SC1004 Part 2

Lectured by Prof Guan Cuntai
(teaching materials by Prof Chng Eng Siong)

Email: ctguan@ntu.edu.sg

Quiz 2 and Exam:

1. **Quiz 2**

- Coverage : Ch 6 ,7, 8
- Time/Date: Week 13, last lecture time (10:30-11.20am, 17th April 2024)

2. Final Exam

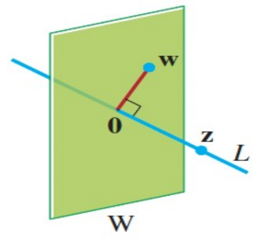
- **Coverage** : Ch 6, 7, 8 (Q3 & Q4)
- **Date/Time**: 2 May 2024 (Thursday), 1.00-3.00pm

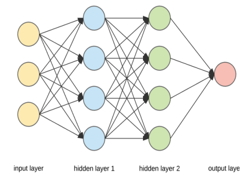
(Ch 9 will not be tested)

Syllabus for Part 2

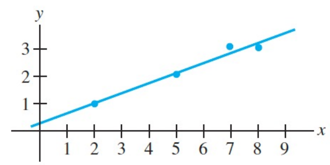
Chapter	Topics	Week (Lecture)	Week (Tut)
6	Orthogonality	8-9	9-10
7	Least Squares	9-10	10-11
8	EigenValue and Eigenvectors	11-12	12-13
9	Singular Value Decomposition (SVD)	13	

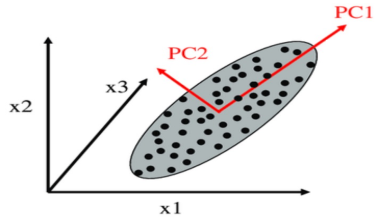
Table 1: schedule





$A\mathbf{x} = \mathbf{b}$





Online Video learning Schedule

https://www.youtube.com/channel/UCBzG5jg3huxiPkCt-Serrjw

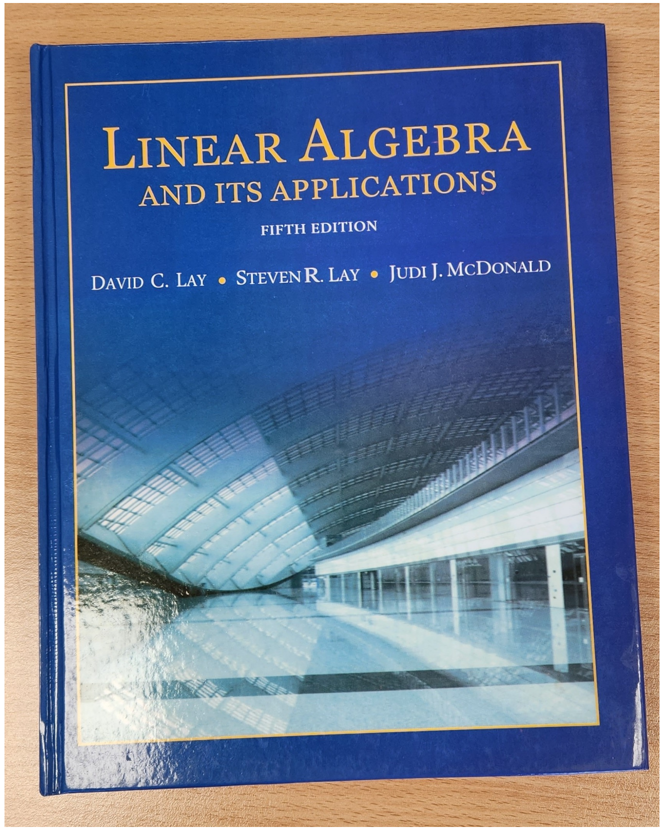
Week	Part	Topic	Notes
8	6.1.1-6.2.3	Orthogonality, Normalization, Dot-Product, Inequalities,	Lecture 1: **6.1.1 - 6.1.3** Lecture 2: **6.1.4 - 6.2.3**
9	6.2.4-6.3.2	Orthogonal/Orthonormal Sets, Basis, Gram Schmidt and QR Decomposition	Lecture 3: **6.2.4** Lecture 4: **6.2.5 – 6.3.2**
10	7.1.1-7.2.1	Least Squares and Normal Eqn, Projection Matrix, Applications	Lecture 5: **7.1.1 – 7.1.3** Lecture 6: **7.1.4 – 7.2.1**
11	8.1.1-8.1.2	Eigenvectors, Eigen-values, Characteristics Eqn	Lecture 7: **8.1.1** Lecture 8: **8.1.2**
12	8.1.3-8.1.5	Diagonalisation, Power of A, Change of basis	Lecture 9: **8.1.3** Lecture 10: **8.1.4 – 8.1.5**
13	9.1.1-9.2	Introduction to SVD and PCA (Not examined in quiz/exam)	Lecture 11: **9.1.1 – 9.2** Lecture 12: **Quiz 2**

How will we conduct the course?

1) Before the lectures, watch the videos according to the schedule in Table 1
   - You can watch past years zoom video recordings at https://www.youtube.com/@linearalgebra1884/playlists?view=50&sort=dd&shelf_id=2

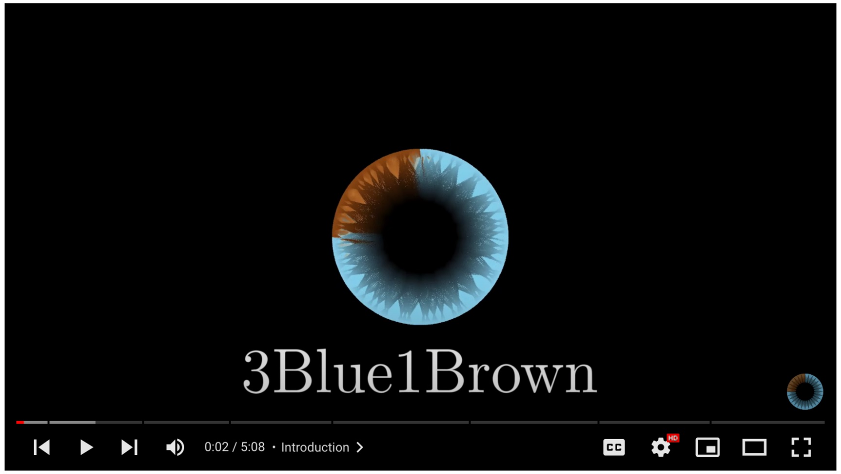
2) During lecture hours –
   - We will summarize the lectures and highlight the key points
   - Q&A.

