

# Numerical Linear Algebra

## L21 - Eigenvalues & Eigenvectors - Gilbert Strang - MIT

Eigenvector: Vector which maintains direction even after being acted upon by the matrix.

- If  $\lambda \neq 0$ ,  $\underline{A}$  is non-singular. The corresponding eigenvector will be part of  $\underline{A}$ 's null space.
- Consider a projection matrix. Only vectors in the column space of  $\underline{P}$ . This is because  $\underline{P}$  will bring a vector into its plane (and therefore changing its direction, if it does not exist in the column space of  $\underline{P}$  already). Hence,  $\underline{P}$  has only two eigenvalues: 0 or 1.
- Why is  $\det(\underline{A} - \lambda \underline{I}) = 0$ ?

$$\underline{A}\underline{x} = \lambda \underline{x}$$

$$(\underline{A} - \lambda \underline{I})\underline{x} = 0$$

Because we know  $\underline{x} \neq 0$ , we know that  $(\underline{A} - \lambda \underline{I})$  has a non-trivial nullspace.  
Hence,  $\det(\underline{A} - \lambda \underline{I}) = 0$ .

→ If  $\underline{A}$  has eigenvalues  $\lambda_1, \dots, \lambda_n$  & eigenvectors  $\underline{x}_1, \dots, \underline{x}_n$ ,  $\underline{A} + n \underline{I}$  will have eigenvalues  $(\lambda_1 + n), \dots, (\lambda_n + n)$ , and eigenvectors  $\underline{x}_1, \dots, \underline{x}_n$ . (Eigenvalues change, but eigenvectors remain the same.)

What if  $\underline{A}$  has eigenvalues  $\lambda_1, \dots, \lambda_n$  &  $\underline{B}$  has eigenvalues  $\alpha_1, \dots, \alpha_n$ , does  $(\underline{A} + \underline{B})$  have eigenvalues  $(\lambda_1 + \alpha_1), \dots, (\lambda_n + \alpha_n)$ ?

$$\begin{aligned} \underline{A}\underline{x} &= \lambda \underline{x} \\ \underline{B}\underline{x} &= \alpha \underline{x} \\ (\underline{A} + \underline{B})\underline{x} &= (\lambda + \alpha) \underline{x} \end{aligned}$$

we are making an assumption here! we are thinking that  $\underline{x}$  is also an eigenvector of  $\underline{B}$ , and that may not be true!

rotate a vector by 90°

- What if  $\underline{A}$  is a rotation matrix? E.g.  $\underline{A} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$

→ Because no vector  $\underline{x}$  can be in the same direction after this rotation, no real

eigen vectors exist for such a matrix.

→ Hence, even real matrices can have complex eigenvalues.

→ If a matrix is anti-symmetric (i.e.  $\tilde{A}^T = -\tilde{A}$ ), eigenvalues shall be purely complex.

→ For a triangular matrix, eigenvalues are the diagonal elements. Because  $\det(\text{triangular matrix}) = \text{product of diagonal elements}$ .

→ If  $\tilde{A}$  has eigenvalues  $\lambda_1, \dots, \lambda_n$  & eigenvectors  $\underline{x}_1, \dots, \underline{x}_n$ ,

→  $\tilde{A}^2$  will have eigenvalues  $\lambda_1^2, \dots, \lambda_n^2$  & eigenvectors  $\underline{x}_1, \dots, \underline{x}_n$ :

$$\text{Pf: } \tilde{A}^2 \underline{x} = \tilde{A}(\tilde{A} \underline{x}) = \tilde{A}(\lambda \underline{x}) = \lambda (\tilde{A} \underline{x}) = \lambda (\lambda \underline{x}) = \lambda^2 \underline{x} \quad (\text{shown})$$

→  $\tilde{A}^{-1}$  will have eigenvalues  $\frac{1}{\lambda_1}, \dots, \frac{1}{\lambda_n}$  & eigenvectors  $\underline{x}_1, \dots, \underline{x}_n$ :

$$\text{Pf: } \tilde{A}^{-1} \underline{v} = \tilde{A} \left( \frac{\underline{A} \underline{v}}{\lambda} \right) = \frac{\underline{v}}{\lambda} = \frac{1}{\lambda} \cdot \underline{v}$$
$$\begin{bmatrix} \underline{A} \underline{v} = \lambda \underline{v} \\ \underline{v} = \frac{\underline{A} \underline{v}}{\lambda} \end{bmatrix}$$

### Defective Eigenvalues & Matrices

14/10/25

$$\text{Let } \tilde{A} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}, \quad \tilde{B} = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}$$

$\tilde{A}$  &  $\tilde{B}$  have  $\lambda=2$ , with algebraic multiplicity 3.

For  $\tilde{A}$ , we can have 3 L.I. eigenvectors, i.e.  $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$  &  $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ , and geometric multiplicity = 3.

For  $\tilde{B}$ , we have only one eigenvector, i.e.  $\underline{e}_1$ . Hence, the geometric multiplicity = 1.

→ An eigenvalue, whose algebraic multiplicity = geometric multiplicity is called

a defective eigenvalue.

→ A matrix that has at least one defective eigenvalue is a defective matrix, i.e. it does not possess a full set of  $m$  L.I. eigenvectors.

No eigen decomposition

→ Diagonal matrix is not defective.

→ Diagonalsability:

If  $\tilde{A} \in \mathbb{R}^{m \times n}$  is not defective iff it has eigenvalue decomposition:

$$\tilde{A} = \tilde{X} \tilde{\Delta} \tilde{X}^{-1}, \text{ where } \tilde{X} \in \mathbb{R}^{n \times m}, \det(\tilde{X}) \neq 0.$$

If a): Given a diagonalsable matrix ( $\tilde{A} = \tilde{X} \tilde{\Delta} \tilde{X}^{-1}$ ),  $\tilde{A}$  is non-defective

$\tilde{A}$  &  $\tilde{\Delta}$  are similar, hence they must have same eigenvalues (and corresponding geometric/algebraic multiplicity). Because  $\tilde{\Delta}$  is not defective,  $\tilde{A}$  is also not defective.

$\tilde{\Delta}$  is a  
diagonal  
matrix

If b): Given  $\tilde{A}$  is not defective,  $\tilde{A}$  can be diagonalsed s.t  $\tilde{A} = \tilde{X} \tilde{\Delta} \tilde{X}^{-1}$ .

Let  $\tilde{X} = \begin{bmatrix} \tilde{x}_1 & \dots & \tilde{x}_n \end{bmatrix}$  eigen vectors of  $\tilde{A}$

$$\text{Then } \tilde{A} \tilde{X} = \begin{bmatrix} \tilde{x}_1 & \dots & \tilde{x}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} = \tilde{X} \tilde{\Delta}$$

$$\therefore \tilde{A} = \tilde{X} \tilde{\Delta} \tilde{X}^{-1}$$

## Unitary (Orthogonal) Diagonalizability

→ Let  $\tilde{Q} \in \mathbb{C}^{m \times m}$  is a unitary matrix if  $\tilde{Q}^\perp \tilde{Q} = \tilde{Q} \tilde{Q}^\perp = I$

$\tilde{Q}^\perp$  represents conjugate transpose of  $\tilde{Q}$  conjugate

If for a non-defective matrix  $A \in \mathbb{R}^{m \times m}$ , we define unitary diagonalizability if there a unitary matrix  $\tilde{Q}$  s.t  $A = \tilde{Q} \tilde{\Delta} \tilde{Q}^{-1} = \tilde{Q} \tilde{\Delta} \tilde{Q}^\perp$ .  
\* not all non-defective matrix are unitary diagonalizable

## Symmetric Matrix & Eigenvalues & Eigenvectors

→ Asymmetric matrix has real eigenvalues & real eigenvectors.

If : A symmetric matrix has real eigenvalues :

$$\sum_{\tilde{x}} \tilde{x} = \lambda \tilde{x} \quad ①$$

$$\sum_{\tilde{x}} \tilde{x}^* = \lambda^* \tilde{x}^* \quad [\text{conjugate both sides}]$$

$$\sum_{\tilde{x}} \sum_{\tilde{x}} \tilde{x}^* = \lambda^* \sum_{\tilde{x}} \tilde{x}^* \tilde{x} \quad ② \quad [\text{pre-multiply } \tilde{x}^T \text{ on both sides}]$$

$$(\tilde{x}^*)^T \sum_{\tilde{x}} \tilde{x} = \lambda \tilde{x}^T \tilde{x} \quad ③$$

$$③ \Rightarrow (\sum_{\tilde{x}})^T \tilde{x}^* = \lambda \tilde{x}^T \tilde{x}$$

$$④ \Rightarrow (\sum_{\tilde{x}})^T \tilde{x}^* = \lambda^* \tilde{x}^T \tilde{x}$$

$$\mathcal{O} = (\lambda - \lambda^*) \underline{x}^* \underline{x}$$

$\lambda = \lambda^*$   $\Rightarrow \lambda$  is real valued

$(\lambda_1, \underline{x}_1)$  &  $(\lambda_2, \underline{x}_2)$  are two eigenpairs, where  $\lambda_1 \neq \lambda_2$

$$\underline{x}_1 = \lambda_1 \underline{x}_1$$

$$\underline{x}_2 = \lambda_2 \underline{x}_2$$

$$\underline{x}_1^T \underline{x}_2 = \lambda_1 \underline{x}_1^T \underline{x}_1$$

$$\underline{x}_1^T \underline{x}_2 = \lambda_2 \underline{x}_1^T \underline{x}_2$$

$$(\delta \underline{x}_2)^T \underline{x}_1 = \lambda_1 \underline{x}_2^T \underline{x}_1 \quad \textcircled{1}$$

$$(\delta \underline{x}_2)^T \underline{x}_1 = \lambda_2 \underline{x}_2^T \underline{x}_1 \quad \textcircled{2}$$

$$\textcircled{1} - \textcircled{2} \Rightarrow (\lambda_1 - \lambda_2) \underline{x}_1^T \underline{x}_2 = 0$$

$$\underline{x}_1^T \underline{x}_2 = 0 \quad [\lambda_1 \neq \lambda_2]$$

What happens if eigenvalues of  $\underline{\Sigma}$  are degenerate? repeated eigenvalues

Let eigenvalue  $\lambda$  be repeated for  $\underline{\Sigma} \in \mathbb{R}^{m \times m}$ , where the geometric multiplicity is  $r < m$ . To prove that  $\underline{\Sigma}$  is not defective, we need to show that algebraic multiplicity of  $\underline{\Sigma}$  is also  $r$ .

