第1章
反問題

1.1 前言

反問題在大氣科學中隨處可見，不只這樣，其他許多物理或數學問題都有反問題存在。反問題甚至在日常生活也經常碰到，例如由一個人的背影是否可認出是哪個人，或者由野獸的足跡是否可判斷出是何種動物。根據我們的經驗，的確是可以的，只要有夠多的資訊。我們很容易辨認出親密的人的背影，獵人也不難由野獸的足跡判斷出它們的種類，進而跟蹤它們。另一個反問題的例子是由某個學系大學入學考試的加權總分計算出原始總分，這個問題只要有其他的輔助知識也可輕易解決。

既然稱為反問題，當然有相應的正問題存在。如何觀察某人而熟悉他的背影，或者如何審視一只野獸而得知其足跡的特徵，這是上面提出的兩個反問題的正問題。當然由原始總分和權因子計算出加權總分，更是簡單的算術而已。一般說來，反問題較難求解，必須有其他的輔助知識才能處理。

正反問題不分的情況

正反問題的關係大致可分為三種情況。第一個情況是正反不分，例如乘法和除法（除了乘以0和除以0外），指數與對數等的關係，還

有四種 Fourier 變換和相應的逆變換、Hankel 變換及其逆變換等等，都屬於正反不分的情況。其他的變換，例如在 GPS（全球定位系統，Global Positioning System）衛星技術中出現的 Abel 變換：

$$\alpha(a) = -2a\int_a^{r_0} \frac{d\ln\mu}{dx}\frac{dx}{\sqrt{x^2-a^2}} \tag{1}$$

$$\ln\mu(x) = \frac{1}{\pi}\int_x^{r_0} \frac{\alpha(a)da}{\sqrt{a^2-x^2}} \tag{2}$$

其中 $\alpha(a)$ 為低軌衛星（low earth orbiting satellite, LEO）觀測到的 GPS 射線彎角（bending angle），μ 為大氣中的折射指數垂直分布，a 和 x 是一種垂直坐標，r_0 為大氣層頂到地心的距離。這個問題可說是正反不分，因為由 (1) 式可導出 (2) 式，反過來也可以，它們同樣是下限處具有奇異點的積分式。關於這個變換，我們將在第 11 章說明。

根據 Planck 定律，一個溫度為 T 的黑體放出的輻射強度（radiance）隨波長 λ 的分布可用下面的 Planck 函數表示：

$$B_\lambda(T) = \frac{2hc^2}{\lambda^5(e^{hc/\lambda KT}-1)} \tag{3}$$

其中 c, h, K 分別為光速、Boltzmann 常數和 Planck 常數。已知黑體的溫度，就可由 (3) 式計算出它放出的輻射強度隨波長的分布。反過來說，若已知一物體放出的輻射強度 I_λ，也可求出它的亮度溫度（brightness temperature）T_B。(3) 式中令

$$T = T_B, \qquad B_\lambda = I_\lambda$$

就可解出亮度溫度如下：

$$T_B = \frac{hc}{\lambda K\ln(1+2hc^2/\lambda^5 I_\lambda)} \tag{4}$$

(4) 式中 T_B 和 I_λ 之間的關係是正反不分的。

還有，飽和水汽壓 e_s 只和氣溫 T 有關，可寫為下面的函數形式：

$$e_s = f(T) \tag{5}$$

若實際水汽壓 e 是已知的，則可按 (5) 式求出露點溫度 T_d：

$$e = f(T_d) \tag{6}$$

假如 $f(T)$ 的函數形式採用 Tetens 公式：

$$f(T) = 6.1078\exp\left[\frac{a(T-273.16)}{T-b}\right] \tag{7}$$

其中水汽壓以 mb（即 hP）為單位，a 和 b 值在平水面上為

$$a = 17.2693882, \qquad b = 35.86\ \mathrm{K}$$

則正反問題是不分的，因為由已知的 $f(T)$ 值按 (7) 式可求出 T 來。不過一般都認為，由溫度求飽和水汽壓是正問題，由實際水汽壓求露點溫度則是反問題。

適定的反問題

第二個情況是正反問題差異相當明顯，但反問題是適定的（well-posed）。所謂適定這個詞的意思包括三個要件，第一解是存在的，第二解是唯一的，第三解是穩定的。這裡所謂穩定是指解連續的依賴於資料，換句話說資料改變一點點，解也只能改變一點點。

做為例子，考慮下面的線性代數方程組：

$$\mathbf{Ax} = \mathbf{b} \tag{8}$$

其中 A 為 $N\times N$ 矩陣，x 和 b 為 $N\times 1$ 向量。假設 (8) 式中矩陣 A 為已知，若要由已知的 x 求出 b，這是正問題；若要由已知的 b 求出 x，則是反問題。正問題只是矩陣乘法而已，運算簡單，反問題卻需求解代數方程組，運算比較複雜。因此，在這裏正反問題差異相當明顯。假如矩陣 A 的行列式不接近於零，則反問題是適定的，利用現有的套裝軟體很容易

求出解來。

接著考慮下面渦度ζ和流函數ψ之間的關係：

$$\nabla^2\psi \equiv \psi_{xx} + \psi_{yy} = \zeta \tag{9}$$

其中下標 x 和 y 分別表示關於 x 和 y 的偏導數，ψ 是 x 和 y 的連續可微函數。若要由已知的ψ決定出ζ，則只要進行微分就夠了，這是很簡單的正問題。若要由已知的ζ決定出ψ，則 (9) 式稱為 Poisson 方程，需要再給定輔助條件才有解，但這個解需用稍微複雜的計算機程式才能求出來，因此這是反問題。對反問題來說，(9) 式的定解條件必須是下面三個邊界條件之一：

(a) ψ 本身在邊界上給定，這稱為 Dirichlet 型邊界條件。

(b) ψ_n 在邊界上給定，其中下標 n 表示外法向導數，這稱為 Neumann 型邊界條件。

(c) $\psi_n + \alpha\psi$ 在邊界上給定，其中α是一個常數，這稱為混合型邊界條件。

若邊界條件是這三者之一，解才是適定的。

另一個例子是第一類 Volterra 積分方程：

$$g(y) = \int_a^y K(x, y) f(x) dx \tag{10}$$

其中核函數 $K(x,\ y)$ 是 x 和 y 的已知函數。若由已知的 $f(x)$ 求 $g(y)$，則 (10) 式是積分運算，這是正問題。若要由已知的 $g(y)$ 求出 $f(x)$，這是積分方程，不過由於上下限之一是 y 的函數，因此解通常是適定的。

最後的例子是求解跟 Wien 位移定律有關的非線性方程。Wien 位移定律說，黑體輻射強度最大處的波長和絕對溫度成反比。因此，將 (3) 式對溫度 T 微分，然後令結果等於零，得到

$$5 - x - 5e^{-x} = y \tag{11}$$

其中 $x=hc/\lambda KT$，$y=0$。假如要由已知 x 求出 y，這是簡單的正問題。若 $y=0$，就像在這裏的情況，要求出 x，則是反問題了，此時 (11) 式是非線性方程，需用疊代法求解，這是稍為複雜的反問題。這個問題的答案

是 x=4.9652，因此

$$\lambda_m = \frac{a}{T} \tag{12}$$

其中 a=0.2897834 cm K。

不適定的反問題

第三個情況是正反問題差異相當明顯，但反問題是不適定的。仍然考慮 (8) 式，如果方陣（方形矩陣）A 的行列式很接近於零，則反問題是不適定的。對氣溫垂直分布的紅外遙測來說，輻射傳遞方程經過線性化後可寫為 (8) 式，但此時 A 為 $L \times M$ 矩陣，x 和 b 分別為 $M \times 1$ 和 $L \times 1$ 向量，其中 L 和 M 分別為頻道和垂直層的個數。由於 L 和 M 通常不相等，因此在這情況下求解 (8) 式的第一個念頭是決定出最小二乘解，不過解仍是不適定的。

當 $\zeta = 0$ 時，(9) 式是 Laplace 方程 $\nabla^2 \psi = 0$，假設輔助條件不是邊界條件，而是下面的初始條件：

$$\psi(0, y) = 0, \qquad \psi_x(0, y) = \frac{1}{k} \sin ky \tag{13a, b}$$

其中 $k > 0$。很容易用變數分離法證明，這個問題的解是

$$\psi(x, y) = \frac{1}{k^2} \sinh kx \sin ky \tag{14}$$

當 k 由 0 增加到 ∞ 時，(13b) 所示的 $\psi_x(0, y)$ 慢慢趨近於零，但當 $x \neq 0$ 時 (14) 式所示的解並不趨近於零，而是趨近於無窮大。顯然的零是一個解，因此對這個輔助條件來說，解不是連續的依賴於資料，在這裏資料是指輔助條件。這是偏微分方程中不適定性的很有名的例子。用最通俗的話說，(9) 式所示的橢圓型偏微分方程，若以初始條件做為輔助條件，則解是不適定的。

接著考慮下面的第一類 Fredholm 積分方程：

$$g(y) = \int_a^b K(x, y) f(x) dx \tag{15}$$

其中 $b>a$。上式類似於 (10) 式，但上下限都是常數。對反問題來說，$f(x)$ 的解並不是適定的，因為若 $\tilde{f}(x)$ 是一個解，則下面的 $f(x)$ 也是解：

$$f(x)=\tilde{f}(x)+C\sin kx \tag{16}$$

其中 C 可以是任意值，波數 k 是很大的數。這是因為當核函數 $K(x,y)$ 為有界（bounded）時，根據 Riemann-Lebesgue 引理（lemma）（證明見曾 1988c 第 542 頁），有

$$\int_a^b K(x,y)\sin kx\,dx \to 0, \qquad \text{當 } k\to\infty \text{ 時} \tag{17}$$

(17) 式用文字說明就是，若被積項中有波數 k 很大的正弦函數的因子，不論從哪裏積到哪裏，由於正面積和負面積會互相抵消，故總面積會趨近於零。顯然的，由 (16) 式可以看出，解不是唯一的，也不是穩定的。

氣象學中的反問題

在氣象學中反問題隨處可見，但最重要的有下面幾類，將分別說明。

(a) 四維資料同化（data assimilation）和參數估計（parameter estimation）問題。四維資料同化是指由氣象變數觀測值的時間序列藉助動力模式決定出最佳的初始值，參數估計問題則是決定出動力方程中的參數值。對這兩個問題來說，相應的正問題則是數值預報，也就是由已知的初始條件和參數值，按動力模式預報氣象變數值。

(b) 客觀分析。這是將分布不規則的測站上觀測到的氣象變數值內插到分布規則的格點上，相應的正問題則是簡單的古典內插，也就是將格點上的分析值內插到測站上。由於測站分布非常不均勻，在海洋上、高山上、極區等地甚至沒有測站，因此反問題比較複雜。

(c) 溫溼垂直分布的衛星紅外遙測。這是指由衛星上的紅外輻射觀測決定氣溫和水汽垂直分布，相應的正問題是輻射傳遞模式，

也就是由已知的氣溫、水汽垂直分布以及其他氣候資料決定衛星上觀測到的輻射強度。

(d) 由 GPS 可降水量觀測決定水汽垂直分布。可降水量的定義為

$$u = \int_0^\infty \rho_v dz \tag{18}$$

這是單位垂直氣柱中水汽的總量，更正確的名稱是水汽積分量。若已知水汽密度 ρ_v 的垂直分布，很容易計算出可降水量。可是若可降水量為已知，就無法計算出水汽垂直分布了，這是由於無窮多個水汽分布相應於一個可降水量值，因此解不是唯一的。如果可降水量觀測足夠準確，還是很有用的，因為藉助水汽分布的背景值可得到最佳的分析值，這是資料同化的目的之一。

(e) 由 GPS 的折射率觀測決定氣象變數。根據 (2) 式，折射指數 μ(真空中光速和介質中光速的比值) 可由彎角分布 α 準確的決定出來。大氣中折射指數比 1 稍大，故使用另一變數 N 比較方便：

$$N = (\mu - 1) \times 10^6 \tag{19}$$

折射率 N 和氣象變數在微波區有下面的關係：

$$N = 77.6\frac{p}{T} + 3.73 \times 10^5 \frac{e}{T^2} \tag{20}$$

上式稱為折射率方程，p 和 e 分別為以 mb 為單位的氣壓和水汽壓，T 為溫度。由氣象變數據 (20) 式計算 N 是正問題，由 N 連同流體靜力方程計算氣象變數是反問題，因為後者少了一個方程。因此，反問題有無窮多個解，必須藉助氣象變數的背景值決定出適當的解。

令 x 為未知數，通常稱為狀態向量（state vector），其元素稱為狀態變數（state variables）；y 為可觀測量（observables），這是可被觀測到的量，下面將舉例說明。它們之間有下面的關係：

$$\mathbf{\Gamma}(\mathbf{x}, \mathbf{y}) = 0 \tag{21}$$

其中 y 和 x 分別為 $L \times 1$ 和 $M \times 1$ 向量，$\mathbf{\Gamma}$ 為 $K \times 1$ 矩陣，K 表示方程的

個數。在這裡暫時不考慮觀測誤差。可觀測量是指儀器觀測到的物理量，它不一定是狀態向量。舉例說，衛星輻射計各個測溫頻道觀測到的物理量是亮度溫度，並不是氣溫垂直分布，因此氣溫垂直分布是狀態向量，可觀測量則是亮度溫度。假如已有 y 的觀測值 $\mathbf{y}_o$，而且不計觀測誤差，則由上式有

$$\boldsymbol{\Gamma}(\mathbf{x}, \mathbf{y}_o) = 0 \tag{22}$$

我們的反問題是由已知的 $\mathbf{y}_o$ 決定出 x 來。例如對四維資料同化問題來說，x 是初始值，$\mathbf{y}_o$ 是氣象變數觀測值的時間序列，Γ 則是動力模式。不過 (22) 式並不容易求解，還需要其他的輔助條件。

在某些情況下，(21) 式可寫為顯函數的形式，稱為向前模式（forward model）：

$$\mathbf{y} = \mathbf{h}(\mathbf{x}) \tag{23}$$

其中 $\mathbf{h}$ 為非線性觀測算符（observation operator），也就是由狀態向量 x 計算可觀測量 y 的過程，它的維數和可觀測量一樣。對氣溫垂直分布的遙測來說，狀態向量 x 是氣溫垂直分布，可觀測量 y 是衛星輻射計測溫頻道可觀測到的亮度溫度，h 則是輻射傳遞模式。若 y 有觀測值 $\mathbf{y}_o$，且不計觀測誤差，則 (23) 式可改寫為

$$\mathbf{y}_o = \mathbf{h}(\mathbf{x}) \tag{24}$$

反問題就是由已知的 $\mathbf{y}_o$ 決定出 x。由於按 (24) 式求出來的解 x 通常是不適定的，故需要其他的輔助資料，例如 x 的背景值 x_b，才能決定出最佳的 x。在氣象學上所謂背景值是指氣候值、持續值或數值預報值。持續值是指 12 小時或 6 小時前的分析。現在數值預報技術飛躍進步，準確度提高不少，故通常以數值預報值當做背景值。

假如觀測算符是線性的，則有

$$\mathbf{y} = \mathbf{H}\mathbf{x} \tag{25}$$

其中 H 為 $L \times M$ 矩陣。對客觀分析問題來說，x 是格點上的變數值，y

是測站上的變數值，H 就是古典內插法了。若有 y 的觀測值 $\mathbf{y}_o$，而且不計觀測誤差，則 (25) 式變為

$$\mathbf{y}_o = \mathbf{H}\mathbf{x} \tag{26}$$

可而由 (26) 式解出 x：

$$\mathbf{x} = \mathbf{H}^{-1}\mathbf{y}_o \tag{27}$$

這是不可行的，因為 H 不是方陣（方形矩陣），故其逆矩陣是沒有定義的。在這情況下，雖然有最小二乘解，但解可能是不適定的，必須藉助背景值 x_b，才能使解穩定下來。

關於反問題，還有幾點必須注意：

(a) 在求解反問題時，必須先求解正問題，雖然正問題看起來較簡單，不過有時還是很複雜的，例如數值預報模式或輻射傳遞模式。

(b) 反問題可視為一種內插過程，例如前面提到的 Wien 定律有關的方程 (11) 式，令 x 以很小的間距由 0 增加到 100，這樣就可知道解在那個範圍內了，然後再進行內插，就可求出使 y=0 的 x 值來。因此，客觀分析問題才能解釋成反問題。

(c) 由於許多反問題是不適定的，解可能不是唯一的，或者不是穩定的，因此必須藉助輔助條件，如背景值等，以便使解穩定下來。

1.2 最小二乘法

求解物理問題時，經常需令某種誤差取極小值，以便得到最佳解。這個誤差當然指某種誤差的度量（measure），因為誤差通常是一個向量，也就是說它不只是一個數。最常用的誤差度量是誤差的平方和，即

$$J = \frac{1}{2}\sum_{k=1}^{K}\varepsilon_k^2 \tag{1}$$

其中 $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_K$ 是 K 個誤差。在上式中為方便起見，多寫了 1/2 的因

子。令 (1) 式取極小值而得到的解一般稱為最小二乘解（least square solution）。上式可改寫

$$J=\frac{1}{2}\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} \tag{2}$$

其中上標 T 表示轉置，$\boldsymbol{\varepsilon}$ 為 $K\times 1$ 單行向量：

$$\boldsymbol{\varepsilon}=[\varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_K]^T$$

(1) 式中已設誤差是無偏的（unbiased），也是不相關的（uncorrelated）。換句話說，

$$\mathcal{E}(\boldsymbol{\varepsilon})=\mathbf{0} \tag{3}$$

$$\mathbf{C}\equiv \mathrm{Cov}[\boldsymbol{\varepsilon}]\equiv \mathcal{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T)=\sigma^2\,\mathbf{I} \tag{4}$$

其中 $\mathcal{E}$ 表示期望值，C 稱為誤差的協方差矩陣（covariance matrix），I 則是單位矩陣。(3) 式表示誤差的期望值等於零，這就是所謂無偏。無偏的假設並不帶有限制性，因為一個隨機變數減去平均數後就變為無偏的隨機數。(4) 式則表示協方差矩陣是對角的，即各個誤差之間不相關，而且對角項，即各誤差的方差（variance），都等於 σ^2。(3) 和 (4) 式也可寫為

$$\mathcal{E}(\varepsilon_k)=0 \tag{5}$$

$$\mathcal{E}(\varepsilon_k\varepsilon_l)=\sigma^2\delta_{kl} \tag{6}$$

其中 δ_{kl} 為 Kronecker 符號，它的定義是

$$\delta_{kl}=1,\qquad k=l;\qquad \delta_{kl}=0,\qquad k\neq l \tag{7}$$

必須指出，$\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}$ 和 $\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T$ 是不同的，因為 $\boldsymbol{\varepsilon}$ 是 $K\times 1$ 矩陣，故 $\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}$ 是一個純量，即

$$\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}=\varepsilon_1^2+\varepsilon_2^2+\cdots+\varepsilon_K^2 \tag{8}$$

