Lecture 13

Monday, 14 February 2022 1:43 PM

[Steepest descent method]

Continued: Recall

$$\min_{\underline{x} \in \mathbb{R}^n} f(\underline{x}) \left(:= \overbrace{\underbrace{\frac{1}{2} \underline{x}^T Q \underline{x} - \underline{b}^T \underline{x} + \underline{c}}}^{\text{quadratic form}} \right)$$

$$\left(\begin{array}{l} Q = Q^T > 0 \\ Q \text{ is SPD} \end{array} \right)$$

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

$$\alpha_k := \frac{(g^{(k)})^T g^{(k)}}{(g^{(k)})^T Q g^{(k)}} \qquad g^{(k)} = \nabla f(x^{(k)}).$$

We proved:

$$\frac{f(\underline{x}^{(k+1)}) - f(\underline{x}^{*})}{f(\underline{x}^{(k)}) - f(\underline{x}^{*})} = 1 - \frac{1}{\beta} \le \delta$$

$\underline{x}^{*}$ (minimum)

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

Kantorovich inequality leads to

$$\boxed{\beta \le \frac{(\lambda_{min} + \lambda_{max})^2}{4 \lambda_{min} \lambda_{max}}} \rightarrow \text{To be proved.}$$

$$\frac{f(\underline{x}^{(k+1)}) - f(\underline{x}^{*})}{f(\underline{x}^{(k)}) - f(\underline{x}^{*})} \le 1 - \frac{4 \lambda_{min} \lambda_{max}}{(\lambda_{min} + \lambda_{max})^2}$$

$$= \left(\frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}} \right)^2$$

$$\boxed{k = \frac{\lambda_{max}}{\lambda_{min}}}$$

condition number

$$= \left(\frac{k-1}{k+1} \right)^2$$

$$= \left(1 - \frac{2}{k+1} \right)^2 =: \delta.$$

$$k+1 \overline{\smash{\big)} k-1} \quad 1, \quad k+1, \quad -2$$

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

$$f(x^{(k+1)}) - f(x^*) \le \delta \left(f(x^{(k)}) - f(x^*) \right)$$
$$\le \delta^2 \left(f(x^{(k-1)}) - f(x^*) \right)$$

Q is SPD $\quad 0 < \lambda_{min} = \lambda_1 \le \lambda_2 \le \dots \le \lambda_n = \lambda_{max}$

$$k = \frac{\lambda_{max}}{\lambda_{min}} \geq 1$$

What we wish: If k is small (not too large than 1), then δ will be much smaller than 1.

If k is large, the convergence is slower.

Discussions: Number of iteration (# of iterations) to reduce the optimality gap by a factor of 10 (10%) can be computed using k. $(*)$

$k = \frac{\lambda_{max}}{\lambda_{min}}$ (given)	$\delta := \left(1 - \frac{2}{k+1}\right)^2$	# of iteration for $(*)$
1.1	0.0023	1
3.0	0.25	2
10.0	0.67	6.

$(0.67)^n = 0.1$

$n \log 0.67 = \log 0.1$

$n = \frac{\log 0.1}{\log 0.67}$

To Prove

$$\beta := \frac{(d^T Q d)(d^T Q^{-1} d)}{(d^T d)(d^T d)} \le \frac{(\lambda_{min} + \lambda_{max})^2}{4 \lambda_{min} \lambda_{max}}.$$

Q is S.P.D. $\qquad Q = R D R^T \quad \dots$

Q is S.P.D.

$$Q = R D R^T \qquad R^T R = I \qquad R^{-1} = R^T$$

$$D = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}$$

$$Q^{-1} = (R D R^T)^{-1} = (R^T)^{-1} D^{-1} R^{-1} = R D^{-1} R^T.$$

$$f = R^T d \qquad f^T = d^T R$$

$$\beta := \frac{(d^T R D \overbrace{R^T d}^{f})(d^T R D^{-1} R^T d)}{(\underbrace{d^T R}_{} R^T d)(d^T R R^T d)}$$

$$= \frac{(f^T D f)(f^T D^{-1} f)}{(f^T f)(f^T f)}$$

$$= \frac{\sum_{i=1}^n \lambda_i f_i^2}{\|f\|^2} \frac{\sum_{i=1}^n \lambda_i^{-1} f_i^2}{\|f\|^2}$$

$$\beta := \left(\sum_{i=1}^n \lambda_i x_i\right)\left(\sum_{i=1}^n \lambda_i^{-1} x_i\right), \text{ where } x_i = \frac{f_i^2}{\|f\|^2}, \quad \sum_{i=1}^n x_i = 1.$$

