Directions in CG I

- For selecting directions, we hope that $p_1, \ldots, p_n$ are linearly independent and

$$x_k = \arg \min_{x \in x_0 + \text{span}\{p_1, \ldots, p_k\}} f(x) \tag{1}$$

- With (1),

$$\text{span}\{p_1, \ldots, p_n\} = R^n,$$

so

$$x_n = \arg \min_{x \in R^n} f(x)$$

Directions in CG II

Then

$$Ax_n = b$$

and the procedure stops at most n iterations

- But how to maintain (1)?
- Let

$$x_k = x_0 + P_{k-1}y + \alpha p_k,$$

where

$$P_{k-1} = \left[p_1, \ldots, p_{k-1}\right], y \in R^{k-1}, \alpha \in R$$

Directions in CG III

$$
\begin{aligned}
& f(x_k) \\
=& \frac{1}{2}(x_0 + P_{k-1}y + \alpha p_k)^T A(x_0 + P_{k-1}y + \alpha p_k) \\
& - b^T(x_0 + P_{k-1}y + \alpha p_k) \\
=& \frac{1}{2}(x_0 + P_{k-1}y)^T A(x_0 + P_{k-1}y) - b^T(x_0 + P_{k-1}y) \\
& + \alpha p_k^T A(x_0 + P_{k-1}y) - b^T(\alpha p_k) + \frac{\alpha^2}{2} p_k^T A p_k \\
=& f(x_0 + P_{k-1}y) + \alpha p_k^T A P_{k-1} y - \alpha p_k^T r_0 + \frac{\alpha^2}{2} p_k^T A p_k
\end{aligned}
$$

Directions in CG IV

$$\min_{x \in x_0 + \text{span}\{p_1, \ldots, p_k\}} f(x)$$
$$= \min_{y, \alpha} f(x_0 + P_{k-1} y + \alpha p_k)$$

is difficult because the term

$$\alpha p_k^T A P_{k-1} y$$

involves both

$$\alpha \text{ and } y$$

Directions in CG V

- If

$$p_k \in \text{span}\{Ap_1, \ldots, Ap_{k-1}\}^\perp,$$

then

$$\alpha p_k^T A P_{k-1} y = 0$$

and

$$\min_{x \in x_0 + \text{span}\{p_1,\ldots,p_k\}} f(x)$$

$$= \min_y f(x_0 + P_{k-1}y) + \min_\alpha (-\alpha p_k^T r_0 + \frac{\alpha^2}{2} p_k^T A p_k),$$

Directions in CG VI

- Then we have two independent optimization problems
- Therefore, we require p_k is A-conjugate to $p_1, \ldots, p_{k-1}$. That is

$$p_i^T A p_k = 0, i = 1, \ldots, k-1$$

- By induction,

$$x_{k-1} = \arg\min_y f(x_0 + P_{k-1} y)$$

See also (1)

Directions in CG VII

- The solution of the second problem is

$$\alpha_k = \frac{p_k^T r_0}{p_k^T A p_k}$$

- Because of A-conjugacy,

$$\begin{aligned} p_k^T r_{k-1} &= p_k^T (b - A x_{k-1}) \\ &= p_k^T (b - A(x_0 + P_{k-1} y_{k-1})) = p_k^T r_0 \end{aligned}$$

Directions in CG VIII

- We have

$$\alpha_k = \frac{p_k^T r_0}{p_k^T A p_k} = \frac{p_k^T r_{k-1}}{p_k^T A p_k} \text{ and}$$

$$x_k = x_{k-1} + \alpha_k p_k$$

- New algorithm

Directions in CG IX

$k = 0$; $x_0 = 0$; $r_0 = b$
while $r_k \neq 0$
    $k = k + 1$
    Choose any $p_k \in \text{span}\{Ap_1, \ldots, Ap_{k-1}\}^\perp$
        such that $p_k^T r_{k-1} \neq 0$
    $\alpha_k = p_k^T r_{k-1} / p_k^T A p_k$
    $x_k = x_{k-1} + \alpha_k p_k$
    $r_k = b - Ax_k$
end

- We still haven't specified p_k yet
- One way is to minimize the distance to r_{k-1}:

Directions in CG X

- Reason: r_{k-1} is now the negative gradient direction
- The algorithm becomes

$k = 0$; $x_0 = 0$; $r_0 = b$
while $r_k \neq 0$
    $k = k + 1$
    if $k = 1$
        $p_1 = r_0$
    else
        Let p_k minimize $\|p - r_{k-1}\|_2$ over all vectors
            $p \in \text{span}\{Ap_1, \ldots, Ap_{k-1}\}^\perp$
    end

Directions in CG XI

$$\alpha_k = p_k^T r_{k-1} / p_k^T A p_k$$
$$x_k = x_{k-1} + \alpha_k p_k$$
$$r_k = b - A x_k$$

end

Directions in CG XII

Lemma

If p_k *minimizes* $\|p - r_{k-1}\|_2$ *over all vectors* $p \in span\{Ap_1, \ldots, Ap_{k-1}\}^{\perp}$, *then*

$$p_k = r_{k-1} - AP_{k-1}z_{k-1} \tag{2}$$

where z_{k-1} *solves*

$$\min_z \|r_{k-1} - AP_{k-1}z\|_2, \tag{3}$$

$P_k = [p_1, \ldots, p_k]$*: an* $n \times k$ *matrix*

Directions in CG XIII

Proof:

- Let z_{k-1} be the solution of (3) and define

$$p = r_{k-1} - AP_{k-1}z_{k-1}$$

- The optimality condition of (3) is

$$P_{k-1}^T A^T (r_{k-1} - AP_{k-1}z) = 0$$

  Thus

$$p^T A p_i = 0, \forall i = 1, \ldots, k-1$$

Directions in CG XIV

- Moreover, p is the orthogonal projection of r_{k-1} into

$$\text{span}\{Ap_1, \ldots, Ap_{k-1}\}^\perp, \tag{4}$$

  (details not shown) so it is the closest vector in (4) to r_{k-1}
- Thus

$$p = p_k$$