Worksheet 19: Orthogonal Projections and Least-Squares (§5.4)

Theorem 1. Let V be any subspace of $\mathbb{R}^n$ and let $\mathbb{R}^n \xrightarrow{\text{proj}_V} \mathbb{R}^n$ be the orthogonal projection onto V. Let $\vec{x} \in \mathbb{R}^n$ be arbitrary. Then the vector $\text{proj}_V(\vec{x})$ is the *closest* vector to $\vec{x}$ that is in the subspace V, in the sense that

$$\|\vec{x} - \text{proj}_V(\vec{x})\| \leq \|\vec{x} - \vec{v}\|$$

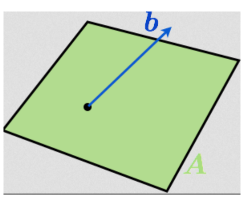
for all $\vec{v} \in V$.

Problem 1. Draw a sketch illustrating Theorem 1, showing $V, \vec{x}$, and $\text{proj}_V(\vec{x})$. Discuss the meaning of "closest vector to $\vec{x}$ in V." Why is the given inequality of lengths relevant? For vectors $\vec{x} \in V$, what is $\text{proj}_V(\vec{x})$? For vectors $\vec{x} \notin V$, discuss how to compute $\text{proj}_V(\vec{x})$.

Solution: See the book for pictures. Note that $||\vec{x}-\text{proj}_V(\vec{x})|| \leq ||\vec{x}-\vec{v}||$ says that the distance between the given vector $\vec{x}$ and an arbitrary vector $\vec{v} \in V$ is *smallest* when $\vec{v} = \text{proj}_V(\vec{x})$. So $\text{proj}_V(\vec{x})$ is the closest vector to $\vec{x}$ in V. If $\vec{x} \in V$, then $\text{proj}_V(\vec{x}) = \vec{x}$. To compute the projection of $\vec{x}$ onto V, we would need to find an orthonormal basis $(\vec{u}_1, \dots, \vec{u}_d)$ for V, then use the formula from Worksheet 16 for the projection: $\text{proj}_V(\vec{x}) = \sum_{i=1}^{d}(\vec{x} \cdot \vec{u}_i)\vec{u}_i$.

Problem 2. Let A be a $m \times n$ matrix and let $\vec{b}$ be a vector in $\mathbb{R}^m$.

(a) Explain why $A\vec{x} = \vec{b}$ is consistent if and only if $\vec{b} \in \text{im}A$.

(b) 

Suppose you came up with the linear system $A\,\vec{x} = \vec{b}$ based on real-life measurements (say, 14 equations in 20 variables), whose solution would make your company millions of dollars. The cartoon figure indicates the span of the columns of your matrix A (the green plane) and your vector $\vec{b}$ (the blue arrow poking out of the green plane). Can this system be solved?

(c) How would you go about finding the vector $\vec{b}'$ in the subspace $\text{im}\,A$ *closest* to $\vec{b}$? Sketch it in the figure. Explain why $A\vec{x} = b'$ is consistent. In what sense is it the *closest* consistent system to your original system $A\vec{x} = \vec{b}$?

(d) Consider the toy case

$$A\vec{x} = \begin{bmatrix} 1 & -1 & 0 \\ 2 & 3 & 5 \\ -5 & 1 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 30 \\ 30 \\ 30 \end{bmatrix}.$$

Find an orthonormal basis for $\text{im}\,A$ and use it compute the "closest" vector $\vec{b}'$ such that $A\vec{x} = \vec{b}'$ is consistent. What is the " closest" consistent system of linear equations to your toy data problem? (You do not have to solve it right now). Its solutions are the **least squares solutions** to this toy system.

(e) Explain why the least squares solutions might *not actually be solutions* to the system. What are they, then? Under what circumstances are the least square solutions *actually solutions*? If $A\vec{x} = \vec{b}$ is consistent, what are its least squares solutions?

