Linear Algebra 2

Lecture #13
Orthogonal projection onto a subspace. Least-squares method

Irena Penev

March 2, 2023

Theorem 1.1

Let V be a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, 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 $\exists! \mathbf{x}_U \in U$ that

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

Moreover, if $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal basis of U, then

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

Theorem 1.1

Let V be a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, 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 $\exists! \mathbf{x}_U \in U$ that

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

Moreover, if $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal basis of U, then

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

- $\mathbf{x}_U$ is called the *orthogonal projection* of $\mathbf{x}$ onto U.

Theorem 1.1

Let V be a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, 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 $\exists! \mathbf{x}_U \in U$ that

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

Moreover, if $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal basis of U, then

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

- $\mathbf{x}_U$ is called the *orthogonal projection* of $\mathbf{x}$ onto U.
- If the basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is **orthonormal**, then

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

Theorem 1.1

Let V be a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, 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 $\exists! \mathbf{x}_U \in U$ that

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

Moreover, if $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal basis of U, then

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

- $\mathbf{x}_U$ is called the *orthogonal projection* of $\mathbf{x}$ onto U.
- If the basis $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is **orthonormal**, then

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

- If $\mathbf{x} \in U$, then $\mathbf{x}_U = \mathbf{x}$.

Theorem 1.1

Let V be a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, 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 $\exists! \mathbf{x}_U \in U$ s.t.

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

Moreover, if $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is an orthogonal basis of U, then

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

Corollary 1.2

Let V be a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $|| \cdot ||$. Let $\mathbf{u}$ be any non-zero vector in V, and set $U := \mathrm{Span}(\mathbf{u})$. Then for every $\mathbf{x} \in V$, we have that $\mathbf{x}_U = \mathrm{proj}_{\mathbf{u}}(\mathbf{x}) = \tfrac{\langle \mathbf{x}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \mathbf{u}$.

Theorem 1.1

Let V be a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, 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 $\exists ! \mathbf{x}_U \in U$ s.t.

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

Moreover, if $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal basis of U, then

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

Corollary 1.3

Let V be a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, 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 $U^\perp$.

Proof of Theorem 1.1.

Proof of Theorem 1.1. Using Corollary 2.5 from Lecture Notes 12, we fix any orthogonal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ of U, and we extend it to an orthogonal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{u}_{k+1}, \dots, \mathbf{u}_n\}$ of V. By Theorem 3.3(a) from Lecture Notes 12, $\{\mathbf{u}_{k+1}, \dots, \mathbf{u}_n\}$ is an orthogonal basis of $U^\perp$.

Set

$$\mathbf{u}^* \quad := \quad \sum_{i=1}^{k} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i.$$

Proof of Theorem 1.1. Using Corollary 2.5 from Lecture Notes 12, we fix any orthogonal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ of U, and we extend it to an orthogonal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{u}_{k+1}, \dots, \mathbf{u}_n\}$ of V. By Theorem 3.3(a) from Lecture Notes 12, $\{\mathbf{u}_{k+1}, \dots, \mathbf{u}_n\}$ is an orthogonal basis of $U^\perp$.

Set

$$\mathbf{u}^* \quad := \quad \sum_{i=1}^{k} \tfrac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i.$$

Since $\mathbf{u}^*$ is a linear combination of the vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$, which form a basis of U, we see that $\mathbf{u}^* \in U$.

Proof of Theorem 1.1. Using Corollary 2.5 from Lecture Notes 12, we fix any orthogonal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ of U, and we extend it to an orthogonal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{u}_{k+1}, \dots, \mathbf{u}_n\}$ of V. By Theorem 3.3(a) from Lecture Notes 12, $\{\mathbf{u}_{k+1}, \dots, \mathbf{u}_n\}$ is an orthogonal basis of $U^\perp$.

Set

$$\mathbf{u}^* \quad := \quad \sum_{i=1}^{k} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i.$$

Since $\mathbf{u}^*$ is a linear combination of the vectors $\mathbf{u}_1, \dots, \mathbf{u}_k$, which form a basis of U, we see that $\mathbf{u}^* \in U$.
Now, fix any $\mathbf{u} \in U$. We must show that $||\mathbf{x} - \mathbf{u}^*|| \leq ||\mathbf{x} - \mathbf{u}||$, and that equality holds iff $\mathbf{u}^* = \mathbf{u}$. Clearly, this is sufficient to prove the theorem.

Proof of Theorem 1.1 (continued). Reminder: $\mathbf{u}^* = \sum_{i=1}^{k} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i$.

Proof of Theorem 1.1 (continued). Reminder: $\mathbf{u}^* = \sum\limits_{i=1}^{k} \frac{\langle \mathbf{x},\mathbf{u}_i \rangle}{\langle \mathbf{u}_i,\mathbf{u}_i \rangle} \mathbf{u}_i$.

Claim. $(\mathbf{u}^* - \mathbf{u}) \perp (\mathbf{x} - \mathbf{u}^*)$.

Proof of the Claim.

Proof of Theorem 1.1 (continued). Reminder: $\mathbf{u}^* = \sum\limits_{i=1}^{k} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i$.

Claim. $(\mathbf{u}^* - \mathbf{u}) \perp (\mathbf{x} - \mathbf{u}^*)$.

Proof of the Claim. Since, $\mathbf{u}^*, \mathbf{u} \in U$, and since U is a subspace of V, it is clear that $\mathbf{u}^* - \mathbf{u} \in U$.

Proof of Theorem 1.1 (continued). Reminder: $\mathbf{u}^* = \sum_{i=1}^{k} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i$.

Claim. *$(\mathbf{u}^* - \mathbf{u}) \perp (\mathbf{x} - \mathbf{u}^*)$.*

Proof of the Claim. Since, $\mathbf{u}^*, \mathbf{u} \in U$, and since U is a subspace of V, it is clear that $\mathbf{u}^* - \mathbf{u} \in U$. So, it suffices to show that $\mathbf{x} - \mathbf{u}^* \in U^\perp$.

