# NLA Short Notes

→ Angle bet. two vectors $= \cos^{-1}\left[\dfrac{(x, y)}{\|x\|_2 \|y\|_2}\right]$

## Norms

→ Properties of a valid norm :

→ Should be real & non-negative

→ $\|\underset{\sim}{x}\| = 0 \Leftrightarrow \underset{\sim}{x} = \underset{\sim}{0}$

→ $\|\alpha \underset{\sim}{x}\| = \alpha \|\underset{\sim}{x}\|$

→ $\|\underset{\sim}{x} + \underset{\sim}{y}\| \leq \|\underset{\sim}{x}\| + \|\underset{\sim}{y}\|$

→ $|\underset{\sim}{x}^T \underset{\sim}{y}| \leq \|\underset{\sim}{x}\|_2 \|\underset{\sim}{y}\|_2$ (Cauchy-Shwarz Inequality)

→ $\|\underset{\sim}{A}\|_1 = \max\limits_{1 \leq i \leq m} (\|\underset{\sim}{a_j}\|_1)$

→ $\|\underset{\sim}{A}\|_\infty =$ max. row absolute sum

→ If $\underset{\sim}{A} = \underset{\sim}{u}\underset{\sim}{v}^T$, $\|A\|_2 = \|u\|_2 \|v\|_2$ (when $\underset{\sim}{x} = \dfrac{\underset{\sim}{v}}{\|\underset{\sim}{v}\|}$)

→ $\|\underset{\sim}{A}\underset{\sim}{B}\|_p \leq \|\underset{\sim}{A}\|_p \|\underset{\sim}{B}\|_p$

→ $\|\underset{\sim}{A}\|_F = \sqrt{\sum\limits_{i=0}^{m}\sum\limits_{j=0}^{n} |a_{ij}|^2} = \sqrt{tr(\underset{\sim}{A}\underset{\sim}{A}^T)} = \sqrt{tr(\underset{\sim}{A}^T\underset{\sim}{A})} = \sqrt{\sum\limits_{i=0}^{m} \|\underset{\sim}{a_i}\|^2}$

# Conditioning & Stability

→ Absolute cond. no, $\hat{\kappa} = \frac{\|\delta f\|}{\|\delta x\|} = \frac{\|J(x)\,\delta x\|}{\|\delta x\|} = \|J(x)\|$

→ Relative cond. no. $= \frac{\|\delta f\|}{\|f\|} \div \frac{\|\delta x\|}{\|x\|} = \frac{\|J(x)\|\,\|x\|}{\|f\|}$

→ $\hat{\kappa}(\underset{\sim}{A}) = \frac{\sigma_1}{\sigma_m} = \sqrt{\frac{\lambda_{max}(\underset{\sim}{A}^T\underset{\sim}{A})}{\lambda_{min}(A^TA)}}$

→ $\hat{\kappa}^R(\underset{\sim}{A}) = \|\underset{\sim}{A}\|\,\|\underset{\sim}{A}^+\|$, where $\|\underset{\sim}{A}^+\| = (\underset{\sim}{A}^T\underset{\sim}{A})^{-1}\underset{\sim}{A}^T$

→ Backwards stable $\Rightarrow$ Stable $\Rightarrow$ Forward stable
$\quad\quad\quad\quad\quad\quad\quad \not\Leftarrow \quad\quad\quad\quad \not\Leftarrow$

→ For a backward stable algo, $f(\tilde{x}) = \tilde{f}(x)$

Backward stability:
$\tilde{f}(x) = f(x + \delta x)$

Forward stability:
$\|\tilde{f}(x) - f(x)\| = O(\epsilon_m)$

# SVD

$\underset{\sim}{A} = \underset{\sim}{U}\,\underset{\sim}{\Sigma}\,\underset{\sim}{V}^T$, $\quad \underset{\sim}{U} \in \mathbb{R}^{m \times m}$, orthogonal

$\underset{\sim}{\Sigma} \in \mathbb{R}^{m \times n}$, diagonal

$\underset{\sim}{V}^T \in \mathbb{R}^{n \times n}$, orthogonal

→ $\|A\|_2 = \sigma_1$, $\|A\|_F = \sqrt{\sigma_1^2 + \dots \sigma_n^2}$

# Principal Component Analysis

→ Orthogonal transformation of the feature space.

