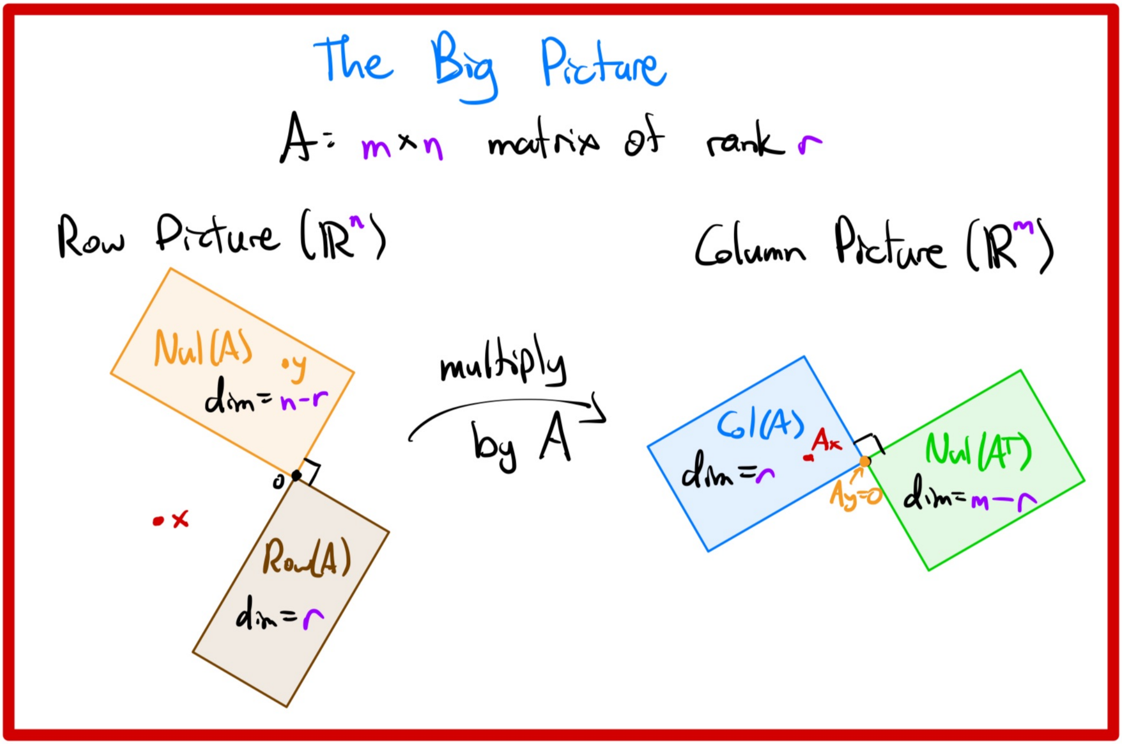
The Big Picture

L10

Last time: we discussed orthogonality of the 4 subspaces
Here is a cartoon summary:



NB: The dimensions match up with $\dim V + \dim V^\perp$:

- $\dim \mathrm{Nul}(A) + \dim \mathrm{Row}(A) = (n-r) + r = n$
- $\dim \mathrm{Col}(A) + \dim \mathrm{Nul}(A^T) = r + (m-r) = m$

The Matrix A^TA

Recall: A^TA is the matrix of column dot products.
Its (i,j) entry is $(\text{col } i)\cdot(\text{col } j)$. It is symmetric.

$$A = \begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix} \rightsquigarrow A^TA = \begin{pmatrix} — v_1^T — \\ — v_2^T — \\ — v_3^T — \end{pmatrix}\begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix} = \begin{pmatrix} v_1\cdot v_1 & v_1\cdot v_2 & v_1\cdot v_3 \\ v_2\cdot v_1 & v_2\cdot v_2 & v_2\cdot v_3 \\ v_3\cdot v_1 & v_3\cdot v_2 & v_3\cdot v_3 \end{pmatrix}$$

equal

$(3,1)$ entry

Important Fact that will keep coming up:

$$\boxed{\text{Nul}(A^TA) = \text{Nul}(A)}$$

We can prove this using orthogonality of the 4 subspaces.

Proof: Showing $\text{Nul}(A^TA) = \text{Nul}(A)$ means showing that every vector in $\text{Nul}(A^TA)$ is also in $\text{Nul}(A)$, & vice-versa.

