7-oct-2021

114

$$\underset{\sim}{X}_{i+1} = \underset{\sim}{x}_i - [H_i]^{-1} \nabla f_i$$

Quasi-Newton Methods

$$\underset{\sim}{X}_{i+1} = \underset{\sim}{x}_i - \lambda_i^* [H_i]^{-1} \nabla f_i$$

$$\Rightarrow \underset{\sim}{X}_{i+1} = \underset{\sim}{X}_i + \lambda_i^* \underset{\sim}{S}_i$$

$$\therefore \underset{\sim}{S}_i = -[H_i]^{-1} \nabla f_i$$

If $[H_i]^{-1}$ is identity matrix $\longrightarrow$ steepest descent method.

Either approximate $[H_i]$ with $[A_i]$ or

" $[H_i]^{-1}$ with $[B_i]$

$$\therefore \underset{\sim}{X}_{i+1} = \underset{\sim}{x}_i - \lambda_i^* [B_i] \nabla f_i$$

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

How to compute $[B_i]$ $\quad [[B]_i \approx [H_i]^{-1}]$

Expand $\nabla f(\underset{\sim}{x})$ about $\underset{\sim}{X}_0$ using Taylor series.

$$\nabla f(\underset{\sim}{X}) \simeq \nabla f(\underset{\sim}{X}_0) + [H_0](\underset{\sim}{X} - \underset{\sim}{X}_0) \qquad [A_i] \simeq [H_0]$$

$$\nabla f_{i+1} = \nabla f(\underset{\sim}{X}_0) + [A_i](\underset{\sim}{X}_{i+1} - \underset{\sim}{X}_0)$$

$$\nabla f_i = \nabla f(\underset{\sim}{X}_0) + [A_i](\underset{\sim}{X}_i - \underset{\sim}{X}_0)$$

$$\underbrace{\nabla f_{i+1} - \nabla f_i}_{\underset{\sim}{g}_i} = [A_i] \underbrace{(\underset{\sim}{X}_{i+1} - \underset{\sim}{X}_i)}_{\underset{\sim}{d}_i}$$

$$\therefore \quad [A_i]\, \underset{\sim}{d}_i = \underset{\sim}{g}_i \qquad \Bigg| \begin{array}{l} \underset{\sim}{d}_i = \underset{\sim}{x}_{i+1} - \underset{\sim}{x}_i \\ \underset{\sim}{g}_i = \nabla f_{i+1} - \nabla f_i \end{array}$$

$$\therefore \quad \underset{\sim}{d}_i = [B_i]\, \underset{\sim}{g}_i \qquad [B_i] = [A_i]^{-1}$$

n equations in n^2 unknowns.

the choice for the elements of $[B_i]$ is not unique.

we would like to choose a $[B_i]$ that is close to $[H_0]^{-1}$

- Let's say we know $[B_i]$ then how to find $[B_{i+1}]$?

$$[B_{i+1}] = [B_i] + [\Delta B_i]$$

<u>Rank 1 updates</u>

$$\underbrace{[B_{i+1}]} = [B_i] + [\Delta B_i]$$

$\Delta B_i \leftarrow$ Rank 1 or Rank 2 update.

$$[\Delta B_i] = c\, \underset{\sim}{z}\, \underset{\sim}{z}^T \qquad \underset{\sim}{z}_{n\times 1}$$

$$[B_{i+1}] = [B_i] + c\, \underset{\sim}{z}\, \underset{\sim}{z}^T$$

$$\underset{\sim}{d}_i = [B_{i+1}]\, \underset{\sim}{g}_i$$

$$\underset{\sim}{d}_i = [B_i]\, \underset{\sim}{g}_i + c\, \underset{\sim}{z}\, \underbrace{\underset{\sim}{z}^T \underset{\sim}{g}_i}_{\text{scalar.}}$$

$$c\, \underset{\sim}{z} = \frac{\underset{\sim}{d}_i - [B_i]\, \underset{\sim}{g}_i}{\underset{\sim}{z}^T \underset{\sim}{g}_i}$$

