บทที่ 6
Chemometric

ในปัจจุบัน การวัดแสง NIR เป็นเทคนิคการวิเคราะห์ ซึ่งส่วนใหญ่นำมาประยุกต์ใช้กับ Chemometric Chemometric คือการใช้คณิตศาสตร์และเทคนิคทางสถิติสำหรับการแปลผลข้อมูลจากข้อมูลที่ทำการวิเคราะห์ กรณีนี้คือสเปกตรัม NIR [12] ทั้ง Chemometric และเทคโนโลยี NIR เอื้อประโยชน์ซึ่งกันและกัน โครงร่างในการสร้างแบบจำลองข้อมูลทางเคมีคือแนวความคิดที่ว่าสัญญาณที่วัดได้แสดงถึงปรากฏการณ์ทางเคมี (ข้อมูล) สามารถแบ่งได้เป็นสองส่วน คือ ระบบ (เกิดจากจุดเด่นทางเคมี) และไม่ใช่ระบบ (สัญญาณรบกวน) จากแนวความคิดนี้ทำให้สร้างแบบจำลองได้ง่ายขึ้น โดยมีพื้นฐานอยู่บนข้อมูลที่วัดได้ ซึ่งได้แก่

ข้อมูล = แบบจำลองเคมี + สัญญาณรบกวน

แบบจำลองส่วนใหญ่ที่ถูกนำมาใช้สำหรับการสอบเทียบสเปกตรัมข้อมูลคือ แบบจำลองกำลังสองน้อยที่สุด (Least Squares) ใช้อธิบายตัวแปรกับความยาวคลื่นที่เลือกมา โดยการทำให้กำลังสองของค่าความผิดพลาดน้อยที่สุด ตามปกติแบบจำลองกำลังสองน้อยที่สุดจะไม่เป็นไปตามอุดมคติ เนื่องจากการตอบสนองไม่เป็นเชิงเส้น การประมาณค่า Beer-Lambert สัญญาณรบกวนที่เกิดขึ้น ความสัมพันธ์ระหว่างตัวแปร โดยสมการทำนายนั้นสามารถเขียนได้ตามสมการที่ (6.1) [12, 13]

$$C = K_1 + K_2 \log\left(\frac{1}{R_{\lambda 1}}\right) + K_3 \log\left(\frac{1}{R_{\lambda 2}}\right) + \ldots \qquad (6.1)$$

โดยที่ C คือค่าความเข้มข้นขององค์ประกอบที่สนใจ

$R_{\lambda 1}, R_{\lambda 2}, R_{\lambda 3}, \ldots$ คือค่าการดูดกลืนที่แต่ละความยาวคลื่น ซึ่งสอดคล้องกับการประมาณค่าเชิงเส้นตามกฎ Beer-Lambert

$K_1, K_2, K_3, \ldots$ คือค่าคงตัว

วิธีการเพื่อจะอธิบายการแยกข้อมูลเป็นองค์ประกอบที่ตั้งฉากกัน (Orthogonal component) โดยคำว่า Orthogonal หมายถึงไม่มีความสัมพันธ์กันโดยสิ้นเชิง ในกรณีของการสอบเทียบ ปัญหาของการทำ $(XX')^{-1}$ ถูกเฉลยโดยการแปลง X ให้อยู่ในรูปของส่วน orthogonal ขั้นตอนนี้ทำการแปลงให้เป็น XX' หรือทำการรวมกันของ covariance ของเมตริกซ์ หรือการสร้างเมตริกซ์ทแยงมุม ซึ่งทำให้สามารถทำการอินเวอร์สเพื่อหาคำตอบได้ วิธีการโปรเจ็กชันแบบตั้งฉาก (Orthogonal projection) มีประโยชน์ในการแปลผลข้อมูล วิธีการนี้ยังเหมาะในการตรวจสอบในแบบรูปภาพ เพราะข้อมูลถูกฉายลงบนไม่กี่แกนที่ไม่มีความสัมพันธ์กันและมีขอบเขต

ซึ่งง่ายในการแปลผลข้อมูลมากกว่าข้อมูลต้นแบบ ความเป็นไปได้ในการแสดงจุดเด่นของข้อมูลเป็นข้อได้เปรียบหนึ่งของเทคนิคตัวแปรพหุพจน์ (Multivariate) [12]

โดยทั่วไป การสร้างแบบจำลององค์ประกอบเด่นของสเปกตรัมกับสิ่งที่สนใจ ด้วยการสร้าง transfer function f(i) ระหว่างข้อมูลที่สังเกตได้และที่วัดได้ แบบจำลองที่เป็นไปได้คือ

$$Y = f(X) + E \tag{6.2}$$

หรือ

$$X = f(Y) + E' \tag{6.3}$$

โดยวิธีการพื้นฐานทั่วไปคือ การทำให้ผลรวมกำลังสองของเศษเหลือ (E, E') เกิดน้อยที่สุด ส่วนใหญ่ การสร้างแบบจำลองข้อมูลสเปกตรัม ยังคงเป็นแบบจำลองกำลังสองน้อยที่สุดเชิงเส้นตัวแปรเดียว โดยเลือกตัวแปรเดียวหรือที่ความยาวคลื่นหนึ่งมาสร้างแบบจำลองเพื่ออธิบายสิ่งที่สนใจ

