็นฐานของนักเรียนชั้นมัธยมศึกษาปีที่ 3 เรื่อง พื้นที่ผิวและปริมาตร ผลการวิจัยพบว่า 1) ร้อยละ 83.33 ของนักเรียนที่เรียนโดยการจัดการเรียนรู้โดยใช้ปัญหาเป็นฐาน มีความสามารถในการแก้ปัญหาทางคณิตศาสตร์ ผ่านเกณฑ์ร้อยละ 75 ของคะแนนเต็ม 2) นักเรียนที่เรียนโดยการจัดการเรียนรู้โดยใช้ปัญหาเป็นฐาน มีความสามารถในการแก้ปัญหาทางคณิตศาสตร์ สูงกว่านักเรียนที่เรียนโดยการจัดการเรียนรู้แบบปกติ อย่างมีนัยสำคัญ

ทางสถิติที่ระดับ .05 และครองทรัพย์ เป็งขวัญ (2560) ผลการวิจัยพบว่า นักเรียนที่ได้รับการจัดกิจกรรมการเรียนรู้โดยใช้ปัญหาเป็นฐานมีทักษะการแก้ปัญหาทางคณิตศาสตร์หลังเรียนสูงกว่าก่อนเรียน และสูงกว่านักเรียนที่ได้รับการจัดการเรียนรู้แบบปกติ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และจากการศึกษาคะแนนพัฒนาการการทำงานเป็นทีม พบว่า ด้านความรับผิดชอบ ด้านการติดต่อสื่อสาร ด้านการวางแผนการทำงานร่วมกัน ด้านความร่วมมือและ ด้านการยอมรับความคิดเห็นซึ่งกันและกัน พบว่านักเรียนมีการทำงานเป็นทีมอยู่ในระดับดีถึงดีมาก แสดงว่าการจัดจัดการเรียนรู้โดยใช้ปัญหาเป็นฐาน ในเรื่องพื้นที่ผิวและปริมาตร สามารถทำให้นักเรียนมีการทำงานเป็นทีมอยู่ในระดับดีถึงดีมากได้

กิตติกรรมประกาศ งานวิจัยครั้งนี้สำเร็จลุล่วงได้ด้วยดี เพราะผู้วิจัยได้รับความช่วยเหลือที่ดีและได้รับความรู้อันมีค่ายิ่งจาก ผศ. ดร.ธีระพล สลีวงศ์ อาจารย์ที่ปรึกษา ผู้วิจัยขอกราบขอบพระคุณ ผศ. ดร.วราภรณ์ จาตนิล, ผศ. ดร.อังกูร หวังวงศ์ชัย, นายณัฐกฤษ จันทร์ตะ, นางถาวร ลักษณะ และนายอาคม นาคน้อย ซึ่งเป็นผู้เชี่ยวชาญ ที่กรุณาเสียสละเวลาในการตรวจสอบเครื่องมือและพิจารณาให้ข้อเสนอแนะต่าง ๆ ในการปรับปรุงเครื่องมือให้มีความถูกต้องสมบูรณ์มากยิ่งขึ้น

เอกสารอ้างอิง

[1] ครองทรัพย์ เป็งขวัญ. (2560). *การพัฒนาทักษะการเชื่อมโยงทางคณิตศาสตร์และทักษะการแก้ปัญหาทางคณิตศาสตร์ด้วยการจัดกิจกรรมการเรียนรู้โดยใช้ปัญหาเป็นฐาน*. (วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต, มหาวิทยาลัยอุบลราชธานี).

[2] บันเย็น เพ็งกระจ่าง. (2561). *การพัฒนาครูด้านการเรียนการสอนในศตวรรษที่ 21 ของโรงเรียนสารสาสน์วิเทศคลองหลวง สังกัดสำนักงานคณะกรรมการส่งเสริมการศึกษาเอกชน*. (วิทยานิพนธ์การศึกษาตามหลักสูตรปริญญาศึกษาศาสตรมหาบัณฑิต, มหาวิทยาลัยเกริก).

[3] พรทิพา เมืองโคตร, นงลักษณ์ วิริยะพงษ์ และมนชยา เจียงประดิษฐ์. (2559). ความสามารถในการแก้ปัญหาทางคณิตศาสตร์โดยการจัดการเรียนรู้ โดยใช้ปัญหาเป็นฐานของนักเรียนชั้นมัธยมศึกษาปีที่ 3 เรื่อง พื้นที่ผิวและปริมาตร. วารสารศึกษาศาสตร์, 27(3), 122-132.

[4] วิจารณ์ พานิช. (2555). *วิถีสร้างการเรียนรู้เพื่อศิษย์ในศตวรรษที่ 21* (พิมพ์ครั้งที่ 2). กรุงเทพฯ: มูลนิธิสดศรี สฤษดิ์วงศ์.

[5] วีระยุทธ พรพจน์ธนมาศ. (2565). การศึกษาเปรียบเทียบการตรวจสอบความเที่ยงตรงเชิงเนื้อหาของเครื่องมือวิจัยด้วยเทคนิค IOC, CVR และ CVI. รังสิตสารสนเทศ, 28(1), 169-192.

[6] สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน). (2566). *คู่มือการจัดสอบการทดสอบทางการศึกษาระดับชาติด้านอาชีวศึกษา (Vocational National Educational Test : V-NET) ด้วยระบบดิจิทัล (Digital Testing) ปีการศึกษา 2566*.

[7] สถาบันส่งเสริมวิทยาศาสตร์และเทคโนโลยี กระทรวงศึกษาธิการ. (2551). *คู่มือการวัดผลประเมินผลคณิตศาสตร์* (พิมพ์ครั้งที่ 2). กรุงเทพฯ : ซี เอ็ดยูเคชั่น.

[8] สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน. (2551). *เอกสารหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 แนวปฏิบัติการวัดและประเมินผลการเรียนรู้*. กรุงเทพฯ : สำนักวิชาการและมาตรฐานการศึกษา สำนักงานฯ.

[9] สำนักบริหารงานการมัธยมศึกษาตอนปลาย สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน และกระทรวงศึกษาธิการ. (2558). *แนวทางการจัดทักษะการเรียนรู้ในศตวรรษที่ 21 ที่เน้นสมรรถนะทางสาขาวิชาชีพ*. โรงพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย จำกัด.

[10] Dickinson, T. L., & McIntyre, R. M. (1997). *A conceptual framework for teamwork measurement*.

[11] National Council of Teacher of Mathematics. (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, Virginia: National Council of Teacher of Mathematics.

[12] Nolan, V. (1989). *Teamwork: The syndetic Co*.

[13] Polya G. (1985). *How To Solve It, New York: Henry Houbleday & Company, 1957*.

[14] Tarricone, P., & Luca, J. (2002). *Successful teamwork: A case study*.

[15] Wang, L. et.al. (2009). *Assessing teamwork and collaboration in high school students A multimethod Approach.: Canadian Journal of School Psychology*.

10. Number Theory

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Divisibility Algorithm of Even Number

Itsara Saenjaroen[1, †] and Apisit Pakapongpun[1, ‡]

[1]Department of Mathematics, Faculty of Science, Burapha University, Chonburi 20131, Thailand

Abstract

In this paper, we prove an algorithm for the divisibility of even numbers. Moreover, we expanded this divisibility test and extended the rule to more digit numbers. There will be another set of appropriate values that can be used for the same characteristics as even divisors. This work will show a new perspective on divisibility by even numbers.

Keywords: divisibility, even integer.
2020 MSC: 11B41

1 Introduction

Divisibility by 2 or 5 is straightforward: if a number ends in an even digit, it's divisible by 2. If it ends in 0 or 5, it's divisible by 5. Divisibility by 3 depends on the sum of its digits. However, all these rules have their roots in modulo arithmetic.

In 2019, Alp and Sarikaya [1] established a division process by cutting the rightmost digit. There will be an appropriate value multiplied by a number cut from the dividend's rightmost digit, and then the result is added to the set of numbers remaining after cutting the rightmost digit. If the new result is a multiple of the divisor, then the original dividend will also be a multiple of the divisor. Throughout the years, researchers such as Khosravi et al. [2] gave an algorithm for the divisibility of numbers. They supposed that $a_n a_{n-1} \dots a_2 a_1$ is an integer dividend and $b_m b_{m-1} \dots b_1$ is a prime divisor. If the difference between $\left(b_m b_{m-1} \dots b_2\right)\left(a_1\right)$ and $a_n a_{n-1} \dots a_2 a_1$ can be divided by the original divisor, then the original dividend is divisible as well. In 2021, Tiebekabe and Diouïf [3] demonstrated the divisibility test of 7 proposed by Chika, a rule for solving number divisibility, especially 7. Chika realized this rule but did not know the analytical proof of it. Throughout this paper, we prove a rule for the divisibility of even integers.

† Speaker. ‡ Corresponding author.
E - mail address: apisit.buu@gmail.com (A. Pakapongpun).

2 Main Results

Theorem 2.1. Let $\ldots a_5a_4a_3a_2a_1 = x$ and p be even natural numbers with $p = 2 \cdot p_1$, $p \neq 2$ and $10 \nmid p$. Then, there is a rational number n_1 so that

$$x \equiv 0 \pmod{p} \text{ if and only if } \ldots a_5a_4a_3a_2 + n_1 \cdot a_1 \equiv 0 \pmod{p_1}.$$

Proof: Let $x \equiv 0 \pmod{p}$ be true. There is an integer k such that

$$\ldots a_5a_4a_3a_2a_1 = 10 \cdot (\ldots a_5a_4a_3a_2) + a_1 = p \cdot k.$$

Thus
$$\ldots a_5a_4a_3a_2 = \frac{p \cdot k - a_1}{10}.$$

On the other hand, there is a rational number $n_1 = \frac{n_0}{2}$ where $n_0 \in \mathbb{Z}$ such that

$$\ldots a_5a_4a_3a_2 + n_1 \cdot a_1 \in \mathbb{Z}.$$

Thus
$$\begin{aligned} \ldots a_5a_4a_3a_2 + n_1 \cdot a_1 &= \frac{p \cdot k - a_1}{10} + n_1 \cdot a_1 \\ &= \frac{2 \cdot p_1 \cdot k - 2 \cdot a_0}{10} + n_1 \cdot 2 \cdot a_0 \\ &= \frac{p_1 \cdot k - a_0 + 10 \cdot n_1 \cdot a_0}{5} \\ &= \frac{p_1 \cdot k + (10 \cdot n_1 - 1) \cdot a_0}{5}, \end{aligned}$$

where $a_1 = 2 \cdot a_0$, $a_0 \in \mathbb{Z}$.

At least one rational number $n_1 = \frac{p_1 \cdot t + 1}{10}$ for some $t \in \mathbb{Z}$. Since, p_1 is a natural number with $p = 2 \cdot p_1$, $p \neq 2$ and $10 \nmid p$ that is $5 \nmid p_1$.

If $p \equiv 2 \pmod{10}$ and $t = 9$, then $n_1 = \frac{9 \cdot p_1 + 1}{10}$.

If $p \equiv 4 \pmod{10}$ and $t = 7$, then $n_1 = \frac{7 \cdot p_1 + 1}{10}$.

If $p \equiv 6 \pmod{10}$ and $t = 3$, then $n_1 = \frac{3 \cdot p_1 + 1}{10}$.

If $p \equiv 8 \pmod{10}$ and $t = 1$, then $n_1 = \frac{1 \cdot p_1 + 1}{10}$.

Therefore, $t =$ 1, 3, 7 or 9 and $n_1 = \frac{p_1 \cdot t + 1}{10}$ is a rational number.

Thus
$$\begin{aligned} \ldots a_5a_4a_3a_2 + n_1 \cdot a_1 &= \frac{p_1 \cdot k + (10 \cdot n_1 - 1) \cdot a_0}{5} \\ &= \frac{p_1 \cdot k + p_1 \cdot t \cdot a_0}{5} \\ &= p_1 \cdot \left(\frac{k + t \cdot a_0}{5} \right) \in \mathbb{Z}. \end{aligned}$$

Hence, $\left(\frac{k+t\cdot a_0}{5}\right)\in\mathbb{Z}$ since $5\nmid p_1$. Therefore, $\ldots a_5a_4a_3a_2+n_1\cdot a_1\equiv 0\pmod{p_1}$.

Conversely, there is a rational number $n_1=\frac{n_0}{2}$ for some $n_0\in\mathbb{Z}$, such that $\ldots a_5a_4a_3a_2+n_1\cdot a_1\equiv 0\pmod{p_1}$.

That is $10\cdot\left(\ldots a_5a_4a_3a_2+n_1\cdot a_1\right)\equiv 0\pmod{p}$

so $\ldots a_5a_4a_3a_2a_1+\left(10\cdot n_1-1\right)\cdot a_1\equiv 0\pmod{p}$.

Choosing $n_1=\frac{p_1\cdot t+1}{10}$ for some $t\in\mathbb{Z}$, we have

$$\ldots a_5a_4a_3a_2a_1+\left(p_1\cdot t\right)\cdot 2\cdot a_0\equiv 0\pmod{p}$$
$$\ldots a_5a_4a_3a_2a_1+p\cdot t\cdot a_0\equiv 0\pmod{p}.$$

Thus $$\ldots a_5a_4a_3a_2a_1\equiv 0\pmod{p}.$$

The proof is completed.

For all even natural number p in Theorem 2.1, $\ldots a_5a_4a_3a_2a_1=x$ is an even number.

If $p=2\cdot p_1$, $p\neq 2$, $10\nmid p$ and $n_1=\frac{p_1\cdot t+1}{10}$, then

Table 1: The relationship between p , p_1, t and n_1

p	p_1	t	n_1	$x\equiv 0\pmod{p}\Leftrightarrow \ldots a_5a_4a_3a_2+n_i\cdot a_1\equiv 0\pmod{p_1}$
4	2	7	$3/2$	$x\equiv 0\pmod{4}\Leftrightarrow \ldots a_3a_2+\frac{3}{2}\cdot a_1\equiv 0\pmod{2}$
6	3	3	1	$x\equiv 0\pmod{6}\Leftrightarrow \ldots a_3a_2+1\cdot a_1\equiv 0\pmod{3}$
8	4	1	$1/2$	$x\equiv 0\pmod{8}\Leftrightarrow \ldots a_3a_2+\frac{1}{2}\cdot a_1\equiv 0\pmod{4}$
12	6	9	$11/2$	$x\equiv 0\pmod{12}\Leftrightarrow \ldots a_3a_2+\frac{11}{2}\cdot a_1\equiv 0\pmod{6}$
14	7	7	5	$x\equiv 0\pmod{14}\Leftrightarrow \ldots a_3a_2+5\cdot a_1\equiv 0\pmod{7}$
16	8	3	$5/2$	$x\equiv 0\pmod{16}\Leftrightarrow \ldots a_3a_2+\frac{5}{2}\cdot a_1\equiv 0\pmod{8}$
18	9	1	1	$x\equiv 0\pmod{18}\Leftrightarrow \ldots a_3a_2+1\cdot a_1\equiv 0\pmod{9}$
22	11	9	10	$x\equiv 0\pmod{22}\Leftrightarrow \ldots a_3a_2+10\cdot a_1\equiv 0\pmod{11}$
24	12	7	$17/2$	$x\equiv 0\pmod{24}\Leftrightarrow \ldots a_3a_2+\frac{17}{2}\cdot a_1\equiv 0\pmod{12}$
26	13	3	4	$x\equiv 0\pmod{26}\Leftrightarrow \ldots a_3a_2+4\cdot a_1\equiv 0\pmod{13}$
28	14	1	$3/2$	$x\equiv 0\pmod{28}\Leftrightarrow \ldots a_3a_2+\frac{3}{2}\cdot a_1\equiv 0\pmod{14}$
32	16	9	$29/2$	$x\equiv 0\pmod{32}\Leftrightarrow \ldots a_3a_2+\frac{29}{2}\cdot a_1\equiv 0\pmod{16}$
34	17	7	12	$x\equiv 0\pmod{34}\Leftrightarrow \ldots a_3a_2+12\cdot a_1\equiv 0\pmod{17}$

p	p_1	t	n_1	$x \equiv 0 \pmod{p} \Leftrightarrow \ldots a_5a_4a_3a_2 + n_i \cdot a_1 \equiv 0 \pmod{p_1}$
36	18	3	$11/2$	$x \equiv 0 \pmod{36} \Leftrightarrow \ldots a_3a_2 + 11/2 \cdot a_1 \equiv 0 \pmod{18}$
38	19	1	2	$x \equiv 0 \pmod{38} \Leftrightarrow \ldots a_3a_2 + 2 \cdot a_1 \equiv 0 \pmod{19}$
42	21	9	19	$x \equiv 0 \pmod{42} \Leftrightarrow \ldots a_3a_2 + 19 \cdot a_1 \equiv 0 \pmod{21}$
44	22	7	$31/2$	$x \equiv 0 \pmod{44} \Leftrightarrow \ldots a_3a_2 + 31/2 \cdot a_1 \equiv 0 \pmod{22}$
46	23	3	7	$x \equiv 0 \pmod{46} \Leftrightarrow \ldots a_3a_2 + 7 \cdot a_1 \equiv 0 \pmod{23}$
48	24	1	$5/2$	$x \equiv 0 \pmod{48} \Leftrightarrow \ldots a_3a_2 + 5/2 \cdot a_1 \equiv 0 \pmod{24}$
52	26	9	$47/2$	$x \equiv 0 \pmod{52} \Leftrightarrow \ldots a_3a_2 + 47/2 \cdot a_1 \equiv 0 \pmod{26}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots \qquad \vdots \qquad \vdots$

Example 2.2. We show that $75624 \equiv 0 \pmod{24}$.

If $p = 24 = 2 \cdot 12$, then $p_1 = 12$, $t = 7$ and $n_1 = \dfrac{7 \cdot 12 + 1}{10} = \dfrac{85}{10} = \dfrac{17}{2}$.

By Theorem 2.1, we have that

$$75624 \equiv 0 \pmod{24} \text{ if and only if } 7562 + \left(\frac{17}{2}\right) \cdot 4 \equiv 7596 \equiv 0 \pmod{12}. \tag{1}$$

If $p = 12 = 2 \cdot 6$, then $p_1 = 6$, $t = 9$ and $n_1 = \dfrac{9 \cdot 6 + 1}{10} = \dfrac{55}{10} = \dfrac{11}{2}$.

$$\text{Thus } 7596 \equiv 0 \pmod{12} \text{ if and only if } 759 + \left(\frac{11}{2}\right) \cdot 6 \equiv 792 \equiv 0 \pmod{6}. \tag{2}$$

Using (1), (2) and the fact that $6 \mid 792$ we conclude that $75624 \equiv 0 \pmod{24}$.

Corollary 2.3. Let $\ldots a_5a_4a_3a_2a_1 = x$ and p be even natural numbers with $p \neq 2$, $10 \nmid p$, $p = 2 \cdot p_1$ and $\gcd(2, p_1) = 1$. There is $n_1 \in \mathbb{Z}$ so that

$$x \equiv 0 \pmod{p} \text{ if and only if } \ldots a_5a_4a_3a_2 + n_1 \cdot a_1 \equiv 0 \pmod{p_1}.$$

Proof: By Theorem 2.1, there is a rational number $n_1 = \dfrac{p_1 \cdot t + 1}{10}$ for some $t \in \mathbb{Z}$ such that $x \equiv 0 \pmod{p}$ if and only if $\ldots a_5a_4a_3a_2 + n_1 \cdot a_1 \equiv 0 \pmod{p_1}$. It remains to show that $n_1 \in \mathbb{Z}$. Since, $p = 2 \cdot p_1$, $10 \nmid p$ and $\gcd(2, p_1) = 1$, we have $\gcd(p_1, 10) = 1$. Then, there exists an integer t such that $p_1 \cdot t \equiv -1 \pmod{10}$ so that $n_1 = \dfrac{p_1 \cdot t + 1}{10} \in \mathbb{Z}$ as desired.

Example 2.4. We show that $292448 \equiv 0 \pmod{74}$.

If $p = 74 = 2 \cdot 37$, then $p_1 = 37$, $t = 7$ and $n_1 = \dfrac{7 \cdot 37 + 1}{10} = \dfrac{260}{10} = 26$.

By Corollary 2.3, we have that

$292448 \equiv 0\left(\mathrm{mod}\,74\right)$ if and only if $29244 + \left(26\right)\cdot 8 \equiv 29452 \equiv 0\left(\mathrm{mod}\,37\right)$.

We can check 37 is a factor of 29452 from Theorem 3 in [1],

$$29452 \equiv 294 + \left(10\right)\cdot 52 \equiv 814 \equiv 0\left(\mathrm{mod}\,37\right).$$

Theorem 2.5. Let $\ldots a_5a_4a_3a_2a_1 = x$ and p be even natural numbers with $p \neq 2$, $10 \nmid p$, $p = 2\cdot p_1$ and $\gcd\left(2, p_1\right) = 1$. There are integer numbers n_1 and n_2 with $n_2 \equiv n_1^2\left(\mathrm{mod}\,p_1\right)$ and $n_2 < p_1$ so that

$$x \equiv 0\left(\mathrm{mod}\,p\right) \text{ if and only if } \ldots a_5a_4a_3 + n_2\cdot a_2a_1 \equiv 0\left(\mathrm{mod}\,p_1\right).$$

Proof: From Corollary 2.3, there is an integer $n_1 = \dfrac{p_1\cdot t + 1}{10}$ for some $t \in \mathbb{Z}$ such that

$$x \equiv 0\left(\mathrm{mod}\,p\right) \text{ if and only if } \ldots a_5a_4a_3a_2 + n_1\cdot a_1 \equiv 0\left(\mathrm{mod}\,p_1\right).$$

Assume that $x \equiv 0\left(\mathrm{mod}\,p\right)$. So, there is an integer k such that

$$\ldots a_5a_4a_3a_2a_1 \;=\; 100\cdot\left(\ldots a_5a_4a_3\right) + a_2a_1 \;=\; p\cdot k.$$

Thus $$\ldots a_5a_4a_3 \;=\; \frac{p\cdot k - a_2a_1}{100}.$$

On the other hand, there is an integer $n_2 = n_1^2 + p_1\cdot k_1$, $k_1 \in \mathbb{Z}$ such that

$$\ldots a_5a_4a_3 + n_2\cdot a_2a_1 \in \mathbb{Z}.$$

Let $a_2a_1 = 2\cdot a_0$, $a_0 \in \mathbb{Z}$ and we have $10n_1 = p_1\cdot t + 1$.

Since $$\ldots a_5a_4a_3 \;=\; \frac{p\cdot k - a_2a_1}{100},$$

$$\begin{aligned}
\ldots a_5a_4a_3 + n_2\cdot a_2a_1 &= \frac{2\cdot p_1\cdot k - 2\cdot a_0}{100} + n_2\cdot 2\cdot a_0\\
&= \frac{p_1\cdot k - a_0 + 100\cdot n_2\cdot a_0}{50}\\
&= \frac{p_1\cdot k + \left(100\cdot n_2 - 1\right)\cdot a_0}{50}\\
&= \frac{p_1\cdot k + \left[100\cdot\left(n_1^2 + p_1\cdot k_1\right) - 1\right]\cdot a_0}{50}\\
&= \frac{p_1\cdot\left(k + 100\cdot k_1\cdot a_0\right) + \left(100\cdot n_1^2 - 1\right)\cdot a_0}{50}\\
&= \frac{p_1\cdot\left(k + 100\cdot k_1\cdot a_0\right) + \left(10\cdot n_1 - 1\right)\left(10\cdot n_1 + 1\right)\cdot a_0}{50}\\
&= \frac{p_1\cdot\left(k + 100\cdot k_1\cdot a_0\right) + \left(p_1\cdot t\right)\left(10\cdot n_1 + 1\right)\cdot a_0}{50}\\
&= p_1\cdot\left(\frac{k + \left[100\cdot k_1 + t\cdot\left(10\cdot n_1 + 1\right)\right]\cdot a_0}{50}\right) \in \mathbb{Z}.
\end{aligned}$$

Hence, $\left(\dfrac{k+\left[100\cdot k_1+t\cdot\left(10\cdot n_1+1\right)\right]\cdot a_0}{50}\right)\in\mathbb{Z}$ since $5\nmid p_1$ and $\gcd\left(2,p_1\right)=1$.

Therefore,

$$\ldots a_5a_4a_3+n_2\cdot a_2a_1\equiv 0\left(\bmod p_1\right).$$

Conversely, assume that $\ldots a_5a_4a_3a_2+n_1\cdot a_1\equiv 0\left(\bmod p_1\right)$.

Thus,
$$\begin{aligned} p_1 &\mid \ldots a_5a_4a_3+n_2\cdot a_2a_1\\ p &\mid 100\cdot\left(\ldots a_5a_4a_3+n_2\cdot a_2a_1\right)\end{aligned}$$

that is
$$\begin{aligned} 100\cdot\left(\ldots a_5a_4a_3+n_2\cdot a_2a_1\right) &\equiv 0\left(\bmod p\right)\\ \ldots a_5a_4a_300+100\cdot n_2\cdot a_2a_1 &\equiv 0\left(\bmod p\right)\\ \ldots a_5a_4a_3a_2a_1+\left(100\cdot n_2-1\right)\cdot a_2a_1 &\equiv 0\left(\bmod p\right).\end{aligned}$$

Choosing $n_2\equiv n_1^2\left(\bmod p_1\right)$ and Theorem 2.1, we know that $10\cdot n_1-1=p_1\cdot t$, $t\in\mathbb{Z}$.

So,
$$\begin{aligned} \ldots a_5a_4a_3a_2a_1+\left(100\cdot n_1^2-1\right)\cdot a_2a_1 &\equiv 0\left(\bmod p\right)\\ \ldots a_5a_4a_3a_2a_1+\left(10\cdot n_1+1\right)\left(10\cdot n_1-1\right)\cdot 2\cdot a_0 &\equiv 0\left(\bmod p\right)\\ \ldots a_5a_4a_3a_2a_1+\left(10\cdot n_1+1\right)\left(p_1\cdot t\right)\cdot 2\cdot a_0 &\equiv 0\left(\bmod p\right)\\ \ldots a_5a_4a_3a_2a_1+p\left(10\cdot n_1+1\right)\left(t\cdot a_0\right) &\equiv 0\left(\bmod p\right),\ \text{where } a_2a_1=2\cdot a_0,\ a_0\in\mathbb{Z}.\end{aligned}$$

Thus,

$$x\equiv 0\left(\bmod p\right).$$

The proof is completed.

Example 2.6. We show that $61732\equiv 0\left(\bmod 46\right)$.

If $p=46=2\cdot 23$, then $p_1=23$, $t=3$ and $n_1=\dfrac{3\cdot 23+1}{10}=\dfrac{70}{10}=7$.

By Corollary 2.3, we have that
$61732\equiv 0\left(\bmod 46\right)$ if and only if $6173+\left(7\right)\cdot 2\equiv 6187\equiv 618+\left(7\right)\cdot 7\equiv 667\equiv 0\left(\bmod 23\right)$
and the fact that $6\left|792\right.$ we conclude that $61732\equiv 0\left(\bmod 46\right)$.

Theorem 2.7. Let $\ldots a_5a_4a_3a_2a_1=x$ and p be even natural numbers with $p\neq 2$, $10\nmid p$, $p=2\cdot p_1$ and $\gcd\left(2,p_1\right)=1$. There are integer numbers n_i with $n_i\equiv n_1^i\left(\bmod p_1\right)$ and $n_i<p_1$, so that

$$x\equiv 0\left(\bmod p\right)\text{if and only if }\left(\ldots a_{i+2}a_{i+1}\right)+n_i\cdot\left(a_ia_{i-1}\ldots a_2a_1\right)\equiv 0\left(\bmod p_1\right).$$

Proof: From Corollary 2.3, there is an integer $n_1=\dfrac{p_1\cdot t+1}{10}$ for some $t\in\mathbb{Z}$ such that

$$x\equiv 0\left(\bmod p\right)\text{ if and only if }\ldots a_5a_4a_3a_2+n_1\cdot a_1\equiv 0\left(\bmod p_1\right).$$

Assume that $x\equiv 0\left(\bmod p\right)$. Then, there is an integer k such that

$$\ldots a_{i+2}a_{i+1}a_i \ldots a_3a_2a_1 \;=\; 10^i \cdot \left(\ldots a_{i+2}a_{i+1}\right) + a_i \ldots a_3a_2a_1 \;=\; p \cdot k\,.$$

Thus
$$\ldots a_{i+2}a_{i+1} \;=\; \frac{p \cdot k - \left(a_i \ldots a_3a_2a_1\right)}{10^i}\,.$$

On the other hand, there is an integer $n_i = n_1^i + p_1 \cdot l\,,\ l \in \mathbb{Z}$ such that

$$\left(\ldots a_{i+2}a_{i+1}\right) + n_i \cdot \left(a_ia_{i-1} \ldots a_2a_1\right) \in \mathbb{Z}\,.$$

Thus $\left(\ldots a_{i+2}a_{i+1}\right) + n_i \cdot \left(a_ia_{i-1} \ldots a_2a_1\right)$

$$= \frac{p \cdot k - \left(a_ia_{i-1} \ldots a_2a_1\right)}{10^i} + n_i \cdot \left(a_ia_{i-1} \ldots a_2a_1\right)$$

$$= \frac{2 \cdot p_1 \cdot k - 2 \cdot a_0}{10^i} + 2 \cdot n_i \cdot a_0$$

$$= \frac{2 \cdot p_1 \cdot k + 2 \cdot \left(10^i \cdot n_i - 1\right) \cdot a_0}{10^i}$$

$$= \frac{2 \cdot p_1 \cdot k + 2 \cdot \left(10^i \cdot \left[n_1^i + p_1 \cdot l\right] - 1\right) \cdot a_0}{10^i}$$

$$= \frac{2 \cdot p_1 \cdot k + 2 \cdot 10^i \cdot n_1^i \cdot a_0 + 2 \cdot 10^i \cdot p_1 \cdot l \cdot a_0 - 2 \cdot a_0}{10^i}$$

$$= \frac{2 \cdot p_1 \cdot \left(k + 10^i \cdot l \cdot a_0\right) + 2 \cdot \left[\left(10 \cdot n_1\right)^i - 1\right] \cdot a_0}{10^i}$$

$$= \frac{2 \cdot \left(p_1 \cdot \left(k + 10^i \cdot l \cdot a_0\right) + \left(10 \cdot n_1 - 1\right)\left[\left(10 \cdot n_1\right)^{i-1} + \left(10 \cdot n_1\right)^{i-2} + \ldots + 1\right] \cdot a_0\right)}{10^i}\,,$$

where $a_i \ldots a_3a_2a_1 = 2 \cdot a_0,\ \ a_0 \in \mathbb{Z}$. Since $10 \cdot n_1 - 1 = p_1 \cdot t$, we have

$\left(\ldots a_{i+2}a_{i+1}\right) + n_i \cdot \left(a_ia_{i-1} \ldots a_2a_1\right)$

$$= \frac{2 \cdot \left(p_1 \cdot \left(k + 10^i \cdot l \cdot a_0\right) + \left(p_1 \cdot t\right)\left[\left(10 \cdot n_1\right)^{i-1} + \left(10 \cdot n_1\right)^{i-2} + \ldots + 1\right] \cdot a_0\right)}{10^i}$$

$$= p_1 \cdot \frac{2 \cdot \left(k + \left\{10^i \cdot l + t \cdot \left[\left(10 \cdot n_1\right)^{i-1} + \left(10 \cdot n_1\right)^{i-2} + \ldots + 1\right]\right\} \cdot a_0\right)}{10^i} \in \mathbb{Z}\,.$$

Hence, $\dfrac{2 \cdot \left(k + \left\{10^i \cdot l + t \cdot \left[\left(10 \cdot n_1\right)^{i-1} + \left(10 \cdot n_1\right)^{i-2} + \ldots + 1\right]\right\} \cdot a_0\right)}{10^i} \in \mathbb{Z}$ since $5 \nmid p_1$ and $\gcd\left(2, p_1\right) = 1$.

Therefore,

$$\left(\ldots a_{i+2}a_{i+1}\right) + n_i \cdot \left(a_ia_{i-1} \ldots a_2a_1\right) \equiv 0 \left(\mathrm{mod}\ p_1\right).$$

Conversely, assume that $\left(\ldots a_{i+2}a_{i+1}\right) + n_i \cdot \left(a_ia_{i-1} \ldots a_2a_1\right) \equiv 0 \left(\mathrm{mod}\ p_1\right)$,

where $n_i \equiv n_1^i \left(\mathrm{mod}\ p_1\right)$, and $n_i \equiv n_1^i + p_1 \cdot l,\ \ l \in \mathbb{Z}$. We have that $10n_1 - 1 = p_1 \cdot t$ so that

$$p_1 \mid \left(\ldots a_{i+2}a_{i+1}\right) + n_i \cdot \left(a_i a_{i-1} \ldots a_2 a_1\right)$$

$$p \mid 10^i \cdot \left[\left(\ldots a_{i+2}a_{i+1}\right) + n_i \cdot \left(a_i a_{i-1} \ldots a_2 a_1\right)\right]$$

such that

$$10^i \cdot \left[\left(\ldots a_{i+2}a_{i+1}\right) + n_i \cdot \left(a_i a_{i-1} \ldots a_2 a_1\right)\right] \equiv 0 \pmod{p}$$

$$\ldots a_5 a_4 a_3 a_2 a_1 + \left(10^i \cdot n_1^i - 1\right) \cdot \left(a_i a_{i-1} \ldots a_2 a_1\right) \equiv 0 \pmod{p}$$

$$\ldots a_5 a_4 a_3 a_2 a_1 + \left(10 \cdot n_1 - 1\right)\left[\left(10 \cdot n_1\right)^{i-1} + \left(10 \cdot n_1\right)^{i-2} + \ldots + 1\right] \cdot \left(a_i a_{i-1} \ldots a_2 a_1\right) \equiv 0 \pmod{p}$$

$$\ldots a_5 a_4 a_3 a_2 a_1 + \left(p_1 \cdot t\right)\left[\left(10 \cdot n_1\right)^{i-1} + \left(10 \cdot n_1\right)^{i-2} + \ldots + 1\right] \cdot 2 \cdot a_0 \equiv 0 \pmod{p}$$

$$\ldots a_5 a_4 a_3 a_2 a_1 + \left(p \cdot t\right)\left[\left(10 \cdot n_1\right)^{i-1} + \left(10 \cdot n_1\right)^{i-2} + \ldots + 1\right] \cdot a_0 \equiv 0 \pmod{p}.$$

Thus, $x \equiv 0 \pmod{p}$.

The proof is completed.

Example 2.8. We show that $476209384194397806 \equiv 0 \pmod{102}$.

If $p = 102 = 2 \cdot 51$ then $p_1 = 51$. By Theorem 2.1 and Theorem 2.7, we obtained the division rule as follows:

$n_i \equiv n_1^i \pmod{p_1}$	n_i	$\left(\ldots a_{i+2}a_{i+1}\right) + \left[n_i\right] \cdot \left(a_i a_{i-1} \ldots a_2 a_1\right) \equiv 0 \pmod{p_1}$
$n_1 = \dfrac{9 \cdot 51 + 1}{10} = 46$	46	$\left(\ldots a_3 a_2\right) + \left[46(or-5)\right] \cdot \left(a_1\right) \equiv 0 \pmod{51}$
$n_{10} \equiv 46^{10} \equiv 43 \pmod{51}$	43	$\left(\ldots a_{12}a_{11}\right) + \left[43(or-8)\right] \cdot \left(a_{10} \ldots a_3 a_2 a_1\right) \equiv 0 \pmod{51}$
		if and only if $x \equiv 0 \pmod{102}$.

Thus $$47620938 + \left(43\right) \cdot \underbrace{4194397806}_{10\,digit} \equiv 180406726596 \equiv 0 \pmod{51}.$$

We can check 51 is a factor of 180406726596 from Theorem 3 in [1],

$$180406726596 \equiv 1804067 + \left(19\right) \cdot \underbrace{265960}_{6\,digit} \equiv 6857307$$

$$\equiv 6857 + \left(28\right) \cdot 307 \equiv 15453$$

$$\equiv 154 + \left(25\right) \cdot 53 \equiv 1479 \equiv 147 + \left(-5\right) \cdot 9 \equiv 102 \equiv 0 \pmod{51}.$$

3 Conclusion

In conclusion, we assess the divisibility of even numbers based on their rightmost digits and find that for each decreasing number of rightmost digits, there always exists at least one suitable rational number. If we consider the divisor that satisfies $p = 2 \cdot p_1$ where

$\gcd\left(2, p_1\right) = 1$, then there will be at least one suitable value that is an integer. This is true for every decreasing number of rightmost digits.

Acknowledgment

This work is supported by Faculty of Science, Burapha University, Thailand.

References

[1] N, Alp and M. Z.Sarikaya, *New divisibility algorithm for natural number*," 2019, [Online]. Available: https://www.researchgate.net/publication/337439296_NEW_DIVISIBILITY_ALGORITHM_FOR_NATURAL_NUMBER. (Accessed May. 14, 2023).

[2] H. Khosravi et al., *A new algorithm for divisibility of numbers*, World Applied Sciences Journal, vol. 18, no. 6, pp. 786 - 787, 2012. doi: 10.5829/idosi.wasj.2012.18.06.306

[3] P. Tiebekabe and I. Diouïf, *New divisibility tests*," Far East Journal of Mathematical Education, Vol. 21 no 1, pp. 31-41, 2021.

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

สมการไดโอแฟนไทน์ $n^x + p^y = z^2$ เมื่อ p เป็นจำนวนเฉพาะ และ $n \equiv 2(\mathrm{mod}\ 3p)$

อนุสรา ประสิทธิ์นอก[1,†] และ วีรยุทธ นิลสระคู [1,‡]

[1]ภาควิชาคณิตศาสตร์ สถิติและคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี 34190

บทคัดย่อ

วัตถุประสงค์ของงานวิจัยนี้ คือ ศึกษาหาผลเฉลย (n, x, y, z) ที่เป็นจำนวนเต็มที่ไม่เป็นลบของสมการไดโอแฟนไทน์ $n^x + p^y = z^2$ โดยที่ p เป็นจำนวนเฉพาะ และ n เป็นจำนวนเต็มบวกซึ่ง $n \equiv 2(\mathrm{mod}\ 3p)$ ในการพิสูจน์จะใช้ข้อคาดการณ์ของคาตาลานและทฤษฎีจำนวนเบื้องต้น จากการศึกษาพบว่า

(i) ถ้า $p \equiv 19(\mathrm{mod}\ 24)$ และ $n \equiv 2(\mathrm{mod}\ 3p)$ แล้วสมการไดโอแฟนไทน์ $n^x + p^y = z^2$ มีผลเฉลยเพียงผลเฉลยเดียว คือ $(n, x, y, z) = (2, 3, 0, 3)$

(ii) ถ้า $p \equiv 13(\mathrm{mod}\ 24)$ และ $n \equiv 2(\mathrm{mod}\ 3p)$ แล้วสมการไดโอแฟนไทน์ $n^x + p^y = z^2$ มีผลเฉลยอยู่ในรูปทั่วไป คือ $(n, x, y, z) \in \{(2, 3, 0, 3)\} \cup \{(n, 1, 0, \sqrt{n+1}) : n+1$ เป็นกำลังสองสมบูรณ์$\}$

คำสำคัญ: จำนวนเฉพาะ, สมภาค, ส่วนตกค้างกำลังสอง, สมการไดโอแฟนไทน์

2020 MSC: 11D61

1 บทนำ

สมการไดโอแฟนไทน์เป็นสมการที่มีหลายลักษณะ หลายเงื่อนไขที่แตกต่างกันไป รูปแบบหนึ่งที่มีผู้สนใจศึกษากันอย่างกว้างขวาง คือ สมการไดโอแฟนไทน์ที่อยู่ในรูป $a^x + b^y = z^2$ โดยที่ a และ b เป็นตัวแปรที่ทราบค่า เช่น

ในปี ค.ศ. 2011 Alongkot Suvarnamani [3] ได้ศึกษาสมการไดโอแฟนไทน์ $2^x + p^y = z^2$ โดยที่ p เป็นจำนวนเฉพาะ เมื่อ x, y และ z เป็นจำนวนเต็มที่ไม่เป็นลบ พบว่ามีผลเฉลย คือ $(x, y, z) = (3, 0, 3)$ แต่มีข้อบกพร่องในการพิสูจน์ ดังนั้นจึงมีนักวิจัยนำเสนอที่แก้ไขประเด็นที่ผิดพลาด ดังต่อไปนี้

ในปี ค.ศ. 2013 Banyat Sroysang [4] ได้ศึกษาสมการไดโอแฟนไทน์ $2^x + 19^y = z^2$ โดยที่ x, y และ z

[†]ผู้นำเสนอ ‡ผู้แต่งหลัก
อีเมล: anusara.pr.63@ubu.ac.th (อนุสรา ประสิทธิ์นอก), weerayuth.ni@ubu.ac.th (วีรยุทธ นิลสระคู).

เป็นจำนวนเต็มที่ไม่เป็นลบ พบว่ามีผลเฉลยเพียงผลเฉลยเดียว คือ $(x, y, z) = (3, 0, 3)$

ในปี ค.ศ. 2018 Nechemia Burshtein [7] ได้ศึกษาสมการไดโอแฟนไทน์ $2^x + p^y = z^2$ เมื่อ $y = 1$ และ $p = 7, 13, 29, 37$ และ 257 พบว่ามีผลเฉลยที่เป็นจำนวนเต็มบวก ดังนี้ กรณีที่ $p = 7$ มีผลเฉลยเพียงผลเฉลยเดียว คือ $(x, y, z) = (1, 1, 3)$ กรณีที่ $p = 13, 29, 37$ ไม่มีผลเฉลย และกรณีที่ $p = 257$ มีผลเฉลยสองผลเฉลย คือ $(x, y, z) \in \{(14, 1, 129), (5, 1, 17)\}$

ในปี ค.ศ. 2019 Gawkhare Mahesh และ Vikita Sinari [6] ได้ศึกษาสมการไดโอแฟนไทน์ $2^x + p^y = z^2$ โดยที่ p เป็นจำนวนเฉพาะคี่ เมื่อ x และ y ไม่เป็นจำนวนเต็มบวกคี่พร้อมกัน พบว่ามีผลเฉลย คือ $(p, x, y, z) \in \{(p, 3, 0, 3)\} \cup \{(2^{m+1} + 1, 2m, 1, 2^m + 1)\,;\, m \in \mathbb{N}$ และ $2^{m+1} + 1$ เป็นจำนวนเฉพาะ$\} \cup \{(2^q - 1, q + 2, 2, p + 2)\,;\, q$ และ $2^q - 1$ เป็นจำนวนเฉพาะ $\}$

ในปี ค.ศ. 2022 Suton Tadee [11] ได้ศึกษาสมการไดโอแฟนไทน์ $2^x + p^y = z^2$ โดยที่ p เป็นจำนวนเฉพาะซึ่ง $p \equiv 3 (\mathrm{mod}\, 4)$ และ $x \neq 1$ เมื่อ x, y และ z เป็นจำนวนเต็มที่ไม่เป็นลบ พบว่ามีผลเฉลยอยู่ในรูปทั่วไป คือ $(p, x, y, z) \in \{(p, 3, 0, 3)\} \cup \{(3, 0, 1, 2)\} \cup \{(p, 2 + \log_2(p + 1), 2, p + 2) : \log_2(p + 1) \in \mathbb{Z}\}$

ในปี ค.ศ. 2021 Nongluk Viriyapong และ Chokchai Viriyapong [8] ได้ศึกษาสมการไดโอแฟนไทน์ $n^x + 13^y = z^2$ โดยที่ n เป็นจำนวนเต็มบวกซึ่ง $n \equiv 2 (\mathrm{mod}\, 39)$ และ $n + 1$ ไม่เป็นกำลังสองสมบูรณ์ เมื่อ x, y และ z เป็นจำนวนเต็มที่ไม่เป็นลบ พบว่ามีผลเฉลยเพียงผลเฉลยเดียว คือ $(n, x, y, z) = (2, 3, 0, 3)$

ในปี ค.ศ. 2022 Nongluk Viriyapong และ Chokchai Viriyapong [9] ได้ศึกษาสมการไดโอแฟนไทน์ $n^x + 19^y = z^2$ โดยที่ n เป็นจำนวนเต็มบวกซึ่ง $n \equiv 2 (\mathrm{mod}\, 57)$ เมื่อ x, y และ z เป็นจำนวนเต็มที่ไม่เป็นลบ พบว่ามีผลเฉลยเพียงผลเฉลยเดียว คือ $(n, x, y, z) = (2, 3, 0, 3)$

ดังนั้น ผู้วิจัยจึงสนใจศึกษาสมการไดโอแฟนไทน์ $n^x + p^y = z^2$ โดยที่ p เป็นจำนวนเฉพาะ และ n เป็นจำนวนเต็มบวกซึ่ง $n \equiv 2 (\mathrm{mod}\, 3p)$ เมื่อ x, y และ z เป็นจำนวนเต็มที่ไม่เป็นลบ

2 ความรู้พื้นฐาน

ให้ $\mathbb{N}$ แทน เซตของจำนวนนับหรือจำนวนธรรมชาติ (set of all natural numbers)
$\mathbb{Z}$ แทน เซตของจำนวนเต็ม (set of all integer numbers)

นิยาม 2.1. [2] ให้ m เป็นจำนวนเต็มบวก a และ b เป็นจำนวนเต็ม แล้วจะเรียกว่า a **สมภาค กับ** b **มอดุโล** m (a **is congruent** b **modulo** m) ซึ่งเขียนแทนด้วย $a \equiv b \pmod m$ ถ้า a และ b มีเศษจากการหารด้วย m เท่ากัน

ทฤษฎีบท 2.2. [2] ให้ $a, b \in \mathbb{Z}$ และ $m \in \mathbb{N}$ จะได้ว่า $a \equiv b \pmod m$ ก็ต่อเมื่อ $m \mid (a - b)$

ทฤษฎีบท 2.3. [2] ให้ $a, b, c, d \in \mathbb{Z}$ และ $m, n \in \mathbb{N}$ แล้ว

1. $a \equiv a \pmod m$
2. ถ้า $a \equiv b \pmod m$ แล้ว $b \equiv a \pmod m$
3. ถ้า $a \equiv b \pmod m$ และ $b \equiv c \pmod m$ แล้ว $a \equiv c \pmod m$
4. ถ้า $a \equiv b \pmod m$ และ $c \equiv d \pmod m$ แล้ว $a \pm c \equiv b \pm d \pmod m$
5. ถ้า $a \equiv b \pmod m$ และ $c \equiv d \pmod m$ แล้ว $ac \equiv bd \pmod m$
6. ถ้า $a \equiv b \pmod m$ แล้ว $a^n \equiv b^n \pmod m$
7. ถ้า $a \equiv b \pmod m$ และ $n \mid m$ แล้ว $a \equiv b \pmod n$
8. ถ้า $a \equiv b \pmod m$ และ $c \neq 0$ แล้ว $ac \equiv bc \pmod{|c|m}$
9. ถ้า $a \equiv b \pmod m$ แล้ว $\gcd(a, m) = \gcd(b, m)$ เมื่อ $\gcd(a, m)$ หมายถึง ห.ร.ม ของ a และ m

นิยาม 2.4. [1] สมภาคในรูป $x^2 \equiv a(\mathrm{mod}\, p)$ เมื่อ p เป็นจำนวนเฉพาะ และ $a \in \mathbb{Z}$ จะเรียกว่า **สมภาคกำลังสอง (quadratic congruence)**

นิยาม 2.5. [1] ให้ p เป็นจำนวนเฉพาะคี่ และ $a \in \mathbb{Z}$ ซึ่ง $\gcd(a,p) = 1$ ถ้า $x^2 \equiv a(\mathrm{mod}\, p)$ มีผลเฉลยแล้วจะเรียก a ว่า **ส่วนตกค้างกำลังสอง (quadratic residues)** ของ p และถ้าไม่มีผลเฉลย แล้วจะเรียก a ว่า **ส่วนไม่ตกค้างกำลังสอง (quadratic non-residue)**

นิยาม 2.6. [1] ให้ p เป็นจำนวนเฉพาะคี่ และ $a \in \mathbb{Z}$ ซึ่ง $\gcd(a,p) = 1$ **สัญลักษณ์เลอช็องดร์** $\left(\frac{a}{p}\right)$ กำหนดโดย

$$\left(\frac{a}{p}\right) = \begin{cases} 1 & \text{ถ้า } a \text{ เป็นส่วนตกค้างกำลังสองของ } p \\ -1 & \text{ถ้า } a \text{ เป็นส่วนไม่ตกค้างกำลังสองของ } p \end{cases}$$

ทฤษฎีบท 2.7. [1] ให้ p เป็นจำนวนเฉพาะคี่ และ $a, b \in \mathbb{Z}$ ซึ่ง $\gcd(ab,p) = 1$

1. ถ้า $a \equiv b(\mathrm{mod}\, p)$ แล้ว $\left(\frac{a}{p}\right) = \left(\frac{b}{p}\right)$
2. $\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right)$
3. $\left(\frac{a^2}{p}\right) = 1$

ทฤษฎีบท 2.8. [1] ถ้า p เป็นจำนวนเฉพาะคี่ แล้ว

$$\left(\frac{2}{p}\right) = \begin{cases} 1 & \text{ถ้า } p \equiv \pm 1(\mathrm{mod}\, 8) \\ -1 & \text{ถ้า } p \equiv \pm 3(\mathrm{mod}\, 8) \end{cases}$$

ทฤษฎีบท 2.9. [5] ถ้า p เป็นจำนวนเฉพาะคี่ ซึ่ง $p \neq 3$ แล้ว

$$\left(\frac{3}{p}\right) = \begin{cases} 1 & \text{ถ้า } p \equiv \pm 1(\mathrm{mod}\, 12) \\ -1 & \text{ถ้า } p \equiv \pm 5(\mathrm{mod}\, 12) \end{cases}$$

บทตั้ง 2.10. [8] ถ้า z เป็นจำนวนเต็ม แล้ว $z^2 \equiv 0, 1(\mathrm{mod}\, 3)$

ทฤษฎีบท 2.11. [10] **ข้อคาดการณ์ของคาตาลาน (Catalan's conjecture)** สมการไดโอแฟนไทน์ $a^x - b^y = 1$ มีผลเฉลยเพียงผลเฉลยเดียว คือ $(a,b,x,y) = (3,2,2,3)$ เมื่อ a, b, x และ y เป็นจำนวนเต็มบวก ซึ่ง $\min\{a,b,x,y\} > 1$

บทตั้ง 2.12. ถ้า p เป็นจำนวนเฉพาะซึ่ง $p \equiv 1(\mathrm{mod}\, 3)$ แล้วสมการไดโอแฟนไทน์ $1 + p^y = z^2$ ไม่มีผลเฉลยเมื่อ y และ z เป็นจำนวนเต็มที่ไม่เป็นลบ

พิสูจน์. ให้ p เป็นจำนวนเฉพาะซึ่ง $p \equiv 1(\mathrm{mod}\, 3)$
สมมติว่า (y,z) เป็นจำนวนเต็มที่ไม่เป็นลบที่สอดคล้องสมการ $1 + p^y = z^2$
กรณีที่ 1. $y = 0$ จะได้ว่า $z^2 = 2$ ซึ่งเป็นไปไม่ได้
กรณีที่ 2. $y = 1$ จะได้ว่า $z^2 = 1 + p$ เนื่องจาก $p \equiv 1(\mathrm{mod}\, 3)$ นั่นคือ $p + 1 \equiv 2(\mathrm{mod}\, 3)$ เพราะฉะนั้น $z^2 \equiv 2(\mathrm{mod}\, 3)$ ซึ่งเกิดข้อขัดแย้งกับบทตั้ง 2.10
กรณีที่ 3. $y > 1$ จะได้ว่า $z^2 > 1 + p$ เนื่องจาก $p \equiv 1(\mathrm{mod}\, 3)$ นั่นคือ $p \geq 7$ จะได้ว่า $z^2 > 8$ เพราะฉะนั้น $z \geq 3$
พิจารณา $z^2 - p^y = 1$
เนื่องจาก $\min\{z,p,2,y\} > 1$ โดยบทตั้ง 2.11 จะได้ว่า $(z,p,2,y) = (3,2,2,3)$ ซึ่งเกิดข้อขัดแย้ง
ดังนั้น จากทั้ง 3 กรณี พบว่าไม่มีผลเฉลยเป็นจำนวนเต็มที่ไม่เป็นลบ □

บทตั้ง 2.13. สมการไดโอแฟนไทน์ $n^x + 1 = z^2$ โดยที่ n เป็นจำนวนเต็มบวก และ x, z เป็นจำนวนเต็มที่ไม่เป็นลบ มีผลเฉลยอยู่ในรูปทั่วไป คือ

$$(n, x, z) \in \{(2, 3, 3)\} \cup \{(n, 1, \sqrt{n+1}) : n+1 \text{ เป็นกำลังสองสมบูรณ์}\}$$

พิสูจน์. ให้ n เป็นจำนวนเต็มบวก และ x, z เป็นจำนวนเต็มที่ไม่เป็นลบ
สมมติว่า (n, x, z) เป็นผลเฉลยของสมการ $n^x + 1 = z^2$
กรณีที่ 1. $x = 0$ จะได้ว่า $z^2 = 2$ ซึ่งเป็นไปไม่ได้
กรณีที่ 2. $x = 1$ จะได้ว่า $z^2 = n + 1$
(i) กรณี $n + 1$ ไม่เป็นกำลังสองสมบูรณ์ จะเกิดข้อขัดแย้ง
(ii) กรณี $n + 1$ เป็นกำลังสองสมบูรณ์ จะได้ $z = \sqrt{n+1}$
กรณีที่ 3. $x > 1$ จะได้ว่า $z^2 > n + 1$
เนื่องจาก $n \geq 1$ จะได้ว่า $z^2 > 2$ เพราะฉะนั้น $z \geq 2$
พิจารณา $z^2 - n^x = 1$
(i) กรณี $n = 1$ จะได้ $z^2 = 2$ ซึ่งเป็นไปไม่ได้
(ii) กรณี $n > 1$ เนื่องจาก $\min\{z, n, 2, x\} > 1$
โดยบทตั้ง 2.11 จะได้ว่า $(z, n, 2, x) = (3, 2, 2, 3)$ เพราะฉะนั้น $(n, x, z) = (2, 3, 3)$
ดังนั้น สมการ $n^x + 1 = z^2$ มีผลเฉลยอยู่ในรูปทั่วไป คือ

$$(n, x, z) \in \{(2, 3, 3)\} \cup \{(n, 1, \sqrt{n+1}) : n+1 \text{ เป็นกำลังสองสมบูรณ์}\}$$

□

3 ผลการศึกษา

บทตั้ง 3.1. ให้ p เป็นจำนวนเฉพาะ และ n เป็นจำนวนเต็มบวก ถ้า $p \equiv 19 (\mod 24)$ และ $n \equiv 2 (\mod 3p)$ แล้ว $n + 1$ ไม่เป็นกำลังสองสมบูรณ์

พิสูจน์. ให้ p เป็นจำนวนเฉพาะซึ่ง $p \equiv 19 (\mod 24)$ และ n เป็นจำนวนเต็มบวกซึ่ง $n \equiv 2 (\mod 3p)$
สมมติให้ $n + 1$ เป็นกำลังสองสมบูรณ์ นั่นคือ จะมีจำนวนเต็ม t ที่ซึ่ง $t^2 = n + 1$
จาก $n \equiv 2 (\mod 3p)$ จะได้ว่า $n \equiv 2 (\mod p)$ เพราะฉะนั้น $t^2 \equiv 3 (\mod p)$ ดังนั้น $\left(\frac{3}{p}\right) = 1$
จาก $p \equiv 19 (\mod 24)$ จะได้ว่า $p \equiv 19 (\mod 12)$ เพราะฉะนั้น $p \equiv -5 (\mod 12)$
โดยทฤษฎีบท 2.9 จะได้ว่า $\left(\frac{3}{p}\right) = -1$ ซึ่งเกิดข้อขัดแย้ง ดังนั้น $n + 1$ ไม่เป็นกำลังสองสมบูรณ์ □

บทตั้ง 3.2. ถ้า p เป็นจำนวนเฉพาะซึ่ง $p \equiv 13, 19 (\mod 24)$ และ x เป็นจำนวนเต็มบวกคี่ แล้ว $\left(\frac{2^x}{p}\right) = -1$

พิสูจน์. ให้ p เป็นจำนวนเฉพาะซึ่ง $p \equiv 13, 19 (\mod 24)$
และให้ x เป็นจำนวนเต็มบวกคี่ นั่นคือ จะมีจำนวนเต็มที่ไม่เป็นลบ m ซึ่ง $x = 2m + 1$
พิจารณา $\left(\frac{2^x}{p}\right) = \left(\frac{2^{2m+1}}{p}\right) = \left(\frac{(2^m)^2 \cdot 2^1}{p}\right) = 1\left(\frac{2}{p}\right)$
เนื่องจาก $p \equiv 13, 19 (\mod 24)$ จะได้ว่า $p \equiv 13, 19 (\mod 8)$
เนื่องจาก $13 \equiv -3 (\mod 8)$ และ $19 \equiv 3 (\mod 8)$ จะได้ $p \equiv \pm 3 (\mod 8)$
โดยทฤษฎีบท 2.8 จะได้ว่า $\left(\frac{2}{p}\right) = -1$ ดังนั้น $\left(\frac{2^x}{p}\right) = -1$ □

ทฤษฎีบท 3.3. ให้ n, x, y เป็นจำนวนเต็มบวก และ z เป็นจำนวนเต็มที่ไม่เป็นลบ ถ้า p เป็นจำนวนเฉพาะซึ่ง $p \equiv 13, 19 (\mod 24)$ แล้วสมการไดโอแฟนไทน์ $n^x + p^y = z^2$ โดยที่ $n \equiv 2 (\mod 3p)$ ไม่มีผลเฉลย

พิสูจน์. ให้ n, x, y เป็นจำนวนเต็มบวก และ z เป็นจำนวนเต็มที่ไม่เป็นลบ
ให้ p เป็นจำนวนเฉพาะซึ่ง $p \equiv 13, 19 (\text{mod } 24)$ และ $n^x + p^y = z^2$ เมื่อ $n \equiv 2 (\text{mod } 3p)$
กรณีที่ 1. x เป็นจำนวนเต็มบวกคู่ นั่นคือ จะมีจำนวนเต็มบวก k ที่ซึ่ง $x = 2k$
จาก $n \equiv 2 (\text{mod } 3p)$ จะได้ว่า $n \equiv 2 (\text{mod } 3)$ ดังนั้น $n^x \equiv 1 (\text{mod } 3)$
เนื่องจาก $p \equiv 13, 19 (\text{mod } 24)$ จะได้ว่า $p \equiv 13, 19 (\text{mod } 3)$
เนื่องจาก $13 \equiv 1 (\text{mod } 3)$ และ $19 \equiv 1 (\text{mod } 3)$ จะได้ว่า $p^y \equiv 1 (\text{mod } 3)$
เพราะฉะนั้น $z^2 \equiv 2 (\text{mod } 3)$ ซึ่งเกิดข้อขัดแย้งกับบทตั้ง 2.10
กรณีที่ 2. x เป็นจำนวนเต็มบวกคี่
เนื่องจาก $n \equiv 2 (\text{mod } 3p)$ จะได้ว่า $n \equiv 2 (\text{mod } p)$ เพราะฉะนั้น $n^x \equiv 2^x (\text{mod } p)$
เนื่องจาก $p \equiv 0 (\text{mod } p)$ จะได้ว่า $p^y \equiv 0 (\text{mod } p)$ ดังนั้น $z^2 \equiv 2^x (\text{mod } p)$
จะได้ว่า $\left(\frac{2^x}{p}\right) = 1$ ซึ่งเกิดข้อขัดแย้งกับบทตั้ง 3.2
ดังนั้น จากทั้ง 2 กรณี พบว่า ถ้า p เป็นจำนวนเฉพาะซึ่ง $p \equiv 13, 19 (\text{mod } 24)$ แล้วสมการไดโอแฟนไทน์ $n^x + p^y = z^2$ โดยที่ $n \equiv 2 (\text{mod } 3p)$ ไม่มีผลเฉลย เมื่อ n, x, y เป็นจำนวนเต็มบวก และ z เป็นจำนวนเต็มที่ไม่เป็นลบ

□

ทฤษฎีบท 3.4. ให้ p เป็นจำนวนเฉพาะ และ n เป็นจำนวนเต็มบวก ถ้า $p \equiv 13 (\text{mod } 24)$ แล้วสมการไดโอแฟนไทน์ $n^x + p^y = z^2$ โดยที่ $n \equiv 2 (\text{mod } 3p)$ มีผลเฉลยอยู่ในรูปทั่วไป เมื่อ x, y และ z เป็นจำนวนเต็มที่ไม่เป็นลบ คือ

$$(n, x, y, z) \in \{(2, 3, 0, 3)\} \cup \{(n, 1, 0, \sqrt{n+1}) : n+1 \text{ เป็นกำลังสองสมบูรณ์}\}$$

พิสูจน์. ให้ p เป็นจำนวนเฉพาะ, n เป็นจำนวนเต็มบวก และ x, y, z เป็นจำนวนเต็มที่ไม่เป็นลบ
ซึ่ง $p \equiv 13 (\text{mod } 24)$, $n \equiv 2 (\text{mod } 3p)$ และ $n^x + p^y = z^2$
กรณีที่ 1. $x = 0$ จะได้ $1 + p^y = z^2$ โดยบทตั้ง 2.12 พบว่าไม่มีผลเฉลย
กรณีที่ 2. $y = 0$ จะได้ $n^x + 1 = z^2$ โดยบทตั้ง 2.13 พบว่า

$$(n, x, z) \in \{(2, 3, 3)\} \cup \{(n, 1, \sqrt{n+1}) : n+1 \text{ เป็นกำลังสองสมบูรณ์}\}$$

กรณีที่ 3. $x \geq 1, y \geq 1$ โดยทฤษฎีบท 3.3 พบว่าไม่มีผลเฉลย
ดังนั้น จากทั้ง 3 กรณี พบว่า ถ้า $p \equiv 13 (\text{mod } 24)$ แล้วสมการไดโอแฟนไทน์ $n^x + p^y = z^2$ โดยที่ $n \equiv 2 (\text{mod } 3p)$ มีผลเฉลยอยู่ในรูปทั่วไป คือ

$$(n, x, y, z) \in \{(2, 3, 0, 3)\} \cup \{(n, 1, 0, \sqrt{n+1}) : n+1 \text{ เป็นกำลังสองสมบูรณ์}\}$$

เมื่อ x, y และ z เป็นจำนวนเต็มที่ไม่เป็นลบ

□

ต่อไปจะเป็นการนำทฤษฎีบท 3.4 ไปประยุกต์ใช้ เมื่อ $p = 37, 61$

บทแทรก 3.5. สมการไดโอแฟนไทน์ $n^x + 37^y = z^2$ โดยที่ $n \equiv 2 (\text{mod } 111)$ มีผลเฉลยอยู่ในรูปทั่วไป เมื่อ x, y และ z เป็นจำนวนเต็มที่ไม่เป็นลบ คือ

$$(n, x, y, z) \in \{(2, 3, 0, 3)\} \cup \{(n, 1, 0, \sqrt{n+1}) : n+1 \text{ เป็นกำลังสองสมบูรณ์}\}$$

พิสูจน์. เนื่องจาก 37 เป็นจำนวนเฉพาะซึ่ง $37 \equiv 13 (\text{mod } 24)$
จาก $n \equiv 2 (\text{mod } 111)$ จะได้ว่า $n \equiv 2 (\text{mod } 3(37))$
เพราะฉะนั้น โดยทฤษฎีบท 3.4 พบว่าสมการไดโอแฟนไทน์ $n^x + 37^y = z^2$ มีผลเฉลยอยู่ในรูปทั่วไป คือ $(n, x, y, z) \in \{(2, 3, 0, 3)\} \cup \{(n, 1, 0, \sqrt{n+1}) : n+1 \text{ เป็นกำลังสองสมบูรณ์}\}$

□

ตัวอย่าง 3.6. พิจารณา $n \equiv 2 (\text{mod } 111)$ เมื่อ $n+1$ เป็นกำลังสองสมบูรณ์
สำหรับ $1 \leq n \leq 20,000$ พบว่า

$$(n,x,y,z) \in \{(2,3,0,3)\} \cup \{(224,1,0,15),(9215,1,0,96),(15875,1,0,126)\}$$

เป็นผลเฉลยของสมการไดโอแฟนไทน์ $n^x + 37^y = z^2$ โดยที่ $n \equiv 2 (\text{mod } 111)$ เมื่อ $1 \leq n \leq 20,000$

บทแทรก 3.7. สมการไดโอแฟนไทน์ $n^x + 61^y = z^2$ โดยที่ $n \equiv 2 (\text{mod } 183)$ มีผลเฉลยอยู่ในรูปทั่วไป เมื่อ x, y และ z เป็นจำนวนเต็มที่ไม่เป็นลบ คือ

$$(n,x,y,z) \in \{(2,3,0,3)\} \cup \{(n,1,0,\sqrt{n+1}) : n+1 \text{ เป็นกำลังสองสมบูรณ์}\}$$

พิสูจน์. เนื่องจาก 61 เป็นจำนวนเฉพาะซึ่ง $61 \equiv 13 (\text{mod } 24)$
จาก $n \equiv 2 (\text{mod } 183)$ จะได้ว่า $n \equiv 2 (\text{mod } 3(61))$
เพราะฉะนั้น โดยทฤษฎีบท 3.4 พบว่าสมการไดโอแฟนไทน์ $n^x + 61^y = z^2$ มีผลเฉลยอยู่ในรูปทั่วไป คือ $(n,x,y,z) \in \{(2,3,0,3)\} \cup \{(n,1,0,\sqrt{n+1}) : n+1 \text{ เป็นกำลังสองสมบูรณ์}\}$

□

ตัวอย่าง 3.8. พิจารณา $n \equiv 2 (\text{mod } 183)$ เมื่อ $n+1$ เป็นกำลังสองสมบูรณ์
สำหรับ $1 \leq n \leq 64,000$ พบว่า

$$(n,x,y,z) \in \{(2,3,0,3)\} \cup \{(4760,1,0,69),(12995,1,0,114),(63503,1,0,252)\}$$

เป็นผลเฉลยของสมการไดโอแฟนไทน์ $n^x + 61^y = z^2$ โดยที่ $n \equiv 2 (\text{mod } 183)$ เมื่อ $1 \leq n \leq 64,000$

ทฤษฎีบท 3.9. ให้ p เป็นจำนวนเฉพาะ และ n เป็นจำนวนเต็มบวก ถ้า $p \equiv 19 (\text{mod } 24)$ แล้วสมการไดโอแฟนไทน์ $n^x + p^y = z^2$ โดยที่ $n \equiv 2 (\text{mod } 3p)$ เมื่อ x, y และ z เป็นจำนวนเต็มที่ไม่เป็นลบ มีผลเฉลยเพียงผลเฉลยเดียว คือ $(n,x,y,z) = (2,3,0,3)$

พิสูจน์. ให้ p เป็นจำนวนเฉพาะ, n เป็นจำนวนเต็มบวก และให้ x, y, z เป็นจำนวนเต็มที่ไม่เป็นลบ
ซึ่ง $p \equiv 19 (\text{mod } 24)$, $n \equiv 2 (\text{mod } 3p)$ และ $n^x + p^y = z^2$
กรณีที่ 1. $x = 0$ จะได้ $1 + p^y = z^2$ โดยบทตั้ง 2.12 พบว่าไม่มีผลเฉลย
กรณีที่ 2. $y = 0$ จะได้ว่า $n^x + 1 = z^2$ โดยบทตั้ง 2.13 และบทตั้ง 3.1 จะได้ว่า $(n,x,z) = (2,3,3)$
กรณีที่ 3. $x \geq 1$ และ $y \geq 1$ โดยทฤษฎีบท 3.3 พบว่าไม่มีผลเฉลย
ดังนั้น จากทั้ง 3 กรณี พบว่า ถ้า $p \equiv 19 (\text{mod } 24)$ แล้วสมการไดโอแฟนไทน์ $n^x + p^y = z^2$ โดยที่ $n \equiv 2 (\text{mod } 3p)$ มีผลเฉลยเพียงผลเฉลยเดียว คือ $(n,x,y,z) = (2,3,0,3)$ เมื่อ x, y และ z เป็นจำนวนเต็มที่ไม่เป็นลบ

□

ต่อไปจะเป็นการนำทฤษฎีบท 3.9 ไปประยุกต์ใช้ เมื่อ $p = 43, 67$

บทแทรก 3.10. สมการไดโอแฟนไทน์ $n^x + 43^y = z^2$ โดยที่ n เป็นจำนวนเต็มบวกซึ่ง $n \equiv 2 (\text{mod } 129)$ มีผลเฉลยเพียงผลเฉลยเดียว เมื่อ x, y และ z เป็นจำนวนเต็มที่ไม่เป็นลบ คือ $(n,x,y,z) = (2,3,0,3)$

พิสูจน์. เนื่องจาก 43 เป็นจำนวนเฉพาะซึ่ง $43 \equiv 19 (\text{mod } 24)$
เนื่องจาก $n \equiv 2 (\text{mod } 129)$ จะได้ว่า $n \equiv 2 (\text{mod } 3(43))$
เพราะฉะนั้น โดยทฤษฎีบท 3.9 พบว่าสมการไดโอแฟนไทน์ $n^x + 43^y = z^2$ มีผลเฉลยเพียงผลเฉลยเดียว คือ $(n,x,y,z) = (2,3,0,3)$

□

บทแทรก 3.11. สมการไดโอแฟนไทน์ $n^x + 67^y = z^2$ โดยที่ n เป็นจำนวนเต็มบวกซึ่ง $n \equiv 2(\text{mod } 201)$ มีผลเฉลยเพียงผลเฉลยเดียว เมื่อ x, y และ z เป็นจำนวนเต็มที่ไม่เป็นลบ คือ $(n, x, y, z) = (2, 3, 0, 3)$

พิสูจน์. เนื่องจาก 67 เป็นจำนวนเฉพาะซึ่ง $67 \equiv 19(\text{mod } 24)$
เนื่องจาก $n \equiv 2(\text{mod } 201)$ จะได้ว่า $n \equiv 2(\text{mod } 3(67))$
เพราะฉะนั้น โดยทฤษฎีบท 3.9 พบว่าสมการไดโอแฟนไทน์ $n^x + 67^y = z^2$ มีผลเฉลยเพียงผลเฉลยเดียว คือ $(n, x, y, z) = (2, 3, 0, 3)$

□

บทแทรก 3.12. ถ้า $p \equiv 13, 19(\text{mod } 24)$ แล้วสมการไดโอแฟนไทน์ $2^x + p^y = z^2$ มีผลเฉลยเพียงผลเฉลยเดียว คือ $(x, y, z) = (3, 0, 3)$

พิสูจน์. เนื่องจาก $p \equiv 13, 19(\text{mod } 24)$ และ $n = 2$ จะได้ว่า $n \equiv 2(\text{mod } 3p)$ ดังนั้น โดยทฤษฎีบท 3.4 และทฤษฎีบท 3.9 พบว่าสมการไดโอแฟนไทน์ $2^x + p^y = z^2$ มีผลเฉลยเพียงผลเฉลยเดียว คือ $(x, y, z) = (3, 0, 3)$

□

4 สรุปผลและข้อเสนอแนะ

4.1 สรุปผลการศึกษา

จากการศึกษาเพื่อหาผลเฉลย (n, x, y, z) ที่เป็นจำนวนเต็มที่ไม่เป็นลบของสมการไดโอแฟนไทน์ $n^x + p^y = z^2$ โดยที่ p เป็นจำนวนเฉพาะ และ n เป็นจำนวนเต็มบวกซึ่ง $n \equiv 2(\text{mod } 3p)$ พบว่า

ถ้า $p \equiv 13(\text{mod } 24)$ และ $n \equiv 2(\text{mod } 3p)$ แล้วสมการไดโอแฟนไทน์ $n^x + p^y = z^2$ มีผลเฉลยอยู่ในรูปทั่วไป คือ $(n, x, y, z) \in \{(2, 3, 0, 3)\} \cup \{(n, 1, 0, \sqrt{n+1}) : n + 1 \text{ เป็นกำลังสองสมบูรณ์}\}$

ถ้า $p \equiv 19(\text{mod } 24)$ และ $n \equiv 2(\text{mod } 3p)$ แล้วสมการไดโอแฟนไทน์ $n^x + p^y = z^2$ มีผลเฉลยเพียงผลเฉลยเดียว คือ $(n, x, y, z) = (2, 3, 0, 3)$

4.2 ข้อเสนอแนะ

ศึกษาสมการไดโอแฟนไทน์ $n^x + p^y = z^2$ โดยที่ $n \equiv 2(\text{mod } 3p)$ ว่า ถ้า p เป็นจำนวนเฉพาะใด ๆ ซึ่ง $p \not\equiv 13, 19(\text{mod } 24)$ มีผลเฉลยที่เป็นจำนวนเต็มที่ไม่เป็นลบหรือไม่

กิตติกรรมประกาศ ผู้แต่งขอขอบคุณผู้ทรงคุณวุฒิทุกท่านที่ได้ให้ข้อคิดเห็นและข้อเสนอแนะต่าง ๆ เพื่อปรับปรุงบทความวิจัยนี้ และขอขอบพระคุณเจ้าของเอกสารและงานวิจัยทุกท่านที่ผู้ศึกษาค้นคว้าได้นำมาอ้างอิงในการทำวิจัย จนกระทั่งงานวิจัยฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี

References

[1] วัลลภ เหมวงษ์, *ทฤษฎีจำนวน (number theory)*, มหาวิทยาลัยราชภัฏอุดรธานี (2564), 150–166.

