Linear Algebra 2

Lecture #17

Orthogonal projection onto subspaces of $\mathbb{R}^n$. The method of least squares

Irena Penev

March 25, 2024

- This lecture has four parts:

- This lecture has four parts:
    1. A brief review of projections onto subspaces

- This lecture has four parts:
  1. A brief review of projections onto subspaces
  2. Orthogonal projection onto subspaces of $\mathbb{R}^n$

- This lecture has four parts:
    1. A brief review of projections onto subspaces
    2. Orthogonal projection onto subspaces of $\mathbb{R}^n$
    3. The method of least squares

- This lecture has four parts:
  1. A brief review of projections onto subspaces
  2. Orthogonal projection onto subspaces of $\mathbb{R}^n$
  3. The method of least squares
  4. Data fitting

1. A brief review of projections onto subspaces

1. A brief review of projections onto subspaces

Theorem 6.5.1

Let V be a finite-dimensional real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $|| \cdot ||$. Let U be a subspace of V, and let $\mathbf{x} \in V$. Then there exists a unique vector $\mathbf{x}_U \in U$ that has the property that

$$||\mathbf{x} - \mathbf{x}_U|| = \min_{\mathbf{u} \in U} ||\mathbf{x} - \mathbf{u}||.$$

Moreover, if $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is an orthogonal basis of U, then this vector $\mathbf{x}_U$ is given by the formula

$$\mathbf{x}_U = \sum_{i=1}^{k} \text{proj}_{\mathbf{u}_i}(\mathbf{x}) = \sum_{i=1}^{k} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i.$$

- **Terminology/Notation:** The vector $\mathbf{x}_U$ from Theorem 6.5.1 is called the *orthogonal projection* of $\mathbf{x}$ onto U.

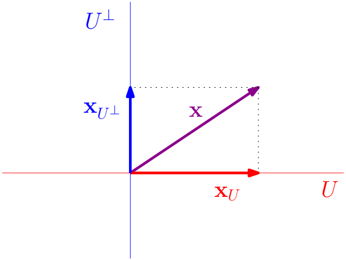
Corollary 6.5.3

Let V be a finite-dimensional real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $|| \cdot ||$. Let U be a subspace of V, and let $\mathbf{x} \in V$. Then

$$\mathbf{x} = \mathbf{x}_U + \mathbf{x}_{U^\perp}.$$

Moreover, this is the unique way of expressing $\mathbf{x}$ as a sum of a vector in U and a vector in $U^\perp$.[a]

[a] This means that for all $\mathbf{y} \in U$ and $\mathbf{z} \in U^\perp$, if $\mathbf{x} = \mathbf{y} + \mathbf{z}$, then $\mathbf{y} = \mathbf{x}_U$ and $\mathbf{z} = \mathbf{x}_{U^\perp}$.



- Suppose that V is a finite-dimensional real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $|| \cdot ||$, and suppose that U is a subspace of V.

- Suppose that V is a finite-dimensional real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $|| \cdot ||$, and suppose that U is a subspace of V.
- We can then define the function $\text{proj}_U : V \to V$ by setting $\text{proj}_U(\mathbf{x}) = \mathbf{x}_U$ for all $\mathbf{x} \in V$ (where $\mathbf{x}_U$ is the orthogonal projection of $\mathbf{x}$ onto U, as in Theorem 6.5.1).

- Suppose that V is a finite-dimensional real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $|| \cdot ||$, and suppose that U is a subspace of V.
- We can then define the function $\text{proj}_U : V \to V$ by setting $\text{proj}_U(\mathbf{x}) = \mathbf{x}_U$ for all $\mathbf{x} \in V$ (where $\mathbf{x}_U$ is the orthogonal projection of $\mathbf{x}$ onto U, as in Theorem 6.5.1).
- Clearly, $\text{proj}_U(\mathbf{u}) = \mathbf{u}$ for all $\mathbf{u} \in U$.

- Suppose that V is a finite-dimensional real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $|| \cdot ||$, and suppose that U is a subspace of V.
- We can then define the function $\text{proj}_U : V \to V$ by setting $\text{proj}_U(\mathbf{x}) = \mathbf{x}_U$ for all $\mathbf{x} \in V$ (where $\mathbf{x}_U$ is the orthogonal projection of $\mathbf{x}$ onto U, as in Theorem 6.5.1).
- Clearly, $\text{proj}_U(\mathbf{u}) = \mathbf{u}$ for all $\mathbf{u} \in U$.
- Moreover, we have that $\text{Im}(\text{proj}_U) = U$ and $\text{proj}_U[U] = U$.

- Suppose that V is a finite-dimensional real or complex vector space, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $|| \cdot ||$, and suppose that U is a subspace of V.
- We can then define the function $\text{proj}_U : V \to V$ by setting $\text{proj}_U(\mathbf{x}) = \mathbf{x}_U$ for all $\mathbf{x} \in V$ (where $\mathbf{x}_U$ is the orthogonal projection of $\mathbf{x}$ onto U, as in Theorem 6.5.1).
- Clearly, $\text{proj}_U(\mathbf{u}) = \mathbf{u}$ for all $\mathbf{u} \in U$.
- Moreover, we have that $\text{Im}(\text{proj}_U) = U$ and $\text{proj}_U[U] = U$.
- Using the formula from Theorem 6.5.1, we can easily see that the function proj_U is linear.

2. Orthogonal projection onto subspaces of $\mathbb{R}^n$

2. Orthogonal projection onto subspaces of $\mathbb{R}^n$

- In what follows, we will assume that $\mathbb{R}^n$ is equipped with the standard scalar product $\cdot$ and the induced norm $||\cdot||$.

2. Orthogonal projection onto subspaces of $\mathbb{R}^n$

- In what follows, we will assume that $\mathbb{R}^n$ is equipped with the standard scalar product $\cdot$ and the induced norm $||\cdot||$.
- Suppose that U is a subspace of $\mathbb{R}^n$.

2. Orthogonal projection onto subspaces of $\mathbb{R}^n$

- In what follows, we will assume that $\mathbb{R}^n$ is equipped with the standard scalar product $\cdot$ and the induced norm $||\cdot||$.
- Suppose that U is a subspace of $\mathbb{R}^n$.
  - As we saw, $\text{proj}_U : \mathbb{R}^n \to \mathbb{R}^n$ is linear (and its image is U).

2. Orthogonal projection onto subspaces of $\mathbb{R}^n$

- In what follows, we will assume that $\mathbb{R}^n$ is equipped with the standard scalar product $\cdot$ and the induced norm $||\cdot||$.
- Suppose that U is a subspace of $\mathbb{R}^n$.
  - As we saw, $\text{proj}_U : \mathbb{R}^n \to \mathbb{R}^n$ is linear (and its image is U).
  - Since proj_U is linear, it has a standard matrix.

2. Orthogonal projection onto subspaces of $\mathbb{R}^n$

- In what follows, we will assume that $\mathbb{R}^n$ is equipped with the standard scalar product $\cdot$ and the induced norm $||\cdot||$.
- Suppose that U is a subspace of $\mathbb{R}^n$.
  - As we saw, $\text{proj}_U : \mathbb{R}^n \to \mathbb{R}^n$ is linear (and its image is U).
  - Since proj_U is linear, it has a standard matrix.
    - Note that this matrix belongs to $\mathbb{R}^{n\times n}$.

2. Orthogonal projection onto subspaces of $\mathbb{R}^n$

- In what follows, we will assume that $\mathbb{R}^n$ is equipped with the standard scalar product $\cdot$ and the induced norm $||\cdot||$.
- Suppose that U is a subspace of $\mathbb{R}^n$.
  - As we saw, $\text{proj}_U : \mathbb{R}^n \to \mathbb{R}^n$ is linear (and its image is U).
  - Since proj_U is linear, it has a standard matrix.
    - Note that this matrix belongs to $\mathbb{R}^{n \times n}$.
    - By definition, if A is the standard matrix of proj_U, then we have that

$$\text{proj}_U(\mathbf{x}) \quad = \quad A\mathbf{x} \qquad\qquad \forall \mathbf{x} \in \mathbb{R}^n.$$

2. Orthogonal projection onto subspaces of $\mathbb{R}^n$

- In what follows, we will assume that $\mathbb{R}^n$ is equipped with the standard scalar product $\cdot$ and the induced norm $||\cdot||$.
- Suppose that U is a subspace of $\mathbb{R}^n$.
  - As we saw, $\text{proj}_U : \mathbb{R}^n \to \mathbb{R}^n$ is linear (and its image is U).
  - Since proj_U is linear, it has a standard matrix.
    - Note that this matrix belongs to $\mathbb{R}^{n \times n}$.
    - By definition, if A is the standard matrix of proj_U, then we have that

$$\text{proj}_U(\mathbf{x}) \quad = \quad A\mathbf{x} \qquad\qquad \forall \mathbf{x} \in \mathbb{R}^n.$$

  - Our goal is to we give formulas for the standard matrices of orthogonal projections onto various subspaces of $\mathbb{R}^n$.

- In section 3.3 of the Lecture Notes (last semester), we defined the row space of a matrix A to be the span of the rows of A.

- In section 3.3 of the Lecture Notes (last semester), we defined the row space of a matrix A to be the span of the rows of A.
- Proposition 3.3.1(b) states that

$$\mathrm{Row}(A) \;=\; \{\mathbf{u}^T \mid \mathbf{u} \in \mathrm{Col}(A^T)\}.$$

- In section 3.3 of the Lecture Notes (last semester), we defined the row space of a matrix A to be the span of the rows of A.
- Proposition 3.3.1(b) states that

$$\mathrm{Row}(A) \;=\; \{\mathbf{u}^T \mid \mathbf{u} \in \mathrm{Col}(A^T)\}.$$

- In what follows, it will be convenient to slightly modify the definition of the row space, as follows:

$$\mathrm{Row}(A) \;:=\; \mathrm{Col}(A^T).$$

- In section 3.3 of the Lecture Notes (last semester), we defined the row space of a matrix A to be the span of the rows of A.
- Proposition 3.3.1(b) states that

$$\mathrm{Row}(A) \quad = \quad \{\mathbf{u}^T \mid \mathbf{u} \in \mathrm{Col}(A^T)\}.$$

- In what follows, it will be convenient to slightly modify the definition of the row space, as follows:

$$\mathrm{Row}(A) \quad := \quad \mathrm{Col}(A^T).$$

- So, we (re)defined the row space of a matrix to be the span of the transposes of its rows.

- For example, for the matrix

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

we have that

$$A^T = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 1 & 2 & 3 \\ 2 & 3 & 4 \end{bmatrix},$$

and consequently,

$$\mathrm{Row}(A) = \mathrm{Span}\left(\begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \\ 3 \\ 4 \end{bmatrix} \right).$$

- For example, for the matrix

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

we have that

$$A^T = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 1 & 2 & 3 \\ 2 & 3 & 4 \end{bmatrix},$$

and consequently,

$$\mathrm{Row}(A) = \mathrm{Span}\left(\begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \\ 3 \\ 4 \end{bmatrix} \right).$$

- If this change of definition bothers you, then every time you see $\mathrm{Row}(\square)$, mentally replace it with $\mathrm{Col}(\square^T)$.

Theorem 6.6.1

Let $A \in \mathbb{R}^{n \times m}$. Then $\text{Row}(A)^{\perp} = \text{Nul}(A)$ and $\text{Row}(A) = \text{Nul}(A)^{\perp}$.

Proof.

Theorem 6.6.1

Let $A \in \mathbb{R}^{n \times m}$. Then $\text{Row}(A)^\perp = \text{Nul}(A)$ and $\text{Row}(A) = \text{Nul}(A)^\perp$.

Proof. In view of Theorem 6.4.3(c), it suffices to show that $\text{Row}(A)^\perp = \text{Nul}(A)$.

