4.5 節補充//最小二乘法

考慮下面的線性觀測模式：

$$\mathbf{y}_o = \mathbf{H}\mathbf{x} + \boldsymbol{\varepsilon}_r \quad (1)$$

上式中 $\mathbf{H} \in \mathfrak{R}^{L \times M}$ ($L > M$) 表示線性觀測算符。假如要由已知的觀測值 $\mathbf{y}_o \in \mathfrak{R}^L$ 決定狀態變數 $\mathbf{x} \in \mathfrak{R}^M$ ，但沒有背景值，也沒有關於觀測誤差 $\boldsymbol{\varepsilon}_r \in \mathfrak{R}^L$ 的統計量，此時只得使用最小二乘法，也就是說決定最佳的 $\mathbf{x}$ ，使下面觀測誤差的方差取極小值：

$$J = \frac{1}{2} \boldsymbol{\varepsilon}_r^T \boldsymbol{\varepsilon}_r = \frac{1}{2} (\mathbf{H}\mathbf{x} - \mathbf{y}_o)^T (\mathbf{H}\mathbf{x} - \mathbf{y}_o) \quad (2)$$

上式乘出來，有

$$2J = \mathbf{x}^T \mathbf{H}^T \mathbf{H} \mathbf{x} - \mathbf{y}_o^T \mathbf{H} \mathbf{x} - \mathbf{x}^T \mathbf{H}^T \mathbf{y}_o + \mathbf{y}_o^T \mathbf{y}_o \quad (3)$$

求解這個問題有下面幾種方法。

微分等於零

將(3)式對 $\mathbf{x}$ 微分，並令結果等於零，有

$$2 \frac{\partial J}{\partial \mathbf{x}} = 2\mathbf{H}^T \mathbf{H} \mathbf{x} - \mathbf{H}^T \mathbf{y}_o - \mathbf{H}^T \mathbf{y}_o = 2(\mathbf{H}^T \mathbf{H} \mathbf{x} - \mathbf{H}^T \mathbf{y}_o) = \mathbf{0} \quad (4)$$

因而

$$\mathbf{H}^T \mathbf{H} \mathbf{x} = \mathbf{H}^T \mathbf{y}_o \quad (5)$$

由此解出 $\mathbf{x}$ ，得到

$$\mathbf{x} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}_o \quad (6)$$

這個方法需用到下面的矩陣微分公式：

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T \mathbf{H}^T \mathbf{H} \mathbf{x} = 2\mathbf{H}^T \mathbf{H} \mathbf{x}, \quad \frac{\partial}{\partial \mathbf{x}} \mathbf{y}_o^T \mathbf{H} \mathbf{x} = \frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T \mathbf{H}^T \mathbf{y}_o = \mathbf{H}^T \mathbf{y}_o$$

這是缺點。

配方法

考慮下面的表達式：

$$\begin{aligned} & [\mathbf{x} - (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}_o]^T (\mathbf{H}^T \mathbf{H}) [\mathbf{x} - (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}_o] \\ &= \mathbf{x}^T \mathbf{H}^T \mathbf{H} \mathbf{x} - \mathbf{y}_o^T \mathbf{H} \mathbf{x} - \mathbf{x}^T \mathbf{H}^T \mathbf{y}_o + \mathbf{y}_o^T \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}_o \\ &= 2J - \mathbf{y}_o^T \mathbf{y}_o + \mathbf{y}_o^T \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}_o \end{aligned} \quad (7)$$

在導出上式第 2 個等號右邊時已用到(3)式。上式改寫為

$$\begin{aligned} 2J &= [\mathbf{x} - (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}_o]^T (\mathbf{H}^T \mathbf{H}) [\mathbf{x} - (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}_o] \\ &\quad + \mathbf{y}_o^T \mathbf{y}_o - \mathbf{y}_o^T \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}_o \end{aligned} \quad (8)$$

