I.5 Orthogonal Matrices and Subspaces

- 1. Orthogonal vectors x and y

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

- 2. Orthogonal basis for a subspace
  - 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? (square, orthonormal vectors)

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

$$\left(i=\begin{bmatrix}1\\0\\0\end{bmatrix},\ j=\begin{bmatrix}0\\1\\0\end{bmatrix},\ k=\begin{bmatrix}0\\0\\1\end{bmatrix}\right)\text{ in }\mathbb{R}^3,\ \mathbf{H}_2=\begin{bmatrix}1&1\\1&-1\end{bmatrix},\ \mathbf{H}_4=\begin{bmatrix}\mathbf{H}_2&\mathbf{H}_2\\\mathbf{H}_2&-\mathbf{H}_2\end{bmatrix}=\begin{bmatrix}1&1&1&1\\1&-1&1&-1\\1&1&-1&-1\\1&-1&-1&1\end{bmatrix},\ \mathbf{H}_8=\begin{bmatrix}\mathbf{H}_4&\mathbf{H}_4\\\mathbf{H}_4&-\mathbf{H}_4\end{bmatrix}$$

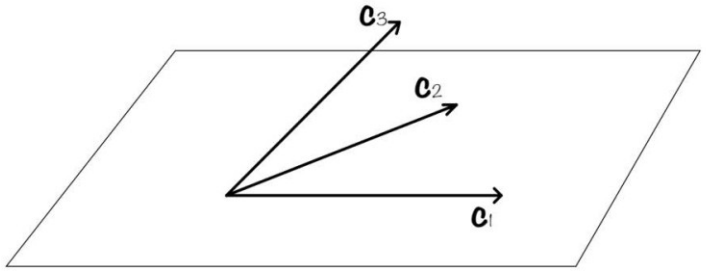
- Every subspace of $\mathbf{R}^n$ has an orthogonal basis: Gram-Schmidt idea
  - Two independent vectors a and b in the plane:

$$\mathbf{c} = \mathbf{b} - \frac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}}\mathbf{a} \rightarrow \mathbf{a}^T\mathbf{c} = \mathbf{0}$$

$$\left.\begin{aligned}
&\mathbf{q}_1 = \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|} \\
&\mathbf{A}_2 = \mathbf{a}_2 - \left(\mathbf{a}_2{}^T\mathbf{q}_1\right)\mathbf{q}_1 \rightarrow \mathbf{q}_2 = \frac{\mathbf{A}_2}{\|\mathbf{A}_2\|} \rightarrow \mathbf{q}_1{}^T\mathbf{A}_2 = 0\,? \\
&\mathbf{A}_3 = \mathbf{a}_3 - \left(\mathbf{a}_3{}^T\mathbf{q}_1\right)\mathbf{q}_1 - \left(\mathbf{a}_3{}^T\mathbf{q}_2\right)\mathbf{q}_2 \rightarrow \mathbf{q}_3 = \frac{\mathbf{A}_3}{\|\mathbf{A}_3\|}
\end{aligned}\right\}$$

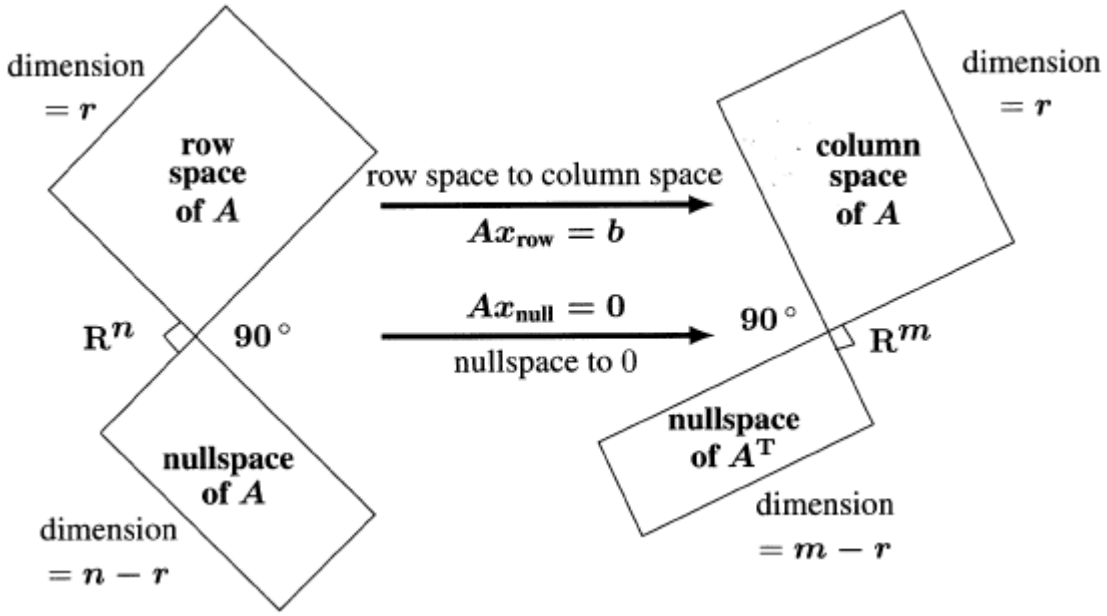
$$\rightarrow \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 \end{bmatrix} \begin{bmatrix} \|\mathbf{a}_1\| & \mathbf{a}_2{}^T\mathbf{q}_1 & \mathbf{a}_3{}^T\mathbf{q}_1 \\ & \|\mathbf{A}_2\| & \mathbf{a}_3{}^T\mathbf{q}_2 \\ & & \|\mathbf{A}_3\| \end{bmatrix}$$

