Introduction to Engineering Optimization (ME6806)

Dr. Yogesh Kumar
Assistant Professor
Mechanical Engineering Department
National Institute of Technology Patna
Bihar - 800 005, India
yogesh.me@nitp.ac.in

Broyden–Fletcher–Goldfarb–Shanno (BFGS) Method

As stated earlier, a major difference between the DFP and BFGS methods is that in the BFGS method, the Hessian matrix is updated iteratively rather than the inverse of the Hessian matrix. The BFGS method can be described by the following steps.

1. Start with an initial point $\mathbf{X}_1$ and a $n \times n$ positive definite symmetric matrix $[B_1]$ as an initial estimate of the inverse of the Hessian matrix of f. In the absence of additional information, $[B_1]$ is taken as the identity matrix $[I]$. Compute the gradient vector $\nabla f_1 = \nabla f(\mathbf{X}_1)$ and set the iteration number as $i = 1$.
2. Compute the gradient of the function, ∇f_i, at point $\mathbf{X}_i$, and set

$$\mathbf{S}_i = -[B_i]\nabla f_i$$

3. Find the optimal step length λ_i^* in the direction $\mathbf{S}_i$ and set

$$\mathbf{X}_{i+1} = \mathbf{X}_i + \lambda_i^* \mathbf{S}_i$$

Broyden–Fletcher–Goldfarb–Shanno (BFGS) Method

4. Test the point $\mathbf{X}_{i+1}$ for optimality. If $\|\nabla f_{i+1}\| \leq \varepsilon$, where ε is a small quantity, take $\mathbf{X}^* \approx \mathbf{X}_{i+1}$ and stop the process. Otherwise, go to step 5.

5. Update the Hessian matrix as

$$[B_{i+1}] = [B_i] + \left(1 + \frac{\mathbf{g}_i^T [B_i] \mathbf{g}_i}{\mathbf{d}_i^T \mathbf{g}_i}\right) \frac{\mathbf{d}_i \mathbf{d}_i^T}{\mathbf{d}_i^T \mathbf{g}_i} - \frac{\mathbf{d}_i \mathbf{g}_i^T [B_i]}{\mathbf{d}_i^T \mathbf{g}_i} - \frac{[B_i] \mathbf{g}_i \mathbf{d}_i^T}{\mathbf{d}_i^T \mathbf{g}_i}$$

where

$$\mathbf{d}_i = \mathbf{X}_{i+1} - \mathbf{X}_i = \lambda_i^* \mathbf{S}_i$$

$$\mathbf{g}_i = \nabla f_{i+1} - \nabla f_i = \nabla f(\mathbf{X}_{i+1}) - \nabla f(\mathbf{X}_i)$$

6. Set the new iteration number as $i = i + 1$ and go to step 2.

Minimize $f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$ from the starting point $\mathbf{X}_1 = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$ using the BFGS method with

$$[B_1] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \varepsilon = 0.01.$$

Solution

Iteration 1 (i = 1)

Here

$$\nabla f_1 = \nabla f(\mathbf{X}_1) = \begin{Bmatrix} 1 + 4x_1 + 2x_2 \\ -1 + 2x_1 + 2x_2 \end{Bmatrix}\Bigg|_{(0,0)} = \begin{Bmatrix} 1 \\ -1 \end{Bmatrix}$$

and hence

$$\mathbf{S}_1 = -[B_1]\nabla f_1 = -\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{Bmatrix} 1 \\ -1 \end{Bmatrix} = \begin{Bmatrix} -1 \\ 1 \end{Bmatrix}$$

Broyden–Fletcher–Goldfarb–Shanno (BFGS) Method

To find the minimizing step length λ_1^* along $\mathbf{S}_1$, we minimize

$$f(\mathbf{X}_1 + \lambda_1 \mathbf{S}_1) = f\left(\begin{Bmatrix} 0 \\ 0 \end{Bmatrix} + \lambda_1 \begin{Bmatrix} -1 \\ 1 \end{Bmatrix}\right) = f(-\lambda_1, \lambda_1) = \lambda_1^2 - 2\lambda_1$$

with respect to λ_1. Since $df/d\lambda_1 = 0$ at $\lambda_1^* = 1$, we obtain

$$\mathbf{X}_2 = \mathbf{X}_1 + \lambda_1^* \mathbf{S}_1 = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} + 1 \begin{Bmatrix} -1 \\ 1 \end{Bmatrix} = \begin{Bmatrix} -1 \\ 1 \end{Bmatrix}$$

