Orthogonality

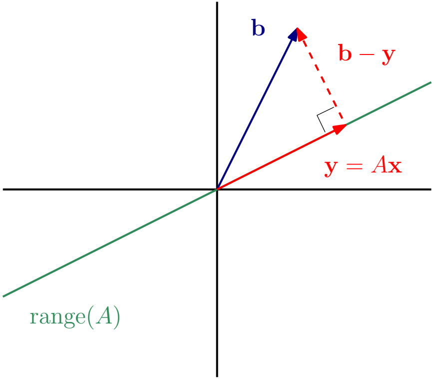
Orthogonality

Normal Equation Revisited

Alternate perspective on the "normal equation":

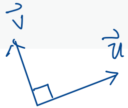
$$A^{\mathrm{T}}(\mathbf{b} - A\mathbf{x}) = \mathbf{0} \quad \Longleftrightarrow \quad \mathbf{z}^{\mathrm{T}}(\underbrace{\mathbf{b} - A\mathbf{x}}_{\text{residual } = \mathbf{r}}) = 0 \quad \text{for all } \mathbf{z} \in \mathcal{R}(A)\,,$$

i.e., $\mathbf{x}$ solves the normal equation if and only if the residual is orthogonal to the range of A.



Orthogonal Vectors

If $\theta = \pm \pi/2$,



$\cos(\theta) = 0$, so

$\vec{u}^T \vec{v} = 0$

Recall that the angle θ between two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ satisfies

$\vec{u} \cdot \vec{v} = \|\vec{u}\| \, \|\vec{v}\| \cos\theta$

$\|$

$\vec{u}^T \vec{v}$

$$\cos(\theta) = \frac{\mathbf{u}^{\mathrm{T}}\mathbf{v}}{\|\mathbf{u}\|_2 \, \|\mathbf{v}\|_2}.$$

Definition 1

- Two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are **orthogonal** if $\mathbf{u}^{\mathrm{T}}\mathbf{v} = 0$.
- Vectors $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_k \in \mathbb{R}^n$ are **orthogonal** if $\mathbf{q}_i^{\mathrm{T}}\mathbf{q}_j = 0$ for all $i \neq j$. i.e., pairwise orthogonal
- Vectors $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_k \in \mathbb{R}^n$ are **orthonormal** if $\mathbf{q}_i^{\mathrm{T}}\mathbf{q}_j = \delta_{i,j} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$

orthogonal → perpendicularity; normal → unit length

Notation. (Kronecker delta function)

$$\delta_{i,j} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

$\vec{q}_i^T \vec{q}_i = 1 = \|\vec{q}_i\|_2^2 \Rightarrow \|\vec{q}_i\|_2 = 1$

i.e., $\vec{q}_i$ is unit.

Matrices with Orthogonal Columns

Note: (i,j)-entry of Q^TQ $= \vec{q_i}^T \vec{q_j}$

Let $Q = \left[\, \mathbf{q}_1 \mid \mathbf{q}_2 \mid \cdots \mid \mathbf{q}_k \,\right] \in \mathbb{R}^{n\times k}$. Note that

$$Q^TQ = \begin{bmatrix} \mathbf{q}_1^T \\ \hline \mathbf{q}_2^T \\ \hline \vdots \\ \hline \mathbf{q}_k^T \end{bmatrix} \left[\begin{array}{c|c|c|c} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_k \end{array}\right] = \begin{bmatrix} \mathbf{q}_1^T\mathbf{q}_1 & \mathbf{q}_1^T\mathbf{q}_2 & \cdots & \mathbf{q}_1^T\mathbf{q}_k \\ \mathbf{q}_2^T\mathbf{q}_1 & \mathbf{q}_2^T\mathbf{q}_2 & \cdots & \mathbf{q}_2^T\mathbf{q}_k \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{q}_k^T\mathbf{q}_1 & \mathbf{q}_k^T\mathbf{q}_2 & \cdots & \mathbf{q}_k^T\mathbf{q}_k \end{bmatrix}.$$

$\vec{q_i}^T \vec{q_j} = 0$ $\Longleftrightarrow$ all off-diagonal terms are 0's.
$i \neq j$

Therefore,

- $\mathbf{q}_1, \ldots, \mathbf{q}_k$ are orthogonal. $\Longleftrightarrow$ Q^TQ is a $k \times k$ diagonal matrix.
- $\mathbf{q}_1, \ldots, \mathbf{q}_k$ are orthonormal. $\Longleftrightarrow$ Q^TQ is the $k \times k$ identity matrix.

