Dot Products, Transposes, and Orthogonal Projections

David Jekel

November 13, 2015

Properties of Dot Products

Recall that the "dot product" or "standard inner product" on $\mathbb{R}^n$ is given by

$$\vec{x} \cdot \vec{y} = x_1 y_1 + \cdots + x_n y_n.$$

Another notation that is used for the inner product is $\langle \vec{x}, \vec{y} \rangle$. The dot product satisfies these three properties:

- It is "symmetric," which means that $\langle \vec{x}, \vec{y} \rangle = \langle \vec{y}, \vec{x} \rangle$.
- It is "linear in each variable." This means that if α and β are scalars,
$$\langle \alpha \vec{x} + \beta \vec{y}, \vec{z} \rangle = \alpha \langle \vec{x}, \vec{z} \rangle + \beta \langle \vec{y}, \vec{z} \rangle .$$
and
$$\langle \vec{z}, \alpha \vec{x} + \beta \vec{y} \rangle = \alpha \langle \vec{z}, \vec{x} \rangle + \beta \langle \vec{z}, \vec{x} \rangle .$$
- "Positivity": For any $\vec{x}$, we have $\langle \vec{x}, \vec{x} \rangle \geq 0$. The only way it can be zero is if $\vec{x} = \vec{0}$.

The first two properties can be checked by direct computation. The third one follows from the fact that

$$\langle \vec{x}, \vec{x} \rangle = x_1^2 + \cdots + x_n^2.$$

Since each x_j^2 is ≥ 0, we know $\langle \vec{x}, \vec{x} \rangle \geq 0$. Moreover, the only way it can be zero if all the terms are zero, which implies $\vec{x} = 0$.

We then define $\|\vec{x}\| = \sqrt{\langle \vec{x}, \vec{x} \rangle}$, and we know that $\|\vec{x}\| = 0$ if and only if $\vec{x} = \vec{0}$.

I draw attention to these three properties because they are the most important for what we are doing in linear algebra. Also, in more advanced math, there is a general idea of inner product (something that satisfies these properties), which is super-important and math and physics.

Properties of Transposes

Recall that the transpose of a matrix is defined by $(A^T)_{i,j} = A_{j,i}$. In other words, to find A^T you switch the row and column indexing. For example, if

$$A = \begin{pmatrix} 6 & -1 & 0 \\ 1 & 2 & 4 \end{pmatrix}, \text{ then } A^T = \begin{pmatrix} 6 & 1 \\ -1 & 2 \\ 0 & 4 \end{pmatrix}.$$

Transposes and Matrix Products: If you can multiply together two matrices A and B, then $(AB)^T = A^T B^T$. To prove this, suppose that A is $n \times k$ and B is $k \times m$. We'll compute each entry of $(AB)^T$ using the definition of matrix multiplication:

$$((AB)^T)_{i,j} = (AB)_{j,i} = A_{j,1}B_{1,i} + A_{j,2}B_{2,i} + \cdots + A_{j,k}B_{k,i}.$$

On the other hand,

$$\begin{aligned}(B^T A^T)_{i,j} &= (B^T)_{i,1}(A^T)_{1,j} + (B^T)_{i,1}(A^T)_{1,j} + \cdots + (B^T)_{i,k}(A^T)_{k,j} \\ &= B_{1,i}A_{j,1} + B_{2,i}A_{j,2} + \cdots + B_{k,i}A_{j,k},\end{aligned}$$

which is the same as $((AB)^T)_{i,j}$.

If this is true when you multiply two matrices together, then it must also be true when you multiply three or more matrices together (formally, we prove it by "mathematical induction"). Thus, for instance,

$$(ABC)^T = C^T B^T A^T.$$

Transposes and Dot Products: Suppose that $\vec{x}$ and $\vec{y}$ are two vectors in $\mathbb{R}^n$. We can view them as column vectors or $n \times 1$ matrices, and then the dot product can be written as

$$(\vec{x})^T \vec{y} = x_1 y_1 + \cdots + x_n y_n = \langle \vec{x}, \vec{y} \rangle.$$

Next, suppose that A is an $n \times m$ matrix and $\vec{x} \in \mathbb{R}^m$ and $\vec{y} \in \mathbb{R}^n$. Then

$$\langle A\vec{x}, \vec{y} \rangle = (A\vec{x})^T \vec{y} = (\vec{x})^T A^T \vec{y} = \left\langle \vec{x}, A^T \vec{y} \right\rangle,$$

or in different notation $(A\vec{x}) \cdot \vec{y} = \vec{x} \cdot (A^T \vec{y})$.

