CS5321 Numerical Optimization Homework 1

Due Oct 28

1. (30%) For a single variable unimodal function $f \in [0, 1]$, we want to find its minimum. We have introduced the binary search algorithm in the class. But in each iteration, we need two function evaluations, $f(x_k)$ and $f(x_k + \epsilon)$. Here is another type of algorithms, called ternary search. Figure 1 illustrates the idea. The initial triplet of x values is $\{x_1, x_2, x_3\}$.

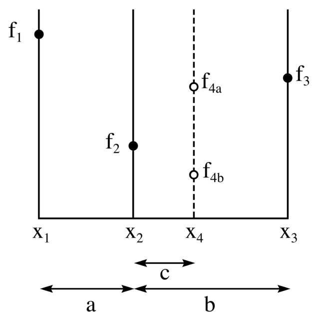

Figure 1: The idea of ternary search.

   (a) (10%) For the search direction, show that to find the minimum point, if $f(x_4) = f_{4a}$, the triplet $\{x_1, x_2, x_4\}$ is chosen for the next iteration. If $f(x_4) = f_{4b}$, the triplet $\{x_2, x_4, x_3\}$ is chosen. (Hint: use the property of unimodal.)

   因為是 unimodel，$\exists x \in [x_1, x_3]$，也就是 x_1 到 x_3 中一定有極值，表示只要 x 往極值移動時，一定為單調遞增或單調遞減。

   反證法：

   若 $f(x_4) = f_{4a}$，假設此函數在 $\exists x_m \in [x_4, x_3]$ 間有極值，但這樣就與 unimodel 定義不合，因為無法單調遞增或單調遞減，不是唯一的 local minimum。

   若 $f(x_4) = f_{4b}$，假設此函數在 $\exists x_m \in [x_1, x_2]$ 間有極值，但這樣就與 unimodel 定義不合，因為無法單調遞增或單調遞減，不是唯一的 local minimum。

   (b) (10%) For either case, we want these three points keep the same ratio, which means

$$\frac{a}{b} = \frac{c}{a} = \frac{c}{b - c}.$$

Show that under this condition, the ratio of $b/a = (\sqrt{5}+1)/2$, which is the golden ratio ϕ. (So this algorithm is called the *Golden-section search*).

假設 $a=b=c=0$ 不成立的情況下，由題目式子得出

$$\frac{a}{b}=\frac{c}{a} \longrightarrow c=\frac{a^2}{b}$$

$$\frac{c}{a}=\frac{c}{b-c} \longrightarrow a=b-c$$

將第一列式子代入第二列可得

$$\frac{b}{a}=\frac{b^2}{a^2}-1 \longrightarrow (\frac{b}{a})^2-\frac{b}{a}-1=0$$

$$\frac{b}{a}=\frac{1\pm\sqrt{5}}{2}$$

又因為 a 和 b 為長度故相除一定大於 0

$$\frac{b}{a}=\frac{1+\sqrt{5}}{2}$$

(c) (10%) If we let each iteration of the algorithm has two function evaluations, show the convergence rate of the Golden-section search is ϕ^{-2}. (This means it is faster than the binary search algorithm under the same number of function evaluations.)

因為 binary search 需要目標值與中間值進行比較，才能進行收斂，但 ternary search 每次收斂都只需要一個點（ex：這題的 x_4）, 這邊將 ternary search 每次也都代入兩個點

$$\frac{b}{a+b}=\frac{\frac{b}{a}}{1+\frac{b}{a}}$$

將 $\frac{b}{a}=\frac{\sqrt{5}+1}{2}$ 代入後可得

$$\frac{b}{a+b}=\frac{\sqrt{5}-1}{2}$$

因此可得知每代一次點可以收斂 $\phi^{-1}(\phi=\frac{b}{a})$，所以收斂兩次等於做平方，故得 ϕ^{-2}

2. (15%) Show that Newton's method for single variables is equivalent to build a quadratic model

$$q(x)=f(x_k)+f'(x_k)(x-x_k)+\frac{f''(x_k)}{2}(x-x_k)^2$$

at the point x_k and use the minimum point of $q(x)$ as the next point. (Hint: to show the next point $x_{k+1}=x_k-f'(x_k)/f''(x_k)$)

