Note Title Arnoldi algorithm 2024-11-27

Want to find orthonormal basis $q_1, q_2, \dots q_j$ of $K_j(A, y) = \text{span}(y, Ay, A^2y, \dots, A^{j-1}y)$

$$q_i^T q_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Assume you have $q_1, \dots q_j$ basis of $K_j(A, y)$, let us extend it with q_{j+1}

We assume that at this step q_j has degree $j-1$

$$w = Aq_j \quad \in K_{j+1}(A, y)$$

$$w \leftarrow w - q_1(q_1^T w)$$

, in this way

$$q_1^T \left(w - q_1 (q_1^T w)\right) = q_1^T w - \overbrace{(q_1^T q_1)}^{1}(q_1^T w) = 0$$

$$w \leftarrow w - q_2(q_2^T w)$$

```
function Q = arnoldi(A, y, n)
m = length(y);
Q = zeros(m,n+1);
Q(:,1) = y / norm(y);
for j = 1:n
    w = A * Q(:,j);
    for i = 1:j
        w = w - Q(:,i) * (Q(:,i)'*w);
    end
    nw = norm(w);
    Q(:,j+1) = w / nw;
end
```

Cost: n mat-vec products $+ \ O(n^2 m)$ ops

(typically, $n \ll m$)

At each iteration of the outer "for" loop, we have:

$$A q_j = q_1 \underset{= q_1^T w}{h_{1j}} + q_2 \underset{= q_2^T w}{h_{2j}} + q_3 h_{3j} + \dots + q_{j+1} \underset{= \|w\|}{h_{j+1,j}}$$

$$A \underbrace{\left[q_1 \,\middle|\, q_2 \,\middle|\, \cdots \,\middle|\, q_j \right]} = \left[q_1 \,\middle|\, q_2 \,\middle|\, q_3 \,\cdots\, q_{j+1} \right] \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1j} \\ h_{21} & h_{22} & & h_{2j} \\ 0 & h_{32} & \ddots & \vdots \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \\ 0 & 0 & & h_{jj} \\ 0 & 0 & \cdots\ 0 & h_{j+1,j} \end{bmatrix}$$

$$\underset{m\times m}{A}\ \underset{m\times j}{Q_j} = \underset{m\times(j+1)}{Q_{j+1}}\ \underset{(j+1)\times j}{H_j}$$

$$\begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \end{bmatrix}$$

Breakdown if $H(j+1,j) = 0$ at some step, but this implies

$$\underbrace{A q_j} = q_1 h_{1j} + q_2 h_{2j} + \dots + q_j h_{jj} + q_{j+1} \cdot 0$$

$\downarrow$

$p(A)y$ for p of degree exactly j $\qquad q_1, \dots q_j$ have degree at most $j-1$

This implies that $A^j y \in K_j(A, y) = \mathrm{span}(y, Ay, \dots A^{j-1} y)$.

$\Rightarrow y, Ay, \dots A^{j-1} y, A^j y$ are linearly dependent.

Solving $Ax = y$ for large, sparse $A \in \mathbb{R}^{m \times m}$, $y \in \mathbb{R}^m$, using Krylov subspaces: (even when A is not symm. pos. def.)

Idea: let us generate iterates in $K_j(A, y)$ at step j.

$$x_j = \arg\min_{x \in K_j(A,y)} \|Ax - y\|_2$$

GMRES ("generalized minimum residual")

$$x \in K_j(A,y) \iff x = Q_j \cdot c \qquad Q_j = \left[\, q_1 \,\middle|\, q_2 \dots \middle|\, q_j \right]$$

$q_1 c_1 + q_2 c_2 + \dots + q_j c_j$ spans every vector in $K_j(A,y)$ as $c_1, \dots c_j$ change

$$x_j = \arg\min \|Ax - y\|_2 = \arg\min_{c \in \mathbb{R}^j} \|\underbrace{\boxed{AQ_j}}_{m \times j} c - y\|_2$$

$$= \arg\min \|Q_{j+1} \underline{H}_j c - y\|_2$$

take $Q = \left[\, Q_{j+1} \,\middle|\, Q_c \right]$ orthogonal

$$= \arg\min_c \|Q^T (Q_{j+1} \underline{H}_j c - y)\|_2$$

$$Q^T Q_{j+1} = \begin{bmatrix} Q_{j+1}^T Q_{j+1} \\ Q_c^T Q_{j+1} \end{bmatrix} = \begin{bmatrix} I_{j+1} \\ 0 \end{bmatrix}$$

$$= \arg\min_c \left\| \begin{bmatrix} \underline{H}_j c \\ 0 \end{bmatrix} - \begin{bmatrix} \|y\| \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right\|_2$$

$$Q^T y = \begin{bmatrix} \|y\| \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$= \arg\min_c \|\underbrace{\underline{H}_j}_{(j+1) \times j} c - e_1 \|y\| \|_2 \qquad q_1 = \frac{y}{\|y\|}$$

$\uparrow$

a smaller least squares problem!