- Indeed, by Theorem 6.4.3(c), we have that

$$(\text{Row}(A)^\perp)^\perp \quad = \quad \text{Row}(A).$$

- So, if $\text{Row}(A)^\perp = \text{Nul}(A)$, then

$$\text{Nul}(A)^\perp \quad = \quad (\text{Row}(A)^\perp)^\perp = \text{Row}(A).$$

Theorem 6.6.1

Let $A \in \mathbb{R}^{n \times m}$. Then $\text{Row}(A)^\perp = \text{Nul}(A)$ and $\text{Row}(A) = \text{Nul}(A)^\perp$.

Proof. In view of Theorem 6.4.3(c), it suffices to show that $\text{Row}(A)^\perp = \text{Nul}(A)$.

- Indeed, by Theorem 6.4.3(c), we have that

$$(\text{Row}(A)^\perp)^\perp \quad = \quad \text{Row}(A).$$

- So, if $\text{Row}(A)^\perp = \text{Nul}(A)$, then

$$\text{Nul}(A)^\perp \quad = \quad (\text{Row}(A)^\perp)^\perp = \text{Row}(A).$$

Set

$$A \quad = \quad \begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix},$$

so that $\text{Row}(A) = \text{Span}(\mathbf{a}_1, \dots, \mathbf{a}_n)$.

Proof (continued). Now, for all vectors $\mathbf{x} \in \mathbb{R}^m$:

$$\begin{aligned}
\mathbf{x} \in \mathrm{Nul}(A) &\iff A\mathbf{x} = \mathbf{0} \\
&\iff \begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix} \mathbf{x} = \mathbf{0} \\
&\iff \begin{bmatrix} \mathbf{a}_1 \cdot \mathbf{x} \\ \vdots \\ \mathbf{a}_n \cdot \mathbf{x} \end{bmatrix} = \mathbf{0} \\
&\iff \mathbf{a}_i \cdot \mathbf{x} = 0 \ \ \forall i \in \{1, \dots, n\} \\
&\iff \mathbf{a}_i \perp \mathbf{x} \ \ \forall i \in \{1, \dots, n\} \\
&\iff \mathbf{x} \in \{\mathbf{a}_1, \dots, \mathbf{a}_n\}^\perp \\
&\overset{(*)}{\iff} \mathbf{x} \in \mathrm{Span}(\mathbf{a}_1, \dots, \mathbf{a}_n)^\perp \\
&\iff \mathbf{x} \in \mathrm{Row}(A)^\perp,
\end{aligned}$$

where (*) follows from the fact that $\{\mathbf{a}_1, \dots, \mathbf{a}_m\}^\perp = \mathrm{Span}(\mathbf{a}_1, \dots, \mathbf{a}_m)^\perp$ (by Proposition 6.4.2). This proves that $\mathrm{Nul}(A) = \mathrm{Row}(A)^\perp$, and we are done. □

Theorem 6.6.1

Let $A \in \mathbb{R}^{n \times m}$. Then $\text{Row}(A)^\perp = \text{Nul}(A)$ and $\text{Row}(A) = \text{Nul}(A)^\perp$.

Theorem 6.6.1

Let $A \in \mathbb{R}^{n \times m}$. Then $\text{Row}(A)^{\perp} = \text{Nul}(A)$ and $\text{Row}(A) = \text{Nul}(A)^{\perp}$.

Corollary 6.6.2

Let $A \in \mathbb{R}^{n \times m}$. Then all the following hold:

(a) $\text{Nul}(A^T A) = \text{Nul}(A)$;

(b) $\text{Row}(A^T A) = \text{Row}(A)$;

(c) $\text{rank}(A^T A) = \text{rank}(A)$.

Proof.

Theorem 6.6.1

Let $A \in \mathbb{R}^{n \times m}$. Then $\text{Row}(A)^\perp = \text{Nul}(A)$ and $\text{Row}(A) = \text{Nul}(A)^\perp$.

Corollary 6.6.2

Let $A \in \mathbb{R}^{n \times m}$. Then all the following hold:

(a) $\text{Nul}(A^T A) = \text{Nul}(A)$;

(b) $\text{Row}(A^T A) = \text{Row}(A)$;

(c) $\text{rank}(A^T A) = \text{rank}(A)$.

Proof. We first prove (a).

Theorem 6.6.1

Let $A \in \mathbb{R}^{n \times m}$. Then $\text{Row}(A)^{\perp} = \text{Nul}(A)$ and $\text{Row}(A) = \text{Nul}(A)^{\perp}$.

Corollary 6.6.2

Let $A \in \mathbb{R}^{n \times m}$. Then all the following hold:

(a) $\text{Nul}(A^T A) = \text{Nul}(A)$;

(b) $\text{Row}(A^T A) = \text{Row}(A)$;

(c) $\text{rank}(A^T A) = \text{rank}(A)$.

Proof. We first prove (a). Note that $A^T A \in \mathbb{R}^{m \times m}$, and that both $\text{Nul}(A)$ and $\text{Nul}(A^T A)$ are subspaces of $\mathbb{R}^m$.

Theorem 6.6.1

Let $A \in \mathbb{R}^{n \times m}$. Then $\text{Row}(A)^\perp = \text{Nul}(A)$ and $\text{Row}(A) = \text{Nul}(A)^\perp$.

Corollary 6.6.2

Let $A \in \mathbb{R}^{n \times m}$. Then all the following hold:

(a) $\text{Nul}(A^T A) = \text{Nul}(A)$;

(b) $\text{Row}(A^T A) = \text{Row}(A)$;

(c) $\text{rank}(A^T A) = \text{rank}(A)$.

Proof. We first prove (a). Note that $A^T A \in \mathbb{R}^{m \times m}$, and that both $\text{Nul}(A)$ and $\text{Nul}(A^T A)$ are subspaces of $\mathbb{R}^m$. Now, fix any $\mathbf{x} \in \mathbb{R}^m$. WTS $\mathbf{x} \in \text{Nul}(A^T A)$ iff $\mathbf{x} \in \text{Nul}(A)$.

Theorem 6.6.1

Let $A \in \mathbb{R}^{n \times m}$. Then $\text{Row}(A)^\perp = \text{Nul}(A)$ and $\text{Row}(A) = \text{Nul}(A)^\perp$.

Corollary 6.6.2

Let $A \in \mathbb{R}^{n \times m}$. Then all the following hold:

(a) $\text{Nul}(A^T A) = \text{Nul}(A)$;

(b) $\text{Row}(A^T A) = \text{Row}(A)$;

(c) $\text{rank}(A^T A) = \text{rank}(A)$.

Proof. We first prove (a). Note that $A^T A \in \mathbb{R}^{m \times m}$, and that both $\text{Nul}(A)$ and $\text{Nul}(A^T A)$ are subspaces of $\mathbb{R}^m$. Now, fix any $\mathbf{x} \in \mathbb{R}^m$. WTS $\mathbf{x} \in \text{Nul}(A^T A)$ iff $\mathbf{x} \in \text{Nul}(A)$.

Suppose first that $\mathbf{x} \in \text{Nul}(A)$.

Theorem 6.6.1

Let $A \in \mathbb{R}^{n \times m}$. Then $\text{Row}(A)^\perp = \text{Nul}(A)$ and $\text{Row}(A) = \text{Nul}(A)^\perp$.

Corollary 6.6.2

Let $A \in \mathbb{R}^{n \times m}$. Then all the following hold:

(a) $\text{Nul}(A^T A) = \text{Nul}(A)$;

(b) $\text{Row}(A^T A) = \text{Row}(A)$;

(c) $\text{rank}(A^T A) = \text{rank}(A)$.

Proof. We first prove (a). Note that $A^T A \in \mathbb{R}^{m \times m}$, and that both $\text{Nul}(A)$ and $\text{Nul}(A^T A)$ are subspaces of $\mathbb{R}^m$. Now, fix any $\mathbf{x} \in \mathbb{R}^m$. WTS $\mathbf{x} \in \text{Nul}(A^T A)$ iff $\mathbf{x} \in \text{Nul}(A)$.

Suppose first that $\mathbf{x} \in \text{Nul}(A)$. Then $A\mathbf{x} = \mathbf{0}$, and consequently, $A^T A\mathbf{x} = \mathbf{0}$. So, $\mathbf{x} \in \text{Nul}(A^T A)$.

Theorem 6.6.1

Let $A \in \mathbb{R}^{n \times m}$. Then $\text{Row}(A)^\perp = \text{Nul}(A)$ and $\text{Row}(A) = \text{Nul}(A)^\perp$.

Corollary 6.6.2

Let $A \in \mathbb{R}^{n \times m}$. Then all the following hold:

- (a) $\text{Nul}(A^T A) = \text{Nul}(A)$;
- (b) $\text{Row}(A^T A) = \text{Row}(A)$;
- (c) $\text{rank}(A^T A) = \text{rank}(A)$.

Proof (continued). Suppose, conversely, that $\mathbf{x} \in \text{Nul}(A^T A)$.

Theorem 6.6.1

Let $A \in \mathbb{R}^{n \times m}$. Then $\text{Row}(A)^\perp = \text{Nul}(A)$ and $\text{Row}(A) = \text{Nul}(A)^\perp$.

Corollary 6.6.2

Let $A \in \mathbb{R}^{n \times m}$. Then all the following hold:

a. $\text{Nul}(A^T A) = \text{Nul}(A)$;
b. $\text{Row}(A^T A) = \text{Row}(A)$;
c. $\text{rank}(A^T A) = \text{rank}(A)$.

Proof (continued). Suppose, conversely, that $\mathbf{x} \in \text{Nul}(A^T A)$. Then $A^T A\mathbf{x} = \mathbf{0}$, and it follows that $\mathbf{x}^T A^T A\mathbf{x} = \mathbf{0}$.

Theorem 6.6.1

Let $A \in \mathbb{R}^{n \times m}$. Then $\text{Row}(A)^{\perp} = \text{Nul}(A)$ and $\text{Row}(A) = \text{Nul}(A)^{\perp}$.

Corollary 6.6.2

Let $A \in \mathbb{R}^{n \times m}$. Then all the following hold:

(a) $\text{Nul}(A^T A) = \text{Nul}(A)$;

(b) $\text{Row}(A^T A) = \text{Row}(A)$;

(c) $\text{rank}(A^T A) = \text{rank}(A)$.

Proof (continued). Suppose, conversely, that $\mathbf{x} \in \text{Nul}(A^T A)$. Then $A^T A\mathbf{x} = \mathbf{0}$, and it follows that $\mathbf{x}^T A^T A\mathbf{x} = \mathbf{0}$. But note that

$$\mathbf{x}^T A^T A\mathbf{x} = (A\mathbf{x})^T (A\mathbf{x}) = (A\mathbf{x}) \cdot (A\mathbf{x}) = ||A\mathbf{x}||^2;$$

Theorem 6.6.1

Let $A \in \mathbb{R}^{n \times m}$. Then $\text{Row}(A)^\perp = \text{Nul}(A)$ and $\text{Row}(A) = \text{Nul}(A)^\perp$.

Corollary 6.6.2

Let $A \in \mathbb{R}^{n \times m}$. Then all the following hold:

(a) $\text{Nul}(A^T A) = \text{Nul}(A)$;

(b) $\text{Row}(A^T A) = \text{Row}(A)$;

(c) $\text{rank}(A^T A) = \text{rank}(A)$.

Proof (continued). Suppose, conversely, that $\mathbf{x} \in \text{Nul}(A^T A)$. Then $A^T A\mathbf{x} = \mathbf{0}$, and it follows that $\mathbf{x}^T A^T A\mathbf{x} = \mathbf{0}$. But note that

$$\mathbf{x}^T A^T A\mathbf{x} \;=\; (A\mathbf{x})^T (A\mathbf{x}) \;=\; (A\mathbf{x}) \cdot (A\mathbf{x}) \;=\; ||A\mathbf{x}||^2;$$

consequently, $||A\mathbf{x}||^2 = 0$.

