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

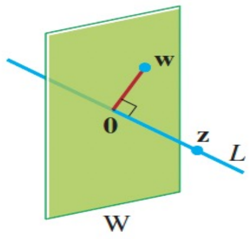
- **Coverage** : Ch 6, 7, 8 (Q3 & Q4)
- **Date/Time**: 2 May 2024 (Thursday), 1.00-3.00pm

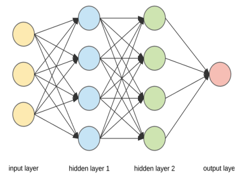
(Ch 9 will not be tested)

Syllabus for Part 2

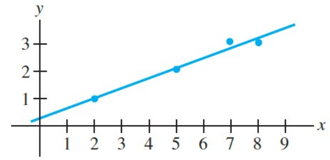
Chapter	Topics	Week (Lecture)	Week (Tut)
6	Orthogonality	8-9	9-10
7	Least Squares	9-10	10-11
8	EigenValue and Eigenvectors	11-12	12-13
9	Singular Value Decomposition (SVD)	13	

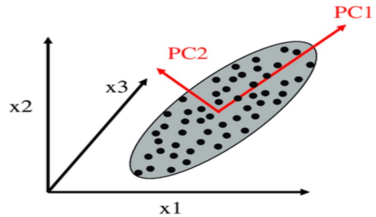
Table 1: schedule





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





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**

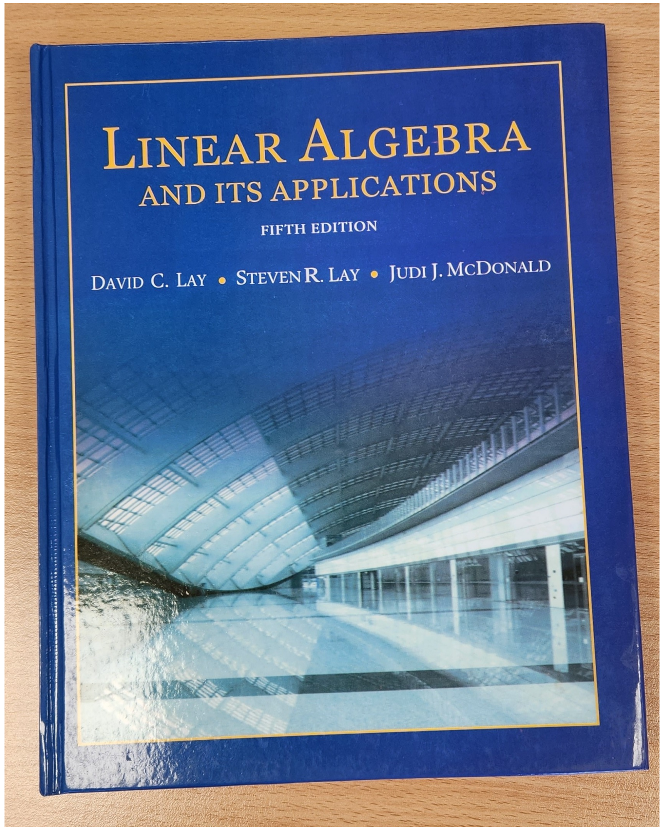
How will we conduct the course?

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

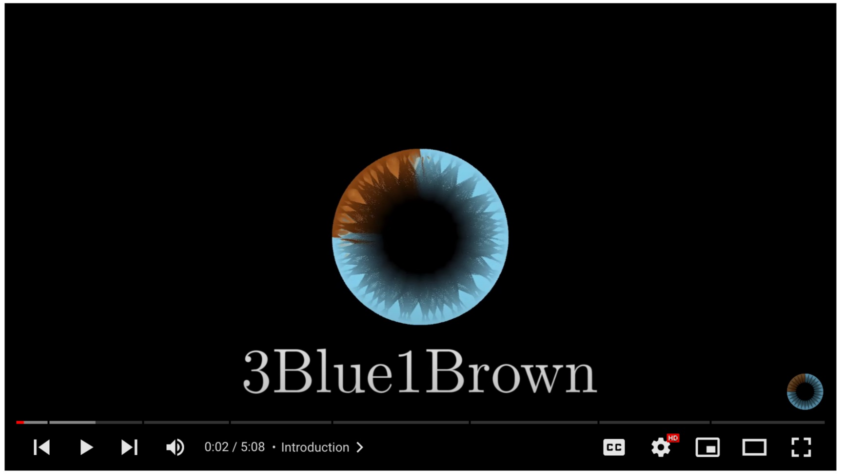
2) During lecture hours –
   - We will summarize the lectures and highlight the key points
   - Q&A.

