# MAT 222 Linear Algebra
# Week 10
# Lecture Notes 2

Murat Karaçayır

Akdeniz University
Department of Mathematics

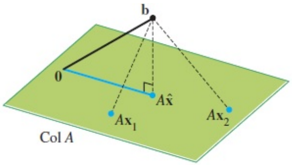
8th May 2025

# Least Squares Problems

- Last week we observed that the orthogonal projection of a vector $\mathbf{y}$ onto a vector space $V$ is the point in $V$ that is closest to $\mathbf{y}$. More explicitly,

$$\|\mathbf{y} - \text{proj}_V \mathbf{y}\| < \|\mathbf{y} - \mathbf{u}\|$$

  for any other vector $\mathbf{u}$ of $V$.
- Now we will discuss the same problem in a similar context.
- If $V$ is a subspace of the Euclidean space $\mathbb{R}^m$, then $V = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ for some vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in \mathbb{R}^m$.
- $V$ can also be written $V = \{A\mathbf{x} : \mathbf{x} \in \mathbb{R}^n\} = C(A)$, where $A$ is the $m \times n$ matrix whose columns are $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$.




- Then, given a vector $\mathbf{b}$ not in $V$, we may be interested in finding the vector in $V$ that is closest to $\mathbf{b}$.
- This amounts to the following: Find a vector $\hat{\mathbf{x}}$ in $\mathbf{R}^n$ such that $\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$ for any $\mathbf{x} \in \mathbb{R}^n$.
- This is called a **least squares problem**.

# Least Square Solutions to $A\mathbf{x} = \mathbf{b}$

- Our previous work shows that, a least squares problem can be reduced to the problem of finding an $\hat{\mathbf{x}}$ such that the vector $A\hat{\mathbf{x}}$ is equal to the projection of $\mathbf{b}$ on $C(A)$.
- In other words, the vector $A\hat{\mathbf{x}}$ is as close to $\mathbf{b}$ as possible.
- Such a problem may arise if we have an inconsistent system $A\mathbf{x} = \mathbf{b}$ and still want to find the "best" solution to this problem.
- Instead of solving $A\mathbf{x} = \mathbf{b}$, we solve $A\mathbf{x} = \hat{\mathbf{b}}$, where $\hat{\mathbf{b}}$ is the projection of $\mathbf{b}$ onto $C(A)$.
- Any vector $\hat{\mathbf{x}}$ that is a solution of $A\mathbf{x} = \hat{\mathbf{b}}$ is called a **least squares solution** of $A\mathbf{x} = \mathbf{b}$.
- The difference $\|\mathbf{b} - A\hat{\mathbf{x}}\|$ is called the **least squares error**. It is the distance of $\mathbf{b}$ to the vector space $C(A)$.
- Note that, if $\mathbf{b} \in C(A)$, then least squares solutions of $A\mathbf{x} = \mathbf{b}$ are its actual solutions and the least squares error is equal to 0.

# Least Squares Problem: Example

**Example:** Given $A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$, find the least squares solutions of $A\mathbf{x} = \mathbf{b}$.

- Any least square solution of $A\mathbf{x} = \mathbf{b}$ satisfies the system $A\mathbf{x} = \hat{\mathbf{b}}$, where $\hat{\mathbf{b}}$ is the projection of $\mathbf{b}$ onto $C(A)$.
- $\hat{\mathbf{b}}$ can be calculated as $\hat{\mathbf{b}} = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix}$. (Exercise)
- The system $A\mathbf{x} = \hat{\mathbf{b}}$ becomes $\begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix}$, whose solution is unique and equal to $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.
- Thus, the system $A\mathbf{x} = \mathbf{b}$ has a unique least squares solution, which is $\hat{\mathbf{x}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

