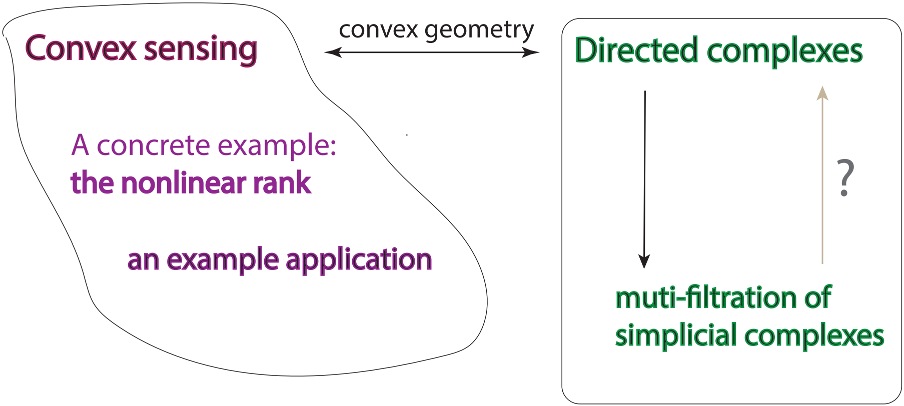
Directed complexes, non-linear rank and convex sensing.

Vladimir Itskov

The Pennsylvania State University

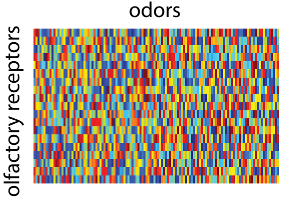
TAGS workshop
MPI, Leipzig, February 23, 2018

Plan of the talk:

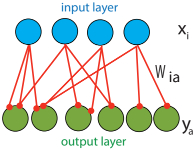


Two inspirations from neuroscience:

1) What is the dimension of the space of smells?

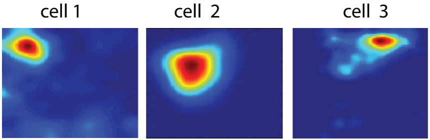


2) Feedforward networks



Why convex sensing?

Receptive fields of neurons:



Example: receptive fields of neurons in hippocampus

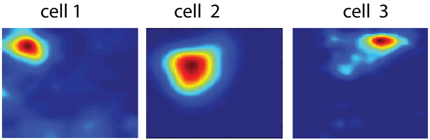
A receptive field is a function

$f : X \to \mathbb{R}$

on a relevant stimulus space X.

Why convex sensing?

Receptive fields of neurons:



Example: receptive fields of neurons in hippocampus

A receptive field is a function

$f : X \to \mathbb{R}$

on a relevant stimulus space X.

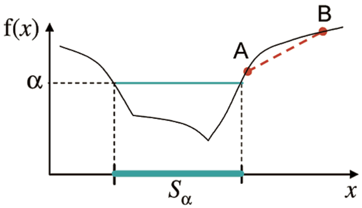
Hippocampal receptive fields are (approximately) quasiconcave.

I.e. the sets $U^\theta = f^{-1}([\theta, +\infty))$ are (approximately) convex.

a quick reminder:

Definition: A function $f : \mathbb{R}^d \to \mathbb{R}$ is quasiconvex if $\forall \alpha \in \mathbb{R}$

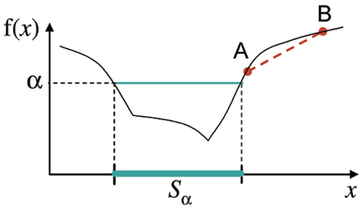
$$S_\alpha = f^{-1}((-\infty, \alpha]) \text{ is convex}$$



a quick reminder:

Definition: A function $f : \mathbb{R}^d \to \mathbb{R}$ is quasiconvex if $\forall \alpha \in \mathbb{R}$

$$S_\alpha = f^{-1}((-\infty, \alpha]) \text{ is convex}$$



Note: If

$$\begin{cases} \phi : \mathbb{R} \to \mathbb{R} \text{ is monotone increasing} \\ f : \mathbb{R}^d \to \mathbb{R} \text{ is quasiconvex} \end{cases} \implies \phi \circ f \text{ is quasiconvex}$$

Natural questions:

A neuroscience meta-observation:

Many/lots/most of brain areas have (classes of) neurons with convex receptive fields . . . [a long list of systems] . . .

Natural questions:

A neuroscience meta-observation:

Many/lots/most of brain areas have (classes of) neurons with convex receptive fields ... [a long list of systems] ...

Assume we do not know anything about the stimulus space X and the receptive fields, but can measure neuronal responses.

Natural questions:

A neuroscience meta-observation:

Many/lots/most of brain areas have (classes of) neurons with convex receptive fields ... [a long list of systems] ...

Assume we do not know anything about the stimulus space X and the receptive fields, but can measure neuronal responses.

Question(s):
Can we use the *neuronal responses* to tell

- the dimension of the underlying space X?
- topological features of the space X? ("easy")
- the (convex) geometry of the space X. (harder)

The dimension inference. . . (an ill-posed problem)

Given data: a matrix M_{ia}

Assumptions:

$$M_{ia} = f_i(x_a),$$

where f_i and x_a are unknown
$x_a \in \mathbb{R}^d$, and
$f_i : \mathbb{R}^d \to \mathbb{R}$ are quasiconvex.