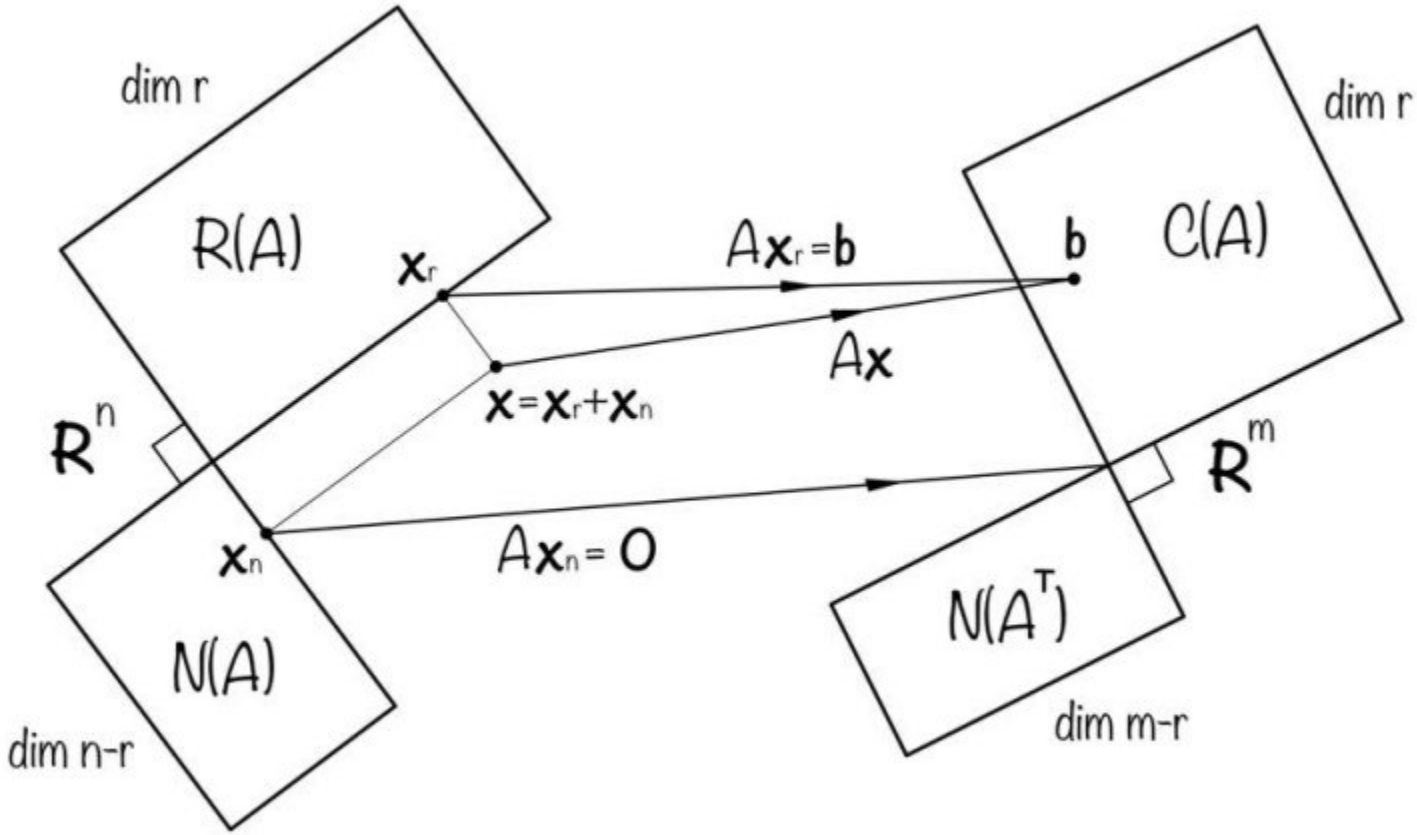
$$= \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ & r_{22} & r_{23} \\ & & r_{33} \end{bmatrix} \Leftrightarrow \mathbf{A} = \mathbf{QR}$$



- 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^Ty=0: The column space of A is orthogonal to the nullspace of A^T

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




dim r
R(A)
x_r
$Ax_r = b$
b
C(A)
dim r
$x = x_r + x_n$
Ax
R^n
R^m
x_n
$Ax_n = 0$
N(A)
$N(A^T)$
dim n-r
dim m-r

