Projections

(Section 4.2)

Recall Q: Why does the best fit polynomial approach work? What exactly is it producing when it gives you an answer?

Example 1: (best fit line from before)

Given the points $(1,3)$, $(2,-1)$, and $(6,5)$, we found the "best-fit" line through these points.

We wrote $y=mx+b$ for the line and plugged in the points:

$$3 = m+b$$
$$-1 = 2m+b$$
$$5 = 6m+b\,,$$

then we rewrote this as a matrix equation $A\vec{x} = \vec{b}$

where

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 6 \end{bmatrix}$$

$$\vec{x} = \begin{bmatrix} b \\ m \end{bmatrix}$$

$$\vec{b} = \begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix}.$$

If there was a solution $\vec{x}$, then

$$\vec{b} = A\vec{x} \in Ran(A).$$

Let's note that $\vec{b} = \begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix}$ is

not in $Ran(A)$!

$\mathrm{Ran}(A) = \mathrm{Col}(A) =$ span of columns of A

Columns of A: $\begin{bmatrix}1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\2\\6\end{bmatrix}$

and rewriting

$$\vec{b} = A\hat{x},$$

$$\begin{bmatrix}3\\-1\\5\end{bmatrix} = \begin{bmatrix}1 & 1\\1 & 2\\1 & 6\end{bmatrix}\begin{bmatrix}b\\m\end{bmatrix}$$

$$= \begin{bmatrix}1 & 1\\1 & 2\\1 & 6\end{bmatrix}\left(\begin{bmatrix}b\\0\end{bmatrix} + \begin{bmatrix}0\\m\end{bmatrix}\right)$$

$$= \begin{bmatrix}1 & 1\\1 & 2\\1 & 6\end{bmatrix}\left(b\begin{bmatrix}1\\0\end{bmatrix} + m\begin{bmatrix}0\\1\end{bmatrix}\right)$$

(all matrices are linear transformations)

$$= b\left(\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 6 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)$$

$$+ m\left(\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 6 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)$$

$$= b\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + m\begin{bmatrix} 1 \\ 2 \\ 6 \end{bmatrix}$$

So a solution $\begin{bmatrix} b \\ m \end{bmatrix}$ to $A\vec{x} = \vec{b}$

would have to be a linear combination of the columns of A. But obtaining the values b and m is the same as solving

$A\vec{x} = \vec{b}$, which we can

Show in Wolfram Alpha has no solution!

$$\mathrm{RREF}\,[A \,\vdots\, \vec{b}\,]$$

$$= \left[\begin{array}{cc|c} 1 & 1 & 3 \\ 1 & 2 & -1 \\ 1 & 6 & 5 \end{array}\right]$$

$$= \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right]$$

no solution

We "solved" the problem by passing to the equation

$$A^t A \vec{x} = A^t \vec{b},$$

then multiplying by $(A^t A)^{-1}$ to get

$$\vec{x} = (A^t A)^{-1} A^t \vec{b}$$

Observe that, multiplying both sides by A on the left,

$$\underbrace{A\vec{x} = A(A^tA)^{-1}A^t\vec{b}}_{\text{automatically in } \mathrm{Ran}(A)!}$$

If we set

$$\vec{b}_1 = A(A^tA)^{-1}A^t\vec{b},$$

what this says is that

$A\vec{x} = \vec{b}_1$ always has solutions, unlike $A\vec{x} = \vec{b}$!

What we're really doing is replacing $\vec{b}$ with $\vec{b}_1$.

Goal: understand this geometrically.

Observation: $\left(A(A^tA)^{-1}A^t \right)$

Set $P = A(A^tA)^{-1}A^t$.

Then $P \cdot P = P$ (HW 7).

Projections

$P \in M_n(\mathbb{R})$ is called a **projection** if $P \cdot P = P$. This is usually shortened to $P^2 = P$.

