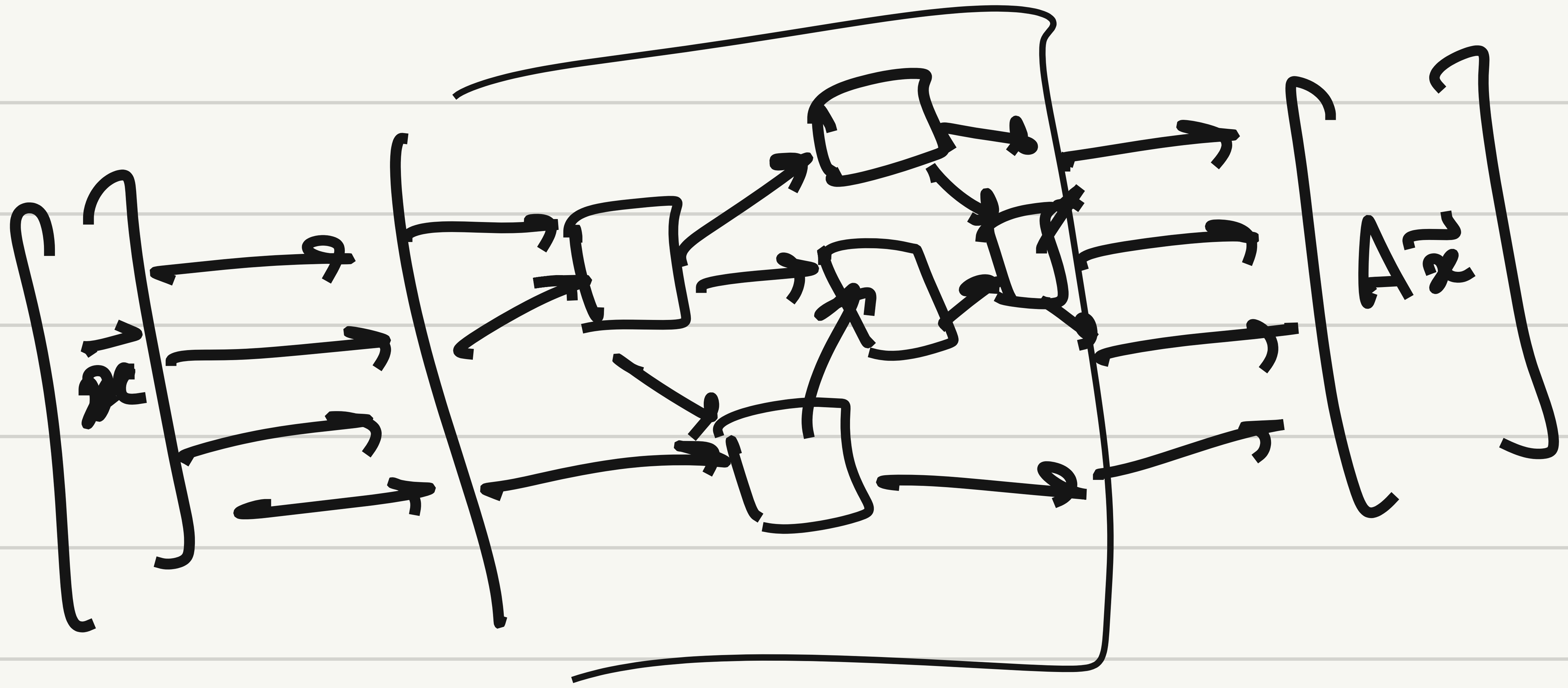
COL726: $\boxed{\text{Iterative Methods}}$

Minor: Monday, 14 Feb

Assg. 2: Tuesday, 15 Feb

Assg. 3: out



Solve $A\vec{x} = \vec{b}$, $A \in \mathbb{C}^{m \times m}$, m very large

but computing $\vec{v} \mapsto A\vec{v}$ is cheap

eg. sparse matrices

low-rank matrices $A = \underbrace{T}_{m \times r}\underbrace{W}_{r \times m}$

$r \ll m$

iterates:

$$\underbrace{\vec{x}_0 \to \vec{x}_1}_{\text{fast}} \to \vec{x}_2 \to \cdots \to \vec{x}_*$$