- $\text{Nul}(A)$ is contained in $\text{Nul}(A^TA)$:

  if $x \in \text{Nul}(A)$ then $Ax = 0 \Rightarrow A^TAx = A^T0 = 0$
  $\Rightarrow x \in \text{Nul}(A^TA)$

  (You showed on the HW that $\text{Nul}(A)$ is in $\text{Nul}(BA)$.)

- $\text{Nul}(A^TA)$ is contained in $\text{Nul}(A)$:

  if $x \in \text{Nul}(A^TA)$ then $A^TAx = 0 \Rightarrow A^T(Ax) = 0$
  $\Rightarrow Ax \in \text{Nul}(A^T) = \text{Col}(A)^\perp$. But $Ax \in \text{Col}(A)$ too, so
  $Ax \in \text{Col}(A)$ & $\text{Col}(A)^\perp \Rightarrow Ax = 0 \Rightarrow x \in \text{Nul}(A)$

//

Orthogonal Projections

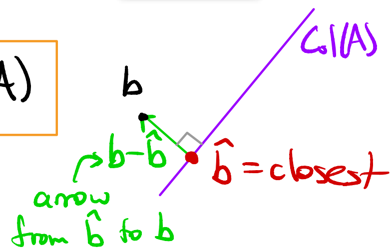
Recall: To find the best approximate solution of $Ax=b$, we need to find the closest vector $\hat{b}$ to b in $\mathrm{Col}(A)=\{Ax : x\in\mathbb{R}^n\}$.

This vector is characterized by the property that

[DEMO]

$$\boxed{b-\hat{b} \text{ is perpendicular to } \mathrm{Col}(A)}$$

or $\quad b-\hat{b}\in \mathrm{Col}(A)^{\perp}$.



Now we give this closest vector a name, with $\mathrm{Col}(A)$ replaced by any subspace.

Def: Let V be a subspace of $\mathbb{R}^n$ and let $b\in\mathbb{R}^n$. The orthogonal projection of b onto V is

$$\boxed{\text{the closest vector to } b \text{ in } V.}$$

Notation: b_V

This vector is characterized by the property

$$\boxed{b-b_V\in V^{\perp}}$$

"the arrow from b_V to b is orthogonal to V"

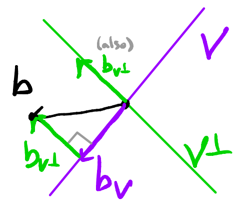
If $b_{V^\perp}=b-b_V$ then $b_{V^\perp}\in V^{\perp}$ and

$$b-b_{V^\perp}=b-(b-b_V)=b_V\in V=(V^{\perp})^{\perp}.$$

It follows that $b_{V^\perp} = b - b_V$ is the orthogonal projection of b onto $V^\perp$.

Def: The **orthogonal decomposition** of b relative to V is

$$\boxed{b = \underbrace{b_V}_{\substack{\text{orthogonal}\\ \text{projection}\\ \text{onto } V}} + \underbrace{b_{V^\perp}}_{\substack{\text{orthogonal}\\ \text{projection}\\ \text{onto } V^\perp}}}$$



[DEMO 1] [DEMO 2] [DEMO 3]

So, like subspaces, orthogonal projections come in complementary pairs. This symmetry is very useful — for instance, sometimes $b_{V^\perp}$ will be easier to compute than b_V!

How to compute b_V?

First you have to express V as a column space or a null space (as always).

NB: The orthogonal projection b_V is the same, no matter how you choose to describe V! Different descriptions have to give you the same answer b_V = closest vector in V. Maybe some descriptions are more convenient!
(Just like when we computed $V^\perp$.)

How to Compute b_V when $V = \mathrm{Col}(A)$:

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

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

Why $\hat{x}$ & not x? Convention.

Then $b_{V^\perp} = b - b_V$ and the orthogonal decomposition is $b = b_V + b_{V^\perp}$.

NB: The distance from b to V is $\|b - b_V\| = \|b_{V^\perp}\|$.

We went through a lot of material to understand why this works (among many other reasons).

Why this works:

If $\hat{x}$ solves $A^TA\hat{x} = A^Tb$ then

$$A\hat{x} \in \mathrm{Col}(A) = V \quad \text{and} \quad A^T(b - A\hat{x}) = A^Tb - A^TA\hat{x} = 0.$$

This means $b - A\hat{x} \in \mathrm{Nul}(A^T) = \mathrm{Col}(A)^\perp = V^\perp$, so $A\hat{x} = b_V$ because $b - b_V \in V^\perp$ characterizes b_V.

Eg: Find b_V for $b = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ and $V = \mathrm{Col}(A)$, $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}$.