The dimension inference. . . (an ill-posed problem)

Given data: a matrix M_{ia}

Assumptions:

$$M_{ia} = f_i(x_a),$$

where f_i and x_a are unknown
$x_a \in \mathbb{R}^d$, and
$f_i : \mathbb{R}^d \to \mathbb{R}$ are quasiconvex.

Question:

Find the minimal embedding dimension d.

The dimension inference. . . (an ill-posed problem)

Given data: a matrix M_{ia}

Assumptions:

$$M_{ia} = f_i(x_a),$$

where f_i and x_a are unknown
$x_a \in \mathbb{R}^d$, and
$f_i : \mathbb{R}^d \to \mathbb{R}$ are quasiconvex.

Question:

Find the minimal embedding dimension d.

Warning: For this to work one needs either

- "good sampling" of points x_a, or
- more assumptions of the functions f_i.

This page is intentionally left blank.

[use blackboard]

Quotienting by monotone functions

$\mathcal{G} \overset{\text{def}}{=} \{f : \mathbb{R} \to \mathbb{R}$ monotone increasing, surjective $\}$
$\mathcal{M}_{m,n} \overset{\text{def}}{=} \{m \times n$ matrices $\}$

The group $\mathcal{G}^m = \mathcal{G} \times \cdots \times \mathcal{G}$ acts on rectangular matrices $\mathcal{M}_{m,n}$.
For $g = (g_1, \ldots, g_m) \in \mathcal{G}^m$ and $M \in \mathcal{M}_{m,n}$,

$$(g \cdot M)_{ia} \overset{\text{def}}{=} g_i(M_{ia}).$$

Observation 1. All the recoverable information about a convex sensing problem is contained in the quotient $\dfrac{\mathcal{M}_{m,n}}{\mathcal{G}^m}$.

Quotienting by monotone functions

$\mathcal{G} \stackrel{\text{def}}{=} \{f : \mathbb{R} \to \mathbb{R} \text{ monotone increasing, surjective } \}$
$\mathcal{M}_{m,n} \stackrel{\text{def}}{=} \{m \times n \text{ matrices } \}$

The group $\mathcal{G}^m = \mathcal{G} \times \cdots \times \mathcal{G}$ acts on rectangular matrices $\mathcal{M}_{m,n}$.
For $g = (g_1, \ldots, g_m) \in \mathcal{G}^m$ and $M \in \mathcal{M}_{m,n}$,

$$(g \cdot M)_{ia} \stackrel{\text{def}}{=} g_i(M_{ia}).$$

Observation 1. All the recoverable information about a convex sensing problem is contained in the quotient $\dfrac{\mathcal{M}_{m,n}}{\mathcal{G}^m}$.

Observation 2. Let $\mathcal{M}^o_{m,n}$ denote matrices with distinct entries in each row, then

$$\frac{\mathcal{M}^o_{m,n}}{\mathcal{G}^m} = (\mathcal{S}_n)^m = \text{pure directed complexes on } n \text{ letters.}$$

directed complexes

A sequence in V is a tuple $s = (v_1, \ldots, v_k)$, where $v_j \in V$ do not repeat.

Definition: A directed complex is a (graded) poset of sequences in V, that contains all the subsequences.

directed complexes

A sequence in V is a tuple $s = (v_1, \ldots, v_k)$, where $v_j \in V$ do not repeat.

Definition: A directed complex is a (graded) poset of sequences in V, that contains all the subsequences.

A pure directed complex is a directed complex complex whose maximal sequences have *maximal possible* length $= |V|$.

directed complexes

A sequence in V is a tuple $s = (v_1, \ldots, v_k)$, where $v_j \in V$ do not repeat.

Definition: A directed complex is a (graded) poset of sequences in V, that contains all the subsequences.

A pure directed complex is a directed complex complex whose maximal sequences have *maximal possible* length $= |V|$.

Example: $M \in \mathcal{M}^o_{m,n}$, $\mathcal{D}(M) = \langle s_1, \ldots, s_m \rangle$ where each sequence $s_i = (v_{a_1}, v_{a_2}, \ldots, v_{a_n})$ is the total order on the i-th row:

$$M_{iv_{a_1}} < M_{iv_{a_2}} < \cdots < M_{iv_{a_n}}.$$

$$\text{e.g.} \qquad \mathcal{D}\begin{pmatrix} 10 & 20 & 30 & 40 \\ 11 & 13 & 14 & 12 \\ 1 & 4 & 2 & 3 \end{pmatrix} = \begin{matrix} \langle(1,2,3,4), \\ (1,4,2,3), \\ (1,3,4,2)\rangle \end{matrix}$$

directed complexes

A sequence in V is a tuple $s = (v_1, \dots, v_k)$, where $v_j \in V$ do not repeat.

Definition: A directed complex is a (graded) poset of sequences in V, that contains all the subsequences.

A pure directed complex is a directed complex complex whose maximal sequences have the same length $= |V|$.

directed complexes

A sequence in V is a tuple $s = (v_1, \dots, v_k)$, where $v_j \in V$ do not repeat.

Definition: A directed complex is a (graded) poset of sequences in V, that contains all the subsequences.

