Lecture-12

Theorems 5.3 and 5.2

Algorithms 5.1, 5.2

Proof

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

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

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

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

Now Conjugacy (4):

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

Assume true for k, prove true for k+1

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

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

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

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

By Theorem 5.2

Proof

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

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

We want to show it is true for i=0,1,2,..,k-1

By applying (3)

$$Ap_i \in A\,\text{span}\{r_0, Ar_0, \dots, A^i r_0\} = \text{span}\{Ar_0, A^2 r_0, \dots, A^{i+1} r_0\}$$
$$\subset \text{span}\{p_0, p_1, \dots, p_{i+1}\} \qquad (C)$$
$$r_{k+1}^T A p_i = 0 \qquad \text{for } i = 0,\dots,k-1 \qquad \text{By (B) \& (C)}$$

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 because of (4), by theorem 5.2

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

By definition

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

$$r_k^T p_i = 0 = r_k^T(-r_i + \beta_i p_{i-1}) = -r_k^T r_i + \beta_i r_k^T p_{i-1} = -r_k^T r_i$$
$$r_k^T r_i = 0 \qquad \text{for } i = 0,\dots,k-1, \quad k = 1,2,\dots,n-1 \qquad \text{QED (1)}$$

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

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

and x_k is minimizer of over the set

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

Proof

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

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

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

Then if Where

Let $h(\sigma) = f(x_0 + \sigma_0 p_0 + \dots + \sigma_{k-1} p_{k-1})$

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

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

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

$r(x)$ is the residual

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

Proof

If

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

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

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

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 $\alpha_0 = -\dfrac{r_0^T p_0}{p_0^T A p_0}$

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

Proof

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

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

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

For i=k-1

$$p_{k-1}^T r_k = p_{k-1}^T r_{k-1} + \alpha_{k-1} p_{k-1}^T A p_{k-1} \qquad \text{By multiplication}$$

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

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

That is 1-D minimizer of quadratic function.

For other vectors p_i

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

induction Conjugacy

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

Therefore $r_k^T p_i = 0 \quad$ 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.

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

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

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

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

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;

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

$While\ r_k \neq 0$

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

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

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

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

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

$$k \leftarrow k+1;$$

$end(while)$

p_0 is steepest descent

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

A practical form of GC

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

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

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

$$r_k^T p_k = -r_k^T r_k + 0;$$

$$r_k^T p_k = -r_k^T r_k \qquad \text{(G)}$$

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

Theorem 5.2

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

A practical form of GC

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

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

Theorem 5.3

$$a_k r_{k+1}^T A p_k = r_{k+1}^T r_{k+1} - 0$$

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

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

A practical form of GC

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

(Already shown)

Now

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

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

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

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

Theorem 5.2

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

From (G)

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

Algorithm 5.2

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

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

$end(while)$

5.2

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

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

$end(while)$

5.1

Algorithm 5.2

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

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

$end(while)$

5.2

We only need to know values of *x*, *p* and *r* only for 2 iterations.

Major computations: matrix-vector product, two inner products, and three vector sums.