GMRES for $Ax = b$

Idea (very simple!): minimise residual in Krylov subspace: [Saad-Schulz 86]

$$x = \text{argmin}_{x \in \mathcal{K}_k(A,b)} \|Ax - b\|_2$$

$$= \text{span}[b, Ab, A^2b, \ldots, A^{k-1}b]$$

$$\Longleftrightarrow \min_y \|AQ_k y - b\|_2 \qquad x = Qy$$

$$AQ_{k+1} = Q_{k+1} H_k$$

$$= \min_y \left\| \begin{bmatrix} Q_k^T \\ Q_{k,\perp}^T \end{bmatrix} A Q_k y - \begin{bmatrix} Q_k^T \\ Q_{k,\perp}^T \end{bmatrix} b \right\|_2$$

$$= \min_y \left\| \begin{bmatrix} \boxed{H_k} \\ 0 \end{bmatrix} y - \begin{bmatrix} b_1 \\ \tilde{b}_2 \end{bmatrix} \right\|_2$$

Hess, O(n) givens rot. for QR.

GMRES for $Ax = b$

Idea (very simple!): minimise residual in Krylov subspace: [Saad-Schulz 86]

$$x = \text{argmin}_{x\in\mathcal{K}_k(A,b)}\|Ax-b\|_2$$

Algorithm: Given $AQ_k = Q_{k+1}\tilde{H}_k$ and writing $x = Q_k y$, rewrite as

$$\begin{aligned}\min_y \|AQ_k y - b\|_2 &= \min_y \|Q_{k+1}\tilde{H}_k y - b\|_2 \\ &= \min_y \left\| \begin{bmatrix}\tilde{H}_k \\ 0\end{bmatrix} y - \begin{bmatrix}\tilde{Q}_k^T \\ \tilde{Q}_{k,\perp}^T\end{bmatrix} b \right\|_2 \\ &= \min_y \left\| \begin{bmatrix}\tilde{H}_k \\ 0\end{bmatrix} y - \|b\|_2 e_1 \right\|_2, \quad e_1 = [1,0,\ldots,0]^T \in \mathbb{R}^n\end{aligned}$$

$$\begin{bmatrix}1\\0\\0\\\vdots\\0\end{bmatrix}\|b\|$$

(where $[\tilde{Q}_k, \tilde{Q}_{k,\perp}]$ orthogonal; same trick as in least-squares)

- Minimised when $\|\tilde{H}_k y - \tilde{Q}_k^T b\| \to \min$; Hessenberg least-squares problem
- Solve via QR (k Givens rotations)+triangular solve, $O(k^2)$ in addition to Arnoldi

GMRES convergence: polynomial approximation

Recall that $x \in \mathcal{K}_k(A,b) \Rightarrow x = p_{k-1}(A)b$. Hence GMRES solution is

$$\tilde p(z) = z \cdot p_{k-1}(z), \quad \tilde p(0) = 0$$

$$p(z) = \tilde p(z) - 1$$

$$\min_{x\in\mathcal{K}_k(A,b)} \|Ax-b\|_2 = \min_{p_{k-1}\in\mathcal{P}_{k-1}} \|Ap_{k-1}(A)b-b\|_2$$

$$= \min_{\tilde p\in\mathcal{P}_k, \tilde p(0)=0} \|(\tilde p(A)-I)b\|_2$$

$$= \min_{p\in\mathcal{P}_k, p(0)=1} \|p(A)b\|_2$$

"$p(0) = -1$" $\Longleftrightarrow$ "$p(0) = 1$"

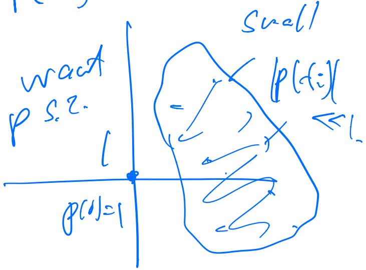
genericity assump.

