Lecture-14

Rate of Convergence of CG

Algorithm 5.2

Given x_0;

set $r_0 \leftarrow Ax_0 - b,\ \ p_0 \leftarrow -r_0, k \leftarrow 0$

While $r_k \neq 0$

$$a_k \leftarrow -\frac{r_k^T r_k}{p_k^T A p_k};$$

$$x_{k+1} \leftarrow x_k + a_k p_k;$$

$$r_{k+1} \leftarrow r_k + a_k A p_k;$$

$$b_{k+1} \leftarrow \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k};$$

$$p_{k+1} \leftarrow -r_{k+1} + b_{k+1} p_k;$$

$$k \quad \leftarrow k+1;$$

end(*while*)

We only need to know values of *x*, *p* and *r* only for 2 iterations.

Major computations: matrix-vector product, two inner products, and three vector sums.

Key points

- According to theorem 5.3 Algorithm 5.2 should converge at most n steps.
- Convergence less than n iterations, depending on the eigenvalues of matrix A.
- If A dose not have favorable eigenvalues, then precondition A to get faster convergence.

Theorem 5.4

If A has only r distinct eigenvalues, then the CG iteration will terminate at the solution in at most r iterations.

Main points

Want to show:

$$\| x_{k+1} - x^* \|_A^2 \leq \min_{P_k} \max_{1 \leq i \leq n} [1 + l_i P_k(l_i)]^2 \| x_0 - x^* \|_A^2$$

Use this:

(Theorem 5.3)

Define polynomial $P_k^*(A) = g_0 I + g_1 A + \ldots + g_k A^k$

Use orthogonal eigenvectors n_i of A.

Show n_i are also eigenvectors of $P_k(A)$

Rate of Convergence

$$x_{k+1} = x_0 + a_0 p_0 + \ldots + a_k p_k$$

By construction

$$x_{k+1} = x_0 + g_0 r_0 + g_1 A r_0 + \ldots + g_k A^k r_0$$

(Theorem 5.3)

Define polynomial:

$$P_k^*(A) = g_0 I + g_1 A + \ldots + g_k A^k$$

Therefore $x_{k+1} = x_0 + P_k^*(A) r_0$ (D)

Now

$$\frac{1}{2} \| x - x^* \|_A^2 = f(x) - f(x^*)$$

$$\| z \|_A^2 = z^T A z$$

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

Rate of Convergence

$$\frac{1}{2}\|x-x^*\|_A^2=\frac{1}{2}(x-x^*)^T A(x-x^*)$$
$$=\frac{1}{2}(x^T-x^{*T})(Ax-Ax^*)$$
$$=\frac{1}{2}x^T Ax-\frac{1}{2}x^{*T}Ax-\frac{1}{2}x^T Ax^*+\frac{1}{2}x^{*T}Ax^*$$
$$=\frac{1}{2}x^T Ax-\frac{1}{2}b^T x-\frac{1}{2}x^T b+\frac{1}{2}x^{*T}Ax^*$$
$$=\frac{1}{2}x^T Ax-\frac{1}{2}b^T x-\frac{1}{2}x^T b+x^{*T}Ax^*-\frac{1}{2}x^{*T}Ax^*$$
$$=\frac{1}{2}x^T Ax-\frac{1}{2}b^T x-\frac{1}{2}x^T b+x^{*T}b-\frac{1}{2}x^{*T}Ax^*$$
$$=\frac{1}{2}x^T Ax-b^T x-(\frac{1}{2}x^{*T}Ax^*-b^T x^*)$$
$$=f(x)-f(x^*)$$

$$\|z\|_A^2=z^T Az$$

$$f(x)=\frac{1}{2}x^T Ax-b^T x$$

$$\frac{1}{2}\|x-x^*\|_A^2=\frac{1}{2}(x-x^*)^T A(x-x^*)=f(x)-f(x^*)$$

$$x_{k+1}=x_0+a_0 p_0+\ldots+a_k p_k$$

By construction

$$f(x)=\frac{1}{2}x^T Ax-b^T x$$

According to Theorem 5.2 x_{k+1} minimizes f, hence $\|x-x^*\|_A^2$

Or

$$\|x_0+P_k^*(A)r_0-x^*\|_A^2$$

$$x_{k+1}=x_0+P_k^*(A)r_0$$ From (D)

