L11

Properties of Orthogonal Projections

Last Time: if V is a subspace of $\mathbb{R}^n$ and $b \in \mathbb{R}^n$ then:

$$b = b_V + b_{V^\perp} \quad \text{is its orthogonal decomposition}$$

where b_V = orthogonal projection of b onto V

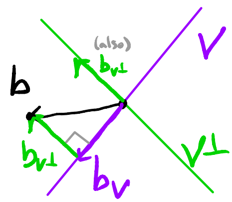
= closest vector to b in V

and $b_{V^\perp}$ = orthogonal projection of b onto $V^\perp$

= closest vector to b in $V^\perp$

$= b - b_V$ = arrow from b_V to b

The distance from b to V is $\|b_{V^\perp}\|$.



The projection b_V is characterized by the property that $\boxed{b - b_V \in V^\perp.}$

[DEMO]

Properties of Orthogonal Projections

(1) $b_V = b \iff b_{V^\perp} = 0 \iff b \in V$

(2) $b_V = 0 \iff b_{V^\perp} = b \iff b \in V^\perp$

(3) $(b_V)_V = b_V$

(4) $b_{\mathbb{R}^n} = b$ (5) $b_{\{0\}} = 0$

What do these equations mean?

Remember, reasoning about orthogonal projections means thinking geometrically.

(1) This says

"b is the closest vector in V to b" $\Longleftrightarrow$ $b \in V$

which is obvious!

$$\boxed{\text{projection onto } V \text{ doesn't move vectors in } V}$$

Then substitute $b = b_V$ into the orthogonal decomposition $b = b_V + b_{V^\perp} \rightsquigarrow b_{V^\perp} = 0$.

(2) This is the same as (1) with $V^\perp$ in place of V. It means

"$V^\perp$ = all vectors in $\mathbb{R}^n$ that are closest to 0 in V."

(3) This says that

"projecting twice is the same as projecting once."

(Orthogonal projection is an idempotent operation.)

Since $b_V \in V$ we have $(b_V)_V = b_V$ by (1).

[demo again]

(4) follows from (1) because $b \in \mathbb{R}^n$

(5) The closest vector in $\{0\}$ to b is 0.

Eg: $b = \begin{pmatrix}1\\1\\1\end{pmatrix}$, $V = \mathrm{Col}(A)$, $A = \begin{pmatrix}1 & -1 & -1\\ 2 & 1 & 4\\ 1 & -1 & -1\end{pmatrix}$

Last time we computed $b_V = b$. That should mean $b \in V = \mathrm{Col}(A)$, i.e. that $Ax = b$ is consistent. Let's check:

$$\left(\begin{array}{ccc|c}1 & -1 & -1 & 1\\ 2 & 1 & 4 & 1\\ 1 & -1 & -1 & 1\end{array}\right) \xrightarrow{\text{rref}} \begin{array}{c}\qquad\qquad \text{consistent} \\ \qquad\qquad\ \downarrow \\ \left(\begin{array}{ccc|c}1 & 0 & 1 & 2/3\\ 0 & 1 & 2 & -1/3\\ 0 & 0 & 0 & 0\end{array}\right)\end{array} \qquad \checkmark$$

Projection Matrices

Recall: If $V = \mathrm{Col}(A)$ then you can compute b_V as follows:

(1) Solve the normal equation $A^TA\hat{x} = A^Tb$

(2) $b_V = A\hat{x}$ for any solution $\hat{x}$.

Fact: A has FCR $\iff$ A^TA is invertible.

(This is a HW problem.)

NB: A^TA is indeed square (symmetric even!).

In this case, $A^TA\hat{x} = A^Tb$ has a unique solution, namely, $\hat{x} = (A^TA)^{-1}A^Tb$, so $b_V = A\hat{x} = A(A^TA)^{-1}A^Tb$. This means you can compute orthogonal projections by multiplying by the $m\times m$ matrix

$$P_V = \boxed{A(A^TA)^{-1}A^T} \leftarrow \text{the Horrible Formula}$$

$$\boxed{P_V b = b_V} \quad \text{for all } b$$

Analogy: This is kind of like computing the solution of $Ax=b$ by multiplying $x = A^{-1}b$ (when A is invertible)

