Topological Feature Selection for Time Series Data

Peter Bubenik
University of Florida

joint work with Johnathan Bush, James Madison University

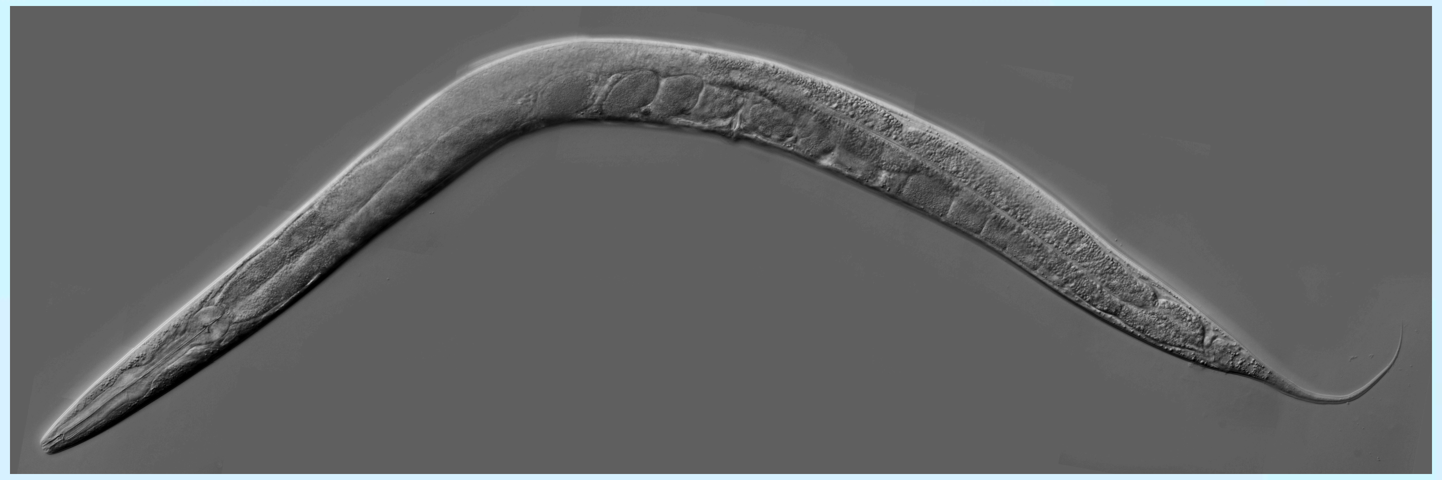
Geometric realization of AATRN
IMSI, Chicago, IL
August 18, 2025

Outline

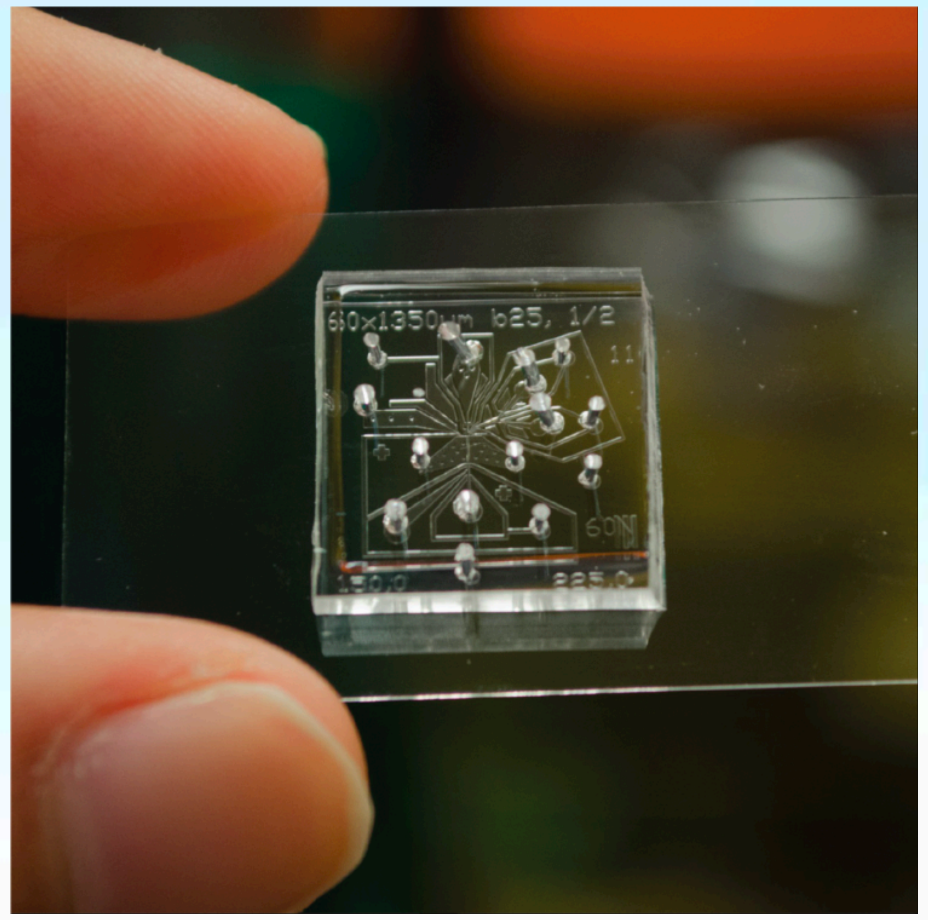
- Background
- Question and Main Idea
- Synthetic Example
- Theory
- Application to Biology

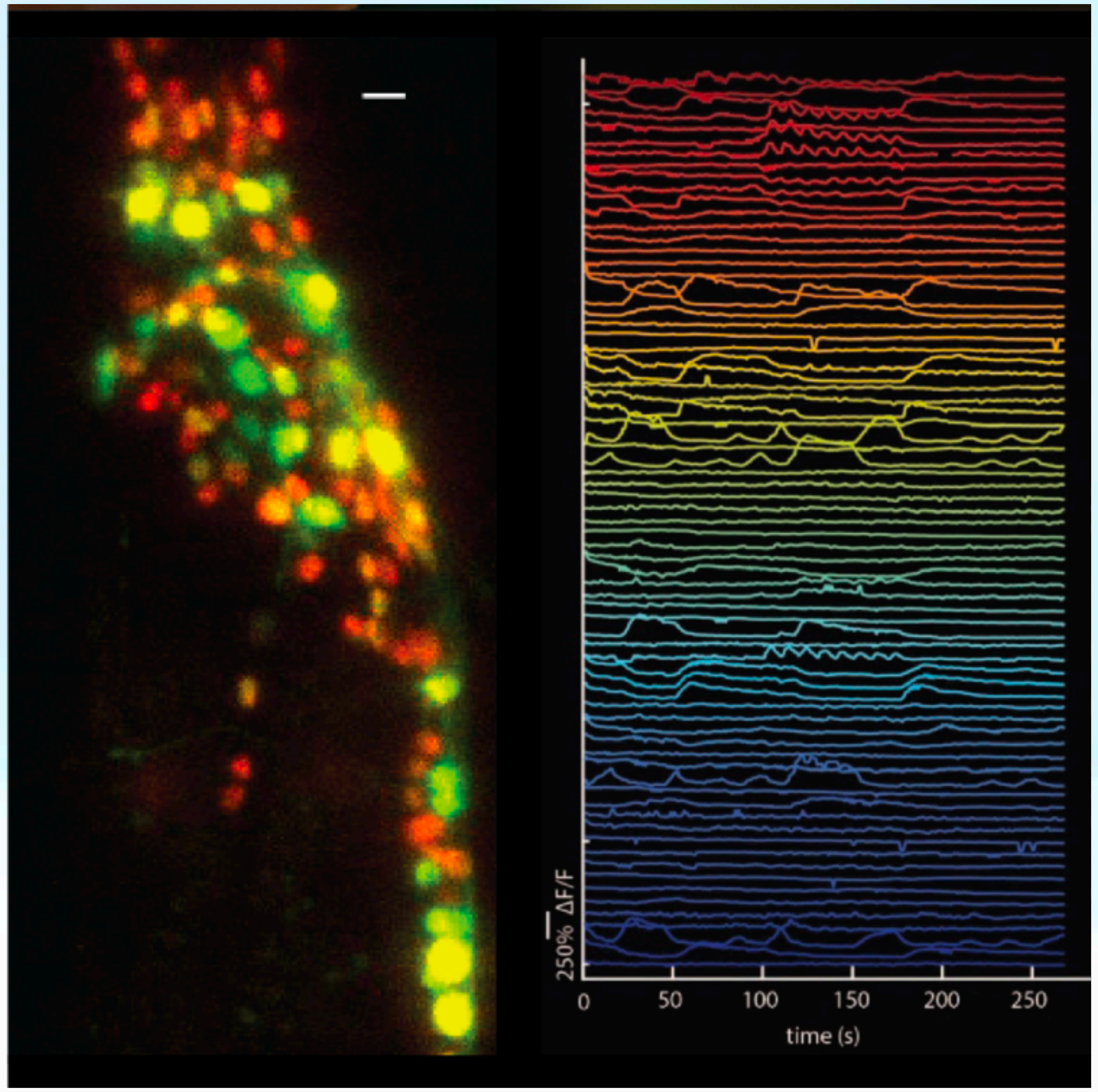
Background

Southeast Center for Mathematics and Biology

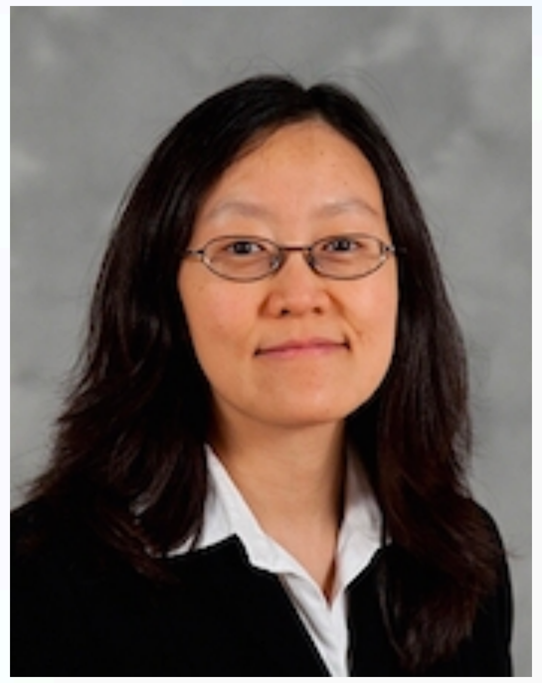


Caenorhabditis elegans

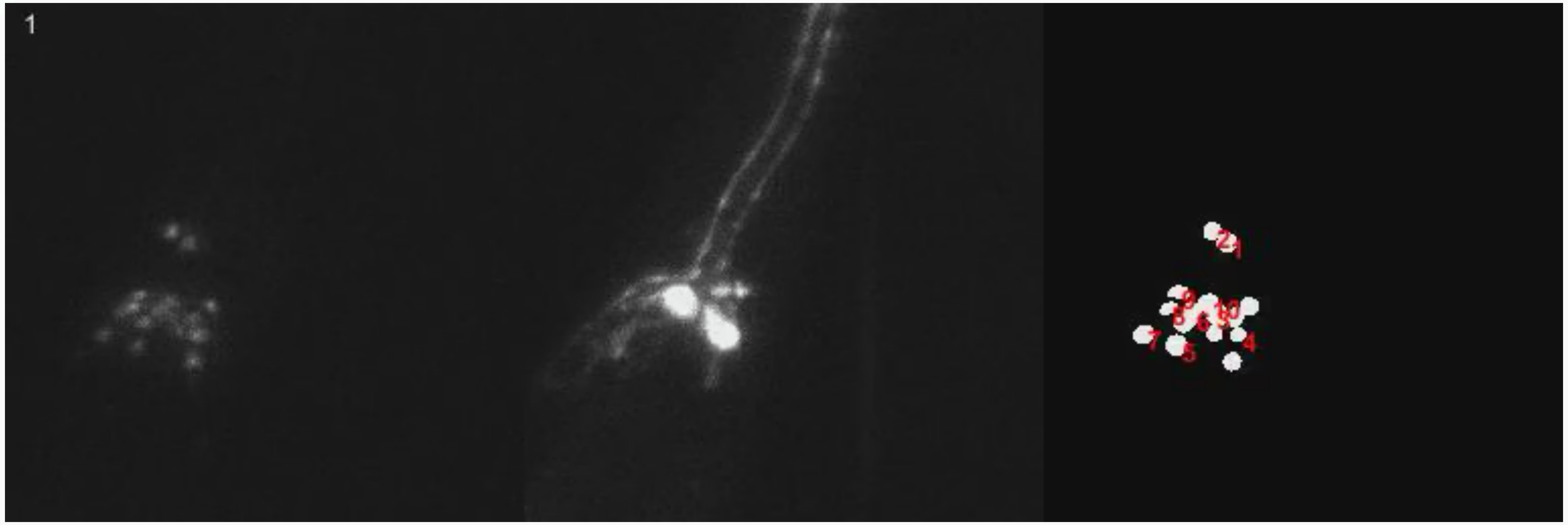




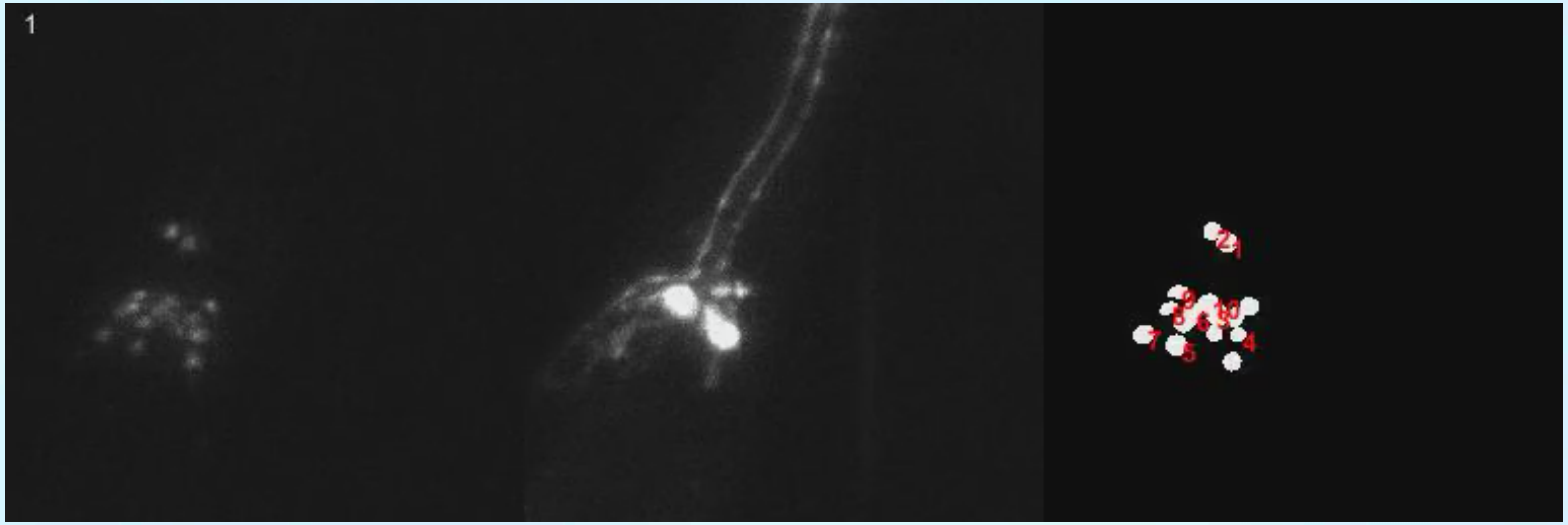
Johnathan Bush



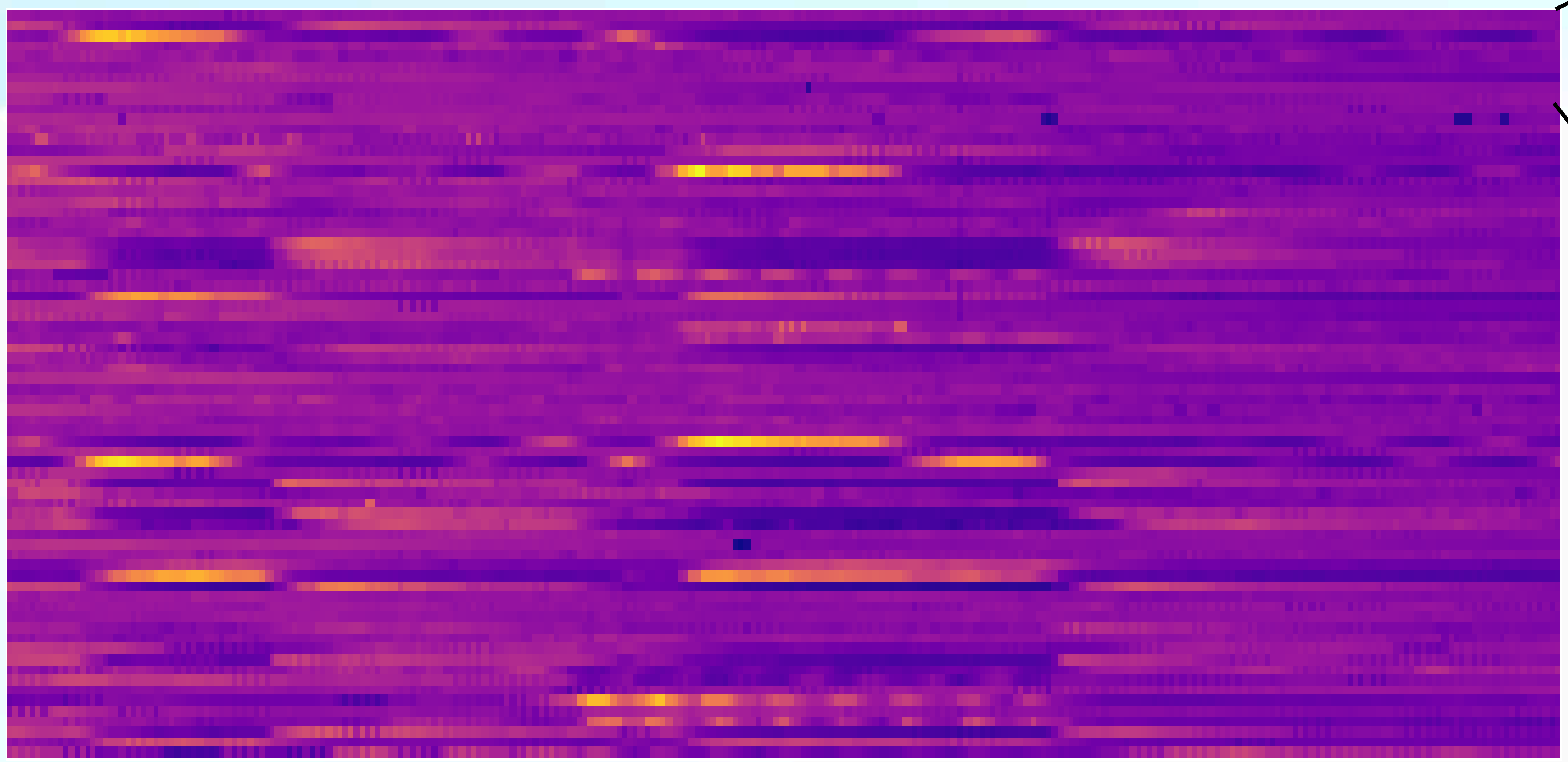
Hang Lu



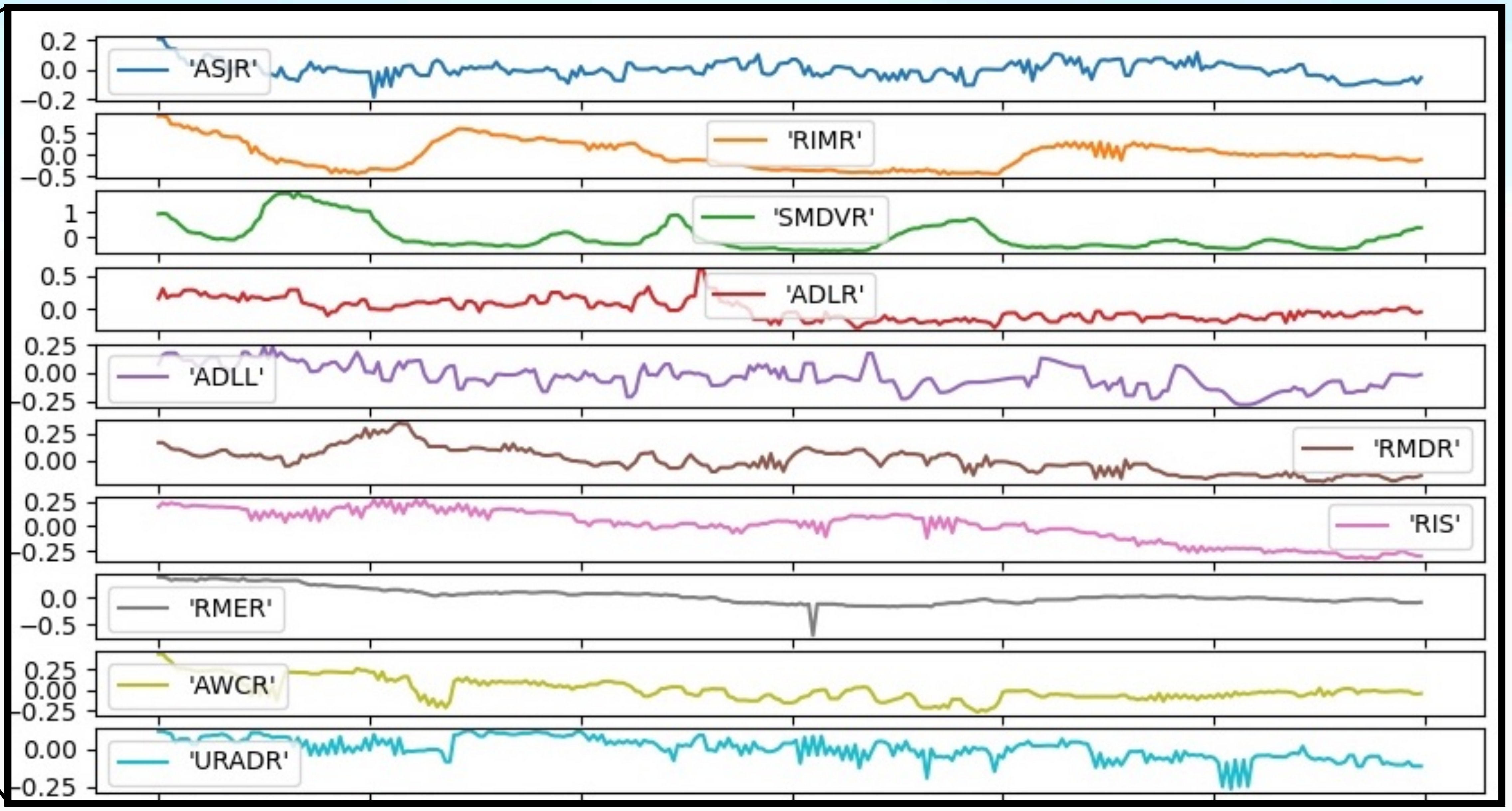
genetically encoded indicators of calcium ions