A pure directed complex is a directed complex complex whose maximal sequences have the same length $= |V|$.

Directed complexes have the usual bells and whistles:

- geometric representations
- homotopy type, homotopy equivalence
- homology $H_*(D)$
- pure directed complex $\mapsto$ a simplicial multi-filtration (via Dowker) .

A take-home message:

The topology of the directed complex $\mathcal{D}(M)$ carries meaningful information about convex sensing problems (e.g. embedding dimension).

A take-home message:

The topology of the directed complex $\mathcal{D}(M)$ carries meaningful information about convex sensing problems (e.g. embedding dimension).

Explicit example: The nonlinear rank problem.

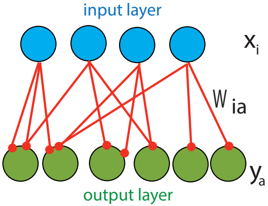
A take-home message:

The topology of the directed complex $\mathcal{D}(M)$ carries meaningful information about convex sensing problems (e.g. embedding dimension).

Explicit example: The nonlinear rank problem.

Here we replace quasiconvex with monotone $\circ$ linear . . .

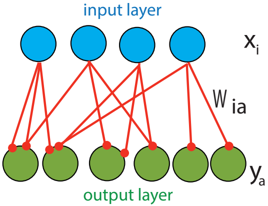
Motivation: one-layer feedforward neural network



$$y_i = \phi_i \left(\sum_{a=1}^{d} W_{ia} x_a - t_i \right),$$

$\phi_i : \mathbb{R} \to \mathbb{R}_{\geq 0}$ are monotone incr.

Motivation: one-layer feedforward neural network



$$y_i = \phi_i \left(\sum_{a=1}^{d} W_{ia} x_a - t_i \right),$$

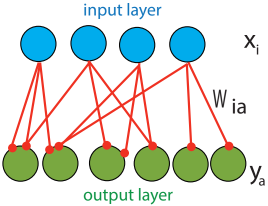
$\phi_i : \mathbb{R} \to \mathbb{R}_{\geq 0}$ are monotone incr.

Assume that all we know is the activity of the output layer, i.e. a collection of points $\{\vec{y}^{\alpha}\} \subseteq \mathbb{R}^n$

$$\vec{y}^{\alpha} = \vec{\Phi}\left(W\vec{x}^{\alpha} - \vec{t}\right) \in \mathbb{R}^n$$

.

Motivation: one-layer feedforward neural network



$$y_i = \phi_i \left(\sum_{a=1}^{d} W_{ia} x_a - t_i \right),$$

$\phi_i : \mathbb{R} \to \mathbb{R}_{\geq 0}$ are monotone incr.

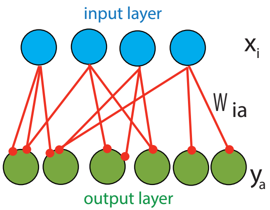
Assume that all we know is the activity of the output layer, i.e. a collection of points $\{\vec{y}^{\alpha}\} \subseteq \mathbb{R}^{n}$

$$\vec{y}^{\alpha} = \vec{\Phi}\left(W\vec{x}^{\alpha} - \vec{t}\right) \in \mathbb{R}^{n}$$

.

Question: Can we use the output of a one-layer network to tell the size d of the input layer?

A non-linear matrix factorization



$$y_i = \varphi_i \left(\sum_{a=1}^{d} W_{ia} x_a - t_i \right),$$

$\varphi_i : \mathbb{R} \to \mathbb{R}_{\geq 0}$ are monotone incr.

Assume that all we know is the activity of the output layer, i.e. a collection of points $\{\vec{y}^{\alpha}\}$, $\vec{y}^{\alpha} \in \mathbb{R}^n$.

Think of this as an $n \times m$ matrix $M = [\vec{y}^1, \ldots, \vec{y}^m] = \vec{\Phi}\,(WX - t)$.

Equivalently, find the minimal d so that the factorization

$$M_{i\alpha} = \phi_i \left(\sum_{a=1}^{d} W_{ia} X_a^{\alpha} \right),$$

is possible with some with monotone increasing $\phi_i : \mathbb{R} \to \mathbb{R}$.

Here the functions ϕ_i are *unknown*.

The non-linear rank

$\mathcal{G} \stackrel{\text{def}}{=} \{f : \mathbb{R} \to \mathbb{R} \text{ monotone increasing, surjective } \}$
$\mathcal{M}_{m,n} \stackrel{\text{def}}{=} \{m \times n \text{ matrices } \}$

The group $\mathcal{G}^m$ acts on rectangular matrices $\mathcal{M}_{m,n}$.
For $g = (g_1, \dots, g_m) \in \mathcal{G}^m$ and $M \in \mathcal{M}_{m,n}$,

$$(g \cdot M)_{ia} \stackrel{\text{def}}{=} g_i(M_{ia}).$$

Definition: The nonlinear rank of an $m \times n$ real-valued matrix $M \in \mathcal{M}_{m,n}$ is defined as the minimum rank of the matrices in the orbit of M, i.e.

$$\text{nrank}(M) \stackrel{\text{def}}{=} \min_{g \in \mathcal{G}^m} \text{rank}(g \cdot M).$$

Simple facts about "nonlinear rank":

- nrank(M) is determined by the ordering of each row. More precisely,

$\mathcal{M}^o_{m,n} \stackrel{\text{def}}{=} m \times n$ matrices without repeating entries in each row.

$$\frac{\mathcal{M}^o_{m,n}}{\mathcal{G}^m} = (\mathcal{S}_n)^m = \text{"pure directed complexes" on } n \text{ letters.}$$

directed complexes

A sequence in V is a tuple $s = (v_1, \ldots, v_k)$, where $v_j \in V$ do not repeat.