Example 2: (projections onto $y=x$)

Let $P = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$. Show

$P^2 = P$.

$$P^2 = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \cdot \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}\begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

$$= \begin{bmatrix} 1/4 + 1/4 & 1/4 + 1/4 \\ 1/4 + 1/4 & 1/4 + 1/4 \end{bmatrix}$$

$$= \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} = P \checkmark$$

If you apply P to a vector

$$\vec{v} = \begin{bmatrix} 7 \\ 3 \end{bmatrix}, \quad \text{then}$$

$$P \begin{bmatrix} 7 \\ 3 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 7 \\ 3 \end{bmatrix}$$

$$= \frac{1}{2} \begin{bmatrix} 10 \\ 10 \end{bmatrix}$$

$$= \begin{bmatrix} 5 \\ 5 \end{bmatrix}$$

x and y coordinates equal

If $\vec{w} = \begin{bmatrix} 8 \\ -2 \end{bmatrix}$,

$$P\begin{bmatrix} 8 \\ -2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 8 \\ -2 \end{bmatrix}$$

$$= \frac{1}{2}\begin{bmatrix} 6 \\ 6 \end{bmatrix}$$

$$= \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

same x and y coordinates

In general, $P\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$

$$= \frac{1}{2}\begin{bmatrix} x+y \\ x+y \end{bmatrix}$$

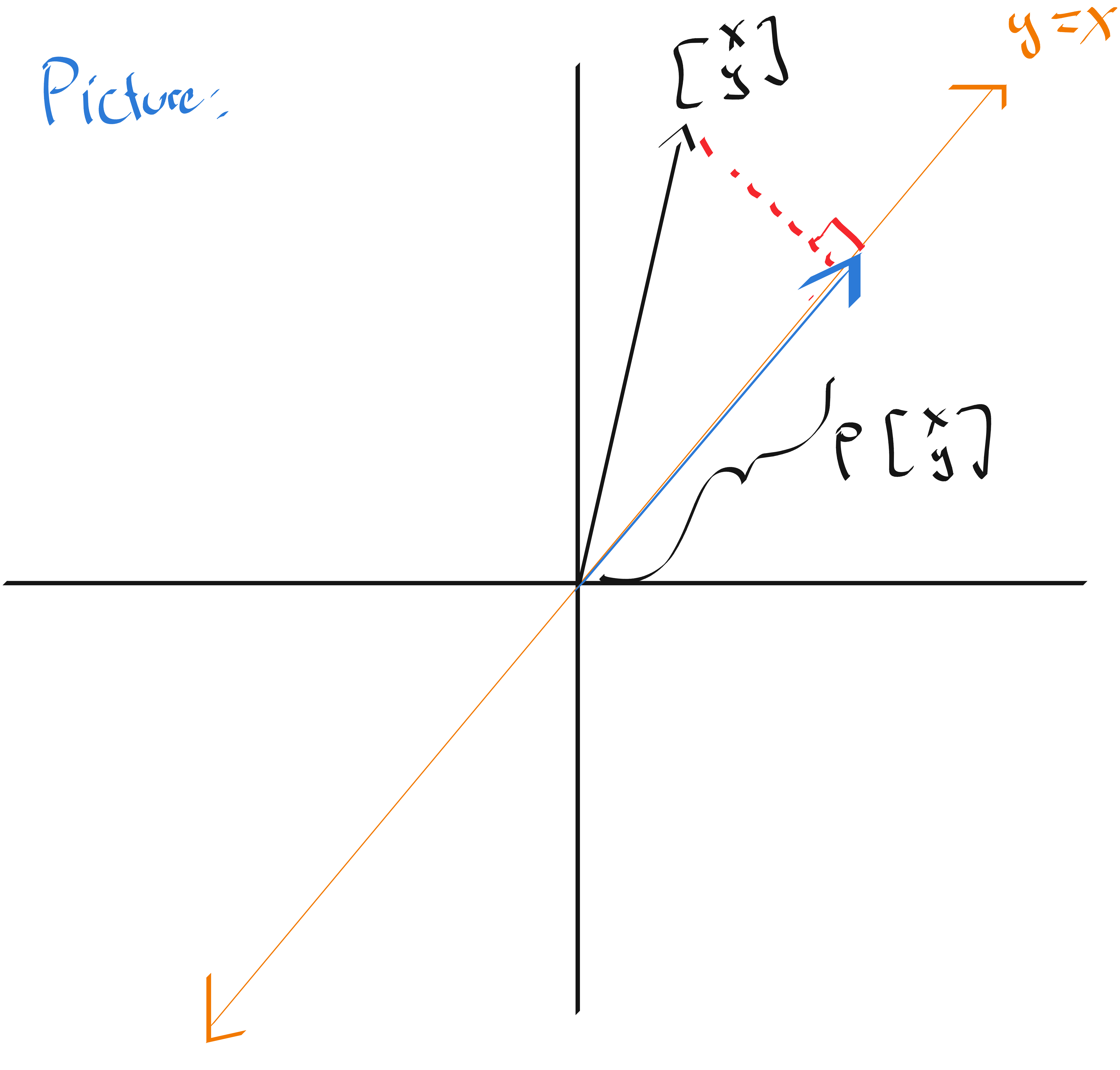
$$= \begin{bmatrix} \frac{x+y}{2} \\ \frac{x+y}{2} \end{bmatrix}$$