Because geometric multiplicity of  $\lambda = r$ , we can find  $r$  L.I. vectors. We can use these to construct a  $r$ -dimensional subspace  $V^E$ .

$V^E = \{\underline{v}_1, \dots, \underline{v}_r\}$ , and let  $\underline{v}_1 \dots \underline{v}_r$  be orthogonal to each other.

Let us construct orthogonal vectors:  $V^\perp = \{v_{r+1}, \dots, v_m\}$ .  $v_{r+1}, \dots, v_m$  are orthogonal to each other, and  $V^\perp$  is orthogonal to  $V^F$ .

Let  $\tilde{V} = \begin{bmatrix} \tilde{v}_1 & \dots & \tilde{v}_r & \tilde{v}_{r+1} & \dots & \tilde{v}_m \end{bmatrix}$ , then:

$$\sum \tilde{v}_i = \sum_{j=1}^m c_{ji} \tilde{v}_j \quad [\sum \tilde{v}_i \text{ can be expressed as a linear combination}]$$

$$\therefore c_{ji} = \lambda \delta_{ji} + 1 \leq i \leq r$$

Also:

$$S \begin{bmatrix} \tilde{v}_1 & \dots & \tilde{v}_r & \tilde{v}_{r+1} & \dots & \tilde{v}_m \end{bmatrix}$$

$$= \begin{bmatrix} \tilde{v}_1 & \dots & \tilde{v}_r & \tilde{v}_{r+1} & \dots & \tilde{v}_m \end{bmatrix}$$

$$\left[ \begin{array}{cccc|cc} \lambda & & & & c_{1,r+1} & c_{1,r+2} & \dots & c_{1,m} \\ & \ddots & & & \vdots & & & \vdots \\ 0 & & \lambda & & c_{r,r+1} & & & c_{r,m} \\ & & & \ddots & \vdots & & & \vdots \\ & & & 0 & c_{r+1,r+1} & & & c_{r+1,m} \\ & & & & \vdots & & & \vdots \\ & & & & c_{m,r+1} & & & c_{m,m} \end{array} \right]$$

$\tilde{M}$   
 $\tilde{M}$  will be symmetric

$$\text{Hence, } \tilde{V} = \tilde{V} \tilde{M} \Rightarrow \tilde{V}^T \tilde{V} = \tilde{M}$$

Hence:

$$\underbrace{V^T S V}_{\sim \sim \sim} = \left[ \begin{array}{ccc|c} \lambda & \dots & 0 & 0 \\ 0 & \dots & \lambda & 0 \\ \hline 0 & & 0 & \zeta \end{array} \right], \text{ where } \zeta \text{ is a matrix with dimensions } (m-r) \times (m-r)$$

represent "characteristic polynomial of"

Hence  $P_{\underbrace{V^T S V}_{\sim \sim \sim}}(z) = P_M(z) = P_\zeta(z)$  Because  $M$  is similar to  $\zeta$

$$\det(\zeta - z I_m) = (\lambda - z)^r \det(\zeta - z I_r)$$

We have to show that  $\lambda$  is not an eigenvalue of  $\zeta$ . If  $\lambda$  is an eigenvalue of  $\zeta$ , then:

$$\zeta u = \lambda u, \quad u \in \mathbb{R}^{m-r} \text{ eigenvector of } \zeta$$

$$\hat{u} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ u \end{bmatrix}$$

$$\underbrace{V^T S V}_{\sim \sim \sim} \hat{u} = \underbrace{M}_{\sim \sim} \hat{u} = \left[ \begin{array}{ccc|c} \lambda & \dots & 0 & 0 \\ 0 & \dots & \lambda & 0 \\ \hline 0 & & 0 & \zeta \end{array} \right] \begin{bmatrix} 0 \\ \vdots \\ 0 \\ u \end{bmatrix} = \lambda \hat{u}$$

Hence  $\lambda$  is an eigenvalue of  $\underbrace{V^T S V}_{\sim \sim \sim}$ :  $\underbrace{V^T S V}_{\sim \sim \sim} \hat{u} = \lambda \hat{u}$

$$\underbrace{S V}_{\sim \sim} \hat{u} = \lambda \underbrace{V^T}_{\sim} \hat{u}$$

By construction,  $\underbrace{V^T}_{\sim} \hat{u}$  is an eigenvector of  $\zeta$  corresponding to eigenvalue  $\lambda$ . It is a vector that lies in the space spanned by  $\{v_{l+1}, \dots, v_m\}$ . This means we are able to find another L.I. eigenvector  $\underbrace{V^T}_{\sim} \hat{u}$  in addition to

what we had, i.e.  $\{\underbrace{v_1 \dots v_r}\}$ . This leads to a contradiction. Hence,  $\lambda$  cannot be an eigenvalue of  $\underbrace{A - zI}_n$ .

$\Rightarrow$  Proven: The algebraic multiplicity = geometric multiplicity for a symmetric matrix. Hence, non-symmetric matrices are not defective.

### Summary

$\rightarrow$  A symmetric matrix  $\underbrace{S} \in \mathbb{R}^{m \times m}$  is always non-defective and is orthogonally diagonalizable with real eigenvalues, i.e.  $\underbrace{S = Q \Delta Q^T}$ , where  $\underbrace{Q}$  is an orthogonal matrix &  $\Delta$  is a diagonal matrix.

$\rightarrow$  A skew-symmetric matrix  $\underbrace{W} \in \mathbb{R}^{m \times m}$  is also non-defective and has purely imaginary eigenvalues.  $\underbrace{W}$  is also unitary diagonalizable:  $\underbrace{W = Q \Delta Q^{-1}}$

$\rightarrow$  Normal matrices: Any matrix  $\underbrace{A} \in \mathbb{C}^{m \times m}$ , s.t.  $\underbrace{A^\perp A = A A^\perp}$

$\rightarrow$  A matrix  $\underbrace{A} \in \mathbb{C}^{m \times m}$  is unitary diagonalizable iff it is normal.



### Schur Factorization

$\downarrow$  SVD of  $T =$  SVD of  $A^\dagger$

$\rightarrow$  A factorization of  $\underbrace{A} \in \mathbb{C}^{m \times m}$  of the form:  $\underbrace{A = Q T Q^\perp}$ , where  $\underbrace{Q}$  is unitary, i.e.  $\underbrace{Q^\perp Q = I}_n$ , and  $\underbrace{T}$  is upper triangular.

$\rightarrow$  Every square matrix admits for every square matrix (even for those of which EVD does not exist). This makes it useful in computations.

$\downarrow$  orthogonal)

$\rightarrow$  If  $\underbrace{A} \in \mathbb{R}^{m \times m}$ , a factorization with real matrices exists, where  $\underbrace{A = V T V^T}$ ,

Where  $\tilde{L}$  is quasi upper triangular!

Either  $1 \times 1$  blocks in the diagonal (normal UTM) or  $2 \times 2$  blocks in the diagonal.

Usually occurs with complex eigenvalues!

→ Schur factorization need not be unique.

→ We can guarantee pure upper triangular for  $\tilde{L}$ , if we allow for complex arithmetic.

### Summary of Decompositions

→ A diagonalization  $A = \tilde{X} \tilde{\Delta} \tilde{X}^{-1}$  exists iff  $\tilde{A}$  is non-defective

$$A = Q \tilde{\Delta} Q^{-1}$$

→ " unitary triangularization called Schur factorization always exists for a q. matrix  $\tilde{L}$  of the form  $A = Q \tilde{T} Q^{-1}$

→ All three of the above are basically similarity transformations. Hence, we know that  $A$  and  $\tilde{\Delta}$  /  $\tilde{T}$  will have same eigenvalues, which is why we use these to get the eigenvalues of  $A$ .

### Eigenvalue Algorithms

→ Solving very characteristic eqn. is unstable!

→ Two classes:

→ Direct: We can get  $\infty$  precision in finite no. of steps

→ Iterative:  $\infty \infty \infty \infty \infty \infty \infty$  iterations

→ For  $m \leq 4$ , we can have a direct method i.e. solving characteristic

eqns). It may be inaccurate, but it's direct.

→ For  $m \geq 5$ , we have no way to deterministically solve a degree- $5$  polynomial eqn. Hence, we shall have to use iterative methods.

→ Monic polynomial:  $p(z) = z^m + a_{m-1}z^{m-1} + \dots + a_1z + a_0$ .

Roots of  $p(z)$  are equal to eigenvalues of this matrix:

$$\tilde{A} = \begin{bmatrix} 0 & 0 & & -a_0 \\ 1 & 0 & & -a_1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -a_{m-2} \\ 1 & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & -a_{m-1} \end{bmatrix}$$

$\tilde{A}$  is the " companion" matrix of  $p(z)$

No algorithm can exactly produce all the roots of an arbitrary polynomial in a finite no. of steps.

→ Most eigen solvers reduce the matrix to Schur factorisation, where they can just read the eigenvalues from the diagonal of  $\tilde{T}$ , because of the similarity transformation that exists between  $\tilde{A}$  &  $\tilde{T}$ .

Schur triangulization: similar to householder tridiagonalisation

$$Q_1^\perp \cdots \tilde{Q}_j^\perp \tilde{A} \tilde{Q}_1 \cdots \tilde{Q}_j = \tilde{T}$$

otherwise it will be a direct method!!!

