Lecture 2 : some geometry of linear maps

Note Title 1/7/2017

What are the linear maps $A : \mathbb{R}^n \to \mathbb{R}^n$ that preserve distance,

i.e. $$\|Ax - Ay\| = \|x - y\| \quad \forall x, y \in \mathbb{R}^n \qquad (\star)$$

linearity of A $\iff$ $\|Ax\|^2 = \|x\|^2 \quad \forall x \in \mathbb{R}^n$

$\|x\|^2 = \langle x, x \rangle$ $\iff$ $x^T A^T A x = x^T x \quad \forall x \in \mathbb{R}^n$

$$\underset{\langle Ax, Ax \rangle}{\overset{\|}{}} \qquad \underset{\langle x, x \rangle}{\overset{\|}{}}$$

polarization identity $\langle x, y \rangle = \frac{1}{4}[\|x+y\|^2 - \|x-y\|^2]$ or $\frac{1}{2}[\|x+y\|^2 - \|x\|^2 - \|y\|^2]$

$\iff$ $\underbrace{\langle Ax, Ay \rangle}_{y^T A^T A x} = \underbrace{\langle x, y \rangle}_{y^T x} \quad \forall x, y \in \mathbb{R}^n$

plug in $y = e_i, x = e_j$ $\iff$ $\boxed{A^T A = I}$

These are, of course, what are usually called "orthogonal matrices" in linear algebra. It's important you think of them as

"LINEAR ISOMETRIES".

maps that preserve distance, i.e. satisfy $(\star)$

There are many linear maps that are NOT isometries.

E.g. $\begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$, $\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$, $\begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix}$,

OR $\underset{\text{orthogonal}}{U}\begin{bmatrix} & \end{bmatrix}U^T$

and their higher-dimensional analogs

$\underbrace{\begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}}_{\substack{\text{scaling/dilating} \\ \text{by } \lambda_i \text{ in the} \\ e_i \text{ direction}}}$ or $\underbrace{\underset{\text{orthogonal}}{U}\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}U^T}_{\substack{\text{scaling/dilating} \\ \text{by } \lambda_i \text{ in the} \\ u_i \text{ direction.}}}$

$U = [u_1 \cdots u_n]$.

Note: $S = U\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}U^T$ satisfies $S^T = S$.

The converse is also true and is a standard result in linear algebra.

🚩 Theorem: If $S^T = S$, then it has real eigenvalues with an orthogonal set of eigenvectors.

[This fact underlies the existence of principal directions and principal curvatures of a regular surface.]

Now, we have two types of linear transformations:

① isometries (orthogonal matrices)

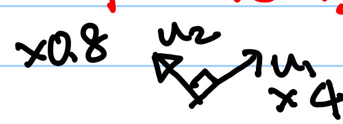
$$R^T R = I$$

They rotate, flip, or a combination of both

They do <u>not</u> scale/dilate.

② scaling/dilation (symmetric matrices) with positive eigenvalues

$$S^T = S$$



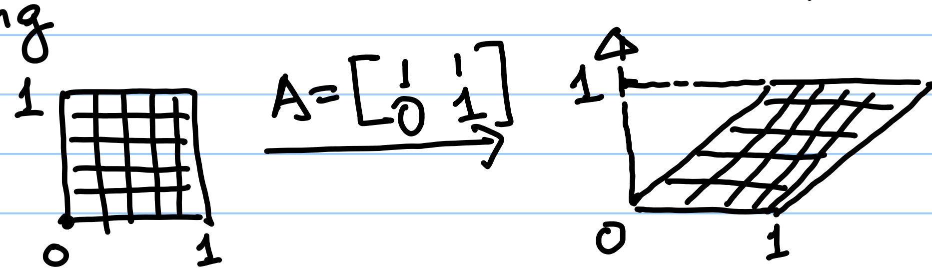
They stretch/dilate/scale.

They do not rotate or flip.