Proof of Theorem 1.1 (continued). Reminder: $\mathbf{u}^* = \sum_{i=1}^{k} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i$.

Claim. *$(\mathbf{u}^* - \mathbf{u}) \perp (\mathbf{x} - \mathbf{u}^*)$.*

Proof of the Claim. Since, $\mathbf{u}^*, \mathbf{u} \in U$, and since U is a subspace of V, it is clear that $\mathbf{u}^* - \mathbf{u} \in U$. So, it suffices to show that $\mathbf{x} - \mathbf{u}^* \in U^{\perp}$.

By Theorem 2.1 from Lecture Notes 12, we have that

$$\mathbf{x} \quad = \quad \sum_{i=1}^{n} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i,$$

Proof of Theorem 1.1 (continued). Reminder: $\mathbf{u}^* = \sum_{i=1}^{k} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i$.

Claim. *$(\mathbf{u}^* - \mathbf{u}) \perp (\mathbf{x} - \mathbf{u}^*)$.*

Proof of the Claim. Since, $\mathbf{u}^*, \mathbf{u} \in U$, and since U is a subspace of V, it is clear that $\mathbf{u}^* - \mathbf{u} \in U$. So, it suffices to show that $\mathbf{x} - \mathbf{u}^* \in U^\perp$.

By Theorem 2.1 from Lecture Notes 12, we have that

$$\mathbf{x} \quad = \quad \sum_{i=1}^{n} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i,$$

and it follows that

$$\mathbf{x} - \mathbf{u}^* \quad = \quad \sum_{i=k+1}^{n} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i.$$

Proof of Theorem 1.1 (continued). Reminder: $\mathbf{u}^* = \sum_{i=1}^{k} \frac{\langle \mathbf{x},\mathbf{u}_i \rangle}{\langle \mathbf{u}_i,\mathbf{u}_i \rangle} \mathbf{u}_i$.

Claim. $(\mathbf{u}^* - \mathbf{u}) \perp (\mathbf{x} - \mathbf{u}^*)$.

Proof of the Claim. Since, $\mathbf{u}^*, \mathbf{u} \in U$, and since U is a subspace of V, it is clear that $\mathbf{u}^* - \mathbf{u} \in U$. So, it suffices to show that $\mathbf{x} - \mathbf{u}^* \in U^\perp$.
By Theorem 2.1 from Lecture Notes 12, we have that

$$\mathbf{x} \quad = \quad \sum_{i=1}^{n} \frac{\langle \mathbf{x},\mathbf{u}_i \rangle}{\langle \mathbf{u}_i,\mathbf{u}_i \rangle} \mathbf{u}_i,$$

and it follows that

$$\mathbf{x} - \mathbf{u}^* \quad = \quad \sum_{i=k+1}^{n} \frac{\langle \mathbf{x},\mathbf{u}_i \rangle}{\langle \mathbf{u}_i,\mathbf{u}_i \rangle} \mathbf{u}_i.$$

So, $\mathbf{x} - \mathbf{u}^*$ is a linear combination of the vectors $\mathbf{u}_{k+1}, \dots, \mathbf{u}_n$; since those $n-k$ vectors form a basis of $U^\perp$, it follows that $\mathbf{x} - \mathbf{u}^* \in U^\perp$. This proves the Claim. ♦

Proof of Theorem 1.1 (continued). Reminder: $\mathbf{u}^* = \sum_{i=1}^{k} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i$;
Claim: $(\mathbf{u}^* - \mathbf{u}) \perp (\mathbf{x} - \mathbf{u}^*)$. WTS $||\mathbf{x} - \mathbf{u}^*|| \leq ||\mathbf{x} - \mathbf{u}||$, and equality holds iff $\mathbf{u}^* = \mathbf{u}$.

Proof of Theorem 1.1 (continued). Reminder: $\mathbf{u}^* = \sum\limits_{i=1}^{k} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i$; Claim: $(\mathbf{u}^* - \mathbf{u}) \perp (\mathbf{x} - \mathbf{u}^*)$. WTS $||\mathbf{x} - \mathbf{u}^*|| \leq ||\mathbf{x} - \mathbf{u}||$, and equality holds iff $\mathbf{u}^* = \mathbf{u}$.

Using the Claim, we can apply the Pythagorean theorem to the vectors $\mathbf{u}^* - \mathbf{u}$ and $\mathbf{x} - \mathbf{u}^*$, as follows:

Proof of Theorem 1.1 (continued). Reminder: $\mathbf{u}^* = \sum_{i=1}^{k} \frac{\langle \mathbf{x},\mathbf{u}_i\rangle}{\langle \mathbf{u}_i,\mathbf{u}_i\rangle}\mathbf{u}_i$; Claim: $(\mathbf{u}^* - \mathbf{u}) \perp (\mathbf{x} - \mathbf{u}^*)$. WTS $||\mathbf{x} - \mathbf{u}^*|| \leq ||\mathbf{x} - \mathbf{u}||$, and equality holds iff $\mathbf{u}^* = \mathbf{u}$.

Using the Claim, we can apply the Pythagorean theorem to the vectors $\mathbf{u}^* - \mathbf{u}$ and $\mathbf{x} - \mathbf{u}^*$, as follows:

$$\begin{aligned}
||\mathbf{x} - \mathbf{u}||^2 &= ||(\mathbf{x} - \mathbf{u}^*) + (\mathbf{u}^* - \mathbf{u})||^2 \\
&= ||\mathbf{x} - \mathbf{u}^*||^2 + ||\mathbf{u}^* - \mathbf{u}||^2 && \text{by the Pythagorean theorem} \\
&\geq ||\mathbf{x} - \mathbf{u}^*||^2,
\end{aligned}$$

Proof of Theorem 1.1 (continued). Reminder: $\mathbf{u}^* = \sum_{i=1}^{k} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i$; Claim: $(\mathbf{u}^* - \mathbf{u}) \perp (\mathbf{x} - \mathbf{u}^*)$. WTS $||\mathbf{x} - \mathbf{u}^*|| \leq ||\mathbf{x} - \mathbf{u}||$, and equality holds iff $\mathbf{u}^* = \mathbf{u}$.