$\quad = I$

because $\vec{q_i}^T \vec{q_i} = 1$.

Matrices with Orthonormal Columns

$$\vec{q}_i^{\,T} \vec{q}_j = \delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

Theorem 2

Let $Q = [\, \mathbf{q}_1 \mid \mathbf{q}_2 \mid \cdots \mid \mathbf{q}_k \,] \in \mathbb{R}^{n \times k}$ *and suppose that* $\mathbf{q}_1, \ldots, \mathbf{q}_k$ *are orthonormal. Then*

1. $Q^{\mathrm{T}} Q = I \in \mathbb{R}^{k \times k}$;
2. $\|Q\mathbf{x}\|_2 = \|\mathbf{x}\|_2$ *for all* $\mathbf{x} \in \mathbb{R}^k$; (2-norm preservation)
3. $\|Q\|_2 = 1$. (exercise ; recall the def'n of matrix p-norm & use 2)

Proof of 2

$$\|Q\vec{x}\|_2^2 = (Q\vec{x})^T (Q\vec{x}) = \vec{x}^T \underbrace{Q^T Q}_{\substack{= I \\ \text{by } 1}} \vec{x} = \vec{x}^T \vec{x} = \|\vec{x}\|_2^2$$

Definition 3

We say that $Q \in \mathbb{R}^{n\times n}$ is an **orthogonal matrix** if $Q^{\mathrm{T}} Q = I \in \mathbb{R}^{n\times n}$.

square matrix

The columns of Q are orthonormal.

- A square matrix with orthogonal columns is not, in general, an orthogonal matrix!

e.g.

- $I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

- $R_\theta = \begin{array}{cc} \vec{r}_1 & \vec{r}_2 \\ \downarrow & \downarrow \end{array} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

$\vec{r}_1^{\,T} \vec{r}_2 = \cos\theta(-\sin\theta) + \sin\theta \cdot \cos\theta = 0$

$\|\vec{r}_1\|_2^2 = \cos^2\theta + \sin^2\theta = 1$

$\|\vec{r}_2\|_2^2 = \sin^2\theta + \cos^2\theta = 1$

- P permutation matrix.

Properties of Orthogonal Matrices

$Q^T Q = I$

Theorem 4

Let $Q \in \mathbb{R}^{n \times n}$ *be orthogonal. Then*

1. $Q^{-1} = Q^{\mathrm{T}}$;
2. Q^{T} *is also an orthogonal matrix;*
3. $\kappa_2(Q) = 1$; the best that we can hope for!
4. *For any* $A \in \mathbb{R}^{n \times n}$, $\|AQ\|_2 = \|A\|_2$;
5. *if* $P \in \mathbb{R}^{n \times n}$ *is another orthogonal matrix, then* PQ *is also orthogonal.*

$$(PQ)^T (PQ) = Q^T \underbrace{P^T P}_{= I} Q = Q^T Q = I$$

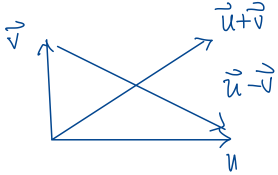
Why Do We Like Orthogonal Vectors?

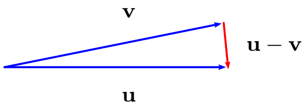
- If $\mathbf{u}$ and $\mathbf{v}$ are orthogonal, then

$$\|\mathbf{u} \pm \mathbf{v}\|_2^2 = \|\vec{u}\|_2^2 + \|\vec{v}\|_2^2 \cancel{\pm 2\vec{u}^T\vec{v}}$$

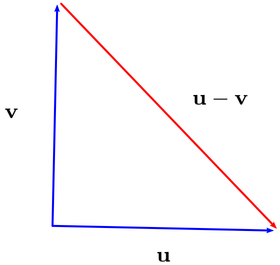
(Pythagorean theorem)

- Without orthogonality, it is possible that $\|\mathbf{u} - \mathbf{v}\|_2$ is much smaller than $\|\mathbf{u}\|_2$ and $\|\mathbf{v}\|_2$.
- The addition and subtraction of orthogonal vectors are guaranteed to be well-conditioned.