Definition: A directed complex is a (graded) poset of sequences in V, that contains all the subsequences.

A pure directed complex is a directed complex complex whose maximal sequences have the same length.

Example: Each $M \in \mathcal{M}^o_{m,n}$ yields a set of sequences s_i in $[n]$ s.t. each sequence $s_i = (v_{i1}, v_{i2}, \ldots, v_{in})$ is the total order on the i-th row:

$$M_{iv_{i1}} < M_{iv_{i2}} < \cdots < M_{iv_{in}}.$$

e.g. $$\mathcal{D}\begin{pmatrix} 10 & 20 & 30 & 40 \\ 11 & 13 & 14 & 12 \\ 1 & 4 & 2 & 3 \end{pmatrix} = \begin{matrix} \{(1,2,3,4), \\ (1,4,2,3), \\ (1,3,4,2)\} \end{matrix}$$

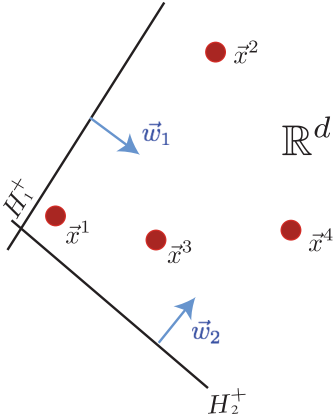
A geometric interpretation of $\mathcal{D}(M)$:

a non-linear matrix factorization $\mapsto$ a pure directed complex

$$\begin{pmatrix} * & * & * & * \\ * & * & * & * \end{pmatrix} \mapsto \begin{matrix} \langle(1,2,3,4), \\ (1,3,4,2)\rangle \end{matrix}$$

$$M_{i\alpha} = f_i\left(\vec{w}_i \cdot \vec{x}^{\alpha}\right)$$

The i-th sequence is the order in which the i-th plane encounters the points $\vec{x}^{\alpha}$.



Simple facts about “nonlinear rank”:

- nrank(M) is determined by the ordering of each row. More precisely,

$$\frac{\mathcal{M}^o_{m,n}}{\mathcal{G}^m} = (\mathcal{S}_n)^m = \text{pure directed complexes on } n \text{ letters.}$$

Simple facts about "nonlinear rank":

- nrank(M) is determined by the ordering of each row. More precisely,

$$\frac{\mathcal{M}^o_{m,n}}{\mathcal{G}^m} = (\mathcal{S}_n)^m = \text{pure directed complexes on } n \text{ letters.}$$

- nrank$(M) \leq \min\{n-1, m\}$ for any $m \times n$ matrix.

Simple facts about "nonlinear rank":

- nrank(M) is determined by the ordering of each row. More precisely,

$$\frac{\mathcal{M}^{o}_{m,n}}{\mathcal{G}^{m}} = (\mathcal{S}_n)^m = \text{pure directed complexes on } n \text{ letters.}$$

- nrank(M) $\leq \min\{n-1, m\}$ for any $m \times n$ matrix.

- There are combinatorial constraints that guarantee that nrank(M) $> d$ for any prescribed d.
  These come from convex geometry...

a quick example: a bound on nrank via Radon's Theorem

Observation: If $M_{i\alpha} = f_i\left(\vec{w}_i \cdot \vec{x}^{\alpha}\right)$ and $s = (v_1, v_2, \ldots, v_n) \in \mathcal{D}(M)$ then for every $l \in [n]$, the convex hulls do not intersect:

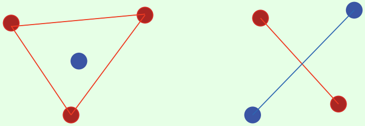
$$\operatorname{conv}\{v^1, v^2, \ldots, v^l\} \cap \operatorname{conv}\{v^{l+1}, v^{l+2}, \ldots v^n\} = \varnothing.$$

a quick example: a bound on nrank via Radon's Theorem

Observation: If $M_{i\alpha} = f_i\left(\vec{w}_i \cdot \vec{x}^{\alpha}\right)$ and $s = (v_1, v_2, \ldots, v_n) \in \mathcal{D}(M)$ then for every $l \in [n]$, the convex hulls do not intersect:

$$\mathrm{conv}\{v^1, v^2, \ldots, v^l\} \cap \mathrm{conv}\{v^{l+1}, v^{l+2}, \ldots v^n\} = \varnothing.$$

Radon's Theorem: Any set of $d+2$ points in $\mathbb{R}^d$ can be partitioned into two disjoint sets whose convex hulls intersect.



Corollary: If $\mathcal{D}(M)$ allows all the partitions on $[n]$ then nrank $M > n-2$.

