5. Orthogonal Matrices and Subspaces

- Orthogonal vectors x and y

$$\left.\begin{array}{c}(\text{test})\\ \mathbf{x}^T\mathbf{y}=0\\ \overline{\mathbf{x}}^T\mathbf{y}=0\end{array}\right\}\rightarrow\begin{cases}\text{Pytagoras Law of right triangles: } \|\mathbf{x}-\mathbf{y}\|^2=\|\mathbf{x}\|^2+\|\mathbf{y}\|^2\\ \text{Law of cosines: } \|\mathbf{x}-\mathbf{y}\|^2=\|\mathbf{x}\|^2+\|\mathbf{y}\|^2-2\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta\end{cases}$$

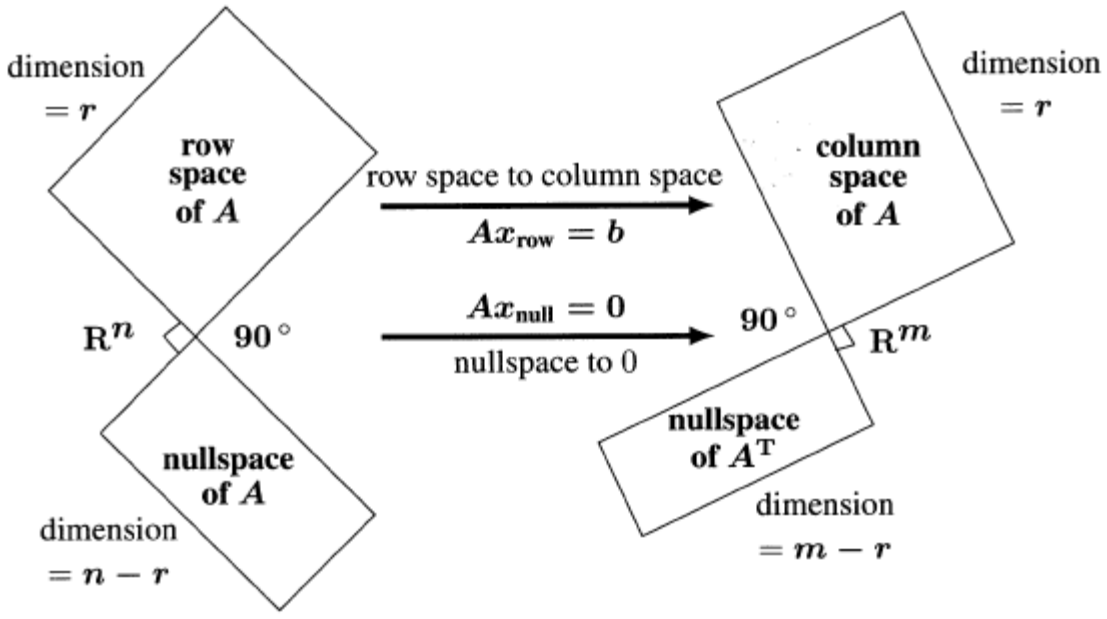
- Orthogonal basis for a subspace

$$\text{orthogonal: } v_i^T v_j=0\xrightarrow{v_i/\|v_i\|}\text{orthonormal: } v_i^T v_i=1$$

  - Standard basis is orthogonal (even orthonormal) in $\mathbf{R}^n$ (*i*, *j*, *k* in $\mathbf{R}^3$)
  - Hadamard matrices H_n containing orthogonal bases of $\mathbf{R}^n$
    - Are those orthogonal matrices?
  - Every subspace of $\mathbf{R}^n$ has an orthogonal basis: Gram-Schmidt idea
    - Two independent vectors a and b in the plane: $a^Tc=0$

- Orthogonal subspace R (row space) and N (null space)
  - Ax=0: The row space of A is orthogonal to the nullspace of A
  - A^Ty=0: The column space of A is orthogonal to the nullspace of A^T

$$\begin{pmatrix} \mathbf{Ax} = \mathbf{0} \text{ means} \\ \text{each row} \cdot \mathbf{x} = \mathbf{0} \end{pmatrix} \longrightarrow \begin{bmatrix} \text{row } 1 \\ \vdots \\ \text{row } m \end{bmatrix} [\mathbf{x}] = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{A}^T\mathbf{y} = \begin{bmatrix} (\text{column } 1)^T \\ \vdots \\ (\text{column } n)^T \end{bmatrix} [\mathbf{y}] = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$



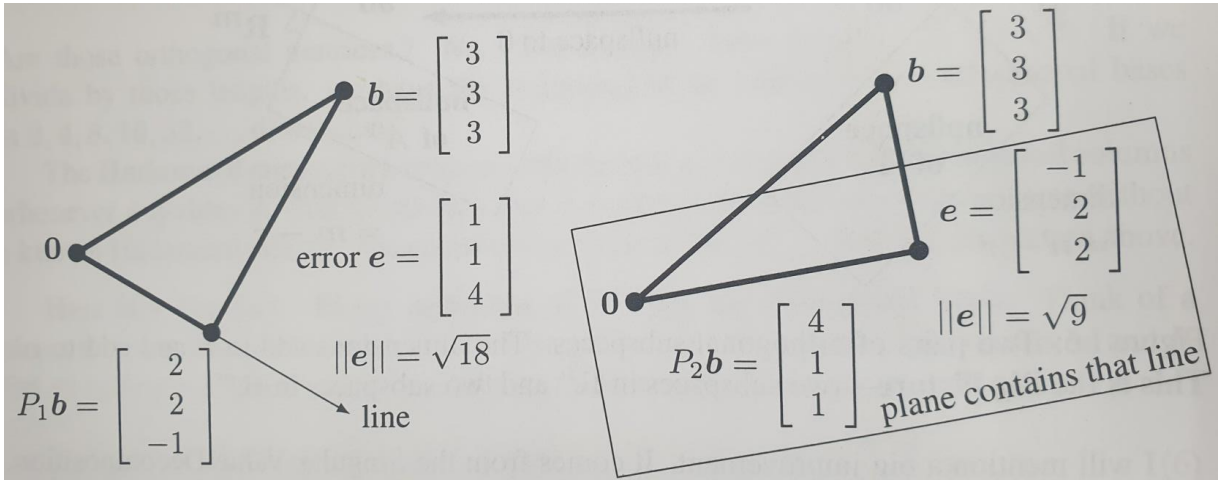
- Tall thin matrices Q with orthonormal columns: $Q^TQ=I$