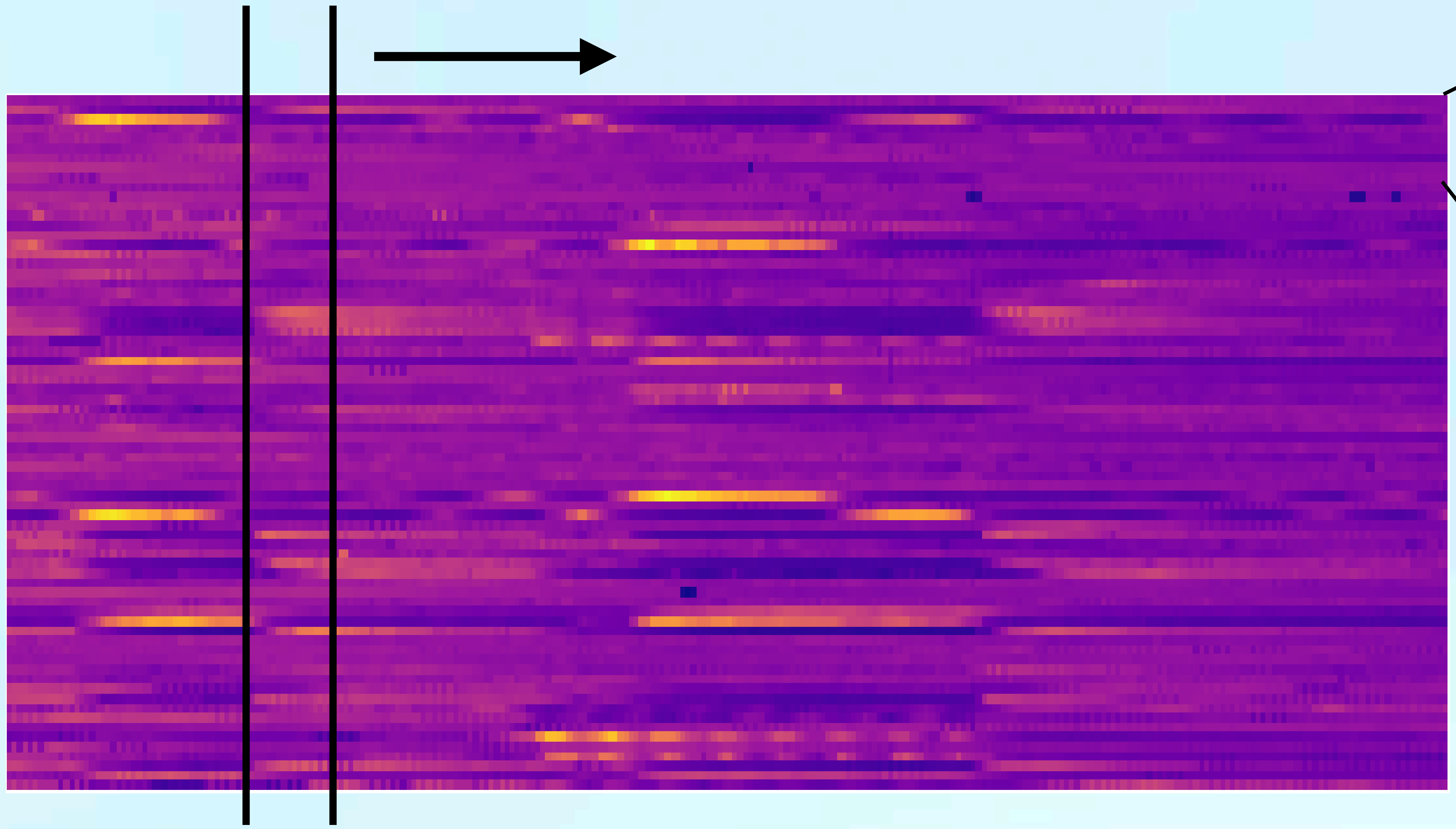
(a lot of hard work)

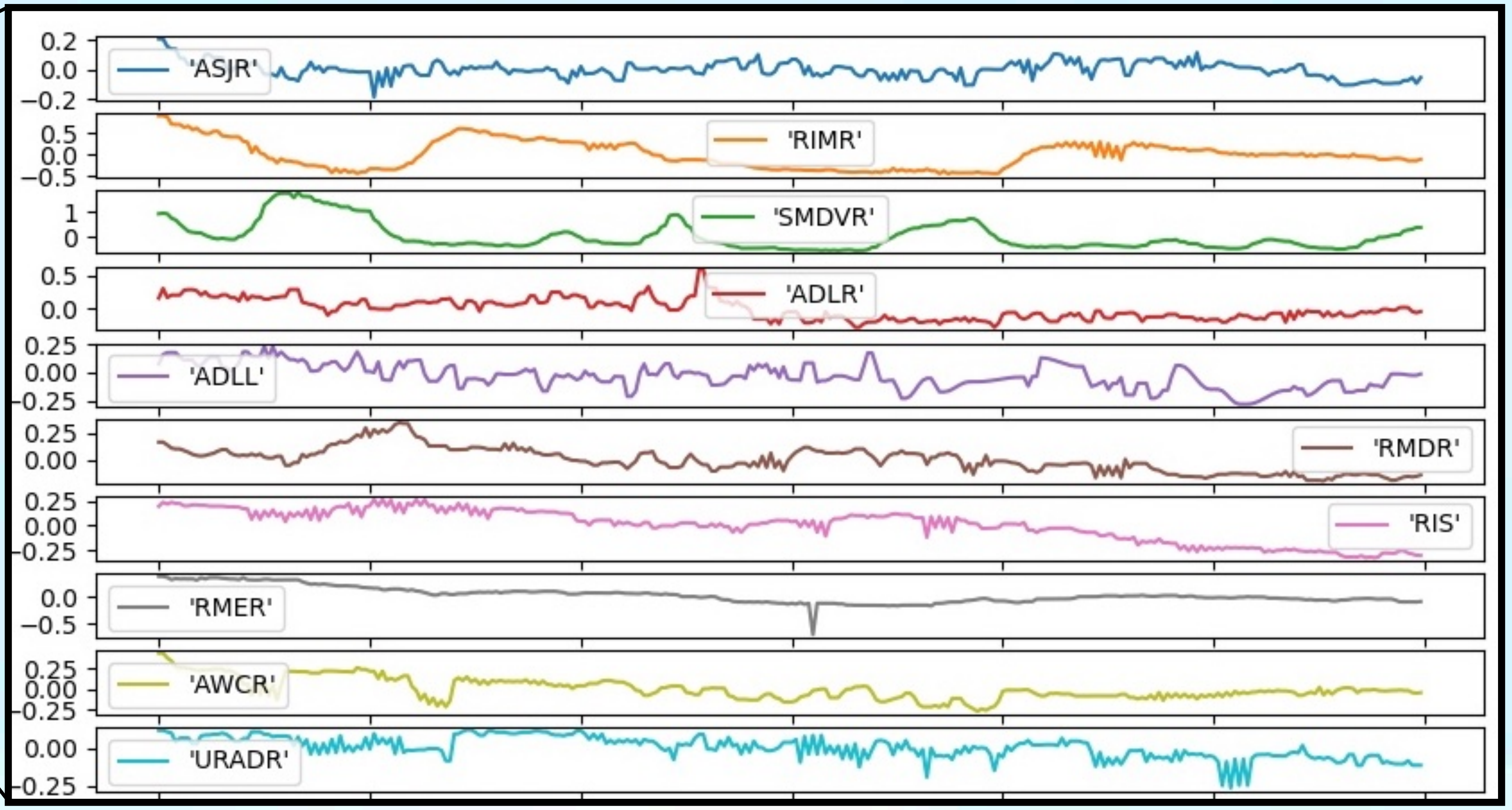


Chaudhary, S., Lee, S. A., Li, Y., Patel, D. S., & Lu, H. (2021). **Graphical-model framework for automated annotation of cell identities in dense cellular images.** *Elife*, *10*, e60321.



time series



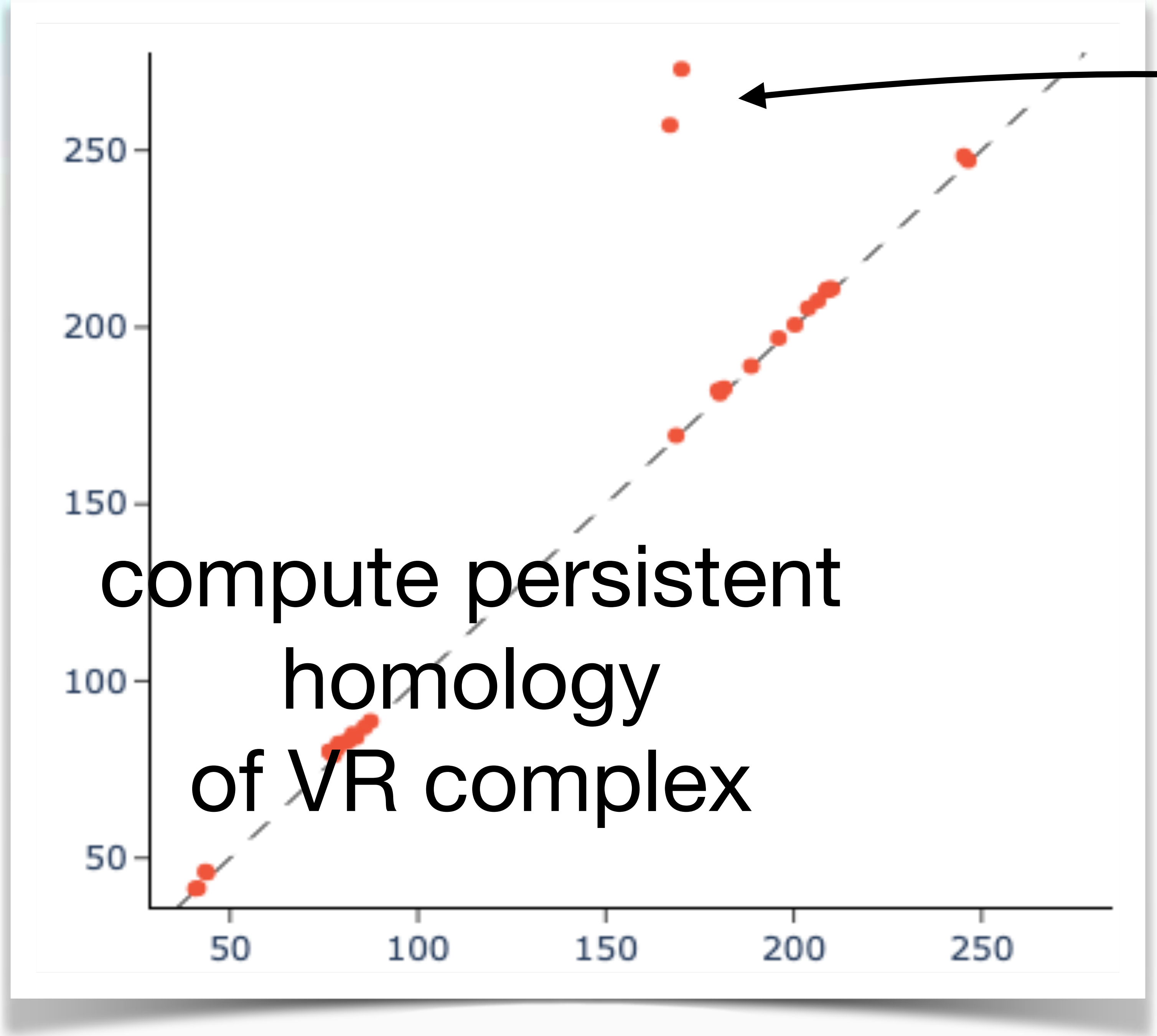


sliding-window embedding,
window length 10

$\Downarrow$

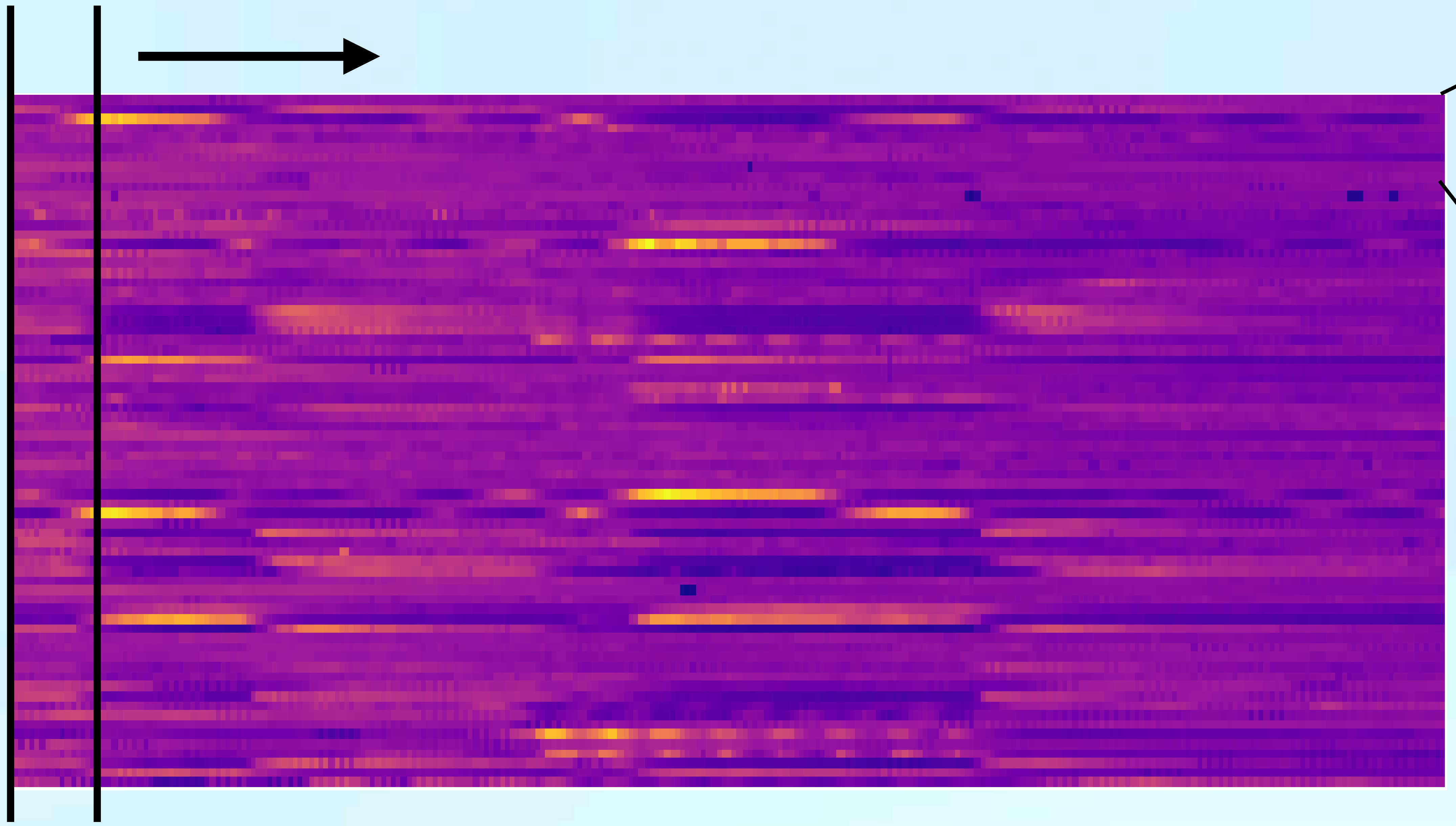
a high-dimensional
point cloud,
269 points in $\mathbb{R}^{72\cdot 10}$

