5. Orthogonal Matrices and Subspaces

- 1. Orthogonal vectors x and y

$$(x-y)^T(x-y) = x^Tx + y^Ty - x^Ty - xy^T$$

$$\|x\|^2 + \|y\|^2 - 2\|x\|\|y\|\cos\theta$$

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

$\theta = 90°$

- 2. Orthogonal basis for a subspace

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad H_4 = \begin{bmatrix} H_2 & H_2 \\ H_2 & -H_2 \end{bmatrix}$$

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

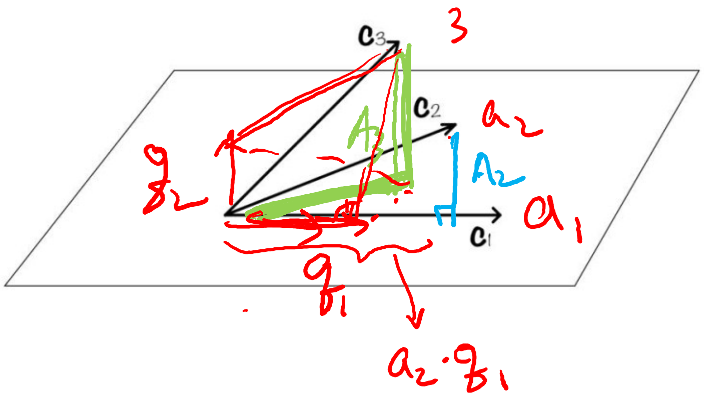
  - Standard basis is orthogonal (even orthonormal) in $\mathbf{R}^n$ (i, j, k in $\mathbf{R}^3$)
  - Hadamard matrices H_n containing orthogonal bases of $\mathbf{R}^n$
    - Are those orthogonal matrices?

$$H_8 = \begin{bmatrix} H_4 & H_4 \\ H_4 & -H_4 \end{bmatrix}$$

  - Every subspace of $\mathbf{R}^n$ has an orthogonal basis: Gram-Schmidt idea
    - Two independent vectors a and b in the plane: $a^Tc=0$

$A_2 \perp q_1?$ $\quad A_2^T q_1 = 0 = [a_2 - (a_2^T q_1) q_1]^T \cdot q_1$

$= a_2^T q_1 - (a_2^T q_1) \cdot q_1^T q_1$



$q_1 = \frac{a_1}{\|a_1\|}$ $\quad A_2 = a_2 - (\underline{a_2^T q_1}) q_1$

$q_2 = \frac{A_2}{\|A_2\|}$ $\quad A_3 = a_3 - \left[(a_3^T q_2) q_2 + (a_3^T q_1) q_1\right]$

