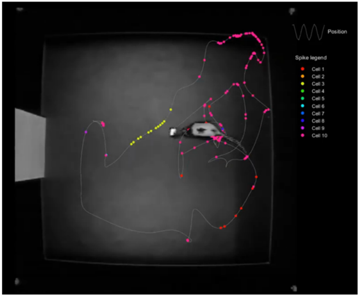
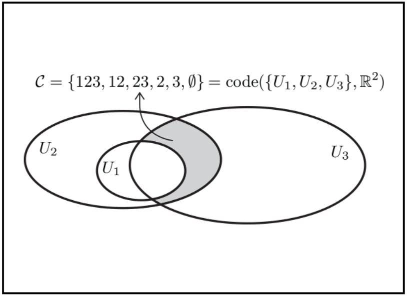
Sunflower Theorems and Convex Codes

Amzi Jeffs
University of Washington, Seattle

UC Davis Algebra & Discrete Mathematics Seminar
April 6th 2020

Part I: Place Cells and Convex Codes

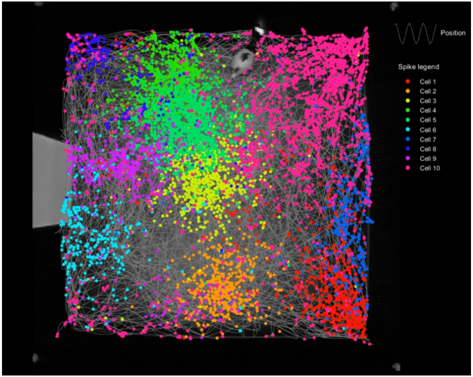




Place Cells

1971: O'Keefe and Dostrovsky describe *place cells* in the hippocampus of rats.

2019 video: `https://youtu.be/puCV1grkdJA`



Main idea: Each place cell fires in a particular region. They "know" where the rat is. How much do they know?

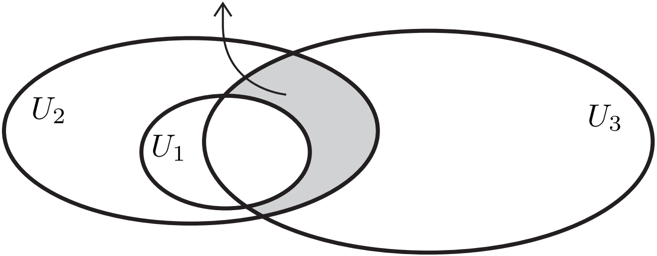
Mathematical Model of Place Cells

2013: Curto et al introduce *convex neural codes*.

- Index your place cells (neurons) by $[n] \stackrel{\text{def}}{=} \{1, 2, \ldots, n\}$.
- Each neuron $i \in [n]$ fires when rat is in convex open U_i in $\mathbb{R}^d$.
- As rat moves, multiple neurons may fire at same time. Write down all the sets of neurons that fire together and get a *convex neural code* $\mathcal{C} \subseteq 2^{[n]}$.

Example with 3 neurons in $\mathbb{R}^2$:

$$\mathcal{C} = \{123, 12, 23, 2, 3, \emptyset\}$$



Formal Definitions

Definition

A *code* is any subset of $2^{[n]}$. Elements of a code are *codewords*.

Definition

Let $\mathcal{U} = \{U_1, \ldots, U_n\}$ be a collection of convex open sets in $\mathbb{R}^d$. The *code* of $\mathcal{U}$ is

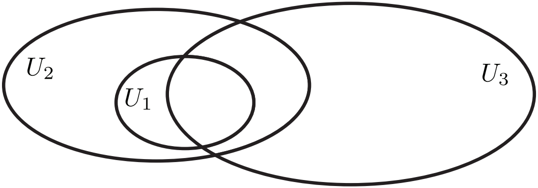
$$\operatorname{code}(\mathcal{U}) \stackrel{\text{def}}{=} \left\{ \sigma \subseteq [n] \,\middle|\, \text{There is } p \in \mathbb{R}^d \text{ with } p \in U_i \Leftrightarrow i \in \sigma \right\}.$$

The collection $\mathcal{U}$ is called a *convex realization*. Codes that have convex realizations are called *convex*.