$\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T$ 則是一個 $K\times K$ 矩陣：

$$\boldsymbol{\varepsilon\varepsilon}^T = \begin{bmatrix} \varepsilon_1\varepsilon_1 & \varepsilon_1\varepsilon_2 & \cdots & \varepsilon_1\varepsilon_K \\ \varepsilon_2\varepsilon_1 & \varepsilon_2\varepsilon_2 & \cdots & \varepsilon_2\varepsilon_K \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_K\varepsilon_1 & \varepsilon_K\varepsilon_2 & \cdots & \varepsilon_K\varepsilon_K \end{bmatrix} \tag{9}$$

純量 $\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon}$ 就是 $K\times K$ 矩陣 $\boldsymbol{\varepsilon\varepsilon}^T$ 的跡(trace)，也就是對角項的和，寫為 $\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = \mathrm{tr}\,\boldsymbol{\varepsilon\varepsilon}^T$ 。

有時某些觀測比其他觀測不可靠，通常這是指各個誤差的方差 $\sigma_k^2 \equiv \mathcal{E}(\varepsilon_k^2)$ 並不相等。若考慮這個情況，則 (1) 式應改寫為

$$J = \frac{1}{2}\sum_{k=1}^{K}\sigma_k^{-2}\varepsilon_k^2 \tag{10}$$

這是因為 σ_k^{-2} 代表 ε_k 的準確度，方差 σ_k 越大，準確度越低，因此 (10) 式是加權平方和，權重為各個誤差的準確度，即方差的倒數。此時最小二乘解就是令加權平方和取極小值而得到的解。

最一般性的加權最小二乘法是令

$$J = \frac{1}{2}\boldsymbol{\varepsilon}^T\mathbf{C}^{-1}\boldsymbol{\varepsilon} \tag{11}$$

取極小值，其中 C 是誤差的協方差矩陣，這是將 (9) 式取平均數而得到的量。在這裡已做了誤差為高斯分布（Gaussian distribution）的假設。假如誤差之間不相關，但方差不一定相等，即協方差矩陣的第 (k, l) 個元素為 $C_{kl} = \sigma_k^2\delta_{kl}$，此時 C 是對角的，對角項表示方差，(11) 式退化成 (10) 式。更進一步說，假如方差都相等，即 $\mathbf{C} = \sigma^2\mathbf{I}$，此時 (11) 式變為

$$J = \frac{1}{2\sigma^2}\boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} \tag{12}$$

這個式子實質上和 (2) 式是相同的。透過線性變換，可把協方差矩陣 C 變為對角的，即 (11) 式一定可改寫為 (10) 式，故最一般性的加權最小二乘法是令 (11) 式取極小值，這將在後面說明。必須指出，若誤差的平均數不等於零，則誤差協方差矩陣的定義是

$$\mathbf{C} = \mathrm{Cov}[\boldsymbol{\varepsilon}] = \mathcal{E}\{[\boldsymbol{\varepsilon} - \mathcal{E}(\boldsymbol{\varepsilon})][\boldsymbol{\varepsilon} - \mathcal{E}(\boldsymbol{\varepsilon})]^T\} \tag{13}$$

將一個平均數為 0 的隨機向量和本身的外積取期望值，就是協方差矩陣。

協方差矩陣的性質

協方差矩陣有下面幾個重要性質：

(a) 協方差矩陣是實數且對稱的，即

$$C_{kl} = C_{lk} \tag{14}$$

其中 C_{kl} 表示協方差矩陣的第 (k,l) 個元素。這個性質可由定義得知。

(b) 協方差矩陣是正定的（positive definite）。也就是說，對元素不全為零的任何向量 $\boldsymbol{\beta}$ 來說，二次形（quadratic form）$\boldsymbol{\beta}^T\mathbf{C}\boldsymbol{\beta}$ 都大於 0，即

$$\boldsymbol{\beta}^T\mathbf{C}\boldsymbol{\beta} > 0 \tag{15}$$

這個性質可用下面的方法證明。設 f 是各個誤差 ε_1, $\varepsilon_2, \cdots, \varepsilon_K$ 的線性組合，即

$$f = \sum_{k=1}^{K} \beta_k \varepsilon_k, \qquad \mathcal{E}(f) = \sum_{k=1}^{K} \beta_k \mathcal{E}(\varepsilon_k) = 0$$

其中 β_k 為向量 $\boldsymbol{\beta}$ 的第 k 個元素，上式中已設誤差向量是無偏的。f 的方差為

$$\begin{aligned}\mathcal{E}\{[f - \mathcal{E}(f)]^2\} &= \mathcal{E}(\sum_{k=1}^{K} \beta_k \varepsilon_k \sum_{l=1}^{K} \beta_l \varepsilon_l) = \sum_{k=1}^{K}\sum_{l=1}^{K} \beta_k \beta_l \mathcal{E}(\varepsilon_k \varepsilon_l) \\ &= \sum_{k=1}^{K}\sum_{l=1}^{K} \beta_k C_{kl} \beta_l = \boldsymbol{\beta}^T\mathbf{C}\boldsymbol{\beta}\end{aligned}$$

因為方差一定大於零，故 $\boldsymbol{\beta}^T\mathbf{C}\boldsymbol{\beta} > 0$。

(c) 實數對稱矩陣 C 的特徵值一定是實數。證明請見附錄 A1.1 節。下面將證明，若 C 也是正定的，則特徵值一定是正數。設 λ_k 和 $\mathbf{u}_k$

分別為協方差矩陣 C 的特徵值和相應的特徵向量，即

$$\mathbf{C}\mathbf{u}_k = \lambda_k \mathbf{u}_k, \qquad k = 1, 2, \cdots, K \tag{16}$$

將 (16) 式兩邊前乘（pre-multiplied）以 $\mathbf{u}_k^T$ ，有

$$\lambda_k = \frac{\mathbf{u}_k^T \mathbf{C}\,\mathbf{u}_k}{\mathbf{u}_k^T \mathbf{u}_k}$$

上式中分母是正值，又因為 C 是正定的，故分子也是正值，從而上式右邊一定是正值，這樣就證明了 C 的特徵值一定是正數。

必須指出，在這裡提到的協方差矩陣，是指同一向量的協方差矩陣，正確的說法應該是自協方差矩陣（auto-covariance matrix）。不同向量的協方差矩陣，稱為互協方差矩陣（cross-covariance matrix），沒有上面提到的性質。

在實際使用時，誤差有不同的定義，完全取決於物理問題。下面將說明做為本書重點的極小方差估計值和變分資料同化。在說明這兩個重點以前，先考慮最簡單的迴歸問題。

迴歸法

做為最小二乘法的例子，首先討論線性迴歸問題。設物理量 y 和另一物理量 x 有下面的線性關係：

$$y = a + bx \tag{17}$$

現在有大量的 x 和 y 的統計樣本 x_k, y_k，我們要由這些觀測值決定出最佳的迴歸係數 a 和 b。對於某個樣本 (x_k, y_k) 來說，(17) 式不會完全滿足，一定會有誤差，即

$$y_k = a + bx_k + \varepsilon_k, \qquad k = 1, 2, \cdots, K \tag{18}$$

其中 K 為樣本個數。對 (18) 式取樣本平均，有

$$\bar{y} = a + b\bar{x} \tag{19}$$

其中上橫線表示樣本平均，舉例說

$$\bar{y} = \frac{1}{K}\sum_{k=1}^{K} y_k \tag{20}$$

(18) 式減去 (19) 式，得到

$$y_k - \bar{y} = b(x_k - \bar{x}) + \varepsilon_k \tag{21}$$

現在要決定最佳的迴歸係數 b，使得 (1) 式所示的誤差平方和取極小值：

$$J = \frac{1}{2}\sum_k \varepsilon_k^2 = \frac{1}{2}\sum_k [y_k - \bar{y} - b(x_k - \bar{x})]^2 \tag{22}$$

將 (22) 式對 b 微分，並令結果等於零，得到

$$\sum_k [y_k - \bar{y} - b(x_k - \bar{x})](x_k - \bar{x}) = 0$$

由上式解出 b，有

$$b = \frac{\sum_k (x_k - \bar{x})(y_k - \bar{y})}{\sum_k (x_k - \bar{x})^2} \tag{23}$$

另一個迴歸係數 a 按 (19) 式決定出來。

必須指出，在實際工作時，(18) 式中自變數 x_k 的間距 $\Delta x = x_{k+1} - x_k$ 必須相等，否則擬合出來的直線將會是錯誤的。因此，假如不相等的話，必須對統計樣本 (x_k, y_k) 做事先處理。先將 x 軸分為許多寬度相等的間距，然後把這個間距內的所有 y_k 做算術平均，當做這個間距中點的 y 樣本。重新定義的樣本是各間距中點的 x 值和這個間距內的 y_k 值的算術平均。

考慮一個實際應用的例子。將風杯風速計 (cup anemometer) 放進環境風速為 v_e 的風洞中，則風速計量出的風速 v 滿足下面的風杯方程：

$$\tau v_t + v = v_e \tag{24}$$

其中下標 t 表示關於時間 t 的導數，τ 為時間常數。現在設有 $(N+1)$ 個觀測值：

$$v_n = v(t_n), \qquad n = 0, 1, 2, \cdots, N \tag{25}$$

上式中 n 表示時間指標。我們要決定出最佳的 τ 和 v_e 值。這個問題可用迴歸法求解。令

$$y = v, \qquad x = v_t \tag{26}$$

則 (24) 式可寫為

$$y = a + bx \tag{27}$$

其中

$$a = v_e, \qquad b = -\tau \tag{28}$$

迴歸係數 a, b 可由 x 和 y 的觀測值決定出來，然後再按 (28) 就可決定出周極方程 (24) 式中的兩個參數 τ, v_e。x 和 y 的樣本值按下式估計：

$$x \equiv v_t \cong \frac{v_n - v_{n-1}}{\Delta t}, \qquad y \equiv v \cong \frac{v_n + v_{n-1}}{2} \qquad (n = 1, 2, \cdots, N) \tag{29}$$

其中 Δt 是時間步長（time step）。這個問題是由動力方程和大量的觀測值決定出動力方程中的參數值，稱為參數估計（parameter estimation）問題。

迴歸法另一個值得提的應用是海面溫度遙測。氣象衛星上攜帶的紅外輻射計 11 μm 和 12 μm 窗區頻道量出的亮度溫度對海面溫度變化比較敏感，故可用來決定這個物理量的值。首先建立海面溫度 T_s 和這兩個頻道亮度溫度 T_{11}, T_{12} 之間的線性關係：

$$T_s = a + bT_{11} + cT_{12} \tag{30}$$

現在已知某時某地的海面溫度，再由衛星資料中找出同一時刻輻射計正好觀測到該地點時的兩個頻道的亮度溫度，這樣由全球全年各時刻的資料就可建立大量的統計樣本，由此可決定出 (30) 式中的迴歸係數 a, b, c。

為了求解這個問題，考慮下面的多元情況：

$$y = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_N x_N \tag{31}$$

對某個樣本來說，上式不會完全滿足，一定有誤差 ε_k：

$$y_k = \beta_1 x_{k1} + \beta_2 x_{k2} + \cdots + \beta_N x_{kN} + \varepsilon_k = \sum_{n=1}^{N} x_{kn}\beta_n + \varepsilon_k \tag{32a}$$

上式中 x 的第一個下標表示不同的樣本，第二個下標表示不同的變數。若寫成矩陣形式，就是

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{32b}$$

其中 $\mathbf{X} = [x_{kn}]$，它是 $K \times N$ 矩陣。現在要決定最佳的 $\boldsymbol{\beta}$，使下面誤差的平方和取極小值：

$$J = \frac{1}{2}\sum_{k=1}^{K}\varepsilon_k^2 = \frac{1}{2}\sum_{k=1}^{K}\left(y_k - \sum_{n=1}^{N} x_{kn}\beta_n\right)^2 \tag{33}$$

將上式對 β_m 微分，並令結果等於零，得到

$$\sum_{k=1}^{K}\left(y_k - \sum_{n=1}^{N} x_{kn}\beta_n\right)\sum_{n=1}^{N} x_{kn}\delta_{mn} = 0, \qquad m = 1, 2, \cdots, N \tag{34}$$

其中 $\partial\beta_n / \partial\beta_m$ 已用 Kronecker 符號 δ_{mn} 表示。(34) 式可進一步簡化為

$$\sum_{k=1}^{K}\sum_{n=1}^{N} x_{km} x_{kn}\beta_n = \sum_{k=1}^{K} x_{km} y_k \tag{35}$$

在導出上式時，已用到下面的 Kronecker 符號的性質：

$$\sum_{n=1}^{N} x_{kn}\delta_{mn} = x_{km} \tag{36}$$

上式中 n 由 1 增加到 N，但只有當 $n = m$ 時，δ_{mn} 才不會等於零，故 x_{kn} 要改為 x_{km}。(35) 式寫成矩陣形式就是

$$\mathbf{X}^T\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^T\mathbf{y} \tag{37}$$

因此，$\boldsymbol{\beta}$ 的最小二乘解為

$$\boldsymbol{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{38}$$

(38) 式的導出雖然要費一番功夫，但結果很簡單。這相當於省略 (32b) 式中的誤差項 $\boldsymbol{\varepsilon}$，然後兩邊前乘以 $\mathbf{X}^T$。若要將 (38) 式所示的結果應用於 (30) 式所示的迴歸模式，只要令

$$x_{k1}=1,\quad x_{k2}=(T_{11})_k,\quad x_{k3}=(T_{12})_k,\quad y_k=(T_s)_k,\quad N=3$$

代入 (38) 式，求出 $\boldsymbol{\beta}$ 後，迴歸係數值就是

$$a=\beta_1,\qquad b=\beta_2,\qquad c=\beta_3$$

極小方差解

接著考慮由某個向量 x 的兩種觀測決定最佳分析值的問題。這個向量以後都稱為狀態向量，它描述一個系統的狀態。對大氣系統來說，這些是指氣溫、氣壓、溼度等氣象變數。這兩個觀測值是實際觀測值和背景值，分別用 $M\times 1$ 向量 $\mathbf{x}_o$ 和 $\mathbf{x}_b$ 表示。前面已說過，背景值是指某一變數的經驗值，在氣象學通常指氣候值、持續值或數值預報值，通常使用較為準確的預報值當做背景值。定義實際觀測誤差（measurement error）（通常稱為量測誤差）和背景誤差（background error）分別為

$$\boldsymbol{\varepsilon}_o=\mathbf{x}_o-\mathbf{x}^{\mathrm{t}},\qquad \boldsymbol{\varepsilon}_b=\mathbf{x}_b-\mathbf{x}^{\mathrm{t}} \tag{39}$$

其中上標 t 表示真值。這兩個誤差都是 $M\times 1$ 向量，假設它們是無偏的，即它們的平均數等於零：

$$\mathcal{E}(\boldsymbol{\varepsilon}_o)=\mathbf{0},\qquad \mathcal{E}(\boldsymbol{\varepsilon}_b)=\mathbf{0} \tag{40}$$

這個假設表示觀測值和背景值的平均數都等於真值 $\mathbf{x}^{\mathrm{t}}$，即

$$\mathcal{E}(\mathbf{x}_o)=\mathbf{x}^{\mathrm{t}},\qquad \mathcal{E}(\mathbf{x}_b)=\mathbf{x}^{\mathrm{t}} \tag{41}$$

現在令分析值 $\mathbf{x}_a$ 是觀測值 $\mathbf{x}_o$ 和背景值 $\mathbf{x}_b$ 的線性組合，即

$$\mathbf{x}_a-\mathbf{x}_b=\mathbf{K}(\mathbf{x}_o-\mathbf{x}_b) \tag{42}$$

其中 $(\mathbf{x}_a-\mathbf{x}_b)$ 稱為分析增量，$(\mathbf{x}_o-\mathbf{x}_b)$ 稱為觀測增量，K 是待求的 $M\times M$ 增益矩陣（gain matrix）。(42) 式可改寫為

$$\boldsymbol{\varepsilon}_a=\boldsymbol{\varepsilon}_b+\mathbf{K}(\boldsymbol{\varepsilon}_o-\boldsymbol{\varepsilon}_b) \tag{43}$$

其中 $\boldsymbol{\varepsilon}_a=\mathbf{x}_a-\mathbf{x}^{\mathrm{t}}$ 是分析誤差。顯然的，(42) 式所示的線性組合會使得分

析誤差也是無偏的。最佳的 K 可令下面的分析誤差的總方差取極小值而得到：

$$J = \mathcal{E}[\boldsymbol{\varepsilon}_a^T \boldsymbol{\varepsilon}_a] \tag{44}$$

將 (43) 式代入 (44) 式，再對 K 微分，然後令結果等於零，就可得到最佳的 K 值。由 (44) 式可知，這個極小方差估計值（minimum variance estimate）是求解最小二乘問題而得到的，將在第四章中詳細說明。

變分資料同化

接著說明變分資料同化。仍然考慮由狀態向量 $\mathbf{x}$ 的觀測值 $\mathbf{x}_o$ 和背景值 $\mathbf{x}_b$ 決定最佳分析值的問題。因為背景值及其統計量（通常指平均數和協方差矩陣）設為已知的，故可視為另一種觀測。定義一個新的 $2M\times 1$ 誤差向量 $\tilde{\boldsymbol{\varepsilon}}$ 如下：

$$\tilde{\boldsymbol{\varepsilon}} = \begin{bmatrix} \boldsymbol{\varepsilon}_o \\ \boldsymbol{\varepsilon}_b \end{bmatrix} = \begin{bmatrix} \mathbf{x}_o - \mathbf{x} \\ \mathbf{x}_b - \mathbf{x} \end{bmatrix} \tag{45}$$

根據最小二乘法，必須令下面的函數取極小值：

$$J = \frac{1}{2}\tilde{\boldsymbol{\varepsilon}}^T \tilde{\mathbf{C}}^{-1} \tilde{\boldsymbol{\varepsilon}} \tag{46}$$

其中 $2M\times 2M$ 矩陣 $\tilde{\mathbf{C}} = \mathcal{E}[\tilde{\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\varepsilon}}^T]$ 是誤差向量 $\tilde{\boldsymbol{\varepsilon}}$ 的協方差矩陣。首先考慮外乘積 $\tilde{\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\varepsilon}}^T$：

$$\tilde{\boldsymbol{\varepsilon}}\tilde{\boldsymbol{\varepsilon}}^T = \begin{bmatrix} \boldsymbol{\varepsilon}_o \\ \boldsymbol{\varepsilon}_b \end{bmatrix} [\boldsymbol{\varepsilon}_o^T \quad \boldsymbol{\varepsilon}_b^T] = \begin{bmatrix} \boldsymbol{\varepsilon}_o\boldsymbol{\varepsilon}_o^T & \boldsymbol{\varepsilon}_o\boldsymbol{\varepsilon}_b^T \\ \boldsymbol{\varepsilon}_b\boldsymbol{\varepsilon}_o^T & \boldsymbol{\varepsilon}_b\boldsymbol{\varepsilon}_b^T \end{bmatrix}$$