# Alternative Way to Compute Least Square Solutions

- Finding the orthogonal projection of $\mathbf{b}$ onto $C(A)$ involves finding an orthogonal basis for $C(A)$ and computing projections onto these basis vectors. This process may be tedious.
- Recall that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to every vector in $C(A)$, in particular it is orthogonal to the columns of $A$.
- This means that $A^T(\mathbf{b} - A\hat{\mathbf{x}}) = \mathbf{0}$, the zero vector.
- Rearranging, We can write this as $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$.
- Therefore, every least squares solution of $A\mathbf{x} = \mathbf{b}$ satisfies the system $A^TA\mathbf{x} = A^T\mathbf{b}$. This system is known as **normal equations** for $A\mathbf{x} = \mathbf{b}$.
- Solving normal equations is generally easier as it does not require computing the projection of $\mathbf{b}$ onto $C(A)$.
- If $A^TA$ is invertible, then there is a unique least squares solution.
- For the previous example, $A^TA = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$ and one can find its inverse as $(A^TA)^{-1} = \dfrac{1}{84}\begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}$.
- Thus, we obtain the least squares solution $\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, same as before. This solution is unique.

# Least Squares Problem: Example

**Example:** Let $A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & -1 \\ 1 & 2 & -3 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$, find the least squares solutions of $A\mathbf{x} = \mathbf{b}$.

- Let us form the normal equations.
- $A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 1 & -1 & -3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & -1 \\ 1 & 2 & -3 \end{bmatrix} = \begin{bmatrix} 3 & 3 & -3 \\ 3 & 5 & -7 \\ -3 & -7 & 11 \end{bmatrix}$ and $A^T\mathbf{b} = \begin{bmatrix} 6 \\ 0 \\ 6 \end{bmatrix}$.
- The augmented matrix $\left[\begin{array}{c|c} A^T A & A^T\mathbf{b} \end{array}\right]$ can be row-reduced as

$$\left[\begin{array}{ccc|c} 3 & 3 & -3 & 6 \\ 3 & 5 & -7 & 0 \\ -3 & -7 & 11 & 6 \end{array}\right] \longrightarrow \left[\begin{array}{ccc|c} 1 & 0 & 1 & 5 \\ 0 & 1 & -2 & -3 \\ 0 & 0 & 0 & 0 \end{array}\right] \longrightarrow \begin{array}{r} x_1 + x_3 = 5 \\ x_2 - 2x_3 = -3 \end{array}.$$

- We have infinitely many solutions given by $\hat{\mathbf{x}} = \begin{bmatrix} 5 \\ -3 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$.
- In this example the columns of $A$ are linearly dependent (Please check). In general, this causes the matrix $A^T A$ to be singular and hence the least squares problem to have infinitely many solutions.

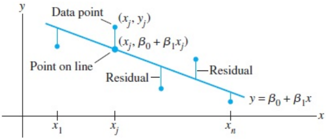
# Least Squares Problem: Example

**Example:** Let $A = \begin{bmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix}$, find the least squares solutions of $A\mathbf{x} = \mathbf{b}$ and compute the least squares error.

- We notice that the columns of $A$ are orthogonal. This makes it easy to compute the orthogonal projection $\hat{\mathbf{b}}$ of $\mathbf{b}$ onto $C(A)$ and thus there is no need to form the normal equations.
- If we name the columns of $A$ by $\mathbf{a}_1$ and $\mathbf{a}_2$, then
$$\hat{\mathbf{b}} = \mathrm{proj}_{\mathbf{a}_1}\mathbf{b} + \mathrm{proj}_{\mathbf{a}_2}\mathbf{b} = \frac{\mathbf{b}\cdot\mathbf{a}_1}{\mathbf{a}_1\cdot\mathbf{a}_1}\mathbf{a}_1 + \frac{\mathbf{b}\cdot\mathbf{a}_2}{\mathbf{a}_2\cdot\mathbf{a}_2}\mathbf{a}_2 = \frac{8}{4}\mathbf{a}_1 + \frac{45}{90}\mathbf{a}_2 = A\begin{bmatrix} 2 \\ 0.5 \end{bmatrix}.$$
- This shows that the least squares solution is $\hat{\mathbf{x}} = \begin{bmatrix} 2 \\ 0.5 \end{bmatrix}$.
- Least squares error is
$$\|\mathbf{b} - \hat{\mathbf{b}}\| = \left\| \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \\ 2.5 \\ 5.5 \end{bmatrix} \right\| = \sqrt{1 + 2.25 + 0.25} = \sqrt{3.5} \approx 1.871$$
- This example shows the advantage of orthogonal columns in $A$.