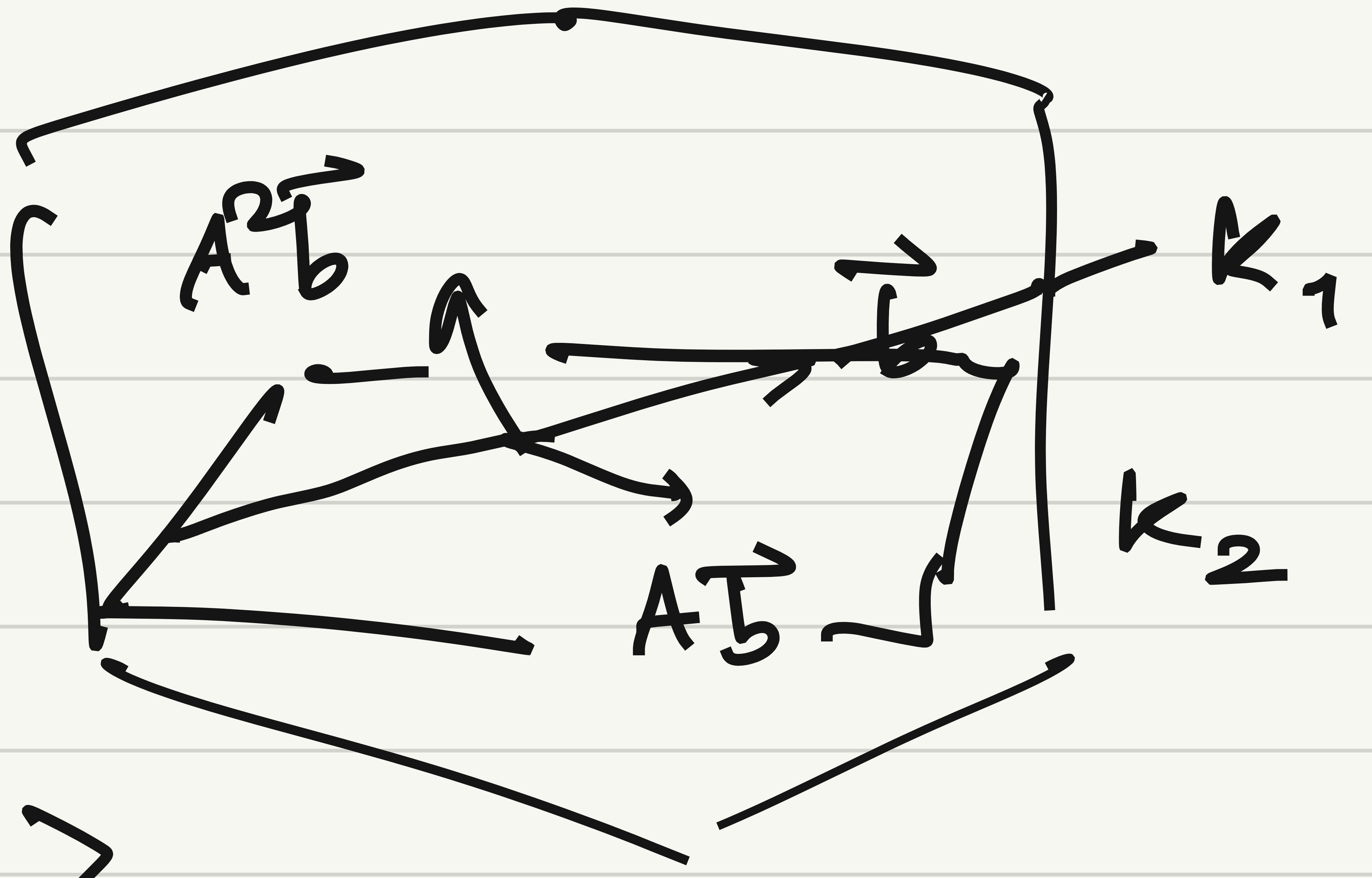
s.t. $A\vec{x}_* = \vec{b}$

Convergence analysis

$A\vec{x} = \vec{b}$

$\vec{b}, \ A\vec{b}, \ A^2\vec{b}, \ A^3\vec{b}, \dots$

$K_1 = \langle \vec{b} \rangle, \quad K_2 = \langle \vec{b}, A\vec{b} \rangle, \ \dots, \ K_n = \langle \vec{b}, A\vec{b}, \dots A^{n-1}\vec{b} \rangle$



