Lecture-10

Theorems 5.2 and 5.3

Algorithms 5.1, 5.2

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

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

$$\{x \mid x = x_0 + span\{p_0, \ldots, p_{k-1}\}\} \qquad (3)$$

Where

Let $h(s) = f(x_0 + s_0 p_0 + \ldots + s_{k-1} p_{k-1})$ $\qquad s = (s_0, s_1, \ldots s_{k-1})$

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

$$\frac{\partial h(s^*)}{\partial s_i} = 0, \qquad i = 0, \ldots, k-1$$

$$\nabla f(x_0 + s_0^* p_0 + \ldots + s^*_{k-1} p_{k-1})^T p_i = 0 \qquad i = 0, \ldots, k-1$$

Chain rule

$r(x)$ is the residual

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

Proof

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

$$r_{k+1} = r_k + a_k A p_k$$

$$r_k = r_{k-1} + a_{k-1} A p_{k-1} \qquad \text{(A)}$$

Use induction:

Prove true for $k=1$:

From (A)

$$r_1 = r_0 + a_0 A p_0$$

$$r_1^T p_0 = (r_0 + a_0 A p_0)^T p_0$$

$$r_1^T p_0 = r_0^T p_0 + a_0 p_0^T A p_0$$

$$r_1^T p_0 = 0$$

Because

$$a_k = -\frac{r_k^T p_k}{p_k^T A p_k}$$

Proof

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

Assume true for k-1

$$r_k = r_{k-1} + a_{k-1} A p_{k-1} \qquad \text{(A)}$$

From (A)

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

True for P_{k-1}

$$a_k = -\frac{r_k^T p_k}{p_k^T A p_k}$$

Definition

And

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

Conjugacy

induction

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

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

Or

$$p_k = -r_k + b_k p_{k-1}$$

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

Therefore

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

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

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

Given x_0;

$set\ r_0 \leftarrow Ax_0 - b,\ \ p_0 \leftarrow -r_0, k \leftarrow 0$

$While\ r_k \neq 0$

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

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

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

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

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

$$k \quad \leftarrow k + 1;$$

$end(while)$

p_0 is steepest descent

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

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:

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

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

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

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

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

Proof

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

Induction: $k=0$

(2) And (3) hold

$$\text{span}\{r_0\} = \text{span}\{r_0\} \qquad (2)$$
$$\text{span}\{p_0\} = \text{span}\{r_0\} \qquad (3) \qquad p_0 = -r_0$$

(4) Holds for $k=1$

$$p_1^T A p_0 = 0 \qquad (4)$$

By construction

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), (3) and (4) are true for k, prove for $k+1$

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

To prove (2), by induction:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

QED (3)

Proof

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

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

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

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

Now Conjugacy (4):

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

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

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

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 $r_{k+1}^T p_i = 0 \quad \text{for } i = 0, \dots, k$ By Theorem 5.2

Proof

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

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

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

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

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

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

By applying (3)

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

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

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

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

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

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

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

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

Now (1)

Since the direction set is conjugate by theorem 5.2

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

$$p_i = -r_i + b_i p_{i-1}$$ By definition

$$r_i \in \text{span}\{p_i, p_{i-1}\} \text{ for } i = 0, \dots, k-1$$

$$r_i = ap_i + bp_{i-1}$$

$$p_{k+1} = -r_{k+1} + b_{k+1} p_k;$$

$$p_i = cr_i + dp_{i-1}$$

$$r_k^T p_i = 0 = r_k^T (cr_i + dp_{i-1}) = cr_k^T r_i + dr_k^T p_{i-1} = cr_k^T r_i$$

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

QED (1)

A practical form of GC

$$p_{k+1} = -r_{k+1} + b_{k+1} p_k;$$

$$p_k = -r_k + b_k p_{k-1};$$

Theorem 5.2

$$r_k^T p_k = -r_k^T r_k + b_k r_k^T p_{k-1}; \qquad r_k^T p_i = 0 \qquad \text{for } i = 0, \dots, k-1$$

$$r_k^T p_k = -r_k^T r_k$$

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

A practical form of GC

$$a_k A p_k = r_{k+1} - r_k$$
$$a_k r_{k+1}^T A p_k = r_{k+1}^T r_{k+1} - r_{k+1}^T r_k$$

$$p_k = -r_k + b_k p_{k-1};$$
$$r_k = p_k + b_k p_{k-1};$$
$$r_{k+1}^T r_k = r_{k+1}^T p_k + b_k r_{k+1}^T p_{k-1}; \quad r_k^T p_i = 0 \quad \text{for } i = 0, \dots, k-1$$
$$r_{k+1}^T r_k = 0$$

$$a_k r_{k+1}^T A p_k = r_{k+1}^T r_{k+1}$$

A practical form of GC

$$a_k r_{k+1}^T A p_k = r_{k+1}^T r_{k+1}$$

Now

$$a_k A p_k = r_{k+1} - r_k$$
$$a_k p_k^T A p_k = p_k^T r_{k+1} - p_k^T r_k$$

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

$$a_k p_k^T A p_k = 0 + r_k^T r_k$$

From a_k

$$a_k p_k^T A p_k = r_k^T r_k$$

$$b_{k+1} \leftarrow \frac{r_{k+1}^T A p_k}{p_k^T A p_k}; \longrightarrow b_{k+1} \leftarrow \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k};$$

Algorithm 5.2

Given x_0;

$set\ r_0 \leftarrow Ax_0 - b,\ \ p_0 \leftarrow -r_0, k \leftarrow 0$

$While\ r_k \neq 0$

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

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

$$r_{k+1} \leftarrow r_k + a_k A p_k;;$$

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

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

$$k \quad \leftarrow k+1;$$

$end(while)$