Solution:

(a) The image of A is spanned by the columns of A, so $\vec{b}$ is in the image of A if and only if $\vec{b}$ is a linear combination of the columns of A. Suppose A has columns $\vec{v}_1, \ldots, \vec{v}_n$. Then the product

$$A\vec{x} = [\vec{v}_1 \ \ \vec{v}_2 \quad \ldots \quad \vec{v}_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\vec{v}_1 + x_2\vec{v}_2 + \ldots x_n\vec{v}_n$$

which says that $A\vec{x} = \vec{b}$ has a solution if and only if $\vec{b}$ is a linear combination of $\vec{v}_1, \ldots, \vec{v}_n$.

(b) No, not exactly. The vector $\vec{b}$ is not in the space spanned by the columns of A, which means the system $A\vec{x} = \vec{b}$ has no solution! Sadly, we don't make millions.

(c) The closest vector in im$A$ to $\vec{b}$ is the projection $\vec{b}'$ of $\vec{b}$ onto im$A$. The system $A\vec{x} = \vec{b}'$ is consistent by (a); a solution $\vec{x}^*$ can be thought of "close" to a solution for $A\vec{x} = \vec{b}$. There are no closer solutions: $||A\vec{x}^* - \vec{b}|| \leq ||A\vec{x} - \vec{b}||$ for all $\vec{x}$.

(d) The green plane is the span of the columns of this matrix which is the span of $\begin{bmatrix} 1 \\ 2 \\ -5 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 3 \\ 1 \end{bmatrix}$, since the third column is dependent on the previous two. Since these two columns (let's call them $\vec{v}_1, \vec{v}_2$) are already perpendicular, it is easy to get an orthonormal basis by scaling each by its length. To compute the projection onto V, we can use the formula

$$\vec{x} \mapsto \frac{(\vec{x} \cdot v_1)}{(\vec{v}_1 \cdot \vec{v}_1)}\vec{v}_1 + \frac{(\vec{x} \cdot v_2)}{(\vec{v}_2 \cdot \vec{v}_2)}\vec{v}_2$$

to find that the projection of $\begin{bmatrix} 30 \\ 30 \\ 30 \end{bmatrix}$ is $\begin{bmatrix} -112/11 \\ 226/11 \\ 200/11 \end{bmatrix}$. So the consistent linear system we need to solve is

$$\begin{bmatrix} 1 & -1 & 0 \\ 2 & 3 & 5 \\ -5 & 1 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -112/11 \\ 226/11 \\ 200/11 \end{bmatrix}.$$

If you solve this, using the methods of Chapter 1 (row reduction), you'll see that the least squares solutions are

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \left\{ \begin{bmatrix} z-2 \\ -(11z-90)/11 \\ z \end{bmatrix} \mid z \in \mathbb{R} \right\}.$$

(e) If $A\vec{x} = \vec{b}$ is inconsistent, the least squares solutions $\vec{x}^*$ are *never* actual solutions, since $A\vec{x}^* = \text{proj}_V(\vec{b}) \neq \vec{b}$. They are merely approximations to solutions, the best we can do. If $A\vec{x} = \vec{b}$ is consistent, then the least squares solutions are the actual solutions.

Lemma. Let A be an $n \times m$ matrix. For all $\vec{x} \in \mathbb{R}^m$ and $\vec{y} \in \mathbb{R}^n$, we have $A\vec{x} \cdot \vec{y} = \vec{x} \cdot A^\top \vec{y}$.

Problem 3.

(a) Verify the lemma for the matrix $A = I_n$ and arbitrary $\vec{x}$ and $\vec{y}$ in $\mathbb{R}^n$.

(b) Verify the lemma for the matrix $A = \begin{bmatrix} 1 & 2 \\ 0 & -1 \end{bmatrix}$ and arbitrary $\vec{x}$ and $\vec{y}$ in $\mathbb{R}^2$.

(c) Prove the lemma. [Hint: Interpret $\vec{w} \cdot \vec{v}$ as a matrix product $\vec{w}^T \vec{v}$. Take the transpose.]