same x & y coordinates

$$\begin{bmatrix} \frac{x+y}{2} \\ \frac{x+y}{2} \end{bmatrix} = \frac{x+y}{2} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

So $\operatorname{Ran}(P) = \{(x,y) \mid y = x\}$,

Picture:



But if $Q = \frac{1}{3}\begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$,

$$Q^2 = \frac{1}{3}\begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} \cdot \frac{1}{3}\begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$$

$$= \frac{1}{9}\begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$$

$$= \frac{1}{9}\begin{bmatrix} 3 & 6 \\ 3 & 6 \end{bmatrix}$$

$$= \frac{1}{3}\begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} = Q$$

So Q is also a projection!

$$Q\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$$

$$= \frac{1}{3}\begin{bmatrix} x+2y \\ x+2y \end{bmatrix}$$

$$= \frac{x+2y}{3}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

This matrix also has range the line $y=x$.

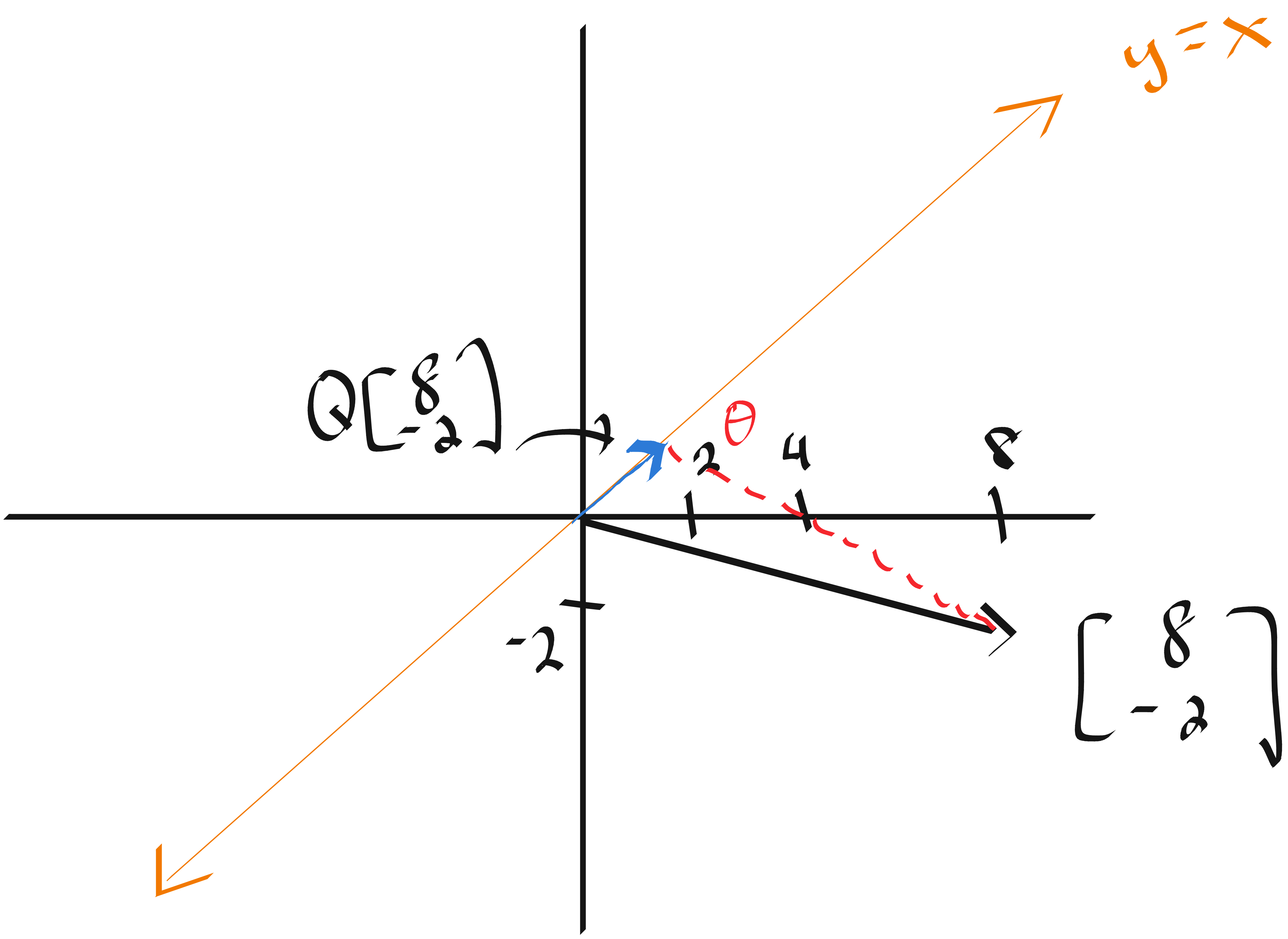
But

$$Q\begin{bmatrix} 8 \\ -2 \end{bmatrix} = \frac{8+2(-2)}{3}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= \frac{4}{3}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\neq \begin{bmatrix} 3 \\ 3 \end{bmatrix} = P\begin{bmatrix} 8 \\ -2 \end{bmatrix}$$

Picture



θ is no longer $90°$!

By varying this angle, we can create infinitely many projections with range $y = x$. How do we pick out the $90°$ one?

If $c > 0$, then the matrix

$$\frac{1}{1+c}\begin{bmatrix} 1 & c \\ 1 & c \end{bmatrix}$$