Therefore, P_k^* solves the following problem:

$$\min_{P_k}\|x_0+P_k(A)r_o-x^*\|_A$$

We know $r_0 = Ax_0 - b = Ax_0 - Ax^* = A(x_0 - x^*)$

$$\begin{aligned} x_{k+1} - x^* &= x_0 + P_k^*(A)r_0 - x^* \\ &= (x_0 - x^*) + P_k^*(A)r_0 \\ &= (x_0 - x^*) + P_k^*(A)A(x_0 - x^*) \\ &= [I + P_k^*(A)A](x_0 - x^*) \qquad \text{(A)} \end{aligned}$$

$x_{k+1} = x_0 + P_k^*(A)r_0$
From (D)

Assume n_i, l_i are eigenvectors & eigenvalues of A

$$x_0 - x^* = \sum_{i=l}^{n} x_i v_i$$

Show n_i are also eigenvectors of $P_k(A)$

$$P_k^*(A) = g_0 I + g_1 A + \ldots + g_k A^k$$

$$P_k(A)n_i = g_0 I n_i + g_1 A n_i + g_2 A^2 n + \ldots + g_k A^k n_i$$

$$P_k(A)n_i = g_0 n_i + g_1 l_i n_i + g_2 l_i A n_i \ldots + g_k A^{k-1} l_i n_i$$

$$P_k(A)n_i = g_0 n_i + g_1 l_i n_i + g_2 l_i^2 n_i \ldots + g_k A^{k-2} l_i^2 n_i$$

$$P_k(A)n_i = g_0 n_i + g_1 l_i n_i + g_2 l_i^2 n_i + \ldots + g_k l_i^k n_i$$

$$P_k(A)n_i = (g_0 + g_1 l_i + g_2 l_i^2 \ldots + g_k l_i^k) n_i$$

$$P_k(A)n_i = P(l_i) n_i$$

Therefore $P_k(A)n_i = P(l_i)n_i \quad \text{for } i = 1,2,\dots,n$

We know $x_0 - x^* = \sum_{i=l}^{n} x_i v_i$

$$x_{k+1} - x^* = [I + P_k^*(A)A](x_0 - x^*) \qquad \text{From (A)}$$

$$x_{k+1} - x^* = \sum_{i=l}^{n} [I + P_k^*(A)A] x_i v_i$$

$$x_{k+1} - x^* = \sum_{i=l}^{n} [x_i v_i + P_k^*(A) A x_i v_i]$$

$$x_{k+1} - x^* = \sum_{i=l}^{n} [x_i v_i + P_k^*(A) l_i x_i v_i]$$

$$x_{k+1} - x^* = \sum_{i=l}^{n} [x_i v_i + l_i P_k^*(l_i) x_i v_i]$$

$$x_{k+1} - x^* = \sum_{i=l}^{n} [1 + l_i P_k^*(l_i)] x_i v_i$$

$$x_{k+1} - x^* = \sum_{i=l}^{n} [1 + l_i P_k^*(l_i)] x_i v_i$$

$$\| x_{k+1} - x^* \|_A^2 = \left(\sum_{i=l}^{n} [1 + l_i P_k^*(l_i)] x_i n_i \right)^T A \left(\sum_{i=l}^{n} [1 + l_i P_k^*(l_i)] x_i n_i \right)$$

$$\| x_{k+1} - x^* \|_A^2 = \left(\sum_{i=l}^{n} [1 + l_i P_k^*(l_i)] x_i n_i \right)^T \left(\sum_{i=l}^{n} [1 + l_i P_k^*(l_i)] x_i A n_i \right)$$