Theorem 6.6.1

Let $A \in \mathbb{R}^{n \times m}$. Then $\text{Row}(A)^\perp = \text{Nul}(A)$ and $\text{Row}(A) = \text{Nul}(A)^\perp$.

Corollary 6.6.2

Let $A \in \mathbb{R}^{n \times m}$. Then all the following hold:

(a) $\text{Nul}(A^T A) = \text{Nul}(A)$;

(b) $\text{Row}(A^T A) = \text{Row}(A)$;

(c) $\text{rank}(A^T A) = \text{rank}(A)$.

Proof (continued). Suppose, conversely, that $\mathbf{x} \in \text{Nul}(A^T A)$. Then $A^T A\mathbf{x} = \mathbf{0}$, and it follows that $\mathbf{x}^T A^T A\mathbf{x} = \mathbf{0}$. But note that

$$\mathbf{x}^T A^T A\mathbf{x} \;=\; (A\mathbf{x})^T (A\mathbf{x}) \;=\; (A\mathbf{x}) \cdot (A\mathbf{x}) \;=\; ||A\mathbf{x}||^2;$$

consequently, $||A\mathbf{x}||^2 = 0$. It follows that $||A\mathbf{x}|| = 0$, and therefore, $A\mathbf{x} = \mathbf{0}$, i.e. $\mathbf{x} \in \text{Nul}(A)$. This proves (a).

Theorem 6.6.1

Let $A \in \mathbb{R}^{n \times m}$. Then $\text{Row}(A)^\perp = \text{Nul}(A)$ and $\text{Row}(A) = \text{Nul}(A)^\perp$.

Corollary 6.6.2

Let $A \in \mathbb{R}^{n \times m}$. Then all the following hold:

(a) $\text{Nul}(A^T A) = \text{Nul}(A)$;

(b) $\text{Row}(A^T A) = \text{Row}(A)$;

(c) $\text{rank}(A^T A) = \text{rank}(A)$.

Proof (continued). For (b), we observe that

$$\begin{aligned} \text{Row}(A^T A) &= \text{Nul}(A^T A)^\perp && \text{by Theorem 6.6.1} \\ &= \text{Nul}(A)^\perp && \text{by (a)} \\ &= \text{Row}(A) && \text{by Theorem 6.6.1.} \end{aligned}$$

Theorem 6.6.1

Let $A \in \mathbb{R}^{n \times m}$. Then $\text{Row}(A)^\perp = \text{Nul}(A)$ and $\text{Row}(A) = \text{Nul}(A)^\perp$.

Corollary 6.6.2

Let $A \in \mathbb{R}^{n \times m}$. Then all the following hold:

(a) $\text{Nul}(A^T A) = \text{Nul}(A)$;

(b) $\text{Row}(A^T A) = \text{Row}(A)$;

(c) $\text{rank}(A^T A) = \text{rank}(A)$.

Proof (continued). Finally, for (c), we have the following:

$$\begin{aligned} \text{rank}(A^T A) &= \dim\big(\text{Row}(A^T A)\big) && \text{by Theorem 3.3.9} \\ &= \dim\big(\text{Row}(A)\big) && \text{by (b)} \\ &= \text{rank}(A) && \text{by Theorem 3.3.9.} \end{aligned}$$

This completes the argument. □

Theorem 6.6.3

Let $A \in \mathbb{R}^{n\times m}$ be a matrix of rank m (i.e. A is a matrix of full column rank). Then the matrix $A(A^TA)^{-1}A^T$ is the standard matrix of orthogonal projection onto $\text{Col}(A)$, that is, for all $\mathbf{x} \in \mathbb{R}^n$, the orthogonal projection of $\mathbf{x}$ onto $C := \text{Col}(A)$ is given by

$$\mathbf{x}_C \;=\; A(A^TA)^{-1}A^T\mathbf{x}.$$

Proof.

Theorem 6.6.3

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank m (i.e. A is a matrix of full column rank). Then the matrix $A(A^T A)^{-1}A^T$ is the standard matrix of orthogonal projection onto $\text{Col}(A)$, that is, for all $\mathbf{x} \in \mathbb{R}^n$, the orthogonal projection of $\mathbf{x}$ onto $C := \text{Col}(A)$ is given by

$$\mathbf{x}_C \quad = \quad A(A^T A)^{-1}A^T\mathbf{x}.$$

Proof. Fix $\mathbf{x} \in \mathbb{R}^n$. We must first check that the expression $A(A^T A)^{-1}A^T\mathbf{x}$ is defined and belongs to $C = \text{Col}(A)$.

Theorem 6.6.3

Let $A \in \mathbb{R}^{n\times m}$ be a matrix of rank m (i.e. A is a matrix of full column rank). Then the matrix $A(A^TA)^{-1}A^T$ is the standard matrix of orthogonal projection onto $\text{Col}(A)$, that is, for all $\mathbf{x} \in \mathbb{R}^n$, the orthogonal projection of $\mathbf{x}$ onto $C := \text{Col}(A)$ is given by

$$\mathbf{x}_C \quad = \quad A(A^TA)^{-1}A^T\mathbf{x}.$$

Proof. Fix $\mathbf{x} \in \mathbb{R}^n$. We must first check that the expression $A(A^TA)^{-1}A^T\mathbf{x}$ is defined and belongs to $C = \text{Col}(A)$.

First, note that $A^TA \in \mathbb{R}^{m\times m}$, and that by Corollary 6.6.2(a), we have that $\text{rank}(A^TA) = \text{rank}(A) = m$.

Theorem 6.6.3

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank m (i.e. A is a matrix of full column rank). Then the matrix $A(A^T A)^{-1} A^T$ is the standard matrix of orthogonal projection onto $\text{Col}(A)$, that is, for all $\mathbf{x} \in \mathbb{R}^n$, the orthogonal projection of $\mathbf{x}$ onto $C := \text{Col}(A)$ is given by

$$\mathbf{x}_C = A(A^T A)^{-1} A^T \mathbf{x}.$$

Proof. Fix $\mathbf{x} \in \mathbb{R}^n$. We must first check that the expression $A(A^T A)^{-1} A^T \mathbf{x}$ is defined and belongs to $C = \text{Col}(A)$.

First, note that $A^T A \in \mathbb{R}^{m \times m}$, and that by Corollary 6.6.2(a), we have that $\text{rank}(A^T A) = \text{rank}(A) = m$. So, by the Invertible Matrix Theorem, $A^T A$ is invertible, and we see that $(A^T A)^{-1}$ is defined and belongs to $\mathbb{R}^{m \times m}$.

Theorem 6.6.3

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank m (i.e. A is a matrix of full column rank). Then the matrix $A(A^T A)^{-1} A^T$ is the standard matrix of orthogonal projection onto $\text{Col}(A)$, that is, for all $\mathbf{x} \in \mathbb{R}^n$, the orthogonal projection of $\mathbf{x}$ onto $C := \text{Col}(A)$ is given by

$$\mathbf{x}_C \;=\; A(A^T A)^{-1} A^T \mathbf{x}.$$

Proof. Fix $\mathbf{x} \in \mathbb{R}^n$. We must first check that the expression $A(A^T A)^{-1} A^T \mathbf{x}$ is defined and belongs to $C = \text{Col}(A)$.

First, note that $A^T A \in \mathbb{R}^{m \times m}$, and that by Corollary 6.6.2(a), we have that $\text{rank}(A^T A) = \text{rank}(A) = m$. So, by the Invertible Matrix Theorem, $A^T A$ is invertible, and we see that $(A^T A)^{-1}$ is defined and belongs to $\mathbb{R}^{m \times m}$. Since $A \in \mathbb{R}^{n \times m}$, $(A^T A)^{-1} \in \mathbb{R}^{m \times m}$, and $A^T \in \mathbb{R}^{m \times n}$, we see that $A(A^T A)^{-1} A^T \in \mathbb{R}^{n \times n}$;

Theorem 6.6.3

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank m (i.e. A is a matrix of full column rank). Then the matrix $A(A^T A)^{-1} A^T$ is the standard matrix of orthogonal projection onto $\text{Col}(A)$, that is, for all $\mathbf{x} \in \mathbb{R}^n$, the orthogonal projection of $\mathbf{x}$ onto $C := \text{Col}(A)$ is given by

$$\mathbf{x}_C = A(A^T A)^{-1} A^T \mathbf{x}.$$

Proof. Fix $\mathbf{x} \in \mathbb{R}^n$. We must first check that the expression $A(A^T A)^{-1} A^T \mathbf{x}$ is defined and belongs to $C = \text{Col}(A)$.

First, note that $A^T A \in \mathbb{R}^{m \times m}$, and that by Corollary 6.6.2(a), we have that $\text{rank}(A^T A) = \text{rank}(A) = m$. So, by the Invertible Matrix Theorem, $A^T A$ is invertible, and we see that $(A^T A)^{-1}$ is defined and belongs to $\mathbb{R}^{m \times m}$. Since $A \in \mathbb{R}^{n \times m}$, $(A^T A)^{-1} \in \mathbb{R}^{m \times m}$, and $A^T \in \mathbb{R}^{m \times n}$, we see that $A(A^T A)^{-1} A^T \in \mathbb{R}^{n \times n}$; since $\mathbf{x} \in \mathbb{R}^n$, we see that $A(A^T A)^{-1} A^T \mathbf{x}$ is defined and belongs to $\mathbb{R}^n$.

Theorem 6.6.3

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank m (i.e. A is a matrix of full column rank). Then the matrix $A(A^T A)^{-1} A^T$ is the standard matrix of orthogonal projection onto $\text{Col}(A)$, that is, for all $\mathbf{x} \in \mathbb{R}^n$, the orthogonal projection of $\mathbf{x}$ onto $C := \text{Col}(A)$ is given by

$$\mathbf{x}_C = A(A^T A)^{-1} A^T \mathbf{x}.$$

Proof. Fix $\mathbf{x} \in \mathbb{R}^n$. We must first check that the expression $A(A^T A)^{-1} A^T \mathbf{x}$ is defined and belongs to $C = \text{Col}(A)$.

First, note that $A^T A \in \mathbb{R}^{m \times m}$, and that by Corollary 6.6.2(a), we have that $\text{rank}(A^T A) = \text{rank}(A) = m$. So, by the Invertible Matrix Theorem, $A^T A$ is invertible, and we see that $(A^T A)^{-1}$ is defined and belongs to $\mathbb{R}^{m \times m}$. Since $A \in \mathbb{R}^{n \times m}$, $(A^T A)^{-1} \in \mathbb{R}^{m \times m}$, and $A^T \in \mathbb{R}^{m \times n}$, we see that $A(A^T A)^{-1} A^T \in \mathbb{R}^{n \times n}$; since $\mathbf{x} \in \mathbb{R}^n$, we see that $A(A^T A)^{-1} A^T \mathbf{x}$ is defined and belongs to $\mathbb{R}^n$. Meanwhile, $(A^T A)^{-1} A^T \mathbf{x}$ is a vector in $\mathbb{R}^m$, and so (next slide):

Theorem 6.6.3

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank m (i.e. A is a matrix of full column rank). Then the matrix $A(A^TA)^{-1}A^T$ is the standard matrix of orthogonal projection onto $\mathrm{Col}(A)$, that is, for all $\mathbf{x} \in \mathbb{R}^n$, the orthogonal projection of $\mathbf{x}$ onto $C := \mathrm{Col}(A)$ is given by

$$\mathbf{x}_C = A(A^TA)^{-1}A^T\mathbf{x}.$$

Proof (continued).

$$A(A^TA)^{-1}A^T\mathbf{x} = \underbrace{A}_{\in \mathbb{R}^{n \times m}} \Big(\underbrace{(A^TA)^{-1}A^T\mathbf{x}}_{\in \mathbb{R}^m} \Big)$$

is a linear combination of the columns of A. By definition, this means that $A(A^TA)^{-1}A^T\mathbf{x} \in \mathrm{Col}(A) = C$.