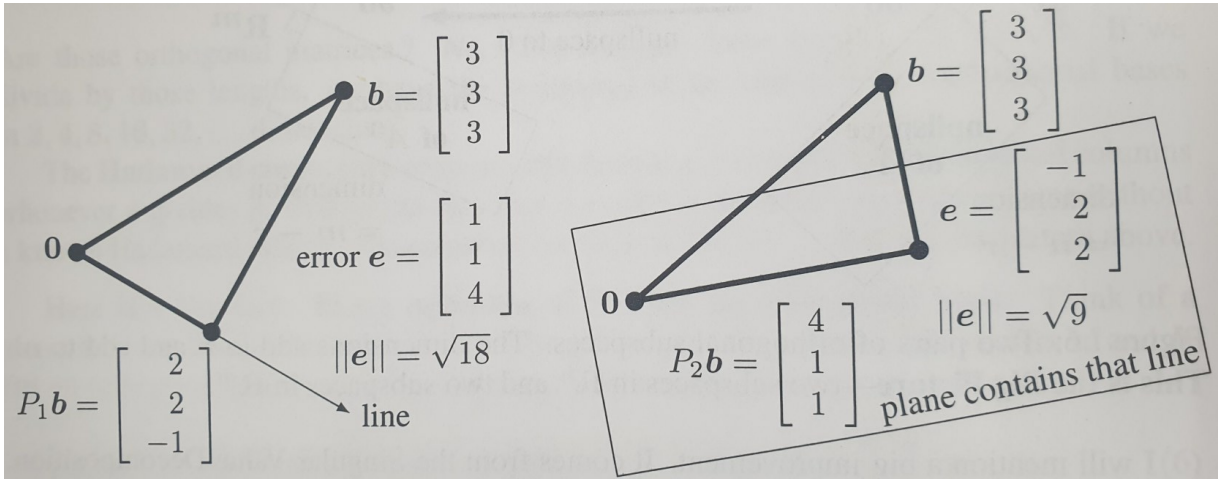
- 4. Tall thin matrices Q with orthonormal columns: $Q^TQ=I$

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

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

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

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



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

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

$$\mathbf{P}_2 = \mathbf{Q}_2 \mathbf{Q}_2^T = \frac{1}{9}\begin{bmatrix} 2 & 2 \\ 2 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 2 & 2 & -1 \\ 2 & -1 & 2 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 8 & 2 & 2 \\ 2 & 5 & -4 \\ 2 & -4 & 5 \end{bmatrix} \rightarrow \mathbf{P}_2\mathbf{b} = \frac{1}{9}\begin{bmatrix} 36 \\ 9 \\ 9 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ 1 \end{bmatrix}$$

$$\mathbf{P}_3 = \mathbf{Q}_3 \mathbf{Q}_3^T = \frac{1}{9}\begin{bmatrix} 2 & 2 & -1 \\ 2 & -1 & 2 \\ -1 & 2 & 2 \end{bmatrix}\begin{bmatrix} 2 & 2 & -1 \\ 2 & -1 & 2 \\ -1 & 2 & 2 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 9 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{bmatrix} = \mathbf{I} \rightarrow \mathbf{P}_3\mathbf{b} = \mathbf{b}$$

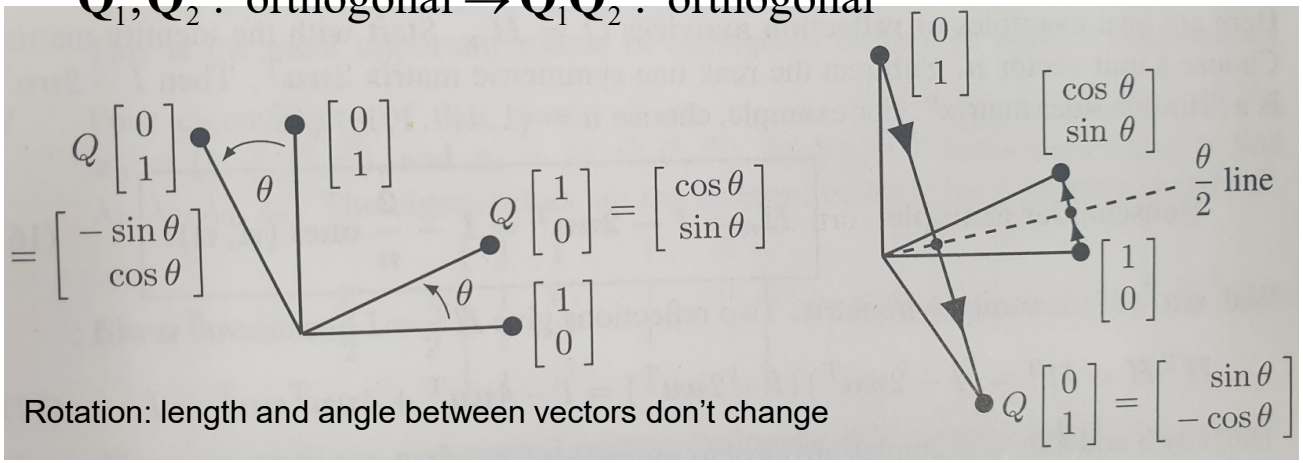
- 5. Orthogonal matrices are square with orthonormal columns: $Q^T=Q^{-1}$

