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

Conjugate Gradient Methods

At a given initial point x_0, the residual is $r_0 = b - Ax_0$.

Step 1: $p_0 = r_0 \Longrightarrow x_1 = x_0 + \alpha_0 p_0 \Longrightarrow r_1 = b - Ax_1;$

Step 2: $p_1 = r_1 - \frac{\langle r_1, p_0 \rangle}{\langle p_0, p_0 \rangle} p_0 \Longrightarrow x_2 = x_1 + \alpha_1 p_1 \Longrightarrow r_2 = b - Ax_2;$

Step 3: $p_2 = r_2 - \cancel{\frac{\langle r_2, p_0 \rangle}{\langle p_0, p_0 \rangle}} p_0 - \frac{\langle r_2, p_1 \rangle}{\langle p_1, p_1 \rangle} p_1 \Longrightarrow x_3 = x_2 + \alpha_2 p_2 \Longrightarrow r_3 = b - Ax_3;$

$\vdots$

Step k: $p_k = r_k - \cancel{\frac{\langle r_k, p_0 \rangle}{\langle p_0, p_0 \rangle}} p_0 - \cdots - \cancel{\frac{\langle r_k, p_{k-2} \rangle}{\langle p_{k-2}, p_{k-2} \rangle}} p_{k-2} - \frac{\langle r_k, p_{k-1} \rangle}{\langle p_{k-1}, p_{k-1} \rangle} p_{k-1};$

$\vdots$

- The computational cost involved in G-S orthogonalization is significantly reduced.

Conjugate Gradient Methods

At a given initial point x_0, the residual is $r_0 = b - Ax_0$.

Step 1: $p_0 = r_0 \implies x_1 = x_0 + \alpha_0 p_0 \implies r_1 = b - Ax_1$

Step 2: $p_1 = r_1 - \dfrac{\langle r_1, p_0\rangle}{\langle p_0, p_0\rangle} p_0 \implies x_2 = x_1 + \alpha_1 p_1 \implies r_2 = b - Ax_2$

Step 3: $p_2 = r_2 - \dfrac{\langle r_2, p_1\rangle}{\langle p_1, p_1\rangle} p_1 \implies x_3 = x_2 + \alpha_2 p_2 \implies r_3 = b - Ax_3$

$\vdots$

Step k: $p_k = r_k - \dfrac{\langle r_k, p_{k-1}\rangle}{\langle p_{k-1}, p_{k-1}\rangle} p_{k-1} \implies x_{k+1} = x_k + \alpha_k p_k \implies r_{k+1} = b - Ax_{k+1}$

$\vdots$

$$\boxed{\left\{\begin{array}{rcl} \alpha_k &=& \frac{r_k^T p_k}{p_k^T A p_k} \\ x_{k+1} &=& x_k + \alpha_k p_k \\ r_{k+1} &=& b - Ax_{k+1} \\ p_{k+1} &=& r_{k+1} - \frac{r_{k+1}^T A p_k}{p_k^T A p_k} p_k \end{array}\right.}$$

This is essentially linear conjugate gradient method. (the inner product used here is always $\langle v_1, v_2\rangle \triangleq v_1^T A v_2$)

Conjugate Gradient Iterations

Algorithm 5.1 (CG–Preliminary Version).

Given x_0;

Set $r_0 \leftarrow Ax_0 - b$, $p_0 \leftarrow -r_0$, $k \leftarrow 0$;

while $r_k \neq 0$

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

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

$$r_{k+1} \leftarrow A x_{k+1} - b;$$

$$\beta_{k+1} \leftarrow \frac{r_{k+1}^T A p_k}{p_k^T A p_k};$$

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

$$k \leftarrow k + 1;$$

end (while)

$\Longleftrightarrow$

Algorithm 5.2 (CG).

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)

CG – Basic Properties

Some notations:

- $\perp$: orthogonal with respect to standard inner product $\langle v_1, v_2 \rangle = v_1^\top v_2$,
- $\perp^A$: orthogonal with respect to inner product $\langle v_1, v_2 \rangle_A = v_1^\top A v_2$,
- $\|v\|_A^2 = v^\top A v$.

The following results are the same as those in Section 5.1, but presented in different formats.

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

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

CG – Basic Properties

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

1. $span\{r_0, r_1, \cdots, r_k\} = span\{p_0, p_1, \cdots, p_k\}$,

Proof: The statement clearly holds when k=0. Suppose at step k the statement is true.

For any $v \in span\{r_0, r_1, \cdots, r_{k+1}\}$, there are scalars $c_0, c_1, \ldots, c_{k+1}$ s.t.

$$\begin{aligned} v &= c_0 r_0 + \cdots + c_k r_k + c_{k+1} r_{k+1} \\ &= c_0 r_0 + \cdots + c_k r_k + c_{k+1}\left(p_{k+1} + \frac{\langle r_{k+1}, p_k \rangle}{\langle p_k, p_k \rangle} p_k\right) \\ &= \boxed{c_0 r_0 + \cdots + c_k r_k + c_{k+1} \frac{\langle r_{k+1}, p_k \rangle}{\langle p_k, p_k \rangle} p_k} + c_{k+1} p_{k+1} \end{aligned}$$

a vector in $span\{r_0, r_1, \cdots, r_k\} = span\{p_0, p_1, \cdots, p_k\}$

$\Longrightarrow$ $v \in span\{p_0, p_1, \cdots, p_k, p_{k+1}\}$

$\Longrightarrow$ $span\{r_0, r_1, \cdots, r_{k+1}\} \subseteq span\{p_0, p_1, \cdots, p_{k+1}\}$

Following similar argument, we can show $span\{p_0, p_1, \cdots, p_{k+1}\} \subseteq span\{r_0, r_1, \cdots, r_{k+1}\}$

