Linear Algebra 2

fredag 27. oktober 2017 08:53

1. We continue to examine the conjugate gradient. First, we place it in the class of Krylov subspace methods

In general, for a given matrix A and vector v, the Krylov subspace of order n is given by

$$K_n(A, v) = \operatorname{span}\{v, Av, \dots, A^{n-1}v\}.$$

Theorem: Let $r^{(k)}$ be the residuals of the conjugate gradient method applied to $Ax = b$, and $p^{(k)}$ the search directions. Then

$$\operatorname{span}\{r^{(0)}, \dots, r^{(n)}\} = K_{n+1}(A, r^{(0)})$$

$$= \operatorname{span}\{p^{(0)}, \dots, p^{(n)})$$

Proof: by induction. Clearly true for $n=0$.

Suppose

$$r^{(k)}, \; p^{(k)} \in K_{k+1}(A, r^{(0)})$$

Note that we have

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

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

$\Rightarrow \; b - Ax^{(k+1)} = b - Ax^{(k)} - \alpha_k A p^{(k)}$)

Now $Ap^{(k)} \in K_{k+2}$

hence $r^{(k)} \in K_{k+2}$ also

i.e. (1) $\mathrm{span}\{r^{(0)}, \dots, r^{(k+1)}\} \subset K_{k+2}(A, r^{(0)})$

To prove the other inclusion, note that

$$A^{k+1} r^{(0)} = A(A^k r^{(0)})$$
$$\in \mathrm{span}(Ap^{(0)}, \dots, Ap^{(k)})$$

(using induction hypotheses that
$\mathrm{span}\{r^{(0)}, \dots, r^{(k)}\} = \mathrm{span}\{p^{(0)}, \dots, p^{(k)}\}$
and $A^k r^{(0)} \in \mathrm{span}\{r^{(0)}, \dots, r^{(k)}\}$)

Now $$A p^{(i)} = \frac{r^{(i+1)} - r^{(i)}}{\alpha_i}$$

$$\Rightarrow A^{k+1} r^{(0)} \in \mathrm{span}\{r^{(0)}, \dots, r^{(k+1)}\}$$

$\Rightarrow$ (2) $\mathrm{span}\{r^{(0)}, \dots, r^{(k+1)}\} \supset K_{k+2}(A, r^{(0)})$

(by induction hypothesis)

combining (1) and (2) gives first result.

For second, we proceed by induction also:

$\mathrm{span}\{p^{(0)}, \dots, p^{(k+1)}\}$

$= \mathrm{span}\{p^{(0)}, \dots, p^{(k)}, r^{(k+1)}\}$

$= \mathrm{span}\{r^{(0)}, \dots, A^k r^{(0)}, r^{(k+1)}\}$ (induction and first part of result)

$= K_{k+2}(A, r^{(0)})$ (first result)

---

2. Krylov subspace methods:

For $Ax = b$ when A is not symmetric, positive definite

we cannot apply conjugate gradient, but can try Krylov subspace methods.

Idea: let $W_k = \{ x^{(0)} + y, \ y \in K_k(A, r^{(0)}) \}$

so that $x^{(k)} \in W_k$, i.e.

$$x^{(k)} = x^{(0)} + q_{k-1}(A)\, r^{(0)}$$

for some polynomials of deg $k-1$. We then choose $x^{(k)}$ so that it satisfies some minimization criteria

- <u>Lemma</u>: $K_m(A, v)$ has dimension equal to degree of v wrt A, i.e. the minimum degree of a nontrivial monic polynomial p: $p(A)\, v = 0$

(NB $\dim K_m(A, v) \leq n$ due to Cayley-Hamilton, where n = dimension of underlying space $\mathbb{R}^n$)

Possible minimization criteria:

a) <u>Full orthogonalization</u> (FOM) ... method

choose $x^{(k)}$: associated residual $r^{(k)}$ is orthogonal to $K_k(A, r^{(0)})$

i.e. $v^T(b - Ax^{(k)}) = 0 \quad \forall v \in K_k(A, r^{(0)})$

b) Minimize residual (GMRES)

(generalized minimum residual)

choose $x^{(k)}$ to minimize $\|r^{(k)}\|$

(Euclidean norm)

i.e. $x^{(k)} = \underset{v \in W_k}{\arg\min} \|b - Av\|$

---

3. Practicalities: To execute a Krylov subspace method, we first need an orthonormal basis of $K_m(A, r^{(0)})$

(the $r^{(k)}$ provide this in conjugate gradient case).

We can generate this by, e.g. Gram-Schmidt

generate first $v_1 = \frac{v}{\|v\|}$, then

$$\hat{v}_k = Av_{k-1}$$

$$\bar{v}_k = \hat{v}_k - \sum_{j=1}^{k-1} v_j^T \hat{v}_k \hat{v}_j$$

$$v_k = \frac{\bar{v}_k}{\|\bar{v}_k\|}$$

If $v_k = 0$, have experienced breakdown

(will mean that the Krylov subspaces have stopped increasing)

Now, we can collect $v_1, \dots, v_m$ into a matrix $V_m = [v_1, \dots, v_m]$

Our methods will generally take the form

$$x^{(k)} = x^{(0)} + V_k z^{(k)}$$

a) <u>Arnoldi method (FOM)</u> :

Here $V_k^T (b - A x^{(k)}) = 0$

$$\Rightarrow V_k^T (b - A x^{(0)} - A V_k z^{(0)}) = 0$$

Now $V_k^T (b - A x^{(0)}) = V_k^T r^{(0)} = \| r^{(0)} \| e_1$

(by construction of V)

hence must solve

$$\underbrace{(V_k^T A V_k)} z^{(k)} = \| r^{(0)} \| e_1$$

we can prove this is upper Hessenberg $\Rightarrow$ system rapidly solvable.

$\rightarrow$ in practice, $H_k = V_k^T A V_k$ is constructed alongside V.

<u>Theorem</u>: Arnoldi method gives exact solution in at most n iterations.

solution in at most n iterations.

as $K_n(A, r^{(0)}) = \mathbb{R}^n$, so $r^{(n)} = 0$.

- in practice, stop when residual is low, but we don't compute this explicitly, use eg

$$\| r^{(k)} \| = h_{k+1,k} \, | e_k^T z^{(k)} |$$

and stop if $\dfrac{\| r^{(k)} \|}{\| r^{(0)} \|} \leq \varepsilon$

tolerance.

- Should also apply with a preconditioner. etc.

b) GMRES: Here we minimize

$\| r^{(k)} \| =$

Now we have

$$r^{(k)} = r^{(0)} - A V_k z^{(k)}$$

$$= V_{k+1} \left(\| r^{(0)} \| e_1 - \hat{H}_k z^{(k)} \right)$$

where $\hat{H}_k = V_{m+1}^T A V_m$

(so H_k is restriction of $\hat{H}_k$ by removing $(k+1)^{th}$ row)

i.e. we solve the least squares

problem to minimize

$$\| \|r^{(0)}\| e_1 - \hat{H}_k z^{(k)} \|,$$ using

standard linear least squares.

- Again, stops after at most n iterations.

c) Lanczos: this is the name for Arnoldi when A is symmetric

⇒ H_k are symmetric, tridiagonal (so linear systems are solved quicker).

→ can be applied in non-symmetric cases in a variant using two different Krylov subspaces