Remarks: The property that $\langle A\vec{x}, \vec{y} \rangle = \left\langle \vec{x}, A^T \vec{y} \right\rangle$ is an essential property of the transpose. In suped-up versions of linear algebra that are used in physics, the property that $\langle A\vec{x}, \vec{y} \rangle = \left\langle \vec{x}, A^T \vec{y} \right\rangle$ is basically taken as the definition of the transpose.

Exercises:

- If the proof that $(AB)^T = B^T A^T$ seemed confusingly general, then on a piece of scratch paper, write down two "random" matrices A and B and compute $(AB)^T$ and $B^T A^T$.

- Prove that if A is an invertible square matrix, then $(A^T)^{-1} = (A^{-1})^T$.
- Let $\vec{e}_j$ be the jth standard basis vector. If A is any matrix, show that $A_{i,j} = \langle \vec{e}_i, A\vec{e}_j \rangle$.
- Suppose that $\langle \vec{x}, A\vec{y} \rangle = \langle \vec{x}, B\vec{y} \rangle$ for all $\vec{x}$ and $\vec{y}$. Prove that $A = B$.
- Let A be some matrix. Then A^T is the only possible value of B that would satisfy $\langle A\vec{x}, \vec{y} \rangle = \langle \vec{x}, B\vec{y} \rangle$ for all values of $\vec{x}$ and $\vec{y}$.

Finding a Basis for Orthogonal Complements

In the homework and tests, you might encounter problems like this:

Problem: Let $V \subset \mathbb{R}^3$ be given by

$$V = \text{Span}\left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 4 \\ 5 \\ 2 \end{pmatrix} \right\}$$

Compute $V^\perp$ (that is, find a basis for $V^\perp$).

Method 1: One way to do this is using the cross product: as explained in lecture the vector

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \times \begin{pmatrix} 4 \\ 5 \\ 2 \end{pmatrix} = \begin{pmatrix} -11 \\ 10 \\ -3 \end{pmatrix}$$

must be orthogonal to $(1,2,4)^T$ and $(4,5,2)^T$, and hence it is orthogonal to V. Since $\dim V = 2$, we know that $\dim V^\perp$ must be one (as stated in lecture notes 7). Thus, $(-11, 10, -3)^T$ has to be a basis for $V^\perp$.

However, the cross product method is special to the case where V is a plane in $\mathbb{R}^3$. For instance, you cannot compute the orthogonal complement of a plane in $\mathbb{R}^4$ using the cross product, since there is no cross product in $\mathbb{R}^4$. But luckily there is a more general way.

Method 2: We are going to interpret $V^\perp$ as the kernel of some matrix. We know (HW 7 problem 7) that a vector $\vec{w}$ is in $V^\perp$ if and only if it is perpendicular to $(1,2,3)^T$ and $(4,5,2)^T$. In other words, it is in $V^\perp$ if and only if

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \cdot \vec{w} = \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \vec{w} = 0$$

and

$$\begin{pmatrix} 4 \\ 5 \\ 2 \end{pmatrix} \cdot \vec{w} = \begin{pmatrix} 4 & 5 & 2 \end{pmatrix} \vec{w} = 0.$$

This is equivalent to saying that

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 2 \end{pmatrix} \vec{w} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Thus, computing $V^\perp$ amounts to computing the set of solutions to this system of equations, or the kernel of this matrix. We already know how to do this using the RREF. After some computation, the RREF is

$$\begin{pmatrix} 1 & 0 & -11/3 \\ 0 & 1 & 10/3 \end{pmatrix}.$$

The first two variables have leading ones and the third variable is free. Setting the free variable to be t, we find that all the solutions have the form

$$\begin{pmatrix} 11t/3 \\ -10t/3 \\ t \end{pmatrix} = t \begin{pmatrix} 11/3 \\ -10/3 \\ 1 \end{pmatrix}.$$

Thus, this vector forms a basis for $V^\perp$. This agrees with our earlier answer from Method 1 because the vector from Method 2 is a scalar multiple of the vector from Method 1. Though Method 2 required explanation, there was not much actual computation–just row reducing one matrix.

