Conjugate Gradient Method

(Recursively construct the conjugate direction vector $\{d_j\}_{j=0}^{n-1}$)

Given x_0, $\vec{d}_0 = -\vec{g}_0 = \vec{b} - Q\vec{x}_0$ (1)

$$\vec{x}_{k+1} = \vec{x}_k + \alpha_k \vec{d}_k \quad (2)$$

$$\alpha_k = -\frac{\vec{g}_k^{\,*}\vec{d}_k}{\vec{d}_k^{\,*} Q \vec{d}_k} \quad (3)$$

$$\vec{d}_{k+1} = -\vec{g}_{k+1} + \beta_k \vec{d}_k \quad (4)$$

$$\beta_k = \frac{\vec{g}_{k+1}^{\,*} Q \vec{d}_k}{\vec{d}_k^{\,*} Q \vec{d}_k} \quad (5)$$

(where $\vec{g}_k = Q\vec{x}_k - \vec{b}$) (6)

Note: only one matrix-vector multiplication: $Q\vec{d}_k$

$$\vec{g}_{k+1} = Q\vec{x}_{k+1} - \vec{b} = Q\vec{x}_k + \alpha_k Q\vec{d}_k - \vec{b} = \vec{g}_k + \alpha_k Q\vec{d}_k \quad (7)$$

C.G. theorem

(a) $\langle \vec{g}_0, \ \ , \vec{g}_k \rangle = \langle \vec{g}_0, Q\vec{g}_0, \cdots, Q^k\vec{g}_0 \rangle = K_k$ (Krylov space of $\vec{g}_0$)

(b) $\langle \vec{d}_0, \cdots, \vec{d}_k \rangle = \langle \vec{g}_0, Q\vec{g}_0, \cdots, Q^k\vec{g}_0 \rangle$

(c) $\vec{d}_k^{\,T} Q \vec{d}_i = 0 \quad \forall\ i \le k-1$ (conjugate direction)

(d) $\alpha_k = \dfrac{\vec{g}_k^{\,T}\vec{g}_k}{\vec{d}_k^{\,T} Q \vec{d}_k}$

(e) $\beta_k = \dfrac{\vec{g}_{k+1}^{\,T}\vec{g}_{k+1}}{\vec{g}_k^{\,T}\vec{g}_k}$ (can be reused from α_k)

Pf: (a),(b),(c) by induction.

for $k=1$: $\langle \vec{g}_0 \rangle = \langle \vec{g}_0 \rangle$
$\langle \vec{d}_0 \rangle = \langle \vec{g}_0 \rangle$ (by definition) both true

Assume true for k:

for $k+1$: (a) $\vec{g}_{k+1} \overset{(7)}{=} \vec{g}_k + \alpha_k Q \vec{d}_k \in K_{k+1}$

$$\underset{K_k}{\vec{g}_k \in} \qquad \underset{QK_k = K_{k+1}}{Q\vec{d}_k \in (b)}$$

$\therefore \langle \vec{g}_0, \ldots, \vec{g}_{k+1} \rangle \subseteq \langle \vec{g}_0, \ldots, Q^{k+1}\vec{g}_0 \rangle = K_{k+1}$

Claim $\vec{g}_{k+1} \notin K_k = \langle \vec{d}_0, \ldots, \vec{d}_k \rangle$

By Corollary: $\vec{g}_{k+1}^{\,*} \vec{d}_j = 0 \quad \forall j = 0, \ldots, k$

$$\Rightarrow \quad \vec{g}_{k+1} = \begin{cases} \vec{0} & (\Leftarrow Q\vec{x}_{k+1} = \vec{b} \text{ and we are done}) \\ \notin \langle \vec{d}_0, \ldots, \vec{d}_k \rangle \end{cases}$$

$\therefore \langle \vec{g}_0, \ldots, \vec{g}_{k+1} \rangle = \langle \vec{g}_0, \ldots, Q^{k+1}\vec{g}_0 \rangle = K_{k+1}$ (8)