is a projection with range $y = x$.

Orthogonal Projections

A projection $P \in M_n(\mathbb{R})$ is called an orthogonal projection if

$$\boxed{\begin{aligned} P^2 &= P \\ P &= P^t \end{aligned}}$$

Observe that, for the previous example,

$$\left(\frac{1}{1+c}\begin{bmatrix} 1 & c \\ 1 & c \end{bmatrix}\right)^t = \frac{1}{1+c}\begin{bmatrix} 1 & 1 \\ c & c \end{bmatrix}$$

if this equals $\frac{1}{1+c}\begin{bmatrix} 1 & c \\ 1 & c \end{bmatrix}$,

then $c = 1$.

Back to best-fit polynomials:

We noted that if

$$P = A(A^tA)^{-1}A^t,$$

then P is a projection.

In fact, P is an orthogonal projection! (HW 7).

What we are doing when we calculate best-fit polynomials is replacing $\vec{b}$ in $A\vec{x} = \vec{b}$ with the orthogonal projection of $\vec{b}$ onto $\mathrm{Ran}(A)$ and solving that equation.

Geometric Interpretation

Let $A \in M_n(\mathbb{R})$. Then $\mathrm{Ran}(A)$ is a subspace of $\mathbb{R}^n$. Of all the projections Q whose range is equal to $\mathrm{Ran}(A)$, the orthogonal projection P minimizes the distance between $x \in \mathbb{R}^n$ and $Px \in \mathrm{Ran}(A) \subseteq \mathbb{R}^n$.

That is, if $y \in \mathrm{Ran}(A)$,

$$\boxed{\|x - y\|_2 \geq \|x - Px\|_2}$$

Why? Take $y \in \mathrm{Ran}(A)$. Then since $\mathrm{Ran}(P) = \mathrm{Ran}(A)$, we know that $Py = y$, as follows:

Since $y \in \mathrm{Ran}(P)$, $y = Px$ for some $x \in \mathbb{R}^n$. Then $y = Px$, so $P \cdot y = P \cdot Px = Px = y$ since $P \cdot P = P$.