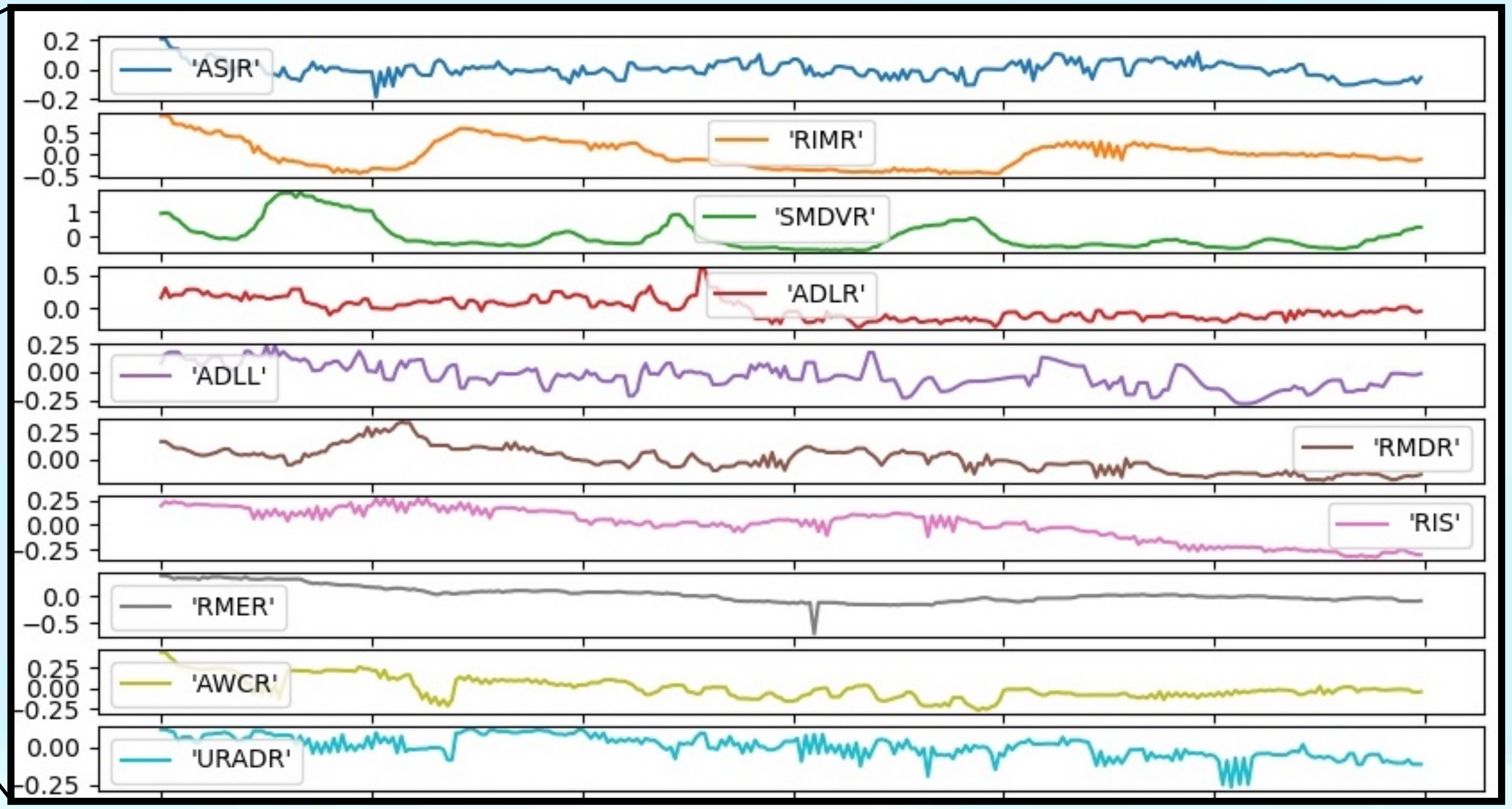
$\Longrightarrow$



compute persistent
homology
of VR complex

two highly persistent cycles

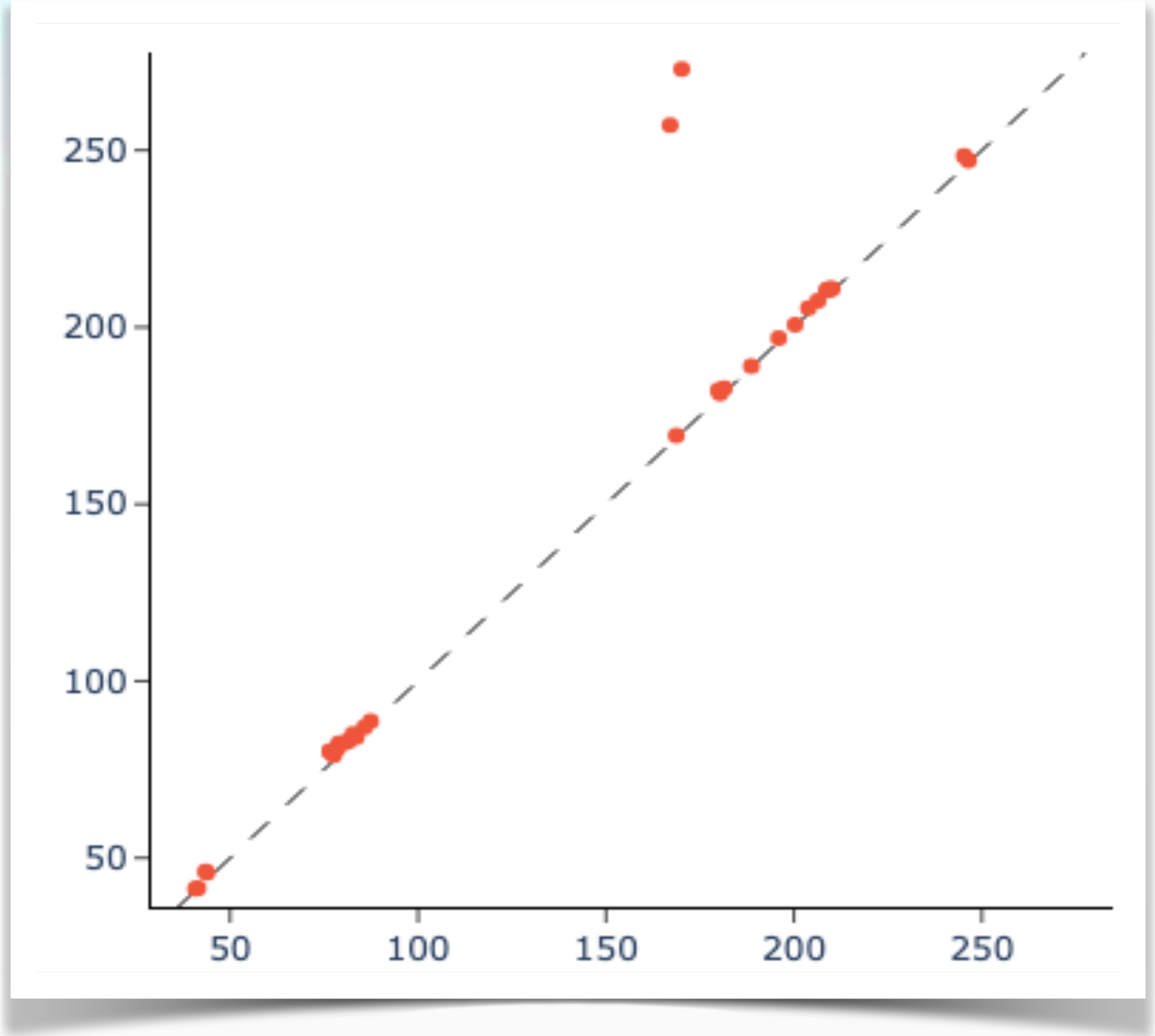


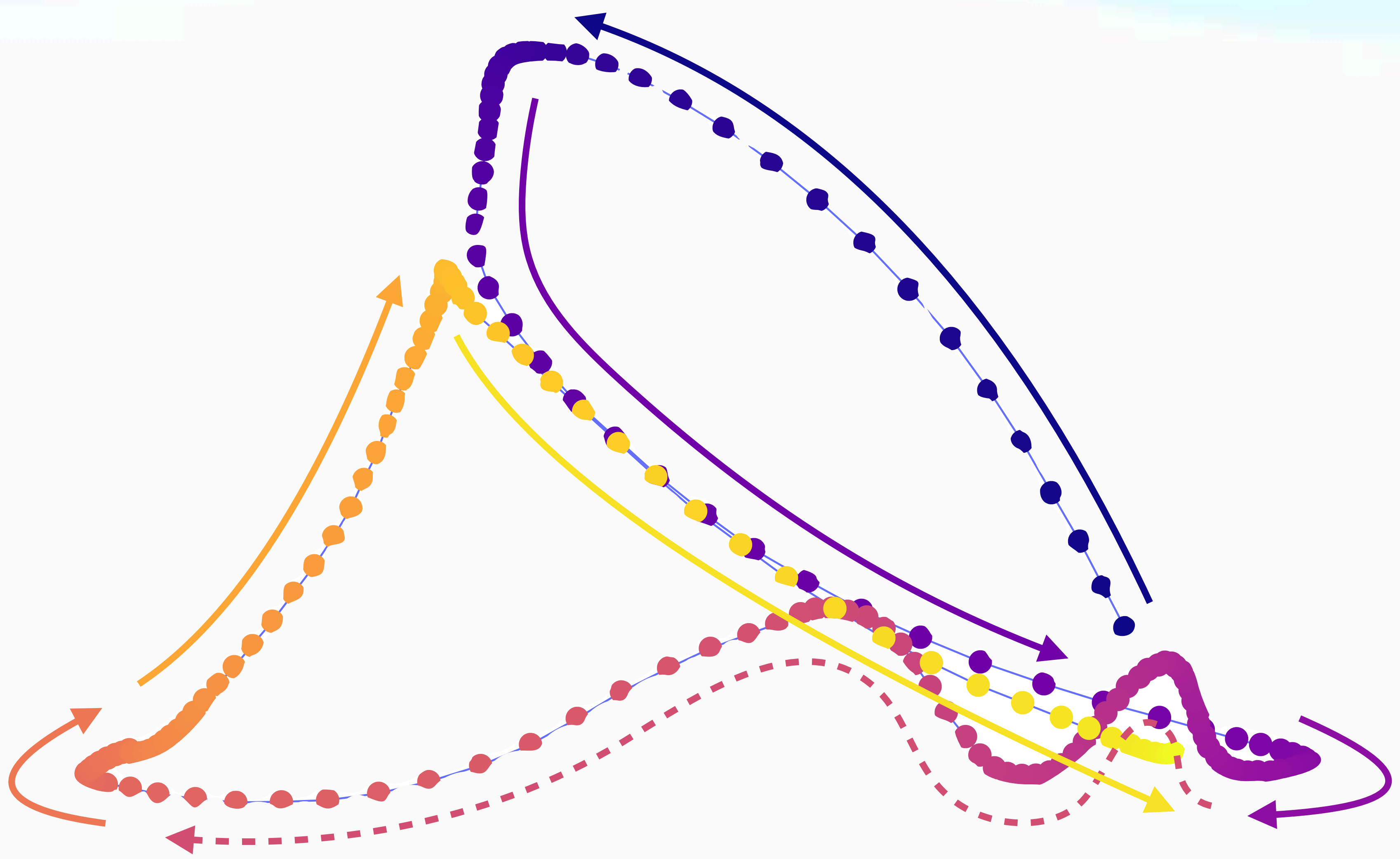


sliding-window embedding,
window length 10

$\Downarrow$

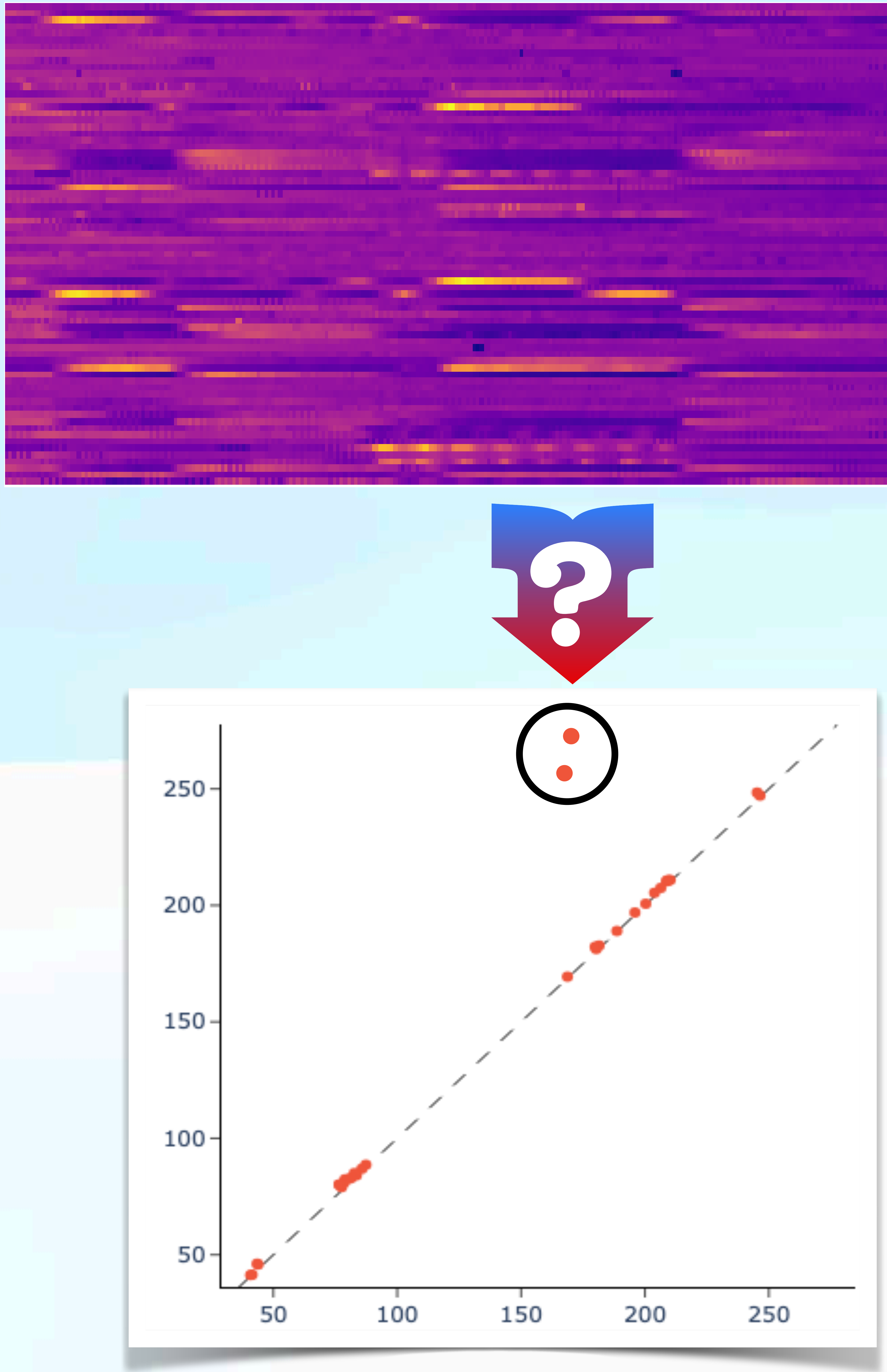
a high-dimensional
point cloud,
289 points in $\mathbb{R}^{72\cdot 10}$





PCA projection

Question and Main Idea



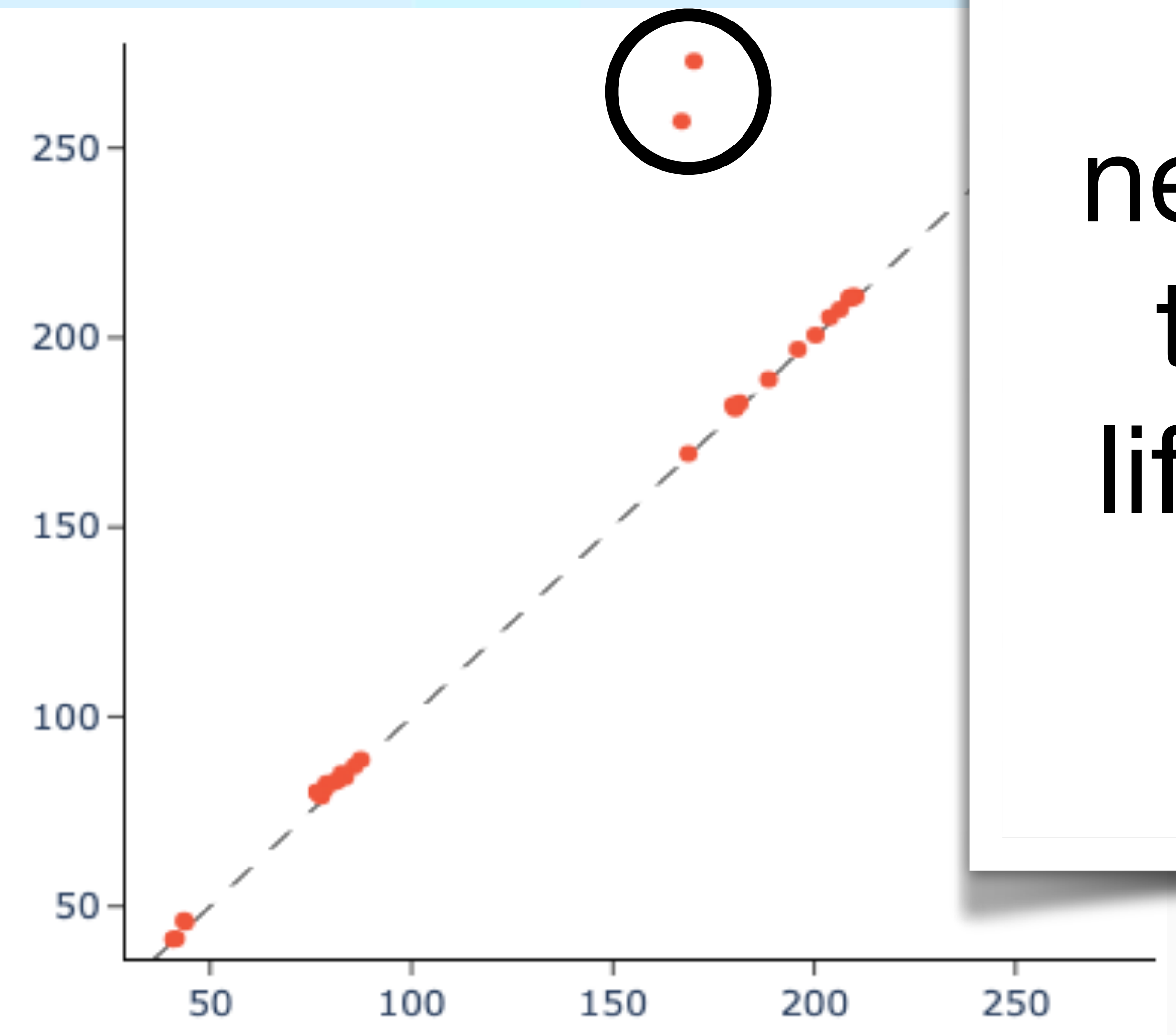
How can topology help us identify the neurons most responsible for driving these global, cyclic brain dynamics?

<u>Idea:</u>
rescale neuron activations and recompute persistent homology

Idea:

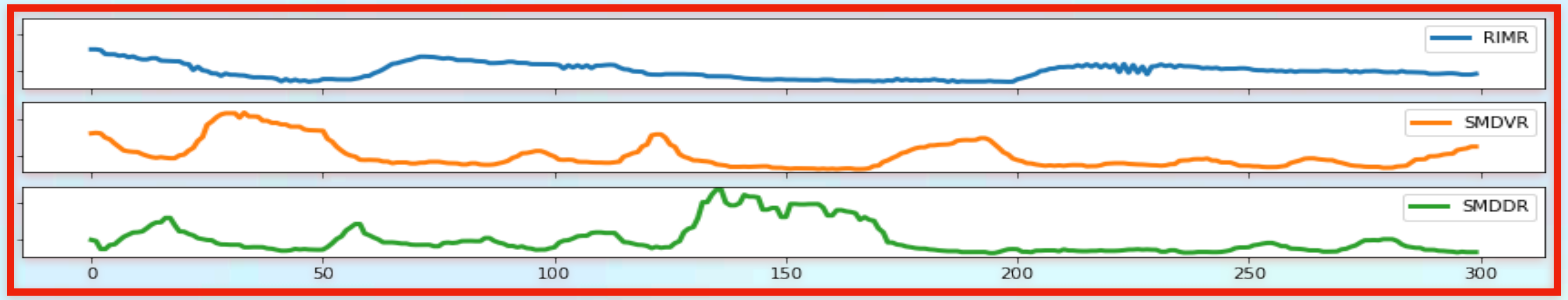
rescale neuron activations and recompute persistent homology

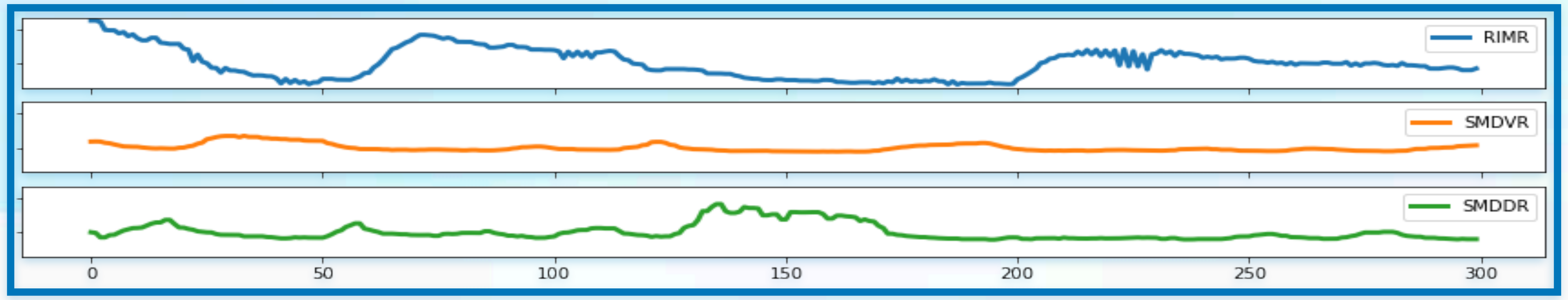
in our setting:

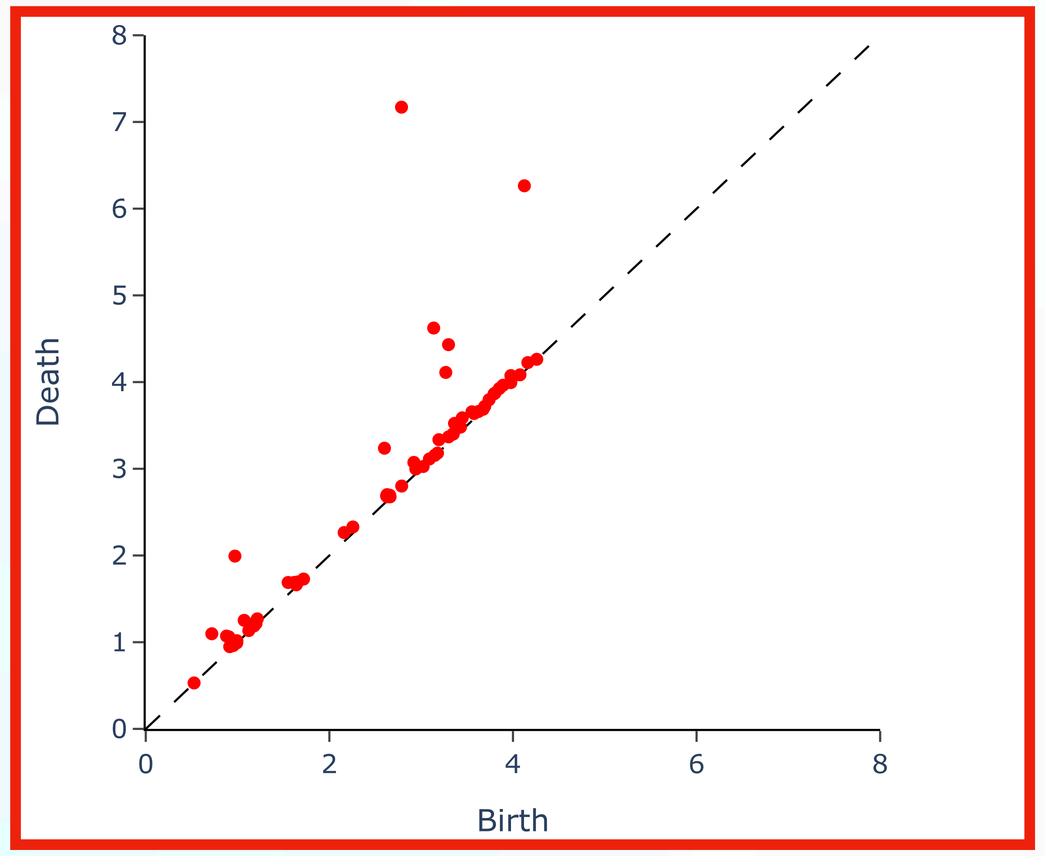


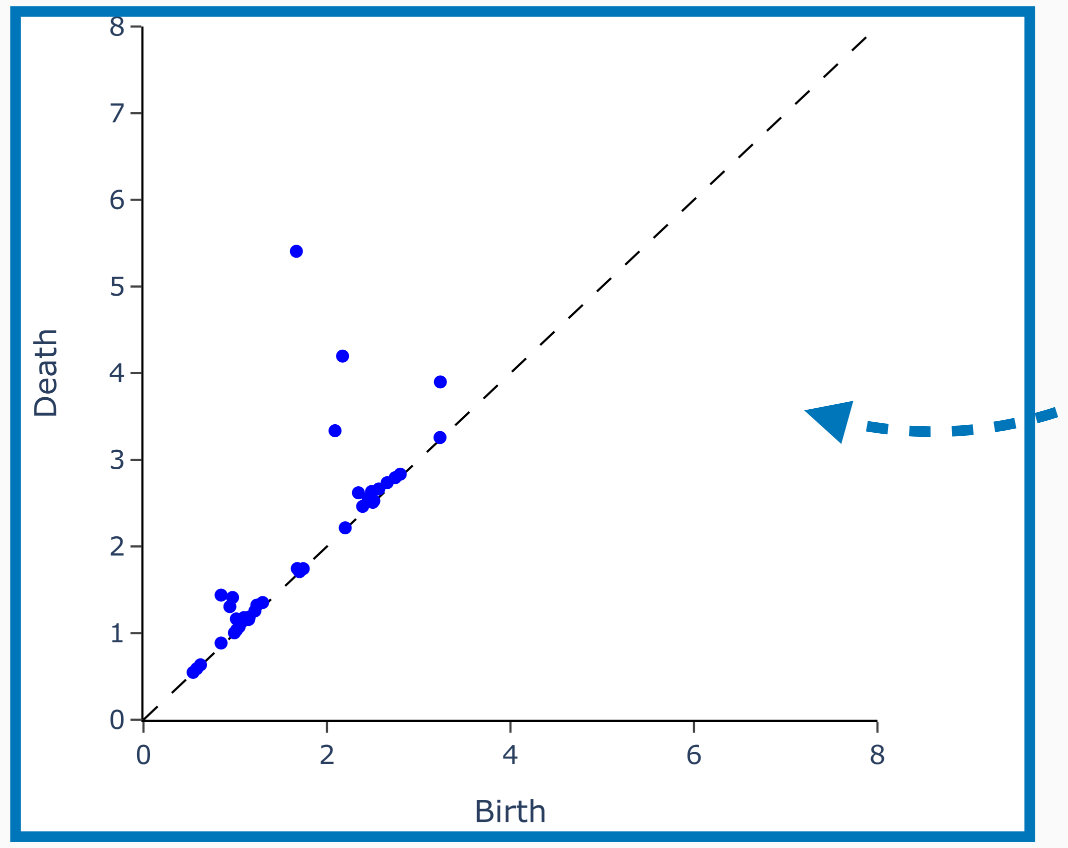
Goal:

rescale the neural activations to increase the lifetimes of these two dominant cycles