References



Linear Algebra and Its Applications
by David Lay, Steven Lay, Judi McDonald

3Blue1Brown on YouTube



Essence of linear algebra preview

https://www.youtube.com/playlist?list=PLZHQObOWTQDPD3MizzM2xVFitgF8hE_ab

Lecture (Week 9)

(Chapter 6.2.3- 6.3.3)

Revision

Key points – 6.1.3 Dot Product/Inner Product (2)

- Properties of dot product

Dot products have many of the same algebraic properties as products of real numbers.

THEOREM 3.2.2 *If* $\mathbf{u}$, $\mathbf{v}$, *and* $\mathbf{w}$ *are vectors in* R^n, *and if* k *is a scalar, then:*

(a) $\mathbf{u}\cdot\mathbf{v}=\mathbf{v}\cdot\mathbf{u}$ **[Symmetry property]**

(b) $\mathbf{u}\cdot(\mathbf{v}+\mathbf{w})=\mathbf{u}\cdot\mathbf{v}+\mathbf{u}\cdot\mathbf{w}$ **[Distributive property]**

(c) $k(\mathbf{u}\cdot\mathbf{v})=(k\mathbf{u})\cdot\mathbf{v}$ **[Homogeneity property]**

(d) $\mathbf{v}\cdot\mathbf{v}\geq 0$ *and* $\mathbf{v}\cdot\mathbf{v}=0$ *if and only if* $\mathbf{v}=\mathbf{0}$ **[Positivity property]**

- Transformation on dot product
  - $A\boldsymbol{u}\cdot\boldsymbol{v}=\boldsymbol{u}\cdot A^T\boldsymbol{v}$
  - $\boldsymbol{u}\cdot A\boldsymbol{v}=A^T\boldsymbol{u}\cdot\boldsymbol{v}$

  (Using $\boldsymbol{u}\cdot\boldsymbol{v}=\boldsymbol{u}^T\boldsymbol{v}$, and $(AB)^T=B^TA^T$ to derive)

Explanation to transformation on dot product:

- Let's write the dot product in matrix form:
  $$A\boldsymbol{u}\cdot\boldsymbol{v}=(A\boldsymbol{u})^T\boldsymbol{v}$$
- Using $(AB)^T=B^TA^T$
  $$(A\boldsymbol{u})^T\boldsymbol{v}=(\boldsymbol{u}^TA^T)\boldsymbol{v}$$
- Using the distributive property of matrix
  $$(\boldsymbol{u}^TA^T)\boldsymbol{v}=\boldsymbol{u}^T(A^T\boldsymbol{v})$$
- Write back to dot product format
  $$\boldsymbol{u}^T(A^T\boldsymbol{v})=\boldsymbol{u}\cdot A^T\boldsymbol{v}$$

So we get: $A\boldsymbol{u}\cdot\boldsymbol{v}=\boldsymbol{u}\cdot A^T\boldsymbol{v}$

Revision

Key points – 6.2.2 Orthogonal Projection

- Projection theorem (projection from one vector to another)
  - Project vector $\boldsymbol{y}$ on to $\boldsymbol{u}$: $\hat{\boldsymbol{y}} = Proj_u\ \boldsymbol{y} = \frac{\boldsymbol{y}\cdot\boldsymbol{u}}{\boldsymbol{u}\cdot\boldsymbol{u}}\boldsymbol{u}$
  - Residual: $\boldsymbol{z} = \boldsymbol{y} - \hat{\boldsymbol{y}} = \boldsymbol{y} - \frac{\boldsymbol{y}\cdot\boldsymbol{u}}{\boldsymbol{u}\cdot\boldsymbol{u}}\boldsymbol{u}$
- Explain
  1) Geometric approach:

     $\hat{\boldsymbol{y}}$ is on the line of $\boldsymbol{u}$ with the length of $\|\hat{\boldsymbol{y}}\|$

     $$\hat{\boldsymbol{y}} = \|\hat{\boldsymbol{y}}\|\frac{\boldsymbol{u}}{\|\boldsymbol{u}\ \|}$$

     From triangle (see figure on the right): $\|\hat{\boldsymbol{y}}\|$= $\|\boldsymbol{y}\|cos\theta$

     From $\boldsymbol{y}\cdot\boldsymbol{u} = \|\boldsymbol{y}\|\|\boldsymbol{u}\|cos\theta$, we get: $\|\boldsymbol{y}\|cos\theta = \frac{\boldsymbol{y}\cdot\boldsymbol{u}}{\|\boldsymbol{u}\|}$

     So, we get $\hat{\boldsymbol{y}} = \|\hat{\boldsymbol{y}}\|\frac{\boldsymbol{u}}{\|\boldsymbol{u}\ \|} = \|\boldsymbol{y}\|cos\theta\frac{\boldsymbol{u}}{\|\boldsymbol{u}\ \|} = \frac{\boldsymbol{y}\cdot\boldsymbol{u}}{\|\boldsymbol{u}\|}\frac{\boldsymbol{u}}{\|\boldsymbol{u}\ \|} = \frac{\boldsymbol{y}\cdot\boldsymbol{u}}{\|\boldsymbol{u}\|^2}\ \boldsymbol{u} = \frac{\boldsymbol{y}\cdot\boldsymbol{u}}{\boldsymbol{u}\cdot\boldsymbol{u}}\boldsymbol{u}$

