Lecture-11

Theorems 5.3 and 5.2

Algorithms 5.1, 5.2

Theorem 5.3

1. The directions are indeed conjugate.
2. Therefore, the algorithm terminates in n steps (from Theorem 5.1).
3. The residuals are mutually orthogonal.
4. Each direction p_k and r_k is contained in Krylov subspace of r_0 degree k.

Theorem 5.3

Suppose that the kth iteration generated by the conjugate gradient method is not the solution point x^*. The following four properties hold:

$$r_k^T r_i = 0 \qquad \text{for } i = 0, \ldots, k-1 \qquad (1)$$

$$\operatorname{span}\{r_0, r_1, \ldots, r_k\} = \operatorname{span}\{r_0, Ar_0, \ldots, A^k r_0\} \qquad (2)$$

$$\operatorname{span}\{p_0, p_1, \ldots, p_k\} = \operatorname{span}\{r_0, Ar_0, \ldots, A^k r_0\} \qquad (3)$$

$$p_k^T A p_i = 0 \qquad \text{for } i = 0, \ldots, k-1 \qquad (4)$$

Therefore, the sequence $\{x_k\}$ converges to x^* in at most n steps.

Proof

$$r_k^T r_i = 0 \qquad \text{for } i = 0, \ldots, k-1 \qquad (1)$$

$$\operatorname{span}\{r_0, r_1, \ldots, r_k\} = \operatorname{span}\{r_0, Ar_0, \ldots, A^k r_0\} \qquad (2)$$

$$\operatorname{span}\{p_0, p_1, \ldots, p_k\} = \operatorname{span}\{r_0, Ar_0, \ldots, A^k r_0\} \qquad (3)$$

$$p_k^T A p_i = 0 \qquad \text{for } i = 0, \ldots, k-1 \qquad (4)$$

- Use induction on (2) and (3)
  - First prove (2)
  - Then prove (3) using (2)
- Prove (4) by induction using (3) and Theorem 5.2
- Prove (1) using (4) and Theorem 5.2

Proof

$$r_k^T r_i = 0 \qquad \text{for } i = 0, \ldots, k-1 \qquad (1)$$

$$\text{span}\{r_0, r_1, \ldots, r_k\} = \text{span}\{r_0, Ar_0, \ldots, A^k r_0\} \qquad (2)$$

$$\text{span}\{p_0, p_1, \ldots, p_k\} = \text{span}\{r_0, Ar_0, \ldots, A^k r_0\} \qquad (3)$$

$$p_k^T A p_i = 0 \qquad \text{for } i = 0, \ldots, k-1 \qquad (4)$$

(2) And (3)

Induction: $k=0$

$$\text{span}\{r_0\} = \text{span}\{r_0\} \qquad (2)$$

$$\text{span}\{p_0\} = \text{span}\{r_0\} \qquad (3) \qquad p_0 = -r_0$$

Proof

$$r_k^T r_i = 0 \qquad \text{for } i = 0, \ldots, k-1 \qquad (1)$$

$$\text{span}\{r_0, r_1, \ldots, r_k\} = \text{span}\{r_0, Ar_0, \ldots, A^k r_0\} \qquad (2)$$

$$\text{span}\{p_0, p_1, \ldots, p_k\} = \text{span}\{r_0, Ar_0, \ldots, A^k r_0\} \qquad (3)$$

$$p_k^T A p_i = 0 \qquad \text{for } i = 0, \ldots, k-1 \qquad (4)$$

Assume (2) and (3) are true for k, prove for $k+1$

To prove (2), by induction:

$$r_k \in \text{span}\{r_0, Ar_0, \ldots, A^k r_0\}, \qquad p_k \in \text{span}\{r_0, Ar_0, \ldots, A^k r_0\}$$

$$Ap_k \in \text{span}\{Ar_0, A^2 r_0, \ldots, A^{k+1} r_0\} \qquad \text{By multiplying with } A$$

$$r_{k+1} = r_k + \alpha_k A p_k \qquad \text{Therefore} \qquad r_{k+1} \in \text{span}\{r_0, Ar_0, \ldots, A^{k+1} r_0\}$$