$$\begin{cases} \text{if } \mathbf{Q} \text{ multiplies any vector } \mathbf{x}, \text{ the length of the vector does not change: } \|\mathbf{Qx}\| = \|\mathbf{x}\| \\ \text{if } m > n \text{ mthe } m \text{ rows cannot be orthogonal in } \mathbf{R}^n\text{: } \mathbf{QQ}^T \neq \mathbf{I} \end{cases}$$

$$\mathbf{Q}_1 = \frac{1}{3}\begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}, \mathbf{Q}_2 = \frac{1}{3}\begin{bmatrix} 2 & 2 \\ 2 & -1 \\ -1 & 2 \end{bmatrix}, \mathbf{Q}_3 = \frac{1}{3}\begin{bmatrix} 2 & 2 & -1 \\ 2 & -1 & 2 \\ -1 & 2 & 2 \end{bmatrix} \rightarrow \mathbf{Q}_i\mathbf{Q}_i^T = \mathbf{I}\,?$$

$\mathbf{P} = \mathbf{QQ}^T \rightarrow$ projection matrix: $\mathbf{P}^2 = \mathbf{P} = \mathbf{P}^T \xrightarrow{\text{"least squares"}}$

Pb is the orthogonal projection of **b** onto the column space of **P**

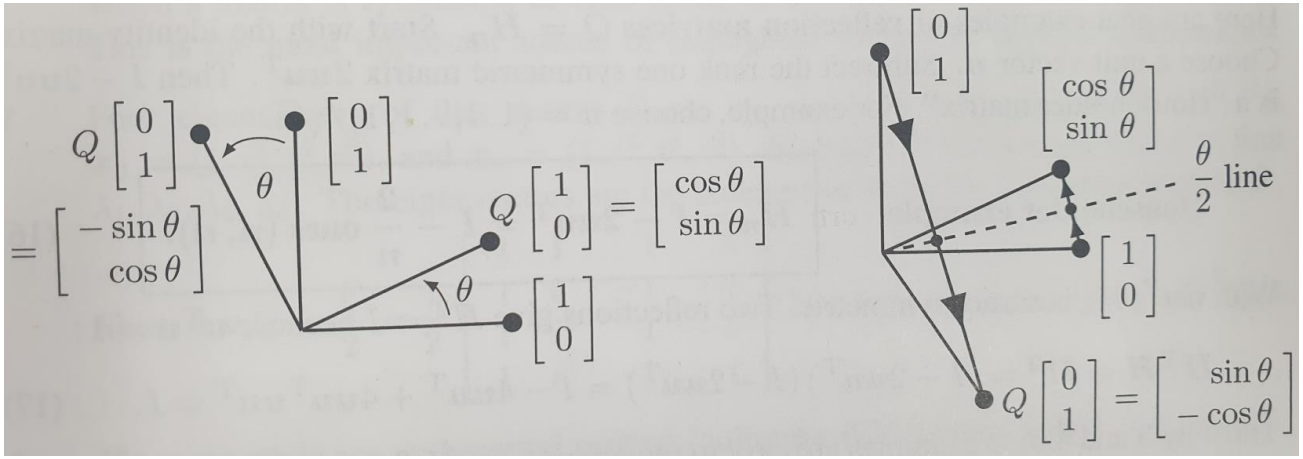


- Orthogonal matrices are square with orthonormal columns: $Q^T=Q^{-1}$

$$\mathbf{Q} \text{ is square} \rightarrow \left\{ \begin{matrix} \mathbf{Q}^T\mathbf{Q} = \mathbf{I} \\ \mathbf{Q}\mathbf{Q}^T = \mathbf{I} \end{matrix} \right\} \rightarrow \mathbf{Q}^{-1} = \mathbf{Q}^T$$

$$\underbrace{\mathbf{Q}_{\text{rotate}} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}}_{\text{rotation through an angle } \theta}, \quad \underbrace{\mathbf{Q}_{\text{reflect}} = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}}_{\text{reflection across the } \frac{\theta}{2} \text{ line}}$$

$$\mathbf{Q}_1, \mathbf{Q}_2 : \text{ orthogonal} \rightarrow \mathbf{Q}_1\mathbf{Q}_2 : \text{ orthogonal}$$



Orthogonal basis = orthogonal axes in $\mathbf{R}^n$

orthogonal $\mathbf{Q}(n\times n)$: $\mathbf{v} = c_1\mathbf{q}_1 + \cdots + c_n\mathbf{q}_n \rightarrow c_i = \mathbf{q}_i{}^T\mathbf{v}$

$\mathbf{v} = \mathbf{Qc} \rightarrow \mathbf{Q}^T\mathbf{v} = \mathbf{Q}^T\mathbf{Qc} = \mathbf{c}$

Householer reflections: $\mathbf{Q} = \mathbf{H}_n = \mathbf{I} - 2\mathbf{u}\mathbf{u}^T$

$\mathbf{u} = (1,1,\ldots,1)/\sqrt{n} \rightarrow \mathbf{Q} = \mathbf{H}_n = \mathbf{I} - \frac{2}{n}\mathbf{ones}(n,n) \xrightarrow{\mathbf{u}\perp\mathbf{w}} \begin{cases} \mathbf{H}_n\mathbf{u} = -\mathbf{u} \\ \mathbf{H}_n\mathbf{w} = +\mathbf{w} \end{cases}$