We have to pre & post multiply to maintain the eigenvalues (similarity transformation)

If  $\tilde{A}$  is real & symmetric,  $\tilde{T}$  will be a diagonal matrix

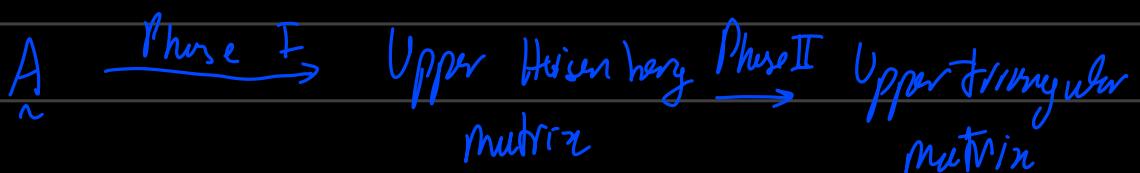
→ Two phases of eigenvalue computation:

1	4	2	3
3	4	1	7
0	2	3	4
0	0	1	3

→ Direct method to reduce  $\underline{A}$  to upper Hessenberg matrix

↳  $UTM$ , last with m  
addl. line of non-zero  
entries parallel to  
diagonal

→ An iterative method to produce a sequence of similarity transformations  
that reduce  $\underline{A}$  to converge to  $UTM$ .



→ Phase I :  $O(m^3)$  flops

→ " II :  $O(m^3)$ "

→  $O(m)$  no. of iterations are reqd.

→ Each iteration requires  $O(m^2)$

to reduce a Hessenberg matrix. Otherwise, we may have to reduce a full matrix, requiring  $O(m^3)$  flops per iteration, thereby bringing the total TC to  $O(m^4)$ . This is why 2-phased soln. is employed.

### Reduction to upper Hessenberg form using Householder

→ Householder transformation =  $Q_1 \underline{A}$

→ To preserve the eigenvalues, we need to post-multiply  $Q_1^T$  to  $Q_1 \underline{A}$  so that there is a similarity transformation b/t  $\underline{A} \Leftarrow Q_1 A Q_1^T$ . However, doing this will result in  $Q_1 \underline{A} Q_1^T$  becoming a dense matrix

again  $\textcircled{2}$ .

again this will result in determinate algo

→ Hence we cannot reduce  $\tilde{A}$  to a UTM / Diagonal matrix using this method.  
Therefore, we use upper Hessenberg matrix:

$$\begin{bmatrix} - & - & - & - & - \\ - & - & - & - & - \\ - & - & - & - & - \\ - & - & - & - & - \\ - & - & - & - & - \end{bmatrix} \xrightarrow{\text{(pre-multiply) } Q_1^\perp} \begin{bmatrix} - & - & - & - & - \\ - & - & - & - & - \\ 0 & - & - & - & - \\ 0 & - & - & - & - \\ 0 & - & - & - & - \end{bmatrix}$$

$$\begin{array}{c} \text{post-multiply} \\ \downarrow \end{array} \quad \begin{array}{c} \text{pre-multiply} \\ \downarrow \end{array} \quad \begin{array}{c} \text{Q}_2 \text{ (pre-multiply)} \\ \downarrow \end{array}$$

$$\begin{bmatrix} - & - & - & - & - \\ - & - & - & - & - \\ 0 & - & - & - & - \\ 0 & 0 & - & - & - \\ 0 & 0 & - & - & - \end{bmatrix} \xleftarrow{Q_2^\perp Q_2} \begin{bmatrix} - & - & - & - & - \\ - & - & - & - & - \\ 0 & - & - & - & - \\ 0 & - & - & - & - \\ 0 & - & - & - & - \end{bmatrix}$$

We repeat this process  $(m-2)$  times, after which  $\tilde{A}$  shall be reduced to upper Hessenberg form:

$$\underbrace{Q_{m-2}^\perp \cdots Q_1^\perp}_{Q_\sim^\perp} \underbrace{A}_{Q_\sim} \underbrace{Q_1 \cdots Q_{m-2}}_{Q_\sim} = \tilde{H} \Rightarrow \tilde{A} = \underbrace{Q_\sim}_{\sim} \tilde{H} \underbrace{Q_\sim^\perp}_{\sim}$$

For a real symmetric matrix with  $\lambda_1, \dots, \lambda_m$  real eigenvalues, and  $\tilde{q}_1, \dots, \tilde{q}_m$  orthogonal eigenvectors.

If  $\underline{x}$  is some vector close to an eigenvector, we have to find  $\alpha$  s.t.  $\|\underline{A}\underline{x} - \alpha \underline{x}\|_2$  is minimised. This is a least sq. problem.

Hence,

$$\begin{aligned} & \underline{x}\alpha = \underline{A}\underline{x} \\ \underline{A}\rightarrow \underline{x} & \quad \underline{x}^T \underline{x} \alpha = \underline{x}^T \underline{A}\underline{x} \\ \underline{x} \rightarrow \alpha & \quad \therefore \boxed{\alpha = \frac{\underline{x}^T \underline{A}\underline{x}}{\underline{x}^T \underline{x}}} \\ \underline{b} \rightarrow \underline{A}\underline{x} & \quad \text{reduces quotient, very useful} \end{aligned}$$

→ This  $\alpha$  is the natural eigenvalue estimate to consider if  $\underline{x}$  is approximately equal to an eigenvector.

21/10/2020

Given  $\underline{x} \Rightarrow$  we can say  $\underline{x} = \sum_{j=1}^m c_j \underline{v}_j$ , where  $c_j$ 's are basically eigenvalues of  $\underline{A}$ .

→ If  $\underline{x}$  is close to one of the eigenvectors  $\underline{v}_j$ , then  $\frac{|c_j|}{|c_i|} \leq \epsilon_j < 1$

Let  $\epsilon = \max_j (\epsilon_j) + \epsilon_i$ ,

then

$$\frac{|c_j|}{|c_i|} < \epsilon \quad \forall j \neq i$$

estimated eigenvector  
actual eigenvector

We can show that  $|r(\underline{x}) - r(\underline{v}_j)| = O(\epsilon^2)$

(Exam question possibility !!)

H.W.: Prove  $|r(\underline{x}) - r(\underline{v}_j)| = O(\epsilon^2)$  using the following:

$$r(\underline{x}) = \frac{\underline{x}^T \underline{A} \underline{x}}{\underline{x}^T \underline{x}} = \underline{x}^T \underline{A} \underline{x} = \sum_{j=1}^m c_j^2 \lambda_j^2$$

Hint: Estimate O.B. for  $\left| \sum_{j=1}^m \zeta_j^2 \lambda_j^2 - \text{r}(A_S) \right| = O(\varepsilon^2)$

" $\lambda$ : Use triangular inequality"

### Eigenvalue Algorithms

#### Power iteration

by magnitude

→ Objective: Find the largest eigenvalue

→ Method:

Let  $A \in \mathbb{R}^{m \times m}$ ,  $v^{(0)}$  is a random vector

1) Get  $v^{(0)}$

2) Get  $\frac{\tilde{A} \tilde{v}^{(0)}}{\|\tilde{A} \tilde{v}^{(0)}\|}$

3) Get  $\frac{\tilde{A} \tilde{v}^{(0)}}{\|\tilde{A} \tilde{v}^{(0)}\|}$  and then normalize it

Repeat this n times. Eventually we get  $\frac{\tilde{A}^n \cdot \tilde{v}^{(0)}}{\|\tilde{A}^n \cdot \tilde{v}^{(0)}\|}$ .

This approaches the longest eigenvector.

by magnitude

Proof:  $\tilde{v}^{(0)} = \underbrace{a_1}_{\sim} \underbrace{q_1}_{\sim} + \dots + \underbrace{a_m}_{\sim} \underbrace{q_m}_{\sim}$

(Because  $\tilde{v}^{(0)}$  is not an eigenvector, it should be having some components of  $q_1, q_2, \dots, q_m$ )

$$\tilde{A}^n \tilde{v}^{(0)} = a_1 \lambda_1^n \underbrace{q_1}_{\sim} + \dots + a_m \lambda_m^n \underbrace{q_m}_{\sim}$$

Let  $|\lambda_1| > |\lambda_2| > \dots > |\lambda_m|$ , then :

$$\underset{n}{\underbrace{A^n}} \underset{\sim}{v}^{(0)} = \alpha_1 \lambda_1^n \left( \underset{\sim}{q}_1 + \frac{\alpha_2}{\alpha_1} \cdot \left( \frac{\lambda_2}{\lambda_1} \right)^n \underset{\sim}{q}_2 + \dots + \frac{\alpha_m}{\alpha_1} \cdot \left( \frac{\lambda_m}{\lambda_1} \right)^n \underset{\sim}{q}_m \right)$$

Because  $\left| \frac{\lambda_j}{\lambda_1} \right| < 1 \quad \forall j=1, \dots, m$ , all these terms tend to 0 as  $n \rightarrow \infty$ .

$$\therefore \underset{n}{\underbrace{A^n}} \underset{\sim}{v}^{(0)} = \alpha_1 \lambda_1^n \underset{\sim}{q}_1$$

Hence  $\underset{n}{\underbrace{v}}^{(n)} = \frac{\underset{n}{\underbrace{A^n}} \underset{\sim}{v}^{(0)}}{\| \underset{n}{\underbrace{A^n}} \underset{\sim}{v}^{(0)} \|} \rightarrow \underset{\sim}{q}_1 \text{ as } n \rightarrow \infty$

→ This method converges very slowly for repeated eigenvalues.

→ For this method  $\underset{\sim}{v}^{(0)}$  must not be orthogonal to  $\underset{\sim}{q}_j$ . However, the chance of this is very low (because we are using a random vector).

