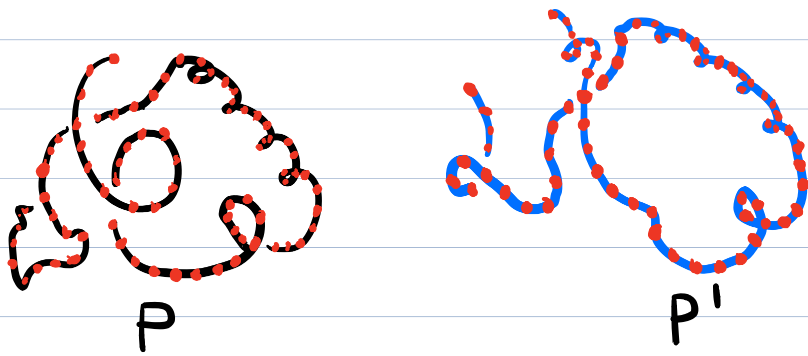
AMAT 342 Lec 14 10/17/19

Today: RMSD, continued
Topology of metric spaces

We start with some review:

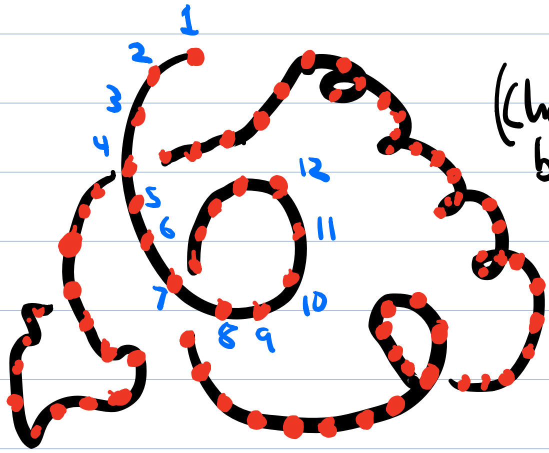
__Question__: Suppose I know the folded structure P of a protein. How do I measure the accuracy of a predicted structure P'?




__Standard Answer__: Compute a metric called __RMSD__ (root mean squared deviation)

# How to represent the 3-D structure of a protein P mathematically

- Fix an order on the atoms of the amino acid sequence

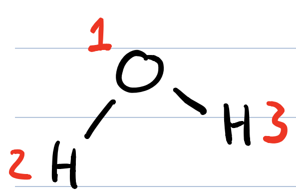



(Choice of order doesn't matter, but when comparing 2 3-D structures, we need to use the same order for both.)

P

- Let $O^n$ denote the set of all ordered lists of points in $\mathbb{R}^3$ of size $n$.
- We represent the 3-D structure of a protein with $n$ atoms as an element of $O^n$.

**Example** Let's represent a water molecule as an element of $O^3$.




Suppose the atom centers are

$$\left.\begin{array}{l} 1: (0,0,0) \\ 2: (1,0,0) \\ 3: (0,1,0). \end{array}\right\} \text{Not chemically accurate}$$

Then we represent the water molecule as the ordered list $((0,0,0),(1,0,0),(0,1,0)) \in O^3$.

We will define RMSD as a function

$$\text{RMSD}: O^n \times O^n \to [0,\infty)$$

This will almost be a metric, but not quite, so we will fix the definition to get a metric

<u>Note</u>: We can represent an element of $O^n$ as a single point in a high-dimensional space.

- For $P \in O^n$, denote the $i^{\text{th}}$ point in $P$ by $(x_i, y_i, z_i)$
- Define a function $V: O^n \to \mathbb{R}^{3n}$ by
$$V(P) = (x_1, y_1, z_1, x_2, y_2, z_2, \ldots, x_n, y_n, z_n).$$

<u>Example</u>: For $P = ((0,0,0),(1,0,0),(0,1,0)) \in O^3$,

$$V(P) = (0,0,0,1,0,0,0,1,0) \in \mathbb{R}^9.$$

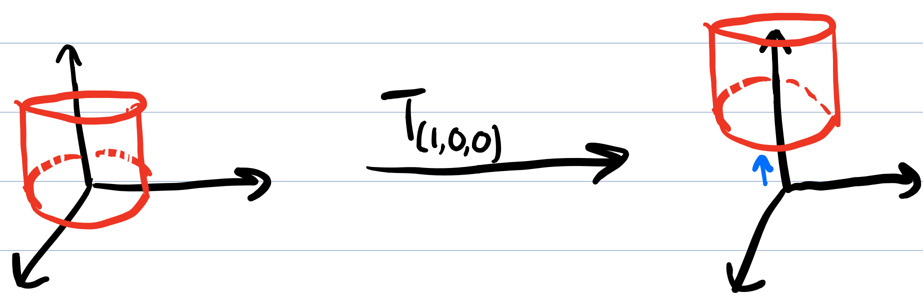
Exercise: Is $V$ invertible?

