Topological Data Analysis

Goal Learn the topology of data sets in $\mathbb{R}^D$

topology: Captures "shape" of objects up to continuous deformations

Benefit
- ~ independent of coordinate choice
- ~ gives geometry even in high-dim'l data
- ~ resilient to noise

Applications
- Blue Brain Project (reconstructing mouse brain)
- Population genomics
- Epidemiology

Mainly in biology.

Problem: TDA has relatively high mathematical overhead.

$\boxed{\text{Our focus}}$ → Detecting n-dimensional holes in data via linear algebra and homology

↳ persistent homology

TDA process

① **Input** data set in $\mathbb{R}^D$.

② Assign a simplicial complex to the data

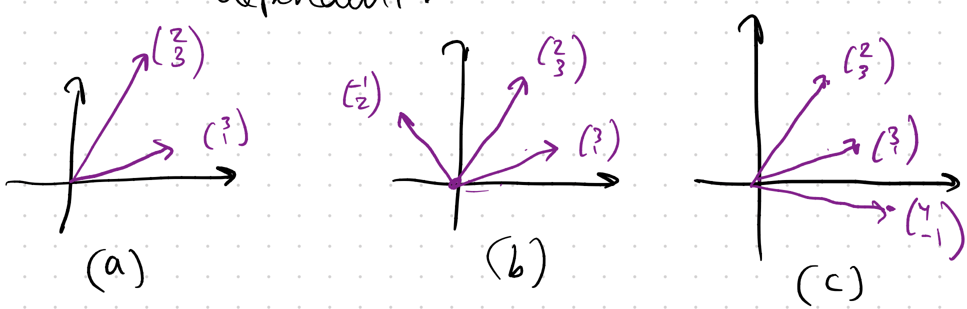
③ **Output** homology ("hole information") of simplicial complex.

★ <u>Disclaimer</u> we'll focus on very low dimensional computations to de-mystify the jargon.

Definition A collection of $n+1$ points $\{x_0, x_1, \dots, x_n\} \subseteq \mathbb{R}^D$ are <u>affinely independent</u> if for all $t_0, \dots, t_n \in \mathbb{R}^n$ with $\sum_{i=0}^{n} t_i = 0$ and

$$\sum_{i=0}^{n} t_i x_i = 0 \quad \text{implies} \quad t_0 = \dots = t_n = 0.$$

exercise Determine if the following are affinely independent:



Definition Let $\{x_0, \dots, x_n\} \subseteq \mathbb{R}^D$ be affinely independent.

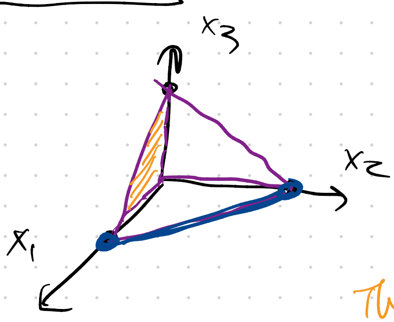
The <u>n-simplex spanned by the x_i</u> is

$$\Delta = \left\{ x \in \mathbb{R}^D \,\middle|\, x = \sum_{i=0}^{n} t_i x_i,\ \sum t_i = 1,\ t_i \geq 0 \right\}$$

The <u>dimension</u> of Δ is n.

example Compute simplex of affinely independent sets from above.

example Consider $\underset{x_0}{\begin{pmatrix}0\\0\\0\end{pmatrix}}, \underset{x_1}{\begin{pmatrix}1\\0\\0\end{pmatrix}}, \underset{x_2}{\begin{pmatrix}0\\1\\0\end{pmatrix}}, \underset{x_3}{\begin{pmatrix}0\\0\\1\end{pmatrix}}$



The simplex Δ is called a <u>tetrahedron</u>. (aka 3-simplex)

The 2-simplex associated to (x_0, x_1, x_3) is a triangle, also called a <u>facet</u> of Δ

The 1-simplex associated to (x_1, x_2) is a line, called a <u>face</u> of Δ.

<u>Definition</u> Let Δ = n-simplex of $\{x_0, \dots, x_n\}$.

For $I \subsetneq \{0, \dots, n\}$ if $|I| = p$, the simplex spanned by $\{x_i \mid i \in I\}$ is denoted σ_I, and called a p-<u>face</u> of Δ.

