Lecture 12

Thursday, 10 February 2022 1:43 PM

Steepest descent for $\boxed{f(\underline{x}) = \frac{1}{2}\underline{x}^T Q \underline{x} - \underline{b}^T \underline{x}}$

$\left[\begin{array}{l} Q = Q^T \\ Q > 0 \end{array}\right]$

$\hookrightarrow \left.\begin{array}{l} \nabla f(\underline{x}) = Q\underline{x} - \underline{b} \\ F = Q \end{array}\right]$

$\underline{x}^{(k+1)} = \underline{x}^{(k)} - \alpha_k \nabla f(\underline{x}^{(k)})$ where

$\alpha_k = \arg\min_{\alpha \geq 0} f(x^{(k)} - \alpha \nabla f(x^{(k)}))$

Qn: Can we find α_k explicitly?

$g^{(k)} = \nabla f(x^{(k)})$
$= Q\underline{x}^{(k)} - \underline{b}.$

Recall $\phi_k(\alpha) = f(\underbrace{x^{(k)} - \alpha g^{(k)}})$

FONC to $\phi_k(\alpha_k)$ yields

$$0 = \phi_k'(\alpha_k) = \nabla f(x^{(k)} - \alpha g^{(k)}) \cdot (-g^{(k)})$$

$$0 = -(g^{(k)})^T (Q(x^{(k)} - \alpha_k g^{(k)}) - \underline{b})$$

$$\Rightarrow -(g^{(k)})^T Q x^{(k)} + \underbrace{\alpha_k (g^{(k)})^T Q g^{(k)}} + (g^{(k)})^T \underline{b} = 0$$

$$\alpha_k (g^{(k)})^T Q g^{(k)} = (g^{(k)})^T [\underbrace{Q x^{(k)} - \underline{b}}_{g^{(k)}}]$$

$$\boxed{\alpha_k = \frac{(g^{(k)})^T g^{(k)}}{(g^{(k)})^T Q g^{(k)}}} \quad \text{— (1)}$$

$$\boxed{\underline{x}^{(k+1)} = \underline{x}^{(k)} - \frac{(g^{(k)})^T g^{(k)}}{(g^{k})^T Q g^{(k)}} g^{(k)} \qquad Q \text{ is SPD.}}$$

Ex. $\min f(x_1, x_2) = x_1^2 + x_2^2$

$\nabla f = \begin{pmatrix} 2x_1 \\ \end{pmatrix}$ $\quad Q = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = 2I$

$g^{(0)} = \nabla f(\underline{x}^{(0)})$
$= Q\underline{x}^{(0)}$

$$\nabla f = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} \qquad Q = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = 2I$$

$$g^{(0)} = \nabla f(x^{(0)}) = Q x^{(0)} = 2x^{(0)}$$

$$Q g^{(0)} = 2 g^{(0)}$$

$$\underline{x}^{(1)} = \underline{x}^{(0)} - \frac{(g^{(0)})^T g^{(0)}}{(g^{(0)})^T Q g^{(0)}} g^{(0)}$$

$$= \underline{x}^{(0)} - \frac{\cancel{(g^{(0)})^T g^{(0)}}}{\cancel{2 g_0^T g^{(0)}}} \times 2 x^{(0)}$$

$$\boxed{\underline{x}^{(1)} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}}$$

Ex. $f(x_1, x_2) = \frac{x_1^2}{5} + x_2^2$ Zig-zag path ✓

Chapter 8 → Gradient methods → (Deviation in the convergence analysis).

Convergence of the steepest descent method [Quadratic forms Q SPD].

$$x^{(k+1)} = x^{(k)} - \alpha_k g^{(k)}$$

$g^{(k)} = \nabla f(x^{(k)})$

$\alpha = \ldots$ in ①

Relation between $f(x^{(k+1)})$ and $f(x^{(k)})$

[Skip superscripts, subscripts]

[$\underline{d} = -g = -\nabla f$]

$$f(\underline{x} + \alpha \underline{d}) = \frac{1}{2}(\underline{x} + \alpha d)^T Q (\underline{x} + \alpha d) - b^T(\underline{x} + \alpha \underline{d})$$

$$= f(\underline{x}) + \alpha \underline{x}^T Q \underline{d} + \frac{\alpha^2}{2} \underline{d}^T Q \underline{d} - \alpha \underline{b}^T \underline{d}$$

$$= f(\underline{x}) + \alpha d^T [Q\underline{x} - \underline{b}] + \frac{\alpha^2}{2} \underline{d}^T Q \underline{d}$$

$$= f(x) - \alpha \underline{d}^T \underline{d} + \frac{\alpha^2}{2} \underline{d}^T Q \underline{d}$$

$$= f(x) - \frac{(\underline{d}^T \underline{d})^2}{\ldots} + \frac{1}{2} \frac{(\underline{d}^T \underline{d})^2}{(\underline{d}^T Q \underline{d})^2} \cancel{(\underline{d}^T Q \underline{d})}$$

α from ①

$$\alpha = \frac{\underline{d}^T \underline{d}}{\underline{d}^T Q \underline{d}}$$

$$= f(x) - \frac{(d^T d)^2}{d^T Q d} + \frac{1}{2}\frac{(d^T d)^2 \cancel{(d^T Q d)}}{(d^T Q d)^{2}}$$

$$f(x+\alpha d) = f(x) - \frac{1}{2}\frac{(d^T d)^2}{d^T Q d}$$

[We are writing $f(x^{k+1})$ in terms of $f(x^k)$]

Compare $f(x^{(k+1)})$ and $f(x^{(k)})$

$$\frac{f(x^{k+1}) - f(x^*)}{f(x^{(k)}) - f(x^*)} \leq \delta$$

$$\left(\Rightarrow f(x^{(k+1)}) - f(x^*) \leq \delta\left(f(x^{(k)}) - f(x^*)\right)\right).$$

We want 'δ' to be smaller rather than larger.

$x^* = Q^{-1} b$

['δ' depends on the smallest and largest eigenvalue of the Hessian]

$$\frac{f(x^{(k+1)}) - f(x^*)}{f(x^{(k)}) - f(x^*)} = \frac{f(x^{(k)}) - \frac{1}{2}\frac{(d^T d)^2}{d^T Q d} - f(x^*)}{f(x^{(k)}) - f(x^*)}$$

$$= 1 - \frac{\frac{1}{2}\frac{(d^T d)^2}{d^T Q d}}{\underbrace{f(x^{(k)}) - f(x^*)}}.$$

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

Exercise $\rightarrow$ $f(x^k) - f(x^*)$

HW $= \frac{1}{2} d^T Q^{-1} d$

$$= 1 - \frac{(d^T d)^2}{(d^T Q d)(d^T Q^{-1} d)}$$

$$= \overbrace{1 - \frac{1}{\beta}}^{\text{small}}$$

$$\beta := \frac{(\underline{d}^T Q \underline{d})(\underline{d}^T Q^{-1} \underline{d})}{(\underline{d}^T \underline{d})^2}$$

$\lambda_{min}, \lambda_{max} \rightarrow Q$

Kantorovich inequality $\rightarrow$

$$\beta \leq \frac{(\lambda_{min} + \lambda_{max})^2}{4 \lambda_{min} \lambda_{max}}$$