$$\| x_{k+1} - x^* \|_A^2 = \left(\sum_{i=l}^{n} [1 + l_i P_k^*(l_i)] x_i n_i \right)^T \left(\sum_{i=l}^{n} [1 + l_i P_k^*(l_i)] x_i l_i n_i \right)$$

$$\| x_{k+1} - x^* \|_A^2 = \sum_{i=l}^{n} l_i [1 + l_i P_k^*(l_i)]^2 x_i^2 \qquad \text{Orthogonal eigenvectors}$$

$$\| x_{k+1} - x^* \|_A^2 = \sum_{i=l}^{n} l_i [1 + l_i P_k^*(l_i)]^2 x_i^{\ 2}$$

Since polynomial generated by GC is optimal

$$\| x_{k+1} - x^* \|_A^2 = \min_{P_k} \sum_{i=1}^{n} l_i [1 + l_i P_k(l_i)]^2 x_j^{\ 2}$$

$$\| x_{k+1} - x^* \|_A^2 \le \min_{P_k} \max_{1 \le i \le n} [1 + l_i P_k(l_i)]^2 \left(\sum_{j=1}^{n} l_j x_j^{\ 2} \right)$$

(C) $$\| x_{k+1} - x^* \|_A^2 \le \min_{P_k} \max_{1 \le i \le n} [1 + l_i P_k(l_i)]^2 \| x_0 - x^* \|_A^2$$

$$x_0 - x^* = \sum_{i=l}^{n} x_i v_i$$

(B) $$\min_{P_k} \max_{1 \le i \le n} [1 + l_i P_k(l_i)]^2$$

Convergence

$$x_0 - x^* = \sum_{i=1}^{n} x_i v_i$$

$$\| x_0 - x^* \|_A^2 = \sum_{i=1}^{n} x_i v_i^T A \sum_{i=1}^{n} x_i v_i$$

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

$$= \sum_{i=1}^{n} x_i v_i^T \sum_{i=1}^{n} x_i l_i v_i$$

$$= \sum_{i=1}^{n} x_i^{\ 2} l_i$$

Theorem 5.4

If A has only r distinct eigenvalues, then the CG iteration will terminate at the solution in at most r iterations.

Proof

Define polynomial:

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

$Q_r(\lambda_i) = 0$ for $i = 1,2,\ldots,n$

$Q_r(0) = 1$

$Q_r(\lambda) - 1$ Is polynomial of degree r with root at $\lambda = 0$

$\tilde{P}_{r-1} = \frac{(Q_r(\lambda) - 1)}{\lambda}$ Degree r-1

$$\min_{P_k} \max_{1 \le i \le n} [1 + \lambda_i P_k(\lambda_i)]^2 \quad \text{(B)}$$

$$0 \le \min_{P_{r-1}} \max_{1 \le i \le n} [1 + \lambda_i P_{r-1}(\lambda_i)]^2 \le \max_{1 \le i \le n} [1 + \lambda_i \tilde{P}_{r-1}(\lambda_i)]^2 = \max_{1 \le i \le n} Q_r(\lambda_i) = 0$$

$$\min_{P_{r-1}} \max_{1\le i\le n} [1 + l_i P_{r-1}(l_i)]^2 = 0$$ For k=r-1

From (C)

$$\| x_{k+1} - x^* \|_A^2 \le \min_{P_k} \max_{1\le i\le n} \left[1 + l_i P_k(l_i)\right]^2 \| x_0 - x^* \|_A^2 = 0$$

$$\| x_r - x^* \|_A^2 = 0$$

Therefore

$$x_r = x^*$$

QED

Theorem 5.5

If A has distinct eigenvalues $l_1 \le l_2 \le \ldots \le l_n$ we have

$$\| x_{k+1} - x^* \|_A^2 \le \left(\frac{l_{n-k} - l_1}{l_{n-k} + l_1} \right)^2 \| x_0 - x^* \|_A^2$$

