Properties of Orthogonal Projections

Recall: if V is a subspace of $\mathbb{R}^n$ and $b \in \mathbb{R}^n$,

$$b = b_V + b_{V^\perp}$$

is its orthogonal decomposition with respect to V.

b_V = orthogonal projection of b onto V

= closest vector in V to b

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

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

The distance from b to V is

$$\|b - b_V\| = \|b_{V^\perp}\|.$$

[demos]

Properties of Projections:

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

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

(3) $(b_V)_V = b_V$

(1) says:

"b is the closest vector in V to itself"

$$\Updownarrow$$

"b is already in V"

In this case, the distance from b to V is zero, so $\|b_{V^\perp}\| = 0 \implies b_{V^\perp} = 0$.

Or: since $b = b_V + b_{V^\perp}$, $b = b_V \iff b_{V^\perp} = 0$.

projection onto V doesn't move the vectors in V.

(2) says:

"0 is the closest vector in V to b"

$$\Updownarrow$$

[demo]

"b is orthogonal to V"

Or: since $b = b_V + b_{V^\perp}$, $b_V = 0 \iff b = b_{V^\perp}$.

Of course (1) $\iff$ (2) by switching $V \longleftrightarrow V^\perp$.

(3) says

"projecting twice is the same as projecting once"

This follows from (1) because $b_V \in V$.

Eg: last time: if $b = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ $V = \mathrm{Col}\begin{pmatrix} 1 & -1 & -1 \\ 2 & 1 & 4 \\ 1 & -1 & -1 \end{pmatrix}$

then we computed $b_V = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, so we should have $b \in V$. Let's check:

$$\left(\begin{array}{ccc|c} 1 & -1 & -1 & 1 \\ 2 & 1 & 4 & 1 \\ 1 & -1 & -1 & 1 \end{array}\right) \xrightarrow{\text{PVF}} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2/3 \\ -1/3 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix}$$

Taking $x_3 = 0$ gives a solution of the vector eq^n:

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \frac{2}{3}\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} - \frac{1}{3}\begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix}$$

So b is indeed in $V = \mathrm{Col}\begin{pmatrix} 1 & -1 & -1 \\ 2 & 1 & 4 \\ 1 & -1 & -1 \end{pmatrix}$.

Projection Matrices

Recall: If $V = Col(A)$ then you 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 $\hat{x}$.

Lemma: A has full column rank if & only if A^TA is invertible.

Proof: Note A^TA is square.

A has FCR

$\iff Nul(A) = \{0\}$ (FCR criteria)

$\iff Nul(A^TA) = \{0\}$ $(Nul(A) = Nul(A^TA))$

$\iff A^TA$ has FCR (FCR criteria)

$\iff A^TA$ is invertible (invertibility criteria) //

In this case, $A^TA\hat{x} = A^Tb$ has the unique solution $\hat{x} = (A^TA)^{-1}A^Tb$, so $b_V = A\hat{x} = A(A^TA)^{-1}A^Tb$.

$$\boxed{\text{If } A \text{ has FCR and } V=\mathrm{Col}(A) \text{ then } b_V = A(A^TA)^{-1}A^Tb. \leftarrow \text{"Horrible Formula"}}$$

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

$$A^TA=\begin{pmatrix}1&1&1\\1&1&0\end{pmatrix}\begin{pmatrix}1&1\\1&1\\1&0\end{pmatrix}=\begin{pmatrix}3&2\\2&2\end{pmatrix}$$

$$(A^TA)^{-1}=\frac{1}{6-4}\begin{pmatrix}2&-2\\-2&3\end{pmatrix}=\begin{pmatrix}1&-1\\-1&3/2\end{pmatrix}$$

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

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

So if $b=\begin{pmatrix}1\\0\\0\end{pmatrix}$ then

$$b_V=\begin{pmatrix}1/2&1/2&0\\1/2&1/2&0\\0&0&1\end{pmatrix}\begin{pmatrix}1\\0\\0\end{pmatrix}=\begin{pmatrix}1/2\\1/2\\0\end{pmatrix}$$ (cf. L11) ✓

Observation: $P_V=\overset{\text{FCR}}{A}(A^TA)^{-1}A^T$ is an $m\times m$ matrix that computes orthogonal projections onto $V=\mathrm{Col}(A)$: $P_Vb=b_V$ for all $b\in\mathbb{R}^m$.

Def: Let V be a subspace of $\mathbb{R}^m$. The projection matrix onto V is the $m \times m$ matrix P_V such that $P_V b = b_V$ for all $b \in \mathbb{R}^m$.