Therefore, $span\{r_0, r_1, \cdots, r_{k+1}\} = span\{p_0, p_1, \cdots, p_{k+1}\}$

CG – Basic Properties

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

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

$$\begin{cases} \alpha_k = \frac{r_k^T p_k}{p_k^T A p_k} \\ x_{k+1} = x_k + \alpha_k p_k \\ r_{k+1} = b - A x_{k+1} \\ p_{k+1} = r_{k+1} - \frac{r_{k+1}^T A p_k}{p_k^T A p_k} p_k \end{cases}$$

Proof: It's easy to show $r_1 \perp p_0$, and $p_1 \perp^A p_0$. Suppose

$r_k \perp span\{p_0, p_1, \cdots, p_{k-1}\}$ and $p_k \perp^A span\{p_0, p_1, \cdots, p_{k-1}\}$. At k+1:

- For all $i = 0, \ldots, k-1$, we have

$$r_{k+1}^T p_i = (r_k - \alpha_k A p_k)^T p_i = r_k^T p_i - \alpha_k p_k^T A p_i = 0$$

since $r_k \perp span\{p_0, \cdots, p_{k-1}\}$ and $p_k \perp^A span\{p_0, \cdots, p_{k-1}\}$. Also

$$r_{k+1}^T p_k = (r_k - \alpha_k A p_k)^T p_k = r_k^T p_k - \frac{r_k^T p_k}{p_k^T A p_k} p_k^T A p_k = 0$$

Therefore, $r_{k+1} \perp span\{p_0, p_1, \cdots, p_k\}$.

CG – Basic Properties

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

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

$$\begin{cases} \alpha_k &= \frac{r_k^T p_k}{p_k^T A p_k} \\ x_{k+1} &= x_k + \alpha_k p_k \\ r_{k+1} &= b - A x_{k+1} \\ p_{k+1} &= r_{k+1} - \frac{r_{k+1}^T A p_k}{p_k^T A p_k} p_k \end{cases}$$

- For all $i = 0, \ldots, k-1$,

$$\begin{aligned} p_{k+1}^T A p_i &= (r_{k+1} - \frac{r_{k+1}^T A p_k}{p_k^T A p_k} p_k)^T A p_i = r_{k+1}^T A p_i - \frac{r_{k+1}^T A p_k}{p_k^T A p_k} p_k^T A p_i \\ &= r_{k+1}^T A p_i = r_{k+1}^T \frac{r_i - r_{i+1}}{\alpha_i} = 0 \end{aligned}$$

since $r_k \perp span\{p_0, \cdots, p_{k-1}\} = span\{r_0, \cdots, r_{k-1}\}$. Moreover

$$p_{k+1}^T A p_k = (r_{k+1} - \frac{r_{k+1}^T A p_k}{p_k^T A p_k} p_k)^T A p_k = r_{k+1}^T A p_k - r_{k+1}^T A p_k = 0$$

Hence, $p_{k+1} \perp^A span\{p_0, p_1, \cdots, p_k\}$.

CG – Basic Properties

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

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

$r_{k+1} = r_k - \alpha_k A p_k$

Proof: For all $i = 0, 1, \ldots, k-1$

$$r_{k+1}^T A p_i = r_{k+1}^T \frac{r_i - r_{i+1}}{\alpha_i} = 0,$$

since by statement 1 and 2, $r_{k+1} \perp span\{p_0, \cdots, p_k\} = span\{r_0, \cdots, r_k\}$.

CG – Finite Step Convergence

For the remaining part of analysis, without loss of generality, we assume $x_0=0$. We also 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.
3.

CG – Finite Step Convergence

For all $x \in span\{p_0, p_1, \ldots, p_k\}$, denote $\Delta x = x - x_{k+1}$. Then,

1. $\Delta x \in span\{p_0, p_1, \ldots, p_k\}$, since

$$\begin{aligned} x_{k+1} &= x_k + \alpha_k p_k = x_{k-1} + \alpha_{k-1} p_{k-1} + \alpha_k p_k = \cdots \\ &= \alpha_0 p_0 + \cdots + \alpha_k p_k \in span\{p_0, p_1, \ldots, p_k\}. \end{aligned}$$

2. By T5.3, $r_{k+1} \perp span\{p_0, p_1, \cdots, p_k\} \Longrightarrow \Delta x^T r_{k+1} = 0$.

3. Therefore,

$$\begin{aligned} \|x - x^*\|_A^2 &= \|x_{k+1} + \Delta x - x^*\|_A^2 \\ &= \|x_{k+1} - x^*\|_A^2 + \|\Delta x\|_A^2 + 2\Delta x^T A(x_{k+1} - x^*) \\ &= \|x_{k+1} - x^*\|_A^2 + \|\Delta x\|_A^2 - 2\Delta x^T r_{k+1} \\ &= \|x_{k+1} - x^*\|_A^2 + \|\Delta x\|_A^2 \\ &\geq \|x_{k+1} - x^*\|_A^2 \end{aligned}$$

   and equality holds if $\Delta x = 0$.

CG – Finite Step Convergence

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. $\boxed{\|x_{k+1} - x^*\|_A \leq \|x_k - x^*\|_A}$,
3.

since $span\{p_0, p_1, \ldots, p_{k-1}\} \subseteq span\{p_0, \ldots, p_{k-1}, p_k\}$

CG – Finite Step Convergence

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. the iteration converges to x^* in at most n steps.

By T5.3, $r_{k+1} \perp span\{p_0, p_1, \cdots, p_k\} \Longrightarrow$ dimension of $span\{p_0, p_1, \cdots, p_k\}$ keeps increasing as long as $r_{k+1} \neq 0$.