最小平方法的探討

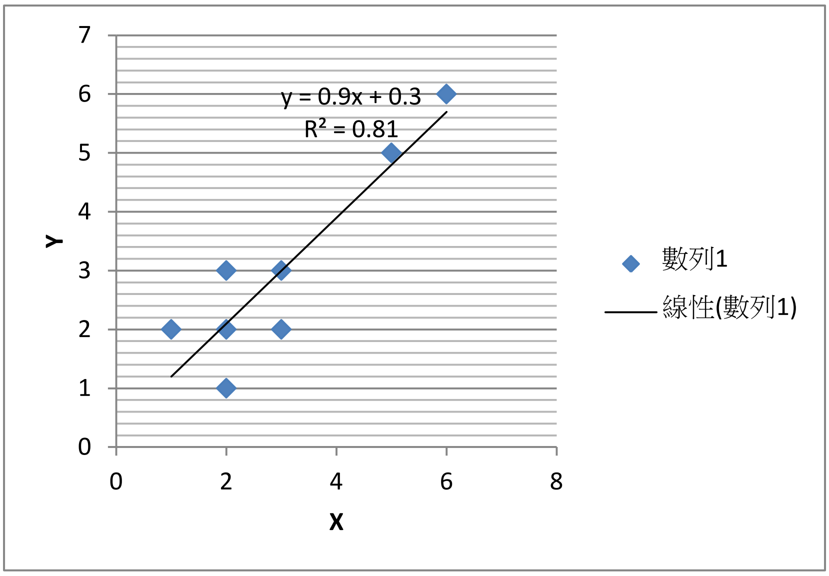
數學科　李益安老師

壹、研究動機

某節上數學課的時候，檢討到一個例題：

X=1,2,2,2,3,3,5,6。 Y=2,1,2,3,2,3,5,6。請畫出其散佈圖並求其迴歸直線方程式：

經計算：$n=8$，$S_{XX}=20$、$S_{XY}=18$、$S_{YY}=20$，利用牛頓最小平方法所得到的迴歸直線方程式為$L:y=0.3+0.9x$，最小平方和為 3.8，有同學發現，這是一個圖形對稱的例子，我們要使用一條直線來代表最可以代表數值”趨勢”的線，我們心目中理想的**”最適合直線”**應該是$L:y=x$，不是嗎？難道是計算錯誤嗎？也沒有！這引發了我對這個問題研究的興趣。



所以，為什麼要定義$F(a,b)=\sum_{i=1}^{n}(y_i-a-bx_i)^2$呢? 如果我們換個方式，把每一個數對$(x_i,y_i)$都往$L:y=a+bx$去做垂直線，那對某一直線所得到的平方和會不會更小呢? 嗯~~直覺上會！

另外，是否有其他的方法也可以找到所謂的**”最適合直線”或”趨勢線”**呢？

貳、研究目的

一、試著用 x 的離差與 y 的離差所造出的函數，找出”函數的最小值”發生時的代表直線，用來表現”趨勢線”的特性。

二、因點到直線的距離為最短，我們重新定義一個”最小平方和”，並找出這個最小平方和的最小值。

進一步探討點到直線的垂直距離的平方和計算所得的 $G(a,b)=\sum_{i=1}^{n}(d_i)^2=\sum_{i=1}^{n}(\frac{|y_i-(a+bx_i)|}{\sqrt{1+b^2}})^2=\sum_{i=1}^{n}\frac{(y_i-a-bx_i)^2}{1+b^2}$ 在不同的 a,b 所產生的最小值，是否會比離差的平方和所得到的 $F(a,b)=\sum_{i=1}^{n}(y_i-a-bx_i)^2$ 在不同的 a,b 所產生的最小值來的小。

三、探討直線 $L:y=a+bx$ 去做水平線，必有實數 $\Delta x_i=x_i-\frac{(y_i-a)}{b}$ ，這些殘差的平方和 $F(a,b)=\sum_{i=1}^{n}(x_i-\frac{y_i-a}{b})^2$ 在不同的 a,b 所產生的最小值。

四、探討兩個方向（x 方向的水平距離和 y 方向的鉛垂距離）的平方和都取，則 $H(a,b)=\sum_{i=1}^{n}(\Delta {y_i}^2+\Delta {x_i}^2)=\sum_{i=1}^{n}((y_i-a-bx_i)^2+(x_i-\frac{y_i-a}{b})^2)$ 在不同的 a,b 所產生的最小值。

五、探討數對 (x_i,y_i) 對直線 L，y 方向的殘差 $\Delta y_i=y_i-(a+bx_i)$ 和對 x 方向的殘差 $\Delta x_i=x_i-\frac{(y_i-a)}{b}$ 的乘積之和 $K(a,b)=\sum_{i=1}^{n}\Delta x_i\Delta y_i=(x_i-\frac{(y_i-a)}{b})(y_i-(a+bx_i))$ (其中 $b\neq 0$)在不同的 a,b 所產生的最小值。

六、討論是否有其他的方法也可以找到所謂的”趨勢線”呢？

參、研究方法步驟

本次作品的研究過程分為三部份。

第一部份，設 $L:y=a+bx$ ，定義 $G(a,b)=\sum_{i=1}^{n}(d_i)^2=\sum_{i=1}^{n}(\frac{|y_i-(a+bx_i)|}{\sqrt{1+b^2}})^2=\sum_{i=1}^{n}\frac{(y_i-a-bx_i)^2}{1+b^2}$ ，找出適當的 a,b ，使得 $G(a,b)$ 的值為最小。

第二部份，設 $L:y=a+bx$ ，定義 $H(a,b)=\sum_{i=1}^{n}(\Delta {y_i}^2+\Delta {x_i}^2)=\sum_{i=1}^{n}((y_i-a-bx_i)^2+(x_i-\frac{y_i-a}{b})^2)$ ，找出適當的 a,b ，使得 $H(a,b)$ 的值為最小。

第三部份，設 $L: y=a+bx$ ，定義 $K(a,b)=\sum_{i=1}^{n}\Delta x_i \Delta y_i=(x_i-\frac{(y_i-a)}{b})(y_i-(a+bx_i))$ ，找出適當的 a,b ，使得 $K(a,b)$ 的值為最小。