A: Yes. $V^{-1}(x_1, x_2, \ldots, x_{3n}) =$

$$((x_1, x_2, x_3), (x_4, x_5, x_6), \ldots, (x_{3n-2}, x_{3n-1}, x_{3n})).$$

## Rigid motions

- A translation in $\mathbb{R}^3$ is a function $T_{\vec{v}}: \mathbb{R}^3 \to \mathbb{R}^3$ given by $T_{\vec{v}}(\vec{x}) = \vec{x} + \vec{v}$ for some fixed $\vec{v} \in \mathbb{R}^3$

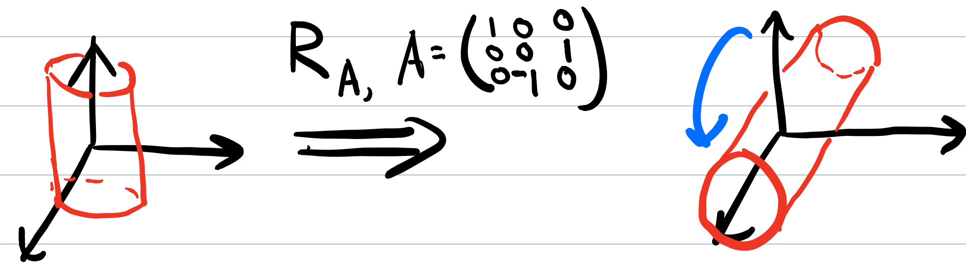



Interpretation: $T_{\vec{v}}$ shifts a geometric object in the direction $\vec{v}$ without rotating.

- A rotation in $\mathbb{R}^3$ is a function $R_A: \mathbb{R}^3 \to \mathbb{R}^3$ of the form

$R_A(\vec{x}) = A\vec{x}$ where $A$ is a $3\times 3$ matrix with determinant $1$

<u>**Interpretation**</u>: $R_A$ rotates a geometric object about the origin in $\mathbb{R}^3$.



$$R_A,\ A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}$$

<u>**Note**</u>: Both translations and rotations are invertible:

$$(T_{\vec{v}})^{-1} = T_{-\vec{v}}. \quad (R_A)^{-1} = R_{A^{-1}}.$$

<u>**Definition**</u>:

A <u>rigid motion</u> in $\mathbb{R}^3$ is a rotation followed by a translation, i.e., a function

$$\varphi : \mathbb{R}^3 \to \mathbb{R}^3 \text{ of the form}$$

$$\varphi = \underset{\substack{\uparrow \\ \text{translation}}}{T_{\vec{v}}} \circ \underset{\substack{\uparrow \\ \text{rotation}}}{R_A}.$$

**Facts:**
- The inverse of a rigid motion is a rigid motion
- The composition of two rigid motions is a rigid motion.

Let $E$ be the set of all rigid motions in $\mathbb{R}^3$.

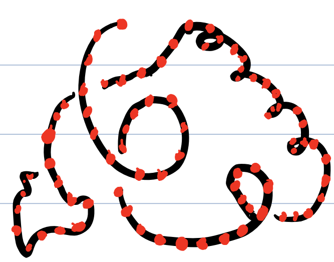
Define $RMSD : O^n \times O^n \to [0, \infty)$ by

$$RMSD(P, P') = \min_{\varphi \in E} \frac{1}{\sqrt{n}} d_2\big(V(P), V(\underbrace{\varphi(P')}_{\text{rigid motion of } P'})\big).$$

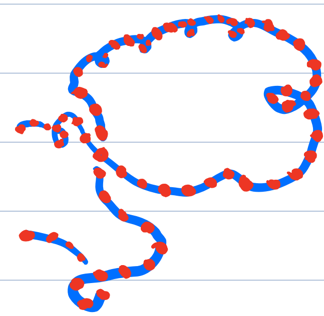
ordinary Euclidean distance

**Interpretation:** To compute $RMSD(P, P')$,

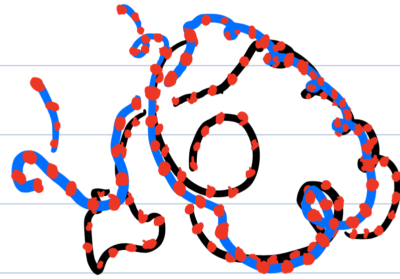
1) Align $P$ and $P'$ as well as possible via a rigid motion $\varphi$



