4-oct-2021

Conjugate Gradient Method.

Fletcher - Reeves Method.

$$f(\underset{\sim}{X}) = \frac{1}{2}\underset{\sim}{X}^T A \underset{\sim}{X} + B^T\underset{\sim}{X} + C$$

$$\nabla f(\underset{\sim}{X}) = A\underset{\sim}{X} + B^T = 0$$

$$A\underset{\sim}{X} = -\underset{\sim}{B}^T$$

Start with an arbitrary point $\underset{\sim}{X}_1$

$$\underset{\sim}{S}_1 = -\nabla f_1$$

$$\lambda_1^* = -\frac{\underset{\sim}{S}_1^T \nabla f_1}{\underset{\sim}{S}_1^T A \underset{\sim}{S}_1}$$

$\underset{\sim}{S}_2$ should be found in a way that is conjugate to $\underset{\sim}{S}_1 = -\nabla f_1$ & $-\nabla f_2$

$$\underset{\sim}{S}_2 = -\nabla f_2 + \beta_2 \underset{\sim}{S}_1 \qquad \underset{\sim}{S}_1 = -\nabla f_1$$

$$\beta_2 = -\frac{\nabla f_2^T \nabla f_2}{\nabla f_1^T \underset{\sim}{S}_1} = \frac{\nabla f_2^T \nabla f_2}{\nabla f_1^T \nabla f_1}$$

$$\boxed{\underset{\sim}{S}_i = -\nabla f_i + \beta_i \underset{\sim}{S}_{i-1}}$$

where $$\beta_i = \frac{\nabla f_i^T \nabla f_i}{\nabla f_{i-1}^T \nabla f_{i-1}} = \frac{|\nabla f_i|^2}{|\nabla f_{i-1}|^2}$$

F-R

Algorithm:

1. Start with $\underset{\sim}{X}_1$
2. set $\underset{\sim}{S}_1 = -\nabla f(\underset{\sim}{X}_1) = -\nabla f_1$
3. $\underset{\sim}{X}_2 = \underset{\sim}{X}_1 + \lambda_1^* \underset{\sim}{S}_1$; set $i = 2$
4. $\underset{\sim}{S}_i = -\nabla f_i + \underbrace{\dfrac{|\nabla f_i|^2}{|\nabla f_{i-1}|^2}}_{\beta_i} \underset{\sim}{S}_{i-1}$
5. $\underset{\sim}{X}_{i+1} = \underset{\sim}{X}_i + \lambda_i^* \underset{\sim}{S}_i$
6. Test $\underset{\sim}{X}_{i+1}$ for optimality. If X_{i+1} is optimum stop, else $i = i+1$ & go to step 4

How do you test optimality? Check $\nabla f_{i+1} \overset{?}{=} 0$

. Minimize $f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$

$\underset{\sim}{X}_1 = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$

$\underline{i = 1}$

$$\nabla f = \begin{Bmatrix} 1 + 4x_1 + 2x_2 \\ -1 + 2x_1 + 2x_2 \end{Bmatrix}$$

$$\nabla f_1 = \begin{Bmatrix} 1 \\ -1 \end{Bmatrix} \qquad \therefore \underset{\sim}{S}_1 = -\nabla f_i \begin{Bmatrix} -1 \\ 1 \end{Bmatrix}$$

λ_1^* is found by minimizing f along $\underset{\sim}{S}_1$

$$f(\underset{\sim}{X}_1 + \lambda_1 \underset{\sim}{S}_1) = \lambda_1^2 - 2\lambda_1 \Rightarrow \frac{df}{d\lambda_1} = 0$$

$$\Rightarrow \lambda_1^* = 1$$

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

$\underline{\underline{i=2}}$

$$\nabla f_2 = \begin{Bmatrix} -1 \\ -1 \end{Bmatrix}$$

$$\underset{\sim}{S}_i = -\nabla f_i + \beta_i \underset{\sim}{S}_{i-1}$$

$$\therefore \underset{\sim}{S}_2 = \begin{Bmatrix} 1 \\ 1 \end{Bmatrix} + \frac{2}{2} \begin{Bmatrix} -1 \\ 1 \end{Bmatrix}$$