Using the Claim, we can apply the Pythagorean theorem to the vectors $\mathbf{u}^* - \mathbf{u}$ and $\mathbf{x} - \mathbf{u}^*$, as follows:

$$\begin{aligned} ||\mathbf{x} - \mathbf{u}||^2 &= ||(\mathbf{x} - \mathbf{u}^*) + (\mathbf{u}^* - \mathbf{u})||^2 \\ &= ||\mathbf{x} - \mathbf{u}^*||^2 + ||\mathbf{u}^* - \mathbf{u}||^2 \qquad \text{by the Pythagorean theorem} \\ &\geq ||\mathbf{x} - \mathbf{u}^*||^2, \end{aligned}$$

and equality holds iff $||\mathbf{u}^* - \mathbf{u}|| = 0$, i.e. iff $\mathbf{u}^* = \mathbf{u}$. Q.E.D.

Theorem 1.1

Let V be a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, 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 $\exists! \mathbf{x}_U \in U$ s.t.

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

Moreover, if $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal basis of U, then

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

Corollary 1.3

Let V be a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, 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 $U^\perp$.

Proof.

Corollary 1.3

Let V be a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, 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 $U^\perp$.

Proof. By Corollary 2.5 from Lecture Notes 12, U has an orthogonal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$, and moreover, this basis can be extended to an orthogonal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{u}_{k+1}, \dots, \mathbf{u}_n\}$ of V.

Corollary 1.3

Let V be a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, 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 $U^\perp$.

Proof. By Corollary 2.5 from Lecture Notes 12, U has an orthogonal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$, and moreover, this basis can be extended to an orthogonal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{u}_{k+1}, \dots, \mathbf{u}_n\}$ of V. By Theorem 3.3(a) from Lecture Notes 12, we have that $\{\mathbf{u}_{k+1}, \dots, \mathbf{u}_n\}$ is an orthogonal basis for $U^\perp$. Now, by Theorem 1.1, we have that

$$\mathbf{x}_U = \sum_{i=1}^{k} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i \quad \text{and} \quad \mathbf{x}_{U^\perp} = \sum_{i=k+1}^{n} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i.$$

Corollary 1.3

Let V be a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, 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 $U^\perp$.

Proof. By Corollary 2.5 from Lecture Notes 12, U has an orthogonal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$, and moreover, this basis can be extended to an orthogonal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{u}_{k+1}, \dots, \mathbf{u}_n\}$ of V. By Theorem 3.3(a) from Lecture Notes 12, we have that $\{\mathbf{u}_{k+1}, \dots, \mathbf{u}_n\}$ is an orthogonal basis for $U^\perp$. Now, by Theorem 1.1, we have that

$$\mathbf{x}_U \quad = \quad \sum_{i=1}^{k} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i \qquad \text{and} \qquad \mathbf{x}_{U^\perp} \quad = \quad \sum_{i=k+1}^{n} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i.$$

On the other hand, by Theorem 2.1 from Lecture Notes 12, we have that $\mathbf{x} = \sum_{i=1}^{n} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i$.

Corollary 1.3

Let V be a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, 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 $U^\perp$.

Proof. By Corollary 2.5 from Lecture Notes 12, U has an orthogonal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$, and moreover, this basis can be extended to an orthogonal basis $\{\mathbf{u}_1, \dots, \mathbf{u}_k, \mathbf{u}_{k+1}, \dots, \mathbf{u}_n\}$ of V. By Theorem 3.3(a) from Lecture Notes 12, we have that $\{\mathbf{u}_{k+1}, \dots, \mathbf{u}_n\}$ is an orthogonal basis for $U^\perp$. Now, by Theorem 1.1, we have that

$$\mathbf{x}_U \quad = \quad \sum_{i=1}^{k} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i \qquad \text{and} \qquad \mathbf{x}_{U^\perp} \quad = \quad \sum_{i=k+1}^{n} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i.$$

On the other hand, by Theorem 2.1 from Lecture Notes 12, we have that $\mathbf{x} = \sum_{i=1}^{n} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i$. So, $\mathbf{x} = \mathbf{x}_U + \mathbf{x}_{U^\perp}$.

Corollary 1.3

Let V be a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, 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 $U^\perp$.

Proof (continued). It remains to prove uniqueness.

Corollary 1.3

Let V be a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, 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 $U^\perp$.

Proof (continued). It remains to prove uniqueness. So, suppose that $\mathbf{y} \in U$ and $\mathbf{z} \in U^\perp$ are s.t. $\mathbf{x} = \mathbf{y} + \mathbf{z}$. WTS $\mathbf{y} = \mathbf{x}_U$ and $\mathbf{z} = \mathbf{x}_{U^\perp}$.

Corollary 1.3

Let V be a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, 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 $U^\perp$.

Proof (continued). It remains to prove uniqueness. So, suppose that $\mathbf{y} \in U$ and $\mathbf{z} \in U^\perp$ are s.t. $\mathbf{x} = \mathbf{y} + \mathbf{z}$. WTS $\mathbf{y} = \mathbf{x}_U$ and $\mathbf{z} = \mathbf{x}_{U^\perp}$. We have that

$$\mathbf{x}_U + \mathbf{x}_{U^\perp} \quad = \quad \mathbf{x} \quad = \quad \mathbf{y} + \mathbf{z},$$

and consequently,

$$\mathbf{x}_U - \mathbf{y} \quad = \quad \mathbf{z} - \mathbf{x}_{U^\perp}.$$

Corollary 1.3

Let V be a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, 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 $U^\perp$.