If A diagonalizable $A = X\Lambda X^{-1}$,

$$p(A) = \sum_{i=0}^{k} c_i A^i = \sum_{i=0}^{k} (X\Lambda X^{-1})^i c_i = \sum_{i=0}^{k} (X\Lambda^i X^{-1}) c_i = X\Big(\sum_{i=0}^{k} \Lambda^i c_i\Big)X^{-1} = X p(\Lambda) X^{-1}$$

$$\|p(A)\|_2 = \|Xp(\Lambda)X^{-1}\|_2 \le \|X\|_2\|X^{-1}\|_2\|p(\Lambda)\|_2$$

$$= \kappa_2(X) \max_{z\in\lambda(A)} |p(z)|$$



Interpretation: find polynomial s.t. $p(0) = 1$ and $|p(\lambda_i)|$ small for all i

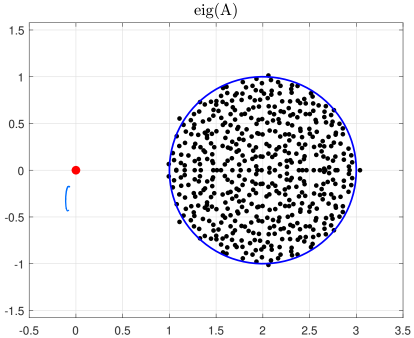
GMRES example

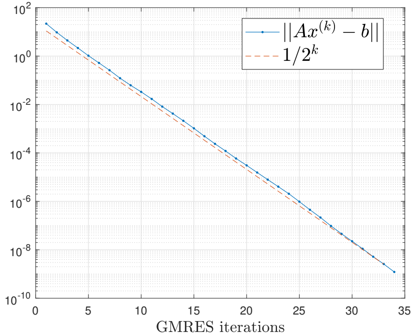
$\in \mathbb{R}^{500\times 500}$

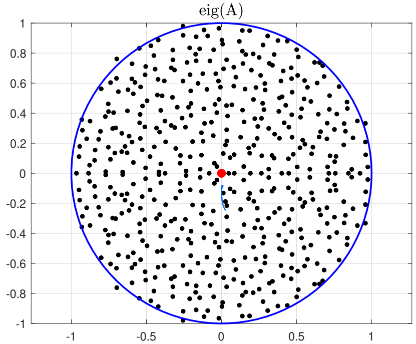
G: Gaussian random matrix ($G_{ij} \sim N(0,1)$, i.i.d.) $G/\sqrt{n}$: eigvals in unit disk

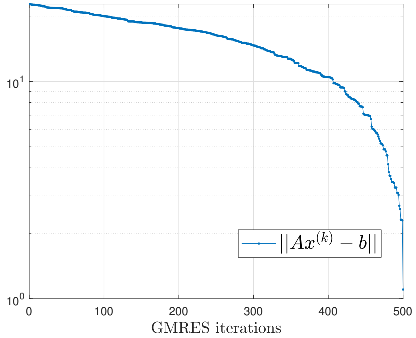
$$A = 2I + G/\sqrt{n},$$
$$p(z) = 2^{-k}(z-2)^k$$

$$A = G/\sqrt{n}$$









Restarted GMRES

For k iterations, GMRES costs k matrix multiplications$+O(nk^2)$ for orthogonalization
$\rightarrow$ Arnoldi eventually becomes expensive.

Practical solution: restart by solving 'iterative refinement':

1. Stop GMRES after $k_{\max}$ (prescribed) steps to get approx. solution $\hat{x}_1$
2. Solve $A\tilde{x} = b - A\hat{x}_1$ via GMRES
3. Obtain solution $\hat{x}_1 + \tilde{x}$

Sometimes multiple restarts needed

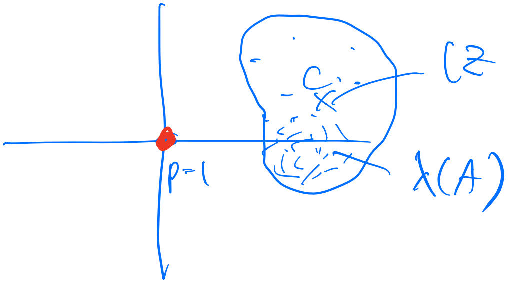
When does GMRES converge fast?