→ Consider a data matrix $\underset{\sim}{A} \in \mathbb{R}^{m \times n}$ no. of experiments particular measurement

1) Move the matrix 0-centered by subtracting the mean of the coln. from each element in the coln.

2) Variance of the coln. $= \dfrac{E(X^2) - \cancel{(E(X))^2}}{m} = \dfrac{\|\underset{\sim}{a_j}\|_2^2}{m}$ 0

3) Hence, total variance, $T = \|a_1\|_2^2 + \cdots + \|a_n\|^2$

$= \|A\|_F^2$

Biggest contributor to the variance

$= \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2$

→ To get the direction with highest variance, we need to find a vector $\underset{\sim}{\hat{w}_1}$ s.t:

$\underset{\sim}{t_1} = \underset{\sim}{A}\,\underset{\sim}{\hat{w}_1}$, $\|t_1\|_2$ is maximised. The solution is $\underset{\sim}{\hat{w}_1} = \underset{\sim}{v_1}$.

$\underset{\sim}{t_1} = \underset{\sim}{A}\,\underset{\sim}{v_1} = \sum_{i=1}^{n} \underset{\sim}{u_i}\,\sigma_i\,\underset{\sim}{v_i^T}\,\underset{\sim}{v_1}$ $\left[ v_i^T v_1 = \begin{cases} 0, & i \neq 1 \\ 1, & i = 1 \end{cases} \right]$

$= u_1 \sigma_1 \underbrace{\cancel{v_1^T v_1}}_{1}$

$= \sigma_1 \underset{\sim}{u_1}$

→ Similarly, $\underset{\sim}{t_2} = \sigma_2 \underset{\sim}{u_2}$, where $\underset{\sim}{t_2}$ is the direction along second highest variance.

## Projection

→ Projector matrix $\underset{\sim}{P}$, must be:

→ $\underset{\sim}{P}^2 = \underset{\sim}{P}$ (idempotent) → $\underset{\sim}{P}\underset{\sim}{v} - \underset{\sim}{v} \in Null(\underset{\sim}{P})$

→ Square matrix → Rank-deficient

→ $\underset{\sim}{P}\underset{\sim}{x} = \underset{\sim}{x}$ $\forall$ $\underset{\sim}{x} \in Range(\underset{\sim}{P})$ → symmetrical

→ Orthogonal projector also fulfils $P = P^T$ (symmetric)

→ If $P$ is a projector, so is $(I-P)$. This the complementary projector of $P$.

→ Let us have $Ax = b$ has no soln., we can solve $A\hat{x} = Pb$ instead. $P$ projects $b$ onto the coln. space of $A$, which means this eqn. solvable.

Hence, $P = A(A^TA)^{-1}A^T$

→ If $P$ is an orthogonal projector, so is $(I-P)$

→ If $A$ is orthogonal, $P = A\overset{I}{\cancel{(A^TA)}}^{-1}A^T = AA^T$

→ Projection & SVD: If $A = U\Sigma V^T$, $U$ forms a orthonormal basis of $A$. Hence, the projection matrix corresponding to $A$, $P = Q\overset{I}{\cancel{(Q^TQ)}}^{-1}Q^T = QQ^T$

## QR Factorisation

→ $Q$: matrix w/ orthonormal colns. → $R$: UTM

→ Gram-Schmidt orthogonalisation:

$$v_j = a_j - (q_1^T a_j)q_1 - \cdots - (q_{j-1}^T a_j)q_{j-1}$$

$$q_j = \frac{v_j}{\|v_j\|_2}$$

Hence $q_n = \dfrac{a_n - \sum_{i=0}^{n-1} r_{in} q_i}{r_{nn}}$, where $r_{ij} = \begin{cases} q_i^T a_j, & i \neq j \\ \|a_j - \sum_{i=1}^{j-1} r_{ij} q_i\|_2, & i = j \end{cases}$

→ Modified Gram Schmidt

→ Gram Schmidt with projector: $q_n = \dfrac{P_n a_n}{\|P_n a_n\|}$, where

$P_n = I - \sum_{i=0}^{n-1} q_i q_i^T = I - \hat{Q}_{j-1} \hat{Q}_{j-1}^T$

→ Modified Gram-Schmidt:

$$P_n a_n = \left(I - \sum_{i=0}^{n-1} q_i q_i^T\right) a_n = \left[\prod_{i=0}^{n-1} (I - q_i q_i^T)\right] a_n$$

Algo: for $j = 1 \to n$:

$v_j^{(1)} = a_j$

$v_j^{(2)} = P_{\perp q_1} a_j = (I - q_1 q_1^T) a_j = v_j^{(1)} - q_1 q_1^T v_j^{(1)}$

$v_j^{(3)} = P_{\perp q_2} a_j = v_j^{(2)} - q_2 q_2^T v_j^{(2)}$

$\vdots$

$v_j^{(j)} = v_j^{(j-1)} - q_{(j-1)} q_{(j-1)}^T v_j^{(j-1)}$

$q_j = \dfrac{v_j^{(j)}}{\|v_j^{(j)}\|_2}$

→ Householder Triangularisation

→ Perform transformations to convert $A$ to $R$:

$Q_n \cdots Q_1 A = R$

ensure, that prev colns. are not changed.

must be an orthogonal matrix

$$Q_k = \begin{bmatrix} I_{(k-1)\times(k-1)} & 0 \\ 0 & \hat{F}_{(m-k+1)(m-k+1)} \end{bmatrix}_{m \times m}$$

→ Let $\hat{x} = \begin{bmatrix} - \\ - \\ - \\ \vdots \\ - \end{bmatrix}$, $F$ should be s.t. $Fx = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

$= \|x\| \cdot e_1$

Hence, $y = Fx = (I - 2uu^T)x$, where $u = -\frac{v}{\|v\|}$,

where $v = \|x\|_2 e_1 - x$

Algo: for $k = 1 \to n$:

$x = A(k:m, k)$ // row $k$ to row $m$ in the $k^{th}$ coln.

$v_k = \text{sgn}(x_1) \cdot \|x\|_2 e_1 + x$

$v_k = \frac{v_k}{\|v_k\|}$

$A(k:m, k:n) -= 2 v_k v_k^T A(k:m, k:n)$

| Algo | FLOPs | Stability | Error |
|---|---|---|---|
| CGS | $2mn^2$ | Unstable | $O(\kappa(A)^2 \cdot \varepsilon_m)$ |
| MGS | " | Backward stable | $O(\kappa(A) \cdot \varepsilon_m)$ |
| Householder | $2mn^2 - \frac{2}{3}n^3$ | " | $O(\varepsilon_m)$ |

## Symmetric +ve Definite Matrices (SPD)

→ Properties

→ $A = A^T$

→ $(x, Ay) = (y, Ax) \quad \forall x, y \in \mathbb{R}^m$

→ $x^T A x > 0, \quad \forall x \in \mathbb{R}^m$

→ If $A$ is S.P.D. & $X \in \mathbb{R}^{m \times n}$ is full rank, then $X^T A X$ is also S.P.D.

→ All eigenvalues are +ve for any S.P.D.

# Cholesky Decomposition

→ If $\underset{\sim}{A}$ is S.P.D., $\underset{\sim}{A}$ can be decomposed s.t. $\underset{\sim}{A} = \underset{\sim}{R}^T \underset{\sim}{R}$ ← VTM

→ $n^3/3$ FLOPs

# Linear Least Squares

→ If $\underset{\sim}{A}\underset{\sim}{x} = \underset{\sim}{b}$ is overdetermined, we can solve for $\boxed{\underset{\sim}{A}^T \underset{\sim}{A} \underset{\sim}{\hat{x}} = \underset{\sim}{A}^T \underset{\sim}{b}}$ ← normal eqn. instead. This shall minimize the residual, i.e $\|\underset{\sim}{r}\|_2$, where $r = \underset{\sim}{A}\underset{\sim}{\hat{x}} - \underset{\sim}{b}$

→ $A\hat{x} = Pb$ // get $b$ to coln. space of $A$

$\underset{\sim}{A}\underset{\sim}{\hat{x}} = \underset{\sim}{A}(\underset{\sim}{A}^T\underset{\sim}{A})^{-1}\underset{\sim}{A}^T\underset{\sim}{b} \Rightarrow \underset{\sim}{\hat{x}} = (\underset{\sim}{A}^T\underset{\sim}{A})^{-1}\underset{\sim}{A}^T\underset{\sim}{b}$

→ Soln. by Cholesky decomposition:

$\underset{\sim}{A}^T\underset{\sim}{A}\underset{\sim}{\hat{x}} = \underset{\sim}{A}^T\underset{\sim}{b}$

$\underset{\sim}{R}^T\underset{\sim}{R}\underset{\sim}{\hat{x}} = \underset{\sim}{A}^T\underset{\sim}{b}$