$K_1 \subseteq K_2 \subseteq K_3 \subseteq \dots$

$\boxed{\text{Krylov subspaces}}$

→ Choose $\underbrace{\vec{x}_n \in K_n}$ which minimizes $\|\vec{b} - A\vec{x}\|_2$

$\vec{x}_n = \underbrace{c_0}\vec{b} + \underbrace{c_1}A\vec{b} + \underbrace{c_2}A^2\vec{b} + \cdots + \underbrace{c_{n-1}}A^{n-1}\vec{b}$

claim: $\dfrac{A^n\vec{b}}{\|A^n\vec{b}\|}$ becomes v. similar to $\dfrac{A^{n+1}\vec{b}}{\|A^{n+1}\vec{b}\|}$

Suppose A is $\boxed{\text{diagonalizable}}$: it has complete set of eigenvectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m$ which are a basis of $\mathbb{C}^m$ and eigenvalues $\lambda_1, \ldots, \lambda_m$

$$A\vec{v}_j = \lambda_j \vec{v}_j$$

$$AV = \left[\lambda_1\vec{v}_1 \,\middle|\, \cdots \,\middle|\, \lambda_m \vec{v}_m\right] = V\Lambda \iff A = V\Lambda V^{-1}$$

$$V = \left[\vec{v}_1 \,\middle|\, \vec{v}_2 \,\middle|\, \cdots \,\middle|\, \vec{v}_m\right] \qquad \Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{bmatrix}$$

$$\vec{b} = s_1\vec{v}_1 + \cdots + s_m\vec{v}_m \Rightarrow A\vec{b} = s_1\lambda_1\vec{v}_1 + \cdots + s_m\lambda_m\vec{v}_m$$

$$\Rightarrow A^k = s_1\lambda_1^k \underbrace{\vec{v}_1} + \cdots + s_m\lambda_m^k \vec{v}_m$$

$$K_n = \langle \vec{b}, A\vec{b}, A^2\vec{b}, \dots, A^{n-2}\vec{b}, A^{n-1}\vec{b} \rangle$$
$$= \langle \vec{q}_1, \vec{q}_2, \vec{q}_3, \dots, \vec{q}_{n-1}, \vec{q}_n \rangle$$

Arnoldi iteration:

$\vec{q}_1 = \vec{b} / \|\vec{b}\|$

for $n = 1, 2, 3, \dots$

$\vec{v}_{n+1} = A\vec{q}_n$ → $O(\text{mat.vec})$

Project $\vec{v}_{n+1}$ onto $\langle \vec{q}_1, \dots, \vec{q}_n \rangle^\perp$ → $O(mn)$

$\vec{q}_{n+1} = \vec{v}_{n+1}^\perp / \|\vec{v}_{n+1}^\perp\|$ → $O(m)$

$$\vec{v}_{n+1} = A\vec{q}_n$$
$$= h_{1n}\vec{q}_1 + h_{2n}\vec{q}_2 + \dots + h_{nn}\vec{q}_n + \underbrace{\vec{v}_{n+1}^\perp}_{h_{n+1,n}\vec{q}_{n+1}}$$

where $h_{n+1,n} = \|\vec{v}_{n+1}^\perp\|$

$$A\vec{q}_n = \left[\vec{q}_1 \middle| \cdots \middle| \vec{q}_{n+1} \right] \begin{bmatrix} h_{1n} \\ h_{2n} \\ \vdots \\ h_{n+1,n} \end{bmatrix}$$

$$Q_n = \begin{bmatrix} \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_n \end{bmatrix}$$