  2) Orthogonal approach:

     As $\hat{\boldsymbol{y}}$ is on the line of $\boldsymbol{u}$, so $\hat{\boldsymbol{y}} = c\boldsymbol{u}$ (c is a scalar to be found)

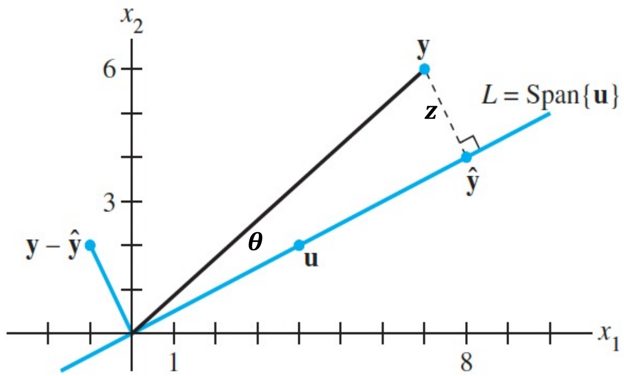
     $$\hat{\boldsymbol{y}} = \boldsymbol{y} - \boldsymbol{z} = c\boldsymbol{u}$$

     Take the dot product with $\boldsymbol{u}$ on both sides: $(\boldsymbol{y} - \boldsymbol{z})\cdot\boldsymbol{u} = c\boldsymbol{u}\cdot\boldsymbol{u}$

     We get: $c\boldsymbol{u}\cdot\boldsymbol{u} = \boldsymbol{y}\cdot\boldsymbol{u} - \boldsymbol{z}\cdot\boldsymbol{u} = \boldsymbol{y}\cdot\boldsymbol{u}$ ($\boldsymbol{z}$ is orthogonal to $\boldsymbol{u}$!) ➔ $c = \frac{\boldsymbol{y}\cdot\boldsymbol{u}}{\boldsymbol{u}\cdot\boldsymbol{u}}$

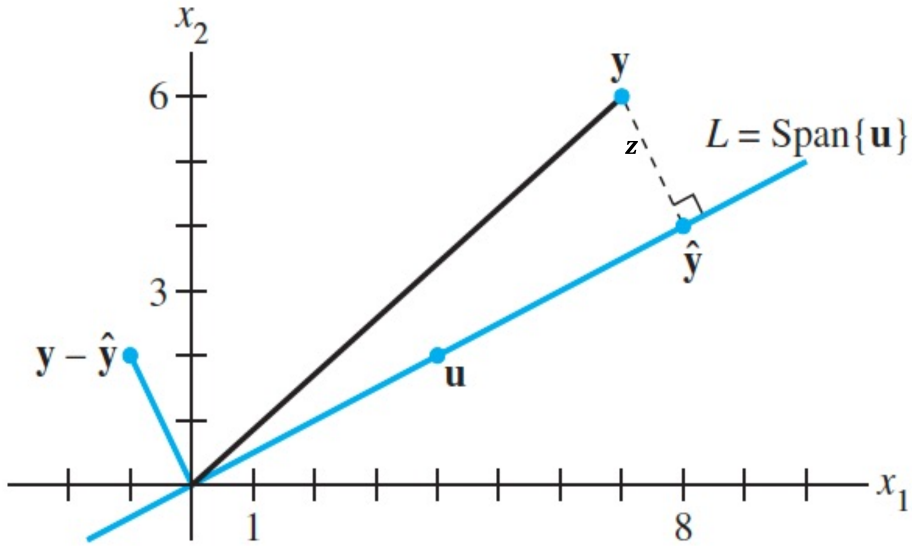
     So we also get: $\hat{\boldsymbol{y}} = c\boldsymbol{u} = \frac{\boldsymbol{y}\cdot\boldsymbol{u}}{\boldsymbol{u}\cdot\boldsymbol{u}}\boldsymbol{u}$

Project $\boldsymbol{y} = \begin{bmatrix}7\\6\end{bmatrix}$ onto vector $\boldsymbol{u} = \begin{bmatrix}2\\1\end{bmatrix}$, and sketch $\hat{\boldsymbol{y}}$.



Example

Project $\boldsymbol{y} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$ onto vector $\boldsymbol{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$



Revision

Key points – 6.2.3 Orthogonal Decomposition

- Project a vector $\boldsymbol{y}$ on to subpace spanned by $\{\boldsymbol{u}_1, \boldsymbol{u}_2 \cdots \boldsymbol{u}_p\}$ in R^n
  - Let W be a subspace of R^n. Then each $\boldsymbol{y}$ in R^n can be written **uniquely** in the form:

$$\boldsymbol{y} = \widehat{\boldsymbol{y}} + \boldsymbol{z}$$

  where $\widehat{\boldsymbol{y}}$ is in W and residual $\boldsymbol{z}$ is in $W^{\perp}$. If $\{\boldsymbol{u}_1, \boldsymbol{u}_2 \cdots \boldsymbol{u}_p\}$ is any orthogonal basis of W, then

$$\widehat{\boldsymbol{y}} = Proj_W \boldsymbol{y} = \frac{\boldsymbol{y} \cdot \boldsymbol{u}_1}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1} \boldsymbol{u}_1 + \cdots + \frac{\boldsymbol{y} \cdot \boldsymbol{u}_p}{\boldsymbol{u}_p \cdot \boldsymbol{u}_p} \boldsymbol{u}_p$$

