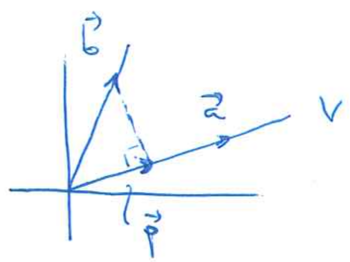
MATH 312
LECTURE 14

: Orthogonal basis and Gram-Schmidt algorithm.

⌈First we did an exercise from a previous exam about least squares⌋

Recall from previous lectures that:

- Projection of a vector $\vec{b}$ onto a line



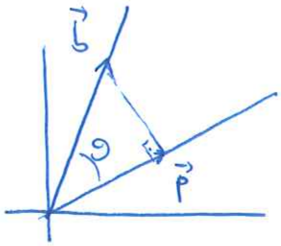
1) We choose $\vec{a}$ (basis for the line V)

2) Then, $\vec{p} = \underbrace{\frac{\vec{a}\vec{a}^T}{\|\vec{a}\|^2}}_{\text{matrix}} \vec{b} = \underbrace{\frac{\vec{a}\cdot\vec{b}}{\|\vec{a}\|^2}}_{\text{number}} \vec{a}$

Remark: Since we choose $\vec{a}$, we can always choose it such that $\|\vec{a}\| = 1$.

In this case, $\vec{p} = (\vec{a}\cdot\vec{b})\vec{a} \rightsquigarrow \|\vec{p}\| = |\vec{a}\cdot\vec{b}|$.

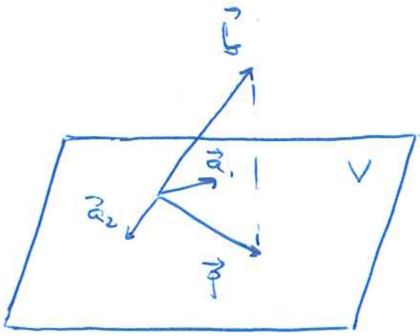
⌈You could obtain this formula from basic trigonometry:



$$\left.\begin{array}{l} \cos\theta = \dfrac{\|\vec p\|}{\|\vec b\|} \Rightarrow \|\vec p\| = \|\vec b\|\cos\theta \\[2ex] \vec a\cdot\vec b = \|\vec a\|\,\|\vec b\|\cos\theta \underset{\uparrow \atop \|\vec a\|=1}{=} \|\vec b\|\cos\theta \end{array}\right\} \Rightarrow$$

$$\Rightarrow \|\vec p\| = \vec a\cdot\vec b \Rightarrow \vec p = \|\vec p\|\,\vec a = (\vec a\cdot\vec b)\,\vec a.$$

∘ Projection of a vector $\vec b$ onto a subspace:



1) Choose a basis for V: $\{\vec a_1, \dots, \vec a_n\}$

2) Create $A = \left[\vec a_1 \,\middle|\, \cdots \,\middle|\, \vec a_n\right]$

3) $\vec p = P\vec b = A(A^TA)^{-1}A^T\,\vec b$.

Remark:

$$A^TA = \begin{bmatrix} \text{---}\ \vec a_1^{\,T}\ \text{---} \\ \vdots \\ \text{---}\ \vec a_n^{\,T}\ \text{---} \end{bmatrix} \begin{bmatrix} | & & | \\ \vec a_1 & \cdots & \vec a_n \\ | & & | \end{bmatrix} = \begin{bmatrix} \vec a_1^{\,T}\vec a_1 & \vec a_1^{\,T}\vec a_2 & \cdots & \vec a_1^{\,T}\vec a_n \\ \vec a_2^{\,T}\vec a_1 & \vec a_2^{\,T}\vec a_2 & \cdots & \vec a_2^{\,T}\vec a_n \\ \vdots & \vdots & \ddots & \vdots \\ \vec a_n^{\,T}\vec a_1 & \vec a_n^{\,T}\vec a_2 & \cdots & \vec a_n^{\,T}\vec a_n \end{bmatrix}.$$

In the case of a plane, $A^TA = \begin{bmatrix} \vec a_1^{\,T}\vec a_1 & \vec a_1^{\,T}\vec a_2 \\ \vec a_2^{\,T}\vec a_1 & \vec a_2^{\,T}\vec a_2 \end{bmatrix}$.

So if we choose $\vec a_1, \vec a_2$ of unit length and perpendicular to each other, i.e., $\vec a_1\cdot\vec a_2 = 0$ and $\|\vec a_1\| = \|\vec a_2\| = 1$, then

$$A^T A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

• Def: Orthonormal vectors.