我們先複習一下高一下第四章的最小平方法：給定兩組資料 $X=x_1,x_2,x_3,...,x_n$ ，$Y=y_1,y_2,y_3,...,y_n$ ，設 $L: y=a+bx$ ，則對任一數對 (x_i,y_i) ，必有實數 $\Delta y_i=y_i-(a+bx_i)$ 稱為數對 (x_i,y_i) 對直線 L 的殘差，當殘差的平方和 $F(a,b)=\sum_{i=1}^{n}(y_i-a-bx_i)^2$ 最小時，所得到的直線 L 即稱為迴歸直線，或稱為最適合直線，而此方法稱為高斯的最小平方法：

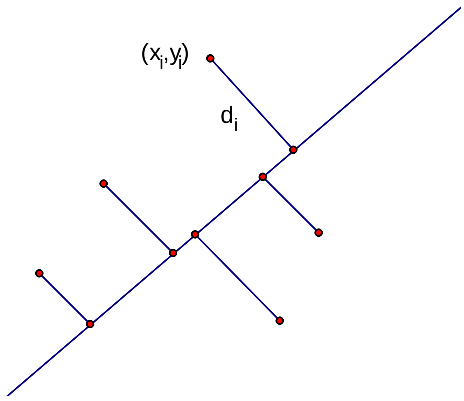
對 a 做偏微分，令 $\frac{F(a,b)}{\partial a}=0$ ，得 $a=\mu_Y-b\mu_X$ ，代入

$$F(a,b)=\sum_{i=1}^{n}(y_i-a-bx_i)^2=\sum_{i=1}^{n}\{(y_i-\mu_Y)-b(x_i-\mu_X)\}^2=b^2S_{XX}-2bS_{XY}+S_{YY}$$

可得到當 $b=\frac{S_{XY}}{S_{XX}}$ 時，$F(a,b)$ 最小值為 $S_{YY}-\frac{S_{XY}{}^2}{S_{XX}}$

註：$S_{YY}=0$ 或 $S_{XX}=0$ 或 $S_{XY}{}^2-S_{XX}S_{YY}=0$ 的情形為所有點在同一直線的特殊情形，以下有碰到上述情形時也先暫時不予討論。

一、第一部份



如果，我們重新定義這些平方和為：每一個點 (x_i,y_i) 到直線 $L: y=a+bx$ 的垂直距離 $d_i=\frac{|y_i-(a+bx_i)|}{\sqrt{1+b^2}}$ 的平方和，即令

$$G(a,b)=\sum_{i=1}^{n}(d_i)^2=\sum_{i=1}^{n}(\frac{|y_i-(a+bx_i)|}{\sqrt{1+b^2}})^2=\sum_{i=1}^{n}\frac{(y_i-a-bx_i)^2}{1+b^2}$$

視為 a 的二次函數，對 a 做偏微分，令

$$\frac{G(a,b)}{\partial a}=\sum_{i=1}^{n}\frac{2(y_i-a-bx_i)(-1)}{1+b^2}=0 ,$$

得 $a=\mu_Y-b\mu_X$ ，

將 a 代入 $G(a,b)$ ，得

$$G(a,b)=\sum_{i=1}^{n}\frac{(y_i-a-bx_i)^2}{1+b^2}=\sum_{i=1}^{n}\frac{(y_i-\mu_Y+b\mu_X-bx_i)^2}{1+b^2}=\sum_{i=1}^{n}\frac{(y_i-\mu_Y-b(x_i-\mu_X))^2}{1+b^2}$$

令

$$G(b)=\sum_{i=1}^{n}\frac{((y_i-\mu_Y)-b(x_i-\mu_X))^2}{1+b^2}=\sum_{i=1}^{n}\frac{b^2(\mu_X-x_i)^2-2b(x_i-\mu_X)(y_i-\mu_Y)+(y_i-\mu_Y)^2}{1+b^2}=\frac{S_{XX}b^2-2S_{XY}b+S_{YY}}{1+b^2}$$

令 $A=S_{XX}\geq 0$ 、 $B=-2S_{XY}$ 、 $C=S_{YY}\geq 0$ ，得 $G(b)=\dfrac{Ab^2+Bb+C}{1+b^2}$

$G(b)$ 對 b 做微分，得 $G'(b)=\dfrac{(2Ab+B)(b^2+1)-2b(Ab^2+Bb+C)}{(1+b^2)^2}=\dfrac{-Bb^2+2(A-C)b+B}{(1+b^2)^2}$ ，

令 $G'(b)=\dfrac{-Bb^2+2(A-C)b+B}{(1+b^2)^2}=0$

(1) 若 $B=-2S_{XY}=0$ ，得 $b=0$ 為唯一解。(不予討論)

(2) 若 $B\neq 0$ ，得 $b=\dfrac{(A-C)\pm\sqrt{(A-C)^2+B^2}}{B}$

$G'(b)$ 再對 b 做微分，得

$$G''(b)=\frac{1}{(1+b^2)^4}\{(1+b^2)[2Bb+2(A-C)]-[-Bb^2+2(A-C)b+B](2(1+b)^2\ \ 2b)\}$$

$$=\frac{2}{(1+b^2)^3}\{(1+b^2)(-Bb+(A-C))-2b(-Bb^2+2(A-C)b+B)\}$$

令 $b_1=\dfrac{(A-C)+\sqrt{(A-C)^2+B^2}}{B}$ ，

$$G''(b_1)=\frac{2}{(1+b_1{}^2)^3}\{(1+b_1{}^2)(-B\frac{(A-C)+\sqrt{(A-C)^2+B^2}}{B}+(A-C))\}=\frac{2}{(1+b_1{}^2)^2}\{-\sqrt{(A-C)^2+B^2}\}<0$$

令 $b_2=\dfrac{(A-C)-\sqrt{(A-C)^2+B^2}}{B}$ ，

$$G''(b_2)=\frac{2}{(1+b_2{}^2)^3}\{(1+b_2{}^2)(-B\frac{(A-C)-\sqrt{(A-C)^2+B^2}}{B}+(A-C))\}=\frac{2}{(1+b_2{}^2)^2}\{\sqrt{(A-C)^2+B^2}\}>0$$

故得 $G(b_2)$ 有最小值，此時的 $b=b_2=\dfrac{(S_{XX}-S_{YY})+\sqrt{(S_{XX}-S_{YY})^2+4S_{XY}{}^2}}{2S_{XY}}$