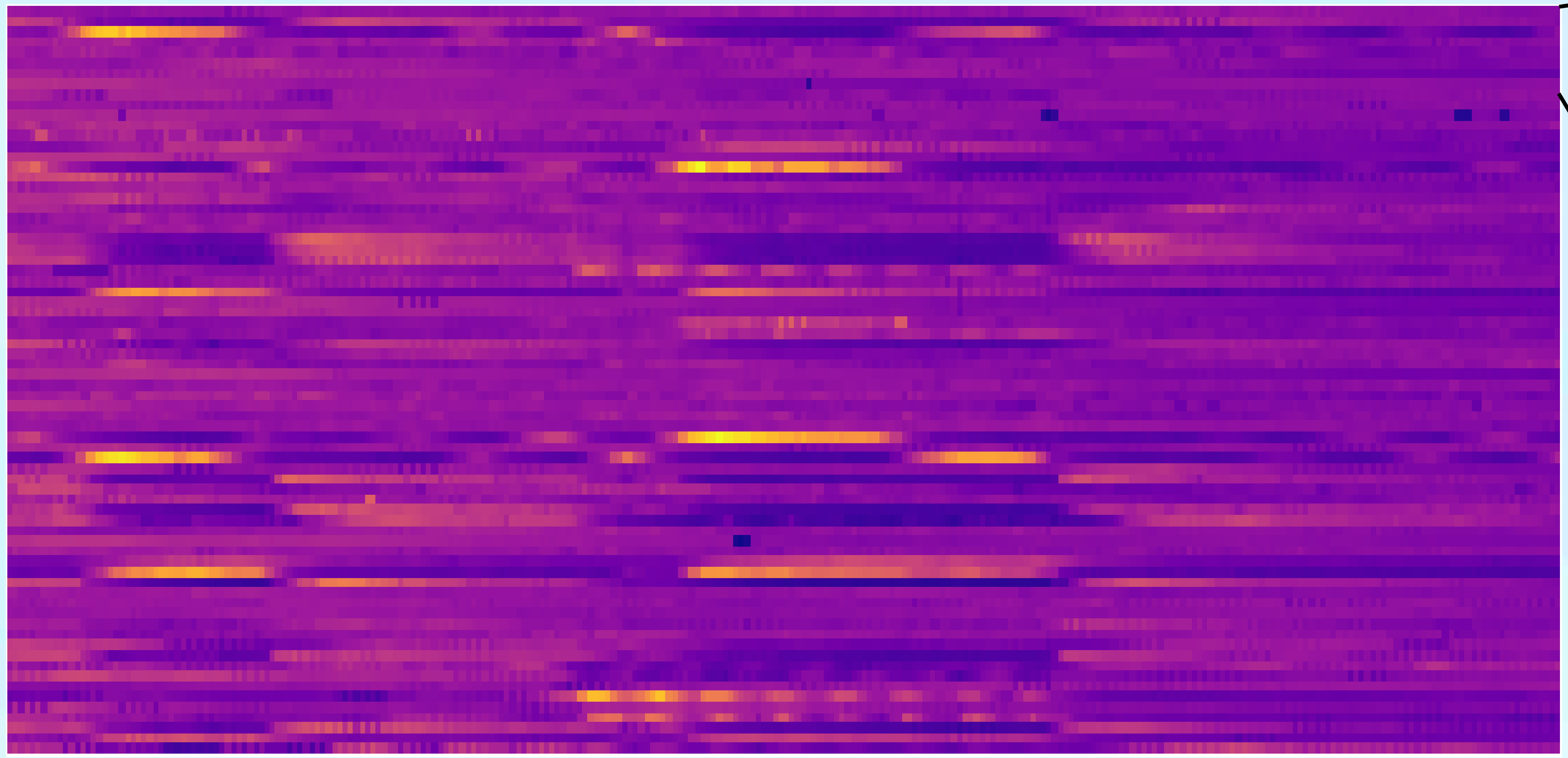


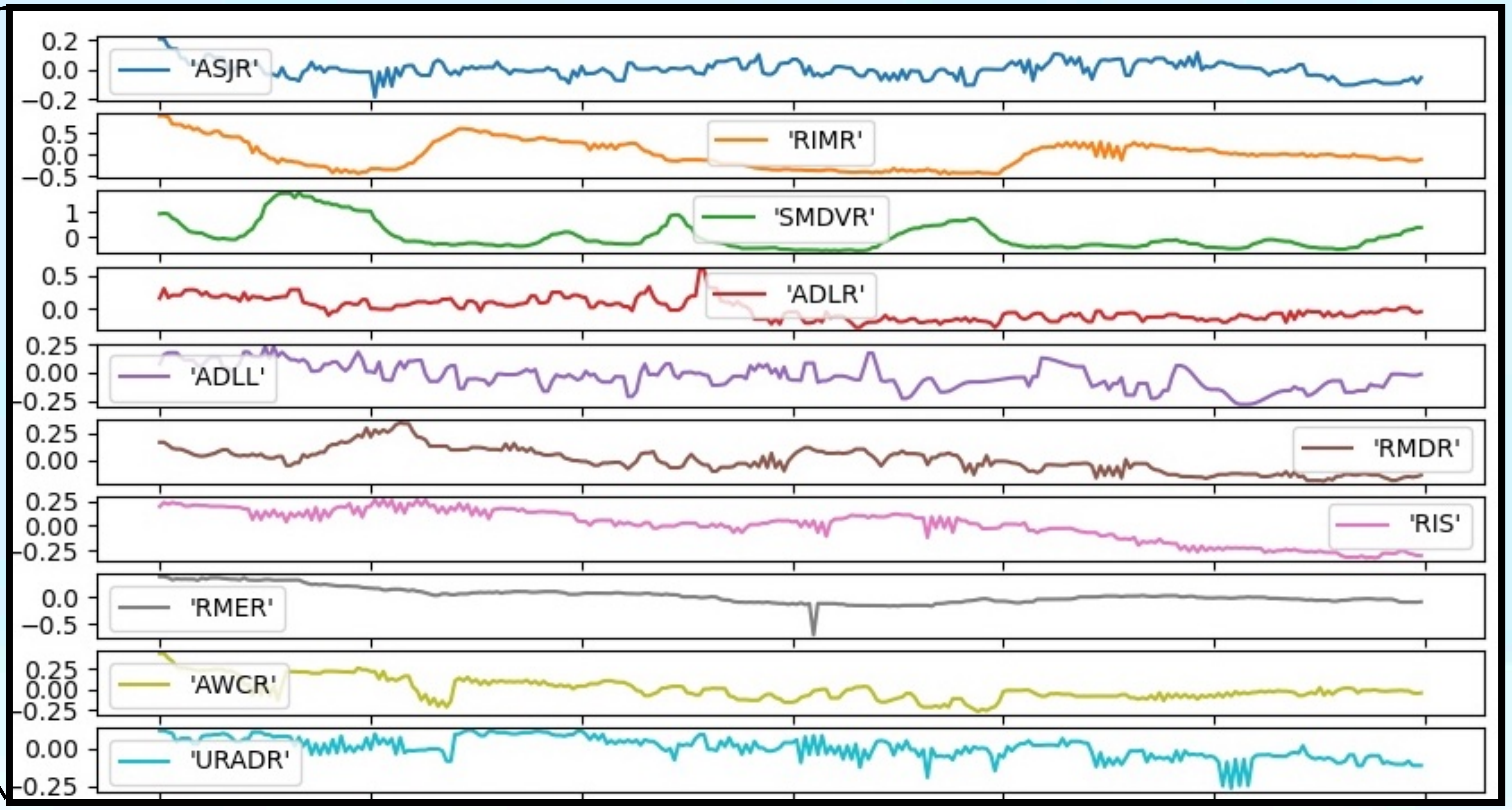


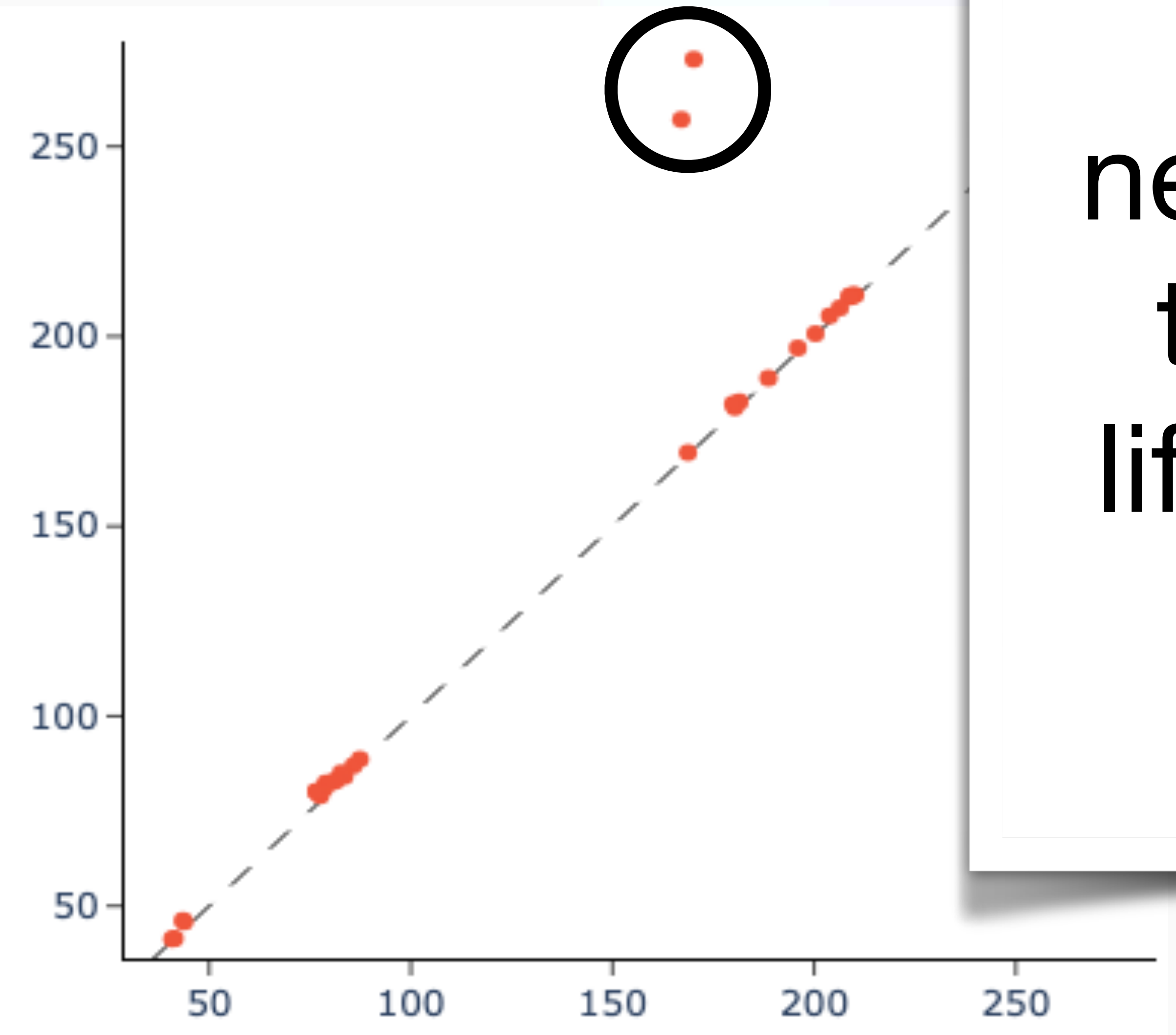
Interpretation:

neurons scaled more highly after optimization contribute more strongly to global dynamics

In our setting:

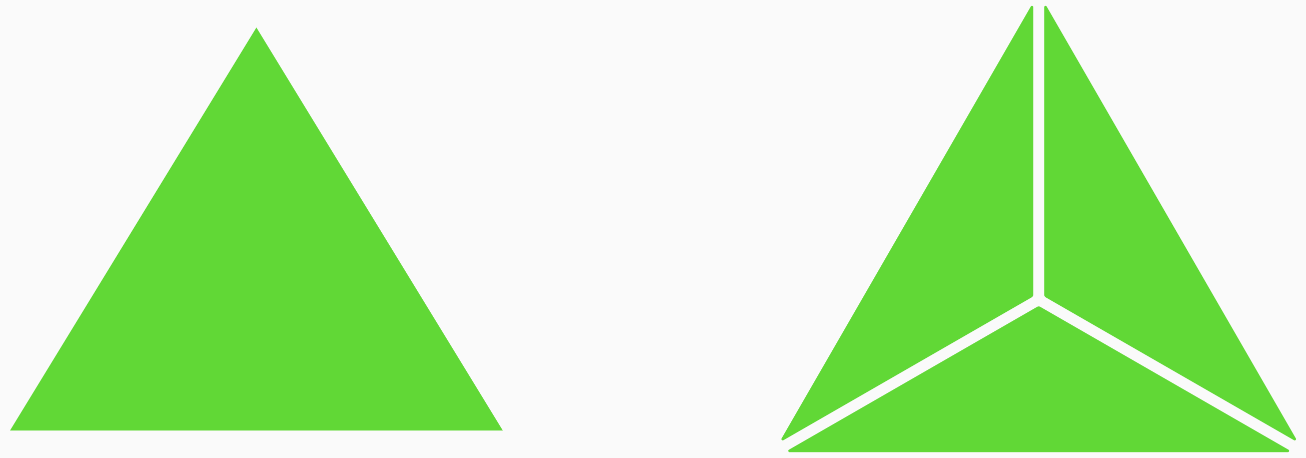






Goal:
rescale the neural activations to increase the lifetimes of these two dominant cycles

restrict to nonnegative scalars that sum to 1
(the standard geometric simplex)



Synthetic Example

Synthetic example

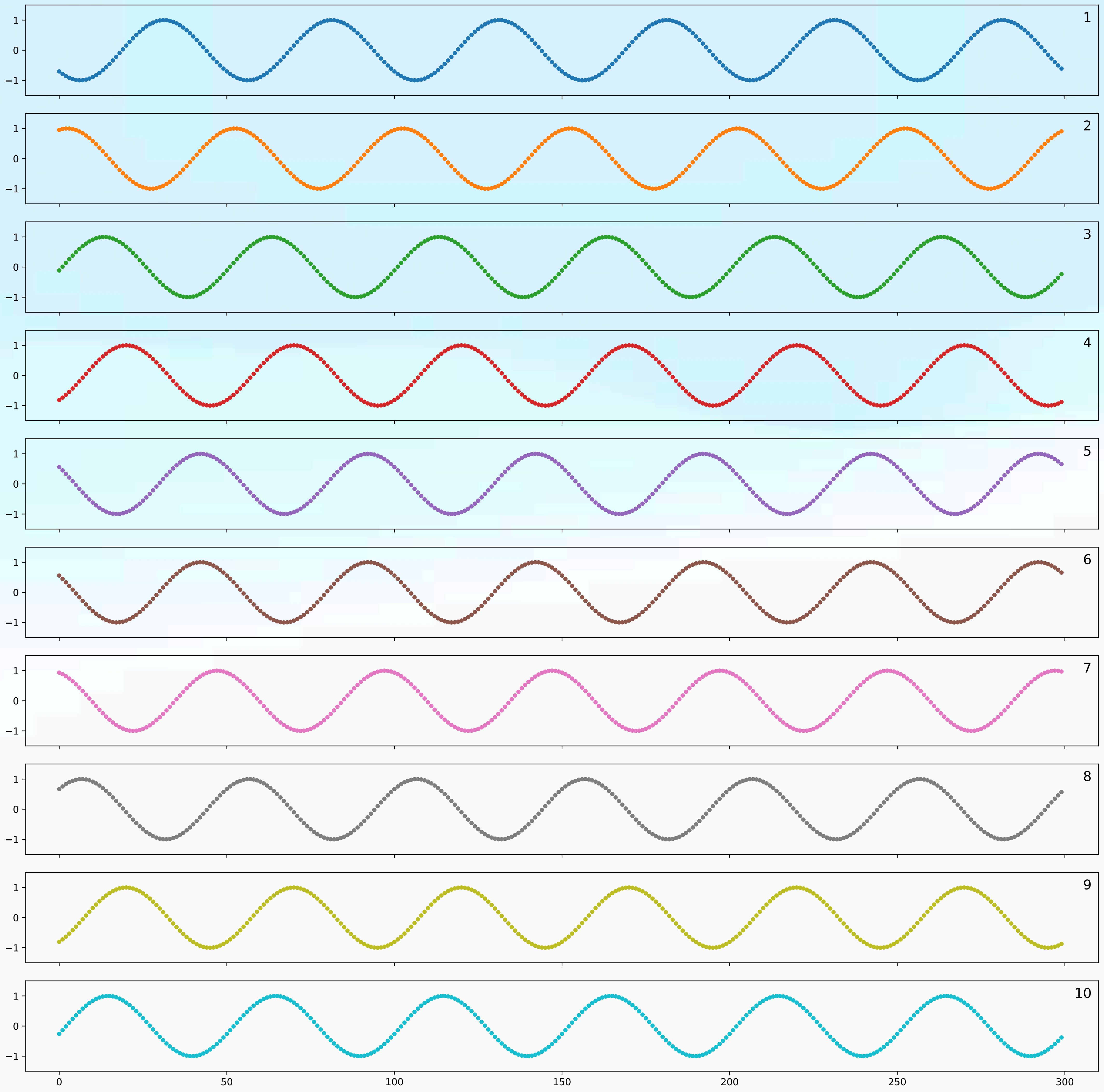
10 sine waves

amplitude $= 1$

period $= 50$

randomly-chosen phase shift

sampled at $0,1,\ldots,300$



Synthetic example

10 sine waves

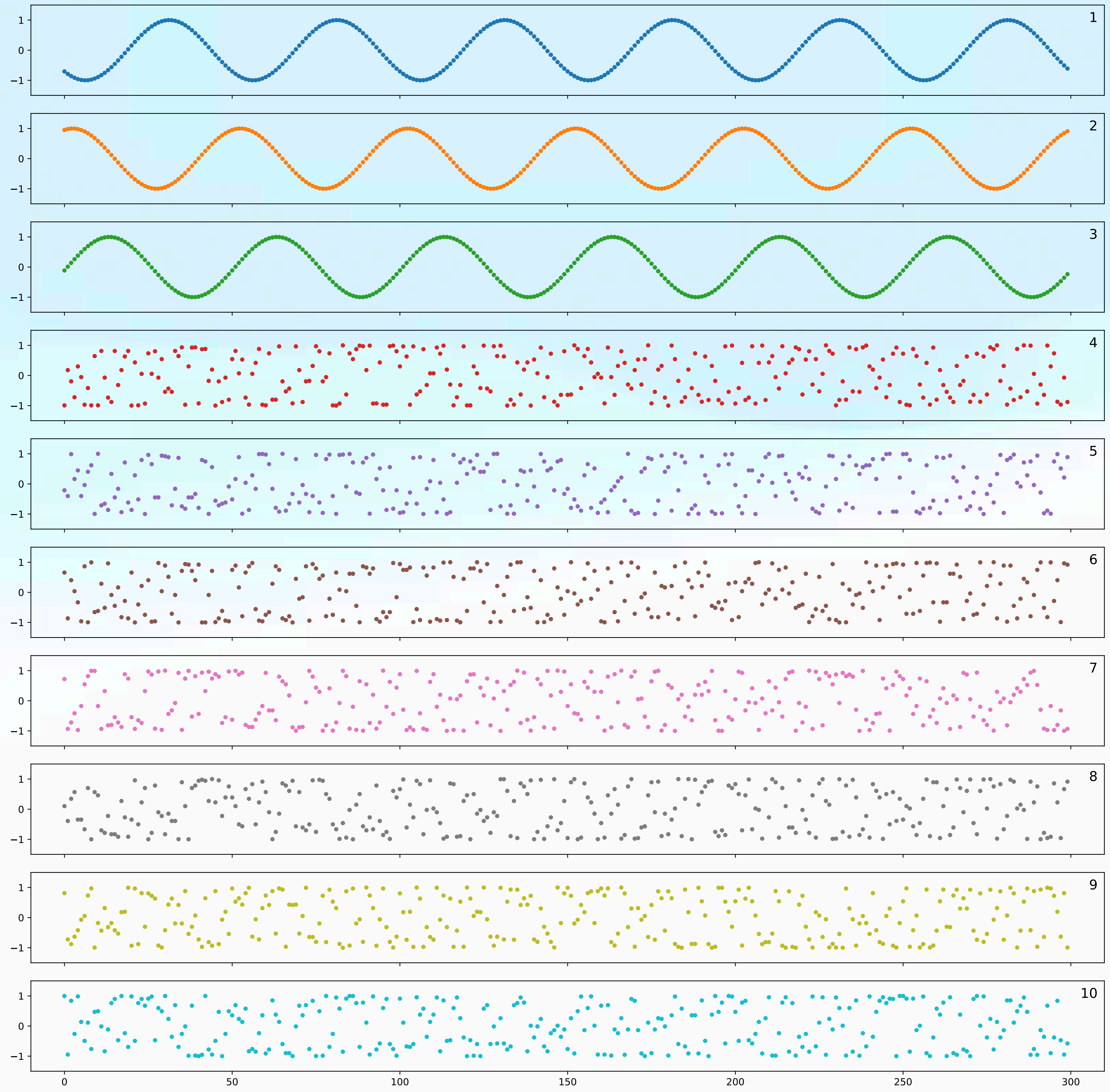
amplitude $= 1$

period $= 50$

randomly-chosen phase shift

sampled at $0, 1, \ldots, 300$

randomly permute
7 of the 10



Synthetic example

10 sine waves

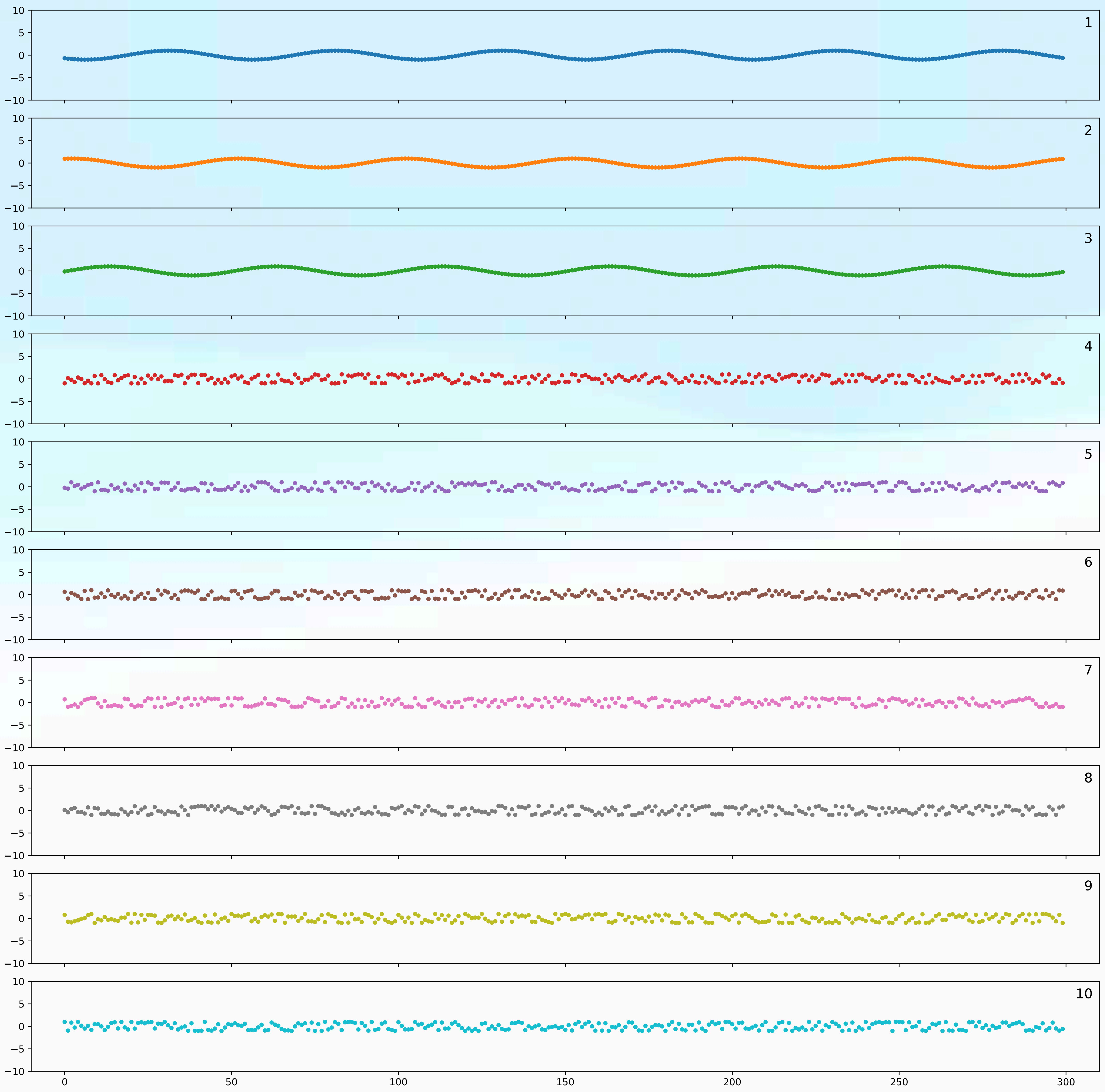
amplitude $= 1$

period $= 50$

randomly-chosen phase shift

sampled at $0, 1, \ldots, 300$

randomly permute
7 of the 10



(zoom out)

