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 $r_k \neq 0$

$$\alpha_k \leftarrow \frac{r_k^T r_k}{p_k^T Ap_k}$$

$$x_{k+1} \leftarrow x_k + \alpha_k p_k$$

$$r_{k+1} \leftarrow r_k + \alpha_k Ap_k$$

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

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

$$k \leftarrow k + 1;$$

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:

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

Use this: 

\[ \text{(Theorem 5.3)} \]

Define polynomial

\[ P_k(A) = \gamma_0 I + \gamma_1 A + \ldots + \gamma_k A^k \]

Use orthogonal eigenvectors \( \mathbf{v}_i \) of \( A \).

Show \( \mathbf{v}_i \) are also eigenvectors of \( P_k(A) \)

Rate of Convergence

\[ x_{k+1} = x_0 + \alpha_0 P_0 + \ldots + \alpha_k P_k \]

By construction

\[ x_{k+1} = x_0 + \gamma_0 r_0 + \gamma_1 Ar_0 + \ldots + \gamma_k A^k r_0 \]

Define polynomial:

\[ P_k(A) = \gamma_0 I + \gamma_1 A + \ldots + \gamma_k A^k \]

(Theorem 5.3)

Therefore

\[ x_{k+1} = x_0 + P_k(A)r_0 \]

(D)

Now

\[ \frac{1}{2} \| x - x^* \|_A^2 = \phi(x) - \phi(x^*) \]

\[ \| z \|_A^2 = z^T A z \]

\[ \phi(x) = \frac{1}{2} x^T A x - b^T x \]
Rate of Convergence

\[
\frac{1}{2} \| x - x^* \|^2 = \frac{1}{2} (x - x^*)^T A (x - x^*) = \frac{1}{2} (x^T - x^*) (Ax - Ax^*) = \frac{1}{2} x^T Ax - \frac{1}{2} x^T Ax^* = \frac{1}{2} x^T Ax - \frac{1}{2} x^T Ax^* - \frac{1}{2} x^T Ax^* + \frac{1}{2} x^T Ax^* = \frac{1}{2} x^T Ax - \phi(x) = \frac{1}{2} x^T Ax - b^T x
\]

According to Theorem 5.2, \( x_{k+1} \) minimizes \( \phi \), hence \( \| x - x^* \|^2 \)

By construction

\[
\phi(x) = \frac{1}{2} x^T Ax - b^T x
\]

Therefore, \( p^*_k \) solves the following problem:

\[
\min_{p_k} \| x_0 + P_k(A)r_0 - x^* \|^2
\]
We know \( r_0 = Ax_0 - b = Ax_0 - Ax^* = A(x_0 - x^*) \)

\[
x_{k+1} - x^* = x_0 + P_k^*(A)r_0 - x^* \\
= (x_0 - x^*) + P_k^*(A)r_0 \\
= (x_0 - x^*) + P_k^*(A)(x_0 - x^*) \\
= [I + P_k^*(A)A](x_0 - x^*) \tag{A}
\]

Assume \( \nu, \lambda \) are eigenvectors & eigenvalues of \( A \)

\[
x_0 - x^* = \sum_{i=1}^{n} \xi_i \nu_i
\]

Show \( \nu_i \) are also eigenvectors of \( P_k(A) \)

\[
P_k^*(A) = \gamma_0 I + \gamma_1 A + \ldots + \gamma_k A^k
\]

\[
P_k(A)\nu_i = \gamma_0 \nu_i + \gamma_1 A \nu_i + \gamma_2 A^2 \nu_i + \ldots + \gamma_k A^{k} \nu_i \\
P_k(A)\nu_i = \gamma_0 \nu_i + \gamma_1 \lambda_i \nu_i + \gamma_2 \lambda_i^2 \nu_i + \ldots + \gamma_k \lambda_i^k \nu_i \\
P_k(A)\nu_i = \gamma_0 \nu_i + \gamma_1 \lambda_i \nu_i + \gamma_2 \lambda_i^2 \nu_i + \ldots + \gamma_k \lambda_i^k \nu_i \\
P_k(A)\nu_i = P_k(A)\nu_i = P(\lambda_i)\nu_i
\]
Therefore \( P_k(A)v_i = P(\lambda_i)v_i \) for \( i = 1, 2, \ldots, n \)

We know \( x_0 - x^* = \sum_{i=1}^{n} \xi_i v_i \)

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

From (A)

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

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

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

\[
x_{k+1} - x^* = \sum_{i=1}^{n} [1 + \lambda_i P_k^* (\lambda_i) \xi_i] v_i
\]

\[
x_{k+1} - x^* = \sum_{i=1}^{n} [1 + \lambda_i P_k^* (\lambda_i)] \xi_i v_i
\]

Orthogonal eigenvectors
\[ \| x_{k+1} - x^* \|_A^2 = \sum_{i=1}^{n} \lambda_i (1 + \lambda_i P_k(\lambda_i))^2 \xi_j^2 \]

