บทที่ ๒

ทฤษฎีที่ใช้ในการวิจัย

ในการวิจัยเรื่องนี้มีความจำเป็นอย่างมากที่ต้องอาศัยความรู้ในหลาย ๆ ด้าน ซึ่งพอจะสรุปเป็นประเภทใหญ่ ๆ ได้ดังนี้

- ทฤษฎีทางสถิติ (Statistical Theory)
- เทคนิคการค้นหาค่าอุตมะ (Optimization Technique)
- วิธีการวิเคราะห์ตัวเลข (Numerical Analysis)

ในบทที่ ๒ นี้ จะได้กล่าวถึงทฤษฎีทางสถิติและเทคนิคการหาค่าอุตมะเสียก่อน และจะได้กล่าวถึงวิธีการวิเคราะห์เชิงตัวเลขต่อไปในบทที่ ๓

สำหรับทฤษฎีทางสถิติที่ใช้ในวิทยานิพนธ์บทนี้ได้แก่

- การทดสอบแบบไคสแควร์ (หัวข้อ ๒.๑)
- การแจกแจงแบบปกติ (หัวข้อ ๒.๒)
- การทดสอบแบบสติวเด้นท์ที (t-Test) (หัวข้อ ๒.๓)

ในหัวข้อ ๒.๔ จะได้กล่าวถึงเทคนิคการค้นหาค่าอุตมะที่ใช้ในวิทยานิพนธ์ฉบับนี้ คือ เทคนิคการค้นหาตัวแปรเดียว (Univariate Search) วิธีทิศทางสังยุคของพาวเวลล์ (Powell Conjugate direction Method) ซึ่งเมื่อได้ดำเนินการศึกษาแล้วค้นพบว่ามีข้อผิดพลาดเกิดขึ้นในวิธีดังกล่าวนั้น และจะได้กล่าวถึงโดยละเอียดในหัวข้อ ๒.๕ และ ๒.๖

๒.๑ การทดสอบแบบไคสแควร์ (Chi-square Test)[๑]

การทดสอบแบบไคสแควร์เป็นสถิติทดสอบที่ใช้ทดสอบการแจกแจงความน่าจะเป็นของข้อมูลที่ได้จากตัวอย่างว่าเป็นชนิดเดียวกันกับการแจกแจงมาตรฐานที่คาดหวังไว้หรือไม่ ค่าสถิติทดสอบไคสแควร์ (χ^2) คือ

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \qquad \text{.........(2.1)}$$

๑"ความน่าจะเป็นและสถิติ" แผนกวิชาคณิตศาสตร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, หน้า ๒๔๐, พ.ศ. ๒๕๒๐.

โดยมี n = จำนวนช่วงของข้อมูล

O_1 = จำนวนความถี่ของข้อมูลในช่วงที่ ๑ คือ $W_1 - W_2$ โดยรวม ค่าที่ W_2 แต่ไม่รวมค่าที่ W_1

O_2 = จำนวนความถี่ของข้อมูลในช่วงที่ ๒ คือ $W_2 - W_3$ โดยรวม ค่าที่ W_3 แต่ไม่รวมค่าที่ W_2

O_i = จำนวนความถี่ของข้อมูลในช่วงที่ i คือ $W_i - W_{i+1}$ โดยรวม ค่าที่ W_{i+1} แต่ไม่รวมค่าที่ W_i

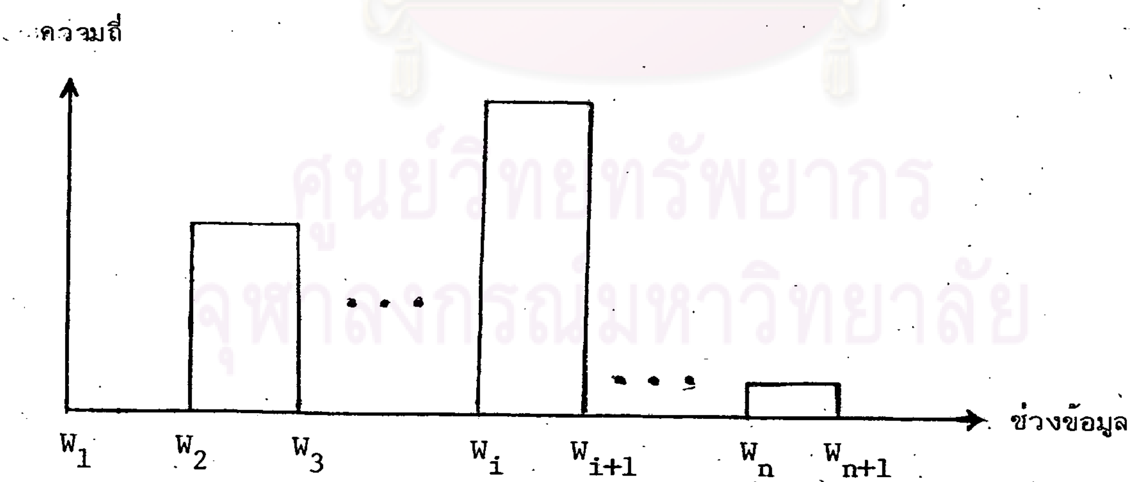
E_i = ค่าคาดหมายของจำนวนความถี่ของข้อมูลในช่วงที่ i คือ ช่วง $W_i - W_{i+1}$ โดยที่ค่าคาดหมายนี้หาได้จากฟังค์ชั่นการแจกแจงที่คาดหวังไว้กับจำนวนข้อมูลทั้งหมดในการทดลอง

ซึ่ง $E_i = N \times [\,F(W_{i+1}) - F(W_i)\,]$

เมื่อ N = จำนวนข้อมูลทั้งหมด (Total number of observations)

รูปที่ ๒.๑

ฮิสโตแกรมที่ได้จากความถี่ของข้อมูล



ถ้า $f(x)$ = ฟังค์ชั่นการแจกแจงมาตรฐานที่นำมาทดสอบและเป็นฟังค์ชั่นแบบต่อเนื่อง (Continuous Function)

$$F(W_{i+1}) = \int_{o}^{W_{i+1}} f(x)dx$$

$$F(W_i) = \int_{o}^{W_i} f(x)dx$$

ถ้า $f(x)$ = ฟังค์ชั่นการแจกแจงมาตรฐานที่นำมาทดสอบและเป็นฟังค์ชั่นแบบไม่ต่อเนื่อง (Discrete Function)

$$F(W_{i+1}) = \sum_{x=W_1}^{W_{i+1}} f(x)$$

$$F(W_i) = \sum_{x=W_1}^{W_i} f(x)$$

ฟังค์ชั่นที่นำมาทดสอบการแจกแจงของข้อมูลบางฟังค์ชั่น ไม่สามารถใช้หาค่าคาดหมายของจำนวนความถี่ของข้อมูลตามสูตรที่กล่าวมาแล้วโดยตรง จึงใช้วิธีการวิเคราะห์เชิงตัวเลข (Numerical Analysis) คำนวณค่าโดยประมาณ สำหรับการวิจัยนี้ฟังค์ชั่นการแจกแจงที่นำมาทดสอบคือ การแจกแจงแบบปกติ ซึ่งไม่สามารถหาค่าคาดหมายตามสูตรได้ รายละเอียดของการหาค่าคาดหมายนี้จะกล่าวถึงโดยละเอียดในบทที่ ๓ หัวข้อ ๓.๑ ถ้าความถี่ที่ได้จากการทดลองมีค่าใกล้เคียงกับความถี่ที่ได้ตามทฤษฎี χ^2 ก็จะมีค่าน้อย ถ้าความถี่ต่างกันมาก ค่าของ χ^2 ก็จะมากด้วย ฉะนั้นขอบเขตที่ยอมรับได้ก็คือ เมื่อ $\chi^2 < \chi^2_\alpha$[๑] โดยที่ χ^2_α เป็นค่าที่ได้จากตารางการแจกแจงไคสแควร์ เมื่อระดับความมีนัยสำคัญคือ α

---

[๑] เรื่องเดียวกัน, หน้า ๒๔๑.