將 $b_2=\dfrac{(S_{XX}-S_{YY})+\sqrt{(S_{XX}-S_{YY})^2+4S_{XY}{}^2}}{2S_{XY}}$ 代入 $G(a,b)=G(b)=\dfrac{S_{XX}b^2-2S_{XY}b+S_{YY}}{1+b^2}$

$$G(b_2)=\frac{A(\frac{(A-C)-\sqrt{(A-C)^2+B^2}}{B})^2+B\frac{(A-C)-\sqrt{(A-C)^2+B^2}}{B}+C}{1+(\frac{(A-C)-\sqrt{(A-C)^2+B^2}}{B})^2}$$

$$=\frac{A((A-C)-\sqrt{(A-C)^2+B^2})^2+B^2((A-C)-\sqrt{(A-C)^2+B^2})+CB^2}{B^2+((A-C)-\sqrt{(A-C)^2+B^2})^2}$$

$$=\frac{A(A-C)^2-2A(A-C)\sqrt{(A-C)^2+B^2}+A(A-C)^2+AB^2+B^2A-B^2C-B^2\sqrt{(A-C)^2+B^2}+CB^2}{B^2+(A-C)^2-2(A-C)\sqrt{(A-C)^2+B^2}+(A-C)^2+B^2}$$

$$=\frac{2A\{B^2+(A-C)^2-(A-C)\sqrt{(A-C)^2+B^2})\}-B^2\sqrt{(A-C)^2+B^2}}{2(B^2+(A-C)^2-(A-C)\sqrt{(A-C)^2+B^2})}$$

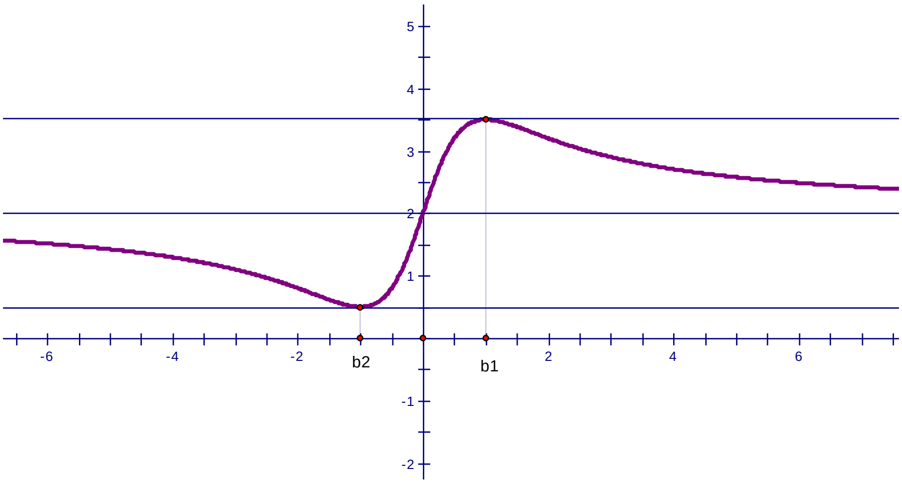
$$=A-B^2\frac{\sqrt{(A-C)^2+B^2}}{2(B^2+(A-C)^2-(A-C)\sqrt{(A-C)^2+B^2})}$$

$$=A-\frac{B^2}{2(\sqrt{(A-C)^2+B^2}-(A-C))}=A-\frac{B^2(\sqrt{(A-C)^2+B^2}+(A-C))}{2((A-C)^2+B^2-(A-C)^2)}$$

$$=A-\frac{B^2(\sqrt{(A-C)^2+B^2}+(A-C))}{2B^2}=A-\frac{(\sqrt{(A-C)^2+B^2}+(A-C))}{2}$$

$$=\frac{(A+C)-\sqrt{(A-C)^2+B^2}}{2}=\frac{(S_{XX}+S_{YY})-\sqrt{(S_{XX}-S_{YY})^2+4S_{XY}{}^2}}{2}$$

(從上式我們可得到 $G(a,b)\geq 0$ ，當 ”=” 成立時， $S_{XX}S_{YY}=S_{XY}{}^2$ ，即所有點在同一直線上)



此時，我們不禁要再問，這樣用垂直距離的平方和計算所得的最小 $G(a,b)$ 值，一定比用離差的平方和所得到的最小 $F(a,b)$ 值來的小嗎?

所以，我們把再 $F(a,b)$ 、 $G(a,b)$ 的兩個最小值相減：

$$(S_{YY}-\frac{S_{XY}{}^2}{S_{XX}})-\frac{(S_{XX}+S_{YY})-\sqrt{(S_{XX}-S_{YY})^2+4S_{XY}{}^2}}{2}$$

$$=\frac{2S_{XX}S_{YY}-2S_{XY}{}^2-S_{XX}{}^2-S_{XX}S_{YY}+S_{XX}\sqrt{(S_{YY}-S_{XX})^2+4S_{XY}{}^2}}{2S_{XX}}$$

$$=\frac{S_{XX}(S_{YY}-S_{XX})-2S_{XY}^{2}+S_{XX}\sqrt{(S_{YY}-S_{XX})^{2}+4S_{XY}^{2}}}{2S_{XX}}$$

$$=\frac{S_{XX}\sqrt{(S_{YY}-S_{XX})^{2}+4S_{XY}^{2}}-(2S_{XY}^{2}-S_{XX}(S_{YY}-S_{XX}))}{2S_{XX}}$$

因為 $S_{XX}\geq 0$，所以我們觀察分子部份：

(1)當 $2S_{XY}^{2}-S_{XX}(S_{YY}-S_{XX})<0$ 時，得明顯的， $F(a,b)-G(a,b)\geq 0$

(2)當 $2S_{XY}^{2}-S_{XX}(S_{YY}-S_{XX})>0$ ，我們把分子的其中兩式平方再相減，得

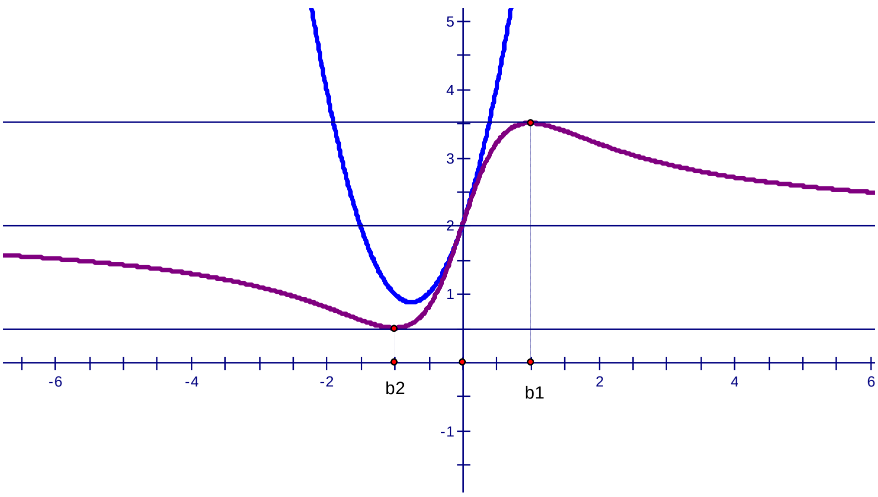
$$(S_{XX}\sqrt{(S_{XX}-S_{YY})^{2}+4S_{XY}^{2}})^{2}-(2S_{XY}^{2}-S_{XX}(S_{YY}-S_{XX}))^{2}=4S_{XY}^{2}(S_{XX}S_{YY}-S_{XY}^{2})\geq 0$$