$q_3 = \frac{A_3}{\|A_3\|}$

$a_1 = \|a_1\| q_1$

$a_2 = (a_2^T q_1) q_1 + \|A_2\| q_2$

$a_3 = (\underline{a_3^T q_1}) q_1 + (\underline{a_3^T q_2}) q_2 + \|A_3\| q_3$

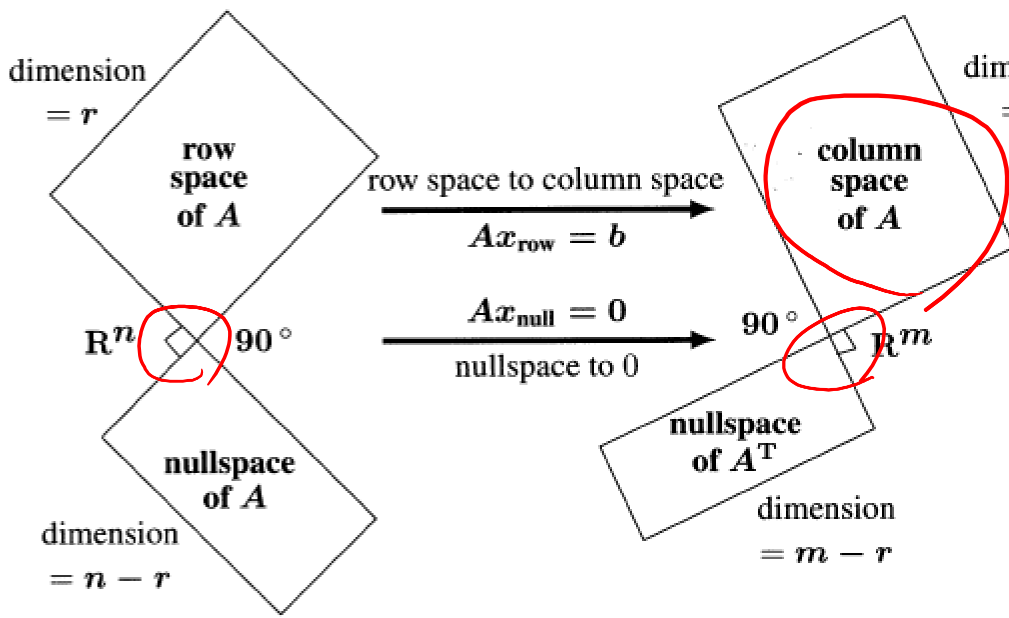
$$\begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \\ 0 & 0 & r_{33} \end{bmatrix}$$

