Thm Let A be an $n\times n$ matrix.

TFAE: ① The columns of A form an orthonormal basis of $\mathbb{R}^n$.

② $A^TA = I_n = AA^T$, i.e. A is invertible & $A^{-1} = A^T$.

③ $Ax \cdot Ay = x \cdot y \quad \forall x,y \in \mathbb{R}^n$.

(This means that A preserves angles & lengths.)

Proof of ②⇒③: First note that, for $u,v \in \mathbb{R}^n$,

$$u\cdot v = u_1v_1 + \dots + u_nv_n = v^Tu = [v_1 \cdots v_n]\begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}.$$

Spse that $A^TA = I_n$ and let $x,y \in \mathbb{R}^n$.

Then:

$$Ax \cdot Ay = (Ay)^T Ax \underset{(AB)^T = B^TA^T}{=} y^T \overbrace{A^TA}^{I_n} x = y^T I_n x = y^Tx = x\cdot y.$$

Defn An $n\times n$ matrix A is orthogonal if $A^TA = I_n (= AA^T)$.

Ex • "Reflections", e.g. $\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

• Rotations $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ (any $\theta \in \mathbb{R}$),

e.g. $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ or $\begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$. (Recall: $\cos\frac{\pi}{4} = \sin\frac{\pi}{4} = \frac{1}{\sqrt{2}}$.)

Problem (from 2018/19 exam paper)

- Give an example of an orthogonal matrix with first row $\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}$.
- Find the orthogonal projection of $\overbrace{\begin{bmatrix} 3 \\ 7 \end{bmatrix}}^{v}$ onto the line passing through $\underbrace{\begin{bmatrix} 1 \\ -1 \end{bmatrix}}_{u}$ and the origin in $\mathbb{R}^2$?

Soln: $$\hat{v} = \frac{v \cdot u}{u \cdot u} u = \frac{-4}{2} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \end{bmatrix}.$$

Check: $$v - \hat{v} = \begin{bmatrix} 3 \\ 7 \end{bmatrix} - \begin{bmatrix} -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 5 \end{bmatrix} \perp \begin{bmatrix} 1 \\ -1 \end{bmatrix} = u.$$

Why should we care about orthogonal projections?

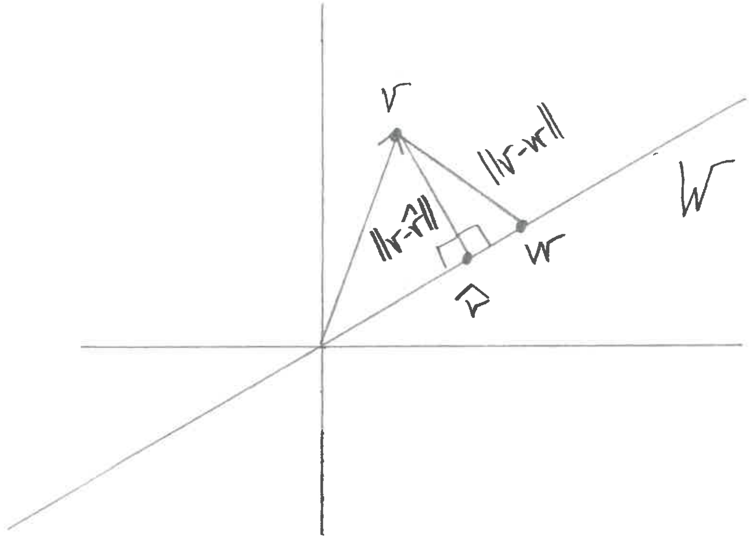
Best Approximation Theorem: Let W be a subspace of $\mathbb{R}^n$.
Let $v \mapsto \hat{v}$ be the orthogonal projection onto W.
Then for each $v \in \mathbb{R}^n$ and $w \in W$,

$$\| v - \hat{v} \| \leq \| v - w \|$$

with equality only for $w = \hat{v}$.

Hence: $\boxed{\hat{v} \text{ is the unique vector in } W \text{ which minimizes the distance from } v.}$

Picture:



Proof: $\|v-w\|^2 = \|\underbrace{(v-\hat{v})}_{\in W^\perp} + \underbrace{(\hat{v}-w)}_{\in W}\|^2 \underset{\text{Pythagoras}}{=} \|v-\hat{v}\|^2 + \|\hat{v}-w\|^2$

$$\geq \|v-\hat{v}\|^2 \text{ with equality iff } \hat{v}=w. \;//$$

We will now apply the Best Approximation Thm to the problem of "data fitting".

Least-squares problems

Suppose that some mathematical model of a phenomenon predicts that a function f of $\begin{bmatrix} x \\ y \end{bmatrix}$ is linear of the form

$$f: \begin{bmatrix} x \\ y \end{bmatrix} \mapsto ax+by \in \mathbb{R}$$

with unknown coefficients $a,b \in \mathbb{R}$. (Can you think of a real-world example of such a function?)