P



P'



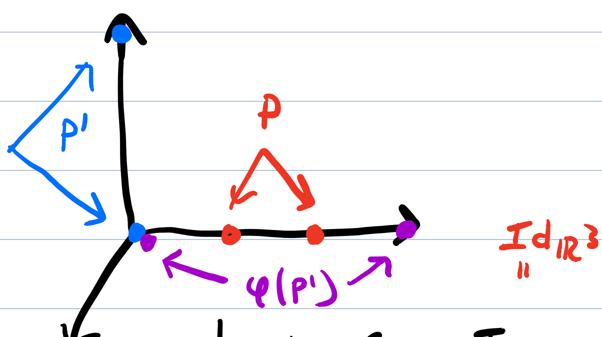
P and $\varphi(P')$

2) Represent P and $\varphi(P')$ as points $V_P, V_{\varphi(P')}$ in $\mathbb{R}^{3n}$.

3) RMSD is the Euclidean distance between these points, normalized so that RMSD doesn't tend to grow as # of atoms grows.

<u>Example</u> $P = ((1,0,0),(2,0,0)) \in O^2$

$$P' = ((0,0,0),(0,0,3)) \in O^2.$$




Let $\varphi = T_{(0,0,0)} \circ R_A = R_A$, where $A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix}$

Then $\varphi(P') = \left(A\begin{pmatrix}0\\0\\0\end{pmatrix}, A\begin{pmatrix}0\\0\\3\end{pmatrix}\right) =$

$((0,0,0),(3,0,0))$

It can be shown that $\varphi$ provides the optimal alignment of $P$ and $P'$, i.e.,

$\varphi$ is the minimizing rigid motion in the expression for RMSD.

Then $\mathrm{RMSD}(P,Q) = \frac{1}{\sqrt{2}}\, d_2((1,0,0,2,0,0),(0,0,0,3,0,0))$

$= \frac{1}{\sqrt{2}} \cdot \sqrt{(1-0)^2+(2-3)^2} = \frac{1}{\sqrt{2}} \cdot \sqrt{1^2+1^2} = 1.$

RMSD is symmetric and satisfies the triangle inequality, but we can have

$\mathrm{RMSD}(P,P') = 0$ if $P \neq P'$ but $\varphi(P) = P'$ for some rigid motion $\varphi$. } So property 1 of a metric is not satisfied.

**Example:** $P = ((1,0,0),(2,0,0))$
$P' = ((0,0,1),(0,0,2))$

Then for $\varphi : \mathbb{R}^3 \to \mathbb{R}^3$ as in the previous example,

$\varphi(P') = P \Rightarrow \mathrm{RMSD}(P, P') = 0.$

Here's how we get a genuine metric:

Define an equivalence relation $\sim$ on $O^n$ by

$P \sim Q$ iff $\exists$ a rigid motion $\varphi : \mathbb{R}^3 \to \mathbb{R}^3$ with $\varphi(P) = Q$.

*Fact: $\mathrm{RMSD}(P,Q) = \mathrm{RMSD}(P',Q')$ if $P \sim P'$ and $Q \sim Q'$

(Exercise: Prove this).

As a consequence, $\mathrm{RMSD} : O^n \times O^n \to [0,\infty)$ descends to a genuine metric on $O^n/\sim$.

Specifically, we define

$$\overline{\mathrm{RMSD}} : O^n/\sim \times O^n/\sim \longrightarrow [0,\infty) \text{ by}$$

$$\overline{\mathrm{RMSD}}([P],[Q]) = \mathrm{RMSD}(P,Q).$$

By the fact* this function is well defined.

It can be shown that $\overline{RMSD}$ is a metric.

**Note:** RMSD can be computed efficiently, even for large examples (on a computer).

This is an example of an optimization problem.

## Metrics and topology

Metric space definition of continuity:

Let $M$ and $N$ be metric spaces with metrics $d_M, d_N$.

A function $f: M \to N$ is continuous at $x \in M$ if $\forall \epsilon > 0, \exists\, \delta > 0$ such that

$$d_M(x,y) < \delta \Rightarrow d_N(f(x), f(y)) < \epsilon.$$

$f$ is said to be continuous if it is continuous at each $x \in M$.

(This definition generalizes the definition for Euclidean subspaces considered earlier).