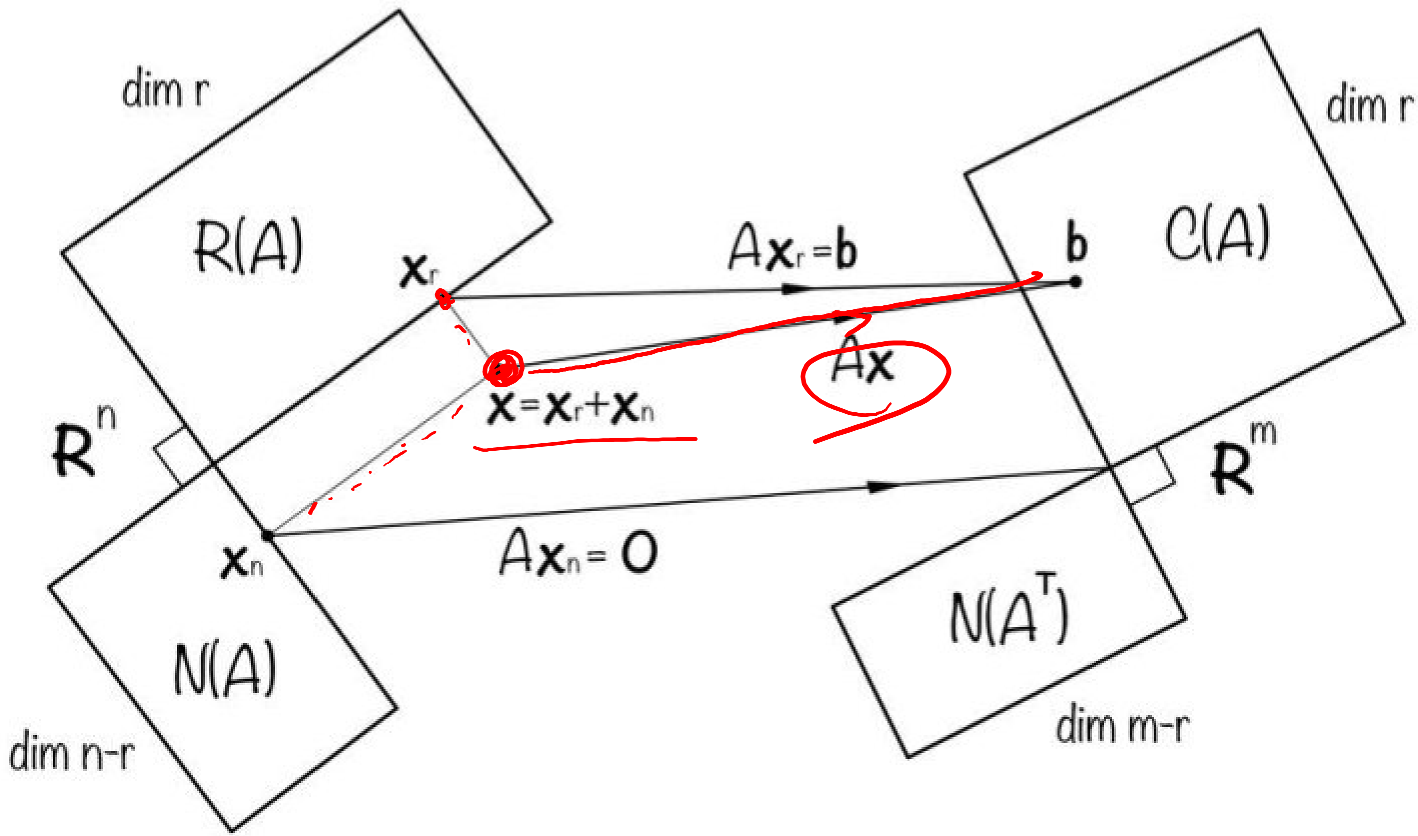
$A = QR$

$\underline{r_{ij} = q_i^T a_j}$

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

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




dim r
R(A)
$\mathbf{x}_r$
$A\mathbf{x}_r = \mathbf{b}$
b
C(A)
dim r
$A\mathbf{x}$
$\mathbf{x} = \mathbf{x}_r + \mathbf{x}_n$
$\mathbf{R}^n$
$\mathbf{R}^m$
$\mathbf{x}_n$
$A\mathbf{x}_n = 0$
N(A)
$N(A^T)$
dim n-r
dim m-r

$m > n$

n

m

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

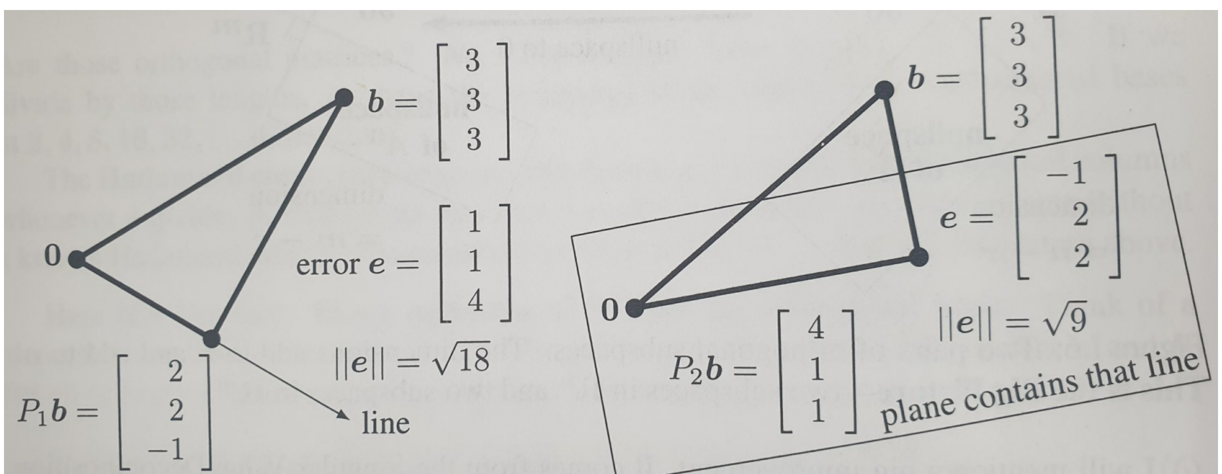
$$\begin{cases} \text{if } \mathbf{Q} \text{ multiplies any vector } \mathbf{x}\text{, the length of the vector does not change: } \|\mathbf{Qx}\| = \|\mathbf{x}\| \\ \text{if } m > n \text{ then } m \text{ rows cannot be orthogonal in } \mathbf{R}^n\text{: } \mathbf{QQ}^T \neq \mathbf{I} \end{cases}$$

