Linear Algebra 2: Lecture 13

Irena Penev

Summer 2023

1 Orthogonal projection onto a subspace

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 there exists a unique vector $\mathbf{x}_U \in U$ that has the property that*

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

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

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

Remark: Note that if $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ happens to be an **orthonormal** basis of U, then we get that $\langle \mathbf{u}_1, \mathbf{u}_1 \rangle = \dots = \langle \mathbf{u}_k, \mathbf{u}_k \rangle = 1$, and so the formula for $\mathbf{x}_U$ from Theorem 1.1 turns into

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

Moreover, we note that if $\mathbf{x} \in U$, then $\mathbf{x}_U = \mathbf{x}$, since in this case, the expression $||\mathbf{x} - \mathbf{u}||$ (for $\mathbf{u} \in U$) is minimized for $\mathbf{u} = \mathbf{x}$.

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. We note that 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

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

(So, $\mathbf{u}^*$ is defined via the formula from from the statement of the theorem. The reason we call it $\mathbf{u}^*$ rather than $\mathbf{x}_U$ is because we have not proven the existence and uniqueness of $\mathbf{x}_U$ yet. However, this is just a minor stylistic matter!) 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 if and only if $\mathbf{u}^* = \mathbf{u}$. Clearly, this is sufficient to prove the theorem.

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

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

and it follows that

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

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. ♦

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 consequently, we have that $||\mathbf{x} - \mathbf{u}^*|| \leq ||\mathbf{x} - \mathbf{u}||$. Moreover, the inequality above is an equality if and only if $||\mathbf{u}^* - \mathbf{u}|| = 0$, i.e. if and only if $\mathbf{u}^* = \mathbf{u}$. This completes the argument. □

Terminology and notation: The vector $\mathbf{x}_U$ from Theorem 1.1 is called the *orthogonal projection* of $\mathbf{x}$ onto U.

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 $U := Span(\mathbf{u})$.*[1] *Then for every $\mathbf{x} \in V$, we have that*

$$\mathbf{x}_U \quad = \quad proj_{\mathbf{u}}(\mathbf{x}) \quad = \quad \tfrac{\langle \mathbf{x}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} \mathbf{u}.$$

Proof. Clearly, $\{\mathbf{u}\}$ is an orthogonal basis of U. So, the result follows immediately from Theorem 1.1. □

[1] So, U is a one-dimensional subspace (i.e. a line) in 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*

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

Moreover, this is the unique way of expressing $\mathbf{x}$ as a sum of a vector in U and a vector in $U^\perp$.[2]

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

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

On the other hand, by Theorem 2.1 from Lecture Notes 12, we have that

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

Consequently,

$$\begin{aligned} \mathbf{x} \quad &= \quad \sum_{i=1}^{n} \tfrac{\langle \mathbf{x},\mathbf{u}_i\rangle}{\langle \mathbf{u}_i,\mathbf{u}_i\rangle}\mathbf{u}_i \\ &= \quad \Big(\sum_{i=1}^{k} \tfrac{\langle \mathbf{x},\mathbf{u}_i\rangle}{\langle \mathbf{u}_i,\mathbf{u}_i\rangle}\mathbf{u}_i\Big) + \Big(\sum_{i=k+1}^{n} \tfrac{\langle \mathbf{x},\mathbf{u}_i\rangle}{\langle \mathbf{u}_i,\mathbf{u}_i\rangle}\mathbf{u}_i\Big) \\ &= \quad \mathbf{x}_U + \mathbf{x}_{U^\perp}. \end{aligned}$$

It remains to prove the uniqueness of this decomposition. So, suppose that $\mathbf{y} \in U$ and $\mathbf{z} \in U^\perp$ are such that $\mathbf{x} = \mathbf{y} + \mathbf{z}$. We must prove that $\mathbf{y} = \mathbf{x}_U$ 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 $\mathbf{x}_U - \mathbf{y} = \mathbf{z} - \mathbf{x}_{U^\perp} = \mathbf{0}$, and consequently, $\mathbf{y} = \mathbf{x}_U$ and $\mathbf{z} = \mathbf{x}_{U^\perp}$. This completes the argument. $\square$

[2] 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 $\mathbf{z} = \mathbf{x}_{U^\perp}$.