Theorem 6.6.3

Let $A \in \mathbb{R}^{n\times m}$ be a matrix of rank m (i.e. A is a matrix of full column rank). Then the matrix $A(A^TA)^{-1}A^T$ is the standard matrix of orthogonal projection onto $\text{Col}(A)$, that is, for all $\mathbf{x} \in \mathbb{R}^n$, the orthogonal projection of $\mathbf{x}$ onto $C := \text{Col}(A)$ is given by

$$\mathbf{x}_C = A(A^TA)^{-1}A^T\mathbf{x}.$$

Proof (continued). In view of Corollary 6.5.3, it is now enough to prove that $(\mathbf{x} - A(A^TA)^{-1}A^T\mathbf{x}) \in C^{\perp}$, for it will then follow that $\mathbf{x}_C = A(A^TA)^{-1}A^T\mathbf{x}$, which is what we need to show.

Theorem 6.6.3

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank m (i.e. A is a matrix of full column rank). Then the matrix $A(A^TA)^{-1}A^T$ is the standard matrix of orthogonal projection onto $\text{Col}(A)$, that is, for all $\mathbf{x} \in \mathbb{R}^n$, the orthogonal projection of $\mathbf{x}$ onto $C := \text{Col}(A)$ is given by

$$\mathbf{x}_C = A(A^TA)^{-1}A^T\mathbf{x}.$$

Proof (continued). In view of Corollary 6.5.3, it is now enough to prove that $(\mathbf{x} - A(A^TA)^{-1}A^T\mathbf{x}) \in C^\perp$, for it will then follow that $\mathbf{x}_C = A(A^TA)^{-1}A^T\mathbf{x}$, which is what we need to show.

- Indeed, if we can show that $(\mathbf{x} - A(A^TA)^{-1}A^T\mathbf{x}) \in C^\perp$, then we get that

$$\mathbf{x} = \underbrace{A(A^TA)^{-1}A^T\mathbf{x}}_{\in C} + (\underbrace{\mathbf{x} - A(A^TA)^{-1}A^T\mathbf{x}}_{\in C^\perp}),$$

  which (by Corollary 6.5.3) implies that $\mathbf{x}_C = A(A^TA)^{-1}A^T\mathbf{x}$ and $\mathbf{x}_{C^\perp} = \mathbf{x} - A(A^TA)^{-1}A^T\mathbf{x}$.

Theorem 6.6.3

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank m (i.e. A is a matrix of full column rank). Then the matrix $A(A^T A)^{-1} A^T$ is the standard matrix of orthogonal projection onto $\text{Col}(A)$, that is, for all $\mathbf{x} \in \mathbb{R}^n$, the orthogonal projection of $\mathbf{x}$ onto $C := \text{Col}(A)$ is given by

$$\mathbf{x}_C = A(A^T A)^{-1} A^T \mathbf{x}.$$

Proof (continued). But note that

$$C^\perp = \text{Col}(A)^\perp = \text{Row}(A^T)^\perp \overset{(*)}{=} \text{Nul}(A^T),$$

where (*) follows from Theorem 6.6.1.

Theorem 6.6.3

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank m (i.e. A is a matrix of full column rank). Then the matrix $A(A^T A)^{-1} A^T$ is the standard matrix of orthogonal projection onto $\text{Col}(A)$, that is, for all $\mathbf{x} \in \mathbb{R}^n$, the orthogonal projection of $\mathbf{x}$ onto $C := \text{Col}(A)$ is given by

$$\mathbf{x}_C = A(A^T A)^{-1} A^T \mathbf{x}.$$

Proof (continued). But note that

$$C^\perp = \text{Col}(A)^\perp = \text{Row}(A^T)^\perp \overset{(*)}{=} \text{Nul}(A^T),$$

where (*) follows from Theorem 6.6.1. So, it in fact suffices to show that the vector $\mathbf{x} - A(A^T A)^{-1} A^T \mathbf{x}$ belongs to $\text{Nul}(A^T)$.

Theorem 6.6.3

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank m (i.e. A is a matrix of full column rank). Then the matrix $A(A^TA)^{-1}A^T$ is the standard matrix of orthogonal projection onto $\text{Col}(A)$, that is, for all $\mathbf{x} \in \mathbb{R}^n$, the orthogonal projection of $\mathbf{x}$ onto $C := \text{Col}(A)$ is given by

$$\mathbf{x}_C = A(A^TA)^{-1}A^T\mathbf{x}.$$

Proof (continued). But note that

$$C^\perp = \text{Col}(A)^\perp = \text{Row}(A^T)^\perp \stackrel{(*)}{=} \text{Nul}(A^T),$$

where (*) follows from Theorem 6.6.1. So, it in fact suffices to show that the vector $\mathbf{x} - A(A^TA)^{-1}A^T\mathbf{x}$ belongs to $\text{Nul}(A^T)$. For this, we compute:

$$A^T\Big(\mathbf{x} - A(A^TA)^{-1}A^T\mathbf{x}\Big) = A^T\mathbf{x} - \underbrace{A^TA(A^TA)^{-1}}_{=I_m}A^T\mathbf{x} = \mathbf{0}.$$

Theorem 6.6.3

Let $A \in \mathbb{R}^{n\times m}$ be a matrix of rank m (i.e. A is a matrix of full column rank). Then the matrix $A(A^TA)^{-1}A^T$ is the standard matrix of orthogonal projection onto $\mathrm{Col}(A)$, that is, for all $\mathbf{x} \in \mathbb{R}^n$, the orthogonal projection of $\mathbf{x}$ onto $C := \mathrm{Col}(A)$ is given by

$$\mathbf{x}_C = A(A^TA)^{-1}A^T\mathbf{x}.$$

Proof (continued). But note that

$$C^\perp = \mathrm{Col}(A)^\perp = \mathrm{Row}(A^T)^\perp \overset{(*)}{=} \mathrm{Nul}(A^T),$$

where (*) follows from Theorem 6.6.1. So, it in fact suffices to show that the vector $\mathbf{x} - A(A^TA)^{-1}A^T\mathbf{x}$ belongs to $\mathrm{Nul}(A^T)$. For this, we compute:

$$A^T\left(\mathbf{x} - A(A^TA)^{-1}A^T\mathbf{x}\right) = A^T\mathbf{x} - \underbrace{A^TA(A^TA)^{-1}}_{=I_m}A^T\mathbf{x} = \mathbf{0}.$$

This proves that $\mathbf{x} - A(A^TA)^{-1}A^T\mathbf{x} \in \mathrm{Nul}(A^T)$, and we are done. □

- **Theorem 6.6.3:** If $A \in \mathbb{R}^{n \times m}$ has full column rank, then $A(A^T A)^{-1} A^T$ is the standard matrix of orthogonal projection onto $\mathrm{Col}(A)$.

- **Theorem 6.6.3:** If $A \in \mathbb{R}^{n \times m}$ has full column rank, then $A(A^T A)^{-1} A^T$ is the standard matrix of orthogonal projection onto $\text{Col}(A)$.
- **Remark:** Suppose that we are given a non-zero matrix $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix}$ in $\mathbb{R}^{n \times m}$, and that we need to compute the standard matrix of orthogonal projection onto $\text{Col}(A)$.

- **Theorem 6.6.3:** If $A \in \mathbb{R}^{n \times m}$ has full column rank, then $A(A^T A)^{-1} A^T$ is the standard matrix of orthogonal projection onto $\text{Col}(A)$.
- **Remark:** Suppose that we are given a non-zero matrix $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix}$ in $\mathbb{R}^{n \times m}$, and that we need to compute the standard matrix of orthogonal projection onto $\text{Col}(A)$.
  - If $\text{rank}(A) = m$ (i.e. A has full column rank), then the matrix that we need is $A(A^T A)^{-1} A^T$, as per Theorem 6.6.3.

- **Theorem 6.6.3:** If $A \in \mathbb{R}^{n \times m}$ has full column rank, then $A(A^T A)^{-1} A^T$ is the standard matrix of orthogonal projection onto $\text{Col}(A)$.
- **Remark:** Suppose that we are given a non-zero matrix $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix}$ in $\mathbb{R}^{n \times m}$, and that we need to compute the standard matrix of orthogonal projection onto $\text{Col}(A)$.
    - If $\text{rank}(A) = m$ (i.e. A has full column rank), then the matrix that we need is $A(A^T A)^{-1} A^T$, as per Theorem 6.6.3.
    - But what if $\text{rank}(A) < m$?

- **Theorem 6.6.3:** If $A \in \mathbb{R}^{n \times m}$ has full column rank, then $A(A^T A)^{-1} A^T$ is the standard matrix of orthogonal projection onto $\text{Col}(A)$.
- **Remark:** Suppose that we are given a non-zero matrix $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix}$ in $\mathbb{R}^{n \times m}$, and that we need to compute the standard matrix of orthogonal projection onto $\text{Col}(A)$.
    - If $\text{rank}(A) = m$ (i.e. A has full column rank), then the matrix that we need is $A(A^T A)^{-1} A^T$, as per Theorem 6.6.3.
    - But what if $\text{rank}(A) < m$?
    - In that case, we let B be the matrix obtained from A by deleting all the non-pivot columns of A.

- **Theorem 6.6.3:** If $A \in \mathbb{R}^{n\times m}$ has full column rank, then $A(A^TA)^{-1}A^T$ is the standard matrix of orthogonal projection onto $\mathrm{Col}(A)$.

- **Remark:** Suppose that we are given a non-zero matrix $A = \left[\begin{array}{ccc} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{array}\right]$ in $\mathbb{R}^{n\times m}$, and that we need to compute the standard matrix of orthogonal projection onto $\mathrm{Col}(A)$.
  - If $\mathrm{rank}(A) = m$ (i.e. A has full column rank), then the matrix that we need is $A(A^TA)^{-1}A^T$, as per Theorem 6.6.3.
  - But what if $\mathrm{rank}(A) < m$?
  - In that case, we let B be the matrix obtained from A by deleting all the non-pivot columns of A.
  - By Theorem 3.3.4, the columns of B form a basis of $\mathrm{Col}(A)$, and we see that $\mathrm{Col}(A) = \mathrm{Col}(B)$.

- **Theorem 6.6.3:** If $A \in \mathbb{R}^{n\times m}$ has full column rank, then $A(A^TA)^{-1}A^T$ is the standard matrix of orthogonal projection onto $\text{Col}(A)$.
- **Remark:** Suppose that we are given a non-zero matrix $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix}$ in $\mathbb{R}^{n\times m}$, and that we need to compute the standard matrix of orthogonal projection onto $\text{Col}(A)$.
  - If $\text{rank}(A) = m$ (i.e. A has full column rank), then the matrix that we need is $A(A^TA)^{-1}A^T$, as per Theorem 6.6.3.
  - But what if $\text{rank}(A) < m$?
  - In that case, we let B be the matrix obtained from A by deleting all the non-pivot columns of A.
  - By Theorem 3.3.4, the columns of B form a basis of $\text{Col}(A)$, and we see that $\text{Col}(A) = \text{Col}(B)$.
  - Moreover, all the columns of B are pivot columns, and so B has full column rank.

- **Theorem 6.6.3:** If $A \in \mathbb{R}^{n \times m}$ has full column rank, then $A(A^T A)^{-1} A^T$ is the standard matrix of orthogonal projection onto $\text{Col}(A)$.

- **Remark:** Suppose that we are given a non-zero matrix $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_m \end{bmatrix}$ in $\mathbb{R}^{n \times m}$, and that we need to compute the standard matrix of orthogonal projection onto $\text{Col}(A)$.
    - If $\text{rank}(A) = m$ (i.e. A has full column rank), then the matrix that we need is $A(A^T A)^{-1} A^T$, as per Theorem 6.6.3.
    - But what if $\text{rank}(A) < m$?
    - In that case, we let B be the matrix obtained from A by deleting all the non-pivot columns of A.
    - By Theorem 3.3.4, the columns of B form a basis of $\text{Col}(A)$, and we see that $\text{Col}(A) = \text{Col}(B)$.
    - Moreover, all the columns of B are pivot columns, and so B has full column rank.
    - But now the matrix B satisfies the hypotheses of Theorem 6.6.3. So, the standard matrix of orthogonal projection onto $\text{Col}(A) = \text{Col}(B)$ is $B(B^T B)^{-1} B^T$.

