EECS 205003 Session 19

Che Lin

Institute of Communications Engineering

Department of Electrical Engineering

Outline

Ch4 Orthogonality

Orthogonal matrices and Gram-Schmidt

Two goals in this SES:

Goal 1: See how orthogonal matrices make calculations of $\hat{\boldsymbol{x}}$, $\boldsymbol{p}$, $\boldsymbol{P}$ easier

Goal 2: See how to obtain orthogonal matrices (Gram-Schmidt process)

Orthonormal vectors

$\boxed{Def}$ The vectors $\boldsymbol{q_1}$, $\boldsymbol{q_2} \cdots \boldsymbol{q_n}$ are orthonormal if

$$\boldsymbol{q_i}^{\mathbf{T}}\boldsymbol{q_j} = \begin{cases} 0 \text{ if } i \neq j \textbf{ (orthogonal)} \\ 1 \text{ if } i = j \textbf{ (unit vectors } \|\mathrm{q_i}\| \textbf{ = 1)} \end{cases}$$

Note: Orthonormal vectors are always independent

Orthonormal matrices

Q is an orthonormal matrix if its columns are orthonormal vectors

(Q can be rectangular)

$\boxed{Fact}$ For orthonormal matrix Q

$$Q^{\mathrm{T}}Q = I$$

Reason:

$$Q^{\mathrm{T}}Q = \begin{bmatrix} - & \boldsymbol{q_1^{\mathrm{T}}} & - \\ & \vdots & \\ - & \boldsymbol{q_n^{\mathrm{T}}} & - \end{bmatrix} \begin{bmatrix} | & & | \\ \boldsymbol{q_1} & \cdots & \boldsymbol{q_n} \\ | & & | \end{bmatrix} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}$$

Note: If Q is square, we call it **orthogonal** matrix

In this case,
$Q^{\mathrm{T}}Q = I \Rightarrow Q^{-1} = Q^{\mathrm{T}}$ **(transpose = inverse)**
To repeat: $Q^{\mathrm{T}}Q = I$ even when Q is rectangular
(Q^{T} is only a left inverse)

For square Q, Q has full rank
$\Rightarrow Q^{-1}$ exist
$\Rightarrow Q^{\mathrm{T}}$ is two-sided inverse
$\Rightarrow$ we also have $QQ^{\mathrm{T}} = I$
(Q also has orthonormal rows)

Import classes of matrix introduced
so far: triangular, diagonal, permutation
symmetric, reduced row echelon,
projection and orthogonal matrices

Ex: Rotation matrix

$$Q = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix}$$

orthogonal + unit vector

$$(cos^2\theta + sin^2\theta = 1)$$

$$Q^{\mathrm{T}} = \underbrace{\begin{bmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{bmatrix}}_{\text{(rotate } \theta\text{)}} = \underbrace{\begin{bmatrix} cos(-\theta) & -sin(-\theta) \\ sin(-\theta) & cos(-\theta) \end{bmatrix}}_{\text{(rotate } -\theta \text{ back)}} = Q^{-1}$$

Ex: permutation matrix **(always orthogonal matrix)**

$$Q = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad Q^{\mathrm{T}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

(Both Q & Q^{T} are orthogonal matrices & $Q^{\mathrm{T}}Q = I$, $Q^{\mathrm{T}}Q = I$)

Ex: $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ not orthogonal matrix

orthogonal but not unit vector

normalize & get $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = Q_1$

Ex: Hadamard matrices

$$Q = \frac{1}{2}\left[\begin{array}{cc} \boxed{\begin{array}{cc} 1 & 1 \\ 1 & -1 \end{array}} & \boxed{\begin{array}{cc} 1 & 1 \\ 1 & -1 \end{array}} \\ \boxed{\begin{array}{cc} 1 & 1 \\ 1 & -1 \end{array}} & \boxed{\begin{array}{cc} -1 & -1 \\ -1 & 1 \end{array}} \end{array}\right] = Q = \frac{1}{2}\begin{bmatrix} Q_1 & Q_1 \\ Q_1 & -Q_1 \end{bmatrix}$$

Ch 4.4 Orthogonal Bases and Gram-Schmidt

Ex: $\begin{bmatrix} 1 & -2 \\ 2 & -1 \\ 2 & 2 \end{bmatrix}$ (rectangular)

(orthogonal but NOT unit vector)

normalize & get

$$Q = \frac{1}{3}\begin{bmatrix} 1 & -2 \\ 2 & -1 \\ 2 & 2 \end{bmatrix} \text{ (not square)}$$

we can add a 3^{rd} column

$$Q = \frac{1}{3}\begin{bmatrix} 1 & -2 & 2 \\ 2 & -1 & -2 \\ 2 & 2 & 1 \end{bmatrix} \text{ (orthogonal matrix)}$$

$\boxed{Fact}$ If Q has orthonormal column $(Q^{\mathrm{T}}Q = I)$

it leaves length unchanged, i.e.,

$\| Q\boldsymbol{x} \| = \| \boldsymbol{x} \| \quad \forall \boldsymbol{x}$ ——— ①