(b) $\vec{d}_{k+1} \overset{(4)}{=} -\vec{g}_{k+1} + \beta_k \vec{d}_k \in K_{k+1} \setminus K_k$

$$-\vec{g}_{k+1} \in K_{k+1}\setminus K_k \ (8), \qquad \vec{d}_k \in K_k$$

$\Rightarrow \langle \vec{d}_0, \ldots, \vec{d}_{k+1} \rangle = K_{k+1}$

(c) $\vec{d}_{k+1} Q \vec{d}_i \overset{(4)}{=} -\vec{g}_{k+1}^{\,*} Q \vec{d}_i + \beta_k \vec{d}_k^{\,*} Q \vec{d}_i \quad i = 1, \ldots, k$

(i) $i = k$ $\quad \vec{d}_{k+1} Q \vec{d}_i \overset{(5)}{=} 0$ by def of β_k

(ii) $i < k$ $\quad \vec{d}_k^{\,*} Q \vec{d}_i = 0$ Conjugate direction

$\vec{d}_i \in K_i \Rightarrow Q\vec{d}_i \in K_{i+1} = \langle \vec{d}_0, \ldots, \vec{d}_{i+1} \rangle \quad i+1 \le k$

By Corollary: $\vec{g}_{k+1}^{\,*} \vec{d}_\ell = 0 \quad \forall \ell \le k$

$\therefore \vec{g}_{k+1}^{\,*} Q \vec{d}_i = 0 \quad \forall i \le k$

(d) $\vec{d}_k \overset{(4)}{=} -\vec{g}_k + \beta_{k-1}\vec{d}_{k-1}$

$$\vec{g}_k^{*}\vec{d}_k = -\vec{g}_k^{*}\vec{g}_k + \beta_{k-1}\underbrace{\cancel{\vec{g}_k^{*}\vec{d}_{k-1}}}_{0 \ \text{Corollary}}$$

$$\therefore \quad \alpha_k = -\frac{\vec{g}_k^{*}\vec{d}_k}{\vec{d}_k^{*}Q\vec{d}_k} = +\frac{\vec{g}_k^{T}\vec{g}_k}{\vec{d}_k^{*}Q\vec{d}_k} \quad ※ \qquad (9)$$

(e) $\vec{g}_k \in K_k = \langle d_0, .., d_k \rangle$

$$\therefore \quad \vec{g}_{k+1}^{*}\vec{g}_k = 0 \qquad (\text{By corollary})$$

By (7) $\quad Q\vec{d}_k = \dfrac{\vec{g}_{k+1} - \vec{g}_k}{\alpha_k}$

$$\therefore \quad \vec{g}_{k+1}^{*}Q\vec{d}_k = \frac{1}{\alpha_k}\left(\vec{g}_{k+1}^{*}\vec{g}_{k+1} - \vec{g}_{k+1}^{*}\vec{g}_k\right) = \frac{1}{\alpha_k}\vec{g}_{k+1}^{*}\vec{g}_{k+1}$$

$$\therefore \quad \beta_k = \frac{\vec{g}_{k+1}^{*}Q\vec{d}_k}{\vec{d}_k^{*}Q\vec{d}_k} = \frac{\vec{g}_{k+1}^{*}\vec{g}_{k+1}}{\alpha_k(\vec{d}_k^{*}Q\vec{d}_k)} \overset{(9)}{=} \frac{\vec{g}_{k+1}^{*}\vec{g}_{k+1}}{\vec{g}_k^{*}\vec{g}_k} \quad ※$$

Recall Expanding Subspace Theorem:

$$x_{k+1} = \min_{\vec{x} \in x_0 + \mathcal{B}_{k+1}} \left\{ \frac{1}{2} \vec{x}^{*} Q \vec{x} - \vec{b}^{*} \vec{x} \right\} \qquad \mathcal{B}_{k+1} = \{\vec{d}_0, \dots, \vec{d}_k\} = K_k = \{\vec{g}_0, \dots, Q^k \vec{g}_0\}$$