[2] อัจฉรา หาญชูวงศ์, *ทฤษฎีจำนวน*, กรุงเทพฯ : โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย (2542).

[3] A. Suvarnamani, *Solutions of the diophantine equations* $2^x + p^y = z^2$, Int. J. Math. Sci. Appl. **1** (2011), no. 3, 1415–1419.

[4] B. Sroysang, *More on the diophantine equation* $2^x + 19^y = z^2$, Int. J. Pure Appl. Math. **88** (2013), no. 1, 157–160.

[5] David M. Burton, *Elementary number theory*, University of New Hampshire (2010), 189.

[6] G. Mahesh and V. Sinari, *On the diophantine equation* $2^x + p^y = z^2$, Bull. Math. Stat. Res. **7** (2019), no. 2, 36–38.

[7] N. Burshtein, *All the solutions to an open problem of s. chotchaisthit on the diophantine equation* $2^x + p^y = z^2$ *when* p *are particular primes and* $y = 1$, Annals of Pure and Applied Mathematics **16** (2018), no. 1, 31–35.

[8] N. Viriyapong and C. Viriyapong, *On the diophantine equation* $n^x + 13^y = z^2$ *where* $n \equiv 2 (\mathrm{mod}\, 39)$ *and* $n + 1$ *is not a square number*, WSEAS Transactions on Mathematics **20** (2021), 442–445.

[9] _____, *On the diophantine equation* $n^x + 19^y = z^2$ *where* $n \equiv 2 (\mathrm{mod}\, 57)$, Int. J. Math. Com. Sci. **17** (2022), no. 4, 1639–1642.

[10] P. Mihailescu, *Primary cycolotomic units and a proof of catalan's conjecture*, J. Reine Angew. Math. **27** (2004), 167–195.

[11] S. Tadee, *On the diophantine equation* $2^x + p^y = z^2$ *where* $x \neq 1$ *and* $p \equiv 3 (\mathrm{mod}\, 4)$, Mathematical Journal by The Mathematical Association of Thailand Under The Patronage of His Majesty The King **67**, no. 707, 13–19.

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

All the Positive Solutions of $p^x - p^y = z^p$ in the Fibonacci and Lucas Numbers when $p = 2$ and $p = 3$

Phitthayathon Phetnun[†,‡]

Department of Mathematics, Faculty of Education
Kamphaeng Phet Rajabhat University, Kamphaeng Phet 62000, Thailand

Abstract

In 2023, Hashim investigated all positive solutions of the equation $2^x + 2^y = z^2$ in the Fibonacci and Lucas numbers. More recently, Tadee conducted a similar study, examining all positive solutions of the equation $3^x + 3^y = z^2$ in the Fibonacci and Lucas numbers. In this paper, we investigate all positive solutions of the equation $p^x - p^y = z^p$ in the Fibonacci and Lucas numbers when $p = 2$ and $p = 3$. We prove that $(x, y, z) \in \{(F_4, F_3, F_3), (F_4, F_3, L_0), (F_4, L_0, F_3), (F_4, L_0, L_0), (F_5, L_3, L_3), (L_2, F_3, F_3), (L_2, F_3, L_0), (L_2, L_0, F_3), (L_2, L_0, L_0)\}$ are the only nine positive solutions in the Fibonacci and Lucas numbers to the equation $2^x - 2^y = z^2$. Finally, we prove that the equation $3^x - 3^y = z^3$ has no positive solution in the Fibonacci and Lucas numbers.

Keywords: Diophantine equation, exponential Diophantine equation, Fibonacci number, Lucas number.

2020 MSC: Primary 11D61; Secondary 11B39.

1 Introduction

The study of the solvability of the Diophantine equations has been one of the most popular topics in mathematics. In 2007, Acu [1] investigated all nonnegative integer solutions of the exponential Diophantine equation $2^x + 5^y = z^2$. He proved that such an equation has exactly two solutions, namely $(x, y, z) \in \{(3, 0, 3), (2, 1, 3)\}$. Since then, there have been increasing interests in studying the solutions of a general form, which is $a^x \pm b^y = z^2$, where a and b are fixed positive integers. Some of these studies can be found in [2–4, 8, 9]. Let p be prime and x, y, z be positive integers. In 2019, Burshtein [2] proved that the Diophantine equation $p^x + p^y = z^2$ has no positive integer solution, except for the cases when $p = 2$ or $p = 3$. Burshtein found that the Diophantine equation $p^x - p^y = z^2$ has infinitely many solutions. In particular, for the case $p = 2$, all positive integer solutions of the equation $p^x - p^y = z^2$ are given by $(p, x, y, z) = (2, 2n + 1, 2n, 2^n)$, where n is a positive integer.

[†]Speaker. ‡Corresponding author.
Email: phitthayathon_p@kpru.ac.th (P. Phetnun)

In another direction, Diophantine equations connected to linear recurrence sequences have been widely studied by many mathematicians (see [6,7]). In 2023, Hashim [5] studied all positive solutions of the equation $p^x + p^y = z^2$ in the Fibonacci and Lucas numbers, when $p = 2$. In 2023, Tadee [10] studied all positive solutions of the equation $p^x + p^y = z^2$ in the Fibonacci and Lucas numbers, when $p = 3$.

Recently, Tadee [11] studied all nonnegative integer solutions of the Diophantine equations $p^x + p^y = z^q$ and $p^x - p^y = z^q$, where p and q are prime numbers. Tadee proved that all nonnegative integer solutions of the equation $p^x + p^y = z^q$ are $(p, q, x, y, z) = (2, q, qt + q - 1, qt + q - 1, 2^{t+1}), (2^q - 1, q, qt + 1, qt, 2(2^q - 1)^t), (2, 2, 2t + 3, 2t, 3 \cdot 2^t)$, where t is a nonnegative integer. All nonnegative integer solutions of the equation $p^x - p^y = z^q$ are $(p, q, x, y, z) = (p, q, t, t, 0), (2, q, qt + 1, qt, 2^t), (4v^2 + 1, 2, 2t + 1, 2t, 2v(4v^2 + 1)^t), (3, 3, 3t + 2, 3t, 2 \cdot 3^t)$, where t is a nonnegative integer and v is a positive integer. We are interested in the equation of the form $p^x - p^y = z^p$, where p is prime. In this paper, we find all positive solutions of the equation $p^x - p^y = z^p$ in the Fibonacci and Lucas numbers when $p = 2$ and $p = 3$.

2 Preliminaries

In this section, we first review the essential preliminaries as follows:

Theorem 2.1. *[2] All positive integer solutions of the equation $2^x - 2^y = z^2$ are given by $(x, y, z) = (2n + 1, 2n, 2^n)$, where n is a positive integer.*

Theorem 2.2. *[11] All positive integer solutions of the equation $3^x - 3^y = z^3$ are given by $(x, y, z) = (3n + 2, 3n, 2 \cdot 3^n)$, where n is a positive integer.*

Proposition 2.3. *[12] Let n be a positive integer. Then*

(i) $F_{n+1} + F_{n-1} = L_n$,

(ii) $F_{n+2} - F_{n-2} = L_n$ *whenever* $n > 1$.

Lemma 2.4. *[10] If i is a positive integer with $i \geq 7$, then $F_i \geq L_j + 2$ for all positive integers $j \leq i - 2$.*

Lemma 2.5. *If i is a positive integer with $i \geq 8$, then $F_i \geq L_j + 3$ for all positive integers $j \leq i - 2$.*

Proof. We prove by induction on i. It is easy to verify that $F_8 \geq L_j + 3$ for all positive integers $j \leq 6$. Assume that $F_i \geq L_j + 3$ for all positive integers $j \leq i - 2$. Then $F_{i+1} \geq F_i \geq L_j + 3$ for all positive integers $j \leq i - 2$. If $j = i - 1$, then it follows from Proposition 2.3(ii) and $i \geq 8$ that $L_j + 3 = L_{i-1} + 3 = F_{i+1} - F_{i-3} + 3 \leq F_{i+1} - 2 < F_{i+1}$.

□

3 Main Results

In this section, we find all positive solutions of the equation $p^x - p^y = z^p$ in the Fibonacci and Lucas numbers when $p = 2$ and $p = 3$. In other words, we first solve the following equations:

$$2^{F_i} - 2^{F_j} = F_k^2 \tag{3.1}$$
$$2^{F_i} - 2^{F_j} = L_k^2 \tag{3.2}$$
$$2^{F_i} - 2^{L_j} = F_k^2 \tag{3.3}$$
$$2^{F_i} - 2^{L_j} = L_k^2 \tag{3.4}$$
$$2^{L_i} - 2^{F_j} = F_k^2 \tag{3.5}$$
$$2^{L_i} - 2^{F_j} = L_k^2 \tag{3.6}$$
$$2^{L_i} - 2^{L_j} = F_k^2 \tag{3.7}$$
$$2^{L_i} - 2^{L_j} = L_k^2 \tag{3.8}$$

and then solve the following equations:

$$3^{F_i} - 3^{F_j} = F_k^3 \tag{3.9}$$
$$3^{F_i} - 3^{F_j} = L_k^3 \tag{3.10}$$
$$3^{F_i} - 3^{L_j} = F_k^3 \tag{3.11}$$
$$3^{F_i} - 3^{L_j} = L_k^3 \tag{3.12}$$
$$3^{L_i} - 3^{F_j} = F_k^3 \tag{3.13}$$
$$3^{L_i} - 3^{F_j} = L_k^3 \tag{3.14}$$
$$3^{L_i} - 3^{L_j} = F_k^3 \tag{3.15}$$
$$3^{L_i} - 3^{L_j} = L_k^3, \tag{3.16}$$

where the indices i, j and k are nonnegative integers, and F_n and L_n represent the nth terms of the Fibonacci and Lucas sequences, respectively, that are defined by the initial values $F_0 = 0, F_1 = 1$ and $L_0 = 2, L_1 = 1$ and the recurrence relations $F_n = F_{n-1} + F_{n-2}$ and $L_n = L_{n-1} + L_{n-2}$, where $n \geq 2$. For the convenience of the reader, we exhibit some values of F_n and L_n for $n \in \{0, 1, 2, ..., 10\}$, as shown in the following table.

n	0	1	2	3	4	5	6	7	8	9	10
F_n	0	1	1	2	3	5	8	13	21	34	55
L_n	2	1	3	4	7	11	18	29	47	76	123

Now, we present our main results as the following theorems.

Theorem 3.1. *The equation $2^x - 2^y = z^2$ has only nine positive solutions in the Fibonacci and Lucas numbers which are $(x, y, z) \in \{(F_4, F_3, F_3), (F_4, F_3, L_0), (F_4, L_0, F_3), (F_4, L_0, L_0), (F_5, L_3, L_3), (L_2, F_3, F_3), (L_2, F_3, L_0), (L_2, L_0, F_3), (L_2, L_0, L_0)\}$.*

Proof. We first consider (3.1) and (3.2). It follows from Theorem 2.1 that $F_i = 2n+1$, $F_j = 2n$, and $F_k = 2^n = L_k$ for some positive integer n. Since $F_i > F_j \geq 2$, and $2n + 1$ and $2n$ are consecutive integers in the Fibonacci sequence, n must be 1. Consequently, $i = 4$, $j = 3$, and $F_k = 2 = L_k$. Note that $F_k = 2$ if and only if $k = 3$. Similarly, $L_k = 2$ if and only if $k = 0$. Thus, we obtain that (F_4, F_3, F_3) and (F_4, F_3, L_0) are solutions of (3.1) and (3.2), respectively.

Next, we consider (3.3) and (3.4). By Theorem 2.1, we obtain that $F_i = 2n + 1$, $L_j = 2n$, and $F_k = 2^n = L_k$ for some positive integer n. Then $F_i = L_j + 1$. This implies that $j \leq i - 1$. If $i \geq 7$, then it follows from Lemma 2.4 that $j = i - 1$. Hence, we have $F_i - 1 = L_{i-1}$. By

Proposition 2.3(i), we obtain $F_i - 1 = F_i + F_{i-2}$, which means $F_{i-2} = -1$, a contradiction. Thus $i < 7$. Since $i > j$, we get $j \leq 5$. Since $L_j = 2n$ and $j \leq 5$, n must be 1 or 2. If $n = 1$, then we obtain that (F_4, L_0, F_3) and (F_4, L_0, L_0) are solutions of (3.3) and (3.4), respectively. If $n = 2$, then we obtain that (F_5, L_3, L_3) is a solution of (3.4).

Now, we consider (3.5) and (3.6). By Theorem 2.1, we obtain that $L_i = 2n + 1$, $F_j = 2n$, and $F_k = 2^n = L_k$ for some positive integer n. Then $L_i = F_j + 1$. By Proposition 2.3(i), we obtain $F_{i+1} + F_{i-1} = F_j + 1$. Since $F_{i+1} = F_i + F_{i-1}$, we consequently have $2F_{i-1} - 1 = F_j - F_i$. If $i \geq j$, then $2F_{i-1} - 1 = F_j - F_i \leq 0$. Thus $i = 1$ and so $n = 0$. This contradicts the fact that n is a positive integer. Hence $i < j$. If $j \geq 7$, then Lemma 2.4 implies that $i = j - 1$ and so $L_{j-1} = F_j + 1$. By Proposition 2.3(i), we get $F_j + F_{j-2} = F_j + 1$, that is, $F_{j-2} = 1$. This is impossible because $j \geq 7$. Hence $j < 7$. Since $i < j$, we get $i \leq 5$. One can verify that it is possible when $i = 2$, $n = 1$, and $j = 3$. Thus, we obtain that (L_2, F_3, F_3) and (L_2, F_3, L_0) are solutions of (3.5) and (3.6), respectively.

Finally, we consider (3.7) and (3.8). By Theorem 2.1, we obtain that $L_i = 2n + 1$, $L_j = 2n$, and $F_k = 2^n = L_k$ for some positive integer n. Then $L_i - L_j = 1$. That is, $i = 2$, $j = 0$, and $n = 1$. Hence $F_k = 2 = L_k$. Thus, we infer that (L_2, L_0, F_3) and (L_2, L_0, L_0) are solutions of (3.7) and (3.8), respectively. This completes the proof. □

Theorem 3.2. *The equation $3^x - 3^y = z^3$ has no positive solution in the Fibonacci and Lucas numbers.*

Proof. We first consider (3.9) and (3.10). It follows from Theorem 2.2 that $F_i = 3n+2$, $F_j = 3n$, and $F_k = 2 \cdot 3^n = L_k$ for some positive integer n. Then $F_i - F_j = 2$. This implies that $i = 5$, $j = 4$, and $n = 1$. This yields $F_k = 6 = L_k$, which is impossible. Hence, (3.9)-(3.10) have no solution.

Next, we consider (3.11) and (3.12). By Theorem 2.2, we obtain that $F_i = 3n + 2$, $L_j = 3n$, and $F_k = 2 \cdot 3^n = L_k$ for some positive integer n. Then $F_i = L_j + 2$. This implies that $j \leq i - 2$. If $i \geq 8$, then by Lemma 2.5, we have $F_i \geq L_j + 3 > L_j + 2$. This contradicts the fact that $F_i = L_j + 2$. If $i = 7$, then we obtain $3n + 2 = F_i = 13$, which is a contradiction. Hence $i < 7$. Since $i > j$, we get $j \leq 5$. For such a case, one can verify that (3.11)-(3.12) have no solution.

Now, we consider (3.13) and (3.14). By Theorem 2.2, we obtain that $L_i = 3n + 2$, $F_j = 3n$, and $F_k = 2 \cdot 3^n = L_k$ for some positive integer n. Then $L_i = F_j + 2$. By Proposition 2.3(i), we obtain $F_{i+1} + F_{i-1} = F_j + 2$. Since $F_{i+1} = F_i + F_{i-1}$, we consequently have $2F_{i-1} - 2 = F_j - F_i$. If $i \geq j$, then $2F_{i-1} - 2 = F_j - F_i \leq 0$. Thus $i \in \{1, 2, 3\}$ and so $3n + 2 = L_i \in \{1, 3, 4\}$, which is impossible. Thus $i < j$. Now, suppose that $j \geq 7$. If $i \leq j - 2$, then by Lemma 2.4, we have $F_j \geq L_i + 2 > L_i$. This contradicts the fact that $L_i = F_j + 2 > F_j$. If $i = j - 1$, then $L_{j-1} = F_j + 2$. By Proposition 2.3(i), we get $F_j + F_{j-2} = F_j + 2$, that is, $F_{j-2} = 2$. This is impossible because $j \geq 7$. Hence $j < 7$. Since $i < j$, we get $i \leq 5$. For such a case, one can verify that (3.13)-(3.14) have no solution.

Finally, we consider (3.15) and (3.16). By Theorem 2.2, we obtain that $L_i = 3n+2$, $L_j = 3n$, and $F_k = 2 \cdot 3^n = L_k$ for some positive integer n. Then $L_i \geq 5$ and $L_i - L_j = 2$. It is obvious that such a case is impossible. Hence, (3.15)-(3.16) have no solution. □

4 Conclusion

To summarize, within the Fibonacci and Lucas numbers, we identified precisely nine positive solutions to the equation $2^x - 2^y = z^2$, as described. Additionally, we confirmed the absence of positive solutions to the equation $3^x - 3^y = z^3$ within such numbers. These results underscore the valuable relationship between the Fibonacci and Lucas numbers and certain exponential Diophantine equations.

5 Suggestion

Let p and q be prime numbers. According to [11], it is known that the Diophantine equations $p^x \pm p^y = z^q$ have infinitely many solutions, as described before. Thus, to extend earlier results in [5] and our own, the reader may study all positive solutions of the equations $p^x \pm p^y = z^q$ in the Fibonacci and Lucas numbers.

Acknowledgment. The author is grateful to the referees for their careful reading of the manuscript and their useful comments.

References

[1] D. Acu, *On a Diophantine equation* $2^x + 5^y = z^2$, Gen. Math. **15**(4) (2007), 145–148.

[2] N. Burshtein, *All the solutions of the Diophantine equations* $p^x + p^y = z^2$ *and* $p^x - p^y = z^2$ *when* $p \geq 2$ *is prime*, Annals of Pure and Applied Mathematics **19**(2) (2019), 111–119.

[3] N. Burshtein, *A short note on solutions of the Diophantine equations* $6^x + 11^y = z^2$ *and* $6^x - 11^y = z^2$ *in positive integers* x, y, z, Annals of Pure and Applied Mathematics **19**(2) (2019), 55–56.

[4] S. Chotchaisthit, *On the Diophantine equation* $4^x + p^y = z^2$ *where* p *is a prime number*, Amer. J. Math. Sci. **1**(1) (2012), 191–193.

[5] H. R. Hashim, *On the solutions of* $2^x + 2^y = z^2$ *in the Fibonacci and Lucas numbers*, J. Prime Res. Math. **19**(1) (2023), 27–33.

[6] H. R. Hashim, *Solutions of the Diophantine equation* $7X^2 + Y^7 = Z^2$ *from recurrence sequences*, Commun. Math. **28**(1) (2020), 55–66.

[7] H. R. Hashim and Sz. Tengely, *Representations pf reciprocals of Lucas sequences*, Miskolc Math. **19**(2) (2018), 865–872.

[8] B. Sroysang, *On the Diophantine equation* $3^x + 5^y = z^2$, Int. J. Pure Appl. Math. **81**(4) (2012), 605–608.

[9] A. Suvarnamani, *Solutions of the Diophantine equation* $2^x + p^y = z^2$, Int. J. Math. Sci. Appl. **1**(3) (2011), 1415–1419.

[10] S. Tadee, *On the positive integer solutions of* $p^x + p^y = z^2$ *in the Fibonacci and Lucas numbers, where* p *is prime*, Int. J. Math. Comput. Sci. **19**(2) (2024), 377–380.

[11] S. Tadee, *The solutions of the Diophantine equations* $p^x + p^y = z^q$ *and* $p^x - p^y = z^q$, Int. J. Math. Comput. Sci. **19**(3) (2024), 621–623.

[12] N. N. Vorobiev, *Fibonacci numbers*, Birkhäuser, Basel, 2002.

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Some Properties of $k-$Narayana Quaternions

Chansouk Sikhammountri[1,†] and Narawadee Phudolsitthiphat[2,‡]

[1]Teaching Mathematics, Faculty of Science,
Chiang Mai University, Chiang Mai 50200, Thailand

[2]Department of Mathematics, Faculty of Science
Chiang Mai University, Chiang Mai 50200, Thailand

Abstract

We introduce $k-$Narayana quaternions and present several properties of these numbers, including but not limited to the Binet formulas, generating functions, and summation formulas. Our results extend and generalize some well-known theorems in this area.

Keywords: $k-$Narayana sequence, quaternions, generating function.

2020 MSC: Primary 05A15; Secondary 11B37, 11B39.

1 Introduction

Numerous researchers have dedicated their time and effort to the study of number sequences due to their widespread utility. Many applications of integer sequences, including Fibonacci, $k-$Fibonacci, Lucas, Jacobsthal, $k-$Jacobthals, Pell, and $k-$Pell have been utilized in various scientific fields such as engineering and architecture.

In the 14th century, Narayana, an Indian mathematician, delved into the exploration of Narayana numbers, as indicated in [8]. Comparable to Fibonacci's rabbit problem, Narayana's cow problem involves the reproductive pattern of cows. In this scenario, cows commence calving in the fourth year, each delivering one calf annually thereafter. The objective revolves around determining the cumulative number of offspring produced within a span of 20 years. This problem can be solved in a similar way to how Fibonacci addressed the rabbit problem. If n is the year, the Narayana problem can be expressed in the form of a recursive relationship as follows:

$$N_n = N_{n-1} + N_{n-3}, \quad n \geq 3$$

where $N_0 = 0, N_1 = 1, N_2 = 1$.
The first 11 Narayana terms are: 0, 1, 1, 1, 2, 3, 4, 6, 9, 13, 19, 28 ...This sequence is called the Narayana sequence, also known as the Fibonacci-Narayana sequence or the Narayana cow

*This research was financially supported by The Royal Thai Government Scholarship under Thailand - Lao PDR Bilateral Development Cooperation.
†Speaker. ‡Corresponding author.
Email: nousikhammountry@gmail.com (C. Sikhammountri), narawadee_n@hotmail.co.th (N. Phudolsitthiphat).

sequence.

For any nonzero integer number k, Ramfrez and Sirvent [7] defined the $k-$Narayana number as follows:

$$N_{k,n} = kN_{k,n-1} + N_{k,n-3}, \quad n \geq 3$$

where $N_{k,0} = 0, N_{k,1} = 1, N_{k,2} = k$.

It is mentioned in [10] that Quaternions are four-dimensional hypercomplex numbers. They was introduced by Sir William Rowan Hamilton and have found widespread use in high-tech areas such as computer graphics, signal processing, and robotics, among others.

Quaternions form a four-dimensional non-commutative associative algebra over the real numbers and are defined as follows:

$$\mathbf{H} = \{q = q_0 + q_1\boldsymbol{e_1} + q_2\boldsymbol{e_2} + q_3\boldsymbol{e_3} | q_0, q_1, q_2, q_3 \in R\}$$

where the imaginary units $\boldsymbol{e_1}, \boldsymbol{e_2}$ and $\boldsymbol{e_3}$ satisfy the following equalities:

$$\boldsymbol{e_1^2} = \boldsymbol{e_2^2} = \boldsymbol{e_3^2} = \boldsymbol{e_1e_2e_3} = -1, \boldsymbol{e_1e_2} = \boldsymbol{e_3} = -\boldsymbol{e_2e_1}, \boldsymbol{e_2e_3} = \boldsymbol{e_1} = -\boldsymbol{e_3e_2},$$
$$\boldsymbol{e_3e_1} = \boldsymbol{e_2} = -\boldsymbol{e_1e_3}.$$

For more details on quaternions, one can refer to, for example, ([3], [9]).
Let $p = p_0 + p_1\boldsymbol{e_1} + p_2\boldsymbol{e_2} + p_3\boldsymbol{e_3}$ and $q = q_0 + q_1\boldsymbol{e_1} + q_2\boldsymbol{e_2} + q_3\boldsymbol{e_3}$ be two quaternions. The addition and subtraction of two quaternions are defined as:

$$p \pm q = (p_0 \pm q_0) + (p_1 \pm q_1)\boldsymbol{e_1} + (p_2 \pm q_2)\boldsymbol{e_2} + (p_3 \pm q_3)\boldsymbol{e_3}.$$

The multiplication of quaternion by a real scalar λ is defined as:

$$\lambda p = \lambda p_0 + \lambda p_1\boldsymbol{e_1} + \lambda p_2\boldsymbol{e_2} + \lambda p_3\boldsymbol{e_3}.$$

The multiplication of two quaternions is defined as:

$$pq = (p_0q_0 - p_0q_1 - p_0q_2 - p_0q_3) + (p_0q_1 + p_1q_0 + p_2q_3 - p_3q_2)\boldsymbol{e_1}$$
$$+ (p_0q_2 - p_1q_3 + p_2q_0 + p_3q_1)\boldsymbol{e_2} + (p_0q_3 + p_1q_2 - p_2q_1 + p_3q_0)\boldsymbol{e_3}.$$

The conjugate of quaternion q is denoted by $\overline{q}$ and defined as:

$$\overline{q} = q_0 - q_1\boldsymbol{e_1} - q_2\boldsymbol{e_2} - q_3\boldsymbol{e_3}.$$

Horadam [4] introduced the complex Fibonacci numbers and Fibonacci quaternions as follows:

$$QF_n = F_n + F_{n+1}\boldsymbol{e_1} + F_{n+2}\boldsymbol{e_2} + F_{n+3}\boldsymbol{e_3},$$

where F_n is the nth Fibonacci number.

In 2012 Halici [2] proved some theorems related to the Fibonacci quaternion. The Fibonacci quaternion sequence has been extensively studied. Ipek [5] introduced the $(p, q)-$Fibonacci quaternion as follows:

$$Q\mathcal{F}_n = \mathcal{F}_n + \mathcal{F}_{n+1}\boldsymbol{e_1} + \mathcal{F}_{n+2}\boldsymbol{e_2} + \mathcal{F}_{n+3}\boldsymbol{e_3},$$

where $\mathcal{F}_n$ is the nth $(p, q)-$Fibonacci number.

Flaut [1] introducted the Narayana quaternion, defined as follows:

$$U_n = N_n + N_{n+1}\boldsymbol{e_2} + N_{n+2}\boldsymbol{e_3} + N_{n+3}\boldsymbol{e_4},$$

where N_n are the nth Narayana number.

In this paper, we introduce $k-$Narayana quaternions and explore various properties, including, but not limited to, Binet formulas, generating functions, and summation formulas. Our results extend and generalize some well-known theorems in this area.

2 Preliminaries

Ramírez and Sirvent [7] determined the Binet formula for the $k-$Narayana number as follows:

$$N_{k,n} = \frac{\alpha_k^{n+1}}{(\alpha_k - \beta_k)(\alpha_k - \gamma_k)} + \frac{\beta_k^{n+1}}{(\beta_k - \alpha_k)(\beta_k - \gamma_k)} + \frac{\gamma_k^{n+1}}{(\gamma_k - \alpha_k)(\gamma_k - \beta_k)}, \tag{2.1}$$

where α_k, β_k, and γ_k are the roots of the characteristic equation $x^3 - kx^2 - 1 = 0$.

Özkan, Kuloğlu, and Peters [6] obtained the formula for the sum of the first $(n+1)$ terms of the $k-$Narayana sequence as follows:

$$\sum_{m=0}^{n} N_{k,m} = \frac{N_{k,1} + N_{k,n} + N_{k,n+1} + N_{k,n+2}}{k} - N_{k,n+1}. \tag{2.2}$$

3 Main Results

First, we will define the $k-$Narayana quaternions.

Definition 3.1. The $k-$Narayana quaternions are defined in the following manner:

$$QN_{k,n} = N_{k,n} + N_{k,n+1}\boldsymbol{e_1} + N_{k,n+2}\boldsymbol{e_2} + N_{k,n+3}\boldsymbol{e_3}, \quad n \geq 0 \tag{3.1}$$

where $\boldsymbol{e_1}$, $\boldsymbol{e_2}$, and $\boldsymbol{e_3}$ are the imaginary units satisfy the following equalities:

$$\boldsymbol{e_1^2} = \boldsymbol{e_2^2} = \boldsymbol{e_3^2} = \boldsymbol{e_1 e_2 e_3} = -1, \boldsymbol{e_1 e_2} = \boldsymbol{e_3} = -\boldsymbol{e_2 e_1}, \boldsymbol{e_2 e_3} = \boldsymbol{e_1} = -\boldsymbol{e_3 e_2},$$
$$\boldsymbol{e_3 e_1} = \boldsymbol{e_2} = -\boldsymbol{e_1 e_3}.$$

The first few k-Narayana quaternions can be written as follows:

$$\begin{aligned}
QN_{k,0} &= N_{k,0} + N_{k,1}\boldsymbol{e_1} + N_{k,2}\boldsymbol{e_2} + N_{k,3}\boldsymbol{e_3} \\
&= \boldsymbol{e_1} + k\boldsymbol{e_2} + k^2\boldsymbol{e_3} \\
QN_{k,1} &= N_{k,1} + N_{k,2}\boldsymbol{e_1} + N_{k,3}\boldsymbol{e_2} + N_{k,4}\boldsymbol{e_3} \\
&= 1 + k\boldsymbol{e_1} + k^2\boldsymbol{e_2} + (k^3+1)\boldsymbol{e_3} \\
QN_{k,2} &= N_{k,2} + N_{k,3}\boldsymbol{e_1} + N_{k,4}\boldsymbol{e_2} + N_{k,5}\boldsymbol{e_3} \\
&= k + k^2\boldsymbol{e_1} + (k^3+1)\boldsymbol{e_2} + (k^4+2k)\boldsymbol{e_3} \\
&\vdots
\end{aligned}$$

For $n, m \geq 0$, let $QN_{k,n} = N_{k,n} + N_{k,n+1}\boldsymbol{e_1} + N_{k,n+2}\boldsymbol{e_2} + N_{k,n+3}\boldsymbol{e_3}$ and $QN_{k,m} = N_{k,m} + N_{k,m+1}\boldsymbol{e_1} + N_{k,m+2}\boldsymbol{e_2} + N_{k,m+3}\boldsymbol{e_3}$ be two $k-$Narayana quaternions. The addition and subtraction of these two $k-$Narayana quaternions are defined as follows:

$$\begin{aligned}
QN_{k,n} \pm QN_{k,m} =& (QN_{k,n} \pm QN_{k,m}) + (QN_{k,n+1} \pm QN_{k,m+1})\boldsymbol{e_1} \\
&+ (QN_{k,n+2} \pm QN_{k,m+2})\boldsymbol{e_2} + (QN_{k,n+3} \pm QN_{k,m+3})\boldsymbol{e_3}.
\end{aligned}$$

The multiplication of a $k-$Narayana quaternion by a real scalar λ is defined as:

$$\lambda QN_{k,n} = \lambda QN_{k,n} + \lambda QN_{k,n+1}\boldsymbol{e_1} + \lambda QN_{k,n+2}\boldsymbol{e_2} + \lambda QN_{k,n+3}\boldsymbol{e_3}.$$

We can see that the set of k-Narayana quaternions forms a vector space over the field $\mathbb{R}$.
The multiplication of two $k-$Narayana quaternions is defined as:

$$\begin{aligned}
QN_{k,n}QN_{k,m} =& (QN_{k,n}QN_{k,m} - QN_{k,n+1}QN_{k,m+1} - QN_{k,n+2}QN_{k,m+2} - QN_{k,n+3}QN_{k,m+3}) \\
&+ (QN_{k,n}QN_{k,m+1} + QN_{k,n+1}QN_{k,m} + QN_{k,n+2}QN_{k,m+3} - QN_{k,n+3}QN_{k,m+2})\boldsymbol{e_1} \\
&+ (QN_{k,n}QN_{k,m+2} - QN_{k,n+1}QN_{k,m+3} + QN_{k,n+2}QN_{k,m} + QN_{k,n+3}QN_{k,m+1})\boldsymbol{e_2} \\
&+ (QN_{k,n}QN_{k,m+3} + QN_{k,n+1}QN_{k,m+2} - QN_{k,n+2}QN_{k,m+1} + QN_{k,n+3}QN_{k,m})\boldsymbol{e_3}.
\end{aligned}$$

The conjugate of a $k-$Narayana quaternion $QN_{k,n}$ is denoted by $\overline{QN_{k,n}}$ and is defined as follows:

$$\overline{QN_{k,n}} = N_{k,n} - N_{k,n+1}\boldsymbol{e_1} - N_{k,n+2}\boldsymbol{e_2} - N_{k,n+3}\boldsymbol{e_3}.$$

The next theorem considers the addition and subtraction of k-Narayana quaternions and their conjugates.

Theorem 3.2. *Let $n \geq 0$. Then*

$$QN_{k,n} + \overline{QN_{k,n}} = 2N_{k,n} \tag{3.2}$$
$$QN_{k,n} - \overline{QN_{k,n}} = 2N_{k,n+1}\boldsymbol{e_1} + 2N_{k,n+2}\boldsymbol{e_2} + 2N_{k,n+3}\boldsymbol{e_3}. \tag{3.3}$$

Proof.

$$\begin{aligned}
QN_{k,n} + \overline{QN_{k,n}} =& (N_{k,n} + N_{k,n+1}\boldsymbol{e_1} + N_{k,n+2}\boldsymbol{e_2} + N_{k,n+3}\boldsymbol{e_3}) \\
&+ (N_{k,n} - N_{k,n+1}\boldsymbol{e_1} - N_{k,n+2}\boldsymbol{e_2} - N_{k,n+3}\boldsymbol{e_3}) \\
=& N_{k,n} + N_{k,n+1}\boldsymbol{e_1} + N_{k,n+2}\boldsymbol{e_2} + N_{k,n+3}\boldsymbol{e_3} + N_{k,n} \\
&- N_{k,n+1}\boldsymbol{e_1} - N_{k,n+2}\boldsymbol{e_2} - N_{k,n+3}\boldsymbol{e_3} \\
=& 2N_{k,n}. \\
QN_{k,n} - \overline{QN_{k,n}} =& (N_{k,n} + N_{k,n+1}\boldsymbol{e_1} + N_{k,n+2}\boldsymbol{e_2} + N_{k,n+3}\boldsymbol{e_3}) \\
&- (N_{k,n} - N_{k,n+1}\boldsymbol{e_1} - N_{k,n+2}\boldsymbol{e_2} - N_{k,n+3}\boldsymbol{e_3}) \\
=& N_{k,n} + N_{k,n+1}\boldsymbol{e_1} + N_{k,n+2}\boldsymbol{e_2} + N_{k,n+3}\boldsymbol{e_3} - N_{k,n} \\
&+ N_{k,n+1}\boldsymbol{e_1} + N_{k,n+2}\boldsymbol{e_2} + N_{k,n+3}\boldsymbol{e_3} \\
=& 2N_{k,n+1}\boldsymbol{e_1} + 2N_{k,n+2}\boldsymbol{e_2} + 2N_{k,n+3}\boldsymbol{e_3}.
\end{aligned}$$

□

The following theorem states the multiplication of a k-Narayana quaternion and its conjugate.

Theorem 3.3. *Let $n \geq 0$ be an integer. The character of the $k-$Narayana quaternion number is given by*

$$QN_{k,n}\overline{QN_{k,n}} = N_{k,n}^2 + N_{k,n+1}^2 + N_{k,n+2}^2 + N_{k,n+3}^2.$$

Proof. From Definition 3.1, we get

$$\begin{aligned}
QN_{k,n}\overline{QN_{k,n}} =& (N_{k,n} + N_{k,n+1}\boldsymbol{e_1} + N_{k,n+2}\boldsymbol{e_2} + N_{k,n+3}\boldsymbol{e_3}) \\
&(N_{k,n} - N_{k,n+1}\boldsymbol{e_1} - N_{k,n+2}\boldsymbol{e_2} - N_{k,n+3}\boldsymbol{e_3}) \\
=& N_{k,n}^2 - N_{k,n}N_{k,n+1}\boldsymbol{e_1} - N_{k,n}N_{k,n+2}\boldsymbol{e_2} - N_{k,n}N_{k,n+3}\boldsymbol{e_3} \\
&+ N_{k,n+1}\boldsymbol{e_1}N_{k,n} - N_{k,n+1}^2\boldsymbol{e_1^2} - N_{k,n+1}\boldsymbol{e_1}N_{k,n+2}\boldsymbol{e_2} \\
&- N_{k,n+1}\boldsymbol{e_1}N_{k,n+3}\boldsymbol{e_3} + N_{k,n+2}\boldsymbol{e_2}N_{k,n} - N_{k,n+2}\boldsymbol{e_2}N_{k,n+1}\boldsymbol{e_1} \\
&- N_{k,n+2}^2\boldsymbol{e_2^2} - N_{k,n+2}\boldsymbol{e_2}N_{k,n+3}\boldsymbol{e_3} + N_{k,n+3}\boldsymbol{e_3}N_{k,n} \\
&- N_{k,n+3}\boldsymbol{e_3}N_{k,n+1}\boldsymbol{e_1} - N_{k,n+3}\boldsymbol{e_3}N_{k,n+2}\boldsymbol{e_2} - N_{k,n+3}^2\boldsymbol{e_3^2} \\
=& N_{k,n}^2 + N_{k,n+1}^2 - N_{k,n+1}N_{k,n+2}\boldsymbol{e_3} + N_{k,n+1}N_{k,n+3}\boldsymbol{e_2} \\
&+ N_{k,n+2}N_{k,n+1}\boldsymbol{e_3} + N_{k,n+2}^2 - N_{k,n+2}N_{k,n+3}\boldsymbol{e_1} \\
&- N_{k,n+3}N_{k,n+1}\boldsymbol{e_2} + N_{k,n+3}N_{k,n+2}\boldsymbol{e_1} + N_{k,n+3}^2 \\
=& N_{k,n}^2 + N_{k,n+1}^2 + N_{k,n+2}^2 + N_{k,n+3}^2.
\end{aligned}$$

□

We can define the norm of k-Narayana quaternions as follows:

Corollary 3.4. *The norm of the $k-$Narayana quaternion sequence is:*

$$\| QN_{k,n} \| = \sqrt{N_{k,n}^2 + N_{k,n+1}^2 + N_{k,n+2}^2 + N_{k,n+3}^2}.$$

Proof. By Theorem 3.3, we have

$$\| QN_{k,n} \| = \sqrt{(QN_{k,n}\overline{QN_{k,n}})} = \sqrt{N_{k,n}^2 + N_{k,n+1}^2 + N_{k,n+2}^2 + N_{k,n+3}^2}.$$

□

Next, we present the Binet formula for the $k-$Narayana quaternion.

Theorem 3.5. *Let $n \geq 0$. Then*

$$\begin{aligned} QN_{k,n} =& \frac{\alpha_k^{n+1}(1 + \alpha_k \boldsymbol{e_1} + \alpha_k^2 \boldsymbol{e_2} + \alpha_k^3 \boldsymbol{e_3})}{(\alpha_k - \beta_k)(\alpha_k - \gamma_k)} + \frac{\beta_k^{n+1}(1 + \beta_k \boldsymbol{e_1} + \beta_k^2 \boldsymbol{e_2} + \beta_k^3 \boldsymbol{e_3})}{(\beta_k - \alpha_k)(\beta_k - \gamma_k)} \\ &+ \frac{\gamma_k^{n+1}(1 + \gamma_k \boldsymbol{e_1} + \gamma_k^2 \boldsymbol{e_2} + \gamma_k^3 \boldsymbol{e_3})}{(\gamma_k - \alpha_k)(\gamma_k - \beta_k)}. \end{aligned}$$

Proof. From Binet's formula for $k-$Narayana number (2.1), we get

$$\begin{aligned} QN_{k,n} =& N_{k,n} + N_{k,n+1}\boldsymbol{e_1} + N_{k,n+2}\boldsymbol{e_2} + N_{k,n+3}\boldsymbol{e_3}. \\ =& \left(\frac{\alpha_k^{n+1}}{(\alpha_k - \beta_k)(\alpha_k - \gamma_k)} + \frac{\beta_k^{n+1}}{(\beta_k - \alpha_k)(\beta_k - \gamma_k)} + \frac{\gamma_k^{n+1}}{(\gamma_k - \alpha_k)(\gamma_k - \beta_k)}\right) \\ &+ \left(\frac{\alpha_k^{n+2}}{(\alpha_k - \beta_k)(\alpha_k - \gamma_k)} + \frac{\beta_k^{n+2}}{(\beta_k - \alpha_k)(\beta_k - \gamma_k)} + \frac{\gamma_k^{n+2}}{(\gamma_k - \alpha_k)(\gamma_k - \beta_k)}\right)\boldsymbol{e_1} \\ &+ \left(\frac{\alpha_k^{n+3}}{(\alpha_k - \beta_k)(\alpha_k - \gamma_k)} + \frac{\beta_k^{n+3}}{(\beta_k - \alpha_k)(\beta_k - \gamma_k)} + \frac{\gamma_k^{n+3}}{(\gamma_k - \alpha_k)(\gamma_k - \beta_k)}\right)\boldsymbol{e_2} \\ &+ \left(\frac{\alpha_k^{n+4}}{(\alpha_k - \beta_k)(\alpha_k - \gamma_k)} + \frac{\beta_k^{n+4}}{(\beta_k - \alpha_k)(\beta_k - \gamma_k)} + \frac{\gamma_k^{n+4}}{(\gamma_k - \alpha_k)(\gamma_k - \beta_k)}\right)\boldsymbol{e_3} \\ =& \frac{\alpha_k^{n+1}}{(\alpha_k - \beta_k)(\alpha_k - \gamma_k)}(1 + \alpha_k \boldsymbol{e_1} + \alpha_k^2 \boldsymbol{e_2} + \alpha_k^3 \boldsymbol{e_3}) + \frac{\beta_k^{n+1}}{(\beta_k - \alpha_k)(\beta_k - \gamma_k)} \\ & (1 + \beta_k \boldsymbol{e_1} + \beta_k^2 \boldsymbol{e_2} + \beta_k^3 \boldsymbol{e_3}) + \frac{\gamma_k^{n+1}}{(\gamma_k - \alpha_k)(\gamma_k - \beta_k)}(1 + \gamma_k \boldsymbol{e_1} + \gamma_k^2 \boldsymbol{e_2} + \gamma_k^3 \boldsymbol{e_3}). \end{aligned}$$

□

The following theorem states that the finite sum of k-Narayana quaternions is (2.2), we get.

Theorem 3.6. *Let $n \geq 0$. Then*

$$\begin{aligned} \sum_{m=0}^{n} QN_{k,m} =& \frac{1 + \boldsymbol{e_1} + \boldsymbol{e_2} + \boldsymbol{e_3} + QN_{k,n} + QN_{k,n+1} + QN_{k,n+2} - kQN_{k,n+1}}{k} \\ & - \boldsymbol{e_2} - \boldsymbol{e_3}(1 + k). \end{aligned}$$

Proof. By (2.2), we have

$$\begin{aligned}
\sum_{m=0}^{n} QN_{k,m} =& QN_{k,0} + QN_{k,1} + QN_{k,2} + \ldots + QN_{k,n} \\
=& (N_{k,0} + N_{k,1}\boldsymbol{e_1} + N_{k,2}\boldsymbol{e_2} + N_{k,3}\boldsymbol{e_3}) \\
&+ (N_{k,1} + N_{k,2}\boldsymbol{e_1} + N_{k,3}\boldsymbol{e_2} + N_{k,4}\boldsymbol{e_3}) + \ldots \\
&+ (N_{k,n} + N_{k,n+1}\boldsymbol{e_1} + N_{k,n+2}\boldsymbol{e_2} + N_{k,n+3}\boldsymbol{e_3}) \\
=& (N_{k,0} + N_{k,1} + N_{k,2} + \ldots + N_{k,n}) \\
&+ (N_{k,1} + N_{k,2} + N_{k,3} + \ldots + N_{k,n+1})\boldsymbol{e_1} \\
&+ (N_{k,2} + N_{k,3} + N_{k,4} + \ldots + N_{k,n+2})\boldsymbol{e_2} \\
&+ (N_{k,3} + N_{k,4} + N_{k,5} + \ldots + N_{k,n+3})\boldsymbol{e_3} \\
=& (\sum_{m=0}^{n} N_{k,m}) + \boldsymbol{e_1}(\sum_{m=0}^{n+1} N_{k,m}) + \boldsymbol{e_2}(\sum_{m=0}^{n+2} N_{k,m} - 1) \\
&+ \boldsymbol{e_3}(\sum_{m=0}^{n+3} N_{k,m} - 1 - k) \\
=& \left(\frac{N_{k,1} + N_{k,n} + N_{k,n+1} + N_{k,n+2}}{k} - N_{k,n+1}\right) \\
&+ \left(\frac{N_{k,1} + N_{k,n+1} + N_{k,n+2} + N_{k,n+3}}{k} - N_{k,n+2}\right)\boldsymbol{e_1} \\
&+ \left(\frac{N_{k,1} + N_{k,n+2} + N_{k,n+3} + N_{k,n+4}}{k} - N_{k,n+3}\right)\boldsymbol{e_2} \\
&+ \left(\frac{N_{k,1} + N_{k,n+3} + N_{k,n+4} + N_{k,n+5}}{k} - N_{k,n+4}\right)\boldsymbol{e_3} \\
&- \boldsymbol{e_2} - (1+k)\boldsymbol{e_3} \\
=& \frac{N_{k,1} + N_{k,1}\boldsymbol{e_1} + N_{k,1}\boldsymbol{e_2} + N_{k,1}\boldsymbol{e_3} + QN_{k,n} + QN_{k,n+1}}{k} \\
&+ \frac{QN_{k,n+2} - kQN_{k,n+1}}{k} - \boldsymbol{e_2} - \boldsymbol{e_3}(1+k) \\
=& \frac{1 + \boldsymbol{e_1} + \boldsymbol{e_2} + \boldsymbol{e_3} + QN_{k,n} + QN_{k,n+1}}{k} \\
&+ \frac{QN_{k,n+2} - kQN_{k,n+1}}{k} - \boldsymbol{e_2} - \boldsymbol{e_3}(1+k).
\end{aligned}$$

□

Theorem 3.7. *The recursive relationship of the k-Narayana quaternion sequence is defined as follows:*

$$QN_{k,n} = kQN_{k,n-1} + QN_{k,n-3}, \quad n \geq 3. \tag{3.4}$$

Proof.

$$\begin{aligned}
kQN_{k,n-1} + QN_{k,n-3} =& k(N_{k,n-1} + N_{k,n}\boldsymbol{e_1} + N_{k,n+1}\boldsymbol{e_2} + N_{k,n+2}\boldsymbol{e_3}) \\
&+ (N_{k,n-3} + N_{k,n-2}\boldsymbol{e_1} + N_{k,n-1}\boldsymbol{e_2} + N_{k,n}\boldsymbol{e_3}) \\
=& (kN_{k,n-1} + N_{k,n-3}) + (kN_{k,n} + N_{k,n-2})\boldsymbol{e_1} \\
&+ (kN_{k,n+1} + N_{k,n-1})\boldsymbol{e_2} + (kN_{k,n+2} + N_{k,n})\boldsymbol{e_3}
\end{aligned}$$

$$
\begin{aligned}
&= N_{k,n} + N_{k,n+1}\boldsymbol{e_1} + N_{k,n+2}\boldsymbol{e_2} + N_{k,n+3}\boldsymbol{e_3} \\
&= QN_{k,n}.
\end{aligned}
$$

□

Theorem 3.8. *The generating function for $k-$Narayana quaternion is:*

$$\sum_{m=0}^{\infty} QN_{k,m}x^m = \frac{QN_{k,0} + (QN_{k,1} - kQN_{k,0})x + (QN_{k,2} - kQN_{k,1})x^2}{1 - kx - x^3}. \tag{3.5}$$

Proof. Let

$$f(x) = \sum_{m=0}^{\infty} QN_{k,n}x^m = QN_{k,0} + QN_{k,1}x + QN_{k,2}x^2 + QN_{k,3}x^3 + \dots \tag{3.6}$$

Multiply Equation (3.6) by kx and x^3, we have

$$kxf(x) = kQN_{k,0}x + kQN_{k,1}x^2 + kQN_{k,2}x^3 + kQN_{k,3}x^4 + \dots \tag{3.7}$$

$$x^3f(x) = QN_{k,0}x^3 + QN_{k,1}x^4 + QN_{k,2}x^5 + QN_{k,3}x^6 + \dots \tag{3.8}$$

Based on the Equation (3.6-3.8) and Theorem 3.7, we obtain

$$
\begin{aligned}
f(x) - kxf(x) - x^3f(x) =& QN_{k,0} + (QN_{k,1} - kQN_{k,0})x + (QN_{k,2} - kQN_{k,1})x^2 \\
&+ (QN_{k,3} - kQN_{k,2} - QN_{k,0})x^3 \\
&+ (QN_{k,4} - kQN_{k,3} - QN_{k,1})x^4 + \dots \\
(1 - kx - x^3)f(x) =& QN_{k,0} + (QN_{k,1} - kQN_{k,0})x + (QN_{k,2} - kQN_{k,1})x^2 \\
&+ 0 + 0 + \dots \\
=& QN_{k,0} + (QN_{k,1} - kQN_{k,0})x + (QN_{k,2} - kQN_{k,1})x^2,
\end{aligned}
$$

which is the desired result. Moreover, we can see that

$$
\begin{aligned}
(1 - kx - x^3)f(x) =& N_{k,0} + N_{k,1}\boldsymbol{e_1} + N_{k,2}\boldsymbol{e_2} + N_{k,3}\boldsymbol{e_3} \\
&+ (N_{k,1} + N_{k,2}\boldsymbol{e_1} + N_{k,3}\boldsymbol{e_2} + N_{k,4}\boldsymbol{e_3} - k(N_{k,0} + N_{k,1}\boldsymbol{e_1} + N_{k,2}\boldsymbol{e_2} + N_{k,3}\boldsymbol{e_3}))x \\
&+ (N_{k,2} + N_{k,3}\boldsymbol{e_1} + N_{k,4}\boldsymbol{e_2} + N_{k,5}\boldsymbol{e_3} - k(N_{k,1} + N_{k,2}\boldsymbol{e_1} + N_{k,3}\boldsymbol{e_2} + N_{k,4}\boldsymbol{e_3}))x^2 \\
=& (N_{k,0} + (N_{k,1} - kN_{k,0})x + (N_{k,2} - kN_{k,1})x^2) \\
&+ (N_{k,1} + (N_{k,2} - kN_{k,1})x + (N_{k,3} - kN_{k,2})x^2)\boldsymbol{e_1} \\
&+ (N_{k,2} + (N_{k,3} - kN_{k,2})x + (N_{k,4} - kN_{k,3})x^2)\boldsymbol{e_2} \\
&+ (N_{k,3} + (N_{k,4} - kN_{k,3})x + (N_{k,5} - kN_{k,4})x^2)\boldsymbol{e_3} \\
=& x + \boldsymbol{e_1} + (k + k^2)x^2\boldsymbol{e_2} + (k^2 + x + kx^2)\boldsymbol{e_3}.
\end{aligned}
$$

□

4 Conclusion

In this study, our aim was to define the $k-$Narayana quaternion and prove some of its properties. These properties encompass the relationship between the $k-$Narayana quaternion and its conjugate, its norm, the Binet formula, finite sum, and generating function.

5 Back Matter

Competing interests. The authors have no conflicts of interest related to the content of the work.

Authors'contributions. Both authors have contributed equally to this paper, and they have both read and approved the final manuscript.

Acknowledgment. The authors express their gratitude to the referees for their valuable comments and suggestions, which improved this paper. Additionally, the authors would like to acknowledge the financial support provided by The Royal Thai Government Scholarship under the Thailand-Lao PDR Bilateral Development Cooperation. This scholarship has afforded the authors the opportunity to express this point.

References

[1] C. Flaut, V. Shpakivskyi, *On generalized Fibonacci quaternions and Fibonacci-Narayana quaternions*, 2012, arXiv preprint arXiv:1209.0584.

[2] S. Halici, *On Fibonacci quaternions. Adv. Appl. Clifford Algebras*, **22**(2) 2012, pp. 321–327.

[3] W.R. Hamilton, *Lectures on quaternions*, Hodges and Smith, Dublin, 1853.

[4] A.F. Horadam, *Complex Fibonacci numbers and Fibonacci quaternions* , The American Mathematical Monthly, **70**(3) 1963, pp. 289–291.

[5] A. Ipek, *On* $(p, q)-$*Fibonacci quaternions and their Binet formulas, generating functions and certain binomial sums*, Advances in Applied Clifford Algebras, **27**, 2017, pp. 1343-1351.

[6] E. Özkan, B. Kuloğlu, J. Peters, $k-$*Narayana sequence self-Similarity*, 2021, hal. archives-ouvertes. fr, 3242990.

[7] J.L. Ramírez and V.F. Sirvent, *A note on the* $k-$*Narayana sequence*, 2015, January, In Annales mathematicae et informaticae pp. 91–105.

[8] Y. Soykan, *On generalized Narayana numbers*, Int. J. Adv. Appl. Math. and Mech, **7**(3) 2020, pp. 43–56.

[9] J.P. Ward, *Quaternions and Cayley Numbers*, Algebra and Applications, Kluwer, London, 1997.

[10] T. Yağmur, *A note on hyperbolic* $(p, q)-$*Fibonacci quaternions*, Commun. Fac. Sci.Uuk.Ser. A1 Math. Stat. **69**(1) 2020, pp. 880–890.

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Some Quadratic and Quartic Diophantine Equations with Solutions Involving Fibonacci and Lucas Numbers

Shayathorn Wanasawat[1,‡], Panida Krongkaew[1,†], Orrawan Prathumwan[1,†] and Onanong Wimolrat[1,†]

[1]Department of Mathematics and Statistics, Faculty of Science and Technology
Thammasat University, Pathum Thani 12120, Thailand

Abstract

This paper presents straightforward methods offering complete solutions to quartic Diophantine equations of various forms expressed as $x^4 - 4x^2y^2 - y^4 = \pm 1, x^4 - 4x^2y^2 - y^4 = \pm 5, x^5 - 5y^4 = \pm 1$, and $x^4 - 5y^4 = \pm 5$. Additionally, we explore analogous quadratic Diophantine equations to such equations. We discover that all solutions to these equations are involving with Fibonacci and Lucas numbers.

Keywords: Fibonacci sequence, Lucas sequence, Diophantine equation, perfect power.

2020 MSC: Primary 11B39; Secondary 11D25.

1 Introduction

Among the vast of sequences, two stand out with distinct significance and hold their special places: the Fibonacci sequence $0, 1, 1, 2, 3, 5, 8, 13, 21, \ldots$, formed by adding the two preceding terms to generate the next and its closed related sequence sharing some counterparts, the Lucas sequence $2, 1, 3, 4, 7, 11, 18, 29, \ldots$, which follows the same recursive pattern. For integers $n \geq 2$, the Fibonacci sequence F_n is precisely defined by the recurrence relation $F_n = F_{n-1} + F_{n-2}$ with the initial conditions $F_0 = 0$ and $F_1 = 1$. Similarly, the Lucas sequence L_n also obeys the recurrence relation $L_n = L_{n-1} + L_{n-2}$ with the initial conditions $L_0 = 2$ and $L_1 = 1$. The n-th terms of these sequences can be expressed using the Binet formulas: $F_n = \frac{1}{\sqrt{5}}(\alpha^n - \beta^n)$ and $L_n = \alpha^n + \beta^n$ where $\alpha = \frac{1+\sqrt{5}}{2}$ and $\beta = \frac{1-\sqrt{5}}{2}$. It is clear that $\alpha + \beta = 1$ and $\alpha\beta = -1$. Moreover, α and β are roots of $t^2 - t - 1 = 0$, we have $\alpha^2 = \alpha + 1$ and $\beta^2 = \beta + 1$. Consequently, $\alpha^3 = \alpha \cdot \alpha^2 = \alpha(\alpha + 1) = \alpha^2 + \alpha = (\alpha + 1) + \alpha = 2\alpha + 1$. Similarly, $\beta^3 = 2\beta + 1$. In general, α and β satisfy the following identities:

$$\alpha^n = \alpha F_n + F_{n-1}, \tag{1.1}$$

$$\beta^n = \beta F_n + F_{n-1}. \tag{1.2}$$

†Speaker. ‡corresponding author.
Email: shayathorn@mathstat.sci.tu.ac.th (S. Wanasawat), panida.kro@dome.tu.ac.th (P. Krongkeaw), orrawan.pra@dome.tu.ac.th (O. Prathumwan), onanong.wim@dome.tu.ac.th (O. Winmorat).

Binet's formula has been the subject of study for decades, yielding numerous elegant and glamorous identities, including the Cassini's identities (see in [4] and [5]):

$$F_n^2 - F_{n+1}F_{n-1} = (-1)^{n+1}, \tag{1.3}$$
$$L_n^2 - L_{n+1}L_{n-1} = 5(-1)^n. \tag{1.4}$$

Manipulating these identities further by replacing the $n+1$-th term with the recursive formulas, we obtain their variations as the differences of squares of Fibonacci and Lucas numbers as follows:

$$F_n^2 - F_nF_{n-1} - F_{n-1}^2 = (-1)^{n+1}, \tag{1.5}$$
$$L_n^2 - L_nL_{n-1} - L_{n-1}^2 = 5(-1)^n. \tag{1.6}$$

Equations (1.5) and (1.6) served inspirations for the work of Keshin and Demirtürk [3] in 2013. They investigated the quadratic Diophantine equation of the many forms such as $x^2-xy-y^2 = \pm 1$, $x^2 - xy - y^2 = \pm 5$, $x^2 - 3xy - y^2 = \pm 1$, $x^2 - 3xy - y^2 = \pm 5$ and more generalized forms like $x^2 - kxy - y^2 \pm x = 0$, $x^2 - kxy - y^2 \pm y = 0$, $x^2 - kxy - y^2 \pm 5x = 0$ and $x^2 - kxy - y^2 \pm 5y = 0$ where $k \geq 3$, etc. Applying Cassini's identities and certain properties of units in the ring $\mathbb{Z}[\alpha] = \{a\alpha + b : a, b \in \mathbb{Z}\}$, they were successfully able to solve all of these quadratic Diophantine equations. It is obvious to see that Fibonacci and Lucas numbers satisfy the first few equations, and remarkably, the solutions to all of these equations are expressed in term of Fibonacci and Lucas numbers. The complete positive integer solutions of some quadratic Diophantine equations of these forms will be presented in the next section and later be used for the rest of this paper. In our work, we extend their ideas to solve quartic Diophantine equations, presenting the positive integer solutions to the following equations:

$$x^4 - 4x^2y^2 - y^4 = -1 \tag{1.7}$$
$$x^4 - 4x^2y^2 - y^4 = 1 \tag{1.8}$$
$$x^4 - 4x^2y^2 - y^4 = 5 \tag{1.9}$$
$$x^4 - 4x^2y^2 - y^4 = -5 \tag{1.10}$$
$$x^4 - 5y^4 = -1 \tag{1.11}$$
$$x^4 - 5y^4 = 1 \tag{1.12}$$
$$x^4 - 5y^4 = -5 \tag{1.13}$$
$$x^4 - 5y^4 = 5 \tag{1.14}$$

It can be shown that most of the above equations are solvable and we obtain their positive integer solutions. Also, we study similar forms of quadratic Diophantine equations.

2 Preliminaries

In this section, a collection of positive integer solutions to some selected quadratic Diophantine equations are presented. Moreover, we discuss the representation of Fibonacci and Lucas numbers as specific forms, followed by the divisibility of Fibonacci and Lucas numbers.

Initially, we highlight some theorems presented in [3], which stand as an important concept and motivation behind this paper. The next following theorem, stated without any proof in [3], prompts us to prove it here for the sake of completeness.

Theorem 2.1. *All positive integer solutions of the equation $x^2 - xy - y^2 = -1$ are given by $(x, y) = (F_{2n}, F_{2n-1})$ for $n \geq 1$.*

Proof. We first notice that $F_{2n}^2 - F_{2n}F_{2n-1} - F_{2n-1}^2 = -1$ holds from the equation (1.5). Next, we solve for x and y in $x^2 - xy - y^2 = -1$. This is equivalent to solving $-x^2 + xy + y^2 = (\alpha\beta)x^2 + (\alpha+\beta)xy + y^2 = 1$. By factorizing this quadratic equation, we find $(\alpha x + y)(\beta x + y) = (\alpha x + y)((1-\alpha)x + y) = (\alpha x + y)(-\alpha x + (x + y)) = 1$. It follows that $\alpha x + y$ must be a unit in $\mathbb{Z}[\alpha]$. Considering that all units in $\mathbb{Z}[\alpha]$ are $\pm\alpha^n$ for any $n \in \mathbb{Z}$ and $\alpha x + y > 0$, we conclude that $\alpha^n = \alpha x + y$. By equation (1.1), we obtain $x = F_n$ and $y = F_{n-1}$. Back substitution leads us to $F_n^2 - F_nF_{n-1} - F_{n-1}^2 = (-1)^{n+1} = -1$. This forces n to be even. We are done. □

Following this, we now introduce other several key theorems from [3] without proofs:

Theorem 2.2. *All positive integer solutions of the equation $x^2 - xy - y^2 = 1$ are given by $(x, y) = (F_{2n+1}, F_{2n})$ for $n \geq 1$.*

Theorem 2.3. *All positive integer solutions of the equation $x^2 - xy - y^2 = -5$ are given by $(x, y) = (L_{2n+1}, L_{2n})$ for $n \geq 0$.*

Theorem 2.4. *All positive integer solutions of the equation $x^2 - xy - y^2 = 5$ are given by $(x, y) = (L_{2n}, L_{2n-1})$ for $n \geq 1$.*

Theorem 2.5. *All positive integer solutions of the equation $x^2 - 3xy + y^2 = -1$ are given by $(x, y) = (F_{2n+1}, F_{2n-1})$ for $n \geq 0$.*

Theorem 2.6. *All positive integer solutions of the equation $x^2 - 3xy + y^2 = 1$ are given by $(x, y) = (F_{2n+2}, F_{2n})$ for $n \geq 1$.*

Theorem 2.7. *All positive integer solutions of the equation $x^2 - 3xy + y^2 = -5$ are given by $(x, y) = (L_{2n+2}, L_{2n})$ for $n \geq 0$.*

Theorem 2.8. *All positive integer solutions of the equation $x^2 - 3xy + y^2 = 5$ are given by $(x, y) = (L_{2n+1}, L_{2n-1})$ for $n \geq 1$.*

Next, all Fibonacci and Lucas numbers that can be written in form of the product of specific perfect power numbers are shown in the following theorem (see Theorems 2 and 3 in [1]).

Theorem 2.9. *Let $a, b, c \in \mathbb{Z}$ with $a \geq 0, b \geq 1$ and $c \geq 2$.*

1. *If $F_n = 2^a b^c$, then $n \in \{1, 2, 3, 6, 12\}$.*

2. *If $L_n = 2^a b^c$, then $n \in \{0, 1, 3, 6\}$.*

We can see that there are a finite numbers of Fibonacci and Lucas numbers which possibly fit into those forms. For further details, interested reader can refer to [1]. Other properties of Fibonacci and Lucas numbers that play an important role in this paper are their divisibility. Some of these properties are listed below.

Theorem 2.10. *Let $m, n \in \mathbb{Z}^+$.*

1. *$F_n \mid F_m$ if and only if $m = kn$ for some $k \in \mathbb{Z}^+$.*

2. *$L_n \mid F_m$ if and only if $m = 2kn$ for some $k \in \mathbb{Z}^+$.*

3. *$L_n \mid L_m$ if and only if $m = (2k-1)n$ for some $k \in \mathbb{Z}^+$.*

3 Main Results

Assume $x > y$ going forward. We start this section by defining certain polynomials in two variables x and y, through appropriate substitution.

Let $P(x, y) = x^2 - xy - y^2$. We compute

$$\begin{aligned} P\left(x^2+y^2, x^2-y^2\right) &= (x^2+y^2)^2 - (x^2+y^2)(x^2-y^2) - (x^2-y^2)^2 \\ &= x^4 + 2x^2y^2 + y^4 - x^4 + y^4 - x^4 + 2x^2y^2 - y^4 \\ &= -x^4 + 4x^2y^2 + y^4, \\ P(x+y, x-y) &= (x+y)^2 - (x+y)(x-y) - (x-y)^2 \\ &= x^2 + 2xy + y^2 - x^2 + y^2 - x^2 + 2xy - y^2 \\ &= -x^2 + 4xy + y^2. \end{aligned}$$

Therefore,

$$-P\left(x^2+y^2, x^2-y^2\right) = x^4 - 4x^2y^2 - y^4, \tag{3.1}$$

$$-P(x+y, x-y) = x^2 - 4xy - y^2. \tag{3.2}$$

Let $Q(x, y) = x^2 - 3xy + y^2$. We have

$$\begin{aligned} Q\left(x^2+y^2, x^2-y^2\right) &= (x^2+y^2)^2 - 3(x^2+y^2)(x^2-y^2) + (x^2-y^2)^2 \\ &= x^4 + 2x^2y^2 + y^4 - 3x^4 + 3y^4 + x^4 - 2x^2y^2 + y^4 \\ &= -x^4 + 5y^4, \\ Q(x+y, x-y) &= (x+y)^2 - 3(x+y)(x-y) + (x-y)^2 \\ &= x^2 + 2xy + y^2 - 3x^2 + 3y^2 + x^2 - 2xy + y^2 \\ &= -x^2 + 5y^2. \end{aligned}$$

Hence,

$$-Q\left(x^2+y^2, x^2-y^2\right) = x^4 - 5y^4, \tag{3.3}$$

$$-Q(x+y, x-y) = x^2 - 5y^2. \tag{3.4}$$

We are now ready to solve the equation (1.7) to (1.14).

3.1 The Equations $x^4 - 4x^2y^2 - y^4 = \pm 1$ and $x^4 - 4x^2y^2 - y^4 = \pm 5$

We first solve the quartic equations $x^4 - 4x^2y^2 - y^4 = \pm 1$. Their positive integer solutions are associated with Fibonacci numbers.

Theorem 3.1. *The only positive integer solution of the equation $x^4 - 4x^2y^2 - y^4 = -1$ is $(x, y) = (2, 1)$.*

Proof. Clearly, $(x, y) = (2, 1)$ satisfies the equation (1.7). Now, we directly solve the equation (1.7), which can be written as $-P\left(x^2+y^2, x^2-y^2\right) = -1$. Then, we apply Theorem 2.2 to the equation

$$(x^2+y^2)^2 - (x^2+y^2)(x^2-y^2) - (x^2-y^2)^2 = 1.$$

This implies that

$$\begin{aligned} x^2 + y^2 &= F_{2n+1}, \\ x^2 - y^2 &= F_{2n}. \end{aligned}$$

By Theorem 2.9, we have $n = 2$, hence $2x^2 = F_{2(2)+2} = F_6 = 8$ and $2y^2 = F_{2(2)-1} = F_3 = 2$. Therefore, $x = 2$ and $y = 1$. □

Theorem 3.2. *The equation $x^4 - 4x^2y^2 - y^4 = 1$ has no positive integer solutions.*

Proof. Equation (1.8) is equivalently to $-P\left(x^2+y^2, x^2-y^2\right) = 1$. Analogously to the previous Theorem 3.1, we establish $x^2 + y^2 = F_{2n}$ and $x^2 - y^2 = F_{2n-1}$ by Theorem 2.1. It follows that $2x^2 = F_{2n+1}$ and $2y^2 = F_{2n-2}$. Then, by Theorem 2.9, the subscripts are either 3 or 6. There is no value of n satisfying these Fibonacci numbers simultaneously. Consequently, no positive integer solution exists for equation (1.8). □

We consider the quadratic equations of the similar form. Let us address these equations

$$x^2 - 4xy - y^2 = -1, \tag{3.5}$$
$$x^2 - 4xy - y^2 = 1. \tag{3.6}$$

Theorem 3.3. *All positive integer solutions of the equation $x^2 - 4xy - y^2 = -1$ are given by $(x, y) = \left(\frac{F_{6k+6}}{2}, \frac{F_{6k+3}}{2}\right)$ for $k \geq 0$.*

Proof. For sufficiency, we aim to solve the equation $-P\left(x+y, x-y\right) = -1$,or simply,

$$(x+y)^2 - (x+y)(x-y) - (x-y)^2 = 1.$$

From this, we derive the following equations

$$x + y = F_{2n+1},$$
$$x - y = F_{2n}.$$

By adding and subtracting these equations, we obtain

$$2x = F_{2n+1} + F_{2n} = F_{2n+2},$$
$$2y = F_{2n+1} - F_{2n} = F_{2n-1}.$$

Hence, $F_3 = 2$ must divides both F_{2n+2} and F_{2n-1}. According to Theorem 2.10, it follows that $3 \mid 2n+2$ and $3 \mid 2n-1$, leading to $n = 3k+2$ for $k \geq 0$. Moreover, the Fibonacci numbers at positions that are multiples of 3 are all even. Therefore, $(x, y) = \left(\frac{F_{6k+6}}{2}, \frac{F_{6k+3}}{2}\right)$ for $k \geq 0$.