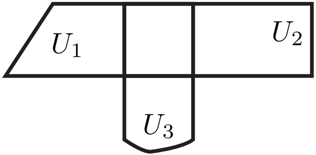
Notice! If the U_i correspond to place cells, we can compute $\operatorname{code}(\mathcal{U})$ from the brain directly (if the rat explores sufficiently...)

Terminology in Practice

Below, $\mathcal{U} = \{U_1, U_2, U_3\}$ is a convex realization of $\mathcal{C} = \{123, 12, 23, 2, 3, \emptyset\}$. Therefore $\mathcal{C}$ is a convex code.



Can you realize $\{123, 1, 2, 3, \emptyset\}$ in $\mathbb{R}^2$? How about $\mathbb{R}^1$?



Cannot be realized in $\mathbb{R}^1$!

The Big Picture

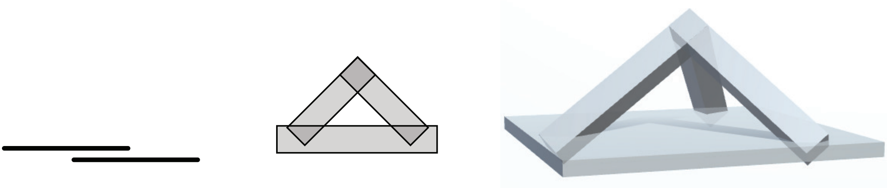
Question: When can we find a convex realization of a given $\mathcal{C}$?
How much does $\mathcal{C}$ tell us about this realization?

Motivation: How much can place cells "know" about a rat's environment?

This Talk: How much can n place cells tell us about the *dimension* of the space that the rat is perceiving?

Motivation: Higher dimensional stimulus like smell, sound, etc.
Also, interesting discrete geometry theory!

Part II: Open Embedding Dimension and Past Results



Formalizing the Dimension Question

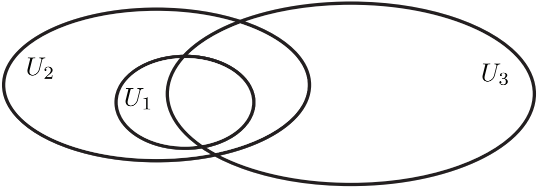
Definition

Let $\mathcal{C} \subseteq 2^{[n]}$ be a code. The *open embedding dimension* of $\mathcal{C}$ is

$$\text{odim}(\mathcal{C}) \overset{\text{def}}{=} \min\{d \mid \mathcal{C} \text{ has a convex realization in } \mathbb{R}^d\}.$$

When no realization exists $\text{odim}(\mathcal{C}) = \infty$.

Example: Recall we can realize $\mathcal{C} = \{123, 12, 23, 2, 3, \emptyset\}$ in $\mathbb{R}^2$. In fact, $\text{odim}(\mathcal{C}) = 1$ since we can "flatten" our realization.



Formalizing the Dimension Question

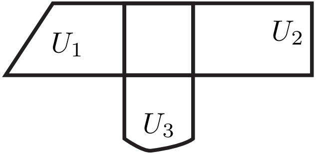
Definition

Let $\mathcal{C} \subseteq 2^{[n]}$ be a code. The *open embedding dimension* of $\mathcal{C}$ is

$$\operatorname{odim}(\mathcal{C}) \stackrel{\text{def}}{=} \min\{d \mid \mathcal{C} \text{ has a convex realization in } \mathbb{R}^d\}.$$

When no realization exists $\operatorname{odim}(\mathcal{C}) = \infty$.

Example: Recall that $\mathcal{C} = \{123, 1, 2, 3, \emptyset\}$ has a realization in $\mathbb{R}^2$ but not $\mathbb{R}^1$. Thus $\operatorname{odim}(\mathcal{C}) = 2$ in this case.



Formalizing the Dimension Question

Definition

Let $\mathcal{C} \subseteq 2^{[n]}$ be a code. The *open embedding dimension* of $\mathcal{C}$ is

$$\operatorname{odim}(\mathcal{C}) \stackrel{\text{def}}{=} \min\{d \mid \mathcal{C} \text{ has a convex realization in } \mathbb{R}^d\}.$$

When no realization exists $\operatorname{odim}(\mathcal{C}) = \infty$.

