

非線性方程組數值解

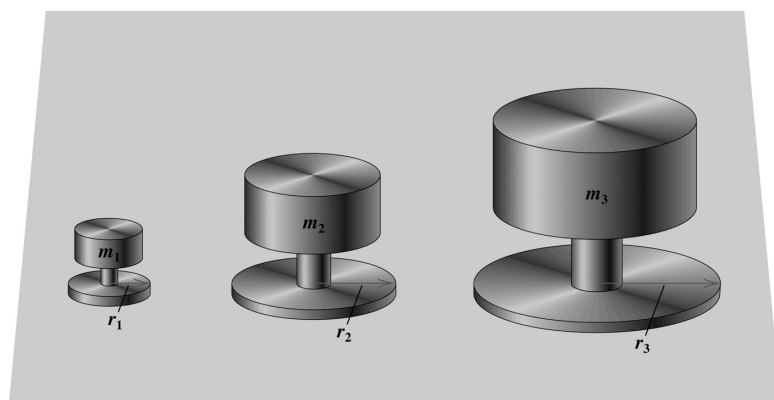
引言

假設在硬質底層之上是一層軟性且均值的土壤，如果要知道，須要多大的壓力才能使一個很大的重物沉入土壤中，我們可以用較小的物體在同樣的土壤上做實驗。明確的說就是，若已知硬質底層位於表面下 $D > d$ 深度處，要使一個半徑為 r 的圓盤沉入土壤表面下 d 深度所須的壓力 p ，可以用下式來近似：

$$p = k_1 e^{k_2 r} + k_3 r$$

其中 k_1 、 k_2 及 k_3 為常數，取決於深度 d 與土壤的一致性，但與圓盤半徑無關。

在此方程式中有 3 個未知常數，所以我們將 3 個不同半徑的小型圓盤，壓入土壤中至同樣深度。如此可決定為能承載大型負重所須的最小面積。記錄對每個圓盤所施加的負載，如下圖所示。



這樣就得到 3 個未知數 k_1 、 k_2 及 k_3 的 3 個非線性方程式

$$m_1 = k_1 e^{k_2 r_1} + k_3 r_1$$

$$m_2 = k_1 e^{k_2 r_2} + k_3 r_2$$

$$m_3 = k_1 e^{k_2 r_3} + k_3 r_3$$

當方程組為非線性時，通常我們必須用數值方法求其近似解。在 10.2 節的習題 12 就是一個這樣的問題。

在可能的情況下我們都設法避免解非線性方程組，通常會使用線性方程組來近似非線性方程組。當這種做法無法獲得滿意的結果時，我們必須直接求解。最直接的方法就是採用我們在第 2 章所介紹的，求解單變數單一非線性方程式的方法，將單變數問題換成包括所有變數的向量問題。

在第 2 章中主要的工具是牛頓法，這種方法通常為二階收斂。這也是我們要修改來解非線性方程組的第一個方法。適用於方程組的牛頓法，須要的計算量極大，所以在 10.3 節中我們將介紹修改的正割法，它可較簡單的求得近似解，但缺點是收斂的比牛頓法慢。

在 10.4 節將說明最陡下降法。這種方法只有線性收斂，但比起其他收斂快速的方法，它不須要精確的初始近似值。所以它常被用來產生牛頓法或類似方法的初始近似值。

在 10.5 節中我們介紹延拓法 (continuation methods)，這種方法利用一個參數，以由一個容易求解的問題轉換到原來非線性問題的解。

在本章中我們省略了大多數理論性部分的證明，因為那通常要用到高等微積分的方法。這部分一般性的參考資料為 Ortega 的著作 *Numerical Analysis—A Second Course* [Or2]。而 [OR] 為更完整的參考。

10.1 多變數函數之固定點

一個非線性方程組，其型式為

$$\begin{aligned} f_1(x_1, x_2, \dots, x_n) &= 0 \\ f_2(x_1, x_2, \dots, x_n) &= 0 \\ &\vdots \\ f_n(x_1, x_2, \dots, x_n) &= 0 \end{aligned} \tag{10.1}$$

其中每個函數 f_i 可以想成是一個，由 n 維空間 $\mathbb{R}^n$ 的向量 $\mathbf{x} = (x_1, x_2, \dots, x_n)^t$ 到實數線 $\mathbb{R}$ 的映射。圖 10.1 則是在 $n = 2$ 時，非線性方程組的幾何表示。

這樣一個包含 n 個未知數 n 個方程式的方程組，也可以經由定義函數 $\mathbf{F}$ 映射 $\mathbb{R}^n$ 到 $\mathbb{R}^n$ 表示成

$$\mathbf{F}(x_1, x_2, \dots, x_n) = (f_1(x_1, x_2, \dots, x_n), f_2(x_1, x_2, \dots, x_n), \dots, f_n(x_1, x_2, \dots, x_n))^t$$

如果用向量的符號來表示變數 $x_1, x_2, \dots, x_n$ ，則方程組 (10.1) 可寫為

$$\mathbf{F}(\mathbf{x}) = \mathbf{0} \tag{10.2}$$

函數 $f_1, f_2, \dots, f_n$ 為 $\mathbf{F}$ 的座標函數 (coordinate functions)。

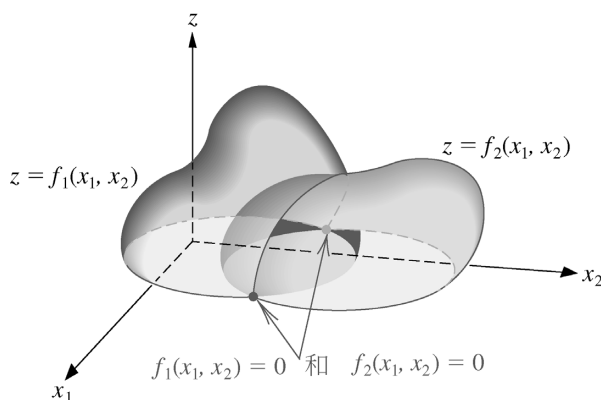


圖 10.1

例題 1 將 3×3 非線性方程組

$$3x_1 - \cos(x_2x_3) - \frac{1}{2} = 0$$

$$x_1^2 - 81(x_2 + 0.1)^2 + \sin x_3 + 1.06 = 0$$

$$e^{-x_1x_2} + 20x_3 + \frac{10\pi - 3}{3} = 0$$

寫成 (10.2) 式的型式。

解 定義由 $\mathbb{R}^3$ 到 $\mathbb{R}$ 的座標函數 f_1 、 f_2 及 f_3 為

$$f_1(x_1, x_2, x_3) = 3x_1 - \cos(x_2x_3) - \frac{1}{2}$$

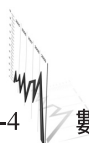
$$f_2(x_1, x_2, x_3) = x_1^2 - 81(x_2 + 0.1)^2 + \sin x_3 + 1.06$$

$$f_3(x_1, x_2, x_3) = e^{-x_1x_2} + 20x_3 + \frac{10\pi - 3}{3}$$

然後定義 $\mathbf{F}$ 由 $\mathbb{R}^3 \rightarrow \mathbb{R}^3$ 為

$$\begin{aligned} \mathbf{F}(\mathbf{x}) &= \mathbf{F}(x_1, x_2, x_3) \\ &= (f_1(x_1, x_2, x_3), f_2(x_1, x_2, x_3), f_3(x_1, x_2, x_3))^t \\ &= \left(3x_1 - \cos(x_2x_3) - \frac{1}{2}, x_1^2 - 81(x_2 + 0.1)^2 \right. \\ &\quad \left. + \sin x_3 + 1.06, e^{-x_1x_2} + 20x_3 + \frac{10\pi - 3}{3} \right)^t \end{aligned}$$

在開始討論求解 (10.1) 式或 (10.2) 式型式問題之前，我們須要一些關於由 $\mathbb{R}^n$ 到 $\mathbb{R}^n$ 函數的連續性與可微性的定義與定理。雖然這些可以用更直接的方式來表達（見習題 12），但我們採用的方式，能夠更深入的呈現由 $\mathbb{R}^n$ 到 $\mathbb{R}$ 之函數的極限與連續性，理論性的概念。



定義 10.1

令 f 為一個定義於集合 $D \subset \mathbb{R}^n$ 的函數且映射到 $\mathbb{R}$ 。我們說函數 f 在 $\mathbf{x}_0$ 處有**極限** (limit) L ，記做

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x}) = L$$

之條件為，對任何已知數 $\varepsilon > 0$ ，存在有一數 $\delta > 0$ ，在 $\mathbf{x} \in D$ 且

$$0 < \|\mathbf{x} - \mathbf{x}_0\| < \delta$$

可使

$$|f(\mathbf{x}) - L| < \varepsilon$$

在 7.1 節已說明過了，極限的存在性與使用何種範數 (norm) 無關。可使用任一種方便的範數以滿足此定義的條件。雖然 δ 的值取決於所選用的範數，但 δ 的存在與否，和所選用的範數無關。

有了極限的概念，我們就可以定義由 $\mathbb{R}^n$ 到 $\mathbb{R}$ 之函數的連續性。雖然可以選擇不同的範數，但連續性同樣與範數的種類無關。

定義 10.2

令 f 為一個由集合 $D \subset \mathbb{R}^n$ 到 $\mathbb{R}$ 的函數。函數 f 在 $\mathbf{x}_0 \in D$ 處為**連續** (continuous) 之前提為，存在有 $\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x})$ 且

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x}) = f(\mathbf{x}_0)$$

此外，如果 f 在 D 的每一點為連續，則 f 在 D 之上為**連續**，記做 $f \in C(D)$ 。

利用由 $\mathbb{R}^n$ 到 $\mathbb{R}$ 的座標函數，我們現在可以定義由 $\mathbb{R}^n$ 到 $\mathbb{R}^n$ 之函數的極限與連續性概念。

定義 10.3

令 $\mathbf{F}$ 為一個由 $D \subset \mathbb{R}^n$ 到 $\mathbb{R}^n$ 的函數，型式為

$$\mathbf{F}(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_n(\mathbf{x}))^t$$

其中對每個 i ， f_i 是一個由 $\mathbb{R}^n$ 到 $\mathbb{R}$ 的映射。我們定義

$$\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \mathbf{F}(\mathbf{x}) = \mathbf{L} = (L_1, L_2, \dots, L_n)^t$$

若且唯若，對每個 $i = 1, 2, \dots, n$ ， $\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f_i(\mathbf{x}) = L_i$ 。

函數 $\mathbf{F}$ 在 $\mathbf{x}_0 \in D$ 處為**連續**之前提為，存在有 $\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \mathbf{F}(\mathbf{x})$ 且 $\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} \mathbf{F}(\mathbf{x}) = \mathbf{F}(\mathbf{x}_0)$ 。另外，如果 $\mathbf{F}$ 在集合 D 中的每一個 $\mathbf{x}$ 為連續，則 $\mathbf{F}$ 在 D 上為連續。此概念記做

$\mathbf{F} \in C(D)$ 。

對於由 $\mathbb{R}$ 到 $\mathbb{R}$ 的函數，通常經由證明其為可微即可說明其為連續 (見定理 1.6)。雖然此定理可一般化到多變數函數，但是多變數函數的導數 (或全導數) 是相當複雜的，我們不在此討論。在此我們用下個定理來取代，下個定理建立了，一個 n 變數函數在某點的連續性，與此函數在該點的偏導數之間的關係。

定理 10.4

令 f 為一個由 $D \subset \mathbb{R}^n$ 到 $\mathbb{R}$ 的函數且 $\mathbf{x}_0 \in D$ 。如果 f 的所有偏導數存在，且有常數 $\delta > 0$ 及 $K > 0$ 使得，在 $\|\mathbf{x} - \mathbf{x}_0\| < \delta$ 且 $\mathbf{x} \in D$ 的時候，我們有

$$\left| \frac{\partial f(\mathbf{x})}{\partial x_j} \right| \leq K, \quad j = 1, 2, \dots, n$$

則 f 在 $\mathbf{x}_0$ 為連續。 ■

■ $\mathbb{R}^n$ 中的固定點

在第 2 章中，我們說明了如何以迭代的程序求解方程式 $f(x) = 0$ ，我們是先將方程式轉換成 $x = g(x)$ 型式的固定點 (fixed-point) 問題。在此我們要探討一個適用於由 $\mathbb{R}^n$ 到 $\mathbb{R}^n$ 函數的類似作法。

定義 10.5

一個由 $D \subset \mathbb{R}^n$ 到 $\mathbb{R}^n$ 的函數 $\mathbf{G}$ ，如果 $\mathbf{G}(\mathbf{p}) = \mathbf{p}$ ，則此函數有固定點 $\mathbf{p} \in D$ 。 ■

下一個定理則將固定點定理 2.4 推廣到 n 維的情況。此定理為壓縮映射定理 (Contraction Mapping Theorem) 的一個特例，在 [Or2], p. 153 有其證明。

定理 10.6

令，對一組常數 $a_1, a_2, \dots, a_n$ 及 $b_1, b_2, \dots, b_n$ ， $D = \{ (x_1, x_2, \dots, x_n)^t \mid a_i \leq x_i \leq b_i, i = 1, 2, \dots, n \}$ 。設 $\mathbf{G}$ 為由 $D \subset \mathbb{R}^n$ 到 $\mathbb{R}^n$ 的連續函數，且在 $\mathbf{x} \in D$ 時 $\mathbf{G}(\mathbf{x}) \in D$ 。則 $\mathbf{G}$ 在 D 中有一個固定點。

再假設， $\mathbf{G}$ 的所有分量函數都有連續偏導數且存在有常數 $K < 1$ ，在 $\mathbf{x} \in D$ 時，對每個 $j = 1, 2, \dots, n$ 及分量函數 g_i ，都有

$$\left| \frac{\partial g_i(\mathbf{x})}{\partial x_j} \right| \leq \frac{K}{n}$$

則由 D 中任選 $\mathbf{x}^{(0)}$ 並依

$$\mathbf{x}^{(k)} = \mathbf{G}(\mathbf{x}^{(k-1)}) \quad , \quad k \geq 1$$

所產生之序列 $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$ ，會收斂到唯一之固定點 $\mathbf{p} \in D$ 且

$$\|\mathbf{x}^{(k)} - \mathbf{p}\|_{\infty} \leq \frac{K^k}{1-K} \|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\|_{\infty} \quad (10.3)$$

■

例題 2 將非線性方程組

$$\begin{aligned} 3x_1 - \cos(x_2x_3) - \frac{1}{2} &= 0 \\ x_1^2 - 81(x_2 + 0.1)^2 + \sin x_3 + 1.06 &= 0 \\ e^{-x_1x_2} + 20x_3 + \frac{10\pi - 3}{3} &= 0 \end{aligned}$$

以第 i 個方程式解 x_i 的方式寫成 $\mathbf{x} = \mathbf{G}(\mathbf{x})$ 的固定點型式，證明在

$$D = \{(x_1, x_2, x_3)^t \mid -1 \leq x_i \leq 1, \quad i = 1, 2, 3\}$$

之上存在有唯一解，並以 $\mathbf{x}^{(0)} = (0.1, 0.1, -0.1)^t$ 為初始值開始迭代，直到 l_{∞} 範數準確到 10^{-5} 為止。

解 以第 i 個方程式解 x_i 可得固定點問題

$$\begin{aligned} x_1 &= \frac{1}{3} \cos(x_2x_3) + \frac{1}{6} \\ x_2 &= \frac{1}{9} \sqrt{x_1^2 + \sin x_3 + 1.06} - 0.1 \\ x_3 &= -\frac{1}{20} e^{-x_1x_2} - \frac{10\pi - 3}{60} \end{aligned} \quad (10.4)$$

令 $\mathbf{G} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ 定義為 $\mathbf{G}(\mathbf{x}) = (g_1(\mathbf{x}), g_2(\mathbf{x}), g_3(\mathbf{x}))^t$ ，其中

$$\begin{aligned} g_1(x_1, x_2, x_3) &= \frac{1}{3} \cos(x_2x_3) + \frac{1}{6} \\ g_2(x_1, x_2, x_3) &= \frac{1}{9} \sqrt{x_1^2 + \sin x_3 + 1.06} - 0.1 \\ g_3(x_1, x_2, x_3) &= -\frac{1}{20} e^{-x_1x_2} - \frac{10\pi - 3}{60} \end{aligned}$$

我們將使用定理 10.4 及 10.6 來證明 $\mathbf{G}$ 在

$$D = \{(x_1, x_2, x_3)^t \mid -1 \leq x_i \leq 1, \quad i = 1, 2, 3\}$$

之上有唯一之固定點。對 D 中的 $\mathbf{x} = (x_1, x_2, x_3)^t$ ，

$$|g_1(x_1, x_2, x_3)| \leq \frac{1}{3} |\cos(x_2x_3)| + \frac{1}{6} \leq 0.50$$

$$|g_2(x_1, x_2, x_3)| = \left| \frac{1}{9} \sqrt{x_1^2 + \sin x_3 + 1.06} - 0.1 \right| \leq \frac{1}{9} \sqrt{1 + \sin 1 + 1.06} - 0.1 < 0.09$$

及

$$|g_3(x_1, x_2, x_3)| = \frac{1}{20} e^{-x_1 x_2} + \frac{10\pi - 3}{60} \leq \frac{1}{20} e + \frac{10\pi - 3}{60} < 0.61$$

所以對每個 $i = 1, 2, 3$ ，我們有

$$-1 \leq g_i(x_1, x_2, x_3) \leq 1$$

因此，在 $\mathbf{x} \in D$ 時 $\mathbf{G}(\mathbf{x}) \in D$ 。

求在 D 之上的偏導數界限得

$$\left| \frac{\partial g_1}{\partial x_1} \right| = 0 \quad , \quad \left| \frac{\partial g_2}{\partial x_2} \right| = 0 \quad , \quad \left| \frac{\partial g_3}{\partial x_3} \right| = 0$$