TPT

$$\le \underbrace{\left(\frac{\lambda_1 + \lambda_n}{2}\right)^2}_{(AM)^2} \underbrace{\frac{1}{\lambda_1 \lambda_n}}_{(G.M)^{-2}}$$

λ_1, λ_n

$\boxed{\text{Kantorovich inequality}}$

$\lambda_1 \le \lambda_2 \cdots \le \lambda_n.$

Kantorovich inequality $\left(\sum_{i=1}^n x_i = 1\right)$

$$\left(\sum_{i=1}^n \lambda_i x_i\right)\left(\sum_{i=1}^n \lambda_i^{-1} x_i\right) \le (AM)^2 (GM)^{-2}$$

$$AM := \frac{\lambda_1 + \lambda_n}{2}$$

$$GM := \sqrt{\lambda_1 \lambda_n}.$$

[American Math Monthly 1995 Pták]

$$\left(\sum_{i=1}^n \lambda_i x_i\right)\left(\sum_{i=1}^n \lambda_i^{-1} x_i\right) \le \left(\frac{\lambda_1 + \lambda_n}{2}\right)^2 \frac{1}{\lambda_1 \lambda_n} = \frac{AM^2}{GM^2}$$

Step 1 The inequality is invariant w.r.t. replacing λ_i by

Step 1. The inequality is invariant w.r.t. replacing λ_i by $c\lambda_i$ $(c>0)$

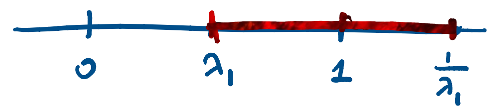
Step 2. W.l.o.g. Choose $GM = 1$.

$$\sqrt{\lambda_1 \lambda_n} = 1, \quad \lambda_n = \frac{1}{\lambda_1}$$

$$0 < \lambda_1 \le \lambda_2 \le \dots \le \lambda_n = \frac{1}{\lambda_1}$$

Each y between x and $\frac{1}{x}$, $0 < x \le 1$ satisfies

$$y + \frac{1}{y} \le x + \frac{1}{x}$$



$$\lambda_i + \frac{1}{\lambda_i} \le \lambda_1 + \frac{1}{\lambda_1} \quad (i=1,\dots n).$$

Step 3. $ab \le \frac{a^2+b^2}{2}$

$(a-b)^2 \ge 0$

$a = \sqrt{\sum_{i=1}^{n} \lambda_i x_i}$

$$\sqrt{\left(\sum_{i=1}^{n} \lambda_i x_i\right)\left(\sum_{i=1}^{n} \lambda_i^{-1} x_i\right)} \le \frac{1}{2}\left(\sum_{i=1}^{n} \lambda_i x_i + \sum_{i=1}^{n} \lambda_i^{-1} x_i\right)$$

$$= \frac{1}{2} \sum_{i=1}^{n} (\lambda_i + \lambda_i^{-1})\, x_i$$

$$\le \frac{1}{2}\left(\lambda_1 + \frac{1}{\lambda_1}\right)\underbrace{\left(\sum_{i=1}^{n} x_i\right)}_{=1}$$

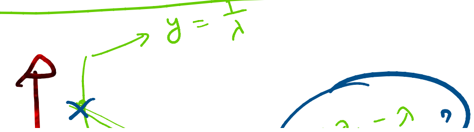
$$\le \frac{1}{2}(\lambda_1 + \lambda_n) = \frac{AM}{GM}$$

$$\Rightarrow \left(\sum_{i=1}^{n} \lambda_i x_i\right)\left(\sum_{i=1}^{n} \lambda_i^{-1} x_i\right) \le \frac{(\lambda_1+\lambda_n)^2}{4\lambda_1\lambda_n}$$