即 $S_{XX}\sqrt{(S_{XX}-S_{YY})^{2}+4S_{XY}^{2}}>2S_{XY}^{2}-S_{XX}(S_{YY}-S_{XX})$ ，故也得 $F(a,b)-G(a,b)\geq 0$

我們得證，用垂直距離的平方和計算所得的最小 $G(a,b)$ 值

$\dfrac{(S_{XX}+S_{YY})-\sqrt{(S_{XX}-S_{YY})^{2}+4S_{XY}^{2}}}{2}$ 一定比用離差的平方和所得到的最小 $F(a,b)$ 值 $S_{YY}-\dfrac{S_{XY}^{2}}{S_{XX}}$

來的小。



另外，如上圖，從幾何的觀點來看， $G(b)=\dfrac{S_{XX}b^{2}-2S_{XY}b+S_{YY}}{1+b^{2}}\leq F(b)=S_{XX}b^{2}-2S_{XY}b+S_{YY}$ ，

得 $G(b_2)\leq G(\dfrac{S_{XY}}{S_{XX}})\leq F(\dfrac{S_{XY}}{S_{XX}})$ 亦可得證。

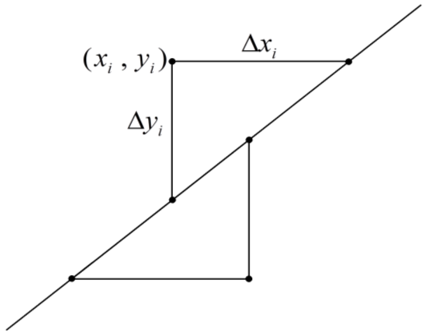
再回到最初的例題，利用本方法”最小垂直距離的平方和”所得到的直線就變成了 $L:y=x$ ，而最小垂直距離的平方和也只剩下了 2。

二、第二部份

如果我們將每一數對(x_i, y_i)，對直線$L: y = a + bx$去做水平線，必有實數$\Delta x_i = x_i - \dfrac{(y_i - a)}{b}$，

將這些殘差的平方和$F(a,b) = \sum_{i=1}^{n}(x_i - \dfrac{y_i - a}{b})^2$去求最小值時，可得當$a = \mu_Y - b\mu_X$，$b = \dfrac{S_{YY}}{S_{XY}}$時，

$F(a,b) = \sum_{i=1}^{n}(x_i - \dfrac{y_i - a}{b})^2$的最小值為$S_{XX} - \dfrac{S_{XY}{}^2}{S_{YY}}$，最初例子得到的直線為$L: y = \dfrac{-1}{3} + \dfrac{10}{9}x$



可見只取水平方向的殘差平方和，或鉛垂方向的殘差平方和，都會對a,b值造成影響。為了公平起見，如果我們兩個平方和都取，令

$H(a,b) = \sum_{i=1}^{n}(\Delta y_i{}^2 + \Delta x_i{}^2) = \sum_{i=1}^{n}((y_i - a - bx_i)^2 + (x_i - \dfrac{y_i - a}{b})^2)$，

$(b \neq 0)$。那結果又會是如何呢?

對a做偏微分，令$\dfrac{H(a,b)}{\partial a} = 0$，得

$\sum_{i=1}^{n}\{(2(y_i - a - bx_i)(-1) + 2(x_i - \dfrac{y_i - a}{b})\dfrac{1}{b})\} = 0$，得$a = \mu_Y - b\mu_X$，代入$H(a,b)$可得

$$H(a,b) = H(b) = \sum_{i=1}^{n}\{(y_i - \mu_Y - b(x_i - \mu_X))^2 + (x_i - \mu_X - \dfrac{y_i - \mu_Y}{b})^2\}$$

$$= S_{YY} - 2bS_{XY} + b^2 S_{XX} + S_{XX} - \dfrac{2S_{XY}}{b} + \dfrac{S_{YY}}{b^2}$$

對$H(b)$微分，$H'(b) = -2S_{XY} + 2bS_{XX} + \dfrac{2S_{XY}}{b^2} - \dfrac{2S_{YY}}{b^3}$

令$H'(b) = 0$，得$H'(b) = b^4 S_{XX} - b^3 S_{XY} + bS_{XY} - S_{YY} = 0$，得 $b^4 - \dfrac{S_{XY}}{S_{XX}}b^3 + \dfrac{S_{XY}}{S_{XX}}b - \dfrac{S_{YY}}{S_{XX}} = 0$

考慮方程式$f(x) = x^4 - Bx^3 + Bx - C$，其中$B = \dfrac{S_{XY}}{S_{XX}}, C = \dfrac{S_{YY}}{S_{XX}} > 0$，令四根$x_1, x_2, x_3, x_4$

則$F(x) = f(x + \dfrac{B}{4}) = (x + \dfrac{B}{4})^4 - B(x + \dfrac{B}{4})^3 + B(x + \dfrac{B}{4}) - C = 0$的四根變成

$x_1 - \dfrac{B}{4}, x_2 - \dfrac{B}{4}, x_3 - \dfrac{B}{4}, x_4 - \dfrac{B}{4}$

$$F(x) = x^4 + (\dfrac{-3}{8}B^2)x^2 + (\dfrac{-1}{8}B^3 + B)x + (\dfrac{-3}{256}B^4 + \dfrac{1}{4}B^2 - C)$$

令$F(x) = x^4 + (\dfrac{-3}{8}B^2)x^2 + (\dfrac{-1}{8}B^3 + B)x + (\dfrac{-3}{256}B^4 + \dfrac{1}{4}B^2 - C) = x^4 + px^2 + qx + r$

將 $p=\dfrac{-3}{8}B^2, q=\dfrac{-1}{8}B^3+B, r=\dfrac{-3}{256}B^4+\dfrac{1}{4}B^2-C$ 代入四次方程式判別式