Solution: For (a), this just says $\vec{x} \cdot \vec{y} = \vec{x} \cdot \vec{y}$ since $I_n = I_n^\top$. For (b):

$$A\vec{x} \cdot \vec{y} = \begin{bmatrix} x_1 + 2x_2 \\ -x_2 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = y_1(x_1 + 2x_2) - x_2 y_2.$$

And

$$\vec{x} \cdot A^\top \vec{y} = \vec{x} \cdot \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ 2y_1 - y_2 \end{bmatrix} = x_1 y_1 + x_2(2y_1 - y_2),$$

which is the same.

For (c): $\vec{A}x \cdot \vec{y} = (A\vec{x})^\top \vec{y} = (\vec{x}^\top) A^\top \vec{y} = \vec{x}^\top (A^\top \vec{y}) = \vec{x} \cdot A^\top \vec{y}$.

Theorem 2. If A is an $m \times n$ matrix, then $\ker A^\top = (\operatorname{im} A)^\perp$.

Problem 4.

(a) Use Theorem 2 above to deduce that $(\ker A^\top)^\perp = (\operatorname{im} A)$ also that $\ker(A)^\perp = \operatorname{im}(A^\top)$. [Hint: Use the fact that $(A^\top)^\top = A$ and $(V^\perp)^\perp = V$.]

(b) Use Theorem 2 to show that A and $A^\top$ have the same rank.

(c) Prove Theorem 2. [Hint: One way uses the lemma. Another way proceeds by writing A as a row of columns $[\vec{v}_1\ \vec{v}_2\ \dots\ \vec{v}_n]$, so that $A^\top$ is a column of rows $\begin{bmatrix} \vec{v}_1^\top \\ \vec{v}_2^\top \\ \vdots \\ \vec{v}_n^\top \end{bmatrix}$, and then thinking about what $\vec{x} \in \mathbb{R}^m$ are in $\ker A^\top$. It's not a bad idea to understand the proof both ways.]

Solution:

For (a), we can "perp" both sides: $\ker A^\top = (\operatorname{im} A)^\perp$ implies $(\ker A^\top)^\perp = ((\operatorname{im} A)^\perp)^\perp = \operatorname{im} A$. For the second statement, we apply this statement to $A^\top$.

For (b), rank $A = \dim \operatorname{im} A = n - \dim \ker A$ by rank nullity, and this is $\dim(\ker A)^\perp = \dim \operatorname{im}(A^\top)$, which is rank $A^\top$.

Now (c): First we do the proof using the lemma. Take $\vec{x} \in \ker A^\top$. We want to show that $\vec{x} \in (\operatorname{im} A)^\perp$. For this, we need $\vec{y} \cdot \vec{x} = 0$ for all $\vec{y} \in \operatorname{im} A$. Writing $\vec{y} = A\vec{z}$, we need $A\vec{z} \cdot \vec{x} = 0$ for all $\vec{z} \in \mathbb{R}^n$. By the lemma, this is the same as $\vec{z} \cdot A^\top \vec{x} = 0$ for all z. Of course, this true because $A^\top x = 0$ (def of kernel). For the reverse inclusion: say $y \in (\operatorname{im} A)^\perp$. This means $Ax \cdot y = 0$ for all x. This is the same as $x \cdot A^\top y = 0$ for all x. This means $A^\top y = 0$, so $y \in \ker A^\top$. QED.

For the other proof: using the notation of the hint, note $\ker A^\top =$

$$\{\vec{x} \in \mathbb{R}^n \mid \begin{bmatrix} \vec{v}_1^\top \\ \vec{v}_2^\top \\ \vdots \\ \vec{v}_n^\top \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = 0\} = \{\vec{x} \in \mathbb{R}^n \mid \vec{v}_j^\top \vec{x} = 0\ \forall j = 1, \dots, n\}$$

$$= \{\vec{x} \in \mathbb{R}^n \mid \vec{v}_j \cdot \vec{x} = 0\ \forall j = 1, \dots, n\}.$$

This is exactly $(\operatorname{im} A)^\perp$ because the subspace $\operatorname{im} A$ is spanned by $\{\vec{v}_1\ \vec{v}_2\ \dots\ \vec{v}_n\}$.