Synthetic example

10 sine waves

amplitude $= 1$

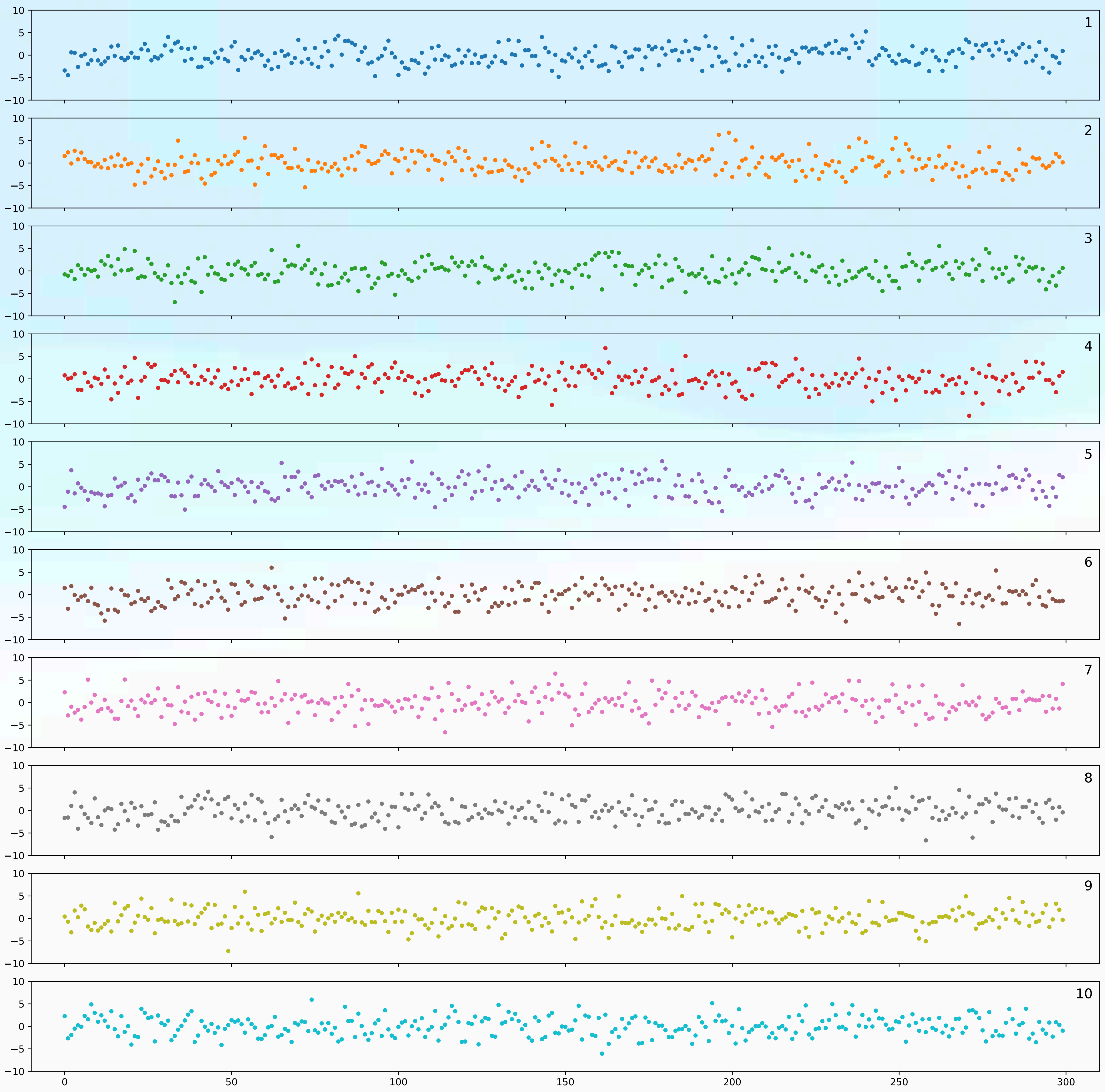
period $= 50$

randomly-chosen phase shift

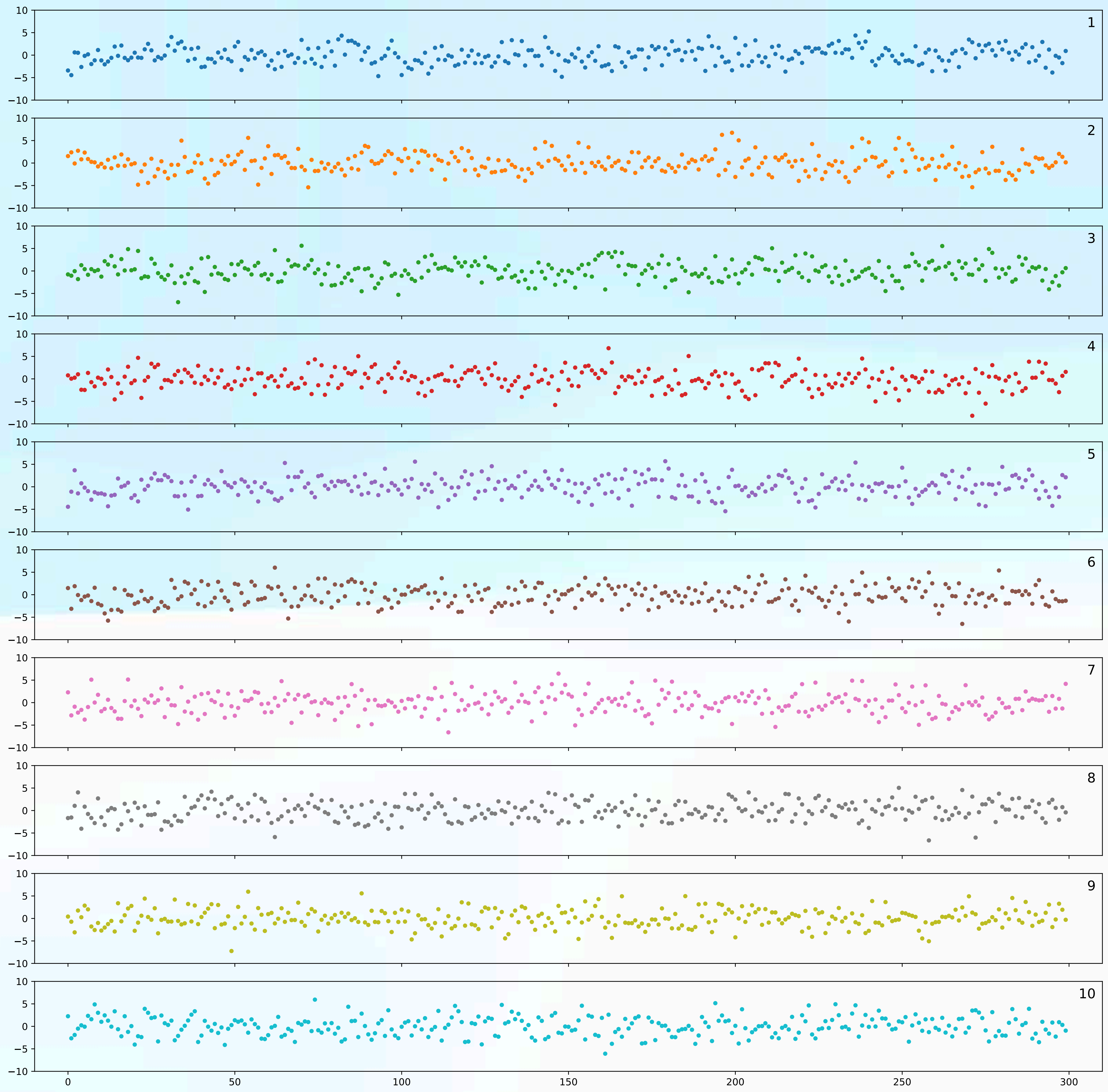
sampled at $0,1,\ldots,300$

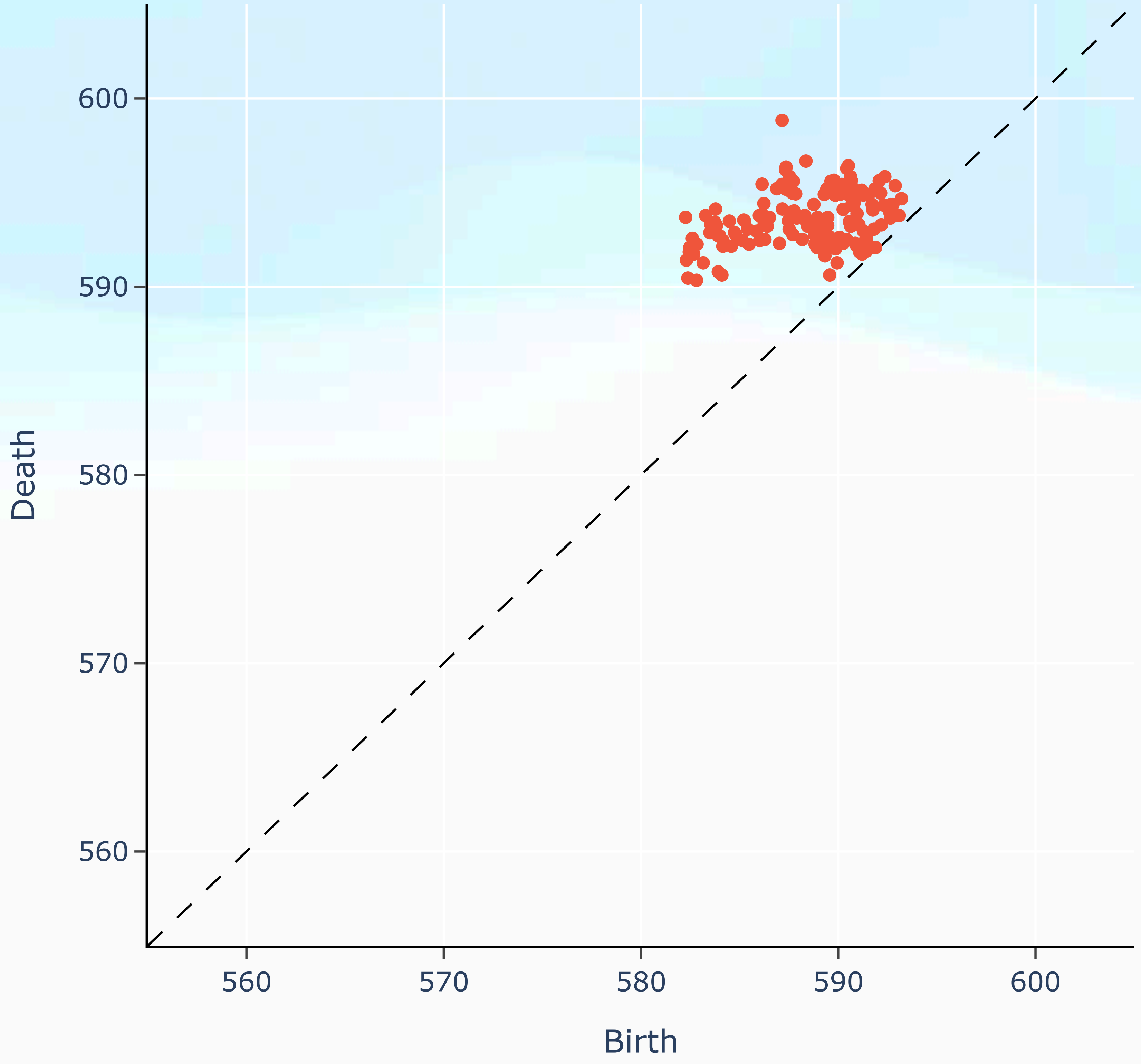
randomly permute
7 of the 10

add iid Gaussian noise
st. dev. $\sigma = 1.5$



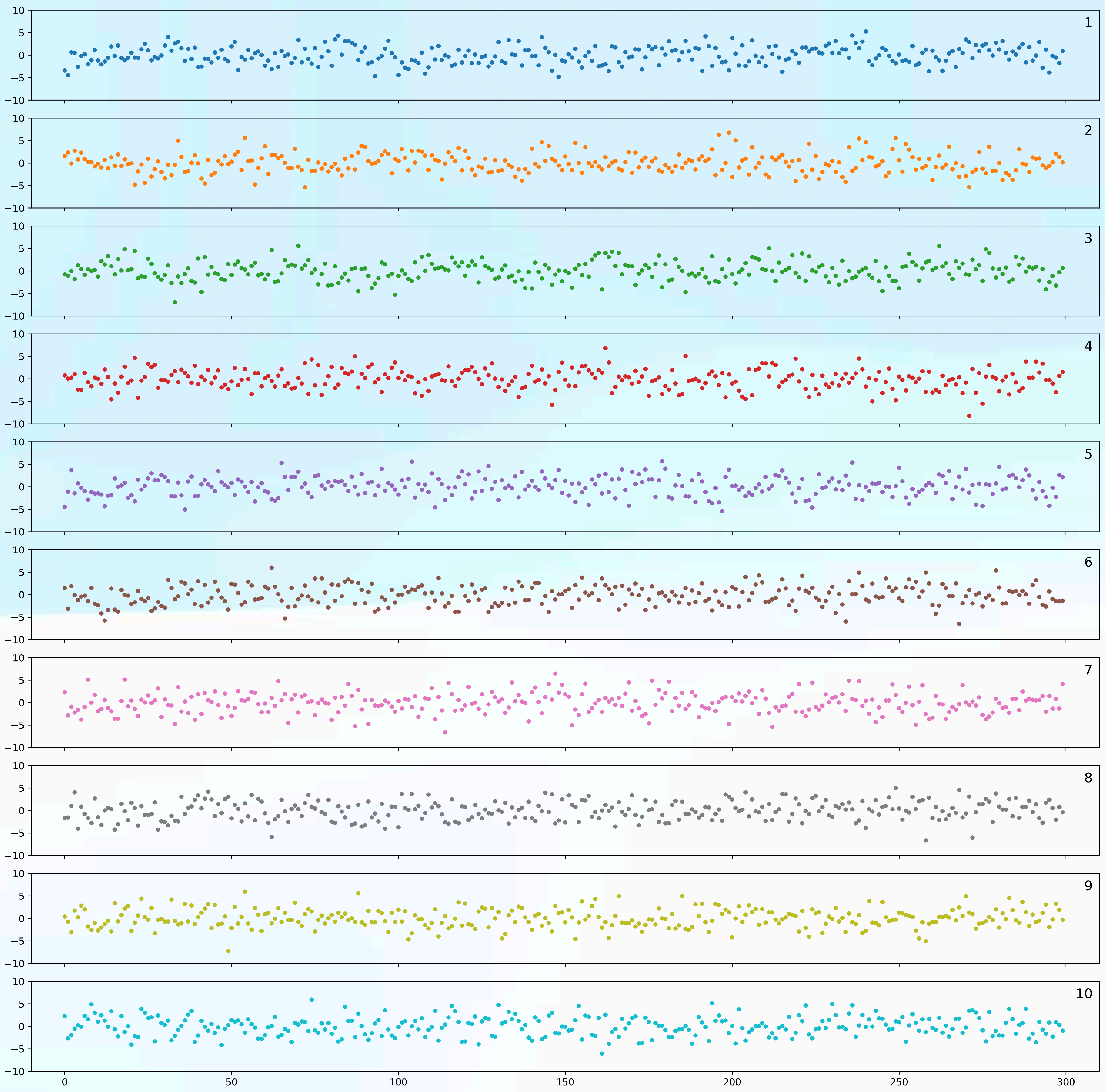
Synthetic example

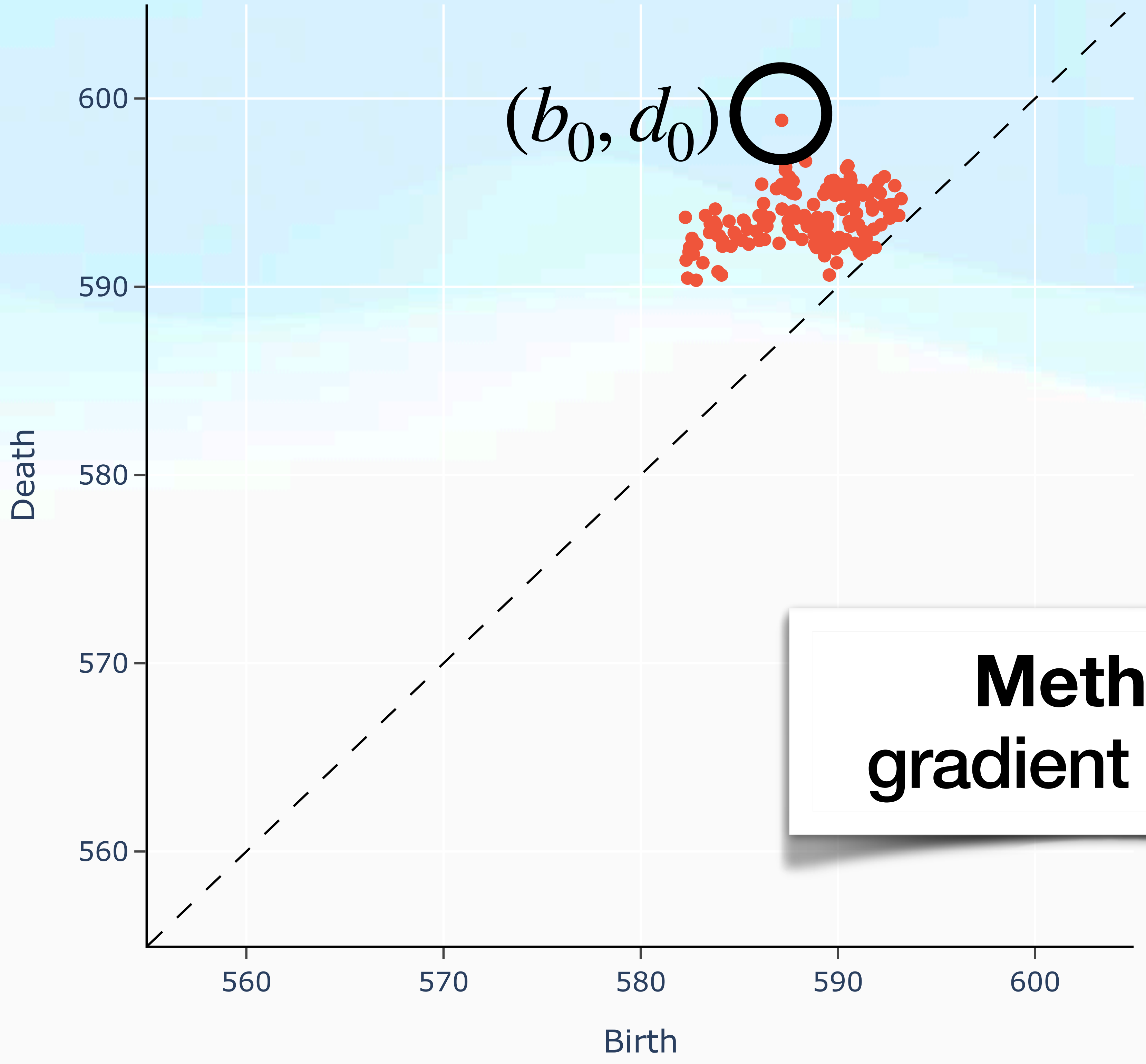




PD of sliding window embedding, window length = 250

Synthetic example



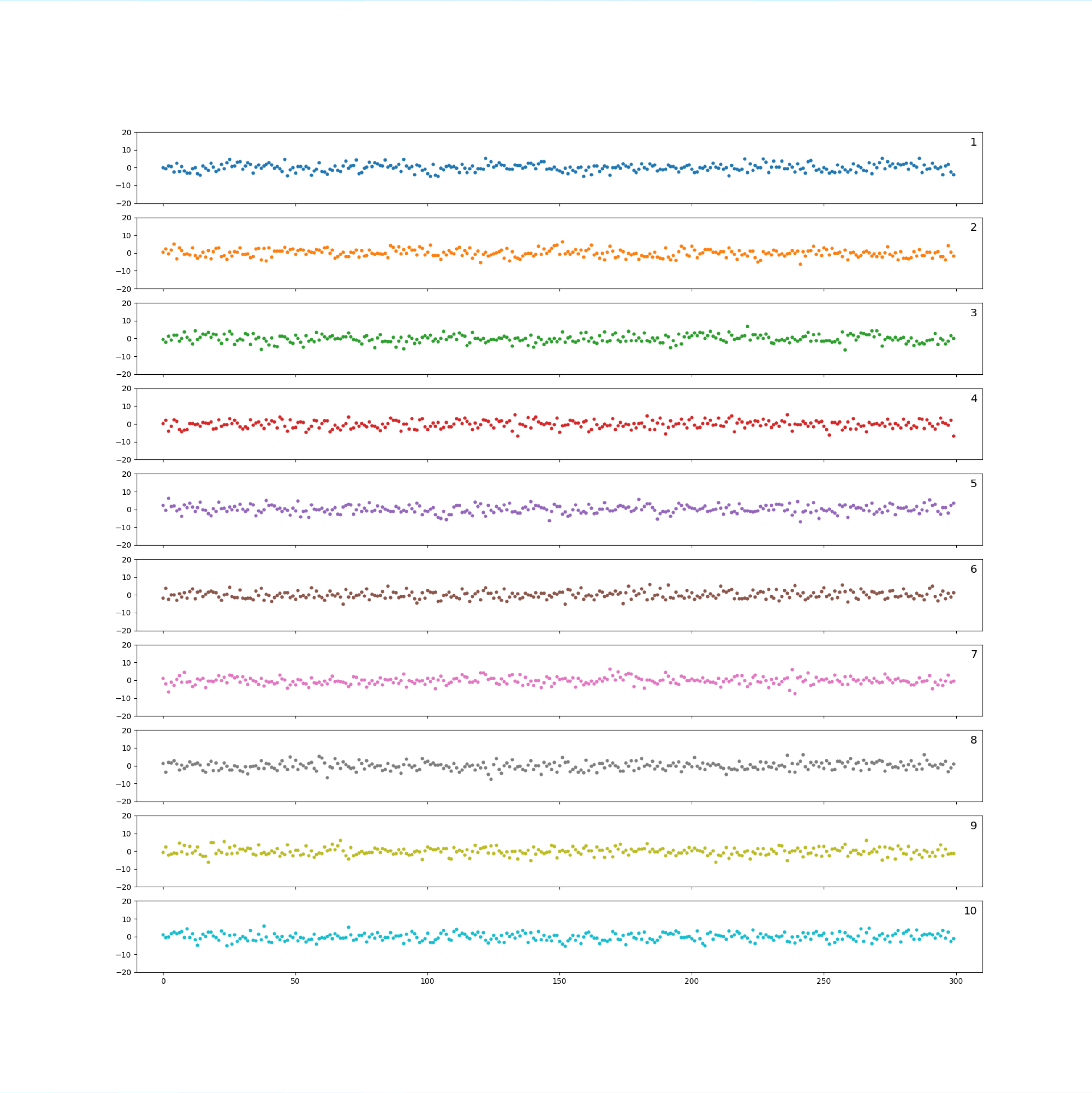