By combining this with induction hypothesis on (2)

$$\text{span}\{r_0, r_1, \ldots, r_{k+1}\} \subset \text{span}\{r_0, Ar_0, \ldots, A^{k+1} r_0\}$$

Proof

$$r_k^T r_i = 0 \qquad \text{for } i = 0, \ldots, k-1 \qquad (1)$$
$$\text{span}\{r_0, r_1, \ldots, r_k\} = \text{span}\{r_0, Ar_0, \ldots, A^k r_0\} \qquad (2)$$
$$\text{span}\{p_0, p_1, \ldots, p_k\} = \text{span}\{r_0, Ar_0, \ldots, A^k r_0\} \qquad (3)$$
$$p_k^T A p_i = 0 \qquad \text{for } i = 0, \ldots, k-1 \qquad (4)$$

To prove the reverse inclusion

$$A^{k+1} r_0 = A(A^k r_0) \in \text{span}\{Ap_0, Ap_1, \ldots, Ap_k\}$$ Induction on (3)

Since

$$Ap_i = \frac{(r_{i+1} - r_i)}{\alpha_i}, \text{ for } i = 0, \ldots, k$$ Because $r_{k+1} = r_k + \alpha_k A p_k$

Therefore $A^{k+1} r_0 \in \text{span}\{r_0, r_1, \ldots, r_{k+1}\}$

$$\text{span}\{r_0, Ar_0, \ldots, A^{k+1} r_0\} \subset \text{span}\{r_0, r_1, \ldots, r_k, r_{k+1}\}$$

$$\text{span}\{r_0, r_1, \ldots, r_k, r_{k+1}\} \subset \text{span}\{r_0, Ar_0, \ldots, A^{k+1} r_0\}$$ Induction hypothesis on (2)

Therefore $\text{span}\{r_0, r_1, \ldots, r_k, r_{k+1}\} = \text{span}\{r_0, Ar_0, \ldots, A^{k+1} r_0\}$ QED (2)

Proof

$$r_k^T r_i = 0 \qquad \text{for } i = 0, \ldots, k-1 \qquad (1)$$
$$\text{span}\{r_0, r_1, \ldots, r_k\} = \text{span}\{r_0, Ar_0, \ldots, A^k r_0\} \qquad (2)$$
$$\text{span}\{p_0, p_1, \ldots, p_k\} = \text{span}\{r_0, Ar_0, \ldots, A^k r_0\} \qquad (3)$$
$$p_k^T A p_i = 0 \qquad \text{for } i = 0, \ldots, k-1 \qquad (4)$$

Show (3) holds if k is replaced by $k+1$

$\text{span}\{p_0, p_1, \ldots, p_k, p_{k+1}\}$

$= \text{span}\{p_0, p_1, \ldots, p_k, r_{k+1}\}$ $\quad p_{k+1} \leftarrow -r_{k+1} + \beta_{k+1} p_k;$

$= \text{span}\{r_0, Ar_0, \ldots, A^k r_0, r_{k+1}\}$ $\quad$ Induction hypo for (3)

$= \text{span}\{r_0, r_1, \ldots, r_k, r_{k+1}\}$ $\quad$ By (2)

$= \text{span}\{r_0, Ar_0, \ldots, A^{k+1} r_0\}$ $\quad$ By (2) for $k+1$

QED (3)

Proof

$$r_k^T r_i = 0 \quad \text{for } i = 0,\dots,k-1 \qquad (1)$$

$$\text{span}\{r_0, r_1,\dots,r_k\} = \text{span}\{r_0, Ar_0,\dots,A^k r_0\} \qquad (2)$$

$$\text{span}\{p_0, p_1,\dots, p_k\} = \text{span}\{r_0, Ar_0,\dots,A^k r_0\} \qquad (3)$$

$$p_k^T A p_i = 0 \quad \text{for } i = 0,\dots,k-1 \qquad (4)$$

Now Conjugacy (4):

(4) Holds for k=1 $\quad p_1^T A p_0 = 0 \quad$ (4)

