Ch 10

Def 10.1 Let Q be a real symmetric $n \times n$ matrix.

$d^{(0)}, \dots, d^{(m)}$ are <u>Q-conjugate</u> if for all $i \neq j$, we have $d^{(i)T} Q d^{(j)} = 0$.

Lemma 10.1 Let Q be a symmetric positive definite $n \times n$ matrix.

If $d^{(0)}, \dots, d^{(k)} \in \mathbb{R}^n$, $k \leq n-1$, are nonzero and Q-conjugate then they are linearly independent.

(proof) Consider $\alpha_0 d^{(0)} + \cdots + \alpha_k d^{(k)} = 0$.
WTS $\alpha_0 = \cdots = \alpha_k = 0$.

Multiply by $d^{(j)T} Q$ $0 \leq j \leq k$, then

$$\alpha_0 d^{(j)T} Q d^{(0)} + \cdots + \alpha_k d^{(j)T} Q d^{(k)} = 0.$$

By Q-conjugacy, $d^{(j)T} Q d^{(i)} = 0$ for all $i \neq j$.

$\Rightarrow$ $\alpha_j d^{(j)T} Q d^{(j)} = 0$ for all $0 \leq j \leq k$.

$\Rightarrow$ $\alpha_j = 0$ for all $0 \leq j \leq k$. $\square$

(Q is positive definite $x^T Q x > 0$ for all $x \neq 0$.

§ 10.2

$f(x) = \frac{1}{2} x^T Q x - x^T b$. $Q = Q^T > 0$ $x \in \mathbb{R}^n$
$\hookrightarrow$ f has a global minimizer.

Basic Conjugate Direction Algorithm.

Given $x^{(0)}$ and Q-conjugate directions $d^{(0)}, d^{(1)}, \dots, d^{(n-1)}$; for $k \geq 0$

$$(*) \quad \begin{cases} g^{(k)} = \nabla f(x^{(k)}) = Qx^{(k)} - b \\ \alpha_k = -\dfrac{g^{(k)T} d^{(k)}}{d^{(k)T} Q d^{(k)}} \\ x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)} \end{cases}$$

Thm 10.1 For any starting point $x^{(0)}$, the basic conjugate direction algorithm converges to the unique x^* (that solves $Qx = b$) in n steps; $x^{(n)} = x^*$.

(proof) Consider $x^* - x^{(0)} \in \mathbb{R}^n$
Since $d^{(0)}, \dots, d^{(n-1)}$ are n linearly independent in $\mathbb{R}^n$,

$$\mathbb{R}^n = \text{span}\{d^{(0)}, \cdots, d^{(n-1)}\}$$

There exist constants β_i, $i=0,\cdots,n-1$ such that

$$x^* - x^{(0)} = \beta_0 d^{(0)} + \cdots + \beta_{n-1} d^{(n-1)} \qquad (0)$$

Multiply by $d^{(k)T}Q$ $(0 \le k < n)$,

$$d^{(k)T}Q\left(x^* - x^{(0)}\right) = d^{(k)T}Q\left(\beta_0 d^{(0)} + \cdots + \beta_{n-1} d^{(n-1)}\right)$$

$$= \beta_0 d^{(k)T}Qd^{(0)} + \cdots + \beta_{n-1} d^{(k)T}Qd^{(n-1)}$$

$$= \beta_k d^{(k)T}Qd^{(k)} \quad \text{(By } Q\text{-conjugacy)}$$

Thus, $$\beta_k = \frac{d^{(k)T}Q(x^* - x^{(0)})}{d^{(k)T}Qd^{(k)}} \qquad (1)$$

From $(*)$, $$x^{(k)} = x^{(k-1)} + \alpha_{k-1} d^{(k-1)}$$
$$= x^{(k-2)} + \alpha_{k-2} d^{(k-2)} + \alpha_{k-1} d^{(k-1)}$$
$$= \cdots$$
$$= x^{(0)} + \alpha_0 d^{(0)} + \cdots + \alpha_{k-1} d^{(k-1)} \qquad (2)$$

Write $$x^* - x^{(0)} = x^* - x^{(k)} + x^{(k)} - x^{(0)}$$

Multiply by $d^{(k)T}Q$,

$$d^{(k)T}Q\left(x^* - x^{(0)}\right) = d^{(k)T}Q\left(x^* - x^{(k)}\right) + \underbrace{d^{(k)T}Q\left(x^{(k)} - x^{(0)}\right)}_{=0 \text{ by } (2)}$$

$$= d^{(k)T}\left(Qx^* - Qx^{(k)}\right)$$

$$= d^{(k)T}\left(b - Qx^{(k)}\right)$$

$$\underbrace{\beta_k d^{(k)T}Qd^{(k)}}_{\text{by } (1)} = -d^{(k)T}g^{(k)} \quad \text{by } (*).$$

$$\Rightarrow \beta_k = -\frac{d^{(k)T}g^{(k)}}{d^{(k)T}Qd^{(k)}} = \alpha_k$$

$$\Rightarrow x^{(k)} = x^{(0)} + \alpha_0 d^{(0)} + \cdots + \alpha_{n-1} d^{(n-1)} \quad \text{from } (2)$$

$$x^* = x^{(0)} + \beta_0 d^{(0)} + \cdots + \beta_{n-1} d^{(n-1)} \quad \text{from } (0). \qquad \square$$