以及

$$\left| \frac{\partial g_1}{\partial x_2} \right| \leq \frac{1}{3} |x_3| \cdot |\sin x_2 x_3| \leq \frac{1}{3} \sin 1 < 0.281 \quad , \quad \left| \frac{\partial g_1}{\partial x_3} \right| \leq \frac{1}{3} |x_2| \cdot |\sin x_2 x_3| \leq \frac{1}{3} \sin 1 < 0.281$$

$$\left| \frac{\partial g_2}{\partial x_1} \right| = \frac{|x_1|}{9\sqrt{x_1^2 + \sin x_3 + 1.06}} < \frac{1}{9\sqrt{0.218}} < 0.238$$

$$\left| \frac{\partial g_2}{\partial x_3} \right| = \frac{|\cos x_3|}{18\sqrt{x_1^2 + \sin x_3 + 1.06}} < \frac{1}{18\sqrt{0.218}} < 0.119$$

$$\left| \frac{\partial g_3}{\partial x_1} \right| = \frac{|x_2|}{20} e^{-x_1 x_2} \leq \frac{1}{20} e < 0.14 \quad , \quad \text{和} \quad \left| \frac{\partial g_3}{\partial x_2} \right| = \frac{|x_1|}{20} e^{-x_1 x_2} \leq \frac{1}{20} e < 0.14$$

因為 g_1 、 g_2 及 g_3 的偏導數在 D 中為有界，由定理 10.4 可知，這些函數在 D 上為連續。因此 $\mathbf{G}$ 在 D 上為連續。同時，對每個 $\mathbf{x} \in D$ ，

$$\left| \frac{\partial g_i(\mathbf{x})}{\partial x_j} \right| \leq 0.281 \quad , \quad i = 1, 2, 3 \quad , \quad \text{及} \quad j = 1, 2, 3$$

且定理 10.6 第二部分的條件成立，並有 $K = 3(0.281) = 0.843$ 。

以同樣的方法也可證明，對每個 $i = 1, 2, 3$ 及 $j = 1, 2, 3$ ， $\partial g_i / \partial x_j$ 在 D 中為連續（此證明留在習題 3）。最後， $\mathbf{G}$ 在 D 中有唯一之固定點，且非線性方程組在 D 中有解。

要留意的是， $\mathbf{G}$ 在 D 中有唯一之固定點並不代表原來的方程組在此定義域中的解也是唯一的，因為在 (10.4) 式中 x_2 是選用主要平方根。習題 7(d) 則會探討在此選負的平方根會如何。

要求出固定點 $\mathbf{p}$ 的近似值，我們選擇 $\mathbf{x}^{(0)} = (0.1, 0.1, -0.1)^t$ 。由

$$x_1^{(k)} = \frac{1}{3} \cos x_2^{(k-1)} x_3^{(k-1)} + \frac{1}{6}$$

$$x_2^{(k)} = \frac{1}{9} \sqrt{\left(x_1^{(k-1)}\right)^2 + \sin x_3^{(k-1)} + 1.06 - 0.1}$$

$$x_3^{(k)} = -\frac{1}{20} e^{-x_1^{(k-1)} x_2^{(k-1)}} - \frac{10\pi - 3}{60}$$

所產生的向量序列收斂到 (10.4) 式的唯一解。直到滿足

$$\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\|_{\infty} < 10^{-5}$$

為止的結果列於表 10.1。

表 10.1

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$\ \mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\ _{\infty}$
0	0.10000000	0.10000000	-0.10000000	
1	0.49998333	0.00944115	-0.52310127	0.423
2	0.49999593	0.00002557	-0.52336331	9.4×10^{-3}
3	0.50000000	0.00001234	-0.52359814	2.3×10^{-4}
4	0.50000000	0.00000003	-0.52359847	1.2×10^{-5}
5	0.50000000	0.00000002	-0.52359877	3.1×10^{-7}

在前例中使用誤差界限 (10.3) 式及 $K = 0.843$ 可得

$$\|\mathbf{x}^{(5)} - \mathbf{p}\|_{\infty} \leq \frac{(0.843)^5}{1 - 0.843} (0.423) < 1.15$$

這個值無法代表 $\mathbf{x}^{(5)}$ 的真實誤差。此方程組的真實解為

$$\mathbf{p} = \left(0.5, 0, -\frac{\pi}{6}\right)^t \approx (0.5, 0, -0.5235987757)^t, \quad \text{所以 } \|\mathbf{x}^{(5)} - \mathbf{p}\|_{\infty} \leq 2 \times 10^{-8}$$

■ 加速收斂

利用類似解線性方程組的高斯-賽德法，在計算 $x_i^{(k)}$ 時，使用新獲得的 $x_1^{(k)}, \dots, x_{i-1}^{(k)}$ 取代 $x_1^{(k-1)}, \dots, x_{i-1}^{(k-1)}$ ，以加速固定點迭代收斂速度。前面例題中的各分量方程式成為

$$x_1^{(k)} = \frac{1}{3} \cos\left(x_2^{(k-1)} x_3^{(k-1)}\right) + \frac{1}{6}$$

$$x_2^{(k)} = \frac{1}{9} \sqrt{\left(x_1^{(k)}\right)^2 + \sin x_3^{(k-1)} + 1.06 - 0.1}$$

$$x_3^{(k)} = -\frac{1}{20} e^{-x_1^{(k)} x_2^{(k)}} - \frac{10\pi - 3}{60}$$

使用 $\mathbf{x}^{(0)} = (0.1, 0.1, -0.1)^t$ ，計算結果列於表 10.2。

表 10.2

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$\ \mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\ _\infty$
0	0.10000000	0.10000000	-0.10000000	
1	0.49998333	0.02222979	-0.52304613	0.423
2	0.49997747	0.00002815	-0.52359807	2.2×10^{-2}
3	0.50000000	0.00000004	-0.52359877	2.8×10^{-5}
4	0.50000000	0.00000000	-0.52359877	3.8×10^{-8}

迭代到 $\mathbf{x}^{(4)}$ 的 l_∞ 範數的誤差值已小於 10^{-7} 。所以對這個問題，高斯-賽德法確實能加速收斂，但不保證在所有情形下都能加速收斂。

在 Maple 中函數 *fsolve* 可求解聯立方程組。例題 2 的固定點問題，可以用以下指令求解：

```
g1 := x1 = 1/3 cos(x2*x3) + 1/6 : g2 := x2 = 1/9 sqrt(x1)^2 + sin(x3) + 1.06 - 0.1 :
g3 := x3 = -1/20 e^(-x1*x2) - 10*pi-3/60 :
fsolve({g1,g2,g3},{x1,x2,x3},{x1=-1..1,x2=-1..1,x3=-1..1});
```

前 3 個指令定義出方程組，最後一個指令則呼叫 *fsolve* 函數。顯示出來的答案為

```
{x1 = 0.5000000000, x2 = -2.079196195 10^-11, x3 = -0.5235987758}
```

在以 *fsolve(eqns,vars,options)* 指令求解方程組時，參數 *eqns* 代表已知的方程式，*vars* 代表要求解的未知數，*options* 為選項條件。在 *options* 選項中我們可以設定求解的區間。但此種設定並非必要的，如果省略此參數，Maple 會自行決定合適的區間。

習題組 10.1 完整習題請見隨書光碟

1. 證明，定義為

$$\mathbf{F}(x_1, x_2, x_3) = (x_1 + 2x_3, x_1 \cos x_2, x_2^2 + x_3)'$$

的函數 $\mathbf{F} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ 在 $\mathbb{R}^3$ 的每一點為連續。

- 寫出一個函數的例子，此函數 $\mathbf{F} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ 在 $\mathbb{R}^2$ 中除了 $(1, 0)$ 外在每一點為連續。
- 證明，例題 2 之一階偏導數在 D 上連續。
- 非線性方程組

$$-x_1(x_1 + 1) + 2x_2 = 18 \quad , \quad (x_1 - 1)^2 + (x_2 - 6)^2 = 25$$

有 2 個解。

- 以圖形法求其近似解。
- 以 (a) 的結果做為初始近似值，以迭代法求其近似解，準確到 l_∞ 範數小於 10^{-5} 。

5. 非線性方程組

$$x_1^2 - 10x_1 + x_2^2 + 8 = 0, \quad x_1x_2^2 + x_1 - 10x_2 + 8 = 0$$

可以轉換成固定點問題

$$x_1 = g_1(x_1, x_2) = \frac{x_1^2 + x_2^2 + 8}{10}, \quad x_2 = g_2(x_1, x_2) = \frac{x_1x_2^2 + x_1 + 8}{10}$$

- a. 用定理 10.6 以證明, $\mathbf{G} = (g_1, g_2)^t$ 映射 $D \subset \mathbb{R}^2$ 到 $\mathbb{R}^2$, 在

$$D = \{(x_1, x_2)^t \mid 0 \leq x_1, x_2 \leq 1.5\}$$

中有唯一之固定點。

- b. 使用泛函迭代求其近似解。
c. 高斯-賽德法是否可加速收斂？

6. 非線性方程組

$$5x_1^2 - x_2^2 = 0, \quad x_2 - 0.25(\sin x_1 + \cos x_2) = 0$$

在 $(\frac{1}{4}, \frac{1}{4})^t$ 附近有解。

- a. 求 $\mathbb{R}^2$ 中的函數 $\mathbf{G}$ 與集合 D 以使得 $\mathbf{G} : D \rightarrow \mathbb{R}^2$ 且 $\mathbf{G}$ 在 D 中有唯一之固定點。
b. 使用泛函迭代求其近似解, 準確到 l_∞ 範數小於 10^{-5} 。
c. 高斯-賽德法是否可加速收斂？
7. 用定理 10.6 以證明 $\mathbf{G} : D \subset \mathbb{R}^3 \rightarrow \mathbb{R}^3$ 在 D 中有唯一固定點。使用泛函迭代求其近似解, 準確到 l_∞ 範數小於 10^{-5} 。

a. $\mathbf{G}(x_1, x_2, x_3) = \left(\frac{\cos(x_2x_3) + 0.5}{3}, \frac{1}{25}\sqrt{x_1^2 + 0.3125} - 0.03, -\frac{1}{20}e^{-x_1x_2} - \frac{10\pi - 3}{60} \right)^t$

$$D = \{(x_1, x_2, x_3)^t \mid -1 \leq x_i \leq 1, i = 1, 2, 3\}$$

b. $\mathbf{G}(x_1, x_2, x_3) = \left(\frac{13 - x_2^2 + 4x_3}{15}, \frac{11 + x_3 - x_1^2}{10}, \frac{22 + x_3^3}{25} \right)$

$$D = \{(x_1, x_2, x_3)^t \mid 0 \leq x_i \leq 1.5, i = 1, 2, 3\}$$

c. $\mathbf{G}(x_1, x_2, x_3) = (1 - \cos(x_1x_2x_3), 1 - (1 - x_1)^{1/4} - 0.05x_3^2 + 0.15x_3, x_1^2 + 0.1x_2^2 - 0.01x_2 + 1)^t$

$$D = \{(x_1, x_2, x_3)^t \mid -0.1 \leq x_1 \leq 0.1, -0.1 \leq x_2 \leq 0.3, 0.5 \leq x_3 \leq 1.1\}$$

d. $\mathbf{G}(x_1, x_2, x_3) = \left(\frac{1}{3}\cos(x_2x_3) + \frac{1}{6}, -\frac{1}{9}\sqrt{x_1^2 + \sin x_3 + 1.06} - 0.1 - \frac{1}{20}e^{-x_1x_2} - \frac{10\pi - 3}{60} \right)^t$

$$D = \{(x_1, x_2, x_3)^t \mid -1 \leq x_i \leq 1, i = 1, 2, 3\}$$

8. 用泛函迭代求下列非線性方程組的近似解, 準確到 l_∞ 範數小於 10^{-5} 。

a. $x_2^2 + x_2^2 - x_1 = 0$

b. $3x_1^2 - x_2^2 = 0$

$x_1^2 - x_2^2 - x_2 = 0$

$3x_1x_2^2 - x_1^3 - 1 = 0$

$$\text{c. } x_1^2 + x_2 - 37 = 0$$

$$x_1 - x_2^2 - 5 = 0$$

$$x_1 + x_2 + x_3 - 3 = 0$$

$$\text{d. } x_1^2 + 2x_2^2 - x_2 - 2x_3 = 0$$

$$x_1^2 - 8x_2^2 + 10x_3 = 0$$

$$\frac{x_1^2}{7x_2x_3} - 1 = 0$$

9. 用高斯-賽德法求習題 7 各固定點的近似值，至 l_∞ 範數小於 10^{-5} 。
10. 用高斯-賽德法重做習題 8。
11. 在 5.9 節的習題 10 中，我們考慮了預測 2 種競爭物種族群數的問題。在此問題中我們假設，可以用方程組

$$\frac{dx_1(t)}{dt} = x_1(t)(4 - 0.0003x_1(t) - 0.0004x_2(t))$$

及

$$\frac{dx_2(t)}{dt} = x_2(t)(2 - 0.0002x_1(t) - 0.0001x_2(t))$$

來預測族群數的變化。在本題中，我們要考慮這 2 種物種平衡狀態下族群數的問題。2 種物種處於平衡狀態的數學條件為

$$\frac{dx_1(t)}{dt} = 0 \quad \text{及} \quad \frac{dx_2(t)}{dt} = 0$$

這種狀態出現於，第 1 種物種滅絕，而第 2 種物種族群數為 20,000，或第 2 種物種滅絕第 1 種物種族群數 13,333。是否還有其他的平衡態？

12. 證明，函數 $\mathbf{F}$ 映射 $D \subset \mathbb{R}^n$ 到 $\mathbb{R}^n$ 在 $\mathbf{x}_0 \in D$ 為連續的條件為，給定任何數 $\varepsilon > 0$ ，必可找到一個數 $\delta > 0$ 可使得，對任何向量範數 $\|\cdot\|$ ，

$$\|\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{x}_0)\| < \varepsilon$$

只要 $\mathbf{x} \in D$ 且 $\|\mathbf{x} - \mathbf{x}_0\| < \delta$ 。

13. 令 A 為 $n \times n$ 矩陣且 $\mathbf{F}$ 為由 $\mathbb{R}^n$ 到 $\mathbb{R}^n$ 的函數，定義為 $\mathbf{F}(\mathbf{x}) = A\mathbf{x}$ 。利用習題 12 的結果以證明 $\mathbf{F}$ 在 $\mathbb{R}^n$ 上為連續。



10.2 牛頓法

在 10.1 節的例題 2 中，我們用代數的方式解出變數 x_1 、 x_2 及 x_3 ，然後將問題轉換成收斂的固定點問題。但是在實際應用中，要找到所有變數之外顯解的機會不大。在本節中，我們將以算則般逐步的方式，說明一般情況下的轉換。

要建構一個一維情形的固定點算則，我們發現，滿足

$$g(x) = x - \phi(x)f(x)$$

的函數 ϕ 可以用二階的速度收斂到函數 g 的固定點 p (見 2.4 節)。利用此條件，牛頓法由選擇 $\phi(x) = 1/f'(x)$ 開始，假設 $f'(x) \neq 0$ 。

用類似的方法， n 維的時候包括了矩陣

$$A(\mathbf{x}) = \begin{bmatrix} a_{11}(\mathbf{x}) & a_{12}(\mathbf{x}) & \cdots & a_{1n}(\mathbf{x}) \\ a_{21}(\mathbf{x}) & a_{22}(\mathbf{x}) & \cdots & a_{2n}(\mathbf{x}) \\ \vdots & \vdots & & \vdots \\ a_{n1}(\mathbf{x}) & a_{n2}(\mathbf{x}) & \cdots & a_{nn}(\mathbf{x}) \end{bmatrix} \quad (10.5)$$

其中每個元素 $a_{ij}(\mathbf{x})$ 為由 $\mathbb{R}^n$ 到 $\mathbb{R}$ 的函數。這須要找到 $A(\mathbf{x})$ 可使得

$$\mathbf{G}(\mathbf{x}) = \mathbf{x} - A(\mathbf{x})^{-1}\mathbf{F}(\mathbf{x})$$

以二階的速度收斂到 $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ 的解，前提是在 $\mathbf{G}$ 的固定點 $\mathbf{p}$ 處 $A(\mathbf{x})$ 為非奇異。

下個定理與 2.4 節的定理 2.8 類似。此定理的證明須要將 $\mathbf{G}$ 表示成相對於 $\mathbf{p}$ 的 n 變數泰勒級數。

定理 10.7

令 $\mathbf{p}$ 為 $\mathbf{G}(\mathbf{x}) = \mathbf{x}$ 的解。設存在一數 $\delta > 0$ 使得

- (i) $\partial g_i / \partial x_j$ 在 $N_\delta = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{p}\| < \delta\}$ 為連續， $i = 1, 2, \dots, n$ 且 $j = 1, 2, \dots, n$ ；
- (ii) 對所有 $i = 1, 2, \dots, n$ 、 $j = 1, 2, \dots, n$ 及 $k = 1, 2, \dots, n$ ，當 $\mathbf{x} \in N_\delta$ ， $\partial^2 g_i(\mathbf{x}) / (\partial x_j \partial x_k)$ 為連續且有某常數 M 使得 $|\partial^2 g_i(\mathbf{x}) / (\partial x_j \partial x_k)| \leq M$ ；
- (iii) 對所有 $i = 1, 2, \dots, n$ 及 $k = 1, 2, \dots, n$ ， $\partial g_i(\mathbf{p}) / \partial x_k = 0$ 。

則存在有一數 $\hat{\delta} \leq \delta$ ，使得 $\mathbf{x}^{(k)} = \mathbf{G}(\mathbf{x}^{(k-1)})$ 所產生的序列以二階方式收斂到 $\mathbf{p}$ ，其 $\mathbf{x}^{(0)}$ 為任意選擇的並滿足 $\|\mathbf{x}^{(0)} - \mathbf{p}\| < \hat{\delta}$ 。另外，在 $k \geq 1$ 時

$$\|\mathbf{x}^{(k)} - \mathbf{p}\|_\infty \leq \frac{n^2 M}{2} \|\mathbf{x}^{(k-1)} - \mathbf{p}\|_\infty^2, \quad k \geq 1$$

要使用定理 10.7，假設 $A(\mathbf{x})$ 是 $n \times n$ 由 $\mathbb{R}^n$ 到 $\mathbb{R}$ 的函數矩陣，如 (10.5) 式的型式，其各元素的值將於稍後決定。再假設在 $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ 的解 $\mathbf{p}$ 的附近， $A(\mathbf{x})$ 為非奇異，並令 $b_{ij}(\mathbf{x})$ 代表 $A(\mathbf{x})^{-1}$ 第 i 列第 j 行的元素。