To establish the necessary condition, we substitute $x = \frac{F_{6k+6}}{2}$ and $y = \frac{F_{6k+3}}{2}$ for $k \geq 0$ into the equation (3.5). By employing Binet's formula, we have

$$\begin{aligned}
x^2 - 4xy - y^2 &= \frac{1}{4}\left(\frac{\alpha^{6k+6} - \beta^{6k+6}}{\sqrt{5}}\right)^2 - \left(\frac{\alpha^{6k+6} - \beta^{6k+6}}{\sqrt{5}}\right)\left(\frac{\alpha^{6k+3} - \beta^{6k+3}}{\sqrt{5}}\right) \\
&\quad - \frac{1}{4}\left(\frac{\alpha^{6k+3} - \beta^{6k+3}}{\sqrt{5}}\right)^2 \\
&= \frac{1}{20}\left(\alpha^{12k+12} - 2\alpha^{6k+6}\beta^{6k+6} + \beta^{12k+12}\right) \\
&\quad - \frac{1}{5}\left(\alpha^{12k+9} - \alpha^{6k+6}\beta^{6k+3} - \alpha^{6k+3}\beta^{6k+6} + \beta^{12k+9}\right) \\
&\quad - \frac{1}{20}\left(\alpha^{12k+6} - 2\alpha^{6k+3}\beta^{6k+3} + \beta^{12k+6}\right) \\
&= \frac{1}{20}\left(\alpha^{12k+12} - 2(\alpha\beta)^{6k+6} + \beta^{12k+12}\right) \\
&\quad - \frac{1}{5}\left(\alpha^{12k+9} - \alpha^{6k+6}\beta^{6k+3} - \alpha^{6k+3}\beta^{6k+6} + \beta^{12k+9}\right) \\
&\quad - \frac{1}{20}\left(\alpha^{12k+6} - 2(\alpha\beta)^{6k+3} + \beta^{12k+6}\right).
\end{aligned}$$

Notice that $\alpha^{6k+6}\beta^{6k+3}+\alpha^{6k+3}\beta^{6k+6}=\left(\alpha^3+\beta^3\right)(\alpha\beta)^{6k+3}=-\left(\alpha^3+\beta^3\right)=-(2\alpha+1+2\beta+1)=-2(\alpha+\beta)-2=-4$. Then the equation becomes

$$\begin{aligned}
x^2-4xy-y^2 &= \frac{1}{20}\left(\alpha^{12k+12}-2+\beta^{12k+12}\right)\\
&\quad-\frac{1}{5}\left(\alpha^{12k+9}+4+\beta^{12k+9}\right)-\frac{1}{20}\left(\alpha^{12k+6}+2+\beta^{12k+6}\right)\\
&= \frac{1}{20}\left(\alpha^{12k+12}-2+\left(-\frac{1}{\alpha}\right)^{12k+12}\right)-\frac{1}{5}\left(\alpha^{12k+9}+4+\left(-\frac{1}{\alpha}\right)^{12k+9}\right)\\
&\quad-\frac{1}{20}\left(\alpha^{12k+6}+2+\left(-\frac{1}{\alpha}\right)^{12k+6}\right)\\
&= \left(-\frac{2}{20}-\frac{4}{5}-\frac{2}{20}\right)+\left(\frac{\alpha^{12k+12}+\alpha^{-12k-12}}{20}\right)+\left(-\frac{\alpha^{12k+9}-\alpha^{-12k-9}}{5}\right)\\
&\quad+\left(-\frac{\alpha^{12k+6}+\alpha^{-12k-6}}{20}\right)\\
&= -1+\left(\frac{\alpha^{12k+12}}{20}-\frac{\alpha^{12k+9}}{5}-\frac{\alpha^{12k+6}}{20}\right)+\left(\frac{\alpha^{-12k-12}}{20}+\frac{\alpha^{-12k-9}}{5}-\frac{\alpha^{-12k-6}}{20}\right)\\
&= -1+\left(\frac{\alpha^{12}}{20}-\frac{\alpha^9}{5}-\frac{\alpha^6}{20}\right)\alpha^{12k}+\left(\frac{\alpha^{-12}}{20}+\frac{\alpha^{-9}}{5}-\frac{\alpha^{-6}}{20}\right)\alpha^{-12k}\\
&= -1+\left(\frac{\alpha^{12}}{20}-\frac{4\alpha^9}{20}-\frac{\alpha^6}{20}\right)\alpha^{12k}+\left(\frac{\alpha^{-12}}{20}+\frac{4\alpha^{-9}}{20}-\frac{\alpha^{-6}}{20}\right)\alpha^{-12k}\\
&= -1+\frac{1}{20}\left(\alpha^6-4\alpha^3-1\right)\alpha^{12k+6}+\frac{1}{20}\left(-1-4\alpha^3+\alpha^6\right)\alpha^{-12k-12}.
\end{aligned}$$

By equation (1.1), we have $\alpha^6-4\alpha^3-1=0$, thus the above equation reduces to -1 as desired. This confirms that $(x,y)=\left(\frac{F_{6k+6}}{2},\frac{F_{6k+3}}{2}\right)$ for $k\geq 0$ is a solution to the equation (3.5). Thus, the theorem is proven. □

Theorem 3.4. *All positive integer solutions of the equation $x^2-4xy-y^2=1$ are given by $(x,y)=\left(\frac{F_{6k+3}}{2},\frac{F_{6k}}{2}\right)$ for $k\geq 1$.*

Proof. Following similar steps as in Theorem 3.3, it suffices to solve the equation $-P(x+y,x-y)=1$. This means that

$$(x+y)^2-(x+y)(x-y)-(x-y)^2=-1.$$

Again by Theorem 2.1, we obtain the system

$$\begin{aligned}
x+y &= F_{2n},\\
x-y &= F_{2n-1}.
\end{aligned}$$

This yields

$$\begin{aligned}
2x &= F_{2n}+F_{2n-1}=F_{2n+1},\\
2y &= F_{2n}-F_{2n-1}=F_{2n-2}.
\end{aligned}$$

Since F_3 equals to 2, we deduce that 3 divides $2n+1$ and $2n-2$. This leads us to $n=3k+1$ for $k\geq 1$. Also, all Fibonacci numbers of indices divisible by 3 are even. Therefore, $(x,y)=\left(\frac{F_{6k+3}}{2},\frac{F_{6k}}{2}\right)$ for $k\geq 1$.

For necessity, we simplify and verify that the equation (3.6) hold true for $k \geq 1$ by substituting $x = \frac{F_{6k+3}}{2}$ and $y = \frac{F_{6k}}{2}$ using Binet's formula as follows:

$$\begin{aligned}
x^2 - 4xy - y^2 &= \frac{1}{4}\left(\frac{\alpha^{6k+3} - \beta^{6k+3}}{\sqrt{5}}\right)^2 - \left(\frac{\alpha^{6k+3} - \beta^{6k+3}}{\sqrt{5}}\right)\left(\frac{\alpha^{6k} - \beta^{6k}}{\sqrt{5}}\right) - \frac{1}{4}\left(\frac{\alpha^{6k} - \beta^{6k}}{\sqrt{5}}\right)^2 \\
&= \frac{1}{20}\left(\alpha^{12k+6} - 2\alpha^{6k+3}\beta^{6k+3} + \beta^{12k+6}\right) \\
&\quad - \frac{1}{5}\left(\alpha^{12k+3} - \alpha^{6k+3}\beta^{6k} - \alpha^{6k}\beta^{6k+3} + \beta^{12k+3}\right) \\
&\quad - \frac{1}{20}\left(\alpha^{12k} - 2\alpha^{6k}\beta^{6k} + \beta^{12k}\right) \\
&= \frac{1}{20}\left(\alpha^{12k+6} - 2(\alpha\beta)^{6k+3} + \beta^{12k+6}\right) \\
&\quad - \frac{1}{5}\left(\alpha^{12k+3} - \alpha^{6k+3}\beta^{6k} - \alpha^{6k}\beta^{6k+3} + \beta^{12k+3}\right) \\
&\quad - \frac{1}{20}\left(\alpha^{12k} - 2(\alpha\beta)^{6k} + \beta^{12k}\right).
\end{aligned}$$

Similarly to the previous proof, we evaluate $\alpha^{6k+3}\beta^{6k} + \alpha^{6k}\beta^{6k+3} = 4$. It follows that

$$\begin{aligned}
x^2 - 4xy - y^2 &= \frac{1}{20}\left(\alpha^{12k+6} + 2 + \beta^{12k+6}\right) - \frac{1}{5}\left(\alpha^{12k+3} - 4 + \beta^{12k+3}\right) - \frac{1}{20}\left(\alpha^{12k} - 2 + \beta^{12k}\right) \\
&= \frac{1}{20}\left(\alpha^{12k+6} + 2 + \left(-\frac{1}{\alpha}\right)^{12k+6}\right) - \frac{1}{5}\left(\alpha^{12k+3} - 4 + \left(-\frac{1}{\alpha}\right)^{12k+3}\right) \\
&\quad - \frac{1}{20}\left(\alpha^{12k} - 2 + \left(-\frac{1}{\alpha}\right)^{12k}\right) \\
&= \left(\frac{2}{20} + \frac{4}{5} + \frac{2}{20}\right) + \frac{1}{20}\left(\alpha^{12k+6} + \alpha^{-12k-6}\right) - \frac{1}{5}\left(\alpha^{12k+3} - \alpha^{-12k-3}\right) \\
&\quad - \frac{1}{20}\left(\alpha^{12k} + \alpha^{-12k}\right) \\
&= 1 + \frac{1}{20}\left(\alpha^{12k+6} - 4\alpha^{12k+3} - \alpha^{12k}\right) + \frac{1}{20}\left(\alpha^{-12k-6} + 4\alpha^{-12k-3} - \alpha^{-12k}\right) \\
&= 1 + \frac{1}{20}\left(\alpha^6 - 4\alpha^3 - 1\right)\alpha^{12k} + \frac{1}{20}\left(-1 - 4\alpha^3 + \alpha^6\right)\left(-\alpha^{-12k-6}\right).
\end{aligned}$$

Since $\alpha^6 - 4\alpha^3 - 1 = 0$, we have $x^2 - 4xy - y^2 = 1$, confirming the validity of the solution $(x, y) = \left(\frac{F_{6k+3}}{2}, \frac{F_{6k}}{2}\right)$ for $k \geq 1$ to the equations (3.6). □

For quartic equations (1.9) and (1.10), we are concerning with Lucas numbers instead of Fibonacci numbers. Applying Theorem 2.3 and 2.4, we have the following theorems.

Theorem 3.5. *The equation $x^4 - 4x^2y^2 - y^4 = -5$ has no positive integer solutions.*

Theorem 3.6. *The equation $x^4 - 4x^2y^2 - y^4 = 5$ has no positive integer solutions.*

The quadratic versions of the equations (1.9) and (1.10) are given by

$$x^2 - 4xy - y^2 = -5, \tag{3.7}$$
$$x^2 - 4xy - y^2 = 5. \tag{3.8}$$

The positive integer solutions to the equations (3.7) and (3.8) are provided by the following theorems.

Theorem 3.7. *All positive integer solutions of* $x^2 - 4xy - y^2 = -5$ *are given by* $(x, y) = \left(\frac{L_{6k+3}}{2}, \frac{L_{6k}}{2}\right)$ *for* $k \geq 0$.

Theorem 3.8. *All positive integer solutions of* $x^2 - 4xy - y^2 = 5$ *are given by* $(x, y) = \left(\frac{L_{6k+6}}{2}, \frac{L_{6k+3}}{2}\right)$ *for* $k \geq 0$.

3.2 The Equations $x^4 - 5y^4 = \pm 1$ and $x^4 - 5y^4 = \pm 5$

We solve the quartic equations (1.11) to (1.14) and the quadratic equations:

$$x^2 - 5y^2 = -1, \tag{3.9}$$
$$x^2 - 5y^2 = 1, \tag{3.10}$$
$$x^2 - 5y^2 = -5, \tag{3.11}$$
$$x^2 - 5y^2 = 5. \tag{3.12}$$

Unlike in Section 3.1, we start by solving the equations (1.13) and (1.14). To solve such equations, we apply Theorems 2.7 and 2.8 which are related to Lucas numbers.

Theorem 3.9. *The equation* $x^4 - 5y^4 = -5$ *has no positive integer solutions.*

Proof. Solving the equation (1.13) is equivalent to solve $-Q\left(x^2 + y^2, x^2 - y^2\right) = -5$. That is $\left(x^2 + y^2\right)^2 - 3\left(x^2 + y^2\right)\left(x^2 - y^2\right) + \left(x^2 - y^2\right)^2 = 5$. By Theorem 2.8, we have $x^2 + y^2 = L_{2n+1}$ and $x^2 - y^2 = L_{2n-1}$ for $n \geq 1$. Subtracting these equations yields $2y^2 = L_{2n+1} - L_{2n-1} = L_{2n}$. Applying Theorem 2.9, we find that the only possible value of n must be 3. However, this leads to a contradiction, as $2x^2 = L_7 + L_5 = 40$, and thus $x = 2\sqrt{10}$, which is non-integer. Therefore, the equation (1.13) has no positive integer solutions. □

Theorem 3.10. *The equation* $x^4 - 5y^4 = 5$ *has no positive integer solutions.*

Proof. To find the solutions to equation (1.14), we must solve $-Q\left(x^2 + y^2, x^2 - y^2\right) = 5$, which yields $\left(x^2 + y^2\right)^2 - 3\left(x^2 + y^2\right)\left(x^2 - y^2\right) + \left(x^2 - y^2\right)^2 = -5$. We then express $x^2 + y^2 = L_{2n+2}$ and $x^2 - y^2 = L_{2n}$ for $n \geq 0$ from the Theorem 2.7. The difference between these equations results in $2y^2 = L_{2n+2} - L_{2n} = L_{2n+1}$. By Theorem 2.9, there are no integer values for n, since the feasible options are either $-\frac{1}{2}$ or $\frac{5}{2}$. Consequently, there exist no positive integer solutions for equation (1.14). □

Next, we solve the equations (3.11) and (3.12) which are equivalently to solve $-Q(x + y, x - y) = -5$ and $-Q(x + y, x - y) = 5$, respectively.

Theorem 3.11. *All positive integer solutions of the equation* $x^2 - 5y^2 = -5$ *are given by* $(x, y) = \left(\frac{L_{6k+1}+L_{6k-1}}{2}, \frac{L_{6k}}{2}\right)$ *for* $k \geq 1$.

Proof. Applying Theorem 2.8 to the following equation

$$(x + y)^2 - 3(x + y)(x - y) + (x - y)^2 = 5.$$

This implies that

$$x + y = L_{2n+1},$$
$$x - y = L_{2n-1}.$$

Solving for y, we find $y = \frac{L_{2n+1}-L_{2n-1}}{2} = \frac{L_{2n}}{2} = \frac{L_{2n}}{F_3}$. By Theorem 2.10, 3 divides $2n$, and hence $n = 3k$ for $k \geq 1$. Now solving for x, we obtain $x = \frac{L_{2n+1}+L_{2n-1}}{2} = \frac{L_{6k+1}+L_{6k-1}}{2}$. Since the $6k+1$-th term and the $6k-1$-th term of the Lucas sequence are always an odd integer. So x is indeed

an integer. Consequently, the solutions to the equation (3.11) are $(x, y) = \left(\frac{L_{6k+1}+L_{6k-1}}{2}, \frac{L_{6k}}{2}\right)$ for $k \geq 1$.

Assume $x = \frac{L_{6k+1}+L_{6k-1}}{2}$ and $y = \frac{L_{6k}}{2}$ for $k \geq 1$. This necessarily leaves us to verify that such x and y satisfy the equation (3.11). Replace the Lucas numbers in x and y with Binet's formula, we have

$$\begin{aligned}
x^2 - 5y^2 &= \frac{1}{4}\left[\left(\alpha^{6k+1} + \beta^{6k+1}\right)^2 + 2\left(\alpha^{6k+1} + \beta^{6k+1}\right)\left(\alpha^{6k-1} + \beta^{6k-1}\right) + \left(\alpha^{6k-1} + \beta^{6k-1}\right)^2\right] \\
&\quad - \frac{5}{4}\left(\alpha^{6k} + \beta^{6k}\right)^2 \\
&= \frac{1}{4}\left[\left(\alpha^{12k+2} + 2\alpha^{6k+1}\beta^{6k+1} + \beta^{12k+2}\right) + 2\left(\alpha^{12k} + \alpha^{6k+1}\beta^{6k-1} + \alpha^{6k-1}\beta^{6k+1} + \beta^{12k}\right)\right. \\
&\quad \left. + \left(\alpha^{12k-2} + 2\alpha^{6k-1}\beta^{6k-1} + \beta^{12k-2}\right)\right] - \frac{5}{4}\left(\alpha^{12k} + 2\alpha^{6k}\beta^{6k} + \beta^{12k}\right) \\
&= \frac{1}{4}\left[\left(\alpha^{12k+2} + 2(\alpha\beta)^{6k+1} + \beta^{12k+2}\right) + 2\left(\alpha^{12k} + \alpha^{6k+1}\beta^{6k-1} + \alpha^{6k-1}\beta^{6k+1} + \beta^{12k}\right)\right. \\
&\quad \left. + \left(\alpha^{12k-2} + 2(\alpha\beta)^{6k-1} + \beta^{12k-2}\right)\right] - \frac{5}{4}\left(\alpha^{12k} + 2(\alpha\beta)^{6k} + \beta^{12k}\right).
\end{aligned}$$

Recall that $\alpha + \beta = -1$ and $\alpha\beta = -1$. Further, we find that $\alpha^{6k+1}\beta^{6k-1} + \alpha^{6k-1}\beta^{6k+1} = \left(\alpha^2 + \beta^2\right)(\alpha\beta)^{6k-1} = -(\alpha + 1 + \beta + 1) = -(\alpha + \beta + 2) = -3$. It follows that

$$\begin{aligned}
x^2 - 5y^2 &= \frac{1}{4}\left[\left(\alpha^{12k+2} - 2 + \beta^{12k+2}\right) + 2\left(\alpha^{12k} - 3 + \beta^{12k}\right) + \left(\alpha^{12k-2} - 2 + \beta^{12k-2}\right)\right] \\
&\quad - \frac{5}{4}(\alpha^{12k} + 2 + \beta^{12k}) \\
&= \frac{1}{4}\left[\left(\alpha^{12k+2} - 2 + \left(-\frac{1}{\alpha}\right)^{12k+2}\right) + 2\left(\alpha^{12k} - 3 + \left(-\frac{1}{\alpha}\right)^{12k}\right)\right. \\
&\quad \left. + \left(\alpha^{12k-2} - 2 + \left(-\frac{1}{\alpha}\right)^{12k-2}\right)\right] - \frac{5}{4}\left(\alpha^{12k} + 2 + \left(-\frac{1}{\alpha}\right)^{12k}\right) \\
&= \left(-\frac{2}{4} - \frac{6}{4} - \frac{2}{4} - \frac{10}{4}\right) + \frac{1}{4}\left(\alpha^{12k+2} + \alpha^{-12k-2} + 2\alpha^{12k} + 2\alpha^{-12k} + \alpha^{12k-2} + \alpha^{-12k+2}\right) \\
&\quad - \frac{5}{4}\left(\alpha^{12k} + \alpha^{-12k}\right) \\
&= -5 + \frac{1}{4}\left(\alpha^{12k+2} - 3\alpha^{12k} + \alpha^{12k-2}\right) + \frac{1}{4}\left(\alpha^{-12k-2} - 3\alpha^{-12k} + \alpha^{-12k+2}\right) \\
&= -5 + \frac{1}{4}\left(\alpha^4 - 3\alpha^2 + 1\right)\alpha^{12k-2} + \frac{1}{4}\left(1 - 3\alpha^2 + \alpha^4\right)\alpha^{-12k-2}.
\end{aligned}$$

By equation (1.1), $\alpha^4 - 3\alpha^2 + 1$ can be simplified into 0. Therefore, $x^2 - 5y^2 = -5$. The proof is completed. □

Theorem 3.12. *All positive integer solutions of the equation $x^2 - 5y^2 = 5$ are given by $(x, y) = \left(\frac{L_{6k+4}+L_{6k+2}}{2}, \frac{L_{6k+3}}{2}\right)$ for $k \geq 0$.*

Proof. Let x, y be positive integers. We apply Theorem 2.9 to the equation

$$(x + y)^2 - 3(x + y)(x - y) + (x - y)^2 = -5.$$

To solve this equation, we add and subtract the equations in the system

$$\begin{aligned}
x + y &= L_{2n+2}, \\
x - y &= L_{2n}.
\end{aligned}$$

This implies that $x = \frac{L_{2n+2}+L_{2n}}{2}$ and $y = \frac{L_{2n+1}}{2}$. Note that $F_3 = 2$. By Theorem 2.10 and the necessity of y being an integer, we determine its value and find that the subscripts of the Lucas number must be divisible by the subscript of Fibonacci number. In other words, 3 divides 2n+1. Thus, $n = 3k+1$ for $k \geq 0$. Then $x = \frac{L_{6k+4}+L_{6k+2}}{2}$. Now, we only need to show that x is certainly an integer. To do this, we use the fact that $3 \nmid 6k+4$ and $3 \nmid 6k+2$, then by Theorem 2.10, $2 \nmid L_{6k+4}$ and $2 \nmid L_{6k+2}$, respectively. This implies that both L_{6k+4} and L_{6k+2} are odd numbers, so their sum is an even number, which is divisible by 2, ensuring that x is an integer.

To verify that $(x, y) = \left(\frac{L_{6k+4}+L_{6k+2}}{2}, \frac{L_{6k+3}}{2}\right)$ for $k \geq 0$ satisfies the equation (3.12), we use Binet's formula to substitute the Lucas numbers into the equation $x^2 - 5y^2 = -5$

$$\begin{aligned}
x^2 - 5y^2 &= \frac{1}{4}\left[\left(\alpha^{6k+4} + \beta^{6k+4}\right)^2 + 2\left(\alpha^{6k+4} + \beta^{6k+4}\right)\left(\alpha^{6k+2} + \beta^{6k+2}\right) + \left(\alpha^{6k+2} + \beta^{6k+2}\right)^2\right] \\
&\quad - \frac{5}{4}\left(\alpha^{6k+3} + \beta^{6k+3}\right)^2 \\
&= \frac{1}{4}\left[\left(\alpha^{12k+8} + 2\alpha^{6k+4}\beta^{6k+4} + \beta^{12k+8}\right) + 2\left(\alpha^{12k+6} + \alpha^{6k+4}\beta^{6k+2} + \alpha^{6k+2}\beta^{6k+4}\right.\right. \\
&\quad \left.\left. + \beta^{12k+6}\right) + \left(\alpha^{12k+4} + 2\alpha^{6k+2}\beta^{6k+2} + \beta^{12k+4}\right)\right] - \frac{5}{4}\left(\alpha^{12k+6} + 2\alpha^{6k+3}\beta^{6k+3}\right. \\
&\quad \left. + \beta^{12k+6}\right) \\
&= \frac{1}{4}\left[\left(\alpha^{12k+8} + 2\left(\alpha\beta\right)^{6k+4} + \beta^{12k+8}\right) + 2\left(\alpha^{12k+6} + \alpha^{6k+4}\beta^{6k+2} + \alpha^{6k+2}\beta^{6k+4}\right.\right. \\
&\quad \left.\left. + \beta^{12k+6}\right) + \left(\alpha^{12k+4} + 2\left(\alpha\beta\right)^{6k+2} + \beta^{12k+4}\right)\right] - \frac{5}{4}\left(\alpha^{12k+6} + 2\left(\alpha\beta\right)^{6k+3}\right. \\
&\quad \left. + \beta^{12k+6}\right).
\end{aligned}$$

We can show that $\alpha^{6k+4}\beta^{6k+2} + \alpha^{6k+2}\beta^{6k+4} = 3$ in a similar way to the proof in the previous theorem. Thus,

$$\begin{aligned}
x^2 - 5y^2 &= \frac{1}{4}\left[\left(\alpha^{12k+8} + 2 + \beta^{12k+8}\right) + 2\left(\alpha^{12k+6} + 3 + \beta^{12k+6}\right) + \left(\alpha^{12k+4} + 2 + \beta^{12k+4}\right)\right] \\
&\quad - \frac{5}{4}\left(\alpha^{12k+6} - 2 + \beta^{12k+6}\right) \\
&= \frac{1}{4}\left[\left(\alpha^{12k+8} + 2 + \left(-\frac{1}{\alpha}\right)^{12k+8}\right) + 2\left(\alpha^{12k+6} + 3 + \left(-\frac{1}{\alpha}\right)^{12k+6}\right)\right. \\
&\quad \left. + \left(\alpha^{12k+4} + 2 + \left(-\frac{1}{\alpha}\right)^{12k+4}\right)\right] - \frac{5}{4}\left(\alpha^{12k+6} - 2 + \left(-\frac{1}{\alpha}\right)^{12k+6}\right) \\
&= \left(\frac{2}{4} + \frac{6}{4} + \frac{2}{4} + \frac{10}{4}\right) + \frac{1}{4}\left(\alpha^{12k+8} + \alpha^{-12k-8} + 2\alpha^{12k+6} + 2\alpha^{-12k-6}\right. \\
&\quad \left. + \alpha^{12k+4} + \alpha^{-12k-4}\right) - \frac{5}{4}\left(\alpha^{12k+6} + \alpha^{-12k-6}\right) \\
&= 5 + \frac{1}{4}\left(\alpha^{12k+8} - 3\alpha^{12k+6} + \alpha^{12k+4}\right) + \frac{1}{4}\left(\alpha^{-12k-8} - 3\alpha^{-12k-6} + \alpha^{-12k-4}\right) \\
&= 5 + \frac{1}{4}\left(\alpha^4 - 3\alpha^2 + 1\right)\alpha^{12k+4} + \frac{1}{4}\left(1 - 3\alpha^2 + \alpha^4\right)\alpha^{-12k-8}.
\end{aligned}$$

Once again, we have $x^2 - 5y^2 = 5$ since $\alpha^4 - 3\alpha^2 + 1 = 0$ as before. We conclude that $(x, y) = \left(\frac{L_{6k+4}+L_{6k+2}}{2}, \frac{L_{6k+3}}{2}\right)$ for $k \geq 0$ is surely a solution to the equation (3.12). □

Lastly, the solutions to the equations (1.11), (1.12), (3.9), and (3.10) are presented in the following theorems, stated without proofs.

Theorem 3.13. *The equation $x^4 - 5y^4 = -1$ has no positive integer solutions.*

Theorem 3.14. *The only positive integer solution of the equation $x^4 - 5y^4 = 1$ is $(x, y) = (3, 2)$.*

Theorem 3.15. *All positive integer solutions of the equation $x^2 - 5y^2 = -1$ are given by $(x, y) = \left(\frac{L_{6k+3}}{2}, \frac{F_{6k+3}}{2}\right)$ for $k \geq 0$.*

Theorem 3.16. *All positive integer solutions of $x^2 - 5y^2 = 1$ are given by $(x, y) = \left(\frac{L_{6k}}{2}, \frac{F_{6k}}{2}\right)$ for $k \geq 1$.*

Acknowledgment. The authors are grateful to the referees for their careful reading of the manuscript and their useful comments.

References

[1] F. Luca and V. Patel, *On perfect powers that are sums of two Fibonacci numbers*, J. Number Theory, **189** (2018), 90–96.

[2] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 4th ed., Oxford University Press, Oxford, 1980.

[3] R. Keskin and B. Demirtürk, *Solutions of some Diophantine equations using generalized Fibonacci and Lucas sequences*, Ars Combinatoria (in press), Ars Combinatoria, **111** (2013), 161–179.

[4] T. Koshy, *Fibonacci and Lucas Numbers with Applications*, Wiley, New York, 2001.

[5] S. Vajda, *Fibonacci and Lucas Numbers and the Golden Section*, E. Horwood Limited, West Essex, 1989.

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Sums of Iterated Partial Sums of the k-Fibonacci Sequence

Supamit Pimsri[1,†], Somthawin Khunkhet[1,‡], and Boonyen Thongkam[1]

[1]Department of Mathematics, Faculty of Science
Ubon Ratchathani Rajabhat University, Ubon Ratchathani 34000, Thailand

Abstract

In this paper, we present sums and alternating sums of the iterated partial sums of the k-Fibonacci sequence. As special case, we give sums of the iterated partial sums of Fibonacci and Pell sequences.

Keywords: iterated partial sums, partial sums, k-Fibonacci sequence.

2020 MSC: 11B39; 11A25.

1 Introduction

Recently, some mathematicians are interested in the iterated partial sums of the sequence as follows. Chu [1], defined $P(F_n) := \left\{\sum_{i=1}^{n} F_i\right\}_{n\geq 1}$; this is, the function P give the sequence of partial sums of Fibonacci sequence $\{F_n\}_{n\geq 1}$. The author gave an identity involving $P^k(F_n)$, which is the resulting sequence from applying P to $\{F_n\}_{n\geq 1}$ k time, and provide a combinatorial interpretation of the number in $P^k(F_n)$. For example, for natural number k ,

$$\sum_{m=1}^{n} a_{k-1}(m) = a_{k-1}(n+2) - \binom{n+k}{k-1}$$

and

$$s_k(n) = a_k(n-2(k-1))$$

where $a_k(n)$ denote the nth number in the sequence $P^k(F_n)$ and
$s_k(n) = |\{S \subseteq \{1,2,\ldots,n\} : |S| \geq k \text{ and } \min S \geq |S|\}|$ is a shift of $a_k(n)$.

†Speaker. ‡Corresponding author.
Email: supamit.p@ubru.ac.th (S. Pimsri), somthawin.k@ubru.ac.th (S. Khunkhet),
boonyen.t@ubru.ac.th (B. Thongkam).

Falcon and Plaza [2], defined the iterated partial sums of the k-Fibonacci sequence, say $S_{k,n}^{r)} = \sum_{j=1}^{n} S_{k,j}^{r-1)}$with initial condition $S_{k,n}^{0)} = F_{k,n}$. They presented the iterated partial sums of the k-Fibonacci numbers are given as a function of k-Fibonacci numbers. For example, for natural number r , they showed that

$$S_{k,n}^{r)} = \sum_{j=0}^{n} \binom{r+j-1}{j} F_{k,n-j} \tag{1.1}$$

and

$$S_{k,n}^{r)} = S_{k,n-1}^{r)} + S_{k,n}^{r-1)}. \tag{1.2}$$

Falcon and Plaza [3, 4], introduced general k-Fibonacci number $\{F_{k,n}\}_{n\geq 0}$ were found by studying the recursive application of two geometrical transformations used in the well-known four-triangle longest-edge (4TLE) partition. From this definition, if $k = 1$ the classical Fibonacci sequence [5] is obtained $0, 1, 1, 2, 3, 5, 8, \ldots$ and if $k = 2$ that is the Pell sequence [6] $0, 1, 2, 5, 12, 29, 70, \ldots$. They presented some properties of these numbers are deduce directly from elementary matrix algebra. For example, They showed some properties for the sum of the k-Fibonacci sequence, obtained by summing up the first n matrices $\left(R^{k-1}L\right)^n$ as following, where $R^{k-1} = \begin{bmatrix} -k+2 & k-1 \\ -k+1 & k \end{bmatrix}$ and $L = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$

$$\begin{aligned} \sum_{i=1}^{n} F_{k,i} &= \frac{1}{k}\left(F_{k,n+1} + F_{k,n} - 1\right) \\ \sum_{i=1}^{n} F_{k,2i} &= \frac{1}{k}\left(F_{k,2n+1} - 1\right) \\ \sum_{i=1}^{n} F_{k,2i+1} &= \frac{1}{k} F_{k,2n+2}. \end{aligned}$$

In combinatorics, these numbers are related to Ramsey-type theorems for subset of $\mathbb{N}$. Our purpose in this paper we investigate some properties of the iterated partial sums of the k-Fibonacci sequence. We give some new identities using (1.1), (1.2) and alternating sums.

2 Preliminaries

In this section, we present the definition of the iterated partial sums of the k-Fibonacci sequence and their properties.

Definition 2.1. The k-Fibonacci numbers are defined as

$$F_{k,n} = kF_{k,n-1} + F_{k,n-2} \quad , n \geq 2$$

with $F_{k,0} = 0$ and $F_{k,1} = 1$.

Note that if $k = 1$ then $F_{1,n} = F_n$ is the classical Fibonacci numbers and if $k = 2$ then $F_{2,n} = P_n$ is the Pell numbers.

Definition 2.2. [2] For $n, r \geq 1$, the iterated partial sums of the k-Fibonacci numbers are defined as

$$S_{k,n}^{r)} = \sum_{j=1}^{n} S_{k,j}^{r-1)}$$

with initial condition $S_{k,n}^{0)} = F_{k,n}$.

Table 1: Iterated partial sums of k-Fibonacci sequences

$r\backslash n$	1	2	3	4
0	$F_{k,1}$	$F_{k,2}$	$F_{k,3}$	$F_{k,4}$
1	$F_{k,1}$	$F_{k,2}+F_{k,1}$	$F_{k,3}+F_{k,2}+F_{k,1}$	$F_{k,4}+F_{k,3}+F_{k,2}+F_{k,1}$
2	$F_{k,1}$	$F_{k,2}+2F_{k,1}$	$F_{k,3}+2F_{k,2}+3F_{k,1}$	$F_{k,4}+2F_{k,3}+3F_{k,2}+4F_{k,1}$
3	$F_{k,1}$	$F_{k,2}+3F_{k,1}$	$F_{k,3}+3F_{k,2}+6F_{k,1}$	$F_{k,4}+3F_{k,3}+6F_{k,2}+10F_{k,1}$
4	$F_{k,1}$	$F_{k,2}+4F_{k,1}$	$F_{k,3}+4F_{k,2}+10F_{k,1}$	$F_{k,4}+4F_{k,3}+10F_{k,2}+20F_{k,1}$

Table 2: Iterated partial sums of k-Fibonacci sequences in power of k.

$r\backslash n$	1	2	3	4	5
0	1	k	k^2+1	k^3+2k	k^4+3k^2+1
1	1	$k+1$	k^2+k+2	k^3+k^2+3k+2	$k^4+k^3+4k^2+3k+3$
2	1	$k+2$	k^2+2k+4	k^3+2k^2+5k+6	$k^4+2k^3+6k^2+8k+9$
3	1	$k+3$	k^2+3k+7	$k^3+3k^2+8k+13$	$k^4+3k^3+9k^2+16k+22$
4	1	$k+4$	$k^2+4k+11$	$k^3+4k^2+12k+24$	$k^4+4k^3+13k^2+28k+46$

Table 3: Iterated partial sums of the classical Fibonacci sequences ($k=1$)

$r\backslash n$	1	2	3	4	5	6	7	8	9	10	11	12
0	1	1	2	3	5	8	13	21	34	55	89	144
1	1	2	4	7	12	20	33	54	88	143	232	376
2	1	3	7	14	26	46	79	133	221	364	596	972
3	1	4	11	25	51	97	176	309	530	894	1490	2462
4	1	5	16	41	92	189	365	674	1204	2098	3588	6050

Table 4: Iterated partial sums of the Pell sequences ($k=2$)

$r\backslash n$	1	2	3	4	5	6	7	8	9	10	11	12
0	1	2	5	12	29	70	169	408	985	2378	5741	13860
1	1	3	8	20	49	119	288	696	1681	4059	9800	23660
2	1	4	12	32	81	200	488	1184	2865	6924	16724	40384
3	1	5	17	49	130	330	818	2002	4867	11791	28515	68899
4	1	6	23	72	202	532	1350	3352	8219	20010	48525	117424

The Table 1 shown the first elements of these sequences. By applying elements in Table 1, the following sequences are obtained in Tables 2-4.

Theorem 2.3. *[2] For $r \geq 1$,*

$$S_{k,n}^{r)} = \sum_{j=0}^{n} \binom{r+j-1}{j} F_{k,n-j}.$$

Theorem 2.4. *[2] For $r \geq 1$,*

$$S_{k,n}^{r)} = S_{k,n-1}^{r)} + S_{k,n}^{r-1)}.$$

3 Main Results

In this section, we study the sums and alternating sums of iterated partial sums of k-Fibonacci sequence.

Theorem 3.1. *For $m \geq 0$,*

$$\sum_{r=0}^{m} S_{k,n}^{r)} = (m+1)F_{k,n} + \sum_{j=1}^{n} \binom{m+j}{j+1} F_{k,n-j}.$$

Proof. From Theorem 2.3, we have

$$\begin{aligned}
\sum_{r=0}^{m} S_{k,n}^{r)} &= S_{k,n}^{0)} + \sum_{r=1}^{m} S_{k,n}^{r)} \\
&= S_{k,n}^{0)} + \sum_{r=1}^{m} \left(\sum_{j=0}^{n} \binom{r+j-1}{j} F_{k,n-j} \right) \\
&= S_{k,n}^{0)} + \sum_{j=0}^{n} \left(\sum_{r=1}^{m} \binom{r+j-1}{j} F_{k,n-j} \right) \\
&= S_{k,n}^{0)} + \sum_{j=0}^{n} \left(F_{k,n-j} \sum_{r=1}^{m} \binom{r+j-1}{j} \right) \\
&= S_{k,n}^{0)} + \sum_{j=0}^{n} \binom{m+j}{j+1} F_{k,n-j} \\
&= F_{k,n} + \sum_{j=0}^{n} \binom{m+j}{j+1} F_{k,n-j} \\
&= F_{k,n} + \binom{m}{1} F_{k,n} + \sum_{j=1}^{n} \binom{m+j}{j+1} F_{k,n-j} \\
&= (m+1)F_{k,n} + \sum_{j=1}^{n} \binom{m+j}{j+1} F_{k,n-j}.
\end{aligned}$$

□

Corollary 3.2. *For $m \geq 0$ and $n \geq 1$,*

1. $\displaystyle\sum_{r=0}^{m} S_{1,n}^{r)} = F_n + \sum_{j=0}^{n} \binom{m+j}{j+1} F_{n-j};$

2. $\displaystyle\sum_{r=0}^{m} S_{2,n}^{r)} = P_n + \sum_{j=0}^{n} \binom{m+j}{j+1} P_{n-j}.$

Corollary 3.3. *For $m \geq 0$,*

1. $\displaystyle\sum_{r=0}^{m} S_{k,1}^{r)} = (m+1)F_{k,1};$

2. $\displaystyle\sum_{r=0}^{m} S_{k,2}^{r)} = (m+1)F_{k,2} + \frac{m(m+1)}{2} F_{k,1};$

3. $\displaystyle\sum_{r=0}^{m} S_{k,3}^{r)} = (m+1)F_{k,3} + \frac{m(m+1)}{2} F_{k,2} + \frac{m(m+1)(m+2)}{3!} F_{k,1}.$

Theorem 3.4. *For $n \geq 1$,*

$$\sum_{r=0}^{n} S_{k,r+1}^{n-r)} = F_{k,n+1} + \sum_{j=1}^{n} 2^{n-j} F_{k,j}.$$

Proof.

$$\begin{aligned}
\sum_{r=0}^{n} S_{k,r+1}^{n-r)} &= S_{k,n+1}^{0)} + \sum_{r=0}^{n-1} S_{k,r+1}^{n-r)} \\
&= F_{k,n+1} + \sum_{r=0}^{n-1} \left[\sum_{i=0}^{r+1} \binom{n-r+i-1}{i} F_{k,r+1-i} \right] \\
&= F_{k,n+1} + \sum_{r=0}^{n-1} \left[\binom{n-r-1}{0} F_{k,r+1} + \binom{n-r}{1} F_{k,r} + \cdots + \binom{n-1}{r} F_{k,1} \right] \\
&= F_{k,n+1} + \sum_{r=0}^{n-1} \binom{n-1}{r} F_{k,1} + \sum_{r=0}^{n-2} \binom{n-2}{r} F_{k,2} + \cdots + \binom{0}{0} F_{k,n} \\
&= F_{k,n+1} + 2^{n-1} F_{k,1} + 2^{n-2} F_{k,2} + \cdots + F_{k,n} \\
&= F_{k,n+1} + \sum_{j=1}^{n} 2^{n-j} F_{k,j}.
\end{aligned}$$

□

Corollary 3.5. *For $n \geq 1$,*

1. $\sum_{r=0}^{n} S_{1,r+1}^{n-r)} = F_{n+1} + \sum_{j=1}^{n} 2^{n-j} F_j$;

2. $\sum_{r=0}^{n} S_{2,r+1}^{n-r)} = P_{n+1} + \sum_{j=1}^{n} 2^{n-j} P_j$.

Example 3.6. For $n = 1, 2, 3$,

$$\begin{aligned}
\sum_{r=0}^{1} S_{k,r+1}^{1-r)} &= F_{k,2} + F_{k,1}; \\
\sum_{r=0}^{2} S_{k,r+1}^{2-r)} &= F_{k,3} + F_{k,2} + 2F_{k,1}; \\
\sum_{r=0}^{3} S_{k,r+1}^{3-r)} &= F_{k,4} + F_{k,3} + 2F_{k,2} + 4F_{k,1}.
\end{aligned}$$

Next, we present alternating sums of iterated partial sums of k-Fibonacci sequence.

Theorem 3.7. *For $n \geq 1$,*

$$\sum_{r=0}^{n} (-1)^r S_{k,r+1}^{n-r)} = \begin{cases} F_{k,n} - F_{k,n+1} & ; \ n \text{ is odd}; \\ F_{k,n+1} - F_{k,n} & ; \ n \text{ is even}. \end{cases}$$

Proof. From Theorem 2.4 we have

$$\begin{aligned}\sum_{r=0}^{n}(-1)^r S_{k,r+1}^{n-r)} &= S_{k,1}^{n)} + \sum_{r=1}^{n-1}(-1)^r \left(S_{k,r}^{n-r)} + S_{k,r+1}^{n-r-1)}\right) + (-1)^n S_{k,n+1}^{0)} \\ &= S_{k,1}^{n)} - \left(S_{k,1}^{n-1)} + S_{k,2}^{n-2)}\right) + \cdots + (-1)^{n-1}\left(S_{k,n-1}^{1)} + S_{k,n}^{0)}\right) + (-1)^n S_{k,n+1}^{0)} \\ &= (-1)^{n-1} S_{k,n}^{0)} + (-1)^n S_{k,n+1}^{0)} \\ &= (-1)^n \left(S_{k,n+1}^{0)} - S_{k,n}^{0)}\right) \\ &= (-1)^n \left(F_{k,n+1} - F_{k,n}\right).\end{aligned}$$

□

Corollary 3.8. *For $n \geq 1$,*

1. $$\sum_{r=0}^{n}(-1)^r S_{1,r+1}^{n-r)} = \begin{cases} -F_{n-1} & ;\ n \text{ is odd;} \\ F_{n-1} & ;\ n \text{ is even.} \end{cases}$$
2. $$\sum_{r=0}^{n}(-1)^r S_{2,r+1}^{n-r)} = \begin{cases} P_n - P_{n+1} & ;\ n \text{ is odd;} \\ P_{n+1} - P_n & ;\ n \text{ is even.} \end{cases}$$

The following theorem present different of alternating sums of iterated partial sum of k-Fibonacci sequence.

Theorem 3.9. *For $n \geq 1$,*

$$\sum_{r=0}^{n-1}(-1)^r \left[k S_{k,r+1}^{n-r)} - S_{k,r+1}^{n-r-1)}\right] = (-1)^n \left[F_{k,n} - F_{k,n+1}\right].$$

Proof. From Theorem 3.7 we have,

$$\sum_{r=0}^{n}(-1)^r k S_{k,r+1}^{n-r)} = (-1)^n k \left(F_{k,n+1} - F_{k,n}\right) \tag{3.1}$$

$$\sum_{r=0}^{n-1}(-1)^r S_{k,r+1}^{n-r-1)} = (-1)^{n-1}\left(F_{k,n} - F_{k,n-1}\right). \tag{3.2}$$

Subtract (3.2) from (3.1) we get

$$\begin{aligned}(-1)^n k F_{k,n+1} + \sum_{r=0}^{n-1}(-1)^r \left[k S_{k,r+1}^{n-r)} - S_{k,r+1}^{n-r-1)}\right] &= (-1)^n \left(k F_{k,n+1} - k F_{k,n} + F_{k,n} - F_{k,n-1}\right) \\ \sum_{r=0}^{n-1}(-1)^r \left[k S_{k,r+1}^{n-r)} - S_{k,r+1}^{n-r-1)}\right] &= (-1)^n \left(F_{k,n} - F_{k,n+1}\right).\end{aligned}$$

□

From Theorems 3.7 and 3.9, we have

$$\sum_{r=0}^{n}(-1)^r S_{k,r+1}^{n-r)} = \sum_{r=0}^{n-1}(-1)^{r+1}\left[k S_{k,r+1}^{n-r)} - S_{k,r+1}^{n-r-1)}\right].$$

Conclusion. In this paper, we present the sums of the iterated partial sums of k-Fibonacci sequence and alternating sums. Moreover, we present results to some special cases such as classical Fibonacci and Pell sequences.

Acknowledgment. The authors would like to thank the referees for their valuable comments.

References

[1] H. V. Chu, *Partial sums of the Fibonacci sequence*, Fibonacci Quart. **59**(2) (2021), 132–135.

[2] S. Falcón and Á. Plaza, *Iterated Partial Sums of the k-Fibonacci Sequences*, Axioms. **11**(10) (2022), 542.

[3] S. Falcón and Á. Plaza, *On the Fibonacci k-numbers*, Chaos, Solitons and Fractals. **32**(5) (2007), 1615–1624.

[4] S. Falcón and Á. Plaza, *The k-Fibonacci sequence and the Pascal 2-trangle*, Chaos, Solitons and Fractals. **33**(1) (2007), 38–49.

[5] T. Koshy, *Fibonacci and Lucas numbers with applications*, Wiley, New York, 2017.

[6] A. F. Horadam, *Pell identities*, Fibonacci Quart. **9**(3) (1971), 245–252.

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

สมบัติบางประการสำหรับลำดับ k-โอเรสเมในรูปแบบเชิงซ้อน*

ชนนิกานต์ คนเพียร[1,†] และ บุญยงค์ ศรีพลแผ้ว[1,‡]

[1]สาขาวิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา 20131

บทคัดย่อ

ในงานวิจัยนี้ได้พิสูจน์สมบัติของลำดับ k-โอเรสเมในรูปแบบเชิงซ้อน โดยการสร้างฟังก์ชันก่อกำเนิด พิสูจน์สูตรไบเนตและเอกลักษณ์บางประการของลำดับ k-โอเรสเมเชิงซ้อน

คำสำคัญ: k-โอเรสเม, k-โอเรสเมเชิงซ้อน, ฟังก์ชันก่อกำเนิด

2020 MSC: ปฐมภูมิ 11B37; ทุติยภูมิ 11B39

1 บทนำ

จากการนำเสนอของ Horadam [4] เกี่ยวกับประวัติความเป็นมาของตัวเลขที่เกิดจากนิโคล โอเรสเม คือลำดับ $\{O_n\}_{n\geq 1} = \left\{\frac{n}{2^n}\right\} = \left\{\frac{1}{2}, \frac{2}{4}, \frac{3}{8}, ..., \frac{n}{2^n}, ...\right\}$ โดยที่ตัวเลขของโอเรสเมสามารถนิยามผ่านความสัมพันธ์เวียนเกิด คือ $O_{n+2} = O_{n+1} - \frac{1}{4}O_n$ มีเงื่อนไขเริ่มต้น คือ $O_0 = 0$ และ $O_1 = \frac{1}{2}$ ซึ่งจำนวนโอเรสเมมีคุณสมบัติที่น่าสนใจเป็นจำนวนมากและมีบทประยุกต์ในหลายสาขาของวิทยาศาสตร์ (ดูตัวอย่างได้จาก [1–3]) จากนั้น Cerda-Morales [5] ได้กล่าวถึงนิยามความสัมพันธ์เวียนเกิดของลำดับ k-โอเรสเมว่า $O_n^{(k)} = O_{n-1}^{(k)} - \frac{1}{k^2}O_{n-2}^{(k)}$ โดยมีเงื่อนไขเริ่มต้นของสองจำนวนแรก คือ $O_0^{(k)} = 0$ และ $O_1^{(k)} = \frac{1}{k}$ ซึ่งได้พิสูจน์สูตรไบเนต เอกลักษณ์ของ Cassini ของลำดับ k-โอเรสเม โดยมีเงื่อนไข คือ $k^2 - 4 > 0$ และได้พิสูจน์สูตรการหาผลรวม รวมถึงการพิสูจน์คุณสมบัติหลายประการของพหุนามของ k-โอเรสเม และ Soykan [7] ได้แนะนำ k -โอเรสเมทั่วไป ได้กล่าวถึงลำดับ k-โอเรสเม และลำดับ k-โอเรสเมลูคัส ซึ่งได้สร้างฟังก์ชันก่อกำเนิด พิสูจน์เอกลักษณ์ของ Cassini พิสูจน์เอกลักษณ์ของ Catalan พิสูจน์เอกลักษณ์ของ d'Ocagne ของลำดับ k-โอเรสเม และได้พิสูจน์สูตรผลรวมรูปแบบต่าง ๆ พร้อมทั้งการหาเมทริกซ์ที่เกี่ยวข้องกับตัวเลข k-โอเรสเม ที่มีเงื่อนไข คือ $k^2 - 4 > 0$

*งานวิจัยเรื่องนี้ได้รับทุนสนับสนุนจากภาควิชาคณิตศาสตร์ และคณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา
†ผู้นำเสนอ ‡ผู้แต่งหลัก
อีเมล: 63030028@go.buu.ac.th (ชนนิกานต์ คนเพียร), boonyong@buu.ac.th (บุญยงค์ ศรีพลแผ้ว).

จากนิยามความสัมพันธ์เวียนเกิดของลำดับฟีโบนักชี ที่กล่าวว่า $F_{n+2} = F_{n+1} + F_n$ ซึ่งมีเงื่อนไขเริ่มต้นของสองจำนวนแรก คือ $F_0 = 0, F_1 = 1$ ตามลำดับ และ Harman [6] ได้สร้างนิยามลำดับฟิโบนักชีเชิงซ้อน เมื่อ n เป็นจำนวนเต็ม คือ

$$G_n = F_n + iF_{n+1}$$

โดย $i = \sqrt{-1}$ และได้สร้างฟังก์ชันก่อกำเนิด พิสูจน์สูตรไบเนตและเอกลักษณ์ต่าง ๆ ของลำดับฟิโบนักชีเชิงซ้อน

งานวิจัยนี้จึงสนใจที่จะศึกษาลำดับ k-โอเรสเม เพื่อนำมาทำเป็นลำดับ k-โอเรสเมเชิงซ้อน พร้อมทั้งสร้างฟังก์ชันก่อกำเนิด พิสูจน์สูตรไบเนต พิสูจน์เอกลักษณ์และสมบัติบางประการ รวมถึงสูตรของผลรวมรูปแบบต่าง ๆ ของลำดับ k-โอเรสเมเชิงซ้อน

2 ความรู้พื้นฐานและงานวิจัยที่เกี่ยวข้อง

ในงานวิจัยของ Cerda-Morales [5] ได้นิยามลำดับทั่วไปของ k-โอเรสเม จากความสัมพันธ์เวียนเกิด คือ $O_n^{(k)} = O_{n-1}^{(k)} - \frac{1}{k^2}O_{n-2}^{(k)}$ โดยมีเงื่อนไขเริ่มต้น คือ $O_0^{(k)} = 0$ และ $O_1^{(k)} = \dfrac{1}{k}$ และกล่าวถึงสูตรไบเนตของลำดับ k-โอเรสเม เมื่อ $k^2 - 4 > 0$ คือ $O_n^{(k)} = \dfrac{1}{\sqrt{k^2-4}}\left(\alpha^n - \beta^n\right)$ เมื่อ $\alpha = \dfrac{k + \sqrt{k^2-4}}{2k}$ และ $\beta = \dfrac{k - \sqrt{k^2-4}}{2k}$ โดยที่ α และ β เป็นผลเฉลยของสมการลักษณะเฉพาะ $r^2 - r + \dfrac{1}{k^2} = 0$

นอกจากนั้นในงานวิจัยของ Cerda-Morales [5] ได้พิสูจน์เอกลักษณ์ของ Cassini ของลำดับ k-โอเรสเม

$$O_{n+1}^{(k)}O_{n-1}^{(k)} - \left(O_n^{(k)}\right)^2 = -\left(\frac{1}{k^2}\right)^n$$

Soykan [7] ได้พิสูจน์เอกลักษณ์ของ Catalan และเอกลักษณ์ของ d'Ocagne ของลำดับ k-โอเรสเม โดยที่ m, n และ r เป็นจำนวนเต็ม เมื่อ $k^2 - 4 > 0$ ตามลำดับ คือ

$$O_{n+r}^{(k)}O_{n-r}^{(k)} - \left(O_n^{(k)}\right)^2 = -\frac{1}{2^{2r}k^{2n}(k^2-4)}\left[\left(k + \sqrt{k^2-4}\right)^r - \left(k - \sqrt{k^2-4}\right)^r\right]^2$$

และ

$$O_{m+1}^{(k)}O_n^{(k)} - O_m^{(k)}O_{n+1}^{(k)} = -\frac{1}{k\sqrt{k^2-4}}\left(\alpha^m\beta^n - \alpha^n\beta^m\right)$$

3 ผลการศึกษา

จากงานวิจัยของ Harman [6] ที่ได้ให้นิยามลำดับทั่วไปของฟิโบนักชีเชิงซ้อน เราได้นำแนวความคิดเดียวกันมาสร้างนิยามของลำดับ k-โอเรสเมเชิงซ้อน ดังนี้

นิยาม 3.1. ลำดับทั่วไปของ k-โอเรสเมเชิงซ้อน $\left(CO_n^{(k)}\right)$ เมื่อ n เป็นจำนวนเต็ม กำหนดโดย

$$CO_n^{(k)} = O_n^{(k)} + iO_{n+1}^{(k)}$$

โดย $O_n^{(k)}$ คือพจน์ที่ n ของลำดับ k-โอเรสเม และ $i = \sqrt{-1}$

เราจะเริ่มจากการพิสูจน์สมการเวียนเกิดของพจน์ที่ n ของลำดับ k -โอเรสเมเชิงซ้อน ดังนี้

บทตั้ง 3.2. สมการเวียนเกิดของ $CO_n^{(k)}$ โดยที่ n เป็นจำนวนเต็ม คือ

$$CO_{n+2}^{(k)} = CO_{n+1}^{(k)} - \frac{1}{k^2}CO_n^{(k)}$$

พิสูจน์. จาก $CO_{n+2}^{(k)} = O_{n+2}^{(k)} + iO_{n+3}^{(k)}$ จะได้

$$\begin{aligned} CO_{n+2}^{(k)} &= \left(O_{n+1}^{(k)} - \frac{1}{k^2}O_n^{(k)}\right) + i\left(O_{n+2}^{(k)} - \frac{1}{k^2}O_{n+1}^{(k)}\right) \\ &= O_{n+1}^{(k)} - \frac{1}{k^2}O_n^{(k)} + iO_{n+2}^{(k)} - \frac{1}{k^2}iO_{n+1}^{(k)} \\ &= O_{n+1}^{(k)} + iO_{n+2}^{(k)} - \frac{1}{k^2}\left(O_n^{(k)} + iO_{n+1}^{(k)}\right) \\ &= CO_{n+1}^{(k)} - \frac{1}{k^2}CO_n^{(k)} \end{aligned}$$

□

เราพิสูจน์สูตรไบเนตของลำดับ k-โอเรสเมเชิงซ้อน ซึ่งเป็นสูตรในการหาค่าพจน์ทั่วไปของลำดับ k-โอเรสเมเชิงซ้อน

ทฤษฎีบท 3.3. สูตรไบเนตของลำดับ k-โอเรสเมเชิงซ้อน เมื่อ n เป็นจำนวนเต็ม และ $k^2 - 4 > 0$ คือ

$$CO_n^{(k)} = \frac{1}{\sqrt{k^2-4}}\left[\alpha^n\tilde{\alpha} - \beta^n\tilde{\beta}\right]$$

เมื่อ $\alpha = \dfrac{k+\sqrt{k^2-4}}{2k}$, $\tilde{\alpha} = 1 + \alpha i$, $\beta = \dfrac{k-\sqrt{k^2-4}}{2k}$ และ $\tilde{\beta} = 1 + \beta i$
โดยที่ α และ β เป็นผลเฉลยของสมการลักษณะเฉพาะ $r^2 - r + \dfrac{1}{k^2} = 0$

พิสูจน์. จากทฤษฎีบทสูตรไบเนตของลำดับ k-โอเรสเม สามารถนำมาพิสูจน์ได้ ดังนี้

$$\begin{aligned} CO_n^{(k)} &= O_n^{(k)} + iO_{n+1}^{(k)} \\ &= \left[\frac{1}{\sqrt{k^2-4}}\left(\alpha^n - \beta^n\right)\right] + i\left[\frac{1}{\sqrt{k^2-4}}\left(\alpha^{n+1} - \beta^{n+1}\right)\right] \\ &= \frac{1}{\sqrt{k^2-4}}\left[\alpha^n - \beta^n + i\alpha^{n+1} - i\beta^{n+1}\right] \\ &= \frac{1}{\sqrt{k^2-4}}\left[\alpha^n + i\alpha^{n+1} - \left(\beta^n + i\beta^{n+1}\right)\right] \\ &= \frac{1}{\sqrt{k^2-4}}\left[\alpha^n(1+\alpha i) - \beta^n(1+\beta i)\right] \\ &= \frac{1}{\sqrt{k^2-4}}\left[\alpha^n\tilde{\alpha} - \beta^n\tilde{\beta}\right] \end{aligned}$$

□

ฟังก์ชันก่อกำเนิดของลำดับ k-โอเรสเมเชิงซ้อน เมื่อ n เป็นจำนวนเต็ม อยู่ในรูปทั่วไปคือ

$$g(t) = \sum_{n=0}^{\infty} CO_n^{(k)}t^n$$

เราจะนำฟังก์ชันก่อกำเนิดของลำดับ k-โอเรสเมเชิงซ้อน มาหาค่าผลบวกให้ได้ผลลัพธ์เป็นทฤษฎีบทดังต่อไปนี้

ทฤษฎีบท 3.4. ฟังก์ชันก่อกำเนิดของลำดับ k-โอเรสเมเชิงซ้อน สามารถเขียนได้ดังนี้

$$g(t) = \frac{(i+t)k}{k^2 - k^2t + t^2}$$

พิสูจน์. โดยใช้ความสัมพันธ์เวียนเกิดของลำดับ k-โอเรสเมเชิงซ้อนจะได้ว่า

$$\begin{aligned}
& g(t)\cdot\left[t^0 - t^1 + \frac{1}{k^2}t^2\right] \\
&= \left(\sum_{n=0}^{\infty} CO_n^{(k)}t^n\right)\left[t^0 - t^1 + \frac{1}{k^2}t^2\right] \\
&= \sum_{n=0}^{\infty} CO_n^{(k)}t^n - \sum_{n=0}^{\infty} CO_n^{(k)}t^{n+1} + \frac{1}{k^2}\sum_{n=0}^{\infty} CO_n^{(k)}t^{n+2} \\
&= \sum_{j=0}^{\infty} CO_j^{(k)}t^j - \sum_{j=1}^{\infty} CO_{j-1}^{(k)}t^j + \frac{1}{k^2}\sum_{j=2}^{\infty} CO_{j-2}^{(k)}t^j \\
&= \left(CO_0^{(k)}t^0 + CO_1^{(k)}t^1 + \sum_{j=2}^{\infty} CO_j^{(k)}t^j\right) - \left(CO_0^{(k)}t^1 + \sum_{j=2}^{\infty} CO_{j-1}^{(k)}t^j\right) + \frac{1}{k^2}\left(\sum_{j=2}^{\infty} CO_{j-2}^{(k)}t^j\right) \\
&= CO_0^{(k)}t^0 + CO_1^{(k)}t - CO_0^{(k)}t + \sum_{j=2}^{\infty}\left[CO_j^{(k)} - CO_{j-1}^{(k)} + \frac{1}{k^2}CO_{j-2}^{(k)}\right]t^j \\
&= CO_0^{(k)}t^0 + CO_1^{(k)}t - CO_0^{(k)}t
\end{aligned}$$

จะได้ว่า

$$\begin{aligned}
g(t) &= \frac{CO_0^{(k)}t^0 + CO_1^{(k)}t - CO_0^{(k)}t}{\left[t^0 - t^1 + \frac{1}{k^2}t^2\right]} \\
&= \frac{CO_0^{(k)}(1-t) + CO_1^{(k)}t}{1 - t + \frac{t^2}{k^2}}
\end{aligned}$$

แทนค่า $CO_0^{(k)}$ และ $CO_1^{(k)}$ จะได้ว่า

$$\begin{aligned}
g(t) &= \frac{\frac{i}{k}\left(1-t\right) + \left(\frac{1+i}{k}\right)t}{1 - t + \frac{t^2}{k^2}} \\
&= \frac{\left(\frac{i - it + t + it}{k}\right)}{\left(\frac{k^2 - k^2t + t}{k^2}\right)} \\
&= \frac{(i+t)k^2}{k\left(k^2 - k^2t + t^2\right)} \\
&= \frac{(i+t)k}{k^2 - k^2t + t^2}
\end{aligned}$$

□

ต่อไปเราจะพิสูจน์เอกลักษณ์ Cassini ของลำดับ k-โอเรสเมเชิงซ้อน โดยใช้เอกลักษณ์ของ Cassini และเอกลักษณ์ของ d'Ocagne ของลำดับ k-โอเรสเม

ทฤษฎีบท 3.5. เอกลักษณ์ Cassini ของลำดับ k-โอเรสเมชิงซ้อน เมื่อ n เป็นจำนวนเต็ม และ $k^2-4>0$ คือ

$$CO_{n+1}^{(k)}\cdot CO_{n-1}^{(k)}-\left(CO_n^{(k)}\right)^2=\left(\frac{1}{k}\right)^{2n-1}\left[\left(\frac{1}{k}\right)^3-\left(\frac{1}{k}\right)-i\left(\frac{1}{k}\right)\right]$$

พิสูจน์. จากนิยามของลำดับทั่วไปของ k-โอเรสเมเชิงซ้อนจะได้ว่า
$CO_{n+1}^{(k)}\cdot CO_{n-1}^{(k)}-\left(CO_n^{(k)}\right)^2$

$$\begin{aligned}
&= \left(O_{n+1}^{(k)}+iO_{n+2}^{(k)}\right)\left(O_{n-1}^{(k)}+iO_n^{(k)}\right)-\left(O_n^{(k)}+iO_{n+1}^{(k)}\right)^2\\
&= O_{n+1}^{(k)}O_{n-1}^{(k)}+iO_{n+1}^{(k)}O_n^{(k)}+iO_{n+2}^{(k)}O_{n-1}^{(k)}+i^2O_{n+2}^{(k)}O_n^{(k)}\\
&\quad -\left[\left(O_n^{(k)}\right)^2+2iO_n^{(k)}O_{n+1}^{(k)}+i^2\left(O_{n+1}^{(k)}\right)^2\right]\\
&= O_{n+1}^{(k)}O_{n-1}^{(k)}+iO_{n+1}^{(k)}O_n^{(k)}+iO_{n+2}^{(k)}O_{n-1}^{(k)}-O_{n+2}^{(k)}O_n^{(k)}\\
&\quad -\left(O_n^{(k)}\right)^2-2iO_n^{(k)}O_{n+1}^{(k)}+\left(O_{n+1}^{(k)}\right)^2\\
&= O_{n+1}^{(k)}O_{n-1}^{(k)}-O_{n+2}^{(k)}O_n^{(k)}-\left(O_n^{(k)}\right)^2+\left(O_{n+1}^{(k)}\right)^2+\left(O_{n+1}^{(k)}O_n^{(k)}+O_{n+2}^{(k)}O_{n-1}^{(k)}-2O_n^{(k)}O_{n+1}^{(k)}\right)i\\
&= \left[O_{n+1}^{(k)}O_{n-1}^{(k)}-\left(O_n^{(k)}\right)^2\right]-\left[O_{n+2}^{(k)}O_n^{(k)}-\left(O_{n+1}^{(k)}\right)^2\right]+\left(O_{n+2}^{(k)}O_{n-1}^{(k)}-O_{n+1}^{(k)}O_n^{(k)}\right)i
\end{aligned}$$

จากเอกลักษณ์ Cassini และ d'Ocagne ของลำดับ k-โอเรสเม และบทตั้ง 3.3 จะได้ว่า
$CO_{n+1}^{(k)}\cdot CO_{n-1}^{(k)}-\left(CO_n^{(k)}\right)^2$

$$\begin{aligned}
&= -\left(\frac{1}{k}\right)^{2n}+\left(\frac{1}{k}\right)^{2(n+1)}-\frac{1}{k\sqrt{k^2-4}}\left[\alpha^{n+1}\beta^{n-1}-\alpha^{n-1}\beta^{n+1}\right]i\\
&= \left(\frac{1}{k}\right)^{2n+2}-\left(\frac{1}{k}\right)^{2n}-\frac{1}{k\sqrt{k^2-4}}\left[\alpha^2(\alpha\beta)^{n-1}-\beta^2(\alpha\beta)^{n-1}\right]i\\
&= \left(\frac{1}{k}\right)^{2n+2}-\left(\frac{1}{k}\right)^{2n}-\left(\frac{1}{k}\right)\frac{1}{\sqrt{k^2-4}}\left[\alpha^2\left(\frac{1}{k^2}\right)^{n-1}-\beta^2\left(\frac{1}{k^2}\right)^{n-1}\right]i\\
&= \left(\frac{1}{k}\right)^{2n+2}-\left(\frac{1}{k}\right)^{2n}-\left(\frac{1}{k}\right)\frac{1}{\sqrt{k^2-4}}\left[\left(\frac{1}{k}\right)^{2(n-1)}\left(\alpha^2-\beta^2\right)\right]i\\
&= \left(\frac{1}{k}\right)^{2n+2}-\left(\frac{1}{k}\right)^{2n}-\left(\frac{1}{k}\right)^{2n-1}\left[\frac{1}{\sqrt{k^2-4}}\left(\alpha^2-\beta^2\right)\right]i\\
&= \left(\frac{1}{k}\right)^{2n-1}\left[\left(\frac{1}{k}\right)^3-\left(\frac{1}{k}\right)-i\left(\frac{\alpha^2-\beta^2}{\sqrt{k^2-4}}\right)\right]
\end{aligned}$$

จากสูตรไบเนต จะได้ว่า

$$CO_{n+1}^{(k)}\cdot CO_{n-1}^{(k)}-\left(CO_n^{(k)}\right)^2 = \left(\frac{1}{k}\right)^{2n-1}\left[\left(\frac{1}{k}\right)^3-\left(\frac{1}{k}\right)-iO_2^{(k)}\right]$$

ดังนั้น

$$CO_{n+1}^{(k)} \cdot CO_{n-1}^{(k)} - \left(CO_n^{(k)}\right)^2 = \left(\frac{1}{k}\right)^{2n-1}\left[\left(\frac{1}{k}\right)^3 - \left(\frac{1}{k}\right) - i\left(\frac{1}{k}\right)\right]$$

□

เราจะพิสูจน์ บทตั้ง 3.5 โดยใช้สูตรไบเนตของลำดับ k-โอเรสเม เพื่อนำสมบัติของบทตั้งนี้ไปใช้ในการพิสูจน์ทฤษฎีบท 3.6

บทตั้ง 3.6. สมบัติบางประการของลำดับ k-โอเรสเม โดยที่ n และ r เป็นจำนวนเต็ม เมื่อ $k^2-4>0$ คือ

$$O_{n+r}^{(k)}O_{n-r+1}^{(k)} + O_{n+r+1}^{(k)}O_{n-r}^{(k)} - 2O_n^{(k)}O_{n+1}^{(k)} = -(k)^{-2n+2r}\left(O_r^{(k)}\right)^2$$

พิสูจน์. จากสูตรของไบเนต จะได้ว่า

$$\begin{aligned}
&O_{n+r}^{(k)}O_{n-r+1}^{(k)} + O_{n+r+1}^{(k)}O_{n-r}^{(k)} - 2O_n^{(k)}O_{n+1}^{(k)} \\
&= \frac{1}{\sqrt{k^2-4}}\left(\alpha^{n+r}-\beta^{n+r}\right)\frac{1}{\sqrt{k^2-4}}\left(\alpha^{n-r+1}-\beta^{n-r+1}\right) \\
&\quad + \frac{1}{\sqrt{k^2-4}}\left(\alpha^{n+r+1}-\beta^{n+r+1}\right)\frac{1}{\sqrt{k^2-4}}\left(\alpha^{n-r}-\beta^{n-r}\right) \\
&\quad -2\left[\frac{1}{\sqrt{k^2-4}}\left(\alpha^{n}-\beta^{n}\right)\frac{1}{\sqrt{k^2-4}}\left(\alpha^{n+1}-\beta^{n+1}\right)\right] \\
&= \frac{1}{k^2-4}\left[\left(\alpha^{n+r}-\beta^{n+r}\right)\left(\alpha^{n-r+1}-\beta^{n-r+1}\right) + \left(\alpha^{n+r+1}-\beta^{n+r+1}\right)\left(\alpha^{n-r}-\beta^{n-r}\right)\right. \\
&\quad \left. -2\left(\alpha^{n}-\beta^{n}\right)\left(\alpha^{n+1}-\beta^{n+1}\right)\right] \\
&= \frac{1}{k^2-4}\left[\alpha^{n+r}\alpha^{n-r+1} - \alpha^{n+r}\beta^{n-r+1} - \alpha^{n-r+1}\beta^{n+r} + \beta^{n+r}\beta^{n-r+1} + \alpha^{n+r+1}\alpha^{n-r}\right. \\
&\quad -\alpha^{n+r+1}\beta^{n-r} - \alpha^{n-r}\beta^{n+r+1} + \beta^{n-r}\beta^{n+r+1} - 2\alpha^{n}\alpha^{n+1} + 2\alpha^{n}\beta^{n+1} + 2\alpha^{n+1}\beta^{n} \\
&\quad \left. -2\beta^{n}\beta^{n+1}\right] \\
&= \frac{1}{k^2-4}\left[\alpha^{2n+1} - \alpha^{n+r}\beta^{n-r+1} - \alpha^{n-r+1}\beta^{n+r} + \beta^{2n+1} + \alpha^{2n+1} - \alpha^{n+r+1}\beta^{n-r}\right. \\
&\quad \left. -\alpha^{n-r}\beta^{n+r+1} + \beta^{2n+1} - 2\alpha^{2n+1} + 2\alpha^{n}\beta^{n+1} + 2\alpha^{n+1}\beta^{n} - 2\beta^{2n+1}\right] \\
&= \frac{1}{k^2-4}\left[-\alpha^{n+r}\beta^{n-r+1} - \alpha^{n-r+1}\beta^{n+r} - \alpha^{n+r+1}\beta^{n-r} - \alpha^{n-r}\beta^{n+r+1} + 2\alpha^{n}\beta^{n+1}\right. \\
&\quad \left. +2\alpha^{n+1}\beta^{n}\right] \\
&= \frac{-(\alpha\beta)^{n-r}}{k^2-4}\left[\alpha^{n+r-(n-r)}\beta^{n-r+1-(n-r)} + \alpha^{n-r+1-(n-r)}\beta^{n+r-(n-r)}\right. \\
&\quad +\alpha^{n+r+1-(n-r)}\beta^{n-r-(n-r)} + \alpha^{n-r-(n-r)}\beta^{n+r+1-(n-r)} - 2\alpha^{n-(n-r)}\beta^{n+1-(n-r)} \\
&\quad \left. -2\alpha^{n+1-(n-r)}\beta^{n-(n-r)}\right] \\
&= \frac{-(\alpha\beta)^{n-r}}{k^2-4}\left[\alpha^{2r}\beta + \alpha\beta^{2r} + \alpha^{2r+1} + \beta^{2r+1} - 2\alpha^{r}\beta^{r+1} - 2\alpha^{r+1}\beta^{r}\right]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{-(\alpha\beta)^{n-r}}{k^2-4}\left[\left(\alpha^{2r+1}+\alpha\beta^{2r}-2\alpha^{r+1}\beta^r\right)+\left(\beta^{2r+1}+\alpha^{2r}\beta-2\alpha^r\beta^{r+1}\right)\right]\\
&= \frac{-(\alpha\beta)^{n-r}}{k^2-4}\left[\alpha\left(\alpha^{2r}+\beta^{2r}-2\alpha^r\beta^r\right)+\beta\left(\beta^{2r}+\alpha^{2r}-2\alpha^r\beta^r\right)\right]\\
&= \frac{-(\alpha\beta)^{n-r}}{k^2-4}\left[\alpha\left(\alpha^r-\beta^r\right)^2+\beta\left(\alpha^r-\beta^r\right)^2\right]\\
&= \frac{-(\alpha\beta)^{n-r}}{k^2-4}\left[\left(\alpha+\beta\right)\left(\alpha^r-\beta^r\right)^2\right]\\
&= -\left(\alpha\beta\right)^{n-r}\left(\alpha+\beta\right)\frac{1}{k^2-4}\left(\alpha^r-\beta^r\right)^2\\
&= -\left(\alpha\beta\right)^{n-r}\left(\alpha+\beta\right)\left(\frac{1}{\sqrt{k^2-4}}\right)^2\left(\alpha^r-\beta^r\right)^2
\end{aligned}
$$