$\mathbf{H}^T\mathbf{H} = \mathbf{H}^2 = \mathbf{I}$

$\mathbf{H}_3 = \mathbf{I} - \frac{2}{3}\mathbf{ones} = \frac{1}{3}\begin{bmatrix} 1 & -2 & -2 \\ -2 & 1 & -2 \\ -2 & -2 & 1 \end{bmatrix}$, $\mathbf{H}_4 = \mathbf{I} - \frac{2}{4}\mathbf{ones} = \frac{1}{2}\begin{bmatrix} 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 \\ -1 & -1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{bmatrix}$

eigenvalues of $\mathbf{H}_n$ are -1 (once) and $+1$ (n-1 times)

All reflection matrices have eigenvalues -1 and 1

Examples

- Rotations
- Reflections
- Hadamard matrices
- Haar wavelets
- Discrete Fourier Transform (DFT)
- Complex inner product

6. Eigenvalues and Eigenvectors

eigenvectors of $\mathbf{A}$ don't change direction when you multiply them by $\mathbf{A}$

$$\left.\begin{array}{l}\mathbf{x}:\ \text{eigenvector of } \mathbf{A}\\ \boldsymbol{\lambda}:\ \text{eigenvector of } \mathbf{A}\end{array}\right\} \rightarrow \mathbf{A}\mathbf{x}=\boldsymbol{\lambda}\mathbf{x} \rightarrow \mathbf{A}\left(\mathbf{A}\mathbf{x}\right)=\mathbf{A}\left(\boldsymbol{\lambda}\mathbf{x}\right)=\boldsymbol{\lambda}\left(\mathbf{A}\mathbf{x}\right)=\boldsymbol{\lambda}^2\mathbf{x} \rightarrow \mathbf{A}^k\mathbf{x}=\boldsymbol{\lambda}^k\mathbf{x}$$

$n\times n$ matrices $\rightarrow$ n independent eigenvectors $\mathbf{x}_1$ to $\mathbf{x}_n$ with n different eigenvalues $\boldsymbol{\lambda}_1$ to $\boldsymbol{\lambda}_n$

$$\mathbf{v}=c_1\mathbf{x}_1+\cdots+c_n\mathbf{x}_n \rightarrow \mathbf{A}\mathbf{v}=c_1\lambda_1\mathbf{x}_1+\cdots+c_n\lambda_n\mathbf{x}_n \rightarrow \mathbf{A}^k\mathbf{v}=c_1{\lambda_1}^k\mathbf{x}_1+\cdots+c_n{\lambda_n}^k\mathbf{x}_n$$

How useful?
- solution of differential equations
- similar matrices $\rightarrow$ same eigenvalues
- diagonalize a matrix

Four properties: matrix $\mathbf{A}\left(\text{real}\right)$, $\underbrace{\mathbf{S}\left(\text{symmetric}\right)}_{\text{like real numbers: } \lambda}$, $\underbrace{\mathbf{Q}\left(\text{orthogonal}\right)}_{\text{like complex numbers: } e^{i\theta}}$

$\left(\text{Trace of } \mathbf{S}\right)$ $\sum_{i=1}^{n}\lambda_i=$ trace of matrix

$\left(\text{Determinant}\right)$ $\prod\lambda_i=$ determinant of matrix

$\left(\text{Real eigenvalues of } \mathbf{S}\right)$ $\mathbf{S}:$ real eigenvalues, orthogonal eigenvectors

$\left(\text{Orthogonal eigenvectors}\right)$ if $\lambda_1\neq\lambda_3$, then $\mathbf{x}_1\cdot\mathbf{x}_2=0$, eigenvectors of $\mathbf{A}$ are orthogonal iff $\mathbf{A}^T\mathbf{A}=\mathbf{A}\mathbf{A}^T$

$$\mathbf{S}=\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix} \rightarrow \left\{\mathbf{S}\begin{bmatrix}1\\1\end{bmatrix}=3\begin{bmatrix}1\\1\end{bmatrix} \text{ and } \mathbf{S}\begin{bmatrix}1\\-1\end{bmatrix}=1\begin{bmatrix}1\\-1\end{bmatrix}\right., \quad \mathbf{Q}=\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix} \rightarrow \left\{\mathbf{Q}\begin{bmatrix}1\\-i\end{bmatrix}=i\begin{bmatrix}1\\-i\end{bmatrix} \text{ and } \mathbf{Q}\begin{bmatrix}1\\i\end{bmatrix}=-i\begin{bmatrix}1\\i\end{bmatrix}\right.$$

$\mathbf{eig}\left(\mathbf{A}+\mathbf{B}\right)\neq\mathbf{eig}\left(\mathbf{A}\right)+\mathbf{eig}\left(\mathbf{B}\right)$

$\mathbf{eig}\left(\mathbf{A}\mathbf{B}\right)\neq\mathbf{eig}\left(\mathbf{A}\right)\mathbf{eig}\left(\mathbf{B}\right)$

$\lambda_1=\lambda_2$ might or might not have two independent eigenvectors

(1) **A** controls a system of linear differential equations: $\frac{d\mathbf{u}}{dt} = \mathbf{A}\mathbf{u}$ with $\mathbf{u}(0)$

$\mathbf{u}(0) = c_1\mathbf{x}_1 + \cdots + c_n\mathbf{x}_n$

$$\mathbf{u}(t) = c_1 e^{\lambda_1 t}\mathbf{x}_1 + \cdots + c_n e^{\lambda_n t}\mathbf{x}_n \xrightarrow{\lambda = a + ibt} \begin{cases} e^{at} = \begin{cases} \operatorname{Re}\lambda > 0: \text{ grow} \\ \operatorname{Re}\lambda < 0: \text{ decay} \end{cases} \\ e^{ibt} = \cos bt + i \sin bt: \text{ oscillate} \end{cases}$$

shift in $\mathbf{A} \rightarrow$ shift in λ: $(\mathbf{A} + \mathbf{sI})\mathbf{x} = \boldsymbol{\lambda}\mathbf{x} + \mathbf{s}\mathbf{x} = (\boldsymbol{\lambda} + \mathbf{s})\mathbf{x}$

(2) **B** similar to $\mathbf{A} \rightarrow \mathbf{B} = \underbrace{\mathbf{M}}_{\text{invertible}} \mathbf{A}\mathbf{M}^{-1} \rightarrow eig(\mathbf{B}) = eig(\mathbf{A})$: compute eigenvalues of large matrices

