Iterative methods

: start with guess, iteratively improve using only matrix-vector

- solve really big problems, where you can't even store matrix:

ex: solving PDE on $\underline{100 \times 100 \times 100}$ grid

$\sim 10^6$ unknowns

$\Rightarrow 10^6 \times 10^6$ matrices

storage: $10^{12} \times 8$ bytes per float

$= 8$ TB (!)

too much to store on anything but a big supercomputer

time:

— if 1000×1000 solve takes 0.1 sec, then $10^6 \times 10^6$ solve takes

$\sim 0.1 \text{ sec.} \times (10^3)^3 \sim 10^8$ sec

$\sim$ years!

totally impractical (!)

Saving grace: matrices are very special

can often multiply $\underset{m \times m}{A}\, x$ in $\Theta(m)$ or $\Theta(m \log m)$ ops, with $\Theta(m)$ storage!

Why: big matrices in practice have special structure

example: • A sparse: only $O(1)$ nonzeros per row (indep. of m) (mostly zero)

6.339

- only store nonzero entries & multiply by nonzero

• A can be expressed in terms of FFTs, fast multipole methods etc. (18.336 and 6.339)

Challenge: solve $Ax = b$ or $Ax = \lambda x$ only using $A \cdot$ vector operations (as few as possible! $\ll m$)

Many methods — which is best is very problem-dependent

... need to know how they work to intelligently choose

Simplest iterative method for $Ax = \lambda x$:

Power method: $x = \text{guess}$

for $i = 1, 2, \ldots$

$$x \leftarrow \frac{Ax}{\|Ax\|_2}$$

$\longrightarrow$ eigenvector for largest $|\lambda|$,

uses A only via Ax

Problem: might converge slowly

if $\underbrace{|\lambda_1| \sim |\lambda_2|}_{\text{nearby magnitudes}} > |\lambda_3| \ldots$

gives us only 1 eigenvector at a time

Ex: suppose $|\lambda_1| = |\lambda_2| > |\lambda_3| > |\lambda_4| > \ldots$

initial $x = \sum_{k=1}^{m} c_k v_k$ where v_k are eigenvectors

$n \gg 1$ iterations $\approx \lambda_1^n c_1 v_1 + \lambda_2^n c_2 v_2$ $\neq$ eigenvector

$\|$ $\|_2$

at n+1 iteration: $\dfrac{\lambda_1^{n+1} c_1 v_1 + \lambda_2^{n+1} c_2 v_2}{\| \quad '' \quad \|_2}$

say $\lambda_2 = -\lambda_1$: n^{th} iteration $\dfrac{c_1 v_1 + c_2 v_2}{\| \quad \|}$

$\lambda_2^n = \lambda_1^n (-1)^n$

$(n+1)$ iteration $\dfrac{c_1 v_1 - c_2 v_2}{\| \quad \|}$

indep

span of n, $n+1$ iterations
gives us both V_1, V_2

looking at just 2 iterations
should give us both eigenvectors

- - -

Real problem: "no memory":
throw out everything but
last iteration

Arnoldi algorithm : like power method, but remembers all iterations

("Krylov method")

suppose initial guess $= b$

$$\text{iterations} = \underbrace{b, Ab, A^2b, \ldots, A^{n-1}b}_{\substack{\text{find best solution on step } n\text{, looking at all of these} \\ \text{(all linear combinations)}}}$$

$$\mathcal{K}_n = \text{span of } \underbrace{\{b, Ab, \ldots, A^{n-1}b\}}_{\substack{\text{very ill-conditioned} \\ \text{(nearly} \parallel \text{ to eigenvector} \\ \text{of biggest } |\lambda|)}}$$

$= \text{Krylov space}$

① need better basis for $\mathcal{K}_n$
$=$ orthonormal
$\Rightarrow$ Arnoldi gives us

② given basis $Q_n = (q_1 \cdots q_n)$, find approximate eigenvectors / eigenvalues
$\Rightarrow$ Rayleigh-Ritz procedure

① Arnoldi : modified Gram-Schmidt on power iterations as we go along :

$q_1 = b/\|b\|_2$

for $n = 1, 2, \ldots$

$v = A q_n$

$\in \mathcal{K}_{n+1}$ $\quad$ $\in \mathcal{K}_n = A^{n-1} b + \ldots$

$A \cdot$

for $j = 1$ to n

$h_{jn} = q_j^* v$

$v = v - h_{jn} q_j$

MGS

$h_{n+1,n} = \|v\|_2$

$q_{n+1} = v / h_{n+1,n}$

basis $\Rightarrow Q_n$, coefficients H_n

$\Rightarrow$ basis vectors $q_1 \ldots q_n$ span $\mathcal{K}_n$

interpret h_{jk} as matrix elements

work backwards :

