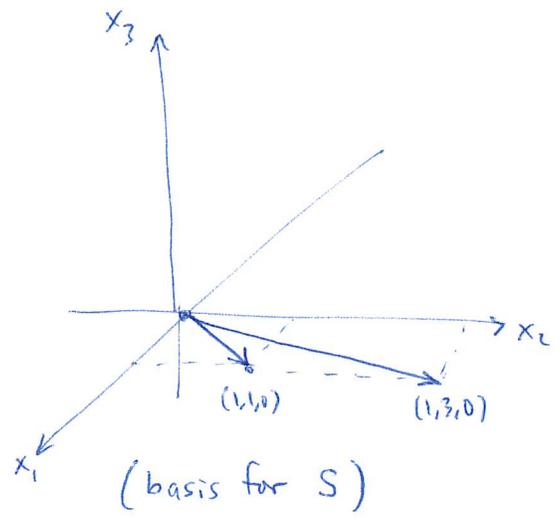


ECE 532 - Lecture 7 - orthogonality and Gram-Schmidt

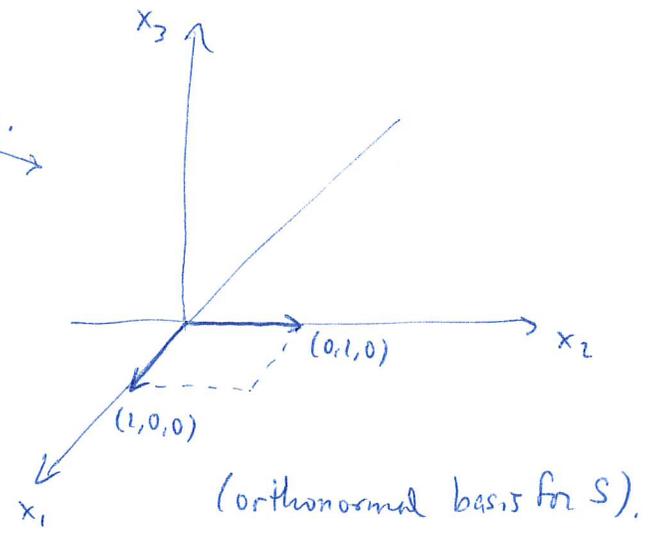
A set of vectors $\{u_1, \dots, u_k\}$ in \mathbb{R}^n is

- i) orthogonal if $u_i^T u_j = 0$ for all $i \neq j$
 - ii) normalized if $\|u_i\| = 1$ for all i
- } "orthonormal" if both are true.

example : $S = \text{span} \left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} \right) = \text{span} \left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right)$



same span.



Theorem : every subspace has an orthonormal basis.

(we'll see how to construct one!)

in other words, given any $A = \underbrace{[a_1, \dots, a_k]}_{\text{lin. indep}}$ and $S = \mathcal{R}(A)$,

there are orthonormal vectors $\{q_1, \dots, q_k\}$ such that if $Q = [q_1, \dots, q_k]$

then $\mathcal{R}(A) = \mathcal{R}(Q)$. A matrix like Q (orthonormal columns)

is called an orthogonal matrix. (really should be called an orthonormal matrix, but that's not the convention).

Orthogonal matrices: $Q = [q_1, \dots, q_n] \in \mathbb{R}^{m \times n}$ is orthogonal if

$$q_i^T q_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases} \quad \text{equivalently, } Q \text{ is orthogonal if:}$$

$$Q^T Q = \begin{bmatrix} q_1^T \\ q_2^T \\ \vdots \\ q_n^T \end{bmatrix} [q_1 \ q_2 \ \dots \ q_n] = \begin{bmatrix} q_1^T q_1 & q_1^T q_2 & \dots & q_1^T q_n \\ q_2^T q_1 & q_2^T q_2 & \dots & q_2^T q_n \\ \vdots & \vdots & \ddots & \vdots \\ q_n^T q_1 & q_n^T q_2 & \dots & q_n^T q_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = I$$

properties [remember: P, Q need not be square!]

i) if P, Q orthogonal, then PQ is orthogonal.

$$\text{proof: } (PQ)^T (PQ) = Q^T \underbrace{P^T P}_I Q = \underbrace{Q^T Q}_I = I. \quad \blacksquare$$

ii) if Q is orthogonal, then 2-norm is preserved: $\|Qx\| = \|x\|$.

$$\text{proof: } \|Qx\|^2 = (Qx)^T (Qx) = x^T \underbrace{Q^T Q}_I x = x^T x = \|x\|^2. \quad \blacksquare$$

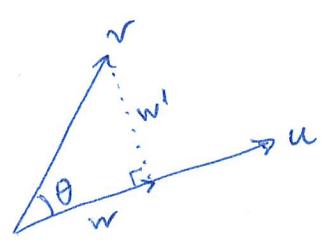
more properties!

i) in general, $QQ^T \neq I$. But if Q is square and orthogonal, $Q^T Q = QQ^T = I$. i.e. Q^T is also orthogonal, and $Q^{-1} = Q^T$.

ii) if $Q_1 \in \mathbb{R}^{m \times n}$ is orthogonal, there exists $Q_2 \in \mathbb{R}^{m \times (m-n)}$ also orthogonal such that $[Q_1 \ Q_2]$ is orthogonal (and square).

$$\text{In fact, } R(Q_2) = R(Q_1)^\perp$$

projections : $w = \text{proj}_u(v)$ (projection of v onto u).