Example:

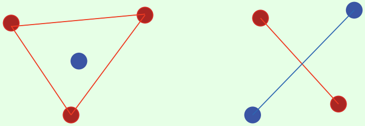
$$\mathcal{D}(M) = \langle 1234, 1423, 1342 \rangle \implies \text{nrank}\, M = 3$$

a quick example: a bound on nrank via Radon's Theorem

Observation: If $M_{i\alpha} = f_i\left(\vec{w}_i \cdot \vec{x}^{\alpha}\right)$ and $s = (v_1, v_2, \ldots, v_n) \in \mathcal{D}(M)$ then for every $l \in [n]$, the convex hulls do not intersect:

$$\text{conv}\{v^1, v^2, \ldots, v^l\} \cap \text{conv}\{v^{l+1}, v^{l+2}, \ldots v^n\} = \varnothing.$$

Radon's Theorem: Any set of $d+2$ points in $\mathbb{R}^d$ can be partitioned into two disjoint sets whose convex hulls intersect.



Corollary: If $\mathcal{D}(M)$ allows all the partitions on $[n]$ then nrank $M > n-2$.

Example:

$$\mathcal{D}(M) = \langle 1234, 1423, 1342 \rangle \implies \text{nrank}\, M = 3$$

Note: The general combinatorial constraints on the nonlinear rank are currently not well-understood, but the topology of $\mathcal{D}(M)$ imposes constraints in two different "good sampling" regimes.

This page is intentionally left blank.

[use blackboard]

The limit of “many/enough” functions

Given a point cloud $X = \{x_a\} \subset \mathbb{R}^d$, $|X| = n$.

For a fixed X, there are only finitely many combinatorially distinct functions, thus define

$$D_{\text{lin}}(X) = \{\text{all sequences in } X \text{ from linear functions } f_w(x) = w \cdot x\}$$

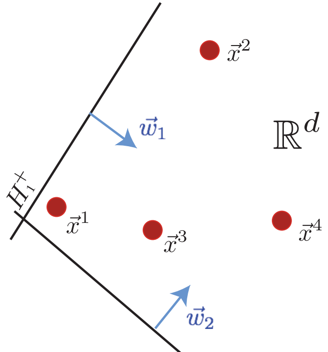
The limit of "many/enough" functions

Given a point cloud $X = \{x_a\} \subset \mathbb{R}^d$, $|X| = n$.

For a fixed X, there are only finitely many combinatorially distinct functions, thus define

$$D_{\text{lin}}(X) = \{\text{all sequences in } X \text{ from linear functions } f_w(x) = w \cdot x\}$$

<u>Lemma</u> [T. Cover, 1967] The number $Q(n, d)$ of maximal sequences in $D_{\text{lin}}(X)$ satisfies
$Q(n+1, d) = Q(n, d) + nQ(n, d-1)$.



The limit of “many/enough” functions

Given a point cloud $X = \{x_a\} \subset \mathbb{R}^d$, $|X| = n$.

For a fixed X, there are only finitely many combinatorially distinct functions, thus define

$$D_{\text{lin}}(X) = \{\text{all sequences in } X \text{ from linear functions } f_w(x) = w \cdot x\}$$

<u>Theorem</u>

The limit of "many/enough" functions

Given a point cloud $X = \{x_a\} \subset \mathbb{R}^d$, $|X| = n$.

For a fixed X, there are only finitely many combinatorially distinct functions, thus define

$$D_{\text{lin}}(X) = \{\text{all sequences in } X \text{ from linear functions } f_w(x) = w \cdot x\}$$

~~Theorem~~ Conjecture. Assume that $n = |X| > d + 2$, then

$$H_* \left(D_{\text{lin}}(X)\right) = H_* \left(\bigvee^{n-1} S^d\right)$$

This page is intentionally left blank.

[use blackboard]

The Dowker multi-filtration of a pure directed complex.

Given M is an $m \times n$ matrix.
and thresholds $\theta = (\theta_1, \ldots, \theta_m)$.

Binary matrix
$B_{ia} \stackrel{\text{def}}{=} (M_{ia} \leq \theta_i)$.

The Dowker multi-filtration of a pure directed complex.

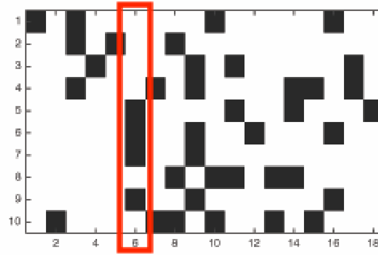
Given M is an $m \times n$ matrix.
and thresholds $\theta = (\theta_1, \dots, \theta_m)$.

The Dowker complex is

$$\mathrm{Dow}(M, \theta) \stackrel{\text{def}}{=} \Delta\left(\sigma_1, \dots, \sigma_n\right)$$

where $\sigma_a \stackrel{\text{def}}{=} \{i \mid M_{ia} \le \theta_i\}$.

Binary matrix
$B_{ia} \stackrel{\text{def}}{=} (M_{ia} \le \theta_i)$.



The Dowker multi-filtration of a pure directed complex.

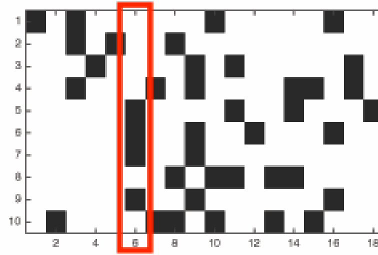
Given M is an $m \times n$ matrix.
and thresholds $\theta = (\theta_1, \ldots, \theta_m)$.