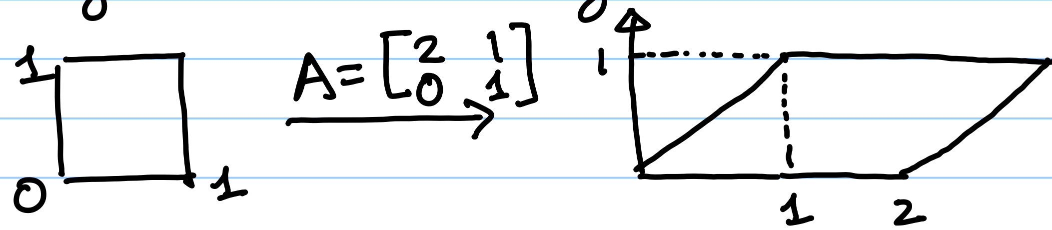
Note: symmetric matrices with only positive eigenvalues are usually called symmetric positive definite matrices.

But what about other linear maps, such as

- shearing



- shearing with some scaling mixed in



It turns out <u>every</u> linear map is a composition of the two types ① and ②:

<u>Theorem (polar decomposition)</u>

For any $A \in \mathbb{R}^{n \times n}$, $\exists\, R, S$

$$A = RS, \quad R^T R = I, \quad S = S^T$$

with positives eigenvalues.

Equivalently, the singular value decomposition (SVD)

$$A = U \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \end{bmatrix} V^T$$

for $\sigma_1 \geq \cdots \geq \sigma_n \geq 0$
U, V orthogonal.

Exercise: show that each of these two decompositions imply the other.

Back to isometry, a subtle question:

- what are all the maps, linear or nonlinear, $T : \mathbb{R}^n \to \mathbb{R}^n$ that preserve distance?

Of course, any translation $x \mapsto x + c$ for a fixed $c \in \mathbb{R}^n$ is an isometry.

So any affine map of the form

$$x \mapsto Rx + c$$

$$R^T R = I, \quad c \in \mathbb{R}^n$$

is an isometry.

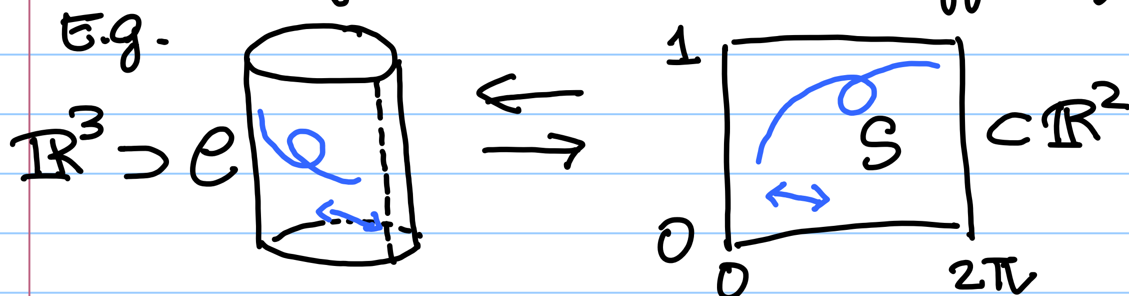
It turns out there are no others. (This is tricky to prove, and I can provide you with a reference.)

But

T: "a curved object" $\to$ "a curved object"

can also be distance preserving. (And in this context it makes no sense to ask if T is linear or affine.)

E.g.



looks like a <u>nonlinear</u> map if you naively think of it as a map from $\mathbb{R}^2$ to $\mathbb{R}^3$

$$\begin{tikzcd} \begin{bmatrix}\cos\theta\\ \sin\theta\\ z\end{bmatrix} \arrow[r, shift right, "T^{-1}"'] & \begin{bmatrix}\theta\\ z\end{bmatrix} \arrow[l, mapsto, shift right, "T"'] \end{tikzcd}$$