(Rotation) times (Rotation) = (Rotation)
(Reflection) times (Reflection) = (Rotation)
(Rotation) times (Reflection) = (Reflection)

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

$$\underbrace{\mathbf{Q}_{\text{rotate}} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}}_{\text{rotation through an angle } \theta,\ \text{Det}(\mathbf{Q}_{\text{rotate}})=+1}, \quad \underbrace{\mathbf{Q}_{\text{reflect}} = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}}_{\text{reflection across the } \frac{\theta}{2} \text{ line},\ \text{Det}(\mathbf{Q}_{\text{reflect}})=-1}$$

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



Rotation: length and angle between vectors don't change

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

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

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

$\rightarrow$ when basis vectors are orthonormal, each coefficient can be found separately!

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

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

$$\mathbf{H}^T\mathbf{H} = \mathbf{H}^2 = \left(\mathbf{I} - 2\mathbf{u}\mathbf{u}^T\right)^T\left(\mathbf{I} - 2\mathbf{u}\mathbf{u}^T\right) = \mathbf{I} - 4\mathbf{u}\mathbf{u}^T + 4\mathbf{u}\mathbf{u}^T\mathbf{u}\mathbf{u}^T = \mathbf{I}$$

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

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

All reflection matrices have eigenvalues -1 and 1

Householder Reflections

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

key point: if $\mathbf{v} = \mathbf{a} - \mathbf{r}$ and $\|\mathbf{a}\| = \|\mathbf{r}\|$, then $\mathbf{H}\mathbf{a} = \mathbf{a} - 2\frac{(\mathbf{a}-\mathbf{r})(\mathbf{a}-\mathbf{r})^T}{(\mathbf{a}-\mathbf{r})^T(\mathbf{a}-\mathbf{r})}\mathbf{a} = \mathbf{r}$

$$\mathbf{H}_k\left[\text{column } k\right] = \begin{bmatrix} \mathbf{I} & \\ & \mathbf{I} - 2\mathbf{u}\mathbf{u}^T \end{bmatrix}\begin{bmatrix} \mathbf{a}_{\text{upper}} \\ \mathbf{a}_{\text{lower}} \end{bmatrix} = r_k$$

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

keep a record of the $\mathbf{H}_j$ by storing only the vectors $\mathbf{v}_j = \mathbf{a}_j - \mathbf{r}_j$, not the matrix

$$\mathbf{A}\mathbf{x} = \mathbf{b} \xrightarrow{\text{multiply by all the } \mathbf{H}\text{'s}} \mathbf{R}\mathbf{x} = \mathbf{Q}^T\mathbf{b}$$

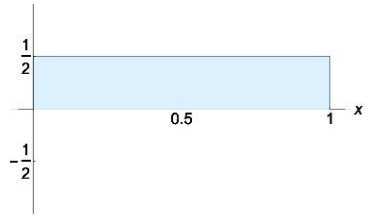
$$\mathbf{A} = \begin{bmatrix} 4 & x \\ 3 & x \end{bmatrix} \rightarrow \mathbf{a} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}, \mathbf{r} = \begin{bmatrix} 5 \\ 0 \end{bmatrix} \rightarrow \mathbf{v} = \mathbf{a} - \mathbf{r} = \begin{bmatrix} -1 \\ 3 \end{bmatrix} \rightarrow \mathbf{u} = \frac{1}{\sqrt{10}}\begin{bmatrix}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

$$\mathbf{H} = \mathbf{I} - 2\mathbf{u}\mathbf{u}^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - 2\frac{1}{10}\begin{bmatrix} 1 & -3 \\ -3 & 9 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 4 & 3 \\ 3 & -4 \end{bmatrix} = \mathbf{Q}^T \rightarrow \mathbf{H}\mathbf{A} = \begin{bmatrix} 5 & x \\ 0 & x \end{bmatrix} = \mathbf{R}$$

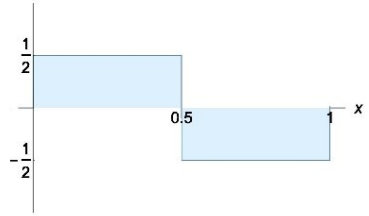
Examples

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

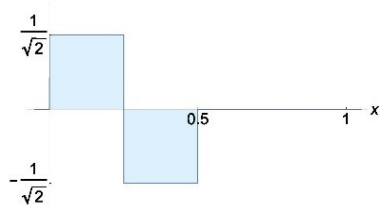
$$\mathbf{W} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & -1 & 0 \\ 1 & -1 & 0 & 1 \\ 1 & -1 & 0 & -1 \end{bmatrix}$$