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

$$P^2 = (Q\underbrace{Q^T)(Q}_{I}Q^T) = QQ^T = P$$

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

$\mathbf{Pb}$ is the orthogonal projection of $\mathbf{b}$ onto the column space of $\mathbf{P}$: $\mathbf{P}_1\mathbf{b},\ \mathbf{P}_2\mathbf{b},\ \mathbf{P}_3\mathbf{b}$



$$\mathbf{P}_1 = \mathbf{Q}_1\mathbf{Q}_1^T = \frac{1}{9}\begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}\begin{bmatrix} 2 & 2 & -1 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 4 & 4 & -2 \\ 4 & 4 & -2 \\ -2 & -2 & 1 \end{bmatrix} \rightarrow \mathbf{P}_1\mathbf{b} = \frac{1}{9}\begin{bmatrix} 18 \\ 18 \\ -9 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}$$

$$\mathbf{P}_2 = \mathbf{Q}_2\mathbf{Q}_2^T = \frac{1}{9}\begin{bmatrix} 2 & 2 \\ 2 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 2 & & \\ & & \\ & & \end{bmatrix}$$

$\rightarrow P_2 b$

$Q_3 Q_3^T = I$

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

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

$\mathbf{Pb}$ is the orthogonal projection of $\mathbf{b}$ onto the column space of $\mathbf{P}$: $\mathbf{P}_1\mathbf{b}$, $\mathbf{P}_2\mathbf{b}$, $\mathbf{P}_3\mathbf{b}$

$P_3 = Q_3 Q_3^T = I \rightarrow P_3 b = b$

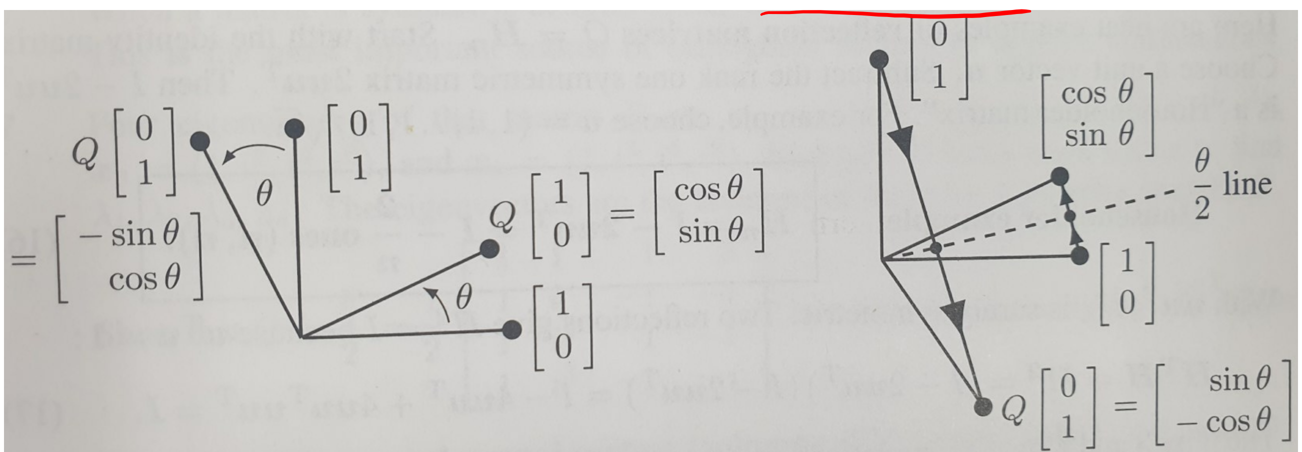
- 5. Orthogonal matrices are square with orthonormal columns: $Q^T = Q^{-1}$