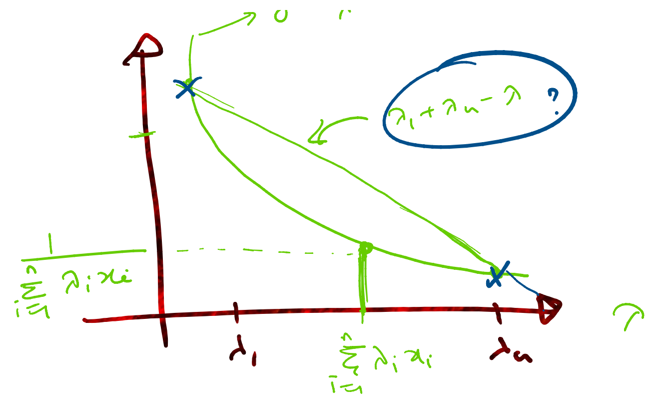
That is,

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

Exercise: $\boxed{\text{3 different proofs for Kantorovich inequality}}$



$\sum_{i=1}^{n} \lambda_i^{-1} x_i$



$$\frac{\sum_{i=1}^n \lambda_i x_i}{\frac{1}{\sum_{i=1}^n \lambda_i x_i}}$$

Gradient method with a fixed step-size:

Choose $\alpha_k = \alpha \in \mathbb{R}$ for all k.

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

$\min_{x \in \mathbb{R}^n} f(\underline{x})$ ↓ (quadratic fn)

→ Simple, no line search

→ Convergence depends on the choice of 'α'.

Theorem: [without proof]

For the fixed step-size gradient algorithm,

$\underline{x}^{(k)} \longrightarrow \underline{x}^{(*)}$ for any $\underline{x}^{(0)}$, iff

$$\boxed{0 < \alpha < \frac{2}{\lambda_{max}(Q)}}$$

Proof is based on Q S.P.D

Rayleigh inequality

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

Ex: (i) Find range of α for which fixed-step gradient method to minimize $f(\underline{x})$ defined by

$$f(x_1, x_2) = \underline{x}^T \underbrace{\begin{bmatrix} 4 & 2\sqrt{2} \\ 0 & 5 \end{bmatrix}}_{A} \underline{x} + \underline{x}^T \begin{bmatrix} 3 & 6 \end{bmatrix} + 24$$

converges.

$$Q = \frac{1}{2}(A + A^T)$$

Q

$\begin{bmatrix} 2 & 2\sqrt{2} \end{bmatrix}$ ... $x^T \begin{bmatrix} 3 & 6 \end{bmatrix} + 24$

converge...

Solution Symmetrize $\rightarrow$ $\frac{1}{2}\underline{x}^T \overbrace{\begin{bmatrix} 8 & 2\sqrt{2} \\ 2\sqrt{2} & 10 \end{bmatrix}}^{Q} \underline{x} + \underline{x}^T [3 \quad 6] + 24$

$$|Q - \lambda I| = 0$$

$$\begin{vmatrix} 8-\lambda & 2\sqrt{2} \\ 2\sqrt{2} & 10-\lambda \end{vmatrix} = 0 \Rightarrow (8-\lambda)(10-\lambda) - 8 = 0$$

$$\Rightarrow (\lambda - 8)(\lambda - 10) - 8 = 0$$

$$\Rightarrow \lambda^2 - 18\lambda + \overbrace{80 - 8}^{72} = 0$$

$$\Rightarrow (\lambda - 6)(\lambda - 12) = 0$$

$$\Rightarrow \lambda = 6, \quad \lambda = 12 \rightarrow \boxed{Q \text{ is SPD}}$$

Range of α : $0 < \alpha < \frac{2}{12} \Rightarrow \boxed{0 < \alpha < 1/6}$

(ii) How many iteration are needed in steepest-descent method to ~~leave~~ an optimality gap of 1%?

$$k = \frac{12}{6} = 2$$

$$\delta = \left(1 - \frac{2}{k+1}\right)^2 = \left(1 - \frac{2}{3}\right)^2 = \frac{1}{9} = \underline{0.11}$$

$$\left(f(x^{(k+1)}) - f(x^*)\right) \le (0.11)^k \ldots\ldots$$

$$(0.11)^k = 0.01$$

$$k \log 0.11 = \log 0.01$$

$$k = \frac{\log 0.01}{\log 0.11} \approx \underline{3}.$$

$$k = \frac{\log 0.01}{\log 0.11} \approx \underline{3}.$$