Notes for Q+A Session 7 (Monday, March 30)

$$p_k = -r_k + \beta_k p_{k-1} \qquad (5.13)$$

Recall $\quad r_k = Ax_k - b \quad$ (r_k is the "residual" associated with x_k)

Recall also that

$$\varphi(x) = \frac{1}{2} x^T A x - b^T x$$

$$\therefore \quad \nabla \varphi(x) = Ax - b$$

$$\therefore \quad r_k = Ax_k - b = \nabla \varphi(x_k)$$

Want p_k + p_{k-1} conjugate : $\quad p_{k-1}^T A p_k = 0$

So from (5.13)

$$0 = p_{k-1}^T A p_k = p_{k-1}^T A(-r_k + \beta_k p_{k-1})$$

$$= -p_{k-1}^T A r_k + \beta_k p_{k-1}^T A p_{k-1}$$

$$\Rightarrow \quad \beta_k = \frac{p_{k-1}^T A r_k}{p_{k-1}^T A p_{k-1}} = \frac{r_k^T A p_{k-1}}{p_{k-1}^T A p_{k-1}}$$

Bottom of p. 111

<u>Claim:</u> $$\alpha_k = \frac{r_k^T r_k}{p_k^T A p_k}$$

Alg. 5.1 $$\alpha_k = \frac{-r_k^T p_k}{p_k^T A p_k}$$

$$-r_k^T p_k = -r_k^T(-r_k + \beta_k p_{k-1}) \qquad \text{(by (5.14e))}$$

$$= r_k^T r_k - \beta_k r_k^T p_{k-1}$$

But $r_k^T p_{k-1} = 0$ by (5.11) in Thm 5.2

$$\therefore \quad -r_k^T p_k = r_k^T r_k$$

$$\therefore \quad \alpha_k = \frac{-r_k^T p_k}{p_k^T A p_k} = \frac{r_k^T r_k}{p_k^T A p_k}$$

Note: $\alpha_k = 0 \Rightarrow r_k = 0 \Rightarrow \nabla \phi(x_k) = 0$

$\Rightarrow x_k$ is the minimizer of $\phi(x)$ so stop.

$\therefore$ If we don't stop, $\alpha_k \neq 0$

Claim: $$\beta_{k+1} = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$$

Alg 5-1: $$\beta_{k+1} = \frac{r_{k+1}^T A p_k}{p_k^T A p_k}$$

$$r_{k+1} = r_k + \alpha_k A p_k \implies A p_k = \frac{r_{k+1} - r_k}{\alpha_k} \qquad (\alpha_k \neq 0)$$

$$\therefore \quad r_{k+1}^T A p_k = r_{k+1}^T \left(\frac{r_{k+1} - r_k}{\alpha_k} \right)$$

$$= \frac{r_{k+1}^T r_{k+1}}{\alpha_k} \qquad \left(\text{since } r_{k+1}^T r_k = 0 \text{ by } (5.16) \right)$$

$$p_k^T A p_k = p_k^T \left(\frac{r_{k+1} - r_k}{\alpha_k} \right)$$

$$= -\frac{p_k^T r_k}{\alpha_k} \qquad \left(\text{since } p_k^T r_{k+1} = 0 \text{ by } (5.11) \right)$$

$$= -\frac{(-r_k + \beta_k p_{k-1})^T r_k}{\alpha_k} \qquad (\text{by } (5.14e))$$

$$= \frac{r_k^T r_k}{\alpha_k} \qquad \left(\text{since } p_{k-1}^T r_k = 0 \text{ by } (5.11) \right)$$

$$\therefore \quad \beta_{k+1} = \frac{r_{k+1}^T A p_k}{p_k^T A p_k} = \frac{r_{k+1}^T r_{k+1} / \alpha_k}{r_k^T r_k / \alpha_k} = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$$

<u>Claim:</u> $\frac{1}{2}\|x-x^*\|_A^2 = \frac{1}{2}(x-x^*)^T A (x-x^*) = \varphi(x) - \varphi(x^*)$

