Q & A Session 8 (April 1)

Proof of Theorem 5.4 (p. 115)

$\lambda_1, \lambda_2, \cdots, \lambda_n$ has r distinct values $\tau_1, \tau_2, \cdots \tau_r$

(E.g. $0 < \lambda_1 = \lambda_2 < \lambda_3 = \lambda_4 = \lambda_5$, $r = 2$

$\tau_1 = \lambda_1$, $\tau_2 = \lambda_3$)

$$Q_r(\lambda) = \frac{(-1)^r}{\tau_1 \tau_2 \cdots \tau_r} (\lambda - \tau_1)(\lambda - \tau_2) \cdots (\lambda - \tau_r)$$

$$Q_r(\tau_i) = \frac{(-1)^r}{\tau_1 \tau_2 \cdots \tau_r} (\tau_i - \tau_1) \cdots \underbrace{(\tau_i - \tau_i)}_{=0} \cdots (\tau_i - \tau_r)$$

$$= 0$$

$$Q_r(0) = \frac{(-1)^r}{\tau_1 \tau_2 \cdots \tau_r} (-\tau_1)(-\tau_2) \cdots (-\tau_r)$$

$$= \frac{\tau_1 \tau_2 \cdots \tau_r}{\tau_1 \tau_2 \cdots \tau_r}$$

$$= 1$$

$$Q_r(\lambda) = a_0 + a_1 \lambda + \cdots + a_r \lambda^r$$

$$Q_r(0) = 1 \Rightarrow a_0 = 1$$

$$\therefore \quad Q_r(\lambda) = 1 + a_1\lambda + \cdots + a_r\lambda^r$$

$$\therefore \quad Q_r(\lambda) - 1 = a_1\lambda_1 + \cdots + a_r\lambda^r$$

$$= \lambda(a_1 + a_2\lambda + \cdots + a_r\lambda^{r-1})$$

$$= \lambda\,\bar{P}_{r-1}(\lambda)$$

$$\Longleftrightarrow \quad \frac{Q_r(\lambda) - 1}{\lambda} = \bar{P}_{r-1}(\lambda)$$

$$\Longleftrightarrow \quad \boxed{Q_r(\lambda) = 1 + \lambda\,\bar{P}_{r-1}(\lambda)}$$

$$0 \le \min_{P_{r-1}} \max_{1\le i\le n} \left[1 + \lambda_i \bar{P}_{r-1}(\lambda)\right]^2$$

$$\le \max_{1\le i\le n} \left[1 + \lambda_i \bar{P}_{r-1}(\lambda)\right]^2$$

$$= \max_{1\le i\le n} \left[\underset{0}{\underset{\shortparallel}{Q_r(\lambda_i)}}\right]^2$$

$$= 0$$

$$\therefore \quad \min_{P_{r-1}} \max_{1\le i\le n} \left[1 + \lambda_i P_{r-1}(\lambda)\right]^2 = 0$$

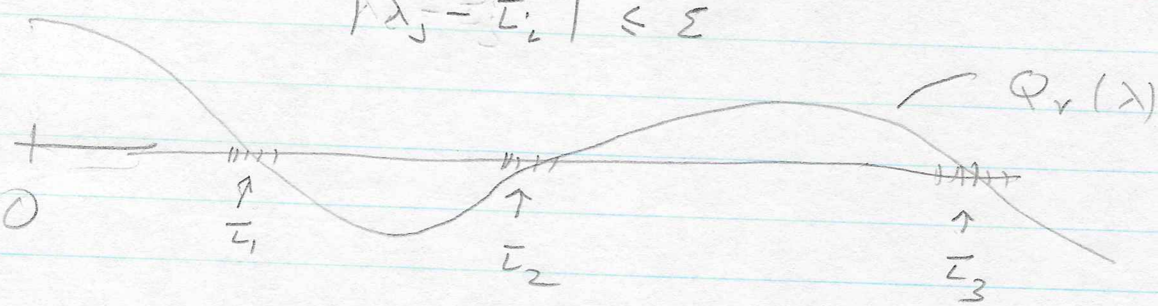
$$\Rightarrow \quad \|x_r - x^*\|_A^2 = \min_{P_{r-1}} \max_{1\le i\le n} \left[1 + \lambda_i P_{r-1}(\lambda_i)\right]^2 \|x_0 - x^*\|_A^2$$

$$= 0$$

Generalizations: Suppose you have r clusters of eigenvalues $\tau_1, \tau_2 \cdots \tau_r$

For each eigenvalue λ_j there is an i s.t.