Theorem 6.6.3

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank m (i.e. A is a matrix of full column rank). Then the matrix $A(A^T A)^{-1} A^T$ is the standard matrix of orthogonal projection onto $\mathrm{Col}(A)$, that is, for all $\mathbf{x} \in \mathbb{R}^n$, the orthogonal projection of $\mathbf{x}$ onto $C := \mathrm{Col}(A)$ is given by

$$\mathbf{x}_C = A(A^T A)^{-1} A^T \mathbf{x}.$$

Theorem 6.6.3

Let $A \in \mathbb{R}^{n\times m}$ be a matrix of rank m (i.e. A is a matrix of full column rank). Then the matrix $A(A^TA)^{-1}A^T$ is the standard matrix of orthogonal projection onto $\mathrm{Col}(A)$, that is, for all $\mathbf{x} \in \mathbb{R}^n$, the orthogonal projection of $\mathbf{x}$ onto $C := \mathrm{Col}(A)$ is given by

$$\mathbf{x}_C = A(A^TA)^{-1}A^T\mathbf{x}.$$

Corollary 6.6.4

Let $\mathbf{a}$ be a non-zero vector in $\mathbb{R}^n$. Then the standard matrix of projection onto the line $L := \mathrm{Span}(\mathbf{a})$ is the matrix

$$\mathbf{a}(\mathbf{a}^T\mathbf{a})^{-1}\mathbf{a}^T = \mathbf{a}(\mathbf{a}\cdot\mathbf{a})^{-1}\mathbf{a}^T = \frac{1}{\mathbf{a}\cdot\mathbf{a}}\mathbf{a}\mathbf{a}^T.$$

Consequently, for every vector $\mathbf{x} \in \mathbb{R}^n$, we have that

$$\mathbf{x}_L = \mathrm{proj}_L(\mathbf{x}) = \frac{1}{\mathbf{a}\cdot\mathbf{a}}\mathbf{a}\mathbf{a}^T\mathbf{x}.$$

Proof. This is a special case of Theorem 6.6.3 for $A = \begin{bmatrix} \mathbf{a} \end{bmatrix}$. $\square$

Theorem 6.6.5

Let U be a subspace of $\mathbb{R}^n$, and let $P \in \mathbb{R}^{n\times n}$ be the standard matrix of proj_U. Then $I_n - P$ is the standard matrix of $\text{proj}_{U^\perp}$, that is, for all $\mathbf{x} \in \mathbb{R}^n$, the orthogonal projection of $\mathbf{x}$ onto $U^\perp$ is given by $\mathbf{x}_{U^\perp} = (I_n - P)\mathbf{x}$.

Proof.

Theorem 6.6.5

Let U be a subspace of $\mathbb{R}^n$, and let $P \in \mathbb{R}^{n \times n}$ be the standard matrix of proj_U. Then $I_n - P$ is the standard matrix of $\text{proj}_{U^\perp}$, that is, for all $\mathbf{x} \in \mathbb{R}^n$, the orthogonal projection of $\mathbf{x}$ onto $U^\perp$ is given by $\mathbf{x}_{U^\perp} = (I_n - P)\mathbf{x}$.

Proof. We observe that for all $\mathbf{x} \in \mathbb{R}^n$, we have that

$$(I_n - P)\mathbf{x} \;=\; I_n\mathbf{x} - P\mathbf{x} \;\stackrel{(*)}{=}\; \mathbf{x} - \mathbf{x}_U \;\stackrel{(**)}{=}\; \mathbf{x}_{U^\perp},$$

where (*) follows from the fact that P is the standard matrix of proj_U, and (**) follows from Corollary 6.5.3. So, $I_n - P$ is indeed the standard matrix of $\text{proj}_{U^\perp}$. □

- Theorem 6.6.5: if $P \in \mathbb{R}^{n \times n}$ is the standard matrix of proj_U, then $I_n - P$ is the standard matrix of $\text{proj}_{U^\perp}$.

- Theorem 6.6.5: if $P \in \mathbb{R}^{n \times n}$ is the standard matrix of proj_U, then $I_n - P$ is the standard matrix of $\text{proj}_{U^\perp}$.

Corollary 6.6.6

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank n (i.e. A is a matrix of full row rank). Then the matrix $I_m - A^T(AA^T)^{-1}A$ is the standard matrix of orthogonal projection onto $N := \text{Nul}(A)$, that is, for all $\mathbf{x} \in \mathbb{R}^m$, the orthogonal projection of $\mathbf{x}$ onto N is given by
$\mathbf{x}_N = (I_m - A^T(AA^T)^{-1}A)\mathbf{x}$.

Proof.

- Theorem 6.6.5: if $P \in \mathbb{R}^{n\times n}$ is the standard matrix of proj_U, then $I_n - P$ is the standard matrix of $\text{proj}_{U^\perp}$.

Corollary 6.6.6

Let $A \in \mathbb{R}^{n\times m}$ be a matrix of rank n (i.e. A is a matrix of full row rank). Then the matrix $I_m - A^T(AA^T)^{-1}A$ is the standard matrix of orthogonal projection onto $N := \text{Nul}(A)$, that is, for all $\mathbf{x} \in \mathbb{R}^m$, the orthogonal projection of $\mathbf{x}$ onto N is given by $\mathbf{x}_N = (I_m - A^T(AA^T)^{-1}A)\mathbf{x}$.

Proof. First, note that

$$\text{Nul}(A) \overset{(*)}{=} \text{Row}(A)^\perp = \text{Col}(A^T)^\perp.$$

where (*) follows from Theorem 6.6.1.

- Theorem 6.6.5: if $P \in \mathbb{R}^{n\times n}$ is the standard matrix of proj_U, then $I_n - P$ is the standard matrix of $\mathrm{proj}_{U^\perp}$.

Corollary 6.6.6

Let $A \in \mathbb{R}^{n\times m}$ be a matrix of rank n (i.e. A is a matrix of full row rank). Then the matrix $I_m - A^T(AA^T)^{-1}A$ is the standard matrix of orthogonal projection onto $N := \mathrm{Nul}(A)$, that is, for all $\mathbf{x} \in \mathbb{R}^m$, the orthogonal projection of $\mathbf{x}$ onto N is given by
$\mathbf{x}_N = (I_m - A^T(AA^T)^{-1}A)\mathbf{x}$.

Proof. First, note that

$$\mathrm{Nul}(A) \overset{(*)}{=} \mathrm{Row}(A)^\perp = \mathrm{Col}(A^T)^\perp.$$

where (*) follows from Theorem 6.6.1. Note further that $A^T \in \mathbb{R}^{m\times n}$ and that $\mathrm{rank}(A^T) = \mathrm{rank}(A) = n$, i.e. A^T has full column rank.

- Theorem 6.6.5: if $P \in \mathbb{R}^{n \times n}$ is the standard matrix of proj_U, then $I_n - P$ is the standard matrix of $\text{proj}_{U^\perp}$.

Corollary 6.6.6

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank n (i.e. A is a matrix of full row rank). Then the matrix $I_m - A^T(AA^T)^{-1}A$ is the standard matrix of orthogonal projection onto $N := \text{Nul}(A)$, that is, for all $\mathbf{x} \in \mathbb{R}^m$, the orthogonal projection of $\mathbf{x}$ onto N is given by
$\mathbf{x}_N = (I_m - A^T(AA^T)^{-1}A)\mathbf{x}$.

Proof. First, note that

$$\text{Nul}(A) \overset{(*)}{=} \text{Row}(A)^\perp = \text{Col}(A^T)^\perp.$$

where (*) follows from Theorem 6.6.1. Note further that $A^T \in \mathbb{R}^{m \times n}$ and that $\text{rank}(A^T) = \text{rank}(A) = n$, i.e. A^T has full column rank. So, by Theorem 6.6.3, the standard matrix of orthogonal projection onto $\text{Col}(A^T)$ is $A^T(AA^T)^{-1}A$.

- Theorem 6.6.5: if $P \in \mathbb{R}^{n \times n}$ is the standard matrix of proj_U, then $I_n - P$ is the standard matrix of $\mathrm{proj}_{U^\perp}$.

Corollary 6.6.6

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank n (i.e. A is a matrix of full row rank). Then the matrix $I_m - A^T(AA^T)^{-1}A$ is the standard matrix of orthogonal projection onto $N := \mathrm{Nul}(A)$, that is, for all $\mathbf{x} \in \mathbb{R}^m$, the orthogonal projection of $\mathbf{x}$ onto N is given by
$\mathbf{x}_N = (I_m - A^T(AA^T)^{-1}A)\mathbf{x}$.

Proof. First, note that

$$\mathrm{Nul}(A) \overset{(*)}{=} \mathrm{Row}(A)^\perp = \mathrm{Col}(A^T)^\perp.$$

where (*) follows from Theorem 6.6.1. Note further that $A^T \in \mathbb{R}^{m \times n}$ and that $\mathrm{rank}(A^T) = \mathrm{rank}(A) = n$, i.e. A^T has full column rank. So, by Theorem 6.6.3, the standard matrix of orthogonal projection onto $\mathrm{Col}(A^T)$ is $A^T(AA^T)^{-1}A$. Finally, by Theorem 6.6.5, the standard matrix of orthogonal projection onto $\mathrm{Col}(A^T)^\perp = \mathrm{Nul}(A)$ is $I_m - A^T(AA^T)^{-1}A$. $\square$

3. The method of least squares

3. The method of least squares

- In some real-world applications, we may be interested in finding the best **approximate** solution to a (possibly inconsistent) matrix-vector equation $A\mathbf{x} = \mathbf{b}$.

3. The method of least squares

- In some real-world applications, we may be interested in finding the best **approximate** solution to a (possibly inconsistent) matrix-vector equation $A\mathbf{x} = \mathbf{b}$.
- More formally, suppose we are given a norm $||\cdot||$ on $\mathbb{R}^n$, a matrix $A \in \mathbb{R}^{n\times m}$, and a vector $\mathbf{b} \in \mathbb{R}^n$.

3. The method of least squares

- In some real-world applications, we may be interested in finding the best **approximate** solution to a (possibly inconsistent) matrix-vector equation $A\mathbf{x} = \mathbf{b}$.
- More formally, suppose we are given a norm $||\cdot||$ on $\mathbb{R}^n$, a matrix $A \in \mathbb{R}^{n\times m}$, and a vector $\mathbf{b} \in \mathbb{R}^n$.
- We would then like to find a vector $\mathbf{x}$ for which

$$||A\mathbf{x} - \mathbf{b}||$$

is as small as possible.

3. The method of least squares

- In some real-world applications, we may be interested in finding the best **approximate** solution to a (possibly inconsistent) matrix-vector equation $A\mathbf{x} = \mathbf{b}$.
- More formally, suppose we are given a norm $||\cdot||$ on $\mathbb{R}^n$, a matrix $A \in \mathbb{R}^{n\times m}$, and a vector $\mathbf{b} \in \mathbb{R}^n$.
- We would then like to find a vector $\mathbf{x}$ for which

$$||A\mathbf{x} - \mathbf{b}||$$

  is as small as possible.
- If $A\mathbf{x} = \mathbf{b}$ is consistent, then any solution of that equation will minimize $||A\mathbf{x} - \mathbf{b}||$.