In order to minimize $\|x-y\|_2^2$, write

$$\|x-y\|_2^2 = \|x - \underbrace{Px + Px}_{= \vec{0}} - y\|_2^2$$

Then thinking of x, y as column vectors,

$$\|x - Px + Px - y\|_2^2 =$$

$$(x - Px + Px - y)^t (x - px + Px - y)$$

$$= \left((x-Px)^t + (Px-y)^t\right)(x - px + Px - y)$$

$$= (x-Px)^t(x-Px) + (Px-y)^t(x-px)$$

$$+ (x-Px)^t(Px-y) + (Px-y)^t(Px-y)$$

$$= \|x - Px\|_2^2 + (Px-y)^t(x-Px)$$

examine, show it is zero

$$+ (x-Px)^t(Px-y) + \|Px-y\|_2^2$$

But $y = Py$, so

$$(Px - y)^t (x - Px)$$

$$= (Px - Py)^t (x - Px)$$

$$= (P(x-y))^t ((I_n - P)x)$$

$$= (x-y)^t P^t (I_n - P) \cdot x$$

$(P = P^t)$ $$= (x-y)^t P \cdot (I_n - P) \cdot x$$

$$= (x-y)^t (P - P^2) \cdot x$$

$(P^2 = P)$ $$= (x-y)^t (\underbrace{P - P}_{\text{zero matrix}}) \cdot x$$

$$= 0$$

Then $(x-Px)^t(Px-y)$

$$= \left((Px-y)^t (x-Px) \right)^t$$

$$= 0^t = 0, \quad \text{so}$$

$$\|x-y\|_2^2 = \|x-Px\|_2^2 + \overset{0}{\cancel{(Px-y)^t(x-Px)}}$$
$$+ \overset{0}{\cancel{(x-Px)^t(Px-y)}} + \|Px-y\|_2^2 .$$

We have that

$$\|x-y\|_2^2 = \underbrace{\|x-Px\|_2^2}_{\text{fixed}} + \underbrace{\|Px-y\|_2^2}_{\text{varies with } y} .$$

Minimizing $\|x-y\|_2^2$ is now the same as minimizing $\|Px-y\|_2^2$.

But this can never be less than zero, and is equal to zero only when $Px-y=\vec{0}$, so $Px=y$ gives the minimum!

Then

$$\|x-y\|_2^2 \geq \|Px-x\|_2^2+\underbrace{\|Px-\tilde{P}x\|_2^2}_{0}$$

Choosing $y=Px$

so $\|x-y\|_2^2 \geq \|Px-x\|_2^2$, and taking square roots,

$$\|x-y\|_2 \geq \|Px-x\|_2=\|x-Px\|_2 \checkmark$$

Observation: (orthogonality of Px and $x-Px$)

If $x \in \mathbb{R}^n$ and P is an orthogonal projection, then regarding x as a column vector,

$$(Px)^t \cdot (x-Px) = x^t P^t (x-Px)$$

$$(P=P^t) \quad = x^t P (x-Px)$$

$$= x^t P (I_n - P) x$$

$$= x^t (P-P^2) x$$

$$(P=P^2) \quad = x^t (\underbrace{P-P}_{0_{n\times 1}}) x$$

$$= 0$$

In the picture for Example 2, the red dashed line is a translated copy of $\vec{x} - P\vec{x}$. This last fact tells you that the angle is always $90°$ between this translate and Px, since if $\vec{x}, \vec{y} \in \mathbb{R}^n$, then regarding x and y as column vectors,