Since $\nabla f_2 = \nabla f(\mathbf{X}_2) = \begin{Bmatrix} -1 \\ -1 \end{Bmatrix}$ and $\|\nabla f_2\| = 1.4142 > \varepsilon$, we proceed to update the matrix $[B_i]$ by computing

$$\mathbf{g}_1 = \nabla f_2 - \nabla f_1 = \begin{Bmatrix} -1 \\ -1 \end{Bmatrix} - \begin{Bmatrix} 1 \\ -1 \end{Bmatrix} = \begin{Bmatrix} -2 \\ 0 \end{Bmatrix}$$

Broyden–Fletcher–Goldfarb–Shanno (BFGS) Method

$$\mathbf{d}_1 = \lambda_1^* \mathbf{S}_1 = 1 \begin{Bmatrix} -1 \\ 1 \end{Bmatrix} = \begin{Bmatrix} -1 \\ 1 \end{Bmatrix}$$

$$\mathbf{d}_1 \mathbf{d}_1^{\mathrm{T}} = \begin{Bmatrix} -1 \\ 1 \end{Bmatrix} \{-1 \quad 1\} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

$$\mathbf{d}_1^{\mathrm{T}} \mathbf{g}_1 = \{-1 \quad 1\} \begin{Bmatrix} -2 \\ 0 \end{Bmatrix} = 2$$

$$\mathbf{d}_1 \mathbf{g}_1^{\mathrm{T}} = \begin{Bmatrix} -1 \\ 1 \end{Bmatrix} \{-2 \quad 0\} = \begin{bmatrix} 2 & 0 \\ -2 & 0 \end{bmatrix}$$

$$\mathbf{g}_1 \mathbf{d}_1^{\mathrm{T}} = \begin{Bmatrix} -2 \\ 0 \end{Bmatrix} \{-1 \quad 1\} = \begin{bmatrix} 2 & -2 \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{g}_1^{\mathrm{T}} [B_1] \mathbf{g}_1 = \{-2 \quad 0\} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{Bmatrix} -2 \\ 0 \end{Bmatrix} = \{-2 \quad 0\} \begin{Bmatrix} -2 \\ 0 \end{Bmatrix} = 4$$

$$\mathbf{d}_1 \mathbf{g}_1^{\mathrm{T}} [B_1] = \begin{bmatrix} 2 & 0 \\ -2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ -2 & 0 \end{bmatrix}$$

$$[B_1] \mathbf{g}_1 \mathbf{d}_1^{\mathrm{T}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & -2 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & -2 \\ 0 & 0 \end{bmatrix}$$

$$[B_2]| = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \left(1 + \frac{4}{2}\right) \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 2 & 0 \\ -2 & 0 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 2 & -2 \\ 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} \frac{3}{2} & -\frac{3}{2} \\ -\frac{3}{2} & \frac{3}{2} \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix} - \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{5}{2} \end{bmatrix}$$

Broyden–Fletcher–Goldfarb–Shanno (BFGS) Method

Iteration 2 (i = 2)

The next search direction is determined as

$$\mathbf{S}_2 = -[B_2]\nabla f_2 = -\begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{5}{2} \end{bmatrix} \begin{Bmatrix} -1 \\ -1 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 2 \end{Bmatrix}$$

To find the minimizing step length λ_2^* along $\mathbf{S}_2$, we minimize

$$f(\mathbf{X}_2 + \lambda_2 \mathbf{S}_2) = f\left(\begin{Bmatrix} -1 \\ 1 \end{Bmatrix} + \lambda_2 \begin{Bmatrix} 0 \\ 2 \end{Bmatrix}\right) = f(-1, 1 + 2\lambda_2) = 4\lambda_2^2 - 2\lambda_2 - 1$$

with respect to λ_2. Since $df/d\lambda_2 = 0$ at $\lambda_2^* = \frac{1}{4}$, we obtain

$$\mathbf{X}_3 = \mathbf{X}_2 + \lambda_2^* \mathbf{S}_2 = \begin{Bmatrix} -1 \\ 1 \end{Bmatrix} + \frac{1}{4} \begin{Bmatrix} 0 \\ 2 \end{Bmatrix} = \begin{Bmatrix} -1 \\ \frac{3}{2} \end{Bmatrix}$$

This point can be identified to be optimum since

$$\nabla f_3 = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} \quad \text{and} \quad \|\nabla f_3\| = 0 < \varepsilon$$

Thanks