CS475 / CM375
Lecture 24: Dec 1, 2011

Conjugate gradient convergence

Minimization formulation

- Recall: $F(x) \equiv \frac{1}{2}x^T A x - b^T x$

$$x^* = argmin_x\, F(x) \Longrightarrow Ax^* = b$$

- At iteration k

$$F\left(x^k\right) = \min\left\{F\left(x^0 + \sum_{i=0}^{k-1} \alpha_i p^i\right) : \alpha_0, \dots, \alpha_{k-1} \in \Re\right\}$$

$$F\left(x^k\right) - F(x) = \min\left\{F\left(x^0 + \sum_{i=0}^{k-1} \alpha_i p^i\right) - F(x) : \alpha_0, \dots, \alpha_{k-1} \in \Re\right\}$$

$$\vdots$$

Algebra $\left(F(y) - F(x) = \frac{1}{2}\left|\left|y - x\right|\right|_A^2\right)$

$$\vdots$$

$$\frac{1}{2}\left|\left|x^k - x\right|\right|_A^2 = \min\left\{\frac{1}{2}\left|\left|e^0 + \sum_{i=0}^{k-1} \alpha_i p^i\right|\right|_A^2 : \alpha_0, \dots, \alpha_{k-1} \in \Re\right\}$$

Error Characterization

- Characterization of the error e^k at step k:
- Let $e^k = x^k - x$. Then

$$\left|\left|e^k\right|\right|_A = \min\left\{\left|\left|e^0 + \sum_{i=0}^{k-1} \alpha_i p^i\right|\right|_A : \alpha_0, \dots, \alpha_{k-1} \in \Re\right\}$$
$$= \min\left\{\left|\left|e^0 + \sum_{i=0}^{k-1} \gamma_i A^i r^0\right|\right|_A : \gamma_0, \dots, \gamma_{k-1} \in \Re\right\}$$

(since $span\{p^0, \dots, p^{k-1}\} = span\{r^0, \dots, A^{k-1}r^0\}$)

Error determined by Polynomial

- Let $Q_{k-1}(x) = \gamma_0 + \gamma_1 x + \cdots + \gamma_{k-1}x^{k-1}$. Then $Q_{k-1}(A) = \gamma_0 + \gamma_1 A + \cdots + \gamma_{k-1}A^{k-1} = \sum_{i=0}^{k-1} \gamma_i A^i$
- Rewrite error as a function of the Q_{k-1} polynomial

$$e^0 + \sum_{i=0}^{k-1} \gamma_i A^i r^0 = e^0 + Q_{k-1}(A) r^0$$
$$= e^0 + Q_{k-1}(A) A e^0 \qquad (2)$$
$$= (I + Q_{k-1}(A)A)e^0$$

For (2): $r^0 = b - Ax^0 = A(x - x^0) = Ae^0$

Error Determined by Polynomial

- Let $P_k(x) = 1 + Q_{k-1}(x)x$. Then $P_k(0) = 1$, $\deg(P_k) \leq k$ and $e^0 + \sum_{i=0}^{k-1} \gamma_i A^i r^0 = P_k(A)e^0$
  $\therefore \left\|e^k\right\|_A = \min\left\{\left\|P_k(A)e^0\right\|_A : P_k(0) = 1, \deg(P_k) \leq k\right\}$
- In other words, let $\tilde{P}_k(x)$ be a poly of $deg \leq k$, $\tilde{P}_k(o) = 1$. Then $\left\|e^k\right\|_A \leq \left\|\tilde{P}_k(A)e^0\right\|_A$

$\therefore$ CG finds the optimal poly $\bar{P}_k$ to minimize the error in the A-norm.

Polynomial characterization

- What is $\left\|P_k(A)e^0\right\|_A$?
- Let $\{v_j\}$ be the eigenvectors of A. Write
  $$e^0 = \sum_{j=1}^{n} \xi_j v_j$$
- Consider $w_k = P_k(A)e^0 = \sum_{j=1}^{n} \xi_j P_k(A) v_j$
  $$= \sum_{j=1}^{n} \xi_j P_k(\lambda_j) v_j \qquad (Av_j = \lambda_j v_j)$$
  $$Aw_k = \sum_{j=1}^{n} \xi_j P_k(\lambda_j) \lambda_j v_j$$

Polynomial Characterization

- Hence:

$$\left||P_k(A)e^0|\right|_A^2 = \left||w_k|\right|_A^2 = (w_k, Aw_k)$$
$$= \left(\sum_{i=1}^n \xi_i P_k(\lambda_i) v_i, \sum_{j=1}^n \xi_j P_k(\lambda_j) \lambda_j v_j\right)$$
$$= \sum_{j=1}^n \xi_j^2 P_k^2(\lambda_j) \lambda_j$$

NB: $(v_i, v_j) = 1, (v_j, v_j) = 1$

Error Bound

- Let $P_k(x)$ be a poly of $deg \leq k$, $P_k(0) = 1$. Then

$$\left||e^k|\right|_A^2 \leq \left||P_k(A)e^0|\right|_A^2$$
$$= \sum_{j=1}^n \xi_j^2 P_k^2(\lambda_j) \lambda_j$$
$$\leq \max_{\lambda_{min} \leq \lambda \leq \lambda_{max}} P_k^2(\lambda) \sum_j \xi_j^2 \lambda_j$$
$$= \max_{\lambda_{min} \leq \lambda \leq \lambda_{max}} P_k^2(\lambda) \left||e^0|\right|_A^2$$

Error Bound

- By Choosing $P_k(x)$ appropriately, one can show:
- <u>Theorem:</u> $\left|\left|x^k - x\right|\right|_A \leq 2\left(\frac{\sqrt{\kappa(A)}-1}{\sqrt{\kappa(A)}+1}\right)^k \left|\left|x^0 - x\right|\right|_A$

where $\kappa(A) = \frac{\lambda_{max}}{\lambda_{min}}$

Notes

1. This is an upper bound only, CG convergence is usually better
2. CG convergence depends on all $\{\lambda_j\}$, not just λ_{min}, λ_{max}.

   E.g. A has 3 distinct eigenvalues: $\lambda_1 < \lambda_2 < \lambda_3$.

   Define Lagrange poly $P_3(x)$ of $deg \leq 3$ such that

   $P_3(0) = 1,\ P_3(\lambda_j) = 0 \quad j = 1,2,3$

   Then $\left|\left|e^3\right|\right|_A^2 \leq \left|\left|P_3(A)e^0\right|\right|_A^2 = \sum_{j=1}^{3} \xi_j^2 P_3^2(\lambda_j)\lambda_j = 0$

   $\Longrightarrow$ CG converges in 3 iterations, independent of $\kappa(A)$.

Notes

3. For Poisson equation, convergence rates for optimal SOR and CG are the same. However, no parameter needs to be optimized for CG.