Eigenvalues

$l_1, \ldots, l_{n-k}, l_{n-k+1}, \ldots, l_n$

Eigenvalues

$l_1, \ldots, l_{n-k}, l_{n-k+1}, \ldots, l_n$

Select polynomial $\overline{P}_k(l)$ of degree k such that

$Q_{k+1}(l) = 1 + l\overline{P}_k(l)$ Has roots at k largest eigenvalues $l_n, l_{n-1}, \ldots, l_{n-k+1}$

As well as at mid point l_1 and l_{n-k}

Maximum value attained by Q on the remaining eigenvalues is precisely $\left(\frac{l_{n-k} - l_1}{l_{n-k} + l_1}\right)$

(C) $\| x_{k+1} - x^* \|_A^2 \le \min_{P_k} \max_{1 \le i \le n} \left[1 + l_i P_k(l_i)\right]^2 \| x_0 - x^* \|_A^2$

$$\| x_{k+1} - x^* \|_A^2 \le \left(\frac{l_{n-k} - l_1}{l_{n-k} + l_1}\right)^2 \| x_0 - x^* \|_A^2$$

Example

$\{l_1, \ldots, l_{n-m}\}\{l_{n-m+1}, \ldots, l_n\}$

m largest eigenvalues

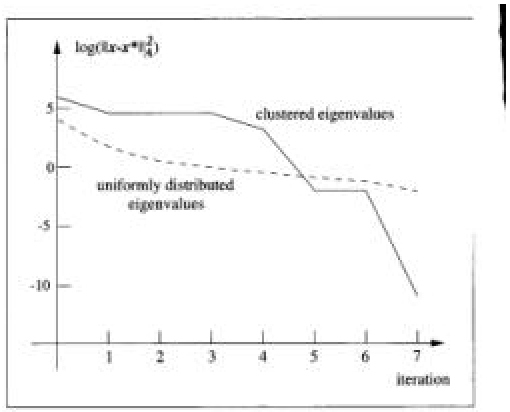
Clustered around 1

$$\| x_{m+1} - x^* \|_A \approx e \| x_0 - x^* \|_A$$

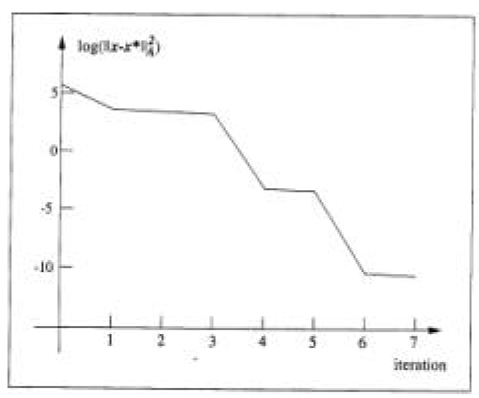
For small value of e CG will converge in only $m+1$ steps.

$$\| x_{k+1} - x^* \|_A^2 \le \left(\frac{l_{n-k} - l_1}{l_{n-k} + l_1}\right)^2 \| x_0 - x^* \|_A^2$$

Example



The matrix has five large eigenvalues with all smaller eigenvalues Clustered around .95 and 1.05



N=14, has four clusters of eigenvalues: single eigenvalues at 140, 120, a cluster of 10 eigenvalues very close to 10 with the remaining eigenvalues clustered between .95 and 1.05.

Convergence Using L2 norm

$$\| x_{k+1} - x^* \|_A \le \left(\frac{\sqrt{k(A)} - 1}{l\sqrt{k(A)} + 1} \right)^2 \| x_0 - x^* \|_A$$

$$k(A) |=\| A \|_2 \| A^{-1} \|_2 = \frac{l_1}{l_n}$$

Convergence Rate of Steepest Descent: Quadratic Function

$$\| x_{k+1} - x^* \|_Q^2 \le \left(\frac{l_n - l_1}{l_n + l_1} \right)^2 \| x_k - x^* \|_Q^2$$

Theorem 3.3

As the condition number increases the contours of the quadratic become more elongated, the zigzags of line search becomes more pronounced.