สำหรับชั้นแห่งความเป็นอิสระ (ν) สำหรับการแจกแจงไคสแควร์ในที่นี้ขึ้นอยู่กับองค์ประกอบ ๒ ประการ

- จำนวนค่าสังเกตในตัวอย่าง (n)
- จำนวนค่าสถิติที่ได้จากตัวอย่างเพื่อนำมาใช้ในการคำนวณหาความถี่ที่คาดว่าจะได้ตามทฤษฎี (k)

โดยที่ $\nu = n - k - 1$

๒.๒ การแจกแจงแบบปกติ (Normal Distribution)

การแจกแจงแบบปกติเป็นการแจกแจงความน่าจะเป็นของตัวแปรสุ่มที่มีค่าต่อเนื่องซึ่งสำคัญที่สุด เพราะข้อมูลส่วนมาก เช่น ข้อมูลที่เกี่ยวกับการวิจัย การอุตสาหกรรม การศึกษา ฯลฯ จะมีการแจกแจงประเภทนี้

ตัวแปรสุ่มที่มีการแจกแจงแบบปกติ เรียกว่า ตัวแปรสุ่มปกติ (Normal random variable) ถ้าค่าเฉลี่ยของตัวแปรสุ่มปกติ x คือ μ และส่วนเบี่ยงเบนมาตรฐานคือ σ ฟังค์ชั่นการแจกแจงแบบปกติของ x เขียนโดย สัญญลักษณ์ $N(x ; \mu , \sigma^2)$

$$N(x ; \mu , \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \text{เมื่อ } -\infty < x < \infty$$

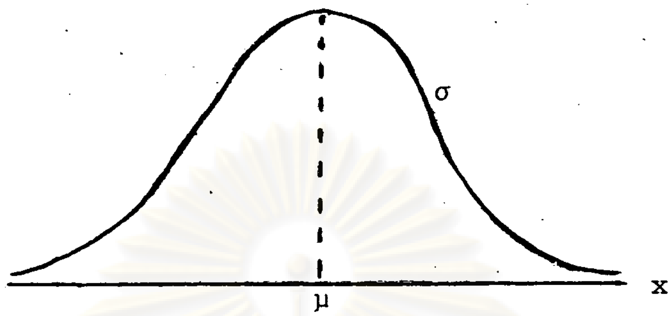
$$\pi = 3.14159...$$

$$e = 2.71828...$$

ลักษณะของกราฟของ $N(x ; \mu , \sigma^2)$ แสดงไว้ในรูปที่ ๒.๒

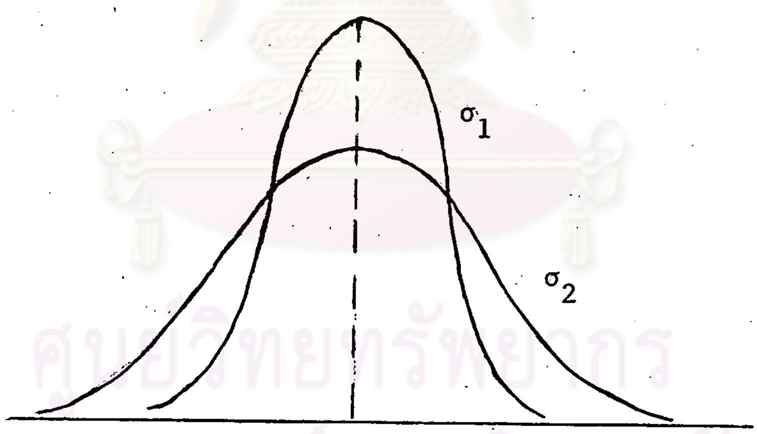
<u>รูปที่ ๒.๒</u>

โค้งปกติ $n\ (x\ ;\ \mu\ ,\ \sigma^2)$



<u>รูปที่ ๒.๓</u>

โค้งปกติที่มี $\mu_1 = \mu_2$ และ $\sigma_1 < \sigma_2$



รูปที่ ๒.๓ แสดงว่าโค้งปกติที่มีค่าเบี่ยงเบนมาตรฐานมาก จะมีความโด่งน้อยกว่าและกระจายออกมากกว่าโค้งปกติที่มีค่าเบี่ยงเบนมาตรฐานน้อย

คุณสมบัติของโค้งปกติมีดังนี้

ก. เส้นโค้งปกติจะมีความถี่สูงรวมอยู่ที่ค่ามัชฌิมเลขคณิต (μ) และจะกระจายลดน้อยลงไปทางค่าสูงและค่าต่ำอย่างสม่ำเสมอ

ข. เส้นโค้งจะมีลักษณะสมมาตรที่ μ ค่าออร์ดิเนตของจุด $x = \mu \pm k$ เมื่อเป็นค่าคงที่จะเท่ากัน เช่น ความสูงของเส้นโค้งที่จุด $x = \mu + \sigma$ จะเท่ากับความสูงของเส้นโค้งที่จุด $x = \mu - \sigma$ จากความสมมาตรของเส้นโค้งนี้เอง จะได้ค่ามัชฌิมเลขคณิตมัธยฐานและฐานนิยมมีค่าเท่ากัน (mean = median = mode)

ค. เส้นโค้งปกติมีลักษณะเป็นรูปประฆัง ปลายเส้นโค้งทั้งสองจะค่อย ๆ ลาดลงสู่แกนนอน แต่ไม่มีโอกาสที่จะสัมผัสกับแกนนอนเลย พื้นที่ภายใต้เส้นโค้งปกติที่มีคะแนนมาตรฐานมากกว่า $\mu \pm 3$ จะมีค่าน้อยมาก

ง. พื้นที่ทั้งหมดใต้เส้นโค้งปกติ และอยู่เหนือแกน x มีค่า = 1

<u>ค่าเฉลี่ยของ X คือ E(X)</u>

พิสูจน์ $$E(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} x e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}\, dx$$

ให้ $$Z = \frac{x-\mu}{\sigma} \quad \therefore\ dx = \sigma dz$$

$$E(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-z^2/2} (\mu + \sigma z)\, dz$$

$$= \frac{\mu}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} e^{-z^2/2}\, dz + \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z e^{-z^2/2}\, dz$$

พิจารณา $$\int_{-\infty}^{+\infty} e^{-z^2/2}\, dz$$

ให้ $$I = \int_{-\infty}^{+\infty} e^{-u^2/2}\, du$$

$$\therefore \quad I^2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-(u^2+v^2)/2}\, du\,dv$$

โดยทฤษฎีเกี่ยวกับการเปลี่ยนตัวแปรจาก u, v ให้มาอยู่ในโพลาโคออดิเนท จะได้

$$I^2 = \int_{0}^{\infty}\int_{0}^{2\pi} e^{-r^2/2}\, r\, d\theta\, dr$$

$$= 2\pi \int_{0}^{\infty} e^{-r^2/2}\, r\, dr$$

$$= 2\pi \left[-e^{-r^2/2} \right]_{0}^{\infty} = 2\pi$$

$$\therefore \quad I = \int_{-\infty}^{+\infty} e^{-u^2/2}\, du = \sqrt{2\pi}$$

และ $\int_{-\infty}^{\infty} z e^{-z^2/2}\, dz = \left[-e^{-z^2/2} \right]_{-\infty}^{+\infty} = 0$

$$\therefore \quad E(X) = \frac{\mu}{\sqrt{2\pi}} \sqrt{2\pi} + \frac{\sigma}{\sqrt{2\pi}} \cdot 0 = \mu$$

<u>ความแปรปรวนของ X คือ $E(X-\mu)^2$</u>

พิสูจน์ $E(X-\mu)^2 = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-\mu)^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$

ให้ $z = \frac{x-\mu}{\sigma}, \quad dx = \sigma dz$

$$\therefore \quad E(X-\mu)^2 = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z^2 e^{-z^2/2}\, dz$$

ให้ $u = z, \quad du = dz$

$$dv = ze^{-z^2/2}\,dz$$

$$\therefore\quad v = -e^{-z^2/2}$$

$$\therefore\quad E(X-\mu)^2 = \frac{\sigma^2}{\sqrt{2\pi}}\left(-z\,e^{-z^2/2}\Big]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} e^{-z^2/2}\,dz\right)$$

$$= \frac{\sigma^2}{\sqrt{2\pi}}\,(0 + \sqrt{2\pi}\,)$$

$$= \sigma^2$$

แสดงว่าตัวแปรสุ่ม X ที่มีฟังก์ชั่นการแจกแจงเป็นแบบปกติ

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\,e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

จะมีค่าเฉลี่ย μ และความเบี่ยงเบน σ

๒.๓ การทดสอบแบบสติวเด้นท์ที (t-test)[๑]

การทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ยของประชากรสองชุดว่ามีความแตกต่างกันหรือไม่ โดยเลือกตัวอย่างมาจำนวนหนึ่งจากประชากรทั้งสองชุด แล้วทำการทดสอบซึ่งโดยใช้ตัวสถิติ t กล่าวคือ

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_{\bar{X}_1 - \bar{X}_2}} \qquad \text{.........(2.2)}$$

เมื่อ $\bar{X}_1$ และ $\bar{X}_2$ เป็นค่าเฉลี่ยจากตัวอย่างที่เลือกมาจากประชากรชุดที่หนึ่ง และชุดที่สองตามลำดับ ซึ่งมีขนาดตัวอย่างเป็น n_1 และ n_2

μ_1 , μ_2 เป็นค่าเฉลี่ยของประชากรชุดที่หนึ่งและชุดที่สอง

$S_{\bar{X}_1 - \bar{X}_2}$ เป็นความคลาดเคลื่อนมาตรฐานของผลต่าง ระหว่างค่าเฉลี่ยจากตัวอย่างที่เลือกจากประชากรทั้งสองชุด

---

[๑] เรื่องเดียวกัน, หน้า ๒๒๕.