6.1 การแปลผลข้อมูล

การแปลผลข้อมูลทำตามแนวความคิดของ Chemometric โดยใช้แบบจำลองกำลังสองน้อยที่สุด ซึ่งสามารถแยกได้เป็น 2 วิธีคือ Classical Least Squares และ Inverse Least Squares โดยตามแนวความคิดของ Classical Least Squares คือ การหาสเปกตรัมจากความเข้มข้นองค์ประกอบของตัวอย่างที่ทำการวิเคราะห์ หรือสร้างแบบจำลองอธิบายสเปกตรัมในรูปตัวแปรความเข้มข้นขององค์ประกอบ ส่วน Inverse Least Squares เป็นการสร้างแบบจำลองอธิบายองค์ประกอบในรูปของสเปกตรัมการดูดกลืน ซึ่งจะนำมาใช้ในการแปลผลข้อมูลความชื้นจากสเปกตรัม

6.1.1 แบบจำลองกำลังสองน้อยที่สุด [12]

คำตอบของแบบจำลองกำลังสองน้อยที่สุดมีความน่าสนใจ เนื่องจากมีการใช้อย่างกว้างขวางในการสร้างแบบจำลองข้อมูลทางคณิตศาสตร์ แม้ว่ากำลังสองน้อยที่สุดจะถูกจัดอยู่ใน Chemometric แต่ไม่ใช่วิธี orthogonal compression ในการวัดและวิเคราะห์สเปกตรัมนั้น มีอยู่สองวิธีในการสอบเทียบข้อมูลแบบหลายตัวแปรโดยตรง คือ Classical Least Squares (CLS) และ Inverse Least Squares (ILS) ทั้งสองวิธีมีสมการทางคณิตศาสตร์ร่วมกัน แต่มีหลักการที่ต่างกันคือ การสร้างแบบจำลองเชิงเส้นเพื่อทำนาย "pure spectrum" หรือการหาค่าสัมประสิทธิ์การถดถอย ซึ่งสามารถทำนายสิ่งที่สนใจได้จากสเปกตรัม ต้นกำเนิดของทฤษฎีกำลังสองน้อยที่สุดสามารถย้อนกลับไปยัง Gauss และ Legendre (ยุค 1700)

6.1.1.1 Classical Least Squares (CLS)

หลักการของ Classical Least Squares (CLS) กล่าวไว้ว่า ตัวแปร X (สเปกตรัม) เป็นฟังก์ชันของสิ่งที่เราสนใจ Y (ค่าความเข้มข้นของสิ่งที่สนใจ) วิธีการคือเพื่อลดเศษเหลือใน X ให้น้อยที่สุด

$$\min([X - YK])^2$$

แบบจำลองของ CLS คือ

$$X = YK + E \tag{6.4}$$

โดยคำตอบจากกำลังสองน้อยที่สุดสามารถหาได้เป็น

$$\hat{K} = (Y'Y)^{-1}Y'X \tag{6.5}$$

โดย $\hat{K}$ ประกอบด้วยส่วนที่เกี่ยวข้องกับสเปกตรัมหรือส่วนประกอบเชิงเส้นของ X เพื่อทำการวิเคราะห์ Y ซึ่งหมายความว่า CLS เป็นแบบหนึ่งของแบบจำลองแฟคเตอร์ (Factor model) เนื่องจากเมตริกซ์ X ถูกแยกตัวประกอบเป็น $\hat{K}$ ได้โดย Y

การทำนายในแบบจำลอง CLS สามารถทำได้โดย

$$c = (KK')^{-1}Kx \tag{6.6}$$

โดยที่ x คือสเปกตรัมของตัวอย่างที่ไม่รู้ค่าที่เราสนใจ K หาได้จากสมการ และ c คือความเข้มข้นของสิ่งที่เราสนใจ

ข้อด้อยที่เห็นได้ชัดเจนคือ ต้องรู้ค่าของสเปกตรัมและสิ่งรบกวนในการใช้วิธีนี้ โดยสมการของ CLS เหมาะในการประมาณค่านอกช่วง CLS นั้นเป็นที่รู้จักกันในชื่อ K-matrix หรือ Reverse- Direct- และ Total calibration

6.1.1.2 Inverse Least Squares (ILS)

ตัวอย่างของ Inverse Least Squares ที่ใช้กันเช่น สมการถดถอยหลายตัวแปร (Multiple Linear Regression) Partial Least Squares (PLS) และอื่นๆ

6.1.1.2.1 สมการถดถอยหลายตัวแปร (MLR)