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

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

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

$$(Q_1Q_2)^T(Q_1Q_2)$$
$$= \underbrace{Q_2^T \underbrace{Q_1^T Q_1}_{} Q_2}_{}$$
$$= I$$



$$\begin{cases} \text{Orthogonal basis} = \text{orthogonal axes in } \mathbf{R}^n \\ \text{orthogonal } \mathbf{Q}(n\times n):\ \mathbf{v} = c_1\mathbf{q}_1 + \cdots + c_n\mathbf{q}_n \to c_i = \mathbf{q}_i^T\mathbf{v} \\ \mathbf{v} = \mathbf{Qc} \to \mathbf{Q}^T\mathbf{v} = \mathbf{Q}^T\mathbf{Qc} = \mathbf{c} \end{cases}$$

$$(I - 2uu^T)u = u - 2uu^Tu$$
$$= -u$$

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

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

$\mathbf{H}^T\mathbf{H} = \mathbf{H}^2 = \mathbf{I}$ $\quad (I-2uu^T)(I-2uu^T) = I - 4uu^T + 4uu^Tuu^T$

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

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

All reflection matrices have eigenvalues -1 and 1

$$\mathbf{H} = \mathbf{I} - 2\frac{\mathbf{v}\mathbf{v}^T}{\|\mathbf{v}\|^2} = \mathbf{I} - 2\mathbf{u}\mathbf{u}^T$$

$u = \frac{v}{\|v\|}$

key point: if $\mathbf{v} = \mathbf{a} - \mathbf{r}$ and $\|\mathbf{a}\| = \|\mathbf{r}\|$, then $\mathbf{Ha} = \mathbf{r}$

$$\mathbf{H}_{n-1} \cdots \mathbf{H}_2\mathbf{H}_1\mathbf{A} = \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \cdots & \mathbf{r}_n \end{bmatrix} \rightarrow \mathbf{Q}^T\mathbf{A} = \mathbf{R}$$

$$\mathbf{A} = \begin{bmatrix} 4 & x \\ 3 & x \end{bmatrix} \rightarrow \mathbf{a} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}, \mathbf{r} = \begin{bmatrix} 5 \\ 0 \end{bmatrix} \rightarrow \mathbf{v} = \mathbf{a} - \mathbf{r} = \begin{bmatrix} -1 \\ 3 \end{bmatrix} \rightarrow \mathbf{u} = \frac{1}{\sqrt{10}}\begin{bmatrix} -1 \\ 3 \end{bmatrix}$$

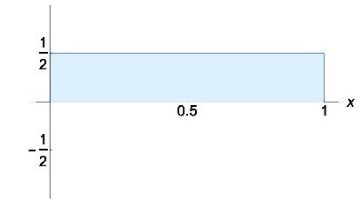
$$\mathbf{H} = \mathbf{I} - 2\mathbf{u}\mathbf{u}^T = \frac{1}{5}\begin{bmatrix} 4 & 3 \\ 3 & -4 \end{bmatrix} = \mathbf{Q}^T \rightarrow \mathbf{HA} = \begin{bmatrix} 5 & x \\ 0 & x \end{bmatrix} = \mathbf{R}$$

$R\hat{x} = Q^T b$

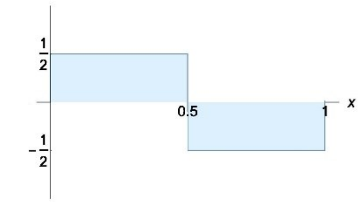
Examples

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

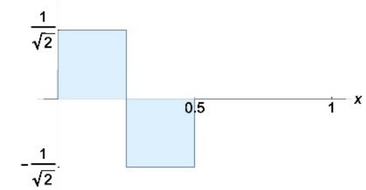
$$\mathbf{W} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & -1 & 0 \\ 1 & -1 & 0 & 1 \\ 1 & -1 & 0 & -1 \end{bmatrix}$$