ในกรณีที่ความแปรปรวนของประชากรชุดที่หนึ่งและชุดที่สอง (σ_1^2 และ σ_2^2) เท่ากัน สมมติให้เท่ากับ σ^2

$$S_{\bar{X}_1 - \bar{X}_2} = \sqrt{S^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

โดยที่
$$S^2 = \left\{ \sum_{j=1}^{n_1} (X_{1j} - \bar{X}_1)^2 + \sum_{j=1}^{n_2} (X_{2j} - \bar{X}_2)^2 \right\} \Big/ (n_1 + n_2 - 2)$$

สำหรับค่าองศาแห่งความเป็นอิสระของ t หาจากสูตร $n_1 + n_2 - 2$

ถ้าความเบี่ยงเบนมาตรฐานของประชากรทั้งสองไม่เท่ากัน $\sigma_1^2 \neq \sigma_2^2$

$$S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

เมื่อ
$$S_1^2 = \sum_{j=1}^{n_1} \frac{(X_{1j} - \bar{X}_1)^2}{n_1 - 1}$$

$$S_2^2 = \sum_{j=1}^{n_2} \frac{(X_{2j} - \bar{X}_2)^2}{n_2 - 1}$$

สำหรับค่าองศาแห่งความเป็นอิสระของ t หาจากสูตร

$$f = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(S_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(S_2^2/n_2\right)^2}{n_2 - 1}}$$

ค่าเฉลี่ยของตัวอย่างและความเบี่ยงเบนมาตรฐานของตัวอย่างนี้ ไม่จำเป็นต้องมีค่าเท่ากับค่าเฉลี่ยของประชากร และส่วนเบี่ยงเบนมาตรฐานของประชากรเสมอไป แต่จะมีค่าใกล้เคียงกัน เพียงไรนั้นก็ต้องขึ้นอยู่กับวิธีการเลือกตัวอย่างและขนาดของตัวอย่างเป็นสำคัญ ถ้านำเอาค่าเฉลี่ยของตัวอย่างจากตัวอย่างทุก ๆ ชุดที่สามารถเลือกขึ้นมาได้จากประชากรนั้น ๆ มาหาค่าเฉลี่ยอีกครั้งหนึ่ง ค่าเฉลี่ยที่ได้ใหม่นี้จะเท่ากับค่าเฉลี่ยของประชากร

$$E(\bar{X}) = \mu$$

ค่าเฉลี่ยจากค่าเฉลี่ยของตัวอย่างทุกชุดที่สามารถเลือกมาได้จากประชากรทั้งหมด จะใช้สัญญลักษณ์เป็น

$$\bar{\bar{X}} = \sum_{i=1}^{a} \bar{X}_i / a$$

เมื่อ a เป็นจำนวนชุดของตัวอย่างขนาด n ซึ่งเลือกมาจากประชากรขนาด N

$\bar{X}_i$ เป็นค่าเฉลี่ยของตัวอย่างชุดที่ i, $i = 1, 2, 3, \ldots, a$

ถ้านำเอาค่าของความแปรปรวนของตัวอย่างทุก ๆ ชุดที่สามารถเลือกขึ้นมาได้จากประชากรนั้น ๆ มาหาค่าเฉลี่ย ค่าเฉลี่ยที่ได้ใหม่นี้จะเท่ากับค่าของความแปรปรวนของประชากร $E(S^2) = \sum_{i=1}^{a} S_i^2/a = \sigma^2$

ค่าเฉลี่ยที่ได้จากตัวอย่างแต่ละชุด $(\bar{X}_i)$ ก็จะมีการเบี่ยงเบนไปจากค่าเฉลี่ยของ $\bar{X}$ $(\bar{\bar{X}})$ เช่นเดียวกันกับที่ X_i มีการเบี่ยงเบนไปจากค่าเฉลี่ยของ X $(\bar{X})$ เมื่อรวมค่าของความคลาดเคลื่อนยกกำลังสองทั้งหมดเข้าด้วยกันแล้วเฉลี่ยจะเรียกว่าเป็นความแปรปรวนของค่าเฉลี่ยของตัวอย่าง $\sigma_{\bar{X}}^2$ ซึ่งเมื่อถอดกรณฑ์ที่สองของ $\sigma_{\bar{X}}^2$ ก็จะได้ค่าส่วนเบี่ยงเบนมาตรฐานของค่าเฉลี่ยของตัวอย่าง $\sigma_{\bar{X}}$

$$\sigma_{\bar{X}}^2 = \sum_{i=1}^{a} (\bar{X}_i - \bar{\bar{X}})^2 / a$$

$$\sigma_{\bar{X}} = \sqrt{\sum_{i=1}^{a} \frac{(\bar{X}_i - \bar{\bar{X}})^2}{a}}$$

ค่าของความแปรปรวนของ $\bar{X}$ ($\sigma^2_{\bar{X}}$) นี้ย่อมมีค่าน้อยกว่าค่าของความแปรปรวนของประชากร (σ^2) ทั้งนี้เนื่องจาก $\sigma^2_{\bar{X}}$ เป็นความแปรปรวนของค่าเฉลี่ยของตัวอย่างชุดต่าง ๆ อีกทีหนึ่ง และ $\sigma^2_{\bar{X}}$ นี้จะมีค่าน้อยกว่า σ^2 เป็นจำนวน n เท่า เมื่อ n เป็นขนาดของตัวอย่างที่เลือกมาจากประชากรทั้งหมด

$$\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$$

ก่อนที่จะพูดถึงวิธีการค้นหาตัวแปรเดียวในหัวข้อ ๒.๔ จะขอกล่าวถึงปัญหาทั่ว ๆ ไปของการหาค่าที่มากที่สุด (หรือน้อยที่สุด) ของฟังก์ชั่นตัวแปรหนึ่งตัวแปรหรือฟังก์ชั่นของตัวแปรหลายตัวแปร วิธีการที่รู้จักกันโดยทั่ว ๆ ไปในการค้นหาตัวแปรที่ทำให้ฟังก์ชั่นต่ำสุดนั้นใช้ Classical optimization technique ซึ่งวิธีการนี้ใช้แคลคูลัสหรือต้องอาศัยอนุพันธ์ (derivative) ของฟังก์ชั่นในการค้นหาค่าที่เหมาะสม (optimize) ของ X ที่ทำให้ f(X) มีค่าต่ำที่สุด หรือใช้ในการหาเวคเตอร์ที่เหมาะสมของ $\underline{X}$ ที่ทำให้ $f(\underline{X})$ ต่ำที่สุด โดยที่ $\underline{X} = (x_1, x_2, \ldots, x_n)$

มีข้อเสียหลายอย่างต่อการใช้ classical optimization technique ซึ่งจำกัดการใช้เทคนิคเหล่านี้ในทางภาคปฏิบัติ เพราะโดยทั่ว ๆ ไปแล้ว classical optimization technique ต้องใช้กับฟังก์ชั่นที่มีความต่อเนื่อง (continuous) หรือต้องหาอนุพันธ์ได้ (differentiable) เงื่อนไขเหล่านี้มักจะไม่มีในปัญหาทางภาคปฏิบัติที่เราต้องแก้ปัญหาเสมอไป เพราะเราไม่สามารถรู้ได้ว่าฟังก์ชั่นที่ต้องแก้ปัญหามีความต่อเนื่อง (Continuous) หรือหาอนุพันธ์ได้ (differentiable)

จากปัญหาที่พบในทางภาคปฏิบัติจึงมีวิธีการหาตัวแปรที่ทำให้ฟังก์ชั่นต่ำสุด (หรือสูงที่สุด) โดยไม่ต้องอาศัยการหาอนุพันธ์ของฟังก์ชั่น ซึ่งเทคนิคการค้นหาตัวแปรโดยไม่ต้องอาศัยอนุพันธ์มีวิธีการหลายวิธีซึ่งแต่ละวิธีก็มีเงื่อนไขของการนำเอาไปใช้ให้มีประสิทธิภาพแตกต่างกัน สำหรับการวิจัยนี้ใช้วิธีการค้นหาตัวแปรเดียว (Univariate Search) ซึ่งเป็นวิธีที่อาจใช้อนุพันธ์ของฟังก์ชั่นก็ได้หรือไม่ต้องใช้อนุพันธ์ของฟังก์ชั่นก็ได้