วัตถุประสงค์ของฟังก์ชันเพื่อจะลดเศษเหลือในตัวแปร Y ให้น้อยที่สุด โดยที่ X คือเมตริกซ์ของค่าการดูดกลืน Y คือเมตริกซ์ความเข้มข้นขององค์ประกอบ ในที่นี้คือเปอร์เซ็นต์ความชื้นในแป้ง

$$\min([Y - X\beta]^2)$$

แบบจำลอง คือ

$$Y = X\beta + E \tag{6.7}$$

โดยที่ E คือเมตริกซ์ของค่าความผิดพลาด ซึ่งสามารถเขียนสมการใหม่ได้เป็น

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{1,2} & X_{1,3} & \cdots & X_{1,j} \\ 1 & X_{2,2} & X_{2,3} & \cdots & X_{2,j} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n,2} & X_{n,3} & \cdots & X_{n,j} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_j \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \tag{6.8}$$

โดยเราสามารถประมาณค่า β ได้ดังนี้

$$\hat{\beta} = (X^T X)^{-1} X^T Y \tag{6.9}$$

เมื่อนำเอา β ซึ่งคือเวกเตอร์ถดถอยมาคูณกับสเปกตรัม ทำให้เราสามารถทำนายเป้าหมายที่สนใจ $\hat{y}$ โดยเป็นไปตามสมการ

$$\hat{y} = \beta(1)x(1) + \beta(2)x(2) + ... + \beta(p)x(p) \tag{6.10}$$

หรือ

$$\hat{y} = x\beta \tag{6.11}$$

โดยที่ $\beta(1)$ เป็นสัมประสิทธิ์ตัวแรกของเวกเตอร์ β และ $x(1)$ เป็นข้อมูลสเปกตรัมความยาวคลื่นแรกของตัวอย่างที่ไม่รู้ค่าขององค์ประกอบที่เราสนใจ ILS สามารถนำมาใช้กับข้อมูลที่มีสิ่งเจือปนอื่นในตัวอย่างที่เราสนใจได้ แต่สิ่งที่เป็นจุดเด่นของข้อมูลที่จะถูกทำนายนั้น ต้องนำมาใช้ในการสร้างแบบจำลองสอบเทียบด้วย วิธี ILS นั้น รู้จักกันเป็น P-matrix MLR Forward- Indirect- และ Partial calibration

6.1.1.2.2 สหสัมพันธ์ร่วม (Co-correlation) และการอินเวอร์ส

ปัญหาหลักในการหาสัมประสิทธิ์ถดถอยเพื่อทำการวิเคราะห์ปัญหาในแบบจำลองคือความเกี่ยวข้องกับธรรมชาติในการอินเวอร์สเมตริกซ์ของ variance-covariance การแก้เมตริกซ์ $(XX')^{-1}$ เป็นกุญแจสำคัญสำหรับการสอบเทียบ ถ้าเมตริกซ์ X มีสหสัมพันธ์ภายในและภายนอก (Intercorrelation, Intracorrelation) แล้ว ข้อมูลจะมีความเป็นเชิงเส้นร่วมกันสูงหรือการเป็นส่วนประกอบเชิงเส้นของกันและกัน ทำให้เกิดปัญหาในการอินเวอร์สเมตริกซ์คือ อาจไม่สามารถทำการอินเวอร์สได้ ส่วนการมีสหสัมพันธ์ร่วม (Co-correlation) นั้นทำให้เมตริกซ์มี rank เพิ่มขึ้นและเกิดตัวเลขที่ค่ามากในแนวทแยงมุม และเกือบเป็นเมตริกซ์ทแยงมุม ทำให้ค่าที่ได้จากการอินเวอร์สเกิดความไม่เสถียร คือค่าที่ได้จะมีความไวต่อสัญญาณรบกวน

แต่จากมุมมองของนักเคมีแล้ว ปัญหาจากการมีความสัมพันธ์เชิงเส้น (Co-linearity) นั้นไม่ใช่ปัญหาเพียงแค่มีอย่างน้อยหนึ่งตัวแปรที่สามารถอธิบายองค์ประกอบที่สนใจได้ ส่วนการมีสหสัมพันธ์ร่วม (Co-correlation) แสดงว่า มีข้อมูลมากพอเพื่อใช้ในการอธิบายโดยที่ต้องใช้

กระบวนการทางคณิตศาสตร์ให้ถูกต้องในการแก้ปัญหา โดยประมาณค่าที่ให้กำลังสองของค่าความผิดพลาดน้อยที่สุด โดยค่าส่วนใหญ่ที่นำมาพิจารณาใช้ในการสร้างสมการทำนายได้แก่