NB: The matrix P_V is defined by the equality

$$P_V b = b_V$$

for all vectors b. This uniquely characterizes P_V by the Fact below. Use the above equation to answer questions about P_V!

(This is the first time we've defined a matrix by its action on $\mathbb{R}^m$.)

Fact: If A & B are $m \times n$ matrices and $Ax = Bx$ for all x, then $A = B$.

Indeed, $Ae_i = i^{th}$ col of A, so actually a matrix is determined by its action on the unit coordinate vectors.

What if $V = \mathrm{Col}(A)$ but A does not have full column rank? How to compute P_V?

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

This A does not have full column rank:

$$A \xrightarrow{\text{ref}} \begin{pmatrix} 1 & -1 & -1 \\ 0 & 3 & 6 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{pivots}$$

This says that $\left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix} \right\}$ is a basis for V. This means:

(1) $V = \mathrm{Span}\left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix} \right\} = \mathrm{Col}\begin{pmatrix} 1 & -1 \\ 2 & 1 \\ 1 & -1 \end{pmatrix}$

(2) $\left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix} \right\}$ is LI

$\rightsquigarrow \begin{pmatrix} 1 & -1 \\ 2 & 1 \\ 1 & -1 \end{pmatrix}$ has full column rank.

So replace A by $B = \begin{pmatrix} 1 & -1 \\ 2 & 1 \\ 1 & -1 \end{pmatrix}$:

$$B^T B = \begin{pmatrix} 1 & 2 & 1 \\ -1 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 2 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 6 & 0 \\ 0 & 3 \end{pmatrix}$$

$$(B^T B)^{-1} = \begin{pmatrix} 1/6 & 0 \\ 0 & 1/3 \end{pmatrix} = \frac{1}{6} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

$$P_V = B(B^TB)^{-1}B^T = \frac{1}{6}\begin{pmatrix} 1 & -1 \\ 2 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} 1 & 2 & 1 \\ -1 & 1 & -1 \end{pmatrix}$$

$$= \frac{1}{6}\begin{pmatrix} 1 & -2 \\ 2 & 2 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} 1 & 2 & 1 \\ -1 & 1 & -1 \end{pmatrix} = \frac{1}{6}\begin{pmatrix} 3 & 0 & 3 \\ 0 & 6 & 0 \\ 3 & 0 & 3 \end{pmatrix}$$

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

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

(cf. L11) ✓

NB: What if A is a 3×3 matrix with FCR?
Then A has FRR too $\Rightarrow V = \mathrm{Col}(A) = \mathbb{R}^3$.
In this case $b_V = b$ for any b (because $b \in V$)
So
$$P_V b = b_V = b \quad \text{for all } b.$$
The only matrix that fixes every vector is the identity matrix:
$$P_V = I_3$$
More on this later.

Procedure for Computing P_V:

(1) Find a basis $\{v_1, \dots, v_n\}$ of V

(2) $B = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix}$

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

for example, if $V = \mathrm{Col}(A)$ then use the pivot columns

Eg: Suppose $V = \mathrm{Span}\{v\}$ is a line.

$B = v$ (matrix with one column)

$B^TB = v \cdot v$ (a scalar)

$B(B^TB)^{-1}B^T = v(v\cdot v)^{-1}v^T = \dfrac{vv^T}{v\cdot v}$

outer product

Projection Matrix onto a Line

If $V = \mathrm{Span}\{v\}$ then $P_V = \dfrac{vv^T}{v\cdot v}$

Eg: $V = \mathrm{Span}\left\{\begin{pmatrix}1\\1\end{pmatrix}\right\}$

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

So if $b = \begin{pmatrix}1\\0\end{pmatrix}$ then $b_V = P_V b = \begin{pmatrix}1/2\\1/2\end{pmatrix}$ (cf L11)

Properties of Projection Matrices:

Let V be a subspace of $\mathbb{R}^m$ and let P_V be its projection matrix.

$$\boxed{\begin{array}{ll} (1)\ \mathrm{Col}(P_V) = V & (3)\ P_V^2 = P_V \\ (2)\ \mathrm{Nul}(P_V) = V^\perp & (4)\ P_V + P_{V^\perp} = I_m \\ \multicolumn{2}{c}{(5)\ P_V = P_V^T} \\ (6)\ P_{\mathbb{R}^n} = I_n & (7)\ P_{\{0\}} = 0 \end{array}}$$

Recall: A (square) matrix S is symmetric if $S = S^T$.

Proofs of the Properties:

This is a translation of properties of projections.

(1) $\mathrm{Col}(P_V) = \{P_V b : b \in \mathbb{R}^m\} = \{b_V : b \in \mathbb{R}^m\}$

This equals V

- because $b_V \in V$ for any b,
- and $b_V = b$ for any $b \in V$.