Proof (continued). It remains to prove uniqueness. So, suppose that $\mathbf{y} \in U$ and $\mathbf{z} \in U^\perp$ are s.t. $\mathbf{x} = \mathbf{y} + \mathbf{z}$. WTS $\mathbf{y} = \mathbf{x}_U$ and $\mathbf{z} = \mathbf{x}_{U^\perp}$. We have that

$$\mathbf{x}_U + \mathbf{x}_{U^\perp} \quad = \quad \mathbf{x} \quad = \quad \mathbf{y} + \mathbf{z},$$

and consequently,

$$\mathbf{x}_U - \mathbf{y} \quad = \quad \mathbf{z} - \mathbf{x}_{U^\perp}.$$

But $\mathbf{x}_U - \mathbf{y} \in U$ and $\mathbf{z} - \mathbf{x}_{U^\perp} \in U^\perp$.

Corollary 1.3

Let V be a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, 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 $U^\perp$.

Proof (continued). It remains to prove uniqueness. So, suppose that $\mathbf{y} \in U$ and $\mathbf{z} \in U^\perp$ are s.t. $\mathbf{x} = \mathbf{y} + \mathbf{z}$. WTS $\mathbf{y} = \mathbf{x}_U$ and $\mathbf{z} = \mathbf{x}_{U^\perp}$. We have that

$$\mathbf{x}_U + \mathbf{x}_{U^\perp} \quad = \quad \mathbf{x} \quad = \quad \mathbf{y} + \mathbf{z},$$

and consequently,

$$\mathbf{x}_U - \mathbf{y} \quad = \quad \mathbf{z} - \mathbf{x}_{U^\perp}.$$

But $\mathbf{x}_U - \mathbf{y} \in U$ and $\mathbf{z} - \mathbf{x}_{U^\perp} \in U^\perp$. Since $U \cap U^\perp = \{\mathbf{0}\}$ (by Theorem 3.3(f) from Lecture Notes 12), it follows that $\mathbf{x}_U - \mathbf{y} = \mathbf{z} - \mathbf{x}_{U^\perp} = \mathbf{0}$. So, $\mathbf{y} = \mathbf{x}_U$ and $\mathbf{z} = \mathbf{x}_{U^\perp}$. Q.E.D.

- For a finite-dimensional vector space V over $\mathbb{R}$ or $\mathbb{C}$, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $|| \cdot ||$, and a subspace U of V, we can define the function $\mathrm{Proj}_U : V \to V$ by

  $$\mathrm{Proj}_U(\mathbf{x}) \quad = \quad \mathbf{x}_U$$

  for all $\mathbf{x} \in V$.

- For a finite-dimensional vector space V over $\mathbb{R}$ or $\mathbb{C}$, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $|| \cdot ||$, and a subspace U of V, we can define the function $\text{Proj}_U : V \to V$ by

  $$\text{Proj}_U(\mathbf{x}) \quad = \quad \mathbf{x}_U$$

  for all $\mathbf{x} \in V$.
- Clearly, $\text{Im}(\text{Proj}_U) = U$ and $\text{Proj}_U[U] = U$.

- For a finite-dimensional vector space V over $\mathbb{R}$ or $\mathbb{C}$, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $|| \cdot ||$, and a subspace U of V, we can define the function $\mathrm{Proj}_U : V \to V$ by

  $$\mathrm{Proj}_U(\mathbf{x}) \quad = \quad \mathbf{x}_U$$

  for all $\mathbf{x} \in V$.
- Clearly, $\mathrm{Im}(\mathrm{Proj}_U) = U$ and $\mathrm{Proj}_U[U] = U$.
- Using the formula from Theorem 1.1, we can easily see that the function Proj_U is linear.

- For a finite-dimensional vector space V over $\mathbb{R}$ or $\mathbb{C}$, equipped with a scalar product $\langle \cdot, \cdot \rangle$ and the induced norm $|| \cdot ||$, and a subspace U of V, we can define the function $\mathrm{Proj}_U : V \to V$ by

  $$\mathrm{Proj}_U(\mathbf{x}) \quad = \quad \mathbf{x}_U$$

  for all $\mathbf{x} \in V$.
- Clearly, $\mathrm{Im}(\mathrm{Proj}_U) = U$ and $\mathrm{Proj}_U[U] = U$.
- Using the formula from Theorem 1.1, we can easily see that the function Proj_U is linear.
- Indeed, if $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is any orthogonal basis of U (this exists by Corollary 2.5 from Lecture Notes 12), then the following hold (next two slides):

(1) for all $\mathbf{x}, \mathbf{y} \in V$, we have that

$$\begin{aligned}
\mathrm{Proj}_U(\mathbf{x}+\mathbf{y}) &\overset{\text{Theorem 1.1}}{=} \sum_{i=1}^{k} \frac{\langle \mathbf{x}+\mathbf{y}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i \\
&= \sum_{i=1}^{k} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle + \langle \mathbf{y}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i \\
&= \Big(\sum_{i=1}^{k} \frac{\langle \mathbf{x}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i \Big) + \Big(\sum_{i=1}^{k} \frac{\langle \mathbf{y}, \mathbf{u}_i \rangle}{\langle \mathbf{u}_i, \mathbf{u}_i \rangle} \mathbf{u}_i \Big) \\
&\overset{\text{Theorem 1.1}}{=} \mathrm{Proj}_U(\mathbf{x}) + \mathrm{Proj}_U(\mathbf{y});
\end{aligned}$$