- Sum of Squares

$$SSR = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 \tag{6.12}$$

$$SST = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \tag{6.13}$$

โดย SSR คือ Sum of Squares of Regression เป็นค่าความแปรผันที่เกิดจากแบบจำลอง SST คือ Sum of Squares of Total เป็นค่าความแปรผันของข้อมูลกับค่าเฉลี่ยของตัวเอง $\bar{Y}$ คือค่าเฉลี่ยของค่าที่ได้จากการทดลองตามวิธีมาตรฐาน $\hat{Y}$ คือค่าที่ทำนายได้ และ Y คือค่าที่ได้จากการทดลองตามวิธีมาตรฐาน เราสามารถนำค่านี้มาใช้ในการหาค่า Coefficient of Determination (R^2) หรือค่าความสัมพันธ์ระหว่างค่า Y ที่ได้จากสมการและค่าที่ได้จากวิธีมาตรฐาน โดยให้ค่าใกล้เคียงกับ 1 ให้มากที่สุด โดยสามารถหาได้จาก

$$R^2 = SSR / SST \tag{6.14}$$

ค่าที่ใช้ในการประเมินความสามารถในการทำนายของสมการ ยกตัวอย่างเช่น Standard error of calibration (SEC) ซึ่งใช้ในการประเมินความสามารถในการทำนายของสมการเมื่อใช้กับกลุ่มที่ทดลอง และ Standard error of prediction (SEP) ซึ่งใช้ในการประเมินความสามารถของสมการ เมื่อเทียบกับตัวอย่างนอกกลุ่มทดลอง โดยทั้ง SEC และ SEP ควรมีค่าน้อยที่สุด

- Standard error of calibration

$$SEC = \sqrt{\sum_{i=1}^{n} \frac{(\hat{Y}_i - Y_i)^2}{n - f - 1}}$$

โดยค่า $\hat{Y}$ คือค่าที่คำนวณได้จากสมการทำนาย n คือจำนวนตัวอย่าง f คือจำนวนตัวแปรหรือจำนวนค่าการดูดกลืนที่ความยาวคลื่นต่างๆ หรือจำนวนแฟคเตอร์ในกรณี Principal Component Analysis (PCA) หรือ Partial Least Squares (PLS)

- Standard error of prediction

$$SEP = \sqrt{\sum_{i=1}^{n} \frac{(\hat{Y}_i - Y_i - bias)^2}{n - 1}}$$

ค่า *bias* นั้นคือค่าเฉลี่ยของค่าความผิดพลาด

เนื่องจากการมีความสัมพันธ์กันระหว่างสิ่งที่สนใจ Y และข้อมูลสเปกตรัม และความสัมพันธ์ระหว่างตัวแปรด้วยกันเป็นไปตามทฤษฎีของ NIR เนื่องจากจุดยอดของสเปกตรัม

นั้นกว้างและซ้อนทับกัน จึงทำให้เกิดปัญหากับทฤษฎีของ ILS ในการสอบเทียบข้อมูลกับสิ่งที่สนใจ PLS เป็นวิธีการเพื่อทำให้ค่าความแปรปวนร่วมกำลังสองระหว่าง X และ Y มีค่าสูงสุด

$$\max_{w}\left[\operatorname{cov}(t,y)\,|\,t=Xw \wedge \|w\|=1\right]$$

โดย PLS จะทำให้ X และ Y อยู่ในรูปขององค์ประกอบที่ตั้งฉาก (Orthogonal component) โดยวิธีนี้ทำให้

$$\|\beta\|_{2,PLS} < \|\beta\|_{2,ILS} \text{ ถ้า a<min(n,p)} \qquad (6.15)$$

จากความสัมพันธ์ จะเห็นได้ว่า สัมประสิทธิ์ถดถอยของ PLS มีความเสถียรกว่า ILS โดยที่ความสัมพันธ์นี้จะเป็นจริงก็ต่อเมื่อมีจำนวนน้อยกว่าจำนวนตัวแปร PLS นั้น มีอยู่สองโครงสร้างแบบจำลองแรก PLS1 สำหรับอธิบายสิ่งที่สนใจเพียงหนึ่งตัว และอีกแบบจำลอง PLS2 สำหรับหลายตัวพร้อมกัน

6.1.1.2.3 Partial Least Squares(PLS) [8, 12]

PLS ใช้หลักการพื้นฐานของวิธีแฟคเตอร์ ทำการแยกข้อมูลออกเป็นองค์ประกอบที่เป็นอิสระจากกัน โดยจะต่างจากวิธีของ MLR ซึ่งอาจมีปัญหาในการอินเวอร์สเพื่อแก้สมการ และสมการที่ได้อาจจะไม่มีเสถียรภาพ PLS นั้นจะทำการประมาณค่าในส่วนของทั้ง X และ Y และสร้างความสัมพันธ์ไปพร้อมกันโดยใช้แฟคเตอร์ในการหาองค์ประกอบหลัก (Principal component) ในการแยกองค์ประกอบหลักเริ่มต้นจาก

$$w = X^T y(y^T y)^{-1} \qquad (6.16)$$

โดยที่ $X = yw^T + E$ เมื่อ w คือ เวกเตอร์ถ่วงน้ำหนัก (Weights vector) E คือเศษเหลือ หลังจากนั้นทำการนอร์มอไลซ์ w และหาสกอร์แฟคเตอร์ (Score factor, t)