ดังนั้น

$$
\begin{aligned}
O^{(k)}_{n+r}O^{(k)}_{n-r+1}+O^{(k)}_{n+r+1}O^{(k)}_{n-r}-2O^{(k)}_{n}O^{(k)}_{n+1} &= -\left(\frac{1}{k^2}\right)^{n-r}(1)\left(O^{(k)}_r\right)^2\\
&= -(k)^{-2n+2r}\left(O^{(k)}_r\right)^2
\end{aligned}
$$

□

ต่อไปเราจะพิสูจน์เอกลักษณ์ Catalan ของลำดับ k-โอเรสเมเชิงซ้อน โดยใช้เอกลักษณ์ของ Catalan และสูตรไบเนต ของลำดับ k-โอเรสเม และบทตั้ง 3.6

ทฤษฎีบท 3.7. เอกลักษณ์ Catalan ของลำดับ k-โอเรสเมเชิงซ้อน โดยที่ n และ r เป็นจำนวนเต็ม เมื่อ $k^2-4>0$ คือ

$$
CO^{(k)}_{n+r}CO^{(k)}_{n-r}-\left(CO^{(k)}_n\right)^2=(k)^{-2n+2r}\left[\frac{1}{k^2}-1-i\right]\left(O^{(k)}_r\right)^2
$$

พิสูจน์. จากนิยามของลำดับ k-โอเรสเมเชิงซ้อน จะได้ว่า

$$
\begin{aligned}
&CO^{(k)}_{n+r}CO^{(k)}_{n-r}-\left(CO^{(k)}_n\right)^2\\
&= \left(O^{(k)}_{n+r}+iO^{(k)}_{n+r+1}\right)\left(O^{(k)}_{n-r}+iO^{(k)}_{n-r+1}\right)-\left(O^{(k)}_n+iO^{(k)}_{n+1}\right)^2\\
&= O^{(k)}_{n+r}O^{(k)}_{n-r}+iO^{(k)}_{n+r}O^{(k)}_{n-r+1}+iO^{(k)}_{n+r+1}O^{(k)}_{n-r}+i^2O^{(k)}_{n+r+1}O^{(k)}_{n-r+1}\\
&\quad -\left[\left(O^{(k)}_n\right)^2+2iO^{(k)}_nO^{(k)}_{n+1}+i^2\left(O^{(k)}_{n+1}\right)^2\right]
\end{aligned}
$$

$$\begin{aligned}
&= O^{(k)}_{n+r}O^{(k)}_{n-r} + iO^{(k)}_{n+r}O^{(k)}_{n-r+1} + iO^{(k)}_{n+r+1}O^{(k)}_{n-r} - O^{(k)}_{n+r+1}O^{(k)}_{n-r+1} \\
&\quad -\left(O^{(k)}_{n}\right)^2 - 2iO^{(k)}_{n}O^{(k)}_{n+1} + \left(O^{(k)}_{n+1}\right)^2 \\
&= O^{(k)}_{n+r}O^{(k)}_{n-r} - O^{(k)}_{n+r+1}O^{(k)}_{n-r+1} - \left(O^{(k)}_{n}\right)^2 + \left(O^{(k)}_{n+1}\right)^2 \\
&\quad +\left[O^{(k)}_{n+r}O^{(k)}_{n-r+1} + O^{(k)}_{n+r+1}O^{(k)}_{n-r} - 2O^{(k)}_{n}O^{(k)}_{n+1}\right]i \\
&= \left[O^{(k)}_{n+r}O^{(k)}_{n-r} - \left(O^{(k)}_{n}\right)^2\right] - \left[O^{(k)}_{n+r+1}O^{(k)}_{n-r+1} - \left(O^{(k)}_{n+1}\right)^2\right] \\
&\quad +\left[O^{(k)}_{n+r}O^{(k)}_{n-r+1} + O^{(k)}_{n+r+1}O^{(k)}_{n-r} - 2O^{(k)}_{n}O^{(k)}_{n+1}\right]i
\end{aligned}$$

จากเอกลักษณ์ Catalan และบทตั้ง 3.6 จะได้ว่า
$CO^{(k)}_{n+r}CO^{(k)}_{n-r} - \left(CO^{(k)}_{n}\right)^2$

$$\begin{aligned}
&= -\frac{1}{2^{2r}k^{2n}\left(k^2-4\right)}\left[\left(k+\sqrt{k^2-4}\right)^r - \left(k-\sqrt{k^2-4}\right)^r\right]^2 \\
&\quad +\frac{1}{2^{2r}k^{2(n+1)}\left(k^2-4\right)}\left[\left(k+\sqrt{k^2-4}\right)^r - \left(k-\sqrt{k^2-4}\right)^r\right]^2 - (k)^{-2n+2r}\left(O^{(k)}_{r}\right)^2 i \\
&= -\frac{1}{2^{2r}k^{2r}\left(k^{2n-2r}\right)}\left(\frac{1}{\sqrt{k^2-4}}\right)^2\left[\left(k+\sqrt{k^2-4}\right)^r - \left(k-\sqrt{k^2-4}\right)^r\right]^2 \\
&\quad +\frac{1}{2^{2r}k^{2r}\left(k^{2n-2r+2}\right)}\left(\frac{1}{\sqrt{k^2-4}}\right)^2\left[\left(k+\sqrt{k^2-4}\right)^r - \left(k-\sqrt{k^2-4}\right)^r\right]^2 \\
&\quad -(k)^{-2n+2r}\left(O^{(k)}_{r}\right)^2 i \\
&= -\frac{1}{k^{2n-2r}}\left(\frac{1}{\sqrt{k^2-4}}\left[\left(\frac{k+\sqrt{k^2-4}}{2k}\right)^r - \left(\frac{k-\sqrt{k^2-4}}{2k}\right)^r\right]\right)^2 \\
&\quad +\frac{1}{k^{2n-2r+2}}\left(\frac{1}{\sqrt{k^2-4}}\left[\left(\frac{k+\sqrt{k^2-4}}{2k}\right)^r - \left(\frac{k-\sqrt{k^2-4}}{2k}\right)^r\right]\right)^2 \\
&\quad -(k)^{-2n+2r}\left(O^{(k)}_{r}\right)^2 i
\end{aligned}$$

จากสูตรของไบเนต จะได้ว่า

$$\begin{aligned}
CO^{(k)}_{n+r}CO^{(k)}_{n-r} - \left(CO^{(k)}_{n}\right)^2 &= -\frac{1}{k^{2n-2r}}\left(O^{(k)}_{r}\right)^2 + \frac{1}{k^{2n-2r+2}}\left(O^{(k)}_{r}\right)^2 - \frac{1}{k^{2n-2r}}\left(O^{(k)}_{r}\right)^2 i \\
&= \frac{1}{k^{2n-2r}}\left[-1+\frac{1}{k^2}-i\right]\left(O^{(k)}_{r}\right)^2 \\
&= k^{-2n+2r}\left[\frac{1}{k^2}-1-i\right]\left(O^{(k)}_{r}\right)^2
\end{aligned}$$

□

ถัดมาเราจะพิสูจน์เอกลักษณ์ d'Ocagne ของลำดับ k-โอเรสเมเชิงซ้อน โดยใช้สูตรไบเนตของลำดับ k-โอเรสเมและสูตรไบเนตของลำดับ k- โอเรสเมเชิงซ้อน

ทฤษฎีบท 3.8. เอกลักษณ์ d'Ocagne ของลำดับ k-โอเรสเมเชิงซ้อน โดยที่ m และ n เป็นจำนวนเต็ม เมื่อ $k^2-4>0$ คือ

$$CO_{n+1}^{(k)}\cdot CO_m^{(k)}-CO_n^{(k)}\cdot CO_{m+1}^{(k)}=\left(\frac{1}{k}\right)^{2n+1}\tilde{\alpha}\tilde{\beta}O_{m-n}^{(k)}$$

พิสูจน์. โดยใช้สูตรของไบเนต จะได้ว่า
$CO_{n+1}^{(k)}\cdot CO_m^{(k)}-CO_n^{(k)}\cdot CO_{m+1}^{(k)}$

$$\begin{aligned}
&= \frac{1}{\sqrt{k^2-4}}\left(\alpha^{n+1}\tilde{\alpha}-\beta^{n+1}\tilde{\beta}\right)\frac{1}{\sqrt{k^2-4}}\left(\alpha^{m}\tilde{\alpha}-\beta^{m}\tilde{\beta}\right)\\
&\quad -\frac{1}{\sqrt{k^2-4}}\left(\alpha^{n}\tilde{\alpha}-\beta^{n}\tilde{\beta}\right)\frac{1}{\sqrt{k^2-4}}\left(\alpha^{m+1}\tilde{\alpha}-\beta^{m+1}\tilde{\beta}\right)\\
&= \frac{1}{k^2-4}\left[\left(\alpha^{n+1}\tilde{\alpha}-\beta^{n+1}\tilde{\beta}\right)\left(\alpha^{m}\tilde{\alpha}-\beta^{m}\tilde{\beta}\right)\right.\\
&\quad \left.-\left(\alpha^{n}\tilde{\alpha}-\beta^{n}\tilde{\beta}\right)\left(\alpha^{m+1}\tilde{\alpha}-\beta^{m+1}\tilde{\beta}\right)\right]\\
&= \frac{1}{k^2-4}\left[\left(\alpha^{n+1}\tilde{\alpha}\alpha^{m}\tilde{\alpha}-\alpha^{n+1}\tilde{\alpha}\beta^{m}\tilde{\beta}-\alpha^{m}\tilde{\alpha}\beta^{n+1}\tilde{\beta}+\beta^{n+1}\tilde{\beta}\beta^{m}\tilde{\beta}\right)\right.\\
&\quad \left.-\left(\alpha^{n}\tilde{\alpha}\alpha^{m+1}\tilde{\alpha}-\alpha^{n}\tilde{\alpha}\beta^{m+1}\tilde{\beta}-\alpha^{m+1}\tilde{\alpha}\beta^{n}\tilde{\beta}+\beta^{n}\tilde{\beta}\beta^{m+1}\tilde{\beta}\right)\right]\\
&= \frac{1}{k^2-4}\left[\alpha^{n+m+1}\tilde{\alpha}^2-\alpha^{n+1}\tilde{\alpha}\beta^{m}\tilde{\beta}-\alpha^{m}\tilde{\alpha}\beta^{n+1}\tilde{\beta}+\beta^{n+m+1}\tilde{\beta}^2-\right.\\
&\quad \left.\alpha^{n+m+1}\tilde{\alpha}^2+\alpha^{n}\tilde{\alpha}\beta^{m+1}\tilde{\beta}+\alpha^{m+1}\tilde{\alpha}\beta^{n}\tilde{\beta}-\beta^{n+m+1}\tilde{\beta}^2\right]\\
&= \frac{1}{k^2-4}\left[-\alpha^{n+1}\tilde{\alpha}\beta^{m}\tilde{\beta}-\alpha^{m}\tilde{\alpha}\beta^{n+1}\tilde{\beta}+\alpha^{n}\tilde{\alpha}\beta^{m+1}\tilde{\beta}+\alpha^{m+1}\tilde{\alpha}\beta^{n}\tilde{\beta}\right]\\
&= \frac{\tilde{\alpha}\tilde{\beta}}{k^2-4}\left[-\alpha^{n+1}\beta^{m}-\alpha^{m}\beta^{n+1}+\alpha^{n}\beta^{m+1}+\alpha^{m+1}\beta^{n}\right]\\
&= \frac{\tilde{\alpha}\tilde{\beta}}{k^2-4}\left[-\alpha^{n}\beta^{m}\left(\alpha-\beta\right)+\alpha^{m}\beta^{n}\left(\alpha-\beta\right)\right]\\
&= \frac{\tilde{\alpha}\tilde{\beta}}{k^2-4}\left[\left(\alpha-\beta\right)\left(\alpha^{m}\beta^{n}-\alpha^{n}\beta^{m}\right)\right]\\
&= \frac{\tilde{\alpha}\tilde{\beta}}{k^2-4}\left[\left(\alpha-\beta\right)\alpha^{n}\beta^{n}\left(\alpha^{m-n}-\beta^{m-n}\right)\right]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\tilde{\alpha}\tilde{\beta}}{k^2-4}\left[\left(\frac{\sqrt{k^2-4}}{k}\right)\left(\frac{1}{k}\right)^{2n}\left(\alpha^{m-n}-\beta^{m-n}\right)\right] \\
&= \frac{1}{\sqrt{k^2-4}}\left[\tilde{\alpha}\tilde{\beta}\left(\frac{\sqrt{k^2-4}}{k}\right)\left(\frac{1}{k}\right)^{2n}\right]\frac{1}{\sqrt{k^2-4}}\left(\alpha^{m-n}-\beta^{m-n}\right) \\
&= \tilde{\alpha}\tilde{\beta}\left(\frac{1}{k}\right)\left(\frac{1}{k}\right)^{2n}\left(\frac{\alpha^{m-n}-\beta^{m-n}}{\sqrt{k^2-4}}\right) \\
&= \tilde{\alpha}\tilde{\beta}\left(\frac{1}{k}\right)^{2n+1}\left(\frac{\alpha^{m-n}-\beta^{m-n}}{\sqrt{k^2-4}}\right)
\end{aligned}
$$

จากสูตรของไบเนตจะได้ว่า

$$
\begin{aligned}
CO_{m+1}^{(k)}\cdot CO_n^{(k)} - CO_m^{(k)}\cdot CO_{n+1}^{(k)} &= \tilde{\alpha}\tilde{\beta}\left(\frac{1}{k}\right)^{2n+1}O_{m-n}^{(k)} \\
&= \left(\frac{1}{k}\right)^{2n+1}\tilde{\alpha}\tilde{\beta}O_{m-n}^{(k)}
\end{aligned}
$$

□

ต่อมาเราจะพิสูจน์เอกลักษณ์ Honsberger ของลำดับ k-โอเรสเมเชิงซ้อน โดยใช้สูตรไบเนตของลำดับ k-โอเรสเมและสูตรไบเนตของลำดับ k-โอเรสเมเชิงซ้อน

ทฤษฎีบท 3.9. เอกลักษณ์ Honsberger ของลำดับ k-โอเรสเมเชิงซ้อน โดยที่ m และ n เป็นจำนวนเต็ม เมื่อ $k^2-4>0$ คือ

$$kO_n^{(k)}CO_{m+1}^{(k)} - \frac{1}{k}O_{n-1}^{(k)}CO_m^{(k)} = CO_{n+m}^{(k)}$$

พิสูจน์. โดยใช้สูตรของไบเนตในการพิสูจน์ เมื่อ $k^2\alpha^2 = k^2\alpha - 1$ และ $k^2\beta^2 = k^2\beta - 1$ จะได้ว่า $kO_n^{(k)}CO_{m+1}^{(k)} - \frac{1}{k}O_{n-1}^{(k)}CO_m^{(k)}$

$$
\begin{aligned}
&= k\left(\frac{\alpha^n-\beta^n}{\sqrt{k^2-4}}\right)\left(\frac{\alpha^{m+1}\tilde{\alpha}-\beta^{m+1}\tilde{\beta}}{\sqrt{k^2-4}}\right) - \frac{1}{k}\left(\frac{\alpha^{n-1}-\beta^{n-1}}{\sqrt{k^2-4}}\right)\left(\frac{\alpha^m\tilde{\alpha}-\beta^m\tilde{\beta}}{\sqrt{k^2-4}}\right) \\
&= \frac{1}{k^2-4}\left[k\left(\alpha^n-\beta^n\right)\left(\alpha^{m+1}\tilde{\alpha}-\beta^{m+1}\tilde{\beta}\right) - \frac{1}{k}\left(\alpha^{n-1}-\beta^{n-1}\right)\left(\alpha^m\tilde{\alpha}-\beta^m\tilde{\beta}\right)\right] \\
&= \frac{1}{k^2-4}\left[k\left(\alpha^n\alpha^{m+1}\tilde{\alpha}-\alpha^n\beta^{m+1}\tilde{\beta}-\alpha^{m+1}\tilde{\alpha}\beta^n+\beta^n\beta^{m+1}\tilde{\beta}\right)\right. \\
&\quad \left.-\frac{1}{k}\left(\alpha^{n-1}\alpha^m\tilde{\alpha}-\alpha^{n-1}\beta^m\tilde{\beta}-\alpha^m\tilde{\alpha}\beta^{n-1}+\beta^{n-1}\beta^m\tilde{\beta}\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{k^2-4}\left[k\left(\alpha^{n+m+1}\tilde{\alpha}\right)-\frac{1}{k}\left(\alpha^{n+m-1}\tilde{\alpha}\right)-k\left(\alpha^{n}\beta^{m+1}\tilde{\beta}\right)+\frac{1}{k}\left(\alpha^{n-1}\beta^{m}\tilde{\beta}\right)\right.\\
&\quad \left.-k\left(\alpha^{m+1}\tilde{\alpha}\beta^{n}\right)+\frac{1}{k}\left(\alpha^{m}\tilde{\alpha}\beta^{n-1}\right)+k\left(\beta^{n+m+1}\tilde{\beta}\right)-\frac{1}{k}\left(\beta^{n+m-1}\tilde{\beta}\right)\right]\\
&= \frac{1}{k^2-4}\left[\frac{1}{k}\left(\alpha^{n+m-1}\tilde{\alpha}\right)\left(k^2\alpha^2-1\right)+\frac{1}{k}\left(\alpha^{n-1}\beta^{m}\tilde{\beta}\right)\left(1-k^2\alpha\beta\right)\right.\\
&\quad \left.+\frac{1}{k}\left(\alpha^{m}\tilde{\alpha}\beta^{n-1}\right)\left(1-k^2\alpha\beta\right)+\frac{1}{k}\left(\beta^{n+m-1}\tilde{\beta}\right)\left(k^2\beta^2-1\right)\right]\\
&= \frac{1}{k^2-4}\left[\frac{1}{k}\left(\alpha^{n+m-1}\tilde{\alpha}\right)\left(k^2\alpha-1-1\right)+\frac{1}{k}\left(\alpha^{n-1}\beta^{m}\tilde{\beta}\right)\left(1-k^2\frac{1}{k^2}\right)\right.\\
&\quad \left.+\frac{1}{k}\left(\alpha^{m}\tilde{\alpha}\beta^{n-1}\right)\left(1-k^2\frac{1}{k^2}\right)+\frac{1}{k}\left(\beta^{n+m-1}\tilde{\beta}\right)\left(k^2\beta-1-1\right)\right]\\
&= \frac{1}{k^2-4}\left[\frac{1}{k}\left(\alpha^{n+m-1}\tilde{\alpha}\right)\left(k^2\alpha-2\right)+\frac{1}{k}\left(\beta^{n+m-1}\tilde{\beta}\right)\left(k^2\beta-2\right)\right]\\
&= \frac{1}{k^2-4}\left[\frac{1}{k}\left(\alpha^{n+m-1}\tilde{\alpha}\right)\left(k^2\left(\frac{k+\sqrt{k^2-4}}{2k}\right)-2\right)\right.\\
&\quad \left.+\frac{1}{k}\left(\beta^{n+m-1}\tilde{\beta}\right)\left(k^2\left(\frac{k-\sqrt{k^2-4}}{2k}\right)-2\right)\right]\\
&= \frac{1}{k^2-4}\left[\left(\alpha^{n+m-1}\tilde{\alpha}\right)\frac{1}{k}\left(\frac{k^2+k\sqrt{k^2-4}-4}{2}\right)\right.\\
&\quad \left.+\left(\beta^{n+m-1}\tilde{\beta}\right)\frac{1}{k}\left(\frac{k^2-k\sqrt{k^2-4}-4}{2}\right)\right]\\
&= \frac{1}{k^2-4}\left[\alpha^{n+m-1}\tilde{\alpha}\left(\frac{k^2+k\sqrt{k^2-4}-4}{2k}\right)+\beta^{n+m-1}\tilde{\beta}\left(\frac{k^2-k\sqrt{k^2-4}-4}{2k}\right)\right]\\
&= \frac{1}{k^2-4}\left[\alpha^{n+m-1}\tilde{\alpha}\left(\frac{k^2-4}{2k}+\frac{k\sqrt{k^2-4}}{2k}\right)+\beta^{n+m-1}\tilde{\beta}\left(\frac{k^2-4}{2k}-\frac{k\sqrt{k^2-4}}{2k}\right)\right]\\
&= \frac{1}{\sqrt{k^2-4}}\left[\alpha^{n+m-1}\tilde{\alpha}\left(\frac{\sqrt{k^2-4}}{2k}+\frac{k}{2k}\right)-\beta^{n+m-1}\tilde{\beta}\left(-\frac{\sqrt{k^2-4}}{2k}+\frac{k}{2k}\right)\right]\\
&= \frac{1}{\sqrt{k^2-4}}\left[\alpha^{n+m-1}\tilde{\alpha}\left(\frac{k+\sqrt{k^2-4}}{2k}\right)-\beta^{n+m-1}\tilde{\beta}\left(\frac{k-\sqrt{k^2-4}}{2k}\right)\right]\\
&= \frac{1}{\sqrt{k^2-4}}\left[\alpha^{n+m-1}\tilde{\alpha}(\alpha)-\beta^{n+m-1}\tilde{\beta}(\beta)\right]\\
&= \frac{1}{\sqrt{k^2-4}}\left[\alpha^{n+m}\tilde{\alpha}-\beta^{n+m}\tilde{\beta}\right]\\
&= CO_{n+m}^{(k)}
\end{aligned}
$$

□

เราจะพิสูจน์สูตรการหาผลรวมของ n พจน์ของลำดับ k-โอเรสเมเชิงซ้อน โดยใช้สูตรไบเนตของลำดับ k-โอเรสเมเชิงซ้อนและสูตรผลบวกจำกัดพจน์ของอนุกรมเรขาคณิต

ทฤษฎีบท 3.10. ผลรวมของ n พจน์สำหรับลำดับ k-โอเรสเมในรูปแบบเชิงซ้อน เมื่อ n เป็นจำนวนเต็ม และ $k^2-4>0$ คือ

$$\sum_{j=1}^{n} CO_j^{(k)} = k^2\Big(CO_1^{(k)} - CO_{n+2}^{(k)}\Big) - CO_0^{(k)}$$

พิสูจน์. จากสูตรของไบเนต จะได้ว่า

$$\begin{aligned}
&\sum_{j=1}^{n} CO_j^{(k)} \\
&= \sum_{j=1}^{n} \frac{1}{\sqrt{k^2-4}}\Big(\alpha^j\tilde{\alpha} - \beta^j\tilde{\beta}\Big) \\
&= \frac{1}{\sqrt{k^2-4}}\Big(\tilde{\alpha}\sum_{j=1}^{n}\alpha^j - \tilde{\beta}\sum_{j=1}^{n}\beta^j\Big) \\
&= \frac{1}{\sqrt{k^2-4}}\left[\tilde{\alpha}\left(\frac{\alpha\big(1-\alpha^n\big)}{1-\alpha}\right) - \tilde{\beta}\left(\frac{\beta\big(1-\beta^n\big)}{1-\beta}\right)\right] \\
&= \frac{1}{\sqrt{k^2-4}}\left[\tilde{\alpha}\left(\frac{\alpha-\alpha^{n+1}}{1-\alpha}\right) - \tilde{\beta}\left(\frac{\beta-\beta^{n+1}}{1-\beta}\right)\right] \\
&= \frac{1}{\sqrt{k^2-4}}\left[\frac{\tilde{\alpha}\big(\alpha-\alpha^{n+1}\big)\big(1-\beta\big) - \tilde{\beta}\big(\beta-\beta^{n+1}\big)\big(1-\alpha\big)}{\big(1-\alpha\big)\big(1-\beta\big)}\right] \\
&= \frac{1}{\sqrt{k^2-4}}\left[\frac{\alpha\tilde{\alpha} - \alpha\tilde{\alpha}\beta - \alpha^{n+1}\tilde{\alpha} + \alpha^{n+1}\tilde{\alpha}\beta - \big(\beta\tilde{\beta} - \alpha\beta\tilde{\beta} - \beta^{n+1}\tilde{\beta} + \alpha\beta^{n+1}\tilde{\beta}\big)}{\alpha - \alpha + \alpha\beta}\right] \\
&= \frac{1}{\sqrt{k^2-4}}\left[\frac{\alpha\tilde{\alpha} - \alpha\tilde{\alpha}\beta - \alpha^{n+1}\tilde{\alpha} + \alpha^{n+1}\tilde{\alpha}\beta - \beta\tilde{\beta} + \alpha\beta\tilde{\beta} + \beta^{n+1}\tilde{\beta} - \alpha\beta^{n+1}\tilde{\beta}}{\alpha\beta}\right] \\
&= \frac{1}{\sqrt{k^2-4}}\left[\frac{\alpha\tilde{\alpha} - \beta\tilde{\beta} - \alpha^{n+1}\tilde{\alpha} + \alpha^{n+1}\tilde{\alpha}\beta + \beta^{n+1}\tilde{\beta} - \alpha\beta^{n+1}\tilde{\beta}}{\alpha\beta}\right] \\
&\quad + \frac{1}{\sqrt{k^2-4}}\left[\frac{-\alpha\tilde{\alpha}\beta + \alpha\beta\tilde{\beta}}{\alpha\beta}\right] \\
&= \frac{1}{\sqrt{k^2-4}}\left(\frac{1}{\alpha\beta}\right)\Big[\alpha\tilde{\alpha} - \beta\tilde{\beta} - \alpha^{n+1}\tilde{\alpha}\Big(1-\beta\Big) + \beta^{n+1}\tilde{\beta}\Big(1-\alpha\Big)\Big] \\
&\quad + \frac{1}{\sqrt{k^2-4}}\left[\frac{\alpha\beta\Big(-\tilde{\alpha}+\tilde{\beta}\Big)}{\alpha\beta}\right] \\
&= \frac{k^2}{\sqrt{k^2-4}}\Big[\alpha\tilde{\alpha} - \beta\tilde{\beta} - \alpha^{n+1}\tilde{\alpha}\Big(\alpha\Big) + \beta^{n+1}\tilde{\beta}\Big(\beta\Big)\Big] - \frac{1}{\sqrt{k^2-4}}\Big(\tilde{\alpha}-\tilde{\beta}\Big) \\
&= k^2\left[\left(\frac{\alpha\tilde{\alpha}-\beta\tilde{\beta}}{\sqrt{k^2-4}}\right) - \left(\frac{\alpha^{n+2}\tilde{\alpha}-\beta^{n+2}\tilde{\beta}}{\sqrt{k^2-4}}\right)\right] - \frac{\tilde{\alpha}-\tilde{\beta}}{\sqrt{k^2-4}} \\
&= k^2\Big(CO_1^{(k)} - CO_{n+2}^{(k)}\Big) - CO_0^{(k)}
\end{aligned}$$

□

เราจะพิสูจน์สูตรการหาผลรวมไม่จำกัดพจน์ของลำดับ k-โอเรสเมเชิงซ้อน โดยใช้สูตรไบเนตของลำดับ k-โอเรสเมเชิงซ้อนและสูตรผลบวกไม่จำกัดพจน์ของอนุกรมเรขาคณิต

ทฤษฎีบท 3.11. ผลรวมไม่จำกัดพจน์สำหรับลำดับ k-โอเรสเมเชิงซ้อน เมื่อ $k^2-4>0$ คือ

$$\sum_{j=1}^{\infty} CO_j^{(k)} = kCO_2^{(k)}$$

พิสูจน์.

$$\begin{aligned}
\sum_{j=1}^{\infty} CO_j^{(k)} &= \sum_{j=1}^{\infty} \frac{1}{\sqrt{k^2-4}}\left(\alpha^j\tilde{\alpha}-\beta^j\tilde{\beta}\right) \\
&= \frac{1}{\sqrt{k^2-4}}\left(\tilde{\alpha}\sum_{j=1}^{\infty}\alpha^j-\tilde{\beta}\sum_{j=1}^{\infty}\beta^j\right) \\
&= \frac{1}{\sqrt{k^2-4}}\left[\tilde{\alpha}\left(\frac{\alpha}{1-\alpha}\right)-\tilde{\beta}\left(\frac{\beta}{1-\beta}\right)\right] \\
&= \frac{1}{\sqrt{k^2-4}}\left[\frac{\alpha\tilde{\alpha}(1-\beta)-\beta\tilde{\beta}(1-\alpha)}{1-\beta-\alpha+\alpha\beta}\right] \\
&= \frac{1}{\sqrt{k^2-4}}\left[\frac{\alpha\tilde{\alpha}(\alpha)-\beta\tilde{\beta}(\beta)}{\alpha\beta}\right] \\
&= \frac{1}{\alpha\beta}\left[\frac{\alpha^2\tilde{\alpha}-\beta^2\tilde{\beta}}{\sqrt{k^2-4}}\right] \\
&= kCO_2^{(k)}
\end{aligned}$$

□

ต่อมาเราจะพิสูจน์ผลรวมกำลังสองของ n พจน์ของลำดับ k-โอเรสเมเชิงซ้อน โดยใช้บทตั้ง 3.2 และสูตรผลบวกจำกัดพจน์ของอนุกรมเรขาคณิต

ทฤษฎีบท 3.12. ผลรวมกำลังสองของลำดับ k-โอเรสเมเชิงซ้อน เมื่อ n เป็นจำนวนเต็ม คือ

$$\begin{aligned}
\sum_{j=1}^{n}\left(CO_j^{(k)}\right)^2 &= \frac{1}{2k^2-1}\Bigg[-k^4\left[\left(CO_1^{(k)}\right)^2+\left(CO_{n+1}^{(k)}\right)^2\right]+\left(CO_0^{(k)}\right)^2-\left(CO_n^{(k)}\right)^2 \\
&\quad +2k^2\left[\left(\frac{1}{k}\right)^3-\left(\frac{1}{k}\right)-i\left(\frac{1}{k}\right)\right]\left(\frac{k^{2n+1}-k}{k^{2n+2}-k^{2n}}\right)\Bigg]
\end{aligned}$$

พิสูจน์. การพิสูจน์ เราจะให้ T แทนผลรวมกำลังสอง ดังนี้

$$\left(CO_1^{(k)}\right)^2+\left(CO_2^{(k)}\right)^2+\left(CO_3^{(k)}\right)^2+\ldots+\left(CO_n^{(k)}\right)^2=\sum_{j=1}^{n}\left(CO_j^{(k)}\right)^2=T$$

จะได้ว่า

$$\begin{aligned}T &= \sum_{j=1}^{n}\left[CO_{j+1}^{(k)}+\frac{1}{k^2}CO_{j-1}^{(k)}\right]^2\\ &= \sum_{j=1}^{n}\left[\frac{k^2CO_{j+1}^{(k)}+CO_{j-1}^{(k)}}{k^2}\right]^2\end{aligned}$$

จะได้

$$\begin{aligned}k^4T &= \sum_{j=1}^{n}\left[k^2CO_{j+1}^{(k)}+CO_{j-1}^{(k)}\right]^2\\ &= \sum_{j=1}^{n}\left[k^4\left(CO_{j+1}^{(k)}\right)^2+2k^2CO_{j+1}^{(k)}CO_{j-1}^{(k)}+\left(CO_{j-1}^{(k)}\right)^2\right]\\ &= k^4\sum_{j=1}^{n}\left(CO_{j+1}^{(k)}\right)^2+2k^2\sum_{j=1}^{n}CO_{j+1}^{(k)}CO_{j-1}^{(k)}+\sum_{j=1}^{n}\left(CO_{j-1}^{(k)}\right)^2\end{aligned}$$

จากทฤษฎีบท 3.5 จะได้ว่า

$$\begin{aligned}k^4T &= k^4\left[\sum_{m=1}^{n}\left(CO_m^{(k)}\right)^2-\left(CO_1^{(k)}\right)^2+\left(CO_{n+1}^{(k)}\right)^2\right]\\ &\quad +2k^2\sum_{m=1}^{n}\left[\left(CO_m^{(k)}\right)^2+\left(\frac{1}{k}\right)^{2m-1}\left[\left(\frac{1}{k}\right)^3-\left(\frac{1}{k}\right)-i\left(\frac{1}{k}\right)\right]\right]\\ &\quad +\left[\sum_{m=1}^{n}\left(CO_m^{(k)}\right)^2+\left(CO_0^{(k)}\right)^2-\left(CO_n^{(k)}\right)^2\right]\\ &= k^4\left[T-\left(CO_1^{(k)}\right)^2+\left(CO_{n+1}^{(k)}\right)^2\right]+2k^2T\\ &\quad +2k^2\sum_{m=1}^{n}\left[\left(\frac{1}{k}\right)^{2m-1}\left[\left(\frac{1}{k}\right)^3-\left(\frac{1}{k}\right)-i\left(\frac{1}{k}\right)\right]\right]\\ &\quad +T+\left(CO_0^{(k)}\right)^2-\left(CO_n^{(k)}\right)^2\\ &= k^4T-k^4\left[\left(CO_1^{(k)}\right)^2+\left(CO_{n+1}^{(k)}\right)^2\right]+2k^2T\\ &\quad +2k^2\left[\left(\frac{1}{k}\right)^3-\left(\frac{1}{k}\right)-i\left(\frac{1}{k}\right)\right]\sum_{m=1}^{n}\left(\frac{1}{k}\right)^{2m-1}\\ &\quad +T+\left(CO_0^{(k)}\right)^2-\left(CO_n^{(k)}\right)^2\end{aligned}$$

$$\begin{aligned} &= T\left[k^4+2k^2+1\right]-k^4\left[\left(CO_1^{(k)}\right)^2+\left(CO_{n+1}^{(k)}\right)^2\right] \\ &\quad +2k^2\left[\left(\frac{1}{k}\right)^3-\left(\frac{1}{k}\right)-i\left(\frac{1}{k}\right)\right]\sum_{m=1}^{n}\left(\frac{1}{k}\right)^{2m-1} \\ &\quad +\left(CO_0^{(k)}\right)^2-\left(CO_n^{(k)}\right)^2 \end{aligned}$$

จะได้

$$\begin{aligned} k^4T-T\left[k^4+2k^2+1\right] &= -k^4\left[\left(CO_1^{(k)}\right)^2+\left(CO_{n+1}^{(k)}\right)^2\right]+\left(CO_0^{(k)}\right)^2-\left(CO_n^{(k)}\right)^2 \\ &\quad +2k^2\left[\left(\frac{1}{k}\right)^3-\left(\frac{1}{k}\right)-i\left(\frac{1}{k}\right)\right]\sum_{m=1}^{n}\left(\frac{1}{k}\right)^{2m-1} \end{aligned}$$

จากการหาอนุกรมเรขาคณิต จะได้ว่า

$$\begin{aligned} T\left[2k^2+1\right] &= -k^4\left[\left(CO_1^{(k)}\right)^2+\left(CO_{n+1}^{(k)}\right)^2\right]+\left(CO_0^{(k)}\right)^2-\left(CO_n^{(k)}\right)^2 \\ &\quad +2k^2\left[\left(\frac{1}{k}\right)^3-\left(\frac{1}{k}\right)-i\left(\frac{1}{k}\right)\right]\frac{\left(\frac{1}{k}\right)-\left(\frac{1}{k}\right)^{2n+1}}{1-\left(\frac{1}{k}\right)^2} \\ &= -k^4\left[\left(CO_1^{(k)}\right)^2+\left(CO_{n+1}^{(k)}\right)^2\right]+\left(CO_0^{(k)}\right)^2-\left(CO_n^{(k)}\right)^2 \\ &\quad +2k^2\left[\left(\frac{1}{k}\right)^3-\left(\frac{1}{k}\right)-i\left(\frac{1}{k}\right)\right]\frac{\left(\frac{k^{2n}-1}{k^{2n+1}}\right)}{\left(\frac{k^2-1}{k^2}\right)} \\ &= -k^4\left[\left(CO_1^{(k)}\right)^2+\left(CO_{n+1}^{(k)}\right)^2\right]+\left(CO_0^{(k)}\right)^2-\left(CO_n^{(k)}\right)^2 \\ &\quad +2k^2\left[\left(\frac{1}{k}\right)^3-\left(\frac{1}{k}\right)-i\left(\frac{1}{k}\right)\right]\left(\frac{k^{2n+1}-k}{k^{2n+2}-k^{2n}}\right) \end{aligned}$$

จะได้

$$\begin{aligned} T &= \frac{1}{2k^2-1}\Bigg[-k^4\left[\left(CO_1^{(k)}\right)^2+\left(CO_{n+1}^{(k)}\right)^2\right]+\left(CO_0^{(k)}\right)^2-\left(CO_n^{(k)}\right)^2 \\ &\quad +2k^2\left[\left(\frac{1}{k}\right)^3-\left(\frac{1}{k}\right)-i\left(\frac{1}{k}\right)\right]\left(\frac{k^{2n+1}-k}{k^{2n+2}-k^{2n}}\right)\Bigg] \end{aligned}$$

ดังนั้น

$$\begin{aligned} \sum_{j=1}^{n}\left(CO_j^{(k)}\right)^2 &= \frac{1}{2k^2-1}\Bigg[-k^4\left[\left(CO_1^{(k)}\right)^2+\left(CO_{n+1}^{(k)}\right)^2\right]+\left(CO_0^{(k)}\right)^2-\left(CO_n^{(k)}\right)^2 \\ &\quad +2k^2\left[\left(\frac{1}{k}\right)^3-\left(\frac{1}{k}\right)-i\left(\frac{1}{k}\right)\right]\left(\frac{k^{2n+1}-k}{k^{2n+2}-k^{2n}}\right)\Bigg] \end{aligned}$$

□

ต่อมาเราจะพิสูจน์สูตรการหาผลรวมของ n พจน์ของลำดับ k-โอเรสเมเชิงซ้อน กรณีที่ดัชนีเป็นจำนวนเต็มคู่ โดยใช้สูตรไบเนตของลำดับ k-โอเรสเมเชิงซ้อนและสูตรผลบวกจำกัดพจน์ของอนุกรมเรขาคณิต

ทฤษฎีบท 3.13. การหาผลรวมของ n พจน์ของลำดับ k-โอเรสเมเชิงซ้อน กรณีที่ดัชนีเป็นจำนวนเต็มคู่ เมื่อ n เป็นจำนวนเต็ม และ $k^2-4>0$ คือ

$$\sum_{j=1}^{n} CO_{2j}^{(k)} = \frac{k^2}{2k^2+1}\left[k^2\left(CO_1^{(k)} - CO_{2n-1}^{(k)}\right) + CO_0^{(k)} - CO_{2n-2}^{(k)}\right] + CO_{2n}^{(k)} - CO_0^{(k)}$$

พิสูจน์. สูตรที่ใช้ในการหาผลรวมกรณีที่ดัชนีเป็นจำนวนเต็มคู่ คือ

$$\sum_{j=1}^{n} CO_{2j}^{(k)} = \sum_{j=0}^{n-1} CO_{2j}^{(k)} + CO_{2n}^{(k)} - CO_0^{(k)}$$

$$\begin{aligned}
\sum_{j=0}^{n-1} CO_{2j}^{(k)} &= \frac{1}{\sqrt{k^2-4}}\left[\tilde{\alpha}\sum_{j=0}^{n-1}\alpha^{2j} - \tilde{\beta}\sum_{j=0}^{n-1}\beta^{2j}\right] \\
&= \frac{1}{\sqrt{k^2-4}}\left[\tilde{\alpha}\left(\frac{1-\alpha^{2n-2}}{1-\alpha^2}\right) - \tilde{\beta}\left(\frac{1-\beta^{2n-2}}{1-\beta^2}\right)\right] \\
&= \frac{1}{\sqrt{k^2-4}}\left[\frac{\tilde{\alpha}+\alpha^{2n-2}\tilde{\alpha}}{(1-\alpha)(1+\alpha)} - \frac{\tilde{\beta}-\beta^{2n-2}\tilde{\beta}}{(1-\beta)(1+\beta)}\right] \\
&= \frac{1}{\sqrt{k^2-4}}\left[\frac{\tilde{\alpha}+\alpha^{2n-2}\tilde{\alpha}}{\beta(1+\alpha)} - \frac{\tilde{\beta}-\beta^{2n-2}\tilde{\beta}}{\alpha(1+\beta)}\right] \\
&= \frac{1}{\sqrt{k^2-4}}\left[\frac{(\tilde{\alpha}+\alpha^{2n-2}\tilde{\alpha})(\alpha+\alpha\beta) - (\tilde{\beta}-\beta^{2n-2}\tilde{\beta})(\beta+\alpha\beta)}{\alpha\beta+\alpha\beta^2+\alpha^2\beta+(\alpha\beta)^2}\right] \\
&= \frac{1}{\sqrt{k^2-4}}\left[\frac{\alpha\tilde{\alpha}+\alpha\tilde{\alpha}\beta-\alpha^{2n-1}\tilde{\alpha}-\alpha^{2n-1}\tilde{\alpha}\beta-\beta\tilde{\beta}-\alpha\beta\tilde{\beta}+\beta^{2n-1}\tilde{\beta}+\alpha\beta^{2n-1}\tilde{\beta}}{\alpha\beta(1+\beta+\alpha+\alpha\beta)}\right] \\
&= \frac{1}{\sqrt{k^2-4}}\left[\left(\frac{\alpha\tilde{\alpha}-\beta\tilde{\beta}}{\alpha\beta(2+\alpha\beta)}\right) - \left(\frac{\alpha^{2n-1}\tilde{\alpha}-\beta^{2n-1}\tilde{\beta}}{\alpha\beta(2+\alpha\beta)}\right) + \left(\frac{\alpha\tilde{\alpha}\beta-\alpha\beta\tilde{\beta}}{\alpha\beta(2+\alpha\beta)}\right)\right. \\
&\quad \left. -\left(\frac{\alpha^{2n-1}\tilde{\alpha}\beta-\alpha\beta^{2n-1}\tilde{\beta}}{\alpha\beta(2+\alpha\beta)}\right)\right] \\
&= \frac{1}{\alpha\beta(2+\alpha\beta)}\left[\frac{\alpha\tilde{\alpha}-\beta\tilde{\beta}}{\sqrt{k^2-4}} - \frac{\alpha^{2n-1}\tilde{\alpha}-\beta^{2n-1}\tilde{\beta}}{\sqrt{k^2-4}}\right] \\
&\quad + \frac{1}{\alpha\beta(2+\alpha\beta)}\left[\frac{\alpha\tilde{\alpha}\beta-\alpha\beta\tilde{\beta}}{\sqrt{k^2-4}} - \frac{\alpha^{2n-1}\tilde{\alpha}\beta-\alpha\beta^{2n-1}\tilde{\beta}}{\sqrt{k^2-4}}\right] \\
&= \frac{1}{\alpha\beta(2+\alpha\beta)}\left[\frac{\alpha\tilde{\alpha}-\beta\tilde{\beta}}{\sqrt{k^2-4}} - \frac{\alpha^{2n-1}\tilde{\alpha}-\beta^{2n-1}\tilde{\beta}}{\sqrt{k^2-4}}\right] \\
&\quad + \frac{1}{\alpha\beta(2+\alpha\beta)}(\alpha\beta)\left[\frac{\tilde{\alpha}-\tilde{\beta}}{\sqrt{k^2-4}} - \frac{\alpha^{2n-2}\tilde{\alpha}-\beta^{2n-2}\tilde{\beta}}{\sqrt{k^2-4}}\right]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\alpha\beta\big(2+\alpha\beta\big)}\left[\frac{\alpha\tilde{\alpha}-\beta\tilde{\beta}}{\sqrt{k^2-4}}-\frac{\alpha^{2n-1}\tilde{\alpha}-\beta^{2n-1}\tilde{\beta}}{\sqrt{k^2-4}}\right]\\
&\quad+\frac{1}{2+\alpha\beta}\left[\frac{\tilde{\alpha}-\tilde{\beta}}{\sqrt{k^2-4}}-\frac{\alpha^{2n-2}\tilde{\alpha}-\beta^{2n-2}\tilde{\beta}}{\sqrt{k^2-4}}\right]\\
&= \frac{1}{\alpha\beta\big(2+\alpha\beta\big)}\left[CO_1^{(k)}-CO_{2n-1}^{(k)}\right]+\frac{1}{\big(2+\alpha\beta\big)}\left[CO_0^{(k)}-CO_{2n-2}^{(k)}\right]\\
&= \frac{1}{2+\alpha\beta}\left[\frac{1}{\alpha\beta}\Big(CO_1^{(k)}-CO_{2n-1}^{(k)}\Big)+CO_0^{(k)}-CO_{2n-2}^{(k)}\right]\\
&= \frac{1}{\frac{2k^2+1}{k^2}}\left[k^2\Big(CO_1^{(k)}-CO_{2n-1}^{(k)}\Big)+CO_0^{(k)}-CO_{2n-2}^{(k)}\right]\\
&= \frac{k^2}{2k^2+1}\left[k^2\Big(CO_1^{(k)}-CO_{2n-1}^{(k)}\Big)+CO_0^{(k)}-CO_{2n-2}^{(k)}\right]
\end{aligned}
$$

ดังนั้น

$$\sum_{j=1}^{n}CO_{2j}^{(k)}=\frac{k^2}{2k^2+1}\left[k^2\Big(CO_1^{(k)}-CO_{2n-1}^{(k)}\Big)+CO_0^{(k)}-CO_{2n-2}^{(k)}\right]+CO_{2n}^{(k)}-CO_0^{(k)}$$

□

ต่อมาเราจะพิสูจน์สูตรการหาผลรวมไม่จำกัดพจน์ของลำดับ k-โอเรสเมเชิงซ้อน กรณีที่ดัชนีเป็นจำนวนเต็มคู่ โดยใช้สูตรไบเนตของลำดับ k-โอเรสเมเชิงซ้อนและสูตรผลบวกไม่จำกัดพจน์ของอนุกรมเรขาคณิต

ทฤษฎีบท 3.14. การหาผลรวมไม่จำกัดพจน์ของลำดับ k-โอเรสเมเชิงซ้อน กรณีที่ดัชนีเป็นจำนวนเต็มคู่เมื่อ $k^2-4>0$ คือ

$$\sum_{j=1}^{\infty}CO_{2j}^{(k)}=\frac{k^2}{2k^2+1}\left[k^2CO_3^{(k)}+CO_2^{(k)}\right]$$

พิสูจน์. จากสูตรของไบเนต จะได้ว่า

$$
\begin{aligned}
\sum_{j=1}^{\infty}CO_{2j}^{(k)} &= \frac{1}{\sqrt{k^2-4}}\left[\tilde{\alpha}\sum_{j=1}^{\infty}\alpha^{2j}-\tilde{\beta}\sum_{j=1}^{\infty}\beta^{2j}\right]\\
&= \frac{1}{\sqrt{k^2-4}}\left[\tilde{\alpha}\left(\frac{\alpha^2}{1-\alpha^2}\right)-\tilde{\beta}\left(\frac{\beta^2}{1-\beta^2}\right)\right]\\
&= \frac{1}{\sqrt{k^2-4}}\left[\frac{\alpha^2\tilde{\alpha}}{(1-\alpha)(1+\alpha)}-\frac{\beta^2\tilde{\beta}}{(1-\beta)(1+\beta)}\right]\\
&= \frac{1}{\sqrt{k^2-4}}\left[\frac{\alpha^2\tilde{\alpha}}{\beta(1+\alpha)}-\frac{\beta^2\tilde{\beta}}{\alpha(1+\beta)}\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{k^2-4}}\left[\frac{\alpha^2\tilde{\alpha}}{\beta+\alpha\beta}-\frac{\beta^2\tilde{\beta}}{\alpha+\alpha\beta}\right]\\
&= \frac{1}{\sqrt{k^2-4}}\left[\frac{\alpha^2\tilde{\alpha}(\alpha+\alpha\beta)-\beta^2\tilde{\beta}(\beta+\alpha\beta)}{\alpha\beta+\alpha\beta^2+\alpha^2\beta+(\alpha\beta)^2}\right]\\
&= \frac{1}{\sqrt{k^2-4}}\left[\frac{\alpha^3\tilde{\alpha}+\alpha^3\tilde{\alpha}\beta-\beta^3\tilde{\beta}-\alpha\beta^3\tilde{\beta}}{\alpha\beta(1+\beta+\alpha+\alpha\beta)}\right]\\
&= \frac{1}{\sqrt{k^2-4}}\left[\frac{\alpha^3\tilde{\alpha}-\beta^3\tilde{\beta}+\alpha^3\tilde{\alpha}\beta-\alpha\beta^3\tilde{\beta}}{\alpha\beta(2+\alpha\beta)}\right]\\
&= \frac{1}{\alpha\beta(2+\alpha\beta)}\left[\frac{\alpha^3\tilde{\alpha}-\beta^3\tilde{\beta}}{\sqrt{k^2-4}}\right]+\frac{1}{\alpha\beta(2+\alpha\beta)}\left[\frac{\alpha^3\tilde{\alpha}\beta-\alpha\beta^3\tilde{\beta}}{\sqrt{k^2-4}}\right]\\
&= \frac{1}{\alpha\beta(2+\alpha\beta)}\left[\frac{\alpha^3\tilde{\alpha}-\beta^3\tilde{\beta}}{\sqrt{k^2-4}}\right]+\frac{1}{\alpha\beta(2+\alpha\beta)}\alpha\beta\left[\frac{\alpha^2\tilde{\alpha}-\beta^2\tilde{\beta}}{\sqrt{k^2-4}}\right]\\
&= \frac{1}{\alpha\beta(2+\alpha\beta)}\left[\frac{\alpha^3\tilde{\alpha}-\beta^3\tilde{\beta}}{\sqrt{k^2-4}}\right]+\frac{1}{2+\alpha\beta}\left[\frac{\alpha^2\tilde{\alpha}-\beta^2\tilde{\beta}}{\sqrt{k^2-4}}\right]\\
&= \frac{1}{2+\alpha\beta}\left[\frac{1}{\alpha\beta}CO_3^{(k)}+CO_2^{(k)}\right]\\
&= \frac{1}{2+\alpha\beta}\left[k^2CO_3^{(k)}+CO_2^{(k)}\right]\\
&= \frac{k^2}{2k^2+1}\left[k^2CO_3^{(k)}+CO_2^{(k)}\right]
\end{aligned}
$$

□

ต่อมาเราจะพิสูจน์สูตรการหาผลรวมของ n พจน์ของลำดับ k-โอเรสเมเชิงซ้อน กรณีที่ดัชนีเป็นจำนวนเต็มคี่ โดยใช้สูตรไบเนตของลำดับ k-โอเรสเมเชิงซ้อนและสูตรผลบวกจำกัดพจน์ของอนุกรมเรขาคณิต

ทฤษฎีบท 3.15. การหาผลรวมของ n พจน์ของลำดับ k-โอเรสเมเชิงซ้อน กรณีที่ดัชนีป็นจำนวนเต็มคี่ เมื่อ n เป็นจำนวนเต็ม และ $k^2-4>0$ คือ

$$\sum_{j=1}^{n} CO_{2j+1}^{(k)} = \frac{1}{2k^2+1}\left[k^4\left(CO_3^{(k)}-CO_{2n+3}^{(k)}\right)+CO_1^{(k)}-CO_{2n+1}^{(k)}\right]$$

พิสูจน์.

$$
\begin{aligned}
\sum_{j=1}^{n} CO_{2j+1}^{(k)} &= \frac{1}{\sqrt{k^2-4}}\left[\tilde{\alpha}\sum_{j=1}^{n}\alpha^{2j+1}-\tilde{\beta}\sum_{j=1}^{n}\beta^{2j+1}\right]\\
&= \frac{1}{\sqrt{k^2-4}}\left[\tilde{\alpha}\left(\alpha^3\frac{(1-\alpha^{2n})}{1-\alpha^2}\right)-\tilde{\beta}\left(\beta^3\frac{(1-\beta^{2n})}{1-\beta^2}\right)\right]\\
&= \frac{1}{\sqrt{k^2-4}}\left[\frac{\left(\alpha^3\tilde{\alpha}-\alpha^{2n+3}\tilde{\alpha}\right)\left(1-\beta^2\right)-\left(\beta^3\tilde{\beta}-\beta^{2n+3}\tilde{\beta}\right)\left(1-\alpha^2\right)}{1-\alpha^2-\beta^2+\left(\alpha\beta\right)^2}\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{k^2-4}}\left[\frac{\alpha^3\tilde{\alpha}-\alpha^3\tilde{\alpha}\beta^2-\alpha^{2n+3}\tilde{\alpha}+\alpha^{2n+3}\tilde{\alpha}\beta^2-\beta^3\tilde{\beta}-\alpha^2\beta^3\tilde{\beta}+\beta^{2n+3}\tilde{\beta}-\alpha^2\beta^{2n+3}\tilde{\beta}}{\left(2k^2+1\right)\left(\alpha\beta\right)^2}\right] \\
&= \frac{1}{\left(2k^2+1\right)\left(\alpha\beta\right)^2}\left[\frac{\alpha^3\tilde{\alpha}-\beta^3\tilde{\beta}}{\sqrt{k^2-4}}-\frac{\alpha^{2n+3}\tilde{\alpha}-\beta^{2n+3}\tilde{\beta}}{\sqrt{k^2-4}}\right] \\
&\quad -\frac{1}{\sqrt{k^2-4}}\left[\frac{\left(\alpha\beta\right)^2\left(\alpha\tilde{\alpha}-\beta\tilde{\beta}-\alpha^{2n+1}\tilde{\alpha}+\beta^{2n+1}\tilde{\beta}\right)}{\left(2k^2+1\right)\left(\alpha\beta\right)^2}\right] \\
&= \frac{1}{\left(2k^2+1\right)\left(\alpha\beta\right)^2}\left[\frac{\alpha^3\tilde{\alpha}-\beta^3\tilde{\beta}}{\sqrt{k^2-4}}-\frac{\alpha^{2n+3}\tilde{\alpha}-\beta^{2n+3}\tilde{\beta}}{\sqrt{k^2-4}}\right] \\
&\quad -\frac{1}{2k^2+1}\left[\frac{\alpha\tilde{\alpha}-\beta\tilde{\beta}}{\sqrt{k^2-4}}-\frac{\alpha^{2n+1}\tilde{\alpha}-\beta^{2n+1}\tilde{\beta}}{\sqrt{k^2-4}}\right] \\
&= \frac{1}{2k^2+1}\left[k^4\left(CO_3^{(k)}-CO_{2n+3}^{(k)}\right)-CO_1^{(k)}+CO_{2n+1}^{(k)}\right]
\end{aligned}
$$

□

4 สรุปผลและข้อเสนอแนะ

จากการศึกษาลำดับ k-โอเรสเมในรูปแบบเชิงซ้อน ผู้วิจัยได้พิสูจน์สมบัติของลำดับ k-โอเรสเมในรูปแบบเชิงซ้อนตามที่เคยปรากฏในรูปแบบของงลำดับ k-โอเรสเมทั่วไปในงานวิจัยของ Soykan [7] นั่นคือได้สร้างฟังก์ชันก่อกำเนิดของลำดับ k- โอเรสเมเชิงซ้อน พิสูจน์สูตรไบเนตของลำดับ k-โอเรสเมเชิงซ้อน เอกลักษณ์ Cassini ของลำดับ k-โอเรสเมเชิงซ้อน เอกลักษณ์ Catalan ของลำดับ k- โอเรสเมเชิงซ้อน และ เอกลักษณ์ d'Ocagne ของลำดับ k-โอเรสเมเชิงซ้อน นอกจากนี้ผู้วิจัยได้พิสูจน์สมบัติเพิ่มเติมของลำดับ k-โอเรสเมในรูปแบบเชิงซ้อน ได้แก่ เอกลักษณ์ Honsherger ของลำดับ k-โอเรสเมเชิงซ้อน สูตรการหาผลรวมของลำดับ k-โอเรสเมเชิงซ้อน สูตรการหาผลรวมของลำดับ k-โอเรสเมเชิงซ้อนยกกำลังสอง สูตรการหาผลรวมพจน์คี่ของลำดับ k-โอเรสเมเชิงซ้อน และ สูตรการหาผลรวมพจน์คู่ของลำดับ k-โอเรสเมเชิงซ้อน ซึ่งผู้วิจัยเสนอแนะให้นำแนวคิดในการสร้างลำดับในรูปแบบเชิงซ้อนจากลำดับบ k-โอเรสเม ไปสร้างและพิสูจน์สมบัติของลำดับในรูปแบบเชิงซ้อนที่เกิดลำดับอื่น ๆต่อไป

กิตติกรรมประกาศ ผู้วิจัยขอขอบคุณผู้ทรงคุณวุฒิทุกท่านที่ได้ให้ข้อคิดเห็นและข้อเสนอแนะต่าง ๆ เพื่อปรับปรุงบทความวิจัยนี้ และขอบคุณภาควิชาคณิตศาสตร์ และคณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา ที่ให้ทุนวิจัยและนำเสนอผลงานนี้

เอกสารอ้างอิง

[1] M. Clagett, *Nicole Oresme and the Medieval Geometry of Qualities and Motions: A Treatise on the Uniformity and Difformity of Intensities Known as Tractatus de confurationibus qualitatum et motuum*, The University of Wisconsin Press, Wisconsin, 1968.

[2] M. Clagett, "Oresme, Nicole", in: C.C. Gillespie (ed.), Dictionary of Scientific Biography, Vol. 9, Charles Scribner's Sons, New York, 1981, pp. 223-230.

[3] M.C.S. Mangueira, R.P.M. Vieira, F.R.V. Alves and P.M.M.C. Catarino, *The Oresme sequence: The generalization of its matrix form and its hybridization process*, Notes on Number Theory and Discrete Mathematics, **27**(1) (2021), 101–111. https://doi.org/10.7546/nntdm.2021.27.1.101-111

[4] A.F. Horadam, *Oresme Numbers*, The Fibonacci Quarterly, **12**(3) (1974), 267-–271.

[5] G. Cerda-Morales, *Oresme polynomials and their derivatives*, Cornell University. (2019), https://arxiv.org/pdf/1904.01165.pdf

[6] C.J. Harman, *Complex Fibonacci number*, The Fibonacci Quarterly. **19**(1) (1981), 82–86.

[7] Y. Soykan, *A study on generalized p-Oresme numbers*, Asian Journal of Advanced Research and Reports. **15**(7) (2021), 1–25.

11. Other Related Topics in Mathematics

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

System of Stochastic Grey Differential Equations with Singular Spectrum Analysis for Precious Metal Prices Forecasting

Rammarat Panadsako[1,†,‡] and Raywat Tanadkithirun[1]

[1]Department of Mathematics and Computer Science, Faculty of Science
Chulalongkorn University, Bangkok 10330, Thailand

Abstract

The precious metals are valuable assets; therefore, price forecasting is one of the interesting tasks that can be conducted by several methods. In this work, the stochastic grey differential equation with singular spectrum analysis (SGDE+SSA) model was developed to forecast monthly prices of gold, silver, platinum, and palladium. Firstly, the one-dimensional SGDE+SSA model was constructed to forecast prices without consideration of their price correlations. However, these prices are supposed to have some relations, so the multidimensional SGDE+SSA model (MSGDE+SSA) was created by considering the historical correlations of those four metals to model their sources of randomness in the diffusion part of the system of stochastic differential equations. For the sensitivity analysis, the approach of parameter selection was developed to improve the model proficiency. Additionally, the expectation and variance of the models were studied. The accuracy of SGDE+SSA and MSGDE+SSA models was compared with historical prices from January 2005 to January 2024 by using the mean absolute percentage error (MAPE). For SGDE+SSA model by plot of logarithm, the MAPEs of predicted prices for gold, silver, platinum and palladium are 2.8485%, 3.9569%, 3.2240% and 4.8571%, respectively, while the MAPEs of forecasted prices are 5.7642%, 11.0591%, 8.5403% and 44.1193%, respectively. For MSGDE+SSA model by plot of logarithm, the MAPEs of predicted prices for gold, silver, platinum and palladium are 2.8485%, 3.5729%, 2.3522% and 3.4241%, respectively, while the MAPEs of forecasted prices are 5.7642%, 11.7033%, 6.2948% and 47.8827%, respectively. This study found that the MSGDE+SSA model has more efficiency in prediction than the SGDE+SSA model. Taken together, the MSGDE+SSA model is highly efficient in predicting gold, silver and platinum prices and might be a useful tool for other metals.

Keywords: precious metal prices, stochastic differential equation, grey model, stochastic grey differential equation, singular spectrum analysis.

2020 MSC: Primary 60H10; Secondary 60H35, 60J65, 65C30.

†Speaker. ‡Corresponding author.
Email: rammarat.panadsa@gmail.com (R. Panadsako), raywat.t@chula.ac.th (R. Tanadkithirun)

1 Introduction

Precious metals are rare elements and have high economic values as they are used in many fields such as jewelry, aerospace, medicine, and electronics. In particular, gold, silver, platinum, and palladium are essential elements in the development of industries and technology. Therefore, they are investable precious metals and investors can purchase physical metals or metal stocks. Consequently, investors are interested in forecasting these prices for planning the investment.

Several models are proposed to predict metal prices. One of the most famous models is the stochastic differential equation (SDE). This model is appropriate to describe highly fluctuated occurrences. In 2011, Issaranusorn et al. developed an SDE model to predict the gold price by considering seasonality. The gold prices are assumed to follow an extended geometric brownian motion with a time-varying drift which describes seasonal variation in gold prices. This work obtained the prices of gold futures and European gold options prices that depend on the seasonality in gold prices. Moreover, The researcher recommends the model can predict the gold prices in the future with appropriately estimated parameters. In 2019, Alipour et al. studied autoregressive integrated moving average (ARIMA), threshold generalized autoregressive conditional heteroskedastic (TGARCH), and Black Sholes Merton SDE for monthly copper price. The history data of copper prices from early 1987 to September 2017 was separated into a training group and a test group. The training group and the test group have periods from early 1987 to 2014 and 2015 to September 2017, respectively. The estimated parameters of the ARIMA, TGARCH, and SDE models were estimated by using a training group. The summary of the results of prediction models for time series of copper prices from EViews software shows that ARIMA(2,1,3) and TGARCH(1,1) are the lowest Durbin-Watson measure, Akaike criterion, and Shwarz criterion. The result of the comparison of ARIMA(2,1,3), TGARCH(1,1), and SDE model shows that the SDE model achieved the highest predictive power for MAPE. However, the prediction of the price problem can be solved by several techniques. The grey system theory which is a technique since the 1980s can be used to develop a grey model (GM). The grey model is effective with small sample sizes and short-term forecasts but the GM is inappropriate for highly fluctuated situations. In 2020, Gligorić et al. merged the SDE and GM(1,1) to solve this copper price problem; the resulting model was called the stochastic grey differential equation (SGDE). Additionally, the singular spectrum analysis (SSA) technique can be applied to an SGDE to create a new model called the SGDE+SSA model. The results showed that SGDE+SSA has better efficiency than ARIMA(2,1,3), TGARCH(1,1), SDE and SGDE alone.

In this work, the one-dimensional SGDE+SSA and multidimensional SGDE+SSA models were developed by using the historical price from January 2005 to December 2020 for monthly price prediction of gold, silver, platinum, and palladium. The one-dimensional SGDE+SSA model was constructed without consideration of their price correlations. Indeed, these precious metal prices have a correlation, so augmenting their correlation into the SGDE+SSA model can make the model more realistic. The model was called the multidimensional SGDE+SSA model. Next, the expectation and variance of the numerical solution of the models will be derived. Furthermore, the sensitivity analysis of some important parameters will be studied as well. Finally, the accuracy of both models will be compared with historical price from January 2021 to January 2024 and the comparison is represented by the mean absolute percentage error (MAPE) as the criterion.

2 Literature Review

2.1 Precious Metals

Precious metals including the gold, silver, platinum, and palladium have high economic values. Therefore, these metals are interested asset for investor. For these four primary precious metals, gold is the most well-known metal because it can be used to standardize assets in many countries.

It is used in accessories, aerospace, medicine, and electronics [6]. It is one of the keys to the development of industry and technology. Silver is the second most common precious metal. It is an important industrial metal used in the electronics and photography industries. In some situations, silver prices can outperform gold during periods of high investor demand for industry. The correlation of gold and silver prices has been studied and it turns out that they have been strongly related in the same direction [10]. For platinum and palladium, they have similar physical and chemical properties since they are platinum-group metals (PGMs) which consist of platinum, palladium, ruthenium, rhodium, osmium and iridium. Platinum has important role in the automotive industry. It is used to make the catalyst for reducing emissions from vehicles. In addition, the computer and petroleum industries have more demand for the use of platinum. Gold and platinum prices are also highly correlated [8]. Palladium is another PGM with important industrial usage. It is used in electronics and industrial products, dentistry, medicine, chemical, and groundwater treatment. The advantages of investing in precious metals are a hedge against inflation, tangible assets, liquid investment, and portfolio diversification.

2.2 Model Development

Several models can be predict the metal prices. In this work, the SGDE+SSA model was adapted to forecast the gold, silver, platinum and palladium prices. The detail of GM(1,1), SGDE, SGDE+SSA, and accuracy of the model was described in this section.

2.2.1 Grey Model

Grey models deal with the series of primarily discrete data and converting difference equations to differential equations. The grey model creates a continuous and dynamic differential equation from a discrete series of data to predict of time series [2]. In this study, we are interested in GM(1,1) which means that the order of the differential equation equals one and there is only one variable. The following methodology explains the GM(1,1) model in detail. Let $x = x(1), x(2), \ldots, x(T)$ be a positive valued time series that is obtained by training data over a specific period. The accumulating generation operator (AGO) is used to smooth out the randomness of the primitive series. Applying the AGO on the original series [7], the new series by AGO is the monotonically increasing series $x^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \ldots, x^{(1)}(T)\}$ where elements of the new series are calculated as follows:

$$x^{(1)}(t) = \sum_{i=1}^{t} x(i), \quad \text{for } t = 1, 2, \ldots, T.$$

The generated mean value series of adjacent values of accumulated series $x^{(1)}$ is defined as:

$$z^{(1)} = \{z^{(1)}(2), z^{(1)}(3), ..., z^{(1)}(T)\},$$

where

$$z^{(1)}(t) = \frac{1}{2}\left(x^{(1)}(t) + x^{(1)}(t-1)\right), \quad \text{for } t = 2, 3, \ldots, T.$$

The grey differential equation modeling the mean value series is:

$$x(t) + az^{(1)}(t) = b. \tag{2.1}$$

Then, the whitened equation of the grey differential equation is defined as:

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b.$$

The solution of the grey differential equation is given by:

$$\tilde{x}^{(1)}(t+1) = \left(x(1) - \frac{b}{a}\right) e^{-at} + \frac{b}{a} \tag{2.2}$$

Moreover, parameters a and b can be obtained by using the least square method to (2.1) as follow:

$$[a, b]' = (B'B)^{-1}B'Y$$

where

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(t) & 1 \end{bmatrix} \quad \text{and} \quad Y = \begin{bmatrix} x(2) \\ x(3) \\ \vdots \\ x(t) \end{bmatrix}.$$

After that, the solution of grey differential equation can be calculated by substituting a and b in (2.2). Then, the predicted value of primitive metal price is shown below:

$$\begin{aligned} \tilde{x}(1) &= x(1) \\ \tilde{x}(t+1) &= \tilde{x}^{(1)}(t+1) - \tilde{x}^{(1)}(t), \quad \text{for } t = 1, 2, \ldots, T-1. \end{aligned}$$

Therefore, we obtain the predicted series $\tilde{x}(1), \tilde{x}(2), \ldots, \tilde{x}(T)$, and the forecasted series is

$$\tilde{x}(T+1), \tilde{x}(T+2), \ldots, \tilde{x}(T+h)$$

where h represents the number of steps ahead.

Next, GM(1,1) can be adapted to develop an SGDE model.

2.2.2 Stochastic Grey Differential Equation

The actual data $x(t)$ is the collection of observations which have been recorded at time $t = 1, 2, \ldots, T$. Assume that, for $t \geq 2$, their values can be expressed as [4]:

$$x(t) = x(t-1) + \omega(t), \quad \text{for } t = 1, 2, \ldots, T$$

where $\omega(t)$ is a discrete-time white noise process. Therefore, $x(t)$ relies on only $x(t-1)$ not $x(t-2), x(t-3), \ldots, x(1)$, i.e., the future value is associated with only the present value. Then, we can consider the grey differential equation of the AGO series as:

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b + k\sigma\omega(t).$$

Hence, we obtain an SGDE of the AGO series:

$$dx^{(1)}(t) = \left(b - ax^{(1)}(t)\right) dt + k\sigma dW(t) \tag{2.3}$$

where $x^{(1)}(0)$ is equal to $x(1)$, k is the coefficient that depends on the time scale of actual values, σ is the standard deviation of AGO series and $W(t)$ is a standard Brownian motion. The value of the coefficient k is 1, $\frac{1}{\sqrt{12}}$ and $\frac{1}{\sqrt{250}}$ for annual, monthly and daily time scales, respectively. Here, 250 is the number of trading days per year in the United State which is our reference market. Applying Ito's lemma with the function $f(t, x) = xe^{at}$, we have that

$$df\left(t, x^{(1)}(t)\right) = be^{at}dt + k\sigma e^{at}dW(t). \tag{2.4}$$

The solution of the previous SDE is represented by the integral equation:

$$x^{(1)}(T) = x^{(1)}(0)e^{-aT} + \frac{b}{a}(1 - e^{-aT}) + k\sigma \int_0^T e^{-a(T-t)}\, dW(t)$$

where the last integral is an Ito stochastic integral. Integrating (2.4) from time $t-1$ to t, we obtain the final discrete-time equation of the reconstructed AGO series:

$$\begin{cases} \hat{x}^{(1)}(1) & = x(1), \\ \hat{x}^{(1)}(t) & = \hat{x}^{(1)}(t-1)e^{-a\Delta t} + \frac{b}{a}(1 - e^{-a\Delta t}) + k\sigma\sqrt{\frac{1-e^{-2a\Delta t}}{2a}}Z_t \end{cases} \tag{2.5}$$

where $Z_t \sim N(0,1)$ and $\Delta t = 1$ (year, month, or day).