(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 $h_3(x)$

I.6 Eigenvalues and Eigenvectors

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

$$\left.\begin{array}{l}\mathbf{x}:\ \text{eigenvector of } \mathbf{A}\\ \boldsymbol{\lambda}:\ \text{eigenvector of } \mathbf{A}\end{array}\right\} \rightarrow \mathbf{A}\mathbf{x}=\boldsymbol{\lambda}\mathbf{x} \rightarrow \mathbf{A}\left(\mathbf{A}\mathbf{x}\right)=\mathbf{A}\left(\boldsymbol{\lambda}\mathbf{x}\right)=\boldsymbol{\lambda}\left(\mathbf{A}\mathbf{x}\right)=\boldsymbol{\lambda}^2\mathbf{x} \rightarrow \mathbf{A}^k\mathbf{x}=\boldsymbol{\lambda}^k\mathbf{x},\ \mathbf{A}^{-1}\mathbf{x}=\frac{1}{\boldsymbol{\lambda}}\mathbf{x} \text{ if } \boldsymbol{\lambda}\neq 0$$

$n\times n$ 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?
- (1) solution of differential equations
- (2) similar matrices $\rightarrow$ same eigenvalues
- (3) diagonalize a matrix

$$\left.\begin{array}{l}|\lambda_1|>1:\ c_1{\lambda_1}^n\mathbf{x}_1 \text{ will grow as } n \text{ increases}\\ |\lambda_2|<1:\ c_2{\lambda_2}^n\mathbf{x}_2 \text{ will steadily disappear as } n \text{ increases}\end{array}\right\} \rightarrow \text{follow each eigenvector separately!}$$

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

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

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

Four properties: matrix $\mathbf{A}$(real), $\underbrace{\mathbf{S}(\text{symmetric})}_{\substack{\text{like real numbers: } \lambda \\ \text{every } \lambda \text{ is real}}}$, $\underbrace{\mathbf{Q}(\text{orthogonal})}_{\substack{\text{like complex numbers: } e^{i\theta} \\ \text{every } |\lambda|=1 \\ \text{powers of } \mathbf{Q} \text{ don't grow or decay}}}$

(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

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

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

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

$$\mathbf{A} = \begin{bmatrix} 8 & 3 \\ 2 & 7 \end{bmatrix} \rightarrow \left\{ \mathbf{A}\begin{bmatrix} 3 \\ 2 \end{bmatrix} = 10\begin{bmatrix} 3 \\ 2 \end{bmatrix} \text{ and } \mathbf{A}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = 5\begin{bmatrix} 1 \\ -1 \end{bmatrix} \right.$$

$3 \rightarrow 30$? complex λ

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

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

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

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

(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

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

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

(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 $\mathbf{A}$ is broken into simple actions (just muliply by λ)

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

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

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

I.8 Singular Value Decomposition (SVD)

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

other matrices (**A** is not square, $m \times n$, matrix of data): 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

$$\underbrace{\mathbf{A}\mathbf{v}_1 = \sigma_1\mathbf{u}_1,\ldots,\mathbf{A}\mathbf{v}_r = \sigma_r\mathbf{u}_r}_{\substack{r=\text{rank}(\mathbf{A}) \\ \sigma_1\geq\sigma_2\geq\cdots\geq\sigma_r>0}},\ \underbrace{\mathbf{A}\mathbf{v}_{r+1} = \mathbf{0},\ldots,\mathbf{A}\mathbf{v}_n = \mathbf{0}}_{\substack{(n-r)\ \mathbf{v}\text{'s in } N(\mathbf{A}) \\ (m-r)\ \mathbf{u}\text{'s in } N(\mathbf{A}^T)}} \xleftarrow{\text{rank}(\mathbf{A})=2} \mathbf{A}\begin{cases} m=2: \text{ age, height} \\ n=1000 \end{cases}$$

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

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

- 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

$$\mathbf{A}_k = \sigma_1 \mathbf{u}_1 \mathbf{v}_1{}^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k{}^T \text{ is the best rank } k \text{ approximation to } \mathbf{A}:$$
$$\text{If } \mathbf{B} \text{ has rank k then } \|\mathbf{A} - \mathbf{A}_k\| \leq \|\mathbf{A} - \mathbf{B}\|$$

Proof of SVD

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

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

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

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

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

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

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

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

Example

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

$$\rightarrow \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 45\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix}\begin{bmatrix} -1 \\ 1 \end{bmatrix} = 5\begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix} \rightarrow \mathbf{A}\mathbf{v}_1 = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{3}{\sqrt{2}}\begin{bmatrix} 1 \\ 3 \end{bmatrix} = \sigma_1\mathbf{u}_1 = \sqrt{45}\,\frac{1}{\sqrt{10}}\begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