๒.๔ วิธีการค้นหาตัวแปรเดียว (Univariate Search Method)

ในการแก้ปัญหาฟังก์ชั่นที่เชื่อมโยงกับตัวแปรทุก ๆ ตัวในขณะเดียวกัน ในทางตรรกวิทยาแล้วเราอาจจะพิจารณาได้หลาย ๆ วิธีเพื่อลดจำนวนการคำนวณ วิธีที่ง่ายวิธีหนึ่งคือ การเปลี่ยนตัวแปรครั้งละตัว ซึ่งเรียกว่าการค้นหาตัวแปรเดียว แนวความคิดของวิธีการนี้ก็คือ เปลี่ยนตัวแปรครั้งละหนึ่งตัวแปร เพื่อว่าฟังก์ชั่นที่มีค่ามากที่สุด (หรือน้อยที่สุด) หาค่าได้ในทิศทางร่วมแต่ละทิศทาง

หลักการของวิธีการค้นหาตัวแปรเดียวมีดังนี้ [๑]

(ก) กำหนดให้ $\bar{X}_0$ เป็นจุดเริ่มต้นในการหาค่า X_1 , $X_2, \ldots, X_n$ ที่จะทำให้ $f(\bar{X})$ มีค่าต่ำสุด โดยที่ $\bar{X}_0 = [X_{01}, X_{02}, \ldots, X_{0n}]$ ทราบค่ามาแล้วจากข้อมูลหรือกำหนดขึ้นมา

(ข) ให้ $\bar{X}_1$ เป็นจุดถัดไป โดยที่ $\bar{X}_1 = \bar{X}_0 + \lambda_1 \bar{e}_1$ และ $\bar{e}_1 = [1,0,0,\ldots,0]$

λ_1 คือเลขจำนวนจริงที่ทำให้ $f(\bar{X}_0 + \lambda_1 \bar{e}_1)$ มีค่าน้อยที่สุด นั่นคือเราจะแก้ปัญหา

$$\min_{\lambda_1} f(\bar{X}_0 + \lambda_1 \bar{e}_1)$$

สูตรโดยทั่ว ๆ ไปในข้อ ข. นี้ ก็คือ ถ้าจะหาจุด $\bar{X}_{k+1}$ เพื่อที่จะทำให้ฟังก์ชั่นมีค่าน้อยที่สุดเทียบกับตัวแปร X_k นั่นคือ

$$\bar{X}_{k+1} = \bar{X}_k + \lambda_{k+1} \bar{e}_{k+1} \quad k = 0, 1, 2, \ldots, n-1$$

เพื่อทำให้ $f(\bar{X}_k + \lambda_{k+1} \bar{e}_{k+1})$ มีค่าต่ำสุด

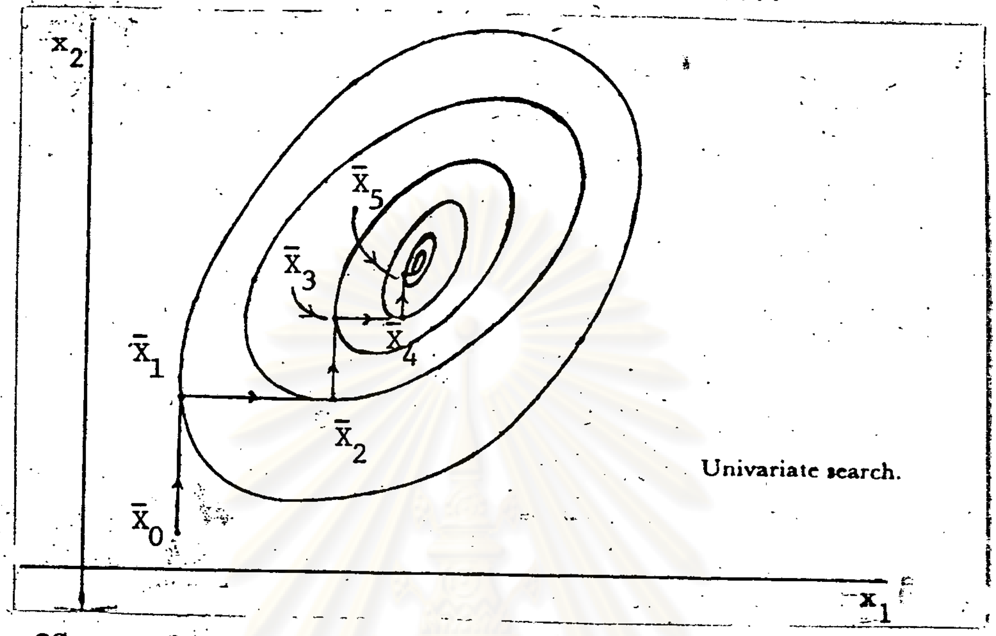
(ค) จะหาจุดถัดไปตามสูตรในข้อ ข. ไปเรื่อย ๆ จนกว่า $|\lambda_k| <$ ค่าใดค่าหนึ่งที่เรากำหนด (Some tolerance value)

---

[๑] Cooper and Steinberg, "Introduction to methods of optimization," W.B. Saunders Company, Philadelphia, London Toronto, pp. 156, 1970.

รูปที่ ๒.๔

แสดงรูปการค้นหาตัวแปรทีละตัวในทิศทางแต่ละทิศทาง



วิธีการหา λ_k

จากข้อ ข. จะพบว่าเราต้องหาค่า λ_k เพื่อที่จะให้ $f(\bar{X}_k)$ มีค่าต่ำสุดนั้น การหาค่า λ_k ขึ้นอยู่กับลักษณะของฟังก์ชั่นที่ต้องการหาค่าต่ำสุด สำหรับการวิจัยนี้ฟังก์ชั่นที่ใช้ในการวิจัยคือ ตัวสถิติทดสอบไคสแควร์ (χ^2) ซึ่งเป็นฟังก์ชั่นกำลังสอง (Quadratric function) เพราะ

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

โดยทั่ว ๆ ไปแล้วการหาค่า λ_k ที่ทำให้ $f(\bar{X}_k)$ ต่ำที่สุดนั้น ทำได้ ๒ วิธีคือ

- วิธีการทางแคลคูลัส (อาศัยอนุพันธ์)
- วิธีการค้นหามิติเดียวโดยไม่ใช้อนุพันธ์ (One-dimensional search method)

๒.๔.๑ วิธีการทางแคลคูลัส

เป็นวิธีการที่ต้องใช้อนุพันธ์ของฟังก์ชั่นที่กำลังสนใจศึกษา

สมมติให้ $f(\bar{X}_k)$ เป็นฟังก์ชั่นกำลังสอง นั่นคือ $f(\bar{X}_k)$ สามารถเขียนให้อยู่ในรูป

$$f(\bar{X}_k) = \bar{X}_k' A\bar{X}_k + \bar{b}' \bar{X}_k + C$$

โดยที่ $\bar{X}_k = \bar{X}_{k-1} + \lambda_k \bar{e}_k$

$$\frac{\partial f(\bar{X}_k)}{\partial \lambda_k} \quad \frac{\partial [\bar{X}_{k-1} + \lambda_k \bar{e}_k]' A [\bar{X}_{k-1} + \lambda_k \bar{e}_k] + \bar{b}' [\bar{X}_{k-1} + \lambda_k \bar{e}_k] + C}{\partial \lambda_k}$$

$$\frac{\partial f(\bar{X}_k)}{\partial \lambda_k} = \frac{\partial}{\partial \lambda_k} \Big\{ \bar{X}_{k-1}' A\bar{X}_{k-1} + \lambda_k \bar{e}_k' A\bar{X}_{k-1} + \bar{X}_{k-1}' A \lambda_k \bar{e}_k + \lambda_k^2 \bar{e}_k' A \bar{e}_k + \bar{b}' \bar{X}_{k-1}$$

$$+ \bar{b}' \lambda_k \bar{e}_k + C \Big\}$$

$$\frac{\partial f(\bar{X}_k)}{\partial \lambda_k} = \bar{e}_k' A \bar{X}_{k-1} + \bar{X}_{k-1}' A \bar{e}_k + 2 \lambda_k \bar{e}_k' A \bar{e}_k + \bar{b}' \bar{e}_k = 0$$

$$2\lambda_k \bar{e}_k' A \bar{e}_k = - [\bar{e}_k' A \bar{X}_{k-1} + \bar{X}_{k-1}' A \bar{e}_k + \bar{b}' \bar{e}_k]$$

$$\lambda_k = -\frac{1}{2} \frac{[\bar{e}_k' A \bar{X}_{k-1} + \bar{X}_{k-1}' A \bar{e}_k + \bar{b}' \bar{e}_k]}{\bar{e}_k' A \bar{e}_k}$$

$\lambda_k = -\frac{1}{2}$ [...