$D=4(p^2+12r)^3-(2p^3-72pr+27q^2)^2$

$=4\{(\dfrac{-3}{8}B^2)^2+12(\dfrac{-3}{256}B^4+\dfrac{1}{4}B^2-C)\}^3-\{2(\dfrac{-3}{8}B^2)^3-72(\dfrac{-3}{8}B^2)(\dfrac{-3}{256}B^4+\dfrac{1}{4}B^2-C)+27(\dfrac{-1}{8}B^3+B)^2\}^2$

經化簡得 $D=27(4B^6+6B^4C+48B^2C^2-256C^3-27B^4-27B^4C^2)$

又 $B^2-C=(\dfrac{S_{XY}}{S_{XX}})^2-\dfrac{S_{YY}}{S_{XX}}=\dfrac{S_{XY}{}^2-S_{XX}S_{YY}}{S_{XX}{}^2}\le 0$ ，所以 $B^2\le C$ ，代入上式得

$D\le 27(-198C^3-27C^2-27C^4)\le 0$

然而 $D=0\Leftrightarrow C=0$ 且 $B^2=C\Leftrightarrow S_{YY}=0$ 且 $S_{XY}{}^2-S_{XX}S_{YY}=0$ (所有點在同一直線，不予討論)

故得 $D<0$ ，所以可得 $H'(b)=0$ 的四根必為二實二虛。

觀察方程式 $H'(x)=f(x)=x^4-Bx^3+Bx-C=0$，其中 $B=\dfrac{S_{XY}}{S_{XX}}, C=\dfrac{S_{YY}}{S_{XX}}>0$ ， $f(0)=-C<0$ ，

因為 $f(x)$ 是四次的，領導係數為 1 的多項式，四根又為二實二虛，故 $x^4-Bx^3+Bx-C=0$ 必得一正根 b_1 及一負根 b_2 。

對於兩根來討論：

$f(0)=-C<0$ ， $f(1)=f(-1)=1-C$ ， $f(B)=B^2-C<0$ ，

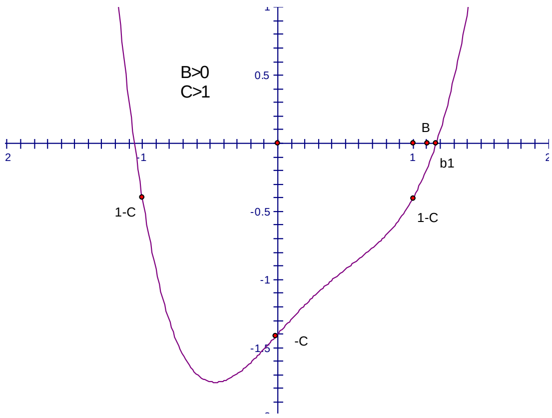
(1) 若 $1-C<0$ ，即 $C>1$ ，即 $S_{YY}>S_{XX}$ ，由勘根定理，正根 $b_1>1$，負根 $b_2<-1$。(圖一、圖四)

(2) 若 $1-C>0$ ，即 $0<C<1$ ，即 $S_{YY}<S_{XX}$ ，由勘根定理，正根 $0<b_1<1$ ，負根 $-1<b_2<0$ 。(圖二、圖三)

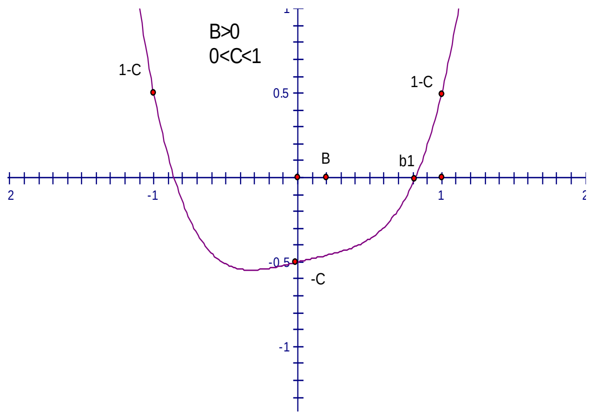
(3) 若 $1-C=0$ ，即 $S_{YY}=S_{XX}$ ，正根 $b_1=1$，負根 $b_2=-1$

(4) 若 $S_{XY}>0$ ，則 $B>0$， $f(B)=B^2-C<0$ ，由勘根定理，正根 $b_1>B=\dfrac{S_{XY}}{S_{XX}}$ 。(圖一、圖二)

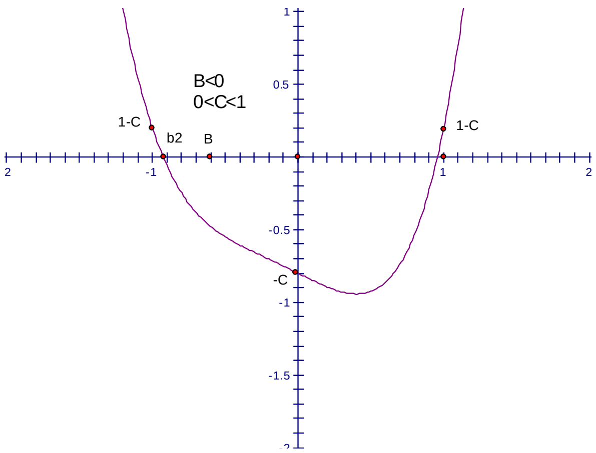
(5) 若 $S_{XY}<0$ ，則 $B<0$， $f(B)=B^2-C<0$ ，由勘根定理，負根 $b_2<B=\dfrac{S_{XY}}{S_{XX}}$ 。(圖三、圖四)



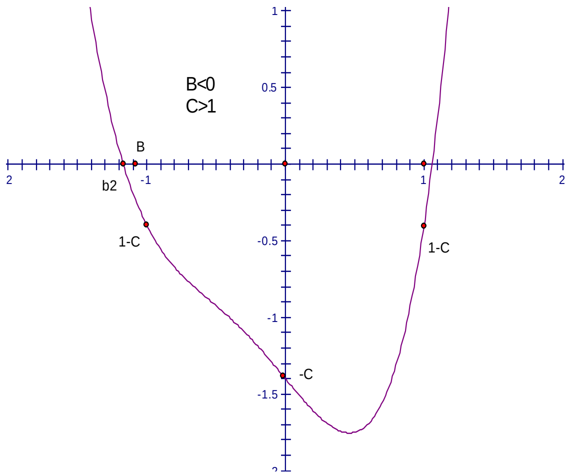
(圖一)