因為 $\mathbf{G}(\mathbf{x}) = \mathbf{x} - A(\mathbf{x})^{-1}\mathbf{F}(\mathbf{x})$ ，我們有 $g_i(\mathbf{x}) = x_i - \sum_{j=1}^n b_{ij}(\mathbf{x}) f_j(\mathbf{x})$ 。所以

$$\frac{\partial g_i}{\partial x_k}(\mathbf{x}) = \begin{cases} 1 - \sum_{j=1}^n \left(b_{ij}(\mathbf{x}) \frac{\partial f_j}{\partial x_k}(\mathbf{x}) + \frac{\partial b_{ij}}{\partial x_k}(\mathbf{x}) f_j(\mathbf{x}) \right), & \text{若 } i = k \\ - \sum_{j=1}^n \left(b_{ij}(\mathbf{x}) \frac{\partial f_j}{\partial x_k}(\mathbf{x}) + \frac{\partial b_{ij}}{\partial x_k}(\mathbf{x}) f_j(\mathbf{x}) \right), & \text{若 } i \neq k \end{cases}$$

由定理 10.7 可知，對所有 $i = 1, 2, \dots, n$ 及 $k = 1, 2, \dots, n$ ，必須要有 $\partial g_i(\mathbf{p}) / \partial x_k = 0$ 。這代表當 $i = k$ 時，

$$0 = 1 - \sum_{j=1}^n b_{ij}(\mathbf{p}) \frac{\partial f_j}{\partial x_i}(\mathbf{p})$$

也就是

$$\sum_{j=1}^n b_{ij}(\mathbf{p}) \frac{\partial f_j}{\partial x_i}(\mathbf{p}) = 1 \quad (10.6)$$

當 $k \neq i$,

$$0 = - \sum_{j=1}^n b_{ij}(\mathbf{p}) \frac{\partial f_j}{\partial x_k}(\mathbf{p})$$

所以

$$\sum_{j=1}^n b_{ij}(\mathbf{p}) \frac{\partial f_j}{\partial x_k}(\mathbf{p}) = 0 \quad (10.7)$$

■ Jacobian 矩陣

定義矩陣 $J(\mathbf{x})$ 為

$$J(\mathbf{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \frac{\partial f_1}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{x}) & \frac{\partial f_2}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial f_2}{\partial x_n}(\mathbf{x}) \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1}(\mathbf{x}) & \frac{\partial f_n}{\partial x_2}(\mathbf{x}) & \cdots & \frac{\partial f_n}{\partial x_n}(\mathbf{x}) \end{bmatrix} \quad (10.8)$$

由 (10.6) 與 (10.7) 式的條件可看出

$$A(\mathbf{p})^{-1}J(\mathbf{p}) = I, \text{ 為單位矩陣, 所以 } A(\mathbf{p}) = J(\mathbf{p}).$$

因此, 我們可以選擇 $A(\mathbf{x}) = J(\mathbf{x})$, 因為這滿足定理 10.7 之 (iii) 的條件。函數 $\mathbf{G}$ 定義為

$$\mathbf{G}(\mathbf{x}) = \mathbf{x} - J(\mathbf{x})^{-1}\mathbf{F}(\mathbf{x})$$

泛函迭代的程序由選擇 $\mathbf{x}^{(0)}$ 開始, 並在 $k \geq 1$ 時產生

$$\mathbf{x}^{(k)} = \mathbf{G}(\mathbf{x}^{(k-1)}) = \mathbf{x}^{(k-1)} - J(\mathbf{x}^{(k-1)})^{-1}\mathbf{F}(\mathbf{x}^{(k-1)}) \quad (10.9)$$

這叫做**非線性方程組的牛頓法** (Newton's method for nonlinear systems), 通常它會以二階收斂, 先決條件為初始近似值夠準確且 $J(\mathbf{p})^{-1}$ 存在。矩陣 $J(\mathbf{x})$ 叫做 **Jacobian 矩陣**, 並且有多種用途。就讀者所熟悉的, 它可用於多變數函數的多重積分, 當須要變數轉換時就會用到它。

牛頓法的缺點是必須在每次迭代時求出 $J(\mathbf{x})$ 矩陣及其逆矩陣。在實用上, 可採用一

種兩階段運算，以避免直接計算 $J(\mathbf{x})^{-1}$ 。首先，求出一個滿足 $J(\mathbf{x}^{(k-1)})\mathbf{y} = -\mathbf{F}(\mathbf{x}^{(k-1)})$ 的向量 $\mathbf{y}$ 。然後將 $\mathbf{y}$ 加上 $\mathbf{x}^{(k)}$ 以得到新的近似值 $\mathbf{x}^{(k+1)}$ 。算則 10.1 即使用這種兩階段程序。

算則 10.1 解方程組之牛頓法 (Newton's Method for Systems)

已知初始近似值 $\mathbf{x}$ ，求非線性方程組 $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ 之近似解：

INPUT 方程式與未知數的個數 n ；初始近似值 $\mathbf{x} = (x_1, \dots, x_n)^t$ ；容許誤差 TOL ；最大迭代次數 N 。

OUTPUT 近似解 $\mathbf{x} = (x_1, \dots, x_n)^t$ 或超過最大迭代次數之訊息。

Step 1 Set $k = 1$.

Step 2 While $(k \leq N)$ do Steps 3–7.

Step 3 Calculate $\mathbf{F}(\mathbf{x})$ and $J(\mathbf{x})$, where $J(\mathbf{x})_{ij} = (\partial f_i(\mathbf{x}) / \partial x_j)$ for $1 \leq i, j \leq n$.

Step 4 Solve the $n \times n$ linear system $J(\mathbf{x})\mathbf{y} = -\mathbf{F}(\mathbf{x})$.

Step 5 Set $\mathbf{x} = \mathbf{x} + \mathbf{y}$.

Step 6 If $\|\mathbf{y}\| < TOL$ then OUTPUT $(\mathbf{x})$;
(程式成功)
STOP.

Step 7 Set $k = k + 1$.

Step 8 OUTPUT ('Maximum number of iterations exceeded');
(程式失敗)
STOP.

例題 1 非線性方程組

$$\begin{aligned} 3x_1 - \cos(x_2x_3) - \frac{1}{2} &= 0 \\ x_1^2 - 81(x_2 + 0.1)^2 + \sin x_3 + 1.06 &= 0 \\ e^{-x_1x_2} + 20x_3 + \frac{10\pi - 3}{3} &= 0 \end{aligned}$$

已經在 10.1 節的例題 2 求得一個近似解 $(0.5, 0, -0.52359877)^t$ 。以 $\mathbf{x}^{(0)} = (0.1, 0.1, -0.1)^t$ 為初始值，用牛頓法求解此問題。

解 定義

$$\mathbf{F}(x_1, x_2, x_3) = (f_1(x_1, x_2, x_3), f_2(x_1, x_2, x_3), f_3(x_1, x_2, x_3))^t$$

其中

$$\begin{aligned} f_1(x_1, x_2, x_3) &= 3x_1 - \cos(x_2x_3) - \frac{1}{2} \\ f_2(x_1, x_2, x_3) &= x_1^2 - 81(x_2 + 0.1)^2 + \sin x_3 + 1.06 \end{aligned}$$

及

$$f_3(x_1, x_2, x_3) = e^{-x_1 x_2} + 20x_3 + \frac{10\pi - 3}{3}$$

此方程組之 Jacobian 矩陣 $J(\mathbf{x})$ 為

$$J(x_1, x_2, x_3) = \begin{bmatrix} 3 & x_3 \sin x_2 x_3 & x_2 \sin x_2 x_3 \\ 2x_1 & -162(x_2 + 0.1) & \cos x_3 \\ -x_2 e^{-x_1 x_2} & -x_1 e^{-x_1 x_2} & 20 \end{bmatrix}$$

令 $\mathbf{x}^{(0)} = (0.1, 0.1, -0.1)^t$ 。則 $\mathbf{F}(\mathbf{x}^{(0)}) = (-0.199995, -2.269833417, 8.462025346)^t$

且

$$J(\mathbf{x}^{(0)}) = \begin{bmatrix} 3 & 9.999833334 \times 10^{-4} & 9.999833334 \times 10^{-4} \\ 0.2 & -32.4 & 0.9950041653 \\ -0.09900498337 & -0.09900498337 & 20 \end{bmatrix}$$

解線性方程組 $J(\mathbf{x}^{(0)})\mathbf{y}^{(0)} = -\mathbf{F}(\mathbf{x}^{(0)})$ 得

$$\mathbf{y}^{(0)} = \begin{bmatrix} 0.3998696728 \\ -0.08053315147 \\ -0.4215204718 \end{bmatrix} \text{ 及 } \mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \mathbf{y}^{(0)} = \begin{bmatrix} 0.4998696728 \\ 0.01946684853 \\ -0.5215204718 \end{bmatrix}$$

依 $k = 2, 3, \dots$ 持續進行，可得

$$\begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \\ x_3^{(k)} \end{bmatrix} = \begin{bmatrix} x_1^{(k-1)} \\ x_2^{(k-1)} \\ x_3^{(k-1)} \end{bmatrix} + \begin{bmatrix} y_1^{(k-1)} \\ y_2^{(k-1)} \\ y_3^{(k-1)} \end{bmatrix}$$

其中

$$\begin{bmatrix} y_1^{(k-1)} \\ y_2^{(k-1)} \\ y_3^{(k-1)} \end{bmatrix} = - \left(J \left(x_1^{(k-1)}, x_2^{(k-1)}, x_3^{(k-1)} \right) \right)^{-1} \mathbf{F} \left(x_1^{(k-1)}, x_2^{(k-1)}, x_3^{(k-1)} \right)$$

因此，在第 k 步時，我們必須解線性方程組 $J(\mathbf{x}^{(k-1)})\mathbf{y}^{(k-1)} = -\mathbf{F}(\mathbf{x}^{(k-1)})$ ，其中

$$J(\mathbf{x}^{(k-1)}) = \begin{bmatrix} 3 & x_3^{(k-1)} \sin x_2^{(k-1)} x_3^{(k-1)} & x_2^{(k-1)} \sin x_2^{(k-1)} x_3^{(k-1)} \\ 2x_1^{(k-1)} & -162 \left(x_2^{(k-1)} + 0.1 \right) & \cos x_3^{(k-1)} \\ -x_2^{(k-1)} e^{-x_1^{(k-1)} x_2^{(k-1)}} & -x_1^{(k-1)} e^{-x_1^{(k-1)} x_2^{(k-1)}} & 20 \end{bmatrix}$$

$$\mathbf{y}^{(k-1)} = \begin{bmatrix} y_1^{(k-1)} \\ y_2^{(k-1)} \\ y_3^{(k-1)} \end{bmatrix}$$

及

$$\mathbf{F}(\mathbf{x}^{(k-1)}) = \begin{bmatrix} 3x_1^{(k-1)} - \cos x_2^{(k-1)} x_3^{(k-1)} - \frac{1}{2} \\ (x_1^{(k-1)})^2 - 81(x_2^{(k-1)} + 0.1)^2 + \sin x_3^{(k-1)} + 1.06 \\ e^{-x_1^{(k-1)} x_2^{(k-1)}} + 20x_3^{(k-1)} + \frac{10\pi-3}{3} \end{bmatrix}$$

使用此一程序所獲得的結果列於表 10.3。

表 10.3

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$\ \mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\ _\infty$
0	0.1000000000	0.1000000000	-0.1000000000	
1	0.4998696728	0.0194668485	-0.5215204718	0.4215204718
2	0.5000142403	0.0015885914	-0.5235569638	1.788×10^{-2}
3	0.5000000113	0.0000124448	-0.5235984500	1.576×10^{-3}
4	0.5000000000	8.516×10^{-10}	-0.5235987755	1.244×10^{-5}
5	0.5000000000	-1.375×10^{-11}	-0.5235987756	8.654×10^{-10}

在前一個例題中，我們看到了，只要有一個夠好的初始值，牛頓法收斂的非常快。但是要找到好的初始值也不容易，而且牛頓法的計算成本相對是較高的。在下一節中我們會介紹一種方法以克服第二個缺點。在 10.4 節則將介紹，以最陡下降法獲得良好的初始值。

■ 用 Maple 求初始近似值

對於 2×2 或某些 3×3 的非線性方程組，可以用 Maple 的繪圖功能獲得良好的初始近似值。例如非線性方程組

$$x_1^2 - x_2^2 + 2x_2 = 0, \quad 2x_1 + x_2^2 - 6 = 0$$

有 2 個解 $(0.625204094, 2.179355825)^t$ 及 $(2.109511920, -1.334532188)^t$ 。在 Maple 中我們先定義 2 個方程式

$$eq1 := x1^2 - x2^2 + 2x2 = 0; \quad eq2 := 2x1 + x2^2 - 6 = 0;$$

要獲得這 2 個方程式在 $-3 \leq x_1, x_2 \leq 3$ 範圍的圖形，輸入指令

$$\text{with}(plots): \text{implicitplot}(\{eq1, eq2\}, x1 = -6..6, x2 = -6..6);$$

由圖 10.2 所示的圖形，我們可以估計出來在 $(2.1, -1.3)^t$ 、 $(0.64, 2.2)^t$ 、 $(-1.9, 3.0)^t$ 與 $(-5.0, -4.0)^t$ 的附近有解。這就足夠做為牛頓法的初始值了。

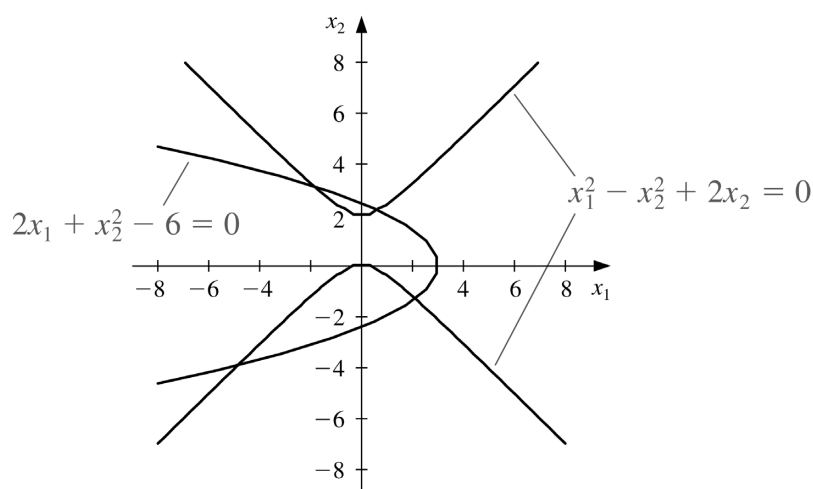


圖 10.2

在 3D 的時候問題複雜得多。考慮以下非線性方程組

$$2x_1 - 3x_2 + x_3 - 4 = 0, \quad 2x_1 + x_2 - x_3 + 4 = 0, \quad x_1^2 + x_2^2 + x_3^2 - 4 = 0$$

用 Maple 指令定義 3 個方程式

$$eq1 := 2x_1 - 3x_2 + x_3 - 4 = 0; eq2 := 2x_1 + x_2 - x_3 + 4 = 0; eq3 := x_1^2 + x_2^2 + x_3^2 - 4 = 0;$$

第 3 個方程式是一個半徑為 2 中心在 (0, 0, 0) 的球，所以 x_1 、 x_2 及 x_3 的範圍是 $[-2, 2]$ 。對這個例子，產生圖形的 Maple 指令為

$$with(plots): implicitplot3d(\{eq1, eq2, eq3\}, x_1 = -2..2, x_2 = -2..2, x_3 = -2..2);$$

在 Maple 中有許多 3D 圖形的選項，可幫助定出非線性方程組的解。例如，我們可以旋轉圖形、放大局部區域、改變座標軸顯示方式等，以幫助我們確認交點的座標。對這個例子，合理的初始近似值為 $(x_1, x_2, x_3)^t = (-0.5, -1.5, 1.5)^t$ 。

習題組 10.2 完整習題請見隨書光碟

1. 用牛頓法及 $\mathbf{x}^{(0)} = \mathbf{0}$ ，求下列非線性方程組的 $\mathbf{x}^{(2)}$ 。

a. $4x_1^2 - 20x_1 + \frac{1}{4}x_2^2 + 8 = 0$

$$\frac{1}{2}x_1x_2^2 + 2x_1 - 5x_2 + 8 = 0$$

c. $x_1(1 - x_1) + 4x_2 = 12$
 $(x_1 - 2)^2 + (2x_2 - 3)^2 = 25$

b. $\sin(4\pi x_1 x_2) - 2x_2 - x_1 = 0$

$$\left(\frac{4\pi - 1}{4\pi}\right)(e^{2x_1} - e) + 4ex_2^2 - 2ex_1 = 0$$

d. $5x_1^2 - x_2^2 = 0$
 $x_2 - 0.25(\sin x_1 + \cos x_2) = 0$

2. 用牛頓法及 $\mathbf{x}^{(0)} = \mathbf{0}$ ，求下列非線性方程組的 $\mathbf{x}^{(2)}$ 。

$$\begin{aligned} \text{a.} \quad & 3x_1 - \cos(x_2x_3) - \frac{1}{2} = 0 \\ & 4x_1^2 - 625x_2^2 + 2x_2 - 1 = 0 \\ & e^{-x_1x_2} + 20x_3 + \frac{10\pi - 3}{3} = 0 \end{aligned}$$

$$\begin{aligned} \text{b.} \quad & x_1^2 + x_2 - 37 = 0 \\ & x_1 - x_2^2 - 5 = 0 \\ & x_1 + x_2 + x_3 - 3 = 0 \end{aligned}$$

$$\begin{aligned} \text{c.} \quad & 15x_1 + x_2^2 - 4x_3 = 13 \\ & x_1^2 + 10x_2 - x_3 = 11 \\ & x_2^3 - 25x_3 = -22 \end{aligned}$$

$$\begin{aligned} \text{d.} \quad & 10x_1 - 2x_2^2 + x_2 - 2x_3 - 5 = 0 \\ & 8x_2^2 + 4x_3^2 - 9 = 0 \\ & 8x_2x_3 + 4 = 0 \end{aligned}$$

3. 利用 Maple 的繪圖功能，求下列非線性方程組的近似解。

$$\begin{aligned} \text{a.} \quad & 4x_1^2 - 20x_1 + \frac{1}{4}x_2^2 + 8 = 0 \\ & \frac{1}{2}x_1x_2^2 + 2x_1 - 5x_2 + 8 = 0 \end{aligned}$$

$$\begin{aligned} \text{b.} \quad & \sin(4\pi x_1x_2) - 2x_2 - x_1 = 0 \\ & \left(\frac{4\pi - 1}{4\pi}\right)(e^{2x_1} - e) + 4ex_2^2 - 2ex_1 = 0 \end{aligned}$$

$$\begin{aligned} \text{c.} \quad & x_1(1 - x_1) + 4x_2 = 12 \\ & (x_1 - 2)^2 + (2x_2 - 3)^2 = 25 \end{aligned}$$

$$\begin{aligned} \text{d.} \quad & 5x_1^2 - x_2^2 = 0 \\ & x_2 - 0.25(\sin x_1 + \cos x_2) = 0 \end{aligned}$$