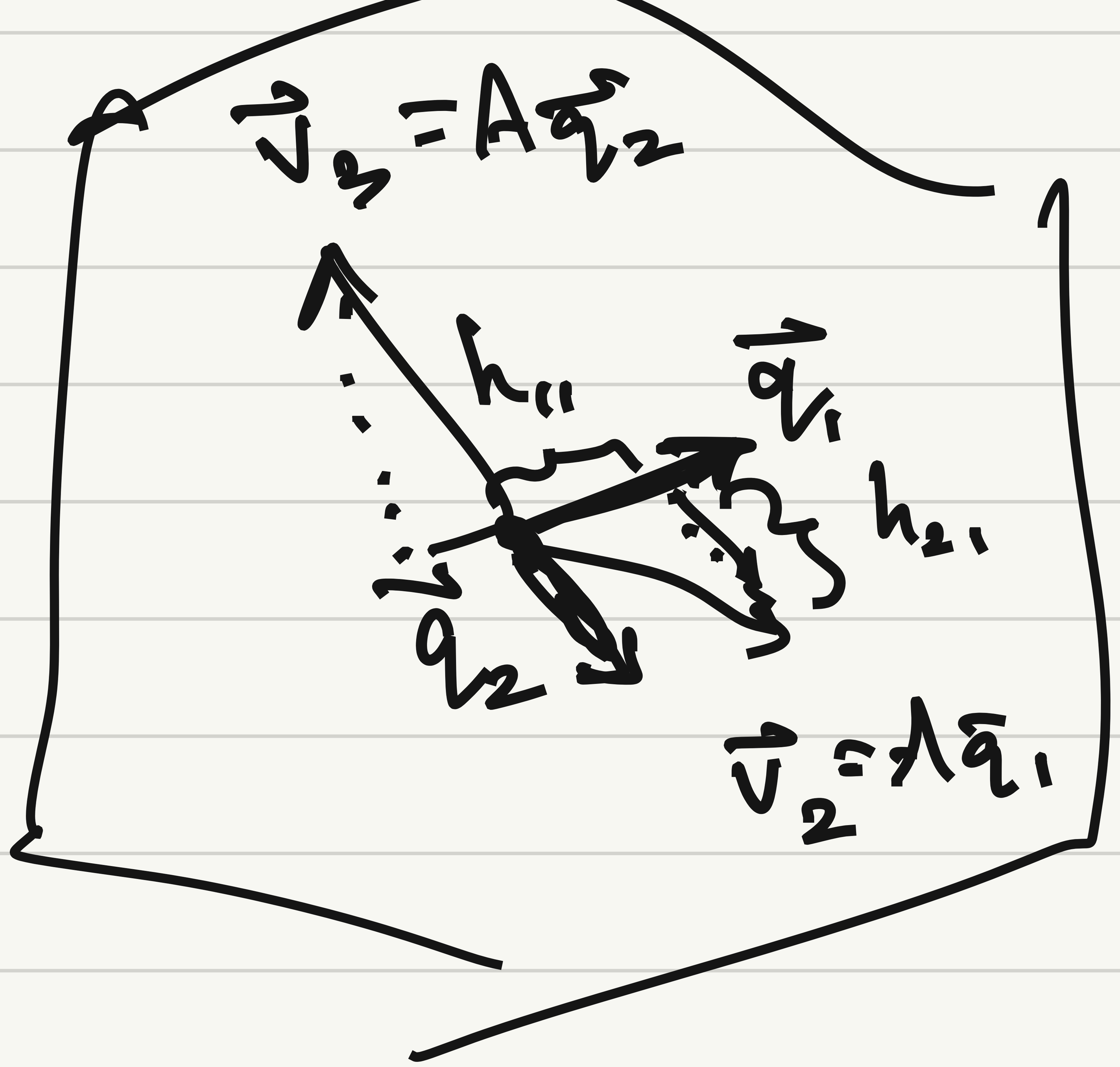
$$AQ_n = \begin{bmatrix} A\vec{q}_1 & \cdots & A\vec{q}_n \end{bmatrix} = Q_{n+1} \cdot H_n$$