Simulating $\hat{x}^{(1)}(t)$ by (2.5) the space of simulation is created which can be represented by the following simulation matrix:

$$\hat{X}^{(1)} = [\hat{x}_{s,t}^{(1)}]_{s,t=1}^{S,T} = \begin{bmatrix} \hat{x}_{1,1}^{(1)} & \hat{x}_{1,2}^{(1)} & \dots & \hat{x}_{1,T}^{(1)} \\ \hat{x}_{2,1}^{(1)} & \hat{x}_{2,2}^{(1)} & \dots & \hat{x}_{2,T}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{x}_{S,1}^{(1)} & \hat{x}_{S,1}^{(1)} & \dots & \hat{x}_{S,T}^{(1)} \end{bmatrix}$$

where S denotes the total number of simulations and T is the number of monitoring periods. Each row of the previous matrix represents one artificial AGO path, while each column represents the set of artificial AGO values at time points. Obviously, for $t \geq 2$, the model generates sequence of expected values of $\widehat{X}$:

$$\begin{aligned} \hat{x}^{(1)}(1) &= x(1) \\ \hat{x}^{(1)}(1) &= \frac{1}{S}\sum_{s=1}^{S} \hat{x}_{s,t}^{(1)}, \quad \text{for } t = 2, 3, \dots, T. \end{aligned}$$

The inverse accumulated generating operation (IAGO) is used to reconstruct a primitive time series of the metal price:

$$\begin{aligned} \hat{x}^{(1)}(1) &= x(1) \\ \hat{x}(t+1) &= \hat{x}^{(1)}(t+1) - \hat{x}^{(1)}(t), \quad t = 1, 2, \dots, T-1. \end{aligned}$$

For simplicity, the reconstructed (predicted) series is expressed as:

$$y = \{\hat{x}(1), \hat{x}(2), \dots, \hat{x}(T)\}$$

and we obtain the forecasted series as:

$$\hat{x}(T+1), \hat{x}(T+2), \dots, \hat{x}(T+h),$$

where h represents the number of steps ahead.

2.2.3 Singular Spectrum Analysis

The objective of SSA is that the residual series can be decomposed into the sum of a small number of independent and interpretable components. The SSA consists of two sections: decomposition and reconstruction [5]. The decomposition section comprises phase1: embedding and phase2: singular value decomposition (SVD). The embedding phase regards a mapping that transfers a one-dimensional time series into a multidimensional series. Suppose $\varepsilon(t), t = 1, 2, ..., T$ is the residual error time series that the real-valued nonzero time series of sufficient length T. We

created a matrix X to collect the residual error series. Let L be an integer such that $2 \leq L \leq T$, and it is called the window length. The selection of L depends on the problem.

Assume that the actual and reconstructed value of the metal price at time t is $x(t)$ and $\hat{x}(t)$, respectively. The residual error is calculated according to the following expression:

$$\varepsilon_t = \hat{x}(t+1) - x(t+1), t = 1, 2, \ldots, T-1.$$

A mapping that transforms residual error series into a multidimensional matrix $[X_1, X_2, \ldots, X_K]$ with the following vectors $X_j = (\varepsilon_j, \ldots, \varepsilon_{j+L-1})' \in \mathbb{R}^L$, where $K = T - L$, is called an embedding. The result of embedding is the trajectory matrix:

$$X = \begin{bmatrix} \varepsilon_1 & \varepsilon_2 & \varepsilon_3 & \ldots & \varepsilon_K \\ \varepsilon_2 & \varepsilon_3 & \varepsilon_4 & \ldots & \varepsilon_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \varepsilon_L & \varepsilon_{L+1} & \varepsilon_{L+2} & \ldots & \varepsilon_{T-1} \end{bmatrix} = [X_1, X_2, \ldots, X_K] = [\varepsilon_{i,j}]_{i,j=1}^{L,K}.$$

Singular value decomposition phase aims to find the eigenvalues of XX' and eigenvectors. Corresponding eigenvalues and eigenvectors are denoted by $\lambda_1, , \lambda_L$ and $U_1, , U_L$. Eigenvalues and corresponding eigenvectors of the matrix XX' must be arranged in decreasing order. Moreover, $U_1, \ldots, U_L$ are orthonormal system, i.e., $\langle U_i, U_j \rangle = 0$ for $i \neq j$ and $\|U_i\| = 1$. Next, we denote $r = max\{i | \lambda_i > 0\} = rankX$. Therefore, the trajectory matrix X can be represented as the following sum of matrices

$$\hat{X} = \sum_{i=1}^{r} U_i U_i' X = \hat{X}_1 + \hat{X}_2 + \ldots \hat{X}_r$$

where $\hat{X}_i = U_i U_i' X$ for all $i = 1, 2, \ldots, r$.

Next, the reconstruction section composes the grouping and diagonal averaging phases. The grouping step corresponds to splitting the elementary matrices X_i into several groups and summing the matrices within each group. Let $I = \{i_1, \ldots, i_q\}$ be a group of indices $i_1, \ldots, i_q$. Then the matrix X_I corresponding to the group I is defined as $X_I = X_{i1} + \cdots + X_{iq}$. The spilt of the set of indices $J = 1, \ldots, q$ into the disjoint subsets $I_1, \ldots, I_m$ corresponds to the representation

$$\hat{X} = \hat{X}_{I_1} + \cdots + \hat{X}_{I_m}.$$

Selecting the value of q is important. The first approach to select the appropriate value of q is based on the plot of the logarithms of eigenvalues. The point where there is a significant drop in logarithm value is adopted as the appropriate value of q. The Second approach is the selecting q by w-correlation. This method is a measure of dependence between two reconstructed residual error series $X_T^{(1)}$ and $X_T^{(2)}$. This method can separate the series into groups for diagonal averaging. The w-correlation method is denoted as follows:

$$\rho_{12}^{(w)} = \frac{\left(X_T^{(1)}, X_T^{(2)}\right)_w}{\|X_T^{(1)}\|_w \|X_T^{(2)}\|_w} \tag{2.6}$$

where $\|X_T^{(i)}\|_w = \sqrt{\left(X_T^{(i)}, X_T^{(i)}\right)_w}$, $\left(X_T^{(i)}, X_T^{(j)}\right)_w = \sum_{k=1}^{T} w_k x_k^{(i)} x_k^{(j)}$, $w_k = \min\{k, L, T-k\}$ (assume $L \leq T/2$). The value of w-correlations is measured on a scale that varies from 0 to 1. If two reconstructed residual error series have zero w-correlation, it means that these two components are separable. Large values of w-correlations between reconstructed components indicate that the series should possibly be gathered into one group and correspond to the same series. Moreover, the rule of thumb for interpreting the size of a correlation coefficient is shown in Table 1.

Table 1: The rule of thumb for interpreting the size of a correlation coefficient

Size of correlation	Interpretation
0.90 to 1.00(−0.90 to −1.00)	Very high positive (negative) correlation
0.70 to 0.90(−0.70 to −0.90)	High positive (negative) correlation
0.50 to 0.70(−0.50 to −0.70)	Moderate positive (negative) correlation
0.30 to 0.50(−0.30 to −0.50)	Low positive (negative) correlation
0 to 0.30(0.00 to −0.30)	Negligible correlation

Therefore, each reconstructed matrix decomposition $\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_r$

$$\hat{X}_n[\hat{\varepsilon}_{i,j}]_{i,j=1}^{L,K} = \begin{bmatrix} \hat{\varepsilon}_{1,1}^{(n)} & \hat{\varepsilon}_{1,2}^{(n)} & \hat{\varepsilon}_{1,3}^{(n)} & \cdots & \hat{\varepsilon}_{1,K}^{(n)} \\ \hat{\varepsilon}_{2,1}^{(n)} & \hat{\varepsilon}_{2,2}^{(n)} & \hat{\varepsilon}_{2,3}^{(n)} & \cdots & \hat{\varepsilon}_{2,K+1}^{(n)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{\varepsilon}_{L,1}^{(n)} & \hat{\varepsilon}_{L+1,2}^{(n)} & \hat{\varepsilon}_{L+2,3}^{(n)} & \cdots & \hat{\varepsilon}_{L,T-1}^{(n)} \end{bmatrix} \quad \text{for } n = 1, 2, \ldots, r$$

is transformed into a new one-dimensional residual error time series of length $T-1$ by making the anti-diagonal averaging over the matrix elements:

$$\delta_n = (\delta_{n,1}, \delta_{n,2}, \ldots, \delta_{n,T-1})$$

where $\delta_{n,1} = \hat{\varepsilon}_{1,1}^{(n)}, \epsilon_{n,2} = \frac{\hat{\varepsilon}_{1,2}^{(n)} + \hat{\varepsilon}_{2,1}^{(n)}}{2}, \delta_{n,3} = \frac{\hat{\varepsilon}_{1,3}^{(n)} + \hat{\varepsilon}_{2,2}^{(n)} + \hat{\varepsilon}_{3,1}^{(n)}}{3}, \ldots, \delta_{n,T-1} = \hat{\varepsilon}_{L,T-1}^{(n)}$ for all $n = 1, 2, \ldots, r$. Accordingly, the original residual error time series is reconstructed by the sum of selected groups:

$$\epsilon(t) = \sum_{i=1}^{r} \delta_{1,t} = \delta_{1,t} + \delta_{2,t} + \cdots + \delta_{r,t}, t = 1, 2, \ldots, T-1.$$

Therefore, we can use the reconstructed residual error series $\epsilon = \epsilon(1), \epsilon(2), \ldots, \epsilon(T-1)$ to forecast the future values of residual error. Next, the future reconstructed residual error was generated based on the linear recurrent equation. Calculation of the linear vector of coefficients is performed in the following way:

$$R = \frac{1}{1-\nu^2} \sum_{i=1}^{r} \alpha_i U_i^{(L-1)} = (\beta_{L-1}, \beta_{L-2}, \ldots, \beta_1)' \tag{2.7}$$

where $U_i^{(L-1)}$ is the vector composed of the first $L-1$ values of the eigenvectors U_i, α_i is the last value of the eigenvectors $U_i, i = 1, 2, \ldots, r$ and ν^2 is the verticality coefficient which must satisfy the condition that $\nu^2 < 1$. The verticality coefficient is calculated as follows:

$$\nu^2 = \sum_{i=1}^{r} \alpha_i^2 = \alpha_1^2 + \alpha_2^2 + \cdots + \alpha_r^2.$$

The $h-$step ahead forecasting of the residual error is based on the following equation:

$$\epsilon(t) = R' \epsilon_h(t), \quad t = T+1, T+2, \ldots, T+h$$

where $\epsilon_h(T+i) = (\epsilon(T+i-L+1), \ldots, \epsilon(T+i-1))'$ for $i = 1, 2, \ldots, h$. Outcomes of the residual error model can be represented as:

$$\begin{cases} \{\epsilon(1), \epsilon(2), \ldots, \epsilon(T-1)\}, \\ \{\epsilon(T+1), \epsilon(T+2), \ldots, \epsilon(T+h)\}. \end{cases}$$

2.3 SGDE+SSA Model

The SGDE+SSA is the model that combines the SGDE model and the SSA technique. The model consisted of reconstructed residual error, the reconstructed (predicted) series, and the forecasted series from the SGDE model. The reconstructed (predicted) series is expressed as:

$$y = \{\hat{x}(1), \hat{x}(2), \ldots, \hat{x}(T)\}$$

and the forecasted series is form as:

$$y = \{\hat{x}(T+1), \hat{x}(T+2), \ldots, \hat{x}(T+h)\}$$

Accordingly, the reconstructed (predicted) series with reconstructed residual was obtained from the SGDE+SSA model. The series is expressed as follows:

$$y = \{\hat{x}(1), \hat{x}(2) + \epsilon(1), \ldots, \hat{x}(T) + \epsilon(T-1)\}$$

and we obtain the forecasted series with reconstructed residual as follows:

$$\hat{x}(T+1) + \epsilon(T+1), \hat{x}(T+2) + \epsilon(T+2), \ldots, \hat{x}(T+h) + \epsilon(T+h).$$

2.4 Accuracy of the Model

The mean absolute percentage error is a measure describing the degree of deviation of predicted values from actual values to percentages. The MAPE is one of the most popular measures of forecasting accuracy which is defined by

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\frac{|\hat{x}(t) - x(t)|}{x(t)} \times 100\%$$

where $\hat{x}(t)$ and $x(t)$ represent the predicted value and actual values.

However, the MAPE is not appropriate for all data. If the actual value of data is close to zero, MAPE produces infinite or undefined values for some occurrence. Additionally, for the actual values less than one or very small, their MAPEs have extremely large percentage errors (outliers), while zero actual values result in infinite MAPEs [9].

3 Main Results

The one-dimensional SGDE+SSA and multidimensional SGDE+SSA models of gold, silver, platinum, and palladium prices were constructed to forecast these prices by using the historical prices from January 2005 to January 2024 [11]. A one-dimensional model means that the models were created without the correlation between four precious metals. Accordingly, the precious metals are not related to other metals. In contrast, the correlations of four prices were considered for constructing a multidimensional model because these prices have relations in real occur. For the construction of the models, the one-dimensional SGDE and multidimensional SGDE models were created. After that, the SSA technique was applied to improve the SGDE models to be SGDE+SSA models. For this work, the one-dimensional SGDE and multidimensional SGDE models were simplified called SGDE and MSGDE models. Moreover, the one-dimensional SGDE+SSA and multidimensional SGDE+SSA models were defined as SGDE+SSA and MSGDE+SSA models.

Firstly, the monthly historical prices of gold, silver, platinum, and palladium were collected from January 2005 to January 2024. This data was separated into two subsets including training data and testing data. The training data ranged from January 2005 to December 2020 (192 months) which was used for model construction. Next, the testing data ranged from January

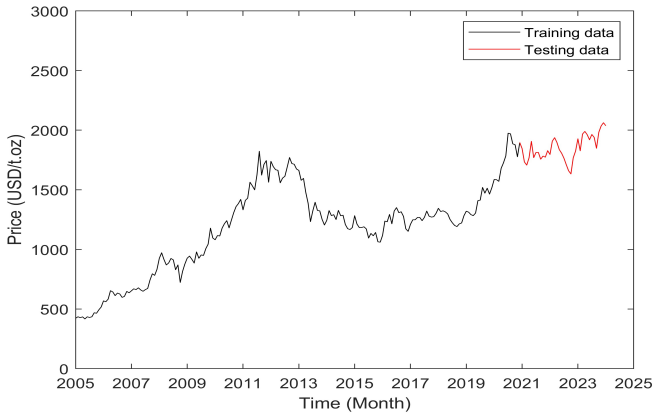


Figure 1: Historical gold price from January 2005 to January 2024

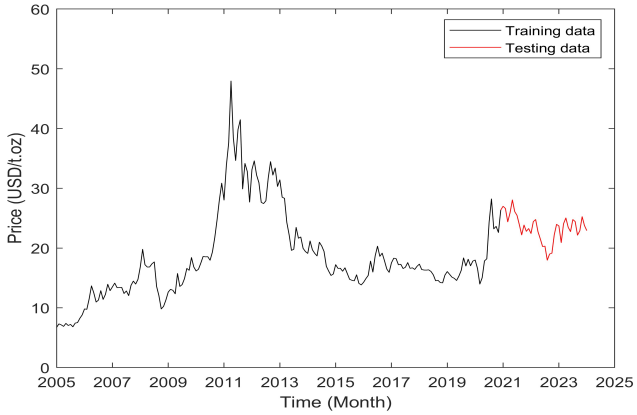


Figure 2: Historical silver price from January 2005 to January 2024

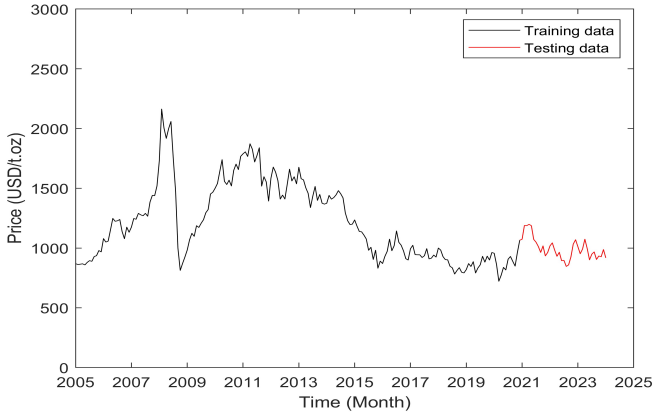


Figure 3: Historical platinum price from January 2005 to January 2024

2021 to January 2024 (37 months). This data was compared with forecasted prices to indicate the proficiency of model prediction. Additionally, historical prices are shown in Figure 1 to Figure 4 where black and red lines represent training and testing data, respectively.

The historical price graphs exposed two interesting behaviors of structures. The first interested behavior of prices appeared from 2008 to 2012. The gold, silver, platinum, and palladium prices decreased from 2008 to early 2009 although these prices explicitly increased after mid-2009. After that, all prices from the middle of 2011 to 2012 reversed to a downtrend again. This

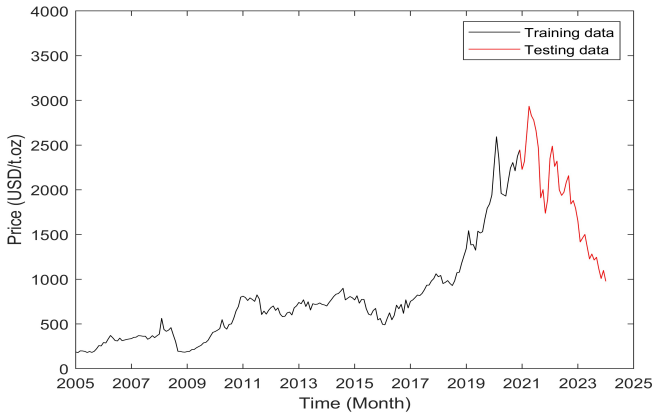


Figure 4: Historical palladium price from January 2005 to January 2024

behavior of prices was affected by the global financial crisis that started in 2008. The crisis was caused by the housing market crisis in the United States. The Housing Market refers to the supply and demand for houses, usually in a particular country or region. The event was the worst housing crisis since the Great Depression (1926-1947). Several people in the world lost their occupations, residences, and businesses. The crucial factors of this crisis consisted of the subprime mortgage crisis, high levels of doubt and a lack of regulation in the financial system. The subprime mortgage crisis was the primary problem of the housing market crisis. As a result, the precious metals including gold, silver, platinum, and palladium prices continually increased after intermediate 2009. However, these prices decreased after the end of 2012 because the housing market crisis was unraveled by government and financial institutions. The second interesting structure appeared from 2019 till to the present. The Covid-19 pandemic wildly spreads around the world. This situation affected the confidence of investors, financial institutions, and the stability of the economic system. The event made a similar result in the housing market crisis for four precious metal prices which their prices increased from previous years. The gold, silver, and platinum prices slightly increased, although the palladium prices sharply rose because of high demand for production. Several industries desired to use palladium for constructing their product. The platinum was replaced by palladium for autocatalyst curbing harmful emission. The reason is the prices between palladium and platinum which the palladium price is cheaper than platinum price. However, the palladium demand suddenly reduced in 2021. This situation affected to rapidly decreased the values of palladium from 2021 to the present.

3.1 One-Dimensional and Multidimensional SGDE Models

The primitive series are defined $x_g = \{x_g(t)\}, x_s = \{x_s(t)\}, x_p = \{x_p(t)\}$ and $x_l = \{x_l(t)\}$. Series represents the historical prices of gold, silver, platinum, and palladium, respectively. Applying the method in section 2, the AGO series of precious metals were generated to convert primitive series into monotonically increasing series. Accordingly, the whitened grey differential equations are constructed as follows:

$$\frac{dx_g^{(1)}(t)}{dt} + a_g z_g^{(1)}(t) = b_g, \tag{3.1}$$

$$\frac{dx_s^{(1)}(t)}{dt} + a_s z_s^{(1)}(t) = b_s, \tag{3.2}$$

$$\frac{dx_p^{(1)}(t)}{dt} + a_p z_p^{(1)}(t) = b_p, \tag{3.3}$$

$$\frac{dx_l^{(1)}(t)}{dt} + a_l z_l^{(1)}(t) = b_l \tag{3.4}$$

where x_g, x_s, x_p and x_l are primitive price at time t,
z_g, z_s, z_p and z_l are mean valued series of adjacent values of AGO series,
a_g, a_s, a_p and a_l parameters are defined by least square method,
b_g, b_s, b_p and b_l parameters are defined by least square method.

Moreover, the estimated parameters by least square method are represented in Table 2. Next,

Table 2: Estimated parameters by least square method for the model

Parameter	Value	Parameter	value
a_g	-3.7141×10^{-3}	b_g	8.1710×10^{2}
a_s	-1.2053×10^{-3}	b_s	16.4278
a_p	1.8199×10^{-3}	b_p	1.4451×10^{3}
a_l	-1.2118×10^{-2}	b_l	1.6224×10^{2}

GM(1,1) and SDE models were combined to construct the SGDE model. Consequently, the (3.1) to (3.4) are appended the additional term. This term is a discrete-time white noise process. Therefore, the future value is associated with only the present value as follows:

$$\begin{aligned}
\frac{dx_g^{(1)}(t)}{dt} + ax_g^{(1)}(t) &= b_g + k\sigma_g\omega_g(t),\\
\frac{dx_s^{(1)}(t)}{dt} + ax_s^{(1)}(t) &= b_s + k\sigma_s\omega_s(t),\\
\frac{dx_p^{(1)}(t)}{dt} + ax_p^{(1)}(t) &= b_p + k\sigma_p\omega_p(t),\\
\frac{dx_l^{(1)}(t)}{dt} + ax_l^{(1)}(t) &= b_l + k\sigma_l\omega_l(t).
\end{aligned}$$

where $x_g^{(1)}(1) = x_g(1), x_s^{(1)}(1) = x_s(1), x_p^{(1)}(1) = x_p(1)$ and $x_l^{(1)}(1) = x_l(1)$,
k is time scale of historical values,
$\sigma_g, \sigma_s, \sigma_p$ and σ_l are the standard deviation of AGO series for precious metals,
$\omega_g, \omega_s, \omega_p$ and ω_l are white noise.

Hence, SGDEs of the AGO series are defined as follows:

$$dx_g^{(1)}(t) = \left(b_g - a_g x_g^{(1)}(t)\right) dt + k\sigma_g dW_g(t), \tag{3.5}$$

$$dx_s^{(1)}(t) = \left(b_s - a_s x_s^{(1)}(t)\right) dt + k\sigma_s dW_s(t), \tag{3.6}$$

$$dx_p^{(1)}(t) = \left(b_p - a_p x_p^{(1)}(t)\right) dt + k\sigma_p dW_p(t), \tag{3.7}$$

$$dx_l^{(1)}(t) = \left(b_l - a_l x_l^{(1)}(t)\right) dt + k\sigma_l dW_l(t) \tag{3.8}$$

where $W_g(t), W_s(t), W_p(t)$ and $W_l(t)$ are standard Brownian motion.

Next, the MSGDEs of the AGO series were constructed. Firstly, the correlations of four prices were analysed which are shown below:

$$C = \begin{bmatrix} 1 & \rho_{gs} & \rho_{gp} & \rho_{gl} \\ \rho_{gs} & 1 & \rho_{sp} & \rho_{sl} \\ \rho_{gp} & \rho_{sp} & 1 & \rho_{pl} \\ \rho_{gl} & \rho_{sl} & \rho_{pl} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0.7707 & 0.1601 & 0.6892 \\ 0.7707 & 1 & 0.6109 & 0.2419 \\ 0.1601 & 0.6109 & 1 & -0.3196 \\ 0.6892 & 0.2419 & -0.3196 & 1 \end{bmatrix}$$

where ρ_{gs}, ρ_{gp}, ρ_{gl}, ρ_{sp}, ρ_{sl} and ρ_{pl} are correlation coefficient.

The coefficient correlations indicate the direction, relation, and strength of prices. In this case, the coefficients of correlation results were separated into four types which are displayed in Table 1. The first range is a very highly positive correlation. The correlation coefficient is 0.7707 for the relationship between gold and silver prices. These prices have a very highly positive correlation, so there appears to be a considerable association between the two variables. For silver and platinum prices, the correlation coefficient is 0.6109 which has the similarity between gold and silver prices. Therefore, gold and silver prices are strongly changing in the same direction. Moreover, platinum price strongly changes with a strong change in silver price as well. The second range is a moderate positive correlation. The correlation coefficient is 0.6892 for the relationship between gold and palladium prices. These prices moderately change in the same direction. Next, the third range is low correlation. The correlation coefficient is 0.1601 for the relationship between gold and platinum prices. For platinum and silver prices, the correlation coefficient is 0.2419. These prices slowly change in the same direction. In contrast, platinum and palladium prices change in the opposite direction and the correlation coefficient is -0.3196. Secondly, the AGO series of MSGDE was recreated because the Itô term of the AGO series in the SGDE model was adjusted to a new term that considers correlations. The cholesky factorization method was applied to generate the new term which the process of the MSGDE model is

$$\begin{bmatrix} dx_g^{(1)}(t) \\ dx_s^{(1)}(t) \\ dx_p^{(1)}(t) \\ dx_l^{(1)}(t) \end{bmatrix} = \begin{bmatrix} b_g - a_g x_g^{(1)}(t) \\ b_s - a_s x_s^{(1)}(t) \\ b_p - a_p x_p^{(1)}(t) \\ b_l - a_l x_l^{(1)}(t) \end{bmatrix} dt + kc \begin{bmatrix} \sigma_g & 0 & 0 & 0 \\ 0 & \sigma_s & 0 & 0 \\ 0 & 0 & \sigma_p & 0 \\ 0 & 0 & 0 & \sigma_l \end{bmatrix} \begin{bmatrix} dW_g(t) \\ dW_s(t) \\ dW_p(t) \\ dW_l(t) \end{bmatrix}$$

where c is lower matrix. After that, c was calculated by applying cholesky factorization to correlation matrix (C). Therefore, the MSGDE model defines as follows:

$$\begin{bmatrix} dx_g^{(1)}(t) \\ dx_s^{(1)}(t) \\ dx_p^{(1)}(t) \\ dx_l^{(1)}(t) \end{bmatrix} = \begin{bmatrix} b_g - a_g x_g^{(1)}(t) \\ b_s - a_s x_s^{(1)}(t) \\ b_p - a_p x_p^{(1)}(t) \\ b_l - a_l x_l^{(1)}(t) \end{bmatrix} dt + k \begin{bmatrix} \sigma_g dW_g(t) \\ \rho_1 \sigma_g dW_g(t) + \rho_2 \sigma_s dW_s(t) \\ \rho_3 \sigma_g dW_g(t) + \rho_4 \sigma_s dW_s(t) + \rho_5 \sigma_p dW_p(t) \\ \rho_6 \sigma_g dW_g(t) + \rho_7 \sigma_s dW_s(t) + \rho_8 \sigma_p dW_p(t) + \rho_9 \sigma_l dW_l(t). \end{bmatrix} \tag{3.9}$$

where $\rho_1 = 0.7707$, $\rho_2 = 0.6371$, $\rho_3 = 0.1601$, $\rho_4 = 0.7651$, $\rho_5 = 0.6237$, $\rho_6 = 0.6892$, $\rho_7 = -0.4540$, $\rho_8 = -0.1324$ and $\rho_9 = 0.5490$.

After that, Ito's lemma was applied to solve the solution of (3.5) to (3.9). The final discrete time of the reconstructed AGO series of these metal prices for SGDEs is

$$\begin{aligned} &\hat{x}_g^{(1)}(1) = x_g(1), \\ &\hat{x}_g^{(1)}(t) = \hat{x}_g^{(1)}(t-1)e^{-a_g \Delta t} + \frac{b_g}{a_g}(1 - e^{-a_g \Delta t}) + k\sigma_g \sqrt{\frac{1 - e^{-2a_g \Delta t}}{2a_g}} Z_t \qquad (3.10) \\ &\hat{x}_s^{(1)}(1) = x_s(1), \\ &\hat{x}_s^{(1)}(t) = \hat{x}_s^{(1)}(t-1)e^{-a_s \Delta t} + \frac{b_s}{a_s}(1 - e^{-a_s \Delta t}) + k\sigma_s \sqrt{\frac{1 - e^{-2a_s \Delta t}}{2a_s}} Z_t, \qquad (3.11) \end{aligned}$$

$$\hat{x}_p^{(1)}(1) = x_p(1),$$
$$\hat{x}_p^{(1)}(t) = \hat{x}_p^{(1)}(t-1)e^{-a_p\Delta t} + \frac{b_p}{a_p}(1-e^{-a_p\Delta t}) + k\sigma_p\sqrt{\frac{1-e^{-2a_p\Delta t}}{2a_p}}Z_t, \tag{3.12}$$
$$\hat{x}_l^{(1)}(1) = x_l(1),$$
$$\hat{x}_l^{(1)}(t) = \hat{x}_l^{(1)}(t-1)e^{-a_l\Delta t} + \frac{b_l}{a_l}(1-e^{-a_l\Delta t}) + k\sigma_l\sqrt{\frac{1-e^{-2a_l\Delta t}}{2a_l}}Z_t \tag{3.13}$$

where $Z_t \sim N(0,1)$ and $\Delta t = 1$ represents month. Moreover, the discrete-time of the reconstructed AGO series for MSGDEs is

$$\hat{x}_g^{(1)}(1) = x_g(1),$$
$$\hat{x}_g^{(1)}(t) = \hat{x}_g^{(1)}(t-1)e^{-a_g\Delta t} + \frac{b_g}{a_g}(1-e^{-a_g\Delta t}) + k\sigma_g\sqrt{\frac{1-e^{-2a_g\Delta t}}{2a_g}}Z_t^{(1)} \tag{3.14}$$
$$\hat{x}_s^{(1)}(1) = x_s(1),$$
$$\hat{x}_s^{(1)}(t) = \hat{x}_s^{(1)}(t-1)e^{-a_s\Delta t} + \frac{b_s}{a_s}(1-e^{-a_s\Delta t}) + \rho_1 k\sigma_g\sqrt{\frac{1-e^{-2a_g\Delta t}}{2a_g}}Z_t^{(1)} + \rho_2 k\sigma_s\sqrt{\frac{1-e^{-2a_s\Delta t}}{2a_s}}Z_t^{(2)}, \tag{3.15}$$
$$\hat{x}_p^{(1)}(1) = x_p(1),$$
$$\hat{x}_p^{(1)}(t) = \hat{x}_p^{(1)}(t-1)e^{-a_p\Delta t} + \frac{b_p}{a_p}(1-e^{-a_p\Delta t}) + \rho_3 k\sigma_g\sqrt{\frac{1-e^{-2a_g\Delta t}}{2a_g}}Z_t^{(1)} + \rho_4 k\sigma_s\sqrt{\frac{1-e^{-2a_s\Delta t}}{2a_s}}Z_t^{(2)} + \rho_5 k\sigma_p\sqrt{\frac{1-e^{-2a_p\Delta t}}{2a_p}}Z_t^{(3)}, \tag{3.16}$$
$$\hat{x}_l^{(1)}(1) = x_l(1),$$
$$\hat{x}_l^{(1)}(t) = \hat{x}_l^{(1)}(t-1)e^{-a_l\Delta t} + \frac{b_l}{a_l}(1-e^{-a_l\Delta t}) + \rho_6 k\sigma_g\sqrt{\frac{1-e^{-2a_g\Delta t}}{2a_g}}Z_t^{(1)} + \rho_7 k\sigma_s\sqrt{\frac{1-e^{-2a_s\Delta t}}{2a_s}}Z_t^{(2)} + \rho_8 k\sigma_p\sqrt{\frac{1-e^{-2a_p\Delta t}}{2a_p}}Z_t^{(3)} + \rho_9 k\sigma_l\sqrt{\frac{1-e^{-2a_l\Delta t}}{2a_l}}Z_t^{(4)} \tag{3.17}$$

where $Z_t^{(1)}, Z_t^{(2)}, Z_t^{(3)}, Z_t^{(4)} \overset{iid}{\sim} N(0,1)$ and $\Delta t = 1$ represents month.

Next, the solutions of the reconstructed AGO series were used to simulate 100,000 paths for each price by (3.10) to (3.17). The parameters of simulations are shown in Table 2. T equals 192 which represents the number of monthly training data. k is equal to $\frac{1}{\sqrt{12}} \approx 0.2887$ for monthly time scale. Moreover, the 100,000 simulated paths of the reconstructed AGO series were computed mean.

Then, the mean path was calculated to generate the predicted and forecasted series of the SGDE and MSGDE models by using the inverse accumulated generating operation (IAGO). The predicted price is the price of the model in the period of training data and The forecasted price is the price of the model in the period of testing data. These prices are shown in Figure 5 to Figure 8 where black, blue, and red dash lines are prices of history, SGDE and MSGDE models, respectively. Moreover, the black dashed line is the separated line to split the training

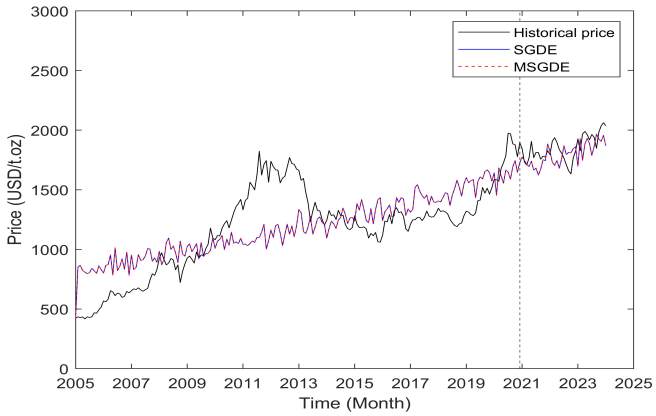


Figure 5: Monthly gold price of history, SGDE and SGDE models

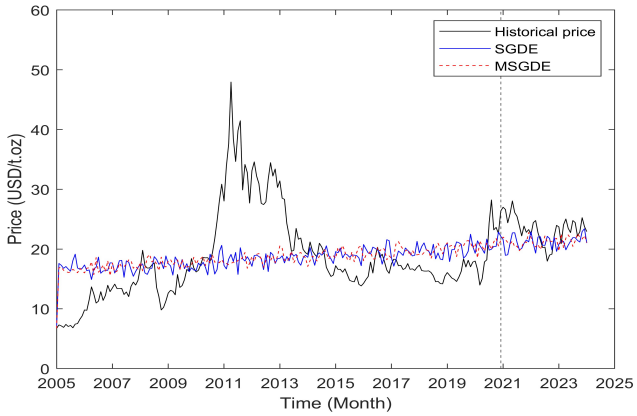


Figure 6: Monthly silver price of history, SGDE and SGDE models

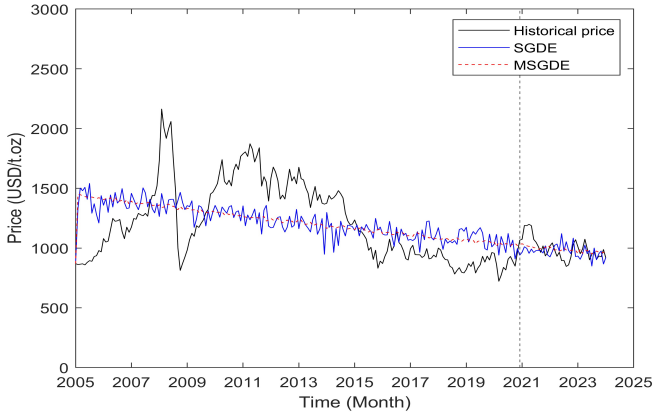


Figure 7: Monthly platinum price of history, SGDE and SGDE models

and testing period. The figures indicate the prices of SGDE and MSGDE models are not significantly different, although these prices do not fit with the historical prices. This situation is related to the results of MAPE which are in Table 3.

The results of MAPE indicate that the SGDE and MSGDE models can forecast precious metal prices. However, the efficiencies of the models should improve because the MAPE values more than 20% for predicted prices. These prices are similar in short periods because historical prices are more fluctuate and rapidly change due to many factors such as politics, geography, and

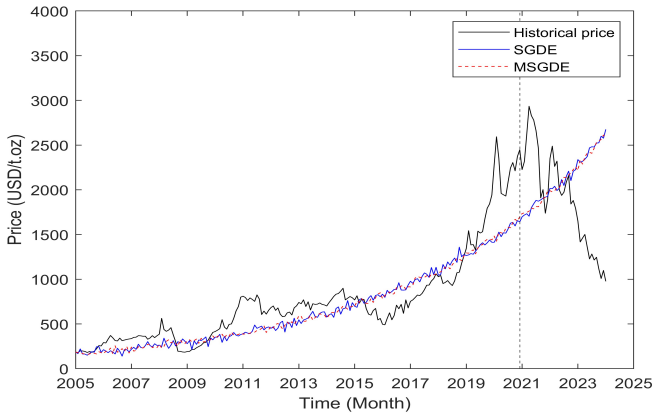


Figure 8: Monthly palladium price of history, SGDE and SGDE models

Table 3: Accuracy of SGDE for predicted and forecasted prices

Model	**Price**	**MAPE**			
		Gold	**Silver**	**Platinum**	**Palladium**
SGDE	Predicted	21.5219	30.7918	21.7233	26.2394
	Forecasted	5.7029	11.2589	8.3864	45.5903
MSGDE	Predicted	21.5219	30.5373	21.1318	25.5757
	Forecasted	5.7029	11.5225	6.1698	45.5810

economics. The explicit examples of this situation are the housing market crisis, the COVID-19 pandemic, and high palladium demand which were described. Moreover, the predicted prices of SGDE are slightly changed as a result of the first term of the right-hand side in (3.5), (3.6), (3.7) and (3.8) has linear character. Moreover, the standard deviation of AGO series ($\sigma_g, \sigma_s, \sigma_p$ and σ_l) are constant values that are not enough to describe the historical prices. Therefore, the SGDE and MSGDE can not cover all situations that suddenly happened in the historical price. For forecasted prices, this model has satisfied efficiency in predicting gold and platinum prices. The MAPE values are less than 10%. Moreover, the MSGDE model has a better result of prediction than the SGDE model which the MAPE of the MSGDE model is less than the MAPE of the SGDE model for platinum price. The results of silver price prediction show that the efficiency of SGDE and MSGDE models are not significantly different at 11%. In contrast, the capability of platinum price prediction is not as expected where MAPE is 45.5903% and 45.5810% for SGDE and MSGDE. This platinum situation is affected by the high palladium demand. Consequently, the SGDE and MSGDE models can predict gold, silver, and platinum prices but the model is not suitable for palladium price prediction. However, the efficiency of SGDE and MSGDE models can improve to increase more accuracy of prediction. This problem can use the SSA technique to reduce the error described in the next section 3.3. Next, the expectation and variance of the numerical solution were derived in the next section.

3.2 Expectations and Variances of Numerical Solution

The solution of the SGDE and MSGDE model, the simulations, means and the results of predictions were shown in the previous section. In this section, the expectation and variance of models were derived. Firstly, the solutions of SGDEs represented by the integral equation were

recalled as follows:

$$
\begin{aligned}
x_g^{(1)}(T) &= x_g^{(1)}(0)e^{-a_g T} + \frac{b_g}{a_g}\left(1 - e^{-a_g T}\right) + k\sigma_g \int_0^T e^{-a_g(T-t)} dW_g(t),\\
x_s^{(1)}(T) &= x_s^{(1)}(0)e^{-a_s T} + \frac{b_s}{a_s}\left(1 - e^{-a_s T}\right) + k\sigma_s \int_0^T e^{-a_s(T-t)} dW_s(t),\\
x_p^{(1)}(T) &= x_p^{(1)}(0)e^{-a_p T} + \frac{b_p}{a_p}\left(1 - e^{-a_p T}\right) + k\sigma_p \int_0^T e^{-a_p(T-t)} dW_p(t),\\
x_l^{(1)}(T) &= x_l^{(1)}(0)e^{-a_l T} + \frac{b_l}{a_l}\left(1 - e^{-a_l T}\right) + k\sigma_l \int_0^T e^{-a_l(T-t)} dW_l(t),
\end{aligned}
$$

and the solutions of MSGDEs are

$$
\begin{aligned}
x_g^{(1)}(T) =& x_g^{(1)}(0)e^{-a_g T} + \frac{b_g}{a_g}\left(1 - e^{-a_g T}\right) + k\sigma_g \int_0^T e^{-a_g(T-t)} dW_g(t),\\
x_s^{(1)}(T) =& x_s^{(1)}(0)e^{-a_s T} + \frac{b_s}{a_s}\left(1 - e^{-a_s T}\right) + \rho_1 k\sigma_g \int_0^T e^{-a_g(T-t)} dW_g(t)\\
&+ \rho_2 k\sigma_s \int_0^T e^{-a_s(T-t)} dW_s(t),\\
x_p^{(1)}(T) =& x_p^{(1)}(0)e^{-a_p T} + \frac{b_p}{a_p}\left(1 - e^{-a_p T}\right) + \rho_3 k\sigma_g \int_0^T e^{-a_g(T-t)} dW_g(t)\\
&+ \rho_4 k\sigma_s \int_0^T e^{-a_s(T-t)} dW_s(t) + \rho_5 k\sigma_p \int_0^T e^{-a_p(T-t)} dW_p(t),\\
x_l^{(1)}(T) =& x_l^{(1)}(0)e^{-a_l T} + \frac{b_l}{a_l}\left(1 - e^{-a_l T}\right) + \rho_6 k\sigma_g \int_0^T e^{-a_g(T-t)} dW_g(t)\\
&- \rho_7 k\sigma_s \int_0^T e^{-a_s(T-t)} dW_s(t) - \rho_8 k\sigma_p \int_0^T e^{-a_p(T-t)} dW_p(t)\\
&+ \rho_9 k\sigma_l \int_0^T e^{-a_l(T-t)} dW_l(t).
\end{aligned}
$$

After that, the expectations were taken in these integral equations. The last integrals represent stochastic integrals that have the centered Gaussian distribution. Then, the expectation of stochastic integrals is equal to zero. Consequently, the expectations of SGDE and MSGDE are

$$
\begin{aligned}
\mathbb{E}\left[x_g^{(1)}(T)\right] &= x_g^{(1)}(0)e^{-a_g T} + \frac{b_g}{a_g}\left(1 - e^{-a_g T}\right),\\
\mathbb{E}\left[x_p^{(1)}(T)\right] &= x_p^{(1)}(0)e^{-a_p T} + \frac{b_p}{a_p}\left(1 - e^{-a_p T}\right),\\
\mathbb{E}\left[x_p^{(1)}(T)\right] &= x_p^{(1)}(0)e^{-a_p T} + \frac{b_p}{a_p}\left(1 - e^{-a_p T}\right),\\
\mathbb{E}\left[x_l^{(1)}(T)\right] &= x_l^{(1)}(0)e^{-a_l T} + \frac{b_l}{a_l}\left(1 - e^{-a_l T}\right).
\end{aligned}
$$

Additionally, Itô isometry was applied to solve variances which variances of SGDE are

$$\begin{aligned}
\mathsf{Var}\left[x_g^{(1)}(T)\right] &= k^2\sigma_g^2\left(\frac{1-e^{-2a_gT}}{2a_g}\right),\\
\mathsf{Var}\left[x_s^{(1)}(T)\right] &= k^2\sigma_s^2\left(\frac{1-e^{-2a_sT}}{2a_s}\right),\\
\mathsf{Var}\left[x_p^{(1)}(T)\right] &= k^2\sigma_p^2\left(\frac{1-e^{-2a_pT}}{2a_p}\right),\\
\mathsf{Var}\left[x_l^{(1)}(T)\right] &= k^2\sigma_l^2\left(\frac{1-e^{-2a_lT}}{2a_l}\right),
\end{aligned}$$

and variances of MSGDE are

$$\begin{aligned}
\mathsf{Var}\left[x_g^{(1)}(T)\right] =& k^2\sigma_g^2\left(\frac{1-e^{-2a_gT}}{2a_g}\right),\\
\mathsf{Var}\left[x_s^{(1)}(T)\right] =& \rho_1^2k^2\sigma_g^2\left(\frac{1-e^{-2a_gT}}{2a_g}\right)+\rho_2^2k^2\sigma_s^2\left(\frac{1-e^{-2a_sT}}{2a_s}\right),\\
\mathsf{Var}\left[x_p^{(1)}(T)\right] =& \rho_3^2k^2\sigma_g^2\left(\frac{1-e^{-2a_gT}}{2a_g}\right)+\rho_4^2k^2\sigma_s^2\left(\frac{1-e^{-2a_sT}}{2a_s}\right)\\
&+\rho_5^2k^2\sigma_p^2\left(\frac{1-e^{-2a_pT}}{2a_p}\right),\\
\mathsf{Var}\left[x_l^{(1)}(T)\right] =& \rho_6^2k^2\sigma_g^2\left(\frac{1-e^{-2a_gT}}{2a_g}\right)+\rho_7^2k^2\sigma_s^2\left(\frac{1-e^{-2a_sT}}{2a_s}\right)\\
&+\rho_8^2k^2\sigma_p^2\left(\frac{1-e^{-2a_pT}}{2a_p}\right)+\rho_9^2k^2\sigma_l^2\left(\frac{1-e^{-2a_lT}}{2a_l}\right).
\end{aligned}$$

Next, the SGDE+SSA and MSGDE+SSA models were developed and shown in the next section.

3.3 SGDE+SSA and MSGDE+SSA Models

SSA is a high-performance technique to reduce the error of model prediction. This technique has two sections. In the first section, the X matrix was constructed to collect errors of a model by considering the important parameter L. The idea of selecting L has many ideas such as the behavior of prices and time period. In this work, L equals $12, 24, 36, 48, 60, 72$, and 84 which represents a period of the year. Next, the $S = XX'$ was generated and computed eigenvalues (λ) and eigenvectors (U). Moreover, eigenvectors were applied to construct several $\hat{X}$ matrices for each eigenvector which were calculated by $UU'X$. For the second section, the idea of generating predicted price and forecasted price has two ways. Firstly, several $\hat{X}$ matrices were grouped into the signal group and noise group by considering the plot of logarithms of eigenvalues. The signal group was further calculated in the diagonal averaging method. Secondly, several matrices were calculated by the diagonal averaging method. Then, the w-correlation was applied to the group results of the diagonal averaging method. Consequently, the predicted prices of SGDE+SSA and MSGDE+SSA were completed from two ideas, and the forecasted prices were generated by (2.7). Four logarithm plots were created for each L value. The varying of q was considered from these plots. The examples of logarithm plots $(L = 12)$ were shown in Figure 9.

For the logarithm plot of gold, the significant drop in values occurs around component 8 which the next component could be interpreted as the start of the noise group. The noise components produced a slowly decreasing sequence of singular values. Accordingly, the only components were organized into noise groups. In contrast, the components from 1 to 8 were grouped into signal groups. Therefore, q values of gold are equal to 8 which were shown in

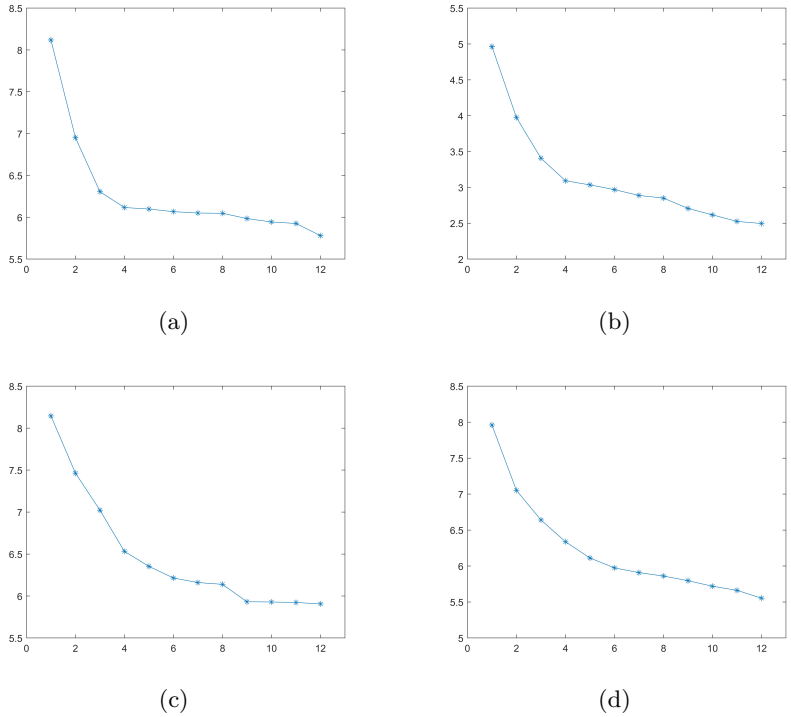


Figure 9: The plot of logarithm for eigenvalues ($L = 12$)

Figure 9(a). In the same way, q values of silver, platinum, and palladium are equal to 8, 8, and 8 which can be observed in the Figure. 9(b), 9(c) and 9(d), respectively.

Table 4: Accuracy of SGDE+SSA by plot of logarithm for predicted price

L	**MAPE**							
	q	**Gold**	**q**	**Silver**	**q**	**Platinum**	**q**	**Palladium**
SGDE	-	21.5219	-	30.7918	-	21.7233	-	26.2394
12	8	2.8485	8	3.9569	8	3.2240	8	5.3470
24	7	4.9144	11	6.2195	13	4.7597	10	7.3112
	19	2.8778	19	4.3454	20	3.3955	19	4.5871
36	9	5.9469	15	7.7577	13	7.0360	13	7.9336
	20	4.8041	22	6.9360	17	6.3362	27	5.7109
48	12	5.7028	18	7.7757	16	7.5091	16	7.3556
	22	5.5334	32	6.7516	27	6.4364	26	7.3556
	36	4.2334	40	5.7820	39	5.4904	32	6.7924
60	11	6.8142	8	10.6824	10	9.0262	9	12.1925
	17	6.3522	19	8.9443	15	8.1724	18	10.4624
	36	4.9380	27	8.2822	27	7.2263	35	8.7255
	42	4.3073	48	6.4964	54	5.5388	45	7.6218
72	11	4.9345	10	8.1801	17	6.6952	15	9.2353
	29	5.3666	26	8.2531	29	6.5426	20	8.8537
	38	5.4169	34	8.1514	45	6.4726	30	8.9812
	56	4.5812	49	8.0076	64	6.1434	62	7.7587
84	9	7.4499	12	11.7264	13	9.5236	11	13.2503
	30	6.6674	28	10.4351	22	8.8116	20	12.1548
	39	6.0493	34	10.0509	37	7.9043	37	10.6370
	70	4.5524	70	7.3364	68	6.7466	68	8.6626

Table 5: Accuracy of MSGDE+SSA by plot of logarithm for predicted price

L	MAPE							
	q	Gold	q	Silver	q	Platinum	q	Palladium
SGDE	-	21.5219	-	30.5373	-	21.1318	-	25.5757
12	8	2.8485	8	3.5729	8	2.3522	8	4.2624
24	7	4.1944	11	5.8410	14	3.4916	10	5.9845
	19	2.8778	17	4.6267	21	2.6572	22	3.4241
36	9	5.9469	17	7.2299	15	5.4984	12	6.7683
	20	4.8041	27	5.7957	22	4.8514	22	5.4927
48	6	7.0241	10	9.2221	18	6.2619	14	7.6661
	22	5.5334	23	7.7021	26	5.6173	31	5.9785
	36	4.2334	41	5.8449	38	4.8778	41	5.0759
60	11	6.8142	13	9.3156	20	6.7476	9	10.8568
	17	6.3522	23	8.5397	29	6.1757	16	9.0417
	36	4.9380	31	7.8563	35	5.8955	28	8.2708
	42	4.3073	44	6.7292	48	5.2846	50	6.4193
72	11	7.3138	13	10.2455	21	7.2968	15	10.2714
	29	6.2853	28	9.4680	29	6.8323	32	8.9938
	38	5.7817	44	8.4774	53	5.9776	51	7.7571
	56	4.5812	58	7.3221	62	5.7178	62	7.0036
84	12	7.3751	11	10.6437	20	7.7859	15	10.9075
	30	6.6674	28	10.1433	35	7.1399	28	10.1340
	39	6.0493	56	8.1190	41	6.9202	49	8.7593
	50	5.2332	64	7.5681	66	6.2468	72	7.6768

Table 6: Accuracy of SGDE+SSA and MSGDE+SSA by plot of logarithm for forecasted prices

Model	Price	MAPE			
		Gold	Silver	Platinum	Palladium
SGDE+SSA	Predicted	2.8485	3.9569	3.2240	4.8571
	Forecasted	5.7642	11.0591	8.5403	44.1193
MSGDE+SSA	Predicted	2.8485	3.5729	2.3522	3.4241
	Forecasted	5.7642	11.7033	6.2948	47.8827

The results of SGDE+SSA and MSGDE+SSA by plot of logarithm are shown in Table 4 and Table 5, respectively. The MAPE results show that the highest efficiency strategy of SGDE+SSA are $(L = 12, q = 8), (L = 12, q = 8), (L = 12, q = 8)$ and $(L = 24, q = 19)$ for gold, silver, platinum and palladium price. These strategies predicted the prices with the lowest MAPE values which are equal to $2.8485\%, 3.9569\%, 3.2240\%$, and 4.5871%, respectively. For the highest efficiency strategies of MSGDE+SSA, the lowest MAPE are 2.8485%, 3.5729%, 2.3522% and 3.4241% which happened from $(L = 12, q = 8), (L = 12, q = 8), (L = 12, q = 8)$ and $(L = 24, q = 22)$. These results show that the SGDE+SSA and MSGDE+SSA models can reduce the error of predicted price in the SGDE and MSGDE models where MAPE approximated 20% to 30%. Furthermore, the efficiency of SGDE+SSA and MSGDE+SSA models were compared in Table 6 for best L and q values for forecasted prices. The results indicated the MSGDE+SSA model can predict gold, silver, and platinum prices which the MAPE values less than the SGDE+SSA model. In contrast, the error of forecasted palladium prices is 44.1193% and 47.8827% because the model can not describe the many fluctuations of palladium from the high demand of industries. Consequently, the MSGDE+SSA model is a more suitable model than SGDE, MSGDE, and SGDE+SSA for gold, silver, and platinum prices, although these models

are not suitable models to forecast the palladium price. Finally, the prices of SGDE+SSA and MSGDE+SSA for the highest strategy shown in Figure 10 to Figure 13.

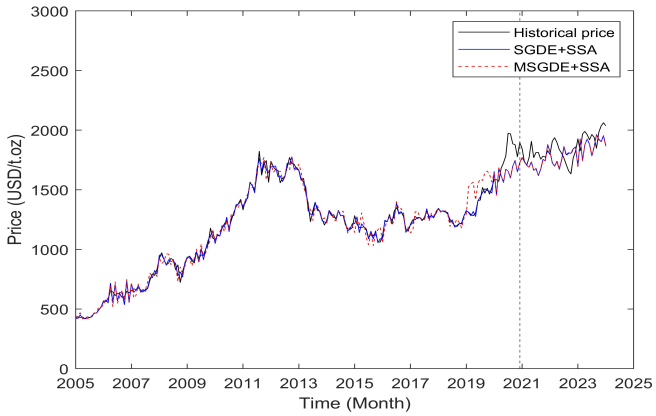


Figure 10: The gold prices of SGDE+SSA and MSGDE+SSA by plot of logarithm

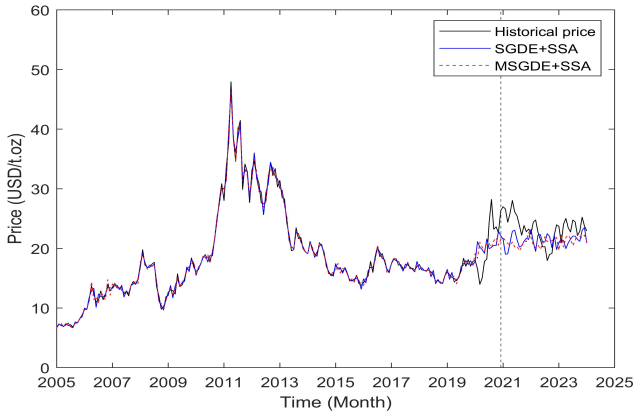


Figure 11: The silver prices of SGDE+SSA and MSGDE+SSA by plot of logarithm

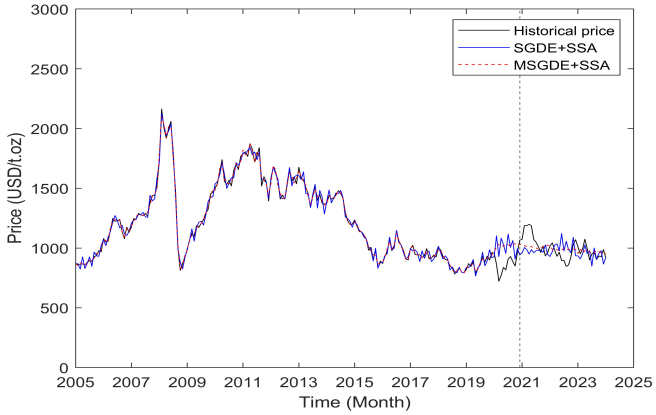


Figure 12: The platinum prices of SGDE+SSA and MSGDE+SSA by plot of logarithm

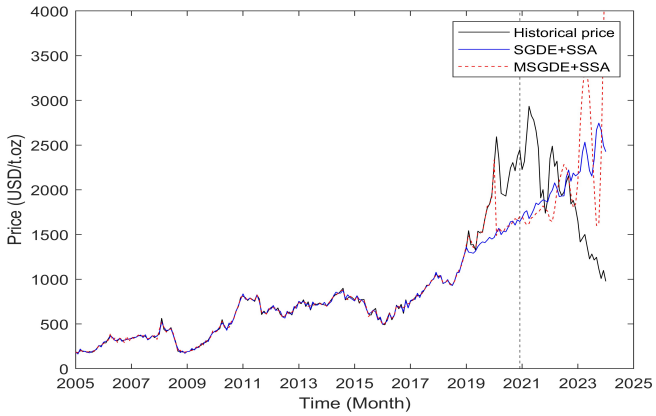
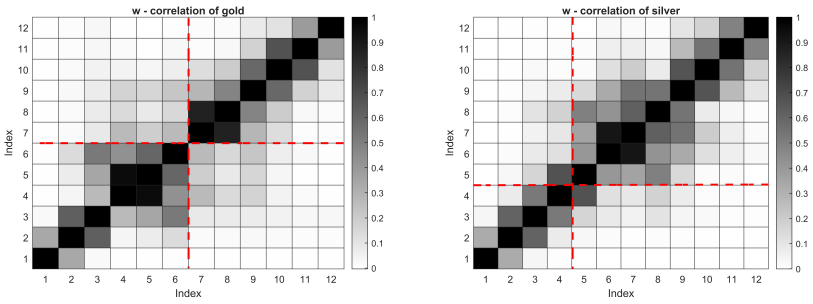


Figure 13: The palladium prices of SGDE+SSA and MSGDE+SSA by plot of logarithm



(a) (b)

(c) (d)

Figure 14: The heatmap of w-correlation ($L = 12$)

For the w-correlation method, the results are displayed in the grey heatmap which color represents the correlation of two series. The values of w-correlation are measured on a scale that varies from 0 to 1. Large values of w-correlation indicate that the series should possibly be gathered into one group and correspond to the same series. The examples of heatmap for $L = 12$ are depicted in Figure 14 for SGDE+SSA. The q values of precious metal prices were selected by observing many boxes of correlation. The red line represents the separated point that denoted the q value. The components before the red line were combined into the signal group and the other components were combined into the noise group. Next, the components in the signal group were summed to generate predicted and forecasted prices. These results of SGDE+SSA and MSGDE+SSA by the w-correlation method are shown in Table 7 and Table 8, respectively. For predicted prices, the $(L = 24, q = 16), (L = 24, q = 14), (L = 12, q = 8)$ and $(L = 12, q = 7)$ are the highest strategies of SGDE+SSA model and $(L = 24, q = 16), (L =$

$12, q = 5), (L = 12, q = 6)$ and $(L = 12, q = 7)$ are the highest strategies of MSGDE+SSA model as well. The comparison of capability between SGDE+SSA and MSGDE+SSA shown in Table 9 and the prices shown in Figure 15 to Figure 18. The results indicated that the MSGDE+SSA model has more efficiency of prediction than SGDE+SSA but the models are not suitable to forecast the palladium price.

Table 7: Accuracy of SGDE+SSA by w-correlation method for predicted price

L	**MAPE**							
	q	**Gold**	**q**	**Silver**	**q**	**Platinum**	**q**	**Palladium**
SGDE	-	21.5219	-	30.7918	-	21.7233	-	26.2394
12	6	3.7188	4	6.2874	8	3.2240	7	5.9453
24	11	4.7968	9	6.5649	12	4.9164	5	8.5298
	16	3.4349	14	5.4655	19	3.4911	12	6.6051
36	10	5.7731	9	8.9442	8	7.9820	7	9.2523
	20	4.8041	19	7.3320	19	6.1453	20	6.6950
48	12	6.4016	18	8.1610	10	8.4670	12	9.4755
	22	5.5334	28	7.2122	16	7.6826	21	7.9847
	30	4.7239	32	6.7516	25	6.6946	32	6.7924
60	13	6.6420	8	10.6824	11	8.7387	14	10.7706
	26	5.7361	23	8.5898	25	7.3642	28	9.4499
	36	4.9380	32	7.9495	39	6.4366	35	8.7255
72	13	7.1275	14	10.6619	12	9.2940	15	12.0536
	29	6.2853	23	9.9448	29	7.8065	30	10.4453
	38	5.7817	34	9.2321	38	7.2788	37	9.8613
	49	4.9054	57	7.3873	43	7.0196	48	9.0601
84	13	7.2713	10	12.3394	18	9.0796	15	12.5461
	28	6.7600	28	10.4351	38	7.8298	36	10.6464
	41	5.8237	49	8.7002	69	6.7149	50	10.0875

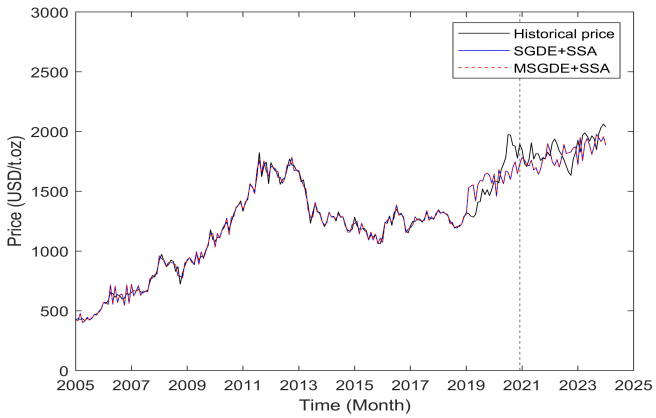


Figure 15: The gold prices of SGDE+SSA and MSGDE+SSA by w-correlation method

Table 8: Accuracy of MSGDE+SSA by w-correlation method for predicted price

L	MAPE							
	q	Gold	q	Silver	q	Platinum	q	Palladium
SGDE	-	21.5219	-	30.5373	-	21.1318	-	25.5757
12	6	3.7188	5	4.6845	6	2.8232	7	4.7458
24	9	4.4906	9	6.0812	11	3.9306	5	6.7758
	16	3.4349	16	4.8708	14	3.4916	15	5.0494
36	11	5.7466	12	7.7828	15	5.4984	12	6.7683
	22	4.4725	22	6.5116	22	4.8514	22	5.4927
48	12	6.4016	9	9.2759	18	6.2619	12	7.8661
	22	5.5334	24	7.6190	31	5.3240	21	7.1172
	31	4.6357	31	6.9042	37	4.9220	30	6.1951
60	13	6.6420	15	9.1996	22	6.5837	15	9.1065
	27	5.6118	23	8.5397	29	6.1757	24	8.6440
	37	4.7975	31	7.8563	37	5.8383	36	7.5647
	49	3.8893	41	6.9512	44	5.4785	45	6.9700
72	13	7.1275	10	10.9947	11	7.8405	16	9.9125
	28	6.3350	21	9.8928	21	7.2968	26	9.5027
	38	5.7817	33	9.1630	38	6.4844	49	7.8854
	49	4.9054	50	8.0510	53	5.9776	52	7.6688
84	23	6.9375	17	10.7315	20	7.7859	20	10.3713
	34	6.3242	30	10.1386	36	7.0768	38	9.4805
	42	5.7693	45	9.1004	51	6.6132	55	8.3406
	57	4.9957		8.2585	62	6.3179	66	7.8281

Table 9: Accuracy of SGDE+SSA and MSGDE+SSA by w-correlation for forecasted prices

Model	Price	MAPE			
		Gold	Silver	Platinum	Palladium
SGDE+SSA	Predicted	3.4349	5.4655	3.2240	5.9453
	Forecasted	5.4418	11.1913	8.5403	46.6540
MSGDE+SSA	Predicted	3.4349	4.6845	2.8232	4.7458
	Forecasted	5.4418	11.8667	6.2756	46.6051

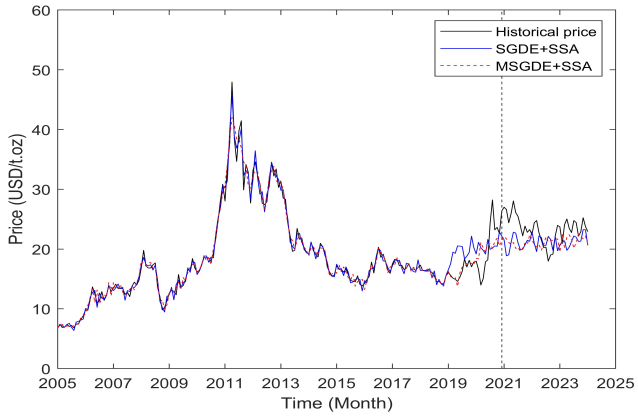


Figure 16: The silver prices of SGDE+SSA and MSGDE+SSA by w-correlation method

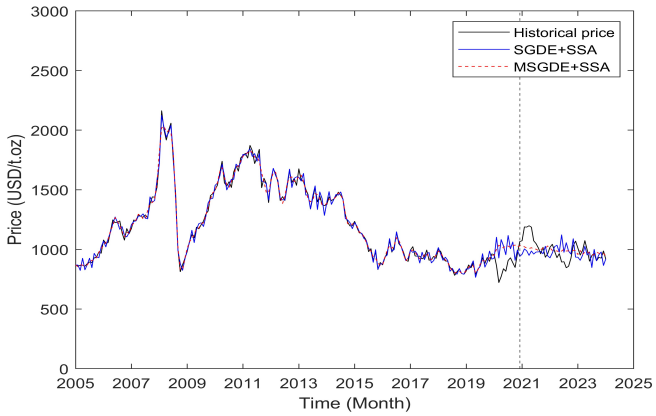


Figure 17: The platiunm prices of SGDE+SSA and MSGDE+SSA by w-correlation method

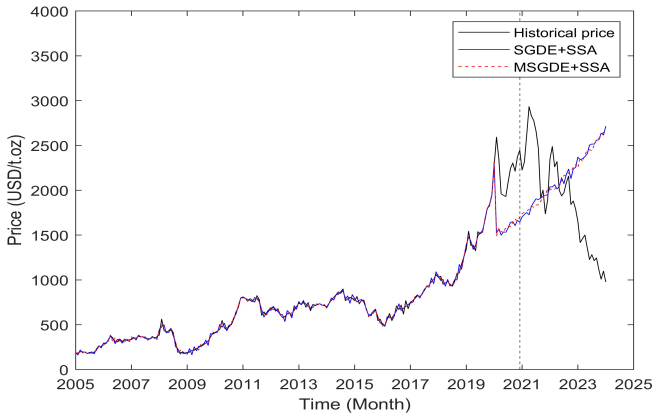


Figure 18: The palladium prices of SGDE+SSA and MSGDE+SSA by w-correlation method

4 Discussion and Conclusion

The one-dimensional and multidimensional models were constructed in this work. The SGDE and MSGDE models can predict the gold, silver, platinum, and palladium prices but the accuracy of the models is not satisfied. Then, the SGDE+SSA and MSGDE+SSA models were created to improve the efficiency of the SGDE and MSGDE models. The SGDE+SSA and MSGDE+SSA models have higher performance than SGDE or MSGDE models for predicted prices of gold, silver, platinum, and palladium although the accuracy of the forecasted prices are not significantly different for four precious metals. Interestingly, this study also found that the MSGDE+SSA has a higher capacity than the SGDE+SSA model. However, the error of the forecasted palladium price is not as expected as it is higher than the error of the other three precious metals. This is because of the decreasing palladium demand after 2021. Therefore, a more appropriate model needs to be explored for predicting palladium prices. Taken together, the results from this study confirmed that the MSGDE+SSA model is the best model to forecast gold, silver, and platinum prices comparing with SGDE, SGDE+SSA and MSGDE models.

References

[1] A. Alipour, A. A. Khodaiari and A. Jafari, *Modeling and prediction of time-series of monthly copper prices*, Int. J. Min. Geo-Eng. **53**(1) (2019), 91–97.

[2] S. Balochian and H. Baloochian, *Improving Grey Prediction Model and Its Application in Predicting the Number of Users of a Public Road Transportation System*, J. Intell. Syst. **30** (2021), 104–114.

[3] H. Bilgil, *New grey forecasting model with its application and computer code*, AIMS Math. **6**(2) (2021), 1497–1514.

[4] Z. Gligorić, M. Gligorić, D. Halilović, Ć. Beljić and K. Urošević, *Hybrid stochastic-grey model to forecast the behavior of metal price in the mining industry*, Sustainability. **12**(16) (2020), 6533.

[5] H. Hassani, *Singular spectrum analysis: methodology and comparison*, J. Data Sci. **5** (2007), 239–257.

[6] N. Issaranusorn, S. Rujivan and K. Mekchay, *Stochastic model for gold prices and its application for no-arbitrage gold derivative pricing*, JNAO. **2**(1) (2011), 9–14.

[7] E. Kayacan, B. Ulutas and O. Kaynak, *Grey system theory-based models in time series prediction*, Expert Syst. Appl. **37**(2) (2010), 1784–1789.

[8] A. A Kearney and R. E. Lombra, *Gold and platinum: Toward solving the price puzzle*, Q. REV. ECON. FINAC. **49** (2009), 884–892.

[9] S. Kim and H. Kim, *A new metric of absolute percentage error for intermittent demand forecasts*, Int. J. Forecasting **32** (2016), 669–679.

[10] J. Sami , *Has the long-run relationship between gold and silver prices really disappeared? Evidence from an emerging market*, Resour. Policy **74** (2021), 102292.

[11] Trading Economics. (2024) *Trading economics: commodities*. Retrieved February 12, 2024, from https://tradingeconomics.com/commodities.

12. Probability Theory and Statistics

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Non-uniform Bound on Translated Poisson Approximation for Poisson Binomial Random Variables via Exchangeable Pair Coupling

Kamonrat Kamjornkittikoon[1,†] and Suporn Jongpreechaharn[2,‡]

[1]Department of Mathematics and Statistics, Faculty of Science and Technology
Kanchanaburi Rajabhat University, Kanchanaburi 71190, Thailand

[2]Department of Mathematics and Computer Science, Faculty of Science
Chulalongkorn University, Bangkok 10330, Thailand

Abstract

It is known that a Poisson binomial distribution, which represents a sum of non-identically Bernoulli distributed random variable, can be approximated by normal or Poisson distribution. In this study, we focus on the approximation of a Poisson binomial distribution through a translated Poisson distribution. To achieve this, we introduce a non-uniform bound for the approximation, utilizing Stein's method and exchangeable pair coupling. Furthermore, we provide an illustrative example to compare the sharpness of our derived bound with previous result.