If $\alpha : (a,b) \to S$ is a curve on S

$T\circ\alpha : (a,b) \to \mathcal{C}$ is a curve on $\mathcal{C}$

Write $\alpha(t) = (\theta(t), z(t))$

$$\text{Length}(\alpha) = \int_a^b \sqrt{\theta'(t)^2 + z'(t)^2}\, dt$$

$$\text{Length}(T\circ\alpha) = \int_a^b \sqrt{(\cos\theta(t))'^2 + (\sin\theta(t))'^2 + z'(t)^2}\, dt$$

$$= \int_a^b \sqrt{(-\sin\theta(t)\,\theta'(t))^2 + (\cos\theta(t)\,\theta'(t))^2 + z'(t)^2}\, dt$$

$\sin^2\theta + \cos^2\theta = 1$

$$= \text{length}(\alpha)$$

In other words, T preserves distance.

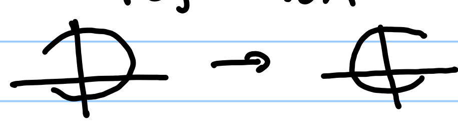
We shall come back to this, in a big way.

Now, back to the geometry of linear maps.

$$\begin{bmatrix}x_1\\ x_2\end{bmatrix} \longmapsto \begin{bmatrix}-x_1\\ x_2\end{bmatrix}$$

preserves distance, but it does not preserve orientation

"reflection"



What are the linear maps $A : \mathbb{R}^n \to \mathbb{R}^n$ that "preserve orientation"?

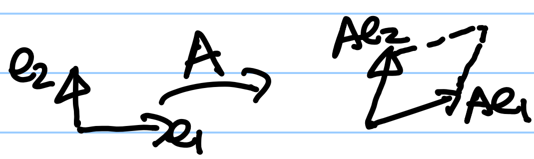
Below, I will

(i) try to convince you that it makes sense to define "orientation preserving" by the condition

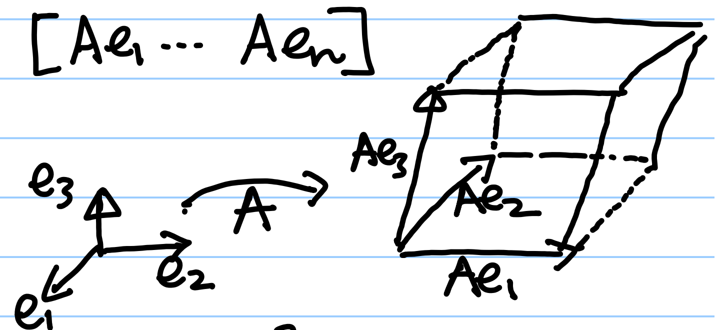
$$\boxed{\det(A) > 0}$$

(ii) at the same time, give you a self-contained presentation of what the scalar quantity $\det(A)$ is supposed to measure about the map A.

$$A = A[e_1 \cdots e_n] = [Ae_1 \cdots Ae_n]$$



$n=2$



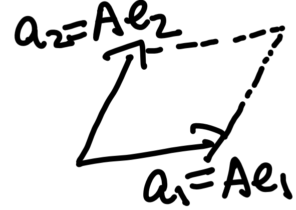
$n=3$

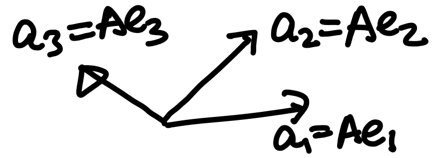
The <u>parallelepiped</u> with sides $Ae_1, \cdots, Ae_n$ is the set

$$P := \left\{ \sum_{i=1}^{n} \alpha_i A e_i \; : 0 \leq \alpha_i \leq 1 \right\}$$

Q : $\underbrace{\text{volume}_n}_{}(P) = ?$

n-dimensional 'volume' based on extending our usual notion of area and volume





You probably have seen that :

$n=2$, $\text{area}(P) = \|a_1 \times a_2\|$

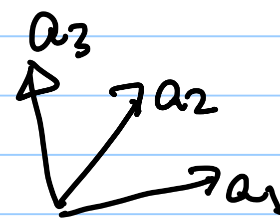
$$
\begin{aligned}
n=3 \qquad \text{volume}(P) &= |a_1 \cdot (a_2 \times a_3)| \\
&= |a_2 \cdot (a_3 \times a_1)| \\
&= |a_3 \cdot (a_1 \times a_2)|
\end{aligned}
$$

called "triple products"

It's useful for you to recall where these come from.