將其整形成 quadratic model 的一般標準式

$$q(x)=\frac{f''(x_k)}{2}x^2+[f'(x_k)-f''(x_k)x_k]x+[f(x_k)-f'(x_k)x_k+\frac{f''(x_k)}{2}{x_k}^2]$$

利用配方法，K 為常數，不需完整寫出

$$q(x) = \frac{f''(x_k)}{2}[x + (\frac{f'(x_k)}{f''(x_k)} - x_k)]^2 + K$$

$$x = x_k - \frac{f'(x_k)}{f''(x_k)}$$

可以得到最小值 $q(x)$，因為平方後必為正數，所以設其為 0 必定為最小值

Newton's method 定義：

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

所以從兩者結論可得出，Newton's method 對於單一變數而言，等同於建立一個 quadratic model

3. (15%) Matrix A is an $n \times n$ symmetric matrix. Show that all A's eigenvalues are positive if and only if A is positive definite.

當 A 為實對稱矩陣，A 可以正交對角化，所以必定存在一個正交矩陣 Q 使得 $Q^TAQ = D$，D 為對角矩陣，對角線上的值為 A 的特徵值，又 Q 為正交矩陣，所以其 $Q^{-1} = Q^T$

又

$$A = QDQ^T$$

左右同乘 x^Tx，可得

$$x^TAx = x^TQDQ^Tx$$

令

$$y = Q^Tx$$

$$y^T = x^TQ$$

可得

$$x^TAx = y^TDy = \lambda_1 y_1 + \lambda_2 y_2 + \lambda_3 y_3 + ... + \lambda_n y_n$$

假設所有 λ_i 皆為正數，且 x 不會等於零向量，所以 y 也不會是零向量，上式即可得出 $x^TAx > 0$，符合正定矩陣的定義，故得證其為正定矩陣。

反之，

$$Ax = \lambda x$$

左右同乘 x^T

$$x^TAx = \lambda x^Tx$$

當左邊的 A 為正定，等式左邊皆為正值（建立在特徵向量不等於零時），又正定的性質 $x^TAx > 0$，所以等式右側也一定大於零，且 x^Tx 可以看成 x 的 norm，所以其特徵值必定為正。

4. (50%) Consider a function $f(x_1, x_2) = (x_1 - x_2)^3 + 2(x_1 - 1)^2$.

(a) Suppose $\vec{x}_0 = (1, 2)$. Compute $\vec{x_1}$ using the steepest descent step with the optimal step length.

$$\frac{\partial f(x_1, x_2)}{\partial x_1} = 3(x_1 - x_2)^2 + 4(x_1 - 1)$$

$$\frac{\partial f(x_1, x_2)}{\partial x_2} = -3(x_1 - x_2)^2$$

$$\nabla f = \begin{bmatrix} 3(x_1 - x_2)^2 + 4(x_1 - 1) \\ -3(x_1 - x_2)^2 \end{bmatrix}$$

$$\frac{\partial^2 f(x_1, x_2)}{\partial x_1^2} = 6(x_1 - x_2) + 4$$

$$\frac{\partial^2 f(x_1, x_2)}{\partial x_2^2} = 6(x_1 - x_2)$$

$$\frac{\partial^2 f(x_1, x_2)}{\partial x_1 \partial x_2} = -6(x_1 - x_2)$$

$$H = \begin{bmatrix} 6(x_1 - x_2) + 4 & -6(x_1 - x_2) \\ -6(x_1 - x_2) & 6(x_1 - x_2) \end{bmatrix}$$

$$\vec{g_0} = \nabla f(1, 2) = \begin{bmatrix} 3 \\ -3 \end{bmatrix}$$

$$\vec{p_0} = -\nabla f(1, 2) = \begin{bmatrix} -3 \\ 3 \end{bmatrix}$$

$$H(1, 2) = \begin{bmatrix} -2 & 6 \\ 6 & -6 \end{bmatrix}$$

$$\alpha = \frac{-\vec{g_0}^T \vec{p_0}}{\vec{p_0}^T H \vec{p_0}} = \frac{-\begin{bmatrix} 3 & -3 \end{bmatrix} \begin{bmatrix} -3 \\ 3 \end{bmatrix}}{\begin{bmatrix} -3 & 3 \end{bmatrix} \begin{bmatrix} -2 & 6 \\ 6 & -6 \end{bmatrix} \begin{bmatrix} -3 \\ 3 \end{bmatrix}} = -\frac{1}{10}$$

