Optimization

Fourth Class

2020 - 2021

By

Dr. Jawad Mahmoud Jassim

Dept. of Math.

Education College for Pure Sciences

University of Basrah

Iraq

Chapter Four

Conjugate Direction Methods

Lecture 2

Conjugate Direction Method:

Let us apply the general descent algorithm to the objective function

$$f: R^n \to R \ defined\ by: \ f(X) = \frac{1}{2} X^T A X + b^T X + c,$$

Where A *is an* $n \times n$ symmetric positive definite matrix, b *is an* $n \times 1$ vector and c is a real number.

Now f has a gradient vector g given by:

$$g(X) = AX + b.$$

$If\ f\ has\ the\ strict\ global\ minimizer\ X^*\ in\ R^n$, then $g(X^*) = 0$.

Suppose the descent directions p_k , $k = 0, 1, 2, \cdots, n-1$, used in the general descent algorithm are $A - conjugate$ and the α_k are determined so that

$$f(X_k + \alpha_k p_k) = \min_{\alpha} f(X_k + \alpha\, p_k)\ is\ satisfied.$$

We then have the general conjugate direction algorithm.

Algorithm (1): (General Conjugate Direction Algorithm)

It is assumed that an estimate X_0 of X^* is given.

1: Compute $g_0\ from\ g_0 = g(X_0)$, where $g\ is\ the\ gradient\ vector$.

2: Compute $p_0\ such\ that\ g_0^T p_0 < 0$.

3: Set $k = 0$.

4: Compute $\alpha_k\ and\ X_{k+1}$ such that

$$f(X_k + \alpha_k p_k) = \min_{\alpha} f(X_k + \alpha\, p_k) \text{ and } X_{k+1} = X_k + \alpha_k p_k \,.$$

5: Compute $p_{k+1}\ such\ that\ \ p_{k+1}^T A p_j = 0\ , j = 0, 1, 2, \cdots , k$.

6: Set $k = k + 1$ and go to step 4.

Example:

Let $f: R^2 \to R$ be defined by $f(X) = X^T A X - b^T X$,

$$where\ A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, b = \begin{bmatrix} 3 \\ 3 \end{bmatrix}, X_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, p_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}\ and\ p_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

Verify that the iteration X_2 generated from general conjugate direction method is the minimizer of f.

Solution:

We can write $f(X)$ as follows:

$$f(X) = [x_1 , x_2] \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - [3 , 3] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= [2x_1 + x_2 , x_1 + 2x_2] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - 3x_1 - 3\, x_2$$

$$= 2x_1^2 + x_1 x_2 + x_1 x_2 + 2x_2^2 - 3x_1 - 3\, x_2$$

$$= 2x_1^2 + 2x_1 x_2 + 2x_2^2 - 3x_1 - 3\, x_2$$

Now, we compute the gradient vector $g(X)$ as follows:

$$g(X) = [\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}]^T = [4x_1 + 2x_2 - 3,\ 2x_1 + 4x_2 - 3]^T.$$

$$\therefore g_0 = g(X_0) = [-3, -3]^T.$$

Now, we compute α *and* X_1 as follows:

$$X_1 = X_0 + \alpha p_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \alpha \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \alpha \\ 0 \end{bmatrix}.$$

$$f(X_1) = 2\alpha^2 - 3\alpha$$

$$Set\ \frac{df}{d\alpha} = 0 \rightarrow 4\alpha - 3 = 0 \rightarrow \alpha = \frac{3}{4}.$$

Since $\frac{d^2 f}{d\alpha^2} = 4 > 0 \rightarrow \alpha = \frac{3}{4}$ is a minimizer of $f(X_1)$.

$$\therefore X_1 = \begin{bmatrix} \frac{3}{4} \\ 0 \end{bmatrix}.$$

$$g(X_1) = [\,3+0-3, \tfrac{3}{2}+0-3]^T = [0, -\tfrac{3}{2}]^T.$$

In the same way find X_2:

$$X_2 = X_1 + \alpha p_1 = \begin{bmatrix} \frac{3}{4} \\ 0 \end{bmatrix} + \alpha \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} \frac{3}{4} + \alpha \\ -2\alpha \end{bmatrix}.$$

$$\therefore f(X_2) = 2\left(\tfrac{3}{4}+\alpha\right)^2 - 4\left(\tfrac{3}{4}+\alpha\right)\alpha + 8\alpha^2 - 3\left(\tfrac{3}{4}+\alpha\right) + 6\alpha$$

$$= 2\left(\frac{9}{16} + \frac{3}{2}\,\alpha + \alpha^2\right) - 3\alpha - 4\alpha^2 + 8\alpha^2 - \frac{9}{4} - 3\alpha + 6\alpha$$

$$= \frac{9}{8} + 3\alpha + 2\alpha^2 - 3\alpha - 4\alpha^2 + 8\alpha^2 - \frac{9}{4} - 3\alpha + 6\alpha$$

$$= 6\alpha^2 + 3\alpha - \frac{9}{8}\ .$$

Set $\frac{df}{d\alpha} = 0 \rightarrow 12\alpha + 3 = 0 \rightarrow \alpha = -\frac{1}{4}$.

Since $\frac{d^2f}{d\alpha^2} = 12 > 0 \rightarrow \alpha = -\frac{1}{4}$ is a minimizer of $f(X_2)$.

$$\therefore X_2 = \begin{bmatrix} \frac{3}{4} + \alpha \\ -2\alpha \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}.$$

$\therefore g(X_2) = [0, 0]^T$.

$\therefore X_2$ is the minimizer of f.

Theorem (4):

If

1: $X_0 \epsilon R^n$ **is arbitrary.**

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

3: $p_k\ , k = 0, 1, 2, \cdots, n-1$ **are** $A - conjugate$.

4: $X_k\ , k = 1, 2, \cdots$ **are generated from general conjugate direction** $algorithm$.

Then $X_m = X^*\ for\ some\ m \leq n$.

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

Theorem (4) says that the minimizer X^* of f *defined by* $f(X) = \frac{1}{2}X^T AX + b^T X + c$, is attained exactly in at most n iterations by using general conjugate direction algorithm.