(see picture) — it's a decomposition of v into $w + w'$ where w is aligned with u and w' is orthogonal to u .

v has length $\|v\|$. Therefore, $\|w\| = \|v\| \cos \theta$
 $\Rightarrow \|w\| = \frac{\|u\| \cdot \|v\| \cdot \cos \theta}{\|u\|} = \left(\frac{u^T v}{\|u\|} \right)$.

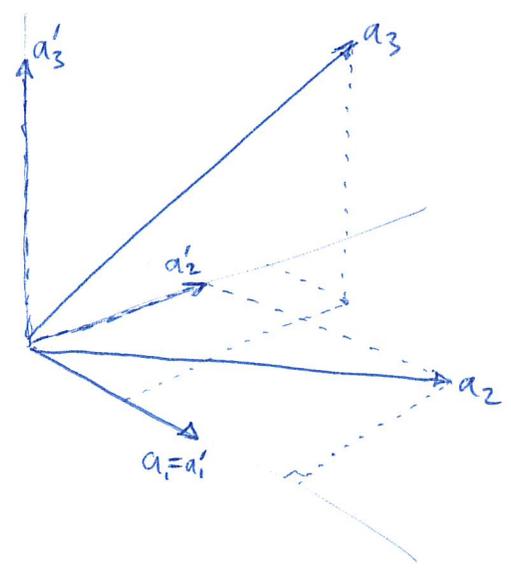
and the direction (normalized) of w should be $\frac{u}{\|u\|}$.

therefore, $w = \|w\| \cdot \frac{u}{\|u\|} = \frac{u^T v}{\|u\|^2} u$

so: $\text{proj}_u(v) = \frac{u^T v}{\|u\|^2} u$. We also have $w' = v - \text{proj}_u(v)$.

Gram-Schmidt orthogonalization given $\{a_1, a_2, \dots, a_n\}$,

$a'_1 = a_1$
 $a'_2 = a_2 - \text{proj}_{a'_1}(a_2)$
 $a'_3 = a_3 - \text{proj}_{a'_1}(a_3) - \text{proj}_{a'_2}(a_3)$
 \vdots
 $a'_n = a_n - \sum_{i=1}^{n-1} \text{proj}_{a'_i}(a_n)$



then normalize: $u_i = \frac{a'_i}{\|a'_i\|}$ for all i .

$\Rightarrow \{u_1, \dots, u_n\}$ is an orthonormal basis for $\text{span}\{a_1, \dots, a_n\}$.

what if the $\{a_1, \dots, a_n\}$ are not linearly independent?

in this case, some a_i is a linear combination of the previous $\{a_1, \dots, a_{i-1}\}$.

so $a'_i = a_i - \sum_{j=1}^{i-1} \text{proj}_{a'_j}(a_i) = 0$. simply move on and ignore

this a_i . Net result:

$$\underbrace{\{a_1, \dots, a_n\}}_{\text{any set of vectors}} \xrightarrow{\text{Gram-Schmidt}} \underbrace{\{u_1, \dots, u_r\}}_{\text{orthonormal basis for } R(A)}$$

any set of vectors

orthonormal basis for $R(A)$.
where $r = \text{rank}(A)$.

Note: $\{u_1, \dots, u_r\}$ is not unique! (could rearrange a_j 's and get a different result). There are many possible orthonormal bases in general.

in Matlab:

$\text{orth}(A)$: produces a matrix that is orthogonal and for which the range equals the range of A .

$\text{null}(A)$: produces a matrix that is orthogonal whose columns are an orthonormal basis for $N(A)$.

Gram - Schmidt examples

ex 1 $A = \begin{bmatrix} -1 & 1 & 2 \\ 1 & -1 & 2 \end{bmatrix}$: $a_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$, $a_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, $a_3 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$.

$$a'_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

$$a'_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} - \text{proj}_{\begin{pmatrix} -1 \\ 1 \end{pmatrix}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} - \left(\frac{-2}{2}\right) \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ [ignore!]}$$

$$a'_3 = \begin{pmatrix} 2 \\ 2 \end{pmatrix} - \text{proj}_{\begin{pmatrix} -1 \\ 1 \end{pmatrix}} \begin{pmatrix} 2 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} - \left(\frac{0}{2}\right) \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$$

so $a'_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$, ~~$a'_2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$~~ , $a'_3 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$

$$\Rightarrow u_1 = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, u_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \implies$$

$$U = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

ex 2 $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}$: $a_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, $a_2 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

$$a'_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$a'_2 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} - \text{proj}_{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} - \left(\frac{1+2+3}{1^2+1^2+1^2}\right) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

$$\Rightarrow u_1 = \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}, u_2 = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix} \implies$$

$$U = \begin{bmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & 0 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

⑥

if S is any subspace, we can find orthogonal $U_1 \in \mathbb{R}^{n \times r}$ such that $\mathcal{R}(U_1) = S$. Then, find U_2 such that $[U_1, U_2]$ is orthogonal and square. [how? one way is to apply G.S. to $[U_1, I]$...].

