Lecture 12

Rate of Convergence

Theorem 5.4

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

Assume eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ take r distinct values:

$$\tau_1 < \tau_2, \ldots, < \tau_r$$

Define polynomial:

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

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

$$Q_r(0) = 1$$

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

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

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

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

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

From (C)

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

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

Therefore

$$x_r = x^*$$

QED

Theorem 5.5

If A has eigenvalues $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$ we have

$$\| x_{k+1} - x^* \|_A^2 \le \left(\frac{\lambda_{n-k} - \lambda_1}{\lambda_{n-k} + \lambda_1} \right)^2 \| x_0 - x^* \|_A^2$$

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 of degree k such that

Q has roots at k largest eigenvalues

$\lambda_n, \lambda_{n-1}, \ldots, \lambda_{n-k+1}$

As well as at mid point λ_1 and λ_{n-k}

$$Q_{k+1}(\lambda) = 1 + \lambda \overline{P}_k(\lambda)$$

Maximum value attained by Q on the remaining eigenvalues is precisely

$$\left(\frac{\lambda_{n-k} - \lambda_1}{\lambda_{n-k} + \lambda_1} \right)$$

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

$$\| x_{k+1} - x^* \|_A^2 \le \left(\frac{\lambda_{n-k} - \lambda_1}{\lambda_{n-k} + \lambda_1} \right)^2 \| x_0 - x^* \|_A^2$$

Example

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

m largest eigenvalues

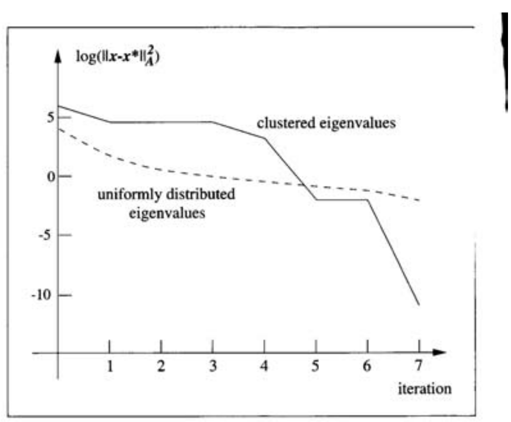
Clustered around 1

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

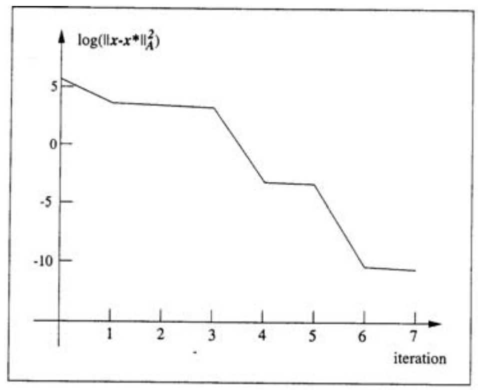
$$\| x_{k+1} - x^* \|_A^2 \leq \left(\frac{\lambda_{n-k} - \lambda_1}{\lambda_{n-k} + \lambda_1} \right)^2 \| x_0 - x^* \|_A^2$$

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

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 Condition number

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

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

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.

Preconditioning

- If the matrix *A* dose not have favorable eigenvalues, we can transform the problem such that eigenvalue distribution improves.

Preconditioning

Original problem:

$$\phi(x) = \frac{1}{2} x^T A x - b^T x \qquad \text{or} \qquad Ax = b$$

Transformation:

$$\hat{x} = Cx \qquad C^{-1}\hat{x} = x$$

Transformed problem:

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

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

Quadratic Function

$$(C^{-T} A C^{-1})\hat{x} = (C^{-T} b)$$

Linear system

Select C such that:

condition number of $C^{-T} A C^{-1}$ is much smaller than the original Matrix A.

The eigenvalues of $C^{-T} A C^{-1}$ are clustered

One possible preconditioner is

$$C^{-T} A C^{-1} = L^{-1} A L^{-T} = L^{-1} L L^T L^{-T} = I$$

Algorithm 5.3 (Preconditioned CG)

Given x_0, preconditioner $M = C^T C$;
set $r_0 \leftarrow Ax_0 - b$,
solve $My_0 = r_0$, for y_0;
$p_0 \leftarrow -r_0, k \leftarrow 0$
While $r_k \neq 0$

$$\alpha_k \leftarrow -\frac{r_k^T y_k}{p_k^T A p_k};$$
$$x_{k+1} \leftarrow x_k + \alpha_k p_k;$$
$$r_{k+1} \leftarrow r_k + \alpha_k A p_k;$$
$$My_{k+1} = r_{k+1}$$
$$\beta_{k+1} \leftarrow \frac{r_{k+1}^T y_{k+1}}{r_k^T y_k};$$
$$p_{k+1} \leftarrow -y_{k+1} + \beta_{k+1} p_k;$$
$$k \leftarrow k+1;$$

end(*while*)