- Explain:
  - Since $\widehat{\boldsymbol{y}}$ is in the subspace W spanned by $\{\boldsymbol{u}_1, \boldsymbol{u}_2 \cdots \boldsymbol{u}_p\}$, we can write

$$\widehat{\boldsymbol{y}} = \boldsymbol{y} - \boldsymbol{z} = c_1 \boldsymbol{u}_1 + c_2 \boldsymbol{u}_2 + \cdots + c_p \boldsymbol{u}_p$$

  Take dot product with $\boldsymbol{u}_i$ on both sides:

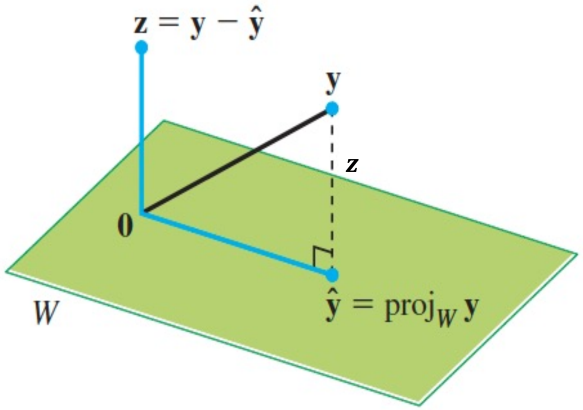
$$(\boldsymbol{y} - \boldsymbol{z}) \cdot \boldsymbol{u}_i = (c_1 \boldsymbol{u}_1 + c_2 \boldsymbol{u}_2 + \cdots + c_p \boldsymbol{u}_p) \cdot \boldsymbol{u}_i,\ i = 1, \cdots, p$$

  Since $\boldsymbol{u}_i \cdot \boldsymbol{u}_j = 0, if\ i \neq j$, and $\boldsymbol{z} \cdot \boldsymbol{u}_i = 0$, so we have

$$c_i \boldsymbol{u}_i \cdot \boldsymbol{u}_i = (\boldsymbol{y} - \boldsymbol{z}) \cdot \boldsymbol{u}_i = \boldsymbol{y} \cdot \boldsymbol{u}_i - \boldsymbol{z} \cdot \boldsymbol{u}_i = \boldsymbol{y} \cdot \boldsymbol{u}_i$$

$$\therefore \qquad c_i = \frac{\boldsymbol{y} \cdot \boldsymbol{u}_i}{\boldsymbol{u}_i \cdot \boldsymbol{u}_i}$$

  Finally: $\widehat{\boldsymbol{y}} = c_1 \boldsymbol{u}_1 + c_2 \boldsymbol{u}_2 + \cdots + c_p \boldsymbol{u}_p = \frac{\boldsymbol{y} \cdot \boldsymbol{u}_1}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1} \boldsymbol{u}_1 + \frac{\boldsymbol{y} \cdot \boldsymbol{u}_2}{\boldsymbol{u}_2 \cdot \boldsymbol{u}_2} \boldsymbol{u}_2 + \cdots + \frac{\boldsymbol{y} \cdot \boldsymbol{u}_p}{\boldsymbol{u}_p \cdot \boldsymbol{u}_p} \boldsymbol{u}_p$



Key points – 6.2.4 Orthonormal Sets

- Definition
  - If $\{\boldsymbol{u}_1, \boldsymbol{u}_2 \cdots \boldsymbol{u}_p\}$ is called an **orthonormal basis** for subspace W if the basis vectors are orthogonal with unit length ($\boldsymbol{u}_i \cdot \boldsymbol{u}_j = 0, if\ i \neq j$, and $\|\boldsymbol{u}\| = 1$)
  - Let $U_{n\times p} = [\boldsymbol{u}_1\ \boldsymbol{u}_2 \cdots \boldsymbol{u}_p]$, $\boldsymbol{u}_i \in R^n$

    Then, $U^T U = I$ (I is a $p\times p$ identify matrix).

    Explain: $U^T = \begin{bmatrix} \boldsymbol{u}_1{}^T \\ \vdots \\ \boldsymbol{u}_p{}^T \end{bmatrix}$ is a $p\times n$ matrix, So $U^T U = \begin{bmatrix} \boldsymbol{u}_1{}^T \\ \vdots \\ \boldsymbol{u}_p{}^T \end{bmatrix} [\boldsymbol{u}_1\ \boldsymbol{u}_2 \cdots \boldsymbol{u}_p] = \begin{bmatrix} \boldsymbol{u}_1{}^T\boldsymbol{u}_1 & \cdots & \boldsymbol{u}_1{}^T\boldsymbol{u}_p \\ \vdots & \ddots & \vdots \\ \boldsymbol{u}_p{}^T\boldsymbol{u}_1 & \cdots & \boldsymbol{u}_p{}^T\boldsymbol{u}_p \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} = I$

- Properties

  a. $\|U\boldsymbol{x}\| = \|\boldsymbol{x}\|$ (preserve the length of vector)

  b. $U\boldsymbol{x} \cdot U\boldsymbol{y} = \boldsymbol{x} \cdot \boldsymbol{y}$

  c. $U\boldsymbol{x} \cdot U\boldsymbol{y} = \boldsymbol{0}$, if and only if $\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{0}$

- Re-write projection equation using U: $\hat{\boldsymbol{y}} = Proj_W\boldsymbol{y} = UU^T\boldsymbol{y}$