4. 利用 Maple 的繪圖功能，求下列非線性方程組的近似解。

$$\begin{aligned} \text{a.} \quad & 3x_1 - \cos(x_2x_3) - \frac{1}{2} = 0 \\ & 4x_1^2 - 625x_2^2 + 2x_2 - 1 = 0 \\ & e^{-x_1x_2} + 20x_3 + \frac{10\pi - 3}{3} = 0 \\ & -1 \leq x_1 \leq 1, -1 \leq x_2 \leq 1, -1 \leq x_3 \leq 1 \end{aligned}$$

$$\begin{aligned} \text{b.} \quad & x_1^2 + x_2 - 37 = 0 \\ & x_1 - x_2^2 - 5 = 0 \\ & x_1 + x_2 + x_3 - 3 = 0 \\ & -4 \leq x_1 \leq 8, -2 \leq x_2 \leq 2, -6 \leq x_3 \leq 0 \end{aligned}$$

$$\begin{aligned} \text{c.} \quad & 15x_1 + x_2^2 - 4x_3 = 13 \\ & x_1^2 + 10x_2 - x_3 = 11 \\ & x_2^3 - 25x_3 = -22 \\ & 0 \leq x_1 \leq 2, 0 \leq x_2 \leq 2, 0 \leq x_3 \leq 2 \\ & \text{及 } 0 \leq x_1 \leq 2, 0 \leq x_2 \leq 2, -2 \leq x_3 \leq 0 \end{aligned}$$

$$\begin{aligned} \text{d.} \quad & 10x_1 - 2x_2^2 + x_2 - 2x_3 - 5 = 0 \\ & 8x_2^2 + 4x_3^2 - 9 = 0 \\ & 8x_2x_3 + 4 = 0 \\ & 0 \leq x_1 \leq 2, -2 \leq x_2 \leq 0, 0 \leq x_3 \leq 2 \end{aligned}$$

5. 用習題 3 的結果做為牛頓法的初始近似值，迭代到 $\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\|_\infty < 10^{-6}$ 為止。

6. 用習題 4 的結果做為牛頓法的初始近似值，迭代到 $\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\|_\infty < 10^{-6}$ 為止。

7. 用牛頓法求下列非線性方程組在指定條件下的近似解。迭代到 $\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\|_\infty < 10^{-6}$ 為止。

$$\begin{aligned} \text{a.} \quad & 3x_1^2 - x_2^2 = 0 \\ & 3x_1x_2^2 - x_1^3 - 1 = 0 \\ & \text{用 } \mathbf{x}^{(0)} = (1, 1)^t \end{aligned}$$

$$\begin{aligned} \text{b.} \quad & \ln(x_1^2 + x_2^2) - \sin(x_1x_2) = \ln 2 + \ln \pi \\ & e^{x_1 - x_2} + \cos(x_1x_2) = 0 \\ & \text{用 } \mathbf{x}^{(0)} = (2, 2)^t \end{aligned}$$

$$\begin{aligned} \text{c.} \quad & x_1^3 + x_1^2x_2 - x_1x_3 + 6 = 0 \\ & e^{x_1} + e^{x_2} - x_3 = 0 \\ & x_2^2 - 2x_1x_3 = 4 \\ & \text{用 } \mathbf{x}^{(0)} = (-1, -2, 1)^t \end{aligned}$$

$$\begin{aligned} \text{d.} \quad & 6x_1 - 2\cos(x_2x_3) - 1 = 0 \\ & 9x_2 + \sqrt{x_1^2 + \sin x_3} + 1.06 + 0.9 = 0 \\ & 60x_3 + 3e^{-x_1x_2} + 10\pi - 3 = 0 \\ & \text{用 } \mathbf{x}^{(0)} = (0, 0, 0)^t \end{aligned}$$

8. 非線性方程組

$$\begin{aligned} E_1 : 4x_1 - x_2 + x_3 &= x_1x_4, & E_2 : -x_1 + 3x_2 - 2x_3 &= x_2x_4, \\ E_3 : x_1 - 2x_2 + 3x_3 &= x_3x_4, & E_4 : x_1^2 + x_2^2 + x_3^2 &= 1 \end{aligned}$$

有 6 個解。

- a. 證明，若 $(x_1, x_2, x_3, x_4)^t$ 是一個解，則 $(-x_1, -x_2, -x_3, x_4)^t$ 也是它的解。
- b. 用牛頓法及 3 個初始近似值，求出所有的解並滿足 $\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\|_\infty < 10^{-5}$ 。

9. 非線性方程組

$$\begin{aligned} 3x_1 - \cos(x_2x_3) - \frac{1}{2} &= 0 \\ x_1^2 - 625x_2^2 - \frac{1}{4} &= 0 \\ e^{-x_1x_2} + 20x_3 + \frac{10\pi - 3}{3} &= 0 \end{aligned}$$

在其解的地方的 Jacobian 矩陣為奇異。用牛頓法及 $\mathbf{x}^{(0)} = (1, 1, 1)^t$ 。留意其收斂速度可能很慢，或在相當的迭代次數內都不收斂。

10. 對於線性方程組 $A\mathbf{x} = \mathbf{b}$

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned}$$

牛頓法可做怎樣的簡化？其中 A 為非奇異矩陣。

11. 證明，在 $n = 1$ 時，(10.9) 式的牛頓法會簡化成 2.3 節的牛頓法。
12. 假設在硬質底層之上是一層軟性且均質的土壤，要知道須要多大的壓力才能使一個很大的重物沉入土壤中，我們可以用較小的物體在同樣的土壤上做實驗。要使一個半徑為 r 的圓盤沉入土壤表面下 d 深度所須的壓力，可以用下式來近似，已知硬質底層位於表面下 $D > d$ 深度處，

$$p = k_1 e^{k_2 r} + k_3 r$$

其中 k_1 、 k_2 及 k_3 為常數，取決於深度 d 與土壤的一致性，但與圓盤半徑無關。(見 [Bek, pp. 89-94])

- a. 假設我們經由實驗知道，使半徑 1 in 的圓盤下沉 1 ft 須要 10 lb/in² 的壓力，使半徑 2 in 的圓盤下沉 1 ft 須要 12 lb/in² 的壓力，使半徑 3 in 的圓盤下沉 1 ft 須要 15 lb/in² 的壓力，求 k_1 、 k_2 及 k_3 的值為何。(已知軟土深度超過 1 ft)
- b. 用 (a) 的結果，預估可承載 500 lb 負荷且下沉不超過 1 ft 的圓盤最小半徑為何。
13. 為了設計重力式傾卸滑槽，使得粒狀物體的傳輸時間最短，C. Chiarella、W. Charlton 及 A.W. Roberts [CCR] 等人用牛頓法解以下方程組：

$$(i) \quad f_n(\theta_1, \dots, \theta_N) = \frac{\sin \theta_{n+1}}{v_{n+1}}(1 - \mu w_{n+1}) - \frac{\sin \theta_n}{v_n}(1 - \mu w_n) = 0, \quad n = 1, 2, \dots, N-1.$$

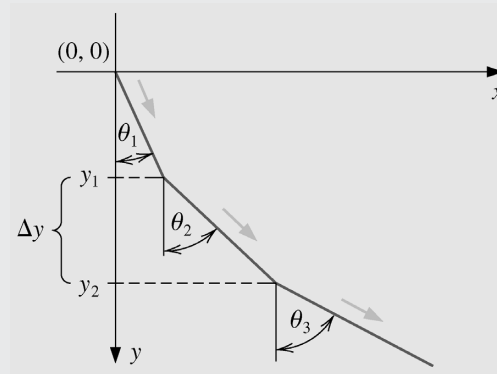
$$(ii) \quad f_N(\theta_1, \dots, \theta_N) = \Delta y \sum_{i=1}^N \tan \theta_i - X = 0, \quad \text{其中}$$

$$(a) \quad v_n^2 = v_0^2 + 2gn\Delta y - 2\mu\Delta y \sum_{j=1}^n \frac{1}{\cos \theta_j}, \quad n = 1, 2, \dots, N,$$

$$(b) \quad w_n = -\Delta y v_n \sum_{i=1}^N \frac{1}{v_i^3 \cos \theta_i}, \quad n = 1, 2, \dots, N.$$

常數 v_0 為粒狀貨物的初始速度， X 為滑槽終點的 x 座標， μ 為磨擦力， N 代表滑槽分為幾段，而 $g = 32.17 \text{ ft/s}^2$ 為重力常數。變數 θ_i 代表第 i 段滑槽與垂直線的夾角，如下圖所示， v_i 為顆

粒在第 i 段滑槽的速度。已知 $\mu = 0$ 、 $X = 2$ 、 $\Delta y = 0.2$ 、 $N = 20$ 及 $v_0 = 0$ ，解 (i) 及 (ii) 式以
 求出 $\theta = (\theta_1, \dots, \theta_N)^t$ ，其中 v_n 及 w_n 的值可直接由 (a) 及 (b) 獲得。迭代到 $\|\theta^{(k)} - \theta^{(k-1)}\|_\infty <$
 10^{-2} 。



14. 有一個很有趣的生物學實驗 (見 [Schr2])，是關於如何決定各種水螅可以存活且不減短預期壽命的
 最大水溫 X_M 。有一個方法，是對收集到的數據使用 $f(x) = y = a/(x - b)^c$ 型式的加權最小平方
 擬合。數據中 x 的值代表水溫。常數 b 為 f 圖形的漸近線，所以也是 X_M 的近似值。

a. 證明，選擇 a 、 b 及 c 以使得

$$\sum_{i=1}^n \left[w_i y_i - \frac{a}{(x_i - b)^c} \right]^2$$

為最小的問題，可簡化為求解非線性方程組

$$\begin{aligned} a &= \sum_{i=1}^n \frac{w_i y_i}{(x_i - b)^c} \bigg/ \sum_{i=1}^n \frac{1}{(x_i - b)^{2c}} \\ 0 &= \sum_{i=1}^n \frac{w_i y_i}{(x_i - b)^c} \cdot \sum_{i=1}^n \frac{1}{(x_i - b)^{2c+1}} - \sum_{i=1}^n \frac{w_i y_i}{(x_i - b)^{c+1}} \cdot \sum_{i=1}^n \frac{1}{(x_i - b)^{2c}} \\ 0 &= \sum_{i=1}^n \frac{w_i y_i}{(x_i - b)^c} \cdot \sum_{i=1}^n \frac{\ln(x_i - b)}{(x_i - b)^{2c}} - \sum_{i=1}^n \frac{w_i y_i \ln(x_i - b)}{(x_i - b)^c} \cdot \sum_{i=1}^n \frac{1}{(x_i - b)^{2c}} \end{aligned}$$

b. 利用以下數據求解此非線性方程組。使用 $w_i = \ln y_i$ 為權重。

i	1	2	3	4
y_i	2.40	3.80	4.75	21.60
x_i	31.8	31.5	31.2	30.2



10.3 準牛頓法

用於求解非線性方程組，牛頓法有一個明顯的缺點，也就是它在每次迭代，都要求出 Jaco-
 bian 矩陣，並解一個包含此矩陣的 $n \times n$ 線性方程組。分析一下牛頓法每次迭代所須的計

算量。表示成 $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ 型式的 n 個非線性方程式，共須求出 $\mathbf{F}$ 的 n 個分量函數的 n^2 個偏導數。即使是使用如 Maple 這種可執行符式運算的軟體，要求出這些偏導數的確解也不容易。

如果無法使用確解，我們可以用有限差分法來近似偏導數。例如

$$\frac{\partial f_j}{\partial x_k}(\mathbf{x}^{(i)}) \approx \frac{f_j(\mathbf{x}^{(i)} + \mathbf{e}_k h) - f_j(\mathbf{x}^{(i)})}{h} \quad (10.10)$$

其中 h 是一個絕對值很小的數，而 $\mathbf{e}_k$ 是一個向量，它只有第 k 個分量為 1 其餘都是 0。這樣的近似方式仍須要計算 n^2 個純量函數值，而且也無法減少解線性方程組所須的 $O(n^3)$ 的計算量。

所以牛頓法每次迭代，最少要計算 $n^2 + n$ 次的純量函數值（求 Jacobian 矩陣要 n^2 次，求 $\mathbf{F}$ 要 n 次）再加上解線性方程組所須的 $O(n^3)$ 的算術運算。如此大的計算成本，只有在 n 相當小且純量函數非常簡單的情形下才可行。

在本節中我們要將正割法 (Secant method) 一般化，以用於非線性方程組，這種方法叫做 Broyden 法 (見 [Broy])。此方法每次迭代只須要 n 次純量函數計算，總算數運算量也減為 $O(n^2)$ 。此一方法屬於一類叫做最小變量正割修正 (least-change secant update) 的方法，由此產生的算則叫做準牛頓 (quasi-Newton) 算則。這類方法，使用一個在每次迭代做修正的近似矩陣以取代牛頓法中的 Jacobian 矩陣。

此方法的缺點是，它無法像牛頓法的二階收斂，它的收斂速度叫做超線性 (superlinear) 收斂。這表示

$$\lim_{i \rightarrow \infty} \frac{\|\mathbf{x}^{(i+1)} - \mathbf{p}\|}{\|\mathbf{x}^{(i)} - \mathbf{p}\|} = 0$$

其中 $\mathbf{p}$ 為方程組 $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ 的解，而 $\mathbf{x}^{(i)}$ 及 $\mathbf{x}^{(i+1)}$ 為 2 個連續的近似解。

在大多數的應用中，計算量的減少所得的好處遠超過收斂速度的減慢。但準牛頓法有另外一個缺點，它們不像牛頓法可以自我修正。牛頓法在連續的迭代中，通常能夠自我修正捨入誤差，但 Broyden 法必須加入特殊防護機制才行。

簡單的描述 Broyden 法，假設對於 $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ 的解 $\mathbf{p}$ ，我們先有一個初始近似值 $\mathbf{x}^{(0)}$ 。我們用牛頓法計算下一個近似值 $\mathbf{x}^{(1)}$ ，如果不容易獲得 $J(\mathbf{x}^{(0)})$ ，我們可用 (10.10) 式的有限差分來近似偏導數。計算 $\mathbf{x}^{(2)}$ 的方法就與牛頓法不同了。我們先看一下用於單一非線性方程式的正割法。正割法用近似值

$$f'(x_1) \approx \frac{f(x_1) - f(x_0)}{x_1 - x_0}$$

取代單變數牛頓法中的 $f'(x_1)$ 。

對非線性方程組， $\mathbf{x}^{(1)} - \mathbf{x}^{(0)}$ 是一個向量，而它的除式沒有定義。但此方法進行的方

式類似於在方程組的牛頓法中，將 $J(\mathbf{x}^{(1)})$ 換成具有以下特性的矩陣 A_1

$$A_1(\mathbf{x}^{(1)} - \mathbf{x}^{(0)}) = \mathbf{F}(\mathbf{x}^{(1)}) - \mathbf{F}(\mathbf{x}^{(0)}) \quad (10.11)$$

在 $\mathbb{R}^n$ 中的任何非 0 向量，可以寫成是向量 $\mathbf{x}^{(1)} - \mathbf{x}^{(0)}$ 的倍數，和垂直於 $\mathbf{x}^{(1)} - \mathbf{x}^{(0)}$ 之向量的倍數的和。所以，要唯一的定義出 A_1 ，我們還必須指定它對垂直於 $\mathbf{x}^{(1)} - \mathbf{x}^{(0)}$ 之向量的作用方式。因為沒有任何 $\mathbf{F}$ 在垂直於 $\mathbf{x}^{(1)} - \mathbf{x}^{(0)}$ 之方向變化的資訊，故我們指定在此方向上沒有變化，亦即

$$\text{在 } (\mathbf{x}^{(1)} - \mathbf{x}^{(0)})^t \mathbf{z} = 0 \text{ 時, } A_1 \mathbf{z} = J(\mathbf{x}^{(0)}) \mathbf{z} \quad (10.12)$$

因此任何垂直於 $\mathbf{x}^{(1)} - \mathbf{x}^{(0)}$ 的向量，在我們將 $J(\mathbf{x}^{(0)})$ 換成 A_1 時不受影響，其中 $J(\mathbf{x}^{(0)})$ 是用來求 $\mathbf{x}^{(1)}$ ，而 A_1 則是用來求 $\mathbf{x}^{(2)}$ 。

(10.11) 式及 (10.12) 式的條件可唯一的定義出 A_1 (見 [DM]) 為

$$A_1 = J(\mathbf{x}^{(0)}) + \frac{[\mathbf{F}(\mathbf{x}^{(1)}) - \mathbf{F}(\mathbf{x}^{(0)}) - J(\mathbf{x}^{(0)})(\mathbf{x}^{(1)} - \mathbf{x}^{(0)})](\mathbf{x}^{(1)} - \mathbf{x}^{(0)})^t}{\|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\|_2^2}$$

我們用這個矩陣取代 $J(\mathbf{x}^{(1)})$ 來計算 $\mathbf{x}^{(2)}$ 為

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} - A_1^{-1} \mathbf{F}(\mathbf{x}^{(1)})$$

求得了 $\mathbf{x}^{(2)}$ 之後，重複此方法以求 $\mathbf{x}^{(3)}$ ，用 A_1 取代 $A_0 \equiv J(\mathbf{x}^{(0)})$ ，用 $\mathbf{x}^{(2)}$ 和 $\mathbf{x}^{(1)}$ 取代 $\mathbf{x}^{(1)}$ 和 $\mathbf{x}^{(0)}$ 。

以通式表示，獲得 $\mathbf{x}^{(i)}$ 之後， $\mathbf{x}^{(i+1)}$ 的算法為

$$A_i = A_{i-1} + \frac{\mathbf{y}_i - A_{i-1} \mathbf{s}_i}{\|\mathbf{s}_i\|_2^2} \mathbf{s}_i^t \quad (10.13)$$

及

$$\mathbf{x}^{(i+1)} = \mathbf{x}^{(i)} - A_i^{-1} \mathbf{F}(\mathbf{x}^{(i)}) \quad (10.14)$$