這就是說，(2)或(3)式右邊加減一項 $\mathbf{y}_o^T \mathbf{H} (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}_o$ ，就變為(8)式。上式右邊第 2 和第 3 項跟 $\mathbf{x}$ 無關，第 1 項一定是正值 ($L \times M$ 矩陣 $\mathbf{H}$ 若是滿秩的，則 $\mathbf{H}^T \mathbf{H}$ 是正定的)，因而第 1 項等於 $\mathbf{0}$ 時，則 J 取極小值，即

$$\mathbf{x} - (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}_o = \mathbf{0} \quad (9)$$

這個結果和(6)式一樣。

上面這種寫法還不夠清楚，最好用下面的方法說明。(3)式可寫為

$$\begin{aligned} 2J &= \mathbf{x}^T \mathbf{H}^T \mathbf{H} \mathbf{x} - 2\mathbf{x}^T \mathbf{H}^T \mathbf{y}_o + \mathbf{y}_o^T \mathbf{y}_o \\ &= \mathbf{x}^T \mathbf{H}^T \mathbf{H} \mathbf{x} - 2\mathbf{x}^T (\mathbf{H}^T \mathbf{H}) (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}_o + \mathbf{y}_o^T \mathbf{y}_o \\ &= [\mathbf{x} - (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}_o]^T \mathbf{H}^T \mathbf{H} [\mathbf{x} - (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}_o] + \text{const.} \end{aligned}$$

上式中 const.是和 $\mathbf{x}$ 無關的量。因而當

$$\mathbf{x} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}_o$$

時， J 取極小值。

變分法

將(2)式中的 $\mathbf{x}$ 取變分，有

$$\begin{aligned}\delta J &= \frac{1}{2}(\mathbf{H}\delta\mathbf{x})^T(\mathbf{H}\mathbf{x} - \mathbf{y}_o) + \frac{1}{2}(\mathbf{H}\mathbf{x} - \mathbf{y}_o)^T \mathbf{H}\delta\mathbf{x} \\ &= \delta\mathbf{x}^T \mathbf{H}^T(\mathbf{H}\mathbf{x} - \mathbf{y}_o)\end{aligned}\quad (10)$$

必須記得，由於 δJ 是純量，因此上式第 1 個等號右邊兩項是相等的。令上式等於零，由於 $\delta\mathbf{x}$ 是任意的，因此 $\mathbf{H}^T(\mathbf{H}\mathbf{x} - \mathbf{y}_o) = \mathbf{0}$ ，這個結果和(6)式一樣。變分法的理論基礎在於擾動法，將在下面說明。

擾動法

令

$$\mathbf{x} = \hat{\mathbf{x}} + \varepsilon\mathbf{e}\quad (11)$$

其中 $\hat{\mathbf{x}}$ 代表使(2)式取極小值的最佳解， ε 是無限小正數， $\mathbf{e}$ 為任意向量。代入(2)式，得到

$$\begin{aligned}2J(\varepsilon) &= [\mathbf{H}(\hat{\mathbf{x}} + \varepsilon\mathbf{e}) - \mathbf{y}_o]^T [\mathbf{H}(\hat{\mathbf{x}} + \varepsilon\mathbf{e}) - \mathbf{y}_o] \\ &= (\mathbf{H}\hat{\mathbf{x}} - \mathbf{y}_o + \varepsilon\mathbf{H}\mathbf{e})^T (\mathbf{H}\hat{\mathbf{x}} - \mathbf{y}_o + \varepsilon\mathbf{H}\mathbf{e})\end{aligned}$$

將上式乘出來，有

$$\begin{aligned}2J(\varepsilon) &= 2\hat{J} + \varepsilon[\mathbf{e}^T \mathbf{H}^T(\mathbf{H}\hat{\mathbf{x}} - \mathbf{y}_o) + (\mathbf{H}\hat{\mathbf{x}} - \mathbf{y}_o)^T \mathbf{H}\mathbf{e}] + \varepsilon^2 \mathbf{e}^T \mathbf{H}^T \mathbf{H}\mathbf{e} \\ &= 2\hat{J} + 2\varepsilon \mathbf{e}^T \mathbf{H}^T(\mathbf{H}\hat{\mathbf{x}} - \mathbf{y}_o) + \varepsilon^2 \mathbf{e}^T \mathbf{H}^T \mathbf{H}\mathbf{e}\end{aligned}\quad (12)$$

