Lecture-10

Theorems 5.2 and 5.3
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 $k$th iteration generated by the conjugate gradient method is not the solution point $x^*$. The following four properties hold:

\begin{align*}
    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_k\} &= \text{span } \{r_0, A r_0, \ldots, A^k r_0\} \quad (3) \\
    p_i^T A p_i &= 0 \quad \text{for } i = 0, \ldots, k - 1 \quad (4)
\end{align*}

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

Proof

\begin{align*}
    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_k\} &= \text{span } \{r_0, A r_0, \ldots, A^k r_0\} \quad (3) \\
    p_i^T A p_i &= 0 \quad \text{for } i = 0, \ldots, k - 1 \quad (4)
\end{align*}

• 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

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

\[ \text{span } \{r_0, r_1, \ldots, r_k\} = \text{span } \{r_0, Ar_0, \ldots, A^k r_0\} \quad (2) \]

\[ \text{span } \{p_0, p_1, \ldots, p_k\} = \text{span } \{r_0, Ar_0, \ldots, A^k r_0\} \quad (3) \]

\[ p_i^k Ap_i = 0 \quad \text{for } i = 0, \ldots, k - 1 \quad (4) \]

(2) And (3)

**Induction: k=0**

\[ \text{span } \{r_0\} = \text{span } \{r_0\} \quad (2) \]

\[ \text{span } \{p_0\} = \text{span } \{r_0\} \quad (3) \quad p_0 = -r_0 \]

Proof

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

\[ \text{span } \{r_0, r_1, \ldots, r_k\} = \text{span } \{r_0, Ar_0, \ldots, A^k r_0\} \quad (2) \]

\[ \text{span } \{p_0, p_1, \ldots, p_k\} = \text{span } \{r_0, Ar_0, \ldots, A^k r_0\} \quad (3) \]

\[ p_i^k Ap_i = 0 \quad \text{for } i = 0, \ldots, k - 1 \quad (4) \]

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

\[ \text{span } \{r_0, r_1, \ldots, r_k, r_{k+1}\} = \text{span } \{r_0, Ar_0, \ldots, A^{k+1} r_0\} \quad (2) \]

To prove (2), by induction:

\[ r_k \in \text{span } \{r_0, Ar_0, \ldots, A^k r_0\} \quad p_k \in \text{span } \{r_0, Ar_0, \ldots, A^k r_0\} \]

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

\[ r_{k+1} = r_k + \alpha_k Ap_k \quad \text{Therefore } \quad r_{k+1} \in \text{span } \{r_0, Ar_0, \ldots, A^{k+1} r_0\} \]

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

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

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
\[ Ap_i = \frac{(r_i - r_{i-1})}{\alpha_i} \quad \text{for } i = 0, \ldots, k \]

Because
\[ r_{k+1} = r_k + \alpha_k Ap_k \]

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

Span 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

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

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

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

\[ = \text{span} \{ p_0, p_1, \ldots, p_k, r_{k+1} \} \]

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

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

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

QED (3)
Proof

\( r_i^T r_i = 0 \) for \( i = 0, \ldots, k - 1 \) \hspace{1cm} (1)

\( \text{span} \{r_0, r_1, \ldots, r_k\} = \text{span} \{r_0, A r_0, \ldots, A^i r_0\} \) \hspace{1cm} (2)

\( \text{span} \{p_0, p_1, \ldots, p_k\} = \text{span} \{r_0, A r_0, \ldots, A^i r_0\} \) \hspace{1cm} (3)

\( p_i^T A p_i = 0 \) for \( i = 0, \ldots, k - 1 \) \hspace{1cm} (4)

Now Conjugacy (4):

(4) Holds for \( k = 1 \)

\( p_i^T A p_0 = 0 \) \hspace{1cm} (4)

By definition:

\[ p_{k+1} = -r_{k+1} + \beta_{k+1} p_k; \]

\[ p_k^T A p_i = -r_k^T A p_i + \beta_k p_k^T A p_i \] for \( i = 0, 1, \ldots, k \) \hspace{1cm} (F)

By definition:

\[ \beta_k = \frac{r_k^T A p_k}{p_k^T A p_k}; \]

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

\( r_{k+1}^T p_i = 0 \) for \( i = 0, \ldots, k \)

By Theorem 5.2

Proof

\( r_i^T r_i = 0 \) for \( i = 0, \ldots, k - 1 \) \hspace{1cm} (1)

\( \text{span} \{r_0, r_1, \ldots, r_k\} = \text{span} \{r_0, A r_0, \ldots, A^i r_0\} \) \hspace{1cm} (2)

\( \text{span} \{p_0, p_1, \ldots, p_k\} = \text{span} \{r_0, A r_0, \ldots, A^i r_0\} \) \hspace{1cm} (3)

\( p_i^T A p_i = 0 \) for \( i = 0, \ldots, k - 1 \) \hspace{1cm} (4)

\[ p_{k+1}^T A p_i = -r_{k+1}^T A p_i + \beta_{k+1} p_k^T A p_i \] for \( i = 0, 1, \ldots, k \) \hspace{1cm} (F)

\( r_{k+1}^T p_i = 0 \) for \( i = 0, \ldots, k \) \hspace{1cm} (B)

By applying (3)

\[ A p_i \in A \text{span} \{r_0, A r_0, \ldots, A^i r_0\} = \text{span} \{A r_0, A^2 r_0, \ldots, A^{i+1} r_0\} \]

\( \subset \text{span} \{p_0, p_1, \ldots, p_{i+1}\} \) \hspace{1cm} (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