Make **B** gradually into a triangular matrix $\rightarrow$ Gradually show up on the main diagonal

$\mathbf{B}\mathbf{y} = \boldsymbol{\lambda}\mathbf{y} \rightarrow \mathbf{M}\mathbf{A}\mathbf{M}^{-1}\mathbf{y} = \boldsymbol{\lambda}\mathbf{y} \rightarrow \mathbf{A}(\mathbf{M}^{-1}\mathbf{y}) = \boldsymbol{\lambda}(\mathbf{M}^{-1}\mathbf{y})$

(3) diagonalize a matrix

$$\mathbf{A}\begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} \mathbf{A}\mathbf{x}_1 & \cdots & \mathbf{A}\mathbf{x}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{x}_1 & \cdots & \lambda_n\mathbf{x}_n \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_n \end{bmatrix}\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \rightarrow \begin{cases} \mathbf{AX} = \mathbf{X\Lambda} \\ \mathbf{A} = \mathbf{X\Lambda X}^{-1} \\ \mathbf{A}^2 = \mathbf{X\Lambda}^2\mathbf{X}^{-1} \end{cases}$$

$$\mathbf{A}^k\mathbf{v} = \mathbf{X\Lambda}^k\mathbf{X}^{-1}\mathbf{v}: \ \mathbf{v} = \mathbf{Xc} \rightarrow \underbrace{\mathbf{c} = \mathbf{X}^{-1}\mathbf{v}}_{c_i} \rightarrow \underbrace{\Lambda^k\mathbf{X}^{-1}\mathbf{v}}_{c_i\lambda_i^k} \rightarrow \underbrace{\mathbf{X}\Lambda\mathbf{X}^{-1}\mathbf{v}}_{\sum c_i\lambda_i^k x_i}$$

Example: $\mathbf{A} = \begin{bmatrix} 8 & 3 \\ 2 & 7 \end{bmatrix} \xrightarrow{\text{divide by 10}} \underbrace{\mathbf{A} = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix}}_{\substack{\text{Markov matrix with} \\ \text{positive columns adding to 1}}} \rightarrow \begin{cases} \mathbf{A}^k\mathbf{v} = c_1(1)^k\,\mathbf{x}_1 + c_1\left(\frac{1}{2}\right)^k\mathbf{x}_2 \\ \text{as } k \text{ increases } \mathbf{A}^k\mathbf{v} \text{ approaches to } c_1\mathbf{x}_1 \end{cases}$

the action of the whole matrix **A** is broken into simple actions (just muliply by λ)

[nondiagonalizable matrices: when GM < AM, **A** is not diagonalizable]

$$\begin{cases} (\text{Geometric Multiplicity} = \text{GM}): \text{count the independent eigenvectors, } \dim N(\mathbf{A} - \boldsymbol{\lambda}\mathbf{I}) \\ (\text{Algebraic Multiplicity} = \text{AM}): \text{count the repetitions of eigenvalues, } \det(\mathbf{A} - \boldsymbol{\lambda}\mathbf{I}) = 0 \end{cases}$$

$$\mathbf{A} = \begin{bmatrix} 5 & 1 \\ 0 & 5 \end{bmatrix}, \begin{bmatrix} 6 & -1 \\ 1 & 4 \end{bmatrix}, \begin{bmatrix} 7 & 2 \\ -2 & 3 \end{bmatrix}$$

8. Singular Value Decomposition (SVD)

best matrices (real symmetric matrices **S**): real eigenvalues and orthogonal eigenvectors

other matrices (**A** is not square, $m \times n$): complex eigenvalues and not orthogonal eigenvectors

key point: two sets of singular vectors $\begin{cases} n \text{ right singular vectors } (\mathbf{v}_1, \ldots, \mathbf{v}_n) \text{ orthogonal in } \mathbf{R}^n \\ m \text{ left singular vectors } (\mathbf{u}_1, \ldots, \mathbf{u}_m) \text{ orthogonal in } \mathbf{R}^m \end{cases}$

connection between n **v**'s and m **u**'s

$$\underbrace{\mathbf{A}\mathbf{v}_1 = \sigma_1\mathbf{u}_1, \ldots, \mathbf{A}\mathbf{v}_r = \sigma_r\mathbf{u}_r}_{\substack{r=\text{rank}(\mathbf{A}) \\ \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0}}, \underbrace{\mathbf{A}\mathbf{v}_{r+1} = \mathbf{0}, \ldots, \mathbf{A}\mathbf{v}_n = \mathbf{0}}_{\substack{(n-r) \text{ } \mathbf{v}\text{'s in } N(\mathbf{A}) \\ (m-r) \text{ } \mathbf{u}\text{'s in } N(\mathbf{A}^T)}}$$

$$\mathbf{A}\underbrace{\begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_r & \cdots & \mathbf{v}_n \end{bmatrix}}_{\text{square orthogonal unit vector: } \mathbf{V}^T = \mathbf{V}^{-1}} = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_r & \cdots & \mathbf{u}_m \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & 0 \end{bmatrix} \rightarrow \begin{cases} \mathbf{AV} = \mathbf{U}\boldsymbol{\Sigma} \\ \mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T \\ = \sigma_1\mathbf{u}_1\mathbf{v}_1^T + \cdots + \sigma_r\mathbf{u}_r\mathbf{v}_r^T \\ (\mathbf{A} \rightarrow r \text{ pieces of rank 1}) \end{cases}$$

$$\mathbf{AV} = \mathbf{U}\boldsymbol{\Sigma} \xrightarrow{\text{reduced form}} \mathbf{AV}_r = \mathbf{U}_r\boldsymbol{\Sigma}_r \rightarrow \mathbf{A}\underbrace{\begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_r \end{bmatrix}}_{\text{row space}} = \underbrace{\begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_r \end{bmatrix}}_{\text{column space}} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix}$$