Binary matrix
$B_{ia} \stackrel{\text{def}}{=} (M_{ia} \leq \theta_i)$.

The Dowker complex is

$$\mathrm{Dow}(M, \theta) \stackrel{\text{def}}{=} \Delta\left(\sigma_1, \ldots, \sigma_n\right)$$

where $\sigma_a \stackrel{\text{def}}{=} \{i \mid M_{ia} \leq \theta_i\}$.



Note: An increasing chain of thresholds induces a filtration.
This is because

$$\theta_i \leq \tilde{\theta}_i \quad \forall i \in [m] \implies \mathrm{Dow}(M, \theta) \subseteq \mathrm{Dow}(M, \tilde{\theta})$$

A fun fact:

Theorem (C. H. Dowker, 1952)

Let B be a binary matrix, then the following two complexes are homotopy equivalent:

$$\mathrm{Dow}(B,1) \sim \mathrm{Dow}(B^T,1)$$

Question: What does Dowker complex have to do with nonlinear rank?

$M_{ia} = \phi_i \left(\sum_{a=1}^{d} W_{ia} x_a^{\alpha} \right)$, where ϕ_i are monotone increasing.

$$M_{ia} \leq \theta_i,$$

$$\phi_i \left(\sum_{a=1}^{d} W_{ia} x_a^{\alpha} \right) \leq \theta_i,$$

$$\sum_{a=1}^{d} W_{ia} x_a^{\alpha} \leq \phi_i^{-1}(\theta_i),$$

$$\vec{x}^{\alpha} \in H_i^{+},$$

where

$H_i^{+} \stackrel{\text{def}}{=} \{x | \sum_{a=1}^{d} W_{ia} x_a - \phi_i^{-1}(\theta_i) \leq 0\}$

Question: What does Dowker complex have to do with nonlinear rank?

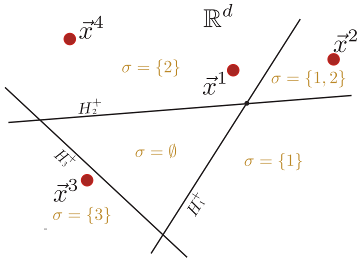
$M_{ia} = \phi_i\left(\sum_{a=1}^d W_{ia} x_a^\alpha\right)$, where ϕ_i are monotone increasing.

$$M_{ia} \leq \theta_i,$$

$$\phi_i\left(\sum_{a=1}^d W_{ia} x_a^\alpha\right) \leq \theta_i,$$

$$\sum_{a=1}^d W_{ia} x_a^\alpha \leq \phi_i^{-1}(\theta_i),$$

$$\vec{x}^\alpha \in H_i^+,$$

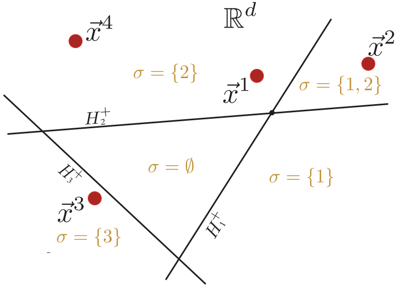


where
$H_i^+ \overset{\text{def}}{=} \{x | \sum_{a=1}^d W_{ia} x_a - \phi_i^{-1}(\theta_i) \leq 0\}$

Question: What does Dowker complex have to do with nonlinear rank (and thus feedforward networks)?

<u>Lemma:</u> Under a condition of "enough sampling"(all the chambers that correspond to the facets of the nerve are sampled) ,

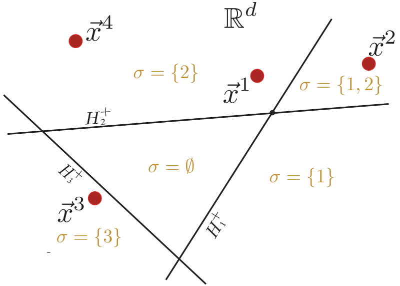
$$Dow\left(M, \vec{\theta}\right) = \text{nerve}\left\{H_i^+\right\} \sim \cup_{i=1}^m H_i^+$$



Question: What does Dowker complex have to do with nonlinear rank (and thus feedforward networks)?

Lemma: Under a condition of "enough sampling"(all the chambers that correspond to the facets of the nerve are sampled) ,

$$Dow\left(M, \vec{\theta}\right) = \text{nerve}\left\{H_i^+\right\} \sim \cup_{i=1}^m H_i^+$$



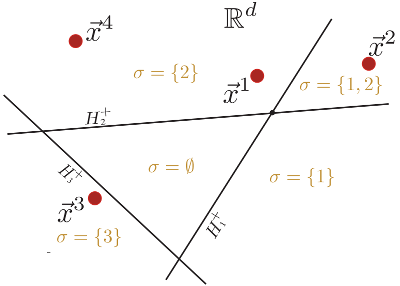
Thus if we had a perfect sampling, then

$$Dow\left(M, \vec{\theta}\right) \sim \mathbb{R}^d \setminus \text{ a polyhedron.}$$

Question: What does Dowker complex have to do with nonlinear rank (and thus feedforward networks)?