Keywords: translated Poisson approximation, Stein's method, non-uniform bound.

2020 MSC: Primary 60F05.

1 Introduction and Main Result

Let $X_1, X_2, X_3, \ldots, X_n$ be a sequence of independent Bernoulli random variables such that

$$p_i = P(X_i = 1), \qquad q_i = P(X_i = 0) = 1 - p_i$$

for each i, and let

$$W = \sum_{i=1}^{n} X_i, \quad \lambda = \sum_{i=1}^{n} p_i \quad \text{and} \quad \sigma^2 = \sum_{i=1}^{n} p_i q_i. \tag{1.1}$$

This distribution of W is formally known as the Poisson binomial distribution, parameterized by $\mathbf{p} = (p_1, ..., p_n)$. In instances where all p_i are uniform and equal to p, the distribution simplifies to the binomial distribution with parameters n and p. It is established that the distribution

†Speaker. ‡Corresponding author.
Email: kamonrat.k@kru.ac.th (K. Kamjornkittikoon), suporn.j@chula.ac.th (S. Jongpreechaharn).

of W can be approximated by the Poisson distribution with a mean λ, expressed as $\mathbf{Po}(\lambda)$, particularly when the probabilities p_i are sufficiently small.

Many authors have developed the error bound for approximating the distribution of W. One of the known uniform bound is obtained by Barbour and Hall [1] as follows:

$$\left|P(W \in A) - \mathbf{Po}(\lambda)(A)\right| \leq \lambda^{-1}(1-e^{-\lambda}) \sum_{i=1}^{n} p_i^2 \leq \min\left\{1, \lambda^{-1}\right\} \sum_{i=1}^{n} p_i^2, \tag{1.2}$$

where $\mathbf{Po}(\lambda)(A) = \sum\limits_{k \in A} \frac{\lambda^k e^{-\lambda}}{k!}$ and $A \subseteq \mathbb{Z}$.

In the case that p_i's are not all small, the estimation of the Poisson binomial distribution with a translated Poisson distribution introduced by Kruopis [6] is investigated to obtain a closer approximation than using the Poisson distribution. We say that an integer-valued random variable Y has a *translated Poisson distribution* with parameters λ and σ^2 written by

$$Y \sim \mathbf{TP}(\lambda, \sigma^2)$$

if $Y - \lambda + \sigma^2 + \gamma \sim \mathbf{Po}(\sigma^2 + \gamma)$, where $\gamma = \langle \lambda - \sigma^2 \rangle$ and $\langle x \rangle = x - \lfloor x \rfloor$ denotes the fractional part of x. Note that $\mathbb{E}Y = \lambda \quad \text{and} \quad \sigma^2 \leq \mathrm{Var}Y = \sigma^2 + \gamma \leq \sigma^2 + 1$. Note also that $\mathbf{Po}(\sigma^2) = \mathbf{TP}(\sigma^2, \sigma^2)$.

Let $Z \sim \mathbf{TP}(\lambda, \sigma^2)$. Čekanavičius [14] proved an error bound for W by utilizing the method of characteristic function. For $0 \leq p_i \leq \frac{1}{2}$ and any Borel sets A, the bound is expressed as follows:

$$|P(W \in A) - P(Z \in A)| \leq C \min\left\{\sigma^{-3} \sum_{i=1}^{n} p_i^2 + \gamma\sigma^{-2}, \sum_{i=1}^{n} p_i^2\right\}, \tag{1.3}$$

where C is a positive constant.

In 2001, Čekanavičius and Vaitkus [15] used Stein's method to estimate the total variation distance between the distribution of W and a translated Poisson distribution, leading to the following inequality. For $0 \leq p_i \leq 1$ and $\sigma > 0$, and for any Borel set A, the inequality is given by:

$$\left|P(W \in A) - P(Z \in A)\right| \leq \sum_{i=1}^{n} p_i^2 q_i \min\left\{2, b^{-1}\tau^{-\frac{1}{2}}\right\} + \gamma \min\left\{1, b^{-1}\right\} + e^{-\frac{\sigma^2}{4}}, \tag{1.4}$$

where $b = \sigma^2 + \gamma$ and $\tau = \sigma^2 - \max\limits_{1 \leq j \leq n} \{p_j q_j\}$.

Afterward, Barbour and Čekanavičius [3] also illustrated the the distance between the distribution of W and a translated Poisson distribution using Stein's method. This demonstration is expressed as follows: for any sets $A \subseteq \mathbb{Z}$,

$$\left|P(W \in A) - P(Z \in A)\right| \leq \min\left\{1, \sigma^{-2}\right\} \left(\gamma + 2d \sum_{i=1}^{n} p_i^2 q_i\right) + \sigma^{-2}, \tag{1.5}$$

where $d \leq \left[\left(\sum\limits_{i=1}^{n} \frac{1-|p_i - q_i|}{2}\right) - 1\right]^{-\frac{1}{2}}$ (see Proposition 4.6 in [4]). In the binomial case where $p \leq \frac{1}{2}$, the convergence rate of (1.5) has order $O(p^{\frac{1}{2}} n^{-\frac{1}{2}} + (np)^{-1})$. Further details can be found in [7] and [8].

In addition to Stein's method, there is also an approach known as the exchangeable pair method, introduced by Stein [12], which is commonly utilized. The exchangeable pair method involves identifying a suitable transformation of random variables, typically seeking a new pair of variables, to ensure that the joint distribution of the transformed pair remains invariant under permutations. For a given random variable Y, we call (Y, Y') an *exchangeable pair* if there is

another random variable Y' such that $\mathcal{L}(Y, Y') = \mathcal{L}(Y', Y)$. As in [10] and [9], the exchangeable pair (Y, Y') is constructed in such a way that

$$\mathbb{E}^Y(Y' - \mathbb{E}Y) = (1 - \nu)(Y - \mathbb{E}Y) + R \tag{1.6}$$

for some $\nu \in (0, 1)$ and R is a random variable of small order. In general, random variable Y and Y' exhibit slight differences stemming from their property of being an exchangable pair. As a result, when we consider the integer-valued random variable Y, it is reasonable to assume that $Y - Y' \in \{-1, 0, 1\}$.

By following the construction method presented by Röllin [10], let $X_1^*, \ldots, X_n^*$ be independent copies of the X_i, and an exchangeable pair (W, W') is constructed to satisfy

$$W' = W - X_K + X_K^*,$$

where K is uniformly distributed over $\{1, 2, \ldots, n\}$. This construction satisfies the condition described in (1.6), with $R \equiv 0$ and $\nu = \frac{1}{n}$. Additionally, Röllin [10] utilized the technique of exchangeable pairs to establish an error bound for translated Poisson approximation, which is presented as follows:

$$\left|P(W \in A) - P(Z \in A)\right| \leq \frac{2 + \sqrt{\sum_{i=1}^{n} p_i^3 q_i}}{\sum_{i=1}^{n} p_i q_i} \tag{1.7}$$

for all $A \subseteq \mathbb{Z}$

Some remarks in the case $p_i = p \leq \frac{1}{2}$ are shown in the following details.

1. Estimates (1.2), (1.3) and (1.7) have order $O(\min\{p, np^2\})$, $O\left(\min\left\{np^2, p^{\frac{1}{2}}n^{-\frac{1}{2}} + (np)^{-1}\right\}\right)$ and $O(p^{\frac{1}{2}}n^{-\frac{1}{2}} + (np)^{-1})$, respectively. When the probability p is significantly smaller than a positive power of n, represented as $p = O(n^{-\delta})$, (1.5) and (1.7) exhibit the same order, following $O\left(n^{-\frac{1}{2}(\delta+1)}\right)$ when $0 < \delta < \frac{1}{3}$, and $O\left(n^{\delta-1}\right)$ when $\delta \geq \frac{1}{3}$.
2. In the case where $p = O(n^{-\delta})$, it can be observed that estimate (1.2) has order $O(n^{-\delta})$ for $0 < \delta < 1$. When $0 < \delta < \frac{1}{3}$, the error bounds presented in (1.3), (1.4), and (1.7) all have the same order of magnitude, which is $O(n^{-\frac{1}{2}(\delta+1)})$. These error bounds converge to zero at a faster rate compared to the error bound presented in (1.2) as n approaches infinity. Similarly, for $\frac{1}{3} \leq \delta < \frac{1}{2}$, the error bounds in (1.3), (1.4), and (1.7) share the same order of $O(n^{\delta-1})$, which also tend to zero faster than the error bound in (1.2) as n approaches infinity. Furthermore, in the case of $\frac{1}{2} \leq \delta < 1$, the approximation using $\mathbf{Po}(\lambda)$ converges to zero more rapidly than the approximation using $\mathbf{TP}(\lambda, \sigma^2)$. For $\frac{1}{2} \leq \delta < \frac{2}{3}$, all estimates (1.3), (1.4), and (1.7) have the same order denoted as $O(n^{\delta-1})$. Lastly, for $\frac{2}{3} \leq \delta < 1$, estimate (1.4) and (1.7) still share the same order of $O(n^{\delta-1})$, while estimate (1.3) exhibits an order of $O(n^{1-2\delta})$.
3. In the scenario of $p = O(n^{-\delta})$ for $\delta \geq 1$, both estimate (1.2) and (1.3) have the same order of $O(n^{1-2\delta})$, which tends to zero. However, the estimates for (1.4) and (1.7) with an order of $O(n^{\delta-1})$ does not tend to zero.

Therefore, when $0 < \delta < \frac{1}{2}$, the approximation using $\mathbf{TP}(\lambda, \sigma^2)$ converges to zero more rapidly than $\mathbf{Po}(\lambda)$. In our work, we use Stein's method and exchangeable pairs to achieve our main result. To find a non-uniform bound for approximating the distribution of W, we concentrate on the translated Poisson distribution as approximation which is shown in the following theorem. Set $b = \sigma^2 + \gamma$ and $s = \lambda - \sigma^2 - \gamma = \lambda - b$.

Theorem 1.1. *Suppose that* $w_0 \in \{1, \ldots, n-1\}$ *such that* $w_0 \leq \frac{b}{\sigma^3}\left(\sum_{i=1}^{n} p_i^2\right)^2$. *We have the following results.*

1. *If* $w_0 > s$, *then*

$$|P(W \leq w_0) - P(Z \leq w_0)| \leq \left(\sqrt{\sum_{i=1}^{n} p_i^3 q_i} + 1\right) b^{-1} \min\left\{1 - e^{-b}, \frac{2b^{-1}(e^b - b - 1)}{w_0 - s + 1}, \frac{b}{w_0 - s}\right\} + \frac{1}{w_0 \sigma^3}\left(\sum_{i=1}^{n} p_i^2\right)^2. \tag{1.8}$$

2. *If* $w_0 = s$, *then*

$$|P(W \leq w_0) - P(Z \leq w_0)| \leq \frac{2}{w_0 \sigma^3}\left(\sum_{i=1}^{n} p_i^2\right)^2. \tag{1.9}$$

3. *If* $w_0 < s$, *then*

$$|P(W \leq w_0) - P(Z \leq w_0)| \leq \frac{1}{w_0 \sigma^3}\left(\sum_{i=1}^{n} p_i^2\right)^2. \tag{1.10}$$

It is evident that the convergence rate of (1.7) which is a uniform bound, is $O(n^{-\frac{1}{2}(\delta+1)})$, whereas the convergence rate of (1.8) which is a non-uniform bound, is $O\left(\frac{n^{-\frac{1}{2}(\delta+1)}}{w_0}\right)$ when $0 < \delta < \frac{1}{3}$. This indicates that the bound provided by (1.8) is sharper than the bound (1.7) as w_0 increases. Furthermore, when $\frac{1}{3} \leq \delta < \frac{1}{2}$, the convergence rate in (1.7) and (1.8) are $O(n^{\delta-1})$ and $O\left(\frac{n^{-\frac{1}{2}(\delta+1)}}{w_0}\right)$, respectively. This further confirms that our non-uniform bound refines the previous bound. Additionally, for $0 < \delta < \frac{1}{3}$, the error bounds in (1.9) and (1.10) with a rate of $O\left(\frac{n^{-\frac{1}{2}(\delta+1)}}{w_0}\right)$ are more accurate than (1.7) with a rate of $O(n^{-\frac{1}{2}(\delta+1)})$. In another case, when $\frac{1}{3} \leq \delta < \frac{1}{2}$, the convergence rate in (1.7) is $O(n^{\delta-1})$, whereas (1.9) and (1.10) exhibit $O\left(\frac{n^{-\frac{1}{2}(\delta+1)}}{w_0}\right)$.

The subsequent sections of this work are organized as follows. Section 2 displays numerical results demonstrating the contrast between uniform and non-uniform bounds. Lastly, the proof of the main result is presented.

2 Example

Let $X_1, X_2, X_3, \ldots, X_n$ be a sequence of independent Bernoulli random variables with parameter p. Then, $W = \sum_{i=1}^{n} X_i$ has a binomial distribution with mean np and variance $np(1-p)$. In this section, we provide a comparison of a uniform bound in (1.7) with our result in Theorem 1.1. By applying the main theorem, the non-uniform bounds are sharper than uniform bounds for sufficiently large w_0 under the condition that $w_0 \leq \frac{b}{\sigma^3}\left(\sum_{i=1}^{n} p_i^2\right)^2$. It should be noted that the non-uniform bounds are more precise compared to the uniform bounds, attributed to the impact of w_0. When $n = 1000$ and $p = 0.3$, the outcomes are illustrated in Table 1.

Table 1: Comparison of uniform bound and non-uniform bound for translated Poisson approximation when $n = 1000$ and $p = 0.3$

s	w_0	Uniform Bound	Non-Uniform Bound
90	80	0.03023	0.03327
	85	0.03023	0.03131
	87	0.03023	0.03059
	89	0.03023	0.02991
90	90	0.03023	0.05915
90	100	0.03023	0.05208
	150	0.03023	0.04321
	200	0.03023	0.03877
	250	0.03023	0.03611
	300	0.03023	0.03434
	350	0.03023	0.02817
	400	0.03023	0.02390
	500	0.03023	0.01837

3 Proof of Main Theorem

Our main result in Theorem 1.1 will be proved by Stein's method. The method was originally introduced by Stein [11] for normal approximation, and the idea was adapted to the Poisson distribution by Chen [5].

Proof of Theorem 1.1. Let $w_0 \in \{1, \ldots, n-1\}$ be such that $w_0 \leq \frac{b}{\sigma^3}\left(\sum_{i=1}^{n} p_i^2\right)^2$. First, we notice that

$$P(W \leq s) = P(W - \lambda \leq -b) \leq P(|W - \lambda| \geq b) \leq \frac{\sigma^2}{b^2},$$

where we utilize the Markov inequality in the last inequality. From this fact and the formula for b that $b = \sigma^2 + \gamma$, we obtain that

$$P(W \leq s) \leq \frac{1}{b}. \tag{3.1}$$

By the condition $w_0 \leq \frac{b}{\sigma^3}\left(\sum_{i=1}^{n} p_i^2\right)^2$, we obtain that

$$P(W \leq s) \leq \frac{1}{w_0\sigma^3}\left(\sum_{i=1}^{n} p_i^2\right)^2. \tag{3.2}$$

1. Assume that $w_0 > s$. Since $Z - s \sim \mathbf{Po}(b)$,

$$\begin{aligned}
&|P(W \leq w_0) - P(Z \leq w_0)| \\
&= |P(W - s \leq w_0 - s) - P(Z - s \leq w_0 - s)| \\
&= |P(W - s \leq w_0 - s) - P(\widetilde{Z} \leq w_0 - s)| \\
&\leq |P(0 < W - s \leq w_0 - s) - P(\widetilde{Z} \leq w_0 - s))| + P(W - s \leq 0) \\
&= |R_1| + R_2,
\end{aligned} \tag{3.3}$$

where $R_1 = P(0 < W - s \leq w_0 - s) - P(\widetilde{Z} \leq w_0 - s)$,

$R_2 = P(W - s \leq 0)$,

and $\widetilde{Z} = Z - s$.

By (3.1) and (3.2), we immediately obtain that

$$R_2 \leq \frac{1}{w_0\sigma^3}\left(\sum_{i=1}^{n} p_i^2\right)^2. \tag{3.4}$$

In order to derive an upper bound for $|R_1|$, we can express the term involving R_1 in a similar manner as demonstrated in the proof by Röllin [10], using Stein's method of exchangeable pair for the Poisson distribution with a parameter b. This method is initiated by Stein's equation, defined for a given function h as:

$$h(w) - \mathbf{Po}(b)(h) = bg(w+1) - wg(w), \tag{3.5}$$

where $\mathbf{Po}(b)(h) = e^{-b}\sum_{l=0}^{\infty} h(l)\frac{b^l}{l!}$ and g is a bounded real–valued functions defined on $\mathbb{N}$ depending on the given function h. In this case, we consider $h_{w_0} : \mathbb{N} \to \mathbb{R}$ defined by

$$h_{w_0}(w) = \begin{cases} 1, & \text{if } 0 < w \leq w_0 - s, \\ 0, & \text{if } w > w_0 - s. \end{cases}$$

Following [2], g_{w_0} which is the solution of (3.5) for the above function h_{w_0} is

$$g_{w_0}(w) = \begin{cases} (w-1)!b^{-w}e^b[\mathbf{Po}(b)(h_{w_0})\mathbf{Po}(b)(1-h_{w-1})], & \text{if } w > w_0 - s, \\ (w-1)!b^{-w}e^b[\mathbf{Po}(b)(h_{w-1})\mathbf{Po}(b)(1-h_{w_0})], & \text{if } w \leq w_0 - s. \end{cases} \tag{3.6}$$

Let $\Delta g_{w_0}(w) = g_{w_0}(w+1) - g_{w_0}(w)$. From (3.6), it follows that

$$\Delta g_{w_0}(w) = \begin{cases} \frac{(w-1)!e^b\mathbf{Po}(b)(h_{w_0})[w\mathbf{Po}(b)(1-h_w) - b\mathbf{Po}(b)(1-h_{w-1})]}{b^{w+1}}, & \text{if } w \geq w_0 - s + 1, \\ \frac{(w-1)!e^b\mathbf{Po}(b)(1-h_{w_0})[w\mathbf{Po}(b)(h_{w_0}) - b\mathbf{Po}(b)(h_{w-1})]}{b^{w+1}}, & \text{if } w \leq w_0 - s. \end{cases}$$

According to Lemma 2.2 in [13], for $w \geq 1$, it holds that

$$|\Delta g_{w_0}(w)| \leq b^{-1}\min\left\{1 - e^{-b}, \frac{2b^{-1}(e^b - b - 1)}{w_0 - s + 1}, \frac{b}{w_0 - s}\right\}. \tag{3.7}$$

From (3.5), we obtain

$$\begin{aligned} R_1 &= b\mathbb{E}(g_{w_0}(W - s + 1)) - \mathbb{E}((W - s)g_{w_0}(W - s)) \\ &= (\sigma^2 + \gamma)\mathbb{E}(g_{w_0}(W - s + 1)) - (\sigma^2 + \gamma)\mathbb{E}(g_{w_0}(W - s)) + (\sigma^2 + \gamma)\mathbb{E}(g_{w_0}(W - s)) \\ &\quad - \mathbb{E}((W - \lambda + \sigma^2 + \gamma)g_{w_0}(W - s)) \\ &= \mathbb{E}\left(\sigma^2\Delta g_{w_0}(W - s) - (W - \lambda)g_{w_0}(W - s)\right) + \mathbb{E}\left(\gamma\Delta g_{w_0}(W - s)\right) \\ &=: K_1 + K_2. \end{aligned} \tag{3.8}$$

By using (3.7), we get

$$|K_2| \leq \gamma\mathbb{E}|\Delta g_{w_0}(W - s)| \leq b^{-1}\min\left\{1 - e^{-b}, \frac{2b^{-1}(e^b - b - 1)}{w_0 - s + 1}, \frac{b}{w_0 - s}\right\}. \tag{3.9}$$

To establish a bound for K_1, we initiate the process by modifying the proof as originally presented in [10], p.1603. From (1.6), we have $\mathbb{E}^W(W' - \lambda) = \left(1 - \frac{1}{n}\right)(W - \lambda)$. Define the anti-symmetric function $F(w, w') := (w' - w)(g_{w_0}(w' - s) + g_{w_0}(w - s))$. By using exchangeability, we can see that $\mathbb{E}^W(W' - W) = -\frac{1}{n}(W - \lambda)$ and $\mathbb{E}F(W - W') = 0$. This implies that

$$\begin{aligned} 0 &= \mathbb{E}F(W - W') \\ &= \mathbb{E}\left\{(W' - W)\left(2g_{w_0}(W - s) + g_{w_0}(W' - s) - g_{w_0}(W - s)\right)\right\} \\ &= -\frac{2}{n}\mathbb{E}\left\{(W - \lambda)g_{w_0}(W - s)\right\} + \mathbb{E}\left\{(W' - W)\left(g_{w_0}(W' - s) - g_{w_0}(W - s)\right)\right\}. \end{aligned}$$

By using exchangeability, we observe that

$$\begin{aligned}\mathbb{E}\left\{\mathbb{I}(W'-W=-1)\Delta g_{w_0}(W-s-1)\right\} &= \mathbb{E}\left\{\mathbb{I}(W-W'=1)\Delta g_{w_0}(W'-s)\right\}\\ &= \mathbb{E}\left\{\mathbb{I}(W'-W=1)\Delta g_{w_0}(W-s)\right\},\end{aligned}$$

which implies that

$$\begin{aligned}&\mathbb{E}\left\{(W'-W)\left(g_{w_0}(W'-s)-g_{w_0}(W-s)\right)\right\}\\ &= \mathbb{E}\left\{\mathbb{I}(W'-W=1)\Delta g_{w_0}(W-s)\right\}+\mathbb{E}\left\{\mathbb{I}(W'-W=-1)\Delta g_{w_0}(W-s-1)\right\}\\ &= 2\mathbb{E}\left\{\mathbb{I}(W'-W=1)\Delta g_{w_0}(W-s)\right\}.\end{aligned}$$

Now, we obtain

$$\mathbb{E}\left((W-\lambda)g_{w_0}(W-s)\right) = n\mathbb{E}\left(\mathbb{I}(W'-W=1)\Delta g_{w_0}(W-s)\right).$$

This implies that

$$\begin{aligned}|K_1| &= \left|\mathbb{E}\left(\left(\mathbb{I}\left(W'-W=1\right)n-\sigma^2\right)\Delta g_{w_0}\left(W-s\right)\right)\right|\\ &\leq \mathbb{E}|\Delta g_{w_0}(W-s)|n\sqrt{\mathrm{Var}S},\end{aligned} \tag{3.10}$$

where $S := \mathbb{E}^W\mathbb{I}(W'-W=1) = P(W'=W+1|W)$. To bound $\mathrm{Var}S$, we introduce a sequence of random variables $X=(X_1,\ldots,X_n)$ that satisfies the properties specified in (1.1). We then consider

$$S^* := E^X\mathbb{I}(W'-W=1) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}^X(X_i=0,X_i^*=1) = \frac{1}{n}\sum_{i=1}^{n}(1-X_i)p_i.$$

Note that $\mathrm{Var}S^* = n^{-2}\sum_{i=1}^{n}p_i^3q_i$. As X is a random variable with corresponding σ-algebras satisfying $\sigma(W)\subset\sigma(X)$, it follows that $\mathrm{Var}S\leq\mathrm{Var}S^*$. From (3.7) and (3.10), we obtain

$$|K_1| \leq \left(b^{-1}\sqrt{\sum_{i=1}^{n}p_i^3q_i}\right)\min\left\{1-e^{-b},\frac{2b^{-1}(e^b-b-1)}{w_0-s+1},\frac{b}{w_0-s}\right\}. \tag{3.11}$$

By combining equations (3.3), (3.4), (3.8), (3.9) and (3.11), we obtain (1.8) as required.

2. Assume that $w_0=s$. We derive the target distribution to

$$\begin{aligned}|P(W\leq w_0)-P(Z\leq w_0)| &= |P(W\leq s)-P(\widetilde{Z}\leq w_0-s)|\\ &\leq P(W\leq s)+P(\widetilde{Z}=0).\end{aligned}$$

By the fact that $\widetilde{Z}$ is a Poisson random variable with parameter b and (3.1), we obtain that

$$|P(W\leq w_0)-P(Z\leq w_0)| \leq \frac{1}{b}+e^{-b}\leq\frac{2}{b}.$$

Therefore, (1.9) holds by utilizing (3.2).

3. Suppose that $w_0<s$. Since $w_0-s<0$ and $\widetilde{Z}\sim\mathbf{Po}(b)$, $P(\widetilde{Z}\leq w_0-s)=0$. By (3.1) and (3.2), we obtain that

$$\begin{aligned}|P(W\leq w_0)-P(Z\leq w_0)| &= |P(W\leq w_0)-P(\widetilde{Z}\leq w_0-s)|\\ &= P(W\leq w_0)\\ &\leq P(W<s)\\ &\leq \frac{1}{w_0\sigma^3}\left(\sum_{i=1}^{n}p_i^2\right)^2.\end{aligned}$$

This results in (1.10), thus concluding the proof. □

Acknowledgment. The authors would like to thank the reviewers for their valuable comments and suggestions.

References

[1] A. D. Barbour and P. Hall, *On the rate of Poisson convergence*, Mathematical Proceedings of the Cambridge Philosophical Society **95** (1984), no. 3, 473–480.

[2] A. D. Barbour, L. Holst, and S. Janson, *Poisson approximation*, Oxford University Press, 1992.

[3] A. D. Barbour and V. Čekanavičius, *Total variation asymptotics for sums of independent integer random variables*, The Annals of Probability **30** (2002), no. 2, 509–545.

[4] A. D. Barbour and A. Xia, *Poisson perturbations*, ESAIM: Probability and Statistics **3** (1999), 131–150.

[5] L. H. Y. Chen, *Poisson Approximation for Dependent Trials*, The Annals of Probability **3** (1975), no. 3, 534–545.

[6] J. Kruopis, *Precision of approximation of the generalized binomial distribution by convolutions of Poisson measures*, Lithuanian Mathematical Journal **26** (1986), 37–49.

[7] S. Y. Novak, *On the accuracy of Poisson approximation*, Extremes **22** (2019), 729–748.

[8] ———, *Poisson approximation*, Probability Surveys **16** (2019), no. none, 228–276.

[9] Y. Rinott and V. Rotar, *On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted U-statistics*, The Annals of Applied Probability **7** (1997), no. 4, 1080–1105.

[10] A. Röllin, *Translated Poisson approximation using exchangeable pair couplings*, The Annals of Applied Probability **17** (2007), no. 5-6, 1596–1614.

[11] C. Stein, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Berkeley, Calif.) (M. Lucien, Le Cam, Jerzy Neyman, and Elizabeth L. Scott, eds.), vol. 2, University of California Press, 1972, pp. 583–602.

[12] ———, *Approximate computation of expectations*, Institute of Mathematical Statistics, Hayward, USA, 1986.

[13] K. Teerapabolarn, *An improvement of poisson approximation for sums of dependent bernoulli random variables*, Communications in Statistics - Theory and Methods **43** (2014), no. 8, 1758–1777.

[14] V. Čekanavičius, *Poisson approximations for sequences of random variables*, Statistics & Probability Letters **39** (1998), no. 2, 101–107.

[15] V. Čekanavičius and P. Vaitkus, *Centered Poisson approximation via Stein's method*, Lithuanian Mathematical Journal **41** (2001), 319–329.

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

การแจกแจงความน่าจะเป็นของความเร็วลมในพื้นที่ที่มีศักยภาพในการตั้งฟาร์มลม: ความเร็วลม*

วนิดา พงษ์ศักดิ์ชาติ[1,‡] และ พรหมพร ธรรมสาร[1,†]
[1]ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา

บทคัดย่อ

พลังงานไฟฟ้ามีความสำคัญอย่างมากทั้งในแง่ของการดำรงชีวิต และคุณภาพชีวิต นอกจากนั้นไฟฟ้ายังเป็นปัจจัยสำคัญในการพัฒนาประเทศทั้งทางคมนาคม เศรษฐกิจ อุตสาหกรรม เกษตรกรรม และการบริการ ซึ่งการผลิตไฟฟ้าจำเป็นต้องใช้เชื้อเพลิงในการผลิต พลังงานลมเป็นแหล่งพลังงานหมุนเวียนที่ใช้ผลิตไฟฟ้า โดยเป็นพลังงานที่ใช้แล้วไม่หมดไป อีกทั้งยังเป็นพลังงานสะอาดไม่ก่อให้เกิดมลพิษกับสิ่งแวดล้อม งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อศึกษาหาการแจกแจงความน่าจะเป็นที่เหมาะสมกับข้อมูลความเร็วลมเฉลี่ยรายวันเพื่อประเมินศักยภาพในการตั้งฟาร์มลมเพื่อผลิตไฟฟ้าพลังงานลม โดยศึกษาในพื้นที่ 4 จังหวัด คือ ขอนแก่น อุบลราชธานี ชัยภูมิ และนครราชสีมา ใช้ข้อมูลความเร็วลมเฉลี่ยรายวันจากสถานีตรวจอากาศกรมอุตุนิยม-วิทยา จำนวน 6 สถานี ที่ตั้งอยู่ในจังหวัดเหล่านี้ การแจกแจงความน่าจะเป็นที่นำมาศึกษา 7 ชนิด คือ การแจกแจงปรกติ การแจกแจงไวบูล การแจกแจงล็อกนอร์มัล การแจกแจงแกมมา การแจกแจงปรกติแบบผสม การแจกแจงไวบูลแบบผสม และการแจกแจงแกมมาแบบผสม จากการศึกษาพบว่าการแจกแจงล็อกนอร์มัลเป็นการแจกแจงความน่าจะเป็นที่เหมาะสมกับข้อมูลความเร็วลมเฉลี่ยรายวันของสถานีตรวจอากาศขอนแก่น สถานีอุตุนิยมวิทยาชัยภูมิ และสถานีตรวจอากาศอุบลราชธานี ส่วนสถานีอุตุนิยมวิทยาเกษตรอุบลราชธานี และสถานีอุตุนิยมวิทยานครราชสีมา การแจกแจงความน่าจะเป็นที่เหมาะสมคือการแจกแจงปรกติ สำหรับสถานีอุตุนิยมวิทยาท่าพระการแจกแจงความน่าจะเป็นที่เหมาะสมคือการแจกแจงไวบูลแบบผสม นอกจากนั้นพื้นที่ในบริเวณสถานีเหล่านี้ยังเป็นพื้นที่ที่มีศักยภาพในการตั้งฟาร์มลมเพื่อผลิตไฟฟ้าเนื่องจากมีความเร็วลมเฉลี่ยรายวันอยู่ในช่วงที่เหมาะสม

†ผู้นำเสนอ ‡ผู้แต่งหลัก
อีเมล: vanida@buu.ac.th (วนิดา พงษ์ศักดิ์ชาติ), 63030176@go.buu.ac.th (พรหมพร ธรรมสาร).

คำสำคัญ: พลังงานลม, ความเร็วลม, การแจกแจงความน่าจะเป็น, การแจกแจงความน่าจะเป็นแบบผสม
2020 MSC: 62P12

1 บทนำ

พลังงานไฟฟ้ามีความสำคัญอย่างมากทั้งในแง่ของการดำรงชีวิต คุณภาพชีวิต และการอำนวยความสะดวกในการดำรงชีวิต นอกจากนั้นไฟฟ้ายังเป็นปัจจัยสำคัญในการพัฒนาประเทศทั้งทางคมนาคม เศรษฐกิจ อุตสาหกรรม เกษตรกรรม และการบริการ ซึ่งการผลิตไฟฟ้าจำเป็นต้องใช้เชื้อเพลิงในการผลิต ปัจจุบันการผลิตไฟฟ้าในระบบของการไฟฟ้าฝ่ายผลิตห่งประเทศไทย (กฟผ.) มีการใช้เชื้อเพลิงอยู่หลายประเภท ได้แก่ ก๊าซธรรมชาติ ถ่านหิน และพลังงานหมุนเวียน (เช่น พลังน้ำ พลังลม และพลังแสงอาทิตย์) โดยเชื้อเพลิงที่ใช้มากที่สุดคือ ก๊าซธรรมชาติ รองลงมาคือถ่านหิน และพลังงานหมุนเวียน ตามลำดับ และเป็นที่ทราบกันดีว่าความต้องการพลังงานไฟฟ้านั้นมีเพิ่มขึ้นในทุก ๆ ปี ทำให้มีความจำเป็นในการใช้เชื้อเพลิงเพิ่มขึ้นด้วยเช่นเดียวกัน อย่างไรก็ตามก๊าซธรรมชาติ และถ่านหิน (เชื้อเพลิงฟอสซิล) ซึ่งเป็นเชื้อเพลิงหลักในการผลิตไฟฟ้ามีปริมาณลดลงเรื่อย ๆ สวนทางกับความต้องการในการใช้ไฟฟ้าในปัจจุบันและในอนาคต ประเทศไทยจึงมีนโยบายและการวางแผนในการพัฒนาพลังงานหมุนเวียนอื่น ๆ มาใช้เชื้อเพลิงฟอสซิล เช่น พลังงานแสงอาทิตย์ พลังงานน้ำ พลังงานชีวมวล และพลังงานลม เป็นต้น

พลังงานลมเป็นหนึ่งในพลังงานหมุนเวียนที่ใช้ผลิตไฟฟ้า และเป็นพลังงานสะอาดที่มีอยู่ตามธรรมชาติ ใช้แล้วไม่หมดไป อีกทั้งยังไม่สร้างมลพิษแก่สิ่งแวดล้อม ถือได้ว่าเป็นพลังงานที่มีประโยชน์แก่ประเทศ ไม่ว่าจะเป็นลดการใช้เชื้อเพลิงฟอสซิลที่ต้องนำเข้ามาจากต่างประเทศ และก่อให้เกิดมลพิษกับสิ่งแวดล้อมอีกด้วย อย่างไรก็ตามลมที่จะสามารถนำมาใช้ผลิตไฟฟ้าได้ต้องมีความเร็วอยู่ในช่วงที่เหมาะสม ดังนั้นการที่จะตั้งฟาร์มลม (wind farm) เพื่อผลิตไฟฟ้าพลังงานลมในพื้นใดจำเป็นต้องมีการประเมินศักยภาพของพลังงานลมในพื้นที่นั้น ๆ ก่อน [4, 5]

การพัฒนาพลังงานลมของประเทศไทยได้เริ่มต้นจากที่การไฟฟ้าฝ่ายผลิตแห่งประเทศไทยได้ติดตั้งกังหันลมตัวแรกที่มีขนาด 150 กิโลวัตต์ ที่แหลมพรหมเทพ จังหวัดภูเก็ต ในปี พ.ศ. 2539 ที่มีความเร็วเฉลี่ยรายปี 5 เมตรต่อวินาที ส่วนการติดตั้งกังหันลมผลิตไฟฟ้าในเชิงพาณิชย์แห่งแรก ติดตั้งที่ยอดเขายายเที่ยงเหนือเขื่อนลำตะคอง จังหวัดนครราชสีมา โดยการไฟฟ้าฝ่ายผลิตแห่งประเทศไทยในปี พ.ศ. 2552 ซึ่งได้ติดตั้งกังหันลมจำนวน 2 ตัว ขนาด 1,250 กิโลวัตต์ ในบริเวณนั้นมีความเร็วลมเฉลี่ยที่ 6.7 เมตรต่อวินาที [6] นอกจากนั้นยังพบว่าในบริเวณภาคตะวันออกเฉียงเหนือ ภาคใต้ และบริเวณนอกชายฝั่งทะเลฝั่งอ่าวไทย เป็นบริเวณที่มีศักยภาพในการผลิตไฟฟ้าจากพลังงานลม การตั้งฟาร์มลมเพื่อผลิตไฟฟ้าในประเทศไทยมี 2 รูปแบบ คือการติดตั้งกังหันลมบนบก ซึ่งส่วนใหญ่เป็นพื้นที่บนภูเขาสูงและชายฝั่งทะเล มีความเร็วลมเฉลี่ยรายวันอยู่ที่ 6 - 7 เมตรต่อวินาที ที่ระดับความสูง 50 เมตร และการติดตั้งกังหันลมนอกชายฝั่งซึ่งจะเป็นพื้นที่ที่เหมาะสมในการตั้งฟาร์มลมเนื่องจากไม่มีสิ่งกีดขวาง พื้นที่กว้างขวางมีกำลังลมที่แรงและสม่ำเสมอตลอดทั้งปี พบว่าเป็นพื้นที่ที่มีศักยภาพคือพื้นที่ที่ความเร็วลมเฉลี่ยรายวัน 6.5 เมตรต่อวินาที [1]

จากการศึกษาที่ผ่านมาได้มีนักวิจัยที่นำวิธีเชิงสถิติคือการแจกแจงความน่าจะเป็นของความเร็วลมมาใช้ในการประเมินศักยภาพของพื้นที่ที่มีศักยภาพในการผลิตไฟฟ้าด้วยพลังงานลมจากฟาร์มลม ซึ่งการแจกแจงความน่าจะเป็นที่เหมาะสมกับความเร็วลมในพื้นที่ต่าง ๆ ที่มีการวิจัยผ่านมา ได้แก่ การแจกแจงไวบูล (Weibull distribution) การแจกแจงแกมมา (Gamma distribution) [7] การแจกแจงไวบูลแบบผสม (Mixture Weibull distribution) การแจกแจงล็อกนอร์มอล (Log-normal distribution) [9] นอกจากนั้นในการศึกษาของ Koca et al. [8] ในปี 2019 ยังพบว่าการแจกแจงความน่าจะเป็นผสมมีความเหมาะสมกับความเร็วลมมากกว่าการแจกแจงฐานนิยมเดี่ยว

ในประเทศไทยยังมีพื้นที่อีกหลายแห่งที่มีสภาพแวดล้อมเหมาะต่อการตั้งฟาร์มลม การศึกษานี้จึงมีวัตถุประสงค์ที่จะศึกษาการแจกแจงความน่าจะเป็นของความเร็วลมและประเมินศักยภาพในการตั้งฟาร์มลมในพื้นที่จังหวัดชัยภูมิ นครราชสีมา อุบลราชธานี และขอนแก่น ทั้งนี้เนื่องจากจังหวัดชัยภูมิ และนครราชสีมา มีการติด

ตั้งกังหันลมผลิตกระแสไฟฟ้าอยู่แล้ว สำหรับจังหวัดอุบลราชธานี และขอนแก่นเป็นจังหวัดที่อยู่ในภาคตะวันออกเฉียงเหนือซึ่งเป็นภูมิภาคที่เป็นพื้นที่ราบสูงและใกล้แม่น้ำโขงซึ่งจะทำให้ได้รับอิทธิพลจากร่องเขาด้วย จึงเป็นพื้นที่ที่น่าสนใจในการตั้งฟาร์มลม

2 ความรู้พื้นฐาน

2.1 ลม

ลมเป็นแหล่งพลังงานสะอาดชนิดหนึ่งที่มีอยู่ตามธรรมชาติสามารถใช้ได้อย่างไม่มีวันหมดสิ้น ซึ่งการใช้ลมเป็นแหล่งพลังงานในการผลิตไฟฟ้าต้องพิจารณาถึงความเร็วลมในพื้นที่นั้นต้องอยู่ในช่วงที่เหมาะสม การนำพลังลมมาใช้ประโยชน์จะต้องอาศัยกังหันลมในการเปลี่ยนพลังงานจลน์จากการเคลื่อนที่ของลมไปเป็นพลังงานกลก่อนนำไปใช้ประโยชน์ โดยพื้นที่ที่มีความเร็วลม 2 - 5 เมตรต่อวินาที ควรติดตั้งกังหันลมขนาด 200 วัตต์ พื้นที่ที่มีความเร็วลม 5 - 7 เมตรต่อวินาที ควรติดตั้งกังหันลมขนาด 500 วัตต์ และ พื้นที่ที่มีความเร็วลม 7 - 12 เมตรต่อวินาที ควรติดตั้งกังหันลมขนาด 1000 วัตต์ [2, 3]

2.2 การแจกแจงความน่าจะเป็น

การแจกแจงความน่าจะเป็นที่นำมาศึกษาเพื่อหาการแจกแจงความน่าจะเป็นที่เหมาะสมของความเร็วลมในพื้นที่ที่ศึกษามีจำนวน 7 ชนิด ได้แก่

1. **การแจกแจงปรกติ (Normal distribution)**

   เมื่อตัวแปรสุ่ม X มีการแจกแจงปรกติ เขียนแทนด้วย $X \sim N(\mu, \sigma^2)$ [8] มีฟังก์ชันความหนาแน่นความน่าจะเป็น

   $$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{2.1}$$

   โดยที่ μ คือ พารามิเตอร์บ่งตำแหน่ง และ σ^2 คือ พารามิเตอร์บ่งรูปร่าง และมีฟังก์ชันความน่าจะเป็นสะสม คือ

   $$F(x;\mu,\sigma) = \frac{1}{2}\left[1 + erf(\frac{x-\mu}{\sqrt{2\sigma^2}})\right] \tag{2.2}$$

2. **การแจกแจงไวบูล (Weibull distribution)**

   เมื่อตัวแปรสุ่ม X มีการแจกแจงไวบูล เขียนแทนด้วย $X \sim W(k, c)$ [8] มีฟังก์ชันความหนาแน่นความน่าจะเป็น

   $$f(x;k,c) = \frac{k}{c}\left(\frac{x}{c}\right)^{k-1} exp\left[-\left(\frac{x}{c}\right)^{k}\right] \tag{2.3}$$

   โดยที่ k คือ พารามิเตอร์บ่งรูปร่าง และ c คือ พารามิเตอร์บ่งขนาด มีฟังก์ชันความน่าจะเป็นสะสม คือ

   $$F(x;k,c) = 1 - exp\left[-\left(\frac{x}{c}\right)^{k}\right] \tag{2.4}$$

3. **การแจกแจงล็อกนอร์มัล**

   เมื่อตัวแปรสุ่ม X มีการแจกแจงล็อกนอร์มอล เขียนแทนด้วย $X \sim Lognormal(\mu, \sigma^2)$ [8] มีฟังก์ชันความหนาแน่นความน่าจะเป็น

   $$f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \tag{2.5}$$

โดยที่ μ คือ พารามิเตอร์บ่งตำแหน่ง และσ^2 คือ พารามิเตอร์บ่งรูปร่าง และมีฟังก์ชันความน่าจะเป็นสะสม คือ

$$F(x;\mu,\sigma) = \frac{1}{2}\left[1 + erf\left(\frac{(\ln x - \mu)}{\sigma\sqrt{2}}\right)\right] \tag{2.6}$$

4. **การแจกแจงแกมมา (Gamma distribution)**

เมื่อตัวแปรสุ่ม X มีการแจกแจงแกมมา เขียนแทนด้วย $X \sim \Gamma(\alpha,\beta)$ [8] มีฟังก์ชันความหนาแน่นความน่าจะเป็น

$$f(x;\alpha,\beta) = \frac{x^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}} exp\left[-\frac{x}{\beta}\right] \tag{2.7}$$

โดยที่ α คือ พารามิเตอร์บ่งรูปร่าง , $\alpha > 0$ และ β คือ พารามิเตอร์บ่งขนาด , $\beta > 0$ และมีฟังก์ชันความน่าจะเป็นสะสม คือ

$$F(x;\alpha,\beta) = \int \frac{x^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}} exp\left[-\frac{x}{\beta}\right] dx \tag{2.8}$$

5. **การแจกแจงปรกติแบบผสม (Mixture normal distribution)**

เมื่อตัวแปรสุ่ม X มีการแจกแจงปรกติแบบผสม เขียนแทนด้วย $X \sim NN(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2)$ [8] มีฟังก์ชันความหนาแน่นความน่าจะเป็น

$$f(x;p,\mu_1,\sigma_1,\mu_2,\sigma_2) = p\frac{1}{\sigma_1\sqrt{2\pi}} exp\left[\frac{-(x-\mu_1)^2}{2\sigma_1^2}\right] + (1-p)\frac{1}{\sigma_2\sqrt{2\pi}} exp\left[\frac{-(x-\mu_2)^2}{2\sigma_2^2}\right] \tag{2.9}$$

โดยที่ p คือ พารามิเตอร์ถ่วงน้ำหนัก μ_1,μ_2 คือ พารามิเตอร์บ่งตำแหน่ง และ σ_1^2,σ_2^2 คือ พารามิเตอร์บ่งรูปร่าง และมีฟังก์ชันความน่าจะเป็นสะสม คือ

$$F(x;p,\mu_1,\sigma_1,\mu_2,\sigma_2) = \frac{p}{2}\left[1 + erf\left(\frac{x-\mu_1}{\sigma_1\sqrt{2}}\right)\right] + \frac{(1-p)}{2}\left[1 + erf\left(\frac{x-\mu_2}{\sigma_2\sqrt{2}}\right)\right] \tag{2.10}$$

6. **การแจกแจงไวบูลแบบผสม (Mixture Weibull distribution)**

เมื่อตัวแปรสุ่ม X มีการแจกแจงไวบูลแบบผสม เขียนแทนด้วย $X \sim WW(p,k_1,c_1,k_2,c_2)$ [8] มีฟังก์ชันความหนาแน่นความน่าจะเป็น

$$f(x;p,k_1,c_1,k_2,c_2) = p\frac{k_1}{c_1}\left(\frac{x}{c_1}\right)^{k_1-1} exp\left[-\left(\frac{x}{c_1}\right)^{k_1}\right] + (1-p)\frac{k_2}{c_2}\left(\frac{x}{c_2}\right)^{k_2-1} exp\left[-\left(\frac{x}{c_2}\right)^{k_2}\right] \tag{2.11}$$

โดยที่ p คือ พารามิเตอร์ถ่วงน้ำหนัก k_1,k_2 คือ พารามิเตอร์บ่งรูปร่าง และ c_1,c_2 คือ พารามิเตอร์บ่งขนาด และมีฟังก์ชันความน่าจะเป็นสะสม

$$F(x;p,k_1,c_1,k_2,c_2) = p\left\{1 - exp\left[-\left(\frac{x}{c_1}\right)^{k_1}\right]\right\} + (1-p)\left\{1 - exp\left[-\left(\frac{x}{c_2}\right)^{k_2}\right]\right\} \tag{2.12}$$

7. **การแจกแจงแกมมาแบบผสม (Mixture gamma distribution)**

เมื่อตัวแปรสุ่ม X มีการแจกแจงแกมมาแบบผสม เขียนแทนด้วย $X \sim GG(p,\alpha_1,\beta_1,\alpha_2,\beta_2)$ [8] มี

ฟังก์ชันความหนาแน่นความน่าจะเป็น

$$f(x;p,\alpha_1,\beta_1,\alpha_2,\beta_2) = p\frac{x^{\alpha_1-1}}{\Gamma(\alpha_1)\beta_1^{\alpha_1}}exp\left[-\frac{x}{\beta_1}\right] + (1-p)\frac{x^{\alpha_2-1}}{\Gamma(\alpha_2)\beta_2^{\alpha_2}}exp\left[-\frac{x}{\beta_2}\right] \quad (2.13)$$

โดยที่ p คือ พารามิเตอร์ถ่วงน้ำหนัก α_1, α_2 คือ พารามิเตอร์แสดงรูปร่าง และ β_1, β_2 คือ พารามิเตอร์แสดงขนาด และมีฟังก์ชันความน่าจะเป็นสะสม คือ

$$F(x;p,\alpha_1,\beta_1,\alpha_2,\beta_2) = p\int \frac{x^{\alpha_1-1}}{\Gamma(\alpha_1)\beta_1^{\alpha_1}}exp\left[-\frac{x}{\beta_1}\right]dv + (1-p)\int \frac{x^{\alpha_2-1}}{\Gamma(\alpha_2)\beta_2^{\alpha_2}}exp\left[-\frac{x}{\beta_2}\right]dx \quad (2.14)$$

3 วิธีการศึกษา

ในการศึกษาการแจกแจงความน่าจะเป็นที่เหมาะสมกับความเร็วลม เพื่อประเมินศักยภาพในการผลิตไฟฟ้าในพื้นที่ที่ศึกษา มีขั้นตอนการศึกษาดังนี้

1. **ข้อมูลความเร็วลม**

   งานวิจัยครั้งนี้ได้รับความอนุเคราะห์ข้อมูลความเร็วลมจากกรมอุตุนิยมวิทยา ข้อมูลที่ศึกษาเป็นข้อมูลความเร็วลมเฉลี่ยรายวัน ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2560 จนถึง 21 กันยายน พ.ศ. 2566 ใน 4 จังหวัด จำนวน 6 สถานี ดังนี้

   - 381201: สถานีตรวจอากาศขอนแก่น(ขอนแก่น)
   - 381301: สถานีอุตุนิยมวิทยาเกษตรท่าพระ (ขอนแก่น)
   - 403201: สถานีอุตุนิยมวิทยาชัยภูมิ (ชัยภูมิ)
   - 407301: สถานีอุตุนิยมวิทยาเกษตรอุบลราชธานี (อุบลราชธานี)
   - 407501: สถานีตรวจอากาศอุบลราชธานี (อุบลราชธานี)
   - 431301: สถานีอุตุนิยมวิทยานครราชสีมา (นครราชสีมา)

   เมื่อได้ข้อมูลความเร็วลมเฉลี่ยรายวันแล้ว ได้แบ่งข้อมูลออกเป็น 2 ส่วน ดังนี้

   - ข้อมูลในวันที่ 1 มกราคม พ.ศ. 2560 - 20 กันยายน พ.ศ. 2565 เป็นชุดข้อมูลฝึกฝน (training data set) จะถูกใช้เพื่อหาการแจกแจงความน่าจะเป็นที่เหมาะสมของข้อมูล
   - ข้อมูลในวันที่ 21 กันยายน พ.ศ. 2565 - 21 กันยายน พ.ศ. 2566 เป็นชุดข้อมูลตรวจสอบ (test data set) จะถูกใช้เพื่อประมาณความคลาดเคลื่อนของการทำนายของการแจกแจงความน่าจะเป็นที่เลือกได้จากการใช้ข้อมูลในส่วนที่หนึ่ง

   โดยจำนวนข้อมูลความเร็วลมเฉลี่ยรายวันของแต่ละสถานีที่ใช้ในการศึกษาแสดงในตารางที่ 1

2. นำชุดข้อมูลฝึกฝนมาประมาณค่าพารามิเตอร์ของการแจกแจงความน่าจะเป็นที่ศึกษาทั้งหมด 7 ชนิด ได้แก่ การแจกแจงปรกติ การแจกแจงไวบูล การแจกแจงล็อกนอร์มอล การแจกแจงแกมมา การแจกแจงปรกติแบบผสม การแจกแจงไวบูลแบบผสม และการแจกแจงแกมมาแบบผสม ด้วยวิธีภาวะน่าจะเป็นสูงสุด (Maximum likelihood method) โดยใช้โปรแกรม R และโปรแกรมสำเร็จ mixR fitdistrplus และ extraDistr

ตารางที่ 1: จำนวนข้อมูลความเร็วลมเฉลี่ยรายวันของชุดข้อมูลฝึกฝนและชุดข้อมูลตรวจสอบ

สถานี	จำนวนข้อมูล (วัน)	
	ชุดข้อมูลฝึกฝน	ชุดข้อมูลตรวจสอบ
สถานีตรวจอากาศขอนแก่น	2038	361
สถานีอุตุนิยมวิทยาเกษตรท่าพระ	2087	364
สถานีอุตุนิยมวิทยาชัยภูมิ	2073	360
สถานีอุตุนิยมวิทยาเกษตรอุบลราชธานี	2017	366
สถานีตรวจอากาศอุบลราชธานี	2087	365
สถานีอุตุนิยมวิทยานครราชสีมา	2085	365

3. พิจารณาการแจกแจงความน่าจะเป็นที่เหมาะสมกับข้อมูลความเร็วลมเฉลี่ยรายวันที่เป็นข้อมูลชุดฝึกฝน ด้วยการทดสอบแอนเดอร์สัน-ดาร์ลิง (โดยใช้ค่าประมาณพารามิเตอร์จากข้อ 2) หากการแจกแจงความน่าจะเป็นชนิดใดมีค่าพี (p-value) ของการทดสอบมากกว่าระดับนัยสำคัญ 0.05 แสดงว่าการแจกแจงที่พิจารณาเป็นการแจกแจงที่เหมาะสมกับชุดข้อมูล

4. นำการแจกแจงความน่าจะเป็นและค่าประมาณพารามิเตอร์ของการแจกแจงนั้นที่ได้จากข้อ 3 มาตรวจสอบกับชุดข้อมูลทดสอบ และประเมินความเหมาะสมของการแจกแจงความน่าจะเป็นกับชุดข้อมูลทดสอบด้วยค่า

   - รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root mean square error: RMSE) [8]

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(\widehat{F}(x_i) - \frac{i}{n+1}\right)^2}{n}} \tag{3.1}$$

   เมื่อ
   $\widehat{F}(x_i)$ คือ ตัวประมาณค่าฟังก์ชันความน่าจะเป็นสะสม
   x_i คือ ความเร็วลมเฉลี่ยรายวัน ณ เวลาที่ i
   n คือ จำนวนข้อมูลความเร็วลมเฉลี่ยรายวัน

   - ค่าสัมประสิทธิ์การกำหนด (Coefficient of determination: R^2) โดย $0 \leq R^2 \leq 1$ [8]

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(\widehat{F}(x_i) - \frac{i}{n+1}\right)^2}{\sum_{i=1}^{n}\left(\widehat{F}(x_i) - \bar{\widehat{F}}(x_i)\right)^2} \tag{3.2}$$

   โดยที่

$$\bar{\widehat{F}}(x_i) = \sum_{i=1}^{n}\frac{\widehat{F}(x_i)}{n} \tag{3.3}$$

   เมื่อ
   $\widehat{F}(x_i)$ คือ ตัวประมาณค่าฟังก์ชันความน่าจะเป็นสะสม
   x_i คือ ความเร็วลมเฉลี่ยรายวัน ณ เวลาที่ i
   n คือ จำนวนข้อมูลความเร็วลมเฉลี่ยรายวัน

หากการแจกแจงความน่าจะเป็นใดมีค่าเป็นไปตามเกณฑ์ทั้ง 2 เกณฑ์นี้มากที่สุด นั่นคือมีค่า RMSE ต่ำ

ที่สุด และ R^2 สูงที่สุด จะถือว่าการแจกแจงความน่าจะเป็นชนิดนั้นเป็นการแจกแจงความน่าจะเป็นที่เหมาะสมกับข้อมูลความเร็วลมเฉลี่ยรายวันในพื้นที่นั้นมากที่สุด

5. นำการแจกแจงความน่าจะเป็นที่เหมาะสมที่สุดที่ได้จากข้อ 4 มาหาความน่าจะเป็นที่ความเร็วลมเฉลี่ยรายวันจะมีค่าอยู่ในช่วง 2 - 12 เมตรต่อวินาที ที่เป็นความเร็วลมที่เหมาะสมในการตั้งฟาร์มลม

4 ผลการศึกษา

4.1 ค่าประมาณพารามิเตอร์ของการแจกแจงความน่าจะเป็นทั้ง 7 ชนิด

จากการนำชุดข้อมูลฝึกฝนของแต่ละสถานีมาประมาณค่าพารามิเตอร์ของการแจกแจงความน่าจะเป็นทั้ง 7 ชนิด ได้ค่าประมาณพารามิเตอร์ดังตารางที่ 2 - 7

ตารางที่ 2: ค่าประมาณพารามิเตอร์ของการแจกแจง 7 ชนิด ณ 381201: สถานีตรวจอากาศขอนแก่น

การแจกแจง	ค่าประมาณพารามิเตอร์
ปรกติ	μ = 5.1925 σ = 1.1314
ไวบูล	k = 4.6266 c = 5.6512
ล็อกนอร์มอล	μ = 1.6241 σ = 0.2144
แกมมา	α = 21.8497 β = 4.2079
ปรกติแบบผสม	μ_1 = 4.8184 σ_1 = 0.8214 μ_2 = 6.1650 σ_2 = 1.2426 p = 0.7221 $(1-p)$ = 0.2778
ไวบูลแบบผสม	k_1 = 6.6477 c_1 = 5.1877 k_2 = 5.2144 c_2 = 6.7203 p = 0.7425 $(1-p)$ = 0.2574
แกมมาแบบผสม	α_1 = 28.1287 β_1 = 5.6717 α_2 = 32.5023 β_2 = 4.8931 p = 0.8614 $(1-p)$ = 0.1385

ตารางที่ 3: ค่าประมาณพารามิเตอร์ของการแจกแจง 7 ชนิด ณ 381301: สถานีอุตุนิยมวิทยาเกษตรท่าพระ

การแจกแจง	ค่าประมาณพารามิเตอร์
ปรกติ	μ = 6.2295 σ = 2.0399
ไวบูล	k = 3.0128 c = 6.9333
ล็อกนอร์มอล	μ = 1.7801 σ = 0.3134
แกมมา	α = 10.3438 β = 1.6604
ปรกติแบบผสม	μ_1 = 5.9157 σ_1 = 1.5521 μ_2 = 9.4995 σ_2 = 3.2714 p = 0.9124 $(1-p)$ = 0.0875
ไวบูลแบบผสม	k_1 = 4.2487 c_1 = 6.4625 k_2 = 2.8155 c_2 = 9.7520 p = 0.8765 $(1-p)$ = 0.1234
แกมมาแบบผสม	α_1 = 13.7578 β_1 = 2.2721 α_2 = 4.1578 β_2 = 0.5386 p = 0.8951 $(1-p)$ = 0.1048

ตารางที่ 4: ค่าประมาณพารามิเตอร์ของการแจกแจง 7 ชนิด ณ 403201: สถานีอุตุนิยมวิทยาชัยภูมิ

การแจกแจง	ค่าประมาณพารามิเตอร์
ปรกติ	μ = 7.8601 σ = 2.0540
ไวบูล	k = 4.0159 c = 8.6417
ล็อกนอร์มอล	μ = 2.0262 σ = 0.2731
แกมมา	α = 14.2154 β = 1.8086
ปรกติแบบผสม	μ_1 = 7.4688 σ_1 = 1.7407 μ_2 = 9.7177 σ_2 = 2.3841 p = 0.8259 $(1-p)$ = 0.1740
ไวบูลแบบผสม	k_1 = 5.0930 c_1 = 8.0478 k_2 = 4.2128 c_2 = 10.3391 p = 0.7697 $(1-p)$ = 0.2302
แกมมาแบบผสม	α_1 = 10.1685 β_1 = 1.3343 α_2 = 23.1302 β_2 = 2.8648 p = 0.4713 $(1-p)$ = 0.5286

ตารางที่ 5: ค่าประมาณพารามิเตอร์ของการแจกแจง 7 ชนิด ณ 407301: สถานีอุตุนิยมวิทยาเกษตรอุบลราชธานี

การแจกแจง	ค่าประมาณพารามิเตอร์
ปรกติ	μ = 4.5826 σ = 1.9022
ไวบูล	k = 2.5463 c = 5.1720
ล็อกนอร์มอล	μ = 1.4399 σ = 0.4072
แกมมา	α = 6.2364 β = 1.3609
ปรกติแบบผสม	μ_1 = 3.4343 σ_1 = 0.9057 μ_2 = 5.8849 σ_2 = 1.8975 p = 0.5314 $(1-p)$ = 0.4685
ไวบูลแบบผสม	k_1 = 4.3851 c_1 = 3.7330 k_2 = 2.9734 c_2 = 6.1730 p = 0.4405 $(1-p)$ = 0.5594
แกมมาแบบผสม	α_1 = 11.8851 β_1 = 3.4847 α_2 = 10.4806 β_2 = 1.7821 p = 0.5255 $(1-p)$ = 0.4744

ตารางที่ 6: ค่าประมาณพารามิเตอร์ของการแจกแจง 7 ชนิด ณ 407501: สถานีตรวจอากาศอุบลราชธานี

การแจกแจง	ค่าประมาณพารามิเตอร์
ปรกติ	μ = 6.7796 σ = 2.2773
ไวบูล	k = 3.0573 c = 7.5637
ล็อกนอร์มอล	μ = 1.8585 σ = 0.3377
แกมมา	α = 9.1924 β = 1.3560
ปรกติแบบผสม	μ_1 = 6.1699 σ_1 = 1.65507 μ_2 = 9.1042 σ_2 = 2.7740 p = 0.7922 $(1-p)$ = 0.2077
ไวบูลแบบผสม	k_1 = 4.2798 c_1 = 6.7553 k_2 = 3.1278 c_2 = = 9.4180 p = 0.7239 $(1-p)$ = 0.2760
แกมมาแบบผสม	α_1 = 12.1192 β_1 = 1.8430 α_2 = 5.6468 β_2 = 0.7621 p = 0.7553 $(1-p)$ = 0.2446

ตารางที่ 7: ค่าประมาณพารามิเตอร์ของการแจกแจง 7 ชนิด ณ 431301: สถานีอุตุนิยมวิทยานครราชสีมา

การแจกแจง	ค่าประมาณพารามิเตอร์
ปรกติ	μ = 7.3586 σ = 1.8269
ไวบูล	k = 4.0509 c = 8.0643
ล็อกนอร์มอล	μ = 1.9653 σ = 0.2492
แกมมา	α = 16.5341 β = 2.2469
ปรกติแบบผสม	μ_1 = 7.0748 σ_1 = 7.0748 μ_2 = 9.6671 σ_2 = 2.4129 p = 0.8905 $(1-p)$ = 0.1094
ไวบูลแบบผสม	k_1 = 5.5448 c_1 = 7.5786 k_2 = 3.9951 c_2 = = 9.9817 p = 0.8268 $(1-p)$ = 0.1731
แกมมาแบบผสม	α_1 = 20.3186 β_1 = 2.7901 α_2 = 8.4453 β_2 = 1.0855 p = 0.8464 $(1-p)$ = 0.1535

4.2 การทดสอบแอนเดอร์สัน-ดาร์ลิง

ผลการทดสอบความหมาะสมของการแจกแจงความน่าจะเป็นทั้ง 7 ชนิด ที่มีต่อความเร็วลมเฉลี่ยรายวัน แสดงดังตารางที่ 8 โดยจะเห็นได้ว่าการแจกแจงความน่าจะเป็นทั้ง 7 ชนิด เป็นการแจกแจงความน่าจะเป็นที่เหมาะสมกับความเร็วลมเฉลี่ยรายวันในทุกสถานี เนื่องจากมีค่าพีของการทดสอบแอนเดอร์สัน-ดาร์ลิงมากกว่าระดับนัยสำคัญ 0.05

ตารางที่ 8: ผลการทดสอบแอนเดอร์สัน-ดาร์ลิง ของการแจกแจงความน่าจะเป็นทั้ง 7 ชนิด และ 6 สถานี

การแจกแจง	การทดสอบแอนเดอร์สัน-ดาร์ลิง					
	381201	381301	403201	407301	407501	431301
ปรกติ	4.4231 (0.2196)	4.8895 (0.1403)	3.5268 (0.5020)	5.3062 (0.0895)	3.172 (0.6504)	3.8632 (0.3783)
ไวบูล	4.8201 (0.1476)	3.8578 (0.3801)	3.3919 (0.5570)	5.4399 (0.0779)	4.7141 (0.1677)	5.5447 (0.0713)
ล็อกนอร์มอล	3.339 (0.5711)	3.8187 (0.3933)	3.2124 (0.6330)	4.3488 (0.2362)	4.8411 (0.1475)	4.1004 (0.3049)
แกมมา	3.2987 (0.5881)	3.6363 (0.4594)	3.533 (0.4995)	4.0756 (0.3065)	5.7301 (0.0588)	5.1448 (0.1080)
ปรกติแบบผสม	4.2024 (0.2718)	4.6432 (0.1801)	2.6225 (0.8676)	4.7551 (0.1577)	3.8871 (0.3703)	3.5917 (0.4765)
ไวบูลแบบผสม	3.5823 (0.4726)	3.3418 (0.5779)	2.0336 (0.9858)	3.6202 (0.4582)	3.0865 (0.6870)	5.8173 (0.0537)
แกมมาแบบผสม	3.5891 (0.4700)	6.459 (0.0576)	4.475 (0.2128)	4.3418 (0.2378)	3.4468 (0.5343)	3.6228 (0.4646)

ค่าใน ( ) คือค่าพี

4.3 การหาการแจกแจงความน่าจะเป็นที่เหมาะสมที่สุด

เมื่อนำการแจกแจงความน่าจะเป็นที่เหมาะสมที่ได้จากชุดข้อมูลฝึกฝนมาตรวจสอบกับชุดข้อมูลทดสอบเพื่อหาการแจกแจงความน่าจะเป็นที่เหมาะสมที่สุดกับข้อมูลความเร็วลมเฉลี่ยรายวันในแต่ละสถานีโดยพิจารณาจาก

ค่า RMSE และ R^2 ได้ผลดังตารางที่ 9 ซึ่งจะเห็นได้ว่าการแจกแจงความน่าจะเป็นที่เหมาะสมกับข้อมูลความเร็วลมเฉลี่ยรายวันของแต่ละสถานีสรุปได้ดังนี้

- 381201: สถานีตรวจอากาศขอนแก่น การแจกแจงที่เหมาะสมคือการแจกแจงล็อกนอร์มัล
- 381301: สถานีอุตุนิยมวิทยาท่าพระ การแจกแจงที่เหมาะสมคือการแจกแจงไวบูลแบบผสม
- 403201: สถานีอุตุนิยมวิทยาชัยภูมิ การแจกแจงที่เหมาะสมคือการแจกแจงล็อกนอร์มัล
- 407301: สถานีอุตุนิยมวิทยาแกษตรอุบลราชธานี การแจกแจงที่เหมาะสมคือการแจกแจงปรกติ
- 407501: สถานีตรวจอากาศอุบลราชธานี การแจกแจงที่เหมาะสมคือการแจกแจงล็อกนอร์มัล
- 431301: สถานีอุตุนิยมวิทยานครราชสีมา การแจกแจงที่เหมาะสมคือการแจกแจงปรกติ

ตารางที่ 9: ค่า RMSE และ R^2 ของการแจกแจงความน่าจะเป็นทั้ง 7 ชนิด

สถานี		การแจกแจงความน่าจะเป็น						
		ปรกติ	ไวบูล	ล็อกนอร์มัล	แกมมา	ปรกติแบบผสม	ไวบูลแบบผสม	แกมมาแบบผสม
381201	RMSE	0.1317	0.1218	0.1003*	0.1107	0.1116	0.1121	0.1047
	R^2	0.6937	0.6955	0.8558*	0.8160	0.8291	0.8177	0.8518
381301	RMSE	0.0615	0.0751	0.0544	0.0549	0.0524	0.0521*	0.0538
	R^2	0.9446	0.9033	0.9583	0.9573	0.9624	0.9625*	0.9599
403201	RMSE	0.0930	0.0893	0.0724*	0.0789	0.0863	0.0885	0.0852
	R^2	0.9016	0.9006	0.9417*	0.9309	0.9195	0.9144	0.9210
407301	RMSE	0.1336*	0.1420	0.1725	0.1595	0.1690	0.1688	0.1696
	R^2	0.7529*	0.6952	0.5574	0.6364	0.5534	0.5504	0.5526
407501	RMSE	0.0759	0.0626	0.0557*	0.0624	0.0687	0.0671	0.0645
	R^2	0.9375	0.9534	0.9682*	0.9601	0.9535	0.9547	0.9588
431301	RMSE	0.1118*	0.1186	0.1343	0.1271	0.1311	0.1287	0.1317
	R^2	0.8459*	0.8354	0.7698	0.7981	0.7967	0.8050	0.7891

* หมายถึงการแจกแจงที่มีค่า RMSE น้อยที่สุด และ R^2 มากที่สุด

4.4 การหาความน่าจะป็นที่ความเร็วลมจะมีค่าในช่วงที่เหมาะสม

เมื่อนำการแจกแจงความน่าจะเป็นที่เหมาะสมที่สุดของความเร็วลมเฉลี่ยรายวันในแต่ละสถานีมาหาความน่าจะเป็นที่ความเร็วลมเฉลี่ยรายวันจะมีค่าในช่วง 2 - 12 เมตรต่อวินาที ซึ่งเป็นช่วงของความเร็วลมที่เหมาะสมในการตั้งฟาร์มลม ได้ผลดังตารางที่ 10 ซึ่งจะเห็นได้ว่าความน่าจะเป็นที่ความเร็วลมเฉลี่ยรายวันจะมีค่าอยู่ในช่วงที่เหมาะสมสำหรับการตั้งฟาร์มลมของแต่ละสถานีนั้นมีค่ามากกว่า 0.95 ยกเว้นสถานี 407301: สถานีอุตุนิยมวิทยาเกษตรอุบลราชธานีเท่านั้นที่มีความน่าจะเป็นน้อยกว่า 0.95 แต่ยังคงมีความน่าจะเป็นมากกว่า 0.90 จึงถือได้ว่าพื้นที่ในบริเวณสถานีทั้ง 6 สถานี มีศักยภาพในการตั้งฟาร์มลมเพื่อผลิตไฟฟ้า

5 สรุปและอภิปรายผล

จากผลการศึกษาพบว่าการแจกแจงล็อกนอร์มัลเป็นการแจกแจงความน่าจะเป็นที่เหมาะสมกับข้อมูลความเร็วลมเฉลี่ยรายวันของสถานีตรวจอากาศขอนแก่น สถานีอุตุนิยมวิทยาชัยภูมิ และสถานีตรวจอากาศอุบลราชธานี

ตารางที่ 10: ความน่าจะเป็นที่ความเร็วลมจะมีค่าอยู่ในช่วง 2-12 เมตรต่อวินาที

สถานี	381201	381301	403201	407301	407501	431301
การแจกแจง	ล็อกนอร์มัล	ไวบูลแบบผสม	ล้อกนอร์มัล	ปรกติ	ล็อกนอร์มัล	ปรกติ
ความน่าจะเป็น	0.9874	0.9982	0.9534	0.9126	0.9679	0.9927

ส่วนสถานีอุตุนิยมวิทยาเกษตรอุบลราชธานี และสถานีอุตุนิยมวิทยานครราชสีมา การแจกแจงความน่าจะเป็นที่เหมาะสมคือการแจกแจงปรกติ สำหรับสถานีอุตุนิยมวิทยาท่าพระการแจกแจงความน่าจะเป็นที่เหมาะสมคือการแจกแจงไวบูลแบบผสม ซึ่งการแจกแจงความน่าจะเป็นที่เหมาะสมเหล่านี้สอดคล้องกับการแจกแจงความน่าจะเป็นที่ได้จากงานวิจัยที่ผ่านมา เช่น งานของ Filom et al. [7] ในปี 2021 Kantar et al. [10] ในปี 2016 และ Koca et al. [8] ในปี 2019 นอกจากนั้นพื้นที่ในบริเวณสถานีเหล่านี้ยังเป็นพื้นที่ที่มีศักยภาพในการตั้งฟาร์มลมเพื่อผลิตไฟฟ้าเนื่องจากมีความเร็วลมเฉลี่ยรายวันอยู่ในช่วงที่เหมาะสมอีกด้วย

ในประเทศไทยยังมีพื้นที่ที่มีความน่าสนใจในการศึกษาถึงศักยภาพในการติดตั้งฟาร์มลมเพื่อผลิตไฟฟ้า และยังมีการแจกแจงความน่าจะเป็นชนิดอื่น ๆ ที่อาจจะมีความเหมาะสมกับข้อมูลความเร็วลมมากกว่าการแจกแจงความน่าจะเป็นที่นำมาศึกษา ดังนั้นงานวิจัยในครั้งต่อไปอาจทำการศึกษาในพื้นที่อื่น ๆ หรือนำการแจกแจงความน่าจะเป็นชนิดอื่นมาศึกษาต่อไป

เอกสารอ้างอิง

[1] กนกวรรณ สุวรรณมุข, *แนวทางการส่งเสริมการผลิตไฟฟ้าจากพลังงานลมชายฝั่งและนอกชายฝั่งในประเทศไทย*, 2561.