$$t = Xw \qquad (6.17)$$

ซึ่ง $ww^T = I$ และ $X = tw^T + E$ และหลังจากนั้นทำการสอบเทียบในส่วนของบล็อก Y

$$q = y^T t(t^T t)^{-1} \qquad (6.18)$$

เมื่อ $y = tq^T + e$ โดย q คือเวกเตอร์ใช้เป็น y-loading หลังจากนี้ทำการสอบเทียบสเปกตรัมกับส่วนของสกอร์แฟคเตอร์

$$p = X^T t(t^T t)^{-1} \qquad (6.19)$$

โดยที่ $X = tp^T + E$ ซึ่ง p คือตัวโหลดดิ้ง (Loading) หรือองค์ประกอบบริสุทธิ์ (Pure component) หลังจากนั้น ทำการลบองค์ประกอบออกจากข้อมูลโดยสมการที่ (6.20), (6.21)

$$X = X - tp^T \qquad (6.20)$$

$$y = y - tq^T \qquad (6.21)$$

โดยทำต่อเนื่องไปเรื่อยๆจนกว่าจะได้จำนวนองค์ประกอบตามต้องการ ค่าสัมประสิทธ์ของสมการเชิงเส้นนั้นสามารถหาได้จากสมการที่ (6.22) โดยอาศัยหลักการของ Inverse Least Squares

$$\hat{\beta} = W(P^T W)^{-1} Q^T \qquad (6.22)$$

สมการที่ได้นั้นสามารถพิจารณาจาก PRESS (Predicted Residual Sum of Squares) RMSEC หรือ SEC RMSEP หรือ SEP และค่า Coefficient of Determination โดยค่า RMSEC RMSEP และ PRESS ควรมีค่าน้อยๆ

- PRESS

$$\sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2$$

- Root mean square error of calibration (RMSEC)

$$\sqrt{\sum_{i=1}^{n} \frac{(\hat{Y}_i - Y_i)^2}{n-f-1}}$$

- Root mean square error of prediction (RMSEP)

$$\sqrt{\sum_{i=1}^{n} \frac{(\hat{Y}_i - Y_i)^2}{n}}$$

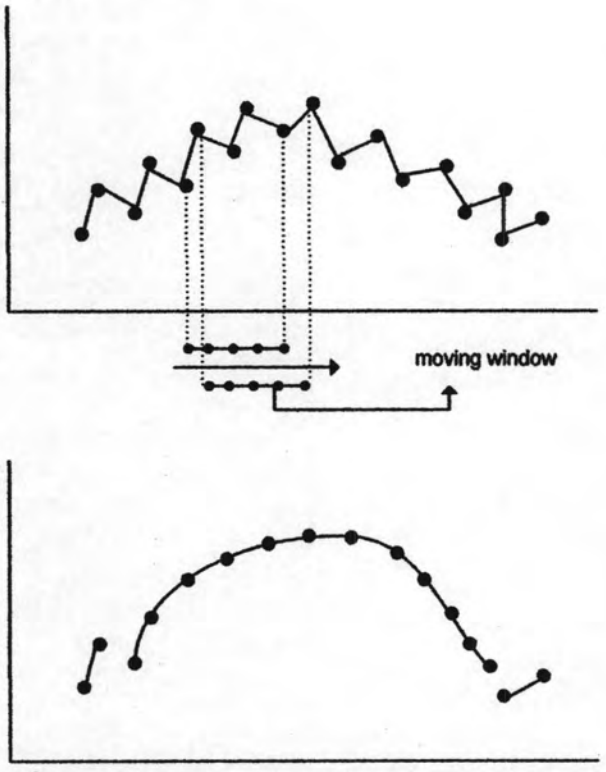
6.2 Digital smoothing และ Filtering methods [14]

โดยทั่วไป การเฉลี่ยถูกใช้อย่างกว้างขวางเพื่อปรับปรุง signal-to-noise ratio ของสัญญาณ ผลของสัญญาณรบกวนสามารถลดได้ เนื่องจากการกระจายแบบปกติของข้อมูล ในการเฉลี่ยกลุ่มข้อมูล x_i ค่าเฉลี่ย $\bar{x}$ สามารถคำนวณได้โดย

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad (6.23)$$

จะสังเกตได้ว่า ค่าความแปรปรวนของ $\bar{x}$ มีค่าเป็น $1/\sqrt{n}$ ของข้อมูลเดิม ดังนั้น การเฉลี่ยทำให้ SNR เพิ่มขึ้น เนื่องจากสเปกตรัมที่วัดได้เกิดจากการเฉลี่ยของการวัดหลายครั้ง