(2) for all $\mathbf{x} \in V$ and scalars α, we have that

$$\begin{aligned}
\mathrm{Proj}_U(\alpha\mathbf{x}) &\overset{\text{Theorem 1.1}}{=} \sum_{i=1}^{k} \frac{\langle \alpha\mathbf{x}, \mathbf{u}_i\rangle}{\langle \mathbf{u}_i, \mathbf{u}_i\rangle}\mathbf{u}_i \\
&= \sum_{i=1}^{k} \frac{\alpha\langle \mathbf{x}, \mathbf{u}_i\rangle}{\langle \mathbf{u}_i, \mathbf{u}_i\rangle}\mathbf{u}_i \\
&= \alpha\sum_{i=1}^{k} \frac{\langle \mathbf{x}, \mathbf{u}_i\rangle}{\langle \mathbf{u}_i, \mathbf{u}_i\rangle}\mathbf{u}_i \\
&\overset{\text{Theorem 1.1}}{=} \alpha\mathrm{Proj}_U(\mathbf{x}).
\end{aligned}$$

- From now on, we discuss projections onto subspaces of $\mathbb{R}^n$.
- We assume that $\mathbb{R}^n$ is equipped with the standard scalar product $\cdot$ and the induced norm $||\cdot||$.

- From now on, we discuss projections onto subspaces of $\mathbb{R}^n$.
- We assume that $\mathbb{R}^n$ is equipped with the standard scalar product $\cdot$ and the induced norm $||\cdot||$.
  - So, for all vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ and $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$ in $\mathbb{R}^n$, we have that

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} x_i y_i.$$

- From now on, we discuss projections onto subspaces of $\mathbb{R}^n$.
- We assume that $\mathbb{R}^n$ is equipped with the standard scalar product $\cdot$ and the induced norm $||\cdot||$.
  - So, for all vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ and $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$ in $\mathbb{R}^n$, we have that

$$\mathbf{x} \cdot \mathbf{y} \quad = \quad \sum_{i=1}^{n} x_i y_i.$$

  - For all vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ in $\mathbb{R}^n$, we have that

$$||\mathbf{x}|| \quad = \quad \sqrt{\mathbf{x} \cdot \mathbf{x}} \quad = \quad \sqrt{\sum_{i=1}^{n} x_i^2}.$$

- From now on, we discuss projections onto subspaces of $\mathbb{R}^n$.
- We assume that $\mathbb{R}^n$ is equipped with the standard scalar product $\cdot$ and the induced norm $||\cdot||$.
  - So, for all vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ and $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^T$ in $\mathbb{R}^n$, we have that

$$\mathbf{x} \cdot \mathbf{y} \quad = \quad \sum_{i=1}^{n} x_i y_i.$$

  - For all vectors $\mathbf{x} = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix}^T$ in $\mathbb{R}^n$, we have that

$$||\mathbf{x}|| \quad = \quad \sqrt{\mathbf{x} \cdot \mathbf{x}} \quad = \quad \sqrt{\sum_{i=1}^{n} x_i^2}.$$

- Recall that if we identify 1×1 matrices with scalars, then we have that $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

- Now, suppose U is a subspace of $\mathbb{R}^n$.

- Now, suppose U is a subspace of $\mathbb{R}^n$.
- As we saw, $\mathrm{Proj}_U : \mathbb{R}^n \to \mathbb{R}^n$ is linear (with image/range U).

- Now, suppose U is a subspace of $\mathbb{R}^n$.
- As we saw, $\mathrm{Proj}_U : \mathbb{R}^n \to \mathbb{R}^n$ is linear (with image/range U).
- Since Proj_U is linear, it has a standard matrix (note that this matrix belongs to $\mathbb{R}^{n \times n}$).

- Now, suppose U is a subspace of $\mathbb{R}^n$.
- As we saw, $\text{Proj}_U : \mathbb{R}^n \to \mathbb{R}^n$ is linear (with image/range U).
- Since Proj_U is linear, it has a standard matrix (note that this matrix belongs to $\mathbb{R}^{n\times n}$).
- Our goal is to give a formula for the standard matrix of projections onto subspaces of $\mathbb{R}^n$.

- In Lecture Notes 7, we defined the row space of a matrix to be the span of its rows.

- In Lecture Notes 7, we defined the row space of a matrix to be the span of its rows.
- In this lecture, it will be convenient to modify the definition somewhat, and we define the row space of a matrix A to be span of the transposes of its rows, i.e. $\text{Row}(A) := \text{Col}(A^T)$.

- In Lecture Notes 7, we defined the row space of a matrix to be the span of its rows.
- In this lecture, it will be convenient to modify the definition somewhat, and we define the row space of a matrix A to be span of the transposes of its rows, i.e. $\mathrm{Row}(A) := \mathrm{Col}(A^T)$.
- 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

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

- In Lecture Notes 7, we defined the row space of a matrix to be the span of its rows.
- In this lecture, it will be convenient to modify the definition somewhat, and we define the row space of a matrix A to be span of the transposes of its rows, i.e. $\mathrm{Row}(A) := \mathrm{Col}(A^T)$.
- 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

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

- If you don't like changing the definition, then every time you see $\mathrm{Row}(\Box)$, mentally replace it by $\mathrm{Col}(\Box^T)$.

Theorem 2.1

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

Proof.

Theorem 2.1

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

Proof. It suffices to show that $\mathrm{Row}(A)^{\perp} = \mathrm{Nul}(A)$, for it will then follow that $\mathrm{Nul}(A)^{\perp} = (\mathrm{Row}(A)^{\perp})^{\perp} = \mathrm{Row}(A)$.

Theorem 2.1

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

Proof. It suffices to show that $\mathrm{Row}(A)^\perp = \mathrm{Nul}(A)$, for it will then follow that $\mathrm{Nul}(A)^\perp = (\mathrm{Row}(A)^\perp)^\perp = \mathrm{Row}(A)$.

Set $A = \begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix}$, so that $\mathrm{Row}(A) = \mathrm{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$.

