Properties of CG directions I

Theorem

After j iterations, we have

$$
\begin{aligned}
r_j &= r_{j-1} - \alpha_j A p_j \\
P_j^T r_j &= 0 \\
span\{p_1, \ldots, p_j\} &= span\{r_0, \ldots, r_{j-1}\} \\
&= span\{b, Ab, \ldots, A^{j-1} b\}
\end{aligned} \tag{1}
$$

$$r_i^T r_j = 0 \text{ for all } i \neq j$$

Properties of CG directions II

- We prove only the first result while others are omitted

$$
\begin{aligned}
r_j &= b - Ax_j \\
&= b - Ax_{j-1} + A(x_{j-1} - x_j) \\
&= r_{j-1} - \alpha_j A p_j
\end{aligned}
$$

- From this theorem, r_i, r_j are mutually orthogonal

Conjugate Gradient Method I

- 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

$$
\begin{aligned}
p_k &= r_{k-1} - AP_{k-1}z_{k-1} && (2) \\
&= r_{k-1} - AP_{k-2}w - \mu Ap_{k-1} \\
&= (1 + \frac{\mu}{\alpha_{k-1}})r_{k-1} + s_{k-1} && (3)
\end{aligned}
$$

- From the earlier theorem

$$r_{k-1} = r_{k-2} - \alpha_{k-1}Ap_{k-1}$$

Conjugate Gradient Method III

- We have

$$s_{k-1} = -\frac{\mu}{\alpha_{k-1}} r_{k-1} - AP_{k-2} w - \mu A p_{k-1}$$
$$= -\frac{\mu}{\alpha_{k-1}} r_{k-2} - AP_{k-2} w \quad (4)$$

- We have

$$r_i^T r_j = 0, \forall i \neq j \quad (5)$$

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 $r_{k-1}^T(AP_{k-2}w) = 0$. 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

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

is minimized

Conjugate Gradient Method V

- The reason of minimizing

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

instead of

$$\|p - r_{k-1}\|_2, p \in \text{span}\{Ap_1, \ldots, Ap_{k-1}\}^{\perp} \tag{7}$$

is that (7) is constrained.

Conjugate Gradient Method VI

- From (2) and (3) we select μ and w to minimize

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

From (6), this is equivalent to minimizing

$$\|(1 + \frac{\mu}{\alpha_{k-1}})r_{k-1}\|^2 + \| - \frac{\mu}{\alpha_{k-1}} r_{k-2} - AP_{k-2}w\|^2$$

Conjugate Gradient Method VII

- If an optimal solution is (μ^*, w^*), then

$$
\begin{aligned}
&\| -\frac{\mu^*}{\alpha_{k-1}} r_{k-2} - AP_{k-2} w^* \| \\
=&| -\frac{\mu^*}{\alpha_{k-1}} | \| r_{k-2} - AP_{k-2} \frac{w^*}{-\mu^*/\alpha_{k-1}} \|
\end{aligned}
$$

and

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

must be the solution of

$$
\min_z \| r_{k-2} - AP_{k-2} z \|
$$

Conjugate Gradient Method VIII

- 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),

$$p_k \in \text{span}\{r_{k-1}, p_{k-1}\}$$

Conjugate Gradient Method IX

- Assume

$$p_k = r_{k-1} + \beta_k p_{k-1} \tag{8}$$

  This assumption is fine as we will adjust the step size α 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

$$p_{k-1}^T A p_k = 0, p_{k-1}^T r_{k-1} = 0$$

Conjugate Gradient Method X

- With (8),

$$
\begin{aligned}
Ap_k &= Ar_{k-1} + \beta_k Ap_{k-1} \\
0 = p_{k-1}^T Ap_k &= p_{k-1}^T Ar_{k-1} + \beta_k p_{k-1}^T Ap_{k-1} \\
\beta_k &= -\frac{p_{k-1}^T Ar_{k-1}}{p_{k-1}^T Ap_{k-1}} \\
\alpha_k = \frac{p_k^T r_{k-1}}{p_k^T Ap_k} &= \frac{(r_{k-1} + \beta_k p_{k-1})^T r_{k-1}}{p_k^T Ap_k} = \frac{r_{k-1}^T r_{k-1}}{p_k^T Ap_k}
\end{aligned}
$$

- 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
        $\beta_k = -p_{k-1}^T A r_{k-1} / p_{k-1}^T A p_{k-1}$
        $p_k = r_{k-1} + \beta_k p_{k-1}$
    end
    $\alpha_k = r_{k-1}^T r_{k-1} / p_k^T A p_k$
    $x_k = x_{k-1} + \alpha_k p_k$

Conjugate Gradient Method XII

$$r_k = b - Ax_k$$

end

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

$$Ar_{k-1}, Ap_{k-1}, Ax_k$$