Example: Let $\mathcal{C} = \{12, 23, \emptyset\}$. In any realization, U_2 is the disjoint union of U_1 and U_3. This is impossible with open sets. Thus $\operatorname{odim}(\mathcal{C}) = \infty$.

Past Results

- **[Curto et al. 17]** Many examples with $\mathrm{odim}(\mathcal{C}) = \infty$ based on topological arguments.
- **[CGIK16] Upper bound:** If $\mathcal{C}$ is convex then $\mathrm{odim}(\mathcal{C}) \leq$ number of inclusion-maximal codewords.
- **Lower bounds:** None larger than n. For example, [CV16] looks for nonzero Betti numbers.

Definition

$\mathcal{C}$ is *intersection complete* if $c_1, c_2 \in \mathcal{C}$ implies $c_1 \cap c_2 \in \mathcal{C}$.

- **[CGIK16]** If $\mathcal{C}$ is intersection complete, $\mathrm{odim}(\mathcal{C}) < \infty$.

This talk: Among all intersection complete codes $\mathcal{C} \subseteq 2^{[n]}$, what is the largest embedding dimension (in terms of n)?

Topological intuition and "general position" reasoning indicate the answer should be n...

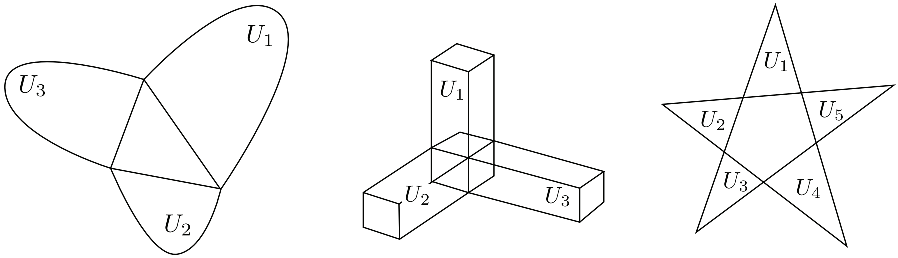
Part III: Sunflowers of Convex Open Sets

Sunflowers of Convex Open Sets

Notation: If $\sigma \subseteq [n]$, then $U_\sigma = \bigcap_{i \in \sigma} U_i$.

Definition

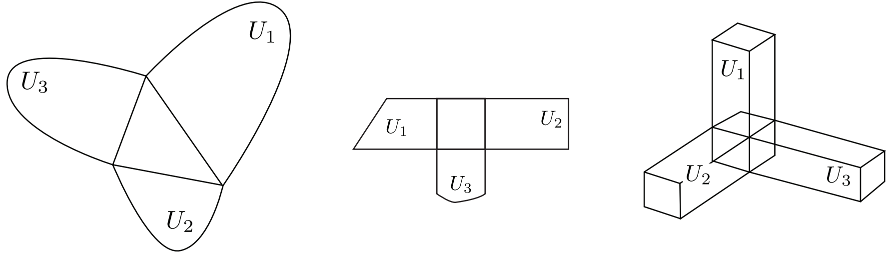
$\mathcal{U}$ is a *sunflower* if $\text{code}(\mathcal{U}) = \{[n], \text{singletons}, \emptyset\}$. The U_i are called *petals* and $U_{[n]}$ is called the *center*.



Rigidity in Sunflowers

Lemma (LSW15)

If $\mathcal{U} = \{U_1, U_2, U_3\}$ *is a sunflower in* $\mathbb{R}^2$*, and* $p_i \in U_i$*, then* $\operatorname{conv}\{p_1, p_2, p_3\}$ *contains a point in the center of* $\mathcal{U}$*.*
This fails in $\mathbb{R}^3$*.*



Consequence: $\mathcal{C}$ with $\operatorname{odim}(\mathcal{C}) = \infty$ for *non-topological* reasons.

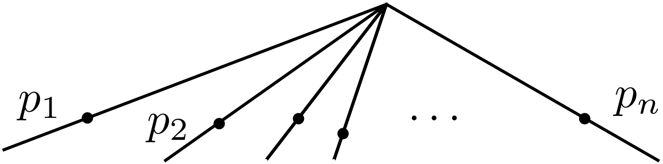
Rigidity in General

Theorem (J18)

If $\mathcal{U} = \{U_1, U_2, \ldots, U_{d+1}\}$ is a sunflower in $\mathbb{R}^d$, and $p_i \in U_i$, then $\operatorname{conv}\{p_1, p_2, \ldots, p_{d+1}\}$ contains a point in the center of $\mathcal{U}$.
This fails in $\mathbb{R}^{d+1}$.