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 $\text{Proj}_U(\mathbf{x}) = \mathbf{x}_U$ for all $\mathbf{x} \in V$ (where $\mathbf{x}_U$ is the orthogonal projection of $\mathbf{x}$ onto U, as in Theorem 1.1). Clearly, $\text{Im}(\text{Proj}_U) = U$ and $\text{Proj}_U[U] = U$. 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:

- for all $\mathbf{x}, \mathbf{y} \in V$, we have that

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

  where both instances of (*) follow from Theorem 1.1;

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

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

  where both instances of (*) follow from Theorem 1.1.

2 Projection onto subspaces of $\mathbb{R}^n$

In this section, we assume that $\mathbb{R}^n$ is equipped with the standard scalar product $\cdot$ and the induced norm $||\cdot||$. 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 $\mathbb{R}^n$. As we saw above (see the comment following the proof of Corollary 1.3), $\text{Proj}_U : \mathbb{R}^n \to \mathbb{R}^n$ is linear (with image/range U). Since Proj_U is linear, it has a standard matrix (note that this matrix belongs to $\mathbb{R}^{n\times n}$). In this section, 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 this lecture, it will be convenient to modify the definition somewhat, and we define the row space of a matrix A, denoted by $\text{Row}(A)$, to be span of the transposes of its rows, i.e. $\text{Row}(A) := \text{Col}(A^T)$. 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

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

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

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. In view of Theorem 3.3(e) from Lecture Notes 12, it suffices to show that $\text{Row}(A)^\perp = \text{Nul}(A)$.[3] 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 the following sequence of

[3] Indeed, by Theorem 3.3(e) from Lecture Notes 12, we have that $(\text{Row}(A)^\perp)^\perp = \text{Row}(A)$. So, if $\text{Row}(A)^\perp = \text{Nul}(A)$, then $\text{Nul}(A)^\perp = (\text{Row}(A)^\perp)^\perp = \text{Row}(A)$.

equivalences:

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

where (*) follows from the fact that $\{\mathbf{a}_1, \dots, \mathbf{a}_m\}^\perp = \text{Span}(\mathbf{a}_1, \dots, \mathbf{a}_m)^\perp$ (by Proposition 3.2 from Lecture Notes 12). This proves that $\text{Nul}(A) = \text{Row}(A)^\perp$, and we are done. $\square$

Corollary 2.2. *Let $A \in \mathbb{R}^{n \times m}$. Then all the following hold:*

(a) $Nul(A^T A) = Nul(A)$;

(b) $Row(A^T A) = Row(A)$;

(c) $rank(A^T A) = rank(A)$.

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 $\mathbf{x} \in \mathbb{R}^m$. We must show that $\mathbf{x} \in \text{Nul}(A^T A)$ if and only if $\mathbf{x} \in \text{Nul}(A)$.

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).

For (b), we observe that

$$\begin{aligned}
\text{Row}(A^T A) &= \text{Nul}(A^T A)^\perp && \text{by Theorem 2.1} \\
&= \text{Nul}(A)^\perp && \text{by (a)} \\
&= \text{Row}(A) && \text{by Theorem 2.1.}
\end{aligned}$$

Finally, for (c), we have that

$$\text{rank}(A^T A) \quad = \quad \dim(\text{Row}(A^T A)) \quad = \quad \dim(\text{Row}(A)) \quad = \quad \text{rank}(A).$$

This completes the argument. $\square$

Theorem 2.3. *Let $A \in \mathbb{R}^{n\times m}$ be a matrix of rank m.*[4] *Then the matrix $A(A^T A)^{-1}A^T$ is the standard matrix of the orthogonal projection onto $Col(A)$, that is, for all $\mathbf{x} \in \mathbb{R}^n$, the orthogonal projection of $\mathbf{x}$ onto $C := Col(A)$ 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 $C = \text{Col}(A)$. First, note that $A^T A \in \mathbb{R}^{m\times m}$, and that by Corollary 2.2(a), we have that $\text{rank}(A^T A) = \text{rank}(A) = m$. So, $A^T A$ is invertible,[5] 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 $A(A^T A)^{-1}A^T\mathbf{x} = \underbrace{A}_{\in\mathbb{R}^{n\times m}} \Big(\underbrace{(A^T A)^{-1}A^T\mathbf{x}}_{\in\mathbb{R}^m}\Big)$ is a linear combination of the columns of A, i.e. $A(A^T A)^{-1}A^T\mathbf{x} \in \text{Col}(A) = C$.