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

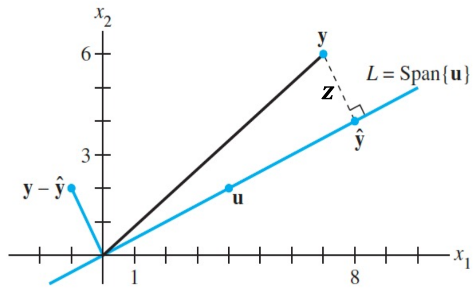
Lecture (Week 10)

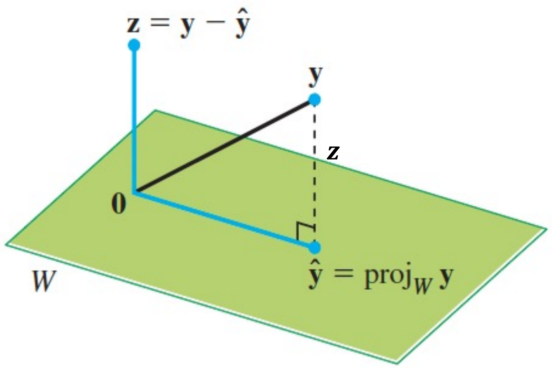
(Chapter 7.1.1-7.2.1)

Revision

Key points – Ch 6: Orthogonal Projection

- Project a vector to a line (1-d subspace): $\widehat{\boldsymbol{y}} = Proj_{\boldsymbol{u}}\, \boldsymbol{y} = \frac{\boldsymbol{y}\cdot\boldsymbol{u}}{\boldsymbol{u}\cdot\boldsymbol{u}}\boldsymbol{u}$
- Project a vector to a subspace spanned by $\{\boldsymbol{u}_1, \boldsymbol{u}_2 \cdots \boldsymbol{u}_p\}$:
  - $\widehat{\boldsymbol{y}} = Proj_W \boldsymbol{y} = \frac{\boldsymbol{y}\cdot\boldsymbol{u}_1}{\boldsymbol{u}_1\cdot\boldsymbol{u}_1}\boldsymbol{u}_1 + \cdots + \frac{\boldsymbol{y}\cdot\boldsymbol{u}_p}{\boldsymbol{u}_p\cdot\boldsymbol{u}_p}\boldsymbol{u}_p$
    - where $\{\boldsymbol{u}_1, \boldsymbol{u}_2 \cdots \boldsymbol{u}_p\}$ is an **orthogonal** basis
  - $\widehat{\boldsymbol{y}} = Proj_W \boldsymbol{y} = \frac{\boldsymbol{y}\cdot\boldsymbol{u}_1}{\boldsymbol{u}_1\cdot\boldsymbol{u}_1}\boldsymbol{u}_1 + \cdots + \frac{\boldsymbol{y}\cdot\boldsymbol{u}_p}{\boldsymbol{u}_p\cdot\boldsymbol{u}_p}\boldsymbol{u}_p = (\boldsymbol{y}\cdot\boldsymbol{u}_1)\boldsymbol{u}_1 + \cdots + (\boldsymbol{y}\cdot\boldsymbol{u}_p)\boldsymbol{u}_p = (\boldsymbol{u}_1{}^T\boldsymbol{y})\boldsymbol{u}_1 + \cdots + (\boldsymbol{u}_p{}^T\boldsymbol{y})\boldsymbol{u}_p = [\boldsymbol{u}_1\ \boldsymbol{u}_2 \cdots \boldsymbol{u}_p]\begin{bmatrix}\boldsymbol{u}_1{}^T \\ \vdots \\ \boldsymbol{u}_p{}^T\end{bmatrix}\boldsymbol{y} = UU^T\boldsymbol{y}$
    - where $\{\boldsymbol{u}_1, \boldsymbol{u}_2 \cdots \boldsymbol{u}_p\}$ is an **orthonormal** basis. U spans the subspace W.