Algo :

$$\text{Initialise } \underset{\sim}{v}^{(0)}, \text{ s.t. } \| \underset{\sim}{v}^{(0)} \| = 1$$

for  $n=1 \rightarrow \infty$

$$\underset{\sim}{w} = \underset{n}{\underbrace{A}} \underset{\sim}{v}^{(n-1)}$$

$$\underset{\sim}{v}^{(n)} = \frac{\underset{\sim}{w}}{\| \underset{\sim}{w} \|}$$

$$\lambda^{(n)} = (\underset{\sim}{v}^{(n)})^T \underset{n}{\underbrace{A}} \underset{\sim}{v}^{(n)}$$

Result: Let  $|\lambda_1| \geq |\lambda_2| \cdots \geq |\lambda_m|$  and  $\tilde{q}_1^\top \tilde{v}^{(0)} \neq 0$ , then iteration of the power iteration shall satisfy:

$$\|\tilde{v}^{(n)} - (\pm \tilde{q}_1)\| = O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^n\right)$$

$\tilde{v}^{(n)}$  tends to  $\tilde{q}_1$  or  $-\tilde{q}_1$  and

$$\text{ratio of } \left|\lambda^{(n)} - \lambda_1\right| = O\left(\left|\frac{\lambda_2}{\lambda_1}\right|^{2n}\right)$$

Because of ratio of eigenvalues, error in eigenvalue is square of error in eigenvector

If  $n$  is even,  $\tilde{v}^{(n)} \rightarrow \tilde{q}_1$   
Else,  $\tilde{v}^{(n)} \rightarrow -\tilde{q}_1$

Shortcomings of this method:

1) We are finding eigenvector corresponding to longest eigenvalue only. What about the remaining eigenvalues / eigenvectors?

2) Convergence is linear with error being reduced by  $\left|\frac{\lambda_2}{\lambda_1}\right|$  at every iteration

$\rightarrow$  if  $\lambda_2 \approx \lambda_1$ , convergence can be very slow.

### Inverse Power Iteration

$\rightarrow \mu \in \mathbb{R}$  that is not eigenvalue of  $\tilde{A}$ , the eigenvectors of  $(\tilde{A} - \mu \tilde{I})^{-1}$  are some of  $\tilde{A}$ . The corresponding eigenvalues are  $\left\{\frac{1}{\lambda_j - \mu}\right\}_{j=1}^m$ , where  $\{\lambda_j\}$  are eigenvalues of  $\tilde{A}$ .

→ Suppose  $\mu$  is close to  $\lambda_3$  of  $\tilde{A}$ , then  $\frac{1}{\lambda_3 - \mu}$  will be much larger  
 then  $\frac{1}{\lambda_j - \mu}$  if  $j \neq 3$

→ Apply power iteration to  $(\tilde{A} - \mu \tilde{I})^{-1}$ , then the process converges to  $q_3$

Algo:

- Initialise  $\mu = \text{some value near } \lambda_3$ ,
- $\tilde{v}^{(0)} = \text{"random vector"}$
- for  $n = 1 \rightarrow \infty$ :
- $\tilde{w} = (\tilde{A} - \mu \tilde{I})^{-1} \tilde{v}^{(n-1)}$

We consider for  $\tilde{w}$  in  $\tilde{v}^{(n)} = \frac{\tilde{w}}{\|\tilde{w}\|}$

$$(\tilde{A} - \mu \tilde{I}) \tilde{w} = \tilde{v}^{(n-1)} \quad \lambda^{(n)} = (\tilde{v}^{(n)})^T \tilde{A} \tilde{v}^{(n)}$$

Although this is badly conditioned, this does not matter so much due to the normalisation.

Suppose  $\lambda_3$  is the closest eigenvalue to  $\mu$  and  $\lambda_n$  is the second closest:

$$|\mu - \lambda_3| < |\mu - \lambda_n| \leq \dots \leq |\mu - \lambda_i|,$$

$\forall j \neq 3$  and  $\tilde{q}_j^T \tilde{v}^{(0)} = 0$ , then iterates of inverse iterations  $\tilde{v}^{(n)}$  satisfy:

$$\|\tilde{v}^{(n)} - (\pm q_3)\| = O\left(\left|\frac{\mu - \lambda_3}{\mu - \lambda_n}\right|^n\right)$$

and

$$|\lambda^{(n)} - \lambda_3| = O\left(\left|\frac{\mu - \lambda_3}{\mu - \lambda_n}\right|^{2n}\right)$$

## Summary:

- Power iteration: eigen vector estimate → eigen value estimate  
 → Inverse power iteration: eigen value estimate → eigen vector estimate

## Rayleigh quotient iteration

Algo:  
 Initialize  $\tilde{v}^{(0)}$  to some vector s.t  $\|\tilde{v}^{(0)}\|_2 = 1$   
 $\lambda^{(0)} = (\tilde{v}^{(0)})^\top \tilde{A} \tilde{v}^{(0)}$

for  $n=1 \rightarrow \infty$

$$(\tilde{A} - \lambda^{(n-1)} \tilde{I}) \tilde{w} = \tilde{v}^{(n-1)}$$

$$\tilde{q}^{(n)} = \frac{\tilde{w}}{\|\tilde{w}\|}$$

$$\lambda^{(n)} = (\tilde{q}^{(n)})^\top \tilde{A} \tilde{q}^{(n)}$$

23/10/25

Convergence is very fast! We can get eigenvalues accurate to 10 d.p. in 3-4 iterations.

Hence  $\|\tilde{v}^{(n+1)} - (\pm q_j)\| = O(\|\tilde{v}^{(n)} - (\pm q_j)\|^3)$

and  $|\lambda^{(n+1)} - \lambda_j| = O(|\lambda^{(n)} - \lambda_j|^3)$  Because we are relating the current & prev. iteration of the eigenvalue.

## Operation counts

Let  $A \in \mathbb{R}^{m \times m}$  be dense.

→ Each step of power iteration has one matrix-vector multiplication, thereby requiring

$O(m^2)$  flops.

- Each step of inverse power iteration requires a soln. of linear system of eqns, which has a cost of  $O(m^3)$  flops.
  - We can solve for  $(\tilde{A} - \mu \tilde{I})^{-1}$  once (outside the loop) because  $(\tilde{A} - \mu \tilde{I})^{-1}$  is a constant. We can reduce the per-iteration time complexity to  $O(m^2)$
- Rayleigh quotient iteration requires  $O(m^5)$  time per operation, but it is still worth it because lesser iterations are reqd.
- \* If  $\tilde{A}$  has been reduced to tridiagonal / upper Hessenberg, time complexity per-iterations shall be reduced to  $O(m^2)$

### Multiple Eigenvectors

#### Subspace iterations (Simultaneous iterations)

- Instead of one vector, we have multiple vectors which are linearly independent to each other.
- \* If we have an  $\infty$  precision computer, all of these vectors shall converge to different eigen vectors. Otherwise, they shall all converge to the "largest" eigenvector.

Let  $\tilde{V}^{(0)} = \begin{bmatrix} | & & | \\ V_1^{(0)} & \dots & V_n^{(0)} \\ | & & | \end{bmatrix}$ ,  $\tilde{V}^{(n)} = \begin{bmatrix} | & & | \\ \tilde{A}^{(n)} V^{(0)} & & V^{(0)} \\ | & & | \end{bmatrix}$

$$= \begin{bmatrix} | & & | \\ V_1^{(n)} & \dots & V_n^{(n)} \\ | & & | \end{bmatrix}$$

Let the eigenvectors of  $\tilde{A}$  be denoted by  $\tilde{q}_1, \dots, \tilde{q}_n$ , and a matrix  $\tilde{Q}$  is

$$\tilde{Q} = \begin{bmatrix} 1 & 1 \\ \tilde{q}_1 & \cdots & \tilde{q}_n \\ \tilde{1} & & \tilde{1} \end{bmatrix}$$

Proof:  $V^{(0)} = \begin{bmatrix} 1 \\ \tilde{1} \end{bmatrix}$

$$= \tilde{Q} \tilde{Q}^T V^{(0)} \quad \left[ \tilde{C} = \tilde{Q}^T \tilde{V}^{(0)} \right]$$

$$= \tilde{V}^{(0)} \quad (\text{shown})$$

Hence  $V^{(k)} = \tilde{A}^k \cdot V^{(0)}$

$$= (\tilde{Q} \tilde{\Delta} \tilde{Q}^T)^k \cdot V^{(0)}$$

$$= \tilde{Q} \tilde{\Delta}^k \tilde{Q}^T \cdot V^{(0)} \quad [\text{Q \& Q}^T \text{ are orthogonal}]$$

$$= \tilde{Q} \tilde{\Delta}^k \tilde{Q}^T \cdot \tilde{Q} \tilde{C} \quad \left[ V^{(0)} = \tilde{Q} \tilde{C} \right]$$

$$= \tilde{Q} \tilde{\Delta}^k \tilde{C}$$

"unwanted eigenvalues"

To extract first  $n$  eigenvalues:

$$\text{Let } \tilde{Q} = \begin{bmatrix} \tilde{Q}_1 & \tilde{Q}_2 \end{bmatrix}, \quad \tilde{Q}_1 = \begin{bmatrix} \tilde{q}_1 & \cdots & \tilde{q}_n \end{bmatrix}, \quad \tilde{Q}_2 = \begin{bmatrix} \tilde{q}_{n+1} & \cdots & \tilde{q}_m \end{bmatrix}$$

$$\text{Let } \tilde{\Delta} = \begin{bmatrix} \tilde{\Delta}_1 & 0 \\ 0 & \tilde{\Delta}_2 \end{bmatrix} \quad \text{and} \quad \tilde{C} = \begin{bmatrix} C_1 \\ \vdots \\ C_n \end{bmatrix}$$

$$\text{Hence, } \underline{V}_{\sim}^{(n)} = \underline{Q}_{\sim} \underbrace{\Delta_{\sim 1}^n}_{\sim} \zeta_1 + \underline{Q}_{\sim} \underbrace{\Delta_{\sim 2}^n}_{\sim} \zeta_2$$

$$= \left( \underline{Q}_{\sim} \underbrace{\Delta_{\sim 1}^n}_{\sim} + \underline{Q}_{\sim} \underbrace{\Delta_{\sim 2}^n}_{\sim} \zeta_2 \zeta_1^{-1} \right) \zeta_1 \quad [\text{Factor out } \zeta_1]$$

\* Assumption:  $\zeta_1$  is invertible

$$= \left( \underbrace{\underline{Q}_{\sim} \Delta_{\sim 1}^n}_{\text{dominant spectrum}} + O\left(\|\lambda_{n+1}\|^n\right) \right) \zeta_1$$

$\nearrow$  "unwanted" spectrum

As  $n \rightarrow \infty$ , we are magnifying the first term.