**Example:** Let $M$ be any metric space and take $N$ to be $\mathbb{R}$ with the Euclidean metric.

For any $x \in M$, the function $d^x : M \to \mathbb{R}$ given by $d^x(y) = d_M(x,y)$ is a continuous function.

<u>Pf</u>: Exercise.

With this definition of continuity, the definition of homeomorphism extends immediately to metric spaces:

For metric spaces $M$ and $N$, $f : M \to N$ is a <u>homeomorphism</u> if

1) $f$ is a continuous bijection
2) $f^{-1}$ is also continuous.

<u>Example</u>: Consider the metric $d$ on $[0, 2\pi)$ given by $d(x,y) = \min(|x-y|, |(x+2\pi)-y|, |(x-2\pi)-y|)$


take $S^1$ to have usual Euclidean metric

Then the function $f : ([0, 2\pi), d) \to S^1$ given by $f(t) = (\cos t, \sin t)$ is a homeomorphism.

The definition of isotopy also extends, but we'll not get into the details of this.

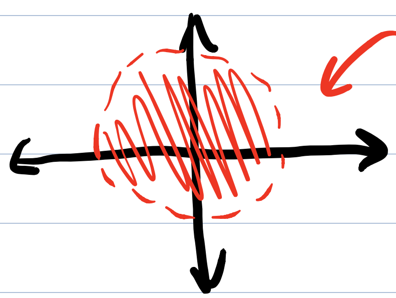
## An alternate discription of continuity

### Open Sets

Let $M$ be a metric space. For $x \in M$ and $r > 0$, the open ball in $M$ of radius $r$, centered at $x$, is the set

$$B(x,r) = \{ y \in M \mid d_M(x,r) < r \}.$$

Example: For $M = \mathbb{R}^2$ with the Euclidean distance. $B(\vec{0}, 1)$ looks like this



disc of radius 1 centered at the origin, with the boundary not included.

A subset of $M$ is called open if it is a union of (possibly infinitely many) open balls.

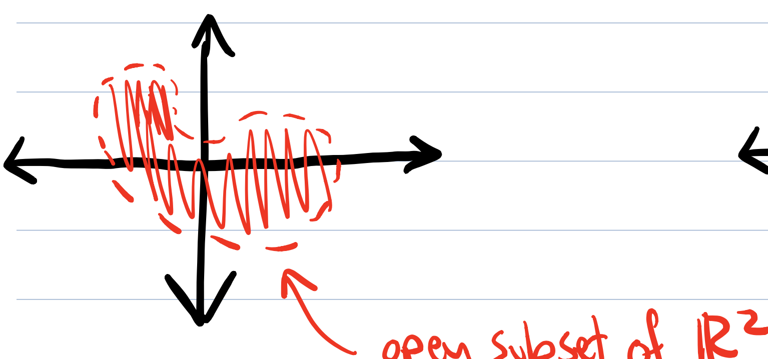
The empty set is always considered open.

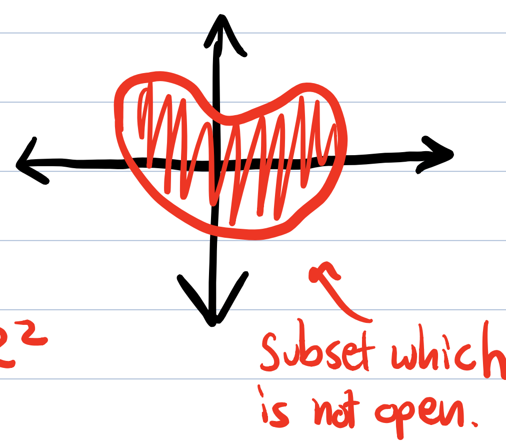
M itself is open: $M = \bigcup_{x \in M} B(x,1)$

**Fact**: A region in $\mathbb{R}^n$ is open if it contains none of its $\underbrace{\text{boundary points}}$.

this is an informal statement because I haven't defined "boundary points." It can be made formal, but I will not go into the details.

**Illustration**: Dashed line = boundary not included
Solid line = boundary included




**Fundamental Fact**: Whether a function of of metric spaces $f: M \to N$ is continuous depends only on the open sets of $M, N$ and not on otherwise on the metric!

<u>Def</u>: For $f: S \to T$ any function and $U \subset T$, $f^{-1}(U) = \{x \in S \mid f(x) \in U\}$.

<u>Proposition</u>: A function $f: M \to N$ of metric spaces is continuous if and only if $f^{-1}(U)$ is open for every open subset of $N$.