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

\[ \text{span } \{r_0, r_1, \ldots, r_k\} = \text{span } \{r_0, A r_0, \ldots, A^k r_0\} \quad (2) \]

\[ \text{span } \{p_0, p_1, \ldots, p_k\} = \text{span } \{r_0, A r_0, \ldots, A^k r_0\} \quad (3) \]

\[ p_i^T A p_i = 0 \quad \text{for } i = 0, \ldots, k - 1 \quad (4) \]

Now (1)

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

\[ r_k^T p_i = 0 \quad \text{for } i = 0, \ldots, k - 1, \quad k = 1, 2, \ldots, n - 1 \]

By definition

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

\[ r_k^T p_i = 0 = r_k^T (-r_i + \beta_i p_{i-1}) = -r_k^T r_i + \beta_i r_k^T p_{i-1} = -r_k^T r_i \]

\[ r_k^T r_i = 0 \quad \text{for } i = 0, \ldots, k - 1, \quad k = 1, 2, \ldots, n - 1 \quad \text{QED (1)} \]

A practical form of GC

\[ p_{k+1} = -r_{k+1} + \beta_{k+1} p_k; \]

\[ p_k = -r_k + \beta_k p_{k-1}; \quad \quad \quad \text{Theorem 5.2} \]

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

\[ r_k^T p_k = -r_k^T r_k \]

\[ \alpha_k \leftarrow \frac{-r_k^T p_k}{p_k^T A p_k}; \quad \quad \quad \alpha_k \leftarrow \frac{r_k^T r_k}{p_k^T A p_k}; \]
A practical form of GC

\[
\beta_{k+1} \leftarrow \frac{r_{k+1}^T A p_k}{p_k^T A p_k}; \quad \beta_{k+1} \leftarrow \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k};
\]

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

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

Theorem 5.3

\[
\alpha_k r_{k+1}^T A p_k = r_{k+1}^T r_{k+1}
\]

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

Theorem 5.3

\[
\alpha_k r_{k+1}^T A p_k = r_{k+1}^T r_{k+1}
\]

Now

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

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

\[
\alpha_k p_k^T A p_k = 0 + r_k^T r_k
\]

\[
\alpha_k p_k^T A p_k = r_k^T r_k
\]

\[
\beta_{k+1} \leftarrow \frac{r_{k+1}^T A p_k}{p_k^T A p_k}; \quad \beta_{k+1} \leftarrow \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k};
\]
Algorithm 5.2

Given $x_0$:

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

While $r_k \neq 0$

\[
\begin{align*}
\alpha_k & \leftarrow -\frac{r_k^T r_k}{p_k^T Ap_k}; \\
x_{k+1} & \leftarrow x_k + \alpha_k p_k; \\
r_{k+1} & \leftarrow r_k + \alpha_k Ap_k; \\
\beta_{k+1} & \leftarrow \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}; \\
p_{k+1} & \leftarrow -r_{k+1} + \beta_{k+1} p_k; \\
k & \leftarrow k + 1;
\end{align*}
\]

end(while)

5.2

5.1

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.

Algorithm 5.2

Given $x_0$:

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

While $r_k \neq 0$

\[
\begin{align*}
\alpha_k & \leftarrow -\frac{r_k^T r_k}{p_k^T Ap_k}; \\
x_{k+1} & \leftarrow x_k + \alpha_k p_k; \\
r_{k+1} & \leftarrow Ax_{k+1} - b; \\
\beta_{k+1} & \leftarrow \frac{r_{k+1}^T Ap_{k+1}}{p_k^T Ap_k}; \\
p_{k+1} & \leftarrow -r_{k+1} + \beta_{k+1} p_k; \\
k & \leftarrow k + 1;
\end{align*}
\]

end(while)

5.2
Proof

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

Now (1)

Since the direction set is conjugate by theorem 5.2

\[ r_k^T p_i = 0 \quad \text{for } i = 0, \ldots, k-1, \quad k = 1, 2, \ldots, n-1 \]
\[ p_i = -r_i + \beta_i p_{i-1} \quad \text{By definition} \]
\[ r_i \in \text{span } \{p_i \} \quad \text{for } i = 0, \ldots, k-1 \]
\[ r_i = a p_i + b p_{i-1} \]
\[ p_{k+1} = -r_{k+1} + \beta_{k+1} p_k; \]
\[ r_i = a p_i + b p_{i-1} \]
\[ r_k^T p_i = 0 = r_k^T (cr_i + dp_{i-1}) = cr_k^T r_i + dr_k^T p_{i-1} = cr_k^T r_i \]
\[ r_i^T r_i = 0 \quad \text{for } i = 0, \ldots, k-1, \quad k = 1, 2, \ldots, n-1 \quad \text{QED} (1) \]

A practical form of GC

\[ \beta_{k+1} \leftarrow \frac{r_{k+1}^T A p_k}{p_k^T A p_k}; \quad \beta_k \leftarrow \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}; \]

\[ \alpha_k A p_k = r_{k+1} - r_k \]
\[ \alpha_k r_{k+1}^T A p_k = r_{k+1}^T r_{k+1} - r_{k+1}^T r_k \quad \text{Theorem 5.3} \]
\[ p_k = -r_k + \beta_k p_{k-1}; \]
\[ r_k = p_k + \beta_k p_{k-1}; \]
\[ r_{k+1}^T r_k = r_{k+1}^T p_k + \beta_k r_k^T p_{k-1}; \]
\[ r_k^T p_i = 0 \quad \text{for } i = 0, \ldots, k-1 \]
\[ r_{k+1}^T r_k = 0 \]
\[ \alpha_k r_{k+1}^T A p_k = r_{k+1}^T r_{k+1} \]