其中 $\hat{J}$ 是極小值，即

$$\hat{J} = \frac{1}{2}(\mathbf{H}\hat{\mathbf{x}} - \mathbf{y}_o)^T (\mathbf{H}\hat{\mathbf{x}} - \mathbf{y}_o)\quad (13)$$

必須記得，由於 $J(\varepsilon)$ 是純量，因此(12)式第 1 個等號右邊中括弧內的兩項是相等的。令

$$\left. \frac{dJ(\varepsilon)}{d\varepsilon} \right|_{\varepsilon=0} = 0$$

我們得到

$$\left. \frac{dJ(\varepsilon)}{d\varepsilon} \right|_{\varepsilon=0} = \mathbf{e}^T \mathbf{H}^T (\mathbf{H}\hat{\mathbf{x}} - \mathbf{y}_o) = 0 \quad (14)$$

因為向量 $\mathbf{e}$ 是任意的，故由上式有 $\mathbf{H}^T (\mathbf{H}\hat{\mathbf{x}} - \mathbf{y}_o) = \mathbf{0}$ ，這個結果和前面一樣。

由上面的說明可知，擾動法把對向量 $\mathbf{x}$ 的微分，如(4)式所示，化為我們比較熟悉的對純量 ε 的微分，如(14)式所示。

(14)式兩邊乘以 ε ，有

$$\varepsilon J'(0) = \varepsilon \mathbf{e}^T \mathbf{H}^T (\mathbf{H}\hat{\mathbf{x}} - \mathbf{y}_o) = 0 \quad (15)$$

令 $\delta \mathbf{x} = \varepsilon \mathbf{e}$ ， $\delta J = \varepsilon J'(0)$ 分別代表 $\mathbf{x}$ 和 J 的變分，則(15)式變為

$$\delta J = \delta \mathbf{x}^T \mathbf{H}^T (\mathbf{H}\hat{\mathbf{x}} - \mathbf{y}_o) = 0 \quad (16)$$

這就是(10)式。

變分學原理

我們再複述變分學原理。變分學是求泛函的極值，泛函的最簡單的形式為

$$J[\varphi] = \int_{x_1}^{x_2} F(x, \varphi, \varphi_x) dx \quad (17)$$

函數的求極值問題已為人所熟知，因此要先把泛函求極值問題先化為函數求極值問題。假設 $\hat{\varphi}(x)$ 是會使 J 取極值的函數。現在令

$$\varphi(x) = \hat{\varphi}(x) + \varepsilon \eta(x) \quad (18)$$

其中 ε 是一個無限小正數， $\eta(x)$ 是 x 的任意已知函數。變動 ε 的值，可得到不同的 $\varphi(x)$ ，從而給出不同的 J 值。因此這時 J 就只依賴於 ε ，即

$$J(\varepsilon) \equiv J[\hat{\varphi} + \varepsilon \eta] = \int_{x_1}^{x_2} F(x, \hat{\varphi} + \varepsilon \eta, \hat{\varphi}_x + \varepsilon \eta_x) dx \quad (19)$$

此外根據(18)式，當 $\varepsilon=0$ 時， J 取極小值。在這情況下，原來泛函求極值的問題就簡化為函數求極值的問題。於是 J 取極值的條件就是

$$\left. \frac{\partial J}{\partial \varepsilon} \right|_{\varepsilon=0} = 0 \quad (20)$$

根據(19)式所示的定義，(20)式可寫為

$$J'(0) = \int_{x_1}^{x_2} \left[\frac{\partial F}{\partial \hat{\varphi}} \eta + \frac{\partial F}{\partial \hat{\varphi}_x} \eta_x \right] dx = 0 \quad (21)$$