Theorem 2.1

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. It suffices to show that $\text{Row}(A)^\perp = \text{Nul}(A)$, for it will then follow that $\text{Nul}(A)^\perp = (\text{Row}(A)^\perp)^\perp = \text{Row}(A)$.

Set $A = \begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix}$, so that $\text{Row}(A) = \text{Span}(\mathbf{a}_1, \ldots, \mathbf{a}_n)$.

Now, for all $\mathbf{x} \in \mathbb{R}^m$, we have that:

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

and so $\text{Nul}(A) = \text{Row}(A)^\perp$. Q.E.D.

Theorem 2.1

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 2.1

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

Corollary 2.2

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

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

Proof.

Theorem 2.1

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

Corollary 2.2

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

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

Proof. Part (a) implies (b) because

$$\mathrm{Row}(A^T A) \overset{Thm.\ 2.1}{=} \mathrm{Nul}(A^T A)^{\perp} \overset{(a)}{=} \mathrm{Nul}(A)^{\perp} \overset{Thm.\ 2.1}{=} \mathrm{Row}(A),$$

Theorem 2.1

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

Corollary 2.2

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

ⓐ $\mathrm{Nul}(A^TA) = \mathrm{Nul}(A)$;

ⓑ $\mathrm{Row}(A^TA) = \mathrm{Row}(A)$;

ⓒ $\mathrm{rank}(A^TA) = \mathrm{rank}(A)$.

Proof. Part (a) implies (b) because

$$\mathrm{Row}(A^TA) \overset{\textit{Thm. } 2.1}{=} \mathrm{Nul}(A^TA)^\perp \overset{(a)}{=} \mathrm{Nul}(A)^\perp \overset{\textit{Thm. } 2.1}{=} \mathrm{Row}(A),$$

and part (b) implies (c) because

$$\mathrm{rank}(A^TA) = \dim(\mathrm{Row}(A^TA)) \overset{(b)}{=} \dim(\mathrm{Row}(A)) = \mathrm{rank}(A).$$

Theorem 2.1

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

Corollary 2.2

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

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

Proof. Part (a) implies (b) because

$$\mathrm{Row}(A^TA) \overset{Thm.\ 2.1}{=} \mathrm{Nul}(A^TA)^\perp \overset{(a)}{=} \mathrm{Nul}(A)^\perp \overset{Thm.\ 2.1}{=} \mathrm{Row}(A),$$

and part (b) implies (c) because

$$\mathrm{rank}(A^TA) = \dim(\mathrm{Row}(A^TA)) \overset{(b)}{=} \dim(\mathrm{Row}(A)) = \mathrm{rank}(A).$$

So, we just need to prove part (a).

Corollary 2.2

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

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

Proof of (a).

Corollary 2.2

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

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

Proof of (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 $\mathbb{R}^m$.

Corollary 2.2

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 of (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, $\mathbf{x} \in \text{Nul}(A^T A)$.

Corollary 2.2

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 of (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, $\mathbf{x} \in \text{Nul}(A^T A)$.

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

Corollary 2.2

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 of (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, $\mathbf{x} \in \text{Nul}(A^T A)$.

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

Corollary 2.2

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 of (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, $\mathbf{x} \in \text{Nul}(A^T A)$.

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

Corollary 2.2

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 of (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, $\mathbf{x} \in \text{Nul}(A^T A)$.

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

Corollary 2.2

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 of (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, $\mathbf{x} \in \text{Nul}(A^T A)$.

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 2.3

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank m. Then the matrix $A(A^T A)^{-1} A^T$ is the standard matrix of the 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 $\mathbf{x}_C = A(A^T A)^{-1} A^T \mathbf{x}$.

Proof.

Theorem 2.3

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank m. Then the matrix $A(A^T A)^{-1} A^T$ is the standard matrix of the 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 $\mathbf{x}_C = A(A^T A)^{-1} A^T \mathbf{x}$.

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 $C = \mathrm{Col}(A)$.

Theorem 2.3

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank m. Then the matrix $A(A^T A)^{-1} A^T$ is the standard matrix of the 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 $\mathbf{x}_C = A(A^T A)^{-1} A^T \mathbf{x}$.

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 $C = \mathrm{Col}(A)$.

First, note that $A^T A \in \mathbb{R}^{m \times m}$, and that by Corollary 2.2(a), we have that $\mathrm{rank}(A^T A) = \mathrm{rank}(A) = m$.

Theorem 2.3

Let $A \in \mathbb{R}^{n\times m}$ be a matrix of rank m. Then the matrix $A(A^TA)^{-1}A^T$ is the standard matrix of the 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. 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 $C = \mathrm{Col}(A)$.

First, note that $A^TA \in \mathbb{R}^{m\times m}$, and that by Corollary 2.2(a), we have that $\mathrm{rank}(A^TA) = \mathrm{rank}(A) = m$.

So, A^TA is invertible, and we see that $(A^TA)^{-1}$ is defined and belongs to $\mathbb{R}^{m\times m}$.

Theorem 2.3

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank m. Then the matrix $A(A^T A)^{-1} A^T$ is the standard matrix of the 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 $\mathbf{x}_C = A(A^T A)^{-1} A^T \mathbf{x}$.

Proof (continued). Reminder: $(A^T A)^{-1}$ is defined and belongs to $\mathbb{R}^{m \times m}$.

Theorem 2.3

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank m. Then the matrix $A(A^T A)^{-1} A^T$ is the standard matrix of the 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 $\mathbf{x}_C = A(A^T A)^{-1} A^T \mathbf{x}$.

