Steepest Descent

$$f(X) = \frac{1}{2}X^T Q X - b^T X$$

where $X = \begin{bmatrix} x \\ y \end{bmatrix}$ $\quad b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$ $\quad Q = \begin{bmatrix} a & c \\ c & b \end{bmatrix}$

$$f(x,y) = \frac{1}{2}\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} a & c \\ c & b \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} b_1 & b_2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

$$f(x,y) = \frac{1}{2}\begin{bmatrix} ax+cy & cx+by \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} b_1 x + b_2 y \end{bmatrix}$$

$$= \frac{1}{2}\left[ax^2 + cxy + cxy + by^2\right] - \left[b_1 x + b_2 y\right]$$

$$\boxed{f(x,y) = \frac{a}{2}x^2 + cxy + \frac{b}{2}y^2 - b_1 x - b_2 y}$$

$$\nabla f(X) = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix} = \begin{pmatrix} ax + cy - b_1 \\ cx + by - b_2 \end{pmatrix}$$

$$= \begin{pmatrix} a & c \\ c & b \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

$$= QX - b$$

$$\nabla f(X) = -(b - QX)$$

$$\{\nabla f(X)\}^T = [-(b - QX)]^T$$

$$= -[b^T - X^T Q^T]$$

$$\{\nabla f(X)\}^T = -[b^T - X^T Q] \quad \text{since } Q = Q^T$$

$$-\{\nabla f(X)\}^T = b^T - X^T Q$$

Iterative procedure.

$$X_{k+1} = X_k - \eta_k \nabla f(X_k)$$

where η_k is the step size

$$\nabla f(X_k) = S, \quad Q = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} = \begin{bmatrix} a & c \\ c & b \end{bmatrix}$$

Derivation of η_k

$$g(\eta_k) = f(X_{k+1})$$

$$= f(X_k - \eta_k \nabla f(X_k))$$

$$= \frac{1}{2}[X_k - \eta_k \nabla f(X_k)]^T Q [X_k - \eta_k \nabla f(X_k)] - b^T[X_k - \eta_k \nabla f(X_k)]$$

$$= \frac{1}{2} X_k^T Q X_k - \frac{1}{2} \eta_k \{\nabla f(X_k)\}^T Q X_k - \frac{1}{2} \eta_k X_k^T Q \nabla f(X_k) \quad \text{same}$$

$$+ \frac{1}{2} \{\nabla f(X_k)\}^T Q \nabla f(X_k) \eta_k^2 - b^T X_k + b^T \eta_k \nabla f(X_k)$$

$$= \frac{1}{2}\{\nabla f(X_k)\}^T Q \nabla f(X_k) \eta_k^2 + b^T \nabla f(X_k) \eta_k - X_k^T Q \nabla f(X_k) \eta_k$$

$$+ \frac{1}{2} X_k^T Q X_k - b^T X_k$$

$$= \frac{1}{2}\{\nabla f(X_k)\}^T Q \nabla f(X_k) \eta_k^2 + (b^T - X_k^T Q) \nabla f(X_k) \eta_k$$

$$+ \frac{1}{2} X_k^T Q X_k - b^T X_k$$

$$= \frac{1}{2}\{\nabla f(X_k)\}^T Q \nabla f(X_k) \eta_k^2 + \left[-\{\nabla f(X_k)\}^T\right] \nabla f(X_k) \eta_k$$

$$+ \frac{1}{2} X_k^T Q X_k - b^T X_k$$

$$g(\eta_k) = A \eta_k^2 + B \eta_k + C$$

where

$$A = \frac{1}{2}\{\nabla f(X_k)\}^T Q \nabla f(X_k)$$

$$B = -\{\nabla f(X_k)\}^T \nabla f(X_k)$$

$$C = \frac{1}{2} X_k^T Q X_k - b^T X_k.$$

$$g(\eta_k) = A\eta_k^2 + B\eta_k + C$$

$$\frac{dg}{d\eta_k} = 2A\eta_k + B$$

Equate to zero $\frac{dg}{d\eta_k}$ for finding step size.

$$2A\eta_k + B = 0$$

$$\eta_k = \frac{-B}{2A}$$

$$= \frac{\{\nabla f(X_k)\}^T \nabla f(X_k)}{\{\nabla f(X_k)\}^T Q \nabla f(X_k)}$$

$$= \frac{S^T S}{S^T Q S}$$

where $S = \nabla f(X_k) = \begin{bmatrix} \partial f/\partial x \\ \frac{\partial f}{\partial y} \end{bmatrix}$

$Q = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$ is Hassian matrix