$v = A q_n = h_{n+1,n} q_{n+1} + h_{nn} q_n + h_{n-1,n} q_{n-1} + \ldots + h_{1n} q_1$

in matrix form: upper-Hessenberg!

$$\underset{m\times m}{A}\ \underset{m\times n}{Q_n} = Q_n \underbrace{\begin{pmatrix} h_{11} & h_{12} & \cdots \\ h_{21} & h_{22} & \ddots \\ & h_{32} & \ddots \end{pmatrix}}_{n\times n} + q_{n+1} h_{n+1,n} \underbrace{\overbrace{(0\,0\,0\,0 \cdots 0\,1)}^{n \text{ cols}}}_{e_n^*}$$

$$Q_n = (q_1 \cdots q_n) \qquad H_n$$

$$= (Aq_1\ Aq_2 \cdots Aq_n)$$

$$\Rightarrow \boxed{AQ_n = Q_n H_n + q_{n+1} h_{n+1,n} e_n^*}$$

$$\Rightarrow Q_n^* A Q_n = H_n + \cancel{Q_n^* q_{n+1}} h_{n+1,n} e_n^*$$

$$Q_n^* q_{n+1} = 0$$

$$Aq_1 = h_{21} q_2 + h_{11} q_1$$

$$Aq_2 = h_{32} q_3 + h_{22} q_2 + h_{12} q_1$$

$$\vdots$$

Arnoldi = column-by-column upper Hessenberg factorization of A

= "nice" orthonormal basis of $\mathcal{K}_n$ (our "memory")

② Rayleigh-Ritz method:

approximate eigenvectors from Q_n

- find approximate eigenvectors $\in \mathcal{K}_n$

$$= \underset{m\times n}{Q_n} \underset{n}{z} \quad \text{for some } z \in \mathbb{C}^n$$

(linear combination of $q_1 \cdots q_n$)

where Ritz vectors, approximate eigenvalue = Ritz values

$$\text{residual} = A(Q_n z) - \nu(Q_n z)$$

$$= \text{"small"}$$

$$\perp \mathcal{K}_n$$

$$\Rightarrow Q_n^*(AQ_n z - \nu Q_n z) = 0$$

$$= \underbrace{\underset{n\times m}{Q_n^*} \underset{m\times m}{A} \underset{m\times n}{Q_n}}_{n\times n} z - \nu z \qquad (Q_n^* Q_n = I)$$

$$= H_n z - \nu z$$

$$\Rightarrow \boxed{H_n z = \nu z}$$

$\Rightarrow$ n solutions $z_1 \cdots z_n$, $\nu_1 \cdots \nu_n$

(easy!) $n\times n$ eigenproblem with eigenvals ν, eigenvectors z!

Does at <u>least</u> as well as power method
(= last column of Q_n)

Optimality of Ritz vectors for $A = A^*$

various ways to see that these are "optimal" eigenvector estimates in $\mathcal{K}_n$

Recall min-max theorem:

$$\lambda_{max} = \max_{\substack{x \neq 0 \\ \in \mathbb{C}^m}} \frac{x^* A x}{x^* x} \quad \text{for } A^* = A$$

Claim: $\nu_{max} = \max_{\substack{x \neq 0 \\ \in \mathcal{K}_n}} \frac{x^* A x}{x^* x} = \max_{\substack{z \neq 0 \\ \in \mathbb{C}^n}} \frac{(Q_n z)^* A Q_n z}{(Q_n z)^* Q_n z}$

$= Q_n z$ for some $z \neq 0$

Q.E.D.

$$= \frac{z^* Q_n^* A Q_n z}{z^* \underbrace{\cancel{Q_n^* Q_n}}_{I} z} = \frac{z^* H_n z}{z^* z}$$

= max eigenvalue of H_n by min-max thm.

= Rayleigh quotient for $H_n = H_n^*$

similarly for ν_{min} etc.

• Alternatively, one can show for $A = A^*$

that Ritz vectors $\underset{m\times n}{Q_n}\,\underset{n\times n}{Z} = \underset{m\times n}{V}$, $\quad Z^* = Z^{-1}$ since $H_n = H_n^*$ (orthonormal eigenvectors)

$Z = \underset{n\times n}{(z_1\, z_2 \cdots z_n)}$

and Ritz values $D = \begin{pmatrix} \nu_1 & & & \\ & \nu_2 & & \\ & & \ddots & \\ & & & \nu_n \end{pmatrix}$ $\qquad \Rightarrow V^*V = I$

$$\min_{V,D} \|AV - VD\|_2$$

over all possible orthonormal bases V of $\mathcal{K}_n$ and all possible diagonal D

to be continued