  Explain: $\hat{\boldsymbol{y}} = Proj_W\boldsymbol{y} = \frac{\boldsymbol{y}\cdot\boldsymbol{u}_1}{\boldsymbol{u}_1\cdot\boldsymbol{u}_1}\boldsymbol{u}_1 + \cdots + \frac{\boldsymbol{y}\cdot\boldsymbol{u}_p}{\boldsymbol{u}_p\cdot\boldsymbol{u}_p}\boldsymbol{u}_p = (\boldsymbol{y}\cdot\boldsymbol{u}_1)\boldsymbol{u}_1 + \cdots + (\boldsymbol{y}\cdot\boldsymbol{u}_p)\boldsymbol{u}_p$

  $$= (\boldsymbol{u}_1{}^T\boldsymbol{y})\boldsymbol{u}_1 + \cdots + (\boldsymbol{u}_p{}^T\boldsymbol{y})\boldsymbol{u}_p = [\boldsymbol{u}_1\ \boldsymbol{u}_2 \cdots \boldsymbol{u}_p] \begin{bmatrix} \boldsymbol{u}_1{}^T\boldsymbol{y} \\ \vdots \\ \boldsymbol{u}_p{}^T\boldsymbol{y} \end{bmatrix} = [\boldsymbol{u}_1\ \boldsymbol{u}_2 \cdots \boldsymbol{u}_p] \begin{bmatrix} \boldsymbol{u}_1{}^T \\ \vdots \\ \boldsymbol{u}_p{}^T \end{bmatrix} \boldsymbol{y} = UU^T\boldsymbol{y}$$

- Note: if U is a square, it is called "**orthogonal matrix**". In this case, $U^{-1} = U^T$

Key points – 6.2.5 Orthogonal Decomposition.

- Geometric interpretation of the orthogonal projection (see figure right-top)
- The best approximation theorem (see figure right-bottom)

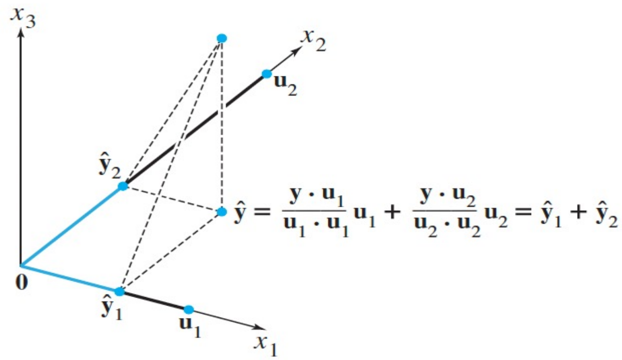
$$\|\boldsymbol{y}-\hat{\boldsymbol{y}}\| < \|\boldsymbol{y}-\boldsymbol{v}\|$$

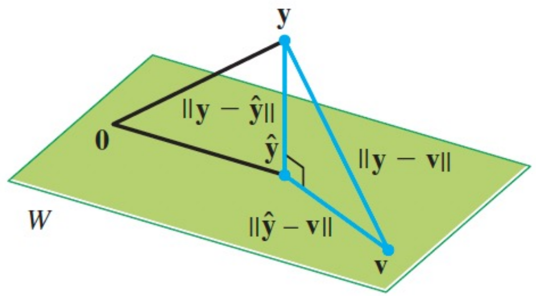
  $\hat{\boldsymbol{y}}$ is the orthogonal projection of $\boldsymbol{y}$ onto W. $\boldsymbol{v}$ is any vector in W **distinct** from $\hat{\boldsymbol{y}}$.

  Explain: $\boldsymbol{y}-\boldsymbol{v} = (\boldsymbol{y}-\hat{\boldsymbol{y}}) + (\hat{\boldsymbol{y}}-\boldsymbol{v})$,

  So, according to Pythagorean theorem: $\|\boldsymbol{y}-\boldsymbol{v}\|^2 = \|\boldsymbol{y}-\hat{\boldsymbol{y}}\|^2 + \|\hat{\boldsymbol{y}}-\boldsymbol{v}\|^2 \Rightarrow \|\boldsymbol{y}-\hat{\boldsymbol{y}}\|^2 = \|\boldsymbol{y}-\boldsymbol{v}\|^2 - \|\hat{\boldsymbol{y}}-\boldsymbol{v}|$

  Therefore we have: $\|\boldsymbol{y}-\hat{\boldsymbol{y}}\| < \|\boldsymbol{y}-\boldsymbol{v}\|$

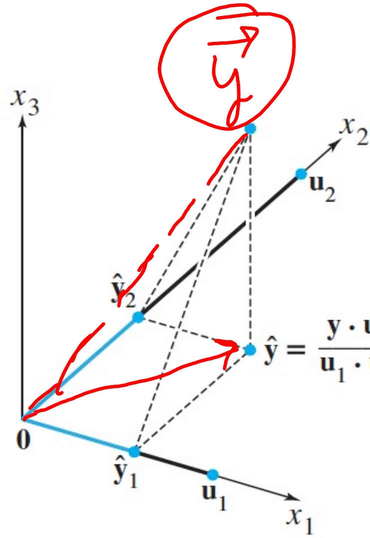




- Example: $\boldsymbol{u}_1 = \begin{bmatrix}3\\1\\1\end{bmatrix}$, $\boldsymbol{u}_2 = \begin{bmatrix}-1\\2\\1\end{bmatrix}$, $\boldsymbol{y} = \begin{bmatrix}3\\0\\1\end{bmatrix}$
  - $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ are orthogonal?
  - $\hat{\boldsymbol{y}} = \frac{[3\ 0\ 1]\begin{bmatrix}3\\1\\1\end{bmatrix}}{9+1+1}\boldsymbol{u}_1 + \frac{[3\ 0\ 1]\begin{bmatrix}-1\\2\\1\end{bmatrix}}{1+4+1}\boldsymbol{u}_2 = \frac{9+0+1}{11}\boldsymbol{u}_1 + \frac{-3+0+1}{1+4+1}\boldsymbol{u}_2 = \frac{10}{11}\begin{bmatrix}3\\1\\1\end{bmatrix} - \frac{1}{3}\begin{bmatrix}-1\\2\\1\end{bmatrix} = \begin{bmatrix}101/33\\8/33\\19/33\end{bmatrix}$

- What if $p = n$? That is, when W is the full space, what will be $\hat{\boldsymbol{y}} = Proj_W\boldsymbol{y}$?

