CS5321 Numerical Optimization Homework 1

Annabella Putri Dirgo - 109006238

October 25, 2022

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

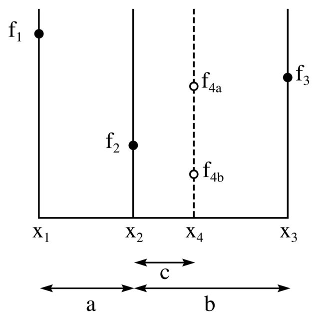

Figure 1: The idea of ternary search.

   (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.)

   (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*).

   (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.)

   (a) Uni-modal has a unique property, in this case is having only one global minimum. Since there are two conditions here, we are going to go step-by-step on how Ternary Search finds the minimum point.

i. When $f(x4) = f4a$, then minimum point is located at $f2$. Thus, the new triplet is $x1, x2, x4$ in Figure 2. We do so in order to preserve the Uni-modal property, and we can ignore the region between $x4, x3$.

ii. When $f(x4) = f4b$, then minimum point is located at f4b. Thus, the new triplet is $x2, x4, x3$ in Figure 3. We do so in order to preserve Uni-modal property, and we can ignore the region between $x1, x2$.

(b) Mathematically, to ensure that the spacing after evaluating $f(x4)$ is proportional to the spacing prior to that evaluation, if $f(x4)$ is $f4a$ and our new triplet of points is $x1, x2, x4$, then we want

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

However, if $f(x4) is f4b$ and our new triplet of points is $x2, x4, x3$, then we want

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

Eliminating c from these two simultaneous equations and substitute each other we get

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

$$b^2 - ab - a^2 = 0$$

Thus, we got

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

where ϕ is the golden ratio:

$$\phi = \frac{1+\sqrt{5}}{2} = 1.618033988...$$

(c) Golden ratio minimal polynomial conjugate root is

$$-\frac{1}{\phi} = 1 - \phi = \frac{1-\sqrt{5}}{2} = -0.61803$$

and we want only its absolute value, thus we have the golden ratio conjugate or silver ratio

$$\Phi = \frac{1}{\phi} = \phi^{-1} = 0.61803$$

By having two function evaluations, the convergence rate of the Golden section search is

$$\phi^{-2} = \frac{1}{\phi^{-2}} = \frac{1}{0.618033988^2} \approx 0.3819660116... < 0.5$$

The convergence rate of the Golden-section search is faster than the convergence binary search, which is only $\frac{1}{2} = 0.5$

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

Let,

$$x_{k+1} = xk + (x - x_k)$$

By substituting $(x - x_k)$ with t, then we have

$$x_{k+1} = x_k + t$$

$$0 = \frac{d}{dt}(f(x_k) + f'(x_k)t + \frac{1}{2}f''(x_k)t^2)$$

Then we have,

$$0 = f'(x_k) + f''(x_k)t$$

$$-f'(x_k) = f''(x_k)$$

Thus,

$$t = -\frac{f'(x_k)}{f''(x_k)}$$

Then we feed t to $x_{k+1} = x_k + t$, therefore we have

$$x_{k+1} = x_k + t$$

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

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.

If A is symmetric and has positive eigenvalues, then, by the spectral theorem for symmetric matrices, there is an orthogonal matrix Q such that $A = Q^\top \Lambda Q$, with $\Lambda = diag(\lambda_1, ..., \lambda_n)$. if x is any non zero vector, then $y := Qx \neq 0$ and

$$x^\top Ax = x^\top \left(Q^\top \Lambda Q\right) x = \left(x^\top Q^\top\right) \Lambda(Qx) = y^\top \Lambda y = \sum_{i=1}^{n} \lambda_i y_i^2 > 0$$

because y is nonzero and A has a positive eigenvalues.

Conversely, suppose that A is positive definite and that $Ax = \lambda x$, with $x \neq 0$. WLOG, we may assume that $x \top x = 1$. Thus,

$$0 < x \top Ax = x \top (\lambda x) = \lambda x \top x = \lambda$$

as desired.

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.

The formula for the Steepest Descent step for $\vec{x_1}$ is

$$\vec{x_1} = \vec{x_0} + \alpha_k \vec{p_k}$$

First we need to find α_k

$$\begin{aligned}\alpha_k &= \frac{-\vec{g}_k^T \vec{p}_k}{\vec{p}_k^T H \vec{p}_k} \\ &= \frac{\begin{pmatrix} -3 & 3 \end{pmatrix}\begin{pmatrix} -3 \\ 3 \end{pmatrix}}{\begin{pmatrix} -3 \\ 3 \end{pmatrix}\begin{pmatrix} -2 & 6 \\ 6 & -6 \end{pmatrix}\begin{pmatrix} -3 & 3 \end{pmatrix}} \\ &= \frac{1}{-20} = -0.05\end{aligned}$$