其中使用了 $\mathbf{y}_i = \mathbf{F}(\mathbf{x}^{(i)}) - \mathbf{F}(\mathbf{x}^{(i-1)})$ 及 $\mathbf{s}_i = \mathbf{x}^{(i)} - \mathbf{x}^{(i-1)}$ 的符號以簡化公式。

如果依照 (10.13) 式及 (10.14) 式來執行計算，每次迭代所須進行的純量函數計算由 $n^2 + n$ 次減為 n 次 (計算 $\mathbf{F}(\mathbf{x}^{(i)})$) 所須的，但求解 $n \times n$ 的線性方程組

$$A_i \mathbf{s}_{i+1} = -\mathbf{F}(\mathbf{x}^{(i)}) \quad (10.15)$$

仍須要 $O(n^3)$ 的算術運算 (見算則 10.1 的 Step 4)。若僅是這樣，付出由二階收斂減為超線性收斂的代價是不值得的。

■ Sherman-Morrison 公式

但如果使用了 Sherman 和 Morrison 的矩陣求逆公式 (見，例如，[DM], p. 55) 則可獲得大幅改進。

定理 10.8 Sherman-Morrison 公式

設 A 為非奇異矩陣，且 $\mathbf{x}$ 及 $\mathbf{y}$ 為向量，並有 $\mathbf{y}^t A^{-1} \mathbf{x} \neq -1$ 。則 $A + \mathbf{x}\mathbf{y}^t$ 為非奇異，且

$$(A + \mathbf{x}\mathbf{y}^t)^{-1} = A^{-1} - \frac{A^{-1}\mathbf{x}\mathbf{y}^t A^{-1}}{1 + \mathbf{y}^t A^{-1} \mathbf{x}}$$

Sherman-Morrison 公式讓我們可以直接由 A_{i-1}^{-1} 求得 A_i^{-1} ，避免了在每次迭代進行矩陣求逆。

令 $A = A_{i-1}$ 、 $\mathbf{x} = (\mathbf{y}_i - A_{i-1}\mathbf{s}_i)/\|\mathbf{s}_i\|_2^2$ 及 $\mathbf{y} = \mathbf{s}_i$ ，由 (10.13) 式可得

$$\begin{aligned} A_i^{-1} &= \left(A_{i-1} + \frac{\mathbf{y}_i - A_{i-1}\mathbf{s}_i}{\|\mathbf{s}_i\|_2^2} \mathbf{s}_i^t \right)^{-1} \\ &= A_{i-1}^{-1} - \frac{A_{i-1}^{-1} \left(\frac{\mathbf{y}_i - A_{i-1}\mathbf{s}_i}{\|\mathbf{s}_i\|_2^2} \mathbf{s}_i^t \right) A_{i-1}^{-1}}{1 + \mathbf{s}_i^t A_{i-1}^{-1} \left(\frac{\mathbf{y}_i - A_{i-1}\mathbf{s}_i}{\|\mathbf{s}_i\|_2^2} \right)} \\ &= A_{i-1}^{-1} - \frac{(A_{i-1}^{-1} \mathbf{y}_i - \mathbf{s}_i) \mathbf{s}_i^t A_{i-1}^{-1}}{\|\mathbf{s}_i\|_2^2 + \mathbf{s}_i^t A_{i-1}^{-1} \mathbf{y}_i - \|\mathbf{s}_i\|_2^2}, \end{aligned}$$

所以

$$A_i^{-1} = A_{i-1}^{-1} + \frac{(\mathbf{s}_i - A_{i-1}^{-1} \mathbf{y}_i) \mathbf{s}_i^t A_{i-1}^{-1}}{\mathbf{s}_i^t A_{i-1}^{-1} \mathbf{y}_i} \quad (10.16)$$

此計算每一步都只有矩陣－向量乘法，所以只須要 $O(n^2)$ 的算術運算。以上的計算不須要求 A_i ，也不須要解 (10.15) 式的線性方程組。

算則 10.2 即使用此計算程序，將 (10.16) 式用於 (10.14) 式的迭代運算中。

算則 10.2 Broyden

已知初始近似值 $\mathbf{x}$ ，求非線性方程組 $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ 的近似解：

INPUT 方程式與未知數的數目 n ；初始近似值 $\mathbf{x} = (x_1, \dots, x_n)^t$ ；容許誤差 TOL ；最大迭代次數 N 。

OUTPUT 近似解 $\mathbf{x} = (x_1, \dots, x_n)^t$ 或最大迭代次數超過之訊息。

Step 1 Set $A_0 = J(\mathbf{x})$ where $J(\mathbf{x})_{ij} = \frac{\partial f_i}{\partial x_j}(\mathbf{x})$ for $1 \leq i, j \leq n$;

$\mathbf{v} = \mathbf{F}(\mathbf{x})$. (註： $\mathbf{v} = \mathbf{F}(\mathbf{x}^{(0)})$.)

Step 2 Set $A = A_0^{-1}$. (用高斯消去法)

Step 3 Set $\mathbf{s} = -A\mathbf{v}$; (註： $\mathbf{s} = \mathbf{s}_1$.)
 $\mathbf{x} = \mathbf{x} + \mathbf{s}$; (註： $\mathbf{x} = \mathbf{x}^{(1)}$.)
 $k = 2$.

Step 4 While $(k \leq N)$ do Steps 5–13.

Step 5 Set $\mathbf{w} = \mathbf{v}$; (儲存 $\mathbf{v}$.)
 $\mathbf{v} = \mathbf{F}(\mathbf{x})$; (註: $\mathbf{v} = \mathbf{F}(\mathbf{x}^{(k)})$.)
 $\mathbf{y} = \mathbf{v} - \mathbf{w}$. (註: $\mathbf{y} = \mathbf{y}_k$.)
Step 6 Set $\mathbf{z} = -A\mathbf{y}$. (註: $\mathbf{z} = -A_{k-1}^{-1}\mathbf{y}_k$.)
Step 7 Set $p = -\mathbf{s}'\mathbf{z}$. (註: $p = \mathbf{s}_k^t A_{k-1}^{-1} \mathbf{y}_k$.)
Step 8 Set $\mathbf{u}' = \mathbf{s}'A$.
Step 9 Set $A = A + \frac{1}{p}(\mathbf{s} + \mathbf{z})\mathbf{u}'$. (註: $A = A_k^{-1}$.)
Step 10 Set $\mathbf{s} = -A\mathbf{v}$. (註: $\mathbf{s} = -A_k^{-1}\mathbf{F}(\mathbf{x}^{(k)})$.)
Step 11 Set $\mathbf{x} = \mathbf{x} + \mathbf{s}$. (註: $\mathbf{x} = \mathbf{x}^{(k+1)}$.)
Step 12 If $\|\mathbf{s}\| < TOL$ then OUTPUT ($\mathbf{x}$);
 (程式成功)
 STOP.

Step 13 Set $k = k + 1$.

Step 14 OUTPUT ('Maximum number of iterations exceeded');
 (程式失敗)
 STOP.

例題 1 用 Broyden 法及 $\mathbf{x}^{(0)} = (0.1, 0.1, -0.1)^t$ 求非線性方程組

$$\begin{aligned}
 3x_1 - \cos(x_2x_3) - \frac{1}{2} &= 0 \\
 x_1^2 - 81(x_2 + 0.1)^2 + \sin x_3 + 1.06 &= 0 \\
 e^{-x_1x_2} + 20x_3 + \frac{10\pi - 3}{3} &= 0
 \end{aligned}$$

的近似解。

解 在 10.2 節的例題 1 中，我們用牛頓法解過一次這個方程組。此方程組的 Jacobian 矩陣為

$$J(x_1, x_2, x_3) = \begin{bmatrix} 3 & x_3 \sin x_2x_3 & x_2 \sin x_2x_3 \\ 2x_1 & -162(x_2 + 0.1) & \cos x_3 \\ -x_2e^{-x_1x_2} & -x_1e^{-x_1x_2} & 20 \end{bmatrix}$$

令 $\mathbf{x}^{(0)} = (0.1, 0.1, -0.1)^t$ 且

$$\mathbf{F}(x_1, x_2, x_3) = (f_1(x_1, x_2, x_3), f_2(x_1, x_2, x_3), f_3(x_1, x_2, x_3))^t$$

其中

$$\begin{aligned}
 f_1(x_1, x_2, x_3) &= 3x_1 - \cos(x_2x_3) - \frac{1}{2} \\
 f_2(x_1, x_2, x_3) &= x_1^2 - 81(x_2 + 0.1)^2 + \sin x_3 + 1.06
 \end{aligned}$$

且

$$f_3(x_1, x_2, x_3) = e^{-x_1 x_2} + 20x_3 + \frac{10\pi - 3}{3}$$

則

$$\mathbf{F}(\mathbf{x}^{(0)}) = \begin{bmatrix} -1.199950 \\ -2.269833 \\ 8.462025 \end{bmatrix}$$

因為

$$\begin{aligned} A_0 &= J(x_1^{(0)}, x_2^{(0)}, x_3^{(0)}) \\ &= \begin{bmatrix} 3 & 9.999833 \times 10^{-4} & -9.999833 \times 10^{-4} \\ 0.2 & -32.4 & 0.9950042 \\ -9.900498 \times 10^{-2} & -9.900498 \times 10^{-2} & 20 \end{bmatrix} \end{aligned}$$

我們可得

$$\begin{aligned} A_0^{-1} &= J(x_1^{(0)}, x_2^{(0)}, x_3^{(0)})^{-1} \\ &= \begin{bmatrix} 0.3333332 & 1.023852 \times 10^{-5} & 1.615701 \times 10^{-5} \\ 2.108607 \times 10^{-3} & -3.086883 \times 10^{-2} & 1.535836 \times 10^{-3} \\ 1.660520 \times 10^{-3} & -1.527577 \times 10^{-4} & 5.000768 \times 10^{-2} \end{bmatrix} \end{aligned}$$

所以

$$\begin{aligned} \mathbf{x}^{(1)} &= \mathbf{x}^{(0)} - A_0^{-1} \mathbf{F}(\mathbf{x}^{(0)}) = \begin{bmatrix} 0.4998697 \\ 1.946685 \times 10^{-2} \\ -0.5215205 \end{bmatrix} \\ \mathbf{F}(\mathbf{x}^{(1)}) &= \begin{bmatrix} -3.394465 \times 10^{-4} \\ -0.3443879 \\ 3.188238 \times 10^{-2} \end{bmatrix} \\ \mathbf{y}_1 &= \mathbf{F}(\mathbf{x}^{(1)}) - \mathbf{F}(\mathbf{x}^{(0)}) = \begin{bmatrix} 1.199611 \\ 1.925445 \\ -8.430143 \end{bmatrix} \\ \mathbf{s}_1 &= \begin{bmatrix} 0.3998697 \\ -8.053315 \times 10^{-2} \\ -0.4215204 \end{bmatrix} \end{aligned}$$

$$\mathbf{s}_1^t A_0^{-1} \mathbf{y}_1 = 0.3424604$$

$$\begin{aligned} A_1^{-1} &= A_0^{-1} + (1/0.3424604) [(\mathbf{s}_1 - A_0^{-1} \mathbf{y}_1) \mathbf{s}_1^t A_0^{-1}] \\ &= \begin{bmatrix} 0.3333781 & 1.11050 \times 10^{-5} & 8.967344 \times 10^{-6} \\ -2.021270 \times 10^{-3} & -3.094849 \times 10^{-2} & 2.196906 \times 10^{-3} \\ 1.022214 \times 10^{-3} & -1.650709 \times 10^{-4} & 5.010986 \times 10^{-2} \end{bmatrix} \end{aligned}$$

及

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} - A_1^{-1} \mathbf{F}(\mathbf{x}^{(1)}) = \begin{bmatrix} 0.4999863 \\ 8.737833 \times 10^{-3} \\ -0.5231746 \end{bmatrix}$$

其餘迭代的結果列於表 10.4。Broyden 法第 5 次迭代的結果略遜於前節例題中牛頓法第 4 次迭代的結果。 ■

表 10.4

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$\ \mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\ _2$
3	0.5000066	8.672157×10^{-4}	-0.5236918	7.88×10^{-3}
4	0.5000003	6.083352×10^{-5}	-0.5235954	8.12×10^{-4}
5	0.5000000	-1.448889×10^{-6}	-0.5235989	6.24×10^{-5}
6	0.5000000	6.059030×10^{-9}	-0.5235988	1.50×10^{-6}

也有方法可以在大幅減少計算量的同時，維持二階收斂。這類方法最早由 Brown [Brow, K] 提出。在 [MC] 中對一些常用的這類分法做了討論與比較。但通常這類方法在使用上都遠比 Broyden 法困難。

習題組 10.3 完整習題請見隨書光碟

1. 用 Broyden 法及 $\mathbf{x}^{(0)} = \mathbf{0}$ ，求下列非線性方程組的 $\mathbf{x}^{(2)}$ 。

a. $4x_1^2 - 20x_1 + \frac{1}{4}x_2^2 + 8 = 0$

$$\frac{1}{2}x_1x_2^2 + 2x_1 - 5x_2 + 8 = 0$$

c. $3x_1^2 - x_2^2 = 0$

$$3x_1x_2^2 - x_1^3 - 1 = 0$$

使用 $\mathbf{x}^{(0)} = (1, 1)^T$ 。

b. $\sin(4\pi x_1x_2) - 2x_2 - x_1 = 0$

$$\left(\frac{4\pi - 1}{4\pi}\right)(e^{2x_1} - e) + 4ex_2^2 - 2ex_1 = 0$$

d. $\ln(x_1^2 + x_2^2) - \sin(x_1x_2) = \ln 2 + \ln \pi$

$$e^{x_1 - x_2} + \cos(x_1x_2) = 0$$

使用 $\mathbf{x}^{(0)} = (2, 2)^T$ 。

2. 用 Broyden 法及 $\mathbf{x}^{(0)} = \mathbf{0}$ ，求下列非線性方程組的 $\mathbf{x}^{(2)}$ 。

a. $3x_1 - \cos(x_2x_3) - \frac{1}{2} = 0$

$$4x_1^2 - 625x_2^2 + 2x_2 - 1 = 0$$

$$e^{-x_1x_2} + 20x_3 + \frac{10\pi - 3}{3} = 0$$

c. $x_1^3 + x_1^2x_2 - x_1x_3 + 6 = 0$

$$e^{x_1} + e^{x_2} - x_3 = 0$$

$$x_2^2 - 2x_1x_3 = 4$$

使用 $\mathbf{x}^{(0)} = (-1, -2, 1)^T$ 。

b. $x_1^2 + x_2 - 37 = 0$

$$x_1 - x_2^2 - 5 = 0$$

$$x_1 + x_2 + x_3 - 3 = 0$$

d. $6x_1 - 2\cos(x_2x_3) - 1 = 0$

$$9x_2 + \sqrt{x_1^2 + \sin x_3 + 1.06} + 0.9 = 0$$

$$60x_3 + 3e^{-x_1x_2} + 10\pi - 3 = 0$$

使用 $\mathbf{x}^{(0)} = (0, 0, 0)^T$ 。

3. 用 Broyden 法，及以下各初始近似值 $\mathbf{x}^{(0)}$ ，求習題 1 各非線性方程組的近似解。
- a. $(0, 0)^t$ b. $(0, 0)^t$ c. $(1, 1)^t$ d. $(2, 2)^t$
4. 用 Broyden 法，及以下各初始近似值 $\mathbf{x}^{(0)}$ ，求習題 2 各非線性方程組的近似解。
- a. $(1, 1, 1)^t$ b. $(2, 1, -1)^t$ c. $(-1, -2, 1)^t$ d. $(0, 0, 0)^t$
5. 用 Broyden 法求下列各非線性方程組的近似解。迭代到 $\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\|_\infty < 10^{-6}$ 為止。
- a. $x_1(1 - x_1) + 4x_2 = 12$
 $(x_1 - 2)^2 + (2x_2 - 3)^2 = 25$
- b. $5x_1^2 - x_2^2 = 0$
 $x_2 - 0.25(\sin x_1 + \cos x_2) = 0$
- c. $15x_1 + x_2^2 - 4x_3 = 13$
 $x_1^2 + 10x_2 - x_3 = 11$
 $x_2^3 - 25x_3 = -22$
- d. $10x_1 - 2x_2^2 + x_2 - 2x_3 - 5 = 0$
 $8x_2^2 + 4x_3^2 - 9 = 0$
 $8x_2x_3 + 4 = 0$

6. 非線性方程組

$$\begin{aligned} 4x_1 - x_2 + x_3 &= x_1x_4 \\ -x_1 + 3x_2 - 2x_3 &= x_2x_4 \\ x_1 - 2x_2 + 3x_3 &= x_3x_4 \\ x_1^2 + x_2^2 + x_3^2 &= 1 \end{aligned}$$

有 6 個解。

- a. 證明，若 $(x_1, x_2, x_3, x_4)^t$ 是一個解，則 $(-x_1, -x_2, -x_3, x_4)^t$ 也是它的解。
- b. 用 Broyden 法 3 次，求出所有的解並迭代到 $\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\|_\infty < 10^{-5}$ 。
7. 非線性方程組
- $$3x_1 - \cos(x_2x_3) - \frac{1}{2} = 0, \quad x_1^2 - 625x_2^2 - \frac{1}{4} = 0, \quad e^{-x_1x_2} + 20x_3 + \frac{10\pi - 3}{3} = 0$$
- 在其解的地方的 Jacobian 矩陣為奇異。用 Broyden 法及 $\mathbf{x}^{(0)} = (1, 1, 1)^t$ 。留意其收斂速度可能很慢，或在相當的迭代次數內都不收斂。
8. 證明，若 $\mathbf{0} \neq \mathbf{y} \in \mathbb{R}^n$ 且 $\mathbf{z} \in \mathbb{R}^n$ ，則 $\mathbf{z} = \mathbf{z}_1 + \mathbf{z}_2$ ，其中 $\mathbf{z}_1 = (\mathbf{y}^t \mathbf{z} / \|\mathbf{y}\|_2^2) \mathbf{y}$ 平行於 $\mathbf{y}$ 且 $\mathbf{z}_2$ 垂直於 $\mathbf{y}$ 。
9. 證明若 $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ ，則 $\det(I + \mathbf{u}\mathbf{v}^t) = 1 + \mathbf{v}^t \mathbf{u}$ 。
10. a. 用習題 9 的結果以證明，若 A^{-1} 存在且 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ ，則若且唯若 $\mathbf{y}^t A^{-1} \mathbf{x} \neq -1$ ，則 $(A + \mathbf{x}\mathbf{y}^t)^{-1}$ 存在。
- b. 利用由右側乘 $A + \mathbf{x}\mathbf{y}^t$ ，證明當 $\mathbf{y}^t A^{-1} \mathbf{x} \neq -1$ 時，我們有

$$(A + \mathbf{x}\mathbf{y}^t)^{-1} = A^{-1} - \frac{A^{-1} \mathbf{x}\mathbf{y}^t A^{-1}}{1 + \mathbf{y}^t A^{-1} \mathbf{x}}$$