[2] กรมพัฒนาพลังงานทดแทน และอนุรักษ์พลังงาน กระทรวงพลังงาน, *คู่มือการพัฒนาและการลงทุนผลิตพลังงานทดแทน*, 1 ed., บริษัท เอเบิล คอนซัลแตนท์ จำกัด, กรุงเทพฯ, 2554.

[3] ชมพูนุท ทับเจริญ และ ฆนัทนันท์ ทวีวัฒน์, *การศึกษาความเป็นไปได้ในการลงทุนติดตั้งกังหันลมเพื่อผลิตกระแสไฟฟ้าของภาคครัวเรือน*, Journal of Science & Technology MSU **12** (2560), no. 2.

[4] พนิดา สุขสมพร้อม, สุพิชชา ถวิลไพร, สุภชัย พลน้ำเที่ยง, ศิโรรัตน์ พัฒนไพโรจน์ และ เกียรติฟ้า ตั้งใจจิต, *การประเมินลักษณะความเร็วลมเฉพาะแหล่งในเขตพื้นที่จังหวัดนครพนม*, วารสารวิจัย มข. (ฉบับบัณฑิตศึกษา) **21** (2564), no. 1, 122–132.

[5] พนิดา สุขสมพร้อม, เกียรติฟ้า ตั้งใจจิต และ สุภชัย พลน้ำเที่ยง, *การวิเคราะห์ข้อมูลความเร็วลมในเขตพื้นที่จังหวัดกาฬสินธุ์*, Journal of Science & Technology MSU **38** (2562), no. 5.

[6] สุภชัย พลน้ำเที่ยง และ เกียรติฟ้า ตั้งใจจิต, *การวิเคราะห์ข้อมูลพลังงานลมในเขตพื้นที่จังหวัดหนองคาย*, วิศวสารลาดกระบัง **34** (2560), no. 2, 29–36.

[7] Siyavash Filom, Soheil Radfar, and Roozbeh Panahi, *Exploring wind energy potential as a driver of sustainable development in the southern coasts of iran: The importance of wind speed statistical distribution model*, Sustainability **13** (2021), no. 14, 7702.

[8] Melih Burak Koca, Muhammet Burak Kiliç, and YUSUF ŞAHİN, *Assessing wind energy potential using finite mixture distributions*, Turkish Journal of Electrical Engineering and Computer Sciences **27** (2019), no. 3, 2276–2294.

[9] Ravindra Kollu, Srinivasa Rao Rayapudi, SVL Narasimham, and Krishna Mohan Pakkurthi, *Mixture probability distribution functions to model wind speed distributions*, International Journal of energy and environmental engineering **3** (2012), 1–10.

[10] Yeliz Mert Kantar, Ilhan Usta, Ismail Yenilmez, and Ibrahim Arik, *A study on estimation of wind speed distribution by using the modified weibull distribution*, INTERNATIONAL JOURNAL OF INFORMATICS TECHNOLOGIES **9** (2016), 63.

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

การศึกษาความแกร่งของสถิติทดสอบความแตกต่างของค่าเฉลี่ยประชากรสองกลุ่มอิสระกัน เมื่อข้อมูลมีการแจกแจงปรกติแบบผสมและการแจกแจงแกมมาแบบผสม*

ภัทราภรณ์ กิจผลเจริญ† สุวิมล ชูเปรม และ บำรุงศักดิ์ เผื่อนอารีย์‡

ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา 20131

บทคัดย่อ

การศึกษาครั้งนี้มีวัตถุประสงค์เพื่อศึกษาความแกร่งของการทดสอบที (t-test) และการทดสอบของเวลช์ (Welch's test) ภายใต้ข้อมูลที่มีการแจกแจงผสม ได้แก่ การแจกแจงปรกติแบบผสม และ การแจกแจงแกมมาแบบผสม โดยการจำลองด้วยเทคนิคมอนติคาร์โล ด้วยโปรแกรม R และกำหนดขนาดตัวอย่างสองกลุ่มเท่ากัน คือ 10, 30, 50, 100 และ 200 ที่ระดับนัยสำคัญ 0.05 เกณฑ์การเปรียบเทียบประสิทธิภาพของความแกร่งคือ ความสามารถในการควบคุมความผิดพลาดแบบที่ 1 ผลการศึกษาพบว่าการทดสอบทีและการทดสอบของเวลช์สามารถควบคุมความผิดพลาดแบบที่ 1 ได้ นั่นคือ สถิติทดสอบทั้งสองมีความแกร่งภายใต้ข้อมูลที่มีการแจกแจงปรกติแบบผสมและการแจกแจงแกมมาแบบผสมทุกกรณีที่ศึกษา

คำสำคัญ: ความผิดพลาดแบบที่ 1, การทดสอบที, การทดสอบของเวลช์, การแจกแจงปรกติแบบผสม, การแจกแจงแกมมาแบบผสม

2020 MSC: ปฐมภูมิ 62F35 ทุติยภูมิ 62F03

1 บทนำ

การศึกษาความแตกต่างของพารามิเตอร์ของประชากรสองกลุ่มอิสระกัน คือการเปรียบเทียบค่าเฉลี่ยของประชากรทั้งสองกลุ่มที่เป็นอิสระกัน โดยในปัจจุบันการทดสอบทียังคงเป็นหนึ่งในสถิติทดสอบที่นักวิจัยนิยมใช้ในการทดสอบสมมุติฐานสำหรับกรณีนี้ โดยมีข้อสมมุติเบื้องต้นของการใช้การทดสอบที คือ ข้อมูลจะต้องถูกสุ่มมาจาก

*งานวิจัยเรื่องนี้ได้รับทุนสนับสนุนจากคณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา

†ผู้นำเสนอ ‡ผู้แต่งหลัก

อีเมล: pattara@go.buu.ac.th (ภัทราภรณ์ กิจผลเจริญ), 62030267@go.buu.ac.th (สุวิมล ชูเปรม), bumrungsak@buu.ac.th (บำรุงศักดิ์ เผื่อนอารีย์).

ประชากรที่มีการแจกแจงปรกติ โดยเมื่อประชากรมีความแปรปรวนเท่ากัน จะใช้การทดสอบที (Independent samples t-test) ในการทดสอบ และเมื่อประชากรมีความแปรปรวนไม่เท่ากัน จะใช้การทดสอบของเวลช์ (Welch's test) ในการทดสอบ โดยจากงานวิจัยของ Derrick, Toher และ White [1] ได้ศึกษาเปรียบเทียบความแกร่ง (Robustness) ของการทดสอบของเวลช์และการทดสอบทีในการควบคุมความผิดพลาดแบบที่ 1 ภายใต้ข้อมูลที่มีการแจกแจงปรกติ ผลการวิจัยพบว่า การทดสอบของเวลช์มีความแกร่งในการควบคุมความผิดพลาดแบบที่ 1 ได้ดีในทุกกรณีที่ศึกษา โดยที่การทดสอบทีสามารถควบคุมความผิดพลาดแบบที่ 1 ได้ดีในบางกรณีเท่านั้น Delacre, Lakens และ Lays [2] ได้ศึกษาความแกร่งของการทดสอบทีและการทดสอบของเวลช์สำหรับใช้กับการทำวิจัยทางด้านจิตวิทยา พบว่า การทดสอบของเวลช์สามารถควบคุมความผิดพลาดแบบที่ 1 ได้ดีกว่าเมื่อความแปรปรวนของข้อมูลสองกลุ่มแตกต่างกัน ส่วนอีกกรณีหนึ่งที่ความแปรปรวนของข้อมูลเท่ากัน การทดสอบทีมีประสิทธิภาพดีกว่า อย่างไรก็ตามในความเป็นจริงแล้วข้อมูลบางประเภท ไม่ได้เป็นไปตามข้อสมมุติเบื้องต้น นั่นคือ ข้อมูลไม่มีการแจกแจงปรกติ ตัวอย่างเช่น ข้อมูลที่มีการแจกแจงปรกติแบบผสม ได้แก่ ข้อมูลเกี่ยวกับการเงิน การวิเคราะห์ความเสี่ยงทางด้านการเงิน [3] ข้อมูลที่มีการแจกแจงแกมมาแบบผสม ได้แก่ ข้อมูลทางด้านอุทกวิทยา [4] ข้อมูลปริมาณน้ำฝน [5] เป็นต้น โดยข้อมูลที่กล่าวมานั้นล้วนแต่เป็นข้อมูลที่มีความสำคัญ ซึ่งหากต้องการวิเคราะห์ข้อมูลเหล่านี้ โดยการทดสอบสมมุติฐานเกี่ยวกับค่าเฉลี่ย สถิติทดสอบพื้นฐานทั้งสองวิธีที่กล่าวมายังมีความแกร่งที่จะใช้ทดสอบข้อมูลที่มีการแจกแจงปรกติแบบผสมและการแจกแจงแกมมาแบบผสมในกรณีนี้ได้หรือไม่ นอกจากนี้ยังสามารถอ้างอิงไปถึงข้อมูลประเภทอื่นที่มีการแจกแจงแบบผสมในลักษณะเดียวกันได้อีกด้วย

ดังนั้น ผู้วิจัยจึงสนใจที่จะศึกษาความแกร่งของการทดสอบของเวลช์และการทดสอบที ในการควบคุมความผิดพลาดแบบที่ 1 (Type I error) กรณีที่ข้อมูลนั้นไม่เป็นไปตามข้อสมมุติเบื้องต้นของวิธีการทดสอบ โดยทำการศึกษาภายใต้ข้อมูลที่มีการแจกแจงผสม ได้แก่ การแจกแจงปรกติแบบผสม และการแจกแจงแกมมาแบบผสม

2 วิธีดำเนินการวิจัย

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาความแกร่งของการทดสอบที (t-test) และการทดสอบของเวลช์ (Welch's test) ภายใต้ข้อมูลที่มีการแจกแจงผสม ได้แก่ การแจกแจงปรกติแบบผสม และการแจกแจงแกมมาแบบผสม การศึกษาครั้งนี้ได้ทำการจำลองข้อมูลที่มีการแจกแจงผสมที่กำหนดด้วยเทคนิคมอนติคาร์โล (Monte Carlo Simulation Technique) ด้วยโปรแกรม R [6] ทดลองทำซ้ำ 10,000 ครั้ง กำหนดขนาดตัวอย่าง คือ 10, 30, 50, 100 และ 200 โดยทั้งสองกลุ่มมีขนาดเท่ากัน และกำหนดพารามิเตอร์ถ่วงน้ำหนักเท่ากับ 0.5 และ 0.8 ซึ่งมีรายละเอียดของวิธีการทดสอบที่ศึกษาและการกำหนดการแจกแจง ดังนี้

2.1 สถิติทดสอบที่ใช้ในการศึกษา

2.1.1 การทดสอบที (Independent samples t-test) เป็นวิธีการทดสอบค่าเฉลี่ยสองประชากรเมื่อข้อมูลถูกสุ่มมาจากประชากรที่มีการแจกแจงปรกติที่เป็นอิสระกันและมีความแปรปรวนเท่ากัน โดยมีการใช้ในตำราสถิติอย่างแพร่หลาย [7-9] โดยมีตัวสถิติทดสอบ ดังนี้

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \quad (2.1)$$

ซึ่งมีการแจกแจงโดยประมาณแบบที ที่องศาเสรีคือ; $df = n_1 + n_2 - 2$

โดยที่ $\bar{X}_i$ แทนค่าเฉลี่ยของตัวอย่างของกลุ่มที่ $i; i = 1, 2$

n_i แทนขนาดตัวอย่างของกลุ่มที่ $i; i = 1, 2$

S_p^2 แทนความแปรปรวนรวม คำนวณได้จาก ; $S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$

S_i^2 แทนความแปรปรวนตัวอย่างของกลุ่มที่ $i; i = 1, 2$

2.1.2 การทดสอบของเวลช์ (Welch's test) เป็นวิธีการทดสอบค่าเฉลี่ยสองประชากรเมื่อข้อมูลถูกสุ่มมาจากประชากรที่มีการแจกแจงปรกติที่เป็นอิสระกันแต่มีความแปรปรวนไม่เท่ากัน [7-9] โดยมีตัวสถิติทดสอบ ดังนี้

$$W = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}} \quad (2.2)$$

ซึ่งมีการแจกแจงโดยประมาณแบบที ที่องศาเสรีคือ; $df = \frac{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)^2}{\frac{1}{n_1-1}\left(\frac{S_1^2}{n_1}\right)^2 + \frac{1}{n_2-1}\left(\frac{S_2^2}{n_2}\right)^2}$

โดยที่ $\bar{X}_i$ แทนค่าเฉลี่ยของตัวอย่างที่ $i; i = 1, 2$

n_i แทนขนาดตัวอย่างของกลุ่มที่ $i; i = 1, 2$

S_i^2 แทนความแปรปรวนตัวอย่างที่ $i; i = 1, 2$

2.2 การแจกแจงที่ใช้ในการศึกษา

ในการศึกษาครั้งนี้ได้จำลองข้อมูลภายใต้การแจกแจงผสม (Mixture Distributions) โดยมีฟังก์ชันความน่าจะเป็น ดังนี้

$$f(x, p) = (p)f_1(x) + (1-p)f_2(x) \ ; 0 < p < 1 \quad (2.3)$$

เมื่อ p แทน พารามิเตอร์ถ่วงน้ำหนัก และ $f_i(x)$ แทน ฟังก์ชันความน่าจะเป็นขององค์ประกอบที่ $i; i = 1, 2$

2.2.1 การแจกแจงปรกติแบบผสม (Mixed normal distribution: MN) มีฟังก์ชันการแจกแจงความน่าจะเป็นดังนี้

$$f(x) = p\left(\frac{1}{\sqrt{2\pi\sigma_1^2}}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\right)+(1-p)\left(\frac{1}{\sqrt{2\pi\sigma_2^2}}e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}\right) \quad (2.4)$$

เมื่อ μ_i, σ_i^2 แทนพารามิเตอร์บ่งตำแหน่งและบ่งรูปร่างขององค์ประกอบที่ $i; i = 1,2$

2.2.2 การแจกแจงแกมมาแบบผสม (Mixed gamma distribution: MG) มีฟังก์ชันการแจกแจงความน่าจะเป็นดังนี้

$$f(x) = p\left(\frac{x^{k_1-1}e^{-\frac{x}{\theta_1}}}{\Gamma(k_1)\theta_1^{k_1}}\right)+(1-p)\left(\frac{x^{k_2-1}e^{-\frac{x}{\theta_2}}}{\Gamma(k_2)\theta_2^{k_2}}\right) \quad (2.5)$$

เมื่อ k_i, θ_i แทนพารามิเตอร์บ่งรูปร่างและบ่งขนาดขององค์ประกอบที่ $i; i = 1,2$ และสามารถกำหนดค่าพารามิเตอร์ได้ดังตารางที่ 1

ตารางที่ 1: ค่าพารามิเตอร์ของการแจกแจงที่ศึกษา

การแจกแจง	สัญลักษณ์	p	ประชากรกลุ่มที่ 1		p	ประชากรกลุ่มที่ 2	
			องค์ประกอบที่ 1	องค์ประกอบที่ 2		องค์ประกอบที่ 1	องค์ประกอบที่ 2
การแจกแจงปรกติแบบผสม (μ, σ)	MN1	0.5	(0,1)	(1,1)	0.5	(0,1)	(1,1)
	MN2	0.8	(0,1)	(1,1)	0.8	(0,1)	(1,1)
	MN3	0.5	(0,1)	(2,1)	0.5	(0,1)	(2,1)
	MN4	0.8	(0,1)	(2,1)	0.8	(0,1)	(2,1)
	MN5	0.5	(0,1)	(1,1)	0.5	(0,2)	(1,2)
	MN6	0.8	(0,1)	(1,1)	0.8	(0,2)	(1,2)
	MN7	0.5	(0,1)	(2,1)	0.5	(0,2)	(2,2)
	MN8	0.8	(0,1)	(2,1)	0.8	(0,2)	(2,2)
การแจกแจงแกมมาแบบผสม (k, θ)	MG1	0.5	(1,1)	(1/4,4)	0.5	(1,1)	(1/4,4)
	MG2	0.8	(1,1)	(1/4,4)	0.8	(1,1)	(1/4,4)
	MG3	0.5	(1,1)	(1/2,2)	0.5	(1,1)	(1/2,2)
	MG4	0.8	(1,1)	(1/2,2)	0.8	(1,1)	(1/2,2)
การแจกแจงแกมมาแบบผสม (k, θ)	MG5	0.5	(1,1)	(1/2,2)	0.5	(1,1)	(1/4,4)
	MG6	0.8	(1,1)	(1/2,2)	0.8	(1,1)	(1/4,4)
	MG7	0.5	(1,1)	(1/4,4)	0.5	(1/2,2)	(1/4,4)
	MG8	0.8	(1,1)	(1/4,4)	0.8	(1/2,2)	(1/4,4)

2.3 ความสามารถในการควบคุมความผิดพลาดแบบที่ 1 (Type I error: α)

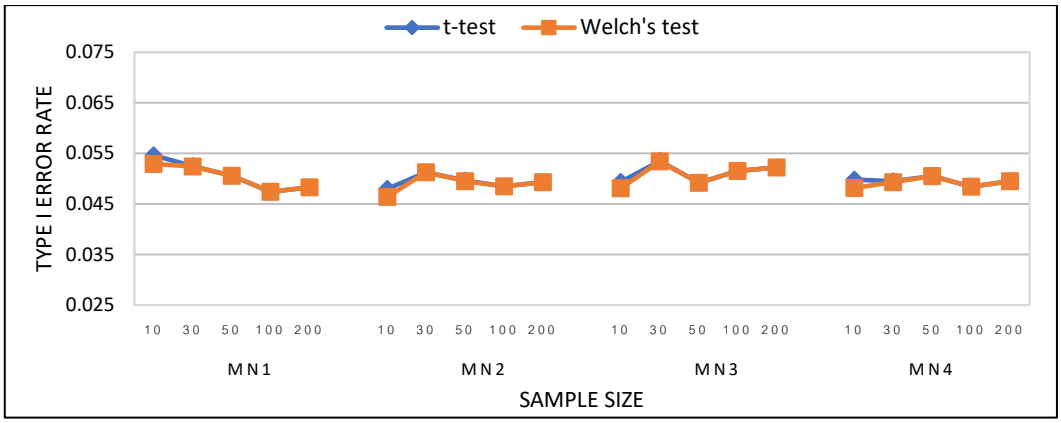
กรณีข้อมูลที่นำมาวิเคราะห์ไม่เป็นไปตามข้อสมมุติเบื้องต้นแต่วิธีการทดสอบยังคงมีประสิทธิภาพและสามารถใช้ทดสอบสมมุติฐานได้ แสดงว่าการทดสอบมีความแกร่ง โดยเกณฑ์ที่ใช้ในการศึกษาความแกร่งของการทดสอบทีและการทดสอบของเวลช์ จะพิจารณาจากความสามารถในการควบคุมความผิดพลาดแบบที่ 1 ซึ่งหมายถึงการปฏิเสธสมมุติฐานหลักเมื่อสมมุติฐานหลักเป็นจริง โดยจำลองข้อมูลภายใต้สมมุติฐานหลักเป็นจริงและนับจำนวนครั้งของการปฏิเสธสมมุติฐานหลัก (I) โดยค่าประมาณความน่าจะเป็นของการเกิดความผิดพลาดแบบที่ 1 คำนวณจาก

$$\hat{\alpha} = \frac{I}{10{,}000}$$

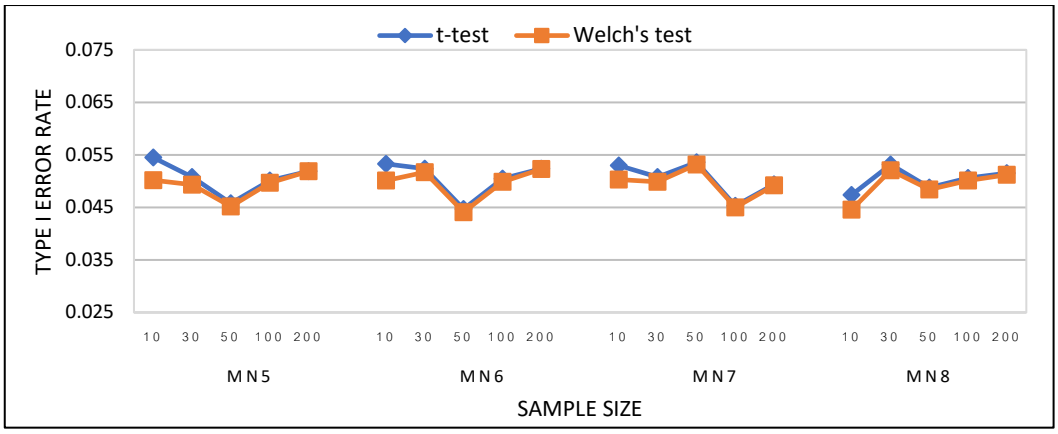
จากนั้นเปรียบเทียบกับเกณฑ์ของ Bradley [0.025,0.075] [10] โดยงานวิจัยนี้กำหนดระดับนัยสำคัญเท่ากับ 0.05 หาก $\hat{\alpha}$ ตกอยู่ในเกณฑ์ดังกล่าวจะสรุปว่าสถิติทดสอบสามารถควบคุมความผิดพลาดแบบที่ 1 ได้

3 ผลการศึกษา

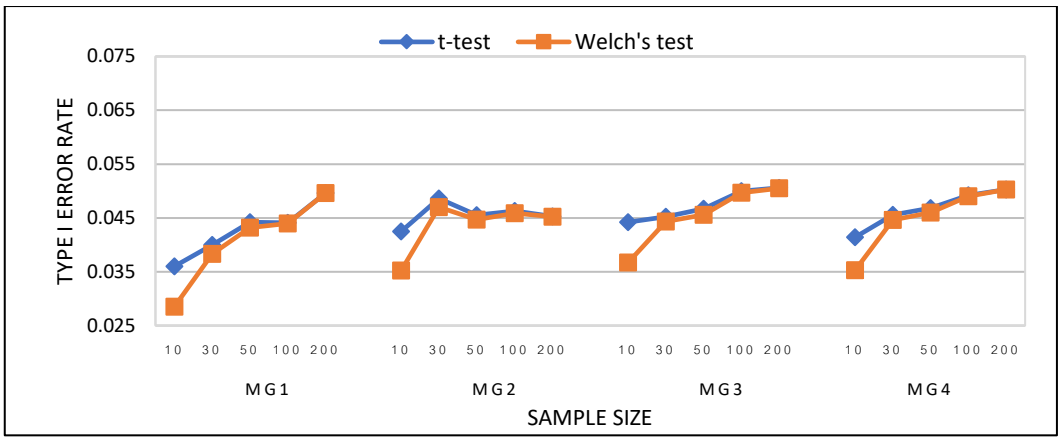
พิจารณาจากค่าประมาณความน่าจะเป็นของการเกิดความผิดพลาดแบบที่ 1 ของการทดสอบทีและการทดสอบของเวลช์ เมื่อข้อมูลมีการแจกแจงแบบผสมทั้งสองการแจกแจง พบว่า ค่าประมาณดังกล่าวของการทดสอบทีและการทดสอบของเวลช์ ตกอยู่ในเกณฑ์ของ Bradley ในทุกกรณีที่ศึกษา ดังนั้น การทดสอบทั้งสองสามารถควบคุมความผิดพลาดแบบที่ 1 ได้ อย่างไรก็ตามค่าประมาณความน่าจะเป็นของการเกิดความผิดพลาดแบบที่ 1 ของการทดสอบทั้งสองวิธีของข้อมูลที่มีการแจกแจงปรกติแบบผสมมีค่าใกล้เคียงกับระดับนัยสำคัญที่กำหนดมากกว่าข้อมูลที่มีการแจกแจงแกมมาแบบผสม แสดงดังภาพที่ 1 – 4



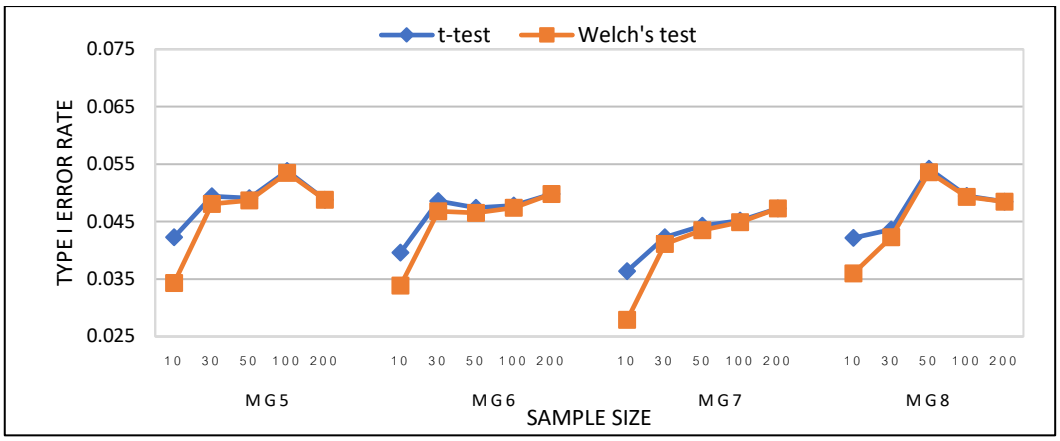
ภาพที่ 1: ค่าประมาณความน่าจะเป็นของการเกิดความผิดพลาดแบบที่ 1 เมื่อข้อมูลมีการแจกแจงปรกติแบบผสม กรณีความแปรปรวนคงที่



ภาพที่ 2: ค่าประมาณความน่าจะเป็นของการเกิดความผิดพลาดแบบที่ 1 เมื่อข้อมูลมีการแจกแจงปรกติแบบผสม กรณีความแปรปรวนไม่คงที่



ภาพที่ 3: ค่าประมาณความน่าจะเป็นของการเกิดความผิดพลาดแบบที่ 1 เมื่อข้อมูลมีการแจกแจงแกมมาแบบผสม กรณีความแปรปรวนคงที่



ภาพที่ 4: ค่าประมาณความน่าจะเป็นของการเกิดความผิดพลาดแบบที่ 1 เมื่อข้อมูลมีการแจกแจงแกมมาแบบผสม กรณีความแปรปรวนไม่คงที่

4 สรุปผลและอภิปรายผลการศึกษา

จากผลการวิจัยโดยการศึกษาความแกร่งของการทดสอบทีและการทดสอบของเวลช์ในการควบคุมความผิดพลาดแบบที่ 1 พบว่าการทดสอบทีและการทดสอบของเวลช์สามารถควบคุมความผิดพลาดแบบที่ 1 ได้ในทุกสถานการณ์ที่ศึกษา เมื่อข้อมูลมีการแจกแจงปรกติแบบผสมและการแจกแจงแกมมาแบบผสมทั้งกรณีที่ตัวอย่างมีขนาด เล็ก ปานกลางและใหญ่ รวมถึงทุกขนาดของพารามิเตอร์ถ่วงน้ำหนักของการแจกแจงผสมที่นำมาศึกษาและเมื่อพิจารณาจากขนาดตัวอย่างที่เพิ่มขึ้นจะเห็นได้ว่าค่าประมาณความน่าจะเป็นของการเกิดความผิดพลาดแบบที่ 1 ของทั้งสองวิธีจะมีค่าเข้าใกล้ระดับนัยสำคัญที่กำหนดมากขึ้น ดังนั้น หากข้อมูลที่นำมาวิเคราะห์มีการแจกแจงปรกติแบบผสมและการแจกแจงแกมมาแบบผสมแล้ว สถิติทดสอบทั้งสองยังคงมีความแกร่งและสามารถใช้ทดสอบสมมุติฐานได้อย่างมีประสิทธิภาพเช่นเดิม

5 ข้อเสนอแนะ

เนื่องจากการวัดประสิทธิภาพของวิธีการทดสอบอาจพิจารณากำลังการทดสอบร่วมด้วย ดังนั้นควรทำการศึกษาเพิ่มเติมเกี่ยวกับกำลังการทดสอบ เพื่อทราบถึงประสิทธิภาพของการทดสอบทีและการทดสอบของเวลช์ภายใต้ข้อมูลที่มีการแจกแจงปรกติแบบผสมและการแจกแจงแกมมาแบบผสม

กิตติกรรมประกาศ ขอขอบคุณคณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา ที่สนับสนุนการทำวิจัยครั้งนี้

เอกสารอ้างอิง

[1] B. Derrick, D. Toher and P. White, *Why Welch's test is type I error robust,* The Quantitative Methods for Psychology. 12(1) (2016), 30-38. DOI:10.20982/tpmp.12.1.

[2] M. Delacre, D. Lakens and C. Lays, *Why Psychologists Should by Default Use Welch's t-test Instead of Student's t-test,* International Review of Social Psychology. 30(1) (2017), 92-101. DOI: https://doi.org/10.5334/irsp.82.

[3] K. Kocak, N. Calis and D. Unal, *Mixture Distribution Approach In Financial Risk Analysis,* Journal of Business Economics and Finance. 2(3) (2013), 13-24.

[4] J. Jiao and W. Cheng, *Tolerance Limits Under Gamma Mixtures: Application in Hydrology,* Journal of Systems Science and Complexity. 36 (2023), 1285-1301.

[5] J. Suhaila, K. Ching-Yee, Y. Fadhilah and F. Hui-Mean, *Introducing the Mixed Distribution in Fitting Rainfall Data,* Open Journal of Modern Hydrology. 1(2) (2011), 11-22. DOI:10.4236/ojmh.2011.12002.

[6] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, 2022.

[7] F. M. Dekking, C. Kraaikamp, H. P. Lopuhaä and L. E. Meester, *A Modern Introduction to Probability and Statistics : Understanding Why and How*, Springer, 2005.

[8] Z. B. Alfassi, Z. Boger and Y. Ronen, *Statistical treatment of analytical data*, CRC Press, 2005.

[9] J. Miles and P. Banyard, *Understanding and using statistics in psychology: a practical introduction*, Sage, 2007.

[10] J.V. Bradley, *Robustness?* British Journal of Mathematical and Statistical Psychology. 31(2) (1978),144-152. DOI:10.1111/j.2044-8317.1978.tb00581.x

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

Hidden Population Size Estimator of Poisson Lognormal Distribution for Capture-Recapture Data*

Orasa Nunkaw[1,†] and Jutamas Boonradsamee [2,‡]
[1]Department of Mathematics and Statistics, Faculty of Science and Digital Innovation,
Thaksin University, Banpro, Papayom, Phattalung, 93210, Thailand
[2]Faculty of Business Administration, Rajamangala University of Technology Srivijaya, Mung District,
Songkla, 90000, Thailand

Abstract

This research introduced the Expectation-Maximization (EM) algorithm approach to estimate the two parameters of the zero-truncated Poisson lognormal (ZTPLN) distribution. A population size estimator derived from the Poisson lognormal distribution was also proposed, offering a robust framework for modeling over-dispersed count data with heterogeneity. Comparisons with the maximum likelihood estimator of the Poisson distribution (MLEPoi), the maximum likelihood geometric distribution (Geo), and Chao's estimators revealed that the new estimator can be beneficially used as a true model.

Keywords: poisson lognormal distribution, capture-recapture, hidden population
2020 MSC: Primary 62F10; Secondary 62F40.

1 Introduction

Knowledge of the size of the target population is applicable across various domains. In ecology, population size data play a pivotal role in guiding decisions regarding habitat protection, fragmentation management, and the promotion of connectivity to sustain viable populations. In the social sciences, population size informs decision-making processes and addresses challenges in social, economic, and public health domains. However, in reality, certain groups of individuals are not easily accessible or identifiable through conventional means of data collection such as census surveys or official records. Capture-Recapture (CR) is a powerful tool for estimating the size of populations, particularly when dealing with elusive or difficult-to-count species or groups. This method has been successfully used in ecology and wildlife conservation to estimate the population sizes of animal species and study population dynamics [1-2]. Recently, this approach has been applied in various fields including the social

* This research was financially supported by the Ministry of Higher Education, Science, Research and Innovation
† Speaker. ‡ Corresponding author.
E-mail address: aorasa@tsu.ac.th (O. Nunkaw), Jutamas.r@rmutsv.ac.th (J. Boonradsamee).

sciences [3-7], public health, and epidemiology [8-10] to estimate the size of specific populations.

The CR approach involves repeated counts of a target population. Suppose that X_i, $i=1,2,3,\ldots,N$ is the number of times individual i is captured over m sampling occasions, and let $p_x = P(X_i = x)$. Similarly, let f_x denote the frequency of individuals captured exactly x times, where $x_i = 1,2,3,\ldots,m$ and m is the largest count. If $X_i = 0$ is not observed, the corresponding f_0 is unknown and might be estimated by its expected value Np_0, while p_0 is the probability of non-identifying in the sample process and needs to be estimated within the appropriate model. The Poisson distribution with parameters λ, and $p_x = \frac{\exp(\lambda)\lambda^x}{x!}$, can be selected as a basic model for counting data but it may not be suitable for CR data in heterogeneous populations. To account for a heterogeneous population in CR data, the Poisson parameter is often considered as an unobserved random variable, with a mixed distribution $h(\lambda)$ and a marginal distribution as:

$$p_x = \int_0^{\infty} \frac{\exp(-\lambda)\lambda^x}{x!} h(\lambda) dx ,$$

where the mixing distribution density $h(\lambda)$ is unknown [11]. One approach to model overdispersion involves exploring established parametric Poisson mixture models like the negative binomial distribution [12], geometric distribution [13], the Conway-Maxwell-Poisson distribution [14] or the Poisson lognormal (PLN) distribution [15-16]. The PLN distribution provides a robust framework for modeling over-dispersed count data with heterogeneity and highly skewed distributions but involves complex numerical methods for estimating the parameters. This study estimated the size of the target population using CR data. In a CR study, the unobserved counts may disappear during the counting procedure. Therefore, here, the Expectation-Maximization (EM) algorithm approach was introduced to estimate the two parameters of the zero-truncated Poisson Lognormal (ZTPLN) distribution. A population size estimator based on the PLN distribution was also proposed, which offered a robust framework for modeling over-dispersed count data and highly skewed distributions.

2 The Poisson Lognormal Distribution for Capture-Recapture Data

The Poisson distribution assumes that the variance is equal to the mean. However, in many real-world scenarios, particularly in count data, the variance often exceeds the mean, indicating overdispersion. The PLN distribution allows for overdispersion by introducing variability in the rate parameter via the lognormal distribution.

2.1 The Poisson Lognormal Distribution

Let the Poisson parameter λ of each individual follow the lognormal distribution as $h(\lambda;\mu,\sigma)=\frac{1}{\lambda\sigma\sqrt{2\pi}}\exp\left\{-\frac{(\log\lambda-\mu)^2}{2\sigma^2}\right\}$, where μ is the mean and σ is the standard deviation of the normal distribution Y, and $Y=\log(\lambda)$. The Poisson lognormal distribution (PLN) probability function can be depicted as:

$$p_x=\frac{1}{x!\sigma\sqrt{2\pi}}\int_0^{\infty}\lambda^{x-1}\exp(-\lambda)\exp\left\{-\frac{(\log\lambda-\mu)^2}{2\sigma^2}\right\}d\lambda,\quad x=0,1,2,... \tag{2.1}$$

The range of parameters for the PLN distribution are $\mu>0$ and $\sigma>0$. To explore the characteristics of the PLN model, plots illustrating the probability mass function of the PLN distribution with parameters μ and σ are shown in Figure 1.

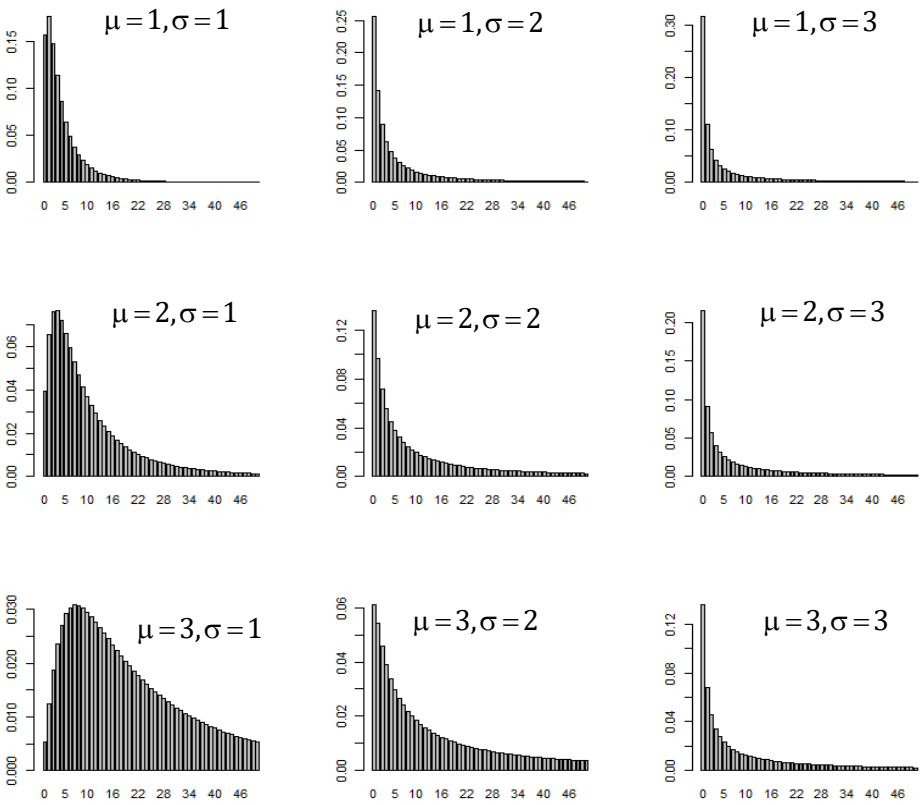


Figure 1: Simulated frequency distributions based on the PLN distribution with $PLN(\mu,\sigma)$

The expected value and variance of X are shown as follows:

$$E(X)=\exp(\mu+\frac{\sigma^2}{2}),$$

and

$$\mathrm{Var}(X) = \exp(\mu + \frac{\sigma^2}{2})M^*,$$

where $M^* = \left[1 + \left\{\exp\left(\mu + \frac{\sigma^2}{2}\right)\right\}\left\{\exp(\sigma^2) - 1\right\}\right]$. The point estimators of μ and σ^2 are obtained by the first and the second moments as $\hat{\mu} = \log\left(\frac{a^2}{\sqrt{b-a}}\right)$ and $\hat{\sigma}^2 = \log\left(\frac{b-a}{a^2}\right)$, where $a = \frac{1}{N}\sum_{x=0}^{m} x f_x$ and $b = \frac{1}{N}\sum_{x=0}^{m} x^2 f_x$ [15]. Let f_x be the frequency counts with value x times and m is the largest count. Then, the completeness likelihood function of the PLN distribution for capture-recapture data is given by

$$\begin{aligned} L_c(x;\mu,\sigma) &= \prod_{x=0}^{m} p_x^{f_x} \\ &= \prod_{x=0}^{m}\left[\frac{1}{x!\sigma\sqrt{2\pi}}\int_0^{\infty}\lambda^{x-1}\exp(-\lambda)\exp\left\{-\frac{(\log\lambda-\mu)^2}{2\sigma^2}\right\}\right]^{f_x} d\lambda. \end{aligned} \tag{2.2}$$

Evaluating the likelihood equation requires numerical integration, which in turn necessitates using an optimization technique to obtain the maximum likelihood function. Fortunately, the R programming offers the **poilogMLE()** function within the **poilog package** to streamline this process.

2.2 Population Size Estimator Based on the Zero-Truncated Poisson Lognormal Distribution

In capture-recapture data, zero counts are unknown and require estimation. Let $x = 0$ denote an individual that cannot be captured from the target population with probability p_0. Then, under the assumption of the zero-truncated Poisson lognormal (ZTPLN) distribution

$$p_x^+ = \frac{p_x}{1-p_0},$$

where

$$p_0 = \frac{1}{x!\sigma\sqrt{2\pi}}\int_0^{\infty}\frac{\lambda^x}{\exp(\lambda)-1}\exp\left\{-\frac{(\log\lambda-\mu)^2}{2\sigma^2}\right\}d\lambda. \tag{2.3}$$

Applying this equation to the Horvitz-Thompson method leads to a population size estimator based on the PLN distribution ($\hat{N}_{PLN}$) as:

$$\hat{N}_{PLN} = \frac{n}{1-\hat{p}_0}, \tag{2.4}$$

where $n=\sum_{x=1}^{m} xf_x$, and $\hat{p}_0$ can be calculated from (2.3) with estimators $\hat{\mu}$ and $\hat{\sigma}$. The two parameters are obtained by maximizing the log-likelihood function of the ZTPLN distribution.

$$\log L_{obs}(x;\mu,\sigma)=\sum_{x=1}^{m}\log\left[\frac{\frac{1}{x!\sigma\sqrt{2\pi}}\int_0^{\infty}\lambda^{x-1}\exp(-\lambda)B^{*}d\lambda}{1-\frac{1}{x!\sigma\sqrt{2\pi}}\int_0^{\infty}\frac{\lambda^{x}}{\exp(\lambda)-1}B^{*}d\lambda}\right]^{f_x}, \tag{2.5}$$

where $B^{*}=\exp\left\{-\frac{(\log\lambda-\mu)^2}{2\sigma^2}\right\}$. Equation (2.5) is then differentiated with respect to μ and σ, followed by setting both equations to zero. However, this does not yield a closed-form expression. One effective method for solving missing or hidden data is the Expectation-Maximization (EM) algorithm [17]. This iterative algorithm consists of two components: the Expectation step (E-step) and the Maximization step (M-step). In the first step, the algorithm attempts to estimate the missing data (zero counts) by replacing them with the conditional expected values, given frequency counts and the current parameters as:

$$\begin{aligned}\hat{f}_0 &= E(f_0\,|\,\text{observed};\mu,\sigma)\\ &= E(f_0\,|\,f_1,f_2,f_3,\ldots,f_m;\mu,\sigma)\\ &= Np_0.\end{aligned}$$

The size of population N can be estimated by $n+\hat{f}_0$, and the expected value of unobserved data can be estimated by $\hat{f}_0=(n+\hat{f}_0)p_0=(np_0)/(1-p_0)$. In the M-step, the associated completeness log-likelihood function of the PLN distribution, which includes unobserved $\hat{f}_0$ and observed data $(f_1,f_2,f_3,\ldots f_m)$, is required to estimate the new iterative parameters and the new population size estimator. Since there are no closed-form solutions for μ and σ under the PLN distribution, the **poilog package** in R programming is used for the EM algorithm.

2.3 The EM Algorithm

The EM algorithm technique under the maximum likelihood estimation of the PLN distribution is given as follows:

Step 0: Set $l=0$ and use the empirical moments $\hat{\mu}^{(1)}=\log\left(\frac{a^2}{\sqrt{b-a}}\right)$, $\hat{\sigma}^{2(1)}=\log\left(\frac{b-a}{a^2}\right)$, and $\hat{\sigma}^{(1)}=\sqrt{\hat{\sigma}^{2(1)}}$ as the initial parameters. Then, use the **poilog package** in R programming to estimate $p_0^{(1)}$ and achieve $\hat{f}_0^{(1)}=\frac{np_0^{(1)}}{1-p_0^{(1)}}$. The algorithm is repeated until the log-likelihood

function of the ZTPLN distribution in (2.5) converges to a constant with an acceptable error. Therefore, the initial value is set as

$$\log L_{obs}(x;\mu,\sigma)^{(l)} = -\infty. \tag{2.6}$$

Step 1: Substitute $\hat{f}_0^{(l)}$ in the completed frequency distribution table (Table 1) to calculate the new parameters $\hat{\mu}^{(l+1)}$ and $\hat{\sigma}^{(l+1)}$ using the maximum likelihood estimation.

Table 1: The frequency distribution

x	0	1	2	...	m
f_x	$\hat{f}_0^{(l)}$	f_1	f_2	...	f_m

The maximum likelihood estimators are computed using the **poilog package** in R programming, with the **poilogMLE** () function leading to new log-likelihood maximum estimators.

Step 2: Estimate the new unobserved probability, $\hat{p}_0^{(l+1)}$, using the **dpoilog** () function in R programming, and then estimate the new unobserved frequency and population size:

$$f_0^{(l+1)} = \frac{n\hat{p}_0^{(l+1)}}{1-\hat{p}_0^{(l+1)}} \tag{2.7}$$

and

$$\hat{N}_{PLN}^{(l+1)} = \frac{n}{1-\hat{p}_0^{(l+1)}}. \tag{2.8}$$

Step 3: Check the algorithm condition by substituting $\hat{\mu}^{(l+1)}$ and $\hat{\sigma}^{(l+1)}$ into the log-likelihood of the ZTPLN distribution, $\log L_{obs}(x;\mu,\sigma)^{(l+1)}$, and compare the difference in values as

$$dif = |\log L_{obs}(x;\mu,\sigma)^{(l)} - \log L_{obs}(x;\mu,\sigma)^{(l+1)}| < 0.0001. \tag{2.9}$$

Then, setting $l = l+1$. If $dif \geq 0.0001$ return to step 1 to update the new maximum likelihood estimators. The algorithm is repeated until the log-likelihood function of the ZTPLN converges to a constant with an acceptable error.

2.4 Confidence Interval Estimation

The confidence interval of the PLN estimator is constructed using the imputed bootstrapping method introduced by [18]. One advantage of utilizing the bootstrap confidence interval is its flexibility and robustness against outliers, skewed distributions, or small sample sizes.

Moreover, this method does not require a variance formula estimator. The procedure for constructing a 95% imputed bootstrap confidence interval is as follows:

$$\hat{N} \pm Z_{0.975} \widehat{SE}(\hat{N}), \tag{2.10}$$

where $\widehat{SE}(\hat{N})$ denotes the standard error of $\hat{N}$ from the imputed bootstrap method, and $Z_{0.975} = 1.96$.

2.5 Alternative Estimators

Comparison with some well-known estimators is then performed to achieve a better judgment of the proposed estimator.

2.5.1 Maximum Likelihood Estimator under the Poisson Distribution

Suppose that the capture-recapture count X follows a Poisson distribution with density. The population size ($\hat{N}_{MLEPoi}$) can then be estimated as

$$\hat{N}_{MLEPoi} = \frac{n}{1-\hat{p}_0} = \frac{n}{1-\exp(-\hat{\lambda}_{ZTPoi})}, \tag{2.11}$$

The maximum likelihood estimator $\hat{\lambda}_{ZTPoi}$ can be calculated using the EM algorithm approach based on the zero-truncated Poisson lognormal distribution. A variance estimation of $\hat{N}_{MLEPoi}$, $\widehat{Var}(\hat{N}_{MLEPoi})$, is given as:

$$\widehat{Var}(\hat{N}_{MLEPoi}) = \frac{\hat{N}_{MLEPoi}}{\left[\exp\frac{\sum_{x=1}^{m} x f_x}{\hat{N}_{MLEPoi}} - \frac{\sum_{x=1}^{m} x f_x}{\hat{N}_{MLEPoi}} - 1\right]}. \tag{2.12}$$

For further details, see [19,20].

2.5.2 Chao's Lower Bound Estimator

The Chao estimator [11] estimates species richness based on a vector or matrix of abundance data for an unobserved heterogeneous population. If counts (X) are assumed to be modeled from a mixed Poisson distribution with arbitrary mixing density $h(\lambda)$; then,

$$p_x = \int_0^{\infty} \frac{\exp(-\lambda)\lambda^x}{x!} h(\lambda) d\lambda, \text{ where } x = 0,1,2,... \tag{2.13}$$

Using the Cauchy-Schwarz inequality for any two random variables X and Y, gives $E[(XY)]^2 \le E(X^2)E(Y^2)$. The Chao lower bound population size estimator ($\hat{N}_{Chao}$) can be written as:

$$\hat{N}_{Chao} = n + \frac{f_1^2}{2f_2}. \tag{2.14}$$

Only the observed frequencies f_1 and f_2 are used in the Chao lower bound estimator. A modified version of variance of the Chao estimator, $\widehat{Var}(\hat{N}_{Chao})$ [21] can be given as:

$$\widehat{\mathrm{Var}}(\hat{N}_{\mathrm{Chao}}) = \left(\frac{1}{4}\right)^2 \frac{f_1^2}{f_2^3} + \frac{f_1^3}{f_2^2} + \left(\frac{1}{2}\right) \frac{f_1^2}{f_2}. \tag{2.15}$$

2.5.3 Maximum Likelihood Estimator under the Geometric Distribution

The geometric distribution comprises a mixture of the Poisson distribution with an exponential density, leading to the associated marginal density with parameter p, and obtained as $p_x(p) = (1-p)^x p$, where $x = 0,1,2,\ldots$. A maximum likelihood estimator based on the geometric distribution for a heterogeneous population ($\hat{N}_{\mathrm{Geo}}$) was proposed by [22] as:

$$\hat{N}_{\mathrm{Geo}} = \frac{n}{1-\hat{p}_0} = \frac{n}{1-\hat{p}} = \frac{nS}{S-n}, \tag{2.16}$$

where $S = \sum_{x=1}^{m} x f_x$. The variance estimation of this estimator can be calculated from

$$\widehat{\mathrm{Var}}(\hat{N}_{\mathrm{Geo}}) = \frac{S^2 n^2}{(S-n)^3}. \tag{2.17}$$

3 Simulation Study

3.1 Simulation Scenarios

A simulation study is used to assess the performance of population size estimators across simulated data sets. The proposed estimator was compared with some well-known estimators highlighted in the previous section. The simulation scheme was designed by generating data following the Poisson lognormal distribution with two parameters $\mu = \{1, 1.5, 2, 2.5, 3, 3.5\}$ and $\sigma = \{1, 2, 3\}$. The population size was fixed as $N = 500$ for a small, $N = 1{,}000$ for a medium, and $N = 5{,}000$ for a large size study. $T = 1{,}000$ data sets were drawn from each simulation scenario and any occurrences of zero counts were truncated before estimating the population size, with the proposed estimator evaluated in terms of accuracy and precision. Let $\hat{N}_{(t)}$ denote the population size estimated value from replication t^{th}, where $t = 1,2,3,\ldots,T$. Then, the expected value of the population size estimator can be achieved by $E(\hat{N}) = \frac{1}{T}\sum_{t=1}^{T} \hat{N}_{(t)}$. The relative bias of population size estimator was selected to investigate accuracy, defined as $\mathrm{Rbias}(\hat{N}) = \frac{1}{N}\left\{E(\hat{N}) - N\right\}$. The precision criteria were defined as the relative variance; $\mathrm{R\,var}(\hat{N}) = \frac{1}{N^2}\left[\frac{1}{T-1}\sum_{t=1}^{T}\left(\hat{N}_{(t)} - E(\hat{N})\right)^2\right]$. Bias and precision describe the estimator performance, and the relative root mean square error was used to measure the performance of the population size estimator as $\mathrm{RRMSE}(\hat{N}) = \frac{1}{N}\sqrt{\mathrm{Var}(\hat{N}) + \left\{\mathrm{bias}(\hat{N})\right\}^2}$, where $\mathrm{bias}(\hat{N}) = E(\hat{N}) - N$ and $\mathrm{Var}(\hat{N}) = \frac{1}{T}\sum_{t=1}^{T}\left\{\hat{N}_t - E(\hat{N})\right\}^2$.

3.2 Simulation Results

The PLN estimator provides an asymptotically unbiased estimate of the population size N, with results displayed in Figures 2 and 3. Both the MLEPoi and Chao estimators underestimated the population size across all scenarios under the PLN distribution although their bias reduced when the location parameter, μ, increased. The Geo estimator also tended to slightly underestimate when the dispersion parameter was small. However, it showed promise in estimating population size under the PLN distribution for $\sigma = 1$. The relative variance of all estimators showed less dispersion when the parameter μ and the population size increased. To select the most suitable estimator under the Poisson lognormal distribution, a balance between bias and precision is necessary. The RRMSE was used to measure and compare the four estimators. As illustrated in Figure 4, the PLN was the optimal choice for estimating the target population size, as it consistently exhibited the lowest RRMSE across all scenarios. The Chao estimator showed promise as an alternative choice for small to medium-sized populations, as its estimated results closely aligned with those of the proposed estimator.

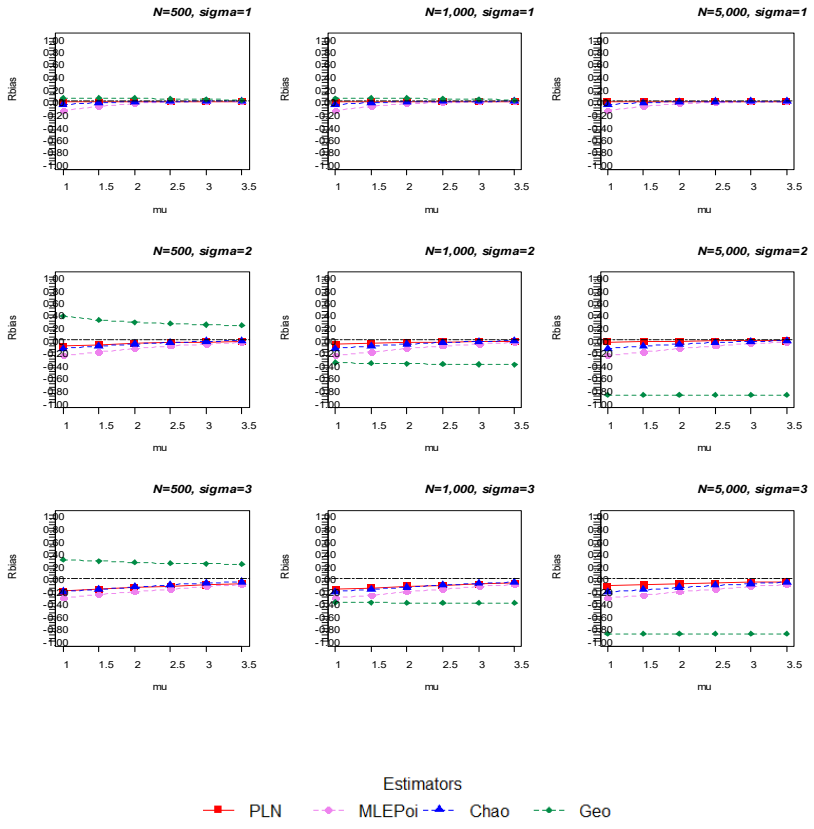


Figure 2: Relative bias (Rbias) of the four estimators for counts drawn from $PLN(\mu,\sigma)$

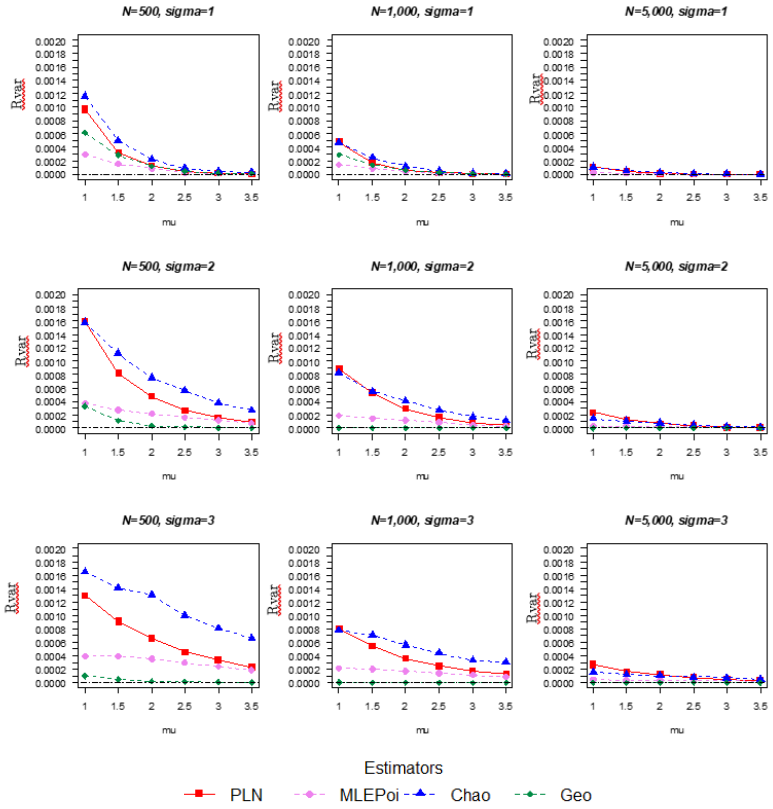


Figure 3: Relative variance (Rvar) of the four estimators for counts drawn from PLN(μ,σ)

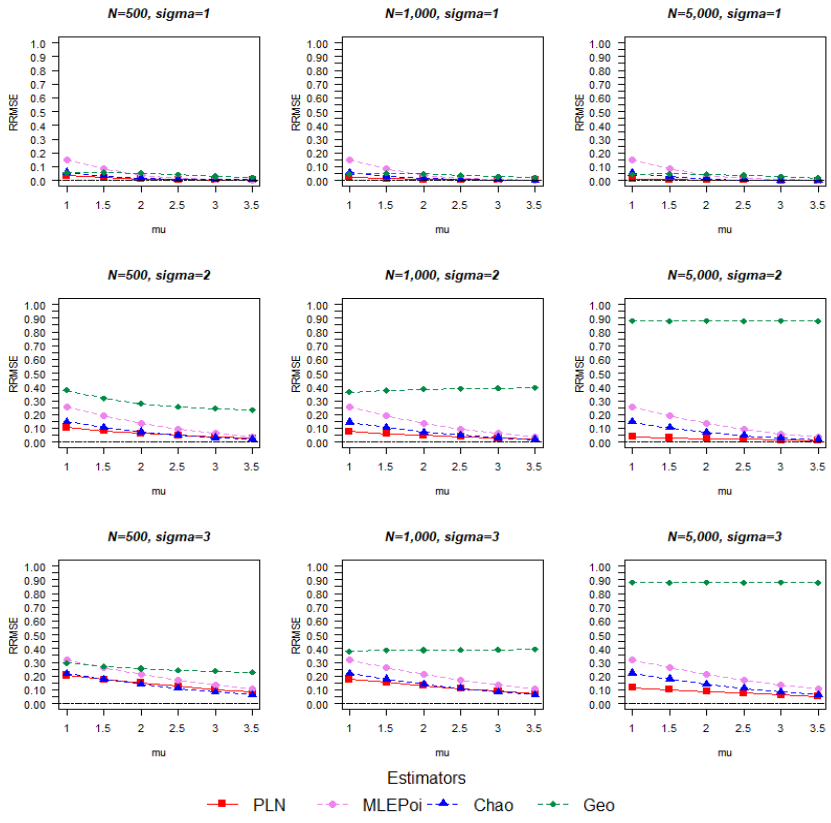


Figure 4: Relative root mean error (RRMSE) of the four estimators for counts drawn from PLN(μ,σ)

4 Real Data Example

In this section, different estimators were applied to a real-data example using Malaysian butterfly's data. The dataset originated from [23] and has been explored by many authors [15,16]. Malaysian butterflies comprise 620 species, totaling 9,029 individuals, as presented in Table 2. The EM algorithm was employed to estimate the two parameters of the ZTPLN distribution, with initial values set as the empirical moments: $\hat{\mu}^{(0)} = 1.7485$ and $\hat{\sigma}^{(0)} = 0.8908$. The algorithm involved 18 repeated steps, during which the log-likelihood function of the ZTPLN distribution converged to a constant with an acceptable error. The maximum log-likelihood estimation with the EM algorithm provided results as $\hat{\mu} = 1.1219$ and $\hat{\sigma} = 1.4817$. The estimated number of species in the population under the PLN estimator was 721. This was lower than the estimate obtained using a numerical integral method (742) [15] and an extension of MacArthur's broken stick model (816) [16].

Table 2: Frequency distribution of Malaysian butterflies

x	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
fx	118	74	44	24	29	22	20	19	20	15	12	14	6	12	6	9	9	6	10	10	
x	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41+
fx	11	5	3	3	5	4	8	3	3	2	5	4	7	4	5	3	3	3	3	1	56

The results of population size estimates across various mixed Poisson distributions are presented in Table 3 and Figure 5. The MLEPoi estimator yielded the lowest number of species because it is designed for homogeneous populations and often underestimates population size heterogeneity [24]. The PLN estimator demonstrated a superior fit compared to the Chao and Geo estimators.

Table 3: Population size estimates

Estimator	$\hat{f}_0$	$\hat{N}$	$\widehat{SE(N)}$	95% of CI
MLEPoi ($\hat{\lambda}$ =8.5435)	1	621	0.35	620 -- 622
Chao	95	659	20.27	619 -- 699
Geo ($\hat{p}$ =0.106)	67	631	9.18	613 -- 649
PLN ($\hat{\mu}$ =1.1219, $\hat{\sigma}$ =1.4817)	157	721	19.70	683 -- 760

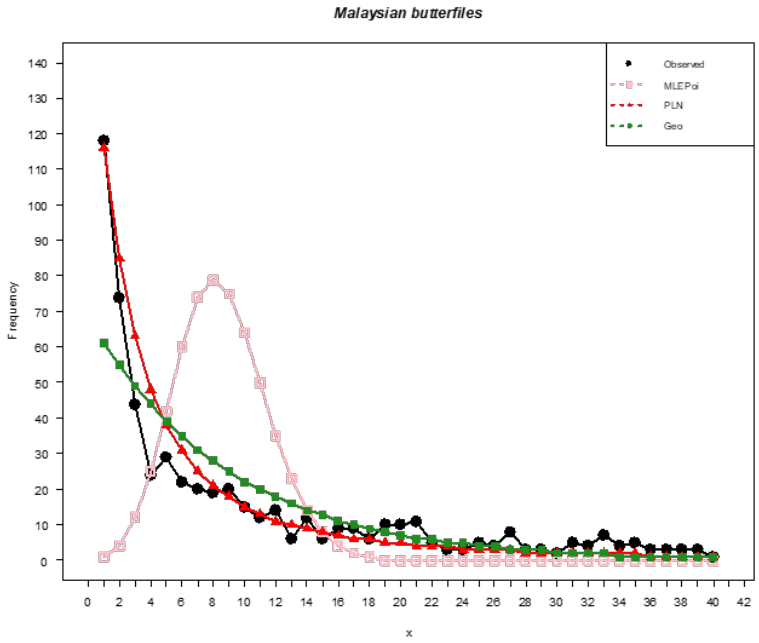


Figure 5: Estimated population size versus observed frequencies

5 Conclusions

Various estimators in the capture-recapture field have widespread applications in various domains. This study proposed a modified approach for estimating population size under a specific of heterogeneity based on the Poisson lognormal distribution. The EM algorithm procedure was employed to estimate the two parameters of the ZTPLN distribution, while the accuracy and precision of the method compared to other well-known estimators were also assessed. The PLN estimator emerged as the optimal choice for estimating the target population size, consistently demonstrating the lowest RRMSE values across all scenarios. The Chao estimator showed promise as an alternative choice for small to medium-sized populations, with estimated results that closely aligned with the proposed estimator. For the future work, it would be worthwhile to explore additional data structures, including those featuring varied covariate variable types or assuming different sampling distributions.

Acknowledgment. The authors would like to thank the reviewers for their very helpful comments and suggestions, which considerably improved this paper. Thanks are also due to the Ministry of Higher Education, Science, Research and Innovation for financial support.

References

[1] J. D. Nichols, *Capture-recapture models*, BioScience, **42**(2).(1992), 94-102.

[2] A. Chao, P. K. Tsay, S. H. Lin, S. H., Shau, W. Y, and D. Y. Chao. *The applications of capture-recapture models to epidemiological data*. Statistics in medicine, **20**(20) (2001), 3123-3157.

[3] A. M. Coumans, M. J. L. F. Cruyff, P. G. Van der Heijden, J. R. L. M. Wolf and H. J. S. I. R. Schmeets. *Estimating homelessness in the Netherlands using a capture-recapture approach*. Social Indicators Research, **130** (2017), 189-212.

[4] D. DI CECCO, A. Tancredi and T. Tuoto, *Analyzing different causes of one–inflation in capture recapture models for criminal populations*. In Sis 2022. 51st meeting of the Italian Statistical Society. Book of short papers. (2022), 1607-1612

[5] M. Mwale, K. Mwangilwa, E. Kakoma, and K. Iaych, *Estimation of the completeness of road traffic mortality data in Zambia using a three source capture recapture method*. Accident Analysis & Prevention, **186** (2023), 107048.

[6] M. E. Piatek, and D. Böhning, *Deriving a zero-truncated modelling methodology to analyse capture–recapture data from self-reported social networks*. Metron, (2023), 1-26.

[7] P. G. Van der Heijden, M. Cruyff, and D. Böhning, *Capture recapture to estimate criminal populations*. Encyclopedia of criminology and criminal justice. Berlin: Springer, (2014), 267-276.

[8] D. Böhning, B. Suppawattanabodee, W. Kusolvisitkul, and C. Viwatwongkasem, *Estimating the number of drug users in Bangkok 2001: A capture–recapture approach using repeated entries in one list*. European Journal of Epidemiology, **19**, (2004). 1075-1083.

[9] A. Barchuk, R. Tursun-Zade, E. Nazarova, Y. Komarov, E. Tyurina, Y. Tumanova, and A. Znaor, *Completeness of regional cancer registry data in Northwest Russia 2008-2017*. BMC cancer, **23**(1) (2023). 994.

[10] A. D. Kiakalayeh, M. R. Taramsari, R. Mohammadi, S. D. Kiakalayeh, and H. Kavakpour, *Comparison of the capture-recapture method and seroprevalence survey for estimation of COVID-19 prevalence in the Islamic Republic of Iran*. Eastern Mediterranean Health Journal, **29**(2) (2023), 126-131.

[11] A. Chao, *Estimating the population size for capture-recapture data with unequal catchability*. *Biometrics*, (1987), 783-791.

[12] K. Lanumteang and D. Böhning, *An extension of Chao's estimator of population size based on the first three capture frequency counts*. Computational Statistics and Data Analysis, **55**(7) (2011), 2302-2311.

[13] O. Anan, D. Böhning and A. Maruotti *On the Turing estimator in capture–recapture count data under the geometric distribution*. Metrika, **82** (2019), 149-172.

[14] O. Anan D. Böhning and A. Maruotti, *Population size estimation and heterogeneity in capture–recapture data: a linear regression estimator based on the Conway–Maxwell–Poisson distribution*, Statistical Methods and Applications, **26** (2017), 49-79.

[15] R. Izsák, *Maximum likelihood fitting of the Poisson lognormal distribution,* Environmental and Ecological Statistics, **15** (2008), 143-156.

[16] M. G. Bulmer, *On fitting the Poisson lognormal distribution to species-abundance data. Biometrics*, **30(**1) (1974), 101-110.

[17] A. P. Dempster, N. M. Laird and, D. B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm.* Journal of the royal statistical society: series B (methodological), **39**(1) (1977), 1-22.

[18] O. Anan, D. Böhning and A. Maruotti, *Uncertainty estimation in heterogeneous capture–recapture count data.* Journal of Statistical Computation and Simulation, **87**(10) (2017), 2094-2114.

[19] A. Chao, and S. M. Lee, *Estimating the number of classes via sample coverage.* Journal of the American statistical Association, **87**(417) (1992), 210-217.

[20] A. Chao, S. M. Lee, and S. L. Jeng, *Estimating population size for capture-recapture data when capture probabilities vary by time and individual animal.* International Biometric Society, **48**(1) (1992), 201-216.

[21] D. Böhning, *A simple variance formula for population size estimators by conditioning.* Statistical Methodology, **5**(5) (2008), 410-423.

[22] S. Niwitpong and D. Böhning and P. G. Van der Heijden, *Capture–recapture estimation based upon the geometric distribution allowing for heterogeneity*, Metrika,**76**, (2013),495-519.

[23] A.S. Corbet, *The distribution of butterflies in Malay Peninsula. Proc Roy Entomol Soc Lond (A)*, **16** (1942), 101-116.

[24] O. Anan, *Capture-recapture modelling for zero-truncated count data allowing for heterogeneity*, Doctoral dissertation, University of Southampton, 2016.