Result: Assumption #1: The first  $n$  eigenvalues are distinct and well-separated from others, i.e.  $|\lambda_1| > |\lambda_2| \dots > |\lambda_n| > |\lambda_{n+1}| \geq \dots \geq |\lambda_m|$

#2: If  $\underline{Q}_{\sim} = [\underline{q}_1 \dots \underline{q}_n] \in \mathbb{R}^{m \times n}$ ,  $\underline{Q}_{\sim}^T \underline{V}_{\sim}^{(0)}$  is non-singular, but all principal submatrices of  $\underline{Q}_{\sim}^T \underline{V}_{\sim}^{(0)}$  are also non-singular.

$$\underline{V}_{\sim}^{(n)} = \underline{A}_{\sim}^n \underline{V}_{\sim}^{(0)} \implies \underline{Q}_{\sim}^{(n)} \underline{R}_{\sim}^{(n)} = \underline{V}_{\sim}^{(n)}$$

QR factorization

Then as  $n \rightarrow \infty$ , the cols. of  $\underline{Q}_{\sim}^{(n)}$  converges linearly to the eigenvectors of  $\underline{A}_{\sim}$ .

$$\left\| \underline{q}_{\sim j}^{(n)} - (\pm \underline{q}_{\sim j}) \right\| = O(C^n) \text{ for } j = 1 \rightarrow m, \text{ where}$$

$$C = \max_{1 \leq n \leq n} \left| \frac{\lambda_{n+1}}{\lambda_n} \right| < 1$$

This is unnormalised simultaneous Lanczos iterations, which is numerically unstable, due to loss of orthogonality.

Let  $\tilde{V}_n^{(0)}$  be a set of random vectors, which are linearly independent.

$$\tilde{V}_n^{(0)} = \tilde{A} \tilde{V}_n^{(0)} = \begin{bmatrix} V_1^{(1)} & V_2^{(1)} & \dots & V_n^{(1)} \end{bmatrix}$$

$$= \tilde{Q}_n^{(1)} \tilde{R}_n^{(1)}$$

↑ This ensures that  $\tilde{V}_1^{(1)}, \tilde{V}_2^{(1)}, \dots, \tilde{V}_n^{(1)}$  are orthonormal

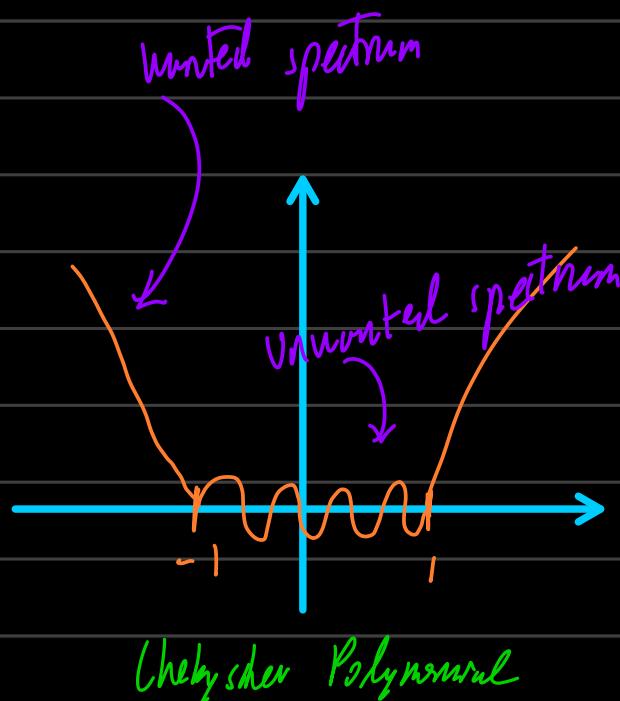
Hence, at every step we are orthonormalizing.

Alg: Choose  $\tilde{Q}_n^{(0)} \in \mathbb{R}^{m \times n}$

for  $n = 1 \rightarrow \infty$

$$\tilde{Z}_n^{(n)} = \tilde{A} \tilde{Q}_n^{(n-1)}$$

$$\tilde{Q}_n^{(n)} \tilde{R}_n^{(n)} = \tilde{Z}_n^{(n)}$$



We can construct  $\hat{A}$  from  $\tilde{A}$  such that  $\hat{A}$  has the wanted spectrum to the left of  $-1$ , and the unwanted spectrum between  $-1$  &  $1$ .

This is a practically applicable method for large sparse matrices.

### QR Algorithm

→ Used mostly with dense matrices

Alg0:  $\tilde{A}^{(0)} = A$  ↑ This is mathematically same as subspace iteration, but makes it better conditioned.

for  $k = 1 \rightarrow \infty$

$$Q^{(n)} R^{(n)} = A^{(n-1)} \quad // \text{orthogonalise}$$

$$A^{(n)} = R^{(n)} Q^{(n)}$$

diagonalised  
↓

Eventually, for a real symmetric matrix,  $A^{(n)}$  will approach Schur form, where eigenvalues can simply be read.

This shall also be useful in SVD, as SVD is essentially an eigenproblem.

Proof:  $Q^{(n)} R^{(n)} = A^{(n-1)}$

$$R^{(n)} = (Q^{(n)})^T A^{(n-1)}$$

28/10/25

Substituting,  $A^{(n)} = R^{(n)} Q^{(n)}$   
 $= (Q^{(n)})^T A^{(n-1)} Q^{(n)}$  (similarly transformation)

→ Simultaneous iteration is equivalent to QR algorithm.

If: Let  $\tilde{Q}_R^{(n)} = \underbrace{\tilde{Q}^{(1)} \cdot \tilde{Q}^{(2)} \cdots \cdot \tilde{Q}^{(n)}}_{\sim} = \prod_{i=1}^n \tilde{Q}^{(i)}$

"  $\tilde{R}_R^{(n)} = \underbrace{\tilde{R}^{(n)} \cdot \tilde{R}^{(n-1)} \cdots \cdot \tilde{R}^{(1)}}_{\sim} = \prod_{i=k}^1 \tilde{R}^{(i)}$

What happens to  $\tilde{A}^n$ ?

When  $n=1$ :  $\tilde{A} = \tilde{A}^{(0)} = \underbrace{\tilde{Q}^{(1)}}_{\sim} \underbrace{\tilde{R}^{(1)}}_{\sim} = \underbrace{\tilde{Q}_R^{(1)}}_{\sim} \underbrace{\tilde{R}_R^{(1)}}_{\sim}$

"  $n=2$ :  $\tilde{A}^2 = \underbrace{\tilde{A} \cdot \tilde{A}}_{\sim} = (\underbrace{\tilde{Q}_R^{(1)} \tilde{R}_R^{(1)}}_{\sim}) (\underbrace{\tilde{Q}_R^{(1)} \tilde{R}_R^{(1)}}_{\sim})$



Simultaneous subspace iteration: Let  $\tilde{Q}_\sim^{(0)} = I$

for  $n = 1 \rightarrow \infty$ :

$$Z = \tilde{A} \tilde{Q}_\sim^{(n-1)}$$

$$Z = \tilde{Q}_\sim^{(n)} \tilde{R}_S^{(n)}$$

$$\tilde{A}_\sim^{(n)} = (\tilde{Q}_\sim^{(n)})^\top \tilde{A} \tilde{Q}_\sim^{(n)}$$

$$\tilde{R}_S^{(n)} = R_S^{(n)} \circ R_T^{(n-1)} \circ \dots \circ R_S^{(1)}$$

What is  $\tilde{A}_\sim^n$  in simultaneous iteration?

$$\text{When } n=1: \quad \tilde{A}_\sim^1 = \tilde{A} \tilde{Q}_\sim^{(0)} = \tilde{Q}_\sim^{(1)} \tilde{R}_S^{(1)}$$

$$\text{When } n=2: \quad \tilde{A}_\sim^2 = \tilde{A} \left( \tilde{Q}_\sim^{(1)} \tilde{R}_S^{(1)} \right)$$

$$= \tilde{A} \tilde{Q}_\sim^{(1)} \tilde{R}_S^{(1)}$$

$$= \tilde{Q}_\sim^{(2)} R_S^{(2)} \tilde{R}_S^{(1)}$$

$$= \tilde{Q}_\sim^{(2)} \tilde{R}_S^{(2)}$$

$$\text{Hence, } \tilde{A}_\sim^n = \tilde{Q}_\sim^{(n)} \tilde{R}_S^{(n)} \quad ③$$

 QR factorization is unique for full-rank matrices

From eqn ① & ③, and assuming  $\tilde{A}_\sim^n$  is full-rank, we know that  $\tilde{Q}_\sim^{(n)} = \tilde{Q}_R^{(n)}$  and  $\tilde{R}_S^{(n)} = R_R^{(n)}$ .

$$\text{We can also state that } \tilde{A}_\sim^n = (\tilde{Q}_\sim^{(n)})^\top \tilde{A} \tilde{Q}_\sim^{(n)}$$



$$\text{Hence, } \underbrace{(A^{-1})^n}_{\sim} P = \underbrace{\left(\bar{Q}^{(n)}\right)}_{\sim} \underbrace{\left(P\left(\bar{R}^{(n)}\right)^{-T}P\right)}_{\sim}$$

Therefore, at the  $n^{\text{th}}$  iteration, we can do the following to speed up convergence.