11. 在 8.1 節的習題 13 我們討論了一個問題，如何以 $R = bw^a$ 型式的最小平方關係，來近似一組天蛾幼蟲的重量與呼吸律的數據。在那個習題中，我們將問題轉換成 log-log 關係求解，並在它的 (b) 小題中加入一個二次項，以改善近似程度。在此我們不進行轉換，使用 8.1 節習題 13 的數據，求出可使 $\sum_{i=1}^n (R_i - bw_i^a)^2$ 為最小的常數 a 及 b 。算出此近似值的誤差，並與之前求得的誤差做比較。



10.4 最陡下降法

解非線性方程組時，牛頓法與準牛頓法的最大好處是，只要有了夠接近的初始值，它們收斂的非常快。但是它們的缺點就是，初始值必須夠接近才能確保收斂。在本節中要介紹**最陡下降** (Steepest Descent) 法，它只有線性收斂，但即使初始值很差，它通常也都會收斂。因此之故，此方法常被用來替其他牛頓型的方法找夠好的初始值，類似於解單一方程式的二分法。

對一個型式為 $g: \mathbb{R}^n \rightarrow \mathbb{R}$ 的多變數函數，最陡下降法求它的局部最小值。這個方法除了在求解非線性方程組時產生初始值，還有許多其他的用途。(在習題中包含了一些其他的應用)

將一個由 $\mathbb{R}^n$ 到 $\mathbb{R}$ 的函數最小化，和非線性方程組的解，兩者的關係是基於，一個型式為

$$\begin{aligned} f_1(x_1, x_2, \dots, x_n) &= 0 \\ f_2(x_1, x_2, \dots, x_n) &= 0 \\ &\vdots \\ f_n(x_1, x_2, \dots, x_n) &= 0 \end{aligned}$$

的方程組，只有在定義為

$$g(x_1, x_2, \dots, x_n) = \sum_{i=1}^n [f_i(x_1, x_2, \dots, x_n)]^2$$

的函數 g 有最小值 0 時，方程組有解 $\mathbf{x} = (x_1, x_2, \dots, x_n)^t$ 。

對任意由 $\mathbb{R}^n$ 到 $\mathbb{R}$ 的函數 g ，以最陡下降法求局部最小值的程序可簡述如下：

1. 利用初始近似值 $\mathbf{x}^{(0)} = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})^t$ 計算 g 的函數值。
2. 由 $\mathbf{x}^{(0)}$ 找出一個可使 g 的函數值下降的方向。
3. 沿此方向移動適當的距離，並令此新值為 $\mathbf{x}^{(1)}$ 。
4. 將 $\mathbf{x}^{(0)}$ 換成 $\mathbf{x}^{(1)}$ 重複步驟 1 到 3。

■ 函數的梯度

在說明如何決定適當的方向與移動距離之前，我們要再回顧一下微積分。極值定理 (Extreme Value Theorem) 指出，一個單變數函數只有在導數為 0 時有相對最小值。我們用以下定義將此推廣到多變數函數。

定義 10.9

對於 $g: \mathbb{R}^n \rightarrow \mathbb{R}$ ，其在 $\mathbf{x} = (x_1, x_2, \dots, x_n)^t$ 處的**梯度** (gradient) 記做 $\nabla g(\mathbf{x})$ ，並定義為

$$\nabla g(\mathbf{x}) = \left(\frac{\partial g}{\partial x_1}(\mathbf{x}), \frac{\partial g}{\partial x_2}(\mathbf{x}), \dots, \frac{\partial g}{\partial x_n}(\mathbf{x}) \right)^t \quad \blacksquare$$

多變數函數的梯度，可類比於單變數函數的導數，一個多變數函數只有當它在 $\mathbf{x}$ 處的梯度為 $\mathbf{0}$ 向量時，它在 $\mathbf{x}$ 處有相對最小值。梯度有另一個重要性質與多變數函數的最小化有關。設 $\mathbf{v} = (v_1, v_2, \dots, v_n)^t$ 是 $\mathbb{R}^n$ 中的單位向量；亦即

$$\|\mathbf{v}\|_2^2 = \sum_{i=1}^n v_i^2 = 1$$

g 在 $\mathbf{x}$ 處相對於方向 $\mathbf{v}$ 的**方向導數** (directional derivative) 代表了，自變數沿 $\mathbf{v}$ 方向改變時， g 的函數值的變量。定義為

$$D_{\mathbf{v}}g(\mathbf{x}) = \lim_{h \rightarrow 0} \frac{1}{h} [g(\mathbf{x} + h\mathbf{v}) - g(\mathbf{x})] = \mathbf{v}^t \cdot \nabla g(\mathbf{x})$$

當 g 為可微，選擇 $\mathbf{v}$ 平行於 $\nabla g(\mathbf{x})$ 的方向，可得最大的方向導數值，前提為 $\nabla g(\mathbf{x}) \neq \mathbf{0}$ 。所以，在 $\mathbf{x}$ 處能使 g 的函數值有最大減量的方向為 $-\nabla g(\mathbf{x})$ 。圖 10.3 顯示了 g 為雙變數函數的情形。

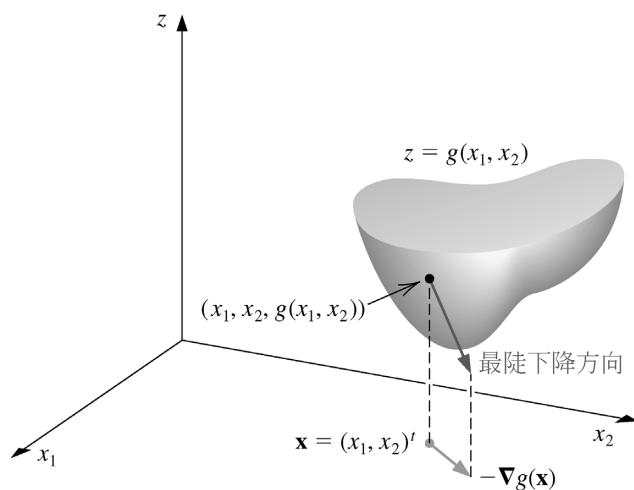


圖 10.3

目的在於將 $g(\mathbf{x})$ 的值減到最小值 0，所以 $\mathbf{x}^{(1)}$ 的適當選擇為，讓它沿著可使 $g(\mathbf{x})$ 有最大幅度下降的方向離開 $\mathbf{x}^{(0)}$ 。因此我們令

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} - \alpha \nabla g(\mathbf{x}^{(0)}) \quad , \quad \alpha > 0 \text{ 為某個常數。} \quad (10.17)$$

故問題簡化為，如何選擇 α 值，使得 $g(\mathbf{x}^{(1)})$ 比 $g(\mathbf{x}^{(0)})$ 小很多。

要決定適當的 α 值，我們考慮單變數函數

$$h(\alpha) = g(\mathbf{x}^{(0)} - \alpha \nabla g(\mathbf{x}^{(0)})) \quad (10.18)$$

能夠使 h 有最小值的 α ，就是 (10.17) 式所要的 α 值。

要直接獲得 h 的最小值，我們必須先對 h 微分，然後解一個求根問題，以找出 h 的轉折點 (critical point)。此過程的計算成本通常是太高了。所以我們採用其他方法，我們先選 3 個數 $\alpha_1 < \alpha_2 < \alpha_3$ ，希望它們很接近 $h(\alpha)$ 的最小值。然後我們建立一個在 α_1 、 α_2 及 α_3 內插 h 的二次多項式 $P(x)$ 。利用類似於 2.6 節的繆勒法，很容易求得一個二次多項式的最小值。

定義 $\hat{\alpha}$ 位於 $[\alpha_1, \alpha_3]$ 之間，可使得 $P(\hat{\alpha})$ 在 $[\alpha_1, \alpha_3]$ 之間為最小，並以 $P(\hat{\alpha})$ 做為 $h(\alpha)$ 之最小值的近似值。然後將 $\hat{\alpha}$ 用於新的迭代以求 g 的最小值：

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} - \hat{\alpha} \nabla g(\mathbf{x}^{(0)})$$

因為已經有了 $g(\mathbf{x}^{(0)})$ ，我們先選 $\alpha_1 = 0$ ，以減少計算量。然後找一個 α_3 使得 $h(\alpha_3) < h(\alpha_1)$ 。(因為 α_1 不能使 h 為最小，故一定存在有這樣的 α_3) 最後，選擇 α_2 等於 $\alpha_3/2$ 。

P 在 $[\alpha_1, \alpha_3]$ 間的最小值，應出現在 P 的唯一轉折點或是右側端點 α_3 ，因為我們的假設是 $P(\alpha_3) = h(\alpha_3) < h(\alpha_1) = P(\alpha_1)$ 。由於 P 是二次多項式，所以解一個線性方程式就可求得轉折點。

例題 1 用最陡下降法及 $\mathbf{x}^{(0)} = (0, 0, 0)^t$ 以找出合理的初始近似值，可用於求解非線性方程組

$$f_1(x_1, x_2, x_3) = 3x_1 - \cos(x_2x_3) - \frac{1}{2} = 0$$

$$f_2(x_1, x_2, x_3) = x_1^2 - 81(x_2 + 0.1)^2 + \sin x_3 + 1.06 = 0$$

$$f_3(x_1, x_2, x_3) = e^{-x_1x_2} + 20x_3 + \frac{10\pi - 3}{3} = 0$$

解 令 $g(x_1, x_2, x_3) = [f_1(x_1, x_2, x_3)]^2 + [f_2(x_1, x_2, x_3)]^2 + [f_3(x_1, x_2, x_3)]^2$ 。則

$$\begin{aligned} \nabla g(x_1, x_2, x_3) &\equiv \nabla g(\mathbf{x}) = \begin{pmatrix} 2f_1(\mathbf{x}) \frac{\partial f_1}{\partial x_1}(\mathbf{x}) + 2f_2(\mathbf{x}) \frac{\partial f_2}{\partial x_1}(\mathbf{x}) + 2f_3(\mathbf{x}) \frac{\partial f_3}{\partial x_1}(\mathbf{x}) \\ 2f_1(\mathbf{x}) \frac{\partial f_1}{\partial x_2}(\mathbf{x}) + 2f_2(\mathbf{x}) \frac{\partial f_2}{\partial x_2}(\mathbf{x}) + 2f_3(\mathbf{x}) \frac{\partial f_3}{\partial x_2}(\mathbf{x}) \\ 2f_1(\mathbf{x}) \frac{\partial f_1}{\partial x_3}(\mathbf{x}) + 2f_2(\mathbf{x}) \frac{\partial f_2}{\partial x_3}(\mathbf{x}) + 2f_3(\mathbf{x}) \frac{\partial f_3}{\partial x_3}(\mathbf{x}) \end{pmatrix} \\ &= 2\mathbf{J}(\mathbf{x})^t \mathbf{F}(\mathbf{x}) \end{aligned}$$

由 $\mathbf{x}^{(0)} = (0, 0, 0)^t$ ，我們得

$$g(\mathbf{x}^{(0)}) = 111.975 \text{ 及 } z_0 = \|\nabla g(\mathbf{x}^{(0)})\|_2 = 419.554$$

令

$$\mathbf{z} = \frac{1}{z_0} \nabla g(\mathbf{x}^{(0)}) = (-0.0214514, -0.0193062, 0.999583)^t$$

由 $\alpha_1 = 0$ ，我們得 $g_1 = g(\mathbf{x}^{(0)} - \alpha_1 \mathbf{z}) = g(\mathbf{x}^{(0)}) = 111.975$ 。我們任意選擇 $\alpha_3 = 1$ 可得

$$g_3 = g(\mathbf{x}^{(0)} - \alpha_3 \mathbf{z}) = 93.5649$$

因為 $g_3 < g_1$ 所以 α_3 可以接受，我們設 $\alpha_2 = \alpha_3/2 = 0.5$ 。因此

$$g_2 = g(\mathbf{x}^{(0)} - \alpha_2 \mathbf{z}) = 2.53557$$

我們現在求內插 $(0, 111.975)$ 、 $(1, 93.5649)$ 、和 $(0.5, 2.53557)$ 的二次多項式。此處使用牛頓前向均差 (forward divided-difference) 內插多項式最方便，其型式為

$$P(\alpha) = g_1 + h_1\alpha + h_3\alpha(\alpha - \alpha_2)$$

此多項式內插

$$g(\mathbf{x}^{(0)} - \alpha \nabla g(\mathbf{x}^{(0)})) = g(\mathbf{x}^{(0)} - \alpha \mathbf{z})$$

於 $\alpha_1 = 0$ 、 $\alpha_2 = 0.5$ 及 $\alpha_3 = 1$ 如下：

$$\alpha_1 = 0, \quad g_1 = 111.975$$

$$\alpha_2 = 0.5, \quad g_2 = 2.53557, \quad h_1 = \frac{g_2 - g_1}{\alpha_2 - \alpha_1} = -218.878$$

$$\alpha_3 = 1, \quad g_3 = 93.5649, \quad h_2 = \frac{g_3 - g_2}{\alpha_3 - \alpha_2} = 182.059, \quad h_3 = \frac{h_2 - h_1}{\alpha_3 - \alpha_1} = 400.937$$

因此，

$$P(\alpha) = 111.975 - 218.878\alpha + 400.937\alpha(\alpha - 0.5)$$

在 $\alpha = \alpha_0 = 0.522959$ ，我們有 $P'(\alpha) = 0$ 。因為 $g_0 = g(\mathbf{x}^{(0)} - \alpha_0 \mathbf{z}) = 2.32762$ 小於 g_1 及 g_3 ，我們設定

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} - \alpha_0 \mathbf{z} = \mathbf{x}^{(0)} - 0.522959 \mathbf{z} = (0.0112182, 0.0100964, -0.522741)^t$$

及

$$g(\mathbf{x}^{(1)}) = 2.32762$$

表 10.5 列出了其餘的計算結果。此非線性方程組有一個真實解 $(0.5, 0, -0.5235988)^t$ ，所以要做為牛頓法或 Broyden 法的初始值， $\mathbf{x}^{(2)}$ 應該就足夠了。由此開始應該要換用這些快速收斂的方法，因為最陡下降法要收斂到 $\|\mathbf{x}^{(k)} - \mathbf{x}\|_\infty < 0.01$ 須要 70 次迭代。 ■

表 10.5

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$g(x_1^{(k)}, x_2^{(k)}, x_3^{(k)})$
2	0.137860	-0.205453	-0.522059	1.27406
3	0.266959	0.00551102	-0.558494	1.06813
4	0.272734	-0.00811751	-0.522006	0.468309
5	0.308689	-0.0204026	-0.533112	0.381087
6	0.314308	-0.0147046	-0.520923	0.318837
7	0.324267	-0.00852549	-0.528431	0.287024

算則 10.3 使用最陡下降法以近似 $g(\mathbf{x})$ 的最小值。開始迭代時，設定 α_1 為 0 並且設 α_3 為 1。如果 $h(\alpha_3) \geq h(\alpha_1)$ ，則持續將 α_3 除 2，直到 $h(\alpha_3) < h(\alpha_1)$ 且 $\alpha_3 = 2^{-k}$ 為止。

要用此方法求方程組

$$f_1(x_1, x_2, \dots, x_n) = 0$$

$$f_2(x_1, x_2, \dots, x_n) = 0$$

$$\vdots$$

$$f_n(x_1, x_2, \dots, x_n) = 0$$

的近似解，只要用 $\sum_{i=1}^n f_i^2$ 取代函數 g 即可。

算則 10.3 最陡下降法 (Steepest Descent)

已知初始近似值 $\mathbf{x}$ ，求最小化問題

$$g(\mathbf{p}) = \min_{\mathbf{x} \in \mathbb{R}^n} g(\mathbf{x})$$

之解 $\mathbf{p}$ 的近似值：

INPUT 變數個數 n ；初始近似值 $\mathbf{x} = (x_1, \dots, x_n)^t$ ；容許誤差 TOL ；最大迭代次數 N 。

OUTPUT 近似解 $\mathbf{x} = (x_1, \dots, x_n)^t$ 或超過最大迭代次數的訊息。

Step 1 Set $k = 1$.

Step 2 While $(k \leq N)$ do Steps 3–15.

Step 3 Set $g_1 = g(x_1, \dots, x_n)$; (∇ : $g_1 = g(\mathbf{x}^{(k)})$.)
 $\mathbf{z} = \nabla g(x_1, \dots, x_n)$; (∇ : $\mathbf{z} = \nabla g(\mathbf{x}^{(k)})$.)
 $z_0 = \|\mathbf{z}\|_2$.

Step 4 If $z_0 = 0$ then OUTPUT ('Zero gradient');
 OUTPUT $(x_1, \dots, x_n, g_1)$;
 (程序完成，可能有最小值)
 STOP.

Step 5 Set $\mathbf{z} = \mathbf{z}/z_0$; (使 $\mathbf{z}$ 成為單位向量)

$$\alpha_1 = 0;$$

$$\alpha_3 = 1;$$

$$g_3 = g(\mathbf{x} - \alpha_3 \mathbf{z}).$$

Step 6 While ($g_3 \geq g_1$) do Steps 7 and 8.

Step 7 Set $\alpha_3 = \alpha_3/2$;

$$g_3 = g(\mathbf{x} - \alpha_3 \mathbf{z}).$$

Step 8 If $\alpha_3 < TOL/2$ then

OUTPUT ('No likely improvement');

OUTPUT ($x_1, \dots, x_n, g_1$);

(程序完成，可能有最小值)

STOP.

