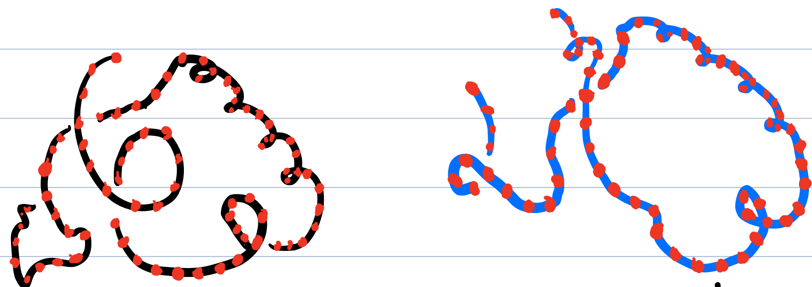
# AMAT 583 Lecture 13 10/8/19

**Today:** RMSD continued
Metrics spaces and topology
Open sets and continuity.

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



P P'

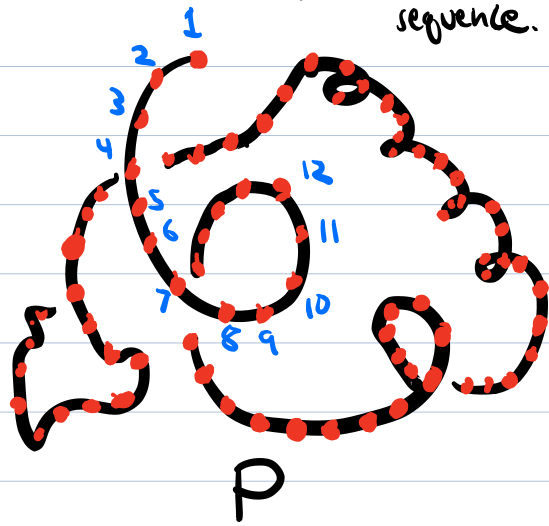
review

**Standard Answer:** Compute a metric called **RMSD** (root mean squared deviation) between P and P'.

RMSD is a fundamental tool in the study of molecules.

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

- Fix an order on the atoms of the amino acid sequence. (Choice of order doesn't matter).




- Let $O^n$ denote the set of all ordered subsets of $\mathbb{R}^3$ of size $n$.
- We represent the 3-D structure of a protein as an element of $O^n$.
- For $P \in O^n$, denote the $i^{th}$ point in $S$ by $(x_i, y_i, z_i)$
- Define a function $V: O^n \to \mathbb{R}^{3n}$ by

  $$V(P) = (x_1, y_1, z_1, x_2, y_2, z_2, \ldots, x_n, y_n, z_n).$$

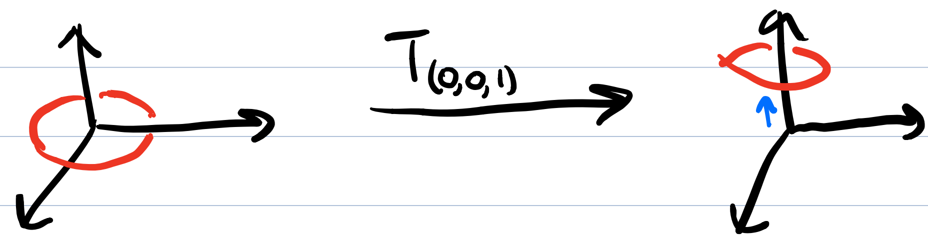
  $\boxed{V \text{ is invertible!}}$

This represents the proteins' 3-D structure as a single point in a high-dimensional space!

Note: This representation throws away a lot of info about the protein (atom type, bond info), but for many applications that is ok.

## 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 $\vec{v} \in \mathbb{R}^3$

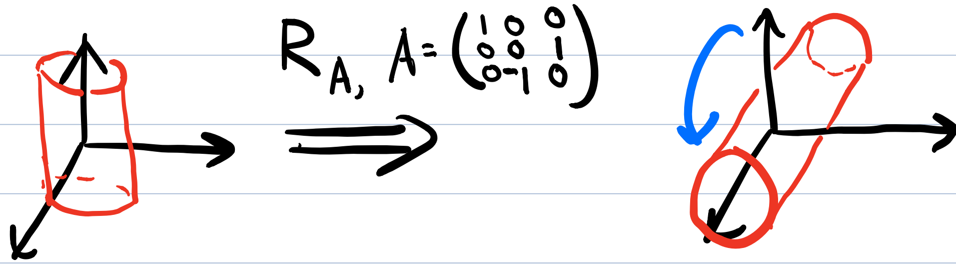



Interpretation: $T_{\vec{v}}$ shifts a geometric object in the direction $\vec{v}$ without rotating.

- 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 $1$

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




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

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

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

> Note: A rigid motion $\varphi$ is invertible, and $\varphi^{-1}$ is also a rigid motion.