3. The method of least squares

- In some real-world applications, we may be interested in finding the best **approximate** solution to a (possibly inconsistent) matrix-vector equation $A\mathbf{x} = \mathbf{b}$.
- More formally, suppose we are given a norm $||\cdot||$ on $\mathbb{R}^n$, a matrix $A \in \mathbb{R}^{n\times m}$, and a vector $\mathbf{b} \in \mathbb{R}^n$.
- We would then like to find a vector $\mathbf{x}$ for which

$$||A\mathbf{x} - \mathbf{b}||$$

  is as small as possible.
- If $A\mathbf{x} = \mathbf{b}$ is consistent, then any solution of that equation will minimize $||A\mathbf{x} - \mathbf{b}||$.
- However, what if the equation $A\mathbf{x} = \mathbf{b}$ is inconsistent?

3. The method of least squares

- In some real-world applications, we may be interested in finding the best **approximate** solution to a (possibly inconsistent) matrix-vector equation $A\mathbf{x} = \mathbf{b}$.
- More formally, suppose we are given a norm $||\cdot||$ on $\mathbb{R}^n$, a matrix $A \in \mathbb{R}^{n\times m}$, and a vector $\mathbf{b} \in \mathbb{R}^n$.
- We would then like to find a vector $\mathbf{x}$ for which

$$||A\mathbf{x} - \mathbf{b}||$$

  is as small as possible.
- If $A\mathbf{x} = \mathbf{b}$ is consistent, then any solution of that equation will minimize $||A\mathbf{x} - \mathbf{b}||$.
- However, what if the equation $A\mathbf{x} = \mathbf{b}$ is inconsistent?
- Then the answer will obviously depend on which norm that we are using.

- In what follows, we will work only with the **norm induced by the standard scalar product** in $\mathbb{R}^n$, i.e. the standard Euclidean norm.
  - Recall that this is the norm $||\cdot||$ given by

$$||\mathbf{x}|| = \sqrt{\mathbf{x}\cdot\mathbf{x}} = \sqrt{x_1^2+\cdots+x_n^2}$$

    for all vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ in $\mathbb{R}^n$.

Theorem 6.7.1

Let $A \in \mathbb{R}^{n \times m}$ and $\mathbf{b} \in \mathbb{R}^n$. Then the matrix-vector equation

$$A^T A\mathbf{x} \;=\; A^T\mathbf{b}$$

is consistent, and moreover, its solution set is precisely the set of vectors $\mathbf{x}$ in $\mathbb{R}^m$ that minimize the expression

$$||A\mathbf{x} - \mathbf{b}||.$$

- **Terminology:** Suppose we are given a matrix $A \in \mathbb{R}^{n \times m}$ and a vector $\mathbf{b} \in \mathbb{R}^n$.
    - Vectors $\mathbf{x} \in \mathbb{R}^m$ that minimize the expression $||A\mathbf{x} - \mathbf{b}||$ are called the *least-squares solutions* of the equation $A\mathbf{x} = \mathbf{b}$ (such solutions are often denoted by $\hat{\mathbf{x}}$), whereas the number

      $$\min_{\mathbf{x} \in \mathbb{R}^m} ||A\mathbf{x} - \mathbf{b}||$$

      is called the *least-squares error* for the equation $A\mathbf{x} = \mathbf{b}$.
    - By Theorem 6.7.1, the equation $A\mathbf{x} = \mathbf{b}$ has at least one least-squares solution $\hat{\mathbf{x}}$, and consequently, the least-squares error is defined and is equal to $||A\hat{\mathbf{x}} - \mathbf{b}||$.

Theorem 6.7.1

Let $A \in \mathbb{R}^{n \times m}$ and $\mathbf{b} \in \mathbb{R}^n$. Then the matrix-vector equation

$$A^T A\mathbf{x} \;=\; A^T\mathbf{b}$$

is consistent, and moreover, its solution set is precisely the set of vectors $\mathbf{x}$ in $\mathbb{R}^m$ that minimize the expression

$$||A\mathbf{x} - \mathbf{b}||.$$

- **Remark:** Obviously, if $A\mathbf{x} = \mathbf{b}$ is consistent, then the least-squares solutions of $A\mathbf{x} = \mathbf{b}$ are precisely the solutions of the equation $A\mathbf{x} = \mathbf{b}$ itself.

Theorem 6.7.1

Let $A \in \mathbb{R}^{n \times m}$ and $\mathbf{b} \in \mathbb{R}^n$. Then the matrix-vector equation

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

is consistent, and moreover, its solution set is precisely the set of vectors $\mathbf{x}$ in $\mathbb{R}^m$ that minimize the expression

$$||A\mathbf{x} - \mathbf{b}||.$$

- **Remark:** Obviously, if $A\mathbf{x} = \mathbf{b}$ is consistent, then the least-squares solutions of $A\mathbf{x} = \mathbf{b}$ are precisely the solutions of the equation $A\mathbf{x} = \mathbf{b}$ itself.
  - This is because if $A\mathbf{x} = \mathbf{b}$ is consistent, then the solutions of that equation minimize the expression $||A\mathbf{x} - \mathbf{b}||$ (indeed, $||A\mathbf{x} - \mathbf{b}|| = 0$ iff $A\mathbf{x} = \mathbf{b}$).

Theorem 6.7.1

Let $A \in \mathbb{R}^{n \times m}$ and $\mathbf{b} \in \mathbb{R}^n$. Then the matrix-vector equation

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

is consistent, and moreover, its solution set is precisely the set of vectors $\mathbf{x}$ in $\mathbb{R}^m$ that minimize the expression

$$||A\mathbf{x} - \mathbf{b}||.$$

- **Remark:** Obviously, if $A\mathbf{x} = \mathbf{b}$ is consistent, then the least-squares solutions of $A\mathbf{x} = \mathbf{b}$ are precisely the solutions of the equation $A\mathbf{x} = \mathbf{b}$ itself.
  - This is because if $A\mathbf{x} = \mathbf{b}$ is consistent, then the solutions of that equation minimize the expression $||A\mathbf{x} - \mathbf{b}||$ (indeed, $||A\mathbf{x} - \mathbf{b}|| = 0$ iff $A\mathbf{x} = \mathbf{b}$).
  - Moreover, the matrix-vector equation $A\mathbf{x} = \mathbf{b}$ is consistent iff the least-squares error of this equation is zero.

Theorem 6.7.1

Let $A \in \mathbb{R}^{n\times m}$ and $\mathbf{b} \in \mathbb{R}^n$. Then the matrix-vector equation

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

is consistent, and moreover, its solution set is precisely the set of vectors $\mathbf{x}$ in $\mathbb{R}^m$ that minimize the expression

$$||A\mathbf{x} - \mathbf{b}||.$$

- First an example, then a proof.

Example 6.7.2

Let

$$A = \begin{bmatrix} 1 & -2 \\ -1 & 2 \\ 0 & 3 \\ 2 & 5 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 3 \\ 1 \\ -4 \\ 2 \end{bmatrix},$$

with entries understood to be in $\mathbb{R}$. Find all least-squares solutions $\hat{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$, as well as the least-squares error. Is the equation $A\mathbf{x} = \mathbf{b}$ consistent?

Solution.

Example 6.7.2

Let

$$A = \begin{bmatrix} 1 & -2 \\ -1 & 2 \\ 0 & 3 \\ 2 & 5 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 3 \\ 1 \\ -4 \\ 2 \end{bmatrix},$$

with entries understood to be in $\mathbb{R}$. Find all least-squares solutions $\hat{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$, as well as the least-squares error. Is the equation $A\mathbf{x} = \mathbf{b}$ consistent?

Solution. We apply Theorem 6.7.1.

Example 6.7.2

Let

$$A = \begin{bmatrix} 1 & -2 \\ -1 & 2 \\ 0 & 3 \\ 2 & 5 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 3 \\ 1 \\ -4 \\ 2 \end{bmatrix},$$

with entries understood to be in $\mathbb{R}$. Find all least-squares solutions $\hat{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$, as well as the least-squares error. Is the equation $A\mathbf{x} = \mathbf{b}$ consistent?

Solution. We apply Theorem 6.7.1. So, we need to find the solutions of the equation $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$.

Example 6.7.2

Let

$$A = \begin{bmatrix} 1 & -2 \\ -1 & 2 \\ 0 & 3 \\ 2 & 5 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 3 \\ 1 \\ -4 \\ 2 \end{bmatrix},$$

with entries understood to be in $\mathbb{R}$. Find all least-squares solutions $\hat{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$, as well as the least-squares error. Is the equation $A\mathbf{x} = \mathbf{b}$ consistent?

Solution. We apply Theorem 6.7.1. So, we need to find the solutions of the equation $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$. We first compute

$$A^T A = \begin{bmatrix} 6 & 6 \\ 6 & 42 \end{bmatrix} \quad \text{and} \quad A^T\mathbf{b} = \begin{bmatrix} 6 \\ -6 \end{bmatrix},$$

Example 6.7.2

Let

$$A = \begin{bmatrix} 1 & -2 \\ -1 & 2 \\ 0 & 3 \\ 2 & 5 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 3 \\ 1 \\ -4 \\ 2 \end{bmatrix},$$

with entries understood to be in $\mathbb{R}$. Find all least-squares solutions $\hat{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$, as well as the least-squares error. Is the equation $A\mathbf{x} = \mathbf{b}$ consistent?

Solution. We apply Theorem 6.7.1. So, we need to find the solutions of the equation $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$. We first compute

$$A^T A = \begin{bmatrix} 6 & 6 \\ 6 & 42 \end{bmatrix} \quad \text{and} \quad A^T\mathbf{b} = \begin{bmatrix} 6 \\ -6 \end{bmatrix},$$

and then we compute

$$\text{RREF}\left(\left[\begin{array}{c:c} A^T A & A^T\mathbf{b} \end{array} \right] \right) = \left[\begin{array}{cc:c} 1 & 0 & 4/3 \\ 0 & 1 & -1/3 \end{array} \right].$$

Solution (continued). Reminder: We need to solve $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$; we computed $\text{RREF}\Big(\left[\begin{array}{c:c} A^T A & A^T\mathbf{b}\end{array}\right]\Big) = \left[\begin{array}{cc:c} 1 & 0 & 4/3 \\ 0 & 1 & -1/3\end{array}\right]$.

Solution (continued). Reminder: We need to solve $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$; we computed $\text{RREF}\Big(\left[\begin{array}{c:c} A^T A & A^T\mathbf{b} \end{array}\right]\Big) = \left[\begin{array}{cc:c} 1 & 0 & 4/3 \\ 0 & 1 & -1/3 \end{array}\right]$.

It follows that

$$\hat{\mathbf{x}} \quad = \quad \begin{bmatrix} 4/3 \\ -1/3 \end{bmatrix}$$

is the unique solution of the matrix-vector equation $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$, and consequently, the unique least-squares solution of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$.

Solution (continued). Reminder: We need to solve $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$; we computed $\text{RREF}\left(\left[\begin{array}{c:c} A^TA & A^T\mathbf{b}\end{array}\right]\right) = \left[\begin{array}{cc:c} 1 & 0 & 4/3 \\ 0 & 1 & -1/3 \end{array}\right]$.

It follows that

$$\hat{\mathbf{x}} = \begin{bmatrix} 4/3 \\ -1/3 \end{bmatrix}$$

is the unique solution of the matrix-vector equation $A^TA\hat{\mathbf{x}} = A^T\mathbf{b}$, and consequently, the unique least-squares solution of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$.

The least-squares error of $A\mathbf{x} = \mathbf{b}$ is

$$||A\hat{\mathbf{x}} - \mathbf{b}|| = ||\begin{bmatrix} 1 & -2 \\ -1 & 2 \\ 0 & 3 \\ 2 & 5 \end{bmatrix}\begin{bmatrix} 4/3 \\ -1/3 \end{bmatrix} - \begin{bmatrix} 3 \\ 1 \\ -4 \\ 2 \end{bmatrix}|| = ||\begin{bmatrix} -1 \\ -3 \\ 3 \\ -1 \end{bmatrix}|| = 2\sqrt{5}.$$

Solution (continued). Reminder: We need to solve $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$; we computed $\text{RREF}\Big(\left[\begin{array}{c|c} A^T A & A^T\mathbf{b}\end{array}\right]\Big) = \left[\begin{array}{cc|c} 1 & 0 & 4/3 \\ 0 & 1 & -1/3\end{array}\right]$.