$\underset{\sim}{R}^T\underset{\sim}{w} = \underset{\sim}{A}^T\underset{\sim}{b}$ // Let $\underset{\sim}{w} = \underset{\sim}{R}\underset{\sim}{\hat{x}}$

↓ solve for $\underset{\sim}{w}$

$\underset{\sim}{R}\underset{\sim}{\hat{x}} = \underset{\sim}{w}$

↓ solve for $\underset{\sim}{x}$

$\underset{\sim}{\hat{x}}$

→ Soln. by QR factorisation

$\underset{\sim}{A}\underset{\sim}{x} = \underset{\sim}{P}\underset{\sim}{b}$

$\underset{\sim}{Q}\underset{\sim}{R}\underset{\sim}{x} = \underset{\sim}{Q}\underset{\sim}{Q}^T\underset{\sim}{b}$

$\underset{\sim}{R}\underset{\sim}{x} = \underset{\sim}{Q}^T\underset{\sim}{b}$

↓ Solve for $\underset{\sim}{x}$

$\underset{\sim}{x}$

→ Soln with SVD

$\underset{\sim}{A}\underset{\sim}{x} = \underset{\sim}{P}\underset{\sim}{b}$

$\underset{\sim}{U}\underset{\sim}{\Sigma}\underset{\sim}{V}^T\underset{\sim}{x} = \underset{\sim}{U}\underset{\sim}{U}^T\underset{\sim}{b}$

$\underset{\sim}{\Sigma}\underset{\sim}{V}^T\underset{\sim}{x} = \underset{\sim}{U}^T\underset{\sim}{b}$

$\underset{\sim}{\Sigma}\underset{\sim}{y} = U^T b$ // Let $\underset{\sim}{V}^T\underset{\sim}{x} = \underset{\sim}{y}$

↓ Solve for $\underset{\sim}{y}$

$\underset{\sim}{V}^T\underset{\sim}{x} = \underset{\sim}{y}$

↓ Solve for $\underset{\sim}{x}$

$\underset{\sim}{x}$

| Algo | Work |
|---|---|
| Cholesky | $mn^2 + n^3/3$ |
| QR (Householder) | $2mn^2 - \frac{2n^3}{3}$ |
| SVD | $2mn^2 + 11n^3$ |

$\boxed{\text{Either } m \geq n \text{ \& } \operatorname{rank}(A) < n \text{, or } m < n}$

→ If $A$ is close to rank-deficient,

$$Ax = b$$

$$U \Sigma V^T x = b$$

$\boxed{U^T U = I \text{, whereas } U U^T \text{ will be a projector matrix}}$

$$\cancel{U^T U}^{\,I} \Sigma V^T x = U^T b$$

$$V^T x = \Sigma^{-1} U^T b$$

Let $x = V_1 y + V_2 z$, where $V = \begin{bmatrix} \underset{\text{range}(A)}{V_1} & \underset{\text{Null}(A)}{V_2} \end{bmatrix}$

Hence $y = \Sigma^{-1} U^T b$

Therefore $x = V_1 \Sigma^{-1} U^T b + V_2 z$

$$\|x\|_2 = \|V_1 \Sigma^{-1} U^T b\|_2 + \|V_2 z\|_2 - \cancel{\|(V_1 \Sigma^{-1} U^T b, V_2 z)\|_2} \; 0$$

inner product

To minimise $\|x\|_2$, we set $z = 0$

→ If $b \notin \text{Range}(A)$, then we solve $Ax = Pb$.

In this case, $x = V_1 \Sigma^{-1} U^T P b + V_2 z$ $\quad \boxed{P = UU^T}$

$$= V_1 \Sigma^{-1} U^T b + V_2 z$$

# Eigen Decomposition

→ $\underset{\sim}{A} = \underset{\sim}{X}\,\underset{\sim}{\Lambda}\,\underset{\sim}{X}^{-1}$, $\underset{\sim}{\Lambda}$ is a diagonal matrix containing eigenvalues and $\underset{\sim}{X}$ is a matrix comprising of the respective eigenvectors

→ Characteristic polynomial $(P_A)$ : $P_A(z) = \det(\underset{\sim}{A} - z\underset{\sim}{I})$

→ Eigenvectors are essentially Nullspace of $(\underset{\sim}{A} - \lambda\underset{\sim}{I})$. Hence, $(\underset{\sim}{A} - \lambda\underset{\sim}{I})$ must be a rank-deficient (singular) matrix, with $\det(\underset{\sim}{A} - \lambda\underset{\sim}{I}) = 0$

→ Geometric multiplicity of $\lambda$: No. of L.I. eigenvectors associated with an eigenvalue $\lambda$.