Let $\vec{x}_{*}$ be the solution $Q\vec{x}_{*} = \vec{b}$. The error in Q-norm:

$$\| \vec{x} - \vec{x}_{*} \|_Q^2 \equiv \frac{1}{2} (\vec{x} - \vec{x}_{*})^{*} Q (\vec{x} - \vec{x}_{*})$$

$$= \frac{1}{2} \vec{x}^{*} Q \vec{x} - \vec{x}^{*} \vec{b} + \underbrace{\frac{1}{2} \vec{x}_{*}^{*} Q \vec{x}_{*}}_{\text{fixed constant}}$$

$$\therefore \quad \vec{x}_{k+1} = \arg\min_{\vec{x} \in x_0 + \mathcal{B}_{k+1}} \| \vec{x} - \vec{x}_{*} \|_Q^2$$

$$= \arg\min_{\vec{x} \in x_0 + K_k} \| \vec{x} - \vec{x}_{*} \|_Q^2 \quad ; \quad E_Q^2(\vec{x}_{k+1}) \equiv \| \vec{x}_{k+1} - \vec{x}_{*} \|_Q^2$$

$$\vec{x}_{k+1} \overset{(2)}{=} \vec{x}_k + \alpha_k \vec{d}_k \overset{(2)}{=} \vec{x}_0 + \alpha_0 \vec{d}_0 + \dots + \alpha_k \vec{d}_k = \vec{x}_0 + \gamma_0 \vec{g}_0 + \dots + \gamma_k Q^k \vec{g}_0 = \vec{x}_0 + P_k(Q) \vec{g}_0$$

$$\langle \vec{d}_0, \dots, \vec{d}_k \rangle \overset{(a)}{=} \langle \vec{g}_0, \dots, Q^k \vec{g}_0 \rangle$$

for some $P_k(Q)$ kth degree polynomials

$$\vec{x}_{k+1} - \vec{x}_{*} = P_k(Q) \vec{g}_0 + \vec{x}_0 - \vec{x}_{*}$$

$$= P_k(Q)(Q\vec{x}_0 - \vec{b}) + \vec{x}_0 - \vec{x}_{*}$$

$$= P_k(Q)(Q\vec{x}_0 - Q\vec{x}_{*}) + \vec{x}_0 - \vec{x}_{*}$$

$$= (P_k(Q) Q + I)(\vec{x}_0 - \vec{x}_{*})$$

$$\therefore \quad \underbrace{\| \vec{x}_{k+1} - \vec{x}_{*} \|_Q^2}_{E_Q^2(\vec{x}_{k+1})} = \frac{1}{2} (\vec{x}_0 - \vec{x}_{*})^{*} Q (I + Q P_k(Q))^2 (\vec{x}_0 - \vec{x}_{*})$$

By Expanding Subspace Thm: $E_Q^2(\vec{x}_{k+1}) = \min_{P_k} \frac{1}{2} (\vec{x}_0 - \vec{x}_{*})^{*} Q (I + Q P_k(Q))^2 (\vec{x}_0 - \vec{x}_{*})$

Let $U^T Q U = \Lambda$, $U^T U = I$, $\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$

write $(\vec{x}_0 - \vec{x}^*) = \sum_{i=1}^{n} \xi_i \vec{u}_i$

$$\|\vec{x}_{k+1} - \vec{x}^*\|_Q^2 = \frac{1}{2} \sum_{i=1}^{n} \lambda_i (1 + \lambda_i P_k(\lambda_i))^2 \xi_i^2$$

$$\leq \max_{\lambda_i \in S_Q} (1 + \lambda_i P_k(\lambda_i))^2 \frac{1}{2} \sum_{i=1}^{n} \lambda_i \xi_i^2$$

S_Q : spectrum of Q

$$= \max_{\lambda_i \in S_Q} (1 + \lambda_i P_k(\lambda_i))^2 \|\vec{x}_0 - \vec{x}^*\|_Q^2$$

$$\therefore \quad E_Q^2(\vec{x}_{k+1}) \leq \min_{P_k} \max_{\lambda_i \in S_Q} (1 + \lambda_i P_k(\lambda_i))^2 E_Q^2(\vec{x}_0)$$

