Lecture-12

Theorems 5.3 and 5.2
Algorithms 5.1, 5.2

Proof

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

Now Conjugacy (4):

(4) Holds for \( k = 1 \) \hspace{1cm} p_1^T A p_0 = 0 \hspace{1cm} (4)

Assume true for \( k \), prove true for \( k+1 \)

By definition:
\[ p_{k+1} = -r_{k+1} + \beta_{k+1} p_k \]
\[ p_{k+1}^T A p_i = -r_{k+1}^T A p_i + \beta_{k+1} p_k^T A p_i \quad \text{for } i = 0, 1, \ldots, k \quad (F) \]

By definition:
\[ \beta_{k+1} = \frac{r_{k+1}^T A p_k}{p_k^T A p_k} \]

Due to this the right side becomes Zero for \( i \neq k \)

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

Therefore
\[ r_{k+1}^T p_i = 0 \quad \text{for } i = 0, \ldots, k \]

By Theorem 5.2
Proof

\[ r^T_i r_i = 0 \quad \text{for } i = 0, \ldots, k - 1 \quad (1) \]
\[ \text{span} \{ r_i, r_{i+1}, \ldots, r_k \} = \text{span} \{ A r_i, A^2 r_{i+1}, \ldots, A^k r_k \} \quad (2) \]
\[ \text{span} \{ p_0, p_1, \ldots, p_i \} = \text{span} \{ A r_0, A^2 r_1, \ldots, A^i r_i \} \quad (3) \]
\[ p_i^T A p_i = 0 \quad \text{for } i = 0, \ldots, k - 1 \quad (4) \]

\[ p^T_{k+1} A p_i = -r^T_{k+1} A p_i + \beta_k r^T_k p_{i+1} p_i \quad \text{for } i = 0, \ldots, k \quad (F) \]
\[ r^T_{k+1} p_i = 0 \quad \text{for } i = 0, \ldots, k \quad (B) \]

We want to show it is true for \( i = 0, 1, 2, \ldots, k - 1 \)

By applying (3)
\[ A p_i \in \text{span} \{ r_0, r_1, \ldots, A^i r_i \} = \text{span} \{ A r_0, A^2 r_1, \ldots, A^i r_i \} \quad (C) \]
\[ r^T_{k+1} A p_i = 0 \quad \text{for } i = 0, \ldots, k - 1 \quad \text{By (B) & (C)} \]

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

\[ r^T_i r_i = 0 \quad \text{for } i = 0, \ldots, k - 1 \quad (1) \]
\[ \text{span} \{ r_i, r_{i+1}, \ldots, r_k \} = \text{span} \{ A r_i, A^2 r_{i+1}, \ldots, A^k r_k \} \quad (2) \]
\[ \text{span} \{ p_0, p_1, \ldots, p_i \} = \text{span} \{ A r_0, A^2 r_1, \ldots, A^i r_i \} \quad (3) \]
\[ p_i^T A p_i = 0 \quad \text{for } i = 0, \ldots, k - 1 \quad (4) \]

Now (1)

Since the direction set is conjugate because of (4), by theorem 5.2
\[ r^T_k p_i = 0 \quad \text{for } i = 0, \ldots, k - 1, \quad k = 1, 2, \ldots, n - 1 \]

By definition
\[ p_i = -r_i + \beta_i p_{i-1} \quad p_{k+1} = -r_{k+1} + \beta_{k+1} p_k \]

\[ r^T_k p_i = 0 = r^T_k (-r_i + \beta_i p_{i-1}) = -r^T_k r_i + \beta_i r^T_k p_{i-1} = -r^T_k r_i \]
\[ r^T_i r_i = 0 \quad \text{for } i = 0, \ldots, k - 1, \quad k = 1, 2, \ldots, n - 1 \quad \text{QED (1)} \]
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

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

and $x_k$ is minimizer of \{x \in A \mid x = x_0 + \text{span} \{p_0, \ldots, p_{k-1}\}\} over the set

\[ \{x \mid x = x_0 + \text{span} \{p_0, \ldots, p_{k-1}\}\} \quad (3) \]

Proof

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

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

Then if

\[ h(\sigma) = \phi(x_0 + \sigma_0 p_0 + \ldots + \sigma_{k-1} p_{k-1}) \]

Let

\[ \frac{\partial h(\sigma)}{\partial \sigma_i} = 0, \quad i = 0, \ldots, k - 1 \]

\[ \nabla \phi(x_0 + \sigma_0 p_0 + \ldots + \sigma_{k-1} p_{k-1}) p_i = 0 \quad \text{for } i = 0, \ldots, k - 1 \]

Chain rule

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

\[ r(x) \text{ is the residual} \]
Proof

If
\[ \nabla_\phi(x) = Ax - b = r(x) \]
\[ x_{k+1} = x_k + \alpha_k p_k \]
\[ r_{k+1} = r_k + \alpha_k Ap_k \]
\[ r_k = r_{k-1} + \alpha_{k-1} Ap_{k-1} \]

Use induction:

Prove true for \( k=1 \):
From (A)
\[ r_1 = r_0 + \alpha_0 Ap_0 \]
\[ r_1^T p_0 = (r_0 + \alpha_0 Ap_0)^T p_0 \]
if \( r_1^T p_0 = 0 \) then \( \alpha_0 = -\frac{r_0^T p_0}{p_0^T Ap_0} \)

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

---

Proof

\[ r_i^T p_i = 0 \quad \text{for } i = 0, \ldots, k-1 \]

Assume true for \( k-1 \)
\[ r_{k-1}^T p_i = 0 \quad \text{for } i = 0, \ldots, k-2 \]

For \( i=k-1 \)
\[ r_k = r_{k-1} + \alpha_{k-1} Ap_{k-1} \]
\[ p_{k-1}^T r_k = p_{k-1}^T r_{k-1} + \alpha_{k-1} p_{k-1}^T Ap_{k-1} \quad \text{By multiplication} \]
If \( p_{k-1}^T r_k = 0 \) then \( \alpha_{k-1} \) is given \[ \alpha_{k-1} = -\frac{r_{k-1}^T p_{k-1}}{p_{k-1}^T Ap_{k-1}} \]
That is 1-D minimizer of quadratic function.

For other vectors \( p_i \)
\[ p_i^T r_k = p_i^T r_{k-1} + \alpha_{k-1} p_i^T Ap_{k-1} = 0 + 0 \quad i = 0, \ldots, k-2 \]
induction
Conjugacy
This implies we have minimized quadratic function in \( k-1 \) variables
Therefore \( r_i^T p_i = 0 \quad \text{for } i = 0, \ldots, k-1 \)

Implies we have minimized quadratic function in \( k \) variables \( \text{QED} \)
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.

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

Or

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

Where $\beta_k$ 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

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

While $r_k \neq 0$

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

end(while)

$p_0$ is steepest descent
$\nabla \phi(x) = Ax - b = r(x)$

A practical form of GC

\[
p_{k+1} = -r_{k+1} + \beta_{k+1} p_k;
\]
\[
p_k = -r_k + \beta_k p_{k-1};
\]
\[
r_k^T p_k = -r_k^T r_k + \beta_k r_k^T p_{k-1};
\]
\[
r_k^T p_j = 0 \quad \text{for} \ i=0, \ldots, k-1
\]
\[
r_k^T p_k = -r_k^T r_k + 0;
\]
\[
r_k^T p_k = -r_k^T r_k \quad \text{(G)}
\]
\[
\alpha_k \leftarrow -\frac{r_k^T p_k}{p_k^T A p_k}; \quad \alpha_k \leftarrow \frac{r_k^T r_k}{p_k^T A p_k};
\]

Theorem 5.2
A practical form of GC

\[ \alpha_k \nabla P_k = r_{k+1} - r_k \]

\[ \alpha_k r_{k+1}^T \nabla P_k = r_{k+1}^T r_{k+1} - r_{k+1}^T r_k \]  
Theorem 5.3

\[ \alpha_k r_{k+1}^T \nabla P_k = r_{k+1}^T r_{k+1} - 0 \]

\[ \alpha_k r_{k+1}^T \nabla P_k = r_{k+1}^T r_{k+1} \]

Now

\[ \alpha_k \nabla P_k = r_{k+1} - r_k \]

\[ \alpha_k p_k^T \nabla P_k = p_k^T r_{k+1} - p_k^T r_k \]

\[ \alpha_k p_k^T \nabla P_k = 0 - p_k^T r_k \]

\[ \alpha_k p_k^T \nabla P_k = 0 + r_k^T r_k \]

\[ \alpha_k p_k^T \nabla P_k = r_k^T r_k \]  
(Already shown)
Algorithm 5.2

Given $x_0$;

1. $r_0 \leftarrow A x_0 - b$, $p_0 \leftarrow -r_0$, $k \leftarrow 0$

While $r_k \neq 0$

2. $\alpha_k \leftarrow \frac{r_k^T r_k}{p_k^T A p_k}$
3. $x_{k+1} \leftarrow x_k + \alpha_k p_k$
4. $r_{k+1} \leftarrow r_k + \alpha_k A p_k$
5. $\beta_{k+1} \leftarrow \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$
6. $p_{k+1} \leftarrow -r_{k+1} + \beta_{k+1} p_k$
7. $k \leftarrow k + 1$

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.

Given $x_0$;

1. $r_0 \leftarrow A x_0 - b$, $p_0 \leftarrow -r_0$, $k \leftarrow 0$

While $r_k \neq 0$

2. $\alpha_k \leftarrow \frac{r_k^T p_k}{p_k^T A p_k}$
3. $x_{k+1} \leftarrow x_k + \alpha_k p_k$
4. $r_{k+1} \leftarrow A x_{k+1} - b$
5. $\beta_{k+1} \leftarrow \frac{r_{k+1}^T A p_k}{p_k^T A p_k}$
6. $p_{k+1} \leftarrow -r_{k+1} + \beta_{k+1} p_k$
7. $k \leftarrow k + 1$

end(while)

5.1

5.2