Full algo (modified QR): (most used algo in eigenvalues)

Define  $\underbrace{A^{(0)}}_{\sim}$  s.t.  $\underbrace{(Q^{(0)})^T}_{\sim} \underbrace{A^{(0)}}_{\sim} \underbrace{Q^{(0)}}_{\sim} = A$   $\| A^{(0)} \|$  is tridiagonalization of  $\underbrace{A}_{\sim}$

for  $n = 1 \rightarrow \infty$ :

Pick a shift  $\mu^{(n)}$  // there are many methods, e.g.  $\mu^{(n)} = A_{mm}^{(n-1)}$

$$\underbrace{Q^{(n)} R^{(n)}}_{\sim} = \underbrace{A^{(n-1)}}_{\sim} - \mu^{(n)} I \quad \text{// shifted QR factorization}$$

$$\underbrace{A^{(n)}}_{\sim} = \underbrace{R^{(n)}}_{\sim} \underbrace{Q^{(n)}}_{\sim} + \mu^{(n)} I \quad \text{// deflating strategy}$$

If any off-diagonal entries is sufficiently close to 0, set  $A_{i,j+1} = A_{j+1,j} = 0$

Split  $\underbrace{A^{(n)}}_{\sim}$  s.t.  $\underbrace{A^{(n)}}_{\sim} = \begin{bmatrix} \underbrace{A_1}_{\sim} & 0 \\ 0 & \underbrace{A_2}_{\sim} \end{bmatrix}$  // decoupling the matrix  
 $\| A^{(n)} \|$  need not be tridiagonal

Apply QR algo on  $\underbrace{A_1}_{\sim}$  &  $\underbrace{A_2}_{\sim}$  // start again from tri-diagonalisation

So far, the discussion on eigenvalue algorithms is half-direct, half-iterative. Let us now discuss purely iterative algorithms.

### Fully Iterative Methods

## Advantages of fully iterative methods :

- It can be faster
  - If a matrix exceeds a particular size, we can devise a much more efficient eigensolver; especially if the matrix is sparse.
  - Based on subspace methods
- 30 / 60 / 25
- subspace must be rich in the soln. that we are seeking. In this case, the subspace must be rich in eigenvectors. Such a subspace is called a Krylov subspace.
- Krylov subspace :  $\tilde{A} \in \mathbb{R}^{m \times m}$ ,  $b \in \mathbb{R}^m$ . The Krylov sequence is the set of vectors  $\tilde{b}, \tilde{A}\tilde{b}, \tilde{A}^2\tilde{b}, \dots$ . A subspace generated in this sequence is a Krylov subspace.   Similar to power iterations

for this to be an actual subspace,  $\tilde{b}, \tilde{A}\tilde{b}, \tilde{A}^2\tilde{b}$  must be L.I.. This is true when  $\tilde{A}$  is full rank, but may break down when  $\tilde{A}$  is not full rank

Constructing the subspace in this method is computationally unstable, because  $\tilde{A}^k\tilde{b}$  approaches the largest eigenvector of  $\tilde{A}$  as  $k \rightarrow \infty$ .

- Arnoldi Iteration : A procedure to construct a Krylov subspace.

Phase I:  $\tilde{A} = \tilde{Q} \tilde{H} \tilde{Q}^T$  ① // reduce  $\tilde{A}$  to upper Hessenberg form

$$\tilde{A} \tilde{Q} = \tilde{Q} \tilde{H} \quad ②$$

We need to break this into a iterative problem from a direct problem.







useful fashion.

Let  $\theta_i = \{\text{eigenvalues of } \tilde{A}_n\}$  || these are called Arnoldi eigenvalue estimators (at step  $n$ )

|| these are also called or Ritz values



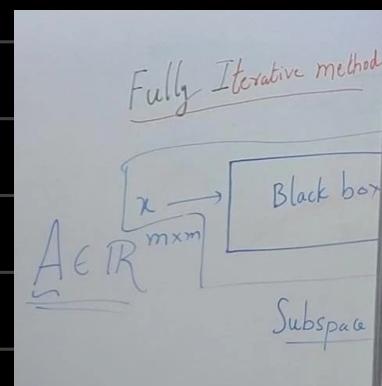
Arnoldi iterations can be viewed as a polynomial approximation

Let  $x \in K_n$ , where  $K_n = \langle b, \tilde{A}b, \dots, \tilde{A}^{n-1}b \rangle$ , then:

$$x = c_0 b + \dots + c_{n-1} \tilde{A}^{n-1} b$$

$$\text{Let } q(z) = c_0 + c_1 z + c_2 z^2 + \dots + c_{n-1} z^{n-1}.$$

$$\text{Hence } x = (q(A))b$$



Phase I

Arnoldi iteration

$$\tilde{A} = Q \tilde{H} Q^T \quad \text{--- ①}$$

$$\tilde{A}Q = Q \tilde{H} \quad \text{--- ②}$$

$$\tilde{A} \hat{Q}_n = \hat{Q}_{n+1} \tilde{H}_n \quad \boxed{\text{--- ③}}$$

$$\tilde{H}_n = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_m \\ 0 & h_{22} & \cdots & 0 \\ \vdots & \ddots & \ddots & h_{nn} \\ 0 & 0 & \cdots & h_{(n+1)(n+1)} \end{bmatrix}_{(n+1) \times n}$$

$$\hat{Q}_n = \begin{bmatrix} 1 & & & \\ q_1 & q_2 & \cdots & q_n \end{bmatrix}$$

$$\hat{Q}_{n+1} = \begin{bmatrix} 1 & & & \\ q_1 & q_2 & \cdots & q_{n+1} \end{bmatrix}$$

$$\begin{aligned} \tilde{A}q_1 &= h_{11}q_1 + h_{21}q_2 \\ \tilde{A}q_2 &= h_{12}q_1 + h_{22}q_2 + h_{32}q_3 \\ &\vdots \\ \tilde{A}q_n &= h_{1n}q_1 + h_{2n}q_2 + \cdots + h_{nn}q_n \end{aligned} \quad \text{--- ③}$$

Algo:-  
 $b$  = arbitrary  
 $q_1 = \frac{b}{\|b\|}$   
 for  $n = 1, 2, 3, \dots$   
 $v = Aq_n$   
 for  $j = 1 \dots n$   
 $h_{jn} = q_j^T v$   
 $v = v - h_{jn}q_j$   
 $h_{n+1,n} = \|v\|$   
 $q_{n+1} = \frac{v}{h_{n+1,n}}$

$$q_1 = \frac{b}{\|b\|} \quad K_n = \langle q_1, q_2, \dots, q_n \rangle = \langle b, Ab, \dots, A^{n-1}b \rangle$$

$$\begin{aligned} b &= \beta_1 q_1 + \beta_2 q_2 \\ &= \beta_1 \frac{b}{\|b\|} + \frac{\beta_2}{h_{21}} \left( Aq_1 - h_{11}q_1 \right) \\ &= \beta_1 \frac{b}{\|b\|} + \frac{\beta_2}{h_{21}} \left( \frac{Ab}{\|b\|} - h_{11} \frac{b}{\|b\|} \right) \\ &= b \underbrace{\left[ \frac{\beta_1}{\|b\|} - \frac{\beta_2 h_{11}}{h_{21} \|b\|} \right]}_{d_1} + Ab \underbrace{\left[ \frac{\beta_2}{h_{21}} \right]}_{d_2} \end{aligned}$$

Arnoldi approximation problem: Find  $p^n \in P^n$   
s.t.  $\|p^n(\tilde{A})b\|_2$  is minimum.

\*  $P^n$  is the set of all monic polynomials of degree  $n$ .

polynomials with coefficient of highest power = 1

Hence, we can restate the problem as:  $\min_{P^n \in P^n} \|p^n(\tilde{A})b\|_2$

The solution to this problem is actually  $P_{H_n}(z) = \det(zI - H_n)$

As we increase  $n$ , we shall essentially be trying to solve the characteristic polynomial of  $\tilde{A}$ .

$$q_1 = \frac{b}{\|b\|} \quad K_n = \langle q_1, q_2, \dots, q_n \rangle = \langle b, Ab, \dots, A^{n-1}b \rangle$$

$$\begin{aligned} b &= \beta_1 q_1 + \beta_2 q_2 \\ &= \beta_1 \frac{b}{\|b\|} + \frac{\beta_2}{h_{21}} \left( Aq_1 - h_{11}q_1 \right) \\ &= \beta_1 \frac{b}{\|b\|} + \frac{\beta_2}{h_{21}} \left( \frac{Ab}{\|b\|} - h_{11} \frac{b}{\|b\|} \right) \\ &= b \underbrace{\left[ \frac{\beta_1}{\|b\|} - \frac{\beta_2 h_{11}}{h_{21} \|b\|} \right]}_{d_1} + Ab \underbrace{\left[ \frac{\beta_2}{h_{21}} \right]}_{d_2} \end{aligned}$$

$$\begin{aligned} \tilde{A} \tilde{Q}_n &= \tilde{Q}_{n+1} \tilde{H}_n \\ \tilde{A}, \tilde{Q}_n &\in \langle q_1, q_2, \dots, q_n \rangle \\ \tilde{A}_p &= \tilde{Q}_n \tilde{Q}_n^T \tilde{A} \\ \tilde{A}_{p,j} &= q_j^T \tilde{A} q_j \\ \hat{A}_p &= \tilde{Q}_n^T (\tilde{Q}_n \tilde{Q}_n^T \tilde{A}) \tilde{Q}_n \\ &= \tilde{Q}_n^T \tilde{A} \tilde{Q}_n \end{aligned}$$

$$\begin{aligned} B_{ij} &= q_i^T B q_j \\ c_i^T B e_j \end{aligned}$$

Projection of  $\tilde{A}$  onto  $K_n = \langle q_1, \dots, q_n \rangle$

$$\hat{A}_p = \tilde{Q}_n^T (\tilde{A}) \tilde{Q}_n$$

$$\hat{A}_p = \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1n} \\ h_{21} & h_{22} & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ h_{nn} & h_{n1} & \dots & h_{nn} \end{bmatrix} = \boxed{H_n}$$

$$\begin{aligned} \tilde{A} \tilde{Q}_n &= \tilde{Q}_{n+1} \tilde{H}_n \\ \Rightarrow \tilde{Q}_n^T \tilde{A} \tilde{Q}_n &= \tilde{Q}_n^T \tilde{Q}_{n+1} \tilde{H}_n \\ &= \begin{bmatrix} 1 & & & 0 \\ 0 & \ddots & & 0 \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix}_{n \times (n+1)} \tilde{H}_n \end{aligned}$$

$$K_n = \langle q_1, \dots, q_n \rangle$$

$$\underline{A} \hat{\underline{Q}}_n = \hat{\underline{Q}}_{n+1} \tilde{H}_n$$

$$\Rightarrow \hat{\underline{Q}}_n^T \underline{A} \hat{\underline{Q}}_n = \hat{\underline{Q}}_n^T \hat{\underline{Q}}_{n+1} \tilde{H}_n \\ = \begin{bmatrix} 1 & & & \\ 0 & \ddots & & \\ & & 0 & \\ & & & \ddots & 1 & 0 \\ & & & & \ddots & 0 \end{bmatrix}_{n \times (n+1)} \tilde{H}_n$$

Since  $\tilde{H}_n (n \times n)$  is a projection of  $A$  onto  $K_n$ , there is a possibility that eigenvalues of  $\tilde{H}_n$  are related to eigenvalues of  $A$  in a useful fashion!