But I want to give you an argument that works for any dimension.

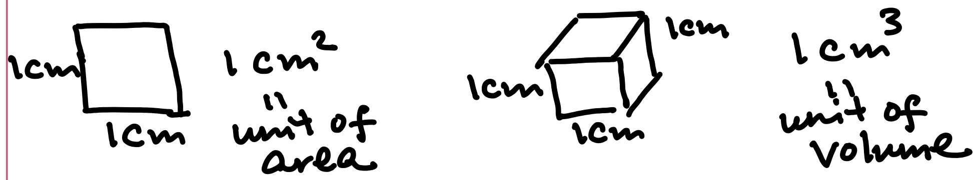
$$A = \underbrace{\begin{bmatrix} | & & | \\ a_1 & \cdots & a_n \\ | & & | \end{bmatrix}}_{\text{the } n \text{ `sides' of } P}$$



($a_i = Ae_i =$ i-th column of A.)

Imagine if a number "$D(A)$" is going to give the volume of the parallelepiped P.

According to our geometric intuition, $D(A)$ must satisfy the following properties :

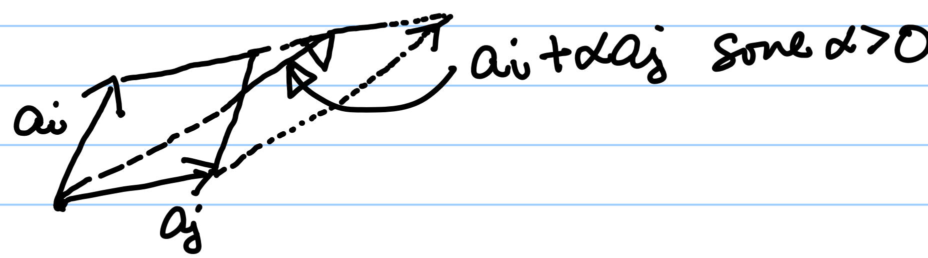


⓪ $D(I) = 1$ (volume of $[0,1]^n = 1$)

① $D\left(\begin{bmatrix} a_1 \cdots \alpha a_i \cdots a_n \end{bmatrix}\right) = \alpha D([a_1 \cdots a_i \cdots a_n]) \quad \forall \alpha > 0$

② $D([a_1 \cdots a_j \cdots a_i \cdots a_n]) = D([a_1 \cdots a_i \cdots a_j \cdots a_n])$

③ $D[a_1 \cdots a_i + \alpha a_j \cdots a_n]) = D[a_1 \cdots a_i \cdots a_n]) \quad \forall \alpha > 0$



The axioms ① and ② should be perfectly reasonable.

I let you convince yourself that ③ is also reasonable, based on "adding the other $n-2$ dimensions" to the figure above.

Now, an ingeneous observation :

⓪,①,②,③ almost completely determine what $D(A)$ is.

Part of this ingeneous observation is also to extend the meaning of $D(A)$ 'a little bit' so that the axioms ⓪,①,②,③ are more convenient to work with.

First, we would like to abandon the $\alpha > 0$ restriction in ① and ③.