The Normal Equation. The least-squares solutions of the system $A\vec{x} = \vec{b}$ are the exact solutions of the (consistent) system

$$A^\top A\vec{x} = A^\top \vec{b}.$$

The system $A^\top A\vec{x} = A^\top \vec{b}$ is called the normal equation of $A\vec{x} = \vec{b}$.

Problem 5. Practice with the Normal Equation. Use the normal equation to find the least-squares solutions of the linear system $A\vec{x} = \vec{b}$ where

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \qquad \text{and} \qquad \vec{b} = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}.$$

Are your least squares solutions actually solutions to $A\vec{x} = \vec{b}$? How can you use your answer to find the projection of $\begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$ onto the plane spanned by the columns of A?

Solution: The least-squares solutions of $A\vec{x} = \vec{b}$ are the solutions of $A^\top A\vec{x} = A^\top\vec{b}$; in this case there is only one such solution, namely $\begin{bmatrix} 5/3 \\ 0 \end{bmatrix}$. By plugging it into $A\vec{x}$, we get $\begin{bmatrix} 5/3 \\ 5/3 \\ 5/3 \end{bmatrix}$, not $\vec{b}$, so it is not an actual solution. Instead, this is the projection of $\vec{b}$ onto the plane spanned by the columns of A.

Problem 6. Proof of the Normal Equation

(a) Use Theorem 2 to show that for any $m \times n$ matrix A, $\ker(A) = \ker(A^\top A)$.
[Hint: For the harder direction, apply the fact that $V \cap V^\perp = \{0\}$ to the subspace $V = \text{im}A$.]

(b) Prove the normal equation.
[Hint: Notice that $A^\top A\vec{x} = A^\top\vec{b}$ if and only if $A\vec{x} - \vec{b} \in \ker A^\top$. Use Theorem 2.]

Solution:

(a) If $\vec{x} \in \ker(A)$, then $A^\top A\vec{x} = A^\top\vec{0} = \vec{0}$, so $\vec{x} \in \ker(A^\top A)$. Conversely, if $\vec{x} \in \ker(A^\top A)$:

$$A^\top A\vec{x} = \vec{0} \quad \Rightarrow \quad A\vec{x} \in \text{im}(A) \cap \ker(A^\top) = \text{im}(A) \cap (\text{im}A)^\perp \quad \Rightarrow \quad A\vec{x} = \vec{0}.$$

So $\vec{x} \in \ker(A)$.

(b) We need to show $A\vec{x}^* = \text{proj}_V(\vec{b})$ (where $V = \text{im}A$) if and only if $A^\top A\vec{x}^* = A^\top\vec{x}^*$. But since $V^\perp = (\text{im}A)^\perp = \ker(A^\top)$, we have

$$\begin{aligned} A^\top A\vec{x}^* = A^\top\vec{b} \quad &\iff \quad A^\top(A\vec{x}^* - \vec{b}) = \vec{0} \\ &\iff \quad A\vec{x}^* - \vec{b} \in \ker(A^\top) \\ &\iff \quad A\vec{x}^* - \vec{b} \in V^\perp \\ &\iff \quad \text{proj}_V\left(A\vec{x}^* - \vec{b}\right) = 0 \\ &\iff \quad \text{proj}_V\left(A\vec{x}^*\right) - \text{proj}_V\left(\vec{b}\right) = 0. \end{aligned}$$

The last implication is due to the linearity of the projection map. Now, since $A\vec{x}^* \in V$ and the projection map restricts to the identity on V, this says precisely that $A\vec{x}^* = \mathrm{proj}_{\mathrm{V}}\vec{b} = 0$.