Proof of SVD

$$\mathbf{AX} = \mathbf{X\Lambda} \Leftrightarrow \mathbf{AV} = \mathbf{U\Sigma}$$

$$\underbrace{\begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}}_{\substack{\text{not symmetric} \\ \mathbf{V} \neq \mathbf{U}}} \underbrace{\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}}_{\mathbf{V}^T = \mathbf{V}^{-1}} = \underbrace{\frac{1}{\sqrt{10}}\begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix}}_{\mathbf{U}^T = \mathbf{U}^{-1}} \underbrace{\begin{bmatrix} 3\sqrt{5} & \\ & \sqrt{5} \end{bmatrix}}_{\substack{rank(\mathbf{A})=2 \rightarrow \sigma_1, \sigma_2 \\ \sigma_1 \sigma_2 = \det(\mathbf{A})}}$$

$$\sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T = \frac{3\sqrt{5}}{\sqrt{10}\sqrt{2}}\begin{bmatrix} 1 \\ 3 \end{bmatrix}\begin{bmatrix} 1 & 1 \end{bmatrix} + \frac{\sqrt{5}}{\sqrt{10}\sqrt{2}}\begin{bmatrix} -3 \\ 1 \end{bmatrix}\begin{bmatrix} -1 & 1 \end{bmatrix} = \frac{3}{2}\begin{bmatrix} 1 & 1 \\ 3 & 3 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 3 & -3 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} = \mathbf{A}$$

- $\mathbf{V}$ contains orthonormal eigenvectors of $\mathbf{A}^T\mathbf{A}$
- $\mathbf{U}$ contains orthonormal eigenvectors of $\mathbf{A}\mathbf{A}^T$
- σ_1^2 to σ_r^2 are the nonzero eigenvalues of both $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$

SVD requires that $\mathbf{A}\mathbf{v}_k = \sigma_k \mathbf{u}_k$: $\mathbf{v}$'s $\left(\mathbf{A}^T\mathbf{A}\mathbf{v}_k = \sigma_k^2 \mathbf{v}_k\right) \rightarrow \mathbf{u}$'s ($\mathbf{u}_k = \frac{\mathbf{A}\mathbf{v}_k}{\sigma_k}$: sign, multiple eigenvalues)

check 1: $\mathbf{u}$'s are eigenvectors of $\mathbf{A}\mathbf{A}^T \rightarrow \mathbf{A}\mathbf{A}^T\mathbf{u}_k =$

check 2: $\mathbf{u}$'s are also orthonormal $\rightarrow \mathbf{u}_j^T\mathbf{u}_k =$

choose $(n-r)$ $\mathbf{v}$'s in $N(\mathbf{A})$ and $(m-r)$ $\mathbf{u}$'s in $N(\mathbf{A}^T)$

- Columns of V are orthogonal eigenvectors of A^TA
- $Av=\sigma u$ gives orthonormal eigenvectors u of AA^T
- σ^2 = eigenvalue of A^TA = eigenvalue of $AA^T \neq 0$

- Why is the SVD so important?
  - It separates the matrix into rank one pieces like the other factorizations A=LU, A=QR, $S=Q\Lambda Q^T$
  - Those pieces come in order of importance
  - First piece $\sigma_1 u_1 v_1^T$ is the closest rank one matrix to A
  - Sum of the first k pieces is best possible for rank k

$$\mathbf{A}_k = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T \text{ is the best rank } k \text{ approximation to } \mathbf{A}:$$

$$\text{If } \mathbf{B} \text{ has rank k then } \|\mathbf{A} - \mathbf{A}_k\| \leq \|\mathbf{A} - \mathbf{B}\|$$

Example

$$\text{Find the matrices } \mathbf{U}, \mathbf{\Sigma}, \mathbf{V} \text{ for } \mathbf{A} = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} \rightarrow \mathbf{A}^T\mathbf{A} = \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix}, \mathbf{A}\mathbf{A}^T = \begin{bmatrix} 9 & 12 \\ 12 & 41 \end{bmatrix}$$

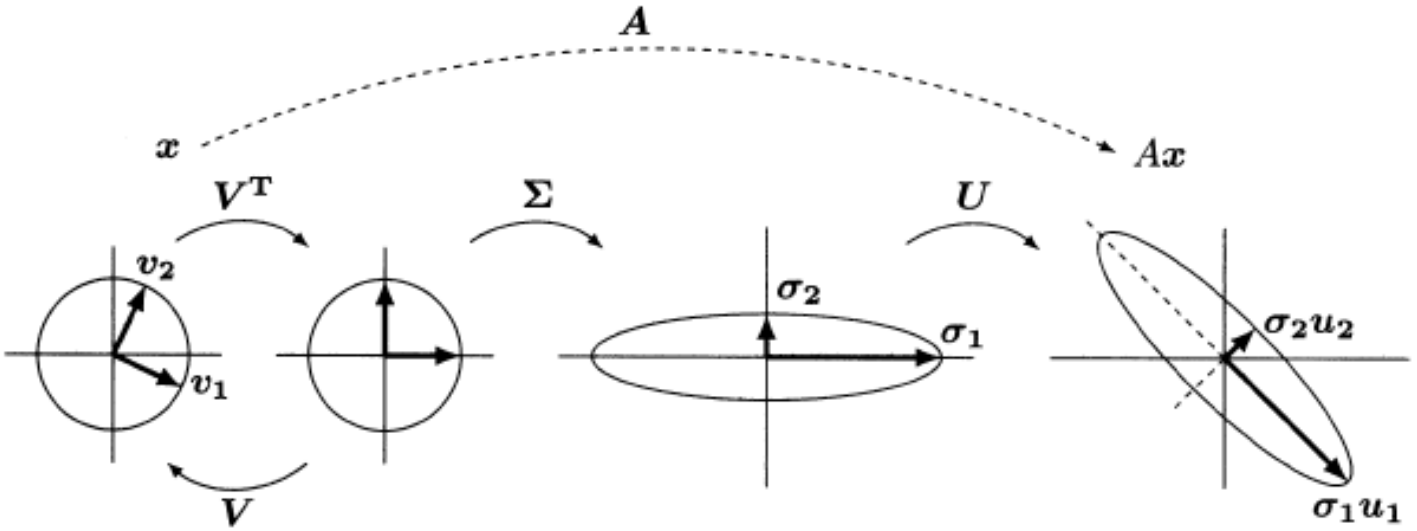
$$\mathbf{U} = \frac{1}{\sqrt{10}}\begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix}, \mathbf{\Sigma} = \begin{bmatrix} \sqrt{45} & \\ & \sqrt{5} \end{bmatrix}, \mathbf{V} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

$$\sigma_1\mathbf{u}_1\mathbf{v}_1^T + \sigma_2\mathbf{u}_2\mathbf{v}_2^T = \mathbf{A}$$