- $\widehat{\boldsymbol{y}}$ is the best approximation of $\boldsymbol{y}$ on W ←→ $\|\boldsymbol{z}\| = \|\boldsymbol{y} - \widehat{\boldsymbol{y}}\|$ is the minimal distance from $\boldsymbol{y}$ to W.
- Think of a linear system: $A\boldsymbol{x} = \boldsymbol{b}$. If A span the subspace W, what solutions we can get when $\boldsymbol{b}$ is on W or not?

Key points – 7.1.1 Consistency in a System of Equations

- Definition:
  - For a linear system: $A\boldsymbol{x} = \boldsymbol{b}$ $\qquad A \in R^{m\times n}, \boldsymbol{x} \in R^{n}, \boldsymbol{b} \in R^{n}$
  - If no solution exists, it is an inconsistent system

- Explain: inconsistency happens when one of the following conditions is true
  - b is not in column space of A: b is not formed by linear combinations of A's columns.
  - The rows of A are dependent, but their corresponding b values are not consistent.
  - Rank (A) < Rank $(A|b)$: rank of A is less than that of the augmented matrix.

- In most cases, inconsistency occurs when $M \gg N$ (**over-determined**), where there are more equations than unknowns.

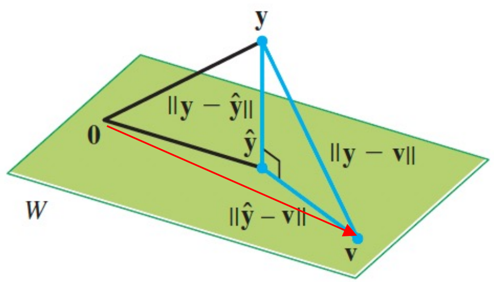
Example of an inconsistent system

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 4 & 6 \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} 3 \\ 2 \\ 6 \end{bmatrix}$$

Key points – 7.1.2 The Least Square Problem

- Definition
  - If there is no solution for system: $A\boldsymbol{x} = \boldsymbol{b}$, we can find an $\widehat{\boldsymbol{x}}$, which is the closet approximation: $A\widehat{\boldsymbol{x}} = \widehat{\boldsymbol{b}}$, such that
$$\|A\widehat{\boldsymbol{x}} - \boldsymbol{b}\| < \|A\boldsymbol{x} - \boldsymbol{b}\|$$

- Explain:
  - Columns of A spans a subspace W
  - $\widehat{\boldsymbol{b}} = A\widehat{\boldsymbol{x}}$ is the linear combination of columns of A, so $\widehat{\boldsymbol{b}}$ is in subspace W
  - If $\widehat{\boldsymbol{b}}$ is the orthogonal projection of $\boldsymbol{b}$ onto W, then $\|A\widehat{\boldsymbol{x}} - \boldsymbol{b}\| = \|\widehat{\boldsymbol{b}} - \boldsymbol{b}\|$ (residual) is orthogonal to W
  - So, $\|A\widehat{\boldsymbol{x}} - \boldsymbol{b}\|$ is the least distance from $\boldsymbol{b}$ to W



- Recall 6.5.2 (see graph above) Best Approximation Theorem : $\|\boldsymbol{y} - \widehat{\boldsymbol{y}}\| < \|\boldsymbol{y} - \boldsymbol{v}\|$
  - $\boldsymbol{y} = \boldsymbol{b}$
  - $\widehat{\boldsymbol{y}} = \widehat{\boldsymbol{b}}$
  - $\boldsymbol{v}$ (red color) is an any vector in W

Key points – 7.1.3 Norm Equation (LS Solution)

- Definition
  - From $A\boldsymbol{x} = \boldsymbol{b}$, define "normal equation": $A^T A\hat{\boldsymbol{x}} = A^T\boldsymbol{b}$

$$A = [\boldsymbol{a_1} \quad \boldsymbol{a_2} \quad \cdots \quad \boldsymbol{a_i} \quad \cdots \quad \boldsymbol{a_n}]$$
$$A \in R^{m\times n}, \boldsymbol{a_i} \in R^m$$