Step 9 Set $\alpha_2 = \alpha_3/2$;

$$g_2 = g(\mathbf{x} - \alpha_2 \mathbf{z}).$$

Step 10 Set $h_1 = (g_2 - g_1)/\alpha_2$;

$$h_2 = (g_3 - g_2)/(\alpha_3 - \alpha_2);$$

$$h_3 = (h_2 - h_1)/\alpha_3.$$

(註：用牛頓前向除差公式求，在 $\alpha = 0$ 、 $\alpha = \alpha_2$ 、 $\alpha = \alpha_3$

內插 $h(\alpha)$ 的二次式 $P(\alpha) = g_1 + h_1\alpha + h_3\alpha(\alpha - \alpha_2)$)

Step 11 Set $\alpha_0 = 0.5(\alpha_2 - h_1/h_3)$; (P 的轉折點出現在 α_0 .)

$$g_0 = g(\mathbf{x} - \alpha_0 \mathbf{z}).$$

Step 12 Find α from $\{\alpha_0, \alpha_3\}$ so that $g = g(\mathbf{x} - \alpha \mathbf{z}) = \min\{g_0, g_3\}$.

Step 13 Set $\mathbf{x} = \mathbf{x} - \alpha \mathbf{z}$.

Step 14 If $|g - g_1| < TOL$ then

OUTPUT ($x_1, \dots, x_n, g$);

(程式成功)

STOP.

Step 15 Set $k = k + 1$.

Step 16 OUTPUT ('Maximum iterations exceeded');

(程式失敗)

STOP. ■

最陡下降法有許多不同的版本，有的用很複雜的方法求 α 的值，以獲得 (10.18) 式所定義之單變數函數 h 的最小值。有的版本則是用多維泰勒多項式取代多變數函數 g ，然後求多項式的最小值。雖然這些方法各有其優點，但所有的最陡下降法都是線性收斂，且其收斂與初始值無關。不過在某些情況下它可能不會收斂到函數 g 的絕對最小值。

關於最陡下降法更詳細的探討可參考 [OR] 或 [RR]。

習題組 10.4 完整習題請見隨書光碟

1. 用最陡下降法及 $TOL = 0.05$ 求下列各非線性方程組的近似解。

a. $4x_1^2 - 20x_1 + \frac{1}{4}x_2^2 + 8 = 0$

$$\frac{1}{2}x_1x_2^2 + 2x_1 - 5x_2 + 8 = 0$$

c. $\ln(x_1^2 + x_2^2) - \sin(x_1x_2) = \ln 2 + \ln \pi$
 $e^{x_1 - x_2} + \cos(x_1x_2) = 0$

b. $3x_1^2 - x_2^2 = 0$

$$3x_1x_2^2 - x_1^3 - 1 = 0$$

d. $\sin(4\pi x_1x_2) - 2x_2 - x_1 = 0$

$$\left(\frac{4\pi - 1}{4\pi}\right)(e^{2x_1} - e) + 4ex_2^2 - 2ex_1 = 0$$

2. 用最陡下降法及 $TOL = 0.05$ 求下列各非線性方程組的近似解。

a. $15x_1 + x_2^2 - 4x_3 = 13$

$$x_1^2 + 10x_2 - x_3 = 11$$

$$x_2^3 - 25x_3 = -22$$

c. $x_1^3 + x_1^2x_2 - x_1x_3 + 6 = 0$

$$e^{x_1} + e^{x_2} - x_3 = 0$$

$$x_2^2 - 2x_1x_3 = 4$$

b. $10x_1 - 2x_2^2 + x_2 - 2x_3 - 5 = 0$

$$8x_2^2 + 4x_3^2 - 9 = 0$$

$$8x_2x_3 + 4 = 0$$

d. $x_1 + \cos(x_1x_2x_3) - 1 = 0$

$$(1 - x_1)^{1/4} + x_2 + 0.05x_3^2 - 0.15x_3 - 1 = 0$$

$$-x_1^2 - 0.1x_2^2 + 0.01x_2 + x_3 - 1 = 0$$

3. 用習題 1 的結果及牛頓法，求習題 1 各非線性方程組的近似解，準確至 10^{-6} 。

4. 用習題 2 的結果及牛頓法，求習題 2 各非線性方程組的近似解，準確至 10^{-6} 。

5. 用最陡下降法求下列函數最小值的近似解，準確至 0.005。

a. $g(x_1, x_2) = \cos(x_1 + x_2) + \sin x_1 + \cos x_2$

b. $g(x_1, x_2) = 100(x_1^2 - x_2)^2 + (1 - x_1)^2$

c. $g(x_1, x_2, x_3) = x_1^2 + 2x_2^2 + x_3^2 - 2x_1x_2 + 2x_1 - 2.5x_2 - x_3 + 2$

d. $g(x_1, x_2, x_3) = x_1^4 + 2x_2^4 + 3x_3^4 + 1.01$

6. a. 證明二次多項式

$$P(\alpha) = g_1 + h_1\alpha + h_3\alpha(\alpha - \alpha_2)$$

內插 (10.18) 式所定義之函數 h ：

$$h(\alpha) = g(\mathbf{x}^{(0)} - \alpha \nabla g(\mathbf{x}^{(0)}))$$

於 $\alpha = 0$ 、 α_2 及 α_3 這 3 點。

b. 證明 P 的轉折點 (critical point) 出現於

$$\alpha_0 = \frac{1}{2} \left(\alpha_2 - \frac{h_1}{h_3} \right)$$

10.5 同倫及延拓法

用於非線性方程組的同倫法或延拓法(*homotopy or continuation*)，是將要求解的問題植入在一組問題中。明確的說，如果要解

$$\mathbf{F}(\mathbf{x}) = \mathbf{0}$$

型的問題，其未知解為 $\mathbf{x}^*$ ，我們考慮由參數 λ 所描述的一族問題， λ 值在 $[0, 1]$ 之間。當問題的解 $\mathbf{x}(0)$ 為已知時則 $\lambda = 0$ ，問題的未知解 $\mathbf{x}(1) \equiv \mathbf{x}^*$ 則相當於 $\lambda = 1$ 。

舉例來說，設 $\mathbf{x}(0)$ 是 $\mathbf{F}(\mathbf{x}^*) = \mathbf{0}$ 的初始近似解。定義

$$\mathbf{G} : [0, 1] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$$

為

$$\mathbf{G}(\lambda, \mathbf{x}) = \lambda \mathbf{F}(\mathbf{x}) + (1 - \lambda) [\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{x}(0))] = \mathbf{F}(\mathbf{x}) + (\lambda - 1) \mathbf{F}(\mathbf{x}(0)) \quad (10.19)$$

我們要求出不同 λ 值時

$$\mathbf{G}(\lambda, \mathbf{x}) = \mathbf{0}$$

的解。當 $\lambda = 0$ ，方程式的型式為

$$\mathbf{0} = \mathbf{G}(0, \mathbf{x}) = \mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{x}(0))$$

而 $\mathbf{x}(0)$ 為其一解。當 $\lambda = 1$ ，此方程式的型式為

$$\mathbf{0} = \mathbf{G}(1, \mathbf{x}) = \mathbf{F}(\mathbf{x})$$

而 $\mathbf{x}(1) = \mathbf{x}^*$ 為其解。

函數 $\mathbf{G}$ 與參數 λ 提供了一族的函數，可以由已知的 $\mathbf{x}(0)$ 到要求的 $\mathbf{x}(1) = \mathbf{x}^*$ 。函數 $\mathbf{G}$ 叫做，函數 $\mathbf{G}(0, \mathbf{x}) = \mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{x}(0))$ 與函數 $\mathbf{G}(1, \mathbf{x}) = \mathbf{F}(\mathbf{x})$ 之間的同倫函數 (homotopy)。

延拓法

延拓問題則是：

- 找出一條，由 $\mathbf{G}(0, \mathbf{x}) = \mathbf{0}$ 的已知解 $\mathbf{x}(0)$ 到 $\mathbf{G}(1, \mathbf{x}) = \mathbf{0}$ 的未知解 $\mathbf{x}(1) = \mathbf{x}^*$ 的路徑，也就是找到 $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ 的解。

我們先假設，對所有的 $\lambda \in [0, 1]$ ， $\mathbf{x}(\lambda)$ 是方程式

$$\mathbf{G}(\lambda, \mathbf{x}) = \mathbf{0} \quad (10.20)$$

的唯一解。則集合 $\{\mathbf{x}(\lambda) \mid 0 \leq \lambda \leq 1\}$ 可以被視為是 $\mathbb{R}^n$ 中的一條曲線，它由 $\mathbf{x}(0)$ 到 $\mathbf{x}(1) = \mathbf{x}^*$ 並以 λ 為參數。延拓法就是要在不同的 λ 值，沿著這條曲線求出序列 $\{\mathbf{x}(\lambda_k)\}_{k=0}^m$ ，其中 $\lambda_0 = 0 < \lambda_1 < \cdots < \lambda_m = 1$ 。

若函數 $\lambda \rightarrow \mathbf{x}(\lambda)$ 和 $\mathbf{G}$ 為可微，則將 (10.20) 式對 λ 微分可得

$$\mathbf{0} = \frac{\partial \mathbf{G}(\lambda, \mathbf{x}(\lambda))}{\partial \lambda} + \frac{\partial \mathbf{G}(\lambda, \mathbf{x}(\lambda))}{\partial \mathbf{x}} \mathbf{x}'(\lambda)$$

並由其解 $\mathbf{x}'(\lambda)$ 得

$$\mathbf{x}'(\lambda) = - \left[\frac{\partial \mathbf{G}(\lambda, \mathbf{x}(\lambda))}{\partial \mathbf{x}} \right]^{-1} \frac{\partial \mathbf{G}(\lambda, \mathbf{x}(\lambda))}{\partial \lambda}$$

這是一個聯立微分方程組，初始條件為 $\mathbf{x}(0)$ 。

因為

$$\mathbf{G}(\lambda, \mathbf{x}(\lambda)) = \mathbf{F}(\mathbf{x}(\lambda)) + (\lambda - 1)\mathbf{F}(\mathbf{x}(0))$$

我們可以求出 Jacobian 矩陣

$$\frac{\partial \mathbf{G}}{\partial \mathbf{x}}(\lambda, \mathbf{x}(\lambda)) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}(\lambda)) & \frac{\partial f_1}{\partial x_2}(\mathbf{x}(\lambda)) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}(\lambda)) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{x}(\lambda)) & \frac{\partial f_2}{\partial x_2}(\mathbf{x}(\lambda)) & \cdots & \frac{\partial f_2}{\partial x_n}(\mathbf{x}(\lambda)) \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1}(\mathbf{x}(\lambda)) & \frac{\partial f_n}{\partial x_2}(\mathbf{x}(\lambda)) & \cdots & \frac{\partial f_n}{\partial x_n}(\mathbf{x}(\lambda)) \end{bmatrix} = J(\mathbf{x}(\lambda))$$

及

$$\frac{\partial \mathbf{G}(\lambda, \mathbf{x}(\lambda))}{\partial \lambda} = \mathbf{F}(\mathbf{x}(0))$$

因此，原微分方程組成為

$$\mathbf{x}'(\lambda) = -[J(\mathbf{x}(\lambda))]^{-1} \mathbf{F}(\mathbf{x}(0)) \quad , \quad 0 \leq \lambda \leq 1 \quad (10.21)$$

及初始條件 $\mathbf{x}(0)$ 。以下定理 (見 [OR], pp. 230-231) 為延拓法可行之條件。

定理 10.10

令 $\mathbf{F}(\mathbf{x})$ 對 $\mathbf{x} \in \mathbb{R}^n$ 為連續可微。假設對所有的 $\mathbf{x} \in \mathbb{R}^n$ ，Jacobian 矩陣 $J(\mathbf{x})$ 為非奇異，且存在有常數 M 使得 $\|J(\mathbf{x})^{-1}\| \leq M$ 。則對任何屬於 $\mathbb{R}^n$ 的 $\mathbf{x}(0)$ ，存在有唯一的函數 $\mathbf{x}(\lambda)$ ，對所有在 $[0, 1]$ 中的 λ 都有

$$\mathbf{G}(\lambda, \mathbf{x}(\lambda)) = \mathbf{0}$$

此外， $\mathbf{x}(\lambda)$ 為連續可微，且對每個 $\lambda \in [0, 1]$ ，

$$\mathbf{x}'(\lambda) = -J(\mathbf{x}(\lambda))^{-1} \mathbf{F}(\mathbf{x}(0)) \quad \blacksquare$$

以下說明題顯示了，併隨非線性方程組之微分方程組的型式。

說明題 考慮非線性方程組

$$f_1(x_1, x_2, x_3) = 3x_1 - \cos(x_2x_3) - 0.5 = 0$$

$$f_2(x_1, x_2, x_3) = x_1^2 - 81(x_2 + 0.1)^2 + \sin x_3 + 1.06 = 0$$

$$f_3(x_1, x_2, x_3) = e^{-x_1x_2} + 20x_3 + \frac{10\pi - 3}{3} = 0$$

其 Jacobian 矩陣為

$$J(\mathbf{x}) = \begin{bmatrix} 3 & x_3 \sin x_2x_3 & x_2 \sin x_2x_3 \\ 2x_1 & -162(x_2 + 0.1) & \cos x_3 \\ -x_2e^{-x_1x_2} & -x_1e^{-x_1x_2} & 20 \end{bmatrix}$$

令 $\mathbf{x}(0) = (0, 0, 0)^t$ ，所以

$$\mathbf{F}(\mathbf{x}(0)) = \begin{bmatrix} -1.5 \\ 0.25 \\ 10\pi/3 \end{bmatrix}$$

微分方程組為

$$\begin{bmatrix} x'_1(\lambda) \\ x'_2(\lambda) \\ x'_3(\lambda) \end{bmatrix} = - \begin{bmatrix} 3 & x_3 \sin x_2x_3 & x_2 \sin x_2x_3 \\ 2x_1 & -162(x_2 + 0.1) & \cos x_3 \\ -x_2e^{-x_1x_2} & -x_1e^{-x_1x_2} & 20 \end{bmatrix}^{-1} \begin{bmatrix} -1.5 \\ 0.25 \\ 10\pi/3 \end{bmatrix} \quad \blacksquare$$

在延拓問題中，我們要求解的微分方程組的通式為

$$\begin{aligned} \frac{dx_1}{d\lambda} &= \phi_1(\lambda, x_1, x_2, \dots, x_n) \\ \frac{dx_2}{d\lambda} &= \phi_2(\lambda, x_1, x_2, \dots, x_n) \\ &\vdots \\ \frac{dx_n}{d\lambda} &= \phi_n(\lambda, x_1, x_2, \dots, x_n) \end{aligned}$$

其中

$$\begin{bmatrix} \phi_1(\lambda, x_1, \dots, x_n) \\ \phi_2(\lambda, x_1, \dots, x_n) \\ \vdots \\ \phi_n(\lambda, x_1, \dots, x_n) \end{bmatrix} = -J(x_1, \dots, x_n)^{-1} \begin{bmatrix} f_1(\mathbf{x}(0)) \\ f_2(\mathbf{x}(0)) \\ \vdots \\ f_n(\mathbf{x}(0)) \end{bmatrix} \quad (10.22)$$

要使用四階 Runge-Kutta 法求解此方程組，我們先選一整數 $N > 0$ 並令 $h = (1 - 0)/N$ 。將區間 $[0, 1]$ 劃分為 N 個次區段，且格點為

$$\lambda_j = jh \quad , \quad j = 0, 1, \dots, N$$

我們用符號 w_{ij} 代表 $x_i(\lambda_j)$ 的近似值，其中 $j = 0, 1, \dots, N$ 且 $i = 1, \dots, n$ 。對於初始條件，設定

$$w_{1,0} = x_1(0), \quad w_{2,0} = x_2(0), \quad \dots, \quad w_{n,0} = x_n(0)$$

設已求得 $w_{1,j}, w_{2,j}, \dots, w_{n,j}$ 。要進一步求 $w_{1,j+1}, w_{2,j+1}, \dots, w_{n,j+1}$ ，我們使用方程式

$$k_{1,i} = h\phi_i(\lambda_j, w_{1,j}, w_{2,j}, \dots, w_{n,j}) \quad , \quad i = 1, 2, \dots, n$$

$$k_{2,i} = h\phi_i\left(\lambda_j + \frac{h}{2}, w_{1,j} + \frac{1}{2}k_{1,1}, \dots, w_{n,j} + \frac{1}{2}k_{1,n}\right) \quad , \quad i = 1, 2, \dots, n$$

$$k_{3,i} = h\phi_i\left(\lambda_j + \frac{h}{2}, w_{1,j} + \frac{1}{2}k_{2,1}, \dots, w_{n,j} + \frac{1}{2}k_{2,n}\right) \quad , \quad i = 1, 2, \dots, n$$

$$k_{4,i} = h\phi_i(\lambda_j + h, w_{1,j} + k_{3,1}, w_{2,j} + k_{3,2}, \dots, w_{n,j} + k_{3,n}) \quad , \quad i = 1, 2, \dots, n$$

以及，

$$w_{i,j+1} = w_{i,j} + \frac{1}{6}(k_{1,i} + 2k_{2,i} + 2k_{3,i} + k_{4,i}) \quad , \quad i = 1, 2, \dots, n$$

我們用向量的符號

$$\mathbf{k}_1 = \begin{bmatrix} k_{1,1} \\ k_{1,2} \\ \vdots \\ k_{1,n} \end{bmatrix}, \quad \mathbf{k}_2 = \begin{bmatrix} k_{2,1} \\ k_{2,2} \\ \vdots \\ k_{2,n} \end{bmatrix}, \quad \mathbf{k}_3 = \begin{bmatrix} k_{3,1} \\ k_{3,2} \\ \vdots \\ k_{3,n} \end{bmatrix}, \quad \mathbf{k}_4 = \begin{bmatrix} k_{4,1} \\ k_{4,2} \\ \vdots \\ k_{4,n} \end{bmatrix}, \quad \text{及} \quad \mathbf{w}_j = \begin{bmatrix} w_{1,j} \\ w_{2,j} \\ \vdots \\ w_{n,j} \end{bmatrix}$$

以簡化表示方式。由 (10.22) 式得 $\mathbf{x}(0) = \mathbf{x}(\lambda_0) = \mathbf{w}_0$ ，且對每個 $j = 0, 1, \dots, N$ ，

$$\begin{aligned} \mathbf{k}_1 &= h \begin{bmatrix} \phi_1(\lambda_j, w_{1,j}, \dots, w_{n,j}) \\ \phi_2(\lambda_j, w_{1,j}, \dots, w_{n,j}) \\ \vdots \\ \phi_n(\lambda_j, w_{1,j}, \dots, w_{n,j}) \end{bmatrix} = h [-J(w_{1,j}, \dots, w_{n,j})]^{-1} \mathbf{F}(\mathbf{x}(0)) \\ &= h [-J(\mathbf{w}_j)]^{-1} \mathbf{F}(\mathbf{x}(0)) \end{aligned}$$

$$\mathbf{k}_2 = h \left[-J \left(\mathbf{w}_j + \frac{1}{2} \mathbf{k}_1 \right) \right]^{-1} \mathbf{F}(\mathbf{x}(0))$$

$$\mathbf{k}_3 = h \left[-J \left(\mathbf{w}_j + \frac{1}{2} \mathbf{k}_2 \right) \right]^{-1} \mathbf{F}(\mathbf{x}(0))$$

$$\mathbf{k}_4 = h [-J(\mathbf{w}_j + \mathbf{k}_3)]^{-1} \mathbf{F}(\mathbf{x}(0))$$