Lemma 10.2 In the conjugate direction algorithm,

$$g^{(k+1)T} d^{(i)} = 0 \quad \text{for all } k,\ 0 \le k \le n-1, \text{ and } 0 \le i \le k.$$

(proof) Note from $(*)$, $x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)}$

$$\begin{aligned} g^{(k+1)} &= Qx^{(k+1)} - b \\ &= Q(x^{(k)} + \alpha_k d^{(k)}) - b \\ &= Qx^{(k)} - b + \alpha_k Q d^{(k)} \\ &= g^{(k)} + \alpha_k Q d^{(k)} \end{aligned} \qquad (1)$$

Prove by induction.

Suppose $k=0$. WTS $g^{(1)T} d^{(0)} = 0$

$$\begin{aligned} g^{(1)T} d^{(0)} &= (g^{(0)} + \alpha_k Q d^{(0)})^T d^{(0)} \quad \text{by (1)} \\ &= g^{(0)T} d^{(0)} + \alpha_0 d^{(0)T} Q d^{(0)} \\ &= g^{(0)T} d^{(0)} + \left(- \frac{g^{(0)T} d^{(0)}}{d^{(0)T} Q d^{(0)}}\right) d^{(0)T} Q d^{(0)} \\ &= 0 \end{aligned}$$

Suppose the result is true for $k-1$. i.e.

$$g^{(k)T} d^{(i)} = 0 \quad \text{for} \quad 0 \le i \le k-1 \qquad (2)$$

WTS it is true for k.

$$\begin{aligned} g^{(k+1)T} d^{(i)} &= (g^{(k)} + \alpha_k Q d^{(k)})^T d^{(i)} \quad (\text{by } (1)) \\ &= g^{(k)T} d^{(i)} + \alpha_k d^{(k)T} Q d^{(i)} \end{aligned}$$

if $0 \le i \le k-1$, then

$g^{(k)T} d^{(i)} = 0$ by (2)

$\alpha_k d^{(k)T} Q d^{(i)} = 0$ since d is Q-conjugate.

if $i = k$, then

$$g^{(k)T} d^{(k)} + \alpha_k d^{(k)T} Q d^{(k)} = g^{(k)T} d^{(k)} + \left(- \frac{g^{(k)T} d^{(k)}}{d^{(k)} Q d^{(k)}}\right) d^{(k)T} Q d^{(k)} = 0.$$

By induction, the result is true for all $0 \le k \le n-1$ and $0 \le i \le k$.

$\square$

Chapter 11

Quasi-Newton algorithm

$$(*)\quad \begin{cases} d^{(k)} = -H_k g^{(k)} \\ \alpha_k = \underset{\alpha \ge 0}{\arg\min}\, f(x^{(k)} + \alpha d^{(k)}) \\ x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)} \end{cases} \qquad (**)\quad \begin{cases} \Delta x^{(i)} = x^{(i+1)} - x^{(i)} = \alpha_i d^{(i)} \\ \Delta g^{(i)} = g^{(i+1)} - g^{(i)} = Q \Delta x^{(i)} \end{cases}$$

$H_0, H_1, \cdots$ are symmetric.

Thm 11.1 Consider a quasi-Newton algorithm applied to a quadratic function with Hessian $Q = Q^T$ such that for $0 \le k \le n-1$,

$$H_{k+1} \Delta g^{(i)} = \Delta x^{(i)} \qquad 0 \le i \le k \qquad (1)$$

where $H_{k+1} = H_{k+1}^T$. If $\alpha_k \ne 0$, $0 \le i \le k$, then $d^{(0)}, \cdots, d^{(k+1)}$ are Q-conjugate.

(proof) Prove by induction.