โดยที่ $\bar{e}_k = [\ 0, 0, \ldots, 1, 0, 0, \ldots, 0\ ]$

ตัวที่ k

๒.๔.๑ วิธีการค้นหามิติเดี่ยวโดยไม่ใช้อนุพันธ์ (one-dimensional search method)

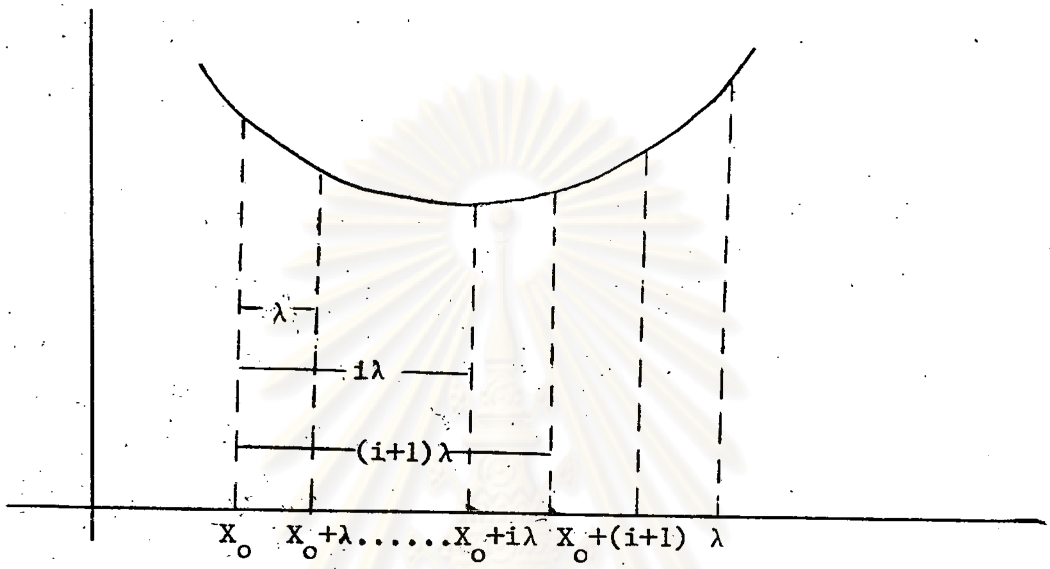
เป็นวิธีที่ไม่ต้องอาศัยอนุพันธ์ของฟังก์ชั่นเหมือนวิธีแรก โดยมีหลักง่าย ๆ ดังนี้ กำหนดช่วงก้าวแรกของ $\bar{X}_o$ ที่จะถูกเปลี่ยนไปในแต่ละทิศทาง โดยการกำหนดค่าของ λ ในตอนแรก การกำหนดค่าของ λ ตัวแรกเพื่อที่จะหา λ_k ตัวใหม่ที่ทำให้ $f(\bar{X}_k)$ มีค่าต่ำที่สุดในทิศทางที่ k นั้น กำหนดค่า λ ได้หลายค่า เช่น

$$\lambda = 0.001 , 0.01 , 0.1 , 0.5 , 1.0$$

นั่นคือ $f(X_o + i\lambda, X_1, X_2, \ldots, X_{n-1}) < f(X_o + (i + 1)\lambda, X_1, X_2 \ldots, X_{n-1})$

จะได้ว่า $\lambda_1 = i\lambda$ ที่ทำให้ $\min\limits_{\lambda_1} f(\bar{X}_o + \lambda_1 \bar{e}_1)$

รูปที่ ๒.๕



ถ้า $f(X_o, X_1, X_2, \ldots, X_{n-1}) < f(X_o + \lambda, X_1, X_2, \ldots, X_{n-1})$

คำนวณค่าของฟังก์ชั่น $f(\bar{X}_o)$ ใหม่ โดยใช้ตัวแปรตัวแรกเป็น

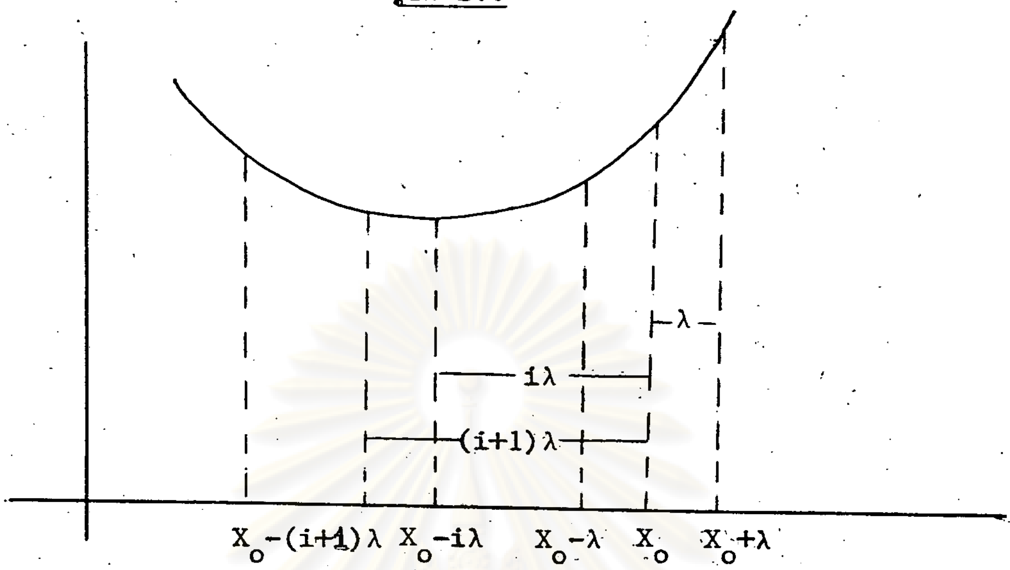
$X_o - \lambda,\ X_o - 2\lambda,\ X_o - 3\lambda, \ldots,\ X_o - i\lambda,\ X_o - (i+1)\ \lambda, \ldots, i = 1,2,3,\ldots$

จนกว่าค่าของฟังก์ชั่น $f(\bar{X}_o)$ เก่ามีค่าน้อยกว่าค่าของฟังก์ชั่น $f(\bar{X}_o)$ ใหม่

นั่นคือ $f(X_o - i\lambda, X_1, X_2, \ldots, X_n) < f(X_o - (i+1)\lambda, X_1, X_2, \ldots, X_n)$

จะได้ว่า $\lambda_1 = -i\lambda$ ที่ทำให้ $\min\limits_{\lambda_1} f(\bar{X}_o + \lambda_1 \bar{e}_1)$

รูปที่ ๒.๖



จะใช้ λ ค่าใดนั้นขึ้นอยู่กับความเหมาะสมของเวลาและผลที่ได้ ถ้าใช้ λ ที่มีค่าน้อย ๆ ความถูกต้องมีมากแต่เสียเวลาในการ iterate ถ้าใช้ λ ที่มีค่าใหญ่ขึ้นความถูกต้องมีน้อยลง แต่ไม่เสียเวลาในการ iterate มาก สำหรับการวิจัยนี้ ผู้วิจัยใช้ $\lambda = 0.01$ รายละเอียดในการเลือกค่าของ λ อยู่ในภาคผนวก ก

ในการหาค่าของตัวแปร $\bar{X}$ ในทิศทางใดทิศทางหนึ่ง ฟังก์ชั่นที่ต้องการหาค่าต่ำสุดนั้นมีตัวแปรเพียงตัวเดียวเท่านั้นที่ต้องการหาค่า ส่วนตัวแปรตัวอื่น ๆ ให้ถือว่าเป็นค่าคงที่ เช่น

$\bar{X} = [X_o, X_1, X_2, \ldots, X_{n-1}]$

ในทิศทางแรก ตัวแปรคือ X_0 ใน $f(X_0, X_1, X_2, \ldots, X_{n-1})$ สำหรับ $X_1, X_2, \ldots, X_{n-1}$ ให้ถือว่าคงที่ หา λ_1 ที่ทำให้ $\min_{\lambda_1} f(\bar{X}_0 + \lambda_1 \bar{e}_1)$

การหา λ_1

จากตัวแปร X_0 หา $f(X_0, X_1, X_2, \ldots, X_{n-1})$

เปลี่ยนค่า X_0 ให้ก้าวไปเป็นระยะ λ นั่นคือ $X_0 + \lambda$ แทนลงใน X_0 เดิม

คำนวณค่า $f(X_0 + \lambda, X_1, X_2, \ldots, X_{n-1})$

เปรียบเทียบ $f(X_0, X_1, X_2, \ldots, X_{n-1})$ กับ $f(X_0 + \lambda, X_1, X_2, \ldots, X_{n-1})$