$$c = \frac{1}{\underset{\sim}{z}^T \underset{\sim}{g}_i} \quad \& \quad \underset{\sim}{z} = \underset{\sim}{d}_i - [B_i]\underset{\sim}{g}_i$$

$$\begin{aligned}
[B_{i+1}] &= [B_i] + [\Delta B_i] \\
&= [B_i] + c\,\underset{\sim}{z}\,\underset{\sim}{z}^T \\
&= [B_i] + \frac{1}{\underset{\sim}{z}^T \underset{\sim}{g}_i}\left[[\underset{\sim}{d}_i - [B_i]\underset{\sim}{g}_i]\,[\underset{\sim}{d}_i - [B_i]\underset{\sim}{g}_i]^T\right]
\end{aligned}$$

$$[B_{i+1}] = [B_i] + \frac{(\underset{\sim}{d}_i - [B_i]\underset{\sim}{g}_i)(\underset{\sim}{d}_i - [B_i]\underset{\sim}{g}_i)^T}{(\underset{\sim}{d}_i - [B_i]\underset{\sim}{g}_i)^T\,\underset{\sim}{g}_i} \quad \text{(A)}$$

Broyden's formula

To implement (A) we should start with an initial symmetric positive definite matrix $[B_1]$

$$\underset{\sim}{X}_2 = \underset{\sim}{X}_1 - \lambda_1^*[B_1]\nabla f_1$$

$$\underset{\sim}{X}_3 = \underset{\sim}{X}_2 - \lambda_2^*[B_2]\nabla f_2$$

$$[B_2] = [B_1] + \frac{(\underset{\sim}{d}_1 - [B_1]\underset{\sim}{g}_1)(\underset{\sim}{d}_1 - [B_1]\underset{\sim}{g}_1)^T}{(\underset{\sim}{d}_1 - [B_1]\underset{\sim}{g}_1)^T\,\underset{\sim}{g}_1} \checkmark$$

Eq. (A) ensures symmetry of $[B_{i+1}]$ if $[B_i]$ is symmetric.

17

Rank 2 updates

$$[\Delta B_i] = c_1 \underset{\sim}{z}_1 \underset{\sim}{z}_1^T + c_2 \underset{\sim}{z}_2 \underset{\sim}{z}_2^T$$

$$[B_{i+1}] = [B_i] + c_1 \underset{\sim}{z}_1 \underset{\sim}{z}_1^T + c_2 \underset{\sim}{z}_2 \underset{\sim}{z}_2^T$$

$$\underset{\sim}{d}_i = [B_i]\underset{\sim}{g}_i + c_1 \underset{\sim}{z}_1 (\underset{\sim}{z}_1^T \underset{\sim}{g}_i) + c_2 \underset{\sim}{z}_2 (\underset{\sim}{z}_2^T \underset{\sim}{g}_i)$$

$(\underset{\sim}{X}_{i+1} - x_i)$

$$c_1 = \frac{1}{\underset{\sim}{z}_1^T \underset{\sim}{g}_i} \; ; \; c_2 = -\frac{1}{\underset{\sim}{z}_2^T \underset{\sim}{g}_i}$$

$$\underset{\sim}{z}_1 = \underset{\sim}{d}_i \quad \& \quad \underset{\sim}{z}_2 = [B_i]\underset{\sim}{g}_i$$

$$\underset{\sim}{d}_i = [B_i]\underset{\sim}{g}_i + c_1 \underset{\sim}{z}_1 (\underset{\sim}{z}_1^T \underset{\sim}{g}_i) + c_2 \underset{\sim}{z}_2 (\underset{\sim}{z}_2^T \underset{\sim}{g}_i)$$

$$\underset{\sim}{d}_i - [B_i]\underset{\sim}{g}_i = \frac{1}{\cancel{(\underset{\sim}{z}_1^T \underset{\sim}{g}_i)}} \underset{\sim}{z}_1 \cancel{(\underset{\sim}{z}_1^T \underset{\sim}{g}_i)} - \frac{1}{\cancel{(\underset{\sim}{z}_2^T \underset{\sim}{g}_i)}} \underset{\sim}{z}_2 \cancel{(\underset{\sim}{z}_2^T \underset{\sim}{g}_i)}$$

$$\therefore \; \underset{\sim}{d}_i - [B_i]\underset{\sim}{g}_i = \underset{\sim}{z}_1 - \underset{\sim}{z}_2$$