A word on the proof: Original proof quite messy. A nicer proof uses Radon's theorem, as suggested by Zvi Rosen.

Note: Result fails badly when the sets are not open.



Application to Codes

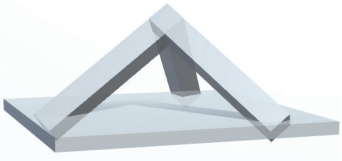
Theorem (J19)

Let $\mathcal{S}_n \subseteq 2^{[n+1]}$ be the code whose codewords are every singleton, every 2-set containing $n+1$, the empty set, and $[n]$. Then $\operatorname{odim}(\mathcal{S}_n) = n$.

Proof.

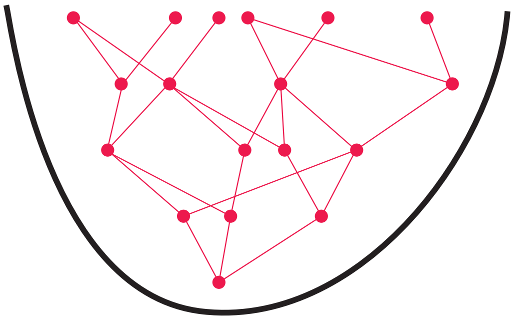
The sets $U_1, \ldots, U_n$ form a sunflower. The set U_{n+1} touches every petal but not the center. Impossible when $d < n$. On the other hand, can get a realization in $\mathbb{R}^n$ by thickening coordinate axes and hyperplane with normal vector $\mathbf{1}$. $\square$

Example: $\mathcal{S}_3$ in $\mathbb{R}^3$



Still smaller than number of neurons. **Where's the surprise?**

Part IV: Code Minors and Exponential Embedding Dimension



Morphisms of Codes

Definition (J18)

Let $\mathcal{C} \subseteq 2^{[n]}$ and $\sigma \subseteq [n]$. The *trunk* of σ in $\mathcal{C}$ is

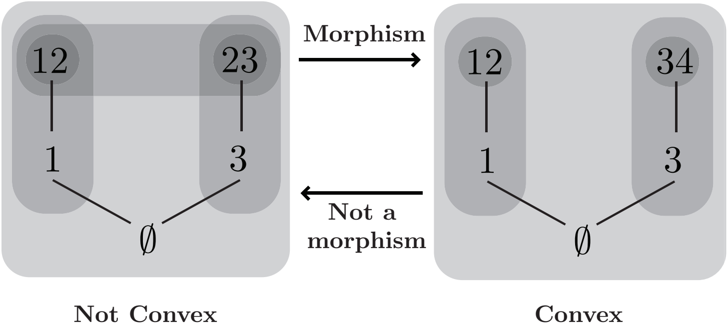
$$\mathrm{Tk}_{\mathcal{C}}(\sigma) \stackrel{\text{def}}{=} \{c \in \mathcal{C} \mid \sigma \subseteq c\}.$$

Geometric note: Codewords in $\mathrm{Tk}_{\mathcal{C}}(\sigma)$ come from points in U_σ.

Definition (J18)

A function $f : \mathcal{C} \to \mathcal{D}$ is called a *morphism* if the preimage of every trunk in $\mathcal{D}$ is a trunk in $\mathcal{C}$.

Morphism Example



Note: Morphisms should not be thought of as simply “continuous,” the trunks here generate the same topology.

Minors of Codes

Definition (J18)

We say $\mathcal{C}$ is a *minor* of $\mathcal{D}$ if there is a surjective morphism $f : \mathcal{D} \to \mathcal{C}$. Codes can be partially ordered by "is a minor of" in a poset $\mathbf{P_{Code}}$.

Note: We are really partially ordering *isomorphism classes*.