# Least Square Lines

- Suppose there are two variables $x$ and $y$ with experimental data $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.
- In many cases this data resembles a line.
- We want to find a line $y = \beta_0 + \beta_1 x$ close to the data points.
- The vertical (not perpendicular!) distance between the data points and points on the line (predicted $y$ values) are called the **residual**$s$.




- The most common approach is to minimize the sum of the squares of these residuals.
- This becomes a least squares problem. We will write it as $X\beta = \mathbf{y}$. $X$ is called the **design matrix**, $\beta$ the **parameter vector** and $\mathbf{y}$ the **observation vector**. $\beta_1$ and $\beta_2$ are called **regression coefficients**.
- The goal is to determine the parameter vector.

# Least Squares Line

| Predicted $y$ value | Observed $y$ value |
|---|---|
| $\beta_0 + \beta_1 x_1$ | $y_1$ |
| $\beta_0 + \beta_1 x_2$ | $y_2$ |
| $\vdots$ | $\vdots$ |
| $\beta_0 + \beta_1 x_n$ | $y_n$ |

- In the above table, if the left column was equal to the right column, we would have the system $X\beta = \mathbf{y}$, where

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

.

- Since not all the observed data lie on the line $y = \beta_0 + \beta_1 x$, the system $X\beta = \mathbf{y}$ is inconsistent.
- What we can do is to find a parameter vector $\beta$ that minimizes the distance $\|\mathbf{y} - X\beta\|$. This is a least squares problem.

# Least Squares Line: Example

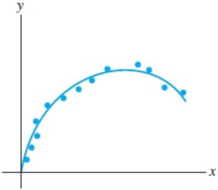
**Example:** The data points $(2,1),(5,2),(7,3)$ and $(8,3)$ are given. Find the line $\beta_0+\beta_1 x$ that best fits the data (in the least squares sense).

- The design matrix and observation vector are $X=\begin{bmatrix}1 & 2\\ 1 & 5\\ 1 & 7\\ 1 & 8\end{bmatrix}$, $\mathbf{y}=\begin{bmatrix}1\\2\\3\\3\end{bmatrix}$.
- Let us construct the normal equations. $X^TX=\begin{bmatrix}4 & 22\\ 22 & 142\end{bmatrix}$ and $X^T\mathbf{y}=\begin{bmatrix}9\\57\end{bmatrix}$.
- The parameter vector is obtained by $\beta=(X^TX)^{-1}X^T\mathbf{y}=\frac{1}{84}\begin{bmatrix}142 & -22\\ -22 & 4\end{bmatrix}\begin{bmatrix}9\\57\end{bmatrix}=\begin{bmatrix}2/7\\5/14\end{bmatrix}$.
- Thus, the regression coefficients are $\beta_0=\frac{2}{7}$ and $\beta_1=\frac{5}{14}$. The least squares line is given by

$$y=\frac{2}{7}+\frac{5}{14}x$$

- It can be used to make predicitions at unknown $x$ values. For example $y(4)$ is predicted as $\frac{18}{7}$.

# Least Squares Parabola




- Sometimes the observed data does not look like a line. In such cases, a different fitting curve should be used.
- An example is a parabola $y = \beta_0 + \beta_1 x + \beta_2 x^2$. (See the figure)
- Then we have the equalities
$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \epsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_2 + \beta_2 x_2^2 + \epsilon_2 \\ \vdots &= \vdots \\ y_n &= \beta_0 + \beta_1 x_n + \beta_2 x_n^2 + \epsilon_n \end{aligned},$$
where $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are the residuals.
- It can be written as $\mathbf{y} = X\beta + \epsilon$, where
$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$
- Finding $\beta$ that minimizes $\|\epsilon\|^2$ is a least squares problem.