It follows that

$$\hat{\mathbf{x}} = \begin{bmatrix} 4/3 \\ -1/3 \end{bmatrix}$$

is the unique solution of the matrix-vector equation $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$, and consequently, the unique least-squares solution of the matrix-vector equation $A\mathbf{x} = \mathbf{b}$.

The least-squares error of $A\mathbf{x} = \mathbf{b}$ is

$$||A\hat{\mathbf{x}} - \mathbf{b}|| = ||\begin{bmatrix} 1 & -2 \\ -1 & 2 \\ 0 & 3 \\ 2 & 5 \end{bmatrix}\begin{bmatrix} 4/3 \\ -1/3 \end{bmatrix} - \begin{bmatrix} 3 \\ 1 \\ -4 \\ 2 \end{bmatrix}|| = ||\begin{bmatrix} -1 \\ -3 \\ 3 \\ -1 \end{bmatrix}|| = 2\sqrt{5}.$$

Since the least-squares error of the equation $A\mathbf{x} = \mathbf{b}$ is strictly positive, we see that the equation is inconsistent. $\square$

Theorem 6.7.1

Let $A \in \mathbb{R}^{n \times m}$ and $\mathbf{b} \in \mathbb{R}^n$. Then the matrix-vector equation

$$A^T A\mathbf{x} \quad = \quad A^T\mathbf{b}$$

is consistent, and moreover, its solution set is precisely the set of vectors $\mathbf{x}$ in $\mathbb{R}^m$ that minimize the expression

$$||A\mathbf{x} - \mathbf{b}||.$$

Proof.

Theorem 6.7.1

Let $A \in \mathbb{R}^{n \times m}$ and $\mathbf{b} \in \mathbb{R}^n$. Then the matrix-vector equation

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

is consistent, and moreover, its solution set is precisely the set of vectors $\mathbf{x}$ in $\mathbb{R}^m$ that minimize the expression

$$||A\mathbf{x} - \mathbf{b}||.$$

Proof. We are looking for vectors $\mathbf{x} \in \mathbb{R}^m$ that minimize the expression $||A\mathbf{x} - \mathbf{b}||$.

Theorem 6.7.1

Let $A \in \mathbb{R}^{n \times m}$ and $\mathbf{b} \in \mathbb{R}^n$. Then the matrix-vector equation

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

is consistent, and moreover, its solution set is precisely the set of vectors $\mathbf{x}$ in $\mathbb{R}^m$ that minimize the expression

$$||A\mathbf{x} - \mathbf{b}||.$$

Proof. We are looking for vectors $\mathbf{x} \in \mathbb{R}^m$ that minimize the expression $||A\mathbf{x} - \mathbf{b}||$. Our goal is to show is that the vectors we are looking for are precisely those that satisfy $A^T A\mathbf{x} = A^T\mathbf{b}$.

Proof (continued). By Proposition 3.3.2(a), we have that $C := \text{Col}(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^m\}$.

Proof (continued). By Proposition 3.3.2(a), we have that $C := \text{Col}(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^m\}$. So, we are in fact looking for the solutions $\mathbf{x}$ of the equation $A\mathbf{x} = \mathbf{b}_C$, because by the definition of $\mathbf{b}_C$, such $\mathbf{x}$'s are precisely the ones for which $||A\mathbf{x} - \mathbf{b}||$ is minimized.

Proof (continued). By Proposition 3.3.2(a), we have that $C := \text{Col}(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^m\}$. So, we are in fact looking for the solutions $\mathbf{x}$ of the equation $A\mathbf{x} = \mathbf{b}_C$, because by the definition of $\mathbf{b}_C$, such $\mathbf{x}$'s are precisely the ones for which $||A\mathbf{x} - \mathbf{b}||$ is minimized.

Moreover, by Corollary 6.5.3, $\mathbf{b} = \mathbf{b}_C + \mathbf{b}_{C^\perp}$ is the only way to decompose $\mathbf{b}$ as a sum of a vector in C and a vector in $C^\perp$.

Proof (continued). By Proposition 3.3.2(a), we have that $C := \text{Col}(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^m\}$. So, we are in fact looking for the solutions $\mathbf{x}$ of the equation $A\mathbf{x} = \mathbf{b}_C$, because by the definition of $\mathbf{b}_C$, such $\mathbf{x}$'s are precisely the ones for which $||A\mathbf{x} - \mathbf{b}||$ is minimized.

Moreover, by Corollary 6.5.3, $\mathbf{b} = \mathbf{b}_C + \mathbf{b}_{C^\perp}$ is the only way to decompose $\mathbf{b}$ as a sum of a vector in C and a vector in $C^\perp$. So, we are looking for those $\mathbf{x}$'s for which $\mathbf{b} - A\mathbf{x} \in C^\perp$.

Proof (continued). By Proposition 3.3.2(a), we have that $C := \text{Col}(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^m\}$. So, we are in fact looking for the solutions $\mathbf{x}$ of the equation $A\mathbf{x} = \mathbf{b}_C$, because by the definition of $\mathbf{b}_C$, such $\mathbf{x}$'s are precisely the ones for which $||A\mathbf{x} - \mathbf{b}||$ is minimized.

Moreover, by Corollary 6.5.3, $\mathbf{b} = \mathbf{b}_C + \mathbf{b}_{C^\perp}$ is the only way to decompose $\mathbf{b}$ as a sum of a vector in C and a vector in $C^\perp$. So, we are looking for those $\mathbf{x}$'s for which $\mathbf{b} - A\mathbf{x} \in C^\perp$. But note that

$$C^\perp \;=\; \text{Col}(A)^\perp \;=\; \text{Row}(A^T)^\perp \;\overset{(*)}{=}\; \text{Nul}(A^T),$$

where (*) follows from Theorem 6.6.1.

Proof (continued). By Proposition 3.3.2(a), we have that $C := \mathrm{Col}(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^m\}$. So, we are in fact looking for the solutions $\mathbf{x}$ of the equation $A\mathbf{x} = \mathbf{b}_C$, because by the definition of $\mathbf{b}_C$, such $\mathbf{x}$'s are precisely the ones for which $||A\mathbf{x} - \mathbf{b}||$ is minimized.

Moreover, by Corollary 6.5.3, $\mathbf{b} = \mathbf{b}_C + \mathbf{b}_{C^\perp}$ is the only way to decompose $\mathbf{b}$ as a sum of a vector in C and a vector in $C^\perp$. So, we are looking for those $\mathbf{x}$'s for which $\mathbf{b} - A\mathbf{x} \in C^\perp$. But note that

$$C^\perp \;=\; \mathrm{Col}(A)^\perp \;=\; \mathrm{Row}(A^T)^\perp \;\stackrel{(*)}{=}\; \mathrm{Nul}(A^T),$$

where (*) follows from Theorem 6.6.1. So, we in fact looking for vectors $\mathbf{x}$ for which $\mathbf{b} - A\mathbf{x} \in \mathrm{Nul}(A^T)$, i.e. those that satisfy $A^T(\mathbf{b} - A\mathbf{x}) = \mathbf{0}$, which is obviously equivalent to $A^TA\mathbf{x} = A^T\mathbf{b}$.

Proof (continued). By Proposition 3.3.2(a), we have that $C := \text{Col}(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^m\}$. So, we are in fact looking for the solutions $\mathbf{x}$ of the equation $A\mathbf{x} = \mathbf{b}_C$, because by the definition of $\mathbf{b}_C$, such $\mathbf{x}$'s are precisely the ones for which $||A\mathbf{x} - \mathbf{b}||$ is minimized.

Moreover, by Corollary 6.5.3, $\mathbf{b} = \mathbf{b}_C + \mathbf{b}_{C^\perp}$ is the only way to decompose $\mathbf{b}$ as a sum of a vector in C and a vector in $C^\perp$. So, we are looking for those $\mathbf{x}$'s for which $\mathbf{b} - A\mathbf{x} \in C^\perp$. But note that

$$C^\perp \;=\; \text{Col}(A)^\perp \;=\; \text{Row}(A^T)^\perp \;\overset{(*)}{=}\; \text{Nul}(A^T),$$

where (*) follows from Theorem 6.6.1. So, we in fact looking for vectors $\mathbf{x}$ for which $\mathbf{b} - A\mathbf{x} \in \text{Nul}(A^T)$, i.e. those that satisfy $A^T(\mathbf{b} - A\mathbf{x}) = \mathbf{0}$, which is obviously equivalent to $A^T A\mathbf{x} = A^T\mathbf{b}$.

It remains to show that the equation $A^T A\mathbf{x} = A^T\mathbf{b}$ is consistent.

Proof (continued). By Proposition 3.3.2(a), we have that $C := \text{Col}(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^m\}$. So, we are in fact looking for the solutions $\mathbf{x}$ of the equation $A\mathbf{x} = \mathbf{b}_C$, because by the definition of $\mathbf{b}_C$, such $\mathbf{x}$'s are precisely the ones for which $||A\mathbf{x} - \mathbf{b}||$ is minimized.

Moreover, by Corollary 6.5.3, $\mathbf{b} = \mathbf{b}_C + \mathbf{b}_{C^\perp}$ is the only way to decompose $\mathbf{b}$ as a sum of a vector in C and a vector in $C^\perp$. So, we are looking for those $\mathbf{x}$'s for which $\mathbf{b} - A\mathbf{x} \in C^\perp$. But note that

$$C^\perp \;=\; \text{Col}(A)^\perp \;=\; \text{Row}(A^T)^\perp \;\stackrel{(*)}{=}\; \text{Nul}(A^T),$$

where (*) follows from Theorem 6.6.1. So, we in fact looking for vectors $\mathbf{x}$ for which $\mathbf{b} - A\mathbf{x} \in \text{Nul}(A^T)$, i.e. those that satisfy $A^T(\mathbf{b} - A\mathbf{x}) = \mathbf{0}$, which is obviously equivalent to $A^TA\mathbf{x} = A^T\mathbf{b}$.

It remains to show that the equation $A^TA\mathbf{x} = A^T\mathbf{b}$ is consistent. By our argument above, a vector $\mathbf{x} \in \mathbb{R}^m$ satisfies $A^TA\mathbf{x} = A^T\mathbf{b}$ iff it satisfies the equation $A\mathbf{x} = \mathbf{b}_C$.

Proof (continued). By Proposition 3.3.2(a), we have that $C := \text{Col}(A) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^m\}$. So, we are in fact looking for the solutions $\mathbf{x}$ of the equation $A\mathbf{x} = \mathbf{b}_C$, because by the definition of $\mathbf{b}_C$, such $\mathbf{x}$'s are precisely the ones for which $||A\mathbf{x} - \mathbf{b}||$ is minimized.

Moreover, by Corollary 6.5.3, $\mathbf{b} = \mathbf{b}_C + \mathbf{b}_{C^\perp}$ is the only way to decompose $\mathbf{b}$ as a sum of a vector in C and a vector in $C^\perp$. So, we are looking for those $\mathbf{x}$'s for which $\mathbf{b} - A\mathbf{x} \in C^\perp$. But note that

$$C^\perp \;=\; \text{Col}(A)^\perp \;=\; \text{Row}(A^T)^\perp \;\overset{(*)}{=}\; \text{Nul}(A^T),$$

where (*) follows from Theorem 6.6.1. So, we in fact looking for vectors $\mathbf{x}$ for which $\mathbf{b} - A\mathbf{x} \in \text{Nul}(A^T)$, i.e. those that satisfy $A^T(\mathbf{b} - A\mathbf{x}) = \mathbf{0}$, which is obviously equivalent to $A^T A\mathbf{x} = A^T\mathbf{b}$.