Where x^* is the minimizer of $\varphi(x)$

$\Rightarrow \nabla\varphi(x^*) = 0 \Rightarrow Ax^* = b$

$\frac{1}{2}(x-x^*)^T A (x-x^*) - \varphi(x) + \varphi(x^*)$

$= \frac{1}{2}x^TAx - x^TAx^* + \frac{1}{2}(x^*)^TAx^*$

$\quad - \left(\frac{1}{2}x^TAx - b^Tx\right) + \left(\frac{1}{2}(x^*)^TAx^* - b^Tx^*\right)$

$= -x^TAx^* + \frac{1}{2}(x^*)^TAx^* + b^Tx + \frac{1}{2}(x^*)^TAx^* - b^Tx^*$

(replace Ax^* by b)

$= -x^Tb + \frac{1}{2}(x^*)^Tb + b^Tx + \frac{1}{2}(x^*)^Tb - b^Tx^*$

$= 0$

$$x_{k+1} = \underset{x \in x_0 + \text{span}\{p_0, p_1, \cdots p_k\}}{\text{argmin}} \ \varphi(x)$$

$$= \underset{x \in x_0 + \text{span}\{p_0, p_1, \cdots p_k\}}{\text{argmin}} \ \tfrac{1}{2}\| x - x^* \|_A^2$$

$$= \underset{x \in x_0 + \text{span}\{p_0, p_1, \cdots p_k\}}{\text{argmin}} \ \| x - x^* \|_A$$

$$= \underset{x \in x_0 + \text{span}\{r_0, Ar_0, \cdots A^k r_0\}}{\text{argmin}} \ \| x - x^* \|_A$$

$$= \underset{P_k \in \mathcal{P}_k}{\text{argmin}} \ \| \underbrace{x_0 + P_k(A) r_0}_{=x} - x^* \|_A$$

$\mathcal{P}_k$ a polynomial of degree $\leq k$

Now note $\quad r_0 = Ax_0 - b = Ax_0 - Ax^* = A(x_0 - x^*)$

$$\therefore \quad x_0 + P_k(A) r_0 - x^* = \left(I + P_k(A) A\right)(x_0 - x^*)$$

$$\therefore \quad x_{k+1} = \underset{\bar{P}_k \in \mathcal{P}_k}{\text{argmin}} \ \| \bar{P}_k(A)(x_0 - x^*) \|_A$$

$\mathcal{P}_k = \{$ polynomials of the form $1 + x\, p_k(x)$ where $p_k(x)$ is a polynomial of deg $\leq k$ $\}$

$$1 + x\, p_k(x) = 1 + \gamma_0 x + \cdots + \gamma_k x^{k+1}$$

A SPD eigenvalues $0 < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$
orthonormal eigenvectors $v_1, v_2, \cdots v_n$

$$A v_i = \lambda v_i \qquad v_i^T v_j = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{if } i \ne j \end{cases}$$

Essentially same as $Q^T A Q = \Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ & & \ddots & \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$

$$Q = [v_1\ v_2 \cdots v_n]$$

Outer product form

$$A = \sum_{i=1}^{n} \lambda_i v_i v_i^T$$

Why: $A v_k = \lambda_k v_k$

$$\left(\sum_{i=1}^{n} \lambda_i v_i v_i^T\right) v_k = \sum_{i=1}^{n} \lambda_i v_i \underbrace{(v_i^T v_k)}_{= \begin{cases} 1 & \text{if } i=k \\ 0 & \text{if } i \ne k \end{cases}} = \lambda_k v_k$$

If $A v_i = B v_i$ for $i = 1, \cdots, n$

and $\{v_1, v_2, \cdots v_n\}$ is a basis for R^n

then must have $Ax = Bx$ for all $x \in R^n$

$$\Rightarrow A = B$$

Note eigenvalues $\{v_1, v_2, \cdots v_n\}$ is a basis for R^n.