將上式取期望值，有

$$\tilde{\mathbf{C}} = \begin{bmatrix} \mathcal{E}[\boldsymbol{\varepsilon}_o\boldsymbol{\varepsilon}_o^T] & \mathcal{E}[\boldsymbol{\varepsilon}_o\boldsymbol{\varepsilon}_b^T] \\ \mathcal{E}[\boldsymbol{\varepsilon}_b\boldsymbol{\varepsilon}_o^T] & \mathcal{E}[\boldsymbol{\varepsilon}_b\boldsymbol{\varepsilon}_b^T] \end{bmatrix} = \begin{bmatrix} \mathbf{O} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \tag{47}$$

其中 $M\times M$ 矩陣 O 和 B 分別為觀測誤差和背景誤差的協方差矩陣：

$$\mathbf{O} = \mathcal{E}[\boldsymbol{\varepsilon}_o\boldsymbol{\varepsilon}_o^T], \qquad \mathbf{B} = \mathcal{E}[\boldsymbol{\varepsilon}_b\boldsymbol{\varepsilon}_b^T]$$

(47) 式中已設觀測誤差和背景誤差不相關，即它們的互協方差矩陣等於零：

$$\mathcal{E}[\boldsymbol{\varepsilon}_o\boldsymbol{\varepsilon}_b^T] = \mathbf{0}, \qquad \mathcal{E}[\boldsymbol{\varepsilon}_b\boldsymbol{\varepsilon}_o^T] = \mathbf{0}$$

在這情況下，(46) 式可寫為

$$J = \frac{1}{2}[\boldsymbol{\varepsilon}_o^T \quad \boldsymbol{\varepsilon}_b^T]\begin{bmatrix} \mathbf{O}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}^{-1} \end{bmatrix}\begin{bmatrix} \boldsymbol{\varepsilon}_o \\ \boldsymbol{\varepsilon}_b \end{bmatrix} = \frac{1}{2}\boldsymbol{\varepsilon}_o^T\mathbf{O}^{-1}\boldsymbol{\varepsilon}_o + \frac{1}{2}\boldsymbol{\varepsilon}_b^T\mathbf{B}^{-1}\boldsymbol{\varepsilon}_b$$

將 (45) 式代回上式，我們發現

$$J = \frac{1}{2}(\mathbf{x}_o - \mathbf{x})^T\mathbf{O}^{-1}(\mathbf{x}_o - \mathbf{x}) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) \qquad (48)$$

最佳的 x 值，即分析值，可令 (48) 式所示的函數取極小值而得到。這個函數稱為目標函數(objective function)或代價函數(cost function)。

1.1 節中已說過，可觀測量 y 並不一定是狀態向量 x，它們之間有下面的關係：

$$\boldsymbol{\Gamma}(\mathbf{x}, \mathbf{y}) = 0 \qquad (49)$$

這個關係通常是物理模式。此時 (48) 式應改寫為

$$J = \frac{1}{2}(\mathbf{y}_o - \mathbf{y})^T\,\mathbf{O}^{-1}(\mathbf{y}_o - \mathbf{y}) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T\,\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) \qquad (50)$$

其中 O 為觀測誤差的協方差矩陣：

$$\mathbf{O} = \mathrm{Cov}[\mathbf{y} - \mathbf{y}_o] = \mathrm{Cov}[\boldsymbol{\varepsilon}_o] = \mathcal{E}(\boldsymbol{\varepsilon}_o\boldsymbol{\varepsilon}_o^T)$$

在這情況下，在令目標函數 J 取極小值的同時，可觀測量 y 和狀態向量 x 之間必須滿足 (49) 式所示的約束條件 (constraint)，這是約束極小化的問題，將在 1.4 節中說明。設 x 和 y 分別為 $M\times 1$ 和 $L\times 1$ 矩陣，則 B 和 O 分別是 $M\times M$ 和 $L\times L$ 矩陣，$\boldsymbol{\Gamma}$ 則是一個向量。

在大部分情況下，(49) 式可寫為顯函數的形式：

$$\mathbf{y} = \mathbf{h}(\mathbf{x}) \qquad (51)$$

向前模式 h 是由狀態向量 x 到可觀測量 y 的變換。向前模式一定有誤差，故(51)式應改寫為

$$\mathbf{y} = \mathbf{h}(\mathbf{x}) + \boldsymbol{\varepsilon}_F \tag{52}$$

上式中 L×1 向量 $\boldsymbol{\varepsilon}_F$ 是向前模式誤差。在這情況下，

$$\mathbf{y}_o - \mathbf{h}(\mathbf{x}) = \mathbf{y}_o - \mathbf{y} + \mathbf{y} - \mathbf{h}(\mathbf{x}) = \boldsymbol{\varepsilon}_o + \boldsymbol{\varepsilon}_F \equiv \boldsymbol{\varepsilon}_r$$

上式稱為觀測的隨機模式，其中觀測誤差 $\boldsymbol{\varepsilon}_r$ 是實際觀測誤差 $\boldsymbol{\varepsilon}_o$ 和向前模式誤差 $\boldsymbol{\varepsilon}_F$ 的和（為了區別起見，我們把 $\boldsymbol{\varepsilon}_r$ 稱為觀測誤差，把 $\boldsymbol{\varepsilon}_o$ 稱為實際觀測誤差）。此時(50)式變為下面最常見的三維變分同化的目標函數：

$$J = \frac{1}{2}[\mathbf{y}_o - \mathbf{h}(\mathbf{x})]^T \mathbf{R}^{-1}[\mathbf{y}_o - \mathbf{h}(\mathbf{x})] + \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) \tag{53}$$

其中 L×L 矩陣 R 是觀測誤差的協方差矩陣：

$$\mathbf{R} = \mathrm{Cov}[\boldsymbol{\varepsilon}_r] = \mathrm{Cov}[\mathbf{y}_o - \mathbf{h}(\mathbf{x})] = \mathcal{E}(\boldsymbol{\varepsilon}_r \boldsymbol{\varepsilon}_r^T) \tag{54}$$

這個協方差矩陣可寫為

$$\begin{aligned}\mathbf{R} &= \mathrm{Cov}[\boldsymbol{\varepsilon}_o + \boldsymbol{\varepsilon}_F] = \mathcal{E}[(\boldsymbol{\varepsilon}_o + \boldsymbol{\varepsilon}_F)(\boldsymbol{\varepsilon}_o + \boldsymbol{\varepsilon}_F)^T] \\ &= \mathcal{E}(\boldsymbol{\varepsilon}_o \boldsymbol{\varepsilon}_o^T + \boldsymbol{\varepsilon}_o \boldsymbol{\varepsilon}_F^T + \boldsymbol{\varepsilon}_F \boldsymbol{\varepsilon}_o^T + \boldsymbol{\varepsilon}_F \boldsymbol{\varepsilon}_F^T) = \mathbf{O} + \mathbf{F}\end{aligned} \tag{55}$$

上式中 L×L 矩陣 O 和 F 分別為實際觀測誤差和向前模式誤差的協方差矩陣：

$$\mathbf{O} = \mathrm{Cov}[\boldsymbol{\varepsilon}_o], \qquad \mathbf{F} = \mathrm{Cov}[\boldsymbol{\varepsilon}_F] \tag{56}$$

在導出(55)式的過程中已假設實際觀測誤差和向前模式誤差是不相關的，也就是說它們的互協方差矩陣等於零：

$$\mathcal{E}(\boldsymbol{\varepsilon}_o \boldsymbol{\varepsilon}_F^T) = \mathbf{0}, \qquad \mathcal{E}(\boldsymbol{\varepsilon}_F \boldsymbol{\varepsilon}_o^T) = \mathbf{0}$$

(53)式很容易導出，只要令

$$\tilde{\boldsymbol{\varepsilon}} = \begin{bmatrix} \boldsymbol{\varepsilon}_r \\ \boldsymbol{\varepsilon}_b \end{bmatrix} = \begin{bmatrix} \mathbf{y}_o - \mathbf{h}(\mathbf{x}) \\ \mathbf{x}_b - \mathbf{x} \end{bmatrix}$$

然後代入 (46) 式，再做類似的假設就可得到。

假如向前模式是線性的，即 $\mathbf{y}=\mathbf{H}\mathbf{x}+\boldsymbol{\varepsilon}_r$，則目標函數變為

$$J=\frac{1}{2}(\mathbf{y}_o-\mathbf{H}\mathbf{x})^T\mathbf{R}^{-1}(\mathbf{y}_o-\mathbf{H}\mathbf{x})+\frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) \tag{57}$$

其中 $L\times M$ 矩陣 H 是線性觀測算符。

假如現在有 K_m 種不同的觀測系統，而且不同觀測系統的觀測誤差都是不相關的，則 (53) 式應改寫為

$$J=\frac{1}{2}\sum_{k=1}^{K_m}[\mathbf{y}_k^o-\mathbf{h}_k(\mathbf{x})]^T\mathbf{R}_k^{-1}[\mathbf{y}_k^o-\mathbf{h}_k(\mathbf{x})]+\frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) \tag{58}$$

其中下標 k 是表示不同觀測系統的指標，這些可能是無線電探空儀觀測到的常規氣象資料、氣象衛星攜帶的輻射計觀測到的亮度溫度、GPS 掩星觀測得到的彎角等等。(58) 也很容易導出，只要令

$$\tilde{\boldsymbol{\varepsilon}}=[(\boldsymbol{\varepsilon}_1^r)^T\quad(\boldsymbol{\varepsilon}_2^r)^T\quad\cdots\quad(\boldsymbol{\varepsilon}_{K_m}^r)^T\quad(\boldsymbol{\varepsilon}_b)^T]^T$$

然後代入 (46) 式，再做類似假設就可得到。這些假設是

$$\mathcal{E}[\boldsymbol{\varepsilon}_k^r(\boldsymbol{\varepsilon}_{k'}^r)^T]=\mathbf{R}_k\delta_{kk'},\qquad\mathcal{E}[\boldsymbol{\varepsilon}_k^r(\boldsymbol{\varepsilon}_b)^T]=\mathbf{0}$$

換句話說，不同的觀測系統是不相關的，而且各種觀測系統的誤差都和背景誤差不相關。

必須指出，若要由狀態向量 x 的觀測值和背景值決定最佳分析值，可令 (48) 式所示的目標函數取極小值而得到，不過這並不是求解反問題，因為沒用到向前模式。至於令 (53) 式和 (57) 式所示的目標函數取極小值，則是分別求解反問題 $\mathbf{y}_o=\mathbf{h}(\mathbf{x})$ 和 $\mathbf{y}_o=\mathbf{H}\mathbf{x}$ 的最重要方法，這是因為反問題是不適定的，需要借助背景值才能求出解來。

四維變分同化

對四維變分同化問題來說，目標函數仍然可寫為 (53) 式，但最好特別強調時間維，即

$$J = \frac{1}{2}\sum_{n=0}^{N}[\mathbf{y}_n^o - \mathbf{h}_n(\mathbf{x}_n)]^T \mathbf{R}_n^{-1}[\mathbf{y}_n^o - \mathbf{h}_n(\mathbf{x}_n)] + \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{B}_0^{-1}(\mathbf{x}_0 - \mathbf{x}_b) \quad (59)$$

其中下標 n 是時間指標，$\mathbf{x}_0$ 和 $\mathbf{x}_b$ 分別表示初始時刻的狀態向量及其背景值，$\mathbf{B}_0$ 表示初始時刻的背景誤差協方差矩陣。$\mathbf{h}_n(\mathbf{x}_n)$ 表示觀測算符，也就是從狀態向量 $\mathbf{x}_n$ 到可觀測量 $\mathbf{y}_n$ 的變換：

$$\mathbf{y}_n = \mathbf{h}_n(\mathbf{x}_n) + \boldsymbol{\varepsilon}_n^F \quad (60)$$

上式中 $\boldsymbol{\varepsilon}_n^F$ 表示向前模式誤差。觀測算符顯然會隨時間而不同，舉例說對衛星資料來說觀測點一定隨時間而不同，觀測算符自然也就不同了。嚴格的說，$\mathbf{x}_b$ 是上次的分析值，即 t_0 時的分析值。至於 t_n 時的狀態向量 $\mathbf{x}_n$ 則由下面的動力方程預報得到：

$$\mathbf{x}_n = \mathcal{M}_{n,n-1}(\mathbf{x}_{n-1}), \qquad n = 1, 2, \cdots, N \quad (61)$$

其中 $\mathbf{x}_n$ 是（$N+1$）個 M 維向量，其元素有 M 個，它表示時間點 t_n（$n = 0, 1, \cdots, N$）的狀態向量。在目前的業務數值預報模式中，M 的數量級大約為 10^6。至於 $\mathcal{M}_{n,n-1}$，則是作用在 t_{n-1} 時狀態向量 $\mathbf{x}_{n-1}$ 以便得到 t_n 時狀態向量 $\mathbf{x}_n$ 的預報算符，可能是非線性的。對線性預報模式來說，(61) 式可寫為

$$\mathbf{x}_n = \mathbf{M}_{n-1}\mathbf{x}_{n-1} + \mathbf{f}_{n-1} \quad (62)$$

$M \times M$ 矩陣 $\mathbf{M}_{n-1}$ 稱為轉移矩陣（transition matrix），$\mathbf{f}_{n-1}$ 為強迫項。

(59) 式中已假設觀測誤差和背景誤差不相關，而且也假設觀測誤差 $\boldsymbol{\varepsilon}_n^r = \boldsymbol{\varepsilon}_n^o + \boldsymbol{\varepsilon}_n^F$ 是白噪聲（white noise），也就是說不同時刻的觀測誤差是不相關的，即

$$E[\boldsymbol{\varepsilon}_n^r(\boldsymbol{\varepsilon}_{n'}^r)^T] = \mathbf{R}_n \delta_{nn'}$$

其中 $\delta_{nn'}$ 為 Kronecker 符號。

若可觀測量不是狀態向量本身，相應的分析值稱為極大後驗機率估計值（maximum a posteriori probability estimate）。必須指出，由於對稱正定矩陣的逆（inverse）也是對稱正定矩陣，故這些目標函數中的各項都是正值。令上面所說的目標函數取極小值而決定出最佳估

計值的方法在氣象學上都稱為變分資料同化（variational data assimilation），包括一維變分反演（inversion）、三維變分分析和四維變分同化，這些都是本書的重點，將在後面各章中詳細說明。

上面以最小二乘法簡單的介紹了極小方差估計值和極大後驗機率估計值，並且導出了變分資料同化使用的目標函數，但理論並不周全，更嚴謹的方法必須建立在機率理論上。簡單的說，極小方差解相當於條件平均數（conditional mean），極大後驗機率估計值則相當於條件眾數（conditional mode），這些將在 4.4 節中說明。

矩陣的對角化

通過線性變換，可將協方差矩陣變換為對角矩陣，那麼就可了解為什麼 (11) 式是最一般性的加權最小二乘法的目標函數。令協方差矩陣 $\mathbf{C}$ 的特徵為 λ_k，相應的特徵向量為 $\mathbf{u}_k$，如 (16) 式所示。因為 C 是實數對稱矩陣，故其特徵向量是正交的，即

$$\mathbf{u}_k^T \mathbf{u}_l = \delta_{kl} \tag{63}$$

在上式中已將特徵向量標準化了，也就是說當 $k=l$ 時它們的長度等於 1。現在令

$$\mathbf{U} = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_K], \qquad \mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_K) \tag{64}$$

U 是特徵向量組成的方形矩陣，$\mathbf{\Lambda}$ 為對角矩陣，對角項為特徵值。利用 (64) 式，則 (16) 和 (63) 式分別寫為

$$\mathbf{CU} = \mathbf{U\Lambda}, \qquad \mathbf{U}^T\mathbf{U} = \mathbf{I} \tag{65a, b}$$

利用 (65b) 式，則由 (65a) 式有

$$\mathbf{C} = \mathbf{U\Lambda U}^T \quad 或 \quad \mathbf{C}^{-1} = \mathbf{U\Lambda}^{-1}\mathbf{U}^T \tag{66a, b}$$

(66) 式中

$$\mathbf{\Lambda}^{-1} = \mathrm{diag}(\lambda_1^{-1}, \lambda_2^{-1}, \cdots, \lambda_K^{-1})$$

因為協方差矩陣是對稱正定的，故其特徵值都大於零，從而逆矩陣 $\mathbf{\Lambda}^{-1}$

是存在的。(66a) 式稱為對稱正定矩陣 C 的特徵值分解（eigenvalue decomposition）。

另一方面，利用 (66b) 式，(11) 式可改寫為

$$J = \frac{1}{2}\boldsymbol{\varepsilon}^T \mathbf{C}^{-1} \boldsymbol{\varepsilon} = \frac{1}{2}\boldsymbol{\varepsilon}^T \mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}^T \boldsymbol{\varepsilon} = \frac{1}{2}\boldsymbol{\varepsilon}'^T \boldsymbol{\Lambda}^{-1} \boldsymbol{\varepsilon}' \tag{67}$$

其中

$$\boldsymbol{\varepsilon}' = \mathbf{U}^T \boldsymbol{\varepsilon} \tag{68}$$

新誤差 $\boldsymbol{\varepsilon}'$ 的協方差矩陣為

$$\mathbf{C}' = \mathrm{Cov}[\boldsymbol{\varepsilon}'] = \mathcal{E}(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}'^T) = \mathbf{U}^T \mathcal{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T)\mathbf{U} = \mathbf{U}^T \mathbf{C}\mathbf{U} = \boldsymbol{\Lambda} \tag{69}$$

在導出上式時已用到 (66a) 式和 (65b) 式。由 (69) 式可看出，新誤差的協方差矩陣變為對角的，對角項就是新誤差的方差，它們等於 C 的特徵值。(67) 式和 (10) 式在形式上完全一樣。因此，透過線性變換 (68) 式，(11) 式就變為 (10) 式，故令 (11) 式取極小值是最一般性的最小二乘法。也因為這樣，目標函數中做為權重的逆矩陣都代表各項的準確度。

1.3 無約束極小化

在變分資料同化問題中需要求出目標函數的極大值或極小值，現在說明一些基本原理。假設目標函數 J 為 x 的函數：$J = f(x)$，這個函數在 x^* 處有極值，而且 f 在 x^* 處具有有界的（bounded）二階導數。令 α 為無窮小正數，η 為非零任意數。將 $f(x^* + \alpha\eta)$ 對 x^* 做 Taylor 級數展開，有

$$f(x^* + \alpha\eta) = f(x^*) + \alpha\eta f'(x^*) + \frac{1}{2}\alpha^2\eta^2 f''(x^*) + O(\alpha^3) \tag{1}$$

若 x^* 處 f 有極小值，則 $f(x^* + \alpha\eta) \geq f(x^*)$。因此由 (1) 式有

$$\alpha\eta f'(x^*) + \frac{1}{2}\alpha^2\eta^2 f''(x^*) + O(\alpha^3) \geq 0 \tag{2}$$

因為 η 是任意數，故可令它等於 $-f'(x^*)$，然後兩邊除以 α，最後再令

α趨近於零，得到$[f'(x^*)]^2 \le 0$。由於左邊是正數或零，因此這個式子成立的條件是

$$f'(x^*) = 0 \tag{3}$$

這是x^*處 f 有極小值的必要條件。用同樣的方法可以證明，x^*處$f(x)$有極大值的必要條件也是(3)式。(3)式是眾人所熟知的x^*處 f 有極值的條件。至於是極大值或極小值，將在下面討論多元（即多變數）的情況時再說明。