6.2.1 Moving-Window Average Smoothing Method



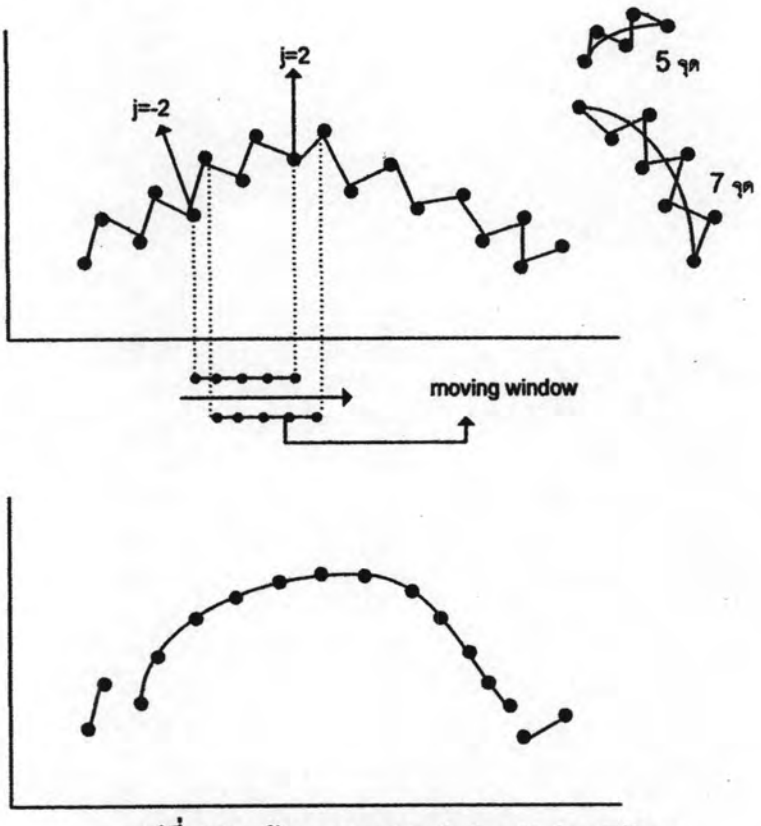
รูปที่ 6.1 หลักการของ moving-window average filter

วิธีการ moving-window average เป็นวิธีการดั้งเดิมและง่ายที่สุดในการทำให้เรียบ และช่วยปรับปรุงค่า SNR ด้วย หลักการของวิธีนี้แสดงในรูป 6.1 สมมุติเวกเตอร์ข้อมูลดิบ $x = [x_1, x_2, \ldots, x_{n-1}, x_n]$ ในทางปฏิบัติขนาดของหน้าต่าง (window size, $2m+1$) ต้องถูกกำหนดก่อนทำการคำนวณ เช่น ในการกรองข้อมูลโดยมีขนาดหน้าต่างเท่ากับ 5 ($m=2$) เริ่มต้นห้าข้อมูลแรกจะถูกใช้ในการหาข้อมูล x_3^* โดยที่ $i=3$ และ $m=2$ จากสมการ

$$x_i^* = \frac{1}{2m+1}\sum_{j=-m}^{m} x_{i+j} \qquad (6.24)$$

โดยที่ x_i^* เป็นค่าที่เกิดจากการกรอง x_{i+j} เป็นข้อมูลดิบ i และ j คือค่าคงตัว จะเห็นได้ว่า ข้อมูลสองค่าแรกไม่สามารถทำการกรองได้ และหลังจากทำ x_3^* ขั้นตอนต่อไปคือ การหา x_4^* และทำต่อไปจนครบข้อมูลทุกค่า

6.2.2 Savitsky-Golay Filter



รูปที่ 6.2 หลักการของ Savitsky-Golay Filter

Savitsky-Golay Filter เป็นการกรองที่อาศัยวิธีการถดถอยพหุพจน์ แทนการใช้เทคนิคการเฉลี่ยอย่างง่าย Savitsky-Golay Filter ใช้ความสามารถของ regression fit ในการปรับปรุงผลของการกรองตามรูป จากรูปที่ 6.2 จะเห็นได้ว่า สามารถทำได้ดีกว่าวิธี moving-window average เพราะว่าได้ใช้ความสามารถในการทำให้เข้ากับข้อมูลของการถดถอยพหุนาม อย่างไรก็ตาม สูตรการคำนวณของ Savitsky-Golay Filter คล้ายกับของ moving-window average แต่สิ่งหลักที่ต่างกันคือ การให้น้ำหนักของข้อมูลที่นำมาคำนวณ

$$x_i^* = \frac{1}{2m+1}\sum_{j=-m}^{m} w_j x_{i+j} \qquad (6.25)$$

สำหรับ Savitsky-Golay Filter ค่าถ่วงน้ำหนัก w_j สามารถหาได้โดยการใช้กำลังสองน้อยที่สุดซึ่งสามารถหาได้ตามสมการ

$$x_j^i = a_0 + a_1 j + a_2 j^2 + \ldots + a_k j^k \qquad (6.26)$$
$$(j = -m, -m+1, \ldots, m-1, m; i = 1, \ldots, n$$