Generalization: Method 2 didn't use the fact that we were in $\mathbb{R}^3$, and so it generalizes to finding bases for orthogonal complements in all dimensions. Suppose that $V \subset \mathbb{R}^n$ is spanned by $\vec{v}_1, \ldots, \vec{v}_m$. Let A be the matrix with columns $\vec{v}_1, \ldots, \vec{v}_m$. Then $V = \operatorname{im} A$ and $V^\perp = (\operatorname{im} A)^\perp$.

By HW 7 problem 7, we know that $\vec{w}$ is in $V^\perp$ if and only if $\vec{v}\cdot\vec{w} = (\vec{v}_j)^T\vec{w} = 0$ for all j. This is equivalent to saying that

$$\begin{pmatrix} (\vec{v}_1)^T \\ \vdots \\ (\vec{v}_m)^T \end{pmatrix} \vec{w} = \vec{0}.$$

That is, $\vec{w}$ is in the kernel of the matrix with row $(\vec{v}_1)^T, \ldots, (\vec{v}_m)^T$. This matrix is just A^T. Thus, we've proved that $V^\perp$ is the kernel of A^T. To summarize, if $\vec{v}_1, \ldots, \vec{v}_m$ are any vectors in $\mathbb{R}^n$, then

$$(\operatorname{Span}(\vec{v}_1, \ldots, \vec{v}_m))^\perp = \ker \begin{pmatrix} (\vec{v}_1)^T \\ \vdots \\ (\vec{v}_m)^T \end{pmatrix}.$$

We can compute a basis for the kernel from the RREF of this matrix.

Orthogonal Complements of Kernel and Image

In the last section, we had a subspace V and wrote $V = \operatorname{im} A$ for a matrix A, and we showed that $V^\perp = \ker A^T$. But the matrix A could have been anything. Thus, we have essentially proved that

$$(\operatorname{im} A)^\perp = \ker A^T \text{ for any } A.$$

Replacing A by A^T, we can conclude that

$$(\operatorname{im} A^T)^\perp = \ker A \text{ for any } A.$$

Another Proof: For good measure, let's give a slightly different proof that $(\operatorname{im} A)^\perp = \ker A^T$. Why give another proof? Well, this proof will relate to what we did in the first section. It is also more "coordinate-free": It does not explicitly refer to the rows and columns of the matrix or require us to choose a basis. Rather it uses the important property of the transpose that $\langle A\vec{x}, \vec{y}\rangle = \langle \vec{x}, A^T\vec{y}\rangle$.

In order to prove that that $(\operatorname{im} A)^\perp = \ker A^T$, we want to show that (1) anything in $\ker A^T$ is in $(\operatorname{im} A)^\perp$ and (2) anything in $(\operatorname{im} A)^\perp$ is in $\ker A^T$.

1. Suppose that $\vec{y} \in \ker A^T$, and we will prove $\vec{y}$ is orthogonal to everything in the image of A. Any element of the image of A can be written as $A\vec{x}$. Then

$$\langle A\vec{x}, \vec{y}\rangle = \langle \vec{x}, A^T\vec{y}\rangle = \left\langle \vec{x}, \vec{0}\right\rangle = 0.$$

   Therefore, $\vec{y} \in (\operatorname{im} A)^\perp$.

2. Suppose that $\vec{y} \in (\operatorname{im} A)^\perp$. Note $AA^T\vec{y}$ is in the image of A, and thus, $\vec{y} \perp AA^T\vec{y}$. Therefore,

$$0 = \langle A(A^T\vec{y}), \vec{y}\rangle = \langle A^T\vec{y}, A^T\vec{y}\rangle = \left\|A^T\vec{y}\right\|^2.$$

   Thus, $\left\|A^T\vec{y}\right\| = 0$ and therefore, $A^T\vec{y} = \vec{0}$, so $\vec{y} \in \ker A^T$.

