Optimization

Fourth Class

2020 - 2021

By

Dr. Jawad Mahmoud Jassim

Dept. of Math.

Education College for Pure Sciences

University of Basrah

Iraq

Chapter Five

Quasi – Newton Methods

Lecture 3

Theorem (3):

If

1: X_0 **in** R^n **is arbitrary.**

2: H_0 **is an arbitrary** $n \times n$ **symmetric positive definite matrix.**

3: $f\colon R^n \to R\ is\ defined\ by\ f(X) = \frac{1}{2} X^T A X + b^T X + c, where\ A\ is\ an$
$n \times n\ symmetric\ positive\ definite\ matrix$, b **is an** $n \times 1$ **vector** $and\ c \quad is\ a\ real\ number$.

4: $X_1, X_2, \cdots, X_k\ and\ H_1, H_2, \cdots, H_k, 1 \le k \le n$ **are generated from** $Algorithm\ (2)$.

Then

a: $s_i^T A s_j = 0, i \neq j, (i, j = 0, 1, \cdots, k)$.

b: $H_k A s_i = s_i\ , (i = 0, 1, \cdots, k-1)$.

<u>Notes (2):</u>

1: Theorem (3) shows that the DFP method is a conjugate direction method.

2: From Theorem $(3 - b)$, with $k = n$, we see that $H_n A$ has n eigenvectors $s_i, (i = 0, 1, \cdots, n - 1)$ each of which corresponds to the eigenvalue unity.

3: From Theorem $(3 - a)$, $we\ see\ that$ $s_i, (i = 0, 1, \cdots, n - 1)$, being $A -\ conjugate$, are linearly independent.

<u>Note (3):</u>

From Theorem (3) and Lemma (1), we conclude that $H_n A = I$, so that $H_n = A^{-1}$.

Example:

Given the objective $function\ f(X) = x_1^2 + 2x_2^2$ **. Compute the first two iterations for DFP method.** $Use\ \ X_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

Solution:

First, we find the gradient vector $g(X)$ **as:**

$$g(X) = [\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}]^T = [2x_1\ , 4x_2]^T.$$

$$\therefore g_0 = g(X_0) = .[2, 4]^T$$

$$\therefore\ p_0 = -H_0 g_0 = -I g_0 = -g_0 = [-2, -4]^T.$$

$$\therefore\ X_1 = X_0 + \alpha p_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \alpha \begin{bmatrix} -2 \\ -4 \end{bmatrix} = \begin{bmatrix} 1 - 2\alpha \\ 1 - 4\alpha \end{bmatrix}.$$

$$f(X_1) = (1 - 2\alpha)^2 + 2(1 - 4\alpha)^2.$$

$$\frac{df}{d\alpha} = 2(1 - 2\alpha)(-2) + 4(1 - 4\alpha)(-4) = 72\alpha - 20.$$

$$\text{Set } \frac{df}{d\alpha} = 0 \rightarrow \alpha = \frac{20}{72} = 0.2778.$$

$$\therefore X_1 = \begin{bmatrix} 0.4444 \\ -0.1112 \end{bmatrix}.$$

$$\therefore g_1 = g(X_1) = \begin{bmatrix} 0,8888 \\ -0.4448 \end{bmatrix}.$$

Now, we compute s_0 , y_0 *and* z_0 as:

$$s_0 = \alpha_0 p_0 = \begin{bmatrix} -1.5556 \\ -1.1111 \end{bmatrix}, \quad y_0 = g_1 - g_0 = \begin{bmatrix} -1.1112 \\ -4.4448 \end{bmatrix}, \quad z_0 = H_0 y_0 = y_0.$$

$$\therefore z_0 = \begin{bmatrix} -1.1112 \\ -4.4448 \end{bmatrix}.$$

***To find the matrix H_1*, we must compute the following:**

$$s_0 s_0^T = \begin{bmatrix} -1.5556 \\ -1.1111 \end{bmatrix} \begin{bmatrix} -1.5556 & -1.1111 \end{bmatrix} = \begin{bmatrix} 2.4199 & 2.6667 \\ 1.7284 & 1.2345 \end{bmatrix}.$$