若目標函數 J 是$M \times 1$向量 x 的函數，即$J = f(\mathbf{x})$，也可用類似的方法導出取極值的必要條件。假設 f 在$\mathbf{x}^*$處有極小值，則極小值$f(\mathbf{x}^*)$滿足下式：

$$f(\mathbf{x}^* + \alpha\boldsymbol{\eta}) \ge f(\mathbf{x}^*) \tag{4}$$

其中α仍然是無窮小的正數，$\boldsymbol{\eta}$為非零的$M \times 1$任意向量。假設 f 具有有界的二階偏導數，將$f(\mathbf{x}^* + \alpha\boldsymbol{\eta})$對$\mathbf{x}^*$做 Taylor 級數展開，有

$$f(\mathbf{x}^* + \alpha\boldsymbol{\eta}) = f(\mathbf{x}^*) + \alpha\,\boldsymbol{\eta}^T\,\frac{\partial f}{\partial \mathbf{x}} + \frac{1}{2}\alpha^2\,\boldsymbol{\eta}^T\mathbf{G}\boldsymbol{\eta} + O(\alpha^3) \tag{5}$$

其中上標 T 表示矩陣的轉置；G 是$M \times M$矩陣，它的第$(i,\ j)$個元素為 $\partial^2 f/\partial x_i \partial x_j$，這個矩陣稱為 Hesse 矩陣。(5)式也可用下式表示：

$$f(x_i^* + \alpha\eta_i) = f(x_i^*) + \alpha\sum_{i=1}^{M}\eta_i\,\frac{\partial f}{\partial x_i} + \frac{1}{2}\alpha^2\sum_{i=1}^{M}\sum_{j=1}^{M}\eta_i\,\frac{\partial^2 f}{\partial x_i \partial x_j}\eta_j + O(\alpha^3)$$

假如$f(\mathbf{x}^*)$是極小值，則由(4)和(5)式有

$$\alpha\boldsymbol{\eta}^T\,\frac{\partial f}{\partial \mathbf{x}} + \frac{1}{2}\alpha^2\,\boldsymbol{\eta}^T\mathbf{G}\boldsymbol{\eta} + O(\alpha^3) \ge 0 \tag{6}$$

因為$\boldsymbol{\eta}$是任意向量，故可令$\boldsymbol{\eta} = -\partial f/\partial \mathbf{x}$，於是(6)式變為

$$-\alpha\left[\frac{\partial f}{\partial \mathbf{x}}\right]^T\left[\frac{\partial f}{\partial \mathbf{x}}\right] + O(\alpha^2) \ge 0$$

將上式兩邊除以α，再令α趨近於零，得到

$$\left[\frac{\partial f}{\partial \mathbf{x}}\right]^T\left[\frac{\partial f}{\partial \mathbf{x}}\right] \le 0$$

上式左邊代表向量 $\partial f/\partial \mathbf{x}$ 長度的平方，因此若要上式成立，則這個向量必須為零向量：

$$\frac{\partial f}{\partial \mathbf{x}} = \mathbf{0} \qquad \text{（極大或極小）} \tag{7}$$

這是 $f(\mathbf{x})$ 取極小值的必要條件，也是取極大值的必要條件（將 (4) 式改為 $f(\mathbf{x}^*+\alpha\boldsymbol{\eta}) \le f(\mathbf{x}^*)$）。必須指出，(7) 式所示的條件相當於令 J 的微分等於零：

$$dJ = d\mathbf{x}^T\,\frac{\partial f}{\partial \mathbf{x}} \equiv \sum_{i=1}^{M}\frac{\partial f}{\partial x_i}dx_i = 0 \tag{8}$$

其中 dx_i 是任意的，故由 (8) 式可得到 (7) 式。

還需另一個條件去分辨極小值或極大值。在 (7) 式所示的條件下，(6) 式變為

$$\frac{1}{2}\alpha^2\boldsymbol{\eta}^T\mathbf{G}\boldsymbol{\eta} + O(\alpha^3) \ge 0 \tag{9}$$

上式兩邊除以α^2，並令α趨近於零，我們發現 $f(\mathbf{x}^*)$ 為極小值的充分條件是

$$\boldsymbol{\eta}^T\mathbf{G}\boldsymbol{\eta} > 0 \ \text{（極小）} \tag{10}$$

也就是說，極小點 $\mathbf{x}^*$ 處的 Hesse 矩陣必須是正定的（positive definite）（嚴格的應該是半正定的）。同理也可以證明，$f(\mathbf{x}^*)$ 為極大值的充分條件為

$$\boldsymbol{\eta}^T\mathbf{G}\boldsymbol{\eta} < 0 \ \text{（極大）} \tag{11}$$

也就是說，極大點 $\mathbf{x}^*$ 處的 Hesse 矩陣必須是負定的（negative definite）（嚴格的說應該是半負定的）。若這個矩陣是正定的，則極值是極小值；若這個矩陣是負定的，則極值是極大值。根據 1.2 節，實

數對稱正定矩陣的特徵值都是正值，因此若 Hesse 矩陣 G 的特徵值全是正數，則極值為極小值。

下降算法

下降算法（descent algorithm）是指求一個函數極小值的疊代法。所有的下降算法，都需用到目標函數的梯度或二階導數，由於二階導數的計算較為複雜，故大型最佳化問題通常不藉由需用到二階導數的下降算法去求解。1.2 節中已提到，變分同化問題是求出使得目標函數取極小值的狀態向量，因此變分資料同化最重要的數值問題是利用下降算法求解。

本節中考慮無約束極小化問題，也就是說目標函數 $J(\mathbf{x})$ 中的 $M\times 1$ 向量 x 的元素之間沒有約束條件。狀態向量 x 的 M 個元素可視為一個 M 維空間的坐標，梯度算符可在這個相空間（phase space）中定義，正如它可在物理空間中定義一樣。梯度 $\nabla J\equiv\partial J/\partial\mathbf{x}$ 決定出來以後，就可用在下降算法中，以便決定出使 J 取極小值的 x。下降算法現在已有大量文獻，可參考 Nocedal 和 Wright 1999，在變分最佳分析上的應用可參考 Navon 和 Legler 1987 以及 Rodgers 2000（第 5 章），有些數值方法的書（例如 Press 等人 1992）也會討論一般的下降算法。

設想我們由某一點 x 點沿某一方向 $\mathbf{e}$ 前往附近的另一點 $\mathbf{x}+\alpha\mathbf{e}$，其中 α 為很小的正數，e 為任意向量。將 $J(\mathbf{x}+\alpha\mathbf{e})$ 對 x 做 Taylor 級數展開，有

$$J(\mathbf{x}+\alpha\mathbf{e})=J(\mathbf{x})+\alpha\,\mathbf{e}^T\mathbf{g}+\cdots \tag{12}$$

其中 $\mathbf{g}\equiv\nabla J\equiv\partial J/\partial\mathbf{x}$，以後都用 g 表示梯度。假如 α 夠小的話，則 (12) 式右邊前兩項佔優勢，更高階項可略去不計。由 (12) 式可知，$\mathbf{e}^T\mathbf{g}$ 的符號決定了正在上升或下降。也就是說當 $\mathbf{e}^T\mathbf{g}>0$ 時，$J(\mathbf{x}+\alpha\mathbf{e})>J(\mathbf{x})$，$\mathbf{x}+\alpha\mathbf{e}$ 處的 J 值較大，此時我們正在上升。相反的，當 $\mathbf{e}^T\mathbf{g}<0$ 時，我們正在下降，此時 $\mathbf{e}$ 稱為 x 處的下降方向。顯然的 $\mathbf{e}=-\mathbf{g}$ 是最陡下降（steepest descent）方向。

假設 $\mathbf{g} \equiv \nabla J \neq 0$，而且 C 是任意對稱正定矩陣。根據對稱正定矩陣的定義，有 $\mathbf{u}^T\mathbf{C}\mathbf{u} > 0$，其中 u 為任意非零向量。令

$$\mathbf{e}_1 \equiv -\mathbf{C}\mathbf{g} \tag{13}$$

則

$$\mathbf{e}_1^T\mathbf{g} = -\mathbf{g}^T\mathbf{C}^T\mathbf{g} = -\mathbf{g}^T\mathbf{C}\mathbf{g} < 0 \tag{14}$$

因此根據(14)式，(13)式所示的 $\mathbf{e}_1$ 也是一個下降方向。

總結的說，下降方向 $\mathbf{e}$ 是使得 $\mathbf{e}^T\mathbf{g} < 0$ 的方向。特別的說，$-\mathbf{C}\mathbf{g}$ 也是下降方向，其中 C 是任意對稱正定矩陣。當 C 為單位矩陣時，$\mathbf{e} = -\mathbf{g}$，這是最陡下降方向。

凸函數的觀念在極小化問題中相當有用。為簡便計只說明一元凸函數。如圖 1-1 所示，考慮函數 $J = f(x)$。令 x_1 和 x_2 為任意兩點，A 和 B 點的坐標分別為 (x_1, J_1) 和 (x_2, J_2)。若曲線 $J = f(x)$ 上兩任意點 A 和 B 點的連線都大於 $f(x)$，則 $f(x)$ 稱為凸函數（convex function）。凸函數的極小點就是全局極小點（global minimum）。當凸函數有多個極小點時，這些極小點必定相等。此外，嚴格凸函數定義為只有唯一的極小點。

下降算法是首先給出 $J(\mathbf{x})$ 的極小點 $\mathbf{x}^*$ 的一個首次猜測值 $\mathbf{x}_0$，然後計算一系列的疊代點 $\mathbf{x}_1$, $\mathbf{x}_2, \cdots$，希望這個序列 $\{\mathbf{x}_n\}$ 的極限就是某一極

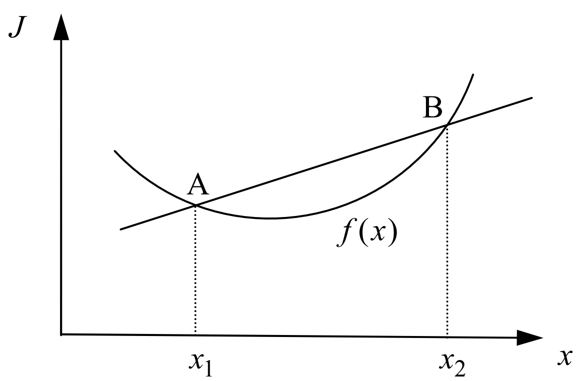

圖 1-1 凸函數。

小點 $\mathbf{x}^*$。許多疊代方案中疊代點以兩個不同的階段趨近於極小點。全局收斂性（global convergence）是指初始點 $\mathbf{x}_0$ 遠離極小點的情況下某個疊代方案的收斂性。在這個疊代的前期階段我們希望疊代點持續的往極小點的附近移動，此時只要求

$$\left\|\mathbf{x}_{n+1}-\mathbf{x}^*\right\| \le \left\|\mathbf{x}_n-\mathbf{x}^*\right\|, \qquad n > n_0$$

就夠了，其中 n_0 為一正整數。上式中 $\|\mathbf{u}\|$ 表示向量 u 的長度，即

$$\|\mathbf{u}\| = \sqrt{\mathbf{u}^T\mathbf{u}}$$

這是一種最常用的範數（norm），範數是描述向量大小的量。在疊代的後期，$\left\|\mathbf{x}_n-\mathbf{x}^*\right\|$ 已變為很小，局部收斂性（local convergence）則告訴我們疊代點收斂於極小點的速率。如果存在著一個與 n 無關的常數 $\beta > 0,\ \ \alpha > 1$ 以及 n_0，使得當 $n > n_0$ 時，

$$\left\|\mathbf{x}_{n+1}-\mathbf{x}^*\right\| \le \beta\left\|\mathbf{x}_n-\mathbf{x}^*\right\|^\alpha$$

成立，則稱這個算法是 α 階收斂的。當 $\alpha = 1$，而且 $0 < \beta < 1$ 時，相應的算法稱為線性收斂的；當 $1 < \alpha < 2$ 時，稱為超線性收斂的；當 $\alpha = 2$ 時，稱為二次收斂的。一般說來，階數越高，收斂得越快。從計算角度看來，一個算法若有超線性或二次收斂，就被認為是較好的算法。對變分同化問題來說，初始點可用背景值，這個值已很接近極小值，因此局部收斂性比全局收斂性重要得多了。

下降算法的收斂性判據一般說來可分為下面幾種：

(a) 梯度判據。這是指目標函數在疊代點的梯度的範數達到充分的小，即 $\left\|\nabla J(\mathbf{x}_n)\right\| < \varepsilon_1$，其中 n 為疊代序號。

(b) 距離判據。相鄰兩次疊代點之間的距離達到充分的小，即 $\left\|\mathbf{x}_{n+1}-\mathbf{x}_n\right\| < \varepsilon_2$。

(c) 函數值判據。相鄰兩次疊代點的函數值或相對值已達到充分的小，即

$$|J(\mathbf{x}_{n+1})-J(\mathbf{x}_n)|<\varepsilon_3 \quad 或 \quad \frac{|J(\mathbf{x}_{n+1})-J(\mathbf{x}_n)|}{|J(\mathbf{x}_n)|}<\varepsilon_4$$

上面幾個式子中 ε_1，ε_2，ε_3，ε_4 都是事先給定的足夠小的正數。這些判據都在一定程度上反映了疊代過程可能已到達極小值。滿足其中任何一個判據，都可視為可能已收斂到極小點，可以終止計算工作。不過即使滿足這些判據，也有可能不是極小點，更不用說是全局極小點了。

最陡下降法

可以想像，為了要到達 J 的極小點 x，可用下面的疊代方案：

$$\mathbf{x}_{n+1}=\mathbf{x}_n-\xi_n\mathbf{g}_n \tag{15}$$

其中 $\mathbf{g}_n\equiv\nabla J(\mathbf{x}_n)$ 為第 n 次疊代時目標函數 J 關於 x 的梯度。$\mathbf{x}_{n+1}$ 應該比 $\mathbf{x}_n$ 更接近極小點，因為 $-\mathbf{g}_n$ 是最陡下降方向。(15) 式所示的疊代方案稱為最陡下降法（method of steepest descent），因為下降方向一直沿著負局部梯度的方向。ξ_n 是正值純量，稱為步長（step size），每次疊代時移動的距離就是步長（因為 $\mathbf{g}_n$ 不是單位向量，故真正的步長是 $\xi_n\|\mathbf{g}_n\|$）。步長假如太小，則走得太慢；假如太大，則 $J(\mathbf{x}_{n+1})$ 的值又可能比 $J(\mathbf{x}_n)$ 大，也就是說沒有往極小點邁進。最佳的步長可令 $J(\mathbf{x}_n-\xi_n\mathbf{g}_n)$ 取極小值而得到：

$$\frac{\partial}{\partial\xi_n}J(\mathbf{x}_{n+1})=\frac{\partial}{\partial\xi_n}J(\mathbf{x}_n-\xi_n\mathbf{g}_n)=0 \tag{16}$$

上式表示沿下降方向尋找極小點。(16) 式相當於

$$\left[\frac{\partial J(\mathbf{x}_{n+1})}{\partial\mathbf{x}_{n+1}}\right]^T\frac{\partial\mathbf{x}_{n+1}}{\partial\xi_n}=0$$

也就是說

$$\mathbf{g}_{n+1}^T\mathbf{g}_n=0 \tag{17}$$

(17) 式表明，在前後兩次疊代的 x 處，其梯度必須互相垂直，如圖 1-2

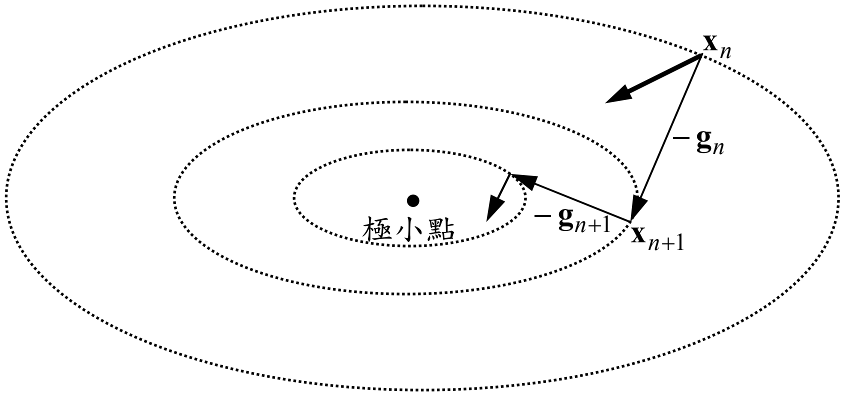

圖 1-2 虛線表示 J 等值線，細線表示最陡下降法，粗線表示最快到達極小點的方向。

所示。(16) 式所示的一維極小化問題稱為線搜尋（line search），這是相當簡單的數值問題，因為下降方向只有一個。(17) 式所示的關係可按下面方法得到。首先由 $\mathbf{x}_n$ 點沿這一點的最陡下降方向 $-\mathbf{g}_n$ 到達一個極小點 $\mathbf{x}_{n+1}$，不要再繼續走，改為沿 $\mathbf{x}_{n+1}$ 處的最陡下降方向 $-\mathbf{g}_{n+1}$ 下去。這個方向垂直於 $\mathbf{x}_{n+1}$ 處的 J 等值線，$-\mathbf{g}_n$ 則平行於這條線，因而得到 (17) 式。最陡下降法開頭幾步的步長較大，但由於相鄰兩次疊代的下降方向是正交的，因此在極小點附近，疊代點逼近極小點的速率越來越慢，從而只具有線性的收斂性。

由於最陡下降法相對的勇健（robust），也就是說對首次猜測值的選擇較不敏感，因此這個方法在疊代初期的表現相當良好。不過在疊代後期收斂速率較慢，故很少單獨用來實際執行極小化工作。

共軛梯度法

大型極小化問題經常使用的下降算法稱為共軛梯度法(conjugate gradient method）。對向量族 $\mathbf{e}_0, \mathbf{e}_1, \cdots$ 來說，假如

$$\mathbf{e}_i^T \mathbf{C} \mathbf{e}_j = 0, \qquad i \neq j \tag{18}$$

則這個向量族稱為關於 C 共軛的，其中 C 為正定對稱矩陣。當 C 是單位

矩陣時，(18) 式就是一般的正交性，因此共軛性是正交性的推廣。

做為例子，考慮下面二次多項式的極小化問題：

$$J(\mathbf{x}) = a + \mathbf{b}^T\mathbf{x} + \frac{1}{2}\mathbf{x}^T\mathbf{G}\mathbf{x} \tag{19}$$

其中 a 為常數，b 和 G 分別為向量和對稱正定矩陣，它們都與 x 無關。G 是 Hesse 矩陣，它的第 (i, j) 個元素是 $\partial^2 J / \partial x_i \partial x_j$。將 (19) 式對 x 微分，有

$$\mathbf{g} \equiv \nabla J = \mathbf{b} + \mathbf{G}\mathbf{x} \tag{20}$$

(20) 式的證明如下。(19) 式可寫為

$$J = a + \sum_i b_i x_i + \frac{1}{2}\sum_i\sum_j x_i G_{ij} x_j$$

將上式對 x_m 微分，有

$$\begin{aligned}\frac{\partial J}{\partial x_m} &= \sum_i b_i\delta_{im} + \frac{1}{2}\sum_i\sum_j \delta_{im}G_{ij}x_j + \frac{1}{2}\sum_i\sum_j x_i G_{ij}\delta_{jm} \\ &= b_m + \frac{1}{2}\sum_j G_{mj}x_j + \frac{1}{2}\sum_i G_{im}x_i\end{aligned}$$