- Explain
  - Since $A\boldsymbol{x} = \boldsymbol{b}$ does not have a solution ($\boldsymbol{b}$ is not a linear combination of columns of A), we project $\boldsymbol{b}$ to W spanned by the columns of A as $\hat{\boldsymbol{b}}$:

$$A\hat{\boldsymbol{x}} = \hat{\boldsymbol{b}}$$

    which has a solution (because $\hat{\boldsymbol{b}}$ is on W)

  - $\boldsymbol{b} - \hat{\boldsymbol{b}} = \boldsymbol{b} - A\hat{\boldsymbol{x}}$ is the residual of $\boldsymbol{b}$ onto W
  - $\boldsymbol{b} - A\hat{\boldsymbol{x}}$ is orthogonal to all columns of A: $\boldsymbol{a_i} \cdot (\boldsymbol{b} - A\hat{\boldsymbol{x}}) = \boldsymbol{0}$
  - Use matrix form: $\boldsymbol{a_i} \cdot (\boldsymbol{b} - A\hat{\boldsymbol{x}})= \boldsymbol{a_i}^T(\boldsymbol{b} - A\hat{\boldsymbol{x}}) = \boldsymbol{0},$ for all $i = 1, \cdots, n$
  - Finally: $\begin{bmatrix} \boldsymbol{a_1}^T \\ \vdots \\ \boldsymbol{a_n}^T \end{bmatrix} (\boldsymbol{b} - A\hat{\boldsymbol{x}}) = \boldsymbol{0} \Rightarrow A^T(\boldsymbol{b} - A\hat{\boldsymbol{x}}) = 0 \Rightarrow A^T A\hat{\boldsymbol{x}} = A^T\boldsymbol{b}$
  - If $A^T A$ is invertible, we get **Least-Square solution**: $\hat{\boldsymbol{x}} = (A^T A)^{-1} A^T\boldsymbol{b}$

This Least-Square solution is derived from the normal equation directly.

Key points – 7.1.3 Find Least Square Solution

- Example: find least square solution using normal equation $\hat{\boldsymbol{x}} = (A^T A)^{-1} A^T \boldsymbol{b}$, **if $\boldsymbol{A^T A}$ is invertible.**
  - Given A and $\boldsymbol{b}$: $A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}$, $\boldsymbol{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$
  - Find $A^T A = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$ (Invertible)
  - We have $(A^T A)^{-1} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}^{-1} = \frac{1}{17\times5-1\times1} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}$
  - Find $A^T \boldsymbol{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$
  - Finally, $\hat{\boldsymbol{x}} = (A^T A)^{-1} A^T \boldsymbol{b} = \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix} \begin{bmatrix} 19 \\ 11 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 84 \\ 168 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$
  - Least square residual: $\boldsymbol{b} - A\hat{\boldsymbol{x}} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} - \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -2 \\ -4 \\ 8 \end{bmatrix}$, least square error: $\|\boldsymbol{b} - A\hat{\boldsymbol{x}}\| = \sqrt{(-2)^2 + (-4)^2 + (8)^2} = \sqrt{84}$

Invert a 2×2 matrix: $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

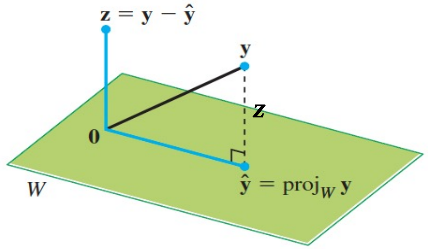
Key points – 7.1.3 Find Least Square Solution(2).