5.3

Given x_0;
set $r_0 \leftarrow Ax_0 - b,\ p_0 \leftarrow -r_0, k \leftarrow 0$
While $r_k \neq 0$

$$\alpha_k \leftarrow -\frac{r_k^T r_k}{p_k^T A p_k};$$
$$x_{k+1} \leftarrow x_k + \alpha_k p_k;$$
$$r_{k+1} \leftarrow r_k + \alpha_k A p_k;$$
$$\beta_{k+1} \leftarrow \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k};$$
$$p_{k+1} \leftarrow -r_{k+1} + \beta_{k+1} p_k;$$
$$k \leftarrow k+1;$$

end(*while*)

5.2

Non-linear CG

- Two changes in linear GC
  - Perform line search for step length
  - Replace residual *r* by gradient of function
- Two algorithms:
  - FR
  - PR
- The difference is only in β

Algorithm 5.4 (FR-CG)

Given x_0;
evaluate $f_0 = f(x_0), \nabla f_0 = \nabla f(x_0)$
set $p_0 \leftarrow -\nabla f_0, k \leftarrow 0$
While $\nabla f_k \neq 0$

compute α_k;
$x_{k+1} \leftarrow x_k + \alpha_k p_k$;
evaluate ∇f_{k+1};
$\beta_{k+1}^{FR} \leftarrow \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{\nabla f_k^T \nabla f_k}$;
$p_{k+1} \leftarrow -\nabla f_{k+1} + \beta_{k+1}^{FR} p_k$;
$k \leftarrow k+1$;

end(*while*)

5.4

Given x_0;
set $r_0 \leftarrow Ax_0 - b,\ p_0 \leftarrow -r_0, k \leftarrow 0$
While $r_k \neq 0$

$\alpha_k \leftarrow -\frac{r_k^T r_k}{p_k^T A p_k}$;
$x_{k+1} \leftarrow x_k + \alpha_k p_k$;
$r_{k+1} \leftarrow r_k + \alpha_k A p_k$;
$\beta_{k+1} \leftarrow \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$;
$p_{k+1} \leftarrow -r_{k+1} + \beta_{k+1} p_k$;
$k \leftarrow k+1$;

end(*while*)

5.2

Choice of step length

$$p_{k+1} \leftarrow -\nabla f_{k+1} + \beta_{k+1}^{FR} p_k$$

The search direction p_k may fail to be a descent direction, unless Step length satisfies certain conditions.

$$p_k = -\nabla f_k + \beta_k^{FR} p_{k-1}$$

$$\nabla f_k^T p_k = -\nabla f_k^T \nabla f_k + \beta_k^{FR} \nabla f_k^T p_{k-1}$$

$$\nabla f_k^T p_k = -\| \nabla f_k \|^2 + \beta_k^{FR} \nabla f_k^T p_{k-1}$$

If $\nabla f_k^T p_{k-1} = 0$, then $\nabla f_k^T p_k < 0$, therefore p_k is a descent direction.

If $\nabla f_k^T p_{k-1} \neq 0$,then the second term may dominate, and $\nabla f_k^T p_k > 0$

Choice of step length

To solve this problem, we will require step length satisfies following Strong Wolf's conditions:

$$f(x_k + \alpha p_k) \le f(x_k) + c_1 \alpha \nabla f_k^T p_k, \quad c_1 \in (0,1)$$

$$|\nabla f(x_k + \alpha p_k)^T p_k| \le c_2 |\nabla f_k^T(x_k) p_k|, \quad 0 < c_1 < c_2 < \frac{1}{2}$$

We will show in Lemma 5.6 that the Wolf's conditions guarantee:

$$\nabla f_k^T p_k < 0$$

Polak-Ribiere

$$\beta_{k+1}^{PR} \leftarrow \frac{\nabla f_{k+1}^T (\nabla f_{k+1} - \nabla f_k)}{\nabla f_k^T \nabla f_k}$$

$$\beta_{k+1}^{FR} \leftarrow \frac{\nabla f_{k+1}^T \nabla f_{k+1}}{\nabla f_k^T \nabla f_k}$$

They are the same if the f is quadratic function, and line search is exact, since gradients (residuals) are mutually orthogonal by Theorem 5.3

For general non-linear functions, numerical experience indicates PR-CG tends to be more robust and efficient.

For PR-CG strong wolf conditions do not guarantee that p_k is always a descent direction.

Other Choices

$\beta_{k+1}^{+} = \max(\beta_{k+1}^{PR}, 0)$ This can satisfy descent property

$$\beta_{k+1}^{HS} \leftarrow \frac{\nabla f_{k+1}^{T}(\nabla f_{k+1} - \nabla f_{k})}{(\nabla f_{k+1} - \nabla f_{k})^{T} p_{k}}$$