Properties of CG directions I

Theorem

After $j$ iterations, we have

\[ r_j = r_{j-1} - \alpha_j A p_j \]

\[ P_j^T r_j = 0 \]

\[ \text{span}\{p_1, \ldots, p_j\} = \text{span}\{r_0, \ldots, r_{j-1}\} = \text{span}\{b, Ab, \ldots, A^{j-1}b\} \]

\[ r_i^T r_j = 0 \text{ for all } i \neq j \]
Properties of CG directions II

- We prove only the first result while others are omitted

\[ r_j = b - Ax_j \]
\[ = b - Ax_{j-1} + A(x_{j-1} - x_j) \]
\[ = r_{j-1} - \alpha_j Ap_j \]

- From this theorem, \( r_i, r_j \) are mutually orthogonal
Recall that

\[ p_k = r_{k-1} - AP_{k-1}z_{k-1} \]

Now we want to find \( z_{k-1} \)

\( z_{k-1} \) is a vector with length \( k - 1 \)

\[ z_{k-1} = \begin{bmatrix} w \\ \mu \end{bmatrix}, \ w : (k - 2) \times 1, \ \mu : 1 \times 1 \]
Conjugate Gradient Method II

\[ p_k = r_{k-1} - AP_{k-1}z_{k-1} \quad (2) \]
\[ = r_{k-1} - AP_{k-2}w - \mu Ap_{k-1} \]
\[ = (1 + \frac{\mu}{\alpha_{k-1}})r_{k-1} + s_{k-1} \quad (3) \]

- From the earlier theorem

\[ r_{k-1} = r_{k-2} - \alpha_{k-1}Ap_{k-1} \]
Conjugate Gradient Method III

- We have

$$s_{k-1} = -\frac{\mu}{\alpha_k} r_{k-1} - AP_{k-2}w - \mu Ap_{k-1}$$

$$= -\frac{\mu}{\alpha_k} r_{k-2} - AP_{k-2}w$$

(4)

- We have

$$r_i^T r_j = 0, \forall i \neq j$$

(5)

and

$$AP_{k-2}w \in \text{span}\{Ap_1, \ldots, Ap_{k-2}\}$$

$$= \text{span}\{Ab, \ldots, A^{k-2}b\}$$

$$\subset \text{span}\{r_0, \ldots, r_{k-2}\}$$
Conjugate Gradient Method IV

- Hence \( r_{k-1}^T (A P_{k-2} w) = 0 \). Further, from (4) and (5)

\[
s_{k-1}^T r_{k-1} = 0 \tag{6}
\]

- Recall from earlier results, our job now is to find \( z_{k-1} \) such that

\[
\| r_{k-1} - A P_{k-1} z \| \]

is minimized
The reason of minimizing

$$\| r_{k-1} - AP_{k-1}z \|$$

instead of

$$\| p - r_{k-1} \|_2, \ p \in \text{span}\{A\rho_1, \ldots, A\rho_{k-1}\}^\perp$$  \hspace{1cm} (7)

is that (7) is constrained.
From (2) and (3) we select $\mu$ and $w$ to minimize

$$\|(1 + \frac{\mu}{\alpha_k})r_{k-1} + s_{k-1}\|^2$$

From (6), this is equivalent to minimizing

$$\|(1 + \frac{\mu}{\alpha_k})r_{k-1}\|^2 + \| - \frac{\mu}{\alpha_k}r_{k-2} - AP_{k-2}w\|^2$$
If an optimal solution is \((\mu^*, w^*)\), then

\[
\left\| - \frac{\mu^*}{\alpha_{k-1}} r_{k-2} - AP_{k-2} w^* \right\| = \left\| - \frac{\mu^*}{\alpha_{k-1}} \right\| \left\| r_{k-2} - AP_{k-2} \frac{w^*}{-\mu^*/\alpha_{k-1}} \right\|
\]

and

\[
\frac{w^*}{-\mu^*/\alpha_{k-1}}
\]

must be the solution of

\[
\min_z \left\| r_{k-2} - AP_{k-2} z \right\|
\]
From the earlier lemma, the solution of

\[ \min_z \| r_{k-2} - AP_{k-2}z \| \]

is

\[ p_{k-1} = r_{k-2} - AP_{k-2}z_{k-2} \]

Therefore, \( s_{k-1} \) is a multiple of \( p_{k-1} \)

From (3),

\[ p_k \in \text{span}\{r_{k-1}, p_{k-1}\} \]
Assume

\[ p_k = r_{k-1} + \beta_k p_{k-1} \]  \hspace{1cm} (8)

This assumption is fine as we will adjust the step size \( \alpha \) for the direction \( p_k \).

Therefore, finding a direction parallel to the real solution of \( \min \| p - r_{k-1} \| \) is enough.

From the A-conjugacy of \( p_i, \forall i \) and (1), we respectively have

\[ p_{k-1}^T A p_k = 0, \quad p_{k-1}^T r_{k-1} = 0 \]
With (8),

\[ Ap_k = Ar_{k-1} + \beta_k Ap_{k-1} \]

\[ 0 = p_{k-1}^T Ap_k = p_{k-1}^T Ar_{k-1} + \beta_k p_{k-1}^T Ap_{k-1} \]

\[ \beta_k = -\frac{p_{k-1}^T Ar_{k-1}}{p_{k-1}^T Ap_{k-1}} \]

\[ \alpha_k = \frac{p_k^T r_{k-1}}{p_k^T Ap_k} = \frac{(r_{k-1} + \beta_k p_{k-1})^T r_{k-1}}{p_k^T Ap_k} = \frac{r_{k-1}^T r_{k-1}}{p_k^T Ap_k} \]

The conjugate gradient method
Conjugate Gradient Method XI

\[ k = 0; \ x_0 = 0; \ r_0 = b \]

while \( r_k \neq 0 \)

\[ k = k + 1 \]

if \( k = 1 \)

\[ p_1 = r_0 \]

else

\[ \beta_k = -p_{k-1}^T Ar_{k-1} / p_{k-1}^T Ap_{k-1} \]

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

end

\[ \alpha_k = r_{k-1}^T r_{k-1} / p_k^T Ap_k \]

\[ x_k = x_{k-1} + \alpha_k p_k \]
Conjugate Gradient Method XII

\[ r_k = b - Ax_k \]
end

- The main cost at each iteration is for three matrix-vector products

\[ Ar_{k-1}, Ap_{k-1}, Ax_k \]