Proof (continued). Reminder: $(A^T A)^{-1}$ 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 2.3

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank m. Then the matrix $A(A^T A)^{-1} A^T$ is the standard matrix of the 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). Reminder: $(A^T A)^{-1}$ 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 $A(A^T A)^{-1} A^T \mathbf{x} = \underbrace{A}_{\in \mathbb{R}^{n \times m}} \Big(\underbrace{(A^T A)^{-1} A^T \mathbf{x}}_{\in \mathbb{R}^m} \Big)$ is a linear combination of the columns of A, i.e. $A(A^T A)^{-1} A^T \mathbf{x} \in \text{Col}(A) = C$.

Theorem 2.3

Let $A \in \mathbb{R}^{n\times m}$ be a matrix of rank m. Then the matrix $A(A^TA)^{-1}A^T$ is the standard matrix of the 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).

Theorem 2.3

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank m. Then the matrix $A(A^T A)^{-1} A^T$ is the standard matrix of the 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 $\mathbf{x}_C = A(A^T A)^{-1} A^T \mathbf{x}$.

Proof (continued). In view of Corollary 1.3, it is now enough to show that $(\mathbf{x} - A(A^T A)^{-1} A^T \mathbf{x}) \in C^\perp$, for it will then follow that $\mathbf{x}_C = A(A^T A)^{-1} A^T \mathbf{x}$, which is what we need to show.

Theorem 2.3

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank m. Then the matrix $A(A^T A)^{-1} A^T$ is the standard matrix of the 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 $\mathbf{x}_C = A(A^T A)^{-1} A^T \mathbf{x}$.

Proof (continued). In view of Corollary 1.3, it is now enough to show that $(\mathbf{x} - A(A^T A)^{-1} A^T \mathbf{x}) \in C^\perp$, for it will then follow that $\mathbf{x}_C = A(A^T A)^{-1} A^T \mathbf{x}$, which is what we need to show. But note that

$$C^\perp \;=\; \mathrm{Col}(A)^\perp \;=\; \mathrm{Row}(A^T)^\perp \;\overset{\text{Thm. 2.1}}{=}\; \mathrm{Nul}(A^T)$$

Theorem 2.3

Let $A \in \mathbb{R}^{n\times m}$ be a matrix of rank m. Then the matrix $A(A^TA)^{-1}A^T$ is the standard matrix of the 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). In view of Corollary 1.3, it is now enough to show 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. But note that

$$C^\perp \;=\; \mathrm{Col}(A)^\perp \;=\; \mathrm{Row}(A^T)^\perp \;\overset{\text{Thm. 2.1}}{=}\; \mathrm{Nul}(A^T)$$

So, it in fact suffices to show that $\mathbf{x} - A(A^TA)^{-1}A^T\mathbf{x} \in \mathrm{Nul}(A^T)$.

Theorem 2.3

Let $A \in \mathbb{R}^{n\times m}$ be a matrix of rank m. Then the matrix $A(A^TA)^{-1}A^T$ is the standard matrix of the 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). In view of Corollary 1.3, it is now enough to show 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. But note that

$$C^\perp = \mathrm{Col}(A)^\perp = \mathrm{Row}(A^T)^\perp \overset{\text{Thm. 2.1}}{=} \mathrm{Nul}(A^T)$$

So, it in fact suffices to show that $\mathbf{x} - A(A^TA)^{-1}A^T\mathbf{x} \in \mathrm{Nul}(A^T)$. For this, we compute:

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

Theorem 2.3

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank m. Then the matrix $A(A^TA)^{-1}A^T$ is the standard matrix of the 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). In view of Corollary 1.3, it is now enough to show 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. But note that

$$C^\perp = \mathrm{Col}(A)^\perp = \mathrm{Row}(A^T)^\perp \overset{\text{Thm. 2.1}}{=} \mathrm{Nul}(A^T)$$

So, it in fact suffices to show that $\mathbf{x} - A(A^TA)^{-1}A^T\mathbf{x} \in \mathrm{Nul}(A^T)$. 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 $\mathbf{x} - A(A^TA)^{-1}A^T\mathbf{x} \in \mathrm{Nul}(A^T)$. Q.E.D.

Theorem 2.4

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 $\mathrm{Proj}_{U^\perp}$.

Proof.

Theorem 2.4

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 $\mathrm{Proj}_{U^\perp}$.

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

$$\begin{aligned} (I_n - P)\mathbf{x} &= I_n\mathbf{x} - P\mathbf{x} \\ &= \mathbf{x} - \mathbf{x}_U && \text{because } P \text{ is the standard matrix of } \mathrm{Proj}_U \\ &= \mathbf{x}_{U^\perp} && \text{by Corollary 1.3,} \end{aligned}$$

and the result follows.

Corollary 2.5

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank n. Then the matrix $I_m - A^T(AA^T)^{-1}A$ is the standard matrix of the 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.

Corollary 2.5

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank n. Then the matrix $I_m - A^T(AA^T)^{-1}A$ is the standard matrix of the 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. Note that

$$\mathrm{Nul}(A) \overset{\text{Thm. 2.1}}{=} \mathrm{Row}(A)^\perp = \mathrm{Col}(A^T)^\perp.$$

Corollary 2.5

Let $A \in \mathbb{R}^{n\times m}$ be a matrix of rank n. Then the matrix $I_m - A^T(AA^T)^{-1}A$ is the standard matrix of the 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. Note that

$$\mathrm{Nul}(A) \overset{\text{Thm. 2.1}}{=} \mathrm{Row}(A)^{\perp} = \mathrm{Col}(A^T)^{\perp}.$$

Note further that $A^T \in \mathbb{R}^{m\times n}$ and $\mathrm{rank}(A^T) = \mathrm{rank}(A) = n$.

Corollary 2.5

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank n. Then the matrix $I_m - A^T(AA^T)^{-1}A$ is the standard matrix of the 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. Note that

$$\mathrm{Nul}(A) \overset{\text{Thm. 2.1}}{=} \mathrm{Row}(A)^{\perp} = \mathrm{Col}(A^T)^{\perp}.$$

Note further that $A^T \in \mathbb{R}^{m \times n}$ and $\mathrm{rank}(A^T) = \mathrm{rank}(A) = n$. So, by Theorem 2.3, the standard matrix of the orthogonal projection onto $\mathrm{Col}(A^T)$ is $A^T(AA^T)^{-1}A$.

