Conjugate Gradient Method I

- Further simplification

$$
\begin{aligned}
r_k &= r_{k-1} - \alpha_k A p_k \\
r_{k-1} &= r_{k-2} - \alpha_{k-1} A p_{k-1} \\
r_{k-1}^T r_{k-1} &= r_{k-1}^T r_{k-2} - \alpha_{k-1} r_{k-1}^T A p_{k-1} \\
&= 0 - \alpha_{k-1} r_{k-1}^T A p_{k-1} \\
r_{k-2}^T r_{k-1} &= r_{k-2}^T r_{k-2} - \alpha_{k-1} r_{k-2}^T A p_{k-1} \\
&= r_{k-2}^T r_{k-2} - \alpha_{k-1} (p_{k-1} - \beta_{k-1} p_{k-2})^T A p_{k-1} \\
&= r_{k-2}^T r_{k-2} - \alpha_{k-1} p_{k-1}^T A p_{k-1} = 0 \\
r_{k-2}^T r_{k-2} &= \alpha_{k-1} p_{k-1}^T A p_{k-1}
\end{aligned}
$$

Conjugate Gradient Method II

$$\beta_k = \frac{-p_{k-1}^T A r_{k-1}}{p_{k-1}^T A p_{k-1}} = \frac{r_{k-1}^T r_{k-1}/\alpha_{k-1}}{r_{k-2}^T r_{k-2}/\alpha_{k-1}} = \frac{r_{k-1}^T r_{k-1}}{r_{k-2}^T r_{k-2}}$$

- A simplified CG procedure:

  $k = 0$; $x_0 = 0$; $r_0 = b$
  while $r_k \neq 0$
      $k = k + 1$
      if $k = 1$
          $p_1 = r_0$
      else
          $\beta_k = r_{k-1}^T r_{k-1} / r_{k-2}^T r_{k-2}$
          $p_k = r_{k-1} + \beta_k p_{k-1}$

Conjugate Gradient Method III

end

$\alpha_k = r_{k-1}^T r_{k-1} / p_k^T A p_k$

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

$r_k = r_{k-1} - \alpha_k A p_k$

end

- One matrix vector product
- $r_k \neq 0$ is not a practical termination criterion
- Also we should avoid calculating

$$r_{k-1}^T r_{k-1}, r_{k-2}^T r_{k-2}$$

at each step

Conjugate Gradient Method IV

- The final version

$$
\begin{aligned}
& k = 0;\ x = 0;\ r = b;\ \rho_0 = \|r\|_2^2 \\
& \text{while } \sqrt{\rho_k} > \epsilon \|b\|_2 \text{ and } k < k_{\max} \\
& \qquad k = k + 1 \\
& \qquad \text{if } k = 1 \\
& \qquad\qquad p = r \\
& \qquad \text{else} \\
& \qquad\qquad \beta = \rho_{k-1}/\rho_{k-2} \\
& \qquad\qquad p = r + \beta p \\
& \qquad \text{end} \\
& \qquad w = Ap
\end{aligned}
$$

Conjugate Gradient Method V

$$\alpha = \rho_{k-1}/p^T w$$
$$x = x + \alpha p$$
$$r = r - \alpha w$$
$$\rho_k = \|r\|_2^2$$

end

- Numerical error will cause the number of iterations $> n$
- For what kind of problems slow convergence occurs?

Conjugate Gradient Method VI

- Convergence properties

Theorem

If $A = I + B$ is an $n \times n$ symmetric positive definite matrix and rank$(B) = r$, then the conjugate gradient method converges in at most $r + 1$ steps

We omit the proof

Conjugate Gradient Method VII

- The case of $A = I$:

$$
\begin{aligned}
& x_0 = 0, r_0 = b \\
& p_1 = r_0 = b \\
& \alpha = r_0^T r_0 / p_1^T A p_1 = b^T b / b^T A b = 1 \\
& x_1 = p_1 = b \\
& r_1 = r_0 - \alpha_1 A p_1 = b - b = 0
\end{aligned}
$$

The conjugate gradient method stops in one iteration

Conjugate Gradient Method VIII

- An error bound in terms of the norm $x^T Ax$:

Theorem

If $Ax = b$,

$$\|x - x_k\|_A \leq 2\|x - x_0\|_A \left(\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1}\right)^k,$$

where

$$\|x\|_A = \sqrt{x^T Ax}$$

- $\sqrt{\kappa}$: the condition number of A

Conjugate Gradient Method IX

- If
$$\sqrt{\kappa} \to 1,$$
then
$$(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1})^k$$
is smaller

Discussion

- In general, if the condition of A is better $\Rightarrow$ fewer CG iterations
- Where is PD of A used? Recall that we have

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

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

For the second part,

$$\frac{1}{2}\alpha^2 p^T A p + \cdots$$

we need $p^T A p > 0$ for the minimization.