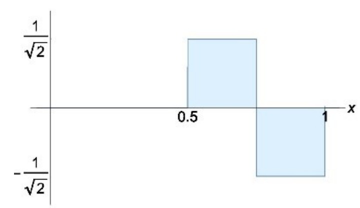
(a) Haar function of $h_0(x)$



(b) Haar function of $h_1(x)$



(c) Haar function of $h_2(x)$



(d) Haar function of $h_3(x)$

6. Eigenvalues and Eigenvectors

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

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

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

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

How useful? $\begin{cases}\text{solution of differential equations} \\ \text{similar matrices} \rightarrow \text{same eigenvalues} \\ \text{diagonalize a matrix}\end{cases}$

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

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

(Determinant) $\prod \lambda_i$ = determinant of matrix

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

$X_1 \cdot X_2 = 0$

(Orthogonal eigenvectors) if $\lambda_1 \neq \lambda_2$, then $\mathbf{x}_1 \cdot \mathbf{x}_2 = 0$, eigenvectors of $\mathbf{A}$ are orthogonal iff $\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T$

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

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

$\operatorname{eig}(\mathbf{AB}) \neq \operatorname{eig}(\mathbf{A})\operatorname{eig}(\mathbf{B})$

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

(1) $\mathbf{A}$ controls a system of linear differential equations: $\dfrac{d\mathbf{u}}{dt} = \boxed{\mathbf{A}}\mathbf{u}$ with $\mathbf{u}(0)$ $u_h = e^{\lambda t}$

$= \lambda u$

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

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

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

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

Make $\mathbf{B}$ gradually into a triangular matrix $\to$ Gradually show up on the main diagonal

$$\mathbf{By} = \boldsymbol{\lambda}\mathbf{y} \to \mathbf{MAM}^{-1}\mathbf{y} = \boldsymbol{\lambda}\mathbf{y} \to \mathbf{A}(\mathbf{M}^{-1}\mathbf{y}) = \boldsymbol{\lambda}(\mathbf{M}^{-1}\mathbf{y})$$

$$\underline{M^{-1}M}A(\underline{M^{-1}y}) = \lambda(\underline{M^{-1}y}) \to Ax = \lambda x$$

$\rightarrow A = X\Lambda X^{-1} = \begin{bmatrix} 3 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 10 & \\ & 5 \end{bmatrix} \begin{bmatrix} -1 & -1 \\ -2 & 3 \end{bmatrix} \frac{1}{-5}$

(3) diagonalize a matrix

$X^{-1}AX = \Lambda$

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

$(X\Lambda X^{-1})(X\Lambda X^{-1})$

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

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

the action of the whole matrix $\mathbf{A}$ is broken into simple actions (just muliply by λ)

[nondiagonalizable matrices: when GM < AM, $\mathbf{A}$ is not diagonalizable]

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

$$\mathbf{A} = \begin{bmatrix} 5 & 1 \\ 0 & 5 \end{bmatrix}, \begin{bmatrix} 6 & -1 \\ 1 & 4 \end{bmatrix}, \begin{bmatrix} 7 & 2 \\ -2 & 3 \end{bmatrix} \rightarrow \begin{cases} \det(\mathbf{A} - \lambda\mathbf{I}) = (\lambda - 5)^2 = 0 \rightarrow \text{AM} = 2 \\ rank(\mathbf{A} - 5\mathbf{I}) = 1 \rightarrow \text{GM} = 1 \end{cases}$$

8. Singular Value Decomposition (SVD)

best matrices (real symmetric matrices $\mathbf{S}$): real eigenvalues and orthogonal eigenvectors

other matrices ($\mathbf{A}$ 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 $\mathbf{v}$'s and m $\mathbf{u}$'s