#

$$E_Q^2(\vec{x}_{k+1}) \leq \min_{P_{k+1},\, P_{k+1}(0)=1} \max_{\lambda_i \in S_Q} (P_{k+1}(\lambda_i))^2 E_Q^2(\vec{x}_0)$$

$$E_Q^2(\vec{x}_k) \leq \min_{P_k,\, P_k(0)=1} \max_{\lambda_i \in S_Q} (P_k(\lambda_i))^2 E_Q^2(\vec{x}_0)$$

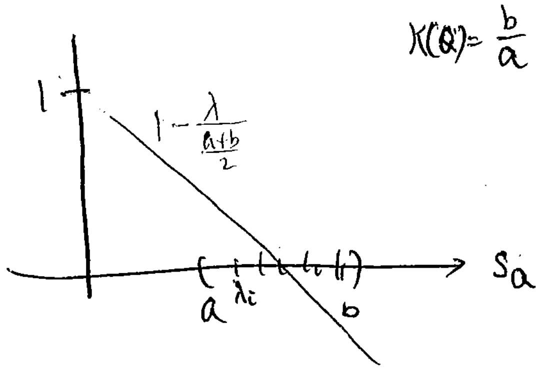
Thm.

$$\frac{\|\vec{x}_k - \vec{x}^*\|_Q}{\|\vec{x}_0 - \vec{x}^*\|_Q} \leq \min_{P_k,\, P_k(0)=1} \max_{\lambda_i \in S_Q} |P_k(\lambda_i)|$$

Eg1 Suppose $S_Q \subseteq [a, b]$

$K(Q) = \frac{b}{a}$

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



$$\max_{\lambda \in [a,b)} |P_k(\lambda)| \leq \max_{\lambda \in [a,b)} \left|1 - \frac{\lambda}{\frac{a+b}{2}}\right|^k$$

$$= \left|\frac{b-a}{a+b}\right|^k = \left|\frac{K(Q) - 1}{K(Q) + 1}\right|^k$$

$$\therefore \quad \frac{\|\vec{x}_k - \vec{x}^*\|_Q}{\|\vec{x}_0 - \vec{x}^*\|_Q} \leq \left|\frac{K(Q) - 1}{K(Q) + 1}\right|^k$$

If $K(Q) = 10^8$

Convergence is slow

The optimal estimate:

$$\min_{P_k,\, P_k(0)=1} \ \max_{a \le \lambda \le b} |P_k(\lambda)|$$

is given by
$$\bar{P}_k(\lambda) = \frac{T_k\left(\frac{b-a-2\lambda}{b-a}\right)}{T_k\left(\frac{b+a}{b-a}\right)}$$

where T_k is the kth Chebyshev polynomial defined recursively by

$$T_k(z) = 2z\, T_{k-1}(z) - T_{k-2}(z)$$

$$\text{with} \begin{cases} T_0(z) = 1 \\ T_1(z) = z \end{cases}$$

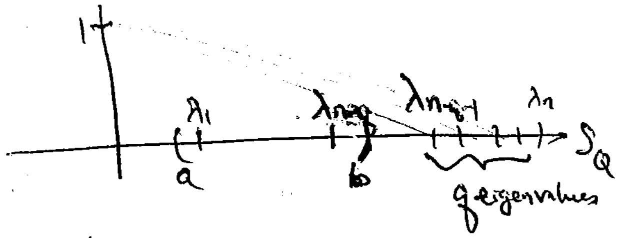
$$\max_{a \le \lambda \le b} |\bar{P}_k(\lambda)| = \frac{1}{T_k\left(\frac{b+a}{b-a}\right)} \le 2\left(\frac{\sqrt{\frac{b}{a}}-1}{\sqrt{\frac{b}{a}}+1}\right)^k$$

$$\therefore \quad \frac{\|\vec{x}_k - \vec{x}_*\|_Q}{\|\vec{x}_0 - \vec{x}_*\|_Q} \le 2\left(\frac{\sqrt{K(Q)}-1}{\sqrt{K(Q)}+1}\right)^k$$

If $K(Q) = 10^{-8}$
Convergence is still slow.