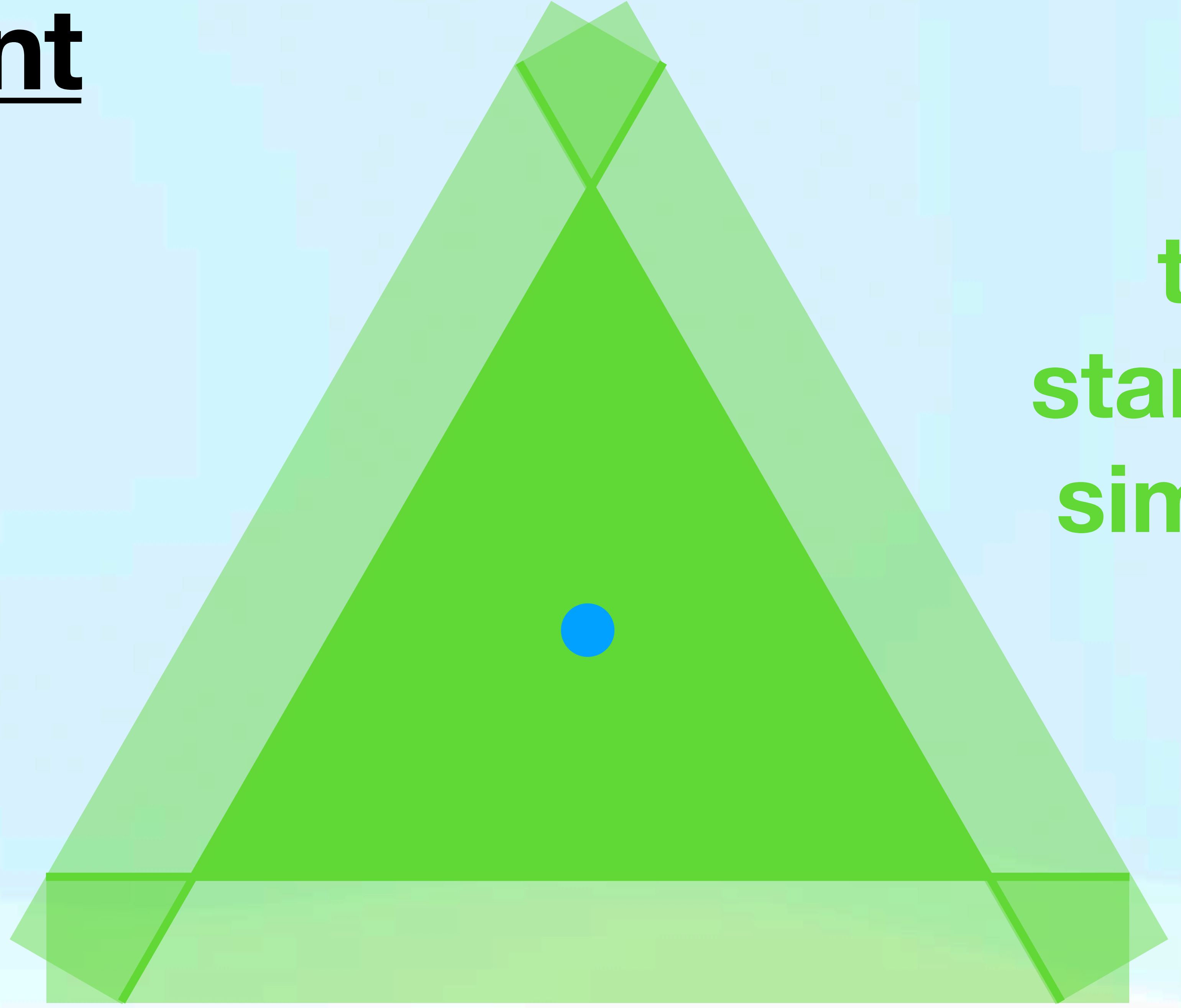
Goal:
rescale the signals
to push (b_0, d_0) further
from the diagonal

Method:
gradient ascent

PD of sliding window embedding, window length = 250

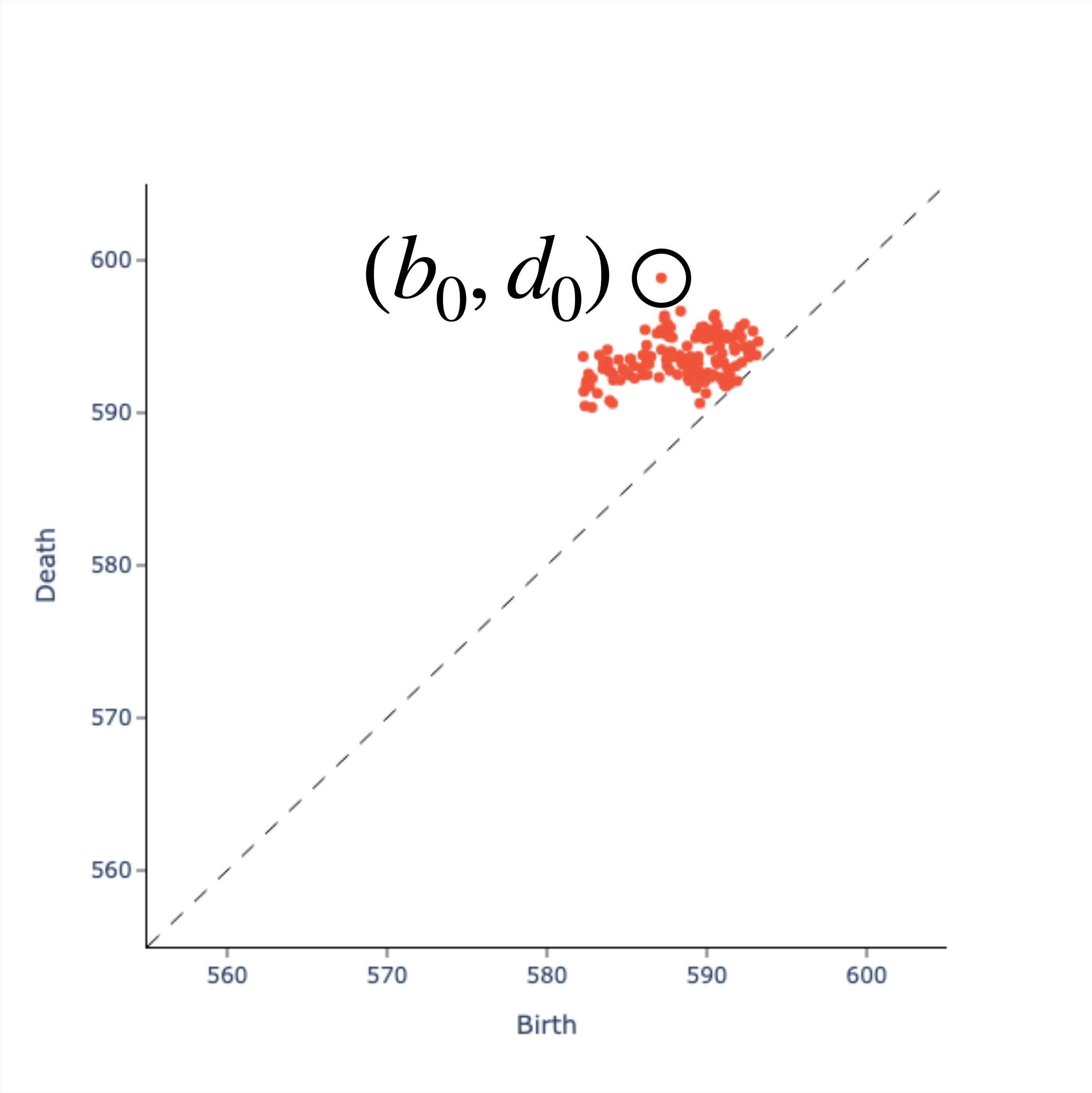
Gradient ascent



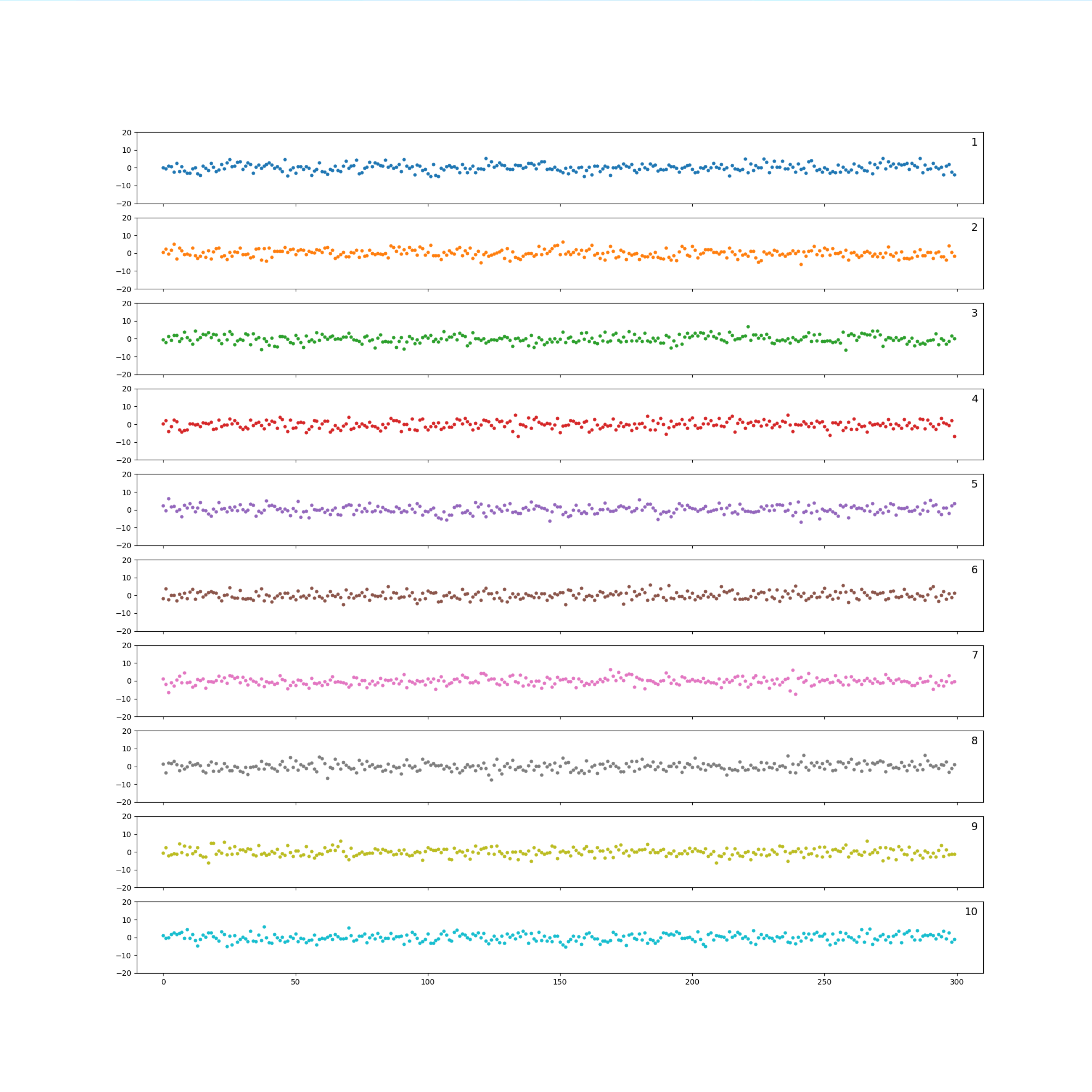


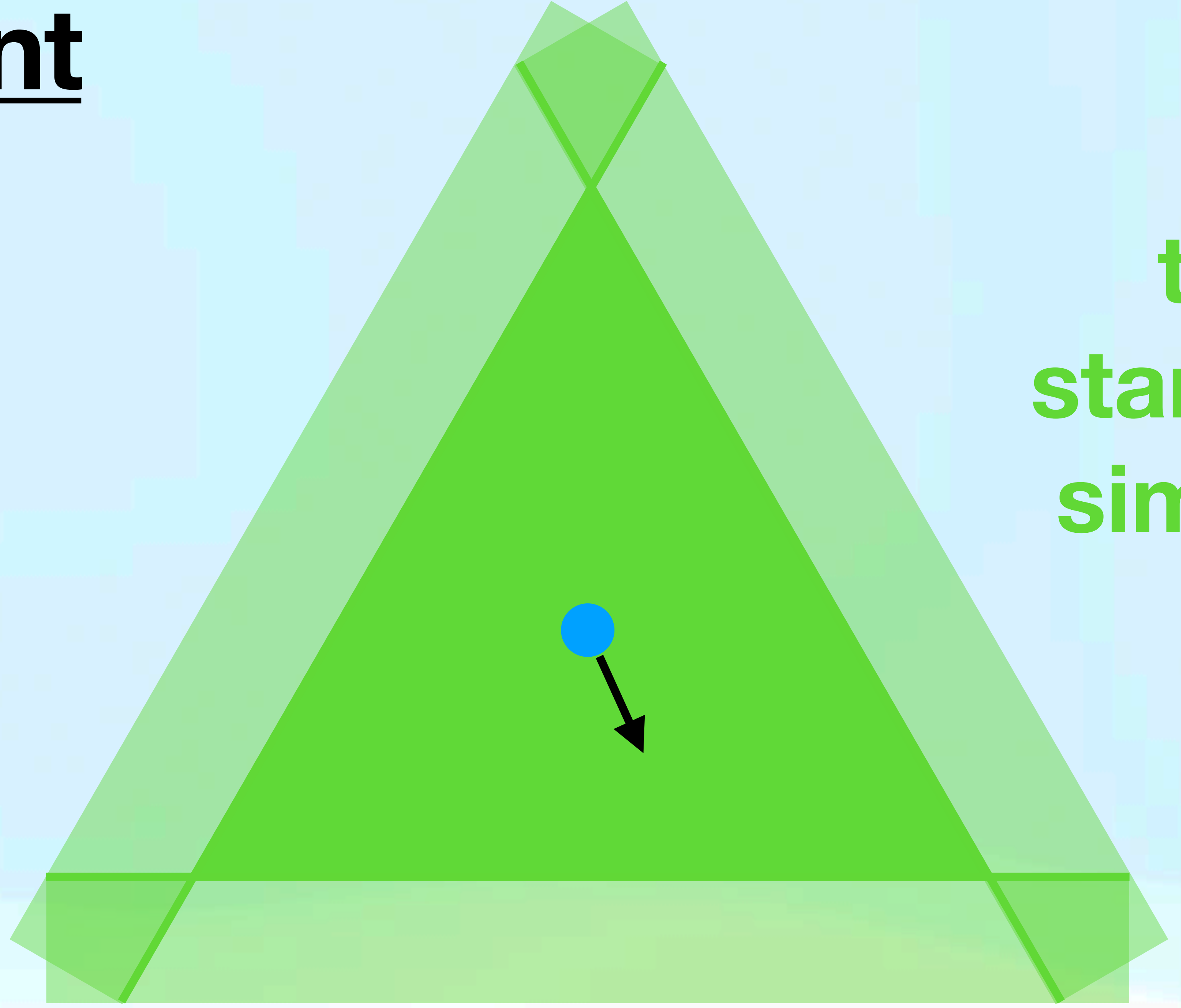
start at the barycenter,

$$\bullet = \left(\frac{1}{10},\frac{1}{10},\frac{1}{10},\frac{1}{10},\frac{1}{10},\frac{1}{10},\frac{1}{10},\frac{1}{10},\frac{1}{10},\frac{1}{10}\right)$$