(ก) ถ้า $f(X_0, X_1, X_2, \ldots, X_{n-1}) > f(X_0, \lambda, X_1, X_2, \ldots, X_{n-1})$

คำนวณค่าของฟังก์ชั่น $f(\bar{X}_0)$ ใหม่ โดยใช้ตัวแปรตัวแรกเป็น

$X_0 + 2\lambda$, $X_0 + 3\lambda$, $X_0 + 4\lambda, \ldots$, $X_0 + i\lambda$, $X_0 + i + 1\lambda, \ldots$ $i = 1,2,3,\ldots$

จนกว่าค่าของฟังก์ชั่น $f(\bar{X}_0)$ ค่าเก่าจะน้อยกว่าค่าของฟังก์ชั่น $f(\bar{X}_0)$ ค่าใหม่

ในทิศทางที่ k ตัวแปรคือ X_k ใน $f(X_0, X_1, X_2, \ldots, X_{n-1})$

สำหรับ $X_1, X_2, \ldots, X_{k-1}, X_{k+1}, \ldots, X_{n-1}$ ให้ถือว่าคงที่ แล้วหา λ_k

ที่ทำให้ $\min\limits_{\lambda_k} f(\bar{X}_{k-1} + \lambda_k \bar{e}_k)$

การหา λ_k

ใช้หลักในการหา λ_k เหมือนกับการหา λ_1

$$\text{โดยที่ } \lambda_k = \begin{cases} i\lambda & \text{ถ้า } f(X_0, X_1, \ldots, X_{k-1}, \ldots, X_{n-1}) > f(X_0, X_1, \ldots, X_{k-1+\lambda}, \ldots X_{n-1}) \\ -i\lambda & \text{ถ้า } f(X_0, X_1, \ldots, X_{k-1}, \ldots, X_{n-1}) < f(X_0, X_1, \ldots, X_{k-1+\lambda}, \ldots X_{n-1}) \end{cases}$$

คุณสมบัติของฟังก์ชั่นที่ใช้วิธีการค้นหาตัวแปรเดียว

- เป็นฟังก์ชั่นที่มีค่าสูงสุดต่ำสุดเพียงจุดเดียว
- ตัวแปรทั้งหลายของฟังก์ชั่นจะต้องไม่มีการกระทบระหว่างกันมาก (Strong interaction)

๒.๔ วิธีทิศทางสังยุคของพาวเวลล์ (Powell's Conjugate Direction Method)

ก่อนอื่นจะขอแนะนำคำนิยามที่เกี่ยวข้องกับทิศทางสังยุคเสียก่อน (Conjugate Direction)[๑]

ให้ $\underline{E}_0$, $\underline{E}_1$, $\underline{E}_2$, ..., $\underline{E}_{n-1}$ เป็นชุดของทิศทาง n ทิศทางที่ไม่เป็นศูนย์ ทิศทางเหล่านี้จะเรียกว่าเป็นสังยุค (Conjugate) ซึ่งกันและกันโดยเทียบกับ "เมทริกซ์สมมาตร A (symmetric matrix)" หรือเรียกสั้น ๆ ว่า A-สังยุค (A-Conjugate)

---

[๑] D.A. Pierre, "Optimization Theory with applications," New York, Jon Welley and Sons, Inc., March, 1961.

ถ้า $\underline{E}_i' A \underline{E}_j = 0$ เมื่อ $i \neq j$

ในที่นี้จะกำหนดให้ A เป็นเมทริกซ์สมมาตร A ที่มีมิติ $n \times n$ เป็น Positive-Definite Symmetric Matrix ก็ต่อเมื่อสำหรับทุก ๆ ค่าจริงที่ไม่เป็นศูนย์ของ $\underline{X}$ ที่มี order $n \times 1$ ทำให้ได้ $\underline{X}' A \underline{X} > 0$

ชุดของทิศทางสังยุคจะมีคุณสมบัติพิเศษดังนี้ "จุดที่ฟังค์ชั่นกำลังสอง (Quadratic function) คือ $Q(\underline{X}) = \underline{X}' A \underline{X} + b' \underline{X} + C$ มีค่าน้อยที่สุด อาจหาได้โดยทำการหาจุดต่ำสุดอย่างอนุบรรพไปตามทิศทางแต่ละทิศทางของ n ทิศทาง A-สังยุค (Sequence of a One-dimensional Search) และจำนวนทิศทางที่ใช้จะไม่มากกว่า n ทิศทางนี้ไม่ว่าจะเริ่มต้น ณ จุดใด ๆ โดยที่ n เป็นมิติของเวคเตอร์ $\underline{X}$"

แต่ถ้าฟังค์ชั่นที่จะหาค่าน้อยที่สุดไม่ได้เป็นสมการกำลังสอง ดังนั้น จำนวนทิศทางที่ใช้หาจุดที่ทำให้ค่าฟังค์ชั่นน้อยที่สุด จึงมากกว่า n ทิศทางเสียส่วนมาก

วิธีทิศทางสังยุคของพาวเวลล์มีหลักกว้าง ๆ ก็คือ เป็นการค้นหาจุดต่ำสุดในลักษณะ-อนุบรรพ (Sequential Search) โดยแบ่งออกเป็นรอบ ๆ (Cycle) ในแต่ละรอบจะทำการค้นหาจุดต่ำสุดไปตามทิศทางแต่ละทิศทางของ n ทิศทาง A-Conjugate ซึ่งเป็นการสลายปัญหาออกมาเป็น ปัญหาของการค้นหาจุดต่ำสุดที่มีมิติเท่ากับ ๑ โดยกำหนดขั้นตอนดังนี้

(ก) สร้างทิศทางสังยุค (Conjugate Direction) $\underline{E}_1, \underline{E}_2, \underline{E}_3, \ldots, \underline{E}_n$

โดยที่ $\underline{E}_1 = [1, 0, 0, \ldots, 0]$ ขนาด $1 \times n$

$\underline{E}_2 = [0, 1, 0, \ldots, 0]$ ขนาด $1 \times n$

$$\underline{E}_3 = [0,\ 0,\ 1, \ldots,\ 0] \quad \text{ขนาด} \quad 1 \times n$$

$$\vdots \qquad\qquad\qquad\qquad \vdots$$

$$\underline{E}_n = [0,\ 0,\ 0, \ldots,\ 0, 1] \quad \text{ขนาด} \quad 1 \times n$$

(ข) กำหนดค่าเริ่มต้นของตัวแปรทั้งหมด

$$\underline{X}_0 = [x_1,\ x_2,\ x_3, \ldots,\ x_n]$$

(ค) สำหรับ $S = 1,2,\ldots, n$ คำนวณค่า λ_s เพื่อให้ได้

$f(\underline{X}_{s-1} + \lambda_s \underline{E}_s)$ มีค่าน้อยที่สุดและให้ $\underline{X}_s = \underline{X}_{s-1} + \lambda_s \underline{E}_s$

(ง) หาเลขจำนวนเต็ม m ซึ่ง $1 \leqslant m \leqslant n$ เพื่อว่า

$f(\underline{X}_{m-1}) - f(\underline{X}_m)$ มีค่ามากที่สุด

และให้ $f(\underline{X}_{m-1}) - f(\underline{X}_m) = \Delta$

(จ) คำนวณค่า $f_3 = f(2\underline{X}_n - \underline{X}_0)$

และให้ $f_1 = f(\underline{X}_0)$

และ $f_2 = f(\underline{X}_n)$

(ฉ) ถ้า $f_3 > f_1$ หรือ $(f_1 - 2f_2 + f_3)(f_1 - f_2 - \Delta)^2 \geqslant \frac{1}{2} \Delta (f_1 - f_3)^2$ อย่างใดอย่างหนึ่ง

ใช้ทิศทางเก่าของ $\underline{E}_1, E_2, \ldots, \underline{E}_n$ สำหรับทำซ้ำในรอบต่อไป และใช้ $\underline{X}_n$ สำหรับเป็นจุดเริ่มต้น $(\underline{X}_0)$ ในรอบถัดไป

(ช) ถ้าเงื่อนไขในข้อ (ฉ) ไม่เป็นจริง

ให้ $\underline{E} = (\underline{X}_n - \underline{X}_0)$ และคำนวณค่า λ เพื่อว่า $f(\underline{X}_n + \lambda\underline{E})$ มีค่าน้อยที่สุดให้ใช้ $\underline{E}_1, \underline{E}_2, \ldots, \underline{E}_{m+1}, \underline{E}_{m+2}, \ldots, \underline{E}_n, \underline{E}$ เป็นทิศทางใหม่ และ $\underline{X}_n + \lambda\underline{E}$ เป็นจุดเริ่มต้น $(\underline{X}_0)$ สำหรับทำซ้ำในรอบถัดไป