(圖二)



(圖三)



(圖四)

對於最小值而言：

(1) 若 $S_{YY}=S_{XX}$ ，正根 $b_1=1$，負根 $b_2=-1$

若 $S_{XY}>0$ ， $H(1)=2(S_{YY}-2S_{XY}+S_{XX})$ 最小；若 $S_{XY}<0$ ， $H(-1)=2(S_{YY}+2S_{XY}+S_{XX})$ 最小

(2) 若 $S_{XY}=0$ ，將 $H'(x)=f(x)=x^4-Bx^3+Bx-C=x^4-C=0$ ， $b=\pm\sqrt[4]{C}$ ，

$H''(b)=2S_{XX}+\dfrac{6S_{YY}}{b^4}=8S_{XX}>0$ ， $b=\pm\sqrt[4]{C}$ 代入 $H(b)$ 皆可得

$H(b)=S_{YY}+2\sqrt{S_{XX}S_{YY}}+S_{XX}=(\sqrt{S_{XX}}+\sqrt{S_{YY}})^2$ 最小

(3) 若 $S_{XY}\neq 0$ 且 $S_{YY}\neq S_{XX}$

$H(b)=S_{YY}-2bS_{XY}+b^2S_{XX}+S_{XX}-\frac{2S_{XY}}{b}+\frac{S_{YY}}{b^2}$ ， $b\neq 0$

$H'(b)=-2S_{XY}+2bS_{XX}+\frac{2S_{XY}}{b^2}-\frac{2S_{YY}}{b^3}=\frac{2S_{XX}}{b^3}(b^4-\frac{S_{XY}}{S_{XX}}b^3+\frac{S_{XY}}{S_{XX}}b-\frac{S_{YY}}{S_{XX}})=\frac{2S_{XX}}{b^3}(b^4-Bb^3+Bb-C)$，

其中 $B=\frac{S_{XY}}{S_{XX}},C=\frac{S_{YY}}{S_{XX}}>0$

$H''(b)=2S_{XX}-\frac{4S_{XY}}{b^3}+\frac{6S_{YY}}{b^4}=2\cdot\frac{b^4S_{XX}-2bS_{XY}+3S_{YY}}{b^4}=\frac{2S_{XX}}{b^4}(b^4-2bB+3C)$

觀察函數 $H(b)$ ，當 $bS_{XY}>0$ ， $H(b)$ 值會較 $bS_{XY}<0$ 時為小，故我們取 $bB>0$ 時，可得較小的 $H(b)$ 值。

因正根 b_1 及負根 b_2 是 $b^4-Bb^3+Bb-C=0$ 的根，故可將 $C={b_1}^4-B{b_1}^3+Bb_1$ 或 $C={b_2}^4-B{b_2}^3+Bb_2$ 代入 $H''(b)$

因此，在 $b=b_1$ 或 $b=b_2$ 時，

$H''(b)=\frac{2S_{XX}}{b^4}(b^4-2bB+3C)=\frac{2S_{XX}}{b^4}(4b^4-3b^3B+Bb)=\frac{2S_{XX}}{b^4}(3b^3(b-B)+b^4+Bb)$

(1)當 $B>0$ 時，取正根 b_1 ，得 $b_1>B>0$ ，代入 $H''(b)$ ，得 $H''(b_1)>0$ ，函數凹口向上，我們有 $H(b_1)$ 為最小值。(圖一、圖二)

(2)當 $B<0$ 時，取負根 b_2 ，得 $b_2<B<0$ ，代入 $H''(b)$ ，得 $H''(b_2)>0$ ，函數凹口向上，我們有 $H(b_2)$ 為最小值。(圖三、圖四)

$B=\frac{S_{XY}}{S_{XX}}$ 和 S_{XY} 值同號，因為 b 代表所求直線的斜率，由 S_{XY} 值和 b 值正負的關係，也符合我們對此直線的期待。

現在我們進一步來找出這兩個實根的值！

方法如下：

利用笛卡爾(Descartes)方法，設 $f(x)=x^4-Bx^3+Bx-C=0$ 的四個根為 x_1,x_2,x_3,x_4 (我們已知二實二虛)，將方程式平移

令 $x=y+\frac{B}{4}$ 代入上式 $f(y+\frac{B}{4})=(y+\frac{B}{4})^4-B(y+\frac{B}{4})^3+B(y+\frac{B}{4})-C=0$

則 $y^4+\frac{-3B^2}{8}y^2+(\frac{-B^3}{8}+B)y+(\frac{-3B^4}{256}+\frac{B^2}{4}-C)=0$ 的四個根為

$y_1,y_2,y_3,y_4=x_1-\frac{B}{4},x_2-\frac{B}{4},x_3-\frac{B}{4},x_4-\frac{B}{4}$

令 $y^4+\frac{-3B^2}{8}y^2+(\frac{-B^3}{8}+B)y+(\frac{-3B^4}{256}+\frac{B^2}{4}-C)=y^4+\alpha y^2+\beta y+\gamma=0$ ，

$y^4+\alpha y^2+\beta y+\gamma=(y^2+ky+l)(x^2-ky+m)=0$

其中，k^2 為 $z^3+2\alpha z^2+(\alpha^2-4\gamma)z-\beta^2=z^3-\frac{3B^2}{4}z^2+(\frac{3}{16}B^4-B^2+4C)z-(\frac{-B^3}{8}+B)^2=0$ 的一實根，且 $l=\frac{-\beta+\alpha k+k^3}{2k}$，$m=\frac{\beta+\alpha k+k^3}{2k}$ (因我們已知根是二實二虛，所以實數 k 值必存在)

設 $z^3-\frac{3B^2}{4}z^2+(\frac{3}{16}B^4-B^2+4C)z-(\frac{-B^3}{8}+B)^2=0$ 的三個根為 $z_1=k^2\,,z_2\,,z_3$ ，再由卡當(Carden)公式，將方程式平移

令 $z=u+\frac{B^2}{4}$ 代入上式 $(u+\frac{B^2}{4})^3-\frac{3B^2}{4}(u+\frac{B^2}{4})^2+(\frac{3}{16}B^4-B^2+4C)(u+\frac{B^2}{4})-(\frac{-B^3}{8}+B)^2=0$

得 $u^3+(-B^2+4C)u+B^2(C-1)=0$ 的三個根為 $u_1,u_2,u_3=z_1-\frac{B^2}{4},z_2-\frac{B^2}{4},z_3-\frac{B^2}{4}$ ，