→ If $\lambda$ corresponds to two eigenvectors $\underset{\sim}{v_1}$ & $\underset{\sim}{v_2}$, any vector $\underset{\sim}{x} = \alpha_1\underset{\sim}{v_1} + \alpha_2\underset{\sim}{v_2}$ will be an eigenvector of A:

$$\text{Pf}: \underset{\sim}{A}(\alpha_1\underset{\sim}{v_1} + \alpha_2\underset{\sim}{v_2}) = \alpha_1\underset{\sim}{A}\underset{\sim}{v_1} + \alpha_2\underset{\sim}{A}\underset{\sim}{v_2} = \alpha_1\lambda\underset{\sim}{v_1} + \alpha_2\lambda\underset{\sim}{v_2}$$
$$= \lambda(\alpha_1\underset{\sim}{v_1} + \alpha_2\underset{\sim}{v_2})$$
$$= \lambda\underset{\sim}{x} \quad \text{(shown)}$$

# Similarity Transformation

→ If $\underset{\sim}{X} \in \mathbb{R}^{m\times m}$ is non-singular, $\underset{\sim}{X}\underset{\sim}{A}\underset{\sim}{X}^{-1}$ is known as similarity transformation of $\underset{\sim}{A}$.

→ Two matrices $\underset{\sim}{A}$ & $\underset{\sim}{B}$ are said to be similarity if there exists a similarity transformation bet. them, i.e $\underset{\sim}{B} = \underset{\sim}{X}\underset{\sim}{A}\underset{\sim}{X}^{-1}$.

→ $\underset{\sim}{A}$ & $\underset{\sim}{B}$ will have same eigenvalues & respective multiplicities

$$P_B(z) = \det(z\underset{\sim}{I} - \underset{\sim}{X}\underset{\sim}{A}\underset{\sim}{X}^{-1})$$
$$= \det(z\underset{\sim}{X}\underset{\sim}{X}^{-1} - \underset{\sim}{X}\underset{\sim}{A}\underset{\sim}{X}^{-1})$$
$$= \det(\underset{\sim}{X}(z\underset{\sim}{I} - \underset{\sim}{A})\underset{\sim}{X}^{-1})$$
$$= \cancel{\det(\underset{\sim}{X})}\det(z\underset{\sim}{I} - \underset{\sim}{A})\cancel{\det(\underset{\sim}{X}^{-1})} \quad \text{(shown)}$$

$\boxed{\text{* Eigenvectors may not be the same for } \underset{\sim}{A} \text{ \& } \underset{\sim}{B}}$

# Defective Eigenvalues & Matrices

→ An eigenvalue, for which algebraic multiplicity > geometric multiplicity is a defective eigenvalue.
→ Any matrix that has a defective eigenvalue is a defective matrix
→ It does not possess a full set of L.I. eigenvectors.

→ Diagonal matrices are not defective

→ Diagonalisability: If $\underset{\sim}{A} \in \mathbb{R}^{m \times n}$ is not defective iff it has eigenvalue decomposition.

Or orthogonal diagonalisability

not all non-defective matrices are unitary diagonalisable

→ Unitary diagonalisability: If a non-defective matrix $\underset{\sim}{A}$ has eigenvalue decomposition $A = \underset{\sim}{Q}\,\underset{\sim}{\Delta}\,\underset{\sim}{Q}^{-1} = \underset{\sim}{Q}\,\underset{\sim}{\Delta}\,\underset{\sim}{Q}^{\perp}$, where $\underset{\sim}{Q}$ is a $\underbrace{\text{unitary matrix}}$.

← tranjugate

$\underset{\sim}{Q} \in \mathbb{C}^{m \times m}$ $\quad \underset{\sim}{Q}^{\perp}\underset{\sim}{Q} = \underset{\sim}{Q}\,\underset{\sim}{Q}^{\perp} = \underset{\sim}{I}$

→ Symmetric matrices have all real eigenvalues & eigenvectors.
→ Skew symmetric matrices have all imaginary eigenvalues. Skew symmetric matrices are also unitary diagonalisable.