Example: Let's work this out in an example and picture it geometrically:

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad A^T = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Verify that

$$\ker A = \operatorname{Span}\left\{\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right\}, \qquad \operatorname{im} A = \operatorname{Span}\left\{\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}\right\}$$

$$\operatorname{im} A^T = \operatorname{Span}\left\{\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right\}, \qquad \ker A^T = \operatorname{Span}\left\{\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right\}.$$

Draw a picture of $\ker A$ and $\operatorname{im} A^T$ in $\mathbb{R}^2$ and $\ker A^T$ and $\operatorname{im} A$ in $\mathbb{R}^3$. Verify that $(\operatorname{im} A)^\perp = \ker A^T$ and $(\operatorname{im} A^T)^\perp = \ker A$.

If you want, see what happens with some other matrices: Draw pictures of the kernel and image of A and A^T.

Formula for Orthogonal Projection

The material in this section is NOT something you need to know for the tests. But it has a lot of good insights and I hope it will be useful for your future study of math and its applications.

Assume that V is a subspace of $\mathbb{R}^n$. We want to prove the following:

- Any $\vec{x} \in \mathbb{R}^n$ can be uniquely written as $\vec{x} = \vec{v} + \vec{w}$, where $\vec{v} \in V$ and $\vec{w} \in V^{\perp}$.
- There is an "orthogonal projection" matrix P such that $P\vec{x} = \vec{v}$ (if $\vec{x}$, $\vec{v}$, and $\vec{w}$ are as above).
- In fact, we can find a nice formula for P.

Setup: Our strategy will be to create P first and then use it to verify all the above statements. We know that any subspace of $\mathbb{R}^n$ has a basis. So let $\vec{v}_1, \ldots, \vec{v}_m$ be a basis for V. Let A be the matrix with columns $\vec{v}_1, \ldots, \vec{v}_m$:

$$A = \begin{pmatrix} \vec{v}_1 & \ldots & \vec{v}_m \end{pmatrix}.$$

I claim that the P that we want is

$$P = A(A^T A)^{-1} A^T.$$

Example: To convince you that this formula is believable, let's see what it tells us in the simple case where V is one-dimensional. Suppose $\vec{v}$ is the line spanned by $\vec{v}$. In that case, there is only one vector in the basis ($m = 1$), and A is just the column vector $\vec{v}$ viewed as an $n \times 1$ matrix. Then we have

$$A^T A = \vec{v}^T \vec{v} = \|\vec{v}\|^2,$$

which is a 1×1 matrix or scalar. Then

$$P = \frac{\vec{v}\vec{v}^T}{\|\vec{v}\|^2},$$

and

$$P\vec{x} = \frac{1}{\|\vec{v}\|^2} \vec{v}(\vec{v}^T \vec{x}) = \frac{\vec{v} \cdot \vec{x}}{\|\vec{v}\|^2} \vec{v}.$$

This agrees with the formula for orthogonal projection onto a line given in class, so the above formula for P is correct in the one-dimensional case.

Organization of Proof: The general formula $P = A(A^T A)^{-1} A^T$ is rather sneaky and not obvious. I am going to keep you in suspense for a while about why it works in general, so this proof will require patience. The proof might seem rather long, but that is because I included fuller explanations of each step than are usually in the lecture notes.

The proof will be organized into short claims so that we don't try to do too many things as once. The first order of business is to prove that $A^T A$ is actually invertible; otherwise, the formula for P doesn't even make sense.

Claim. *$A^T A$ is invertible.*

Proof. First, note that $A^T A$ is square $m \times m$ because A is $n \times m$ and A^T is $m \times n$. We have a theorem given in lecture about equivalent conditions for invertibility of square matrices; we know that $A^T A$ must be invertible if we can prove that $\ker(A^T A) = \{0\}$.