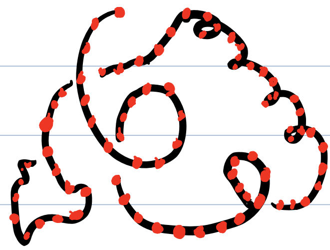
Let $E$ be the set of all rigid motions in $\mathbb{R}^3$.

**Definition:** Let $P$, $P'$ be 3-D structures for a given protien with $n$ atoms, regared as subsets of $\mathbb{R}^3$ of size $n$.

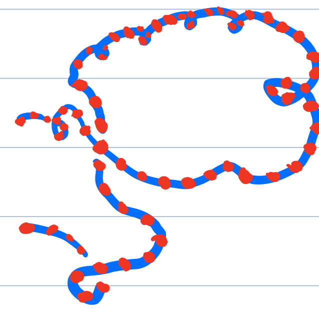
$$\mathrm{RMSD}(P, P') = \min_{\varphi \in E} \frac{1}{\sqrt{n}} \underset{\substack{\text{ordinary} \\ \text{Euclidean} \\ \text{distance}}}{d_2}\Big(V(P), \underbrace{V(\varphi(P'))}_{\text{rigid motion of } P'}\Big).$$

**Interpretation:** To compute $\mathrm{RMSD}(P, P')$,

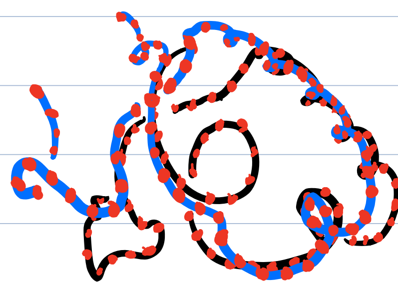
1) Align P and P' as well as possible via a rigid motion $\varphi$



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.

Formally, we regard this as a function

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

This function 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.

Here's how we get a genuine metric here:

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

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

<u>Fact</u>: $\mathrm{RMSD}(P,Q) = \mathrm{RMSD}(P',Q')$ if $P \sim P'$ and $Q \sim Q'$

(<u>Exercise</u>: Prove this).

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

Specifically, we define

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

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

By the fact*, this function is well defined.

<u>Exercise</u>: Prove that $\overline{RMSD}$ is a metric.

## <u>Metrics and topology</u>

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 <u>continuous</u> 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.

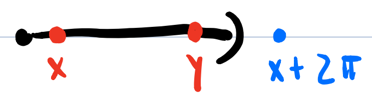
For any $x \in M$, the function $d^x : M \to \mathbb{R}$ given by $d^x(y) = d_M(x,y)$ is a continuous function.

**Pf:** 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 **homeomorphism** if
1) $f$ is a continuous bijection
2) $f^{-1}$ is also continuous.

**Example:** 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.

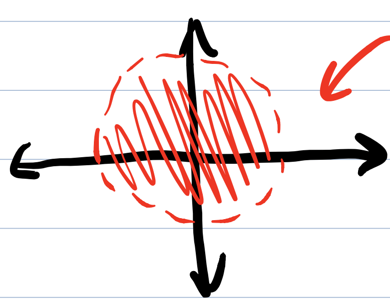
## <u>An alternate discription of continuity</u>

### <u>Open Sets</u>

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 <u>open</u> if it is a union of (possibly infinitely many) open balls.

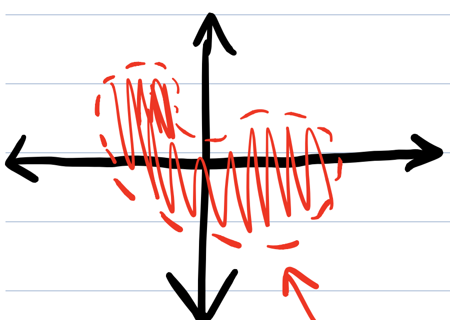
The <u>empty set</u> is always considered open.

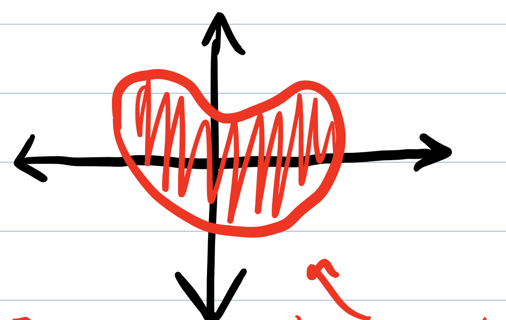
M itself is open: $M = \bigcup_{x \in M} B(x,1)$

<u>Fact</u>: A region in $\mathbb{R}^n$ is open if it contains none of its 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.

<u>Illustration</u>: 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!

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

**Proposition**: 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 $N$.