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

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

$$\sigma_1\mathbf{u}_1\mathbf{v}_1{}^T + \sigma_2\mathbf{u}_2\mathbf{v}_2{}^T = \frac{\sqrt{45}}{\sqrt{20}}\begin{bmatrix} 1 & 3 \\ 1 & 3 \end{bmatrix} + \frac{\sqrt{5}}{\sqrt{20}}\begin{bmatrix} 3 & -3 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix} = \mathbf{A}$$

Questions

- 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 $|\lambda_1| \leq \sigma_1$
- What is the Karhunen-Loève transform and its connection to SVD? stochastic (random) form of PCA

Answers

$(1)\ \mathbf{S} = \mathbf{Q}\Lambda\mathbf{Q}^T = \mathbf{U}\Sigma\mathbf{V}^T \rightarrow \mathbf{U} = \mathbf{V} = \mathbf{Q},\ \Sigma = \Lambda$

$(2)\ \mathbf{Sx} = (-\alpha)\mathbf{x} \leftrightarrow \mathbf{Sv} = \sigma\mathbf{u}:\ \mathbf{u}(\text{or } \mathbf{v}) = -\mathbf{x},\ \sigma = \alpha$

$(3)\ \mathbf{A}^T\mathbf{A} = \mathbf{Q}^T\mathbf{Q} = \mathbf{I} \rightarrow \text{all } \sigma = 1 \rightarrow \mathbf{A} = \mathbf{Q} = \mathbf{U}\Sigma\mathbf{V}^T \leftrightarrow \Sigma = \mathbf{I},\ \mathbf{U} = \mathbf{Q},\ \mathbf{V} = \mathbf{I}$

$(4)\ \left.\begin{array}{l} \|\mathbf{Ax}\| = \left\|\mathbf{U}\Sigma\mathbf{V}^T\mathbf{x}\right\| = \left\|\Sigma\mathbf{V}^T\mathbf{x}\right\| \le \sigma_1\left\|\mathbf{V}^T\mathbf{x}\right\| = \sigma_1\|\mathbf{x}\| \\ \mathbf{Ax} = \lambda\mathbf{x} \rightarrow \|\mathbf{Ax}\| = |\lambda|\|\mathbf{x}\| \end{array}\right\} \rightarrow |\lambda|\|\mathbf{x}\| \le \sigma_1\|\mathbf{x}\|$

$(5)\ \mathbf{A} = \mathbf{xy}^T = \frac{\mathbf{x}}{\|\mathbf{x}\|}\left(\|\mathbf{x}\|\|\mathbf{y}\|\right)\frac{\mathbf{y}^T}{\|\mathbf{y}\|} = \mathbf{u}_1\sigma_1\mathbf{v}_1{}^T,\ \mathbf{Ax} = \left(\mathbf{xy}^T\right)\mathbf{x} = \mathbf{x}\left(\mathbf{y}^T\mathbf{x}\right) = \lambda\mathbf{x} \rightarrow \lambda_1 = \left|\mathbf{y}^T\mathbf{x}\right| \le \sigma_1 = \|\mathbf{x}\|\|\mathbf{y}\|$

(6) KL begins with a covariance matrix $\mathbf{V}$ of a zero-mean random process.

In general $\mathbf{V}$ could be an infinite matrix or a covaraiance function. Then the KL expansion will be an infinite series.

The eigenvectors of $\mathbf{V}$, in order of $\sigma_1{}^2 \ge \sigma_2{}^2 \ge \ldots \ge 0$, are the basis functions $\mathbf{u}_i$ for the KL transform.

The expansion of any vector $\mathbf{v}$ in an orthonormal basis $\mathbf{u}_1, \mathbf{u}_2, \ldots$ is $\mathbf{v} = \sum\left(\mathbf{u}_i{}^T\mathbf{v}\right)\mathbf{u}_i$.