## Defective Matrices

→ Defective eigenvalue: An eigenvalue with geometric multiplicity < algebraic multiplicity. If a matrix has ≥ 1 defective eigenvalue, it is defective.

## Unitary / Orthogonal Diagonalisability

← tranjugate

→ $Q \in \mathbb{C}^{m \times m}$ is a unitary matrix if $\underset{\sim}{Q}^{\perp}\underset{\sim}{Q} = \underset{\sim}{Q}\,\underset{\sim}{Q}^{\perp} = \underset{\sim}{I}$
→ $\underset{\sim}{A}$ is unitary diagonalisable iff $\exists\, \underset{\sim}{Q}$ s.t. $\underset{\sim}{A} = \underset{\sim}{Q}^{\perp}\,\underset{\sim}{\Delta}\,\underset{\sim}{Q}$, $\underset{\sim}{Q}$ is unitary

# Symmetric Matrices

→ A real symmetric matrix is non-defective & unitary diagonalisable, with real eigenvalues.

→ A real skew-symmetric matrix is also non-defective & unitary diagonalisable, with purely complex eigenvalues.

→ Any normal matrix $\underset{\sim}{A} \in \mathbb{C}^{m \times m}$, s.t. $\underset{\sim}{A}^{\perp}\underset{\sim}{A} = \underset{\sim}{A}\underset{\sim}{A}^{\perp}$, will be unitary diagonalisable.

# Schur Factorisation

→ $\underset{\sim}{A} = \underset{\sim}{Q}\,\underset{\sim}{T}\,\underset{\sim}{Q}^{\perp}$, where $\underset{\sim}{Q}$ is unitary, and $\underset{\sim}{T}$ is U.T.M.

→ SVD of $\underset{\sim}{T}$ = SVD of $\underset{\sim}{A}$.

→ Every sq. matrix has a Schur factorisation.

* If $\underset{\sim}{A}$ is real, $\underset{\sim}{A}$ can be decomposed to $\underset{\sim}{U}\,\underset{\sim}{T}\,\underset{\sim}{U}^{T}$, where $\underset{\sim}{U}$ & $\underset{\sim}{T}$ are real, and $\underset{\sim}{T}$ is quasi U.T.M.

→ Schur factorisation need not be unique.

# Eigen solvers

→ Phase 1: Reduce $\underset{\sim}{A}$ to upper Hessenberg matrix $\underset{\sim}{H}$. (UTM but with additional line of non-zero elements parallel to diagonal). $O(m^3)$ flops

→ Phase 2: Reduce $\underset{\sim}{H}$ to U.T.M. $O(m)$ iterations, $O(m^2)$ flops per iteration $\Rightarrow O(m^3)$ flops.

* Without phase 1, we would need $O(m^4)$ flops.

→ Phase 1: Reduce $\underset{\sim}{A}$ to $\underset{\sim}{H}$ as follows: $\underset{\sim}{A} = \underset{\sim}{Q}\,\underset{\sim}{H}\,\underset{\sim}{Q}^{\perp}$, where

$$\underset{\sim}{Q} = \underset{\sim}{Q}_1 \cdots \underset{\sim}{Q}_{m-2}.$$

→ Rayleigh quotient: $\alpha = \dfrac{\underset{\sim}{x}^T \underset{\sim}{A}\,\underset{\sim}{x}}{\underset{\sim}{x}^T \underset{\sim}{x}}$

→ $\alpha$ will be the eigenvalue of $\underset{\sim}{A}$ closest to $\underset{\sim}{x}$. This is the least sq. soln. that minimises $\|\underset{\sim}{A}\underset{\sim}{x} - \alpha\underset{\sim}{x}\|_2$.

# Power Iterations

→ Finds the eigenvector corresponding to largest eigenvector (by magnitude).

Algo: Initialise $\underset{\sim}{v}^{(0)}$ to a random unit vector

for $k = 1 \to \infty$,

$$\underset{\sim}{w} = \underset{\sim}{A}\, \underset{\sim}{v}^{(k-1)}$$

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

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

Results:

$$\| \underset{\sim}{v}^{(k)} - (\pm \underset{\sim}{q}_1) \|_2 = O\left( \left| \frac{\lambda_2}{\lambda_1} \right|^k \right)$$

$$|\lambda^{(k)} - \lambda_1| = O\left( \left| \frac{\lambda_2}{\lambda_1} \right|^{2k} \right)$$

If $k$ is even, $\underset{\sim}{v}^{(k)} \to \underset{\sim}{q}_1$, otherwise, $\underset{\sim}{v}^{(k)} \to -\underset{\sim}{q}_1$

* Convergence is slow if $\lambda_2 \simeq \lambda_1$

# Inverse Power Iterations

Algo: Initialise $\mu$ = some value near $\lambda_J$,
$v^{(0)}$ = " random unit vector

for $k = 1 \to \infty$:

$$\underset{\sim}{w} = (A - \mu I)^{-1} v^{(k-1)}$$ // Solve by turning soln to system of lin eqns.