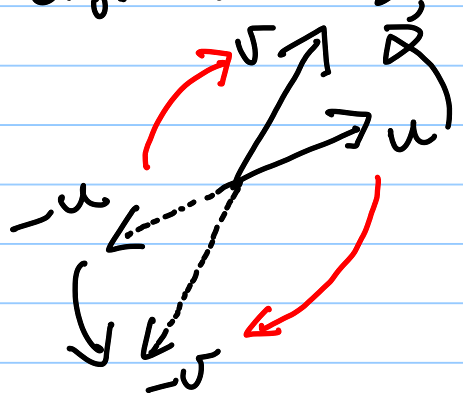
That is, instead of ① and ③, we have

1′ $$D\left(\begin{bmatrix} a_1 \cdots \alpha a_i \cdots a_n \end{bmatrix}\right) = \alpha D([a_1 \cdots a_i \cdots a_n]) \quad \forall \alpha \in \mathbb{R}$$

3′ $$D[a_1 \cdots a_i + \alpha a_j \cdots a_n]) = D[a_1 \cdots a_i \cdots a_n]) \quad i \neq j \quad \forall \alpha \in \mathbb{R}$$

But this would mean $D(A)$ can sometimes be negative. For instance, 1′ means multiplying a '-1' to any column will change the sign of $D(A)$.

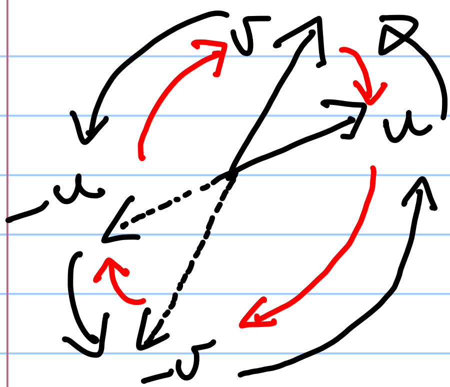
E.g. in 2-D,



$$\begin{aligned} D([u,v]) &= -D([-u,v]) \\ &= -D([u,-v]) \\ &= D([-u,-v]) \end{aligned}$$

With this in mind, and a bit more intuition (related to "orientation"), it appears that when we change ①, ③ to 1′, 3′, we should also change ② to:

(2') $D([a_1 \cdots a_j \cdots a_i \cdots a_n]) = -D([a_1 \cdots a_i \cdots a_j \cdots a_n])$



$D([u,v]) = -D([-u,v])$
$= -D([u,-v])$
$= D([-u,-v])$
$= -D([v,u]) = -D([-v,-u])$
$= D([-v,u]) = D([v,-u])$

Observe that the sign of D corresponds exactly to what we usually call "clockwise" or "anti-clockwise"

The following ordered basis of $\mathbb{R}^2$ are said to have the same orientation:

$[u,v], [v,-u], [-u,-v], [-v,u]$

The following ordered basis of $\mathbb{R}^2$ are said to have the opposite orientation:

$[v,u], [-u,v], [-v,-u], [u,-v]$.

It is harder to make up a sensible notion of 'orientation' in 3- or higher dimensions. The notion of "clockwise" and "anti-clockwise" is inherently a 2-D concept.

Nevertheless, the new axioms (1') + (2') give an extension of the 2-D picture above.

(1') + (2') $\Longrightarrow$

For different choices of signs $\varepsilon_i \in \{+1, -1\}$ and permutation $\sigma : \{1, 2, \ldots, n\} \circlearrowleft$

$$D([\varepsilon_1 a_{\sigma(1)}, \varepsilon_2 a_{\sigma(2)}, \cdots, \varepsilon_n a_{\sigma(n)}])$$

all have the same magnitude. Its sign will be a (difficult-to-interpret) way of defining "the orientation of the ordered basis

$$\{\varepsilon_1 a_{\sigma(1)}, \varepsilon_2 a_{\sigma(2)}, \cdots, \varepsilon_n a_{\sigma(n)}\}."$$

I complained about the unintuitive nature of this notion of orientation in dimension > 2. At the same time we have the following satisfying result:

Theorem:
The (geometrically meaningful!) axioms

0, 1', 2', 3' uniquely determine

$D(A)$, and in a very concrete way:

$$D(A) = \sum_{\sigma} \operatorname{sgn}(\sigma) A_{1,\sigma(1)} A_{2,\sigma(2)} \cdots A_{n,\sigma(n)}$$

Write

$D(a_1, \dots, a_n)$ instead of $D([a_1, \dots, a_n])$

$$D : \mathbb{R}^n \times \cdots \mathbb{R}^n \to \mathbb{R}$$

The proof is easy after we observe that:

Lemma: (1') – (3') imply
$D : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$ is
linear in each argument, i.e.

$$D(v_1, \cdots, \alpha v_i + \beta v_i', \cdots, v_n)$$
$$= \alpha D(v_1, \cdots, v_i, \cdots, v_n)$$
$$+ \beta D(v_1, \cdots, v_i', \cdots, v_n)$$

(we usually call such a map
<u>multi-linear</u>.)