Eg: $V = \mathrm{Col}(A)$, $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}$

$$A^TA = \begin{pmatrix} 3 & 2 \\ 2 & 2 \end{pmatrix} \qquad (A^TA)^{-1} = \frac{1}{2}\begin{pmatrix} 2 & -2 \\ -2 & 3 \end{pmatrix}$$

$$P_V = A(A^TA)^{-1}A^T = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix} \frac{1}{2}\begin{pmatrix} 2 & -2 \\ -2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Last time we computed the projection of $b = \begin{pmatrix}1\\0\\0\end{pmatrix}$ is $b_V = \begin{pmatrix}1/2\\1/2\\0\end{pmatrix}$. Let's check:

$$b_V = P_V b = \begin{pmatrix}1/2 & 1/2 & 0\\ 1/2 & 1/2 & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}1\\0\\0\end{pmatrix} = \begin{pmatrix}1/2\\1/2\\0\end{pmatrix} \checkmark$$

Now it's much easier to compute orthogonal projections of as many b's as we want!

We've produced a "matrix that computes projections." What does this mean?

Fact: A matrix A is determined by the products Ax for all vectors $x \in \mathbb{R}^n$.

Indeed, $Ae_i = i^{th}$ col of A, so A is actually determined by the products Ax for $x = e_1, e_2, \ldots, e_n$.

In other words, A is determined by the function

$$\underset{\text{input}}{x} \rightsquigarrow \underset{\text{output}}{Ax}$$

Eg: If $Ax = x$ for every vector x then $A = I_n$ because I_n does the same thing. ($I_n x = x$ for all x)

Eg: If $Ax = 0$ for every vector x then $A = 0$ (zero matrix) because 0 does the same thing. ($0x = 0$ for all x)

Def: Let V be a subspace of $\mathbb{R}^n$. The projection matrix onto V is the matrix P_V that is defined by

$$\boxed{P_V b = b_V}$$

for all vectors b.

This is the first time we have defined a matrix by its action on all vectors in $\mathbb{R}^n$. It makes sense to do so by the Fact above.

We have a horrible formula for P_V above (when $V = \mathrm{Col}(A)$ and A has FCR), but that's a terrible way to understand / visualize / reason about projection matrices.

Like orthogonal projections, the projection matrix is a geometric construction, so reasoning about it means thinking geometrically.

That said, we also want to compute P_V. We know how when $V = \mathrm{Col}(A)$ and A has FCR. But that means the columns of A are LI $\implies$ they form a basis for V (= their span). So for a general subspace V, first we'll need a basis.

How to Compute P_V in General: (lots of other ways later!)

(1) Find a (any!) basis $\{v_1, v_2, \ldots, v_d\}$ for V.

(2) $B = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_d \\ | & & | \end{pmatrix}$ has FCR & $Col(B) = V$.
(change descriptions)

(3) $P_V = B(B^TB)^{-1}B^T$
(horrible formula)

Eg: Compute P_V for $V = Col\begin{pmatrix} 1 & -1 & -1 \\ 2 & 1 & 4 \\ 1 & -1 & -1 \end{pmatrix}$

We know how to find a basis for a column space: pivot columns!

$$\begin{pmatrix} 1 & -1 & -1 \\ 2 & 1 & 4 \\ 1 & -1 & -1 \end{pmatrix} \xrightarrow{REF} \begin{pmatrix} 1 & -1 & -1 \\ 0 & 3 & 6 \\ 0 & 0 & 0 \end{pmatrix} \xrightarrow{\text{basis}} \left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix} \right\}$$

$$B = \begin{pmatrix} 1 & 1 \\ 2 & -1 \\ 1 & 1 \end{pmatrix} \rightsquigarrow P_V = B(B^TB)^{-1}B^T = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix}$$

Check: $b = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \Rightarrow b_V = P_V b = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ ✓
(again)

Objection! Can't I simplify the horrible formula?

$$B(B^TB)^{-1}B^T = B(B^{-1}(B^T)^{-1})B^T = (BB^{-1})((B^T)^{-1}B^T) \overset{??}{=} I_n$$

This only works if B is square: otherwise B^{-1} doesn't make sense.

Sanity Check: OK, so what if $V = \mathrm{Col}(B)$ for a square matrix B with FCR?

In this case B is invertible, so it has FRR, so $V = \mathrm{Col}(B) = \mathbb{R}^n$. Then $P_V b = P_{\mathbb{R}^n} b = b_{\mathbb{R}^n} = b$

we showed this before

$\Rightarrow P_V = I_n$ ✓

NB: Like orthogonal projections, projection matrices only depend on V and not your description of V. Once V is fixed, then P_V is a matrix with numbers in it, which you can compute in several ways.

→ More on this later.

When $V = \mathrm{Span}\{v\}$ is a line, it's easy to compute P_V.

$V = \mathrm{Col}(A)$ where $A = v$ (matrix with one column).

