Section 5.1 – The Linear Conjugate Gradient Method

AMS 230 Numerical Optimization

Qi Gong
Dept. of Applied Math & Stats.
Baskin School of Engineering
University of California
Santa Cruz, CA

CG – Basic Properties

Theorem 5.3 At the k-th iteration, the following properties hold:

1. $span\{r_0, r_1, \cdots, r_k\} = span\{p_0, p_1, \cdots, p_k\} \boxed{= span\{r_0, Ar_0, \cdots, A^k r_0\}}$,
2. $r_{k+1} \perp span\{p_0, p_1, \cdots, p_k\}$,
3. $p_{k+1} \perp^A span\{p_0, p_1, \cdots, p_k\}$,
4. $r_{k+1} \perp^A span\{p_0, p_1, \cdots, p_{k-1}\}$.

Krylov subspace

Without loss of generality, we assume $x_0=0$, and denote the global min as $x^*=A^{-1}b$.

Theorem 5.2 If at step k, the CG iteration has not already converged, (i.e., $r_k \neq 0$), then

1. x_{k+1} minimizes $\|x - x^*\|_A$ among all $x \in span\{p_0, p_1, \cdots, p_k\}$,
2. $\|x_{k+1} - x^*\|_A \leq \|x_k - x^*\|_A$,
3. $\boxed{\text{the iteration converges to } x^* \text{ in at most n steps.}}$

CG – Convergence Properties

Denote $\mathbb{P}_k(\cdot)$ to be all polynomials of degree less than or equal to k with $P_k(0) = 1$.

Without loss of generality, let $x_0 = 0$. The error at step k satisfies

$$
\begin{aligned}
e_k &= x_k - x^* = x_{k-1} + \alpha_{k-1} p_{k-1} - A^{-1} b = \cdots \\
&= x_0 + \boxed{\alpha_0 p_0 + \cdots + \alpha_{k-1} p_{k-1}} - A^{-1} b
\end{aligned}
$$

by T5.3 $span\{p_0, p_1, \cdots, p_{k-1}\} = span\{b, Ab, \cdots, A^{k-1}b\}$

$$
\begin{aligned}
&= \boxed{c_0 b + c_1 Ab + \cdots + c_{k-1} A^{k-1} b} - A^{-1} b \\
&= (c_0 A + c_1 A^2 + \cdots + c_{k-1} A^k - I) A^{-1} b \\
&= (I - c_0 A - c_1 A^2 - \cdots - c_{k-1} A^k) e_0
\end{aligned}
$$

Lemma: If at step k, the CG iteration has not already converged, (i.e., $r_k \neq 0$), then

1. $e_k = P_k^*(A) e_0$ for some $P_k^* \in \mathbb{P}_k$,

CG – Convergence Properties

Lemma: If at step k, the CG iteration has not already converged, (i.e., $r_k \neq 0$), then

1. $e_k = P_k^*(A)e_0$ for some $P_k^* \in \mathbb{P}_k$,

+

Theorem 5.2 If at step k, the CG iteration has not already converged, (i.e., $r_k \neq 0$), then

1. x_{k+1} minimizes $\|x - x^*\|_A$ among all $x \in \; span\{p_0, p_1, \cdots, p_k\} = \; span\{b, Ab, \cdots, A^k b\}$

$\Longrightarrow$

Lemma: If at step k, the CG iteration has not already converged, (i.e., $r_k \neq 0$), then

1. $e_k = P_k^*(A)e_0$ for some $P_k^* \in \mathbb{P}_k$,
2. x_k minimizes $\|P_k(A)e_0\|_A$ among all $P_k \in \mathbb{P}_k$

(since $\forall x \in \; span\{b, Ab, \cdots, A^{k-1}b\}$, there is a $P_k(A) \in \mathbb{P}_k$, s.t., $x - x^* = P_k(A)e_0$, and vice versa.)

CG – Convergence Properties

Lemma: If at step k, the CG iteration has not already converged, (i.e., $r_k \neq 0$), then

1. $e_k = P_k^*(A)e_0$ for some $P_k^* \in \mathbb{P}_k$,
2. x_k minimizes $\|P_k(A)e_0\|_A$ among all $P_k \in \mathbb{P}_k$
3. $\|e_k\|_A \leq \left(\min_{P_k \in \mathbb{P}_k} \max_{1 \leq i \leq n} |P_k(\lambda_i)| \right) \|e_0\|_A$, where λ_i are eigenvalues of A.