$$\mathbb{C}^{(n+1)\times n}$$

$$H_n = \begin{bmatrix} h_{11} & \cdots & h_{1n} \\ h_{21} & \ddots & h_{2n} \\ & \ddots & \vdots \\ & & h_{nn} \\ & & h_{n+1,n} \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} \vec{q}_1 & \cdots & \vec{q}_n & \vec{q}_{n+1} \end{bmatrix}}_{Q_{n+1}} \begin{bmatrix} h_{1n} \\ h_{2n} \\ \vdots \\ h_{nn} \\ h_{n+1,n} \end{bmatrix}$$

$$A\vec{q}_n = \vec{v}_{n+1} = \underbrace{h_{1n} q_1 + \cdots + h_{nn} q_n}_{P_{\langle q_1 \cdots q_n \rangle} \vec{v}_{n+1}} + \underbrace{h_{n+1,n} q_{n+1}}_{\vec{v}_{n+1}^{\perp}}$$



$$A \begin{bmatrix} \vec{q}_n \end{bmatrix} = \begin{bmatrix} Q_n & \vec{q}_{n+1} \end{bmatrix} \cdot \begin{bmatrix} h_{1n} \\ \vdots \\ h_{nn} \\ \hline h_{n+1,n} \end{bmatrix}$$

$$A \begin{bmatrix} \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_n \end{bmatrix} = \begin{bmatrix} \vec{q}_1 & \vec{q}_2 & \cdots & \vec{q}_n & \vec{q}_{n+1} \end{bmatrix} \cdot \overbrace{\begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ & h_{32} & \ddots & \vdots \\ & & \ddots & h_{nn} \\ & & & h_{n+1,n} \end{bmatrix}}^{H_n}$$

for each n, I get $\vec{q}_1, \vec{q}_2, \dots, \vec{q}_n$

$$A Q_n = Q_{n+1} H_n$$

$H_n \in \mathbb{C}^{(n+1)\times n}$ $\boxed{\text{upper Hessenberg}}$ matrix

$Q_n \in \mathbb{C}^{m\times n}$

Using Q_n, let's do least-squares on $A\vec{x} = \vec{b}$: $\boxed{\text{GMRES}}$

(generalized min. residual)

$\min \|\vec{b} - A\vec{x}_n\|$ over all $\vec{x}_n \in K_n$

$= \text{range}(Q_n)$

$\Leftrightarrow$ over all $\vec{x}_n = Q_n \vec{y}_n$, $\vec{y}_n \in \mathbb{C}^n$

$$\|\vec{b} - A\vec{x}_n\| = \|\vec{b} - AQ_n\vec{y}_n\| = \|\vec{b} - Q_{n+1}H_n\vec{y}_n\|$$

$\|\vec{b} - A\vec{x}_n\| = \|\vec{b} - AQ_n\vec{y}_n\| = \|\vec{b} - Q_{n+1}H_n\vec{y}_n\|$

$\hookrightarrow \stackrel{?}{=} \|Q_{n+1}^*\vec{b} - H_n\vec{y}_n\|$

Lemma: If Q is tall, has orthonormal cols, then

$\|Q\vec{x}\|_2 = \|\vec{x}\|_2$ for all $\vec{x}$, but $\|Q^*\vec{x}\|_2 = \|\vec{x}\|_2$ iff $\vec{x} \in \text{range}(Q)$

Is $\vec{b} \in \text{range}(Q_{n+1})$? Yes!