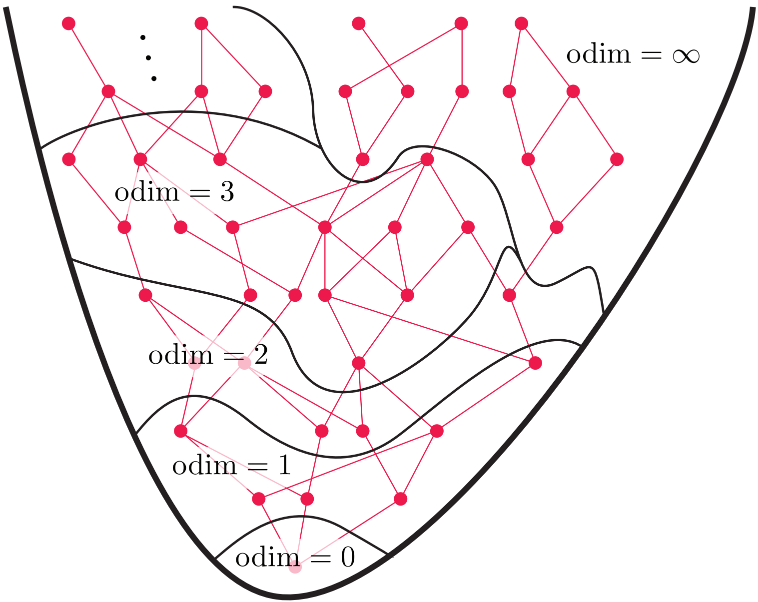
Theorem (J18)

If $\mathcal{C}$ *is a minor of* $\mathcal{D}$*, then* $\operatorname{odim}(\mathcal{C}) \leq \operatorname{odim}(\mathcal{D})$.

Proof idea: Every morphism $f : \mathcal{D} \to \mathcal{C}$ corresponds to a collection of trunks in $\mathcal{D}$. These correspond geometrically to various U_σ, and it turns out $\mathcal{C}$ is the code of these various U_σ.

Main point: $\mathbf{P_{Code}}$ is totally combinatorial, but its structure can inform us about geometry, namely $\operatorname{odim}(\mathcal{C})$.

$\mathbf{P}_{\mathbf{Code}}$ Visually

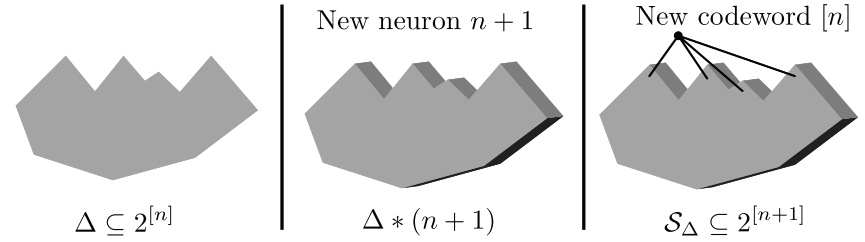


Back to Sunflowers

Definition (J19)

Let $\Delta \subseteq 2^{[n]}$ be a simplicial complex. Define a code $\mathcal{S}_\Delta \subseteq 2^{[n+1]}$ by

$$\mathcal{S}_\Delta \overset{\text{def}}{=} \big(\Delta * (n+1)\big) \cup \{[n]\}.$$



Theorem (J19)

If Δ has m facets, then $\mathcal{S}_m$ is a minor of $\mathcal{S}_\Delta$. Moreover, $\mathcal{S}_\Delta$ is intersection complete.

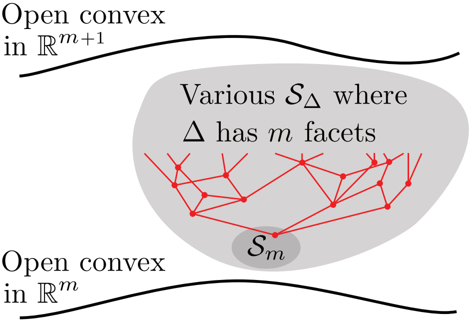
Embedding Dimension for $\mathcal{S}_\Delta$

Corollary (J19)

If Δ has m facets, then $\operatorname{odim}(\mathcal{S}_\Delta) = m$.

Proof.