令 $u^3+(-B^2+4C)\,u+B^2(C-1)=u^3+pu+q=0$ ，因 $B^2<C$ ，所以 $p=-B^2+4C>0$ ，

可得 $D=\frac{q^2}{4}+\frac{p^3}{27}=\frac{B^4(C-1)^2}{4}+\frac{(-B^2+4C)^3}{27}>0$ ，(一實二虛)

$k^2-\frac{B^2}{4}=(\frac{-q}{2}+\sqrt{D})^{\frac{1}{3}}+(\frac{-q}{2}-\sqrt{D})^{\frac{1}{3}}$ ，可得 $k^2=\frac{B^2}{4}+(\frac{-q}{2}+\sqrt{D})^{\frac{1}{3}}+(\frac{-q}{2}-\sqrt{D})^{\frac{1}{3}}$

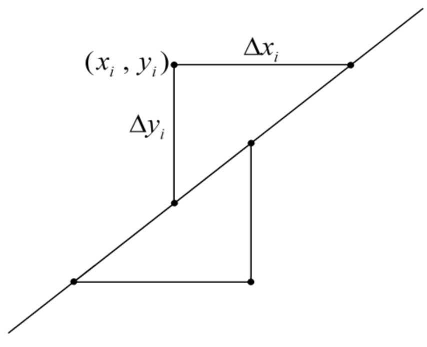
$$k=\pm\sqrt{\frac{B^2}{4}+(\frac{-q}{2}+\sqrt{D})^{\frac{1}{3}}+(\frac{-q}{2}-\sqrt{D})^{\frac{1}{3}}}$$

$$=\pm\sqrt{\frac{B^2}{4}+(\frac{B^2(1-C)}{2}+\sqrt{\frac{B^4(C-1)^2}{4}+\frac{(-B^2+4C)^3}{27}})^{\frac{1}{3}}+(\frac{B^2(1-C)}{2}-\sqrt{\frac{B^4(C-1)^2}{4}+\frac{(-B^2+4C)^3}{27}})^{\frac{1}{3}}}$$

(我們取 k 為正值即可)

再代入求出 $l=\frac{-\beta+\alpha k+k^3}{2k}$ ，$m=\frac{\beta+\alpha k+k^3}{2k}$ ，找出是 $k^2-4l>0$ 或 $k^2-4m>0$ 得出 $y^2+ky+l=0$ 或 $x^2-ky+m=0$ 中的兩實根(其中必只有一個成立)，再將此兩實根加上 $\frac{B}{4}$ ，可以得到我們的正根 b_1 及負根 b_2 。

三、第三部份



我們再考慮數對 (x_i,y_i) 對直線 L，y 方向的殘差 $\Delta y_i=y_i-(a+bx_i)$ 和對 x 方向的殘差 $\Delta x_i=x_i-\frac{(y_i-a)}{b}$ 的乘積和 $K(a,b)=\sum_{i=1}^{n}\Delta x_i\Delta y_i=\sum_{i=1}^{n}(x_i-\frac{(y_i-a)}{b})(y_i-(a+bx_i))$ 其中 $b\neq 0$

由圖形可以看得出來，當斜率 $b>0$ 時，每一個 $\Delta x_i\Delta y_i\leq 0$ ，當斜率 $b<0$ 時，每一個 $\Delta x_i\Delta y_i\geq 0$ 。

因此 $K(a,b)$ 值不是恆正就是恆負。故我們如果要取得適當的 a,b 值，來讓 $K(a,b)$ 值得到最小，

就必須要加上絕對值，即應該要討論$|K(a,b)|$的最小值。

先不考慮$K(a,b)$值的正負，對a做偏微分，令$\frac{K(a,b)}{\partial a}=0$，得$a=\mu_Y-b\mu_X$，將$a$代入$K(a,b)$，

得$K(a,b)=K(b)=\sum_{i=1}^{n}(y_i-\mu_Y-b(x_i-\mu_X))(x_i-\mu_X-\frac{y_i-\mu_Y}{b})=-bS_{XX}+2S_{XY}-\frac{1}{b}S_{YY}$

(1) 若$b>0$時，也可由算幾不等式，$bS_{XX}+\frac{1}{b}S_{YY}\geq 2\sqrt{S_{XX}S_{YY}}\geq 2|S_{XY}|$，故$K(a,b)\leq 0$即

$-K(a,b)\geq 0$

考慮$-K(b)=bS_{XX}-2S_{XY}+\frac{1}{b}S_{YY}$，$-K'(b)=S_{XX}-\frac{1}{b^2}S_{YY}$，$-K''(b)=\frac{2}{b^3}S_{YY}$

若$-K'(b)=S_{XX}-\frac{1}{b^2}S_{YY}=0$，即$b=\pm\sqrt{\frac{S_{YY}}{S_{XX}}}$(取正)，代入$-K''(b)$得$-K''(\sqrt{\frac{S_{YY}}{S_{XX}}})>0$

故得$-K(\sqrt{\frac{S_{XX}}{S_{YY}}})=2(\sqrt{S_{XX}S_{YY}}-S_{XY})$會有最小值

(2) $b<0$時，也可由算幾不等式，$-bS_{XX}-\frac{1}{b}S_{YY}\geq 2\sqrt{S_{XX}S_{YY}}\geq 2|S_{XY}|$，故$K(a,b)\geq 0$

考慮$K(b)=-bS_{XX}+2S_{XY}-\frac{1}{b}S_{YY}$，$K'(b)=-S_{XX}+\frac{1}{b^2}S_{YY}$，$K''(b)=\frac{-2}{b^3}S_{YY}$

若$K'(b)=-S_{XX}+\frac{1}{b^2}S_{YY}=0$，即$b=\pm\sqrt{\frac{S_{YY}}{S_{XX}}}$(取負)，代入$K''(b)$得$K''(-\sqrt{\frac{S_{YY}}{S_{XX}}})>0$

故得$K(-\sqrt{\frac{S_{XX}}{S_{YY}}})=2(\sqrt{S_{XX}S_{YY}}+S_{XY})$會有最小值。

因此，

(1)若$S_{XY}>0$，我們會取$b=\sqrt{\frac{S_{YY}}{S_{XX}}}$，得$|K(a,b)|=-K(\sqrt{\frac{S_{XX}}{S_{YY}}})=2(\sqrt{S_{XX}S_{YY}}-S_{XY})$會有最小值

(2)若$S_{XY}<0$，我們會取$b=-\sqrt{\frac{S_{YY}}{S_{XX}}}$，得$|K(a,b)|=K(\sqrt{\frac{S_{XX}}{S_{YY}}})=2(\sqrt{S_{XX}S_{YY}}+S_{XY})$會有最小值