Since polynomial generated by GC is optimal

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

\[ \| x_{k+1} - x^* \|_A^2 \leq \min_{1 \leq i \leq n} \left( 1 + \lambda_i P_k(\lambda_i) \right) \left( \sum_{j=1}^{n} \lambda_j \xi_j^2 \right) \]

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

(C) \[ x_0 - x^* = \sum_{i=1}^{n} \xi_j v_j \]

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

Convergence

\[ x_0 - x^* = \sum_{i=1}^{n} \xi_j v_j \]

\[ \| x_0 - x^* \|_A^2 = \sum_{i=1}^{n} \xi_j v_i^T A \sum_{j=1}^{n} \xi_j v_j \]

\[ = \sum_{i=1}^{n} \xi_j v_i^T \sum_{j=1}^{n} \xi_j A v_j \]

\[ = \sum_{i=1}^{n} \xi_j v_i^T \sum_{j=1}^{n} \xi_j \lambda_i v_j \]

\[ = \sum_{i=1}^{n} \xi_j^2 \lambda_i \]
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 $r-1$

$$\min_{\lambda_i} \max_{k \leq n} \left[ 1 + \lambda_i P_k(\lambda_i) \right] \geq \max_{\lambda_i} \left[ 1 + \lambda_i \tilde{P}_{r-1}(\lambda) \right] = \max_{\lambda_i \leq \lambda} Q_r(\lambda_i) = 0$$
\[
\min_{P_r} \max_{1 \leq i \leq n} [1 + \lambda_i P_{r-1}(\lambda_i)]^2 = 0 \quad \text{For } k=r-1
\]

From (C)
\[
\| x_{k+1} - x^* \|_A^2 \leq \min_{P_r} \max_{1 \leq i \leq n} [1 + \lambda_i P_{r}(\lambda_i)]^2 \| x_0 - x^* \|_A^2 = 0
\]
\[
\| x_r - x^* \|_A^2 = 0
\]
Therefore
\[
x_r = x^* \quad \text{QED}
\]

**Theorem 5.5**

If \( A \) has distinct eigenvalues \( \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n \) we have
\[
\| x_{k+1} - x^* \|_A^2 \leq \left( \frac{\lambda_{n-k} - \lambda_k}{\lambda_{n-k} + \lambda_k} \right)^2 \| x_0 - x^* \|_A^2
\]

Eigenvalues
\( \lambda_1, \ldots, \lambda_{n-k}, \lambda_{n-k+1}, \ldots, \lambda_n \)
Eigenvalues
\[\lambda_1, \ldots, \lambda_{n-k}, \lambda_{n-k+1}, \ldots, \lambda_n\]

Select polynomial \( \bar{r}(\lambda) \) of degree \( k \) such that
\[Q_{k+1}(\lambda) = 1 + \lambda P_k(\lambda)\]
Has roots at \( k \) largest eigenvalues
\[\lambda_1, \lambda_2, \ldots, \lambda_{n-k+1}\]
As well as at mid point \( \lambda_{n-k} \) and \( \lambda_{n-k+1} \)

Maximum value attained by \( Q \) on the remaining eigenvalues is precisely

\[
\begin{align*}
\| x_{k+1} - x^* \|_A^2 & \leq \min_{1 \leq i \leq n} \max_{1 \leq r \leq k} \left[ \lambda_i, \lambda_i P(\lambda_r) \right] \| x_0 - x^* \|_A^2 \\
\| x_{k+1} - x^* \|_A^2 & \leq \left( \frac{\lambda_{n-k} - \lambda_{n-k+1}}{\lambda_{n-k} + \lambda_{n-k+1}} \right)^2 \| x_0 - x^* \|_A^2
\end{align*}
\]

Example

\[\{\lambda_1, \ldots, \lambda_{n-m}\}, \{\lambda_{n-m+1}, \ldots, \lambda_n\}\]

Clustered around 1

For small value of \( \varepsilon \)
CG will converge in only \( m+1 \) steps.

\[
\begin{align*}
\| x_{m+1} - x^* \|_A & = \varepsilon \| x_0 - x^* \|_A \\
\| x_{k+1} - x^* \|_A^2 & \leq \left( \frac{\lambda_{n-k} - \lambda_{n-k+1}}{\lambda_{n-k} + \lambda_{n-k+1}} \right)^2 \| x_0 - x^* \|_A^2
\end{align*}
\]
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

\[ \| x_{k+1} - x^* \|_A \leq \left( \frac{\sqrt{\kappa(A)} - 1}{\lambda \sqrt{\kappa(A)} + 1} \right)^2 \| x_0 - x^* \|_A \]

\[ \kappa(A) = \| A \|_2 \| A^{-1} \|_2 = \frac{\lambda_1}{\lambda_n} \]

Convergence Rate of Steepest Descent: Quadratic Function

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

Theorem 3.3

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