Q also preserves dot products ,i.e.,

$(Q\boldsymbol{x})^{\mathrm{T}}(Q\boldsymbol{y}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$ ——— ②

Reason:

for ①, $\| Q\boldsymbol{x} \|^2 = (Q\boldsymbol{x})^{\mathrm{T}}(Q\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}Q^{\mathrm{T}}Q\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = \| \boldsymbol{x} \|^2$

for ②, $(Q\boldsymbol{x})^{\mathrm{T}}(Q\boldsymbol{y}) = \boldsymbol{x}^{\mathrm{T}}Q^{\mathrm{T}}Q\boldsymbol{y} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}$

Projection using orthonormal bases: Q replaces A

$$A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b} \quad , \quad \boldsymbol{P} = A\hat{\boldsymbol{x}} \quad , \quad P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$$

$$\Downarrow$$

$$Q^{\mathrm{T}}Q\hat{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{b} \quad , \quad \boldsymbol{P} = Q\hat{\boldsymbol{x}} \quad , \quad P = Q(Q^{\mathrm{T}}Q)^{-1}Q^{\mathrm{T}}$$

$$\Downarrow$$

$$\hat{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{b} \quad , \quad \boldsymbol{P} = Q\hat{\boldsymbol{x}} \quad , \quad P = QQ^{\mathrm{T}}$$

$(\hat{x}_i = \boldsymbol{q_i}^{\mathrm{T}}\boldsymbol{b})$ $\quad \Downarrow \quad$ (Projection is just a dot product)

$$\mathbf{P} = \begin{bmatrix} | & & | \\ \boldsymbol{q_1} & \cdots & \boldsymbol{q_n} \\ | & & | \end{bmatrix} \begin{bmatrix} \boldsymbol{q_1}^{\mathrm{T}}\boldsymbol{b} \\ \vdots \\ \boldsymbol{q_n}^{\mathrm{T}}\boldsymbol{b} \end{bmatrix} = \boldsymbol{q_1}(\boldsymbol{q_1}^{\mathrm{T}}\boldsymbol{b}) + \cdots + \boldsymbol{q_n}(\boldsymbol{q_n}^{\mathrm{T}}\boldsymbol{b})$$

Note: When Q is square, columns of Q span the entire space

$\hat{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{b}$, $\boldsymbol{P} = Q\hat{\boldsymbol{x}}$, $P = QQ^{\mathrm{T}}$

(least square solution) $\Downarrow$

$\boldsymbol{x} = Q^{-1}\boldsymbol{b}$, $\boldsymbol{P} = QQ^{-1}\boldsymbol{b}$, $P = I$

(exact solution) $= \boldsymbol{b}$

or

$$\boldsymbol{P} = \boldsymbol{b} = \boldsymbol{q_1}(\boldsymbol{q_1}^{\mathrm{T}}\boldsymbol{b}) + \cdots + \boldsymbol{q_n}(\boldsymbol{q_n}^{\mathrm{T}}\boldsymbol{b})$$

(Projection onto orthonormal basis & assemble it back)

(Foundation for Fourier series !)

Gram-schmidt process

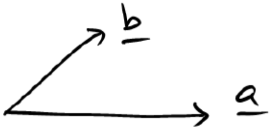
Elimination $\Rightarrow$ make matrix triangular

Gram-Schmidt $\Rightarrow$ make matrix orthonormal

Step 1: construct orthogonal vectors

Step 2: normalize to get orthonormal

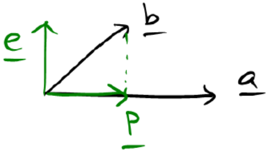
Start with two independent vectors $\boldsymbol{a}, \boldsymbol{b}$



Find orthogonal vectors $\boldsymbol{A}, \boldsymbol{B}$ that span the same space

Q: How do we do that ?

Set $\mathbf{A} = \mathbf{a}$



$\boldsymbol{e} \perp \boldsymbol{P} \Rightarrow \boldsymbol{e} \perp \boldsymbol{a}$, set $\boldsymbol{B} = \boldsymbol{e}$

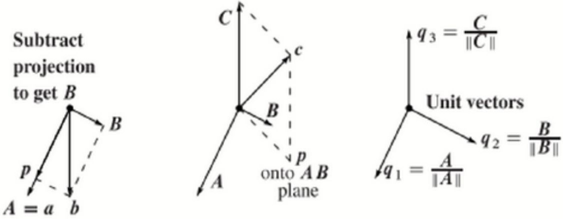
$$\Rightarrow \quad \boldsymbol{B} = \boldsymbol{b} - \boldsymbol{p} = \boldsymbol{b} - \frac{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{A}^{\mathrm{T}}A}\boldsymbol{A}$$

(Check: $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{B} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} - \frac{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{A}^{\mathrm{T}}A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} = 0$)

(indeed orthogonal)

Q: What if we had 3^{rd} independent vector $\boldsymbol{C}$?