$\lambda_j = \{ \text{eigenvalues of } H_n \}$   
are called Arnoldi eigenvalue estimates (at step  $n$ ) or Ritz values

Arnoldi iteration  $\rightarrow$  Polynomial approximation!

$$x \in K_n, K_n = \langle b, Ab, \dots, A^{n-1}b \rangle$$

$$x = c_0 b + c_1 Ab + c_2 A^2 b + \dots + c_{n-1} A^{n-1} b$$

$$q(z) = c_0 + c_1 z + c_2 z^2 + \dots + c_{n-1} z^{n-1}$$

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

Algo:-

$b$  = arbitrary

$$q_1 = \frac{b}{\|b\|}$$

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

$$v = Aq_n$$

for  $j = 1 \dots n$

$$h_{jn} = q_j^T v$$

$$v = v - h_{jn} q_j$$

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

$$q_{n+1} = \frac{v}{h_{n+1,n}}$$

Arnoldi approximation problem:-

Find  $p^n \in P^n$  such that  $\underbrace{\|p^n(A)b\|_2}$   
is minimum!

$P^n$  denotes the space of monic  
polynomials of degree  $n$

$$\min_{p^n \in P^n} \|p^n(A)b\|_2 \quad \rightarrow p_{H^n}(z) = \boxed{\det(zI - H_n)}$$

$\rightarrow$  Krylov subspace  $K_n = \{b, Ab, A^2b, \dots, A^{n-1}b\}$

4/11/2025

\* Assume  $A$  is real & symmetric

$\rightarrow$  Is this really a subspace?

$$\tilde{b} = \sum_{i=1}^m \beta_i \tilde{q}_i \quad \text{eigenvectors of } A$$

$$\text{Let } |\lambda_1| > |\lambda_2| \dots > |\lambda_n| > \dots > |\lambda_m|$$

$$\text{Hence } \underline{\underline{A}} \underline{\underline{b}} = \sum_{i=1}^m \beta_i \lambda_i \underline{\underline{q}_i},$$

$$\underline{\underline{A}}^k \underline{\underline{b}} = \sum_{i=1}^m \beta_i \lambda_i^k \underline{\underline{q}_i}$$

When  $\beta_i \neq 0 \forall i$ , and all eigenvalues are distinct,  $\underline{\underline{b}}, \underline{\underline{A}} \underline{\underline{b}}, \dots$  will be linearly independent.

→ vector in Krylov subspace let  $x \in \mathbb{K}^n$

$$x = c_0 \underline{\underline{b}} + c_1 \underline{\underline{A}} \underline{\underline{b}} + \dots + c_{n-1} \underline{\underline{A}}^{n-1} \underline{\underline{b}}$$

$$= q(\underline{\underline{A}}) \cdot \underline{\underline{b}}$$

$$= q(\underline{\underline{A}}) \cdot \sum_{i=1}^n \beta_i \underline{\underline{q}_i}$$

$$= q(\underline{\underline{A}}) \cdot \underline{\underline{Q}} \cdot \underline{\underline{\beta}}$$

$$= \underline{\underline{Q}} \begin{bmatrix} q(\lambda_1) \cdot \beta_1 \\ \vdots \\ q(\lambda_{n-1}) \cdot \beta_{n-1} \end{bmatrix}$$

$$\text{As written above, } \underline{\underline{\beta}}_{A^n} = \arg \min_{P^m \in P^n} \| p^n(\underline{\underline{A}}) \underline{\underline{b}} \|$$

$$\underline{\underline{Q}}^{\wedge} \begin{bmatrix} \underline{\underline{v}_1} & \cdots & \underline{\underline{v}_m} \end{bmatrix}_{mn} = \underline{\underline{B}}_{m \times n}$$

project back from Krylov subspace to  $\mathbb{R}^m$

→ Cayley Hamilton Theorem: A matrix satisfies its own characteristic polynomial:

$$\det(\tilde{A} - \lambda \tilde{I}) = p(\lambda), \text{ where } p(\lambda) \text{ is a monic polynomial}$$

Because  $\det(\tilde{A} - \lambda \tilde{I}) = 0, p(\lambda) = 0$

Hence  $p(\tilde{A}) = 0$

$$\therefore \tilde{A}^m + c_{m-1} \tilde{A}^{m-1} + \dots + c_0 \tilde{I} = 0$$

$c_0$  cannot be 0 as  
 $\tilde{A}$  cannot be singular,  
and  $c_0$  is the  
determinant of  $\tilde{A}$

How can we solve  $\tilde{A}\tilde{x} = \tilde{b}$  using this?

$$\tilde{A}^m \tilde{b} + c_{m-1} \tilde{A}^{m-1} \tilde{b} + \dots + c_0 \tilde{b} = 0$$

$$\tilde{A}^{m-1} \tilde{b} + c_{m-1} \tilde{A}^{m-2} \tilde{b} + \dots + c_1 \tilde{b} = -c_0 \tilde{A}^{-1} \tilde{b}$$

[Left-multiply  $\tilde{A}^{-1}$ ]

$$\text{Hence } \tilde{x}^* = \tilde{A}^{-1} \tilde{b} = -\frac{1}{c_0} (c_1 \tilde{b} + c_2 \tilde{A} \tilde{b} + \dots + \tilde{A}^{m-1} \tilde{b})$$

Therefore, the Krylov subspace is a good subspace to search for soln. of  $\tilde{A}\tilde{x} = \tilde{b}$ .

We can approximate  $\tilde{x}^*$  to lie in Krylov subspace of dimension  $n-m$ :

GMRes Method

→ generalised minimal residual method (iterative)

→ GMRes tries to solve the problem  $\min_{\underline{z} \in K_n} \|\underline{A}\underline{z} - \underline{b}\|_2$ .

\* Find a vector  $\underline{z}$  in the Krylov subspace s.t. the residual is minimal.

→ Let trial  $\underline{z} = \underline{Q}_n^{\wedge} \underline{y}$

Hence, we now have to minimize:  $\arg \min_{\underline{x} \in \mathbb{R}^n} \|\underline{A} \underline{Q}_n^{\wedge} \underline{x} - \underline{b}\|_2$

$$= \arg \min_{\underline{y} \in \mathbb{R}^n} \|\underline{Q}_{n+1}^{\wedge} \underline{H}_n^{\wedge} \underline{y} - \underline{b}\|_2$$

$$\left[ \underline{A} \underline{Q}_n^{\wedge} + \underline{Q}_{n+1}^{\wedge} \underline{H}_n^{\wedge} \right]$$

$$= \arg \min_{\underline{y} \in \mathbb{R}^n} \|(\underline{Q}_{n+1}^{\wedge})^T (\underline{Q}_{n+1}^{\wedge} \underline{H}_n^{\wedge} \underline{y} - \underline{b})\|_2$$

$$= \arg \min_{\underline{y} \in \mathbb{R}^n} \|\underline{H}_n^{\wedge} \underline{y} - \underline{Q}_{n+1}^{\wedge} \underline{b}\|_2$$

$$= \arg \min_{\underline{y} \in \mathbb{R}^n} \|\underline{H}_n^{\wedge} \underline{y} - \|\underline{b}\| \underline{e}_1\|_2$$

This results in  $\{(n+1) \times n\}$  least square problem. This essentially means that at each step  $n$  of GMRes we solve the least square problem of  $(n+1) \times n$  and  $\underline{q}_n = \underline{Q}_n^{\wedge} \underline{y}$ .

$$\text{Alg}o: \text{ Let } \underline{\underline{y}}_1 = \frac{\underline{\underline{b}}}{\|\underline{\underline{b}}\|}$$

for  $n=1 \rightarrow \infty$

$\leftarrow$  step n of Arnoldi iteration >

$$\underline{\underline{y}} = \underset{\underline{\underline{y}} \in \mathbb{R}^n}{\operatorname{argmin}} \| \underline{\underline{H}}_n^T \underline{\underline{y}} - \|\underline{\underline{b}}\| \underline{\underline{e}}_1 \|_2$$

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

$\rightarrow$  Polynomial approximations of GMRes:

Let  $P_n = \{ \text{polynomials } p \text{ of degree } \leq n \text{ s.t. } p(0) = 1 \}$

$$\text{Let } \underline{\underline{x}}_n \in K_n \Rightarrow \underline{\underline{x}}_n = c_0 \underline{\underline{b}} + \cdots + c_{n-1} \underline{\underline{A}}^{n-1} \underline{\underline{b}}$$

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

$$\therefore \underline{\underline{x}}_n = q_n(A) \cdot \underline{\underline{b}}$$

$$\text{Residual, } \underline{\underline{r}}_n = \underline{\underline{b}} - \underline{\underline{A}} \underline{\underline{x}}_n$$

$$= \underline{\underline{b}} - \underline{\underline{A}} q_n(A) \underline{\underline{b}}$$

$$= \left( \underline{\underline{I}} - \underline{\underline{A}} q_n(A) \right) \underline{\underline{b}}$$

If  $p_n$  is the polynomial,  $p_n(z) = 1 - z q_n(z)$  GMRes approximation problem. Find

$p_n \in P^n$  s.t.  $\|p_n(A)b\|$  is minimised

$\curvearrowleft$  need not be monic polynomial

## Convergence of GMRES

If  $r_n \rightarrow 0$  as  $n \rightarrow \infty$ ,  
GMRES converges.