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

$${}_m\Big[\ \mathbf{U}\ \Big]^{m}\ {}_m\Big[\ \mathbf{A}\ \Big]^{n}\ \Big[\ {}_n\ \mathbf{V}\ \Big]^{n}$$

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

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

Proof of SVD

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

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

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

$$\begin{cases} \mathbf{V} \text{ contains orthonormal eigenvectors of } \mathbf{A}^T\mathbf{A} \\ \mathbf{U} \text{ contains orthonormal eigenvectors of } \mathbf{AA}^T \\ \sigma_1^2 \text{ to } \sigma_r^2 \text{ are the nonzero eigenvalues of both } \mathbf{A}^T\mathbf{A} \text{ and } \mathbf{AA}^T \end{cases}$$

$$AA^T \frac{Av_k}{\sigma_k} = \frac{A\sigma_k^2 v_k}{\sigma_k}$$

$$= \sigma_k A v_k$$

$$\sigma_k u_k$$

SVD requires that $\mathbf{Av}_k = \sigma_k \mathbf{u}_k$: $\mathbf{v}$'s $\left(\mathbf{A}^T\mathbf{Av}_k = \sigma_k^2 \mathbf{v}_k\right) \rightarrow \mathbf{u}$'s $\left(\begin{array}{l} \mathbf{u}_k = \dfrac{\mathbf{Av}_k}{\sigma_k} \text{ for } k = 1, \ldots, r \\ \text{sign, multiple eigenvalues} \end{array} \right)$

$(\text{check } 1)$ $\mathbf{u}$'s are eigenvectors of $\mathbf{AA}^T \rightarrow \mathbf{AA}^T\mathbf{u}_k = \sigma_k^2 \mathbf{u}_k$

$(\text{check } 2)$ $\mathbf{u}$'s are also orthonormal $\rightarrow \mathbf{u}_j^T \mathbf{u}_k = \dfrac{\sigma_k}{\sigma_j} \mathbf{v}_j^T \mathbf{v}_k = \begin{cases} 1 \text{ if } j = k \\ 0 \text{ if } j \neq k \end{cases}$

$$\left(\frac{Av_j}{\sigma_j}\right)^T \left(\frac{Av_k}{\sigma_k}\right) = \frac{v_j^T A^T A v_k}{\sigma_j \sigma_k}$$

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

8. Singular Value Decomposition (SVD)

- 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

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

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

Example

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

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

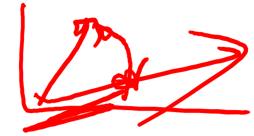
$$\sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T = \mathbf{A}$$

- If $S=Q\Lambda Q^T$ is symmetric positive definite, what is its SVD? $S = U\Sigma V^T$
- 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? $= \alpha(-x)$ $S(-x) = \alpha x$ $= QII^T$ $S(v) = (\sigma)(u)$ $Q^TQ = I$ $A = Q = V\Sigma V^T$
- 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 $|\lambda_1| \leq \sigma_1$
- What is the Karhunen-Loève transform and its connection to SVD?