Sketch of a proof:

Let $A = Q^T diag\{\lambda_1, \ldots, \lambda_n\}Q$, where $Q = [q_1 . q_, \ldots, q_n]$ is an orthogonal matrix.

1. Let $e_0 = c_1 q_1 + \cdots c_n q_n \qquad \Longrightarrow$

$$\begin{aligned} P_k(A)e_0 &= c_1 P_k(A)q_1 + c_2 P_k(A)q_2 + \cdots c_n P_k(A)q_n \\ &= c_1 P_k(\lambda_1)q_1 + c_2 P_k(\lambda_2)q_2 + \cdots + c_n P_k(\lambda_n)q_n \end{aligned}$$

2. $$\begin{aligned} \|P_k(A)e_0\|_A^2 &= c_1^2 P_k^2(\lambda_1)\lambda_1 + c_2^2 P_k^2(\lambda_2)\lambda_2 + \cdots + c_n^2 P_k^2(\lambda_n)\lambda_n \\ &\leq \max_{1 \leq i \leq n} P_k^2(\lambda_i)[c_1^2 \lambda_1 + \cdots + c_n \lambda_n] \\ &= \max_{1 \leq i \leq n} P_k^2(\lambda_i)\|e_0\|_A^2 \end{aligned}$$

CG – Convergence Properties

Lemma: If at step k, the CG iteration has not already converged, (i.e., $r_k \neq 0$), then

1. $e_k = P_k^*(A)e_0$ for some $P_k^* \in \mathbb{P}_k$,
2. x_k minimizes $\|P_k(A)e_0\|_A$ among all $P_k \in \mathbb{P}_k$
3. $\|e_k\|_A \leq \left(\min_{P_k \in \mathbb{P}_k} \max_{1 \leq i \leq n} |P_k(\lambda_i)| \right) \|e_0\|_A$, where λ_i are eigenvalues of A.

Theorem 5.4.

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

1. Construct $\hat{P}(\lambda) = (1 - \frac{\lambda}{\lambda_1})(1 - \frac{\lambda}{\lambda_2}) \cdots (1 - \frac{\lambda}{\lambda_r})$.
2. $\hat{P} \in \mathbb{P}_r$ and $\hat{P}(\lambda_i) = 0$.
3. By the previous Lemma,

$$\begin{aligned} \|e_r\|_A &\leq \left(\min_{P_r \in \mathbb{P}_r} \max_{1 \leq i \leq n} |P_r(\lambda_i)| \right) \|e_0\|_A \\ &\leq \left(\max_{1 \leq i \leq n} |\hat{P}(\lambda_i)| \right) \|e_0\|_A \\ &= 0. \end{aligned}$$

CG – Convergence Properties

Lemma: If at step k, the CG iteration has not already converged, (i.e., $r_k \neq 0$), then

1. $e_k = P_k^*(A)e_0$ for some $P_k^* \in \mathbb{P}_k$,
2. x_k minimizes $\|P_k(A)e_0\|_A$ among all $P_k \in \mathbb{P}_k$
3. $\|e_k\|_A \leq \left(\min_{P_k \in \mathbb{P}_k} \max_{1 \leq i \leq n} |P_k(\lambda_i)| \right) \|e_0\|_A$, where λ_i are eigenvalues of A.

Theorem 5.5.

If A has eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$, we have that

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

1. Construct $\hat{P}_{k+1}(\lambda) = (1 - \frac{\lambda}{\lambda_n})(1 - \frac{\lambda}{\lambda_{n-1}}) \cdots (1 - \frac{\lambda}{\lambda_{n-k+1}})(1 - \frac{2\lambda}{\lambda_1 + \lambda_{n-k}})$.
2. By the previous Lemma,

$$\|e_{k+1}\|_A \leq \left(\min_{P_{k+1} \in \mathbb{P}_{k+1}} \max_{1 \leq i \leq n} |P_{k+1}(\lambda_i)| \right) \|e_0\|_A \leq \left(\max_{1 \leq i \leq n} |\hat{P}_{k+1}(\lambda_i)| \right) \|e_0\|_A \leq \frac{\lambda_{n-k} - \lambda_1}{\lambda_{n-k} + \lambda_1} \|e_0\|_A$$

CG – Convergence Properties

$\kappa(A)$: condition number of A

CG: $\|x_k - x^*\|_A \leq 2\left(\frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1}\right)^k \|x_0 - x^*\|_A$

Much better than SD for ill conditioned problems

SD: $\|x_k - x^*\|_A \leq \left(\frac{\kappa(A) - 1}{\kappa(A) + 1}\right)^k \|x_0 - x^*\|_A$