- If $S=Q\Lambda Q^T$ is symmetric positive definite, what is its SVD?
- If $S=Q\Lambda Q^T$ has a negative eigenvalue($Sx=-\alpha x$), what is the singular value and what are the vectors v and u?
- If A=Q is an orthogonal matrix, why does every singular value equal 1?
- Why are all eigenvalues of a square matrix A less than or equal to σ_1?
- If $A=xy^T$ has rank 1, what are u_1, v_1, σ_1? Check that $|\lambda_1| \leq \sigma_1$

Geometry of SVD

- A = (rotation)(stretching)(rotation) $U\Sigma V^T$ for every A
- If A is m by n and B is n by m, then AB and BA have the same nonzero eigenvalues



$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \underbrace{\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \sigma_1 & \\ & \sigma_2 \end{bmatrix} \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}}_{\text{4 parameters: two angles, two numbers}}$$

First singular vector v_1

Maximize the ratio $\frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|} \rightarrow$ The maximum is σ_1 at the vector $\mathbf{x} = \mathbf{v}_1$

maximizing $\mathbf{x}$ is $\mathbf{v}_1$: $\mathbf{Av}_1 = \sigma_1 \mathbf{u}_1$ (the longest axis of the ellipse), $\|\mathbf{v}_1\| = 1 \rightarrow \|\mathbf{Av}_1\| = \sigma_1$

$\Rightarrow$ Find the maximum value λ of $\frac{\|\mathbf{Ax}\|^2}{\|\mathbf{x}\|^2} = \frac{(\mathbf{Ax})^T \mathbf{Ax}}{\mathbf{x}^T\mathbf{x}} = \frac{\mathbf{x}^T\mathbf{Sx}}{\mathbf{x}^T\mathbf{x}}$

$$\frac{\partial}{\partial x_i}\left(\frac{\mathbf{x}^T\mathbf{Sx}}{\mathbf{x}^T\mathbf{x}}\right) = \left(\mathbf{x}^T\mathbf{x}\right)2\left(\mathbf{Sx}\right)_i - \left(\mathbf{x}^T\mathbf{Sx}\right)2\left(\mathbf{x}\right)_i = 0 \text{ for } i = 1, \ldots, n$$

$$\rightarrow \left(\mathbf{Sx}\right)_i = \left(\frac{\mathbf{x}^T\mathbf{Sx}}{\mathbf{x}^T\mathbf{x}}\right)\left(\mathbf{x}\right)_i \rightarrow \mathbf{Sx} = \lambda\mathbf{x}$$

Maximize $\frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|}$ under the conditon $\mathbf{v}_1{}^T\mathbf{x} = 0 \rightarrow$ The maximum is σ_2 at $\mathbf{x} = \mathbf{v}_2$

Polar decomposition

$$\underset{\text{complex number}}{x+iy} = \underset{\text{polar form}}{re^{i\theta}} \rightarrow \begin{cases} e^{i\theta}: \text{ orthogonal matrix } \mathbf{Q} \\ r \geq 0: \text{ positive semideinite matrix } \mathbf{S} \end{cases}$$

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T = \left(\mathbf{U}\mathbf{V}^T\right)\left(\mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^T\right) = \mathbf{Q}\mathbf{S}$$

if $\mathbf{A}$ is invertible, then $\boldsymbol{\Sigma}$ and $\mathbf{S}$ are also invertible

$$\mathbf{S}^2 = \mathbf{V}\boldsymbol{\Sigma}^2\mathbf{V}^T = \mathbf{A}^T\mathbf{A} \rightarrow \begin{cases} \text{eigenvalues of } \mathbf{S} = \text{singular values of } \mathbf{A} \\ \text{eigenvectors of } \mathbf{S} = \text{singular vectors } \mathbf{v} \text{ of } \mathbf{A} \end{cases}$$

$$\mathbf{A} = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} \rightarrow \mathbf{U} = \frac{1}{\sqrt{10}}\begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} \sqrt{45} & \\ & \sqrt{5} \end{bmatrix}, \mathbf{V} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

$$\mathbf{Q} = \mathbf{U}\mathbf{V}^T = \frac{1}{\sqrt{10}}\begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}$$

$$\mathbf{S} = \mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^T = \frac{\sqrt{5}}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 3 & \\ & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} = \sqrt{5}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

$$\mathbf{A} = \underset{\text{rotation}}{\mathbf{Q}}\ \underset{\text{stretch}}{\mathbf{S}}$$

9. Principal Components and the Best Low Rank Matrix

- major tool in understanding a matrix of data
  - Schmidt(1907)→ Eckart and Young(1936, $\|A\|_F$)→Mirsky(1955)
- Eckart-Young low rank approximation theorem
  - The norm of $A-A_k$ is below the norm of all other $A-B_k$
  - $A_k = \sigma_1 u_1 v_1^T + \ldots + \sigma_k u_k v_k^T$

Eckart-Young: If $\mathbf{B}$ has rank k, then $\|\mathbf{A}-\mathbf{B}\| \geq \|\mathbf{A}-\mathbf{A}_k\|$