Suppose $k=0$. WTS $d^{(0)}$ $d^{(1)}$ are Q-conjugate. i.e.

$$d^{(0)T} Q d^{(1)} = 0.$$

$$\begin{aligned} d^{(1)T} Q d^{(0)} &= (-H_1 g^{(1)})^T Q d^{(0)} && \text{by } (*) \\ &= -g^{(1)T} H_1 Q d^{(0)} \\ &= -g^{(1)T} H_0 Q \left(\frac{1}{\alpha_k} \Delta x^{(0)}\right) && \text{by } (**) \\ &= -\frac{1}{\alpha_k} g^{(1)T} H_1 Q \Delta x^{(0)} \\ &= -\frac{1}{\alpha_k} g^{(1)T} H_1 \Delta g^{(0)} && \text{by } (**) \\ &= -\frac{1}{\alpha_k} g^{(1)T} \Delta x^{(0)} && \text{by } (1) \\ &= -g^{(1)T} d^{(0)} && \text{by } (**) \end{aligned} \qquad (2)$$

Note that $\alpha_0 = \underset{\alpha}{\arg\min}\, f(x^{(0)} + \alpha d^{(0)})$

By FONC, $\nabla f(x^{(0)} + \alpha_0 d^{(0)})^T d^{(0)} = 0$

$$\begin{aligned} &= \left(Q\left(x^{(0)} + \alpha_0 d^{(0)}\right) - b\right)^T d^{(0)} \\ &= \left(Qx^{(0)} - b + \alpha_0 Q d^{(0)}\right)^T d^{(0)} \\ &= \left(g^{(0)} + \alpha_0 Q d^{(0)}\right)^T d^{(0)} \end{aligned}$$

$$= \left(g^{(0)} + Q\,\Delta x^{(0)}\right)^T d^{(0)} \quad \text{by } (**)$$

$$= g^{(1)T} d^{(0)} \quad \text{by } (**)$$

$$= 0$$

Thus, $d^{(1)} Q d^{(0)} = 0.$

Now suppose the result is true for $k-1$ $(k < n-1)$ i.e.

$$d^{(0)}, \cdots, d^{(k)} \text{ are } Q\text{-conjugate} \qquad (3)$$

WTS $d^{(0)}, \cdots, d^{(k+1)}$ are Q-conjugate.

$\hookrightarrow$ $d^{(k+1)T} Q d^{(i)} = 0$ for all $i = 0, \cdots, k$

$$d^{(k+1)T} Q d^{(i)} = -g^{(k+1)T} H_{k+1} Q d^{(i)} \quad \text{by } (*)$$

$$= -g^{(k+1)T} H_{k+1} Q \left(\frac{1}{\alpha_i} \Delta x^{(i)}\right) \quad \text{by } (**)$$

$$= -\frac{1}{\alpha_i} g^{(k+1)T} H_{k+1} Q\, \Delta x^{(i)}$$

$$= -\frac{1}{\alpha_i} g^{(k+1)T} H_{k+1} \Delta g^{(i)} \quad \text{by } (**)$$

$$= -\frac{1}{\alpha_i} g^{(k+1)T} \Delta x^{(i)} \quad \text{by } (1)$$

$$= -g^{(k+1)T} d^{(i)} \quad \text{by } (**)$$

$$= 0 \quad \text{by lemma 10.2.}$$

Thus, $d^{(k+1)T} Q d^{(i)} = 0$ for $i = 0, \cdots, k.$ $\square$

Chapter 8

Lemma 8.2 Let $Q = Q^T > 0$ $Q \in \mathbb{R}^{n \times n}.$

For any $x \in \mathbb{R}^n$,

$$\frac{\lambda_{min}(Q)}{\lambda_{max}(Q)} \le \frac{(x^T x)^2}{(x^T Q x)(x^T Q^{-1} x)} \le \frac{\lambda_{max}(Q)}{\lambda_{min}(Q)}$$

(proof) Rayleigh's inequality

$$\lambda_{min}(Q) \le \frac{x^T Q x}{x^T x} \le \lambda_{max}(Q)$$

$$\lambda_{min}(Q^{-1}) \le \frac{x^TQ^{-1}x}{x^Tx} \le \lambda_{max}(Q^{-1})$$

Q: What is $\lambda_{min}(Q^{-1})$? $\lambda_{max}(Q^{-1})$?

Since Q is symmetric positive definite, Q can be written as

$Q = PDP^T$ where P is orthogonal $P^TP = P^TP = I$
D is a diagonal matrix.

Spectral Theorem?

$$Q = P \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} P^T$$

$$Q^{-1} = \left(P \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} P^T \right)^{-1} = (P^T)^{-1} \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}^{-1} P^{-1}$$

$$= P \begin{pmatrix} \frac{1}{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\lambda_n} \end{pmatrix} P^T$$

$$\Rightarrow \quad \lambda_{min}(Q^{-1}) = \frac{1}{\lambda_{max}(Q)} \qquad \lambda_{max}(Q^{-1}) = \frac{1}{\lambda_{min}(Q)}$$

Thus,

$$\frac{1}{\lambda_{max}(Q)} \le \frac{x^TQ^{-1}x}{x^Tx} \le \frac{1}{\lambda_{min}(Q)}$$

Multiply all together,

$$\frac{\lambda_{min}(Q)}{\lambda_{max}(Q)} \le \frac{(x^TQx)(x^TQ^{-1}x)}{(x^Tx)^2} \le \frac{\lambda_{max}(Q)}{\lambda_{min}(Q)} \qquad \square$$