# Least Squares Parabola: Example

**Example:** Given the data points $(-1, 1/2), (1, -1), (2, -1/2), (3, 2)$, find the parabola that best approximates them.

- We have $X = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 1/2 \\ -1 \\ -1/2 \\ 2 \end{bmatrix}$.
- The normal equations are $X^T X \beta = X^T \mathbf{y}$, where

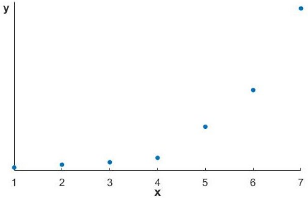
$$X^T X = \begin{bmatrix} 4 & 5 & 15 \\ 5 & 15 & 35 \\ 15 & 35 & 99 \end{bmatrix}, \; X^T \mathbf{y} = \begin{bmatrix} 1 \\ 3.5 \\ 15.5 \end{bmatrix}.$$

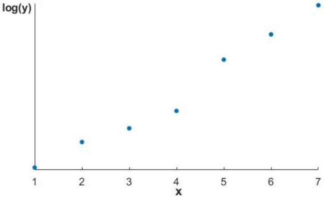
- The solution is given by $\beta = (X^T X)^{-1} X^T \mathbf{y} = \begin{bmatrix} -\frac{41}{44} \\ -\frac{379}{440} \\ \frac{53}{88} \end{bmatrix}$. (Please check)
- Thus, the least squares parabola is $y = -\frac{41}{44} - \frac{379}{440}x + \frac{53}{88}x^2$.
- **Exercise:** Compute the least squares error.

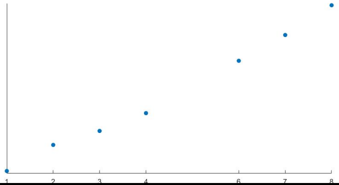
# Distances of Planets to Sun: Example

| Mercury | Venus | Earth | Mars | Jupiter | Saturn | Uranus |
|---|---|---|---|---|---|---|
| 0.39 | 0.72 | 1.00 | 1.52 | 5.20 | 9.54 | 19.2 |

- The above table includes the average distances of the first seven planets to the Sun, using Earth's distance as a unit.
- Labeling the planets from 1 to 7, their graph does not look like a line, it resembles an exponential curve, so the log of $y$ values looks like a line.



- Considering there is a missing planet between Mars and Jupiter that broke apart somewhere in time, if we label Jupiter by 6, the data looks like a line even more.

# Distances of Planets to Sun: Example (Continued)

- Thus, we can find the least square regression line of $\ln(y)$ on $x$, instead of $y$ on $x$. The data is given below.

| 1 | 2 | 3 | 4 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|
| -0.94 | -0.33 | 0 | 0.42 | 1.65 | 2.26 | 2.95 |

- Thus, we have $X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 6 \\ 1 & 7 \\ 1 & 8 \end{bmatrix}$ and $\mathbf{Y} = \begin{bmatrix} -0.94 \\ -0.33 \\ 0 \\ 0.42 \\ 1.65 \\ 2.26 \\ 2.95 \end{bmatrix}$, where $Y = \ln(y)$.
- The solution of this least squares problem is $\beta = \begin{bmatrix} -1.5613 \\ 0.54635 \end{bmatrix}$, giving rise to the least squares line $Y = -1.5613 + 0.54635x$. (Exercise)
- According to this result, the logarithm of the average distance of Neptune to the Sun is $Y(9) \approx 3.36$. So the distance itself is $y(9) = e^{Y(9)} \approx 28.67$.
- Thus, the average distance of Neptune to the Sun is predicted as $(28.67) \cdot 1.496 \times 10^8 \approx 4.289 \times 10^9$ km. Quite accurate!