Suppose that $\vec{x} \in \ker(A^T A)$ and we will prove $\vec{x} = \vec{0}$. The first step is to show $A\vec{x} = \vec{0}$. Because $A^T A\vec{x} = \vec{0}$, we know that

$$\vec{x}^T A^T A\vec{x} = 0.$$

But by the properties of transposes given earlier,

$$\vec{x}^T A^T A\vec{x} = (A\vec{x})^T (A\vec{x}) = (A\vec{x}) \cdot (A\vec{x}) = \|A\vec{x}\|^2 .$$

Therefore, $\|A\vec{x}\| = 0$, and hence $A\vec{x} = \vec{0}$.

If the entries of $\vec{x}$ are $x_1, \ldots, x_m$, then

$$\vec{0} = A\vec{x} = \begin{pmatrix} \vec{v}_1 & \ldots & \vec{v}_m \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} = x_1\vec{v}_1 + \ldots x_m\vec{v}_m.$$

By assumption $\vec{v}_1, \ldots, \vec{v}_m$ are linearly independent. Thus, the only way this linear combination can be zero is if all the x_j's are zero. Therefore, $\vec{x} = \vec{0}$. This completes the proof that $\ker(A^T A) = \{\vec{0}\}$ and hence $A^T A$ is invertible. □

Claim. *Now comes the sneaky part ... P has the following properties:*

a. $PA = A$.

b. $A^T P = A^T$.

c. $P^2 = P$.

Proof. Note $PA = A(A^T A)^{-1} A^T A$. The $(A^T A)^{-1}$ and $A^T A$ cancel, and so $PA = A$, which proves (a). The reason for (b) is similar:

$$A^T P = A^T A(A^T A)^{-1} A^T = A^T.$$

Finally, to prove (c),

$$P^2 = PP = A(A^T A)^{-1} A^T A(A^T A)^{-1} A^T = A(A^T A)^{-1} A^T = P.$$ □

Claim. $\operatorname{im} P = \operatorname{im} A = V$.

Proof. To show $\operatorname{im} P = \operatorname{im} A$, we need to do two things:

- Suppose that $\vec{x} \in \operatorname{im} A$ and prove $\vec{x} \in \operatorname{im} P$.
- Suppose that $\vec{x} \in \operatorname{im} P$ and prove $\vec{x} \in \operatorname{im} P$.

For the first step, assume $\vec{x} \in \operatorname{im} A$. This means that $\vec{x} = A\vec{y}$ for some $\vec{y}$. Then because $PA = A$,

$$P\vec{x} = PA\vec{y} = A\vec{y} = \vec{x}.$$

Thus, $\vec{x} = P\vec{x}$. So $\vec{x}$ is P of *something* which means it is in the image of A.

For the second step, suppose $\vec{x} \in \operatorname{im} P$. Then $\vec{x} = P\vec{y}$ for some $\vec{y}$, which implies

$$\vec{x} = A[(A^TA)^{-1}A^T\vec{y}];$$

thus $\vec{x}$ is A times *something* and hence is in the image of A.

This proves $\operatorname{im} P = \operatorname{im} A$, and the statement $\operatorname{im} A = V$ was proved in the Setup. □

Claim. $\ker P = \ker A^T = V^\perp$.

Proof. The outline is similar to the previous proof. First, suppose $\vec{x} \in \ker A^T$, and we will prove $\vec{x} \in \ker P$. By assumption $A^T\vec{x} = \vec{0}$ and therefore

$$P\vec{x} = A(A^TA)^{-1}A^T\vec{x} = A(A^TA)^{-1}\vec{0} = \vec{0},$$

which means $\vec{x} \in \ker P$.

Next, suppose $\vec{x} \in \ker P$ and we will prove $\vec{x} \in \ker A^T$. Since $\vec{x} \in \ker P$, $P\vec{x} = \vec{0}$. But we showed earlier that $A^TP = A^T$, and hence

$$A^T\vec{x} = A^TP\vec{x} = A^T\vec{0} = \vec{0},$$

so $\vec{x} \in \ker P$.