Then, we need to find $\vec{p_k}$,

$$\begin{aligned}\vec{p}_k(1,2) &= -\nabla f\left(x_0\right) \\ &= -\begin{pmatrix} \frac{\partial f}{\partial x_1}(1,2) \\ \frac{\partial f}{\partial x_2}(1,2) \end{pmatrix} \\ &= -\begin{pmatrix} 3(1-2)^2 + 4(1-1) = 3i \\ -3(1-2)^2 = -3j \end{pmatrix} \\ &= \begin{pmatrix} -3 \\ 3 \end{pmatrix}\end{aligned}$$

Plug them in, we have

$$\vec{x_1} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} + 0.05 \begin{pmatrix} -3 \\ 3 \end{pmatrix} = \begin{pmatrix} 0.85 \\ 2.15 \end{pmatrix}$$

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

i. To compute Newton's Direction, we need

$$\vec{p_k} = -H_k^- 1 \vec{g_k}$$

First we need to compute the Hessian matrix,

$$\begin{aligned}\nabla^2 f(x) &= H(x) \\ &= \begin{pmatrix} \frac{\partial f}{\partial x_1^2} & \frac{\partial f}{\partial x_1 \partial x_2} \\ \frac{\partial f}{\partial x_2 \partial x_1} & \frac{\partial f}{\partial x_2^2} \end{pmatrix} \\ &= \begin{pmatrix} -2 & 6 \\ 6 & -6 \end{pmatrix}\end{aligned}$$

Then, compute the inverse of the Hessian matrix

$$H_k^- 1 = \frac{1}{-24}\begin{pmatrix} -6 & -6 \\ -6 & -2 \end{pmatrix} = \begin{pmatrix} 1/4 & 1/4 \\ 1/4 & 1/12 \end{pmatrix}$$

Finally, we can compute the Newton's Direction

$$\begin{aligned}\vec{p_k} &= -H_k^{-1}\vec{g_k} \\ &= \begin{pmatrix} -1/4 & -1/4 \\ -1/4 & -1/12 \end{pmatrix}\begin{pmatrix} 3 \\ -3 \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ -2/4 \end{pmatrix}\end{aligned}$$

ii. To see whether f is a descent direction at $(x_1, x_2) = (1, 2)$, we need to show that

$$\vec{g_k^T} \times \vec{p_k} < 0$$

$$\begin{pmatrix} 3 & -3 \end{pmatrix} \times \begin{pmatrix} 0 \\ -2/4 \end{pmatrix}$$

In the end we have,

$$\begin{pmatrix} 0 & 3/2 \end{pmatrix}$$

As we can see, $\begin{pmatrix} 0 & 3/2 \end{pmatrix}$ is not less than $\begin{pmatrix} 0 & 0 \end{pmatrix}$. Then we can conclude that f at $(x_1, x_2) = (1, 2)$ is not a descent direction.

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

In order to get the LDL decomposition, we need to find the value of x, d_1, d_2, y.

$$\begin{bmatrix} -2 & 6 \\ 6 & -6 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ x & 1 \end{bmatrix}\begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix}\begin{bmatrix} 1 & y \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} -2 & 6 \\ 6 & -6 \end{bmatrix} = \begin{bmatrix} d_1 & d_1 \times y \\ d_1 \times x & (x \times y \times d_1) + d_2 \end{bmatrix}$$

Solving for the equation, we have

$$\begin{bmatrix} -2 & 6 \\ 6 & -6 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}\begin{bmatrix} -2 & 0 \\ 0 & 12 \end{bmatrix}\begin{bmatrix} 1 & -3 \\ 0 & 1 \end{bmatrix}$$

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

we need to compute $\vec{p} = -L^{-T}D^{-1}L^{-1}g$. Thus,

$$\vec{p} = -\begin{bmatrix} 1 & 3 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} -1/2 & 0 \\ 0 & 1/12 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} 3 \\ -3 \end{bmatrix}$$

In the end we have,

$$\vec{p} = \begin{bmatrix} 0 \\ -6 \end{bmatrix}$$

(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.

For BFGS, We need to compute

$$\vec{p}_1 = -B_1^{-1}\vec{g}_1$$

First we need to compute B_1

$$
\begin{aligned}
B_1 &= I - \frac{\vec{s}_0 \vec{s}_0^T}{\vec{s}_0^T \vec{s}_0} + \frac{\vec{y}_0 \vec{y}_0^T}{\vec{y}_0^T \vec{y}_0} \\
&= \left[\begin{array}{cc} 0.5 & 0.5 \\ 0.5 & -0.5 \end{array} \right]
\end{aligned}
$$

Plugging in B_1 back to the initial formula, we then get

$$
\vec{p}_1 = \left[\begin{array}{c} 0 \\ -9 \end{array} \right]
$$