Lemma: Under a condition of "enough sampling" (all the chambers that correspond to the facets of the nerve are sampled) ,

$$Dow\left(M, \vec{\theta}\right) = \text{nerve}\left\{H_i^+\right\} \sim \cup_{i=1}^m H_i^+$$



Thus if we had a perfect sampling, then

$$Dow\left(M, \vec{\theta}\right) \sim \mathbb{R}^d \setminus \text{ a polyhedron.}$$

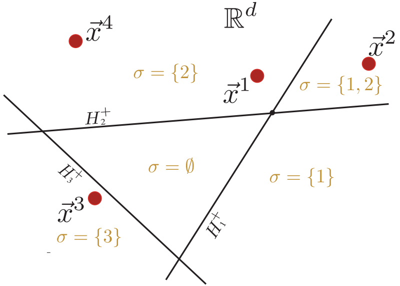
Moreover, (with some trick) manipulating filtrations one can force

$$Dow\left(M, \vec{\theta}\right) \sim \left(\mathbb{R}^d \setminus \text{ polytope}\right) \sim S^{d-1}.$$

Question: What does Dowker complex have to do with nonlinear rank (and thus feedforward networks)?

Lemma: Under a condition of "enough sampling"(all the chambers that correspond to the facets of the nerve are sampled) ,

$$Dow\left(M,\vec{\theta}\right) = \text{nerve}\left\{H_i^+\right\} \sim \cup_{i=1}^m H_i^+$$



Thus if we had a perfect sampling, then

$$Dow\left(M,\vec{\theta}\right) \sim \mathbb{R}^d \setminus \text{ a polyhedron.}$$

Moreover, (with some trick) manipulating filtrations one can force

$$Dow\left(M,\vec{\theta}\right) \sim \left(\mathbb{R}^d \setminus \text{ polytope}\right) \sim S^{d-1}.$$

This page is intentionally left blank.

Homological invariants of a filtered Dowker complex

Given a real $n \times m$ matrix M. An increasing sequence of threshold vectors,

$$\vec{\theta}_1 < \vec{\theta}_2 < \cdots < \vec{\theta}_{p-1} < \vec{\theta}_p, \qquad p \leq mn.$$

Homological invariants of a filtered Dowker complex

Given a real $n \times m$ matrix M. An increasing sequence of threshold vectors,

$$\vec{\theta}_1 < \vec{\theta}_2 < \cdots < \vec{\theta}_{p-1} < \vec{\theta}_p, \qquad p \leq mn.$$

induces a sequence of simplicial complexes $\Delta_j \stackrel{\text{def}}{=} \text{Dow}(M, \vec{\theta}_j)$,

$$\varnothing \hookrightarrow \Delta_1 \hookrightarrow \Delta_2 \hookrightarrow \cdots \hookrightarrow \Delta_{p-1} \hookrightarrow \Delta_p = 2^{[n]}, \qquad p \leq mn$$

Homological invariants of a filtered Dowker complex

Given a real $n \times m$ matrix M. An increasing sequence of threshold vectors,

$$\vec{\theta}_1 < \vec{\theta}_2 < \cdots < \vec{\theta}_{p-1} < \vec{\theta}_p, \qquad p \leq mn.$$

induces a sequence of simplicial complexes $\Delta_j \stackrel{\text{def}}{=} \text{Dow}(M, \vec{\theta}_j)$,

$$\varnothing \hookrightarrow \Delta_1 \hookrightarrow \Delta_2 \hookrightarrow \cdots \hookrightarrow \Delta_{p-1} \hookrightarrow \Delta_p = 2^{[n]}, \qquad p \leq mn$$

as well as homomorphisms in homology (over whatever field)

$$0 \to \tilde{H}_*(\Delta_1) \to \tilde{H}_*(\Delta_2) \to \cdots \to \tilde{H}_*(\Delta_{p-1}) \to 0$$

Homological invariants of a filtered Dowker complex

Given a real $n \times m$ matrix M. An increasing sequence of threshold vectors,

$$\vec{\theta}_1 < \vec{\theta}_2 < \cdots < \vec{\theta}_{p-1} < \vec{\theta}_p, \qquad p \leq mn.$$

induces a sequence of simplicial complexes $\Delta_j \stackrel{\text{def}}{=} \text{Dow}(M, \vec{\theta}_j)$,

$$\varnothing \hookrightarrow \Delta_1 \hookrightarrow \Delta_2 \hookrightarrow \cdots \hookrightarrow \Delta_{p-1} \hookrightarrow \Delta_p = 2^{[n]}, \qquad p \leq mn$$

as well as homomorphisms in homology (over whatever field)

$$0 \to \tilde{H}_*(\Delta_1) \to \tilde{H}_*(\Delta_2) \to \cdots \to \tilde{H}_*(\Delta_{p-1}) \to 0$$

Homological invariants:

- *Betti Curves*, $\qquad \beta_m(\theta) \stackrel{\text{def}}{=} \dim H_m\left(\text{Dow}\left(M, \theta\right)\right)$
- *Persistence intervals*

Why can we infer the nonlinear rank from the Dowker complex?

Given M and $\Theta = (\vec{\theta}_1 < \ldots < \vec{\theta}_p)$
one obtains Betti curves $\beta_k^{\Theta}(\rho)$

The result may depend on the way of thresholding Θ...[use the blackboard]

Why can we infer the nonlinear rank from the Dowker complex?