We say that a set of vectors $\{\vec{q}_1, \dots, \vec{q}_n\}$ is orthonormal if

$$\vec{q}_i^{\,T} \vec{q}_j \equiv \vec{q}_i \cdot \vec{q}_j = \begin{cases} 0 & i \neq j \quad \text{(orthogonal} \equiv \text{perpendicular)} \\ 1 & i = 1 \quad \text{(unit length, i.e. } \|\vec{q}_i\| = 1\text{)}. \end{cases}$$

• Remark:

$$\begin{array}{c} Q \text{ is a matrix with} \\ \text{orthonormal columns} \end{array} \iff Q^T Q = I.$$

Indeed, for $Q = \left[\vec{q}_1 \middle| \cdots \middle| \vec{q}_n \right]$,

$$Q^T Q = \begin{bmatrix} \vec{q}_1^{\,T} \vec{q}_1 & \vec{q}_1^{\,T} \vec{q}_2 & \cdots & \vec{q}_1^{\,T} \vec{q}_n \\ \vdots & \vdots & \ddots & \vdots \\ \vec{q}_n^{\,T} \vec{q}_1 & \vec{q}_n^{\,T} \vec{q}_2 & \cdots & \vec{q}_n^{\,T} \vec{q}_n \end{bmatrix} \overset{\text{orthonormal } (\vec{q}_1, \dots, \vec{q}_n).}{=} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

• Remark: Careful! [*]

For Q rectangular, if the columns are orthonormal, we have that $Q^TQ = I$, but it is false that $QQ^T = I$:

Ex: $Q = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \Rightarrow Q^TQ = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$QQ^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$!!

• Remark: If Q square, then the following statements are equivalent: (i.e., for Q square, if one of the statements is true, then the rest are true too.).

1) $Q^TQ = I$

2) Q has orthonormal columns.

3) $QQ^T = I$

4) Q has orthonormal rows.

5) $Q^{-1} = Q^T$.

Proof: We have already seen 1), 2).

Now, using properties of determinants and 1), we find that

$$\left.\begin{array}{l} Q^TQ = I \Rightarrow \det(Q^TQ) = \det(I) = 1, \\ \det(Q^TQ) = \det(Q^T)\det(Q) = (\det(Q))^2 \end{array}\right\} \Rightarrow \det(Q) = \pm 1 \neq 0 \Rightarrow$$

$\Rightarrow Q^{-1}$ exists.

Thus,

$Q^TQ = I \Rightarrow Q^TQQ^{-1} = Q^{-1} \Rightarrow \| Q^T = Q^{-1} \|$ (this was 5))

Thus,

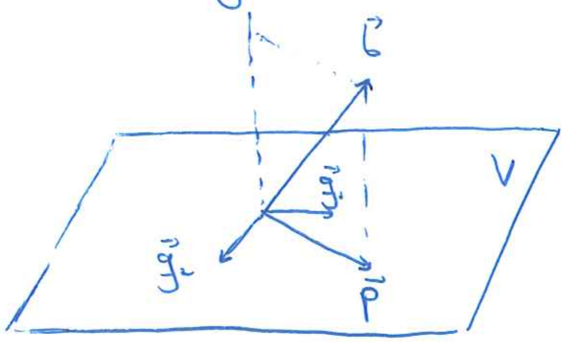
$QQ^{-1} = I = QQ^T$ (this was 3).

Now, denote $\vec{r}_1^{\,T}, \vec{r}_2^{\,T}, \dots, \vec{r}_n^{\,T}$ the rows of Q:

$$QQ^T = \begin{bmatrix} \text{---} & \vec{r}_1^{\,T} & \text{---} \\ & \vdots & \\ \text{---} & \vec{r}_n^{\,T} & \text{---} \end{bmatrix} \begin{bmatrix} | & & | \\ \vec{r}_1 & \cdots & \vec{r}_n \\ | & & | \end{bmatrix} = \begin{bmatrix} \vec{r}_1^{\,T}\vec{r}_1 & \vec{r}_1^{\,T}\vec{r}_2 & \cdots & \vec{r}_1^{\,T}\vec{r}_n \\ \vdots & \vdots & \ddots & \vdots \\ \vec{r}_n^{\,T}\vec{r}_1 & \vec{r}_n^{\,T}\vec{r}_2 & \cdots & \vec{r}_n^{\,T}\vec{r}_n \end{bmatrix} =$$

$$= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \Rightarrow \{\vec{r}_1, \dots, \vec{r}_n\} \text{ are orthonormal (this was 4))}.$$

• Let's go back to projections:



1) We <u>choose</u> an orthonormal basis of V, $\{\vec{q}_1, \dots, \vec{q}_n\}$

2) Create $Q = [\vec{q}_1 | \cdots | \vec{q}_n]$.