- Example to find a least square solution for $A\boldsymbol{x} = \boldsymbol{b}$. If $\boldsymbol{A^T A}$ **is not invertible,** using Gaussian Elimination approach.
- $A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$, $\boldsymbol{b} = \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix}$
- Following normal equation $A^T A\hat{\boldsymbol{x}} = A^T \boldsymbol{b}$
- $A^T A = \begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix}$ (not invertible, rank=3), $A^T \boldsymbol{b} = \begin{bmatrix} 4 \\ -4 \\ 2 \\ 6 \end{bmatrix}$
- Use Gaussian elimination: $\left[\begin{array}{cccc|c} 6 & 2 & 2 & 2 & 4 \\ 2 & 2 & 0 & 0 & -4 \\ 2 & 0 & 2 & 0 & 2 \\ 2 & 0 & 0 & 2 & 6 \end{array}\right]$ ➔ $\left[\begin{array}{cccc|c} 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & -1 & -5 \\ 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$ ➔ $\begin{matrix} x_1 + x_4 = 3 \\ x_2 - x_4 = -5 \\ x_3 - x_4 = -2 \end{matrix}$ ➔ $\begin{matrix} x_1 = 3 - x_4 \\ x_2 = -5 + x_4 \\ x_3 = -2 + x_4 \end{matrix}$
- Finally the least square solutions (infinite): $\hat{\boldsymbol{x}} = \begin{bmatrix} 3 \\ -5 \\ -2 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$

Comments on Least Square Solutions (so far)

- The least-square solution: to find a solution so that $\widehat{\boldsymbol{b}} = A\widehat{\boldsymbol{x}}$ is the orthogonal projection to column space of $A \in R^{m\times n}$.

- When we have orthogonal columns in $A = [\boldsymbol{a}_1, \boldsymbol{a}_{2,} \cdots, \boldsymbol{a}_n]$
  - $\widehat{\boldsymbol{b}} = Proj_w \boldsymbol{b} = \frac{\boldsymbol{b}\cdot\boldsymbol{a}_1}{\boldsymbol{a}_1\cdot\boldsymbol{a}_1}\boldsymbol{a}_1 + \cdots + \frac{\boldsymbol{y}\cdot\boldsymbol{a}_n}{\boldsymbol{a}_n\cdot\boldsymbol{a}_n}\boldsymbol{a}_n$

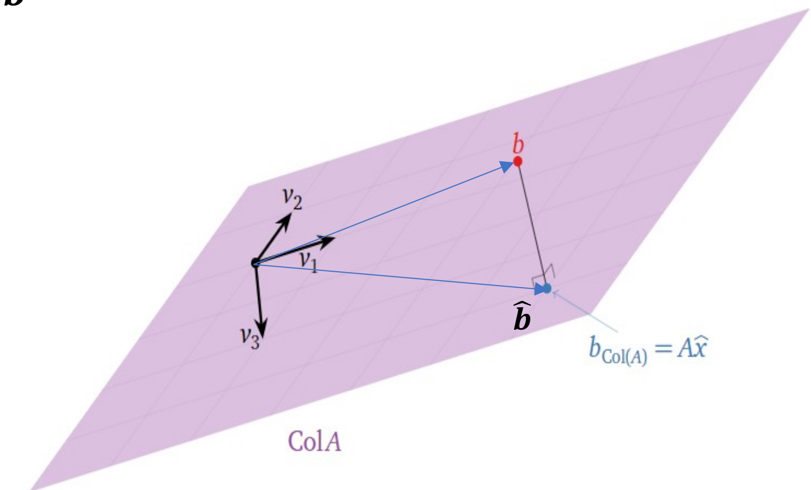


- For any A, the least-square solution can be found using $A^T A\widehat{\boldsymbol{x}} = A^T\boldsymbol{b}$

- If $A^T A$ is invertible (rank(A)=n), unique solution: $\widehat{\boldsymbol{x}} = (A^T A)^{-1}A^T\boldsymbol{b}$

- If $A^T A$ is not invertible (rank(A)<n), there are infinite solutions
  - Is $\widehat{\boldsymbol{x}}$ unique?
  - Is $\widehat{\boldsymbol{b}}$ unique?

- Visualization:
  - A has three columns to form a 2-D subspace (one of the columns is a linear combination of the other two).
  - $\widehat{\boldsymbol{b}}$ can be a linear combination of any two or three vectors.