Now take any vector $x \in \mathbb{R}^n$. Since $U = [U_1, U_2]$ is orthogonal and square, $U^T U = U U^T = I$. so:

$$x = U U^T x$$

$$= [U_1, U_2] [U_1, U_2]^T x$$

$$= U_1 U_1^T x + U_2 U_2^T x.$$

$$= (u_1 u_1^T x + \dots + u_r u_r^T x) + (u_{r+1} u_{r+1}^T x + \dots + u_n u_n^T x)$$

$$= \left(\sum_{i=1}^r \text{proj}_{u_i} x \right) + \left(\sum_{i=r+1}^n \text{proj}_{u_i} x \right)$$

because the u_i are orthonormal.

$$= \underbrace{\text{proj}_{\text{span}(u_1, \dots, u_r)} x}_{\text{projection of } x \text{ onto } S} + \underbrace{\text{proj}_{\text{span}(u_{r+1}, \dots, u_n)} x}_{\text{projection of } x \text{ onto } S^\perp}.$$

projection of x onto S

projection of x onto S^\perp .

Every $x \in \mathbb{R}^n$ can be written as $x_1 + x_2$ where $x_1 \in S$, $x_2 \in S^\perp$ in a unique way. if $\mathcal{R}(U_1) = S$ and $\mathcal{R}(U_2) = S^\perp$, $[U_1, U_2]$ orthogonal, then $x_1 = U_1 U_1^T x = \text{proj}_S x$, $x_2 = U_2 U_2^T x = \text{proj}_{S^\perp} x$.

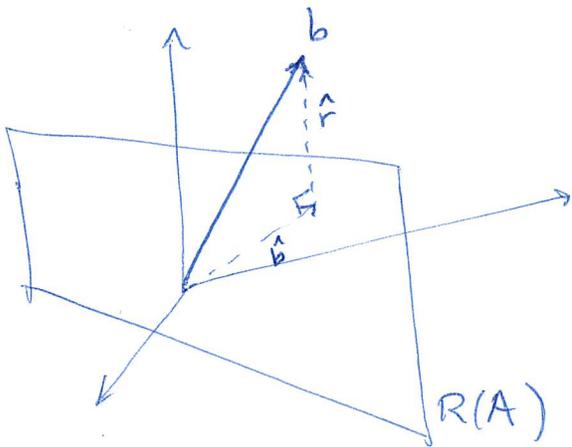
Connection to Least Squares

solving $\min_x \|b - Ax\|$ is the same as finding $\hat{b} \in R(A)$
↑
element of $R(A)$

such that $\|b - \hat{b}\|$ is as small as possible.

ie. $b = \hat{b} + \hat{r}$

where $\hat{b}^T \hat{r} = 0$, $\hat{b} \in R(A)$, $\hat{r} \in R(A)^\perp$.



if $R(U_1) = R(A)$, U_1 orthogonal,

then
$$\begin{cases} \hat{b} = U_1 U_1^T b \\ \hat{r} = U_2 U_2^T b = (I - U_1 U_1^T) b \end{cases}$$

Note: $\hat{b} = \text{proj}_{R(A)} b$, $\hat{r} = \text{proj}_{R(A)^\perp} b$. Also, $\hat{b}^T \hat{r} = b^T \underbrace{U_1 U_1^T U_2 U_2^T}_{0} b = 0$.

★ Note: \hat{b} and \hat{r} are always unique! there may be multiple x 's such that $\hat{b} = Ax$, but there is only one \hat{b} !

We already saw that LS solutions satisfy $A^T A \hat{x} = A^T b$.

If columns of A are independent, $\hat{x} = (A^T A)^{-1} A^T b$.

therefore, $\hat{b} = A \hat{x} = A(A^T A)^{-1} A^T b$.

so:
$$\begin{cases} \hat{b} = \text{proj}_{R(A)} b = A(A^T A)^{-1} A^T b \\ \hat{r} = \text{proj}_{R(A)^\perp} b = (I - A(A^T A)^{-1} A^T) b \end{cases}$$

can easily check that $\hat{b} + \hat{r} = b$ and $\hat{b}^T \hat{r} = 0$.

in Matlab: * if A has full column rank, "A \ b" (backslash) is the same as $(A^T A)^{-1} A^T b$.

* check: $A = \text{rand}(10, 5);$

$U_1 = \text{orth}(A);$

(should be zero!)

→ $U_1 * U_1' - A * \text{inv}(A' * A) * A'$