Proof: exercise.

Proof of theorem:

Once we have this lemma, we can almost immediately derive what $D(\ )$ must be:

$$D(a_1, a_2, \cdots, a_n)$$
$$\overset{\shortparallel}{\sum_{j=1}^n a_{1j} e_j} \quad \overset{\shortparallel}{\sum_{j=1}^n a_{2j} e_j} \quad \overset{\shortparallel}{\sum_{j=1}^n a_{nj} e_j}$$

$$= \sum_{j_1=1}^n a_{1j_1} D(e_{j_1}, a_2, \cdots, a_n)$$

$$= \sum_{j_1=1}^n a_{1j_1} \sum_{j_2=1}^n a_{2j_2} D(e_{j_1}, e_{j_2}, a_3, \cdots, a_n)$$

$$= \cdots$$

$$= \sum_{j_1=1}^n \sum_{j_2=1}^n \cdots \sum_{j_n=1}^n a_{1j_1} a_{2j_2} \cdots a_{nj_n} \underbrace{D(e_{j_1}, e_{j_2}, \cdots, e_{j_n})}_{?} \quad (\star)$$

(n^n terms)

We are almost done, we need just one more standard combinatorics trick to understand the term

$$D(e_{j_1}, e_{j_2}, \ldots, e_{j_n})$$

Note that if $j_1, \ldots, j_n$ are not distinct, then the term must vanish (why?)

But when $j_1, \ldots, j_n$ are distinct, then

$(e_{j_1}, \ldots, e_{j_n})$ is just a <u>permutation</u> of $(e_1, \ldots, e_n)$.

Using the alternating property (2'),

$$D(e_{j_1}, \ldots, e_{j_n}) = (-1)^{N(\sigma)} \underbrace{D(e_1, \ldots, e_n)}_{=1 \text{ by } (0)}$$

where

$N(\sigma)$ = # of interchanging pairs needed to transform $(1, 2, \ldots, n)$ to $(j_1, j_2, \ldots, j_n)$

(This # is not unique, but the <u>parity</u> is, and that is all that matters for $\mathrm{sgn}(\sigma) := (-1)^{N(\sigma)}$.)

So we have collapsed the sum in ⊛
(with n^n terms)
into the following sum:

$$\sum_{\sigma} \operatorname{sgn}(\sigma) A_{1\sigma(1)} A_{2\sigma(2)} \cdots A_{n\sigma(n)} \quad — (*)$$

(with $n!$ terms.)

Of course $(*)$ is the familiar <u>determinant</u> of A. Now we have shown that

$$\boxed{\det(A) = \text{"signed volume" of the parallelepiped } A([0,1]^n)}$$

We say:

A is orientation preserving if $\det(A) > 0$

A is orientation reversing if $\det(A) < 0$

If $\det(A) = 0$, then

$\operatorname{vol}(A[0,1]^n) = 0$, ie the parallelepiped is 'squished' and it must be that some vectors in $A = [a_1, \cdots, a_n]$ are linearly dependent, or that A is singular.

Facts from linear algebra:

- $\det(AB) = \det(A)\det(B)$ — (P_1)

This also has an important geometric interpretation.

We showed earlier that

$$\det(A) = \text{Volume}(A[0,1]^n).$$

But it has an important generalization

$$\frac{\text{Volume}(A(S))}{\text{volume}(S)} = |\det(A)| \quad -(P_2)$$

for any "reasonable enough" set $S \subset \mathbb{R}^n$

(the technical condition for "reasonable enough" is 'measurable'.)