Corollary 2.5

Let $A \in \mathbb{R}^{n \times m}$ be a matrix of rank n. Then the matrix $I_m - A^T(AA^T)^{-1}A$ is the standard matrix of the 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. Note that

$$\text{Nul}(A) \overset{\text{Thm. 2.1}}{=} \text{Row}(A)^\perp = \text{Col}(A^T)^\perp.$$

Note further that $A^T \in \mathbb{R}^{m \times n}$ and $\text{rank}(A^T) = \text{rank}(A) = n$. So, by Theorem 2.3, the standard matrix of the orthogonal projection onto $\text{Col}(A^T)$ is $A^T(AA^T)^{-1}A$.

By Theorem 2.4, the standard matrix of orthogonal projection onto $\text{Col}(A^T)^\perp = \text{Nul}(A)$ is $I_m - A^T(AA^T)^{-1}A$. Q.E.D.

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

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

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

is as small as possible.

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

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

- In some real-world applications, we may be interested in finding the best **approximate** solution to a matrix-vector equation $A\mathbf{x} = \mathbf{b}$ that may possibly be inconsistent.
- More formally, given a matrix $A \in \mathbb{R}^{n \times m}$ and a vector $\in \mathbb{R}^n$, 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 we can use the "least-squares method."

Theorem 3.1

Let $A \in \mathbb{R}^{n \times m}$ and $\mathbf{b} \in \mathbb{R}^n$. Then the 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:

Theorem 3.1

Let $A \in \mathbb{R}^{n \times m}$ and $\mathbf{b} \in \mathbb{R}^n$. Then the 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:
  - 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 $A\mathbf{x} = \mathbf{b}$.

Theorem 3.1

Let $A \in \mathbb{R}^{n \times m}$ and $\mathbf{b} \in \mathbb{R}^n$. Then the 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:
  - 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 3.1, 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.

Proof of Theorem 3.1.

Proof of Theorem 3.1. 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 of Theorem 3.1. 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 $A^T A\mathbf{x} = A^T\mathbf{b}$.

Note 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.

Proof of Theorem 3.1. 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 $A^T A\mathbf{x} = A^T\mathbf{b}$.

Note 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 1.3, $\mathbf{b} = \mathbf{b}_C + \mathbf{b}_{C^\perp}$ is the only way to decompose $\mathbf{b}$ as a sum of a vector in C and a vector in $C^\perp$.

Proof of Theorem 3.1. 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 $A^T A\mathbf{x} = A^T\mathbf{b}$.

Note 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 1.3, $\mathbf{b} = \mathbf{b}_C + \mathbf{b}_{C^\perp}$ 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 of Theorem 3.1. 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 $A^T A\mathbf{x} = A^T\mathbf{b}$.

Note 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 1.3, $\mathbf{b} = \mathbf{b}_C + \mathbf{b}_{C^\perp}$ 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

$$C^\perp \;=\; \text{Col}(A)^\perp \;=\; \text{Row}(A^T)^\perp \;\overset{\text{Thm. 2.1}}{=}\; \text{Nul}(A^T).$$

Proof of Theorem 3.1. 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 $A^T A\mathbf{x} = A^T\mathbf{b}$.

Note 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 1.3, $\mathbf{b} = \mathbf{b}_C + \mathbf{b}_{C^\perp}$ 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

$$C^\perp \;=\; \mathrm{Col}(A)^\perp \;=\; \mathrm{Row}(A^T)^\perp \;\overset{\text{Thm. 2.1}}{=}\; \mathrm{Nul}(A^T).$$

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

Theorem 3.1

Let $A \in \mathbb{R}^{n \times m}$ and $\mathbf{b} \in \mathbb{R}^n$. Then the 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 of Theorem 3.1 We have now shown that vectors $\mathbf{x} \in \mathbb{R}^m$ that minimize $||A\mathbf{x} - \mathbf{b}||$ are precisely those that satisfy $A^T A\mathbf{x} = A^T\mathbf{b}$.

Theorem 3.1

Let $A \in \mathbb{R}^{n \times m}$ and $\mathbf{b} \in \mathbb{R}^n$. Then the 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 of Theorem 3.1 We have now shown that vectors $\mathbf{x} \in \mathbb{R}^m$ that minimize $||A\mathbf{x} - \mathbf{b}||$ are precisely those that satisfy $A^T A\mathbf{x} = A^T\mathbf{b}$.

Moreover, 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 $A\mathbf{x} = \mathbf{b}_C$.

Theorem 3.1

Let $A \in \mathbb{R}^{n \times m}$ and $\mathbf{b} \in \mathbb{R}^n$. Then the 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 of Theorem 3.1 We have now shown that vectors $\mathbf{x} \in \mathbb{R}^m$ that minimize $||A\mathbf{x} - \mathbf{b}||$ are precisely those that satisfy $A^T A\mathbf{x} = A^T\mathbf{b}$.

Moreover, 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. Q.E.D.

Theorem 3.1

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

Example 3.2

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 of Example 3.2.

Solution of Example 3.2. We apply Theorem 3.1. So, we need to find the solutions of the equation $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$.

Solution of Example 3.2. We apply Theorem 3.1. So, we need to find the solutions of the equation $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$.

We first compute

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

Solution of Example 3.2. We apply Theorem 3.1. 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

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

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

Solution of Example 3.2 (continued). The least-squares error of $A\mathbf{x} = \mathbf{b}$ is

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

Solution of Example 3.2 (continued). The least-squares error of $A\mathbf{x} = \mathbf{b}$ is

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

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