通常使用變分算符來演算比較方便。由(18)式可知，假如變動 $\eta(x)$ ，則在 x 處得到不同的 $\varphi(x)$ 。這相當於將 $\hat{\varphi}$ 變動了一個微量 $\varepsilon \eta(x)$ ，這個量可用

$$\delta \varphi \equiv \varepsilon \eta(x) \quad (22)$$

來表示。現在將(21)式兩邊乘以 ε ，並用(22)式，有

$$\varepsilon J'(0) = \int_{x_1}^{x_2} \left[\frac{\partial F}{\partial \hat{\varphi}} \delta \hat{\varphi} + \frac{\partial F}{\partial \hat{\varphi}_x} \delta \hat{\varphi}_x \right] dx = \int_{x_1}^{x_2} \delta F dx = \delta J \quad (23)$$

因為當 J 取極值時上式最左邊等於零，故我們發現

$$\delta J = 0 \quad (24)$$

這是使泛函 J 取極值的條件。

由上面的說明可知，擾動法把「對函數 $\varphi(x)$ 的微分」化為我們比較熟悉的對純量 ε 的微分，如(21)式所示。事實上在微積分學中，一個量對函數 $\varphi(x)$ 的微分是沒有定義的。

分析誤差協方差矩陣

4.2 節中已導出了分析誤差協方差矩陣的表達式如下：

$$\begin{aligned} \mathbf{A} &= \mathbf{B} - \mathbf{KHB} - \mathbf{BH}^T \mathbf{K}^T + \mathbf{K}(\mathbf{HBH}^T + \mathbf{R})\mathbf{K}^T \\ &= \mathbf{B} - \mathbf{KHB} - \mathbf{BH}^T \mathbf{K}^T + \mathbf{KDK}^T \end{aligned} \quad (25)$$

其中 $\mathbf{D} = \mathbf{HBH}^T + \mathbf{R}$ 表示新息協方差矩陣。(25)式是該節(27)式。在該

節中我們要找到最佳的增益矩陣 $\mathbf{K}$ ，使得分析誤差的總方差 $\mathcal{E}(\boldsymbol{\varepsilon}_a^T \boldsymbol{\varepsilon}_a) = \text{tr } \mathbf{A}$ 取極小值。這就是說，令

$$\frac{\partial}{\partial \mathbf{K}} \text{tr}(\mathbf{A}) = \mathbf{0} \quad (26)$$

在這裡我們使用變分法導出最佳增益矩陣 $\mathbf{K}$ 。

將(25)式右邊的增益矩陣取變分，有

$$\begin{aligned} \delta \mathbf{A} &= -\delta \mathbf{K} \mathbf{H} \mathbf{B} - \mathbf{B} \mathbf{H}^T \delta \mathbf{K}^T + \delta \mathbf{K} \mathbf{D} \mathbf{K}^T + \mathbf{K} \mathbf{D} \delta \mathbf{K}^T \\ &= \delta \mathbf{K} (-\mathbf{H} \mathbf{B} + \mathbf{D} \mathbf{K}^T) + (-\mathbf{B} \mathbf{H}^T + \mathbf{K} \mathbf{D}) \delta \mathbf{K}^T \end{aligned} \quad (27)$$

將上式取跡，得到

$$\delta \text{tr } \mathbf{A} = 2 \text{tr} [(-\mathbf{B} \mathbf{H}^T + \mathbf{K} \mathbf{D}) \delta \mathbf{K}^T] \quad (28)$$

導出上式時已用到跡的性質： $\text{tr } \mathbf{C} = \text{tr } \mathbf{C}^T$ 。令 $\delta \text{tr } \mathbf{A} = 0$ ，我們發現 $\mathbf{K} \mathbf{D} = \mathbf{B} \mathbf{H}^T$ ，即

$$\mathbf{K} = \mathbf{B} \mathbf{H}^T \mathbf{D}^{-1} = \mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1} \quad (29)$$