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

$$\boxed{[B_{i+1}] = [B_i] + \frac{\underset{\sim}{d}_i \underset{\sim}{d}_i^T}{\underset{\sim}{d}_i^T \underset{\sim}{g}_i} - \frac{([B_i]\underset{\sim}{g}_i)([B_i]\underset{\sim}{g}_i)^T}{([B_i]\underset{\sim}{g}_i)^T \underset{\sim}{g}_i}}$$

B1

Davidon - Fletcher - Powell (DFP) formula

$$x_{i+1} = x_i + \lambda_i^* S_i$$

$$d_i = x_{i+1} - x_i = \lambda_i^* S_i$$

$$\therefore [B_{i+1}] = [B_i] + \frac{(\lambda_i^*)^{\cancel{2}} S_i S_i^T}{\cancel{\lambda_i^*} S_i^T g_i} - \frac{[B_i]\, g_i\, g_i^T [B_i]}{g_i^T [B_i]\, g_i}$$

$([B_i]^T = [B_i])$

$$\therefore [B_{i+1}] = [B_i] + \frac{\lambda_i^* S_i S_i^T}{S_i^T g_i} - \frac{[B_i]\, g_i\, g_i^T [B_i]}{g_i^T [B_i]\, g_i} \qquad B2$$

A & B are called <u>inverse</u> <u>update formulas</u>

they are approximating the inverse of the Hessian

- It is also possible to develop update formula by approximating the Hessian itself.

$$g_i = [A_i]\, d_i$$

$$[A_{i+1}] = [A_i] + [\Delta A_i]$$

$$\therefore [A_{i+1}] = [A_i] + \frac{g_i\, g_i^T}{g_i^T d_i} - \frac{([A_i]\, d_i)([A_i]\, d_i)^T}{d_i^T [A_i]\, d_i}$$

Broyden - Fletcher - Goldfarb - Shanno (BFGS) formula

119

- DFP & BFGS formulae are rank 2 updates

$$[B_{i+1}] = \rho_i \left\{ [B_i] - \frac{[B_i]\, \underset{\sim}{g_i}\, \underset{\sim}{g_i}^T [B_i]}{\underset{\sim}{g_i}^T [B_i]\, \underset{\sim}{g_i}} + \theta_i\, \underset{\sim}{y_i}\, \underline{y_i^T} \right\} + \frac{\underset{\sim}{d_i}\, \underset{\sim}{d_i}^T}{\underset{\sim}{d_i}^T\, \underset{\sim}{g_i}}$$

$$\underline{y_i} = \left(\underset{\sim}{g_i}^T [B_i]\, \underset{\sim}{g_i}\right)^{1/2} \left\{ \frac{\underset{\sim}{d_i}}{\underset{\sim}{d_i}^T\, \underset{\sim}{g_i}} - \frac{[B_i]\, \underset{\sim}{g_i}}{\underset{\sim}{g_i}^T [B_i]\, \underset{\sim}{g_i}} \right\}$$

DFP: $\rho_i = 1$ & $\theta_i = 0$

BFGS: $\rho_i = 1$ & $\theta_i = 1$

one-dimensional minimization

$$\underset{\sim}{X_{i+1}} = \underset{\sim}{X_i} + \lambda_i^* \underset{\sim}{S_i}$$

If λ_i^* has some numerical errors → DFP method does not ensure positive definiteness of $[B_i]$

DFP is more sensitive to errors in λ_i^* computation compared to BFGS.

Hence, BFGS is more accurate Quasi-Newton method.

- DFP
- BFGS