ตัวอย่างเช่น ถ้าขนาดหน้าต่างเท่ากับ 5 แล้วสมการจะกลายเป็น

$$x_j^i = a_0 + a_1 j + a_2 j^2 \quad (j = -2, -1, 0, 1, 2; i = 1, \ldots, n) \qquad (6.27)$$

โดย i แสดงถึงข้อมูลดั้งเดิม j แสดงถึงตัวแปรของขนาดหน้าต่าง เนื่องจากค่า x_j^i และ j หาค่าได้จากเทคนิคของกำลังสองน้อยที่สุด สามารถนำมาใช้เพื่อหาค่า a โดยการแทนค่าเข้าไปที่สมการ ก็จะได้กลุ่มของสมการเชิงเส้น

$$\begin{cases} x_{-2}^{i-2} = a_0 + a_1(-2) + a_2(-2)^2 \\ x_{-1}^{i-1} = a_0 + a_1(-1) + a_2(-1)^2 \\ x_0^{i} = a_0 + a_1(0) + a_2(0)^2 \\ x_1^{i+1} = a_0 + a_1(1) + a_2(1)^2 \\ x_2^{i+2} = a_0 + a_1(2) + a_2(2)^2 \end{cases} \tag{6.28}$$

หรือ

$$\begin{cases} x_{-2}^{i-2} = a_0 - 2a_1 + 4a_2 \\ x_{-1}^{i-1} = a_0 - a_1 + 4a_2 \\ x_0^{i} = a_0 \\ x_1^{i-1} = a_0 + a_1 + 4a_2 \\ x_2^{i+2} = a_0 + 2a_1 + 4a_2 \end{cases} = \begin{bmatrix} 1 & -1 & 2 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} \tag{6.29}$$

ซึ่งสามารถเขียนได้ในรูป

$$\begin{bmatrix} x_{-2}^{i-2} \\ x_{-1}^{i-1} \\ x_0^{i} \\ x_1^{i-1} \\ x_2^{i+2} \end{bmatrix} = \begin{bmatrix} 1 & -1 & 2 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} \tag{6.30}$$

หรือ

$$x = Ma \tag{6.31}$$

$$x = \begin{bmatrix} x_{-2}^{i-2} \\ x_{-1}^{i-1} \\ x_0^{i} \\ x_1^{i-1} \\ x_2^{i+2} \end{bmatrix}; M = \begin{bmatrix} 1 & -1 & 2 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}; a = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}$$

และสามารถเฉลยได้เป็น

$$a = (M^t M)^{-1} M^t x \tag{6.32}$$

โดยการแทนลงในสมการ เราสามารถทำนายค่าได้ $\hat{x}$

$$\hat{x} = M(M^t M)^{-1} M^t x \tag{6.33}$$

หรือในอีกรูปหนึ่งเป็น

$$\begin{cases} \hat{x}_{-2}^{i-2} = \frac{1}{35}(31x_{-2}^{i-2} + 9x_{-1}^{i-1} - 3x_0^i - 5x_1^{i+1} + 3x_2^{i+2}) \\ \hat{x}_{-1}^{i-1} = \frac{1}{35}(9x_{-2}^{i-2} + 13x_{-1}^{i-1} + 12x_0^i + 6x_1^{i+1} - 5x_2^{i+2}) \\ \hat{x}_0^i = \frac{1}{35}(-3x_{-2}^{i-2} + 12x_{-1}^{i-1} + 17x_0^i + 12x_1^{i+1} - 3x_2^{i+2}) \\ \hat{x}_1^{i+1} = \frac{1}{35}(-5x_{-2}^{i-2} + 6x_{-1}^{i-1} + 12x_0^i + 13x_1^{i+1} + 9x_2^{i+2}) \\ \hat{x}_2^{i+2} = \frac{1}{35}(3x_{-2}^{i-2} - 5x_{-1}^{i-1} - 3x_0^i + 9x_1^{i+1} + 31x_2^{i+2}) \end{cases} \quad (6.34)$$

จะเห็นได้ว่า Savitsky-Golay Filter ใช้เพียงจุดกลางในการทำให้เรียบ ดังนั้น ค่าถ่วงน้ำหนัก [w_{-2}, w_{-1}, w_0, w_1, w_2] สามารถหาได้จากสัมประสิทธิ์ของ x_0^i จากสมการข้างบน ค่าก็คือ $\left[-\frac{3}{35}\ \frac{12}{35}\ \frac{17}{35}\ \frac{12}{35}\ -\frac{3}{35}\right]$ สำหรับขนาดหน้าต่างเท่ากับ 5 และพหุพจน์อันดับ 2 โดยที่ขนาดของหน้าต่างและพหุพจน์อันดับอื่นก็สามารถหาได้เช่นเดียวกัน