$A^T A = v^T v = v \cdot v$ is a 1×1 matrix (number)

$$\Rightarrow (A^T A)^{-1} = \frac{1}{v \cdot v}$$

$$\Rightarrow P_V = A(A^T A)^{-1} A^T = v \frac{1}{v \cdot v} v^T = \frac{1}{v \cdot v} v v^T$$

outer product

Or, if you like,

$$P_V = \frac{v v^T}{v^T v} = \frac{\text{outer product}}{\text{inner product}}$$

How to Compute P_V when $V = \text{Span}\{v\}$ is a Line:

$$\boxed{P_V = \frac{1}{v \cdot v} v v^T}$$

For any vector b,

$$P_V b = \frac{1}{v\cdot v} v v^T \cdot b = \frac{1}{v \cdot v} v (v^T b) = \frac{1}{v\cdot v} v (v \cdot b) = \frac{v\cdot b}{v \cdot v} v$$

so this recovers the formula for projection onto a line from last time.

Eg: Compute P_V for $V = \text{Span}\left\{ \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}$.

$$P_V = \frac{1}{\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$ easy!

The geometric properties of orthogonal projections translate into properties of projection matrices.

Properties of Projection Matrices

V: subspace of $\mathbb{R}^n$ $\quad$ P_V: projection matrix

(1) $\text{Col}(P_V) = V$

(2) $\text{Nul}(P_V) = V^\perp$

(3) $P_V^2 = P_V$ (idempotence)

(4) $P_V + P_{V^\perp} = I_n$

(5) $P_V = P_V^T$ (symmetry)

(6) $P_{\mathbb{R}^n} = I_n$

(7) $P_{\{0\}} = 0$

None of these should be at all mysterious!
(except maybe (5) — but we'll fix that in about 5 weeks)
You'll understand them a lot better if you internalize the

Proofs: Remember, reasoning about projection matrices means thinking geometrically.

(1) $b \in \mathrm{Col}(P_V)$

$\iff P_V x = b$ has a solution

equal

$\iff x_V = b$ has a solution

$\iff b$ is the projection of some vector x.

Any vector in V is the projection of itself, and the projection of any vector is (by definition) in V.

(2) $x \in \mathrm{Nul}(P_V) \iff P_V x = 0$

$\iff x_V = 0 \iff x \in V^\perp$

we showed this before

(3) $P_V^2 b = P_V(P_V b) = P_V(b_V) = (b_V)_V = b_V = P_V b$

Since $P_V^2 b = P_V b$ for every vector b, $P_V^2 = P_V$ (a matrix is determined by its action on every vector).

(4) $(P_V + P_{V^\perp})b = P_V b + P_{V^\perp} b = b_V + b_{V^\perp} = b = I_n b$

orthogonal decomposition

Since $(P_V + P_{V^\perp})b = I_n b$ for every vector b, $P_V + P_{V^\perp} = I_n$.

(5) You can prove this by doing matrix algebra on the horrible formula, but we'll have a better reason (spectral theorem) in about 5 weeks.

(6) $P_{\mathbb{R}^n} b = b_{\mathbb{R}^n} = b = I_n b$ for every vector b

we showed this before

(7) $P_{\{0\}} b = b_{\{0\}} = 0 = 0b$ for every vector b.

Last time: If $V = \mathrm{Nul}(A)$ we computed b_V by first projecting onto $V^\perp = \mathrm{Row}(A)$. We can do the something similar for P_V using (4) above.

Another Way to Compute P_V:

(1) Find a basis for $V^\perp$.

(2) Compute $P_{V^\perp}$.

(3) $P_V = I_n - P_{V^\perp}$

This is faster than the general method when it's easier to compute a basis for $V^\perp$ than for V.

Eg: Compute P_V for $V = \text{Nul}\begin{pmatrix}1 & 2 & 1\end{pmatrix}$.