(3)若$S_{XY}=0$，我們會取$b=\pm\sqrt{\frac{S_{YY}}{S_{XX}}}$，得$|K(a,b)|=-K(\sqrt{\frac{S_{XX}}{S_{YY}}})=2\sqrt{S_{XX}S_{YY}}$會有最小值

肆、研究結果

一、第一部份：設$L:y=a+bx$，定義$G(a,b)=\sum_{i=1}^{n}(d_i)^2=\sum_{i=1}^{n}(\frac{|y_i-(a+bx_i)|}{\sqrt{1+b^2}})^2=\sum_{i=1}^{n}\frac{(y_i-a-bx_i)^2}{1+b^2}$，

$b=\frac{(S_{YY}-S_{XX})+\sqrt{(S_{YY}-S_{XX})^2+4S_{XY}{}^2}}{2S_{XY}}$，$a=\mu_Y-b\mu_X$時，$G(a,b)$的最小值為：

$\dfrac{(S_{XX}+S_{YY})-\sqrt{(S_{XX}-S_{YY})^2+4S_{XY}{}^2}}{2}$ 。

且此最小值會比”高斯最小平方法”的得到的最小值 $S_{YY}-\dfrac{S_{XY}{}^2}{S_{XX}}$ 來的小。

二、第二部份：設 $L: y=a+bx$ ，定義

$$H(a,b)=\sum_{i=1}^{n}(\Delta y_i{}^2+\Delta x_i{}^2)=\sum_{i=1}^{n}((y_i-a-bx_i)^2+(x_i-\frac{y_i-a}{b})^2)$$ ，

(1)若 $S_{XY}>0$，則我們取方程式 $b^4-\dfrac{S_{XY}}{S_{XX}}b^3+\dfrac{S_{XY}}{S_{XX}}b-\dfrac{S_{YY}}{S_{XX}}=0$ 的唯一正根 b_1，$a=\mu_Y-b_1\mu_X$，可使 $H(a,b)$ 的值為最小。

(2)若 $S_{XY}<0$，則我們取方程式 $b^4-\dfrac{S_{XY}}{S_{XX}}b^3+\dfrac{S_{XY}}{S_{XX}}b-\dfrac{S_{YY}}{S_{XX}}=0$ 的唯一負根 b_2，$a=\mu_Y-b_2\mu_X$，可使 $H(a,b)$ 的值為最小。

(3)若 $S_{XY}=0$，則我們取 $b=\pm\sqrt[4]{\dfrac{S_{YY}}{S_{XX}}}$ $a=\mu_Y-(\pm\sqrt[4]{\dfrac{S_{YY}}{S_{XX}}})\mu_X$ ，可得 $H(a,b)=(\sqrt{S_{XX}}+\sqrt{S_{YY}})^2$ 的值為最小。

(4)若 $S_{YY}=S_{XX}$ ，正根 $b_1=1$，負根 $b_2=-1$，

若 $S_{XY}>0$ ， $H(1)=2(S_{YY}-2S_{XY}+S_{XX})$ 為最小；若 $S_{XY}<0$ ， $H(-1)=2(S_{YY}+2S_{XY}+S_{XX})$ 為最小。

三、第三部份：設 $L: y=a+bx$ ，定義 $K(a,b)=\sum_{i=1}^{n}\Delta x_i\Delta y_i=(x_i-\dfrac{(y_i-a)}{b})(y_i-(a+bx_i))$ ，

(1)若 $S_{XY}>0$ ，我們會取 $b=\sqrt{\dfrac{S_{YY}}{S_{XX}}}$ ， $a=\mu_Y-\sqrt{\dfrac{S_{YY}}{S_{XX}}}\mu_X$ ，得

$|K(a,b)|=-K(\sqrt{\dfrac{S_{XX}}{S_{YY}}})=2(\sqrt{S_{XX}S_{YY}}-S_{XY})$ 是最小值。

(2)若 $S_{XY}<0$ ，我們會取 $b=-\sqrt{\dfrac{S_{YY}}{S_{XX}}}$ ， $a=\mu_Y+\sqrt{\dfrac{S_{YY}}{S_{XX}}}\mu_X$ ，得

$|K(a,b)|=K(\sqrt{\dfrac{S_{XX}}{S_{YY}}})=2(\sqrt{S_{XX}S_{YY}}+S_{XY})$ 是最小值。

(3)若 $S_{XY}=0$，我們會取 $b=\pm\sqrt{\dfrac{S_{YY}}{S_{XX}}}$，$a=\mu_Y\pm\sqrt{\dfrac{S_{YY}}{S_{XX}}}\mu_X$，得 $|K(a,b)|=-K(\sqrt{\dfrac{S_{XX}}{S_{YY}}})=2\sqrt{S_{XX}S_{YY}}$ 會是最小值。

伍、討論

以上三個方法所得到的”趨勢線”都會經過(μ_X , μ_Y)。我們嘗試著找出其他的方法，例如：將方法一和方法二的和相加：

$\dfrac{S_{XX}b^2-2S_{XY}b+S_{YY}}{1+b^2}+S_{YY}-2bS_{XY}+b^2S_{XX}+S_{XX}-\dfrac{2S_{XY}}{b}+\dfrac{S_{YY}}{b^2}$ ，或方法二和方法三相加：

$S_{YY}-2bS_{XY}+b^2S_{XX}+S_{XX}-\dfrac{2S_{XY}}{b}+\dfrac{S_{YY}}{b^2}+\left|-bS_{XX}+2S_{XY}-\dfrac{1}{b}S_{YY}\right|$ ，或取其線性組合

$p\dfrac{S_{XX}b^2-2S_{XY}b+S_{YY}}{1+b^2}+q(S_{YY}-2bS_{XY}+b^2S_{XX}+S_{XX}-\dfrac{2S_{XY}}{b}+\dfrac{S_{YY}}{b^2})+r\left|-bS_{XX}+2S_{XY}-\dfrac{1}{b}S_{YY}\right|$等方式，去取其最小值，因為想要求出函數最小值時，解 b 的方程式次數相當高，計算相當繁複，且沒有明確的目的性或必要性，因此沒有再深入探討。

陸、參考資料

一、方程式論，作者：楊重駿。

二、近世代數，作者：康明昌。

三、應用線性代數，作者：Bernard Kolman。

四、代數學專題，作者：薛昭雄，黃志平。

五、大學用書—高等方程式論。