ตามขั้นตอนดังกล่าวแล้วข้างต้น เราต้องหาค่าที่น้อยที่สุด (minimum) ของ $f(\underline{X} + \lambda\underline{E})$ เทียบกับ λ และให้ค่า λ ที่ทำให้ $f(\underline{X} + \underline{E})$ น้อยที่สุดนี้เป็น λ^* สำคัญ $\underline{X}$ และ $\underline{E}$ ต้องทราบค่า ดังนั้นการแก้ปัญหาต้องแก้โดย (one-dimensional minimization)

พาวเวลล์ (Powell) ได้แนะนำวิธีใช้ Quadratic โดยใช้ค่า m ค่า มาใช้ในการหาค่า λ^* ซึ่งเรียกวิธีการอย่างนี้ว่า Quadratic interporation

๒.๕.๑ ขั้นตอนของ Quadratic Interporation

ให้ค่าของส่วนประกอบ (component) ของ $\underline{E}$ เป็น r_i

นั่นคือ $\underline{E} = [r_1, r_2, \ldots, r_n]$

(ก) คำนวณค่า $M = \max |r_i|$ ให้ส่วนประกอบแต่ละส่วนของ r_i มีค่าปกติ (Normalize) โดยการหารแต่ละส่วนประกอบ $\underline{E}$ ด้วย M

ให้ $f(\underline{X} + \lambda E) = F(\lambda)$

(ข) ถ้า $F(1) > F(0)$ คำนวณค่า $F(\lambda)$ สำหรับ $\lambda = \frac{1}{2}, \frac{1}{4}, \ldots$

จนกว่า $F(\lambda) < F(0)$ แล้วจึงให้ $a = 0,\ b = \lambda,\ c = 2\lambda$ หลังจากนี้ข้ามไปที่ข้อ (ง)

(ค) คำนวณค่า $F(\lambda)$ สำหรับ $\lambda = 0,1,2,4,8,\ldots, a,b,c$ (ถ้าข้อ (ข) ไม่เป็นจริง) สิ้นสุดการคำนวณค่า λ เมื่อ $\lambda = c$ เมื่อค่าในตอนนั้น (Current value) ของ $F(\lambda)$ มีค่ามากกว่าค่าของ $F(\lambda)$ ที่ถูกคำนวณก่อนหน้านั้น

(ง) คำนวณค่า λ^* จาก

$$\lambda^* = \frac{1}{2}\,\frac{F_{(a)}(c^2-b^2) + F_{(b)}(a^2-c^2) + F_{(c)}(b^2-a^2)}{F_{(a)}(c-b) + F_{(b)}(a-c) + F_{(c)}(b-a)}$$

๒.๕.๒ การพิสูจน์ว่า λ^* เป็นค่าที่ทำให้ฟังก์ชั่นกำลังสองมีค่าต่ำที่สุด

จาก $f(\underline{X} + \lambda\underline{E})$ ต้องการหา λ ที่ทำให้ $f(\underline{X} + \lambda\underline{E})$ มีค่าต่ำที่สุด โดยที่ $\underline{X}$, $\underline{E}$ ทราบค่ามาก่อนแล้ว ตัวแปร λ จึงเป็นตัวแปรตัวเดียวที่ต้องการหาค่า

ถ้ากำหนดให้ $f(\underline{X} + \lambda\underline{E}) = F(\lambda)$

จะได้ว่า $F(\lambda) = A\lambda^2 + B\lambda + C$

ถ้า a, b, c ที่ได้จากขั้นตอน ๒.๕.๑

ให้ค่าใหม่เป็น $\lambda_1 = a$

$\lambda_2 = b$

$\lambda_3 = c$

โดยที่ $\lambda_1 < \lambda_2 < \lambda_3$ เสมอ

จาก $F(\lambda) = A\lambda^2 + B\lambda + C$ ค่า λ ที่ทำให้ $F(\lambda)$ มีค่าน้อยที่สุด

หาได้โดย $F'(\lambda) = 2A\lambda + B = 0$

$$\lambda_{min}\ (\lambda^*) = \frac{-B}{2A} \quad \text{เมื่อ} \quad A > 0$$

การหาค่า A, B

$$F(\lambda_1) - F(\lambda_2) = A\lambda_1^2 + B\lambda_1 + C - A\lambda_2^2 - B\lambda_2 - C$$

$$= A(\lambda_1^2 - \lambda_2^2) + B(\lambda_1 - \lambda_2)$$

$$F(\lambda_2) - F(\lambda_3) = A(\lambda_2^2 - \lambda_3^2) + B(\lambda_2 - \lambda_3)$$

ถ้าให้ $\lambda_{uv}^2 = \lambda_u^2 - \lambda_v^2$, $\lambda_{uv} = \lambda_u - \lambda_v$

$$\lambda_{12}^2 A + \lambda_{12} B = F(\lambda_1) - F(\lambda_2)$$

$$\lambda_{23}^2 A + \lambda_{23} B = F(\lambda_2) - F(\lambda_3)$$

หาค่า A, B โดย Gramer's rule

$$A = \frac{\begin{vmatrix} F(\lambda_1) - F(\lambda_2) & \lambda_{12} \\ F(\lambda_2) - F(\lambda_3) & \lambda_{23} \end{vmatrix}}{\begin{vmatrix} \lambda_{12}^2 & \lambda_{12} \\ \lambda_{23}^2 & \lambda_{23} \end{vmatrix}}$$

$$= \frac{[F(\lambda_1) - F(\lambda_2)]\lambda_{23} - [F(\lambda_2) - F(\lambda_3)]\lambda_{12}}{\lambda_{12}^2\ \lambda_{23} - \lambda_{23}^2\ \lambda_{12}}$$

$$B = \frac{\begin{vmatrix} \lambda_{12}^2 & F(\lambda_1) - F(\lambda_2) \\ \lambda_{23}^2 & F(\lambda_2) - F(\lambda_3) \end{vmatrix}}{\begin{vmatrix} \lambda_{12}^2 & \lambda_{12} \\ \lambda_{23}^2 & \lambda_{23} \end{vmatrix}}$$

$$= \frac{\lambda_{12}^2\ [F(\lambda_2) - F(\lambda_3)] - \lambda_{23}^2\ [F(\lambda_1) - F(\lambda_2)]}{\lambda_{12}^2\ \lambda_{23} - \lambda_{23}^2\ \lambda_{12}}$$

$$\lambda_{min} = \frac{-B}{2A}$$

$$= \frac{1}{2}\ \frac{\lambda_{23}^2\ [F(\lambda_1) - F(\lambda_2)] - \lambda_{12}^2\ [F(\lambda_2) - F(\lambda_3)]}{\lambda_{23}\ [F(\lambda_1)-F(\lambda_2)]-\lambda_{12}[F(\lambda_2)-F(\lambda_3)]}$$

$$= \frac{1}{2} \frac{\lambda_{23}^2 F(\lambda_1) - F(\lambda_2)(\lambda_{23}^2 + \lambda_{12}^2) + \lambda_{12}^2 F(\lambda_3)}{\lambda_{23} F(\lambda_1) - F(\lambda_2)(\lambda_{23} + \lambda_{12}) + \lambda_{12} F(\lambda_3)}$$

แต่ $\lambda_{23}^2 + \lambda_{12}^2 = \lambda_2^2 - \lambda_3^2 + \lambda_1^2 - \lambda_2^2$

$= \lambda_1^2 - \lambda_3^2$

$= \lambda_{13}^2$ หรือ $-\lambda_{31}^2$

$\lambda_{23} + \lambda_{12} = \lambda_2 - \lambda_3 + \lambda_1 - \lambda_2$

$= \lambda_1 - \lambda_3$

$= \lambda_{13}$ หรือ $-\lambda_{31}$

$$\therefore \lambda_{min} = \frac{1}{2} \frac{F(\lambda_1)\lambda_{23}^2 + F(\lambda_2)\lambda_{31}^2 + F(\lambda_3)\lambda_{12}^2}{F(\lambda_1)\lambda_{23} + F(\lambda_2)\lambda_{31} + F(\lambda_3)\lambda_{12}}$$

$$= \frac{1}{2} \frac{F(\lambda_1)\lambda_{32}^2 + F(\lambda_2)\lambda_{13}^2 + F(\lambda_3)\lambda_{21}^2}{F(\lambda_1)\lambda_{32} + F(\lambda_2)\lambda_{13} + F(\lambda_3)\lambda_{21}}$$