$$= \begin{Bmatrix} 0 \\ 2 \end{Bmatrix}$$

$$\lambda_2^* \rightarrow f(x_2 + \lambda_2 \underset{\sim}{S}_2) = 4\lambda_2^2 - 2\lambda_2 - 1$$

$$\therefore \frac{df}{d\lambda_2} = 8\lambda_2 - 2 = 0 \Rightarrow \lambda_2^* = 1/4$$

$$\therefore \underset{\sim}{X}_3 = \underset{\sim}{X}_2 + \lambda_2^* \underset{\sim}{S}_2$$

$$= \begin{Bmatrix} -1 \\ 1 \end{Bmatrix} + 0.25 \begin{Bmatrix} 0 \\ 2 \end{Bmatrix}$$

$$\underset{\sim}{X}_3 = \begin{Bmatrix} -1 \\ 1.5 \end{Bmatrix}$$

Check for optimality

$$\nabla f_3 = \begin{Bmatrix} 1 + 4(-1) + 2(1.5) \\ -1 + 2(-1) + 2(1.5) \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$$

$$\therefore \underset{\sim}{X}_3 = \underline{\underline{\underset{\sim}{X}^*}}$$

$\nabla f \leq \varepsilon$ (small number; 1×10^{-5})

- Newton's Method.

$$f(\underset{\sim}{X}) = f(\underset{\sim}{x_i}) + \nabla f_i^T(\underset{\sim}{X} - \underset{\sim}{x_i}) + \frac{1}{2}(\underset{\sim}{X} - \underset{\sim}{x_i})^T [H_i] (\underset{\sim}{X} - \underset{\sim}{x_i}) + \cancel{0}$$

Hessian Matrix.

$$\begin{Bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{Bmatrix} = \underset{\sim}{0}$$

$\nabla f = 0$ Necessary condition.

$$\Rightarrow \nabla f_i + [H_i](\underset{\sim}{X} - \underset{\sim}{x_i}) = \underset{\sim}{0}$$

$$f(\underset{\sim}{X}) = \underbrace{\cancel{f(\underset{\sim}{x_i})}}_{0} + \nabla f_i^T (\underset{\sim}{X} - \underset{\sim}{X_i}) + \frac{1}{2}(\underset{\sim}{X} - x_i)^T [H]_i (x_i - x)$$

$$\nabla \underline{f} = \underline{\nabla f_i + [H]_i (\underset{\sim}{X} - \underset{\sim}{x_i})} = 0$$

$$\underset{\sim}{X} = \underset{\sim}{X_{i+1}}$$

$$\therefore \underset{\sim}{X_{i+1}} = \underset{\sim}{x_i} - [H_i]^{-1} \nabla f_i$$

Newton's Method

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

S-D

$$f(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$$

$$\underset{\sim}{X_1} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}$$

$$[H]_i = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}$$

iii

$$\underset{\sim}{X}_2 = \underset{\sim}{X}_1 - [H_1]^{-1} \nabla f_1 \qquad \nabla f_1 = \begin{Bmatrix} 1 \\ -1 \end{Bmatrix}$$

$$= \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} - \frac{1}{4} \begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix} \begin{Bmatrix} 1 \\ -1 \end{Bmatrix} \qquad H^{-1} = \frac{1}{8-4} \begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix}$$

$$= \frac{1}{4} \begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix}$$

$$= \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} - \frac{1}{4} \begin{Bmatrix} 4 \\ -6 \end{Bmatrix} = \begin{Bmatrix} -1 \\ 1.5 \end{Bmatrix}$$

$$\underset{\sim}{X}_2 = \begin{Bmatrix} -1 \\ 1.5 \end{Bmatrix} \simeq \underset{\sim}{X}^*$$