上式用矩陣形式表示，就是

$$\mathbf{g} \equiv \nabla J = \mathbf{b} + \frac{1}{2}\mathbf{G}\mathbf{x} + \frac{1}{2}\mathbf{G}^T\mathbf{x}$$

因為 G 是對稱矩陣，故上式可改寫為 (20) 式。上面已導出兩個矩陣微分公式：

$$\frac{\partial}{\partial \mathbf{x}}\mathbf{b}^T\mathbf{x} = \mathbf{b}, \qquad \frac{\partial}{\partial \mathbf{x}}\mathbf{x}^T\mathbf{G}\mathbf{x} = \mathbf{G}\mathbf{x} + \mathbf{G}^T\mathbf{x}$$

因此，由 (20) 式可知，極小點 $\mathbf{x}^*$ 滿足下式：

$$\mathbf{b} = -\mathbf{G}\mathbf{x}^* \tag{21}$$

由於 (19) 式所示的二次多項式是相當簡單的非線性函數，一般性的目標函數在極小點附近的性質又近似於二次多項式，所以這個函數常

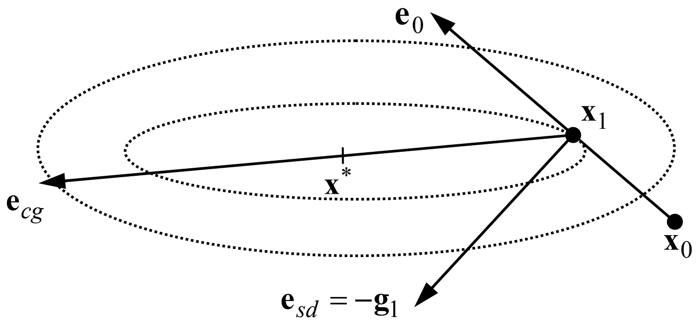

圖 1-3　$\mathbf{e}_{sd}$ 為 $\mathbf{x}$ 點處的最陡下降方向，$\mathbf{e}_{cg}$ 為共軛方向，$\mathbf{x}^*$ 為極小點。

常當做模式之用。現在假設 $\mathbf{x}$ 只有兩個元素。如圖 1-3 所示，設想我們從 $\mathbf{x}_0$ 點沿下降方向 $\mathbf{e}_0$ 抵達 $\mathbf{x}_1$ 點，此處正是 $\mathbf{e}_0$ 方向的極小點。假如再沿最陡下降方向 $\mathbf{e}_{sd}$ 下去，就會耗費太多時間才能抵達極小點。我們有另一個選擇。根據線搜尋，$\mathbf{e}_0$ 和 $\mathbf{e}_{sd}$ 互相垂直，即

$$0 = \mathbf{e}_0^T \mathbf{e}_{sd} = -\mathbf{e}_0^T \mathbf{g}_1 = -\mathbf{e}_0^T (\mathbf{b} + \mathbf{G}\mathbf{x}_1)$$

在導出上式時已用到 (20) 式。將 (21) 式代入上式，得到

$$\mathbf{e}_0^T \mathbf{G}(\mathbf{x}^* - \mathbf{x}_1) = 0 \tag{22}$$

由 (22) 式可知，$(\mathbf{x}^* - \mathbf{x}_1)$ 和 $\mathbf{e}_0$ 關於 G 互為共軛。顯然的，假如由 $\mathbf{x}_0$ 點沿下降方向 $\mathbf{e}_0$ 抵達 $\mathbf{x}$ 點後，然後再沿共軛方向 $(\mathbf{x}^* - \mathbf{x}_1)$ 繼續前行，我們立刻抵達極小點。可以證明，共軛方向是前一次的搜尋方向（即下降方向）$\mathbf{e}_0$ 和這次的最陡下降方向 $\mathbf{e}_{sd} = -\mathbf{g}_1$ 兩者的線性組合，而且在這兩個方向之間。顯然的，對二元二次多項式 $J(\mathbf{x})$ 來說，只要 2 步就可抵達極小點。還可以證明，若 $J(\mathbf{x})$ 是 M 元二次多項式，只要 M 步就可抵達極小點。不過一般的目標函數 $J(\mathbf{x})$ 通常不是二次多項式，故必須更多步才能抵達極小點，甚至無法抵達。

共軛梯度法的疊代方案為

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \xi_n \mathbf{e}_n \tag{23}$$

其中 $\mathbf{e}_n$ 為上次的下降方向。步長 ξ_n 仍由線搜尋決定，即求出一個 ξ_n，使得 $J(\mathbf{x}_n + \xi_n \mathbf{e}_n)$（$\mathbf{x}_n$ 和 $\mathbf{e}_n$ 保持固定）取極小值，即

$$\frac{\partial}{\partial \xi_n} J(\mathbf{x}_n + \xi_n \mathbf{e}_n) = 0 \qquad (24)$$

在這情況下，這次的梯度和上次的搜尋方向（即下降方向）是正交的，即

$$\mathbf{e}_{k-1}^T \mathbf{g}_k = 0, \qquad \text{對所有的}k\text{來說} \qquad (25)$$

另外，令 $\mathbf{e}_0 = -\mathbf{g}_0 \equiv -\nabla J(\mathbf{x}_0)$，也就是說第 0 步是沿著最陡下降方向。圖 1-4 表示各符號的意義。

線搜尋也可不必找到這個下降方向的極小點，此時 (25) 式不成立，這種線搜尋稱為不準確線搜尋（soft line search）。找到極小點的線搜尋稱為準確線搜尋（exact line search）。在實際計算中一般做不到準確線搜尋，實際上也沒必要做到這一點，這是因為準確線搜尋會耗費計算時間，而且對加快收斂的效果也不大，因此現在不少下降算法使用不準確線搜尋，不過不準確線搜尋不在這裡說明。

這次的下降方向則由下式給出：

$$\mathbf{e}_{n+1} = -\mathbf{g}_{n+1} + \beta_n \mathbf{e}_n \qquad (26)$$

(26) 式表明，現在的搜尋方向是上一次的搜尋方向和這次的最陡下降方向的線性組合。考慮下面的表達式：

$$\mathbf{e}_{n+1}^T \mathbf{g}_{n+1} = (-\mathbf{g}_{n+1} + \beta_n \mathbf{e}_n)^T \mathbf{g}_{n+1} = -\mathbf{g}_{n+1}^T \mathbf{g}_{n+1} < 0$$

在導出上式時已使用 (26) 式和 (25) 式。這樣就證明了 (26) 式所示的方向 $\mathbf{e}_{n+1}$ 是下降方向。顯然的，最陡下降法相當於令 (26) 式中 $\beta_n = 0$。

下面將導出若干關係式，然後說明β_n的選擇。我們要求出β_n，使得新的下降方向$\mathbf{e}_{n+1}$和所有以前的下降方向$\mathbf{e}_j$（$j \le n$）關於 G 互為共軛，即

$$\mathbf{e}_{n+1}^T \mathbf{G} \mathbf{e}_j = 0, \qquad j = n, n-1, \cdots, 0 \tag{27}$$

其中 G 為 Hesse 矩陣。

首先考慮下面的表達式：

$$\mathbf{e}_j^T \mathbf{g}_k = \mathbf{e}_j^T (\mathbf{g}_k - \mathbf{g}_{j+1} + \mathbf{g}_{j+1}) = \mathbf{e}_j^T (\mathbf{g}_k - \mathbf{g}_{j+1}), \qquad j = k-2, k-3, \cdots$$

在導出上式第二個等號右邊時已用到 (25) 式。假如$J(\mathbf{x})$以 (19) 式所示的二次函數來近似，則$\mathbf{g}$由 (20) 式給出，再利用 (23) 式，上式變為

$$\begin{aligned}\mathbf{e}_j^T \mathbf{g}_k &= \mathbf{e}_j^T \mathbf{G}(\mathbf{x}_k - \mathbf{x}_{j+1}) = \mathbf{e}_j^T \mathbf{G}(\mathbf{x}_k - \mathbf{x}_{k-1} + \mathbf{x}_{k-1} - \mathbf{x}_{k-2} + \cdots + \mathbf{x}_{j+2} - \mathbf{x}_{j+1}) \\ &= \mathbf{e}_j^T \mathbf{G}(\xi_{k-1}\mathbf{e}_{k-1} + \xi_{k-2}\mathbf{e}_{k-2} + \cdots + \xi_{j+1}\mathbf{e}_{j+1}), \quad j = k-2, k-3, \cdots\end{aligned}$$

上式第 2 個等號右邊的小括弧內已加減一個同樣的量。當$j = k-2$，$j = k-3$，$j = k-3, \cdots$時，利用共軛條件 (27) 式，則由上式分別有

$$\mathbf{e}_j^T \mathbf{g}_k = \mathbf{e}_{k-2}^T \mathbf{G}(\xi_{k-1}\mathbf{e}_{k-1}) = 0$$

$$\mathbf{e}_j^T \mathbf{g}_k = \mathbf{e}_{k-3}^T \mathbf{G}(\xi_{k-1}\mathbf{e}_{k-1} + \xi_{k-2}\mathbf{e}_{k-2}) = 0$$

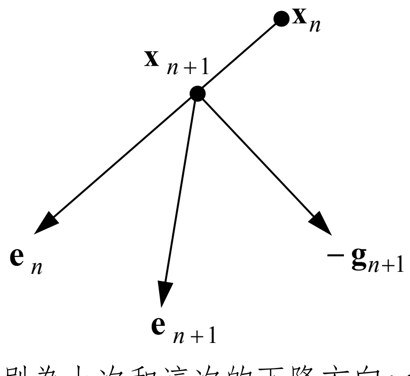

圖 1-4 $\mathbf{x}_n$和$\mathbf{e}_{n+1}$分別為上次和這次的下降方向，$\mathbf{g}_{n+1}$和$\mathbf{e}_n$正交，$\mathbf{e}_{n+1}$是$\mathbf{g}_{n+1}$和$\mathbf{e}_n$的線性組合，它們在同一平面上。

$$\mathbf{e}_j^T \mathbf{g}_k = \mathbf{e}_{k-4}^T \mathbf{G}(\xi_{k-1}\mathbf{e}_{k-1} + \xi_{k-2}\mathbf{e}_{k-2} + \xi_{k-3}\mathbf{e}_{k-3}) = 0$$

$\cdots$

因此，

$$\mathbf{e}_j^T \mathbf{g}_k = 0, \qquad j = k-2, k-3, \cdots$$

由這個式子和 (25) 式可知，現在的梯度不但和上次的搜尋方向正交，而且也和所有以前的搜尋方向都是正交的，也就是說

$$\mathbf{e}_j^T \mathbf{g}_k = 0, \qquad j = k-1, k-2, \cdots \tag{28}$$

將 (26) 式中 n 改為 $j-1$，然後取轉置，再後乘以 $\mathbf{g}_k$，有

$$\mathbf{e}_j^T \mathbf{g}_k = -\mathbf{g}_j^T \mathbf{g}_k + \beta_{j-1}\mathbf{e}_{j-1}^T \mathbf{g}_k, \qquad j = k-1, k-2, \cdots$$

利用 (28) 式，上式變為

$$\mathbf{g}_j^T \mathbf{g}_k = 0, \qquad j = k-1, k-2, \cdots \tag{29}$$

(29) 式表明，所有疊代點上的梯度是互相正交的。

另一方面，考慮下面的表達式：

$$\mathbf{g}_{n+1} - \mathbf{g}_n = \mathbf{G}(\mathbf{x}_{n+1} - \mathbf{x}_n) = \xi_n \mathbf{G}\mathbf{e}_n$$

在導出上式第 2 個等號右邊時已用到 (23) 式。由上式有

$$\mathbf{e}_n = \mathbf{G}^{-1}\xi_n^{-1}(\mathbf{g}_{n+1} - \mathbf{g}_n) \tag{30}$$

為了導出 (26) 式中的 β_n，$j=n$ 時的 (27) 式以 (26) 和 (30) 式代入，有

$$(-\mathbf{g}_{n+1} + \beta_n \mathbf{e}_n)^T \mathbf{G}\mathbf{G}^{-1}\xi_n^{-1}(\mathbf{g}_{n+1} - \mathbf{g}_n) = 0$$

利用 (28) 式，上式變為

$$0 = -\mathbf{g}_{n+1}^T(\mathbf{g}_{n+1} - \mathbf{g}_n) - \beta_n \mathbf{e}_n^T \mathbf{g}_n$$

再將上式中的 $\mathbf{e}_n^T$ 以 (26) 式（(26) 式中 n 先改為 $n-1$）代入，我們發現

$$\begin{aligned}0 &= -\mathbf{g}_{n+1}^T(\mathbf{g}_{n+1}-\mathbf{g}_n) - \beta_n(-\mathbf{g}_n + \beta_{n-1}\mathbf{e}_{n-1})^T\mathbf{g}_n \\ &= -\mathbf{g}_{n+1}^T(\mathbf{g}_{n+1}-\mathbf{g}_n) + \beta_n\mathbf{g}_n^T\mathbf{g}_n = -\mathbf{g}_{n+1}^T\mathbf{g}_{n+1} + \beta_n\mathbf{g}_n^T\mathbf{g}_n\end{aligned}$$

在導出上式第二個等號右邊時用到了(28)式，在導出第三個等號右邊時用到了(29)式。上式第二個等號和第三個等號給出了 Polak-Ribi re（簡稱 PR）算法和 Fletcher-Reeves（簡稱 FR）算法：

$$\beta_n = \frac{\mathbf{g}_{n+1}^T(\mathbf{g}_{n+1}-\mathbf{g}_n)}{\mathbf{g}_n^T\mathbf{g}_n} \text{（PR），} \qquad \beta_n = \frac{\mathbf{g}_{n+1}^T\mathbf{g}_{n+1}}{\mathbf{g}_n^T\mathbf{g}_n} \text{（FR）} \qquad (31)$$

可以證明，當 $j = n-1, n-2, \cdots$ 時，由(27)式所示的共軛條件給出了 β_{n-1}， β_{n-2}，$\cdots$等的表達式，但它們都跟(31)式完全一致。因此(31)式所示的 β_n 會使(26)式所示的 $\mathbf{e}_{n+1}$ 滿足共軛條件(27)式。由(28)式可看出，對(19)式所示的二次多項式來說，FR 算法和 PR 算法完全一樣。不過對一般的函數來說，這兩種算法並不一樣。根據以往的經驗，後者優於前者。

共軛梯度法的疊代步驟如下：

(a) 選擇 $\mathbf{x}_0$，然後以微分、有限差分法或伴隨法（adjoint method）（見 1.4 節）計算 $\mathbf{g}_0 = \nabla J(\mathbf{x}_0)$，令 $\mathbf{e}_0 = -\mathbf{g}_0$。

(b) 令 $\partial J(\mathbf{x}_n + \xi_n\mathbf{e}_n)/\partial\xi_n = 0$，以線搜尋決定 ξ_n。

(c) $\mathbf{x}_{n+1} = \mathbf{x}_n + \xi_n\mathbf{e}_n$，然後以微分、有限差分法或伴隨法計算 $\mathbf{g}_{n+1} = \nabla J(\mathbf{x}_{n+1})$。

(d) $\beta_n = \mathbf{g}_{n+1}^T(\mathbf{g}_{n+1} - \mathbf{g}_n)/\mathbf{g}_n^T\mathbf{g}_n$（PR 算法）或 $\beta_n = \mathbf{g}_{n+1}^T\mathbf{g}_{n+1}/\mathbf{g}_n^T\mathbf{g}_n$（FR 算法）。

(e) $\mathbf{e}_{n+1} = -\mathbf{g}_{n+1} + \beta_n\mathbf{e}_n$。

(f) 回到(b)。

必須指出，在進行一般函數的極小化問題時，需做兩個特別的數值計算，包括線搜尋和求梯度。線搜尋是一維的極小化問題，由於搜尋方向只有一個，因此這只是較為簡單的數值問題。至於求梯度的方法，就比較複雜了，可用微分、有限差分法或伴隨技術求得，對於大型極小問

題來說非用伴隨技術不可。在極小化過程中，為了要計算下降方向，共軛梯度法只需儲存幾個 $M\times 1$ 單行向量，包括現在和前一次的梯度 $\mathbf{g}_{n+1}$, $\mathbf{g}_n$ 以及前一次的下降方向 $\mathbf{e}_n$ 等，不像下面提到的牛頓法至少需儲存 $M\times M$ Hesse 矩陣的元素。共軛梯度法用在 M 維狀態向量的正定二次目標函數時，只需 M 次疊代就能到達極小值。假如同時考慮收斂速率和計算機記憶需求，共軛梯度法也許是求解大型最佳化問題的下降算法中最合適的方法，現有的商業軟體也都有這個算法的程式（IMSL, 1991；Press 等人 1992）

牛頓下降法

另一個常用的方法是牛頓下降法。首先說明如何求出非線性方程 $f(x)=0$ 的解。假定已知第 n 次疊代的解 x_n，接著要求第 $n+1$ 次疊代的解 x_{n+1}。令這個新的解為 $x_{n+1}=x_n+\delta$，其中 δ 為微小的修正量。假設這個新的解已滿足非線性方程，即

$$f(x_n+\delta)=0 \tag{32}$$

將上式對 x_n 做 Taylor 級數展開，有

$$f(x_n+\delta)=f(x_n)+f'(x_n)\delta+\cdots=0$$

對於足夠小的 δ 而且行為良好的函數 $f(x)$ 來說，更高階項可略去不計，因而 $\delta=-f(x_n)/f'(x_n)$。因此疊代方案如下：

$$x_{n+1}=x_n-\frac{f(x_n)}{f'(x_n)} \tag{33}$$

(33) 式是求解非線性方程 $f(x)=0$ 的 Newton-Raphson 疊代方案。

這個方法不限於一元問題，很容易推廣到多元的情況。對多元問題來說，非線性方程組為 $\mathbf{f}(\mathbf{x})=0$，其中 $\mathbf{x}$ 為待求向量。第 $n+1$ 次疊代的解可寫為 $\mathbf{x}_{n+1}=\mathbf{x}_n+\boldsymbol{\delta}$，其中 $\boldsymbol{\delta}$ 為微小的修正向量。假設這個新的解已滿足非線性方程組，即

$$\mathbf{f}(\mathbf{x}_n+\boldsymbol{\delta})=0 \tag{34}$$

將上式對 $\mathbf{x}_n$ 做 Taylor 級數展開，有

$$\mathbf{f}(\mathbf{x}_n+\boldsymbol{\delta})=\mathbf{f}(\mathbf{x}_n)+\mathbf{f}'(\mathbf{x}_n)\boldsymbol{\delta}+\cdots=0 \tag{35}$$