Visualize Least Square Solutions

- When we have orthogonal columns in $A = [\boldsymbol{a}_1, \boldsymbol{a}_{2,} \cdots, \boldsymbol{a}_n]$
  - $\widehat{\boldsymbol{b}} = Proj_w \boldsymbol{b} = \frac{\boldsymbol{b} \cdot \boldsymbol{a}_1}{\boldsymbol{a}_1 \cdot \boldsymbol{a}_1} \boldsymbol{a}_1 + \cdots + \frac{\boldsymbol{y} \cdot \boldsymbol{a}_n}{\boldsymbol{a}_n \cdot \boldsymbol{a}_n} \boldsymbol{a}_n$

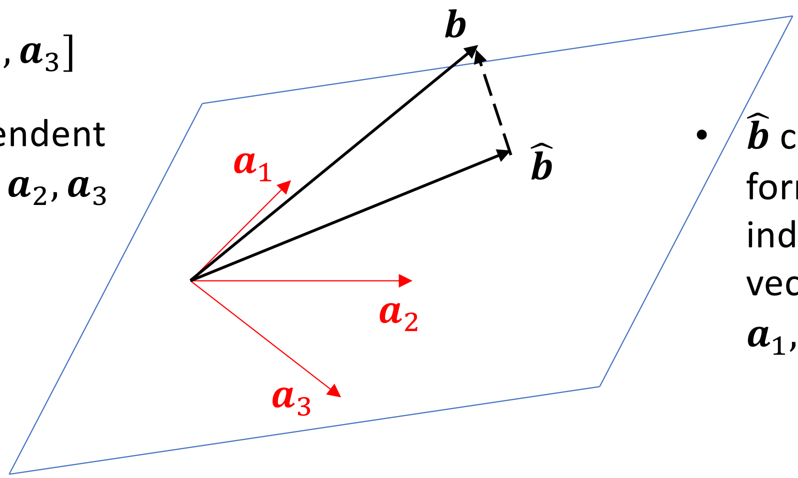
- If $A^T A$ is invertible (rank(A)=n), unique solution: $\widehat{\boldsymbol{x}} = (A^T A)^{-1} A^T \boldsymbol{b}$

- If $A^T A$ is not invertible (rank(A)<n), there are infinite solutions

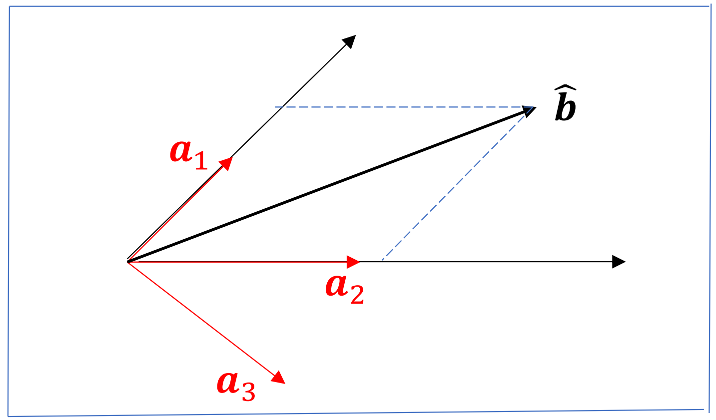
Example: Visualize Least Square Solutions (infinite)

$A = [\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3]$

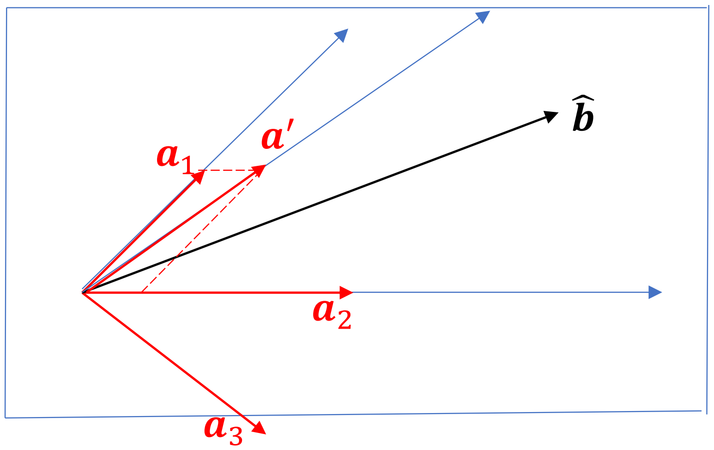
- There are 2 independent vectors among $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$