# Mean-Deviation Form: Example

- Given the values $x_1, x_2, \ldots, x_n$ for the independent variable, let $\bar{x} = \dfrac{x_1 + x_2 + \ldots + x_n}{n}$ be their arithmetic mean.
- Then defining a new variable by $x^* = x - \bar{x}$ substantially simplifies the solution. This new variable is called the **mean-deviation**.

**Example:** The data points $(2,1), (5,2), (7,3)$ and $(8,3)$ are given. Use the mean-deviation of $x$ to find the least squares line $\beta_0 + \beta_1 x$.

- The arithmetic mean is $\bar{x} = 5.5$ and since $x^* = x - \bar{x}$, the values of the mean-deviation become $x_1^* = 2 - 5.5 = -3.5$, $x_2^* = 5 - 5.5 = -0.5$, $x_3^* = 7 - 5.5 = 1.5$, $x_4^* = 8 - 5.5 = 2.5$.
- Thus, data points $(x_i^*, y_i)$ become $(-3.5, 1), (-0.5, 2), (1.5, 3)$ and $(2.5, 3)$.
- Based on this data, the new problem becomes $X^*\beta = \mathbf{y}$, where
$$X^* = \begin{bmatrix} 1 & -3.5 \\ 1 & -0.5 \\ 1 & 1.5 \\ 1 & 2.5 \end{bmatrix} \text{ and } \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}.$$
- Observe that the columns of $X^*$ are orthogonal.
- This means that we can easily compute the projection of $\mathbf{y}$ onto $C(X^*)$.

# Mean-Deviation Form: Example

- Let $\mathbf{a}_1$ and $\mathbf{a}_2$ be the columns of $X^*$, respectively.
- One can show that

$$\hat{\mathbf{y}} = \text{proj}_{C(X^*)}\mathbf{y} = \frac{9}{4}\mathbf{a}_1 + \frac{5}{14}\mathbf{a}_2. \text{ (Exercise)}$$

- Thus, the parameter vector is obtained by $\beta = \begin{bmatrix} \frac{9}{4} \\ \frac{5}{14} \end{bmatrix}$, giving rise to the least squares regression line $\frac{9}{4} + \frac{5}{14}x^*$ in the mean-deviation form.
- In order to return to the original variable $x$, just use $x^* = x - 5.5$. So we have

$$y = \frac{9}{4} + \frac{5}{14}(x - 5.5) = \frac{9}{4} + \frac{5}{14}x - \frac{55}{28} = \frac{2}{7} + \frac{5}{14}x,$$

as before.

**Exercise:** Solve the mean-deviation form of the problem using the normal equations. What do you observe? (Exercise)

# QR-Factorization

- Consider an $m \times n$ matrix $A$ with linearly independent columns. (Note that this implies $m \geq n$.)
- Let $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ be the columns of $A$. $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$ is a basis for $C(A)$ and Gram-Schmidt process can be used to find an orthogonal basis. We can normalize it to construct an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$.
- Let $Q$ be the $m \times n$ matrix whose columns are $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$.
- Gram-Schmidt process ensures that for any $k = 1, 2, \ldots, n$, $\mathrm{Span}\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k\} = \mathrm{Span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$.
- This shows, in particular that $\mathbf{a}_k \in \mathrm{Span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$, so there are constants $r_{1k}, r_{2k}, \ldots, r_{kk}$ such that

$$\mathbf{a}_k = r_{1k}\mathbf{u}_1 + r_{2k}\mathbf{u}_2 + \ldots + r_{kk}\mathbf{u}_k = Q \begin{bmatrix} r_{1k} \\ r_{2k} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

# QR Factorization of a matrix

- But this is valid for all $k = 1, 2, \ldots, n$ so we have $A = QR$, where the columns of $Q$ is an orthonormal basis for $C(A)$ and $R$ is an upper triangular matrix whose $k$-th column consists of the coordinates of $\mathbf{a}_k$ with respect to $\{\mathbf{u}_1, \mathbf{u}_2 \ldots, \mathbf{u}_k\}$.
- Note that if any of the diagonal entries $r_{kk}$ of $R$ is negative, multiplying both it and $\mathbf{u}_k$ does not spoil the relation $A = QR$.
- In summary, we have the following.