What is $Q_{n+1}^*\vec{b}$? $\begin{bmatrix} \vec{q}_1^*\vec{b} \\ \vec{q}_2^*\vec{b} \\ \vdots \end{bmatrix} = \begin{bmatrix} \|\vec{b}\| \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \|\vec{b}\|\vec{e}_1$

$\min \|\vec{b} - A\vec{x}_n\| \iff \min \| \underbrace{\|\vec{b}\|\vec{e}_1} - \underbrace{H_n}\vec{y}_n \|$: $(n+1)\times n$ least squares problem

GMRES alg.

for $n = 1, 2, \ldots$

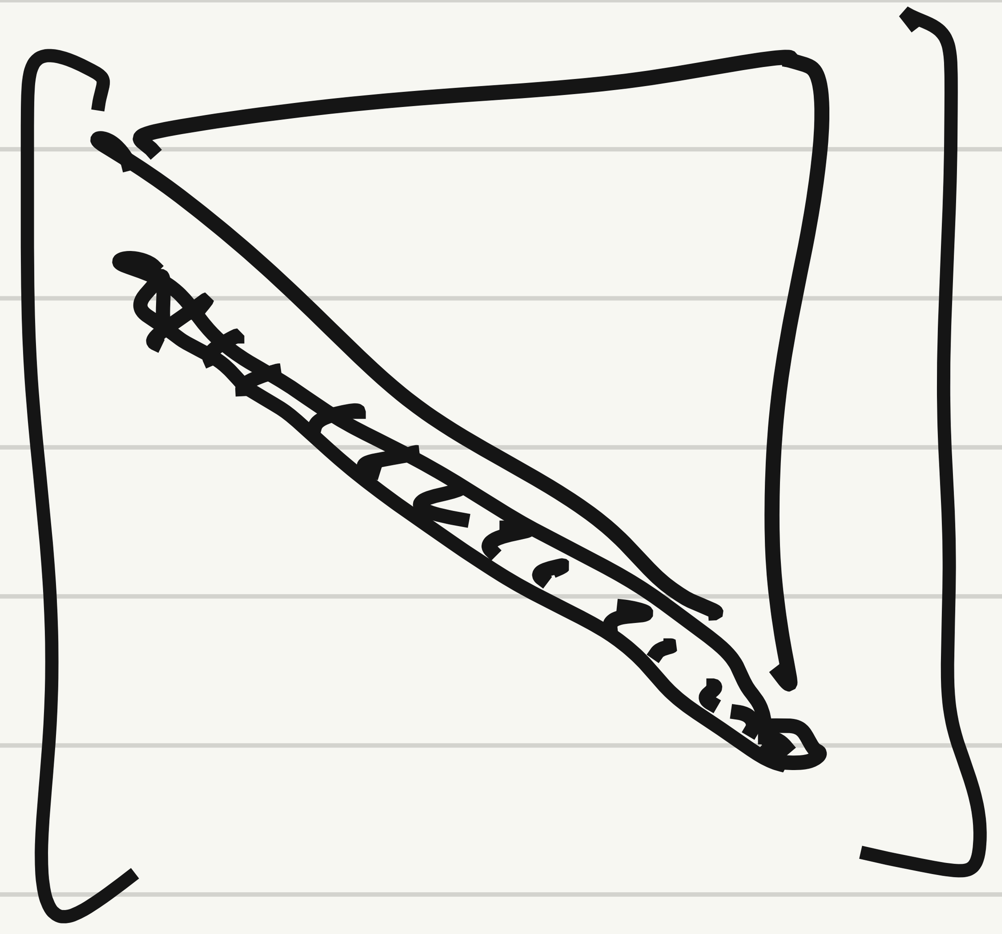
Get $\vec{q}_n$ using Arnoldi $\longrightarrow$ $O(mn$ + mat-vec time$)$

$\vec{y}_n = \text{lsq}(H_n, \|\vec{b}\| \vec{e}_1) \longrightarrow$
- naive QR : $O(n^3)$
- using Hessenberg structure : $O(n^2)$
- by reusing H_{n-1} : $O(n)$

$\vec{x}_n = Q_n \vec{y}_n$

Cost of iterations grows with n.

$O(mn)$ time, $O(mn)$ space for storing Q_n.

$H_n =$  $(n+1) \times n$

Convergence of GMRES

At iter n, $\min \| \underbrace{\vec{b} - A\vec{x}_n}_{\vec{r}_n \text{ (residual)}} \|_2$. Question: How quickly does $\|r_n\| \to 0$ as n increases?