Convergence of GMRES:

$x \in K_j(A,y) \leftrightarrow x = p(A)\, y$ $\quad p(t)$ polynomial of degree $< j$

$$x = y\alpha_0 + Ay\alpha_1 + \dots + A^{j-1} y\, \alpha_{j-1}$$
$$= (\alpha_0 I + \alpha_1 A + \dots + \alpha_{j-1} A^{j-1})\, y$$

$$r_j = \min_{x \in K_j(A,y)} \|Ax - y\| = \min_{\deg p < j} \| A p(A) y - y \|$$

$= (Ap(A) - I)\, y$

$= q(A)\, y$

$q(t) = t \cdot p(t) - 1$

$\deg(q) \leq j$

$q(0) = -1$

$$= \min_{\substack{\deg(q) \leq j \\ q(0) = -1}} \|q(A) y\|$$

If A is diagonalizable,
$$A = V \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots\ \lambda_m \end{bmatrix} V^{-1}$$

$$= \min_q \left\| V \begin{bmatrix} q(\lambda_1) & & & \\ & q(\lambda_2) & & \\ & & \ddots & \\ & & & q(\lambda_m) \end{bmatrix} V^{-1} y \right\|$$

$$\leq \min_q \underbrace{\|V\| \cdot \max_i |q(\lambda_i)| \cdot \|V^{-1}\|}_{\kappa(V)} \cdot \|y\|$$

$$\frac{\|r_j\|}{\|y\|} \leq \boxed{\kappa(V)} \cdot \min_{\substack{q \\ q(0) = -1}} \max_i |q(\lambda_i)|$$

could be large!

At most k different eigenvalues $\Rightarrow$ exact convergence in k steps

K different clusters of eigenvalues $\Rightarrow$ great reduction after K steps

Remark:

- we have to solve a l.s. problem with $\underline{H}_j = \begin{bmatrix} \times&\times&\times&\times&\times \\ \times&\times&\times&\times&\times \\ 0&\times&\times&\times&\times \\ 0&0&\times&\times&\times \\ 0&0&0&\times&\times \\ 0&0&0&0&\times \end{bmatrix}$

  Because of the zeros, computing $qr(H)$ only costs $O(n^2)$ $\ll O(mn^2)$

- computing $\|Ax-y\| = \|\underline{H}c - e_1\|y\|\|$ can be done on the smaller problem
- It's incremental!

$$Q = [q_1 \dots q_j \mid q_{j+1}] \qquad H = \left[\begin{array}{ccc|c} \times&\times&\times&\times \\ \times&\times&\times&\times \\ 0&\times&\times&\times \\ 0&0&\times&\times \\ \hline 0&0&0&\times \end{array}\right]$$

We can merge the two loops: compute step 1 of Arnoldi, solve a 2x1 L.S. problem, check residual, increase j and compute step 2 of Arnoldi, ...

- "lucky breakdown": linear dependence in $K_{j+1}(A,y)$

$\Leftrightarrow h_{j+1,j} = 0 \Leftrightarrow$ $\min \left\| \begin{bmatrix} \times&\times&\times&\times \\ \times&\times&\times&\times \\ 0&\times&\times&\times \\ 0&0&\times&\times \\ 0&0&0&0 \end{bmatrix} c - \begin{bmatrix} \times \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right\|$ is a square linear system and can be solved exactly

$\Rightarrow r_j = 0$.

- What happens if A is symmetric?

  H is symmetric, too ($H_{ji} = H_{ij}$), this follows from noting that $H_{ij} = q_i^T A q_j$.

  $\Rightarrow$ one can restrict the for $i=1:j$ loop to for $i=j-1:j$

Cost: n mat-vec $+ O(mn)$ for n steps

(instead of $O(mn^2)$ for a nonsymmetric A)

<u>Nonsymmetric</u>		<u>Symmetric</u>
Arnoldi	$\to$	Lanczos
GMRES	$\to$	MINRES

Algorithms for large/sparse linear systems:

symm. pos. def. $\to$ conjugate gradient
symmetric $\to$ MINRES
nonsymmetric $\to$ GMRES

<u>Also</u>: direct algorithms (Gaussian elimination)