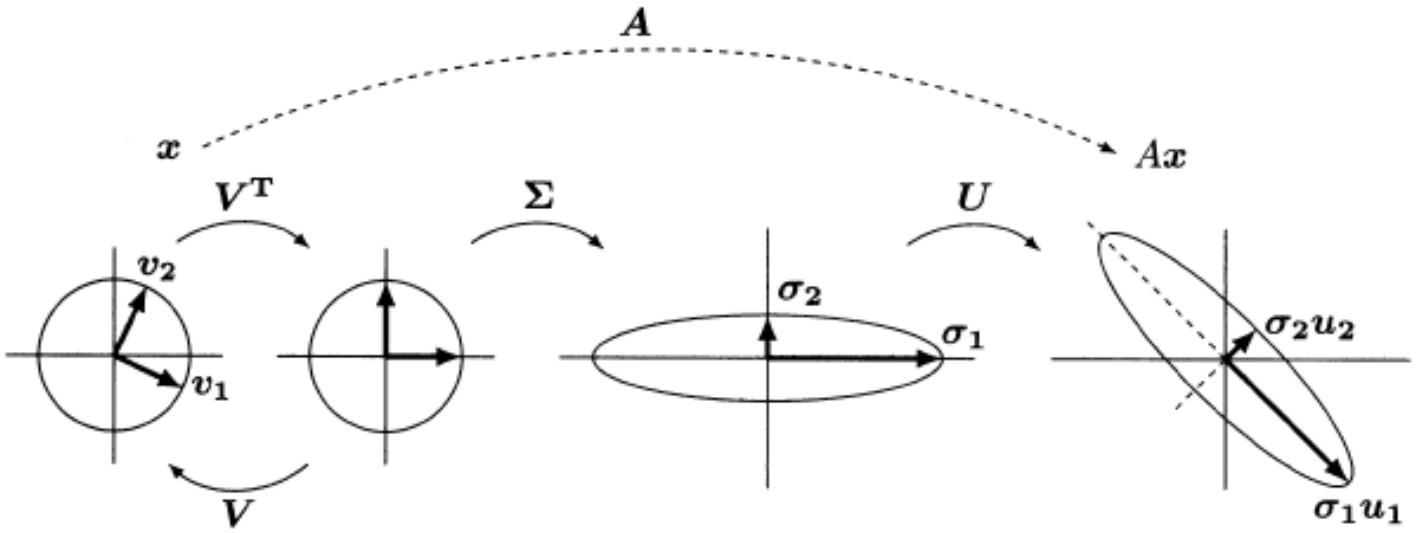
In this stochastic case, that transform decorrelates the random process: the $\mathbf{u}_i$ are independent.

More than that, the ordering of the eigenvalues means that the first k terms, stopping at $\left(\mathbf{u}_k{}^T\mathbf{v}\right)\mathbf{u}_k$,

minimize the expected square error.

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

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

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

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

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

Polar decomposition

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

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

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

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

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

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

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

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

I.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, any norm $||A||$)
- Eckart-Young low rank approximation theorem
  - The norm of $A-A_k$ is below the norm of all other $A-B_k$

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 $\mathbf{A}$

Spectral norm: $\|\mathbf{A}\|_2 = \max_{\mathbf{x}\neq 0} \dfrac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} = \sigma_1 \left(\ell^2 \text{ norm}\right)$

Frobenius norm: $\|\mathbf{A}\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$

Nuclear norm: $\|\mathbf{A}\|_N = \sigma_1 + \cdots + \sigma_r$ (the trace norm)

$\|\mathbf{I}\|_2 = 1 = \|\mathbf{Q}\|_2$, $\|\mathbf{I}\|_F = \sqrt{n} = \|\mathbf{Q}\|_F$, $\|\mathbf{I}\|_N = n = \|\mathbf{Q}\|_N$

$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \xrightarrow{\text{no change in } \|\mathbf{A}\|} \|\mathbf{Q}_1\mathbf{A}\mathbf{Q}_2^T\| = \|\mathbf{Q}_1\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{Q}_2^T\| = \|(\mathbf{Q}_1\mathbf{U})\mathbf{\Sigma}(\mathbf{Q}_2\mathbf{V})^T\| = \|\mathbf{A}\|$

Eckart-Young Theorem: Best approximation by A_k

$$\mathbf{A}=\begin{bmatrix}4&0&0&0\\0&3&0&0\\0&0&2&0\\0&0&0&1\end{bmatrix}\xrightarrow[\text{closest to }\mathbf{A}?]{\text{rank 2 matrix}}\underbrace{\mathbf{A}_2=\begin{bmatrix}4&0&0&0\\0&3&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix}}_{\|\mathbf{A}-\mathbf{A}_2\|_2=2,\ \|\mathbf{A}-\mathbf{A}_2\|_F=\sqrt{5}},\quad \underbrace{\mathbf{B}_2=\begin{bmatrix}3.5&3.5&0&0\\3.5&3.5&0&0\\0&0&1.5&1.5\\0&0&1.5&1.5\end{bmatrix}}_{\mathbf{A}-\mathbf{B}_2\text{ are only 0.5 on the main diagonal}}$$

Eckart-Young in L^2: If $\text{rank}(\mathbf{B}) \le k$, then $\|\mathbf{A}-\mathbf{B}\| = \max_{\mathbf{x}\neq 0}\dfrac{\|(\mathbf{A}-\mathbf{B})\mathbf{x}\|}{\|\mathbf{x}\|} \ge \sigma_{k+1}\left(=\|\mathbf{A}-\mathbf{A}_k\|\right)$

choose $\mathbf{x}\neq\mathbf{0}$ so that $\mathbf{Bx}=\mathbf{0}$ and $\mathbf{x}=\sum_{i=1}^{k+1}c_i\mathbf{v}_i \rightarrow \begin{Bmatrix}\dim N(\mathbf{B})\ge(n-k)\\ \text{combinations of } \mathbf{v}_1 \text{ to } \mathbf{v}_{k+1}:\\ \dim = k+1\end{Bmatrix}$

$$\|(\mathbf{A}-\mathbf{B})\mathbf{x}\|^2 \xrightarrow{\mathbf{Bx}=\mathbf{0}} \|\mathbf{Ax}\|^2 \xrightarrow{\mathbf{Av}_i=\sigma_i\mathbf{u}_i} \left\|\sum_{i=1}^{k+1}c_i\sigma_i\mathbf{u}_i\right\|^2 = \sum_{i=1}^{k+1}c_i^2\sigma_i^2 \ge \left(\sum_{i=1}^{k+1}c_i^2\right)\sigma_{k+1}^2 = \|\mathbf{x}\|^2\sigma_{k+1}^2$$