Ex: relates (P_1) and (P_2)

🚩 Ex: would the ratio on the LHS. of (P_2) still be indifferent to the set S if A is a nonlinear map?

- $\det(A^T) = \det(A)$

$$\left[\begin{array}{l}\text{Geometric meaning is best seen from the polar decomposition of } A \\[6pt] A = \underset{\substack{\uparrow \\ \text{isometry}}}{R}\ \underset{\nwarrow\ \text{scaling}}{\tilde{S}} \qquad \tilde{S} = U\begin{bmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_n\end{bmatrix}U^T \quad \Rightarrow \quad A^T = \underbrace{U\begin{bmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_n\end{bmatrix}U^T}_{\text{scaling}}\ \underset{\substack{\uparrow \\ \text{isometry}}}{R^T}\end{array}\right]$$

Exercise : Experiment with different 2×2 or 3×3 matrices.

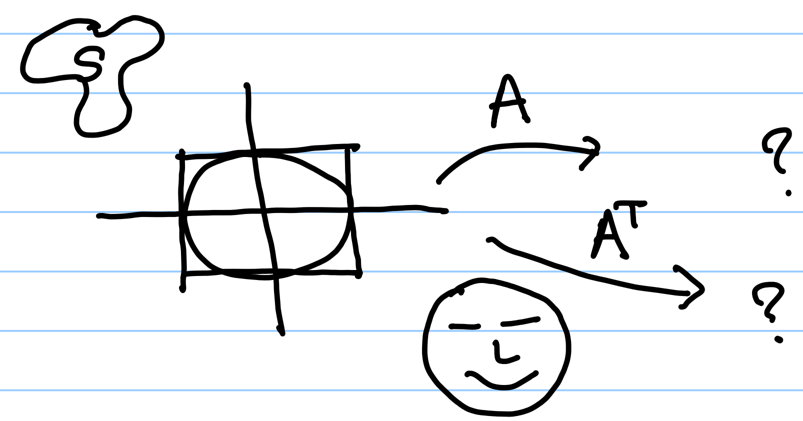
visualize the shapes of

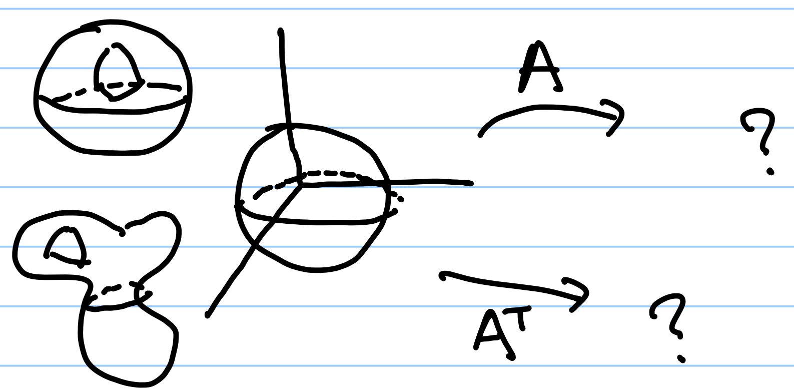
$$A[0,1]^n \quad \text{and} \quad A^T[0,1]^n$$

visualize the shapes of

$$A(S)$$

for different S. (e.g. S = circular disk)





Exercise

- If $R^T R = I$ (isometry)

  what is $\det(R)$?

- If $S^T = S$ (dilation / scaling)

  $$\left(\text{so } S = U \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} U^T \quad U\text{-orthogonal} \right)$$

  what is $\det(R)$?

Solution :

- From the geometric interpretation:

  $\det(R) = 1$ or -1 as it is an isometry.