Therefore, $\ker A^T = \ker P$. But we proved in the previous section that $\ker A^T = (\operatorname{im} A)^\perp = V^\perp$. □

Claim. *If $\vec{x} \in V$, then $P\vec{x} = \vec{x}$, and if $\vec{x} \in V^\perp$, then $P\vec{x} = 0$.*

Proof. We have basically already proved this. If $\vec{x} \in V = \operatorname{im} P$, then $\vec{x} = P\vec{y}$ for some $\vec{y}$, and so $P\vec{x} = PP\vec{y} = P^2\vec{y} = P\vec{y} = \vec{x}$. On the other hand, if $\vec{x} \in V^\perp = \ker P$, then $P\vec{x} = \vec{0}$. □

Claim. *Any $\vec{x} \in \mathbb{R}^n$ can be written uniquely as $\vec{v}+\vec{w}$, where $\vec{v} \in V$ and $\vec{w} \in V^\perp$.*

Proof. Let $\vec{v} = P\vec{x}$ and $\vec{w} = \vec{x} - P\vec{x} = (I - P)\vec{x}$. Then $\vec{v}$ is P of something, so it is in the image of P, which is V. To show that $\vec{w} \in V^\perp$, we use the fact that $P^2 = P$, so

$$P\vec{w} = P(I - P)\vec{w} = (P - P^2)\vec{w} = 0;$$

this implies $\vec{w} \in \ker P = V^\perp$.

This shows that $\vec{x}$ *can* be written as $\vec{v} + \vec{w}$ with $\vec{v} \in V$ and $\vec{w} \in V^\perp$. But we still have to prove that the decomposition is *unique*; that is, this is the *only* way to write $\vec{x}$ as the sum of something in V and something in $V^\perp$. Well, suppose that we have another decomposition $\vec{x} = \vec{v}' + \vec{w}'$ where $\vec{v}' \in V$ and $\vec{w}' \in V^\perp$; our goal is to prove that $\vec{v} = \vec{v}'$ and $\vec{w} = \vec{w}'$. Using the previous claim,

$$P\vec{x} = P\vec{v}' + P\vec{w}' = \vec{v}' + \vec{0}.$$

Therefore, $\vec{v}' = P\vec{x} = \vec{v}$ and $\vec{w}' = \vec{x} - \vec{v}' = \vec{x} - \vec{v} = \vec{w}$, so the decomposition is indeed unique. □

Recap: We have now verified that any $\vec{x}$ can be written uniquely as the sum of some $\vec{v} \in V$ and some $\vec{w} \in V^\perp$, and that $P\vec{x} = \vec{v}$. We have proven the orthogonal projection exists and found a formula for it. The crucial ingredients were the fact that $\ker A^T = (\operatorname{im} A)^\perp$ and that $PA = A$, $A^T P = A^T$, and $P^2 = P$.

Example: Let's see how this formula for P works in practice. Suppose that $V \subset \mathbb{R}^3$ is given by

$$V = \operatorname{Span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -3 \end{pmatrix} \right\}.$$

Then

$$A^T A = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & -3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & -3 \end{pmatrix} = \begin{pmatrix} 5 & -6 \\ -6 & 10 \end{pmatrix}.$$

Then using the formula for the inverse of a 2×2 matrix, we have

$$\begin{aligned} P = A(A^T A)^{-1} A^T &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & -3 \end{pmatrix} \frac{1}{14} \begin{pmatrix} 10 & 6 \\ 6 & 5 \end{pmatrix} \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & -3 \end{pmatrix} \\ &= \frac{1}{14} \begin{pmatrix} 10 & 6 \\ 6 & 5 \\ 2 & -3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & -3 \end{pmatrix} \\ &= \frac{1}{14} \begin{pmatrix} 10 & 6 & 2 \\ 6 & 5 & -3 \\ 2 & -3 & 13 \end{pmatrix} \end{aligned}$$

Verify the above computation and check that $P^2 = P$.

Remarks: In this case, the computation was rather clean because our basis had so many zeros and ones in it. But often in practice you get such nice bases: For instance, if you compute the kernel of some matrix using the RREF, you get a basis with a lot of ones and zeros from parametrizing each of the free variables.

Challenge: In this example P is symmetric (that is, $P^T = P$). Prove that the matrix of *any* orthogonal projection is symmetric.