及

$$\mathbf{x}(\lambda_{j+1}) = \mathbf{x}(\lambda_j) + \frac{1}{6}(\mathbf{k}_1 + 2\mathbf{k}_2 + 2\mathbf{k}_3 + \mathbf{k}_4) = \mathbf{w}_j + \frac{1}{6}(\mathbf{k}_1 + 2\mathbf{k}_2 + 2\mathbf{k}_3 + \mathbf{k}_4)$$

最後， $\mathbf{x}(\lambda_n) = \mathbf{x}(1)$ 即為 $\mathbf{x}^*$ 的近似值。

例題 1 用延拓法及 $\mathbf{x}(0) = (0, 0, 0)'$ 求

$$f_1(x_1, x_2, x_3) = 3x_1 - \cos(x_2 x_3) - 0.5 = 0$$

$$f_2(x_1, x_2, x_3) = x_1^2 - 81(x_2 + 0.1)^2 + \sin x_3 + 1.06 = 0$$

$$f_3(x_1, x_2, x_3) = e^{-x_1 x_2} + 20x_3 + \frac{10\pi - 3}{3} = 0$$

的近似解。

解 其 Jacobian 矩陣為

$$J(\mathbf{x}) = \begin{bmatrix} 3 & x_3 \sin x_2 x_3 & x_2 \sin x_2 x_3 \\ 2x_1 & -162(x_2 + 0.1) & \cos x_3 \\ -x_2 e^{-x_1 x_2} & -x_1 e^{-x_1 x_2} & 20 \end{bmatrix}$$

且

$$F(\mathbf{x}(0)) = (-1.5, 0.25, 10\pi/3)^t$$

使用 $N=4$ 及 $h=0.25$ ，我們得

$$\begin{aligned} \mathbf{k}_1 &= h[-J(\mathbf{x}^{(0)})]^{-1} F(\mathbf{x}(0)) = 0.25 \begin{bmatrix} 3 & 0 & 0 \\ 0 & -16.2 & 1 \\ 0 & 0 & 20 \end{bmatrix}^{-1} \begin{bmatrix} -1.5 \\ 0.25 \\ 10\pi/3 \end{bmatrix} \\ &= (0.125, -0.004222203325, -0.1308996939)^t \\ \mathbf{k}_2 &= h[-J(0.0625, -0.002111101663, -0.06544984695)]^{-1} (-1.5, 0.25, 10\pi/3)^t \\ &= 0.25 \begin{bmatrix} 3 & -0.9043289149 \times 10^{-5} & -0.2916936196 \times 10^{-6} \\ 0.125 & -15.85800153 & 0.9978589232 \\ 0.002111380229 & -0.06250824706 & 20 \end{bmatrix}^{-1} \begin{bmatrix} -1.5 \\ 0.25 \\ 10\pi/3 \end{bmatrix} \\ &= (0.1249999773, -0.003311761993, -0.1309232406)^t; \\ \mathbf{k}_3 &= h[-J(0.06249998865, -0.001655880997, -0.0654616203)]^{-1} (-1.5, 0.25, 10\pi/3)^t \\ &= (0.1249999844, -0.003296244825, -0.130920346)^t; \\ \mathbf{k}_4 &= h[-J(0.1249999844, -0.003296244825, -0.130920346)]^{-1} (-1.5, 0.25, 10\pi/3)^t \\ &= (0.1249998945, -0.00230206762, -0.1309346977)^t \end{aligned}$$

及

$$\begin{aligned} \mathbf{x}(\lambda_1) &= \mathbf{w}_1 = \mathbf{w}_0 + \frac{1}{6}(\mathbf{k}_1 + 2\mathbf{k}_2 + 2\mathbf{k}_3 + \mathbf{k}_4) \\ &= (0.1249999697, -0.00329004743, -0.1309202608)^t \end{aligned}$$

繼續此程序可得

$$\begin{aligned} \mathbf{x}(\lambda_2) &= \mathbf{w}_2 = (0.2499997679, -0.004507400128, -0.2618557619)^t \\ \mathbf{x}(\lambda_3) &= \mathbf{w}_3 = (0.3749996956, -0.003430352103, -0.3927634423)^t \end{aligned}$$

及

$$\mathbf{x}(\lambda_4) = \mathbf{x}(1) = \mathbf{w}_4 = (0.4999999954, 0.126782 \times 10^{-7}, -0.5235987758)^t$$

以上結果非常準確，因為真實解為 $(0.5, 0, -0.52359877)^t$ 。

對於 Runge-Kutta 法要說明一點，類似

$$\mathbf{k}_i = h[-J(\mathbf{x}(\lambda_i) + \alpha_{i-1}\mathbf{k}_{i-1})]^{-1}\mathbf{F}(\mathbf{x}(0))$$

的步驟，可以寫成是對線性方程組

$$J(\mathbf{x}(\lambda_i) + \alpha_{i-1}\mathbf{k}_{i-1})\mathbf{k}_i = -h\mathbf{F}(\mathbf{x}(0))$$

求解 $\mathbf{k}_i$ 。所以在用四階 Runge-Kutta 法計算 $\mathbf{w}_j$ 時，每一個須要解 4 個線性方程組，計算 $\mathbf{k}_1$ 、 $\mathbf{k}_2$ 、 $\mathbf{k}_3$ 及 $\mathbf{k}_4$ 時各一次。因此執行 N 步共須解 $4N$ 個線性方程組。而相較之下，牛頓法每次迭代只須要解一個線性方程組。因此在此程序中四階 Runge-Kutta 法的計算量約相當於牛頓法做 $4N$ 次迭代。

可以用二階 Runge-Kutta 法，例如改良歐拉法或甚至歐拉法，以減少求解線性方程組的數量。另一種可能性是，使用較小的 N 值。下個說明題即說明這些概念。

說明題 對於上面例題中的方程組，使用初始近似值 $\mathbf{x}(0) = (0, 0, 0)^t$ ，表 10.6 列出了歐拉法、中點法及四階 Runge-Kutta 法的結果供比較。最右側的一行列出求得近似解所須求解之線性方程組的數目。

表 10.6

方法	N	$\mathbf{x}(1)$	次數
歐拉法	1	$(0.5, -0.0168888133, -0.5235987755)^t$	1
歐拉法	4	$(0.499999379, -0.004309160698, -0.523679652)^t$	4
中點法	1	$(0.4999966628, -0.00040240435, -0.523815371)^t$	2
中點法	4	$(0.500000066, -0.00001760089, -0.5236127761)^t$	8
Runge-Kutta	1	$(0.4999989843, -0.1676151 \times 10^{-5}, -0.5235989561)^t$	4
Runge-Kutta	4	$(0.4999999954, 0.126783 \times 10^{-7}, -0.5235987758)^t$	16

延拓法可以單獨使用，且不須要特別選取 $\mathbf{x}(0)$ 。但它也可用來替牛頓法或 Broyden 法求初始近似解。例如在以上說明題中，使用歐拉法及 $N = 2$ 即可輕易獲得足以起動牛頓法或 Broyden 法的結果。這也是延拓法較佳的用途，因為它須要的計算量較多。

算則 10.4 延拓法算則 (Continuation Algorithm)

已知初始近似值 $\mathbf{x}$ ，求 $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ 的近似解：

INPUT 方程式與未知數的個數 n ；整數 $N > 0$ ；初始近似值

$$\mathbf{x} = (x_1, x_2, \dots, x_n)^t.$$

OUTPUT 近似解 $\mathbf{x} = (x_1, x_2, \dots, x_n)^t$

Step 1 Set $h = 1/N$;

$$\mathbf{b} = -h\mathbf{F}(\mathbf{x}).$$

Step 2 For $i = 1, 2, \dots, N$ do Steps 3–7.

Step 3 Set $A = J(\mathbf{x})$;

$$\text{Solve the linear system } A\mathbf{k}_1 = \mathbf{b}.$$

Step 4 Set $A = J(\mathbf{x} + \frac{1}{2}\mathbf{k}_1)$;

$$\text{Solve the linear system } A\mathbf{k}_2 = \mathbf{b}.$$

Step 5 Set $A = J(\mathbf{x} + \frac{1}{2}\mathbf{k}_2)$;

$$\text{Solve the linear system } A\mathbf{k}_3 = \mathbf{b}.$$

Step 6 Set $A = J(\mathbf{x} + \mathbf{k}_3)$;

$$\text{Solve the linear system } A\mathbf{k}_4 = \mathbf{b}.$$

Step 7 Set $\mathbf{x} = \mathbf{x} + (\mathbf{k}_1 + 2\mathbf{k}_2 + 2\mathbf{k}_3 + \mathbf{k}_4)/6$.

Step 8 OUTPUT $(x_1, x_2, \dots, x_n)$;

STOP. ■

習題組 10.5 完整習題請見隨書光碟

1. 非線性方程組

$$f_1(x_1, x_2) = x_1^2 - x_2^2 + 2x_2 = 0, \quad f_2(x_1, x_2) = 2x_1 + x_2^2 - 6 = 0$$

有 2 個解 $(0.625204094, 2.179355825)^t$ 及 $(2.109511920, -1.334532188)^t$ 。用延拓法及歐拉法和 $N = 2$ ，在下列條件下求其近似解。

a. $\mathbf{x}(0) = (0, 0)^t$ b. $\mathbf{x}(0) = (1, 1)^t$ c. $\mathbf{x}(0) = (3, -2)^t$

2. 用四階 Runge-Kutta 法及 $N = 1$ 重複習題 1。

3. 用延拓法及歐拉法和 $N = 2$ 求解下列非線性方程組。

a. $4x_1^2 - 20x_1 + \frac{1}{4}x_2^2 + 8 = 0$

$$\frac{1}{2}x_1x_2^2 + 2x_1 - 5x_2 + 8 = 0$$

c. $3x_1 - \cos(x_2x_3) - \frac{1}{2} = 0$

$$4x_1^2 - 625x_2^2 + 2x_2 - 1 = 0$$

$$e^{-x_1x_2} + 20x_3 + \frac{10\pi - 3}{3} = 0$$

b. $\sin(4\pi x_1x_2) - 2x_2 - x_1 = 0$

$$\left(\frac{4\pi - 1}{4\pi}\right)(e^{2x_1} - e) + 4ex_2^2 - 2ex_1 = 0$$

d. $x_1^2 + x_2 - 37 = 0$

$$x_1 - x_2^2 - 5 = 0$$

$$x_1 + x_2 + x_3 - 3 = 0$$

4. 用延拓法及四階 Runge-Kutta 法和 $N = 1$ ，求解下列非線性方程組，給定 $\mathbf{x}(0) = \mathbf{0}$ 。其結果是否可和牛頓法的結果相比，或者可做為牛頓法的初始近似值？

a. $x_1(1 - x_1) + 4x_2 = 12$

$$(x_1 - 2)^2 + (2x_2 - 3)^2 = 25$$

與 10.2(5c) 做比較。

c. $15x_1 + x_2^2 - 4x_3 = 13$

$$x_1^2 + 10x_2 - x_3 = 11$$

$$x_2^3 - 25x_3 = -22$$

與 10.2(6c) 做比較。

b. $5x_1^2 - x_2^2 = 0$

$$x_2 - 0.25(\sin x_1 + \cos x_2) = 0$$

與 10.2(5d) 做比較。

d. $10x_1 - 2x_2^2 + x_2 - 2x_3 - 5 = 0$

$$8x_2^2 + 4x_3^2 - 9 = 0$$

$$8x_2x_3 + 4 = 0$$

與 10.2(6d) 做比較。

5. 重複習題 4，但使用以下來源之初始近似值。

a. 由 10.2(3c)

b. 由 10.2(3d)

c. 由 10.2(4c)

d. 由 10.2(4d)

6. 將延拓法及四階 Runge-Kutta 法和 $N=1$ 用於 10.2 節的習題 7。如此獲得的結果是否和原來的一樣好？

7. 用 $N=2$ 重複習題 5。

8. 用延拓法及四階 Runge-Kutta 法和 $N=1$ 重複 10.2 節的習題 8。

9. 用延拓法及四階 Runge-Kutta 法和 $N=2$ 重複 10.2 節的習題 9。

10. 證明，延拓法及歐拉法和 $N=1$ ，所得的結果與牛頓法第一次迭代的結果相同；也就是，使用 $\mathbf{x}(0) = \mathbf{x}^{(0)}$ 一定會得到 $\mathbf{x}(1) = \mathbf{x}^{(1)}$ 。

11. 證明，延拓法中使用同倫函數 $G(\lambda, \mathbf{x}) = F(\mathbf{x}) - e^{-\lambda}F(\mathbf{x}(0))$ 配合歐拉法及 $h=1$ ，等同於任何 $\mathbf{x}^{(0)}$ 時的牛頓法；也就是，使用 $\mathbf{x}(0) = \mathbf{x}^{(0)}$ 一定會得到 $\mathbf{x}(1) = \mathbf{x}^{(1)}$ 。

12. 將延拓法配合四階 Runge-Kutta 法簡稱為 CMRK4。在完成了習題 4、5、6、7、8 及 9 之後，試回答以下問題。

a. CMRK4 配合 $N=1$ 是否可與牛頓法相比？用之前習題的結果佐證你的答案。

b. 是否應該用 CMRK4 配合 $N=1$ 來求牛頓法的初始近似值？用之前習題的結果佐證你的答案。

c. 用 CMRK4 配合 $N=2$ 重複 (a) 小題。

d. 用 CMRK4 配合 $N=2$ 重複 (b) 小題。



10.6 檢視方法與軟體

在本章中，我們考慮了求非線性方程組

$$f_1(x_1, x_2, \dots, x_n) = 0$$

$$f_2(x_1, x_2, \dots, x_n) = 0$$

$$\vdots$$

$$f_n(x_1, x_2, \dots, x_n) = 0$$

近似解的方法。用於方程組的牛頓法，須要一個很好的初始近似值 $(x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})^t$ ，並產生一個向量序列

$$\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)} - J(\mathbf{x}^{(k-1)})^{-1} \mathbf{F}(\mathbf{x}^{(k-1)})$$

可以快速的收斂到解 $\mathbf{x}$ ，但 $\mathbf{x}^{(0)}$ 必須夠接近 $\mathbf{p}$ 。不過牛頓法的每一次迭代須要執行 n^2 次的偏導數計算或近似，並求解一個 $n \times n$ 的線性方程組，這須要 $O(n^3)$ 的運算次數。

Broyden 法則可以減少每次迭代的計算量，但收斂速度的減緩不會太嚴重。這種方法用 A_{k-1} 取代 Jacobian 矩陣 J ，而每次迭代時可以直接獲得 A_{k-1} 的逆矩陣。這樣就使得所須的算術運算量由 $O(n^3)$ 減到 $O(n^2)$ 。此外，每次迭代只有在計算 f_i 時須要做純量函數求值，所以每一步省了 n^2 次純量函數求值的計算。Broyden 法同樣須要良好的初始近似值。

我們介紹了最陡下降法，它可以為牛頓法或 Broyden 法獲得良好的初始近似值。雖然最陡下降法的收斂速度不夠快，但是它不須要很好的初始近似值。最陡下降法要求出多變數函數 g 的近似最小值。在本章中我們選擇

$$g(x_1, x_2, \dots, x_n) = \sum_{i=1}^n [f_i(x_1, x_2, \dots, x_n)]^2$$

函數 g 的最小值為 0，它出現於函數 f_i 同時為 0 的時候。

同倫函數與延拓法同樣也可用於解非線性方程組，且為目前學界的研究課題 (見 [AG])。在這種方法中，一個問題

$$\mathbf{F}(\mathbf{x}) = \mathbf{0}$$

經由一個位於 $[0, 1]$ 之間的參數 λ ，植入一個單參數問題族中。原來的問題相當於 $\lambda = 1$ ，而一個已知解的問題相當於 $\lambda = 0$ 。例如，在問題組

$$G(\lambda, \mathbf{x}) = \lambda \mathbf{F}(\mathbf{x}) + (1 - \lambda)(\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{x}_0)) = \mathbf{0}, \quad 0 \leq \lambda \leq 1$$

中，固定的 $\mathbf{x}_0 \in \mathbb{R}^n$ 可構成一個同倫函數。當 $\lambda = 0$ ，其解為 $\mathbf{x}(\lambda = 0) = \mathbf{x}_0$ 。原問題的解則相當於 $\mathbf{x}(\lambda = 1)$ 。延拓法則是解相對於 $\lambda_0 = 0 < \lambda_1 < \lambda_2 < \dots < \lambda_m = 1$ 的一序列問題，以求得 $\mathbf{x}(\lambda = 1)$ 。而方程式

$$\lambda_i \mathbf{F}(\mathbf{x}) + (1 - \lambda_i)(\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{x}_0)) = \mathbf{0}$$

的初始值，則是方程式

$$\lambda_{i-1} \mathbf{F}(\mathbf{x}) + (1 - \lambda_{i-1})(\mathbf{F}(\mathbf{x}) - \mathbf{F}(\mathbf{x}_0)) = \mathbf{0}$$

的解。在 netlib 中的 Hompack 程式包可以用多種同倫函數法求解非線性方程組。

在 IMSL 及 NAG 程式庫中有關非線性方程組的副程式，都使用 Levenberg-Marquardt 法，這是一種牛頓法與最陡下降法的加權平均。在一開始，權重偏向最陡下降法，當偵測到開始收斂後，權重移向收斂較快的牛頓法。每個副程式都用有限差分法近似 Jacobian 矩陣，或由使用者提供求 Jacobian 的副程式。

非線性方程組解法的完整討論可見於 Ortega 和 Rheinbolt [OR] 以及 Dennis 和 Schnabel [DenS]。最近對迭代法的研究可見於 Argyros 和 Szidarovszky [AS]，而延拓法之使用可參見 Allgower 和 Georg [AG]。