We have to solve $A^TA\hat{x} = A^Tb$.

$$A^TA = \begin{pmatrix} \text{column dot products} \end{pmatrix} = \begin{pmatrix} 3 & 2 \\ 2 & 2 \end{pmatrix} \qquad A^Tb = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Solve $\left(\begin{array}{cc|c} 3 & 2 & 1 \\ 2 & 2 & 1 \end{array}\right) \xrightarrow{\text{rref}} \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 1/2 \end{array}\right) \rightsquigarrow \hat{x} = \begin{pmatrix} 0 \\ 1/2 \end{pmatrix}$

$$\Rightarrow b_V = A\hat{x} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 \\ 1/2 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix}$$

[DEMO]

Check: $b_{V\perp} = b - b_V = \begin{pmatrix}1\\0\\0\end{pmatrix} - \begin{pmatrix}1/2\\1/2\\0\end{pmatrix} = \begin{pmatrix}1/2\\-1/2\\0\end{pmatrix}$

Is this $\perp$ the columns of A?

$$\begin{pmatrix}1/2\\-1/2\\0\end{pmatrix}\cdot\underbrace{\begin{pmatrix}1\\1\\1\end{pmatrix}}_{\text{1st col}} = 0 \qquad \begin{pmatrix}1/2\\-1/2\\0\end{pmatrix}\cdot\underbrace{\begin{pmatrix}1\\1\\0\end{pmatrix}}_{\text{2nd col}} = 0 \implies b_{V\perp}\in \mathrm{Col}(A)^\perp \checkmark$$

Orthogonal Decomposition:

$$\underbrace{\begin{pmatrix}1\\0\\0\end{pmatrix}}_{b} = \underbrace{\begin{pmatrix}1/2\\1/2\\0\end{pmatrix}}_{b_V} + \underbrace{\begin{pmatrix}1/2\\-1/2\\0\end{pmatrix}}_{b_{V\perp}}$$

Distance to V: $\|b_{V\perp}\| = \frac{1}{2}\|(1,-1,0)\| = \frac{\sqrt{2}}{2}$

NB: Finding b_V for $V=\mathrm{Col}(A)$ means solving the matrix equation $A^TA\hat{x} = A^Tb$, not $Ax=b$!

Def: $A^TA\hat{x} = A^Tb$ is the **normal equation** of $Ax=b$.

The procedure given above wouldn't work if $A^TA\hat{x}=A^Tb$ were inconsistent, so thankfully:

Fact: For any matrix A & any vector $b\in\mathbb{R}^m$, the normal equation $A^TA\hat{x} = A^Tb$ is consistent.

Proof:

$$\mathrm{Col}(A^T) = \mathrm{Row}(A) = \mathrm{Nul}(A)^\perp \overset{\substack{\text{FACT}\\ \mathrm{Nul}(A)=\mathrm{Nul}(A^TA)}}{=} \mathrm{Nul}(A^TA)^\perp$$
$$= \mathrm{Row}(A^TA) = \mathrm{Col}((A^TA)^T) \underset{A^TA \text{ is symmetric}}{=} \mathrm{Col}(A^TA)$$

Since $A^Tb \in Col(A^T)$ this shows $A^Tb \in Col(A^TA)$, which means $A^TA\hat{x} = A^Tb$ is consistent. //

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

We have to solve $A^TA\hat{x} = A^Tb$.

$$A^TA = \left(\text{column dot products} \right) = \begin{pmatrix} 6 & 0 & 6 \\ 0 & 3 & 6 \\ 6 & 6 & 18 \end{pmatrix} \qquad A^Tb = \begin{pmatrix} 4 \\ -1 \\ 2 \end{pmatrix}$$

Solve: $\left(\begin{array}{ccc|c} 6 & 0 & 6 & 4 \\ 0 & 3 & 6 & -1 \\ 6 & 6 & 18 & 2 \end{array}\right) \xrightarrow{\text{rref}} \left(\begin{array}{ccc|c} 1 & 0 & 1 & 2/3 \\ 0 & 1 & 2 & -1/3 \\ 0 & 0 & 0 & 0 \end{array}\right)$

$$\xrightarrow{\text{PVF}} \hat{x} = \begin{pmatrix} 2/3 \\ -1/3 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix}$$

Which solution do we use?

"$b_V = A\hat{x}$ for any solution $\hat{x}$"

Ok, so let's take $x_3 = 0$

$$\rightsquigarrow \hat{x} = \begin{pmatrix} 2/3 \\ -1/3 \\ 0 \end{pmatrix} \rightsquigarrow b_V = A\hat{x} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

NB: $b_V = b$. What does that mean?