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

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

Converges to closest eigenvalue of $\mu$. If $\lambda_J$ is the closest to $\mu$ and $\lambda_K$ is second-most closest to $\mu$,

$$\| v^{(k)} - (\pm q_J) \| = O\left( \left| \frac{\mu - \lambda_J}{\mu - \lambda_K} \right|^k \right)$$

$$|\lambda^{(k)} - \lambda_J| = O\left( \left| \frac{\mu - \lambda_J}{\mu - \lambda_K} \right|^{2k} \right)$$

# Rayleigh Quotient Iteration

Algo: Initialize $\underset{\sim}{v}^{(0)}$ to some random unit vector

$$\lambda^{(0)} = (\underset{\sim}{v}^{(0)})^T \underset{\sim}{A} (\underset{\sim}{v}^{(0)})$$

for $k = 1 \rightarrow \infty$ :

$$\underset{\sim}{w} = (\underset{\sim}{A} - \lambda^{(k-1)} \underset{\sim}{I})^{-1} \underset{\sim}{v}^{(k-1)}$$ // again linear system of eqns.

$$\underset{\sim}{q}^{(k)} = \frac{\underset{\sim}{w}}{\|\underset{\sim}{w}\|}$$

$$\lambda^{(k)} = (\underset{\sim}{q}^{(k)})^T \underset{\sim}{A} (\underset{\sim}{q}^{(k)})$$

→ Very fast convergence:

$$\| \underset{\sim}{v}^{(k+1)} - (\pm \underset{\sim}{q}_i) \| = O\left( \| \underset{\sim}{v}^{(k)} - (\pm \underset{\sim}{q}_i) \|^3 \right)$$

$$|\lambda^{(k+1)} - \lambda_i| = O\left( |\lambda^{(k)} - \lambda_i|^3 \right)$$

## Analysis of Algos. (per iteration)

→ Power iteration: $O(m^2)$ due to matrix-vector multiplication.

→ Inverse power iteration: $O(m^3)$ due to soln. of linear system of eqns.

→ Can be reduced to $O(m^2)$ by solving $(\underset{\sim}{A} - \mu \underset{\sim}{I})^{-1}$ once

→ Rayleigh quotient iteration: $O(m^3)$, but lesser iterations are reqd.

→ Can be reduced to $O(m^2)$ by reducing $\underset{\sim}{A}$ to tridiagonal / upper Hessenberg.

## Multiple Eigen values

→ Subspace / simultaneous iterations

→ Take multiple vectors which are L.I.. Provided we have an $\infty$ precision computer, these will converge to different eigen vectors.

→ Assumption #1: The first n eigenvalues are distinct & well-separated

→ " #2: If $\underset{\sim}{Q_1} = [q_1 \cdots q_n]$, where $\{q_1 \cdots q_n\}$ are eigenvectors of $\underset{\sim}{A}$, $\underset{\sim}{Q_1}^T \underset{\sim}{v}^{(0)}$ is non-singular, and all principal submatrices of $\underset{\sim}{Q_1}^T \underset{\sim}{v}^{(0)}$ are also singular.

→ Orthogonalise at each step to prevent loss of orthogonality, and hence make the algo. stable.

→ Works for large, sparse matrices

Algo: Initialise $\underset{\sim}{\hat{Q}}^{(0)} \in \mathbb{R}^{m \times n}$

$$\text{for } k = 1 \to \infty:$$
$$\underset{\sim}{Z}^{(k)} = \underset{\sim}{A} \underset{\sim}{\hat{Q}}^{(k-1)}$$
$$\underset{\sim}{\hat{Q}}^{(k)}, \underset{\sim}{\hat{R}}^{(k)} = \underset{\sim}{Z} \quad \text{// QR factorisation}$$

→ Pure QR algorithm (dense matrices)

Algo: $\underset{\sim}{A}^{(0)} = \underset{\sim}{A}$

$$\text{for } k = 1 \to \infty$$
$$\underset{\sim}{Q}^{(k)}, \underset{\sim}{R}^{(k)} = \underset{\sim}{A}^{(k-1)} \quad \text{// orthogonalise}$$
$$\underset{\sim}{A}^{(k)} = \underset{\sim}{R}^{(k)} \underset{\sim}{Q}^{(k)}$$

→ As $k \to \infty$, $A^{(k)}$ approaches Schur form.