換句話說，使 $\text{tr } \mathbf{A}$ 取極小值的增益矩陣如(29)式所示。

這個變分法若要說得更清楚，可用下面的方法。令 $\hat{\mathbf{K}}$ 表示最佳增益矩陣， $\mathbf{E}$ 為任意矩陣，再令

$$\mathbf{K} = \hat{\mathbf{K}} + \varepsilon \mathbf{E} \quad (30)$$

其中 ε 為無限小正數。代入(25)式所示的 Joseph 公式，有

$$\mathbf{A} = \mathbf{B} - (\hat{\mathbf{K}} + \varepsilon \mathbf{E}) \mathbf{H} \mathbf{B} - \mathbf{B} \mathbf{H}^T (\hat{\mathbf{K}} + \varepsilon \mathbf{E})^T + (\hat{\mathbf{K}} + \varepsilon \mathbf{E}) \mathbf{D} (\hat{\mathbf{K}} + \varepsilon \mathbf{E})^T \quad (31)$$

上式改寫為

$$\mathbf{A} = \hat{\mathbf{A}} + \varepsilon (\mathbf{E} \mathbf{D} \hat{\mathbf{K}}^T + \hat{\mathbf{K}} \mathbf{D} \mathbf{E}^T - \mathbf{E} \mathbf{H} \mathbf{B} - \mathbf{B} \mathbf{H}^T \mathbf{E}^T) + \varepsilon^2 (\mathbf{E} \mathbf{D} \mathbf{E}^T) \quad (32)$$

其中 $\hat{\mathbf{A}}$ 表示最佳的分析誤差協方差矩陣，即

$$\hat{\mathbf{A}} = \mathbf{B} - \hat{\mathbf{K}}\mathbf{H}\mathbf{B} - \mathbf{B}\mathbf{H}^T\hat{\mathbf{K}}^T + \hat{\mathbf{K}}\mathbf{D}\hat{\mathbf{K}}^T \quad (33)$$

令

$$J(\varepsilon) = \text{tr } \mathbf{A} \quad (34)$$

則因為 $J(\varepsilon)$ 是 ε 的二次方程，故最佳性條件是

$$\left. \frac{dJ(\varepsilon)}{d\varepsilon} \right|_{\varepsilon=0} = 0 \quad (35)$$

將(32)式取跡後對 ε 微分，並令 $\varepsilon = 0$ ，得到

$$\left. \frac{dJ(\varepsilon)}{d\varepsilon} \right|_{\varepsilon=0} = 2 \text{tr}[(\hat{\mathbf{K}}\mathbf{D} - \mathbf{B}\mathbf{H}^T)\mathbf{E}^T] = 0 \quad (36)$$

由上式得到 $\hat{\mathbf{K}} = \mathbf{B}\mathbf{H}\mathbf{D}^{-1}$ ，因此最佳增益矩陣如(29)式所示。將(36)式兩邊乘以 ε ，並令 $\delta\mathbf{K} = \varepsilon\mathbf{E}$ ，(36)式變為

$$\varepsilon \left. \frac{dJ(\varepsilon)}{d\varepsilon} \right|_{\varepsilon=0} = \text{tr}[(\hat{\mathbf{K}}\mathbf{D} - \mathbf{B}\mathbf{H}^T)\delta\mathbf{K}^T] = 0 \quad (37)$$

上式就是(28)式。由上面的說明可知，擾動法把對矩陣 $\mathbf{K}$ 的微分，如(26)式所示，化為我們比較熟悉的對純量 ε 的微分，如(37)式所示。

總結的說，擾動法是變分法的基礎。在導出最小二乘法時，需要將純量對向量微分，這種微分可能我們比較不熟悉，但假如將最佳解加以擾動，如(11)式所示，就可將對向量的微分變為我們比較熟悉的對純量 ε 的微分，如(14)式所示。在求泛函的極值時，也是將最佳解加以擾動，如(18)式所示，從而求函數的極值變為求函數的極值，如(21)式所示。在求最佳的增益矩陣，使得分析誤差的總方差取極小值時，需要將純量對矩陣微分，如(26)式所示，但若使用擾動法，則對矩陣的微分化為我們較熟悉的對純量 ε 的微分，如(36)式所示。了解擾動法以後，就可直接使用變分法了。