6.2.3 Numerical Differentiation

Numerical Differentiation เป็นเครื่องมือหนึ่งที่ใช้ในการวิเคราะห์สัญญาณทางเคมี เนื่องจากส่วนที่มีประโยชน์ในการแปลผลข้อมูลมักถูกซ่อนอยู่ในสัญญาณที่ทำการวัด สัญญาณที่วัดได้มักแสดงเป็นรูปร่างเดียว ที่ไม่แสดงส่วนโค้งหรือจุดยอดของสิ่งที่เป็นจุดเด่น ในการหาอนุพันธ์ทำให้สามารถหาจุดยอดหรือส่วนโค้งที่เป็นจุดเด่นที่ต้องการได้

6.2.3.1 Simple Difference Method

วิธี direct-difference เป็นวิธีการหาอนุพันธ์เชิงตัวเลขของสัญญาณที่ง่ายที่สุด สำหรับข้อมูลสเปกตรัม x_i (i=1,...,n) และความยาวคลื่นที่ทำการวัดสเปกตรัม w_i (i=1,...,n) สมการ direct different method สามารถเขียนได้เป็น

$$y_i = \frac{x_{i+1} - x_i}{w_{i+1} - w_i} \quad (6.35)$$

โดย y_i (i=1,...,n-1) คือกลุ่มข้อมูลสเปกตรัมของ x_i ที่ถูกทำการหาอนุพันธ์

direct different method เป็นวิธีที่ง่าย แต่ข้อด้อยคือ จำนวนจุดข้อมูลสเปกตรัมที่ทำอนุพันธ์นั้นน้อยกว่าข้อมูลต้นฉบับหนึ่งจุด นั่นคือ ทำให้ข้อมูลสเปกตรัมที่หาอนุพันธ์เกิดการเลื่อนไปครึ่งจุดเมื่อเทียบกับข้อมูลต้นฉบับ ดังนั้น จุดสูงสุดต่ำสุดจึงไม่ตรงอย่างแน่นอน วิธีนี้จึงเป็นที่ยอมรับได้เมื่อข้อมูลที่ใช้มีความละเอียดสูง แต่ไม่เหมาะกับข้อมูลที่มีความละเอียดต่ำ

6.2.3.2 Moving-Window Polynomial Least-Squares Fitting Method

ในการกรองด้วยวิธีของ Savitsky-Golay Filter นั้นเป็นวิธีที่อาศัย moving-window polynomial least squares fitting ซึ่งการกรองก็สามารถนำมาใช้ในการหาอนุพันธ์เชิงตัวเลขได้ เนื่องจากอยู่บนพื้นฐานของพหุพจน์ ดังนั้น จึงสามารถทำการหาอนุพันธ์โดยตรง เพื่อให้ได้ค่าถ่วงน้ำหนักที่จุดกลาง เพื่อนำมาใช้ในการหาอนุพันธ์ของสเปกตรัมได้ สิ่งที่น่าสนใจคือ การหาอนุพันธ์โดยวิธีนี้นั้น ไม่เกิดการเลื่อนของจุดข้อมูลเหมือนในกรณีของ direct different method ทุกจุดของสเปกตรัมสามารถเขียนได้เป็น

$$x_j^{i+j} = a_0 + a_1 j + a_2 j^2 + \ldots + a_k j^k \qquad (6.36)$$
$$(i = 1,\ldots,n)(j = -m,\ldots,0,\ldots,m)$$

จากการหาอนุพันธ์อันดับหนึ่งโดยตรง จะได้

$$\frac{d(x_j^{i+j})}{d(j)} = a_1 + 2a_2 j + \ldots + ka_k j^{k-1} \qquad (6.37)$$

ที่ค่า $j = 0$ ได้เป็น

$$\frac{d(x_j^{i+j})}{d(j)\Big|_{j=0}} = a_1 \qquad (6.38)$$

หมายความว่า ที่จุดกลางของหน้าต่างนั้นขึ้นอยู่กับสัมประสิทธิ์ a_1 เพียงค่าเดียว

สำหรับอนุพันธ์อันดับสอง

$$\frac{d^2(x_j^{i+j})}{d(j)^2} = 2a_2 + \ldots + k(k-1)a_k j^{k-2} \qquad (6.39)$$

ที่ค่า $j = 0$ ได้เป็น

$$\frac{d^2(x_j^{i+j})}{d(j)^2\Big|_{j=0}} = 2a_2 \qquad (6.40)$$

ซึ่งสามารถหารูปทั่วไปได้เป็น

$$\frac{d^k(x_j^{i+j})}{d(j)^k\Big|_{j=0}} = k!a_k \qquad (6.41)$$

นั่นคือเราสามารถหาค่าถ่วงน้ำหนักของสเปกตรัมที่ทำการหาอนุพันธ์ได้จากความสัมพันธ์

$$a = (M^t M)^{-1} M^t x \qquad (6.42)$$

หรือ

$$\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ \vdots \\ a_k \end{bmatrix} = (M^t M)^{-1} M^t x \qquad (6.43)$$

ทำให้เราสามารถหาอนุพันธ์ที่อันดับและหน้าต่างขนาดต่างๆ กันได้