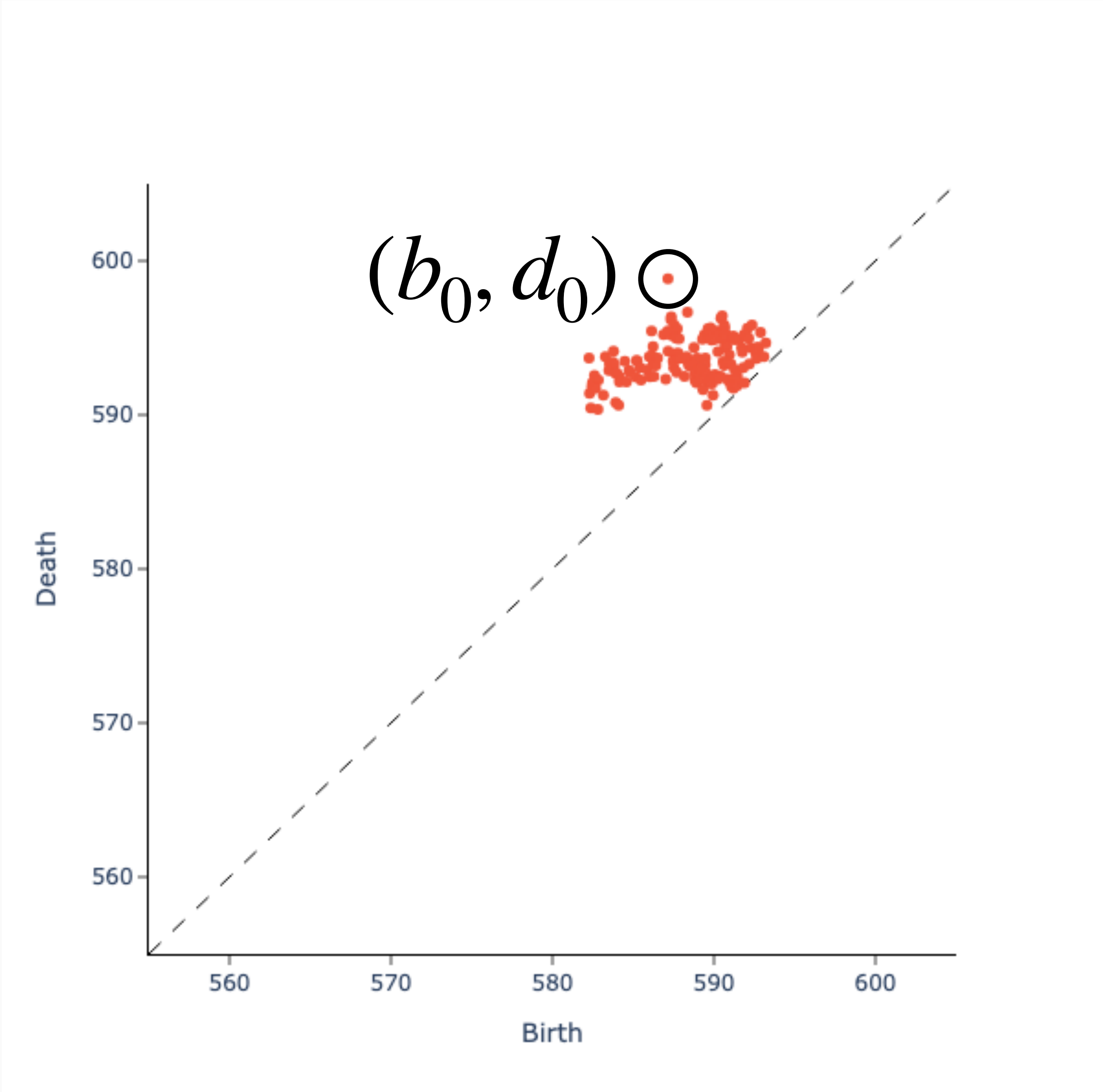


Gradient ascent





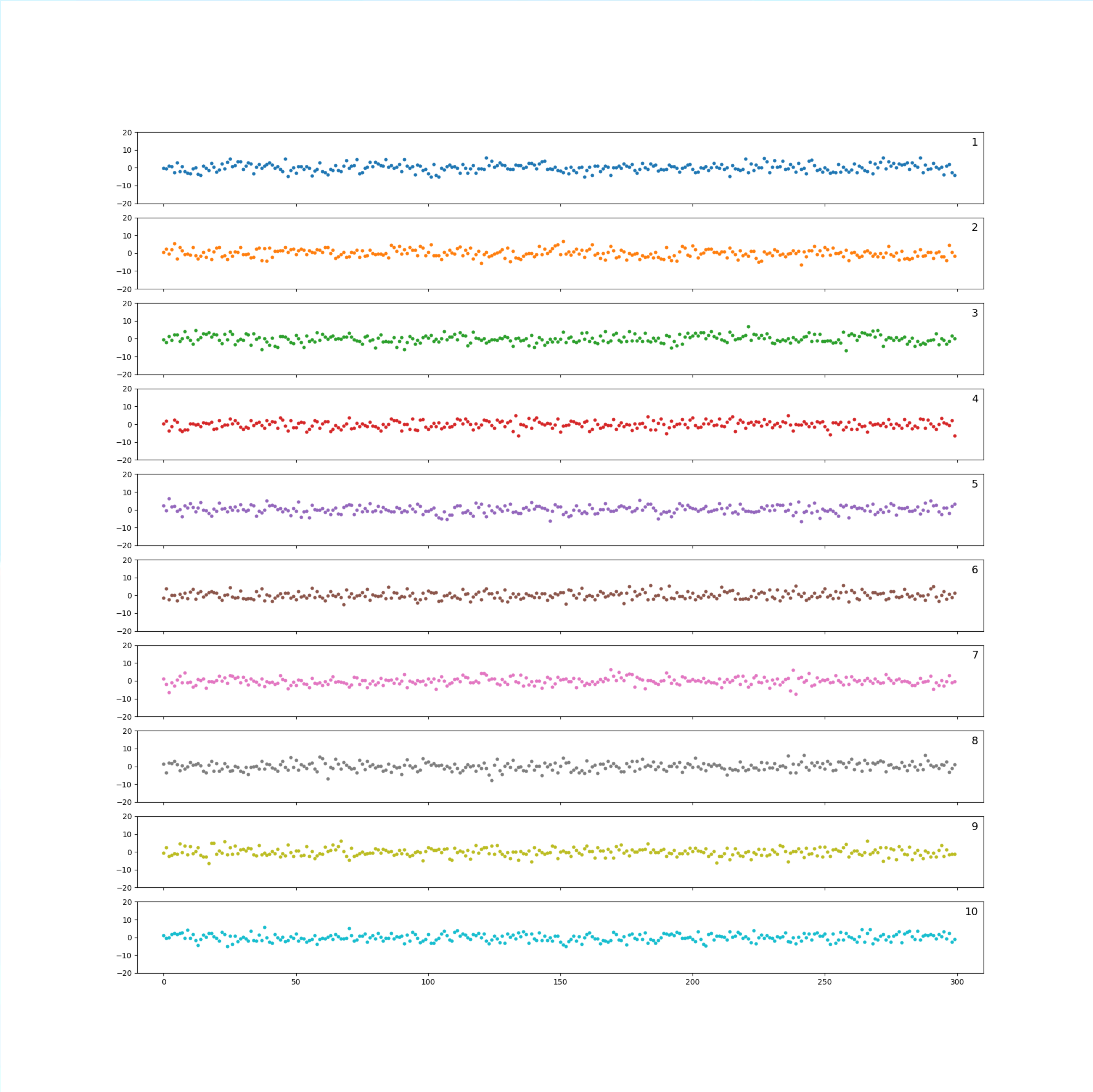
the standard simplex

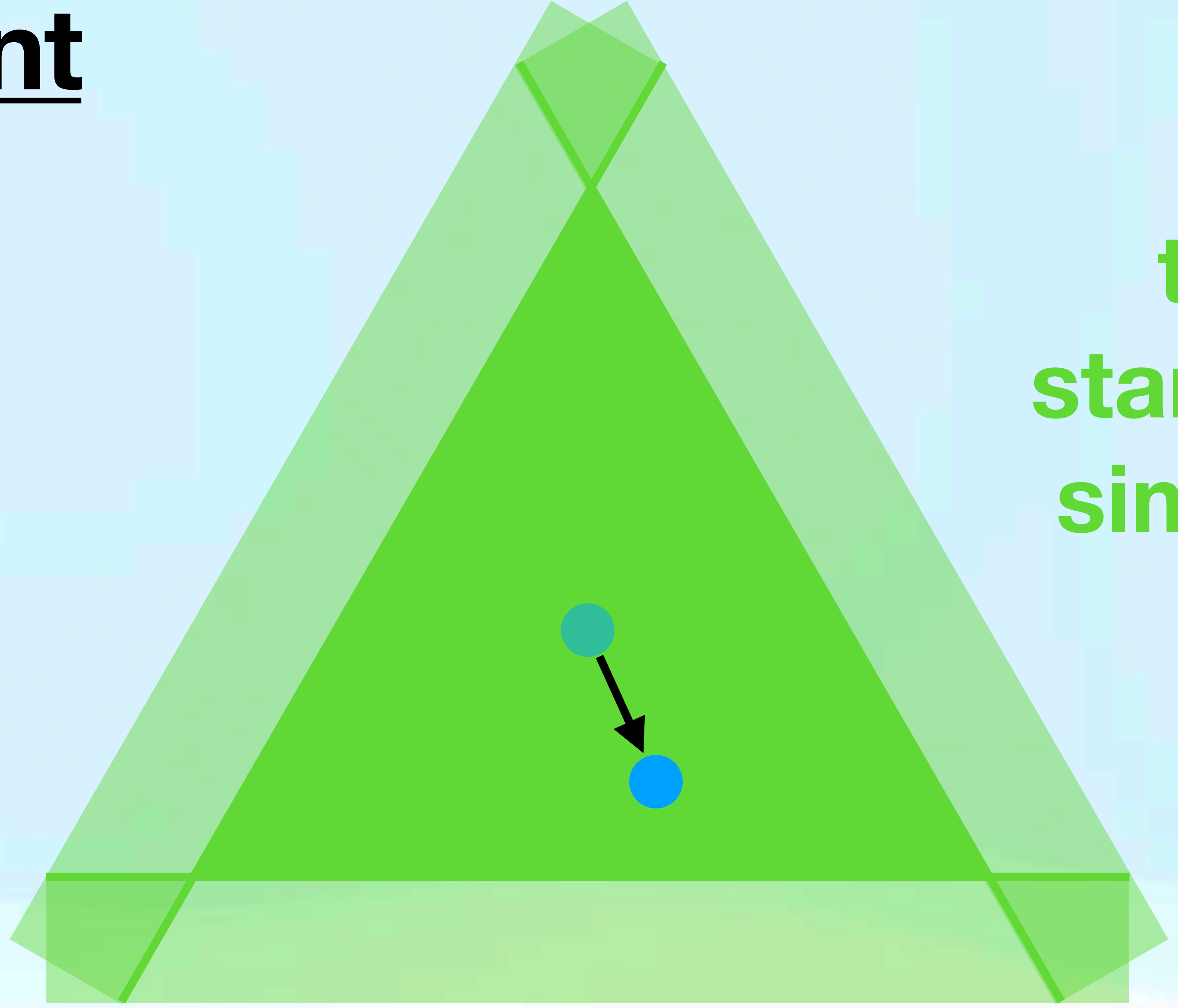


given $f\colon \bullet \mapsto d_0 - b_0$

compute $\nabla f(\bullet)$

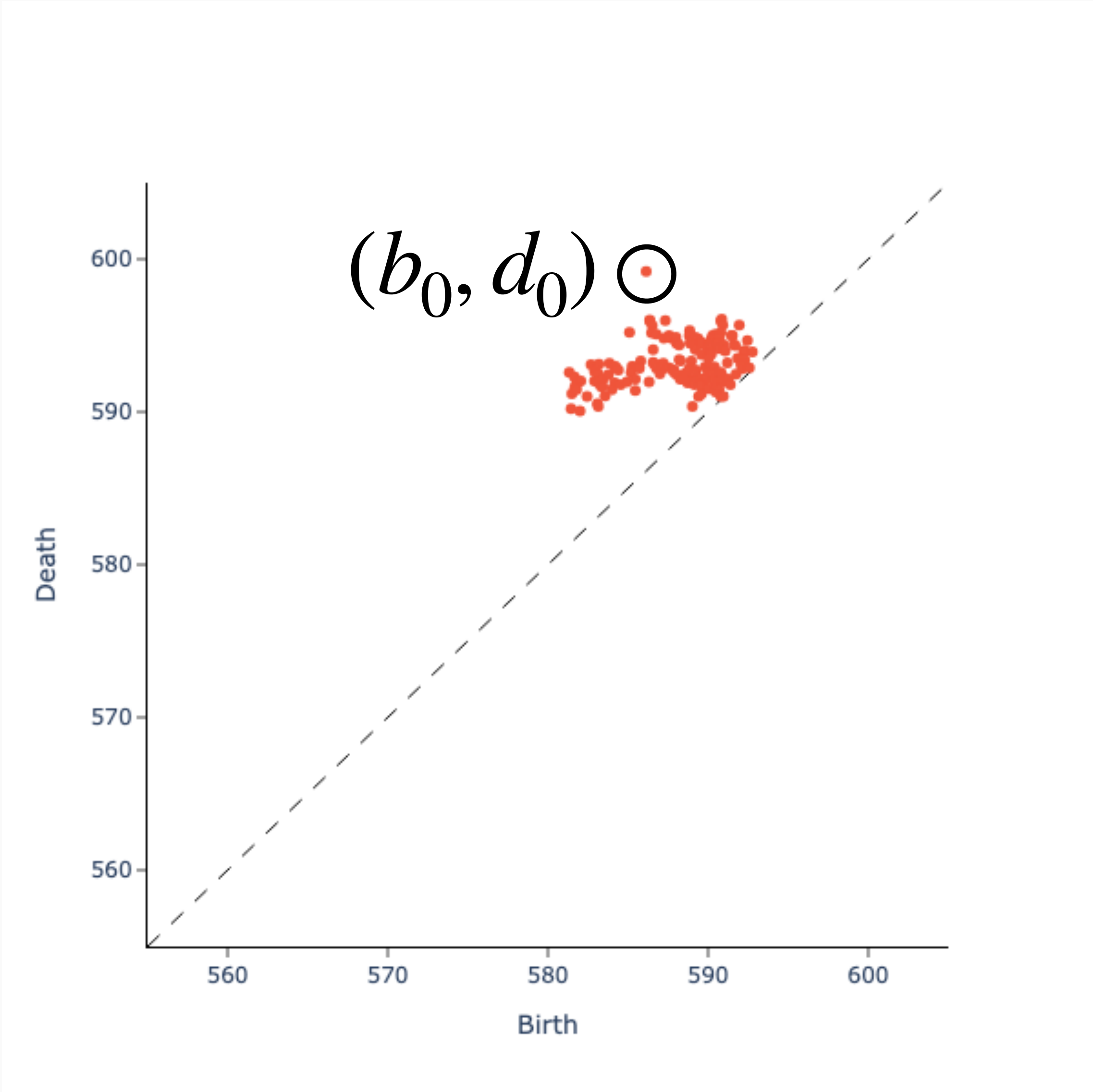
Gradient ascent





the standard simplex

update

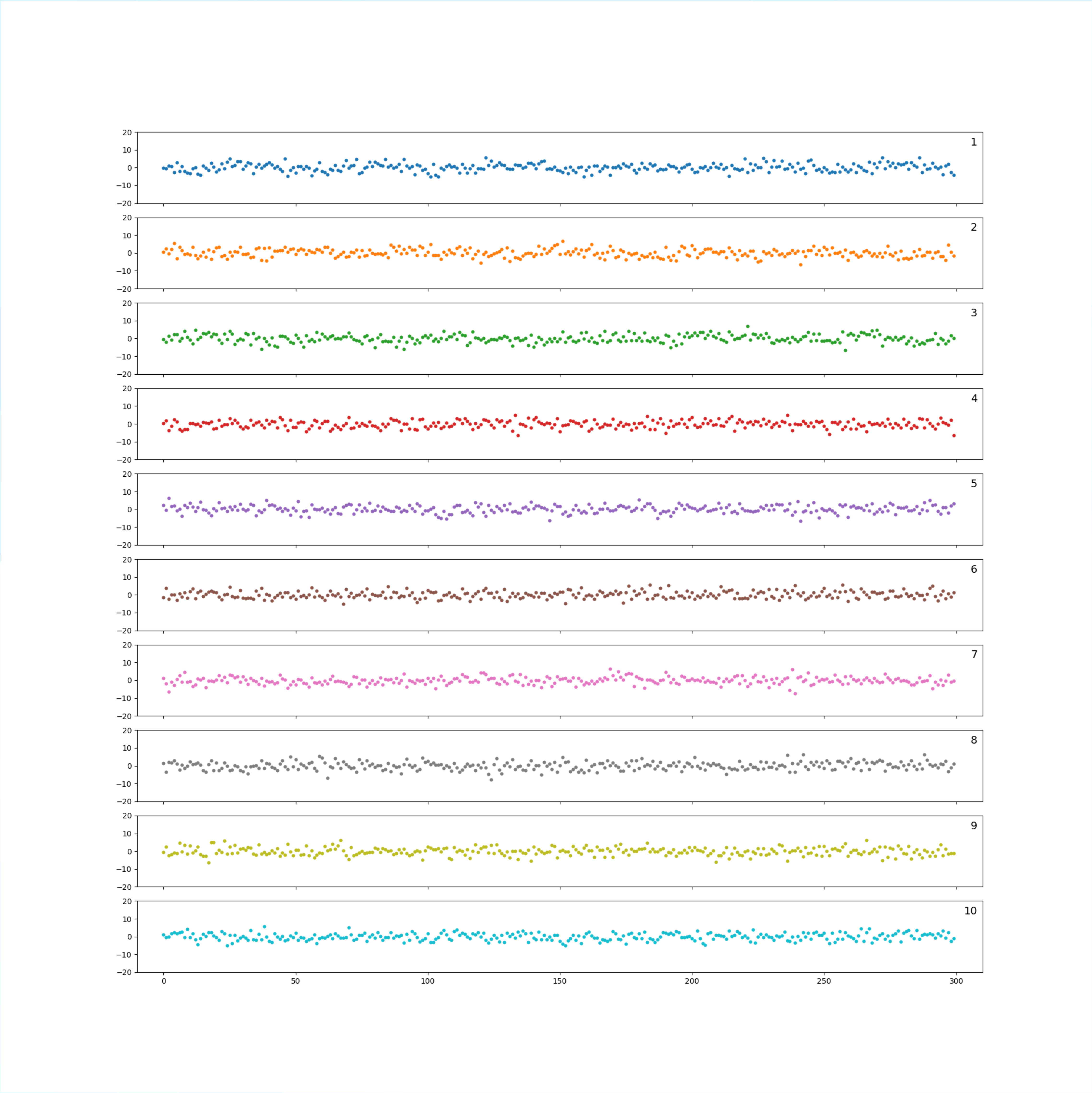


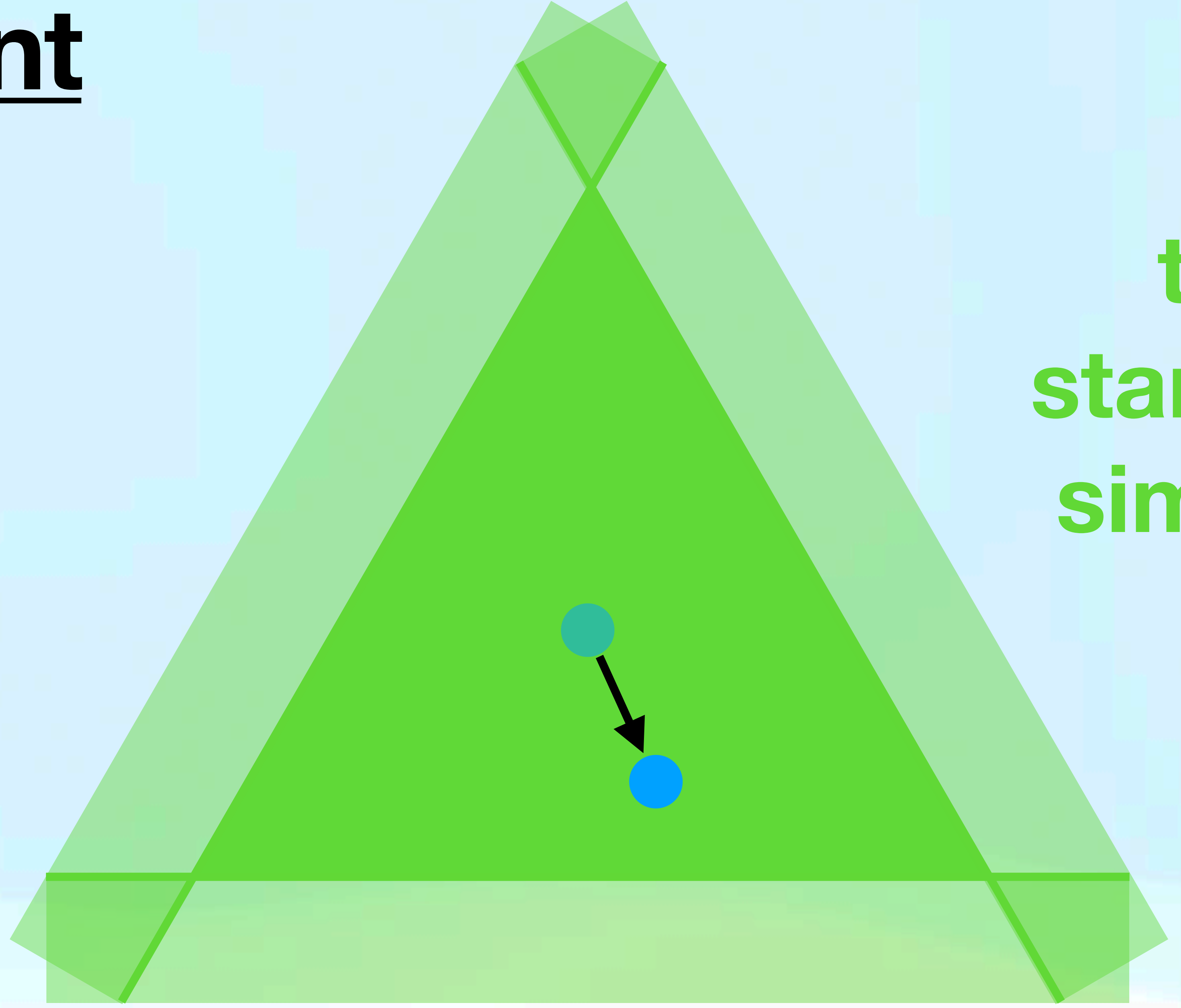
given f: ● $\mapsto d_0 - b_0$

compute $\nabla f($●$)$

update

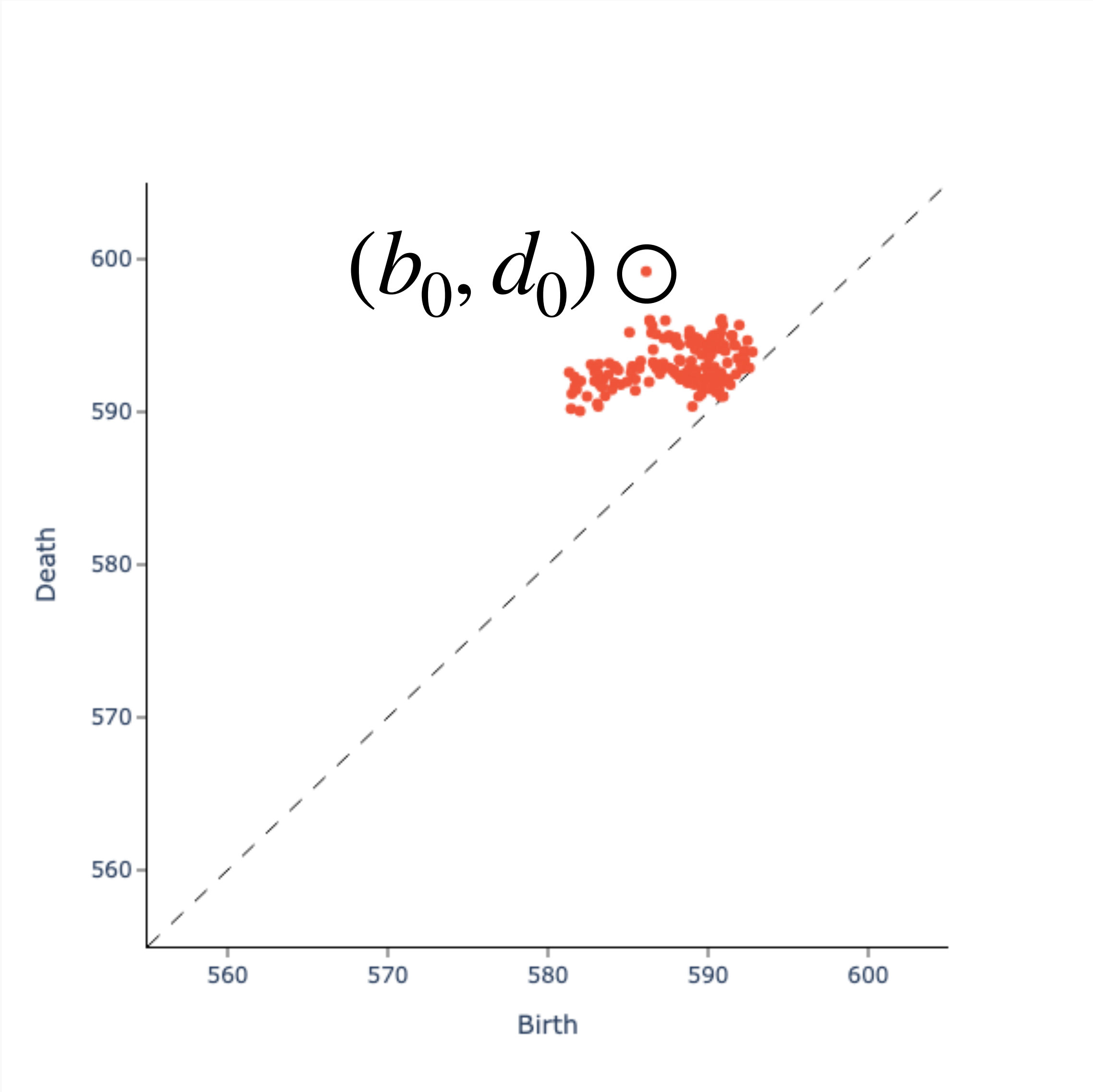
Gradient ascent





the standard simplex

update



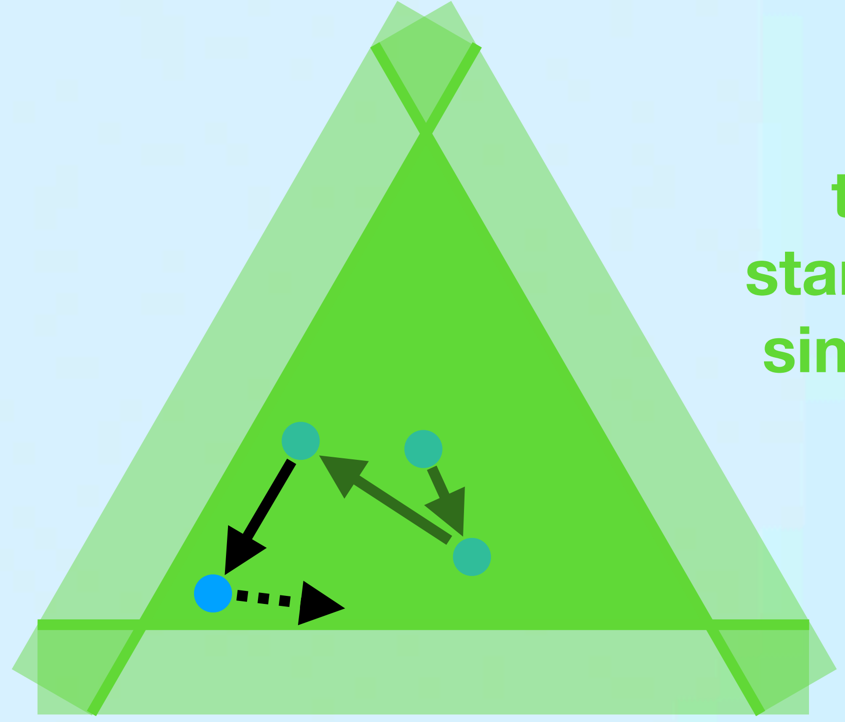
given $f\colon$ ● $\mapsto d_0 - b_0$

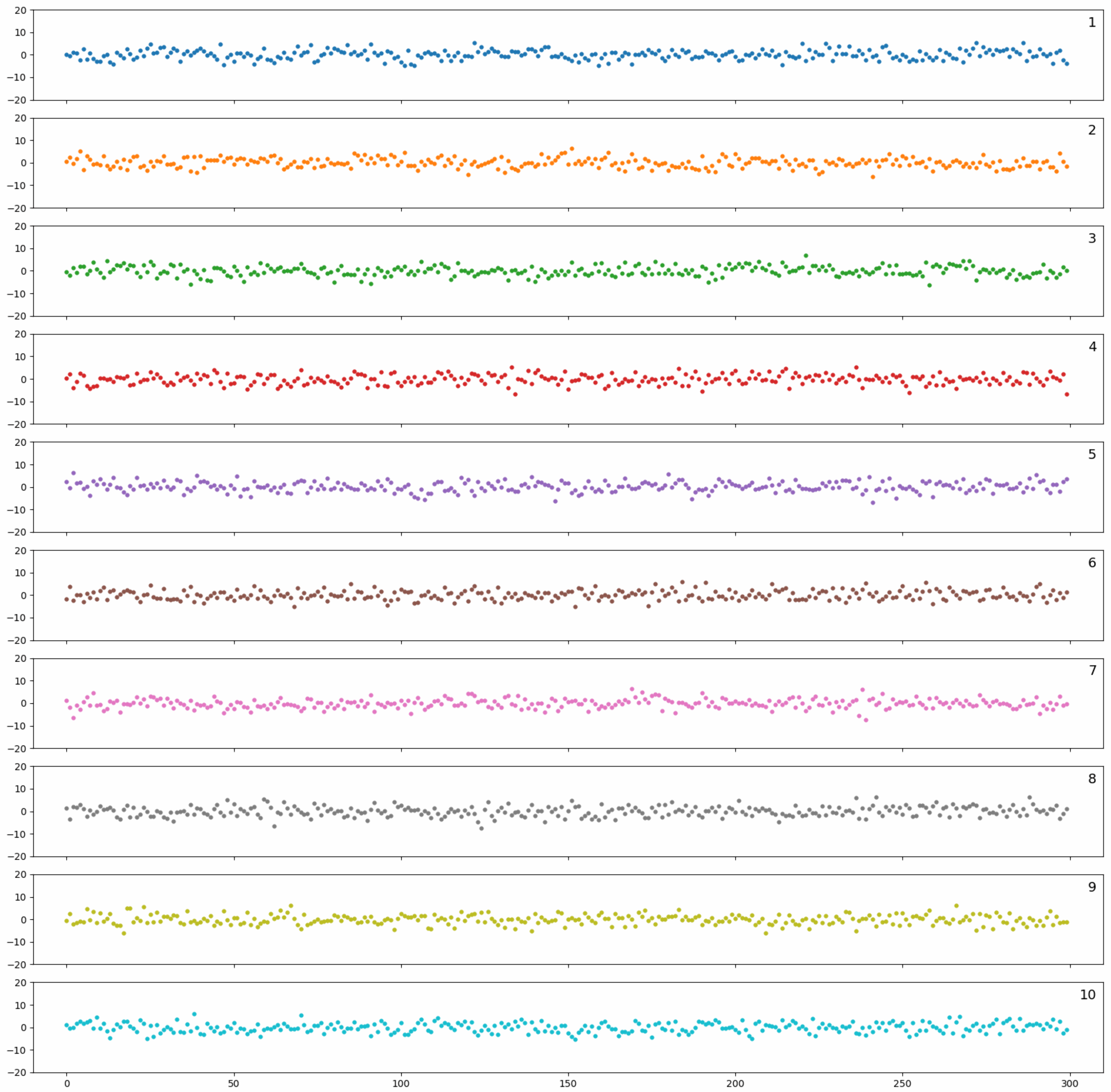
compute $\nabla f($●$)$

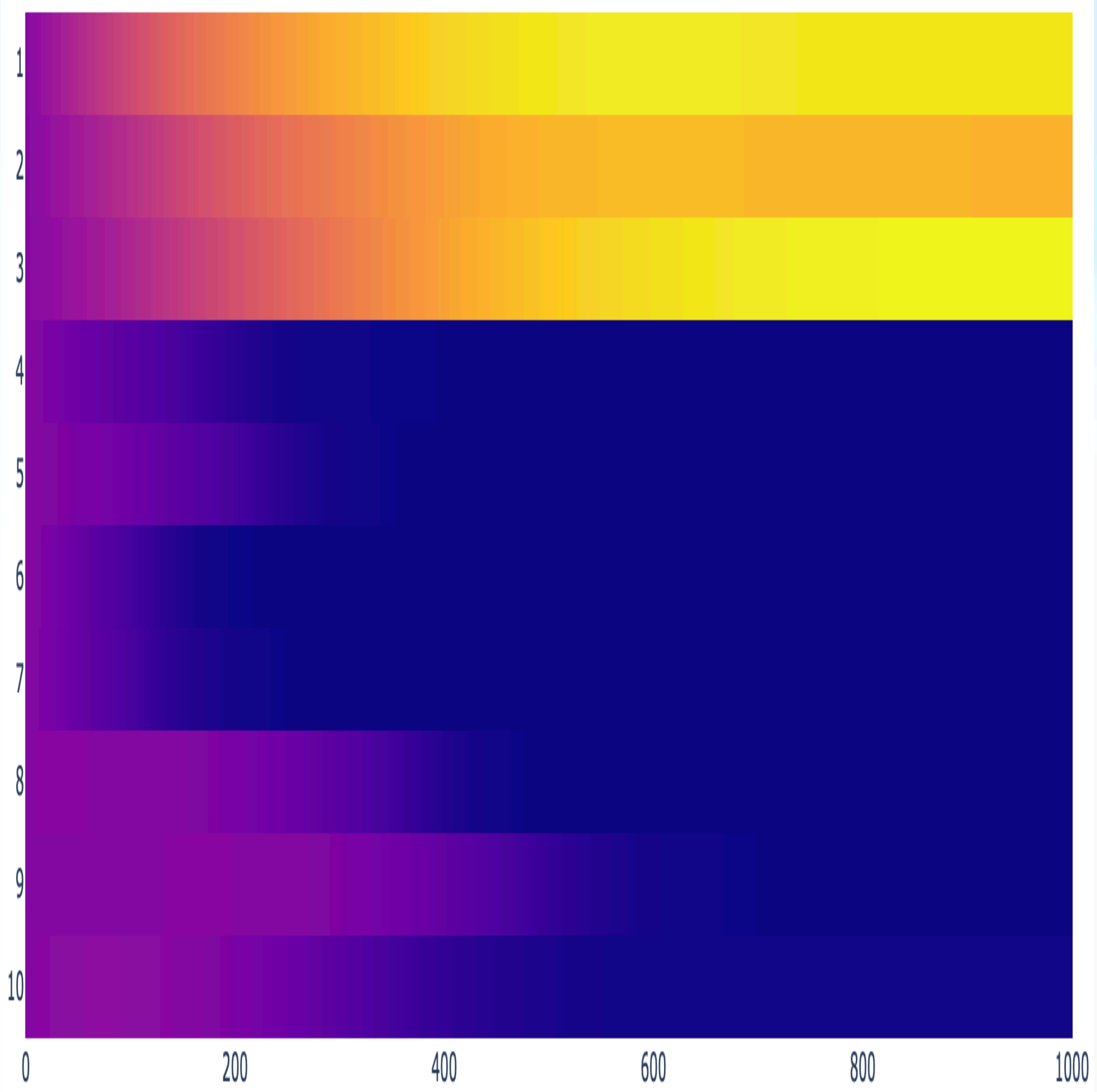
update

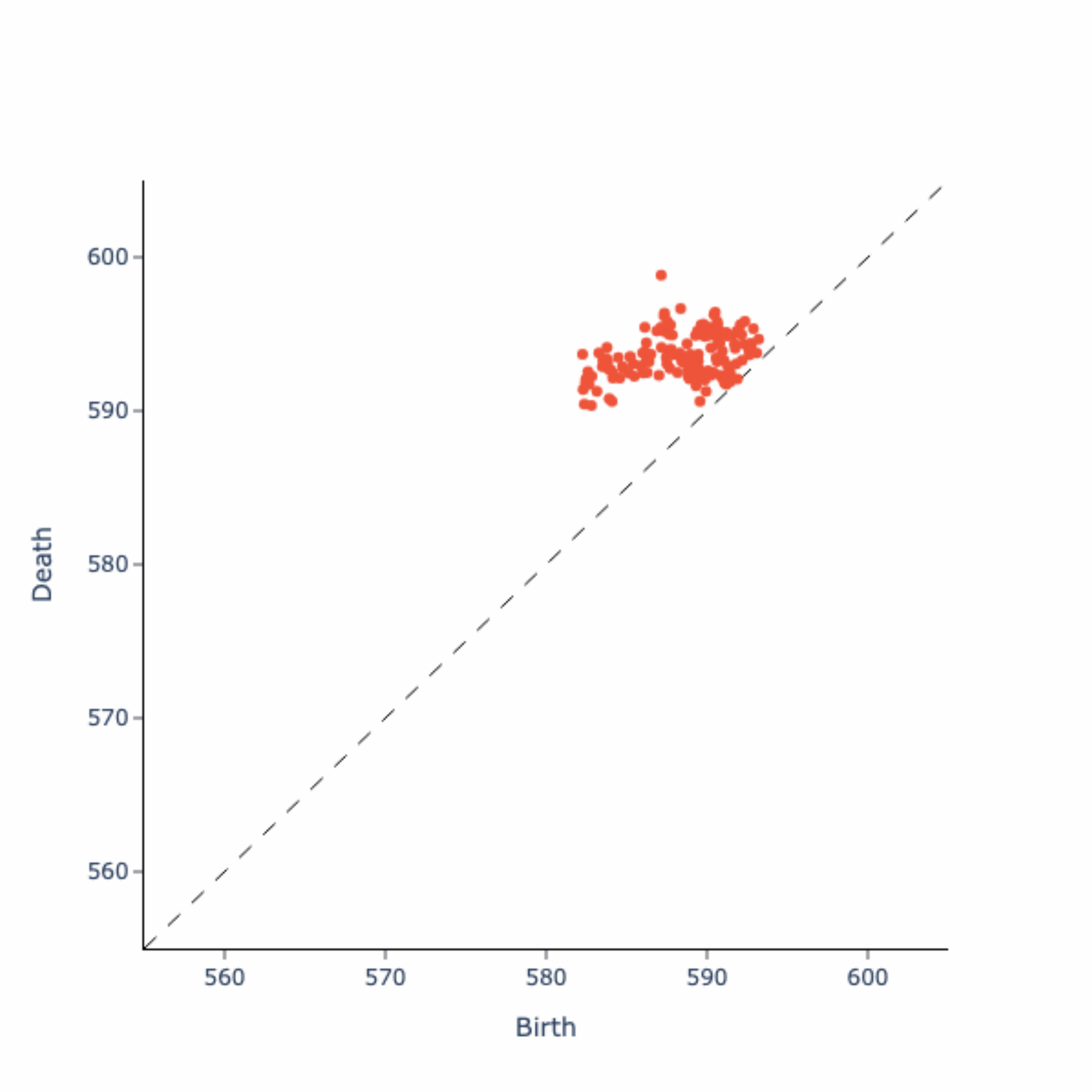
repeat

Synthetic example









path through the simplex

start (barycenter):

$$(0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1)$$

end (score):

$$(\underbrace{0.342, 0.298, 0.358}_{99.7\,\% \text{ of total}}, 0.000, 0.000, 0.000, 0.000, 0.000, 0.000, 0.027)$$

Theory

Topological Feature Selection for Time Series Data

Given time series $X = (x_1, \ldots, x_n)$ in $\mathbb{R}^p$

Consider $v \in \mathbb{R}^p$ with $v_j \geq 0$ and $v_1 + \cdots + v_p = 1$. That is, $v \in |\Delta^{p-1}|$

Define vX : $\forall i, j$, the jth coordinate of x_i is rescaled by v_j

Define $\Phi(v)$ to be the persistence diagram of the VR complex of the SWE of vX

F is a real-valued function on persistence diagrams

Maximize $F(\Phi(v))$

Subject $v \in |\Delta^{p-1}|$

Sliding window embeddings

We will use distances induced by 1 norms

Given time series $X = (x_1, \ldots, x_n)$ in $\mathbb{R}^p$

Recall: for $v \in |\Delta^{p-1}|$ we have the time series vX in $\mathbb{R}^p$

Define vD to be the distance matrix of the SWE of vX

Let vX^j to be the jth component of vX

Define vD^j to be the distance matrix of the SWE of vX^j

Lemma: $vD = vD^1 + \cdots + vD^p$

Topological Feature Selection for Matrices

Given $m \times m$ (distance) matrices $M^1, \ldots, M^p$

Consider $v \in |\Delta^{p-1}|$

Let $vM = v_1M^1 + \cdots + v_pM^p$

Define $\Phi(v)$ to be the persistence diagram of the VR complex of vM

F is a real-valued function on persistence diagrams