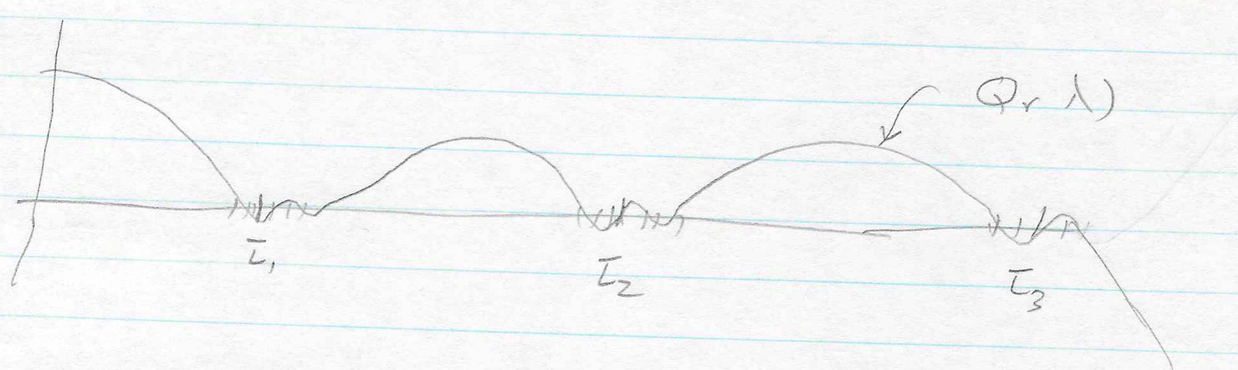
$$|\lambda_j - \tau_i| \leq \varepsilon$$



$$Q_r(\lambda) = \frac{(-1)^3}{\tau_1 \tau_2 \tau_3} (\lambda - \tau_1)(\lambda - \tau_2)(\lambda - \tau_3)$$

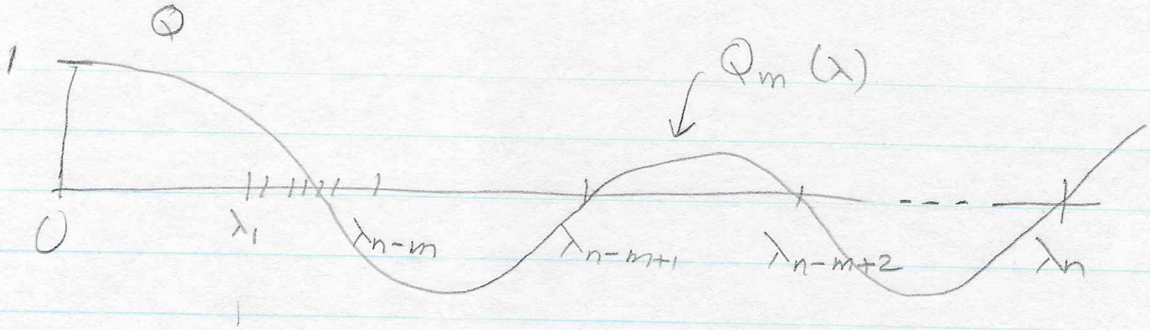
Note $|Q_r(\lambda_i)|$ is small

Could choose $Q_r(\lambda)$ of higher degree to make it smaller in the intervals $[\tau_i - \varepsilon, \tau_i + \varepsilon]$



Get $$\|x_r - x^*\|_A^2 \leq \underbrace{\max_{1 \leq i \leq n} [Q_r(\lambda_i)]^2}_{\text{can make this very small for } r \text{ not too large.}} \|x_0 - x^*\|_A^2$$

Example p. 116



$$\| x_{m+1} - x^* \|_A^2 \le \left(\frac{\lambda_{n-m} - \lambda_1}{\lambda_{n-m} + \lambda_1} \right)^2 \| x_0 - x^* \|_A^2$$

Very useful bound (5.36) p. 117

$$\| x_k - x^* \|_A \le 2 \left(\frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1} \right)^k \| x_0 - x^* \|_A \quad \text{CG}$$

$\kappa(A) = \|A\|_2 \, \|A^{-1}\|_2 = \dfrac{\lambda_n}{\lambda_1}$ is the condition number of A.

Compare to steepest descent (3.29) p. 43

$$\| x_{k+1} - x^* \|_Q^2 \le \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} \right)^2 \| x_k - x^* \|_Q^2$$

$A = Q$

$$\Rightarrow \quad \| x_k - x^* \|_A \le \left(\frac{\kappa(A) - 1}{\kappa(A) + 1} \right)^k \| x_0 - x^* \|_A \quad \text{SD}$$

Note if $K(A) = 100$

$$\frac{K(A)-1}{K(A)+1} = \frac{99}{101} \simeq 0.98$$

$$\frac{\sqrt{K(A)}-1}{\sqrt{K(A)}+1} = \frac{9}{11} \simeq 0.82$$

Also how large do you need to choose k to make

$$\left(\frac{K(A)-1}{K(A)+1}\right)^k \approx \varepsilon$$

$$\Rightarrow \log(\varepsilon) \simeq k \log\left(\frac{K(A)-1}{K(A)+1}\right)$$

$$= k \log\left(\frac{1-1/K(A)}{1+1/K(A)}\right)$$

$$= k\left(\log\left(1-\frac{1}{K(A)}\right) - \log\left(1+\frac{1}{K(A)}\right)\right)$$

note $\log 1+x \simeq x$

$$\simeq k\left(-\frac{1}{K(A)} - \frac{1}{K(A)}\right)$$

$$= \frac{-k}{2K(A)}$$

$$\therefore k_{SD} \simeq 2K(A) \log \varepsilon$$

Similar argument for CG gives

$$K_{CG} \simeq 2\sqrt{K(A)}\,\mathrm{Log}(\varepsilon)$$

$$\therefore \quad \frac{K_{SD}}{K_{CG}} \approx \sqrt{K(A)} \implies K_{SD} \simeq \sqrt{K(A)}\,K_{CG}.$$