其中 $\mathbf{f}'(\mathbf{x})$ 爲 $\mathbf{f}(\mathbf{x})$ 對 x 的 Jacobi 矩陣，這是方形矩陣，它的第 (i, j) 個元素是 $\partial f_i/\partial x_j$。(35) 式若以指標表示，就是

$$f_i(x_j^{(n)})+\sum_j \frac{\partial f_i}{\partial x_j^{(n)}}\delta_j+\cdots=0$$

將 (35) 式中高階項省略掉，然後解出來，得到 $\boldsymbol{\delta}=-[\mathbf{f}'(\mathbf{x}_n)]^{-1}\,\mathbf{f}(\mathbf{x}_n)$，於是牛頓法的疊代方案爲

$$\mathbf{x}_{n+1}=\mathbf{x}_n-[\mathbf{f}'(\mathbf{x}_n)]^{-1}\,\mathbf{f}(\mathbf{x}_n) \tag{36}$$

求目標函數 $J(\mathbf{x})$ 的極小值，相當於求解非線性方程 $\nabla J(\mathbf{x})=0$。跟上面一樣，令 $J_{x_i}(x_j^{(n)}+\delta_j)=0$，然後做 Taylor 級數展開，並省略高階項，有

$$J_{x_i}(x_j^{(n)})+\sum_j \delta_j\frac{\partial J}{\partial x_j\partial x_i}=0$$

在這裡 $J_{x_i}\equiv\partial J/\partial x_i$。上式可寫成矩陣形式：$\nabla J+\mathbf{G}\boldsymbol{\delta}=0$，其中 G 爲 Hesse 矩陣。解出來得到 $\boldsymbol{\delta}=-\mathbf{G}^{-1}\nabla J$，因此，極小化問題的牛頓疊代方案爲

$$\mathbf{x}_{n+1}=\mathbf{x}_n-\mathbf{G}_n^{-1}\mathbf{g}_n \tag{37}$$

(13) 和 (14) 式中已證明，若 $\mathbf{G}_n^{-1}$ 是對稱正定矩陣，則 $-\mathbf{G}_n^{-1}\mathbf{g}_n$ 是下降方向。只要 Hesse 矩陣繼續保持非奇異（nonsingular），而且是正定的對稱矩陣，則牛頓疊代方案似乎沒問題。很容易證明，對 (19) 式所示的二次多項式的極小化問題來說，不論 x 的維數是多少，只要疊代 1 次就到達極小點。

當狀態向量維數 M 不大時，牛頓疊代法經常被使用，牛頓法的收斂速率比最陡下降法快，但需儲存 Hesse 矩陣的 M^2 個元素，因而這個方法不適用於大型極小化問題。

牛頓法疊代方案 (37) 式可以另一方法導出。將目標函數 $J(\mathbf{x})$ 對某次疊代點 $\mathbf{x}_n$ 做 Taylor 級數展開，有

$$J(\mathbf{x}) \cong J(\mathbf{x}_n) + \mathbf{g}^T(\mathbf{x}_n)(\mathbf{x}-\mathbf{x}_n) + \frac{1}{2}(\mathbf{x}-\mathbf{x}_n)^T \mathbf{G}(\mathbf{x}_n)(\mathbf{x}-\mathbf{x}_n) \tag{38}$$

將 (38) 式對 x 微分，並令結果等於零，得到極值處的 x 值如下：

$$\mathbf{x} = \mathbf{x}_n - \mathbf{G}_n^{-1}\mathbf{g}_n \tag{39}$$

然後將 x 當做第 $n+1$ 次疊代的值，就得到 (37) 式所示的疊代方案。可以證明，在適當的條件下，牛頓疊代法至少是二階收斂的。

牛頓疊代法的優點如下：它收斂得較快，方案簡單，而且容易執行。可是付出的代價有下面幾點：

(a) 對首次猜測值比較敏感，也就是說勇健性（robustness）不夠。不過對變分同化並不造成問題，因為這個問題有很好的首次猜測值，那就是背景值。

(b) 另一個缺點是需要計算並且儲存龐大的 Hesse 矩陣，此外也要計算它的逆。

(c) 因為 $\mathbf{x}_{n+1}$ 完全依賴於目標函數 J 在 $\mathbf{x}_n$ 處附近的行為，故 $J(\mathbf{x}_{n+1})$ 不一定會小於 $J(\mathbf{x}_n)$，也就是說這個方法可能不收斂，也許收斂到極大點或到鞍點。我們可以像最陡下降法一樣進行額外的線搜尋，也就是找出一個純量 $\xi_n > 0$，使得 $J(\mathbf{x}_{n+1})$ 取極小值，其中

$$\mathbf{x}_{n+1} = \mathbf{x}_n - \xi_n \mathbf{G}_n^{-1}\mathbf{g}_n \tag{40}$$

(d) Hesse 矩陣 G 在某些疊代點處可能變為奇異的（singular），在這情況下可在 G 的對角項加上一個正值常數 γ，即 (37) 式用下式取代：

$$\mathbf{x}_{n+1} = \mathbf{x}_n - (\mathbf{G}_n + \gamma\mathbf{I})^{-1}\mathbf{g}_n \tag{41}$$

我們選擇適當的 γ，使得 $(\mathbf{G}_n + \gamma\mathbf{I})$ 變為正定矩陣，其中 γ 是無因次正值權因子。這個方法稱為阻尼牛頓法。

(41) 式是最陡下降法和牛頓法的合併方案，也稱為 Levenberg-Marquardt（下面簡稱 LM）疊代法（見 Press 等人 1992）。當 γ 很小時，(41) 式是牛頓疊代法；當 γ 夠大時，它相當於最陡下降法，γ 可

以隨疊代序號而變動。γ的值可這樣選擇：在疊代開始時，γ選用較大的值，以便LM法顯示出最陡下降法對首次猜測值不敏感，從而不需過度謹慎的選擇首次猜測值。然後在每次疊代時，假如$J(\mathbf{x}_{n+1})<J(\mathbf{x}_n)$，則減小$\gamma$的值（舉例說除以 2），以便加快收斂速率。否則增加γ的值，以便擴大搜尋範圍。

準牛頓法

牛頓法收斂速率較快，但因需要計算並且儲存目標函數的 Hesse 矩陣，故計算量太大，難以執行大型極小化問題。相反的，最陡下降法計算雖然簡便，但收斂速率太慢。假如對牛頓法加以改進，不需儲存或計算 Hesse 矩陣，就能加快收斂速率，準牛頓法(quasi-Newton method)就是從這個方向思考的。

牛頓法和最陡下降法可用下式表示：

$$\mathbf{x}_{n+1}=\mathbf{x}_n+\xi_n\mathbf{e}_n\equiv\mathbf{x}_n-\xi_n\mathbf{D}_n\mathbf{g}_n \tag{42}$$

上式中已令$\mathbf{e}_n\equiv-\mathbf{D}_n\mathbf{g}_n$。若$\mathbf{D}_n=\mathbf{I}$，(42)式是最陡下降法。若$\mathbf{D}_n=\mathbf{G}_n^{-1}$，同時$\xi_n=1$，則是牛頓法。為了保持牛頓法的優點，我們希望$\mathbf{D}_n$能近似的等於$\mathbf{G}_n^{-1}$，並且希望$\mathbf{D}_n$能在疊代過程中逐步形成，即

$$\mathbf{D}_{n+1}=\mathbf{D}_n+\mathbf{C}_n \tag{43}$$

其中$\mathbf{C}_n$稱為修正矩陣，$\mathbf{C}_n$不同，就有不同的算法。

為了探討$\mathbf{D}_n$能近似的等於$\mathbf{G}_n^{-1}$的條件，我們仍然考慮M元二次多項式(19)式。由(20)式有

$$\mathbf{G}\,\Delta\mathbf{x}_n=\Delta\mathbf{g}_n \tag{44}$$

其中

$$\Delta\mathbf{g}_n=\mathbf{g}_{n+1}-\mathbf{g}_n,\qquad \Delta\mathbf{x}_n=\mathbf{x}_{n+1}-\mathbf{x}_n \tag{45}$$

(44)式也可寫為

$$\Delta\mathbf{x}_n=\mathbf{G}^{-1}\,\Delta\mathbf{g}_n \tag{46}$$

(44) 或 (46) 式稱為準牛頓條件。由這個條件可知，為了使 $\mathbf{D}_{n+1}$ 能近似的等於 $\mathbf{G}^{-1}$，或 $\mathbf{E}_{n+1} \equiv \mathbf{D}_{n+1}^{-1}$ 近似的等於 $\mathbf{G}$，下式應該成立：

$$\mathbf{D}_{n+1}\,\Delta\mathbf{g}_n = \Delta\mathbf{x}_n, \qquad \mathbf{E}_{n+1}\,\Delta\mathbf{x}_n = \Delta\mathbf{g}_n \tag{47a, b}$$

下面為方便起見，只考慮 (47a) 式所示的準牛頓條件。

為了要使疊代計算簡單一點，(43) 式所示的修正矩陣 $\mathbf{C}_n$ 應選取儘可能簡單的形式，通常是要求 $\mathbf{C}_n$ 的秩（rank）越小越好。若要求 $\mathbf{C}_n$ 是秩 1（rank 1）的矩陣，則可設為

$$\mathbf{C}_n = \alpha\mathbf{u}\mathbf{u}^T \tag{48}$$

其中 α 為純量，$\mathbf{u}$ 為 $M\times 1$ 單行向量。一個矩陣 A 的秩 $r(\mathbf{A})$ 是它的非奇異方形子矩陣（submatrix）的最大階數（一個方形矩陣的階數是指行數或列數，子矩陣是將一個矩陣去掉行或列而得到的新矩陣）。還有，$M\times p$ 矩陣 A 的秩滿足下式：

$$r(\mathbf{A}\mathbf{A}^T) = r(\mathbf{A}^T\mathbf{A}) = r(\mathbf{A}) \le \min(M, p)$$

因為單行向量 $\mathbf{u}$ 的秩等於 1（假如所有的元素都等於零，則秩等於零，但這情況不考慮），故 $\mathbf{u}\mathbf{u}^T$ 是秩 1 矩陣。此時 (43) 式變為

$$\mathbf{D}_{n+1} = \mathbf{D}_n + \alpha\mathbf{u}\mathbf{u}^T \tag{49}$$

將 (49) 式代入準牛頓條件 (47a) 式，得到

$$\alpha\mathbf{u}\mathbf{u}^T\Delta\mathbf{g}_n = \Delta\mathbf{x}_n - \mathbf{D}_n\,\Delta\mathbf{g}_n \tag{50}$$

由於 $\mathbf{u}\mathbf{u}^T\Delta\mathbf{g}_n = \mathbf{u}(\mathbf{u}^T\Delta\mathbf{g}_n)$，其中 $\mathbf{u}^T\Delta\mathbf{g}_n$ 為純量，故 $\mathbf{u}$ 與 $\Delta\mathbf{x}_n - \mathbf{D}_n\,\Delta\mathbf{g}_n$ 成正比。令

$$\mathbf{u} = \Delta\mathbf{x}_n - \mathbf{D}_n\,\Delta\mathbf{g}_n \tag{51}$$

代入 (50) 式，得到

$$\alpha = \frac{1}{\mathbf{u}^T\Delta\mathbf{g}_n} \tag{52}$$

最後，(49) 式以 (52) 式代入，變為

$$\mathbf{D}_{n+1} = \mathbf{D}_n + \frac{\mathbf{u}\mathbf{u}^T}{\mathbf{u}^T \Delta\mathbf{g}_n} \tag{53}$$

其中 u 由 (51) 式給出。

(53) 式稱為秩 1 公式。這個算法存在著明顯的缺點。第一，當 $\mathbf{D}_n$ 是正定矩陣時，由 (53) 式得到的 $\mathbf{D}_{n+1}$ 不一定是正定的，因而不能保證 (42) 式中的 $-\mathbf{D}_{n+1}\mathbf{g}_{n+1}$ 是下降方向。第二，假如 (53) 式中的 $\mathbf{u}^T \Delta\mathbf{g}_n = 0$，但 u 不同時等於 0，則會引起計算不穩定，因而這個方案很少被使用。

DFP 算法（Davidson-Fletcher-Powell algorithm）是一種秩 2 對稱算法。一般的秩 2 對稱算法的修正向量可寫為

$$\mathbf{C}_n = \alpha \mathbf{u}\mathbf{u}^T + \beta \mathbf{v}\mathbf{v}^T \tag{54}$$

$\mathbf{u}$ 和 $\mathbf{v}$ 為 $M \times 1$ 單行向量，α 和 β 為純量。兩個矩陣和的秩小於或等於個別矩陣的和，$\alpha \mathbf{u}\mathbf{u}^T$ 和 $\beta \mathbf{v}\mathbf{v}^T$ 都是秩 1 矩陣，故 $\mathbf{C}_n$ 為秩 2 矩陣（除了在特別情況下）。在將上式代入 (43) 式，再代入 (47a) 式，得到

$$\alpha \mathbf{u}(\mathbf{u}^T \Delta\mathbf{g}_n) + \beta \mathbf{v}(\mathbf{v}^T \Delta\mathbf{g}_n) = \Delta\mathbf{x}_n - \mathbf{D}_n \Delta\mathbf{g}_n \tag{55}$$

滿足上式的 $\mathbf{u}$ 和 $\mathbf{v}$ 不是唯一的，比較簡單的一種取法如下：

$$\mathbf{u} = \mathbf{D}_n \Delta\mathbf{g}_n, \qquad \mathbf{v} = \Delta\mathbf{x}_n \tag{56}$$

$$\alpha = -\frac{1}{\mathbf{u}^T \Delta\mathbf{g}_n}, \qquad \beta = \frac{1}{\mathbf{v}^T \Delta\mathbf{g}_n} \tag{57}$$

因此秩 2 對稱算法如下：

$$\mathbf{D}_{n+1} = \mathbf{D}_n - \frac{\mathbf{D}_n \Delta\mathbf{g}_n (\mathbf{D}_n \Delta\mathbf{g}_n)^T}{\Delta\mathbf{g}_n^T \mathbf{D}_n \Delta\mathbf{g}_n} + \frac{\Delta\mathbf{x}_n \Delta\mathbf{x}_n^T}{\Delta\mathbf{x}_n^T \Delta\mathbf{g}_n} \tag{58}$$

DFP 算法的執行步驟如下：

(a) 給定初始點 $\mathbf{x}_0$，計算準確度 ε 和初始矩陣 $\mathbf{D}_0 = \mathbf{I}$，令 $n = 0$。

(b) 計算 $\mathbf{e}_n = -\mathbf{D}_n \mathbf{g}_n$（見 (42) 式），沿 $\mathbf{e}_n$ 進行精確線搜尋，求出步

長 ξ_n ，使得 $J(\mathbf{x}_n+\xi_n\mathbf{e}_n)$ 取極小值。然後令 $\mathbf{x}_{n+1}=\mathbf{x}_n+\xi_n\mathbf{e}_n$ 。

(c) 若 $\|\mathbf{g}_{n+1}\|<\varepsilon$ ，則取 $\mathbf{x}^*=\mathbf{x}_{n+1}$ ，計算結束。否則由 (58) 式計算 $\mathbf{D}_{n+1}$ ，回到 (b)。

DFP 算法的優點是

(a) 若目標函數 $J(\mathbf{x})$ 為 M 元二次多項式，當初始矩陣取 $\mathbf{D}_0=\mathbf{I}$ 時，DFP 算法會產生關於 Hesse 矩陣共軛的搜尋方向，因此最多只需 M 步疊代，就能達到極小點，這個算法有二階收斂性。

(b) 如果 $J(\mathbf{x})$ 為嚴格凸函數，而且使用精確線搜尋，則 DFP 算法是全局收斂的。

(c) 若 $\mathbf{D}_n$ 是對稱正定矩陣，而且 $\Delta\mathbf{x}_n^T\,\Delta\mathbf{g}_n>0$ ，則 $\mathbf{D}_{n+1}$ 也是對稱正定矩陣，這一點不難證明。

DFP 算法的缺點如下：

(a) 需要的儲存量較大，大約需要 $O(M^2)$ 個儲存單元，因此對於大型問題，不如使用共軛梯度法方便。

(b) 數值計算的穩定性不如下面提到的 BFGS 算法。

(c) 在使用不精確線搜尋時，計算效果不如 BFGS 算法。

另一種更新公式稱為 BFGS 算法，這是 Broyden, Fletcher, Goldfarb 和 Shanno 這四個人不約而同的在 1971 年各自發展出來的，這是秩 2 更新，公式如下

$$\mathbf{D}_{n+1}=\mathbf{D}_n+\kappa_1\,\Delta\mathbf{x}\,\Delta\mathbf{x}^T-\kappa_2(\Delta\mathbf{x}\,\Delta\mathbf{g}^T\mathbf{D}_n+\mathbf{D}_n\,\Delta\mathbf{g}_n\,\Delta\mathbf{x}^T) \tag{59}$$

其中

$$\kappa_2=\frac{1}{\Delta\mathbf{x}^T\,\Delta\mathbf{g}_n},\qquad \kappa_1=\kappa_2(1+\kappa_2\,\Delta\mathbf{g}_n^T\mathbf{D}_n\,\Delta\mathbf{g}_n)$$

BFGS 經常與不準確線搜尋一起使用，也就是說在做線搜尋時不需抵達極小點。它具有前面講過的 DFP 算法的優點，但它的數值穩定性要比 DFP 算法好，而且在使用不精確搜尋時，也能證明它是超線性收斂的。因此 BFGS 算法被認為是目前最好的一種算法，獲得廣泛的應用，現有的商業軟體也都有這個算法的程式（Press 等人 1992）。

1.4 約束極小化

在進行變分同化時，需要求目標函數 $J(\mathbf{x})$ 的極小值，但狀態向量 x 的元素之間可能有某種關係存在：$c_i(\mathbf{x})=0$，這個關係稱爲約束條件（constraint）。考慮一般的約束最佳化問題：

$$\begin{cases} \min J(\mathbf{x}) \\ \text{s.t.} \quad c_i(\mathbf{x})=0, \qquad i=1,2,\cdots,p \end{cases} \tag{1}$$

上式中 s. t. 是 subject to 的簡寫，p（$<M$）爲約束條件個數，M 爲狀態向量 x 的元素個數。以後若不加說明，我們總是假定函數 $J(\mathbf{x})$ 和 $c_i(\mathbf{x})$ 具有求解法所要求的各階偏導數。此外，也不考慮不等式約束條件。在氣象問題中不等式約束條件確實存在，例如在自由大氣中未飽和空氣的位溫要隨高度增加，氣象熱力學變數和水物質參數（如混合比等）必須是正值。在數值模擬時爲避免混合比變爲負值，許多正定方案被構造出來。不過在資料同化中不等式約束條件並不重要，故在這裡完全不提。