If $p = n-1$, σ_I is a <u>facet</u> of Δ.

The <u>boundary</u> of Δ, $\Gamma(\Delta) = \{\sigma_I \text{ facets of } \Delta\}$.

example Compute the boundary of tetrahedron.

exercise A simplicial complex is a finite collection of simplices. Draw some examples, any "bad" ones

Definition A <u>simplicial complex</u> K is a finite collection of simplices so that

(a) for all $\Delta \in K$, if $\sigma \in \Delta$ is a face, then $\sigma \in K$.

(b) If $\Delta, \sigma \in K$ have intersection in their closures: $\bar{\Delta} \cap \bar{\sigma} \neq \phi$, then $\bar{\Delta} \cap \bar{\sigma}$ is a face of both Δ and σ.

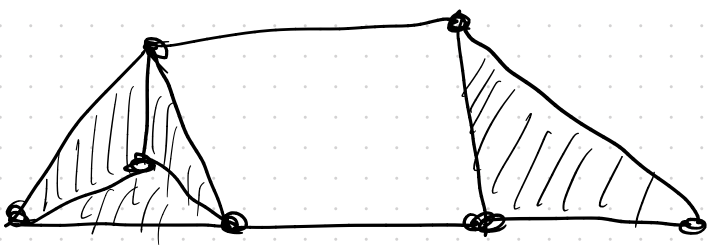
The dimension of K $\dim K = \max_{\Delta \in K}(\dim \Delta)$

the p-skeleton of K is

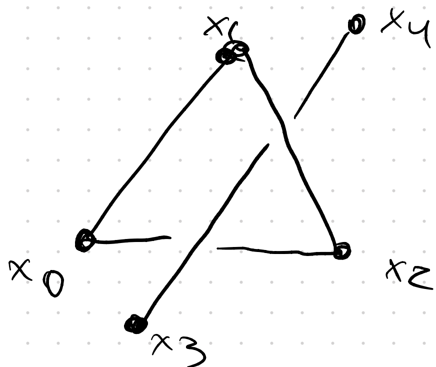
$$K^{(p)} := \{ \Delta \in K \mid \dim(\Delta) \leq p \}$$

The 0-skeleton are called the vertices of K.

$\boxed{\text{example}}$ Shaded ⇒ inside of simplex belongs to K.



$\boxed{\text{example}}$



Not a simplicial complex if we add

$\Delta(\{x_0, x_1, x_2\})$

A connection to graphs

Recall a graph $G = (V, E)$ is a collection of n vertices + edges connecting the vertices.

* **Note** don't need vertices in $\mathbb{R}^D$!

We can generalize this to simplices, as graphs are 1-simplicial complexes.

$\boxed{\text{Def'n}}$ An abstract (not geometric) simplicial complex on vertices $\{0, \dots, n\}$ is a collection of subsets

$\alpha \subseteq \{0, \dots, n\}$ so that $\forall\ \alpha \in A$ if $\beta \subseteq \alpha$, then $\beta \in A$.

$\boxed{\text{Theorem}}$ [All abstract simplicial complexes are geometric]

Let A abstract simplicial complex of dimension d. Then A has a geometric realization in $\mathbb{R}^{2d+1}$.

Exercise ① verify a graph is a 1-dim'l simplicial complex

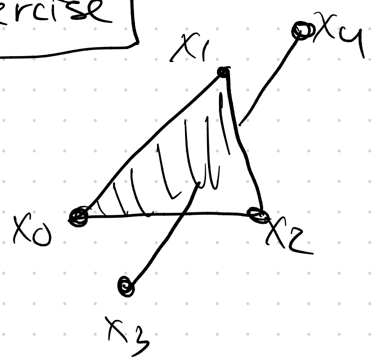
② By Theorem, every graph can be realized geometrically in $\mathbb{R}^3$. give an example of a graph that cannot be realized in $\mathbb{R}^2$? (called <u>non-planar</u>)

<u>Fun fact</u> <u>every</u> non-planar graph contains a K_5 or $K_{3,3}$ subgraph.

K_5: complete on 5 vertices

$K_{3,3}$: complete bipartite graph.

Exercise



was <u>not</u> a simplicial complex. Can it become an abstract simplicial complex? how can you embedd it?