$\mathbf{A}_k = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T$: the closest rank k matrix to $\mathbf{A}$

$$\begin{cases} \text{Spectral norm: } \|\mathbf{A}\|_2 = \max\limits_{\mathbf{x}\neq 0} \dfrac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|} = \sigma_1 \left(\ell^2 \text{ norm}\right) \\ \text{Frobenius norm: } \|\mathbf{A}\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2} \\ \text{Nuclear norm: } \|\mathbf{A}\|_N = \sigma_1 + \cdots + \sigma_r \left(\text{the trace norm}\right) \\ \|\mathbf{I}\|_2 = 1 = \|\mathbf{Q}\|_2, \ \|\mathbf{I}\|_F = \sqrt{n} = \|\mathbf{Q}\|_F, \ \|\mathbf{I}\|_N = n = \|\mathbf{Q}\|_N \end{cases}$$

Eckart-Young Theorem

- Best approximation by A_k

Eckart-Young in L^2:

If $\text{rank}(\mathbf{B}) \leq k$, then $\|\mathbf{A}-\mathbf{B}\| = \max_{\mathbf{x}\neq 0} \dfrac{\|(\mathbf{A}-\mathbf{B})\mathbf{x}\|}{\|\mathbf{x}\|} \geq \sigma_{k+1}$

Eckart-Young in the Frobenius norm:

If $\mathbf{B}$ is closest to $\mathbf{A}$, then $\mathbf{U}^T\mathbf{B}\mathbf{V}$ is closest to $\mathbf{U}^T\mathbf{A}\mathbf{V}$

$$\mathbf{B} = \mathbf{U}\begin{bmatrix} \mathbf{D}_{k\times k} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{V}^T, \mathbf{A} = \begin{bmatrix} \mathbf{L}+\mathbf{E}+\mathbf{R} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{bmatrix}$$

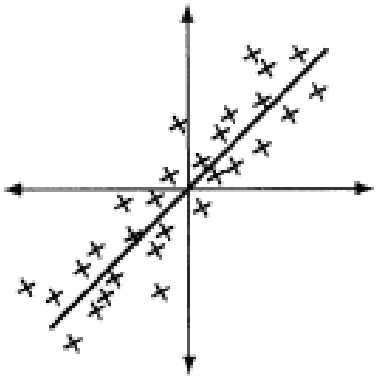
The matrix $\mathbf{D}$ must be the same as $\mathbf{E} = diag(\sigma_1,\ldots,\sigma_k)$

The singular values of $\mathbf{H}$ must be the smallest $(n-k)$ singular values of $\mathbf{A}$

The smallest error $\|\mathbf{A}-\mathbf{B}\|_F$ must be $\|\mathbf{H}\|_F = \sqrt{\sigma_{k+1}^2+\cdots+\sigma_r^2}$

Principal Component Analysis

- Understand *n* sample points in *m*-dimensional space
- Data matrix A_0: *n* samples, *m* variables
  - Find the average (the sample mean) along each row of A_0
  - Subtract that mean from *m* entries in the row
  - Centered matrix $A=A_0$-(mean)
  - How will linear algebra find that closest line through (0,0)? It is in the direction of the first singular vector u_1 of A?



A is $\mathbf{2} \times \boldsymbol{n}$ (large nullspace)

AA^{T} is $\mathbf{2} \times \mathbf{2}$ (small matrix)

$A^{\mathrm{T}}A$ is $\boldsymbol{n} \times \boldsymbol{n}$ (large matrix)

Two singular values $\boldsymbol{\sigma_1 > \sigma_2 > 0}$

- Statistics behind PCA
  - Variances: diagonal entries of the matrix AA^T
  - Covariances: off- diagonal entries of the matrix AA^T
  - Sample covariance matrix: $S=AA^T/(n-1)$
- Geometry behind PCA
  - Sum of squared distances from the data points to the line is a minimum
- Linear algebra behind PCA
  - Singular values σ_i and singular vectors u_i of A
  - Total variance:

$$T = \frac{\|\mathbf{A}\|_F^2}{n-1} = \frac{\sigma_1^2 + \cdots + \sigma_r^2}{n-1}$$

11. Norms of Vectors and Matrices

- The norm of a nonzero vector v is a positive number ||v||
- That number measures the “length” of the vector

every norm for vectors or functions or matrice must share these two properties of the absolute value $|c|$ of a number

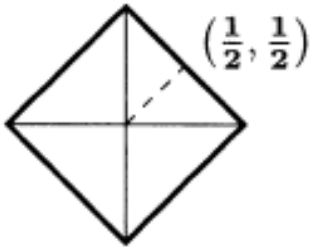
$$\text{All norms} \begin{cases} \text{multiply } \mathbf{v} \text{ by } c \text{ (rescaling)} \rightarrow \|c\mathbf{v}\| = |c|\|\mathbf{v}\| \\ \text{add } \mathbf{v} \text{ to } \mathbf{w} \text{ (Triangle inequality)} \rightarrow \|\mathbf{v}+\mathbf{w}\| \le \|\mathbf{v}\| + \|\mathbf{w}\| \end{cases}$$

$$\begin{cases} \ell^2 \text{ norm} = \text{Euclidean norm: } \|\mathbf{v}\|_2 = \sqrt{|v_1|^2 + \cdots + |v_n|^2} \\ \ell^1 \text{ norm} = 1\text{- norm: } \|\mathbf{v}\|_1 = |v_1| + \cdots + |v_n| \\ \ell^\infty \text{ norm} = \text{max norm: } \|\mathbf{v}\|_\infty = \text{maximum of } |v_1|, \ldots, |v_n| \end{cases}$$

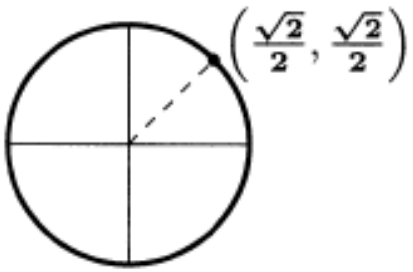
$$\|\mathbf{v}\|_p = \left(|v_1|^p + \cdots + |v_n|^p \right)^{1/p}$$