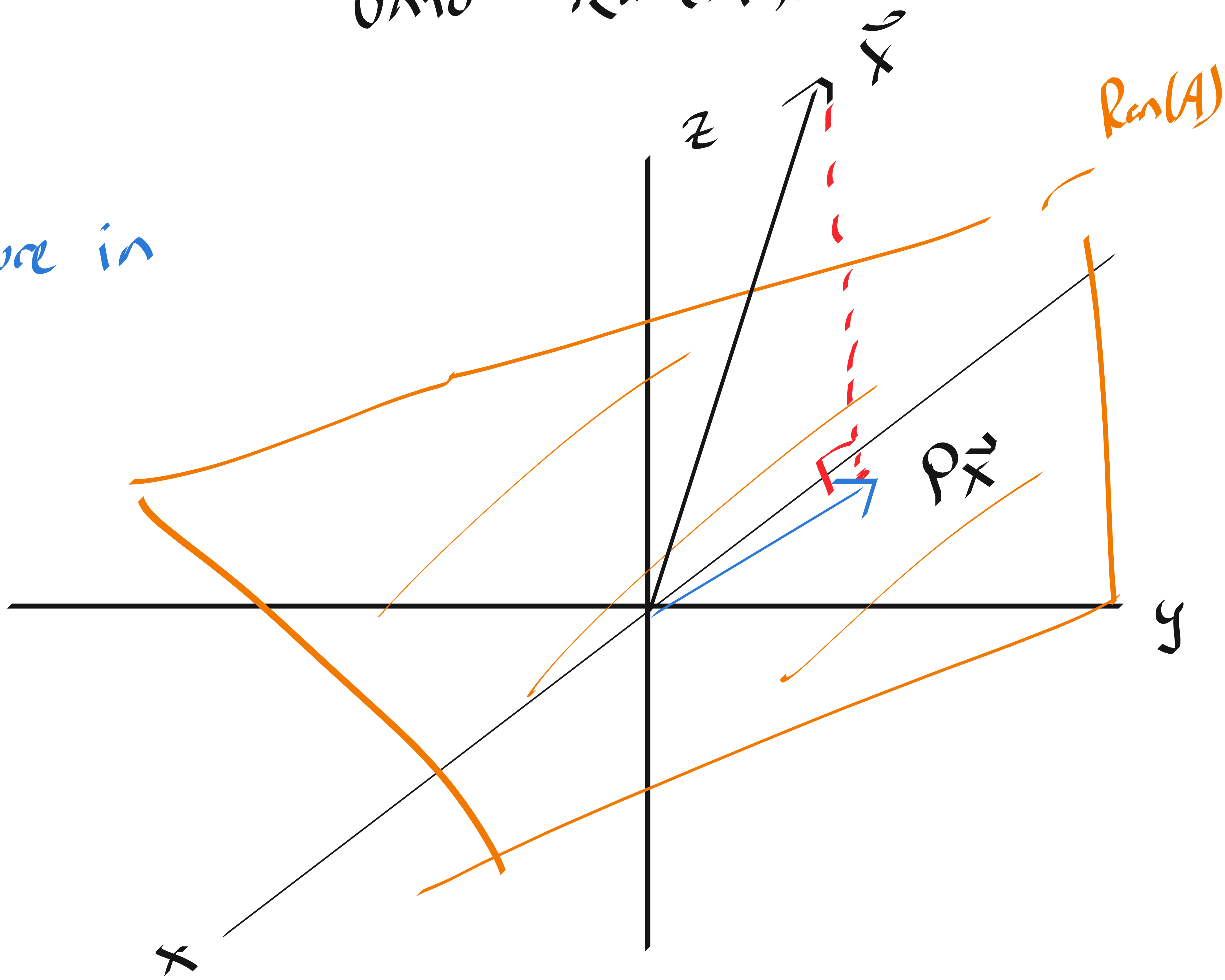
$$0 = \vec{y}^t \vec{x} = \|\vec{x}\|_2 \|\vec{y}\|_2 \cos(\theta)$$

If $\vec{x} \neq \vec{0} \neq \vec{y}$, then $\cos(\theta) = 0$

$$\Rightarrow \theta = 90°.$$

Conclusion: What we do with best-fit polynomials is replace $\vec{b}$ with the vector in $\text{Ran}(A)$ that is closest to $\vec{b}$, in Euclidean length. This occurs via a 90° projection onto $\text{Ran}(A)$.

Picture in $\mathbb{R}^3$



Q: If $A^t A$ is not invertible...

then how do we solve the

best-fit polynomial question?