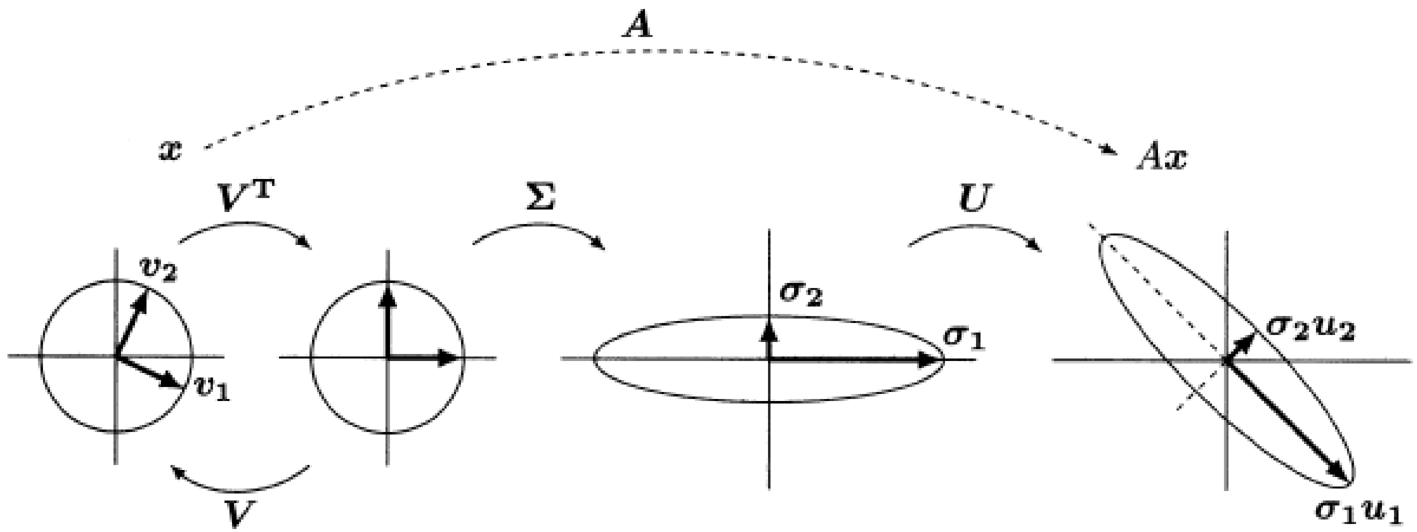
$\|Ax\| = \|U\Sigma V^T x\| = \|\Sigma V^T x\| \leq \sigma_1 \|V^T x\| = \sigma_1 \|x\|$ $\quad \frac{x}{\|x\|} \|x\| \|y\| \frac{y^T}{\|y\|}$

$= \|\lambda x\|$

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 $\mathbf{v}_1$

$\|Ax\| = \|\lambda x\|$

max $\lambda = \dfrac{\|Ax\|}{\|x\|}$

min

Maximize the ratio $\dfrac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|} \rightarrow$ 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), $\|\mathbf{v}_1\| = 1 \rightarrow \|\mathbf{Av}_1\| = \sigma_1$

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

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

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

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

Polar decomposition

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

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

$S = Q \Lambda Q^T$

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

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

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

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

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

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

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 $\mathbf{A}$

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

Eckart-Young Theorem

- Best approximation by A_k

$$\mathbf{A}=\begin{bmatrix}4&0&0&0\\0&3&0&0\\0&0&2&0\\0&0&0&1\end{bmatrix}\xrightarrow[\text{closest to }\mathbf{A}]{\text{rank 2 matrix}}$$

$$\mathbf{A}_2=\begin{bmatrix}4&0&0&0\\0&3&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix},\mathbf{B}_2=\begin{bmatrix}3.5&3.5&0&0\\3.5&3.5&0&0\\0&0&1.5&1.5\\0&0&1.5&1.5\end{bmatrix}$$

Eckart-Young in L^2:

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

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 $\mathbf{U}^T\mathbf{A}\mathbf{V}$

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

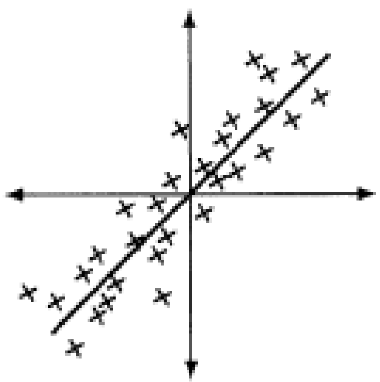
The matrix $\mathbf{D}$ must be the same as $\mathbf{E}=diag(\sigma_1,\ldots,\sigma_k)$

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 $\boldsymbol{\sigma_1} > \boldsymbol{\sigma_2} > \mathbf{0}$

- 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:

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