⭐ b is already in V ⭐

More on this later.

In this case, $b_{V^\perp} = b - b_V = 0$.

What happens if we try a different solution $\hat{x}$?

Try $x_3 = 1/3$:

$$\rightsquigarrow \hat{x} = \begin{pmatrix} 1/3 \\ -1 \\ 1/3 \end{pmatrix} \rightsquigarrow b_V = A\hat{x} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \text{ again!}$$

This always works.

Fact: If $\hat{x}_1$ and $\hat{x}_2$ both solve $A^TA\hat{x} = A^Tb$, then

$$A\hat{x}_1 = A\hat{x}_2 \quad (= b_V)$$

Proof: $A^TA\hat{x}_1 = A^Tb = A^TA\hat{x}_2$

$$\Rightarrow 0 = A^TA\hat{x}_1 - A^TA\hat{x}_2 = A^TA(\hat{x}_1 - \hat{x}_2)$$

$$\Rightarrow \hat{x}_1 - \hat{x}_2 \in \mathrm{Nul}(A^TA) \overset{\text{FACT}}{=} \mathrm{Nul}(A)$$

$$\Rightarrow 0 = A(\hat{x}_1 - \hat{x}_2) = A\hat{x}_1 - A\hat{x}_2$$

$$\Rightarrow A\hat{x}_1 = A\hat{x}_2.$$

//

Now we know how to compute b_V for $V = \mathrm{Col}(A)$.
What about when $V = \mathrm{Nul}(A)$?

→ You could first compute a basis for $\mathrm{Nul}(A)$ (PVF)
This expresses V as a column space: $V = \mathrm{Col}(B)$
(cols of B = vectors in PVF)
Then proceed as above: solve $B^TB\hat{x} = B^Tb$.

But then you have to eliminate twice!

Here's a better way that takes advantage of the symmetry in $b = b_V + b_{V^\perp}$.

How to Compute b_V when $V = \mathrm{Nul}(A)$:

(1) Compute $b_{V^\perp}$ = projection onto $V^\perp = \mathrm{Row}(A) = \mathrm{Col}(A^T)$ by solving $AA^T\hat{x} = Ab$ (i.e. $A^{TT}A^T\hat{x} = A^{TT}b$)

(2) $b_V = b - b_{V^\perp}$.

Eg: Compute b_V for $b = \begin{pmatrix} 7 \\ 3 \\ 11 \end{pmatrix}$, $V = \mathrm{Nul}(A)$ $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$

We'll compute $b_{V^\perp}$ = projection onto $\mathrm{Col}(A^T) = \mathrm{Col}\begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$. ← keep track of which is A & which is A^T

We need to solve $AA^T\hat{x} = Ab$.

$$AA^T = \begin{pmatrix} \text{row dot products} \end{pmatrix} = \begin{pmatrix} 14 & 32 \\ 32 & 77 \end{pmatrix} \qquad Ab = \begin{pmatrix} 46 \\ 109 \end{pmatrix}$$

Solve: $\left(\begin{array}{cc|c} 14 & 32 & 46 \\ 32 & 77 & 109 \end{array}\right) \xrightarrow{\text{rref}} \left(\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 1 \end{array}\right) \rightsquigarrow \hat{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

$$\rightsquigarrow b_{V^\perp} = A^T\hat{x} = \begin{pmatrix} 5 \\ 7 \\ 9 \end{pmatrix} \rightsquigarrow b_V = b - b_{V^\perp} = \begin{pmatrix} 2 \\ -4 \\ 2 \end{pmatrix}$$

Check: b_V should be $\perp$ the rows of A:

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ -4 \\ 2 \end{pmatrix} = 0 \qquad \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ -4 \\ 2 \end{pmatrix} = 0 \quad \checkmark$$

We only had to do elimination once!

When $V = \mathrm{Span}\{v\}$ is a line, projection is easier:

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

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

$A^Tb = v^Tb = v \cdot b$

So the normal equation is

$$(v \cdot v)\hat{x} = v \cdot b \quad \overset{\text{solve}}{\rightsquigarrow} \quad \hat{x} = \frac{v \cdot b}{v \cdot v} \quad \text{(a number)}$$

$$\rightsquigarrow \quad b_V = A\hat{x} = \frac{v \cdot b}{v \cdot v}\, v$$

No elimination necessary! Just two dot products.