**QR-factorization**

If $A$ is an $m \times n$ matrix whose columns are linearly independent, it can be factored as $A = QR$ where $Q$ is $m \times n$ whose columns from an orthonormal basis for $C(A)$ and $R$ is an $n \times n$ upper triangular matrix whose diagonal entries are positive.

# QR Factorization: Example

**Example:** Find the $QR$-factorization of $A = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & -1 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$.

- Let the columns of $A$ be $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$. Then Gram-Schmidt process applied to the set $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ produces the orthogonal basis $\{(0, 1, -1, 0)^T, (0, \frac{1}{2}, \frac{1}{2}, 1)^T, (1, -\frac{1}{3}, -\frac{1}{3}, \frac{1}{3})\}$. (Please check)
- Normalizing gives the orthonormal basis $\{\frac{1}{\sqrt{2}}(0, 1, -1, 0)^T, (0, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}})^T, (\frac{3}{2\sqrt{3}}, -\frac{1}{2\sqrt{3}}, -\frac{1}{2\sqrt{3}}, \frac{1}{2\sqrt{3}})\}$.
- This gives $Q = \begin{bmatrix} 0 & 0 & \frac{3}{2\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & -\frac{1}{2\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & -\frac{1}{2\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{2\sqrt{3}} \end{bmatrix}$.
- As for the entries of $R$, we observe that since the columns of $Q$ are orthonormal we have $Q^T Q = \mathbf{I}$ (we say $Q$ is an **orthogonal matrix**). Thus, $A = QR$ implies $Q^T QR = R \longrightarrow Q^T A = R$.
- **Exercise:** Compute $R$ and obtain the $QR$-factorization of $A$.

- Suppose you want to find a least squares solution to $A\mathbf{x} = \mathbf{b}$, where the columns of $A$ are linearly independent.
- Consider the normal equations $A^T A\mathbf{x} = A^T\mathbf{b}$. Since the columns of $A$ are linearly independent, $A^T A$ is invertible and the unique solution of normal equations is $\hat{\mathbf{x}} = (A^T A)^{-1} A^T\mathbf{b}$.
- Now use the $QR$-factorization of $A$. We have

$$\hat{\mathbf{x}} = ((QR)^T QR)^{-1}(QR)^T\mathbf{b} = (R^T Q^T QR)^{-1} R^T Q^T\mathbf{b} = (R^T \mathbf{I} R)^{-1} R^T Q^T\mathbf{b} = R^{-1}(R^T)^{-1} R^T Q^T\mathbf{b} = R^{-1} Q^T\mathbf{x}.$$

**Example:** Let $A = \begin{bmatrix} 1 & 3 & 5 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 3 & 3 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix}$. Calculate the least squares solution to $A\mathbf{x} = \mathbf{b}$.

- The $QR$-factorization of $A$ is $A = \begin{bmatrix} 1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & -1/2 \\ 1/2 & -1/2 & 1/2 \\ 1/2 & 1/2 & -1/2 \end{bmatrix} \begin{bmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{bmatrix}$.

  (Please check)

# Using QR-Factorization in Least Squares Problems

- The inverse of $R$ can be found as $R^{-1} = \begin{bmatrix} 1/2 & -1 & 1/4 \\ 0 & 1/2 & -3/4 \\ 0 & 0 & 1/2 \end{bmatrix}$.
- Thus, the least squares solution is calculated by

$$\begin{aligned}\hat{\mathbf{x}} &= R^{-1}Q^T\mathbf{b} \\ &= \begin{bmatrix} 1/2 & -1 & 1/4 \\ 0 & 1/2 & -3/4 \\ 0 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & -1/2 & 1/2 \\ 1/2 & -1/2 & 1/2 & -1/2 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix} \\ &= \begin{bmatrix} 10 \\ -6 \\ 2 \end{bmatrix}.\end{aligned}$$