By definition: $\quad p_{k+1} = -r_{k+1} + b_{k+1} p_k;$

$$p_{k+1}^T A p_i = -r_{k+1}^T A p_i + b_{k+1} p_k^T A p_i \quad \text{for } i = 0,1,\dots,k \qquad \text{(F)}$$

By definition: $\quad b_{k+1} \leftarrow \frac{r_{k+1}^T A p_k}{p_k^T A p_k};$

Due to this the right side becomes Zero for $i=k$

By induction hypothesis on (4) the vectors are conjugate up to p_k

Therefore $\quad r_{k+1}^T p_i = 0 \quad \text{for } i = 0,\dots,k$

By Theorem 5.2

Proof

$$r_k^T r_i = 0 \quad \text{for } i = 0,\dots,k-1 \qquad (1)$$

$$\text{span}\{r_0, r_1,\dots,r_k\} = \text{span}\{r_0, Ar_0,\dots,A^k r_0\} \qquad (2)$$

$$\text{span}\{p_0, p_1,\dots, p_k\} = \text{span}\{r_0, Ar_0,\dots,A^k r_0\} \qquad (3)$$

$$p_k^T A p_i = 0 \quad \text{for } i = 0,\dots,k-1 \qquad (4)$$

$$p_{k+1}^T A p_i = -r_{k+1}^T A p_i + b_{k+1} p_k^T A p_i \quad \text{for } i = 0,1,\dots,k \qquad \text{(F)}$$

$$r_{k+1}^T p_i = 0 \quad \text{for } i = 0,\dots,k \qquad \text{(B)}$$

By applying (3)

$$Ap_i \in A\,\text{span}\{r_0, Ar_0,\dots,A^i r_0\} = \text{span}\{Ar_0, A^2 r_0,\dots,A^{i+1} r_0\}$$

$$\subset \text{span}\{p_0, p_1,\dots,p_{i+1}\} \qquad \text{(C)}$$

$$r_{k+1}^T A p_i = 0 \quad \text{for } i = 0,\dots,k-1 \qquad \text{By (B) \& (C)}$$

So the first term vanishes in (F). Due to induction hypothesis on (4) the second term vanishes as well. Hence QED (4).

So the direction set generated by CG method is indeed a conjugate direction set.

According to Theorem 5.1 the algorithm terminates in at most n steps.

Proof

$$r_k^T r_i = 0 \qquad \text{for } i = 0,\ldots,k-1 \qquad (1)$$
$$\text{span}\{r_0, r_1, \ldots, r_k\} = \text{span}\{r_0, Ar_0, \ldots, A^k r_0\} \qquad (2)$$
$$\text{span}\{p_0, p_1, \ldots, p_k\} = \text{span}\{r_0, Ar_0, \ldots, A^k r_0\} \qquad (3)$$
$$p_k^T A p_i = 0 \qquad \text{for } i = 0,\ldots,k-1 \qquad (4)$$

Now (1)

Since the direction set is conjugate because of (3), by theorem 5.2

$$r_k^T p_i = 0 \qquad \text{for } i = 0,\ldots,k-1, \qquad k = 1,2,\ldots,n-1$$

By definition

$$p_i = -r_i + b_i p_{i-1} \qquad\qquad p_{k+1} = -r_{k+1} + b_{k+1} p_k;$$

$$r_k^T p_i = 0 = r_k^T(-r_i + b_i p_{i-1}) = -r_k^T r_i + b_i r_k^T p_{i-1} = -r_k^T r_i$$

$$r_k^T r_i = 0 \qquad \text{for } i = 0,\ldots,k-1, \quad k = 1,2,\ldots,n-1$$

QED (1)

Theorem 5.2

Let x_0 be any starting point and suppose that the sequence $\{x_k\}$ is generated by the conjugate direction algorithm. Then

$$r_k^T p_i = 0 \qquad \text{for } i = 0,\ldots,k-1$$

and x_k is minimizer of over the set

(3)

Proof

First show that a point minimizes over the set (3) if and only if

$r(\tilde{x})^T p_i = 0 \quad \text{for } i = 0, \ldots, k-1$

(3)

Where

Let

Since is strictly convex quadratic, it has a unique minimizer:

$\frac{\partial h(\sigma^*)}{\partial \sigma_i} = 0, \quad i = 0, \ldots, k-1$

$\nabla f(x_0 + \sigma_0^* p_0 + \ldots + \sigma^*_{k-1} p_{k-1})^T p_i = 0 \quad i = 0, \ldots, k-1$ Chain rule

$r(x)$ is the residual

$r(\tilde{x})^T p_i = 0 \quad i = 0, \ldots, k-1$

Proof

$\nabla f(x) = Ax - b = r(x) \qquad x_{k+1} = x_k + \alpha_k p_k$

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

$r_k = r_{k-1} + \alpha_{k-1} A p_{k-1}$ (A)

Use induction:

Prove true for $k=1$:

From (A)

$r_1 = r_0 + \alpha_0 A p_0$

$r_1^T p_0 = (r_0 + \alpha_0 A p_0)^T p_0$

$r_1^T p_0 = r_0^T p_0 + \alpha_0 p_0^T A p_0$

if $r_1^T p_0 = 0$ Then $\alpha_k = -\frac{r_k^T p_k}{p_k^T A p_k}$

But, is a 1-D minimizer of quadratic function.

Proof

$r_k^T p_i = 0 \quad$ for $i = 0, \ldots, k-1$

Assume true for k-1 $\quad r_{k-1}^T p_i = 0 \quad$ for $i = 0, \ldots, k-2$

$$r_k = r_{k-1} + \alpha_{k-1} A p_{k-1}$$

$$p_{k-1}^T r_k = p_{k-1}^T r_{k-1} + \alpha_{k-1} p_{k-1}^T A p_{k-1} = 0$$

By multiplication

If $\quad p_{k-1}^T r_k = 0$

then α_{k-1} is given $\quad \alpha_{k-1} = -\frac{r_{k-1}^T p_{k-1}}{p_{k-1}^T A p_{k-1}}$

That is 1-D minimizer of quadratic function.

For other vectors p_i

$$p_i^T r_k = p_i^T r_{k-1} + \alpha_{k-1} p_i^T A p_{k-1} = 0 \qquad i = 0, \ldots, k-2$$

induction Conjugacy

This implies we have minimized quadratic function in *k-1* variables

Therefore $\quad r_k^T p_i = 0 \quad$ for $i = 0, \ldots, k-1$

Implies we have minimized quadratic function in *k* variables QED

How do we select conjugate directions

- Eigenvalues of *A* are mutually orthogonal and conjugate wrt to *A*.
- Gram-Schmidt process can be modified to produce conjugate directions instead of orthogonal vectors.
- Both approaches are expensive.

Basic Properties of the CG

Each direction is chosen to be a linear combination of the steepest descent direction and the previous direction.

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

Or

$$p_k = -r_k + \beta_k p_{k-1}$$

Where β_k is determined such That p_k and p_{k-1} must be conjugate

Therefore

$$p_{k-1}^T A p_k = -p_{k-1}^T A r_k + \beta_k p_{k-1}^T A p_{k-1}$$

$$\beta_k = \frac{r_k^T A p_{k-1}}{p_{k-1}^T A p_{k-1}}$$

It does not need to know all previous directions, only one previous direction is required.

p_k is automatically conjugate to all previous directions!

Algorithm 5.1

$$
\begin{aligned}
&\text{Given } x_0;\\
&set\ r_0 \leftarrow Ax_0 - b,\ \ p_0 \leftarrow -r_0, k \leftarrow 0\\
&While\ r_k \neq 0\\
&\qquad \alpha_k \leftarrow -\frac{r_k^T p_k}{p_k^T A p_k};\\
&\qquad x_{k+1} \leftarrow x_k + \alpha_k p_k;\\
&\qquad r_{k+1} \leftarrow Ax_{k+1} - b;\\
&\qquad \beta_{k+1} \leftarrow \frac{r_{k+1}^T A p_k}{p_k^T A p_k};\\
&\qquad p_{k+1} \leftarrow -r_{k+1} + \beta_{k+1} p_k;\\
&\qquad k \quad \leftarrow k+1;\\
&end(while)
\end{aligned}
$$

p_0 is steepest descent

$$\nabla f(x) = Ax - b = r(x)$$