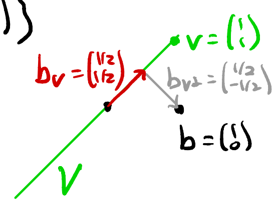
How to Compute b_V when $V = \mathrm{Span}\{v\}$ is a Line:

$$\boxed{b_V = \frac{v \cdot b}{v \cdot v}\, v}$$

Eg: Compute b_V for $b = \begin{pmatrix}1\\0\end{pmatrix}$, $V = \mathrm{Span}\left\{\begin{pmatrix}1\\1\end{pmatrix}\right\}$

$$b_V = \frac{\begin{pmatrix}1\\0\end{pmatrix} \cdot \begin{pmatrix}1\\1\end{pmatrix}}{\begin{pmatrix}1\\1\end{pmatrix} \cdot \begin{pmatrix}1\\1\end{pmatrix}} \begin{pmatrix}1\\1\end{pmatrix} = \frac{1}{2}\begin{pmatrix}1\\1\end{pmatrix}$$

[DEMO]



$$\Rightarrow \quad b_{V^\perp} = b - b_V = \frac{1}{2}\begin{pmatrix}1\\-1\end{pmatrix}$$

Orthogonal decomposition: $\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}1/2\\1/2\end{pmatrix} + \begin{pmatrix}1/2\\-1/2\end{pmatrix}$

Eg: Compute (again) b_V for $b = \begin{pmatrix}7\\3\\11\end{pmatrix}$, $V = \mathrm{Nul}(A)$, $A = \begin{pmatrix}1&2&3\\4&5&6\end{pmatrix}$

Clearly A has rank $2 \Rightarrow 1$ free variable, so $\mathrm{Nul}(A)$ is a line. Let's find a basis for $\mathrm{Nul}(A)$ and use projection onto a line.

$$\begin{pmatrix}1&2&3\\4&5&6\end{pmatrix} \xrightarrow{\text{rref}} \begin{pmatrix}1&0&-1\\0&1&2\end{pmatrix} \xrightarrow{\text{PVF}} \text{basis: } \left\{\begin{pmatrix}1\\-2\\1\end{pmatrix}\right\}$$

OK so $V = \mathrm{Span}\left\{\begin{pmatrix}1\\-2\\1\end{pmatrix}\right\}$ and

$$b_V = \frac{\begin{pmatrix}1\\-2\\1\end{pmatrix}\cdot\begin{pmatrix}7\\3\\11\end{pmatrix}}{\begin{pmatrix}1\\-2\\1\end{pmatrix}\cdot\begin{pmatrix}1\\-2\\1\end{pmatrix}}\begin{pmatrix}1\\-2\\1\end{pmatrix} = \frac{12}{6}\begin{pmatrix}1\\-2\\1\end{pmatrix} = \begin{pmatrix}2\\-4\\2\end{pmatrix} \checkmark$$

[DEMO]

NB: We computed b_V for $b = \begin{pmatrix}7\\3\\11\end{pmatrix}$, $V = \mathrm{Nul}\begin{pmatrix}1&2&3\\4&5&6\end{pmatrix}$ in 2 different ways using 2 different descriptions of V:

- $V = \mathrm{Nul}(A) \rightsquigarrow$ compute $b_{V^\perp} \rightsquigarrow$ solve $AA^T\hat{x} = Ab$
  $\rightsquigarrow b_{V^\perp} = \begin{pmatrix}5\\7\\9\end{pmatrix} \rightsquigarrow b_V = \begin{pmatrix}2\\-4\\2\end{pmatrix}$
- $V = \mathrm{Span}\{v\} \rightsquigarrow b_V = \dfrac{v\cdot b}{v\cdot v}v = \begin{pmatrix}2\\-4\\2\end{pmatrix}$

We got the same answer because the definition

$$b_V = (\text{closest vector to } b \text{ in } V)$$

has nothing to do with how we chose to describe V.

Upshot: You have several ways of computing orthogonal projections. Ask yourself: which is easiest?

- Is it easier to compute $b_{V^\perp}$?
- Should I find a better description of V first?
- Is V or $V^\perp$ a line?

Warning: These are good ways to compute orthogonal projections, but they're terrible ways to understand/reason about them!

Orthogonal projection is a geometric construction. It should be understood geometrically, as in:

$$b_V = (\text{closest vector to } b \text{ in } V)$$

or

$$b - b_V \text{ is orthogonal to } V.$$

the arrow from b_V to b