R^n $\quad W = Span(\vec{u}_1, \vec{u}_2 \cdots \vec{u}_p)$



$\hat{y} = \vec{y}$

(1) $\{\boldsymbol{u}_1, \boldsymbol{u}_2 \cdots \boldsymbol{u}_p\}$ is an **orthogonal basis** :

$$\hat{\boldsymbol{y}} = Proj_W \boldsymbol{y} = \frac{\boldsymbol{y}\cdot\boldsymbol{u}_1}{\boldsymbol{u}_1\cdot\boldsymbol{u}_1}\boldsymbol{u}_1 + \cdots + \frac{\boldsymbol{y}\cdot\boldsymbol{u}_p}{\boldsymbol{u}_p\cdot\boldsymbol{u}_p}\boldsymbol{u}_p$$

(2) $\{\boldsymbol{v}_1, \boldsymbol{v}_2 \cdots \boldsymbol{v}_p\}$ is an **orthonormal basis** :

$UU^T = I$

$$\hat{\boldsymbol{y}} = Proj_W \boldsymbol{y} = (\boldsymbol{y}\cdot\boldsymbol{v}_1)\boldsymbol{v}_1 + \cdots + (\boldsymbol{y}\cdot\boldsymbol{v}_p)\boldsymbol{v}_p$$

$$= (\boldsymbol{v}_1{}^T\boldsymbol{y})\boldsymbol{v}_1 + \cdots + (\boldsymbol{v}_p{}^T\boldsymbol{y})\boldsymbol{v}_p = \underbrace{[\boldsymbol{v}_1\ \boldsymbol{v}_2 \cdots \boldsymbol{v}_p]}_{A}\underbrace{\begin{bmatrix}\boldsymbol{v}_1{}^T\boldsymbol{y}\\ \vdots \\ \boldsymbol{v}_p{}^T\boldsymbol{y}\end{bmatrix}}_{x} = \underbrace{[\boldsymbol{v}_1\ \boldsymbol{v}_2 \cdots \boldsymbol{v}_p]\begin{bmatrix}\boldsymbol{v}_1{}^T\\ \vdots \\ \boldsymbol{v}_p{}^T\end{bmatrix}}\boldsymbol{y} = \boxed{UU^T\boldsymbol{y}}$$

$\hat{y} = \vec{y}$

A $\quad$ A $\quad$ Q $\quad$ x

$W = Span\{\underbrace{\vec{x}_1, \cdots \vec{x}_p}_{A}\} \longrightarrow \{\vec{u}_1 \cdots \vec{u}_p\}$

$\downarrow$ normalizing

V $\quad \{\vec{v}_1 \cdots \vec{v}_p\}$

$W = Span\{\vec{u}_1, \vec{u}_2 \cdots \vec{u}_p\}$
$= Span\{\vec{v}_1, \vec{v}_2 \cdots \vec{v}_p\}$

Key points – 6.3.1 QR Factorization (why)

- Definition of QR factorization
  - Given an $m \times n$ matrix A
  - A can be factorized as $A = QR$,
    - $Q\ (m \times n)$ has orthonormal columns (meaning $Q^T Q = I$)
    - $R\ (n \times n)$ is an "up-triangle" square matrix
- Why QR factorization is useful
  - After factorize A into Q and R, we can easily find the solution for system: $A\boldsymbol{x} = \boldsymbol{b}$ using back substitute only
    - ➢ Explain: $A\boldsymbol{x} = \boldsymbol{b}$ ➔ $QR\boldsymbol{x} = \boldsymbol{b}$ ➔ $Q^T QR\boldsymbol{x} = Q^T\boldsymbol{b}$ ➔ $R\boldsymbol{x} = Q^T\boldsymbol{b} = \boldsymbol{y}$

$$R\boldsymbol{x} = \boldsymbol{y}: \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \\ 0 & 0 & r_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix},$$

$r_{33}\, x_3 = y_3$ ➔ $x_3 = y_3 / r_{33}$

$r_{22}\, x_2 + r_{23}\, x_3 = y_2$ ➔ $x_2 = (y_2 - r_{23}\, x_3)\, /\, r_{22}$

$r_{11}\, x_1 + r_{12}\, x_2 + r_{13}\, x_3 = y_1$ ➔ $x_1 = (y_1 - r_{12}\, x_2 - r_{13}\, x_3)\, /\, r_{11}$

- $\boldsymbol{QR}$ **factorization is an important tool for finding a Least Square solution ($\hat{\boldsymbol{x}} = R^{-1}Q^T\boldsymbol{b}$, in week 10)**

U

$[\vec{u}_1 \cdots \vec{u}_p]^T [\vec{u}_1 \cdots \vec{u}_p]$

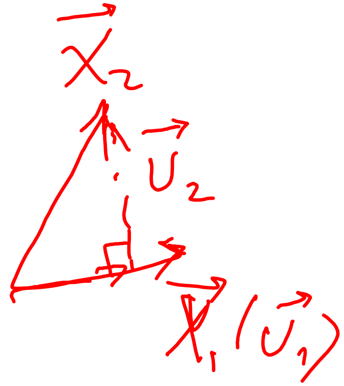
$= \begin{bmatrix} \vec{u}_1^T \\ \vec{u}_p^T \end{bmatrix} [\vec{u}_1 \cdots \vec{u}_p] = \begin{bmatrix} \vec{u}_1^T \vec{u}_1 & 0 & 0 & 0 & 0 \\ 0 & \ddots & & & \\ 0 & & & & \vec{u}_p^T u_p \end{bmatrix}$