Finally, 3) $\vec{p} = P\vec{b}$, with $P = Q\underbrace{(Q^TQ)}_{=I}^{-1}Q^T = \underbrace{QQ^T}_{\text{This is not the identity (in general).}}$

• <u>Remark</u>: Notice the expression

$$\vec{p} \equiv P_V \vec{b} = QQ^T\vec{b} = \left[\vec{q}_1 \middle| \cdots \middle| \vec{q}_n\right]\begin{bmatrix} \text{---} \ \vec{q}_1^{\,T} \ \text{---} \\ \vdots \\ \text{---} \ \vec{q}_n^{\,T} \ \text{---} \end{bmatrix}\vec{b} =$$

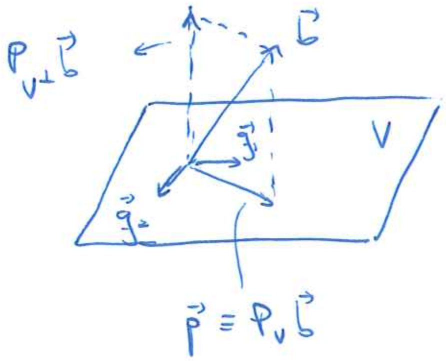
$$= (\underbrace{\vec{q}_1\vec{q}_1^{\,T}}_{\text{matrices}} + \cdots + \underbrace{\vec{q}_n\vec{q}_n^{\,T}}_{\text{matrices}})\vec{b} = \vec{q}_1\vec{q}_1^{\,T}\vec{b} + \cdots + \vec{q}_n\vec{q}_n^{\,T}\vec{b} =$$

$$= \boxed{\underbrace{(\vec{q}_1 \cdot \vec{b})}_{\text{numbers}}\vec{q}_1 + \cdots + \underbrace{(\vec{q}_n \cdot \vec{b})}_{\text{numbers}}\vec{q}_n = P_V\vec{b}}$$

projections of $\vec{b}$ onto the "axis" $\vec{q}_1, \dots, \vec{q}_n$

Using orthonormal basis, the projection of $\vec{b}$ onto a subspace is the sum of the one-dimensional projections!

• Remark: Another way of seeing all this



In the picture it is easy to see that $\left\| \vec{b} = P_V\vec{b} + P_{V^\perp}\vec{b} \right\|$ (always true; V any vector space).

Since $P_V\vec{b} \in V$, $P_V\vec{b} = \alpha_1\vec{g}_1 + \alpha_2\vec{g}_2$.

That is,

$$\vec{b} = \alpha_1\vec{g}_1 + \alpha_2\vec{g}_2 + \underbrace{P_{V^\perp}\vec{b}}_{\in V^\perp,\text{ i.e. it is perpendicular to } V \text{ (and in particular to } \vec{g}_1 \text{ and } \vec{g}_2)}$$

Thus,

$$\left.\begin{array}{l} \vec{b}\cdot\vec{g}_1 = \alpha_1\underbrace{\vec{g}_1\cdot\vec{g}_1}_{=1} + \alpha_2\underbrace{\cancel{\vec{g}_2\cdot\vec{g}_1}}_{=0} + \underbrace{\cancel{P_{V^\perp}\vec{b}\cdot\vec{g}_1}}_{=0} \Rightarrow \alpha_1 = \vec{b}\cdot\vec{g}_1 \\ \vec{b}\cdot\vec{g}_2 = \alpha_1\underbrace{\cancel{\vec{g}_1\cdot\vec{g}_2}}_{=0} + \alpha_2\underbrace{\vec{g}_2\cdot\vec{g}_2}_{=1} + \underbrace{\cancel{P_{V^\perp}\vec{b}\cdot\vec{g}_2}}_{=0} \Rightarrow \alpha_2 = \vec{b}\cdot\vec{g}_2 \end{array}\right\} \Rightarrow$$

$$\Rightarrow P_V\vec{b}_\circ = (\vec{b}\cdot\vec{g}_1)\vec{g}_1 + (\vec{b}\cdot\vec{g}_2)\vec{g}_2 \,/\!\!/ .$$

• Summary (important)

Let V be a vector space with orthonormal basis $\mathcal{V} = \{ \vec{q}_1, \dots, \vec{q}_n \}$. Then,

$$P_V \vec{b} = (\vec{q}_1 \cdot \vec{b})\, \vec{q}_1 + \dots + (\vec{q}_n \cdot \vec{b})\, \vec{q}_n .$$

Two remarks:

1) If $\vec{b} \in V$, then it's clear that $\vec{b} \equiv P_V \vec{b}$, so

$$\vec{b} = (\vec{q}_1 \cdot \vec{b})\, \vec{q}_1 + \dots + (\vec{q}_n \cdot \vec{b})\, \vec{q}_n .$$