What properties of  $A$  determine the size of  $\|r_n\|$ ?

→ GMRES converges monotonically, i.e.  $\|r_{n+1}\| > \|r_n\|$ ,  $K_n \subseteq K_{n+1}$

→ At most  $m$  steps are reqd. for the process to converge  
 \* This leads to a T.C. of  $O(m^2)$  →  $m$  iterations,  $m^2$  ops. per iteration  
 This is not improvement over the direct (Gaussian elimination) method.

→ For this method to be better, we need  $n$  to be much smaller than  $m$ .

$$\text{i.e. } \|r_n\|_2 = \|p_n(\tilde{A})b\|_2 \leq \|p_n(\tilde{A})\| \cdot \|b\|$$

$$\frac{\|r_n\|}{\|b\|} \leq \|p_n(\tilde{A})\| \Rightarrow \frac{\|r_n\|}{\|b\|} \leq \min_{p_n \in P^n} \|p_n(\tilde{A})\|$$

Let us assume  $\tilde{A}$  to be diagonalisable:  $\tilde{A} = \tilde{V} \tilde{\Delta} \tilde{V}^{-1}$ , for some non-singular matrix  $\tilde{V}$ .

$$\text{Hence, } \|p_n(\tilde{A})\| \leq \|\tilde{V}\| \|p_n(\tilde{\Delta})\| \|\tilde{V}^{-1}\|$$

$$\leq \kappa(\tilde{V}) \|p_n(\tilde{\Delta})\|$$

$$\text{cond. no. of } \tilde{V} = \|\tilde{V}\| \|\tilde{V}^{-1}\|$$

$$\|p_n(A)\| \leq \kappa(V) \|p_n(\Delta)\|$$

$$\leq \kappa(\underline{A}) \cdot \max_{\lambda \in \sigma(\underline{A})} |p_n(\lambda)| \quad (\text{Vining-Carey-Hamilton theorem})$$

At step  $n$  of GMRes, the residual  $\underline{r}_n$  satisfies:

$$\frac{\|\underline{r}_n\|}{\|b\|} \leq \min_{p_n \in P^n} \|p_n(\underline{A})\| \leq \kappa(\underline{A}) \cdot \min_{p_n \in P^n} \left( \max_{\lambda \in \sigma(\underline{A})} |p_n(\lambda)| \right)$$

 measure of how far we are from normality

$$\text{If } \underline{A} \text{ is S.P.D., } \|\underline{r}\| \leq \left( \frac{(\kappa(\underline{A}))^2 - 1}{(\kappa(\underline{A}))^2} \right)^{1/2} \|\underline{r}_0\|$$

### Conjugate Gradient Method

→ Another iterative linear system solver using Krylov subspace.

Assumption:  $\underline{A} \in \mathbb{R}^{n \times n}$ ,  $\underline{A}$  is S.P.D.

→ Let  $K_n = \langle b, \dots, \underline{A}^{n-1}b \rangle$  // Krylov subspace

→  $\|\cdot\|_{\underline{A}}$  is defined:  $\|\underline{x}\|_{\underline{A}} := \sqrt{\underline{x}^T \underline{A} \underline{x}}$  //  $\underline{A}$ -norm

→ Conjugate gradient is a set of recurrence formulae that generates a sequence of iterations  $\{\underline{x}_n \in K_n\}$  with the property that at step  $n$ :  $\|\underline{e}_n\|_{\underline{A}}$  is minimised.

$$*\underline{e}_n = \underline{x}^* - \underline{x}_n \quad \text{actual solution}$$

→ Algo: initialize  $\underline{x}_0 = \underline{0}$  // initial 'guess'  
 $\underline{r}_0 = \underline{b}$  // initial residual  
 $\underline{p}_0 = \underline{r}_0$  // \* direction

for  $n=1 \rightarrow \infty$

$$\alpha_n = \frac{\underline{r}_{n-1}^T \underline{r}_{n-1}}{\underline{p}_{n-1}^T A \underline{p}_{n-1}} \quad // \text{step length}$$

$$\underline{x}_n = \underline{x}_{n-1} + \alpha_n \cdot \underline{p}_{n-1} \quad // \text{update approx. soln.}$$

$$\underline{r}_n = \underline{r}_{n-1} - \alpha_n \cdot A \cdot \underline{p}_{n-1} \quad // \text{update residual}$$

$$\beta_n = \frac{\underline{r}_n^T \underline{r}_n}{\underline{r}_{n-1}^T \underline{r}_{n-1}} \quad // \text{ratio of inner prod. of } \underline{r}_n \text{ & } \underline{r}_{n-1}, \text{ represents improvement in search direction}$$

$$\underline{p}_n = \underline{r}_n + \beta_n \underline{p}_{n-1} \quad // \text{get new search direction}$$

Consequences of iterative procedure:

→  $\langle \underline{x}_1, \dots, \underline{x}_n \rangle$  is same as  $\langle \underline{p}_0, \dots, \underline{p}_{n-1} \rangle$  subspace spanned by direction vectors  
 →  $\langle \underline{r}_0, \dots, \underline{r}_{n-1} \rangle$  subspace spanned by residuals  
 →  $\langle \underline{b}, A\underline{b}, \dots, A^{n-1}\underline{b} \rangle$  Krylov subspace

The sequence of iterates also satisfy:  $\underline{r}_n^T \underline{r}_0 = 0 \quad \forall j < n$

The search directions are A-conjugate:  $\underline{p}_n^T A \underline{p}_n = 0 \quad \forall j < n$

(Orthogonal w.r.t to the  $\underline{A}$  direction)

Result : let conjugate gradient iterations be applied to  $\underline{A}$ , where  $\underline{A}$  is an S.P.D., to solve  $\underline{A}\underline{x} = \underline{b}$ . If the iterations are not fully converged, then  $\underline{x}_n$  is the unique point in  $K_n$  that minimises  $\|\underline{e}_n\|_{\underline{A}}$ .

The convergence is monotonic :  $\|\underline{e}_n\|_{\underline{A}} \leq \|\underline{e}_{n+1}\|_{\underline{A}}$

Proof : We need to show that  $\underline{x}_n$  is the unique point that minimises  $\underline{e}_n = \underline{x}^* - \underline{x}_n$ .

$\therefore \underline{x}_n \in K_n$  (Krylov subspace).

Consider an arbitrary point  $\underline{x} \in K_n$ .

We can write  $\underline{x}$  as  $\underline{x} = \underline{x}_n + \Delta \underline{x}$ ,  $\Delta \underline{x} \neq \underline{0}$

If  $\underline{x}$  &  $\underline{x}_n$  are in  $K_n$ ,  $\Delta \underline{x}$  must also be in  $K_n$ .

Let  $\underline{e} = \underline{x}^* - \underline{x}$

$$\begin{aligned} &= \underline{x}^* - (\underline{x}_n + \Delta \underline{x}) \\ &= (\underline{x}^* - \underline{x}_n) - \Delta \underline{x} \\ &= \underline{e}_n - \Delta \underline{x} \end{aligned}$$

$$\text{Hence, } \|\underline{e}\|_{\underline{A}}^2 = (\underline{e}_n - \Delta \underline{x})^T \underline{A} (\underline{e}_n - \Delta \underline{x})$$

$$= \underline{e}_n^T \underline{A} \underline{e}_n + \Delta \underline{x}^T \underline{A} \Delta \underline{x} - 2 \underline{e}_n^T \underline{A} \Delta \underline{x}$$

$$\underline{e_n}^T \underline{A} \underline{\Delta x} = (\underline{x}^* - \underline{x}_n)^T \underline{A} \underline{\Delta x}$$

$$= (\underline{x}^*)^T \underline{A} \underline{\Delta x} - \underline{x}_n^T \underline{A} \underline{\Delta x}$$

$$= \underline{b}^T \underline{\Delta x} - (\underline{b} - \underline{r}_n)^T \underline{A} \underline{\Delta x} \quad \| \underline{x}_n = \underline{b} - \underline{r}_n$$

$$= \underline{b}^T \underline{\Delta x} - \cancel{\underline{b}^T \underline{\Delta x}} + \underline{r}_n^T \underline{\Delta x}$$

$$= \underline{A_n}^T \underline{\Delta x}$$

0

$$\text{Therefore, } \|\underline{e}\|_A^2 = \underline{e_n}^T \underline{A} \underline{e_n} + \underline{\Delta x}^T \underline{A} \underline{\Delta x} - \cancel{2 \underline{e_n}^T \underline{A} \underline{\Delta x}}$$

$$\|\underline{e}\|_A^2 = \underline{e_n}^T \underline{A} \underline{e_n} + \underline{\Delta x}^T \underline{A} \underline{\Delta x}$$

We have to show that  $\underline{x}_n$  minimises  $\|\underline{e}\|_A^2$ .

Because  $\underline{A}$  is s.p.p,  $\|\underline{e}\|_A^2$  is minimum when  $(\underline{\Delta x}^T \underline{A} \underline{\Delta x}) = 0$ . This happens only when  $\underline{\Delta x} = 0$ , i.e.  $\underline{x} = \underline{x}_n$ .

Conjugate gradient vs an optimisation problem

$$\|\underline{e_n}\|_A^2 = \underline{e_n}^T \underline{A} \underline{e_n}$$

$$= (\underline{x}^* - \underline{x}_n)^T \underline{A} (\underline{x}^* - \underline{x}_n)$$

$$= \underline{x}_n^T \underline{A} \underline{x}_n - 2 \underline{x}_n^T \underline{A} \underline{x}^* + (\underline{x}^*)^T \underline{A} \underline{x}^*$$

$$= \underline{x}_n^T \underline{A} \underline{x}_n - 2 \underline{x}_n^T \underline{b} + (\underline{x}^*)^T \underline{A} \underline{x}^*$$