  $\det(S) = \lambda_1 \cdots \lambda_n$

- From results in linear algebra :

$$1 = \det(I) = \det(R^T R) = \underbrace{\det(R^T)}_{\det(R)} \det(R)$$

so $\det(R)^2 = 1$, or $\det(R) = \pm 1$.

$$\det(S) = \underbrace{\det(U)}_{+/-1} \underbrace{\det\left(\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}\right)}_{\lambda_1 \cdots \lambda_n} \underbrace{\det(U^T)}_{+/-1} = \lambda_1 \cdots \lambda_n //$$

$SO(n) := \{$ all linear maps $A: \mathbb{R}^n \to \mathbb{R}^n$ that are (i) isometry and (ii) orientation preserving $\}$

$$= \{A \in \mathbb{R}^{n \times n} \mid A^T A = I, \det(A) = 1\}$$

Also:

$$O(n) := \{A \in \mathbb{R}^{n \times n} \mid A^T A = I\}.$$

$$SE(n) := \left\{ T: \mathbb{R}^n \to \mathbb{R}^n \;\middle|\; \begin{array}{l} Tx = Ax + c, \\ A \in SO(n), \\ c \in \mathbb{R}^n \end{array} \right\}$$

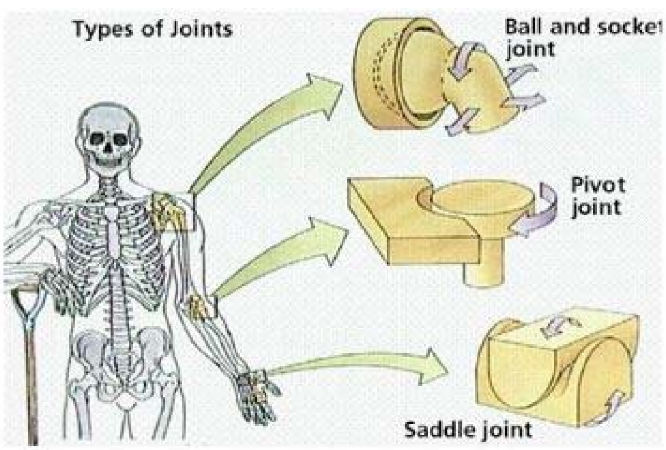
$$E(n) := \left\{ T: \mathbb{R}^n \to \mathbb{R}^n \;\middle|\; \begin{array}{l} Tx \in Ax + c, \\ A \in O(n), \\ c \in \mathbb{R}^n \end{array} \right\}$$

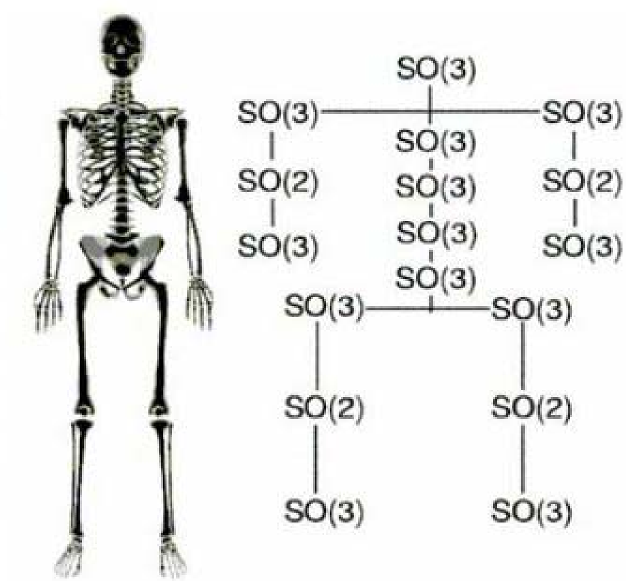
Why do we bother to collect all these specific linear / affine maps and study each of them as a mathematical entity?

- interesting structure?
- useful?
- both?

After Lecture 8, you can tell me if you find the mathematical structures interesting.

Regarding usefulness, I already gave you a hint of how the structure of SO(3) may play a role in the Frenet equation. But 3-D rotations are also ubiquitous in applications. Think of motion sensing, robotics, satellites, drones, fluild mechanics, you name it.





For example, a standard problem in robotics is called Inverse Kinematics (IK) and in this problem it is essential to have a good way to parameterize SO(3). I will share with you such a good way in Lecture 8.