In this case, $V^\perp = \text{Row}\begin{pmatrix}1 & 2 & 1\end{pmatrix} = \text{Span}\left\{\begin{pmatrix}1\\2\\1\end{pmatrix}\right\}$ is a line, so $P_{V^\perp}$ is easy to compute!

$$P_{V^\perp} = \frac{1}{\begin{pmatrix}1\\2\\1\end{pmatrix}\cdot\begin{pmatrix}1\\2\\1\end{pmatrix}}\begin{pmatrix}1\\2\\1\end{pmatrix}\begin{pmatrix}1 & 2 & 1\end{pmatrix} = \frac{1}{6}\begin{pmatrix}1 & 2 & 1\\2 & 4 & 2\\1 & 2 & 1\end{pmatrix} = \begin{pmatrix}1/6 & 1/3 & 1/6\\1/3 & 2/3 & 1/3\\1/6 & 1/3 & 1/6\end{pmatrix}$$

$$\Rightarrow P_V = I_3 - P_{V^\perp} = \begin{pmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{pmatrix} - \begin{pmatrix}1/6 & 1/3 & 1/6\\1/3 & 2/3 & 1/3\\1/6 & 1/3 & 1/6\end{pmatrix} = \begin{pmatrix}5/6 & -1/3 & -1/6\\-1/3 & 1/3 & -1/3\\-1/6 & -1/3 & 5/6\end{pmatrix}$$

That was way easier than the general procedure in this case. To illustrate let's try the general procedure.

(1) Find a basis for $\text{Nul}\begin{pmatrix}1 & 2 & 1\end{pmatrix}$

$$\begin{pmatrix}1 & 2 & 1\end{pmatrix}x = 0 \xrightarrow{\text{PVF}} x = x_2\begin{pmatrix}-2\\1\\0\end{pmatrix} + x_3\begin{pmatrix}-1\\0\\1\end{pmatrix}$$

(2) $B = \begin{pmatrix}-2 & -1\\1 & 0\\0 & 1\end{pmatrix}$

(3) $B^TB = \begin{pmatrix}5 & 2\\2 & 2\end{pmatrix}$ $\qquad (B^TB)^{-1} = \frac{1}{6}\begin{pmatrix}2 & -2\\-2 & 5\end{pmatrix}$

$$P_V = B(B^TB)^{-1}B^T = \cdots = \begin{pmatrix}5/6 & -1/3 & -1/6\\-1/3 & 1/3 & -1/3\\-1/6 & -1/3 & 5/6\end{pmatrix}$$

That was a harder way to compute the same matrix.

Upshots:

(1) Ask yourself: what's the easiest way to compute P_V? You have several options already.

- $\rightarrow$ Is V a line?
- $\rightarrow$ Is it easier to compute $P_{V^\perp}$?

etc.

(2) You get the same matrix no matter which computation you do.

In the above example,

$V \xrightarrow{\text{easy way}}$ describe as $\mathrm{Nul}\begin{pmatrix}1 & 2 & 1\end{pmatrix} \xrightarrow{I_3 - \frac{vv^T}{v\cdot v}} P_V = \begin{pmatrix} 5/6 & -1/3 & -1/6 \\ -1/3 & 1/3 & -1/3 \\ -1/6 & -1/3 & 5/6 \end{pmatrix}$

same matrix

$V \xrightarrow{\text{hard way}}$ describe as $\mathrm{Col}\begin{pmatrix} -2 & -1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \xrightarrow[\text{formula}]{\text{horrible}} P_V = \begin{pmatrix} 5/6 & -1/3 & -1/6 \\ -1/3 & 1/3 & -1/3 \\ -1/6 & -1/3 & 5/6 \end{pmatrix}$

Be careful to distinguish

what P_V is vs. **ways to compute P_V**

(the matrix that computes orthogonal projections)

Here's another example. Let $V = \mathrm{Col}\begin{pmatrix} 1 & -1 & -1 \\ 2 & 1 & 4 \\ 1 & -1 & -1 \end{pmatrix}$.

Before we computed P_V by finding the pivot columns and using the horrible formula:

$$P_V = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix}$$

Eg: Compute P_V in a completely different way.

We know that there were 2 pivot columns $\rightsquigarrow$ rank 2, so V is a plane in $\mathbb{R}^3 \Rightarrow V^\perp$ is a line. So let's find a basis for $V^\perp$ and compute $P_{V^\perp}$.

$$V^\perp = \mathrm{Nul}\begin{pmatrix} 1 & 2 & 1 \\ -1 & 1 & -1 \\ -1 & 4 & -1 \end{pmatrix} \xrightarrow{\text{PVF}} \mathrm{Span}\left\{\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}\right\}$$

$$\Rightarrow P_{V^\perp} = \frac{1}{\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}\cdot\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}}\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}\begin{pmatrix} 1 & 0 & -1 \end{pmatrix} = \begin{pmatrix} 1/2 & 0 & -1/2 \\ 0 & 0 & 0 \\ -1/2 & 0 & 1/2 \end{pmatrix}$$

(we did this in another example)

$$\Rightarrow P_V = I_3 - P_{V^\perp} = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{pmatrix} \checkmark$$

Compare the last example of the previous lecture.