→ Mathematically equivalent to simultaneous iterations

→ Can also be considered a simultaneous inverse iteration applied to a flipped identity matrix.

→ Modified QR (most used by eigenvalues)

Full Algo: Define $\underset{\sim}{A}^{(0)}$ s.t $(\underset{\sim}{Q}^{(0)})^T \underset{\sim}{A}^{(0)} (\underset{\sim}{Q}^{(0)}) = \underset{\sim}{A}$ // tridiagonalisation of $\underset{\sim}{A}$

for $k = 1 \to \infty$:

Pick a shift $\mu^{(k)}$ // many methods for picking, e.g. $\mu^{(k)} = A^{(k-1)}_{mm}$

$Q^{(k)} R^{(k)} = \underset{\sim}{A}^{(k-1)} - \mu^{(k)} \underset{\sim}{I}$ // shifted QR factorisation

$A^{(k)} = \underset{\sim}{R}^{(k)} \underset{\sim}{Q}^{(k)} + \mu^{(k)} \underset{\sim}{I}$

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

Split $\underset{\sim}{A}^{(k)}$ into $\underset{\sim}{A}_1$ & $\underset{\sim}{A}_2$ s.t $A^{(k)} = \begin{bmatrix} \underset{\sim}{A}_1 & 0 \\ 0 & \underset{\sim}{A}_2 \end{bmatrix}$

Apply QR algo (from tridiagonalisation) on $\underset{\sim}{A}_1$ & $\underset{\sim}{A}_2$.

→ Krylov subspace method (fully iterative):

→ Krylov subspace is a subspace rich in eigenvectors. This is the set of vectors $\{\underset{\wedge}{b}, A\underset{\wedge}{b}, A^2\underset{\wedge}{b}, \dots\}$.

→ This seq. is similar to power iteration

→ For this to be an actual subspace $\underset{\wedge}{b}, A\underset{\wedge}{b}, A^2\underset{\wedge}{b}$ etc. must be L.I. They are confirmed to be L.I. if $\underset{\wedge}{A}$ is full-rank.

→ This method is computationally unstable

→ Arnoldi Iteration (To construct Krylov subspace)

Algo: $\underset{\sim}{b}$ = arbitrary vector

$\underset{\sim}{q}_1 = \dfrac{\underset{\sim}{b}}{\|\underset{\sim}{b}\|}$

for $n = 1 \rightarrow \infty$

$\underset{\sim}{v} = \underset{\sim}{A}\,\underset{\sim}{q}_1$

for $j = 1 \rightarrow n$

$h_{jn} = \underset{\sim}{q}_j^T \underset{\sim}{v}$

$\underset{\sim}{v} = \underset{\sim}{v} - h_{jn}\,\underset{\sim}{q}$

$h_{(n+1)n} = \|\underset{\sim}{v}\|$

$q_{n+1} = \dfrac{\underset{\sim}{v}}{h_{(n+1)n}}$

→ At the end of the iterations, we have:

→ constructed a subspace rich in eigenvalues of $\underset{\sim}{A}$

→ Projected $\underset{\sim}{A}$ onto the subspace, to obtain $\underset{\sim}{H}_n$

→ Hence, $\underset{\sim}{H}_n$ is a projection of $\underset{\sim}{A}$ onto $K_n$

→ Eigenvalues of $\underset{\sim}{H}_n$ are Arnoldi eigenvalue estimates, a.k.a. Ritz values

→ Arnoldi iterations can be viewed as polynomial approximation

→ Arnoldi approximation problem: Find $p^n \in P^n$ s.t. $\|p^n(A)\,\underset{\sim}{b}\|_2$ is minimum. $P^n$ is the set of monic polynomials of deg. $n$.

→ Soln. to this problem is actually $P_{\underset{\sim}{H}_n}(z) = \det(z\underset{\sim}{I} - \underset{\sim}{H}_n)$

→ As $n \rightarrow \infty$, the soln. approaches eigenvalues of $\underset{\sim}{A}$.