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 $\{x_1, x_2, x_3\}$.

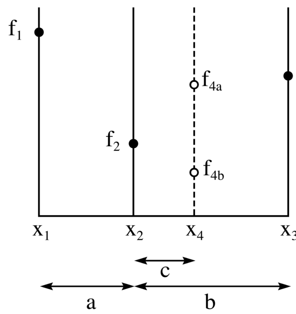

Figure 1: The idea of ternary search.

Answers are put here.
也可以使用中文回答

(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.)

由unimodal可知極值在$[x_1, x_3]$之間，且由任何點往極點靠近都會是單調遞增(若極值為最大值)或單調遞減(若極值為最小值)，而我們在$[x_1, x_3]$之間取兩個端點x_2和x_4，將區間分成3份。

這題要求函數最小值，因此我們可以去判斷x_2和x_4的函數值，它們的函數值和距離極值點有關係的。距離極值點越近，函數值越小(也可能越大視函數而定)。若$f(x_4) = f_{4a}$，可以判斷出$f_{4a} > f_2$，x_2離極點較近，而我們需要縮小區間範圍，因此我們拋棄$[x_4, x_3]$區間，故下一次的iteration會選擇triplet $\{x_1, x_2, x_4\}$

若$f(x_4) = f_{4b}$，可以判斷出$f_{4b} < f_2$，x_4離極點較近，而我們需要縮小區間範圍，因此我們拋棄$[x_1, x_2]$區間，故下一次的iteration會選擇triplet $\{x_2, x_4, x_3\}$

三分搜索主要方法就是，每次通過比較兩個值的大小，縮小三分之一的區間，直到最後區間範圍小於我們設定的閾值為止。

(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 $\neq$ 0或b $\neq$ 0或c $\neq$ 0下，由左邊兩個等式中可以推導出

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

$$\rightarrow c = \frac{a^2}{b} \tag{1}$$

由右邊兩個等式中可以整理推導出

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

$$\rightarrow a = b - c \tag{2}$$

公式(1)帶入公式(2)替換c並且同時乘$\frac{b}{a^2}$可以得到

$$\frac{b}{a} = \frac{b^2}{a^2} - 1$$

$$\rightarrow (\frac{b}{a})^2 - \frac{b}{a} - 1 = 0 \tag{3}$$

公式(3)使用公式解解出兩個值

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

另外因為a和b是長度，所以a和b≥0，因此$\frac{b}{a} \geq 0$ ，所以不可能是負值，最後得到

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

得證。

(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.)

binary search每次都要帶2個點才能進行1次收斂，但ternary search每次收斂都只需要1個點，所以假設ternary search在每次也都帶2個點，等於收斂兩次。

$$\frac{b}{a+b} = \frac{\frac{b}{a}}{1 + \frac{b}{a}} \tag{4}$$

而原先每次收斂的長度都如式(4)，將$b/a = (\sqrt{5}+1)/2$值帶入後化整理得到

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

因此可知每帶一次點可以收斂ϕ^{-1}，所以帶二次點可以收斂兩次ϕ^{-2}，得證。

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

將題目中的多項式以降冪重新排列後

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

利用配方法求極值，公式(5)中C為某一常數使等式成立

$$\to q(x) = \frac{f''(x)}{2}(x - [x_k - \frac{f'(x_k)}{f''(x_k)}])^2 + C \tag{5}$$

$$\arg\min_x \ \mathrm{q}(x) = x_k - \frac{f'(x_k)}{f''(x_k)} \tag{6}$$

根據Newton'method

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)} = \arg\min_x \ \mathrm{q}(x)$$

得證

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為實對稱矩陣時，A是可以正交對角化的，所以存在一個正交矩陣Q讓$Q^TAQ = D$，D為對角矩陣，其中λ_i為A的eigenvalues，其值都為正，$\lambda_i > 0$

$$D = \begin{bmatrix} \lambda_1 & 0 & \dots & \dots & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & \dots & 0 \\ 0 & 0 & \lambda_3 & 0 & \dots & 0 \\ \vdots & \dots & \dots & \ddots & \dots & \vdots \\ \vdots & \dots & \dots & \dots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & 0 & \lambda_n \end{bmatrix}_{n\times n}$$