爲方便說明約束極小化起見，首先考慮求 $J=f(x,y)$ 極值的問題，但另有約束條件，例如 x 和 y 有下面的關係：

$$c(x,y)=0 \tag{2}$$

$J=f(x,y)$ 在 (x^*,y^*) 有極小值的必要條件爲

$$dJ=f_x dx+f_y dy=0 \tag{3}$$

下標 x, y 分別表示關於 x, y 的偏導數。在這情況下，(3) 式中的 dx 和 dy 就不是互相獨立的了，因爲由 (2) 式有

$$c_x dx+c_y dy=0 \tag{4}$$

這是 dx 和 dy 之間的關係。假設 $c_y\neq 0$，則由 (4) 式解出 dy，再代入 (3) 式，得到

$$dJ=(f_x-f_y c_x/c_y)dx=0$$

上式對任何 dx 都成立，故 dx 的係數必須等於零。整理後，有

$$\frac{f_x}{c_x}=\frac{f_y}{c_y}=\lambda \tag{5}$$

其中λ為待求的比例常數，稱為 Lagrange 乘數（Lagrange multiplier）。由(5)式可得到下面兩個方程：

$$f_x-\lambda c_x=0, \qquad f_y-\lambda c_y=0 \tag{6a, b}$$

解出(6a, b)式和原來的約束條件(2)式等 3 個方程，就可求出使 J 取極值的 x 和 y 以及 Lagrange 乘數λ這 3 個未知數的值。顯然的，在約束條件(2)式下求 $J=f(x,y)$ 的極值，相當於求 $J_L=f-\lambda g$ 的極值，但沒有任何約束條件，只要把λ當做控制變數（control variable）就夠了，所謂控制變數是指極小化問題中的未知數。將 J_L 分別對 x, y 和λ微分，然後令結果等於零，就得到(6a)，(6b)和(2)式。新的目標函數 J_L 稱為 Lagrange 函數。

可用另一個說法解釋 Lagrange 乘數λ的作用。(3)式減去(4)式乘以λ，有

$$(f_x-\lambda c_x)dx+(f_y-\lambda c_y)dy=0 \tag{7}$$

上式對任何不等於零的λ都成立。可以選擇一個特別的λ，使得 dx 的係數等於零，由此得到(6a)式。於是(7)式變為

$$(f_y-\lambda c_y)dy=0$$

現在 dy 是任意的，令它的係數等於零就得到(6b)式。Lagrange 乘數本身有幾何或物理意義，引進 Lagrange 乘數可簡化求極值的演算問題，但多了一個未知數。

這種有約束條件的求極值問題，稱為約束極小化（constrained minimization）或約束極大化問題。對於簡單的極小化問題，用 Lagrange 乘數法求解析解比較方便。有時可直接由約束條件解出 y 或 x，然後代入 $J=f(x,y)$，再求 J 的極值，這個方法稱為約束條件消去法或消元法。

根據上面的說明，由(1)式可定義一個 Lagrange（目標）函數：

$$J_{\mathrm{L}}(\mathbf{x},\boldsymbol{\lambda})=J(\mathbf{x})-\sum_{j=1}^{p}\lambda_j c_j(\mathbf{x}) \tag{8}$$

其中$\boldsymbol{\lambda}=[\lambda_1 \quad \lambda_2 \quad \cdots \quad \lambda_p]^T$為 Lagrange 乘數向量。若$\mathbf{x}^*$是(1)式的局部最佳解，而且$\nabla c_j(\mathbf{x})$（$i=1,2,\ \cdots,p$）是線性無關的，則存在一組不全為 0 的實數$\lambda_j^*$（$i=1,2,\ \cdots,p$），使得

$$\nabla J_{\mathrm{L}}(\mathbf{x}^*,\boldsymbol{\lambda}^*)=\nabla J(\mathbf{x}^*)-\sum_{j=1}^{p}\lambda_j^*\nabla c_j(\mathbf{x}^*)=0 \tag{9}$$

其中∇表示關於 x 的梯度。這個必要條件稱為 Lagrange 定理。(9) 式中總和符號前用正號或負號都可以。至於$\mathbf{x}^*$是(1)式的局部極小值的充分條件，仍為$J_{\mathrm{L}}(\mathbf{x},\boldsymbol{\lambda})$的 Hesse 矩陣為正定的，不過需把$\boldsymbol{\lambda}$當做控制變數看待。若把 Lagrange 乘數當做控制變數，不但增加了控制變數的個數，以致增加求解時間，此外 Hesse 矩陣不一定是正定的，以下降算法可能求不出解來。若要以數值方法求解大型約束極小化問題，除了使用約束條件消去法外，最方便的是使用後面提到的伴隨法（adjoint method），絕對不要把 Lagrange 乘數$\boldsymbol{\lambda}$當做控制變數看待。

弱約束條件

上面提到的約束條件(2)式稱為強約束條件（strong constraint）。有時(2)式中的$c(x,y)$只是近似的等於零：

$$c(x,y)\cong 0 \tag{10}$$

(10)式稱為弱約束條件（weak constraint）。求$J=f(x,y)$在弱約束條件(10)式下的極值相當於求罰函數（penalty function）

$$J_{\mathrm{p}}=f(x,y)+\frac{\gamma}{2}c^2(x,y) \tag{11}$$

的極值，其中罰因子（penalty factor）γ是事先給定的正值常數，不是未知數。γ取得越大，則 $c(x,\ y)$越接近於零，得到的解也越接近於強約束條件下的解；取得越小，則 $c(x,\ y)$離零越遠。當γ為有限值時，則約束條件只是近似的滿足而已。弱約束條件法不需引進另

一個未知數，這是優點。

罰約束條件法把極小化問題 (1) 式轉變為一個無約束問題來求解。對 (1) 式這個問題來說，懲罰目標函數（罰函數）是

$$J_{\mathrm{p}}(\mathbf{x},\gamma)=J(\mathbf{x})+\frac{\gamma}{2}\sum_{j=1}^{p}c_j^2(\mathbf{x}) \tag{12}$$

(11) 或 (12) 式右邊第 2 項稱為罰項。當 $\mathbf{x}$ 滿足約束條件時，罰項等於 0；當約束條件被破壞時，罰項變大，而且罰函數也變大，這是 (11) 或 (12) 式稱為罰函數的原因。將 (12) 式取梯度，則在極小點 $\mathbf{x}^*$ 處有

$$\nabla J_{\mathrm{p}}(\mathbf{x}^*,\gamma)=\nabla J(\mathbf{x}^*)+\gamma\sum_{j=1}^{p}c_j(\mathbf{x}^*)\nabla c_j(\mathbf{x}^*)=0$$

根據 Lagrange 定理 (9) 式，若有任何約束條件，則 $\nabla J(\mathbf{x}^*)$ 不一定等於零。若要求 $\nabla J_{\mathrm{p}}(\mathbf{x}^*,\gamma)=0$，則罰因子 γ 必須為無窮大，這是因為 $c_j(\mathbf{x}^*)=0$。因此，當 γ 為有限值時，約束條件只近似的等於 0。

通常以經驗方法估計 (12) 式中的罰因子，然後以下降算法求出使 (12) 式取極小值的 $\mathbf{x}^*$。也可以疊代法選用罰因子，使得 $c_i(\mathbf{x})$ 足夠接近於 0。執行步驟如下：

(a) 選取足夠大的罰因子 $\gamma>0$、準確度 $\varepsilon>0$ 以及初始點 $\mathbf{x}_0$。此時 $n=0$。

(b) 以 $\mathbf{x}_n$ 為初始點，以下降算法求解無約束最佳化問題 $\min J_{\mathrm{p}}(\mathbf{x},\gamma)$，得到最佳解 $\mathbf{x}_{n+1}$。

(c) 令 $\tau=\max\limits_{1\le i\le p}\left|c_i(\mathbf{x}_{n+1})\right|$，若 $\tau<\varepsilon$，則 $\mathbf{x}^*=\mathbf{x}_{n+1}$，否則增加 γ 的值，再回到 (b)。

必須指出，γ 並不是未知數，而是事先給定的值。在實用上都以經驗方法決定，即使以疊代法決定 γ，也只能疊代兩三次，否則太耗費計算時間。

罰函數法的主要缺點之一是，當罰因子 γ 很大時，懲罰目標函數 $J_{\mathrm{p}}(\mathbf{x},\gamma)$ 的 Hesse 矩陣變為病態（ill-conditioned），因而以下降算法求解時收斂速率會變慢。病態的一個度量是條件數（condition number），它是最大特徵值和最小特徵值的比值。假如矩陣的條件數很大的話，則稱為病態矩陣。

例題

做為例子，我們要求定下面函數的極值：

$$J=\frac{1}{2}[(x-2)^2+(y-1)^2+(z-0.1)^2] \tag{13}$$

但受到下面約束條件的限制：

$$x+y+z=3.55 \tag{14}$$

最簡單的方法是消元法，由約束條件解出 z，有 $z=3.55-x-y$，代回(13)式，得到

$$J=\frac{1}{2}[(x-2)^2+(y-1)^2+(3.55-x-y-0.1)^2]$$

現在目標函數只是兩個變數的函數。將 J 分別對 x 和 y 微分，並令結果等於零，有

$$2x+y-5.45=0, \qquad x+2y-4.45=0$$

由上式解出 x 和 y，再代入約束條件，我們發現極小點為

$$x=2.15, \qquad y=1.15, \qquad z=0.25 \tag{15}$$

若使用 Lagrange 乘數法，則需取極值的目標函數（正確的名稱是 Lagrange 函數）為

$$J_{\mathrm{L}}=\frac{1}{2}[(x-2)^2+(y-1)^2+(z-0.1)^2]-\lambda(x+y+z-3.55) \tag{16}$$

將 J 分別對 x, y, z 和 λ 微分，有

$$\frac{\partial J_{\mathrm{L}}}{\partial x}=x-2-\lambda, \qquad \frac{\partial J_{\mathrm{L}}}{\partial y}=y-1-\lambda \tag{17}$$

$$\frac{\partial J_{\mathrm{L}}}{\partial z}=z-0.1-\lambda, \qquad \frac{\partial J_{\mathrm{L}}}{\partial \lambda}=-(x+y+z-3.55) \tag{18}$$

令前三個導數等於零，得到

$$x-2-\lambda=0, \qquad y-1-\lambda=0, \qquad z-0.1-\lambda=0 \tag{19}$$

令 $\partial J_{\mathrm{L}}/\partial\lambda=0$，這就是原來的約束條件 (14) 式。(19) 式中的三個式子相加，並且使用約束條件 (14) 式，得到 $\lambda=0.15$，再代回 (19) 式，我們發現最佳解如 (15) 式所示，顯然的這個解是極小點。至於相應的極小值，則按 (13) 式計算出來。在這裡我們把 Lagrange 乘數 λ 當做控制變數(即未知數）看待。

上面兩個方法都屬於強約束條件法。若用弱約束條件法，則需取極小值的目標函數（正確的名稱是罰函數）為

$$J_{\mathrm{p}}=\frac{1}{2}[(x-2)^2+(y-1)^2+(z-0.1)^2]+\frac{\gamma}{2}(x+y+z-3.55)^2 \tag{20}$$

將上式分別對 x, y, z 微分，並令結果等於零，有

$$x-2+\gamma(x+y+z-3.55)=0 \tag{21}$$

$$y-1+\gamma(x+y+z-3.55)=0 \tag{22}$$

$$z-0.1+\gamma(x+y+z-3.55)=0 \tag{23}$$

上面三個式子相加，得到

$$x+y+z-3.55=-\beta/\gamma$$

其中 $\beta=0.45\gamma/(1+3\gamma)$ 。再代回 (21) 到 (23) 式，我們發現

$$x=2+\beta, \qquad y=1+\beta, \qquad z=0.1+\beta \tag{24}$$

由此可以看出，當 γ 趨近於無窮大時，弱約束條件下的解 (24) 式趨近於強約束條件下的解 (15) 式。

Hesse 矩陣 G 的第 (i, j) 個元素為 $G_{ij}=\partial^2 J/\partial x_i\partial x_j$，其中 J 為目標函數，對約束極小化問題來說 J 是新的目標函數，也就是 Lagrange 函數 J_{L} 或懲罰目標函數 J_{p}。考慮強約束條件法的目標函數 (16) 式。為方便起見，令 $x=x_1$, $y=x_2$, $z=x_3$, $\lambda=x_4$，由此計算出 Hesse 矩陣為

$$\mathbf{G}=\begin{bmatrix}1 & 0 & 0 & -1\\ 0 & 1 & 0 & -1\\ 0 & 0 & 1 & -1\\ -1 & -1 & -1 & 0\end{bmatrix}$$

很容易求出 4 個特徵值如下：$1,1,(1\pm\sqrt{13})/2$。因為其中一個特徵值是負的，故決定出來的極值並不是極小值，也不是極大值。若使用下降算法的程式去進行數值計算，可能求不出解。因此，在進行數值計算時不要把 Lagrange 乘數當做控制變數。

對弱約束條件法的目標函數 (20) 式來說，Hesse 矩陣為

$$\mathbf{G}=\begin{bmatrix}1+\gamma & \gamma & \gamma\\ \gamma & 1+\gamma & \gamma\\ \gamma & \gamma & 1+\gamma\end{bmatrix}$$

這個矩陣的 3 個特徵值如下：$1,1,1+3\gamma$，它們都是正數。因此 G 是正定的，也就是說這個極值的確是極小值。可是當 γ 很大時，Hesse 矩陣變為病態的。

增廣 Lagrange 法

為了求解約束極小化問題，可引進增廣 Lagrange 函數：

$$J_{\mathrm{al}}(\mathbf{x},\boldsymbol{\lambda})=J(\mathbf{x})-\sum_{i=1}^{p}\lambda_j c_j(\mathbf{x})+\frac{\gamma}{2}\sum_{i=1}^{p}c_j^2(\mathbf{x}) \tag{25}$$

將 (25) 式取梯度，有

$$\nabla J_{\mathrm{al}}(\mathbf{x},\boldsymbol{\lambda})=\nabla J(\mathbf{x})-\sum_{i=1}^{p}\lambda_j\nabla c_j(\mathbf{x})+\gamma\sum_{i=1}^{p}c_j(\mathbf{x})\nabla c_j(\mathbf{x})$$

上式右邊最後一項在極小點 $\mathbf{x}^*$ 處的梯度等於零。再利用 (9) 式，得到

$$\nabla J_{\mathrm{al}}(\mathbf{x}^*,\boldsymbol{\lambda})=0 \tag{26}$$

這說明了 $\mathbf{x}^*$ 也是 $J_{\mathrm{al}}(\mathbf{x},\boldsymbol{\lambda})$ 的極小點。必須特別指出，這個方法並不要求罰因子 γ 趨近於無窮大，只要求它大於某個正整數，就能保證無約束最

佳化問題 $\min J_{\mathrm{al}}(\mathbf{x},\boldsymbol{\lambda})$ 的最佳解為原約束最佳化問題(1)式的最佳解。這個方法是罰函數法和二元性算法(duality algorithm)的推廣，它的收斂性已被證明出來，見 Le Dimet 和 Talagrand 1986 的說明，關於二元性算法也請參考這篇文章。

伴隨法

考慮(1)式所示的極小化問題。狀態向量 x 為 $M\times 1$ 向量，因為現在有 p 個約束條件，故個狀態向量的 M 個元素(狀態變數)中只有 $M-p$ 個是獨立的，這種獨立的狀態變數是極小化問題中的未知數，前面已說過，稱為控制變數。其他 p 個元素可由這 $M-p$ 個控制變數透過約束條件計算出來。假如使用約束條件消去法，則只需求解 $M-p$ 個控制變數的最佳化問題，這是最簡便的方法。可是由約束條件通常無法解出其中若干狀態變數，在這情況下只有尋求別的解法。

伴隨法(adjoint method)可用來決定目標函數關於控制變數的梯度，對大型極小化問題來說，這個方法可能是唯一的。這個方法選用 $M-p$ 個控制變數，令這些控制變數為 $x_1,x_2,\cdots,x_{M-p}$。將(8)式對 $x_1,x_2,\cdots,x_M$ 微分，有

$$\frac{\partial J_{\mathrm{L}}}{\partial x_i}=\frac{\partial J(\mathbf{x})}{\partial x_i}-\sum_{j=1}^{p}\lambda_j\frac{\partial c_j(\mathbf{x})}{\partial x_i},\qquad i=1,2,\cdots,M-p \tag{27}$$

$$\frac{\partial J_{\mathrm{L}}}{\partial x_i}=\frac{\partial J(\mathbf{x})}{\partial x_i}-\sum_{j=1}^{p}\lambda_j\frac{\partial c_j(\mathbf{x})}{\partial x_i},\qquad i=M-p+1,M-p+2,\cdots,M \tag{28}$$

後面 p 個 x_i($i=M-p+1,M-p+2,\cdots,M$)不是控制變數，我們令(28)所示的目標函數關於它們的梯度等於零，即

$$\frac{\partial J(\mathbf{x})}{\partial x_i}-\sum_{j=1}^{p}\lambda_j\frac{\partial c_j(\mathbf{x})}{\partial x_i}=0,\qquad i=M-p+1,\ M-p+2,\cdots,\ M \tag{29}$$

上式稱為伴隨方程(adjoint equation)。這是關於 Lagrange 乘數 λ_j

的非線性方程組，若給定$M-p$個控制變數的首次猜測值，按p個約束條件可計算p個非控制變數的狀態變數值。接著由p個伴隨方程求出所有p個λ_j值。λ_j值求出來以後，再按(27)式計算目標函數$J_{\mathrm{L}}(\mathbf{x},\boldsymbol{\lambda})$關於控制變數的梯度。這樣就可使用適當的下降算法，舉例說共軛梯度法，求出極小點。求出極小點的同時，(27)式所示的梯度也趨近於零了。這個方法大幅降低未知數（控制變數）的個數，對求解大型約束最佳化問題相當有利，因此伴隨法是本書的重點之一。

這個方法的疊代步驟如下：

(a) 給定$M-p$個控制變數x_i（$i=1,2,\cdots,M-p$）的首次猜測值。

(b) 由控制變數值按p個約束條件$c_j(\mathbf{x})=0$計算p個非控制變數x_i值（$i=M-p+1,\ M-p+2,\ \cdots,M$），這部分可能需要求解非線性方程組。有了所有的$x_i$以後，計算目標函數$J(\mathbf{x})$。必須指出，因為$c_j(\mathbf{x})=0$，故$J_{\mathrm{L}}(\mathbf{x},\boldsymbol{\lambda})$和$J(\mathbf{x})$完全相等。

(c) 再按p個伴隨方程(29)式計算所有p個λ_j值。有了λ_j以後，按(27)式計算目標函數關於$M-p$個控制變數的梯度。

(d) 使用適合的下降算法（舉例說共軛梯度法或準牛頓法）求出新的控制變數值。

(e) 重複(b)到(d)的步驟，直到解收斂為止。

繼續考慮上面的例題。因為$x,\ y,\ z$以約束條件(14)式連繫起來，故控制變數只有兩個，舉例說選用$x,\ y$。在使用下降算法求解極小化問題時，需要用到目標函數本身和目標函數關於控制變數的梯度值，因此必須令目標函數關於非控制變數z的梯度等於零，由此得到$\lambda=z-0.1$。這個方程稱為伴隨方程。給定控制變數$x,\ y$的首次猜測值，就可由約束條件計算出z，再按伴隨方程計算出λ，然後按(17a, b)式計算目標函數關於控制變數$x,\ y$的梯度，並且按(13)式計算目標函數。有了目標函數和梯度的值以後，就可以下降算法求出新的控制變數，直到收斂為止。伴隨方程一定是線性的，若為偏微分方程，其初始或邊界條件通常是齊次的。