$$\vec{x_1} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \frac{1}{10} \begin{pmatrix} -3 \\ 3 \end{pmatrix} = \begin{pmatrix} \frac{7}{10} \\ \frac{23}{10} \end{pmatrix}$$

(b) What is the Newton's direction of f at $(x_1, x_2) = (1, 2)$? Is it a descent direction?

$$\vec{p_k} = -H_k^{-1} \vec{g_k} = -\begin{bmatrix} -2 & 6 \\ 6 & -6 \end{bmatrix} \begin{bmatrix} 3 \\ -3 \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{1}{2} \end{bmatrix}$$

H^{-1} is not positive definite because it has a negative eigenvalue, so $\vec{p_k}$ is not a descent direction.

(c) Compute the LDL decomposition of the Hessian of f at $(x_1, x_2) = (1, 2)$. (No pivoting)

$$H = \begin{bmatrix} -2 & 6 \\ 6 & -6 \end{bmatrix} \xrightarrow{r_{12}(3)} \begin{bmatrix} -2 & 6 \\ 0 & 12 \end{bmatrix} \xrightarrow{c_{12}(3)} \begin{bmatrix} -2 & 0 \\ 0 & 12 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} -2 & 0 \\ 0 & 12 \end{bmatrix} \begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix} = LDL^T$$

(d) Compute the modified Newton step using LDL modification.

$$\hat{H} = L\hat{D}L^T = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 12 \end{bmatrix} \begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -6 \\ -6 & 30 \end{bmatrix}$$

$$\vec{p} = -\hat{H}^{-1}\vec{g} = -\begin{bmatrix} \frac{5}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{12} \end{bmatrix} \begin{bmatrix} 3 \\ -3 \end{bmatrix} = \begin{bmatrix} -3 \\ -\frac{1}{2} \end{bmatrix}$$

$$\alpha = \frac{-\vec{g}^T\vec{p}}{\vec{p}^T\hat{H}\vec{p}} = \frac{-\begin{bmatrix} 3 & -3 \end{bmatrix} \begin{bmatrix} -3 \\ -\frac{1}{2} \end{bmatrix}}{\begin{bmatrix} -3 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 2 & -6 \\ -6 & 30 \end{bmatrix} \begin{bmatrix} -3 \\ -\frac{1}{2} \end{bmatrix}} = 1$$

(e) Suppose $\vec{x}_0 = (1,1)$ and $\vec{x}_1 = (1,2)$, and the $B_0 = I$. Compute the quasi Newton direction p_1 using BFGS.

$$\vec{s_0} = \vec{x_1} - \vec{x_0} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$B_0 = I$$

$$\vec{y_0} = \nabla f(1,2) - \nabla f(1,1) = \begin{bmatrix} 3 \\ -3 \end{bmatrix}$$

$$B_1 = B_0 - \frac{B_0\vec{s_0}\vec{s_0}^T B_0}{\vec{s_0}^T B_0 \vec{s_0}} + \frac{\vec{y_0}\vec{y_0}^T}{\vec{y_0}^T\vec{s_0}}$$

$$B_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{\begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix}}{\begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}} + \frac{\begin{bmatrix} 3 \\ -3 \end{bmatrix} \begin{bmatrix} 3 & -3 \end{bmatrix}}{\begin{bmatrix} 3 & -3 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}} = \begin{bmatrix} -2 & 3 \\ 3 & -3 \end{bmatrix}$$

$$\vec{p_1} = -B_1^{-1}\vec{g_1} = -B_1^{-1} \cdot \nabla f(\vec{x_1}) = -\begin{bmatrix} 1 & 1 \\ 1 & \frac{2}{3} \end{bmatrix} \begin{bmatrix} 3 \\ -3 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$