$$P_k(A)\, v_i = P_k(\lambda_i)\, v_i$$ for eigenvector v_i of A

$$A v_i = \lambda_i v_i$$

<u>Why</u>? First note $A^j v_i = \lambda_i^j v_i$

Induction on j

$j = 1$ $\quad A v_i = \lambda_i v_i$

Assume true for j and consider

$$\begin{aligned}
A^{j+1} v_i &= A\,(A^j v_i) \\
&= A(\lambda_i^j v_i) \\
&= \lambda_i^j A v_i \\
&= \lambda_i^j (\lambda_i v_i) \\
&= \lambda_i^{j+1} v_i
\end{aligned}$$

$$\begin{aligned}
\therefore \quad P_k(A)\, v_i &= \left(\sum_{j=0}^{k} \alpha_j A^j \right) v_i \\
&= \sum_{j=0}^{k} \alpha_j A^j v_i \\
&= \sum_{j=0}^{k} \alpha_j \lambda_i^j v_i \\
&= \left(\sum_{j=0}^{k} \alpha_j \lambda_i^j \right) v_i \\
&= P_k(\lambda_i)\, v_i
\end{aligned}$$

$$x_{k+1} - x^* = \left(I + P_k^*(A) A \right)(x_0 - x^*) \qquad (5.30)$$

Let $\quad x_0 - x^* = \sum_{i=1}^{n} \xi_i v_i$

$$\therefore \quad x_{k+1} - x^* = \left(I + P_k^*(A) A \right)(x_0 - x^*)$$

$$= \left(I + P_k^*(A) A \right) \sum_{i=1}^{n} \xi_i v_i$$

$$= \sum_{i=1}^{n} \left(I + P_k^*(A) A \right) \xi_i v_i$$

$$= \sum_{i=1}^{n} \left(v_i + P_k^*(A) A v_i \right) \xi_i$$

$$= \sum_{i=1}^{n} \left(v_i + P_k^*(\lambda_i) \lambda_i v_i \right) \xi_i$$

$$= \sum_{i=1}^{n} \left(1 + P_k^*(\lambda_i) \lambda_i \right) \xi_i v_i$$

$$\| z \|_A^2 = z^T A z = z^T \left(\sum_{i=1}^{n} \lambda_i v_i v_i^T \right) z$$

$$= \sum_{i=1}^{n} \lambda_i (z^T v_i)(v_i^T z)$$

$$= \sum_{i=1}^{n} \lambda_i (v_i^T z)^2$$

$$\therefore \quad \| x_{k+1} - x^* \|_A^2 = (x_{k+1} - x^*)^T A \underbrace{(x_{k+1} - x^*)}_{= z = \sum_{i=1}^{n} (1 + P_k^*(\lambda_i) \lambda_i) \xi_i v_i}$$

$$= \sum_{i=1}^{n} \lambda_i \left(1 + \lambda_i P_k^*(\lambda_i) \right)^2 \xi_i^2$$

$$\therefore \quad \|x_{k+1} - x^*\|_A^2 = \min_{P_k} \sum_{i=1}^{n} (1 + P_k(\lambda_i)\lambda_i)^2 \lambda_i \xi_i^2$$

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

Recall $\quad x_0 - x^* = \sum_{i=1}^{n} \xi_i v_i$

$$\therefore \quad \|x_0 - x^*\|_A^2 = \left(\sum_{i=1}^{n} \xi_i v_i\right)^T A \left(\sum_{i=1}^{n} \xi_i v_i\right)$$

$$= \left(\sum_{i=1}^{n} \xi_i v_i\right)^T \left(\sum_{i=1}^{n} \xi_i A v_i\right)$$

$$= \left(\sum_{i=1}^{n} \xi_i v_i\right)^T \left(\sum_{i=1}^{n} \xi_i \lambda_i v_i\right)$$

$$= \sum \lambda_i \xi_i^2$$

$$\therefore \quad \|x_{k+1} - x^*\|_A^2 \leq \left(\min_{P_k} \max_{1 \leq i \leq n} (1 + P_k(\lambda_i)\lambda_i)^2\right) \|x_0 - x^*\|_A^2$$

(5.33) in book

This is the <u>key</u> result.

Speed of Convergence of CG depends on

$$\min_{P_k} \max_{1 \leq i \leq n} (1 + P_k(\lambda_i)\lambda_i)^2$$