Recall GMRES solution satisfies (assuming A diagonalisable+nonsingular)

$$\min_{x\in\mathcal{K}_k(A,b)} \|Ax-b\|_2 = \min_{p\in\mathcal{P}_k,p(0)=1} \|p(A)b\|_2 \le \kappa_2(X) \max_{z\in\lambda(A)} |p(z)|\|b\|_2.$$

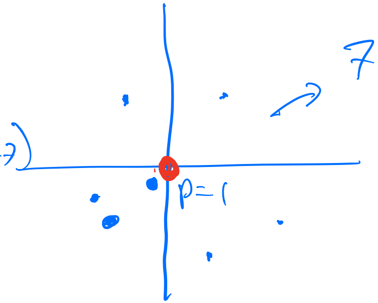
$\max_{z\in\lambda(A)} |p(z)|$ is small when

- $\lambda(A)$ are clustered away from 0
  - a good p can be found quite easily
  - e.g. example 2 slides ago



$(z-c)^k \cdot \tilde{c}$

- When $\lambda(A)$ takes $k(\ll n)$ distinct values
  - Then convergence in k GMRES iterations (why?)



7 iter of GMRES gets you x.

$p(z) = \tilde{c}(z-\lambda_1)(z-\lambda_2)\cdots(z-\lambda_7)$

$\tilde{c}$ s.t. $p(0)=1$.

Preconditioning for GMRES

We've seen that GMRES is great if spectrum clustered away from 0. If not true with

$$Ax = b,$$

then precondition: find $\underbrace{M}_{\text{nonsingular}} \in \mathbb{R}^{n\times n}$ and solve

$$MAx = Mb$$

$[b, Ab, A^2b, \ldots]$

(Aq_k)

Desiderata of M:

- M simple enough s.t. applying M to vector is easy (note that each GMRES iteration requires MA-multiplication), and one of
  1. MA has clustered eigenvalues away from 0
  2. MA has a small number of distinct eigenvalues
  3. MA is well-conditioned $\kappa_2(MA) = O(1)$; then solve normal equation $(MA)^T MAx = (MA)^T Mb$

$M = A^{-1}$ "ideal" preconditioner wrt 1~3

Preconditioners: examples

- ILU (Incomplete LU) preconditioner: $A \approx LU, M = (LU)^{-1} = U^{-1}L^{-1}$, L, U 'as sparse as A' $\Rightarrow MA \approx I$ (hopefully; 'cluster away from 0')

saddle-point matrix

- For $\tilde{A} = \begin{bmatrix} A & B \\ C & \boxed{0} \end{bmatrix}$, set $M = \begin{bmatrix} A^{-1} & \\ & (CA^{-1}B)^{-1} \end{bmatrix}$. Then if M nonsingular, $M\tilde{A}$ has eigvals$\in \{1, \frac{1}{2}(1 \pm \sqrt{5})\} \Rightarrow$ 3-step convergence [Murphy-Golub-Wathen 2000]
- Multigrid-based, operator preconditioning, ...

Finding effective preconditioners is never-ending research topic

Prof. Andy Wathen is our Oxford expert!

Arnoldi for nonsymmetric eigenvalue problems

Arnoldi for eigenvalue problems: Arnoldi iteration+Rayleigh-Ritz (just like Lanczos alg)

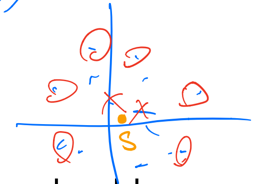
Lanczos sym. eig = Lanczos iter + R-R

1. Compute Q^TAQ
2. Eigenvalue decomposition $Q^TAQ = X\hat{\Lambda}X^{-1}$ (via QR alg + Sylvester eqns) $O(k^3)$ not back. stable (nonexaminable)
3. Approximate eigenvalues diag$(\hat{\Lambda})$ (Ritz values) and eigenvectors QX (Ritz vectors)

As in Lanczos, $Q = Q_k = \mathcal{K}_k(A, b)$, so simply $Q_k^TAQ_k = H_k$ (Hessenberg eigenproblem, ideal for QRalg) $\rightarrow O(k^2)$ flops/QR step

span(b, Ab) = span$(b, (A-sI)b)$ = span(b, Ab)

Which eigenvalues are found by Arnoldi?



- Krylov subspace is invariant under shift: $\mathcal{K}_k(A, b) = \mathcal{K}_k(A - sI, b)$
- Thus any eigenvector that power method applied to $A - sI$ converges to should be contained in $\mathcal{K}_k(A, b)$ $\mathcal{K}_k(A,b)$ #
- To find other (e.g. interior) eigvals, shift-invert Arnoldi: $Q = \mathcal{K}_k((A - sI)^{-1}, b)$