Catastrophic cancellation.



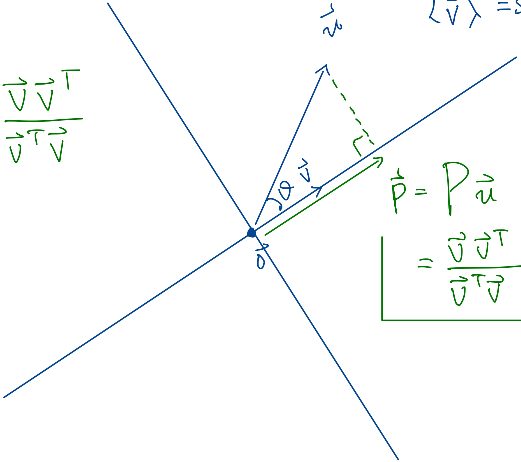
Appendix: Projection and Reflection

(Review of Linear Algebra)

Given $\vec{u}, \vec{v} \in \mathbb{R}^n$

Today: $\|\cdot\|$ for $\|\cdot\|_2$

$\langle \vec{v} \rangle = \text{span}\{\vec{v}\}$



$P = \dfrac{\vec{v}\vec{v}^T}{\vec{v}^T\vec{v}}$

$\vec{p} = P\vec{u}$

$= \dfrac{\vec{v}\vec{v}^T}{\vec{v}^T\vec{v}}\vec{u}$

- $\|\vec{p}\| = \|\vec{u}\| \cos\theta.$
- $\vec{p} = \|\vec{p}\|$ (unit vector in the dir. of $\vec{v}$)

$= \|\vec{p}\| \dfrac{\vec{v}}{\|\vec{v}\|}$

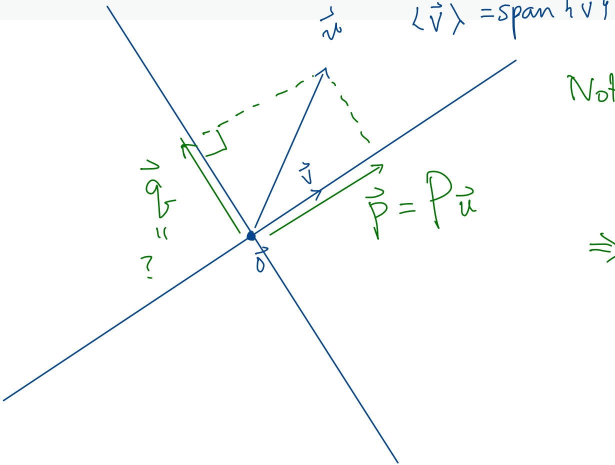
$= \dfrac{\|\vec{u}\| \cos\theta \|\vec{v}\|}{\|\vec{v}\|} \dfrac{\vec{v}}{\|\vec{v}\|}$

$= \dfrac{\|\vec{u}\| \|\vec{v}\| \cos\theta}{\|\vec{v}\|^2} \vec{v}$

$= \dfrac{\vec{u}^T\vec{v}}{\vec{v}^T\vec{v}} \vec{v} = \vec{v} \dfrac{\vec{v}^T\vec{u}}{\vec{v}^T\vec{v}} = P\vec{u}$

P

$\langle \vec{v} \rangle^{\perp}$ = orthogonal complement of $\langle \vec{v} \rangle$.



$\langle \vec{v} \rangle^{\perp}$ = orthogonal complement of $\langle \vec{v} \rangle$.