1. Residual decreases monotonically: $\|\vec{r}_{n+1}\| \leq \|\vec{r}_n\|$

   because $K_n \subseteq K_{n+1}$

2. $\|\vec{r}_n\|$ becomes 0 at some $n \leq m$ because $K_m = \mathbb{C}^m$

Suppose $\vec{x} \in K_n$. Then $\vec{x} = c_0\vec{b} + c_1 A\vec{b} + c_2 A^2\vec{b} + \cdots c_{n-1}A^{n-1}\vec{b}$

$q : \mathbb{C} \to \mathbb{C}$

$$= \underbrace{(c_0 I + c_1 A + c_2 A^2 + \cdots + c_{n-1}A^{n-1})}\vec{b}$$

$q(z) = c_0 + c_1 z + \cdots + c_{n-1} z^{n-1}$

$$= q(A)\vec{b}$$

$\vec{x} = q(A)\vec{b}$

One-to-one corresp. between $\vec{x} \in K_n$ and degree-$(n-1)$ polynomials $q(z)$

At nth iter, $\min \|\vec{r}_n\| = \|\vec{b} - A\vec{x}_n\| = \|\underbrace{\vec{b} - A\,q(A)\vec{b}}\|$

$$= \|\underbrace{(I - A\,q(A))}_{p(A)}\vec{b}\|$$

$\underline{p(z) = 1 - z\,q(z)}$

Choosing $\vec{x} \longleftrightarrow$ choosing polynomial $p(A)$ to minimize $\|p(A)\vec{b}\|$

$\hookrightarrow$ degree n polynomial with const. term $= 1$

$p \in P_n = \{ \ldots \}$

$$\min_{\vec{x}_n \in K_n} \|\vec{r}_n\| = \min_{p \in P_n} \|p(A)\vec{b}\|$$

$$\leq \|\vec{b}\| \min_{p \in P_n} \|p(A)\|$$

GMRES has low residual at iter n for all $\vec{b}$ if

there exists $p \in P_n = \{$ deg. n poly. with $p(0) = 1\}$

s.t. $\| p(A) \|$ is small

Assume $\underline{A = V \Lambda V^{-1}}$

$$V = \begin{bmatrix} \vec{v}_1 \Big| \cdots \Big| \vec{v}_m \end{bmatrix} \text{ eigenvectors}$$

$$\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_m \end{bmatrix} \text{ eigen values}$$

$$A^k = (V \Lambda V^{-1})^k = V \Lambda^k V^{-1}$$

$$\Lambda^k = \begin{bmatrix} \lambda_1^k & & \\ & \ddots & \\ & & \lambda_m^k \end{bmatrix}$$

$$p(A) = V p(\Lambda) V^{-1}$$

$$p(\Lambda) = \begin{bmatrix} p(\lambda_1) & & \\ & \ddots & \\ & & p(\lambda_m) \end{bmatrix}$$

When is $\|p(A)\|$ small?

$$\|p(A)\| = \|V\,p(\Lambda)\,V^{-1}\| \le \|V\|\,\underbrace{\|p(\Lambda)\|}_{\max_j |p(\lambda_j)|}\,\|V^{-1}\| \quad \to \kappa(V)$$

$$\underbrace{\max_j |p(\lambda_j)|}_{\|p\|_{\Lambda(A)}}$$

$$\le \kappa(V)\cdot\|p\|_{\Lambda(A)}$$

where $\Lambda(A) = \{\lambda_j : \lambda_j \text{ is an eigenvalue of } A\}$

$\boxed{\text{spectrum}}$

if $p : \mathbb{C} \to \mathbb{C}$, $S \subseteq \mathbb{C}$

$$\|p\|_S = \sup_{z \in S} |p(z)|$$

Thm: At nth step of GMRES, residual satisfies

$$\frac{\|\vec{r}_n\|}{\|\vec{b}\|} \le \inf_{p_n \in P_n} \|p_n(A)\| \le \inf_{p_n \in P_n} \kappa(V) \, \|p_n\|_{\Lambda(A)}$$

Small if

① eigenvectors are close to orthogonal

and ② there exists $p_n \in P_n$ which is small on the spectrum of A.

↳ degree n poly. with $p(0) = 1$