Eg 2

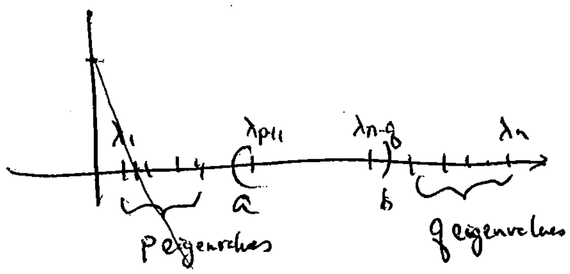
$$\frac{\|\vec{x}_k - \vec{x}_*\|_Q}{\|\vec{x}_0 - \vec{x}_*\|_Q} \le 2\left(\frac{\sqrt{\frac{b}{a}}-1}{\sqrt{\frac{b}{a}}+1}\right)^{\cancel{k-q}\, q}$$



$$\max_{\lambda \in [a,b]} \left|1 - \frac{\lambda}{\lambda_j}\right| < 1 \qquad j \ge n-q$$

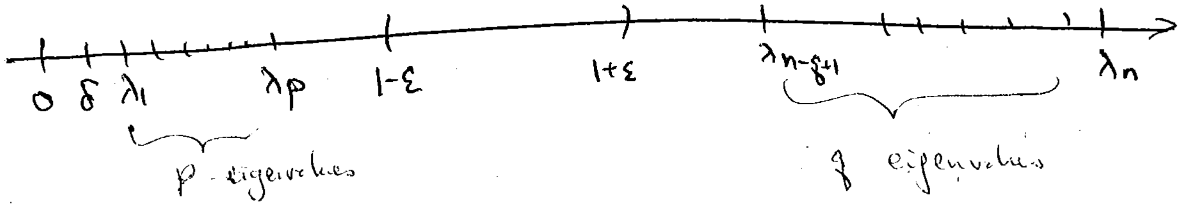
Eg 3

$$\frac{\|\vec{x}_k - \vec{x}_*\|_Q}{\|\vec{x}_0 - \vec{x}_*\|_Q} \le 2\left(\frac{\sqrt{\frac{b}{a}}-1}{\sqrt{\frac{b}{a}}+1}\right)^{k-p-q} \max_{\lambda \in [a,b]} \prod_{j=1}^{p} \left|1 - \frac{\lambda}{\lambda_j}\right|$$



$$\max_{\lambda \in [a,b]} \left|1 - \frac{\lambda}{\lambda_j}\right| \text{ can be large}$$
$$j \le p$$

Clustering of eigenvalues



$$\frac{\|\vec{X}_k-\vec{X}_*\|_Q}{\|\vec{X}_0-\vec{X}_*\|_Q} \le 2\left(\frac{\sqrt{\frac{1+\varepsilon}{1-\varepsilon}}-1}{\sqrt{\frac{1+\varepsilon}{1-\varepsilon}}+1}\right)^{k-p-q} \max_{\lambda\in[1-\varepsilon,\,1+\varepsilon]} \prod_{j=1}^{p}\left|\frac{\lambda_j-\lambda}{\lambda_j}\right|$$

$$\left(\frac{\sqrt{\frac{1+\varepsilon}{1-\varepsilon}}-1}{\sqrt{\frac{1+\varepsilon}{1-\varepsilon}}+1}\right) = \frac{1-\sqrt{1-\varepsilon^2}}{\varepsilon} < \varepsilon$$

and for $\lambda\in[1-\varepsilon,1+\varepsilon]$, & $\lambda_j \ge \delta$

$$\left|\frac{\lambda_j-\lambda}{\lambda_j}\right| \le \frac{1+\varepsilon}{\delta}$$

$\therefore$ $$\frac{\|\vec{X}_k-\vec{X}_*\|_Q}{\|\vec{X}_0-\vec{X}_*\|_Q} \le 2\,\varepsilon^{k-p-q}\cdot\left(\frac{1+\varepsilon}{\delta}\right)^{p}$$

\#