$= I$

$Ax = b$

$Ax = b$

Key points – 6.3.2 QR Factorization (how)



- How do we find Q and R from A – Gram–Schmidt Approach
  - Given any set of p independent columns (basis of non-zero subspace W in R^n): $\{\boldsymbol{x}_1, \boldsymbol{x}_2 \cdots \boldsymbol{x}_p\} \in R^n$ $(A = [\boldsymbol{x}_1\ \boldsymbol{x}_2 \cdots \boldsymbol{x}_p])$
  - Define the following orthogonal set $\{\boldsymbol{v}_1, \boldsymbol{v}_2 \cdots \boldsymbol{v}_p\}$:

$$\boldsymbol{v}_1 = \boldsymbol{x}_1$$

$$\boldsymbol{v}_2 = \boldsymbol{x}_2 - \frac{\boldsymbol{x}_2 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 \quad \text{(so } \boldsymbol{v}_2 \text{ is orthogonal to } \boldsymbol{v}_1\text{)}$$

$$\boldsymbol{v}_3 = \boldsymbol{x}_3 - \frac{\boldsymbol{x}_3 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 - \frac{\boldsymbol{x}_3 \cdot \boldsymbol{v}_2}{\boldsymbol{v}_2 \cdot \boldsymbol{v}_2}\boldsymbol{v}_2 \quad \text{(so } \boldsymbol{v}_3 \text{ is orthogonal to } \boldsymbol{v}_2, \boldsymbol{v}_1\text{)}$$

$$\vdots$$

$$\boldsymbol{v}_p = \boldsymbol{x}_p - \sum_{i=1}^{p-1} \frac{\boldsymbol{x}_p \cdot \boldsymbol{v}_i}{\boldsymbol{v}_i \cdot \boldsymbol{v}_i}\boldsymbol{v}_i \quad \text{(so } \boldsymbol{v}_3 \text{ is orthogonal to } \boldsymbol{v}_{p-1}, \cdots, \boldsymbol{v}_2, \boldsymbol{v}_1\text{)}$$

  - Form a orthonormal basis from $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_p\}$
    - $Q = [\boldsymbol{u}_1\ \boldsymbol{u}_2 \cdots \boldsymbol{u}_p] = [\frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|}\ \frac{\boldsymbol{v}_2}{\|\boldsymbol{v}_2\|} \cdots \frac{\boldsymbol{v}_2}{\|\boldsymbol{v}_2\|}]$

  - Finally, find R
    - Since $A = QR$ and $Q^TQ = I$,

    from $Q^TA = Q^TQR$, we find $R = Q^TA$
    - $R = [\boldsymbol{u}_1\ \boldsymbol{u}_2 \cdots \boldsymbol{u}_p]^T[\boldsymbol{x}_1\ \boldsymbol{x}_2 \cdots \boldsymbol{x}_p]$

$$= \begin{bmatrix} \boldsymbol{u}_1{}^T \\ \vdots \\ \boldsymbol{u}_p{}^T \end{bmatrix} [\boldsymbol{x}_1 \cdots \boldsymbol{x}_p] = \begin{bmatrix} \boldsymbol{u}_1{}^T\boldsymbol{x}_1 & \cdots & \boldsymbol{u}_1{}^T\boldsymbol{x}_p \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{u}_p{}^T\boldsymbol{x}_p \end{bmatrix}$$

where

$$\boldsymbol{u}_2{}^T\boldsymbol{x}_1 = \frac{\boldsymbol{v}_2{}^T}{\|\boldsymbol{v}_2\|}\boldsymbol{v}_1 = 0$$

$$\boldsymbol{u}_3{}^T\boldsymbol{x}_1 = \frac{\boldsymbol{v}_3{}^T}{\|\boldsymbol{v}_3\|}\boldsymbol{v}_1 = 0,\ \boldsymbol{u}_3{}^T\boldsymbol{x}_2 = \frac{\boldsymbol{v}_3{}^T}{\|\boldsymbol{v}_3\|}\boldsymbol{v}_2 = 0$$

….

$$\boldsymbol{v}_1 = \boldsymbol{x}_1$$

$$\boldsymbol{v}_2 = \boldsymbol{x}_2 - \frac{\boldsymbol{x}_2\cdot\boldsymbol{v}_1}{\boldsymbol{v}_1\cdot\boldsymbol{v}_1}\boldsymbol{v}_1$$ (so $\boldsymbol{v}_2$ is orthogonal to $\boldsymbol{v}_1$)

$$\boldsymbol{v}_3 = \boldsymbol{x}_3 - \frac{\boldsymbol{x}_3\cdot\boldsymbol{v}_1}{\boldsymbol{v}_1\cdot\boldsymbol{v}_1}\boldsymbol{v}_1 - \frac{\boldsymbol{x}_3\cdot\boldsymbol{v}_2}{\boldsymbol{v}_2\cdot\boldsymbol{v}_2}\boldsymbol{v}_2$$ (so $\boldsymbol{v}_3$ is orthogonal to $\boldsymbol{v}_2, \boldsymbol{v}_1$)

$\vdots$

$$\boldsymbol{v}_p = \boldsymbol{x}_p - \sum_{i=1}^{p-1}\frac{\boldsymbol{x}_p\cdot\boldsymbol{v}_i}{\boldsymbol{v}_i\cdot\boldsymbol{v}_i}\boldsymbol{v}_i$$ (so $\boldsymbol{v}_3$ is orthogonal to $\boldsymbol{v}_{p-1}, \cdots, \boldsymbol{v}_2, \boldsymbol{v}_1$)