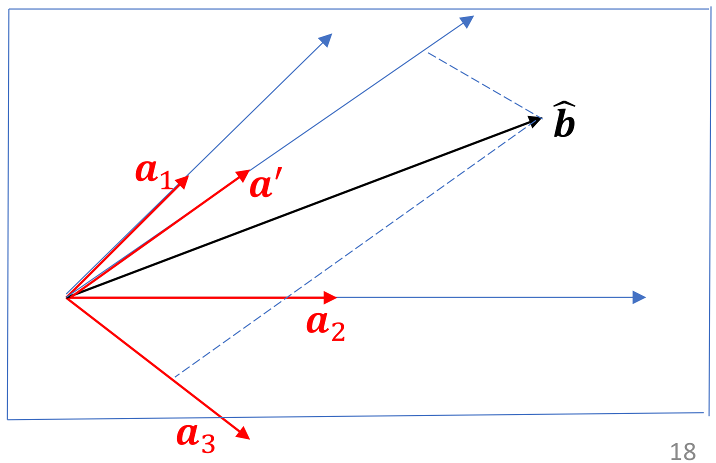
- $\hat{\boldsymbol{b}}$ can be formed by indep vectors $\boldsymbol{a}_1, \boldsymbol{a}_2$



- Any vector $\boldsymbol{a}'$ can be formed by indep vectors $\boldsymbol{a}_1, \boldsymbol{a}_2$



- $\hat{\boldsymbol{b}}$ can be formed by $\boldsymbol{a}'$ and $\boldsymbol{a}_3$.
- There are infinite $\boldsymbol{a}'$s



Key points – 7.1.4 Projection Matrix

- Definition:
  - Project a vector $\boldsymbol{b}$ onto a subspace W, spanned by columns of A. The project matrix is defined as: $P = A(A^T A)^{-1} A^T$

    $\Rightarrow \widehat{\boldsymbol{b}} = P\boldsymbol{b}$

- Explain:
  - $\widehat{\boldsymbol{b}} = Proj_W\ \boldsymbol{b} = A\,\widehat{\boldsymbol{x}}$ is the orthogonal projection of $\boldsymbol{b}$ onto a subspace W
  - Bring in the Least Square solution $\widehat{\boldsymbol{x}} = (A^T A)^{-1} A^T \boldsymbol{b}$ into the above equation
  - $\widehat{\boldsymbol{b}} = A\,\widehat{\boldsymbol{x}} = A\,(\,(A^T A)^{-1} A^T \boldsymbol{b}\,) = A(A^T A)^{-1} A^T \boldsymbol{b} \ \Rightarrow\ P = A(A^T A)^{-1} A^T$

- Properties of project matrix
  - $P^T = P$
  - $P^N = P \times P \times \cdots \times P = P$ (idempotent)

Key points – 7.1.5 Least Square Solution Using QR Factorization

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

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

- Definition:
  - Given $A\boldsymbol{x} = \boldsymbol{b}$
  - Using QR factorization: $A = QR$
  - So we have: $QR\boldsymbol{x} = \boldsymbol{b}$ ➔ multiply Q^Ton both sides $Q^TQR\boldsymbol{x} = Q^T\boldsymbol{b}$
  - Since $Q^TQ = I$, we get: $R\boldsymbol{x} = Q^T\boldsymbol{b}$ ➔ $\boldsymbol{x} = R^{-1}Q^T\boldsymbol{b}$

- Explain why $\boldsymbol{x} = R^{-1}Q^T\boldsymbol{b}$ is a Least Square solution
  - Since $\boldsymbol{x} = R^{-1}Q^T\boldsymbol{b}$, than $A\boldsymbol{x} = A(R^{-1}\ Q^T\boldsymbol{b}) = QR\ (R^{-1}Q^T\boldsymbol{b}) = QQ^T\boldsymbol{b} = \hat{\boldsymbol{b}}$ (Orthogonal Projection of $\boldsymbol{b}$ onto column space of Q and A)
  - Recall: Q is orthonormal. Col (Q) spans the same subspace W as Col (A)
  - So, $\boldsymbol{x}$ is the least square solution.