Let's say $\underset{\sim}{X}_1 = \begin{Bmatrix} 0 \\ 1 \end{Bmatrix}$

$$\nabla f_1 = \begin{Bmatrix} 1 + 4(0) + 2(1) \\ -1 + 2(0) + 2(1) \end{Bmatrix} = \begin{Bmatrix} 3 \\ 1 \end{Bmatrix}$$

$$H = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}$$

$$\therefore \underset{\sim}{X}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \frac{1}{4} \begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix} \begin{Bmatrix} 3 \\ 1 \end{Bmatrix}$$

$$\underset{\sim}{X}_2 = \begin{Bmatrix} 0 \\ 1 \end{Bmatrix} - \frac{1}{4} \begin{Bmatrix} 4 \\ -2 \end{Bmatrix} = \begin{Bmatrix} 0 - 1 \\ 1 + \frac{1}{2} \end{Bmatrix} = \begin{Bmatrix} -1 \\ 1.5 \end{Bmatrix}$$

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

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

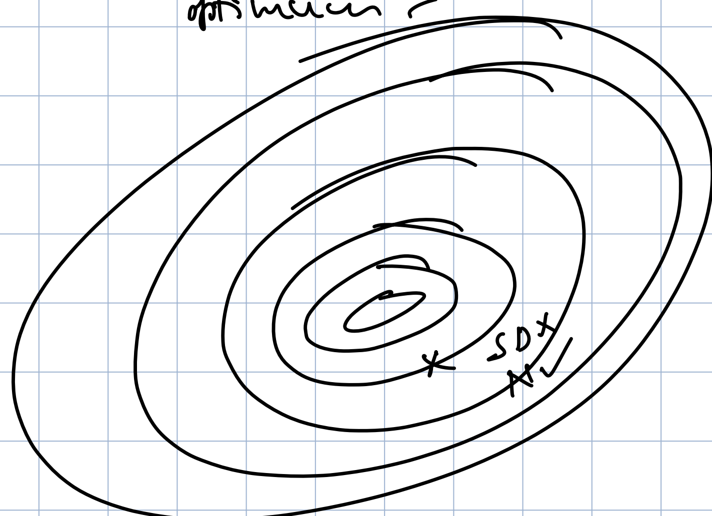
with $\underset{\sim}{S}_i = -[H_i]^{-1} \nabla f_i$

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

$$\lambda_i^* = 1$$

Steepest Descent → Efficient when far away from the optimum

Newton's → Efficient closure to the optimum



SD ✓
N ✗

Marquardt Method.

$[H_i]$ $\quad \boxed{\alpha_i > 0}$

$$[\tilde{H}_i] = [H_i] + \alpha_i [I]$$

$\alpha_i >>> 1 \rightarrow \alpha_i [I] \simeq [\tilde{H}_i]$ $\quad \alpha >>> 1 \quad [\tilde{H}_i] \simeq \alpha_i [I]$

$\alpha_i << 1 \rightarrow [\tilde{H}_i] = [H_i]$

$$\underset{\sim}{S}_i = -\left[\tilde{H}_i\right]^{-1} \nabla f_i$$

$$\left[\tilde{H}_i\right]^{-1} = \left[H_i + \alpha_i I\right]^{-1} \approx \left[\alpha_i I\right]^{-1} \quad \alpha_i >>> 1$$

$$= \frac{1}{\alpha_i} I$$

$$\therefore \boxed{\left[\tilde{H}_i\right]^{-1} \approx \frac{1}{\alpha_i}\left[I\right]}$$

$$\underset{\sim}{S}_i = -\left[\tilde{H}_i\right]^{-1} \nabla f_i$$

$$= -\frac{1}{\alpha_i} \nabla f_i$$

$\alpha_i <<< 1$ $\quad \tilde{H}_i \approx H_i \quad \therefore \tilde{H}_i^{-1} = H_i^{-1}$

$$\underset{\sim}{S}_i = -\left[H_i\right]^{-1} \nabla f_i \quad \leftarrow \text{Newton's method.}$$

- Marquardt method.
  Start with very large α

..