- $R = [\boldsymbol{u}_1\ \boldsymbol{u}_2 \cdots \boldsymbol{u}_p]^T[\boldsymbol{x}_1\ \boldsymbol{x}_2 \cdots \boldsymbol{x}_p]$

$$= \begin{bmatrix} \boldsymbol{u}_1{}^T \\ \vdots \\ \boldsymbol{u}_p{}^T \end{bmatrix}[\boldsymbol{x}_1 \cdots \boldsymbol{x}_p] = \begin{bmatrix} \boldsymbol{u}_1{}^T\boldsymbol{x}_1 & \cdots & \boldsymbol{u}_1{}^T\boldsymbol{x}_p \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{u}_p{}^T\boldsymbol{x}_p \end{bmatrix}$$

$\vec{u}_3 \cdot \vec{x}_1 \rightarrow \vec{v}_3 \cdot \vec{x}_1 = \vec{v}_3 \cdot \vec{v}_1 = 0$

$\vec{u}_3 \cdot \vec{x}_2$

$\vec{u}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|} = \frac{\vec{x}_1}{\|\vec{x}_1\|}$ $\quad \|\vec{x}_1\|^2$

$\Rightarrow \vec{v}_3 \cdot (\vec{v}_2 + c\vec{v}_1) = 0$

$\vec{u}_2 \cdot \vec{x}_1 \Rightarrow \vec{v}_2 \cdot \vec{x}_1 = 0 = \vec{v}_2 \cdot \vec{v}_1 = 0$

$$R = \begin{bmatrix} \vec{u}_1{}^T \\ \vdots \\ \vec{u}_p{}^T \end{bmatrix}[\vec{x}_1 \cdots \vec{x}_p]$$

$$\begin{bmatrix} \vec{u}_1{}^T\vec{x}_1 & \vec{u}_1{}^T\vec{x}_2 & \cdots & \vec{u}_1{}^T\vec{x}_p \\ \vec{u}_2{}^T\vec{x}_1 & \vec{u}_2{}^T\vec{x}_2 & \cdots & \\ \vec{u}_3{}^T\vec{x}_1 & \vec{u}_3{}^T\vec{x}_2 & \cdots & \\ \cdots & 0 & & \end{bmatrix}$$

Key points – 6.3.2 QR Factorization (how)

- Example: $A = \begin{bmatrix} 3 & 8 \\ 0 & 5 \\ -1 & -6 \end{bmatrix}$

  $\boldsymbol{x}_1 = \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}$, $\boldsymbol{x}_2 = \begin{bmatrix} 8 \\ 5 \\ -6 \end{bmatrix}$, $\boldsymbol{x}_1$and $\boldsymbol{x}_2$ are independent

  Find $\boldsymbol{v}_i$: $\boldsymbol{v}_1 = \boldsymbol{x}_1 = \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}$, $\boldsymbol{v}_2 = \boldsymbol{x}_2 - \frac{\boldsymbol{x}_2 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1} \boldsymbol{v}_1 = \begin{bmatrix} 8 \\ 5 \\ -6 \end{bmatrix} - \frac{[8 \quad 5 \quad -6] \cdot \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}}{9+1} \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ 5 \\ -3 \end{bmatrix}$

  Verify: $\boldsymbol{v}_1 \cdot \boldsymbol{v}_2 = [3 \quad 0 \quad -1] \cdot \begin{bmatrix} -1 \\ 5 \\ -3 \end{bmatrix} = 0$

  Normalize $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$: $\|\boldsymbol{v}_1\| = \sqrt{10}$, $\|\boldsymbol{v}_2\| = \sqrt{35}$,

  Q^T

  So we get $\boldsymbol{u}_1 = \frac{\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|} = \begin{bmatrix} 3/\sqrt{10} \\ 0 \\ -1/\sqrt{10} \end{bmatrix}$, $\boldsymbol{u}_2 = \frac{\boldsymbol{v}_2}{\|\boldsymbol{v}_2\|} = \begin{bmatrix} -1/\sqrt{35} \\ 5/\sqrt{35} \\ -3/\sqrt{35} \end{bmatrix}$ ➔ $Q = \begin{bmatrix} 3/\sqrt{10} & -1/\sqrt{35} \\ 0 & 5/\sqrt{35} \\ -1/\sqrt{10} & -3/\sqrt{35} \end{bmatrix}$

  Finally, find R: $R = Q^T A = \begin{bmatrix} 3/\sqrt{10} & 0 & -1/\sqrt{10} \\ -1/\sqrt{35} & 5/\sqrt{35} & -3/\sqrt{35} \end{bmatrix} \begin{bmatrix} 3 & 8 \\ 0 & 5 \\ -1 & -6 \end{bmatrix} = \begin{bmatrix} 10/\sqrt{10} & 30/\sqrt{10} \\ 0 & 35/\sqrt{35} \end{bmatrix}$ ➔ $R = \begin{bmatrix} \sqrt{10} & 3\sqrt{10} \\ 0 & \sqrt{35} \end{bmatrix}$

Key points – 6.3.2 QR Factorization Properties

Q: $n \times p$

$Ax = 0$

$x = 0$

$n \times p$

- Properties of QR factorization
    1) $Q^T Q = I$ ✓ $I_{p \times p}$
    2) Columns of Q is equivalent to columns of A
        1) W =span $\{\boldsymbol{u}_1, \boldsymbol{u}_2 \cdots \boldsymbol{u}_p\}$ = span $\{\boldsymbol{x}_1, \boldsymbol{x}_2 \cdots \boldsymbol{x}_p\}$
        2) Q forms an orthonormal basis to span the same subspace W
    3) QQ^T is the projection matrix onto W (spanned by columns of A or Q)
    4) If A has independent columns, R is invertible, and all the values on the diagonal of R is positive
    5) If A has any dependent columns, simply skip it in Q

$$\begin{cases} Q^T Q = I_{n \times n} \\ Q Q^T = I_{n \times n} \\ Q^{-1} = Q^T \end{cases}$$

$\hat{y} = UU^T \vec{y}$

$R_{p \times p}$

End