總而言之，一個變分資料同化問題的狀態變數，可能會以約束條

件互相聯繫起來，這是約束極小化問題。對物理問題來說，約束條件一定有誤差，因此弱約束條件法反而比較常用。若使用強約束條件法，則有下面 4 個方案求解：

(a) 消元法。這個方法由 p 個約束條件解出其中 p 個狀態變數，然後代入目標函數，此時目標函數只是另外 $M-p$ 個狀態變數（控制變數）的函數，再用求解無約束極小化問題的下降算法決定出極小點。

(b) Lagrange 乘數法，但把乘數當做未知數。這個方法先構造 Lagrange 函數，然後將所有的狀態變數和 Lagrange 乘數當做控制變數，再決定出極小點。必須指出，這個方法只適合求解簡單的約束極小化問題。在求數值解時，Lagrange 乘數一定要避免當做控制變數，否則不但增加了控制變數的個數，而且以下降算法可能求不出解來。

(c) 增廣 Lagrange 函數法把 Lagrange 乘數當做控制變數，但求解法是收斂的，不過對氣象問題來說由於控制變數太多，因此並不適合於求解大型約束極小化問題。

(d) 伴隨法。這個方法仍需構造 Lagrange 函數，已在上面詳細說明了，不但不把 Lagrange 乘數當做未知數，而且大幅降低未知數個數，適合求解大型的約束極小化問題。

Le Dimet 和 Talagrand（1986）說明了 5 種求解約束極小化問題的方法，包括二元性算法在內，除了二元性算法和增廣 Lagrange 函數法外，本節中都已詳細說明了。

1.5 氣象學中的反問題

氣象學中有許多反問題，其中最主要的是下面幾個：

(a) 四維資料同化是由氣象變數觀測值的時間序列連同預報模式決定最佳初始值，以做為分析、預報之用，至於參數估計（parameter estimation）問題，則是決定出模式中參數的最

佳值。

(b) 衛星遙測是由衛星上的輻射觀測反演出氣象變數，例如大氣中的溫溼垂直分布。

(c) 在 GPS 掩星技術中低軌衛星接收到的 GPS 信號可用來進行同化，以便決定出大氣中的氣溫、氣壓以及水汽壓的垂直分布。地基 GPS 可降水量觀測連同水汽含量的背景值可用來決定最佳的水汽垂直分布。

(d) 客觀分析把分布不規則的測站上的氣象變數觀測值內插到規則分布的格點上，以做為分析、預報之用。

(e) 變分最佳分析調整格點上的氣象變數值，使它們滿足動力方程約束條件，從而使得氣象場能保持內部一致。

(f) 由雷達資料反演出氣象變數，這就是由 Doppler 雷達觀測到的反射率因子（reflectivity factor）和徑向風速的時空分布決定出風場以及其他氣象變數值。

這些反問題，舉例說，都可用變分資料同化法求解。1.2 節中已經說過，變分資料同化法的目標函數可寫為

$$J = \frac{1}{2}(\mathbf{y}_o - \mathbf{y})^T \mathbf{O}^{-1}(\mathbf{y}_o - \mathbf{y}) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) \tag{1}$$

其中 x 和 y 分別為狀態向量和可觀測量，下標 o 和 b 分別表示觀測值和背景值，O 和 B 則分別為實際觀測誤差和背景誤差的協方差矩陣。

至於向前模式，也就是可觀測量和狀態向量之間的關係，最一般性的可寫為

$$\mathbf{\Gamma}(\mathbf{x}, \mathbf{y}) = 0 \tag{2}$$

在大部分的情況下，(2) 式可寫為下面的顯函數形式：

$$\mathbf{y} = \mathbf{h}(\mathbf{x}) \tag{3}$$

上式表示兩個向量之間的非線性關係。假如向前模式是線性的，則 (3) 式可簡化為

$$\mathbf{y} = \mathbf{H}\mathbf{x} \tag{4}$$

必須指出，大部分內插法是線性的，故可用這個式子表示。假如 H 是單位矩陣，則 (4) 式可再進一步簡化成

$$\mathbf{y} = \mathbf{x} \tag{5}$$

此時狀態向量本身就是可觀測量。

上述反問題都是本書的主題，可用 (1) 式和約束條件 (2) 到 (4) 式之一來涵蓋，下面將分別做簡單的說明，更詳細的討論將在本書後面各章中出現。

四維變分同化與參數估計

對四維資料同化問題來說，x 是氣象變數的初始值，y 是氣象變數的時間序列，約束條件 (2) 式則是數值預報模式。由於預報模式無法直接寫出初始值和氣象變數時間序列之間的關係，因此它們是以隱函數 (2) 式出現的。在這情況下，目標函數變為

$$J = \frac{1}{2}(\mathbf{y}_o - \mathbf{y})^T \mathbf{O}^{-1}(\mathbf{y}_o - \mathbf{y}) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \boldsymbol{\lambda}^T \boldsymbol{\Gamma}(\mathbf{x}, \mathbf{y}) \tag{6}$$

其中 $\boldsymbol{\lambda}$ 是 Lagrange 乘數。顯然的上式中沒考慮向前模式誤差。(6) 式所示的目標函數更確切的名稱是 Lagrange 函數。

對參數估計問題來說，x 是預報模式中的參數，y 是氣象變數的時間序列。在這情況下，向前模式仍然是數值預報模式，故目標函數仍為 (6) 式。由於這個問題和四維資料同化本質上是相同的，因此通常把它們相提並論，這兩個問題也可同時處理。

做為一個實際應用的例子，考慮下面的風杯方程：

$$\tau v_t + v = v_e \tag{7}$$

其中下標 t 表示關於時間 t 的導數；τ 和 v_e 分別為時間常數和環境風速，這兩個量是風杯方程中的參數。(7) 式的一個有限差分方案為

$$\tau \frac{v_n - v_{n-1}}{\Delta t} + \frac{v_n + v_{n-1}}{2} = v_e, \qquad n = 1, 2, \cdots, N \tag{8}$$

上式中時間導數採用梯形方案加以離散化。現在若有風速 v 的觀測值

$$\tilde{v}_n = \tilde{v}\,(t_n), \qquad n = 0, 1, 2, \cdots, N \tag{9}$$

我們要由這些觀測值決定最佳的初始值和參數，此時控制變數是

$$\mathbf{x} = [v_0 \quad \tau \quad v_e]^T \tag{10}$$

可觀測量及其觀測值為

$$\mathbf{y} = [v_0 \;\; v_1 \;\; \cdots \;\; v_N]^T, \qquad \mathbf{y}_o = [\tilde{v}_0 \;\; \tilde{v}_1 \;\; \cdots \;\; \tilde{v}_N]^T \tag{11}$$

在這情況下，(6) 式變為

$$J = \frac{1}{2}(\mathbf{y}_o - \mathbf{y})^T \mathbf{O}^{-1}(\mathbf{y}_o - \mathbf{y}) + \boldsymbol{\lambda}^T \boldsymbol{\Gamma}(\mathbf{x}, \mathbf{y}) \tag{12}$$

其中約束條件 (2) 式代表 N 個離散化風杯方程 (8) 式。在這裡因為沒有背景值，故上式中已省去了背景項。必須指出，不論有沒有初始風速 v_0 的觀測值 $\tilde{v}_0$，都可決定出最佳的 v_0。

衛星遙測

對氣溫垂直分布的反演（inversion）問題來說，x 是氣溫垂直分布，y 是氣象衛星上輻射計測溫頻道量出的亮度溫度。向前模式則使用 (3) 式，這個式子代表輻射傳遞方程，也就是由氣溫垂直分布及其他氣候資料決定衛星輻射計測溫頻道所應觀測到的亮度溫度。在這情況下，目標函數如 1.2 節 (53) 式所示：

$$J = \frac{1}{2}[\mathbf{y}_o - \mathbf{h}(\mathbf{x})]^T \mathbf{R}^{-1}[\mathbf{y}_o - \mathbf{h}(\mathbf{x})] + \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) \tag{13}$$

假如數值預報模式中垂直方向有 10 層，輻射傳遞模式也使用那 10 層的氣溫，又設某種輻射計有 7 個測溫頻道，則 x, y 和 h 的維數如下：

x：10×1,　　y：7×1,　　h：7×1

一般說來，預報模式和輻射傳遞模式中所用的垂直層數並不一樣，即使一樣，所在氣壓層也不盡相同。若前者用 10 層，後者用 20 層，則向前模式可寫為

$$\mathbf{y} = \mathbf{h}(\mathbf{x}) \equiv \mathbf{h}_2(\mathbf{V}_1\mathbf{x}) \tag{14}$$

其中 V_1 表示垂直內插或外延，也就是將 10 層的氣溫垂直分布內插外延到輻射傳遞模式所需用到的 20 層上，h_2 仍為輻射傳遞模式。在這情況下 (14) 式中各矩陣的維數如下：

x：10×1,　　y：7×1,　　$\mathbf{V}_1$：20×10,　　$\mathbf{h}_2$：7×1

至於水汽垂直分布的反演，則和氣溫的情況大致一樣。

GPS 氣象學

對於 GPS 掩星技術來說，至少有兩種方法可用來進行資料同化。由低軌衛星 LEO 接收到的 GPS 資料可計算出彎角（bending angle）的垂直分布，隨後由彎角可決定出折射率的垂直分布，故可用彎角或折射指數進行同化。折射率和氣象變數以下面的折射率方程聯繫起來：

$$N = 77.6\frac{p}{T} + 3.73\times10^5\frac{e}{T^2} \tag{15}$$

其中 p 和 e 分別是以 mb 為單位的氣壓和水汽壓，T 是絕對溫度。此外，氣溫和氣壓設為滿足下面的流體靜力方程：

$$\frac{\partial p}{\partial z} = -\frac{p}{R_d T_v} g \tag{16}$$

其中 R_d 為空氣的氣體常數，T_v 為虛溫。由於只有 (15) 和 (16) 式兩個方程，無法由已知的折射率 N 決定出三個未知數 p, T, e，除非 e 或 T 已知，例如在極地或平流層中水汽很少，在這裡就可決定出氣溫的垂直分布。還有，假如已有氣溫垂直分布的預報值或觀測值，也可反演出水汽壓。因此，對 GPS 掩星技術來說，一般說來都以資料同化的手法決定出最佳的 p, T, e 值，而不是直接將它們反演出來。

假如使用折射率來進行同化，則向前模式(15)式可寫為(3)式的形式，目標函數仍為(13)式，其中 x 表示所有垂直層上的三個氣象變數 p, T, e，y 則為折射率的垂直分布。假設完全不用(16)式所示的流體靜力方程，數值預報模式在垂直方向有 10 層，折射率有 50 層的觀測，那麼向前模式可寫為

$$\mathbf{y} = \mathbf{h}(\mathbf{x}) \equiv \mathbf{V}_2 \mathbf{h}_1(\mathbf{x}) \tag{17}$$

其中 $\mathbf{h}_1$ 表示由 10 層氣象變數計算 10 層折射率的方程(15)式，$\mathbf{V}_2$ 表示垂直內插或外延，也就是由 10 層的折射率內插外延到 50 層，計算出 50 層的 N，以便跟 50 層的觀測值 N_o 比較。此時各矩陣的維數為

$$\mathrm{x}：30\times1,\quad \mathrm{y}：50\times1,\quad \mathrm{h}_1：10\times1,\quad \mathrm{V}_2：50\times10 \tag{18}$$

假如直接使用彎角來進行同化，則目標函數仍然是(13)式，但向前模式應改寫為

$$\mathbf{y} = \mathbf{h}(\mathbf{x}) \equiv \mathbf{h}_3[\mathbf{V}_2 \mathbf{h}_1(\mathbf{x})] \tag{19}$$

其中 y 代表彎角的垂直分布，h_3 表示由折射率計算出彎角的過程，這個過程稱為 Abel 變換。

此外，由地面站接收到的 GPS 資料可決定出可降水量（precipitable water）

$$u = \int_0^\infty \rho_v dz \tag{20}$$

其中 ρ_v 為水汽密度，因此這個量更適當的名稱是水汽積分量（integrated water vapor）。由於這個觀測法只能決定出氣柱中的水汽總含量，故不能反演出水汽的垂直分布，但可藉由資料同化法利用初景值決定出最佳的水汽垂直分布。在這情況下，向前模式(20)式是線性的，可寫為(4)式，此時目標函數為

$$J = \frac{1}{2}(\mathbf{y}_o - \mathbf{Hx})^T \mathbf{R}^{-1}(\mathbf{y}_o - \mathbf{Hx}) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) \tag{21}$$

其中 y 是可降水量 u，x 則為水汽密度的垂直分布。若數值預報模式在

垂直方向有 10 層，那麼各個矩陣的維數如下：

$$x：10\times1,\qquad y：1\times1,\qquad H：1\times10 \tag{22}$$

客觀分析

考慮 500 mb 層高度場的客觀分析。此時向前模式是線性的，可寫為 (4) 式，矩陣 H 代表向前內插，也就是由格點上的氣象變數值 x 決定測站上的值 y。這個向前模式比較簡單，可用一些古典的內插法，例如 Lagrange 內插多項式或三次樣條函數（cubic spline）。在這情況下，目標函數為 (21) 式。假如要由 66 個測站上觀測到的 500mb 高度決定出 20×19 個格點上的值，那麼各個矩陣的維數如下：

$$x：380\times1,\qquad y：66\times1,\qquad H：66\times380$$

另一個可能性是三維變分分析，這是同時進行氣溫反演和客觀分析的工作。假定要由 1000 個視點上 7 個測溫頻道晴空亮度溫度觀測決定出 20×19 個格點上的 10 層氣溫，而且數值預報和輻射傳遞方程使用的垂直層完全相同，則向前模式可寫為

$$\mathbf{y}=\mathbf{h}(\mathbf{x})\equiv\mathbf{h}_2(\mathbf{H}_1\mathbf{x}) \tag{23}$$

其中 x 表示 20×19 個格點上的 10 層氣溫，y 表示 1000 個視點上的 7 個測溫頻道總共 7000 個亮度溫度觀測，因此 x 和 y 的維數如下：

$$x：3,800\times1,\qquad y：7,000\times1$$

H_1 代表水平內插，將 20×19 個水平格點上的氣溫值內插到 1000 個視點上，h_2 則是輻射傳遞模式，故 H_1 和 $\mathbf{h}_2$ 的維數如下：

$$H_1：10,000\times3,800,\qquad \mathbf{h}_2：7,000\times1$$

$\mathbf{H}_1\mathbf{x}$ 表示 1000 個視點上的 10 層氣溫。在這裏只是說明問題的本質，在實際計算時並不需使用這樣龐大維數的矩陣。這個問題雖然可以先做反演，決定出每個水平格點上的氣溫垂直分布，再進行客觀分析，但在這情況下觀測誤差的協方差矩陣較難估計，因此使用越原始的資料進行同

化，得到的結果越佳。例如對 GPS 掩星技術來說，使用彎角進行資料同化一般說來要比使用折射率更佳。

變分最佳分析

對變分最佳分析（variational optimization analysis）來說，狀態向量本身就是可觀測量，不過它的元素之間有下面的約束條件：

$$\mathbf{c}(\mathbf{x}) = 0 \tag{24}$$

這個關係是氣象學的動力方程。另外，背景値完全不用，此時目標函數可寫為

$$J = \frac{1}{2}(\mathbf{x} - \mathbf{x}_o)^T \mathbf{O}^{-1}(\mathbf{x} - \mathbf{x}_o) + \boldsymbol{\lambda}^T \mathbf{c}(\mathbf{x}) \tag{25}$$

或者寫為

$$J = \frac{1}{2}(\mathbf{x} - \mathbf{x}_o)^T \mathbf{O}^{-1}(\mathbf{x} - \mathbf{x}_o) + \frac{\gamma}{2}\mathbf{c}^T(\mathbf{x})\mathbf{c}(\mathbf{x}) \tag{26}$$

(25) 式和 (26) 式把 (24) 式分別當做強約束和弱約束條件而得到的。至於狀態向量的觀測値 $\mathbf{x}_o$ 是指格點上的分析値。因此這個由 Sasaki（1958）發展出來的變分最佳分析法，狹義的說，是調整格點上的氣象變數的客觀分析値，使它們滿足做為約束條件的動力方程。必須指出，這個問題並不是反問題，因為完全不用向前模式。

單雷達風場反演

Doppler 雷達接收到的反射率因子 η 和徑向風速 v_r 分別滿足下面的回波守恆方程（conservative equation）和徑向速度方程：

$$\eta_t + u\eta_x + v\eta_y + w\eta_z = S \tag{27}$$

$$v_r = \frac{ux + vy + wz}{r} \tag{28}$$

其中 x, y, z 是以雷達站為中心的直角坐標，r 為大氣中反射電磁波的

某個質點群（取樣體積）離雷達站的距離，u, v, w 為質點群沿這三個坐標的速度分量。此外，下標 x, y, z, t 分別表示關於 x, y, z, t 的偏導數。(27) 式是反射率因子的守恆方程，S 則是很小的源函數（source function），由於雷達資料的時空分辨率（resolution）很高，因此反射率因子的時空梯度是已知的。至於 (28) 式，只是幾何關係而已，也就是由三個風速分量計算出離開原點的徑向風速。假設 (27) 式和 (28) 式的離散化形式分別為

$$\mathbf{Q}(\mathbf{x},\mathbf{z})=\mathbf{S}, \qquad \mathbf{y}=\mathbf{H}\mathbf{x} \tag{29}$$

其中 x 表示三個風速分量，y 表示徑向風速，z 表示所有的反射率因子的時空梯度 η_t, η_x, η_y, η_z。在這情況下目標函數為

$$J=\frac{1}{2}(\mathbf{y}_o-\mathbf{H}\mathbf{x})^T\mathbf{R}^{-1}(\mathbf{y}_o-\mathbf{H}\mathbf{x})+\frac{\gamma}{2}\mathbf{Q}^T(\mathbf{x},\mathbf{z}_o)\mathbf{Q}(\mathbf{x},\mathbf{z}_o) \tag{30}$$

其中 γ 為罰因子，R 為徑向風速觀測誤差的協方差矩陣。這個反演問題是決定最佳的速度分量 u, v, w，使得 (30) 式所示的目標函數取極小值。

結語

最後必須指出，在求解氣象學中的反問題時必須注意下面幾點：

(a) 在求解反問題時必須求解正問題，正問題有時也不見得簡單，例如數值預報模式和輻射傳遞模式就特別複雜，通常需使用別人發展出來的模式。

(b) 在求解反問題時，必須使用下降算法，也就是直接由目標函數和它關於控制變數的梯度，決定出使目標函數取極小值的解。這是因為向前模式可能是非線性的，也可能複雜到必須以隱函數表示，無法用微分等於零的方法決定出解來。

(c) 在求解反問題時需要用到代表準確度的觀測誤差和背景誤差的協方差矩陣，它們隨觀測系統和數值預報模式而不同，必須儘可能準確的估計出來。