假設x為任意nonzero vector，又$A = QDQ^T$(因$Q^{-1} = Q^T$)，等是兩邊同時左乘x^T，而右乘x，可以得到

$$x^TAx = x^TQDQ^Tx \tag{7}$$

另$y = Q^T x$替換公式(7)可以得到

$$x^T A x = y^T D y \tag{8}$$

而

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1}$$

把公式(8)乘開得到

$$x^T A x = y^T D y = \begin{bmatrix} y_1 & y_2 & y_3 & \dots & \dots & y_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots & \dots & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & \dots & 0 \\ 0 & 0 & \lambda_3 & 0 & \dots & 0 \\ \vdots & \dots & \dots & \ddots & \dots & \vdots \\ \vdots & \dots & \dots & \dots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & 0 & \lambda_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ \vdots \\ y_n \end{bmatrix}$$
$$= \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 \tag{9}$$

假設λ_i都是正的eigenvalues，又因為x為任意的nonzero vecoter且Q為可逆矩陣，所以$y = Q^T x$是nonzero vector，因此$x^T A x > 0$，滿足正定矩陣定義，因此A為正定矩陣。

A為實對稱矩陣，λ為A的eigenvalue，x為其相應的eigenvector

$$Ax = \lambda x \tag{10}$$

同時左乘x^T

$$\begin{aligned} x^T A x &= \lambda x^T x \\ &= \lambda ||x||^2 \end{aligned} \tag{11}$$

當A為positive definite且x為nonzero vector的eignevector，$x^T A x > 0$，又$||x||^2$為x的長度平方必為正，故其eignevalue λ皆必為正

A's eigenvalues are positive $\iff$ A is positive definite

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.

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

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

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

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

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

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

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

$\Rightarrow H(x_1,x_2) = \begin{bmatrix} 3(x_1-x_2)^2+4(x_1-1) & -6(x_1-x_2) \\ -6(x_1-x_2) & -3(x_1-x_2)^2 \end{bmatrix}$

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

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

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

$\alpha = \frac{-g^T \vec{p_0}}{\vec{p_0}^T H \vec{p_0}} = \frac{18}{-180} = -\frac{1}{10}$

$\vec{x_1} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} - \frac{1}{10} \begin{bmatrix} -3 \\ 3 \end{bmatrix} = \begin{bmatrix} 1.3 \\ 1.7 \end{bmatrix}$

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

$\vec{p_k} = -H_k^{-1} \, \vec{g_k} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{12} \end{bmatrix} \begin{bmatrix} -3 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{1}{2} \end{bmatrix}$

由於H^{-1}並不是正定義矩陣，因為它的特徵值中有負值，所以$\vec{p_k}$不是descent dirction。

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

H(1,2)$= \begin{bmatrix} -2 & 6 \\ 6 & -6 \end{bmatrix} \xrightarrow{r_{12}(3)} \begin{bmatrix} -2 & 6 \\ 0 & 12 \end{bmatrix} \xrightarrow{c_{12}(3)} \begin{bmatrix} -2 & 0 \\ 0 & 12 \end{bmatrix}$

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

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

$\hat{D}$取代D，把-2替換成1 $\begin{bmatrix} -2 & 0 \\ 0 & 12 \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 \\ 0 & 12 \end{bmatrix}$

$\hat{H}_0 = L\hat{D}L^T = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 12 \end{bmatrix} \begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix}$

$\hat{H_0}^{-1} = (L^T)^{-1} \hat{D}^{-1} (L)^{-1} = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{12} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}$

$\vec{p_0} = - \hat{H_0}^{-1} \vec{g_0} = - \begin{bmatrix} \frac{7}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{12} \end{bmatrix} \begin{bmatrix} 3 \\ -3 \end{bmatrix} = \begin{bmatrix} -\frac{9}{2} \\ -\frac{1}{2} \end{bmatrix}$

(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

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

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

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

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