Proceedings of the
28th Annual Meeting in Mathematics (AMM 2024)
Department of Mathematics Statistics and Computer,
Faculty of Science, Ubon Ratchathani University,
Thailand

ความรู้ความเข้าใจและพฤติกรรมการป้องกันโรคโควิด-19 หลังการระบาดใหญ่ของประชาชนในจังหวัดสุราษฎร์ธานี

อัญชุลี ณ ตะกั่วทุ่ง[1] ศุภชัย ดำคำ[1,†,‡] เกตุกนก หนูดี[1] และ กันยากร อ่อนรักษ์[1]

[1]สาขาวิชาคณิตศาสตร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏสุราษฎร์ธานี 84100

บทคัดย่อ

การวิจัยเชิงสำรวจนี้มีวัตถุประสงค์เพื่อ 1) ศึกษาความรู้ความเข้าใจเกี่ยวกับโรคและการป้องกันโรคโควิด-19 2) ศึกษาพฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 และ 3) ศึกษาความสัมพันธ์ระหว่างข้อมูลทั่วไปกับระดับความรู้ความเข้าใจเกี่ยวกับโรคและการป้องกันโรค พฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 ของประชาชนในจังหวัดสุราษฎร์ธานี กลุ่มตัวอย่าง คือประชาชนที่อาศัยและมีภูมิลำเนาในจังหวัดสุราษฎร์ธานี จำนวน 453 คน ใช้วิธีสุ่มตัวอย่างแบบโควตา เครื่องมือที่ใช้เป็นแบบสอบถามความรู้ความเข้าใจและพฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 ที่มีค่า IOC อยู่ในช่วง 0.67-1.00 และมีค่าความเชื่อมั่นด้วยค่าสัมประสิทธิ์แอลฟาของครอนบาค เท่ากับ 0.70 สถิติที่ใช้ในการวิเคราะห์ข้อมูล ได้แก่ การนับความถี่ ค่าร้อยละ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และการทดสอบไคสแควร์

ผลการวิจัยพบว่า กลุ่มตัวอย่างที่ตอบแบบสอบถาม ส่วนใหญ่เป็นเพศหญิง (ร้อยละ 78.15) มีอายุในช่วง 50-59 ปี (ร้อยละ 31.79) สำเร็จการศึกษาในระดับมัธยมศึกษา (ร้อยละ 41.28) มากที่สุด ส่วนใหญ่ประกอบอาชีพเกษตรกรรม (ร้อยละ 29.58) ส่วนใหญ่มีประวัติการได้รับวัคซีนโควิด-19 จำนวน 3 เข็ม (ร้อยละ 47.46) มีความรู้ความเข้าใจเกี่ยวกับโรคและการป้องกันโรคโควิด-19 ระดับปานกลาง (ร้อยละ 79.25) มีพฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 ในภาพรวมมีการปฏิบัติในทุกครั้งมากที่สุด (ร้อยละ 67.33) เมื่อพิจารณารายข้อ พบว่า มีพฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 ดีที่สุดคือ การสวมหน้ากากอนามัยทุกครั้งเมื่อไม่สบายหรือออกจากบ้าน รองลงมาคือ การสังเกตลักษณะอาการที่เกี่ยวข้องกับการติดเชื้อโรคโควิด-19 ของตนเองเป็นประจำ ในการศึกษาความสัมพันธ์ระหว่างข้อมูลทั่วไปกับระดับความรู้ความเข้าใจเกี่ยวกับโรคและการป้องกันโรค พฤติกรรมป้องกันการแพร่ระบาดของโรคโควิด-19 พบว่า เพศและประวัติการได้รับวัคซีนโควิด-19 ส่งผลต่อระดับความรู้ความเข้าใจเกี่ยวกับโรคและการป้องกันโรคโควิด-19 ในขณะที่ข้อมูลเพศ อายุ ระดับการศึกษา ส่งผลต่อระดับพฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 ที่ระดับนัยสำคัญ 0.05

คำสำคัญ: โควิด-19, ความรู้ความเข้าใจ, พฤติกรรมการป้องกัน

---

* งานวิจัยเรื่องนี้ได้รับทุนสนับสนุนการวิจัยจาก มหาวิทยาลัยราชภัฏสุราษฎร์ธานี

†ผู้นำเสนอ ‡ผู้แต่งหลัก

อีเมล: unchulee.nat@sru.ac.th (อัญชุลี ณ ตะกั่วทุ่ง), supachai.dam@sru.ac.th (ศุภชัย ดำคำ), ketkanok.noo@sru.ac.th (เกตุกนก หนูดี), kanyakon.onr@sru.ac.th (กันยากร อ่อนรักษ์)

1 บทนำ

โรคโควิด-19 เกิดขึ้นครั้งแรกในประเทศจีน เมืองอู่ฮั่น มณฑลหูเป่ย เมื่อวันที่ 30 ธันวาคม 2562 จากนั้นพบผู้ป่วยในหลายพื้นที่ของประเทศจีน สำหรับการติดเชื้อนอกประเทศจีนนั้น ได้ถูกรายงานเป็นครั้งแรกในวันที่ 13 มกราคม 2563 ในประเทศไทย และได้กระจายไปยังทวีปต่างๆ ทั่วโลก เมื่อการระบาดขยายวงกว้างออกไปในหลายประเทศมากขึ้น วันที่ 11 มีนาคม 2563 WHO จึงได้ประกาศให้การแพร่ระบาดของโรคโควิด-19 เป็น "การระบาดใหญ่" หรือ Pandemic [1] ซึ่งในปัจจุบันองค์การอนามัยโลกยังคงติดตามและเฝ้าระวังการแพร่ระบาดนี้อย่างต่อเนื่อง โรคโควิด-19 เป็นโรคที่เกิดการติดเชื้อไวรัส ที่มีชื่อว่า Severe Acute Respiratory Syndrome Coronavirus 2 หรือ SARS-CoV-2 ไวรัสสามารถแพร่กระจายจากปากหรือจมูกของผู้ติดเชื้อในอนุภาคของเหลวขนาดเล็ก เมื่อไอ จาม พูด ร้องเพลง หรือหายใจ อนุภาคเหล่านี้มีตั้งแต่ละอองในทางเดินหายใจที่ใหญ่ขึ้นไปจนถึงละอองลอยที่เล็กกว่า ผู้ติดเชื้อไวรัสส่วนใหญ่จะมีอาการป่วยทางเดินหายใจเล็กน้อยถึงปานกลางและหายเป็นปกติโดยไม่ต้องรับการรักษาเป็นพิเศษ แต่บางรายอาจมีอาการป่วยหนักและต้องได้รับการดูแลจากแพทย์ นอกจากนี้ในผู้สูงอายุและผู้ที่มีโรคประจำตัว เช่น โรคหัวใจและหลอดเลือด เบาหวาน โรคทางเดินหายใจเรื้อรัง หรือมะเร็งมีแนวโน้มที่จะเจ็บป่วยรุนแรง อย่างไรก็ตามผู้ป่วยอาจมีอาการป่วยหนักหรือเสียชีวิตได้ทุกวัย [2] การระบาดในไทยพบการแพร่ระบาดหลายสายพันธุ์ โดยสายพันธุ์ที่ระบาดในประเทศไทยในปลายปี 2566 จนถึงต้นปี 2567 นี้ เป็นไวรัส SARS-CoV-2 สายพันธุ์ omicron JN.1 [3]

จากข้อมูลขององค์การอนามัยโลก วันที่ 1 เมษายน 2567 พบผู้ติดเชื้อทั่วโลก 704,539,975 คน ผู้เสียชีวิต 7,008,958 คน โดยพบผู้เชื้อสะสมสูงสุด 3 อันดับแรก คือ สหรัฐอเมริกา (ผู้ติดเชื้อ 111,765,841 คน เสียชีวิต 1,218,840 คน) อินเดีย (ผู้ติดเชื้อ 45,034,146 คน เสียชีวิต 533,547 คน) และฝรั่งเศส (ผู้ติดเชื้อ 40,138,560 คน เสียชีวิต 167,642 คน) สำหรับประเทศไทยอยู่ในลำดับที่ 33 โดยพบผู้ติดเชื้อยืนยันสะสม 4,769,277 ราย เสียชีวิตสะสม 34,581 ราย [4] นอกจากนี้รายงานสถานการณ์โรคโควิด-19 ของกระทรวงสาธารณสุข ในช่วงปี 2566 จนถึงเดือนมีนาคม 2567 พบว่ายังคงพบการติดเชื้อของโรคโควิด-19 อย่างต่อเนื่อง และยังมีผู้ป่วยจำเป็นต้องพักรักษาตัวในโรงพยาบาล ผู้ป่วยปอดอักเสบรุนแรง ผู้ป่วยที่ต้องใช้เครื่องช่วยหายใจ รวมถึงผู้เสียชีวิตเพิ่มจำนวนขึ้น ทั้งนี้ยังไม่มีข้อมูลบ่งชี้ว่าโรคโควิด-19 สายพันธุ์ที่ระบาดในปัจจุบัน ก่อให้เกิดความเจ็บป่วยรุนแรงเพิ่มขึ้น แต่เชื้อมีความสามารถในการแพร่ระบาดได้รวดเร็วขึ้น อีกทั้งประชาชนปฏิบัติตามมาตรการป้องกันตนเองลดลง เป็นเหตุให้จำนวนผู้ป่วยเพิ่มขึ้น สำหรับการป้องกันโรค องค์การอนามัยโลกแนะนำให้บุคคลที่เสี่ยงต่อการติดเชื้อและมีอาการรุนแรง เข้ารับการฉีดวัคซีนโควิด-19 อย่างน้อยหนึ่งเข็มและรับเข็มต่อไปห่างจากเข็มแรก 6-12 เดือน ร่วมกับการปฏิบัติตนตามมาตรการป้องกันการติดเชื้ออย่างสม่ำเสมอ เช่น การสวมหน้ากากอนามัย การปิดปาก ปิดจมูกเมื่อไอหรือจาม และการล้างมือเป็นประจำ [5]

อย่างไรก็ตามการผ่อนคลายมาตรการและข้อจำกัดต่างๆ เพื่อให้ประชาชนและผู้ประกอบการสามารถดำรงชีวิตได้ใกล้เคียงกับปกติ ทั้งที่โรคโควิด-19 ยังเป็นโรคที่มีผู้ป่วยและผู้เสียชีวิตอย่างต่อเนื่อง ทำให้การศึกษาการแพร่ระบาดของโรคโควิด-19 ยังเป็นหัวข้อที่น่าสนใจ ในปี พ.ศ. 2564 จันทิมา ห้าวหาญ และคณะ [6] ได้ศึกษาความรู้ความเข้าใจและพฤติกรรมการปฏิบัติตนเกี่ยวกับการป้องกันโรคโควิด-19 ของประชาชนในจังหวัดภูเก็ต ผลการศึกษาพบว่าระดับความรู้ความเข้าใจเกี่ยวกับการป้องกันโรคโควิด-19 ของประชาชนในจังหวัดภูเก็ต อยู่ในระดับมาก ซึ่งกลุ่มตัวอย่างส่วนใหญ่มีความรู้ใน ประเด็นที่ 1 ด้านความรู้ทั่วไปเกี่ยวกับโรคโควิด-19 มากที่สุด คือ ไวรัสโคโรนา

สายพันธุ์ใหม่ 2019 เหมือนไวรัสโรคทางเดินหายใจตะวันออกกลางและโรคซาร์ส และพฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 ของประชาชนในจังหวัดภูเก็ต อยู่ในระดับมากที่สุด มีค่าเฉลี่ยของข้อที่มีการปฏิบัติน้อยที่สุด คือ การสวมหน้ากากอนามัยทุกครั้งเมื่อไม่สบายหรือออกจากบ้าน ในปีเดียวกัน บงกช โมระสกุลและคณะ [7] ได้ศึกษาความรู้และพฤติกรรมการป้องกันโรคโควิด-19 ของนักศึกษาพยาบาลชั้นปีที่ 1 วิทยาลัยนานาชาติเซนต์เทเรซาและวิทยาลัยเซนต์หลุยส์ ผลการศึกษานี้ใช้เป็นแนวทางในการจัดโปรแกรมการให้ความรู้เกี่ยวกับโรคโควิด-19 ให้กับนักศึกษาพยาบาลชั้นปีที่ 1 ให้ดียิ่งขึ้น เพื่อสามารถป้องกันการติดเชื้อโควิด-19 ได้เมื่อต้องอยู่ในชุมชนระหว่างการเรียนออนไลน์ช่วงที่มีการระบาดของโรคโควิด-19 และแนวทางการเสริมความรู้เมื่อสามารถขึ้นฝึกปฏิบัติการพยาบาลในชั้นปีที่ 2 ต่อมาในปี 2565 กัมปนาท โคตรพันธ์ และคณะ [8] ได้ศึกษาความสัมพันธ์ระหว่างความรอบรู้ด้านสุขภาพกับพฤติกรรมการป้องกันโรคติดเชื้อไวรัสโคโรนา 2019 ของประชาชนในจังหวัดมุกดาหาร จากผลการศึกษาพบว่ากลุ่มตัวอย่างมีความรอบรู้ด้านสุขภาพในการป้องกันควบคุมโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับปานกลาง (ร้อยละ 65.4) พฤติกรรมการป้องกันโรคติดเชื้อไวรัสโคโรนา 2019 อยู่ในระดับปานกลาง (ร้อยละ 55.3) และภาพรวมความรอบรู้ด้านสุขภาพมีความสัมพันธ์ทางบวกระดับปานกลางกับพฤติกรรมการป้องกันโรคติดเชื้อไวรัสโคโรนา 2019 อย่างมีนัยสำคัญทางสถิติ ($r = 0.522$, p-value < 0.001) ผลการวิจัยนี้หน่วยงานสาธารณสุขสามารถนำไปเป็นข้อมูลพื้นฐานในการออกแบบกิจกรรมพัฒนาความรอบรู้ด้านสุขภาพเพื่อส่งเสริมพฤติกรรมการป้องกันโรคสำหรับประชาชนและกลุ่มเสี่ยงได้

จังหวัดสุราษฎร์ธานี เป็นศูนย์กลางทางเศรษฐกิจของกลุ่มจังหวัดภาคใต้ตอนบน มีผู้คนเดินทางจากภายนอกเข้ามาในจังหวัดเป็นจำนวนมาก รวมทั้งยังเป็นเมืองท่องเที่ยวที่มีนักท่องเที่ยวทั้งชาวไทยและชาวต่างชาติเข้ามาตลอดปี ซึ่งเป็นปัจจัยหนึ่งที่ส่งผลต่อการแพร่ระบาดของโรคโควิด-19 การศึกษาครั้งนี้จึงมุ่งศึกษาความรู้ความเข้าใจเกี่ยวกับโรคและการป้องกันโรคโควิด-19 รวมถึงพฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 หลังการระบาดใหญ่ของประชาชนในจังหวัดสุราษฎร์ธานี เพื่อเป็นแนวทางในพัฒนาระบบการป้องกันการแพร่ระบาดของโรคโควิด-19 ต่อไป

2 วัตถุประสงค์การวิจัย

1. ศึกษาความรู้ความเข้าใจเกี่ยวกับโรคและการป้องกันโรคโควิด-19 ของประชาชนในจังหวัดสุราษฎร์ธานี
2. ศึกษาพฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 ของประชาชนในจังหวัดสุราษฎร์ธานี
3. ศึกษาความสัมพันธ์ระหว่างข้อมูลทั่วไปกับความรู้ความเข้าใจเกี่ยวกับโรคและการป้องกันโรค พฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 ของประชาชนในจังหวัดสุราษฎร์ธานี

3 วิธีการดำเนินการวิจัย

การวิจัยเชิงสำรวจนี้ มีวิธีการดำเนินการวิจัย ดังนี้

3.1 ประชากรและตัวอย่าง

ประชากรที่ใช้ในการวิจัย คือ ประชาชนที่อาศัยและมีภูมิลำเนาในอำเภอเมืองสุราษฎร์ธานี จังหวัดสุราษฎร์ธานี จากข้อมูลประชากร ณ เดือนธันวาคม พ.ศ. 2565 จำนวน 189,816 คน

กลุ่มตัวอย่าง คือ ประชาชนที่อาศัยและมีภูมิลำเนาในอำเภอเมืองสุราษฎร์ธานี จังหวัดสุราษฎร์ธานี ไม่น้อยกว่าจำนวน 400 คน ที่ได้จากกำหนดขนาดตัวอย่างโดยใช้ Yamane (1967) [10] และใช้การเลือกตัวอย่างแบบโควต้า (Quota Sampling) ของประชากรจำแนกตามตำบลในเขตอำเภอเมืองสุราษฎร์ธานี

3.2 เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยเป็นแบบสอบถามความรู้ความเข้าใจและพฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 ของประชาชนจังหวัดสุราษฎร์ธานี จำนวน 3 ตอน ประกอบด้วย

ตอนที่ 1 ข้อมูลทั่วไปของผู้ตอบแบบสอบถาม เป็นแบบตรวจสอบรายการ (Checklist) จำนวน 5 ข้อ ได้แก่ เพศ อายุ ระดับการศึกษา อาชีพ และประวัติการได้รับวัคซีนโควิด-19

ตอนที่ 2 ความรู้ความเข้าใจเกี่ยวกับโรคและการป้องกันโรคโควิด-19 มีจำนวน 30 ข้อ แบ่งเป็น 2 ประเด็น คือ 1) ความรู้ความเข้าใจเกี่ยวโรคโควิด-19 และ 2) ความรู้ความเข้าใจเกี่ยวกับการป้องกันโรคโควิด-19 ซึ่งมีลักษณะแบบเลือกตอบ กำหนดเกณฑ์การให้คะแนน คือถูกต้อง เท่ากับ 1 คะแนน และไม่ถูกต้อง เท่ากับ 0 คะแนน

โดยเกณฑ์ในการแปลผลคะแนนใช้การพิจารณาแบ่งระดับคะแนนอิงเกณฑ์ โดยประยุกต์จากหลักเกณฑ์ของ Bloom (1971) [11] ซึ่งแบ่งเกณฑ์คะแนนคนที่ตอบถูกออกเป็น 3 ระดับ ดังนี้

ความรู้ระดับสูง หมายถึง ได้คะแนนตั้งแต่ร้อยละ 80 ขึ้นไป หรือ 24-30 คะแนน

ความรู้ระดับปานกลาง หมายถึง ได้คะแนนตั้งแต่ร้อยละ 60-79 หรือ 18-23 คะแนน

ความรู้ระดับน้อย หมายถึง ได้คะแนนต่ำกว่าร้อยละ 60 หรือ 0-17 คะแนน

ตอนที่ 3 พฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 เป็นมาตราส่วนประมาณค่า (Rating Scale) ชนิด 5 ระดับ จำนวน 10 ข้อ รายละเอียดดังนี้

ทุกครั้ง หมายถึง ฉันปฏิบัติตรงกับข้อความนั้น 6-7 วัน/สัปดาห์

บ่อยครั้ง หมายถึง ฉันปฏิบัติตรงกับข้อความนั้น 4-5 วัน/สัปดาห์

บางครั้ง หมายถึง ฉันปฏิบัติตรงกับข้อความนั้น 3 วัน/สัปดาห์

น้อยครั้ง หมายถึง ฉันปฏิบัติตรงกับข้อความนั้น 1-2 วัน/สัปดาห์

ไม่ปฏิบัติเลย หมายถึง ฉันไม่เคยปฏิบัติตรงกับข้อความนั้นเลย

การตรวจสอบความตรงเชิงเนื้อหา โดยนำแบบสอบถามที่สร้างเสร็จแล้วให้ผู้ทรงคุณวุฒิ จำนวน 3 ท่าน ประกอบด้วย นักวิชาการด้านสาธารณสุข นักวิชาการด้านสถิติ และนักวิชาการด้านการวัดผลและประเมินผล พิจารณาและตรวจสอบความตรงเชิงเนื้อหาเป็นรายข้อ เพื่อตรวจสอบถูกต้องของเนื้อหาและความเหมาะสมของสำนวนภาษา โดยใช้ดัชนีความสอดคล้อง IOC (Index of Congruence) ของ Rovinelli and Hambleton (1977) [12] มีสูตรในการคำนวณ ดังนี้ $IOC = \Sigma R/N$

เมื่อ IOC แทน ดัชนีความสอดคล้อง

ΣR แทน ผลรวมของคะแนนความคิดเห็นของผู้เชี่ยวชาญ

N แทน จำนวนผู้เชี่ยวชาญ

การตรวจสอบคุณภาพของแบบสอบถาม โดยใช้ค่าความเชื่อมั่น (Reliability) ด้วยค่าสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's Alpha Method) Cronbach (1990) [13] มีสูตรในการคำนวณ ดังนี้

$$\alpha = \left[\frac{k}{k-1}\right]\left[1-\frac{\sum S_i^2}{S_t^2}\right]$$

เมื่อ α แทน สัมประสิทธิ์ความเชื่อมั่นของเครื่องมือ

k แทน จำนวนข้อของเครื่องมือ

S_i^2 แทน ความแปรปรวนของคะแนนคำถามแต่ละข้อ

S_t^2 แทน ความแปรปรวนของคะแนนรวมของผู้ตอบทั้งหมด

ทั้งนี้ แบบสอบที่ใช้มีค่าความเชื่อมั่นเป็น 0.70 โดยที่ด้านความรู้ความเข้าใจเกี่ยวกับโรคและการป้องกันโรคโควิด-19 มีค่าความเชื่อมั่นมากกว่า 0.71 และด้านพฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 มีค่าความเชื่อมั่นมากกว่า 0.89 หลังจากนั้นนำแบบสอบถามที่ได้ไปดำเนินการขอจริยธรรมการวิจัยในมนุษย์จากคณะกรรมการจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยราชภัฏสุราษฎร์ธานี เอกสารรับรองเลขที่ SRU-EC2023/081 เมื่อได้รับการรับรองจริยธรรมการวิจัยในมนุษย์แล้ว จึงนำแบบสอบถามไปเก็บรวบรวมข้อมูลกับกลุ่มตัวอย่าง

3.3 การวิเคราะห์ข้อมูล

ผู้วิจัยใช้สูตรในการวิเคราะห์ข้อมูลของชูศรี วงษ์รัตนะ (2562) [14] โดยมีสูตรในการคำนวณ ดังนี้

สูตร ร้อยละ $P = \frac{f}{n} \times 100$

เมื่อ P แทน ค่าร้อยละ

f แทน ความถี่ที่ต้องการแปลงเป็นค่าร้อยละ

n แทน จำนวนข้อมูลทั้งหมด

สูตร ค่าเฉลี่ย $\overline{x} = \frac{\sum x}{n}$

เมื่อ $\overline{x}$ แทน ค่าเฉลี่ย

$\sum x$ แทน ผลรวมของข้อมูลทั้งหมด

n แทน จำนวนข้อมูลทั้งหมด

สูตร ส่วนเบี่ยงเบนมาตรฐาน $S = \frac{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}}{n-1}$

เมื่อ S แทน ส่วนเบี่ยงเบนมาตรฐาน

x_i แทน ข้อมูลแต่ละตัว

$\overline{x}$ แทน ค่าเฉลี่ย

n แทน จำนวนข้อมูลทั้งหมด

$\sum$ แทน ผลรวม

สูตร ไควสแควร์ $\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$

เมื่อ χ^2 แทน ค่าไควสแควร์ มีองศาอิสระ (r-1)(c-1)

O_{ij} แทน ความถี่ที่ได้จากการสังเกตใน cell (i,j)

E_{ij} แทน ความถี่คาดหวังใน cell (i,j)

โดยมีรายละเอียดการวิเคราะห์ข้อมูล ดังนี้

1. การวิเคราะห์ข้อมูลทั่วไปของผู้ตอบแบบสอบถาม โดยการนับความถี่และค่าร้อยละ

2. การวิเคราะห์ข้อมูลความรู้ความเข้าใจเกี่ยวกับโรคและการป้องกันโรค พฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 ของประชาชนในจังหวัดสุราษฎร์ธานี โดยการนับความถี่และค่าร้อยละ

3. การวิเคราะห์ความสัมพันธ์ระหว่างข้อมูลทั่วไปกับความรู้ความเข้าใจเกี่ยวกับโรคและการป้องกันโรค พฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 ด้วยการทดสอบไคสแควร์ (Chi-Square test)

4 ผลการวิจัย

1. ข้อมูลทั่วไปของกลุ่มตัวอย่าง จำนวน 453 คน พบว่าผู้ตอบแบบสอบถามส่วนใหญ่เป็นเพศหญิง (ร้อยละ 78.15) ส่วนใหญ่มีอายุในช่วง 50-59 ปี (ร้อยละ 31.79) ในด้านการศึกษาส่วนใหญ่สำเร็จการศึกษาในระดับมัธยมศึกษา (ร้อยละ 41.28) ส่วนใหญ่ประกอบอาชีพเกษตรกรรม/เกษตรกร (ร้อยละ 29.58) และส่วนใหญ่มีประวัติการได้รับโควิด-19 จำนวน 3 เข็ม มากที่สุด (ร้อยละ 47.46) ดังตารางที่ 1

ตารางที่ 1 จำนวนและร้อยละของผู้ตอบแบบสอบถาม

ข้อมูลทั่วไป	จำนวน	ร้อยละ
1. เพศ		
ชาย	99	21.85
หญิง	354	78.15
รวม	**453**	**100**
2. อายุ		
ต่ำกว่า 20 ปี	21	4.64
20-29 ปี	29	6.40
30-39 ปี	39	8.61
40-49 ปี	103	22.74
50-59 ปี	144	31.79
60 ปีขึ้นไป	117	25.83
รวม	**453**	**100**
3. ระดับการศึกษา		
ประถมศึกษา	151	33.33
มัธยมศึกษา	187	41.28
อนุปริญญาหรือเทียบเท่า	34	7.51
ปริญญาตรี	63	13.91
ปริญญาตรีขึ้นไป	18	3.97
รวม	**453**	**100**

ข้อมูลทั่วไป	จำนวน	ร้อยละ
4. อาชีพ		
รับจ้าง	124	27.37
เกษตรกรรม/เกษตรกร	134	29.58
ค้าขาย/ธุรกิจส่วนตัว	119	26.27
พนักงานของรัฐ/รัฐวิสาหกิจ	26	5.74
ข้าราชการ	8	1.77
นักเรียนนักศึกษา	21	4.64
แม่บ้าน	21	4.64
รวม	**453**	**100**
5. ประวัติการได้รับวัคซีนโควิด-19		
ไม่ได้ฉีด	12	2.65
ฉีด 1 เข็ม	21	4.64
ฉีด 2 เข็ม	139	30.68
ฉีด 3 เข็ม	215	47.46
ฉีด 4 เข็ม	56	12.36
มากกว่า 4 เข็ม	10	2.21
รวม	**453**	**100**

2. ผลการวิเคราะห์ความรู้ความเข้าใจเกี่ยวกับโรคและการป้องกันโรคโควิด-19 พบว่า ผู้ตอบแบบสอบถามส่วนใหญ่มีความรู้ความเข้าใจในระดับปานกลาง (ร้อยละ79.25) รองลงมามีความรู้ความเข้าใจระดับมาก (ร้อยละ 11.92) ดังตารางที่ 2 และตารางที่ 3

ตารางที่ 2 จำนวนและร้อยละของผู้ตอบแบบสอบถามที่ตอบถูกในด้านความรู้ความเข้าใจเกี่ยวกับโรคโควิด-19

ความรู้ความเข้าใจเกี่ยวกับโรคโควิด-19	จำนวนผู้ตอบถูก	ร้อยละ
1. โรคโควิด-19 (COVID-19, ย่อจาก Coronavirus disease 2019) เป็นโรคติดเชื้อทางเดินหายใจที่เกิดจากไวรัสโคโรนา มีชื่อทางการว่า SARS-CoV-2	436	96.25
2. โรคโควิด-19 เริ่มต้นระบาดครั้งแรกที่เมืองอู่ฮั่น สาธารณรัฐประชาชนจีน และแพร่ระบาดไปยังประเทศอื่นๆ ทั่วโลก	444	98.01
3. โรคโควิด-19 สามารถแพร่กระจายผ่านการสัมผัสกับผู้ติดเชื้อ ผ่านทางละอองเสมหะจากการไอ จาม น้ำมูก น้ำลายได้	445	98.23
4. โรคโควิด-19 สามารถแพร่เชื้อผ่านสินค้า ที่ผลิตในประเทศที่มีรายงานการระบาด	183	40.40
5. โรคโควิด-19 แพร่กระจายได้แค่ในอากาศแห้งหนาวและไม่แพร่ในอากาศ ร้อนชื้น	228	50.33
6. โรคโควิด-19 สามารถติดต่อกันโดยการสัมผัสเท่านั้น	146	32.23
7. โรคโควิด-19 เป็นโรคที่อันตรายและแพร่ระบาดไปในวงกว้างได้อย่างรวดเร็ว หากได้รับเชื้อ จะสามารถนำเชื้อไปติดกับยังผู้อื่นได้ทันที	395	87.20

ความรู้ความเข้าใจเกี่ยวกับโรคโควิด-19	จำนวนผู้ตอบถูก	ร้อยละ
8. ระยะฟักตัวของโรคโควิด-19 มีระยะเวลา 2-14 วัน	406	89.62
9. ระยะฟักตัวสามารถแพร่เชื้อได้ ช่วง 2-3 วันก่อนที่จะแสดงอาการ แต่ระยะฟักตัวแต่ละคนอาจต่างกันออกไป จึงควรกักตัวทันทีหลังเสี่ยงรับเชื้อ	30	6.62
10. ผู้ที่ได้รับเชื้อโควิด-19 จะไม่แสดงอาการจนกว่าจะพ้นระยะฟักตัวของเชื้อ	317	69.98
11. อาการของผู้ป่วยโควิด-19 อาการทั่วไปคล้ายไข้หวัดใหญ่ เช่น มีไข้สูง ไอ หายใจลำบาก	420	92.72
12. ผู้ป่วยโควิด-19 เสียชีวิตเนื่องจากระบบหายใจล้มเหลวเพราะปอดถูกทำลาย	419	92.49
13. ผู้ป่วยโควิด-19 จำนวนมากไม่มีอาการใดๆ และสามารถฟื้นตัวจนหายเองได้	239	52.76
14. หลังหายป่วยจากการติดเชื้อโควิด-19 แล้ว ผู้ป่วยจากโควิด-19จะมีภูมิคุ้มกันที่เกิดขึ้นตามธรรมชาติอยู่ และภูมิคุ้มกันนี้ จะป้องกันไม่ให้ติดเชื้อซ้ำตลอดชีวิต	286	63.13
15. การมีระบบภูมิคุ้มกันที่ดี สามารถช่วยรักษาโรคโควิด-19 (COVID-19) ได้โดยไม่จำเป็นต้องทานยา	198	43.71
16. Antigen Test Kit (ATK) หรือชุดตรวจโควิด-19 แบบเร่งด่วน ด้วยการ Swab เก็บตัวอย่างสารคัดหลั่งทางจมูก หรือเก็บจากคอ จำเป็นต้องผ่านการตรวจจากโรงพยาบาลเท่านั้น	243	53.64
17. ควรตรวจหาเชื้อโควิด-19 ด้วย ATK ด่วน เมื่อมีอาการมีไข้ ไอ ลิ้นไม่รับรส ปวดเมื่อยตามร่างกาย ปวดศีรษะ หายใจหอบ หายใจลำบาก หรือมีประวัติเดินทางหรือไปในสถานที่เสี่ยง พักอาศัย หรืออยู่ร่วมบ้านกับผู้ติดเชื้อโควิด-19	431	95.14
18. ปัจจุบันยังไม่มียาที่มีผลในการต้านไวรัสโรคโควิด-19 โดยเฉพาะ	126	27.81
19. การฉีดวัคซีนป้องกันโรคโควิด-19 สามารถลดการแพร่ระบาด ลดความรุนแรงของอาการป่วย และลดการเสียชีวิตได้	407	89.85
20. ควรฉีดวัคซีนโควิด-19 จำนวน 2 เข็มเป็นพื้นฐาน หลังจากนี้สามารถฉีดได้ทุก 4 เดือน เพื่อลดโอกาสติดเชื้อและความรุนแรงของโรค	309	68.21

ความรู้ความเข้าใจเกี่ยวกับการป้องกันโรคโควิด-19	จำนวนผู้ตอบถูก	ร้อยละ
21. การใส่หน้ากากอนามัยที่ถูกต้อง ป้องกันความเสี่ยงจากการติดโรคโควิด-19 ได้	424	93.60
22. การใส่หน้ากากอนามัยที่ไม่ถูกวิธี เป็นสาเหตุให้มีโอกาสติดเชื้อโรคโควิด-19 มากขึ้น	103	22.74
23. หลังจากได้รับการฉีดวัคซีนป้องกันโรคโควิด-19 แล้วไม่จำเป็นต้องสวมหน้ากากอนามัย	357	78.81
24. ควรเว้นระยะห่างจากผู้ป่วยติดเชื้อหรือผู้ที่มีอาการอย่างน้อย 1 เมตร	360	79.47
25. การรักษาระยะห่างจากผู้อื่นในพื้นที่แออัดสามารถลดการแพร่เชื้อได้จึงไม่จำเป็นต้องสวมหน้ากากอนามัย	294	64.90
26. ล้างมือด้วยสบู่และน้ำหรือใช้แอลกอฮอล์สำหรับล้างมือบ่อย ๆ โดยเฉพาะหลังสัมผัสกับผู้ป่วยหรือข้าวของเครื่องใช้ของผู้ป่วย	421	92.94
27. ล้างมือให้สม่ำเสมอด้วยสบู่ หรือแอลกอฮอล์เจลอย่างน้อย 20 วินาที ความเข้มข้นของแอลกอฮอล์ไม่ต่ำกว่า 50%	130	28.70

ความรู้ความเข้าใจเกี่ยวกับการป้องกันโรคโควิด-19	จำนวนผู้ตอบถูก	ร้อยละ
28. การใช้ช้อนกลางประจำตัวเมื่อต้องรับประทานอาหารร่วมกับผู้อื่น เพื่อลดความเสี่ยงในการติดเชื้อไวรัสจากการทานอาหารกับผู้อื่น	431	95.14
29. การรับประทานอาหารที่ปรุงสุกใหม่ๆ ด้วยความร้อน ช่วยลดความเสี่ยงโรคโควิด-19	424	93.60
30. เลือกทานอาหารที่มีประโยชน์ และคุณค่าทางอาหารอยู่เสมอเพื่อช่วยเรื่องระบบภูมิคุ้มกันให้ทำงานได้อย่างมีประสิทธิภาพ	430	94.92

ตารางที่ 3 ระดับความรู้ความเข้าใจเกี่ยวกับโรคและการป้องกันโรคโควิด-1 9 ของผู้ตอบแบบสอบถาม

ระดับความรู้	จำนวน	ร้อยละ
ระดับสูง	54	11.92
ระดับปานกลาง	359	79.25
ระดับน้อย	40	8.83

3. ผลการศึกษาพฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 ของผู้ตอบแบบสอบถาม พบว่า มีพฤติกรรมการปฏิบัติทุกครั้งมากที่สุด (ร้อยละ 67.33) เมื่อพิจารณาเป็นรายข้อ พบว่า มีพฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 ดีที่สุดคือ การสวมหน้ากากอนามัยทุกครั้งเมื่อไม่สบายหรือออกจากบ้าน รองลงมาคือ การสังเกตลักษณะอาการที่เกี่ยวข้องกับการติดเชื้อโรคโควิด-19 ของตนเองเป็นประจำ และพฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 น้อยกว่าทุกข้อคือ การหลีกเลี่ยงการใช้มือสัมผัสใบหน้า ตา จมูก โดยไม่จำเป็น ดังตารางที่ 4 และตารางที่ 5

ตารางที่ 4 จำนวนและร้อยละของกลุ่มตัวอย่างจำแนกตามพฤติกรรมการป้องกันโรคโควิด-19

ข้อปฏิบัติ	ความถี่ในการปฏิบัติ				
	ทุกครั้ง	บ่อยครั้ง	บางครั้ง	นานๆ ครั้ง	ไม่ปฏิบัติเลย
1. สวมหน้ากากอนามัยทุกครั้งเมื่อไม่สบายหรือออกจากบ้าน	307 (67.77)	97 (21.41)	47 (10.38)	2 (0.44)	0 (0.00)
2. เมื่อไอ หรือจามจะใช้กระดาษทิชชู หรือผ้าปิดปากหรือจมูก	273 (60.26)	145 (32.01)	33 (7.28)	2 (0.44)	0 (0.0)
3. หลีกเลี่ยงการใช้มือสัมผัสใบหน้า ตา จมูก โดยไม่จำเป็น	203 (44.81)	159 (35.10)	80 (17.66)	10 (2.21)	1 (0.22)
4. หลีกเลี่ยงที่จะเข้าไปในที่ที่มีผู้คนหนาแน่นแออัดหรือพื้นที่ปิด	228 (50.33)	141 (31.13)	77 (17.00)	7 (1.55)	0 (0.0)
5. เว้นระยะห่างจากผู้ป่วยติดเชื้อหรือผู้ที่มีอาการอย่างน้อย 1 เมตร	275 (60.71)	102 (22.52)	71 (15.67)	4 (0.88)	1 (0.22)
6. ล้างมือด้วยสบู่ หรือแอลกอฮอล์ หลังจากสัมผัสสิ่งของ หรือกลับจากข้างนอก	287 (63.36)	107 (23.62)	56 (12.36)	3 (0.66)	0 (0.0)

ข้อปฏิบัติ	ความถี่ในการปฏิบัติ				
	ทุกครั้ง	บ่อยครั้ง	บางครั้ง	นานๆ ครั้ง	ไม่ปฏิบัติเลย
7. แยกของใช้ส่วนตัว และหลีกเลี่ยงการใช้ของร่วมกับผู้อื่น	282 (62.25)	130 (28.70)	39 (8.61)	2 (0.44)	(0.00) 0
8. เปลี่ยนเสื้อผ้าหรือการอาบน้ำทันที เมื่อกลับมาจากข้างนอก	259 (57.17)	109 (24.06)	81 (17.88)	4 (0.88)	0 (0.00)
9. สังเกตลักษณะอาการที่เกี่ยวข้องกับการติดเชื้อโรคโควิด-19 ของตนเองเป็นประจำ เช่น การไอ เจ็บ คอ มีไข้ และการหายใจหอบเหนื่อย เป็นต้น	298 (65.78)	132 (29.14)	18 (3.97)	1 (0.22)	4 (0.88)
10. ดูแลสุขภาพตัวเองและป้องกันตัวเองอยู่สม่ำเสมอ เช่น ออกกำลังกายเป็นประจำ พกหน้ากากอนามัยและเจลแอลกอฮอล์ติดตัวหลังออกจากบ้าน	238 (52.54)	138 (30.46)	68 (15.01)	4 (0.88)	5 (1.10)

ตารางที่ 5 จำนวนและร้อยละของกลุ่มตัวอย่างจำแนกตามระดับพฤติกรรมการป้องกันการระบาดของโรคโควิด-19

ระดับพฤติกรรมการป้องกัน	จำนวน	ร้อยละ
ปฏิบัติทุกครั้ง	305	67.33
ปฏิบัติบ่อยครั้ง	127	28.04
ปฏิบัติบางครั้ง	21	4.63
ปฏิบัตินานๆ ครั้ง	0	0.00
ไม่ปฏิบัติเลย	0	0.00

4. การศึกษาความสัมพันธ์ระหว่างข้อมูลทั่วไปกับระดับความรู้ความเข้าใจเกี่ยวกับโรคและการป้องกันโรคโควิด-19 ด้วยการทดสอบไคสแควร์ กำหนดสมมติฐานในการทดสอบ ดังนี้

H_0 : ข้อมูลทั่วไปของผู้ตอบแบบสอบถามไม่ส่งผลต่อระดับความรู้ความเข้าใจเกี่ยวกับโรคและการป้องกันโรคโควิด-19

H_1 : ข้อมูลทั่วไปของผู้ตอบแบบสอบถามส่งผลต่อระดับความรู้ความเข้าใจเกี่ยวกับโรคและการป้องกันโรคโควิด-19

ผลการทดสอบ พบว่า ข้อมูลเพศกับประวัติการได้รับวัคซีนโควิด-19 ที่แตกต่างกันจะส่งผลต่อระดับความรู้ความเข้าใจเกี่ยวกับการป้องกันโรคโควิด-19 ส่วนข้อมูลอายุ ระดับการศึกษา และอาชีพ ไม่ส่งผลต่อระดับความรู้ความเข้าใจเกี่ยวกับโรคโควิด-19 อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ดังตารางที่ 6

ตารางที่ 6 ผลการวิเคราะห์ความสัมพันธ์ระหว่างข้อมูลทั่วไปกับระดับความรู้ความเข้าใจเกี่ยวกับโรคและการป้องกันโรคโควิด-19 ด้วยสถิติทดสอบไคสแควร์

ข้อมูลทั่วไป	ระดับความรู้ความเข้าใจ			χ^2 (df)	p-value
	สูง	ปานกลาง	ต่ำ		
เพศ					
ชาย	66 (14.57)	32 (7.06)	1 (0.22)	19.87 (7)	0.006*
หญิง	278 (61.37)	73 (16.11)	3 (0.66)		

ข้อมูลทั่วไป	ระดับความรู้ความเข้าใจ			χ^2 (df)	p-value
	สูง	ปานกลาง	ต่ำ		
อายุ					
ต่ำกว่า 20 ปี	15 (3.31)	6 (1.32)	0 (0.00)	32.42	0.594
20-29 ปี	20 (4.42)	9 (1.99)	0 (0.00)	(35)	
30-39 ปี	32 (7.06)	6 (1.32)	1 (0.22)		
40-49 ปี	83 (18.32)	19 (4.19)	1 (0.22)		
50-59 ปี	108 (23.84)	36 (7.95)	0 (0.00)		
60 ปีขึ้นไป	86 (18.98)	29 (6.40)	2 (0.44)		
ระดับการศึกษา					
ประถมศึกษา	112 (24.72)	39 (8.61)	0 (0.00)	28.09	0.460
มัธยมศึกษา	145 (32.01)	40 (8.83)	2 (0.44)	(28)	
อนุปริญญาหรือเทียบเท่า	28 (6.18)	6 (1.32)	0 (0.00)		
ปริญญาตรี	45 (9.93)	17 (3.75)	1 (0.22)		
ปริญญาตรีขึ้นไป	14 (3.09)	3 (0.66)	1 (0.22)		
อาชีพ					
รับจ้าง	95 (20.97)	28 (6.18)	1 (0.22)	34.72	0.780
เกษตรกรรม/เกษตรกร	104 (22.96)	28 (6.18)	2 (0.44)	(42)	
ค้าขาย/ธุรกิจส่วนตัว	86 (18.98)	32 (7.06)	1 (0.22)		
พนักงานของรัฐ/รัฐวิสาหกิจ	20 (4.42)	6 (1.32)	0 (0.00)		
ข้าราชการ	5 (1.10)	3 (0.66)	0 (0.00)		
นักเรียน/นักศึกษา	16 (3.53)	5 (1.10)	0 (0.00)		
แม่บ้าน	18 (3.97)	3 (0.66)	0 (0.00)		
ประวัติการได้รับวัดซีนโควิด-19					
ไม่ได้ฉีด	10 (2.21)	2 (0.44)	0 (0.00)	66.11	0.001*
ฉีด 1 เข็ม	10 (2.21)	10 (2.21)	1 (0.22)	(35)	
ฉีด 2 เข็ม	78 (17.22)	44 (9.71)	1 (0.22)		
ฉีด 3 เข็ม	176 (38.85)	38 (8.39)	1 (0.22)		
ฉีด 4 เข็ม	46 (10.15)	9 (1.99)	1 (0.22)		
มากกว่า 4 เข็ม	8 (1.77)	2 (0.44)	0 (0.00)		

** ระดับนัยสำคัญทางสถิติที่ระดับ 0.05*

5. การศึกษาความสัมพันธ์ระหว่างข้อมูลทั่วไปกับพฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 ด้วยการทดสอบไคสแควร์ กำหนดสมมติฐานในการทดสอบ ดังนี้

H_0 : ข้อมูลทั่วไปของผู้ตอบแบบสอบถามไม่ส่งผลต่อระดับพฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19

H_1 : ข้อมูลทั่วไปของผู้ตอบแบบสอบถามส่งผลต่อระดับพฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19

ผลการทดสอบพบว่า ข้อมูลเพศ อายุ ระดับการศึกษา และอาชีพที่แตกต่างกันจะส่งผลต่อพฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 ส่วนข้อมูลประวัติการได้รับวัคซีนโควิด-19 ไม่ส่งผลต่อพฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ดังตารางที่ 7

ตารางที่ 7 ผลการวิเคราะห์ความสัมพันธ์ระหว่างข้อมูลทั่วไปกับระดับพฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 ด้วยสถิติทดสอบไคสแควร์

ข้อมูลทั่วไป	ระดับพฤติกรรมการป้องกัน					χ^2 (df)	p-value
	ทุกครั้ง	บ่อยครั้ง	บางครั้ง	นานๆ ครั้ง	ไม่ปฏิบัติเลย		
เพศ							
ชาย	54 (11.92)	40 (8.83)	5 (1.10)	0 (0.00)	0 (0.00)	34.95	0.020*
หญิง	251 (55.41)	87 (19.20)	16 (3.53)	0 (0.00)	0 (0.00)	(20)	
อายุ							
ต่ำกว่า 20 ปี	16 (3.53)	4 (0.66)	1 (0.22)	0 (0.00)	0 (0.00)	145.13	0.002*
20-29 ปี	18 (3.97)	10 (2.21)	1 (0.22)	0 (0.00)	0 (0.00)	(100)	
30-39 ปี	18 (3.97)	20 (4.41)	1 (0.22)	0 (0.00)	0 (0.00)		
40-49 ปี	71 (15.67)	26 (5.74)	6 (1.32)	0 (0.00)	0 (0.00)		
50-59 ปี	102 (22.52)	36 (7.95)	6 (1.32)	0 (0.00)	0 (0.00)		
60 ปีขึ้นไป	80 (17.66)	31 (6.84)	6 (1.32)	0 (0.00)	0 (0.00)		
ระดับการศึกษา							
ประถมศึกษา	103 (22.74)	41 (9.05)	7 (1.54)	0 (0.00)	0 (0.00)	104.19	0.036*
มัธยมศึกษา	124 (27.37)	55 (12.14)	8 (1.77)	0 (0.00)	0 (0.00)	(80)	
อนุปริญญา	20 (4.41)	10 (2.21)	4 (0.88)	0 (0.00)	0 (0.00)		
ปริญญาตรี	43 (9.49)	18 (3.97)	2 (0.44)	0 (0.00)	0 (0.00)		
ปริญญาตรี ขึ้นไป	15 (3.31)	3 (0.66)	0 (0.00)	0 (0.00)	0 (0.00)		
อาชีพ							
รับจ้าง	88 (19.43)	32 (7.06)	4 (0.88)	0 (0.00)	0 (0.00)	212.16	0.000*
เกษตรกร	86 (18.98)	43 (9.49)	5 (1.10)	0 (0.00)	0 (0.00)	(120)	
ค้าขาย/ธุรกิจส่วนตัว	77 (17.00)	34 (7.50)	8 (1.77)	0 (0.00)	0 (0.00)		
พนักงานของรัฐ	17 (3.75)	8 (1.77)	1 (0.22)	0 (0.00)	0 (0.00)		
ข้าราชการ	6 (1.32)	2 (0.44)	0 (0.00)	0 (0.00)	0 (0.00)		
นักเรียน/นักศึกษา	18 (3.97)	2 (0.44)	1 (0.22)	0 (0.00)	0 (0.00)		
แม่บ้าน	13 (2.87)	6 (1.32)	2 (0.44)	0 (0.00)	0 (0.00)		

ข้อมูลทั่วไป	ระดับพฤติกรรมการป้องกัน					χ^2 (df)	p-value
	ทุกครั้ง	บ่อยครั้ง	บางครั้ง	นานๆ ครั้ง	ไม่ปฏิบัติเลย		
ประวัติการได้รับวัคซีนโควิด-19							
ไม่ได้ฉีด	9 (1.99)	3 (0.66)	0 (0.00)	0 (0.00)	0 (0.00)	123.71	0.054
ฉีด 1 เข็ม	11 (2.42)	9 (1.99)	1 (0.22)	0 (0.00)	0 (0.00)	(100)	
ฉีด 2 เข็ม	97 (21.41)	39 (8.61)	3 (0.66)	0 (0.00)	0 (0.00)		
ฉีด 3 เข็ม	141 (31.12)	61 (13.46)	13 (2.87)	0 (0.00)	0 (0.00)		
ฉีด 4 เข็ม	42 (9.27)	10 (2.21)	4 (0.88)	0 (0.00)	0 (0.00)		
มากกว่า 4 เข็ม	5 (0.10)	5 (1.10)	0 (0.00)	0 (0.00)	0 (0.00)		

** ระดับนัยสำคัญทางสถิติที่ระดับ 0.05*

5 สรุป อภิปรายผลและข้อเสนอแนะ

5.1 สรุปและอภิปรายผล

1. ผลการศึกษาความรู้ความเข้าใจเกี่ยวกับโรคและการป้องกันโรคโควิด-19 พบว่า ประชาชนในจังหวัดสุราษฎร์ธานี มีความรู้ความเข้าใจเกี่ยวกับโรคและการป้องกันโรคโควิด-19 ในระดับปานกลาง (ร้อยละ 79.25) อภิปรายผลได้ว่า ช่วงเวลาของการศึกษาครั้งนี้เป็นช่วงหลังการระบาดใหญ่ที่ประชาชนได้รับข้อมูลข่าวสารเกี่ยวกับโรคและการป้องกันโรคไม่เข้มข้น ซึ่งสอดคล้องกับงานวิจัยของบงกช โมระสกุล และคณะ [7] ได้ศึกษาความรู้และพฤติกรรมการป้องกันโรคโควิด-19 ของนักศึกษาพยาบาลชั้นปีที่ 1 วิทยาลัยนานาชาติเซนต์เทเรซา และวิทยาลัยเซนต์หลุยส์ พบว่านักศึกษาพยาบาลชั้นปีที่ 1 มีความรู้ความเข้าใจเกี่ยวกับโรคโควิด-19 ในระดับปานกลาง ซึ่งแตกต่างจากงานวิจัยของจันทิมา ห้าวหาญ และคณะ [6] ได้ศึกษาความรู้ความเข้าใจและพฤติกรรมการปฏิบัติตนเกี่ยวกับการป้องกันโรคโควิด-19 (COVID-19) ของประชาชนในจังหวัดภูเก็ต ผลการศึกษาพบว่า ระดับความรู้ความเข้าใจเกี่ยวกับการป้องกันโรคโควิด-19 (COVID-19) ของประชาชนในจังหวัดภูเก็ต อยู่ในระดับมาก ซึ่งเป็นผลการศึกษาในช่วงการระบาดหนักของโรคโควิด-19

2. ผลการศึกษาพฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 ของประชาชนในจังหวัดสุราษฎร์ธานี พบว่า มีพฤติกรรมการปฏิบัติทุกครั้งมากที่สุด (ร้อยละ 67.33) เมื่อพิจารณาเป็นรายข้อ พบว่า ประชาชนมีพฤติกรรมการป้องกันการแพร่ระบาดที่ดีที่สุดคือ การสวมหน้ากากอนามัยทุกครั้งเมื่อไม่สบายหรือออกจากบ้าน รองลงมาคือ การสังเกตลักษณะอาการที่เกี่ยวข้องกับการติดเชื้อโรคโควิด-19 ของตนเองเป็นประจำ และพฤติกรรมการป้องกันการแพร่ระบาดที่น้อยกว่าทุกข้อคือ การหลีกเลี่ยงการใช้มือสัมผัสใบหน้า ตา จมูก โดยไม่จำเป็น ซึ่งสอดคล้องกับงานวิจัยของธานี กล่อมใจ และคณะ [9] ได้ศึกษาความรู้และพฤติกรรมของประชาชนเรื่องการป้องกันตนเองจากการติดเชื้อไวรัสโคโรนาสายพันธุ์ใหม่ 2019 พบว่า พฤติกรรมการป้องกันตนเองจากการติดเชื้อไวรัสโคโรนาสายพันธุ์ใหม่ 2019 ในภาพรวมอยู่ในระดับมาก

3. ผลศึกษาความสัมพันธ์ระหว่างข้อมูลทั่วไปกับความรู้ความเข้าใจเกี่ยวกับโรค การป้องกันโรค และพฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 ของประชาชนในจังหวัดสุราษฎร์ธานี พบว่า เพศและประวัติการได้รับวัคซีนโควิด-19 ส่งผลต่อระดับความรู้ความเข้าใจเกี่ยวกับโรคและการป้องกันโรคโควิด-19 อภิปรายผลได้ว่า

ประชาชนที่ได้รับการฉีดวัคซีนโควิด-19 จะมีความตระหนักรู้เกี่ยวกับโรคและส่วนใหญ่จะได้รับความรู้เกี่ยวกับโรคและการป้องกันโรคพร้อมๆ กับการได้รับวัคซีนไปด้วย ในขณะที่เพศ อายุ ระดับการศึกษาและอาชีพส่งผลต่อพฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 ส่วนข้อมูลประวัติการได้รับวัคซีนโควิด-19 ไม่ส่งผลต่อพฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 อภิปรายผลได้ว่า ประชาชนที่ได้รับวัคซีนตามจำนวนที่หน่วยงานด้านสาธารณสุขให้ข้อมูล ย่อมมีความมั่นใจต่อประสิทธิภาพของวัคซีนโควิด-19 ที่ได้รับ ทำให้พฤติกรรมการป้องกันการแพร่ระบาดของโรคโควิด-19 ไม่แตกต่างกัน

5.2 ข้อเสนอแนะ

1. การวิจัยครั้งต่อไป ควรเพิ่มการศึกษาตัวแปรเกี่ยวกับการแพร่ระบาดของโรคโควิด-19 ในจังหวัดสุราษฎร์ธานี เช่น ความรู้ความเข้าใจและพฤติกรรมการป้องกันโรคของกลุ่มบุคลากรปฏิบัติงานด้านสาธารณสุขในพื้นที่ เช่น อาสาสมัครสาธารณสุขประจำหมู่บ้าน (อสม.) หรือนักท่องเที่ยวต่างชาติที่เดินทางเข้ามาในจังหวัดสุราษฎร์ธานี เป็นต้น

2. หน่วยงานด้านสาธารณสุขในพื้นที่จังหวัดสุราษฎร์ธานี สามารถนำผลการวิจัยไปวางแผน ส่งเสริมและสนับสนุนการจัดกิจกรรมด้านสาธารณสุขเกี่ยวกับการป้องกันและเฝ้าระวังการแพร่ระบาดของโรคโควิด-19 ในช่วงหลังการระบาดใหญ่ในพื้นที่ได้

เอกสารอ้างอิง

[1] กรมควบคุมโรค, สถานการณ์ผู้ติดเชื้อ COVID-19 (2567), [ออนไลน์] เข้าถึงเมื่อวันที่ 16 กุมภาพันธ์ 2567 เข้าถึงจาก https:// ddc.moph.go.th/viralpneumonia/.

[2] WHO, Coronavirus disease (COVID-19) pandemic. [online]. Available : https://www.who.int/emergencies/diseases/novel-coronavirus-2019, (July 16, 2023).

[3] ศูนย์การแพทย์สมเด็จพระเทพรัตนราชสุดาฯ สยามบรมราชกุมารี, เช็กอาการ โควิดสายพันธุ์ใหม่ XBB 1.16 ร้ายแรงแค่ไหน หวั่น ติดง่ายกว่าเดิม 2 เท่า, [ออนไลน์] เข้าถึงเมื่อวันที่ 10 กุมภาพันธ์ 2567 เข้าถึงจาก http://medicine.swu.ac.th/msmc/?p=7686.

[4] Worldometers. (2023). COVID-19 CORONAVIRUS PANDEMIC [online]. Available: https://www.worldometers.info/coronavirus/, (2024, April 1)

[5] Marra AR, Kobayashi T, Callado GY, et al. The effectiveness of COVID-19 vaccine in the prevention of post-COVID conditions: a systematic literature review and meta-analysis of the latest research. Antimicrobial Stewardship & Healthcare Epidemiology. 2023;3(1):e168. doi:10.1017/ash.2023.447

[6] จันทิมา ห้าวหาญ และพรรณวดี ขำจริง, *ความรู้ความเข้าใจและพฤติกรรมการปฏิบัติตนเกี่ยวกับการป้องกันโรคโควิด-19 (COVID-19) ของประชาชนในจังหวัดภูเก็ต*, รายงานสืบเนื่องจากการประชุมวิชาการระดับชาติ ครั้งที่ 11 หัวข้อ “Community-led Social Innovation in the Era of Global Changes amidst Covid-19 Crisis: นวัตกรรมทางสังคมของชุมชนในยุคของการเปลี่ยนแปลงโลกท่ามกลางวิกฤตโควิด-19”, 2564, นครศรีธรรมราช, 19 กุมภาพันธ์ 2564, pp. 169-178.

[7] บงกช โมระสกุล และพรศิริ พันธสี, *ความรู้และพฤติกรรมการป้องกันโรคโควิด-19 ของนักศึกษาพยาบาลชั้นปีที่ 1 วิทยาลัยนานาชาติเซนต์เทเรซา และวิทยาลัยเซนต์หลุยส์*, วารสารศูนย์อนามัยที่ 9, 15(37) (2564), 179-195.

[8] กัมปนาท โคตรพันธ์ และนิยม จันทร์นวล, *ความสัมพันธ์ระหว่างความรอบรู้ด้านสุขภาพกับพฤติกรรมการป้องกันโรคติดเชื้อไวรัสโคโรนา 2019 ของประชาชนในจังหวัดมุกดาหาร*, รายงานสืบเนื่องจากการประชุมวิชาการระดับชาติ มอบ. วิจัย ครั้งที่ 16, 2565, อุบลราชธานี, 11-12 กรกฎาคม 2565, pp. 148-160.

[9] ธานี กล่อมใจ จรรยา แก้วใจบุญ และทักษิกา ชัชวรัตน์, *ความรู้และพฤติกรรมของประชาชนเรื่องการป้องกันตนเองจากการติดเชื้อไวรัสโคโรนา สายพันธุ์ใหม่ 2019*, วารสารการพยาบาล การสาธารณสุขและการศึกษา, 21(2), (2563), pp. 29-39.

[10] T. Yamane, *Statistics An Introductory Analysis*, 2nd Ed., Harper and Row, New York, 1967.

[11] Bloom, B.S., *Handbook on formative and summative evaluation of student learning*, New York, 1971.

[12] R.J. Rovinelli and R.K. Hambleton, *On the use of content specialists in the assessment of criterion-referenced test item validity*, Tijdschrift voor Onderwijsresearch, 2(2) (1977), 49-60

[13] L.J. Cronbach, *Essentials of psychological testing*, 5th ed. New York, 1990

[14] ชูศรี วงศ์รัตนะ, *เทคนิคการใช้สถิติเพื่อการวิจัย*, พิมพ์ครั้งที่ 14, กรุงเทพมหานคร, 2562

บรรณาธิการ

ฝ่ายวิชาการ

1	รศ. ดร.ฐิตารีย์ วุฒิจิรัฐิติกาล	มหาวิทยาลัยอุบลราชธานี
2	รศ. ดร.รตนกร วัฒนทวีกุล	มหาวิทยาลัยอุบลราชธานี
3	รศ. ดร.ศราวุธ แสนการุณ	มหาวิทยาลัยอุบลราชธานี
4	ผศ. ดร.กนกพร ช่างทอง	มหาวิทยาลัยอุบลราชธานี
5	ผศ. ดร.คณิศา โชติจันทึก	มหาวิทยาลัยอุบลราชธานี
6	ผศ. ดร.พัชรี วงษาสนธิ์	มหาวิทยาลัยอุบลราชธานี
7	ผศ. ดร.ไพรินทร์ สุวรรณศรี	มหาวิทยาลัยอุบลราชธานี
8	ผศ. ดร.วีรยุทธ นิลสระคู	มหาวิทยาลัยอุบลราชธานี
9	ผศ. ดร.สุพจน์ สีบุตร	มหาวิทยาลัยอุบลราชธานี
10	ผศ.รตี โบจรัส	มหาวิทยาลัยอุบลราชธานี
11	อ. ดร.กฤษดา นารอง	มหาวิทยาลัยอุบลราชธานี
12	อ. ดร.จิรัชยา ใจสะอาดซื่อตรง	มหาวิทยาลัยอุบลราชธานี
13	อ. ดร.ธนวิทย์ จีรุพันธ์	มหาวิทยาลัยอุบลราชธานี
14	อ. ดร.นงคราญ สระโสม	มหาวิทยาลัยอุบลราชธานี
15	อ. ดร.ศักดิ์ดา น้อยนาง	มหาวิทยาลัยอุบลราชธานี
16	อ.ธนาตย์ เดโชชัยพร	มหาวิทยาลัยอุบลราชธานี
17	รศ. ดร.ตวงรัตน์ ไชยชนะ	จุฬาลงกรณ์มหาวิทยาลัย
18	รศ. ดร.รตินันท์ บุญเคลือบ	จุฬาลงกรณ์มหาวิทยาลัย
19	ผศ. ดร.ธีรพงษ์ พงษ์พัฒนเจริญ	จุฬาลงกรณ์มหาวิทยาลัย
20	ผศ. ดร.ภัททิรา เรืองสินทรัพย์	มหาวิทยาลัยเกษตรศาสตร์
21	รศ. ดร.จิระศักดิ์ มงคลเคหา	มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน
22	ผศ. ดร.วัชรินทร รักษาศักดิ์ชัย	มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน
23	รศ. ดร.กิตติกร นาคประสิทธิ์	มหาวิทยาลัยขอนแก่น
24	รศ. ดร.เกียรติสุดา นาคประสิทธิ์	มหาวิทยาลัยขอนแก่น
25	รศ. ดร.นรากร คณาศรี	มหาวิทยาลัยขอนแก่น
26	รศ. ดร.บัณฑิต ภิบาลจอมมี	มหาวิทยาลัยขอนแก่น
27	รศ. ดร.พิกุล ภูผาสุข	มหาวิทยาลัยขอนแก่น

28	รศ. ดร.สมนึก วรวิเศษ	มหาวิทยาลัยขอนแก่น
29	ผศ. ดร.ณัฐวุฒิ นุโพธิ์	มหาวิทยาลัยขอนแก่น
30	ผศ. ดร.ทศพร แถลงธรรม	มหาวิทยาลัยขอนแก่น
31	ผศ. ดร.นวรัตน์ เอกก้านตรง	มหาวิทยาลัยขอนแก่น
32	ผศ. ดร.พงศกร ยศแก้ว	มหาวิทยาลัยขอนแก่น
33	ผศ. ดร.สัมพันธ์ ถิ่นเวียงทอง	มหาวิทยาลัยขอนแก่น
34	รศ. ดร.ธเนศร์ โรจน์ศิรพิศาล	มหาวิทยาลัยเชียงใหม่
35	ผศ. ดร.ธีรนุช บุนนาค	มหาวิทยาลัยเชียงใหม่
36	ผศ. ดร.นัยน์รัตน์ กันยะมี	มหาวิทยาลัยเชียงใหม่
37	ผศ. ดร.ภาคภูมิ เพ็ชรประดับ	มหาวิทยาลัยเชียงใหม่
38	ผศ. ดร.กรวิกา ก้องกุล	มหาวิทยาลัยทักษิณ
39	ผศ. ดร.ณภัทณ์จันทร์ ด่านสวัสดิ์	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
40	ผศ. ดร.วริสา ยมเสถียรกุล	มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี
41	อ. ดร.ทศพร คำดวง	มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์
42	อ. ดร.ขวัญชีวา วัฒนตรีภพ	มหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา
43	ผศ.ดิษฐพล มั่นธรรม	มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ ศูนย์-พระนครศรีอยุธยา หันตรา
44	ผศ. ดร.นฤปนาถ เหล็กโคกสูง	มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน วิทยาเขต-ขอนแก่น
45	ผศ. ดร.เบญจวรรณ โรจนดิษฐ์	มหาวิทยาลัยเทคโนโลยีสุรนารี
46	รศ.ศิริจันทร์ เวสารัชศาต	มหาวิทยาลัยธรรมศาสตร์
47	ผศ. ดร.ขจี จันทรขจร	มหาวิทยาลัยธรรมศาสตร์
48	ผศ. ดร.นันทพัทธ์ ตระกูลไตรพฤกษ์	มหาวิทยาลัยธรรมศาสตร์
49	ผศ. ดร.เบญจวรรณ สุขเจริญภิญโญ	มหาวิทยาลัยนเรศวร
50	ผศ. ดร.วนินทร พูนไพบูลย์พิพัฒน์	มหาวิทยาลัยนเรศวร
51	ผศ. ดร.สุภาวรรณ จันทร์ไพแสง	มหาวิทยาลัยนเรศวร
52	ผศ. ดร.สุรีย์พร ชาวแพรกน้อย	มหาวิทยาลัยนเรศวร
53	ผศ. ดร.อุมารินทร์ ปิ่นตบแต่ง	มหาวิทยาลัยนเรศวร
54	ผศ. ดร.สินีนาฏ ศรีมงคล	มหาวิทยาลัยบูรพา
55	รศ. ดร.กรรณิการ์ ขำพึ่งสน	มหาวิทยาลัยพะเยา

56	อ. ดร.นพดล ยศบุญเรือง	มหาวิทยาลัยพะเยา
57	ผศ. ดร.มนตรี ทองมูล	มหาวิทยาลัยมหาสารคาม
58	ผศ. ดร.ชนันท์ ลิ่วเฉลิมวงศ์	มหาวิทยาลัยมหิดล
59	ผศ. ดร.วิทวัชร์ โฆษิตวัฒนฤกษ์	มหาวิทยาลัยมหิดล
60	ผศ. ดร.วรารัตน์ วงศ์เกี่ย	มหาวิทยาลัยมหิดล
61	อ. ดร.ปิยนันท์ ผาโสม	มหาวิทยาลัยมหิดล
62	ผศ. ดร.มัลลิกา ราชกิจ	มหาวิทยาลัยแม่โจ้
63	ผศ. ดร.สิตา ชากฤษณ์	มหาวิทยาลัยแม่โจ้
64	รศ. ดร.วัลลภ เหมวงษ์	มหาวิทยาลัยราชภัฏอุดรธานี
65	ผศ. ดร.สุพรรณี สมพงษ์	มหาวิทยาลัยราชภัฏสกลนคร
66	อ. ดร.อัจฉริยา นิลสระคู	มหาวิทยาลัยราชภัฏอุบลราชธานี
67	รศ. ดร.กิตติพงษ์ ไหลภาภรณ์	มหาวิทยาลัยวลัยลักษณ์
68	ผศ. ดร.พิศุทธวรรณ ศรีภิรมย์ สิรินิลกุล	มหาวิทยาลัยศรีนครินทรวิโรฒ
69	ผศ. ดร.วิศรุต โพธิ์อ้น	มหาวิทยาลัยศรีนครินทรวิโรฒ
70	ผศ. ดร.ศญาพัฒน์ สุขใส	มหาวิทยาลัยศรีนครินทรวิโรฒ
71	ผศ.ปัญญวัฒน์ หาอาษา	มหาวิทยาลัยศรีนครินทรวิโรฒ
72	อ. ดร.ธีรศักดิ์ ฉลาดการณ์	มหาวิทยาลัยศรีนครินทรวิโรฒ
73	รศ. ดร.พรทรัพย์ พรสวัสดิ์	มหาวิทยาลัยศิลปากร
74	ผศ. ดร.สวรรยา ศกุนตะเสฐียร	มหาวิทยาลัยศิลปากร
75	พ.ท.หญิง ผศ.ณหทัย สระกบแก้ว	โรงเรียนนายร้อยพระจุลจอมเกล้า
76	รศ. ดร.อรวรรณ ตรีภักดิ์	มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่
77	รศ. ดร.เอธัสวัฒน์ คำมณี	มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่
78	ผศ. ดร.กิตติศักดิ์ ชุมพงศ์	มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่
79	ผศ. ดร.นัฐดา จิเบ็ญจะ	มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่
80	ผศ. ดร.ภานุพงศ์ วิจิตรคุณากร	มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่
81	รศ. ดร.อนิรุทธ ผลอ่อน	มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตปัตตานี
82	รศ. ดร.อาทิตย์ อินทรสิทธิ์	มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตปัตตานี
83	รศ. ดร.อารียุทธ สมาแอ	มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตปัตตานี
84	ผศ. ดร.นิฟาตมะห์ มะกาเจ	มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตปัตตานี
85	ผศ. ดร.อารีนา ฮะซานี	มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตปัตตานี

86	รศ. ดร.นพรัตน์ โพธิ์ชัย	สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหาร-ลาดกระบัง
87	ผศ. ดร.พุทธา สักกะพลางกูร	สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหาร-ลาดกระบัง

ฝ่ายจัดทำเอกสาร

1 อ. ดร.กฤษดา นารอง
2 อ. ดร.วรยุทธ วงษ์นิล
3 อ. ดร.วิชิต สมบัติ
4 อ.ธวัชชัย สลางสิงห์
5 อ.ธนาตย์ เดโชชัยพร
6 อ. ดร.ไพชยนต์ คงไชย
7 ผศ. ดร.วีรยุทธ นิลสระคู
8 รศ. ดร.รตนกร วัฒนทวีกุล
9 ผศ.รตี โบจรัส
10 นายนัฐพงษ์ สืบสุข
11 นางสาวพูลพิศมัย ไพศาลธรรม

คณะกรรมการจัดการประชุมวิชาการ ทางคณิตศาสตร์ ครั้งที่ 28 ประจำปี 2567

The 28^{th} Annual Meeting in Mathematics 2024

คณะกรรมการที่ปรึกษา

1	รศ. ดร.ชุตินันท์ ประสิทธิ์ภูริปรีชา	อธิการบดีมหาวิทยาลัยอุบลราชธานี
2	ศ. ดร.ศิริพร จึงสุทธิวงษ์	คณบดีคณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี
3	รศ. ดร.ชาญชัย ศุภอรรถกร	หัวหน้าภาควิชาคณิตศาสตร์ สถิติ และคอมพิวเตอร์
4	รศ. ดร.ศจี เพียรสกุล	ผู้อำนวยการศูนย์ส่งเสริมการวิจัยคณิตศาสตร์ แห่งประเทศไทย
5	ศ.กิตติคุณ ดร.พัฒนี อุดมกะวานิช	นายกสมาคมคณิตศาสตร์แห่งประเทศไทย ในพระบรมราชูปถัมภ์

คณะกรรมการฝ่ายดำเนินการ

1. รศ. ดร.ศราวุธ แสนการุณ
2. รศ. ดร.รตนกร วัฒนทวีกุล
3. รศ. ดร.ฐิตารีย์ วุฒิจิรัฐติกาล
4. รศ. ดร.เชิดศักดิ์ บุตรจอมชัย
5. ผศ. ดร.กนกพร ช่างทอง
6. ผศ. ดร.คณิศา โชติจันทึก
7. ผศ. ดร.ชัชวิน นามมั่น
8. ผศ. ดร.ชิดหทัย เพชรช่วย
9. ผศ. ดร.พัชรี วงษาสนธิ์
10. ผศ. ดร.ไพรินทร์ สุวรรณศรี
11. ผศ. ดร.วีรยุทธ นิลสระคู
12. ผศ. ดร.สุพจน์ สีบุตร
13. ผศ.รตี โบจรัส
14. อ. ดร.กฤษดา นารอง
15. อ. ดร.จิรัชยา ใจสะอาดซื่อตรง
16. อ. ดร.ธนวิทย์ จีรุพันธ์
17. อ. ดร.นงคราญ สระโสม
18. อ. ดร.ไพชยนต์ คงไชย
19. อ. ดร.วรยุทธ วงศ์นิล
20. อ. ดร.วิชิต สมบัติ

21. อ. ดร.ศักดิ์ดา น้อยนาง
22. อ. ดร.สมปอง เวฬุวนาธร
23. อ.กุลธรา มหาดิลกรัตน์
24. อ.ธนาตย์ เดโชชัยพร
25. อ.ธวัชชัย สลางสิงห์
26. นางสาวจิราภรณ์ ทองสุด
27. นางสาวดุจฤทัย สหพงษ์
28. นางสาวพูลพิศมัย ไพศาลธรรม
29. นางสาวมลฤดี กาญจนวงษ์
30. นางสาวลลิตภัทรา ริมทอง
31. นางสาววิศัลยา จันทร์เกษมสุข
32. นางสาวศิรดาภักดิ์ พิทักษา
33. นางสาวสุตินทรณ์ อาชญาทา
34. นางสาวสุนิสา นาครินทร์
35. นางสาวอมรรัตน์ วะสุรีย์
36. นางกานต์อนงค์ นิตรักษ์
37. นางเกษมณี โสภาณเวช
38. นางทุติยาภรณ์ วีระกุล
39. นางนันทนา พิมพ์พันธ์
40. นางพิกุล ยิ่งยง
41. นางเรไร กาฬบุตร
42. นางวรุณี ไชยกาล
43. นางศันสนีย์ สืบสุข
44. นางศิริพร ระวี
45. นางสมหญิง บุตรจอมชัย
46. นางสุกัญญา พิมพ์บุญมา
47. นายกมล คำพิบูลย์
48. นายชาติชนะ โมพีชาติ
49. นายธนศิลป์ ทองไทย
50. นายนราธิป ธรรมเรือง
51. นายนัฐพงษ์ สืบสุข
52. นายประจักกิจ ระวี
53. นายภูมรินทร์ ทองแดง
54. นายรัชต์วิภพ มีทรัพย์รุ่งโรจน์
55. นายวิชิต คำภูบาล
56. นายศุภชัย เชื้อพันธ์
57. นายอภัยวรรณ สุระพร
58. นายอภิรักษ์ ทูลภิรมย์


AMM
28TH ANNUAL MEETING
IN MATHEMATICS
2024