Eckart-Young Theorem: Best approximation by A_k

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

$\text{rank}(\mathbf{B}) \le k$ is closest to $\mathbf{A} \to \mathbf{B} = \mathbf{A}_k$

$$\mathbf{B} = \mathbf{U}\begin{bmatrix} \mathbf{D} & \mathbf{0} \\ k\times k & \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{V}^T \to \xrightarrow[\text{diagonalize } \mathbf{A},\ \text{rank}(\mathbf{C})\le k]{\mathbf{U},\mathbf{V} \text{ not necessarily}} \mathbf{A} = \mathbf{U}\begin{bmatrix} \mathbf{L}+\mathbf{E}+\mathbf{R} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{bmatrix}\mathbf{V}^T, \mathbf{C} = \mathbf{U}\begin{bmatrix} \mathbf{L}+\mathbf{D}+\mathbf{R} & \mathbf{F} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{V}^T$$

$$\|\mathbf{A}-\mathbf{B}\|_F^2 = \|\mathbf{A}-\mathbf{C}\|_F^2 + \|\mathbf{L}\|_F^2 + \|\mathbf{R}\|_F^2 + \|\mathbf{F}\|_F^2 \xrightarrow[\mathbf{L}=\mathbf{R}=\mathbf{F}=\mathbf{G}=\mathbf{0}]{\text{as small as possible}} \mathbf{A} = \mathbf{U}\begin{bmatrix} \mathbf{E} & \mathbf{0} \\ \mathbf{0} & \mathbf{H} \end{bmatrix}\mathbf{V}^T$$

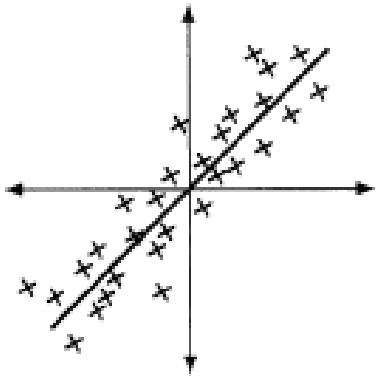
The matrix $\mathbf{D}$ must be the same as $\mathbf{E} = diag(\sigma_1,\ldots,\sigma_k)$

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
    - sum of squares of distances from the mean
  - Covariances: off- diagonal entries of the matrix AA^T
  - Sample covariance matrix: $S=AA^T/(n-1)$
    - One DOF has already been used for mean=0

$$\mathbf{A}=\begin{bmatrix}3 & -4 & 7 & 1 & -4 & -3\\ 7 & -6 & 8 & -1 & -1 & 7\end{bmatrix}\rightarrow \mathbf{S}=\frac{1}{n-1}\mathbf{A}\mathbf{A}^T=\frac{1}{6-1}\begin{bmatrix}20 & 25\\ 25 & 40\end{bmatrix}=\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

$$\mathbf{U}=\begin{bmatrix}-0.5606 & -0.8281\\ -0.8281 & 0.5606\end{bmatrix},\ \mathbf{\Sigma}=\begin{bmatrix}56.9258 & 0\\ 0 & 3.0742\end{bmatrix},\ \mathbf{V}=\begin{bmatrix}-0.5606 & -0.8281\\ -0.8281 & 0.5606\end{bmatrix}$$

- Geometry behind PCA
  - Minimize perpendicular distances: perpendicular least square, orthogonal regression
  - Sum of squared distances from the data points to the line is a minimum

$$\underbrace{\sum_{j=1}^{n}\left\|\mathbf{a}_j\right\|^2}_{\text{fixed by data}} = \underbrace{\sum_{j=1}^{n}\left|\mathbf{a}_j{}^T\mathbf{u}_1\right|^2}_{\rightarrow \mathbf{u}_1{}^T\left(\mathbf{A}\mathbf{A}^T\right)\mathbf{u}_1} + \sum_{j=1}^{n}\left|\mathbf{a}_j{}^T\mathbf{u}_2\right|^2$$

- Linear algebra behind PCA
  - Singular values σ_i and singular vectors u_i of A ← eigenvalues $\sigma_i{}^2$ and eigenvectors of $S=AA^T/(n-1)$
  - Total variance:

$$T = \frac{\left\|\mathbf{A}\right\|_F^2}{n-1} = \frac{\left\|\mathbf{a}_1\right\|^2 + \cdots + \left\|\mathbf{a}_n\right\|^2}{n-1} = \frac{\sigma_1^2 + \cdots + \sigma_r^2}{n-1}$$