(2) $\mathrm{Nul}(P_V) = \{b \in \mathbb{R}^m : P_V b = 0\} = \{b \in \mathbb{R}^m : b_V = 0\}$

But we know $b_V = 0 \iff b \in V^\perp$.

(3) For any vector b,

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

This equals b_V because $b_V \in V$ already

$$= b_V = P_V b$$

Since $P_V^2 b = P_V b$ for all vectors b, $P_V^2 = P_V$.

(4) For any vector b,

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

This equals b because $b = b_V + b_{V^\perp}$ is the orthogonal decomposition.

$$= b = I_m b$$

Since $(P_V + P_{V^\perp})b = I_m b$ for all vectors b, $P_V + P_{V^\perp} = I_m$.

(5) Choose a basis for $V \rightsquigarrow P_V = B(B^T B)^{-1} B^T$.

$$P_V^T = \left(B(B^T B)^{-1} B^T\right)^T = B^{TT}\left((B^T B)^{-1}\right)^T B^T$$

$$= B\left((B^T B)^T\right)^{-1} B^T = B(B^T B)^{-1} B^T = P_V$$

- For any invertible matrix A,

$$(A^{-1})^T = (A^T)^{-1} \quad \text{because}$$

$$(A^{-1})^T A^T = (AA^{-1})^T = I_n^T = I_n$$

(6) If $V = \mathbb{R}^n$ then $b \in V$ for all b, so

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

Also $I_n b = b$ for all b, so $P_V = I_n$.

(7) If $V = \{0\}$ then $P_V b$ must be 0 for every b, because 0 is the only vector in V:

$$P_V b = b_V = 0 \quad \text{for all } b.$$

Also $0 b = 0$ for all b, so $P_V = 0$. ✓

Last time: if $V = \text{Nul}(A)$, we computed b_V by first computing the projection onto $V^\perp = \text{Col}(A^T)$, then using $b_V = b - b_{V^\perp}$.

We can do the same for projection matrices, using (5):

Procedure: To compute P_V for $V = \mathrm{Nul}(A)$:

(1) Compute $P_{V^\perp}$ for $V^\perp = \mathrm{Col}(A^T)$

(2) $P_V = I_m - P_{V^\perp}$

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

In this case, $V^\perp = \mathrm{Col}\begin{pmatrix}1\\2\\1\end{pmatrix}$ is a line:

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

$$P_V = \begin{pmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{pmatrix} - \frac{1}{6}\begin{pmatrix}1 & 2 & 1\\2 & 4 & 2\\1 & 2 & 1\end{pmatrix} = \frac{1}{6}\begin{pmatrix}5 & -2 & -1\\-2 & 2 & -2\\-1 & -2 & 5\end{pmatrix}$$

This was much easier than finding a basis for V using PVF, then using $P_V = B(B^TB)^{-1}B^T$:

$$\begin{aligned} x_1 &= -2x_2 - x_3\\ x_2 &= x_2\\ x_3 &= x_3 \end{aligned} \quad\rightsquigarrow\quad V = \mathrm{Span}\left\{\begin{pmatrix}-2\\1\\0\end{pmatrix}, \begin{pmatrix}-1\\0\\1\end{pmatrix}\right\}$$

$$B = \begin{pmatrix}-2 & -1\\1 & 0\\0 & 1\end{pmatrix} \rightsquigarrow B^TB = \begin{pmatrix}5 & 2\\2 & 2\end{pmatrix}$$

$$\rightsquigarrow (B^TB)^{-1} = \frac{1}{10-4}\begin{pmatrix}2 & -2\\-2 & 5\end{pmatrix} = \frac{1}{6}\begin{pmatrix}2 & -2\\-2 & 5\end{pmatrix}$$

$$\rightsquigarrow B(B^TB)^{-1}B^T = \frac{1}{6}\begin{pmatrix}-2 & -1\\1 & 0\\0 & 1\end{pmatrix}\begin{pmatrix}2 & -2\\-2 & 5\end{pmatrix}B^T$$

$$= \frac{1}{6}\begin{pmatrix} -2 & -1 \\ 2 & -2 \\ -2 & 5 \end{pmatrix}\begin{pmatrix} -2 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

$$= \frac{1}{6}\begin{pmatrix} 5 & -2 & -1 \\ -2 & 2 & -2 \\ -1 & -2 & 5 \end{pmatrix}$$

→ Be intelligent about what you actually have to compute! Ask yourself: "is it easier to compute P_V or $P_{V^\perp}$?"