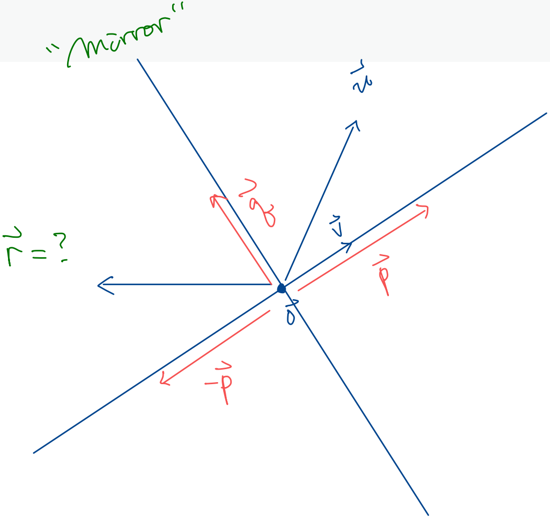
Note that

$$\vec{p} + \vec{q} = \vec{u}$$

$$\Rightarrow \vec{q} = \vec{u} - \vec{p}$$

$$= \vec{u} - P\vec{u}$$

$$= \underbrace{(I - P)}_{\shortparallel \atop Q} \vec{u}$$



$\langle \vec{v} \rangle = \text{span}\{\vec{v}\}$

$$\begin{aligned}\vec{r} &= \vec{q} + (-\vec{p}) \\ &= (I - P)\vec{u} - P\vec{u} \\ &= \underbrace{(I - 2P)}_{= R}\vec{u}\end{aligned}$$

$\langle \vec{v} \rangle^{\perp}$ = orthogonal complement of $\langle \vec{v} \rangle$.

Projection and Reflection Operators

Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^m$ be nonzero vectors.

- Projection of $\mathbf{u}$ onto $\langle \mathbf{v} \rangle = \text{span}(\mathbf{v})$:

$$\frac{\mathbf{v}^{\mathrm{T}}\mathbf{u}}{\mathbf{v}^{\mathrm{T}}\mathbf{v}}\mathbf{v} = \underbrace{\left(\frac{\mathbf{v}\mathbf{v}^{\mathrm{T}}}{\mathbf{v}^{\mathrm{T}}\mathbf{v}}\right)}_{=:P}\mathbf{u} =: P\mathbf{u}.$$

- Projection of $\mathbf{u}$ onto $\langle \mathbf{v} \rangle^{\perp}$, the orthogonal complement of $\langle \mathbf{v} \rangle$:

$$\mathbf{u} - \frac{\mathbf{v}^{\mathrm{T}}\mathbf{u}}{\mathbf{v}^{\mathrm{T}}\mathbf{v}}\mathbf{v} = \left(I - \frac{\mathbf{v}\mathbf{v}^{\mathrm{T}}}{\mathbf{v}^{\mathrm{T}}\mathbf{v}}\right)\mathbf{u} =: (I - P)\mathbf{u}.$$

- Reflection of $\mathbf{u}$ across $\langle \mathbf{v} \rangle^{\perp}$:

$$\mathbf{u} - 2\frac{\mathbf{v}^{\mathrm{T}}\mathbf{u}}{\mathbf{v}^{\mathrm{T}}\mathbf{v}}\mathbf{v} = \left(I - 2\frac{\mathbf{v}\mathbf{v}^{\mathrm{T}}}{\mathbf{v}^{\mathrm{T}}\mathbf{v}}\right)\mathbf{u} =: (I - 2P)\mathbf{u}.$$

Projection and Reflection Operators (cont')

Summary: for given $\mathbf{v} \in \mathbb{R}^m$, a nonzero vector, let

$$P = \frac{\mathbf{v}\mathbf{v}^{\mathrm{T}}}{\mathbf{v}^{\mathrm{T}}\mathbf{v}} \in \mathbb{R}^{m \times m}.$$

Then the following matrices carry out geometric transformations

- Projection onto $\langle \mathbf{v} \rangle$: P
- Projection onto $\langle \mathbf{v} \rangle$: $I - P$
- Reflection across $\langle \mathbf{v} \rangle^{\perp}$: $I - 2P$

Note. If $\mathbf{v}$ were a unit vector, the definition of P simplifies to $P = \mathbf{v}\mathbf{v}^{\mathrm{T}}$.