○ A^TA is sensitive to small errors, so QR method is often used.

Key points – 7.2.1 Applications of Least Square

- Least Square method is used to find a linear regression (linear curve fitting) – try to find a line which fits the discrete data points

$$y = \beta_0 + \beta_1 x$$

  such that $\sum (y_i - \hat{y}_i)^2$ is minimal ($\hat{y}_i$is the estimated value from the linear equation $y = \beta_0 + \beta_1 x$, and β_0, β_1 called regression coefficients)

- Solution:
  - Given n data points, the system equations are:

$$\begin{matrix} y_1 = \beta_0 + \beta_1 x_1 \\ \vdots \\ y_n = \beta_0 + \beta_1 x_n \end{matrix} \rightarrow \boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = X\boldsymbol{\beta}$$

  - The least square solution: $\boldsymbol{\beta} = (X^T X)^{-1} X^T \boldsymbol{y}$ Recall: $\hat{\boldsymbol{x}} = (A^T A)^{-1} A^T \boldsymbol{b}$

- Example:
  - $X = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix}, \boldsymbol{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$
  - $X^T X = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}, X^T \boldsymbol{y} = \begin{bmatrix} 9 \\ 57 \end{bmatrix} \rightarrow \boldsymbol{\beta} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}^{-1} \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix}$

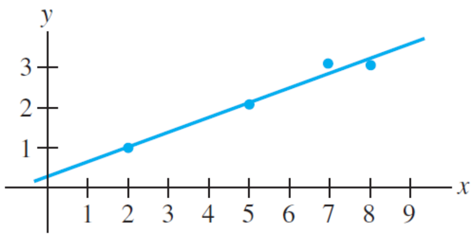


FIGURE 2 The least-squares line $y = \frac{2}{7} + \frac{5}{14}x$.

i	x_i	y_i
1	2	1
2	5	2
3	7	3
4	8	3

Key points – 7.2.1 Applications of Least Square (2)

- Least square fitting of other curves
  - If we can use certain known functions to fit the discrete data points,

    $y = \beta_0 f_0(x) + \beta_1 f_1(x) + \cdots + \beta_k f_k(x)$

    we can use least square method to find regression coefficients $\beta_0, \beta_{1,} \cdots, \beta_k$
  - Example:
    - For data shown on the right, we could fit it with a combination of linear and quadratic functions, i.e.

      $$y = \beta_0 + \beta_1 x + \beta_2 x^2$$

    So we can form the system equations as:

$$\begin{matrix} y_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 \\ \vdots \\ y_n = \beta_0 + \beta_1 x_n + \beta_2 x_n^2 \end{matrix} \rightarrow \boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = X\boldsymbol{\beta} \rightarrow \boldsymbol{\beta} = (X^T X)^{-1} X^T \boldsymbol{y}$$

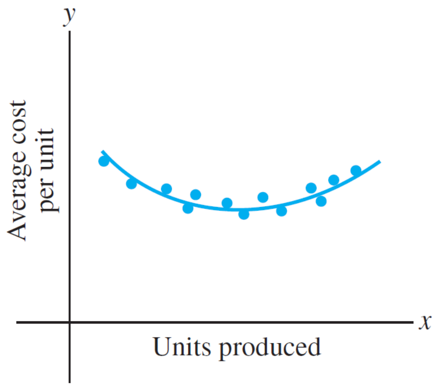


FIGURE 3

Average cost curve.

End

Additional notes:

- Differences between LU and QR factorization
  - LU is applied to any square matrix, QR is applied to a matrix with independent columns
  - LU factorization produces an upper-triangle and a lower-triangle matrix
  - QR factorization produces an orthonormal matrix and an upper-triangle
  - Find LU factorization through Gaussian elimination
  - Find QR factorization using the Gram-Schmidt algorithm
  - Different use cases:
    - LU factorization is used to find solutions of systems of linear equations, matrix inversion, and matrix determinant.
    - QR factorization is used in least-squares, eigenvalue, and signal processing.