Given M and $\Theta = (\vec{\theta}_1 < \ldots < \vec{\theta}_p)$ one obtains Betti curves $\beta_k^{\Theta}(\rho)$

The result may depend on the way of thresholding Θ...[use the blackboard]

However, one can estimate

$$\bar{\beta}_k(\rho) \stackrel{\text{def}}{=} \frac{1}{|\mathcal{S}_m|} \sum_{s \in \mathcal{S}_m} \beta_k^{\Theta_s}(\rho)$$

very well via random sampling of $\mathcal{S}_m$.

Importantly, $\bar{\beta}_k(\rho)$ is invariant under the action of the group $\mathcal{G}^m$.

Why can we infer the nonlinear rank from the Dowker complex?

Given M and $\Theta = (\vec{\theta}_1 < \ldots < \vec{\theta}_p)$ one obtains Betti curves $\beta_k^{\Theta}(\rho)$

The result may depend on the way of thresholding Θ...[use the blackboard]

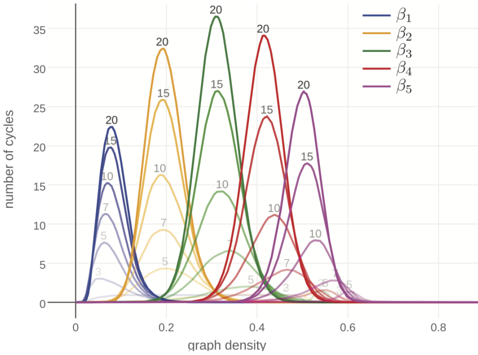
<u>However,</u> one can estimate

$$\bar{\beta}_k(\rho) \stackrel{\text{def}}{=} \frac{1}{|\mathcal{S}_m|} \sum_{s \in \mathcal{S}_m} \beta_k^{\Theta_s}(\rho)$$

very well via random sampling of $\mathcal{S}_m$.

Importantly, $\bar{\beta}_k(\rho)$ is invariant under the action of the group $\mathcal{G}^m$.

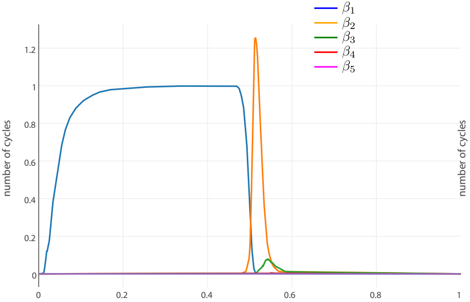
Average Betti curves $\bar{\beta}_k(\rho)$ in various ranks.



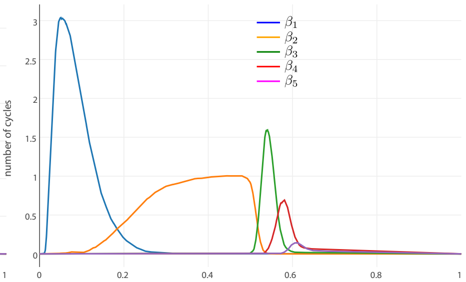
Two different ways of inferring d:

1) A single persistent cycle $\in H_m(\mathrm{Dow}(M, -))$ in dimension $m = (d-1)$.
2) The 'shapes' of the of the Betti curves

[use the board]

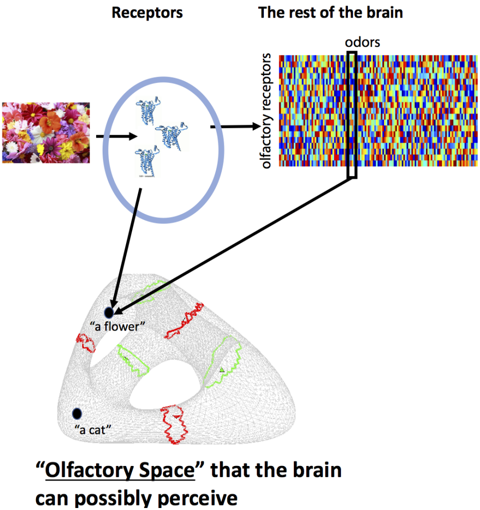


Betti curves for random matrices of
nrank = 2



Betti curves for random matrices of
nrank = 3

Can one estimate the dimension of the space of smells?



A question: Can we tell the dimension of the olfactory space from OR ? ligand response map?

Answer: A "yes". *Preliminary findings:* We've found that the fly olfactory space of likely low-dimensional ($4 \leq d \leq 6$).

This page is intentionally left blank.

[use blackboard]

Quick summary

- Convex sensing problems can be best understood in terms of directed complexes.
- Natural questions about feedforward networks can be restated in terms of "nonlinear matrix factorization" and the nonlinear rank.
- Exact bounds on the nonlinear rank are still poorly understood, but there are hard constraints from geometry and topology.
- Nonlinear rank can be estimated (with high precision) using topological tools.

This page is intentionally left blank.

[use blackboard]

Vielen Dank!

this talk:

Philip Egger | Min-Chun Wu | Aliaksandra Yarosh

Funding: DARPA YFA, NSF IOS , NSF DMS + NIH NIGMS.