Results of [CGIK16] let us construct a realization of $\mathcal{S}_\Delta$ in $\mathbb{R}^m$.
On the other hand, $\operatorname{odim}(\mathcal{S}_\Delta) \geq \operatorname{odim}(\mathcal{S}_m) = m$. $\square$

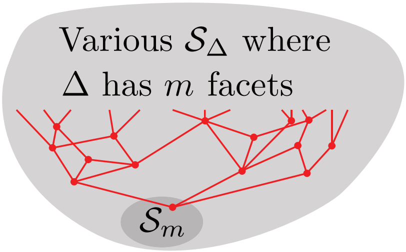


Exponential Embedding Dimension

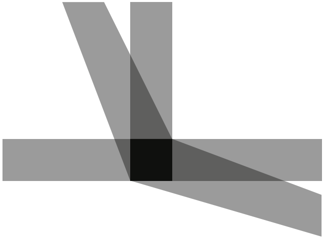
Corollary (J19)

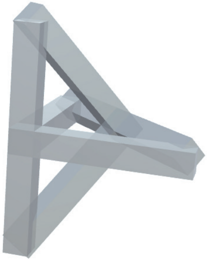
Open embedding dimension may grow ***exponentially*** *in terms of number of neurons n, as large as $\binom{n-1}{\lfloor (n-1)/2 \rfloor}$. Surprising!*

What's behind this? Number of neurons varies wildly throughout layers of $\mathbf{P}_{\mathbf{Code}}$.



Part V: A Generalization and an Open Problem





Beyond Sunflowers

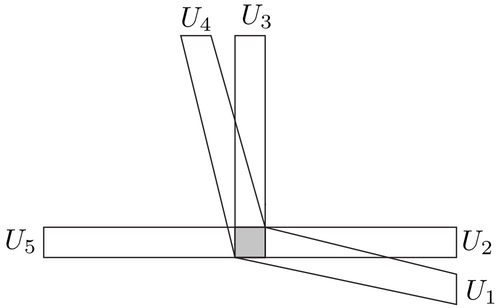
Definition

$\mathcal{U} = \{U_1, \ldots, U_n\}$ is called a *k-flexible sunflower* if

$$\text{code}(\mathcal{U}) = \{[n], \text{codewords containing} \leq k \text{ neurons}, \emptyset\}.$$

Note: When $k = 1$, these are usual sunflowers.

Example: A 2-flexible sunflower with 5 petals in $\mathbb{R}^2$.



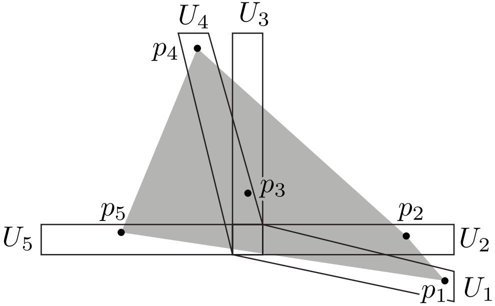
Flexible Sunflower Theorem

Theorem (J19)

Let $\mathcal{U} = \{U_1, \ldots, U_n\}$ be a k-flexible sunflower in $\mathbb{R}^d$. For $i \in [n]$, let $p_i \in U_i$. Then $\operatorname{conv}\{p_1, \ldots, p_n\}$ contains a point in the center of $\mathcal{U}$ as long as $n \geq dk + 1$.

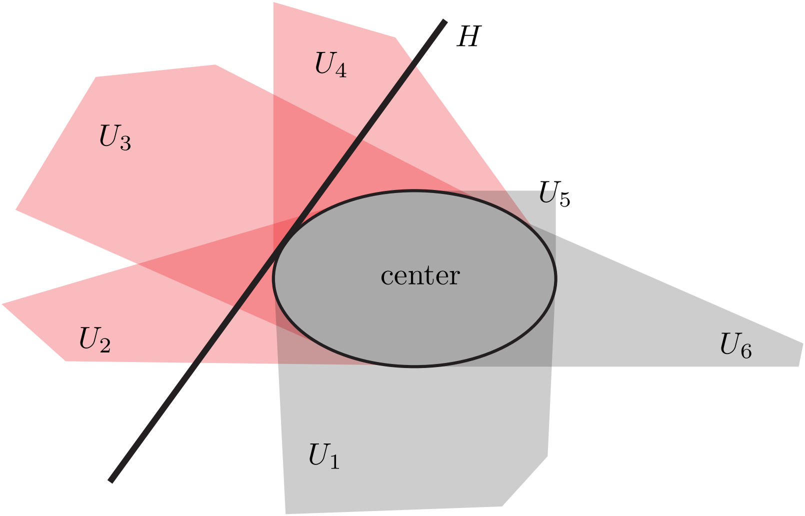
About the proof: Uses Tverberg's theorem. When $k = 1$ Tverberg's theorem becomes Radon's theorem, and we have the original sunflower result.