That is, the coordinates of $\vec{b}$ in an orthogonormal basis are given by the inner products with the basis vectors:

$$\vec{b}\,\big|_{\mathcal{V}} = \begin{bmatrix} \vec{q}_1 \cdot \vec{b} \\ \vdots \\ \vec{q}_n \cdot \vec{b} \end{bmatrix}$$

2) If $\mathcal{U} = \{ \vec{u}_1, \dots, \vec{u}_n \}$ is an orthogonal basis only, i.e., perpendicular vectors but not unit size, then

$$P_{\mathcal{U}} \vec{b} = \left(\frac{\vec{u}_1 \cdot \vec{b}}{\|\vec{u}_1\|^2} \right) \vec{u}_1 + \dots + \left(\frac{\vec{u}_n \cdot \vec{b}}{\|\vec{u}_n\|^2} \right) \vec{u}_n .$$

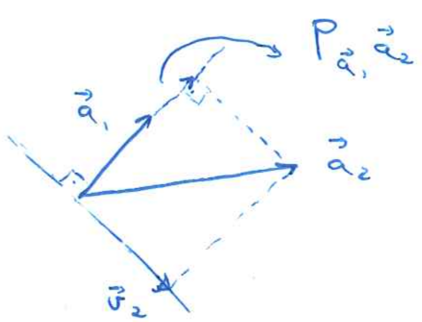
→ <u>Typical problem:</u> Since orthonormal basis seem so convenient, given a subspace, like the plane $x+y+z=0$ in $\mathbb{R}^3$, how do we find an orthonormal basis for the plane?

↳ We only know how to find a (general) basis:

$$x+y+z=0 \to \begin{bmatrix}1 & 1 & 1\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix}=0 \to \Big\{\underbrace{\begin{bmatrix}-1\\1\\0\end{bmatrix}}_{\vec{a}_1},\underbrace{\begin{bmatrix}-1\\0\\1\end{bmatrix}}_{\vec{a}_2}\Big\}$$ (but this is not orthonormal).

<u>Gram-Schmidt algorithm</u>

Idea: (in 2d)



1) Take $\vec{v}_1=\vec{a}_1$

2) Find the component of $\vec{a}_2$ parallel to $\vec{a}_1$: its projection

$$P_{\vec{a}_1}\vec{a}_2=\left(\frac{\vec{a}_1^T\vec{a}_2}{\|\vec{a}_1\|^2}\right)\vec{a}_1$$

3) Then, $\vec{v}_2=\vec{a}_2-\left(\frac{\vec{a}_1^T\vec{a}_2}{\|\vec{a}_1\|^2}\right)\vec{a}_1$ is clearly orthogonal to $\vec{v}_1$

4) Finally, normalise to make them unit vectors: $\vec{q}_1=\frac{\vec{v}_1}{\|\vec{v}_1\|}$, $\vec{q}_2=\frac{\vec{v}_2}{\|\vec{v}_2\|}$

• General: Given a basis $\{\vec{a}_1, \dots, \vec{a}_n\}$, Gram-Schmidt provides a new orthonormal basis $\{\vec{g}_1, \dots, \vec{g}_n\}$ with

$\text{span}\{\vec{a}_1, \dots, \vec{a}_n\} = \text{span}\{\vec{g}_1, \dots, \vec{g}_n\}$,

where we obtain $\vec{g}_1, \dots, \vec{g}_n$ from $\vec{a}_1, \dots, \vec{a}_n$ as follows:

1) $\vec{v}_1 = \vec{a}_1$

2) $\vec{v}_2 = \vec{a}_2 - \left(\dfrac{\vec{v}_1^T \vec{a}_2}{\|\vec{v}_1\|^2}\right)\vec{v}_1$

3) $\vec{v}_3 = \vec{a}_3 - P_{\text{span}\{\vec{v}_1, \vec{v}_2\}}\vec{a}_3 = \vec{a}_3 - \left(\dfrac{\vec{v}_1^T \vec{a}_3}{\|\vec{v}_1\|^2}\right)\vec{v}_1 - \left(\dfrac{\vec{v}_2^T \vec{a}_3}{\|\vec{v}_2\|^2}\right)\vec{v}_2$

$\vdots$

k) $$\vec{v}_k = \vec{a}_k - \underbrace{\left(\frac{\vec{v}_1^T \vec{a}_k}{\|\vec{v}_1\|^2}\right)\vec{v}_1 - \dots - \left(\frac{\vec{v}_{k-1}^T \vec{a}_k}{\|\vec{v}_{k-1}\|^2}\right)\vec{v}_{k-1}}_{P_{\text{span}\{\vec{v}_1, \dots, \vec{v}_{k-1}\}}\vec{a}_k}$$

• To conclude, $\vec{g}_i = \dfrac{\vec{v}_i}{\|\vec{v}_i\|}$ (for $i = 1, \dots, n$).

[→ Two exercises/examples from previous exams]