หรือ $$\lambda^* = \frac{1}{2} \frac{F(\lambda_1)[\lambda_3^2 - \lambda_2^2] + F(\lambda_2)[\lambda_1^2 - \lambda_3^2] + F(\lambda_3)[\lambda_2^2 - \lambda_1^2]}{F(\lambda_1)[\lambda_3 - \lambda_2] + F(\lambda_2)[\lambda_1 - \lambda_3] + F(\lambda_3)[\lambda_2 - \lambda_1]}$$

แทนค่า $a = \lambda_1$, $b = \lambda_2$, $c = \lambda_3$

จะได้ $$\lambda^* = \frac{1}{2} \frac{F(a)[c^2 - b^2] + F(b)[a^2 - c^2] + F(c)[b^2 - a^2]}{F(a)[c - b] + F(b)[a - c] + F(c)[b - a]}$$

๒.๖ ข้อผิดพลาดบางประการในวิธีทิศทางสังยุคของพาวเวลล์

ก่อนอื่นจะขอกล่าวถึงวิธีทิศทางสังยุคของพาวเวลล์บางส่วนอีกครั้งหนึ่งเพื่อจะชี้ให้เห็นถึงข้อผิดพลาดบางประการของพาวเวลล์

ให้ f เป็นฟังก์ชั่นที่มีค่าจริง (real-valued function) บน E^n(Euclidean n-space) ประมาณจุด P* ซึ่งเรียกว่าจุดที่เหมาะสมที่สุด (optimal) ซึ่งทำให้ f บน E^n มีค่าต่ำที่สุด

ขบวนการขั้นแรกของพาวเวลล์ ในขั้นต้นเลือก $\underline{E}_1, \underline{E}_2, \ldots, \underline{E}_n$ เป็นจุดเริ่มต้นของทิศทาง n ทิศทาง (n coordinate directions) และให้ $\underline{X}_0$ เป็นจุดเริ่มต้น และถ้าพิจารณาเฉพาะข้อ ค ถึง ฉ บางส่วนใหม่ในหัวข้อ ๒.๕

(ค) สำหรับ S = 1, 2,..., n คำนวณ λ_s เพื่อให้ค่าที่น้อยที่สุดของ $f(\underline{X}_{s-1} + \lambda_s \underline{E}_s)$ และให้ $\underline{X}_s = \underline{X}_{s-1} + \lambda_s \underline{E}_s$ และทำไปตามขั้นตอนจนถึงข้อ ฉ ซึ่งเป็นข้อที่มีเงื่อนไขจริงกับไม่จริง และถ้าข้อ ฉ เป็นความจริงจะต้องทำเงื่อนไขดังนี้

สำหรับ s = 1, 2,..., n-1 แทน $\underline{E}_s$ โดย $\underline{E}_{s+1}$ และแทน $\underline{E}_n$ โดย $(\underline{X}_n - \underline{X}_0)$

และเลือก λ เพื่อที่จะ $\min\limits_{\lambda} f(\underline{X}_n + \lambda(\underline{X}_n - \underline{X}_0))$ แล้วแทน $\underline{X}_0$ ด้วย $\underline{X}_n + \lambda(\underline{X}_n - \underline{X}_0)$ และเริ่มรอบใหม่ถัดไปตามข้อ ก. กล่าวคือ คำนวณ $\underline{X}_1, \ldots, \underline{X}_n$ โดย minimize ในทิศทาง $\underline{E}_1, \underline{E}_2, \ldots, \underline{E}_n$ หลังจากนั้นกำหนดทิศทางใหม่ โดยเลือกทิศทาง $\underline{E}_1$ ใหม่เป็น $\underline{E}_2$ เก่า หรือ $\underline{E}_s$ ใหม่เป็น $\underline{E}_{s+1}$ อันเก่า S = 1, 2,..., n-1 และ $\underline{E}_n$ ใหม่คือ $\underline{E}_n = \underline{X}_n - \underline{X}_0$

$\underline{X}_0$ ใหม่คือค่าที่ได้จากการ minimize จาก $\underline{X}_0$ เก่า โดยใช้ $\underline{E}_n$ อันใหม่ ขั้นตอนทั้งหมดจาก $\underline{X}_0$ อันแรกจนถึง $\underline{X}_0$ อันใหม่ถัดไปประกอบด้วย 1 iteration

ตัวอย่างแสดงข้อขัดแย้งของทฤษฎีของพาวเวลล์[๑] โดยกำหนดฟังก์ชั่นกำลังสองที่ strictly convex (strictly convex quadratic function) ที่มีจุดต่ำสุดเพียงจุดเดียว ที่ (x, y, z) = (0, 0, 0) ซึ่งมีฟังก์ชั่นดังนี้

$$f(x,y,z) = (x - y + z)^2 + (-x + y + z)^2 + (x + y - z)^2$$

---

[๑] Willard I. Zangwill, "Minimizing a function without calculating derivative", The Computer Journals, p. 293-296, Nov. 1967.

ถ้าเราใช้การแก้สมการโดยใช้อนุพันธ์ของฟังก์ชั่นจะได้ผลดังนี้

$$\frac{\partial f}{\partial x}(x, y, z) = 2(x - y + z) - 2(-x + y + z) + 2(x + y + z) = 0 \quad \ldots\ldots(2.3)$$

$$\frac{\partial f}{\partial y}(x, y, z) = -2(x - y + z) + 2(x - y + z) + 2(x + y - z) \quad \ldots\ldots(2.4)$$

$$\frac{\partial f}{\partial z}(x, y, z) = 2(x - y + z) + 2(-x + y + z) - 2(x + y - z) \quad \ldots\ldots(2.5)$$

ให้สมการ (1), (2), (3) เท่ากับศูนย์

จะได้

$$3x - y - z = 0 \quad \ldots\ldots(2.6)$$

$$-x + 3y - z = 0 \quad \ldots\ldots(2.7)$$

$$-x - y + 3z = 0 \quad \ldots\ldots(2.8)$$

$$x = 0$$

$$y = 0$$

$$z = 0$$

นั่นคือ $\underline{X}^* = (0, 0, 0)$ เป็นจุดที่ทำให้ $f(x, y, z)$ มีค่าต่ำที่สุดเพียงจุดเดียว ถ้าใช้วิธีการของพาวเวลล์ซึ่งไม่ต้องอาศัยอนุพันธ์ของฟังก์ชั่น

ให้เลือก (x, y, z) เป็นจุดเริ่มต้นที่ $(\frac{1}{2}, 1, \frac{1}{2})$

ทิศทาง	จุดที่ฟังก์ชั่นมีค่าต่ำสุด		
จุดเริ่มต้น	$(\frac{1}{2}, 1, \frac{1}{2})$	=	$\underline{X}_0$
ทิศทางตามแกน X	$(\frac{1}{2}, 1, \frac{1}{2})$	=	$\underline{X}_1$
ทิศทางตามแกน Y	$(\frac{1}{2}, \frac{1}{3}, \frac{1}{2})$	=	$\underline{X}_2$
ทิศทางตามแกน Z	$(\frac{1}{2}, \frac{1}{3}, \frac{5}{18})$	=	$\underline{X}_3$

ทิศทางใหม่คือ $\underline{E}_3 = \underline{X}_3 - \underline{X}_0 = (\frac{1}{2}, \frac{1}{3}, \frac{5}{18}) - (\frac{1}{2}, 1, \frac{1}{2})$

$= (0, \frac{-2}{3}, \frac{-2}{9})$

เซทใหม่ของทิศทางทั้งหมดคือ

$\underline{E}_1 = (0, 1, 0)$

$\underline{E}_2 = (0, 0, 2)$

$\underline{E}_3 = (0, \frac{-2}{3}, \frac{-2}{9})$

จะสังเกตเห็นว่าค่าของตัวแปร X ของทิศทางทั้ง ๓ คือ 0

จากผลอันนี้ค่าของตัวแปร X จะเปลี่ยนแปลงไม่ได้อีก จากการคำนวณต่อจะได้ว่า จุดที่ทำให้ฟังก์ชั่น f(x, y, z) มีค่าต่ำสุด โดยใช้วิธีการนี้ คือ $\underline{X}_3 = (\frac{1}{2}, \frac{1}{3}, \frac{5}{18})$ ตามทิศทางของแกน z แต่ในความเป็นจริงแล้ว จุดที่ทำให้ฟังก์ชั่น f(x, y, z) มีค่าต่ำสุด คือ $\underline{X}^* = (0, 0, 0)$ ซึ่งวิธีการของพาวเวลล์ไม่อาจที่จะไปถึงจุดนี้ได้ ถึงแม้ว่าจำนวนรอบของการ iterate จะเข้าใกล้อนันต์ก็ตาม