CS475 / CS675
Lecture 9: May 31, 2016

Conjugate Gradient

Readings: [TB] Chapt 38, p. 293-301

[Saad] Sect 5.3 and 6.7

Method of Conjugate Directions

- Each new direction is “orthogonal” in some sense to previous search directions
- It turns out the best strategy is to require A-orthogonality (or conjugacy)
  - i.e., p, q are A-orthogonal if $p^T A q = 0$
- Definition: the A-inner product is defined as:

$$(p, q)_A \equiv p^T A q$$

- The A-norm is defined as $||p||_A = \sqrt{(p, p)_A}$

Gram-Schmidt Process

- Construct a set of orthogonal vectors
- Suppose the previous search directions $p^0, p^1, \dots, p^{k-1}$ are A-orthogonal. Given the current r^k, construct p^k
- Let $p^k = r^k + \sum_{i=0}^{k-1} \beta_i p^i$ then

$$\left(p^k, p^j\right)_A = 0 \Longrightarrow \left(r^k, p^j\right)_A + \left(\sum_{i=0}^{k-1} \beta_i p^i, p^j\right)_A = 0$$

$$\left(r^k, p^j\right)_A + \beta_j \left(p^j, p^j\right)_A = 0$$

$$\beta_j = -\frac{\left(r^k, p^j\right)_A}{\left(p^j, p^j\right)_A}$$

Conjugate Gradient Picture

Conjugate Gradient Method

- Construct a set of A-orthogonal search vectors $\{p^k\}$ by the residual vectors $\{r^k\}$
  - i.e., $p^k = r^k - \sum_{i=0}^{k-1} \frac{(r^k, p^i)_A}{(p_i, p_i)_A} p^i$

Conjugate Gradient (CG)

$x^0 =$ initial guess; $r^0 = b - Ax^0$

for $k = 0, 1, \dots, n-1$

compute $p^k = r^k - \sum_{i=0}^{k-1} \frac{(r^k, p^i)_A}{(p^i, p^i)_A} p^i$

$x^{k+1} = x^k + \alpha_k p^k$

$r^{k+1} = r^k - \alpha_k A p^k$

end

Notes:

1. $\alpha_k = (r^k, p^k)/(p^k, p^k)_A$
2. $r^{k+1} = b - Ax^{k+1}$

Useful Facts

- Krylov subspace

$$span\{p^0, \dots, p^{k-1}\} = span\{r^0, \dots, r^{k-1}\}$$
$$= span\{r^0, Ar^0, \dots, A^{k-1}r^0\}$$
$$\equiv K_k(A, r^0)$$
$$= k\text{-dim Krylov subspace}$$

- Since $r^k \perp span\{r^0, r^1, \dots, r^{k-1}\}$
  - i.e. $\left(r^k, r^j\right) = 0 \quad j = 0,1, \dots, k-1$

  Then $r^k \perp span\{p^0, \dots, p^{k-1}\}$

Useful Facts

- $\left(r^k, p^k\right) = (r^k, r^k)$
  - Proof:

Useful Facts

- $\left(r^k, p^i\right)_A = 0 \qquad i = 0,1,\ldots,k-2$
  - Proof:

Speed up

- By using the previous facts, we can speed up the computation of p^k by simplifying the summation
- $p^k = r^k - \frac{(r^k, p^{k-1})_A}{(p^{k-1}, p^{k-1})_A} p^{k-1}$
  - Proof:

Speed up

- Speed up the computation of $\left(r^k, p^{k-1}\right)_A$ by avoiding any matrix multiplication
- $\left(r^k, p^{k-1}\right)_A = -\frac{1}{\alpha_{k-1}}(r^k, r^k)$
  - Proof:

Speed up

- Similarly, speed up the computation of $\left(p^{k-1}, p^{k-1}\right)_A$ by avoiding any matrix multiplication
- $\left(p^{k-1}, p^{k-1}\right)_A = \frac{1}{\alpha_{k-1}}(r^{k-1}, r^{k-1})$
  - Proof:

Efficient Conjugate Gradient (CG)

x^0= initial guess; $r^0 = b - Ax^0$
for $k = 0,1,\ldots,n-1$

$$\beta_{k-1} = (r^k, r^k)/(r^{k-1}, r^{k-1}) \qquad (\beta_{-1} = 0)$$
$$p^k = r^k + \beta_{k-1} p^{k-1}$$
$$\alpha_k = (r^k, r^k)/(p^k, Ap^k)$$
$$x^{k+1} = x^k + \alpha_k p^k$$
$$r^{k+1} = r^k - \alpha_k A p^k$$

end

Notes

1. Only 1 matrix-vector multiplication, 2 inner products
2. At most n A-orthogonal vectors in $\Re^n$. Terminate in at most n steps → exact solution