It remains to show that the equation $A^T A\mathbf{x} = A^T\mathbf{b}$ is consistent. By our argument above, a vector $\mathbf{x} \in \mathbb{R}^m$ satisfies $A^T A\mathbf{x} = A^T\mathbf{b}$ iff it satisfies the equation $A\mathbf{x} = \mathbf{b}_C$. Since the latter equation is consistent (this follows from the definition of C and the existence of $\mathbf{b}_C$), so is the former. □

Theorem 6.7.1

Let $A \in \mathbb{R}^{n \times m}$ and $\mathbf{b} \in \mathbb{R}^n$. Then the matrix-vector equation

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

is consistent, and moreover, its solution set is precisely the set of vectors $\mathbf{x}$ in $\mathbb{R}^m$ that minimize the expression

$$||A\mathbf{x} - \mathbf{b}||.$$

4. Data fitting

4. Data fitting

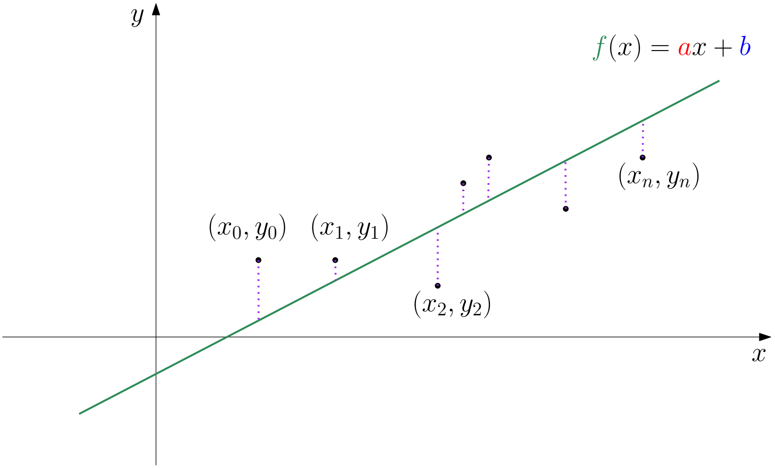
- Suppose we are given a collection of two or more data points, and we wish to find a line that best fits them. How would we do this?

4. Data fitting

- Suppose we are given a collection of two or more data points, and we wish to find a line that best fits them. How would we do this?
    - We will be plotting our data points, say $(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n)$, in $\mathbb{R}^2$.

4. Data fitting

- Suppose we are given a collection of two or more data points, and we wish to find a line that best fits them. How would we do this?
  - We will be plotting our data points, say $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$, in $\mathbb{R}^2$.
  - Most commonly, the x-axis is time (measured in whatever time units happen to be convenient for the problem that we are studying), whereas the y-axis is the quantity that we are measuring, such as population size, the average global temperature, the number of products of a certain type produced or consumed in a given region, etc.
  - We are looking for a line $f(x) = ax + b$ that best fits our data points (picture: next slide).


$f(x) = ax + b$
y
x
(x_0, y_0)
(x_1, y_1)
(x_2, y_2)
(x_n, y_n)

- So, we set up a system of linear equations shown below.

$$
\begin{aligned}
ax_0 + b &= y_0 \\
ax_1 + b &= y_1 \\
&\vdots \\
ax_n + b &= y_n
\end{aligned}
$$

- So, we set up a system of linear equations shown below.

$$\begin{aligned} ax_0 + b &= y_0 \\ ax_1 + b &= y_1 \\ &\vdots \\ ax_n + b &= y_n \end{aligned}$$

- This linear system can be rewritten as the matrix-vector equation below, where the vector $\begin{bmatrix} a & b \end{bmatrix}^T$ is the unknown.

$$\begin{bmatrix} x_0 & 1 \\ x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix}$$

- So, we set up a system of linear equations shown below.

$$\begin{aligned} {\color{red}a}x_0 + {\color{blue}b} &= y_0 \\ {\color{red}a}x_1 + {\color{blue}b} &= y_1 \\ &\vdots \\ {\color{red}a}x_n + {\color{blue}b} &= y_n \end{aligned}$$

- This linear system can be rewritten as the matrix-vector equation below, where the vector $\begin{bmatrix} {\color{red}a} & {\color{blue}b} \end{bmatrix}^T$ is the unknown.

$$\begin{bmatrix} x_0 & 1 \\ x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} {\color{red}a} \\ {\color{blue}b} \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix}$$

- Except in rare cases, the system above will be inconsistent.

- So, we set up a system of linear equations shown below.

$$\begin{aligned} ax_0 + b &= y_0 \\ ax_1 + b &= y_1 \\ &\vdots \\ ax_n + b &= y_n \end{aligned}$$

- This linear system can be rewritten as the matrix-vector equation below, where the vector $\begin{bmatrix} a & b \end{bmatrix}^T$ is the unknown.

$$\begin{bmatrix} x_0 & 1 \\ x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix}$$

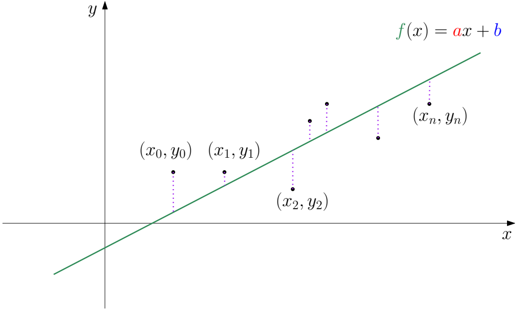
- Except in rare cases, the system above will be inconsistent.
- For this reason, we will look for the least-squares solution(s) $\begin{bmatrix} \hat{a} & \hat{b} \end{bmatrix}^T$ of the system, which yields the line $\hat{f}(x) = \hat{a}x + \hat{b}$.

- This (approximate) solution minimizes the following quantity:

$$\left\| \begin{bmatrix} x_0 & 1 \\ x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} - \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix} \right\| = \left\| \begin{bmatrix} ax_0 + b - y_0 \\ ax_1 + b - y_1 \\ \vdots \\ ax_n + b - y_n \end{bmatrix} \right\| = \left\| \begin{bmatrix} f(x_0) - y_0 \\ f(x_1) - y_1 \\ \vdots \\ f(x_n) - y_n \end{bmatrix} \right\|$$

$$= \sqrt{\sum_{i=0}^{n} \Big(f(x_i) - y_i \Big)^2}.$$

- So, we are effectively minimizing the sum of squares of the vertical distances between our data points and the line.



Example 6.7.3

Using the method of least squares, find the line that best fits the data points $(1, 2)$, $(2, 3)$, $(3, 3)$, $(5, 6)$.

Solution.

Example 6.7.3

Using the method of least squares, find the line that best fits the data points $(1, 2)$, $(2, 3)$, $(3, 3)$, $(5, 6)$.

Solution. We are looking for the function $f(x) = ax + b$ that best fits these four data points.

Example 6.7.3

Using the method of least squares, find the line that best fits the data points $(1, 2)$, $(2, 3)$, $(3, 3)$, $(5, 6)$.

Solution. We are looking for the function $f(x) = ax + b$ that best fits these four data points. We get the linear system below.

$$\begin{aligned} 1a + b &= 2 \\ 2a + b &= 3 \\ 3a + b &= 3 \\ 5a + b &= 6 \end{aligned}$$

Example 6.7.3

Using the method of least squares, find the line that best fits the data points $(1, 2)$, $(2, 3)$, $(3, 3)$, $(5, 6)$.

Solution. We are looking for the function $f(x) = ax + b$ that best fits these four data points. We get the linear system below.

$$\begin{aligned} 1a + b &= 2 \\ 2a + b &= 3 \\ 3a + b &= 3 \\ 5a + b &= 6 \end{aligned}$$

At a glance, we can see that this system is inconsistent;

Example 6.7.3

Using the method of least squares, find the line that best fits the data points $(1, 2)$, $(2, 3)$, $(3, 3)$, $(5, 6)$.

Solution. We are looking for the function $f(x) = ax + b$ that best fits these four data points. We get the linear system below.

$$\begin{aligned} 1a + b &= 2 \\ 2a + b &= 3 \\ 3a + b &= 3 \\ 5a + b &= 6 \end{aligned}$$

At a glance, we can see that this system is inconsistent; so, we will not be able to find an exact solution and will instead have to settle for an approximate one.

Example 6.7.3

Using the method of least squares, find the line that best fits the data points $(1,2)$, $(2,3)$, $(3,3)$, $(5,6)$.

Solution (continued). This system can be rewritten as a matrix-vector equation below, where $\begin{bmatrix} a & b \end{bmatrix}^T$ is the unknown.

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 3 \\ 6 \end{bmatrix}$$

Example 6.7.3

Using the method of least squares, find the line that best fits the data points $(1, 2)$, $(2, 3)$, $(3, 3)$, $(5, 6)$.

Solution (continued). This system can be rewritten as a matrix-vector equation below, where $\begin{bmatrix} a & b \end{bmatrix}^T$ is the unknown.

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 3 \\ 6 \end{bmatrix}$$

We multiply both sides by the transpose of the matrix on the left, and we get the following (where a and b became $\hat{a}$ and $\hat{b}$, respectively, because we are now approximating):

$$\begin{bmatrix} 1 & 2 & 3 & 5 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 5 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 3 \\ 6 \end{bmatrix}.$$

Example 6.7.3

Using the method of least squares, find the line that best fits the data points $(1, 2)$, $(2, 3)$, $(3, 3)$, $(5, 6)$.

Solution (continued). After performing matrix multiplication, we obtain

$$\begin{bmatrix} 39 & 11 \\ 11 & 5 \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} 47 \\ 14 \end{bmatrix}.$$

Example 6.7.3

Using the method of least squares, find the line that best fits the data points $(1,2)$, $(2,3)$, $(3,3)$, $(5,6)$.

Solution (continued). After performing matrix multiplication, we obtain

$$\begin{bmatrix} 39 & 11 \\ 11 & 5 \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} 47 \\ 14 \end{bmatrix}.$$

We now form the augmented matrix of the matrix-vector equation above, and we row reduce to obtain:

$$\text{RREF}\left(\left[\begin{array}{cc:c} 39 & 11 & 47 \\ 11 & 5 & 14 \end{array} \right] \right) = \left[\begin{array}{cc:c} 1 & 0 & 81/74 \\ 0 & 1 & 29/74 \end{array} \right].$$

Example 6.7.3

Using the method of least squares, find the line that best fits the data points $(1, 2)$, $(2, 3)$, $(3, 3)$, $(5, 6)$.

Solution (continued). After performing matrix multiplication, we obtain

$$\begin{bmatrix} 39 & 11 \\ 11 & 5 \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} 47 \\ 14 \end{bmatrix}.$$

We now form the augmented matrix of the matrix-vector equation above, and we row reduce to obtain:

$$\text{RREF}\left(\left[\begin{array}{cc|c} 39 & 11 & 47 \\ 11 & 5 & 14 \end{array} \right] \right) = \left[\begin{array}{cc|c} 1 & 0 & 81/74 \\ 0 & 1 & 29/74 \end{array} \right].$$

This yields the least-squares solution

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} 81/74 \\ 29/74 \end{bmatrix}.$$

Example 6.7.3

Using the method of least squares, find the line that best fits the data points $(1, 2)$, $(2, 3)$, $(3, 3)$, $(5, 6)$.

Solution (continued). After performing matrix multiplication, we obtain

$$\begin{bmatrix} 39 & 11 \\ 11 & 5 \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} 47 \\ 14 \end{bmatrix}.$$

We now form the augmented matrix of the matrix-vector equation above, and we row reduce to obtain:

$$\text{RREF}\left(\left[\begin{array}{cc:c} 39 & 11 & 47 \\ 11 & 5 & 14 \end{array}\right] \right) = \left[\begin{array}{cc:c} 1 & 0 & 81/74 \\ 0 & 1 & 29/74 \end{array}\right].$$

This yields the least-squares solution

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} 81/74 \\ 29/74 \end{bmatrix}.$$

So, the line that best fits our data points is

$$\hat{f}(x) = \frac{81}{74}x + \frac{29}{74}.$$

□