$$s_0^T y_0 = \begin{bmatrix} -1.5556 & -1.1111 \end{bmatrix} \begin{bmatrix} -1.1112 \\ -4.4448 \end{bmatrix} = 6.6672.$$

$$\frac{s_0 s_0^T}{s_0^T y_0} = \begin{bmatrix} 0.3630 & 0.4 \\ 0.2592 & 0.1852 \end{bmatrix}.$$

$$z_0 z_0^T = \begin{bmatrix} -1.1112 \\ -4.4448 \end{bmatrix} \begin{bmatrix} -1.1112 & -4.4448 \end{bmatrix} = \begin{bmatrix} 1.2348 & 4.9391 \\ 4.9391 & 19.7562 \end{bmatrix}.$$

$$z_0 z_0^T = \begin{bmatrix} -1.1112 \\ -4.4448 \end{bmatrix} \begin{bmatrix} -1.1112 & -4.4448 \end{bmatrix} = \begin{bmatrix} 1.2348 & 4.9391 \\ 4.9391 & 19.7562 \end{bmatrix}.$$

$$y_0^T z_0 = \begin{bmatrix} -1.1112 & -4.4448 \end{bmatrix} \begin{bmatrix} -1.1112 \\ -4.4448 \end{bmatrix} = 20.9910 .$$

$$\frac{z_0 z_0^T}{y_0^T z_0} = \begin{bmatrix} 0.0588 & 0.2353 \\ 0.2353 & 0.9412 \end{bmatrix}.$$

$$\therefore H_1 = H_0 + \frac{s_0 s_0^T}{s_0^T y_0} - \frac{z_0 z_0^T}{y_0^T z_0}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0.3630 & 0.4 \\ 0.2592 & 0.1852 \end{bmatrix} - \begin{bmatrix} 0.0588 & 0.2353 \\ 0.2353 & 0.9412 \end{bmatrix}$$

$$= \begin{bmatrix} 1.3042 & 0.1647 \\ 0.0239 & 0.2440 \end{bmatrix}.$$

Now, we compute p_1 as:

$$p_1 = -H_1 g_1 = -\begin{bmatrix} 1.3042 & 0.1647 \\ 0.0239 & 0.2440 \end{bmatrix} \begin{bmatrix} 0.8888 \\ -0.4448 \end{bmatrix} = \begin{bmatrix} -1.2324 \\ -0.1298 \end{bmatrix}.$$

$$\therefore X_2 = X_1 + \alpha p_1 = \begin{bmatrix} 0.4444 \\ -0.1112 \end{bmatrix} + \alpha \begin{bmatrix} -1.2324 \\ -0.1298 \end{bmatrix} = \begin{bmatrix} 0.4444 - 1.2324\alpha \\ -0.1112 - 0.1298\alpha \end{bmatrix}.$$

$$f(X_2) = (0.4444 - 1.2324\alpha)^2 + 2(-0.1112 - 0.1298\alpha)^2$$

$$\therefore \frac{df}{d\alpha} = 2(0.4444 - 1.2324\alpha)(-1.2324) + 4(-0.1112 - 0.1298\alpha)(-0.1298)$$

$$= -1.0954 + 3.0376\alpha + 0.0577 + 0.0674\alpha = -1.8963 + 3.105\alpha$$

Set $\frac{df}{d\alpha} = 0 \;\rightarrow\; -1.8963 + 3.105\alpha = 0 \;\rightarrow\; \alpha = 0.6107.$

$$\therefore X_2 = \begin{bmatrix} 0.4444 - 1.2324\alpha \\ -0.1112 - 0.1298\alpha \end{bmatrix} = \begin{bmatrix} -0.3082 \\ -0.1905 \end{bmatrix}.$$

<u>H.W.</u>

Given the objective $function$

$$f(X) = 4x_1^2 + 3x_2^2 - 5x_1x_2 - 8x_1\ .$$

Compute the first two iterations for DFP method.

$$Use\ \ X_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$