Substract components in the direction of $\boldsymbol{A}$ & $\boldsymbol{B}$ from $\boldsymbol{C}$



First project $\boldsymbol{b}$ onto the line though $\boldsymbol{a}$ and find the orthogonal $\boldsymbol{B}$ as $\boldsymbol{b}$-$\boldsymbol{p}$.

Then project $\boldsymbol{c}$ onto the $\boldsymbol{AB}$ plane and find $\boldsymbol{C}$ as $\boldsymbol{c}$-$\boldsymbol{p}$.

Divide by $\|\boldsymbol{A}\|, \|\boldsymbol{B}\|, \|\boldsymbol{C}\|$.

$$\boldsymbol{C} = \boldsymbol{c} - \frac{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{c}}{\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}}\boldsymbol{A} - \frac{\boldsymbol{B}^{\mathrm{T}}\boldsymbol{c}}{\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}}\boldsymbol{B} \quad (\boldsymbol{C} \perp \boldsymbol{A}\ ,\ \boldsymbol{C} \perp \boldsymbol{B})$$

Step 2: $\boldsymbol{q_1} = \dfrac{\boldsymbol{A}}{\|\boldsymbol{A}\|}$, $\boldsymbol{q_2} = \dfrac{\boldsymbol{B}}{\|\boldsymbol{B}\|}$, $\boldsymbol{q_3} = \dfrac{\boldsymbol{C}}{\|\boldsymbol{C}\|}$

(we can keep doing this to construct more orthonormal vectors)

(read Ex5 in textbook p.235)

QR decomposition

Recall: when we studied Elimination, use Elimination matrices to represent the process ⇒ leads to $A = L\mathcal{U}$

$(EA = \mathcal{U} \Rightarrow A = E^{-1}\mathcal{U} = L\mathcal{U})$

A similar equation $A = QR$ relates

A to Q of the Gram-Schmidt process

$(Q^{\mathrm{T}}QR = R \Rightarrow R = Q^{\mathrm{T}}A)$

$$\begin{array}{ccc} A & Q & R \\ \begin{bmatrix} | & | & | \\ \boldsymbol{a_1} & \boldsymbol{a_2} & \boldsymbol{a_3} \\ | & | & | \end{bmatrix} = & \begin{bmatrix} | & | & | \\ \boldsymbol{q_1} & \boldsymbol{q_2} & \boldsymbol{q_3} \\ | & | & | \end{bmatrix} & \begin{bmatrix} \boldsymbol{q_1^{\mathrm{T}} a_1} & \boldsymbol{q_1^{\mathrm{T}} a_2} & \boldsymbol{q_1^{\mathrm{T}} a_3} \\ & \boldsymbol{q_2^{\mathrm{T}} a_2} & \boldsymbol{q_2^{\mathrm{T}} a_3} \\ & & \boldsymbol{q_3^{\mathrm{T}} a_3} \end{bmatrix} \end{array}$$

(R is upper triangular since later **q**'s are chosen to be orthogonal to earlier **a**'s e.g. $\boldsymbol{q_2^{\mathrm{T}} a_1} = 0$, $\boldsymbol{q_3^{\mathrm{T}} a_1} = 0 \cdots$)

This is Gram-schmidt in a nutshell:

— $\boldsymbol{a_1}$ & $\boldsymbol{q_1}$ are along a single line

— $\boldsymbol{a_1}$, $\boldsymbol{a_2}$ & $\boldsymbol{q_1}$, $\boldsymbol{q_2}$ on the same place

($\boldsymbol{a_1}, \boldsymbol{a_2}$ are combinations of $\boldsymbol{q_1}, \boldsymbol{q_2}$)

— $\boldsymbol{a_1}$, $\boldsymbol{a_2}$, $\boldsymbol{a_3}$ & $\boldsymbol{q_1}$, $\boldsymbol{q_2}$, $\boldsymbol{q_3}$ in one subspace

(dim $=$ 3)($\boldsymbol{a_1}$, $\boldsymbol{a_2}$, $\boldsymbol{a_3}$ are combinations of $\boldsymbol{q_1}$, $\boldsymbol{q_2}$, $\boldsymbol{q_3}$)

In general, $\boldsymbol{a_1}, \cdots, \boldsymbol{a_k}$ are combinations of $\boldsymbol{q_1}, \cdots, \boldsymbol{q_k}$ only

$\Rightarrow$ R is upper triangular

Solving least squares problem

$A\mathbf{x} = \mathbf{b}$ (no solution)

$A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ (using QR)

$(QR)^{\mathrm{T}}QR\hat{\boldsymbol{x}} = R^{\mathrm{T}}Q^{\mathrm{T}}\boldsymbol{b}$

$\Rightarrow R^{\mathrm{T}}R\hat{\boldsymbol{x}} = R^{\mathrm{T}}Q^{\mathrm{T}}\boldsymbol{b} \quad or \quad R\hat{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{b} \quad or \quad \hat{\boldsymbol{x}} = R^{-1}Q^{\mathrm{T}}\boldsymbol{b}$

(can be easily sloved using back substitution)