- Important vector norms and a failure

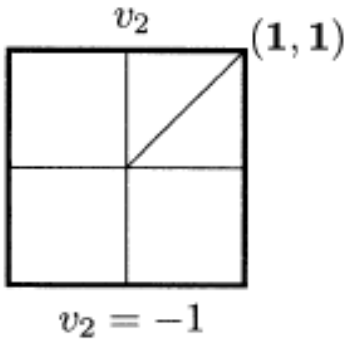
ℓ^1 norm
$|v_1| + |v_2| \leq 1$
diamond



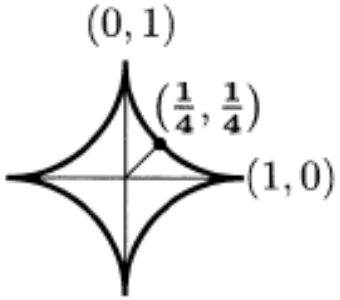
ℓ^2 norm
$v_1^2 + v_2^2 \leq 1$
circle



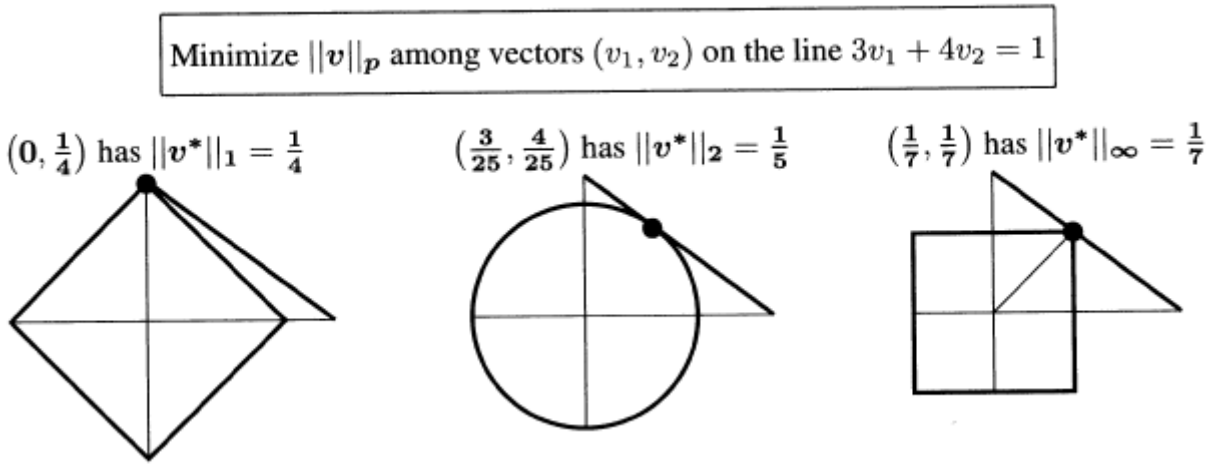
ℓ^∞ norm
$|v_1| \leq 1, |v_2| \leq 1$
square



$\ell^{1/2}$ norm
$\sqrt{|v_1|} + \sqrt{|v_2|} \leq 1$
not convex



- Minimum of $||v||_p$ on the line $a_1v_1 + a_2v_2 = 1$



- Inner products and S=norm

$$\left.\begin{array}{l}\text{Inner product} = \text{length squared}: \mathbf{v}\cdot\mathbf{v} = \mathbf{v}^T\mathbf{v} = \|\mathbf{v}\|^2 \\ \text{Angle } \theta \text{ between vector } v \text{ and } w: \mathbf{v}^T\mathbf{w} = \|\mathbf{v}\|\|\mathbf{w}\|\cos\theta\end{array}\right\} \rightarrow \left\{\begin{array}{l}\text{Cauchy - Schwarz}: \left|\mathbf{v}^T\mathbf{w}\right| \le \|\mathbf{v}\|\|\mathbf{w}\| \\ \text{Triangle Inequality}: \|\mathbf{v}+\mathbf{w}\| \le \|\mathbf{v}\| + \|\mathbf{w}\|\end{array}\right.$$

Choose any symmetric positive definite matrix $\mathbf{S}$

$\|\mathbf{v}\|_S^2 = \mathbf{v}^T\mathbf{S}\mathbf{v}$ gives a norm for $\mathbf{v}$ in $\Re^n$ (called the S - norm)

$(\mathbf{v},\mathbf{w})_S = \mathbf{v}^T\mathbf{S}\mathbf{w}$ gives the S - inner product for $\mathbf{v},\mathbf{w}$ in $\Re^n$

Norm of Matrices

- Frobenius Norm
- Matrix Norm ||A|| from vector norm ||v||
- Nuclear Norm

$$\begin{cases} \|\mathbf{A}\|_F = \sqrt{|a_{11}|^2 + \cdots + |a_{1n}|^2 + \cdots + |a_{mn}|^2} \\ \|\mathbf{A}\|_F = \|\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T\|_F = \|\boldsymbol{\Sigma}\mathbf{V}^T\|_F = \|\boldsymbol{\Sigma}\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2} \\ \|\mathbf{A}\|_F^2 = \text{trace of } \mathbf{A}^T\mathbf{A} = \text{sum of eigenvalues} = \sigma_1^2 + \cdots + \sigma_r^2 \end{cases}$$

$$\begin{cases} \|\mathbf{A}\| = \max_{\mathbf{v}\neq 0} \dfrac{\|\mathbf{A}\mathbf{v}\|}{\|\mathbf{v}\|} = \text{largest growth factor} \\ \ell^2 \text{ norm}: \ \|\mathbf{A}\|_2 = \text{largest singular value } \sigma_1 \text{ of } \mathbf{A} \\ \ell^1 \text{ norm}: \ \|\mathbf{A}\|_1 = \text{largest } \ell^1 \text{ norm of the columns of } \mathbf{A} \\ \ell^\infty \text{ norm}: \ \|\mathbf{A}\|_\infty = \text{largest } \ell^1 \text{ norm of the rows of } \mathbf{A} \end{cases}$$

$$\|\mathbf{A}\|_N = \sigma_1 + \cdots + \sigma_r = \text{trace norm}$$