Example:



What's Special About Open Sets?

Key observation: Supporting hyperplanes on the center of a k-flexible sunflower "cut off" at most k petals. Below $k = 3$.



Tangled Sunflower Codes

Definition (J19)

Let $\mathcal{T}_n \subseteq 2^{[2n]}$ be the code

$$\mathcal{T}_n \stackrel{\text{def}}{=} \{\{\text{odds}\}, \{\text{evens}\}, 12, 34, \ldots, (2n-1)(2n), \text{all singletons}, \emptyset\}.$$

That is, $\mathcal{T}_n$ describes two n-petal sunflowers with touching petals. Define $t_n \stackrel{\text{def}}{=} \text{odim}(\mathcal{T}_n)$.

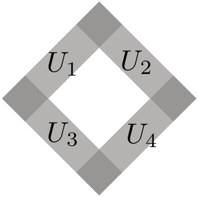
Examples:

- $\mathcal{T}_1 = \{12, 1, 2, \emptyset\}$.
- $\mathcal{T}_2 = \{13, 24, 12, 34, 1, 2, 3, 4, \emptyset\}$.
- $\mathcal{T}_3 = \{135, 246, 12, 34, 56, 1, 2, 3, 4, 5, 6, \emptyset\}$.

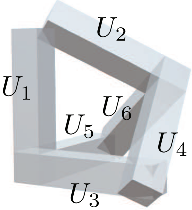
What can we say about t_n?

For small n, we can find t_n by brute force:

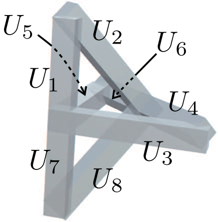
$t_1 = 1$



$t_2 = 2$



$t_3 = 3$



$t_4 = 3$

Moreover, $t_5 = 4$. If you can figure out t_6, **let me know!**

Theorem (J19)

For all n:

- $t_n \leq t_{n+1} \leq t_n + 1$, *and*
- $\lceil \frac{n}{2} \rceil \leq t_n \leq n$.

References

This talk is based on [J19]:

“Embedding Dimension Phenomena in Intersection Complete Codes”
https://arxiv.org/abs/1909.13406

Other papers referenced:

- **[CGIK16]** “On open and closed convex codes”
  https://arxiv.org/abs/1609.03502
- **[Curto et al 13]** “The neural ring: an algebraic tool for analyzing the intrinsic structure of neural codes”
  https://arxiv.org/abs/1212.4201
- **[Curto et al 17]** “What makes a neural code convex?”
  https://arxiv.org/abs/1508.00150
- **[LSW15]** “Obstructions to convexity in neural codes”
  https://arxiv.org/abs/1509.03328

Part VI (Bonus): Closed Convexity and More

Closed Convexity

Definition

We may consider realizations using closed convex sets, and define *closed embedding dimension* $\mathrm{cdim}(\mathcal{C})$ for codes.

Theorem (J19)

For all examples in this talk $\mathcal{C} \subseteq 2^{[n]}$, *we have* $\mathrm{cdim}(\mathcal{C}) \leq n$.

Open question: Is $\mathrm{cdim}(\mathcal{C})$ ever larger than n?

Theorem (J19)

For intersection complete codes, $\mathrm{cdim}(\mathcal{C}) \leq \mathrm{odim}(\mathcal{C})$.

Open question: Can you find $\mathcal{C}$ with $\mathrm{odim}(\mathcal{C}) < \mathrm{cdim}(\mathcal{C}) < \infty$?

The Terrain for Future Work

Theorem

If $\mathcal{C}$ is a minor of $\mathcal{D}$, then $\operatorname{cdim}(\mathcal{C}) \le \operatorname{cdim}(\mathcal{D})$.