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 $\mathbf{x}_C = A(A^T A)^{-1}A^T\mathbf{x}$,[6] which is what we need to show. But note that

$$\begin{aligned} C^\perp \quad &= \quad \text{Col}(A)^\perp \\ &= \quad \text{Row}(A^T)^\perp \\ &= \quad \text{Nul}(A^T) \qquad\quad \text{by Theorem 2.1.} \end{aligned}$$

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

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

This proves that $\mathbf{x} - A(A^T A)^{-1}A^T\mathbf{x} \in \text{Nul}(A^T)$. $\square$

[4]In particular, this implies that $m \leq n$, since $\text{rank}(A) \leq \min\{n, m\}$.

[5]We are using Theorem 4.1 from Lecture Notes 7.

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

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

which (by Corollary 1.3) implies that $\mathbf{x}_C = A(A^T A)^{-1}A^T\mathbf{x}$.

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 $Proj_{U^\perp}$.*

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.*[7] *Then the matrix $I_m - A^T(AA^T)^{-1}A$ is the standard matrix of the orthogonal projection onto $N := Nul(A)$, 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

$$\begin{aligned}\mathrm{Nul}(A) &= \mathrm{Row}(A)^\perp && \text{by Theorem 2.1} \\ &= \mathrm{Col}(A^T)^\perp.\end{aligned}$$

Note further that $A^T \in \mathbb{R}^{m\times n}$ and $\mathrm{rank}(A^T) = \mathrm{rank}(A) = n$.[8] So, by Theorem 2.3, the standard matrix of the orthogonal projection onto $\mathrm{Col}(A^T)$ is $A^T(AA^T)^{-1}A$. By Theorem 2.4, the standard matrix of orthogonal projection onto $\mathrm{Col}(A^T)^\perp = \mathrm{Nul}(A)$ is $I_m - A^T(AA^T)^{-1}A$. This completes the argument. □

3 The method of least squares

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, suppose we are given a norm $||\cdot||$ on $\mathbb{R}^n$, 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 the answer will obviously depend on which norm we are using. In the remainder of this section, we will work only with the norm

[7] In particular, this implies that $n \leq m$, since $\mathrm{rank}(A) \leq \min\{n, m\}$.

[8] We are using Corollary 4.2 from Lecture Notes 7.

induced by the standard scalar product in $\mathbb{R}^n$, i.e. the standard Pythagorean norm. Recall that this is the norm $||\cdot||$ given by

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

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

Theorem 3.1. *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. 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

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

where (*) follows from Theorem 2.1. 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 $A^T A\mathbf{x} = A^T\mathbf{b}$.

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 $A^T A\mathbf{x} = A^T\mathbf{b}$ if and only if 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. $\square$

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 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.

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, $||A\mathbf{x} - \mathbf{b}|| = 0$ if and only if $A\mathbf{x} = \mathbf{b}$).

Example 3.2. *Let*

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

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 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

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

and then we compute

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

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

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. □

Remark: In the example above, the equation $A\mathbf{x} = \mathbf{b}$ had a unique least-squares solution $\hat{\mathbf{x}}$, and we obtained the least-squares error of $A\mathbf{x} = \mathbf{b}$ by computing $||A\hat{\mathbf{x}} - \mathbf{b}||$. But what if $A\mathbf{x} = \mathbf{b}$ had more than one least-squares solution? In that case, we would choose one least squares solution $\hat{\mathbf{x}}$ (any one will do), and compute $||A\hat{\mathbf{x}} - \mathbf{b}||$. By the definition of a least-squares solution, the value of $||A\hat{\mathbf{x}} - \mathbf{b}||$ is the same regardless of which least-squares solution $\hat{\mathbf{x}}$ we choose.