Maximize $F(\Phi(v))$

Subject $v \in |\Delta^{p-1}|$

The augmented combinatorial simplex

Δ^{m-1} consists of all non-empty subsets of $\{1,\ldots,m\}$

Let K be the collection of all subsets of $\{1,\ldots,m\}$

A **weight** is an order-preserving map $w : (K, \subseteq) \to (\mathbb{R}, \leq)$

Example: the VR complex of a $m \times m$ distance matrix

Since Δ^{m-1} is contractible, the homology of K is 0

In the persistent homology of (K, w), all simplices are paired

Topological Feature Selection for Weights

Given weights $w_1, \ldots, w_p$ for the augmented combinatorial simplex K

Consider $v \in |\Delta^{p-1}|$

Have weight $v_1 w_1 + \cdots + v_p w_p$

Define $\Phi(v)$ to be the persistence diagram of K with weight $v_1 w_1 + \cdots + v_p w_p$

F is a real-valued function on persistence diagrams

$$\text{Maximize} \quad F(\Phi(v))$$

$$\text{Subject} \quad v \in |\Delta^{p-1}|$$

PL Vineyards

A vineyard is a one-parameter family of persistence diagrams

Recall: K is the augmented $(m-1)$-combinatorial simplex

Let $w(-)$ be a PL curve of weights on K

Theorem: The vineyard of the persistence diagrams of $(K, w(-))$ is PL

Corollary: Persistence diagrams are stable for the q-Wasserstein distances (Cohen-Steiner–Edelsbrunner–Morozov, Skraba-Turner)

Topological Feature Selection for Weights

Recall: $v \in |\Delta^{p-1}|$, $\Phi(v)$ is the PD of $(K, v_1 w_1 + \cdots + v_p w_p)$

F is a real-valued function on persistence diagrams

Maximize $G(v) := F(\Phi(v))$

Subject $v \in |\Delta^{p-1}|$

Assume F is 1-Lipschitz with respect to q-Wasserstein distance and G is PL
Example: F is total persistence or maximal persistence

Then G has piecewise constant gradients – perform gradient ascent
Get PL path $v(\,-\,) : [0,T] \to |\Delta^{p-1}|$

Theorem: Along the PL gradient path, G is 1-Lipschitz

Mean Gradient Path

Given weights $w_1, \ldots, w_p$ for the augmented combinatorial simplex K,

our gradient ascent algorithm produces a PL path $v_1, v_2, \ldots, v_N \in |\Delta^{p-1}|$

We have a function $v : (\mathbb{R}^K)^p \to |\Delta^{p-1}|^N$

Let $X : (\Omega, \Sigma, P) \to \mathbb{R}^K$ be a random variable

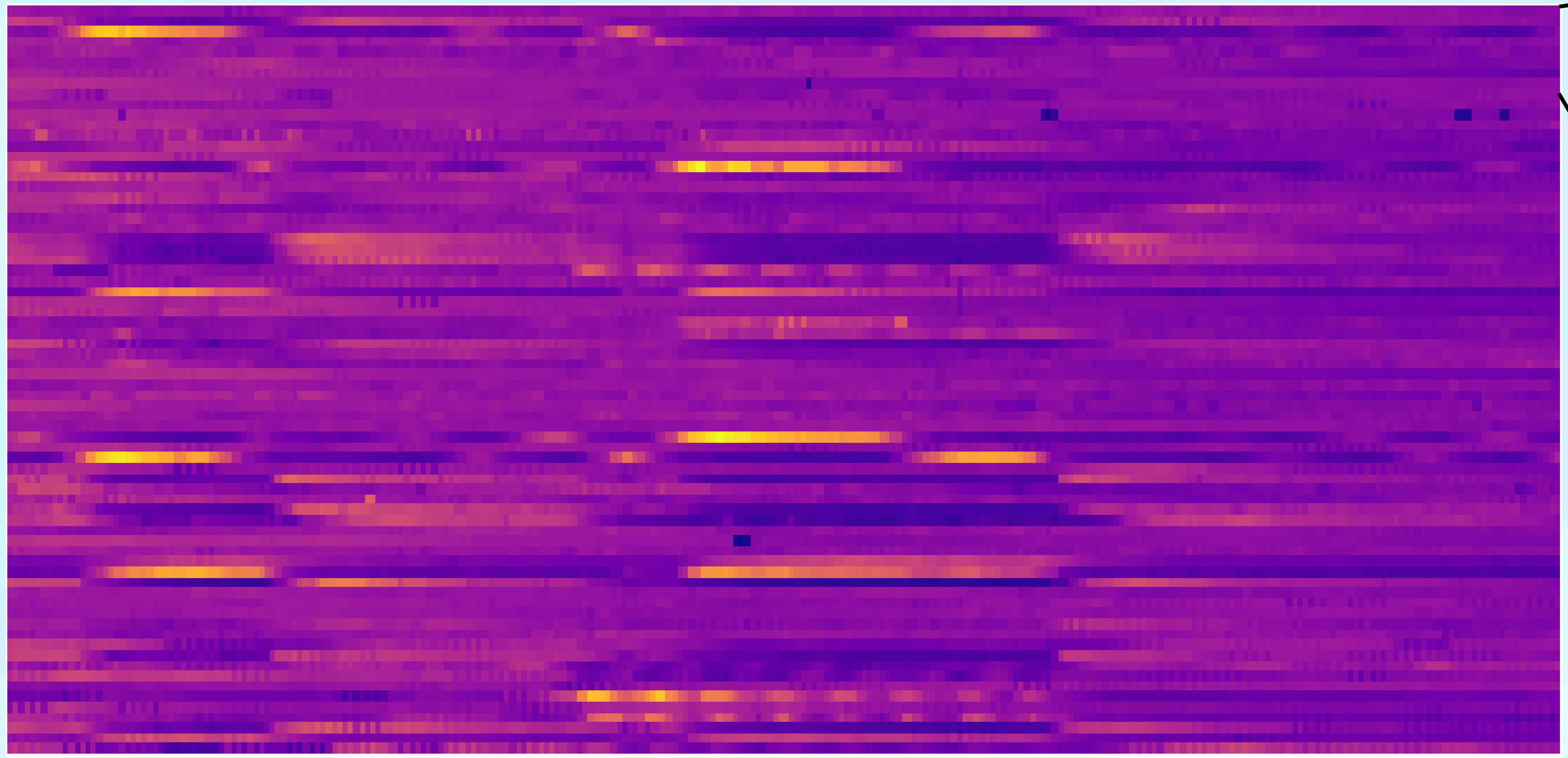
$v \circ X$ is a random gradient path

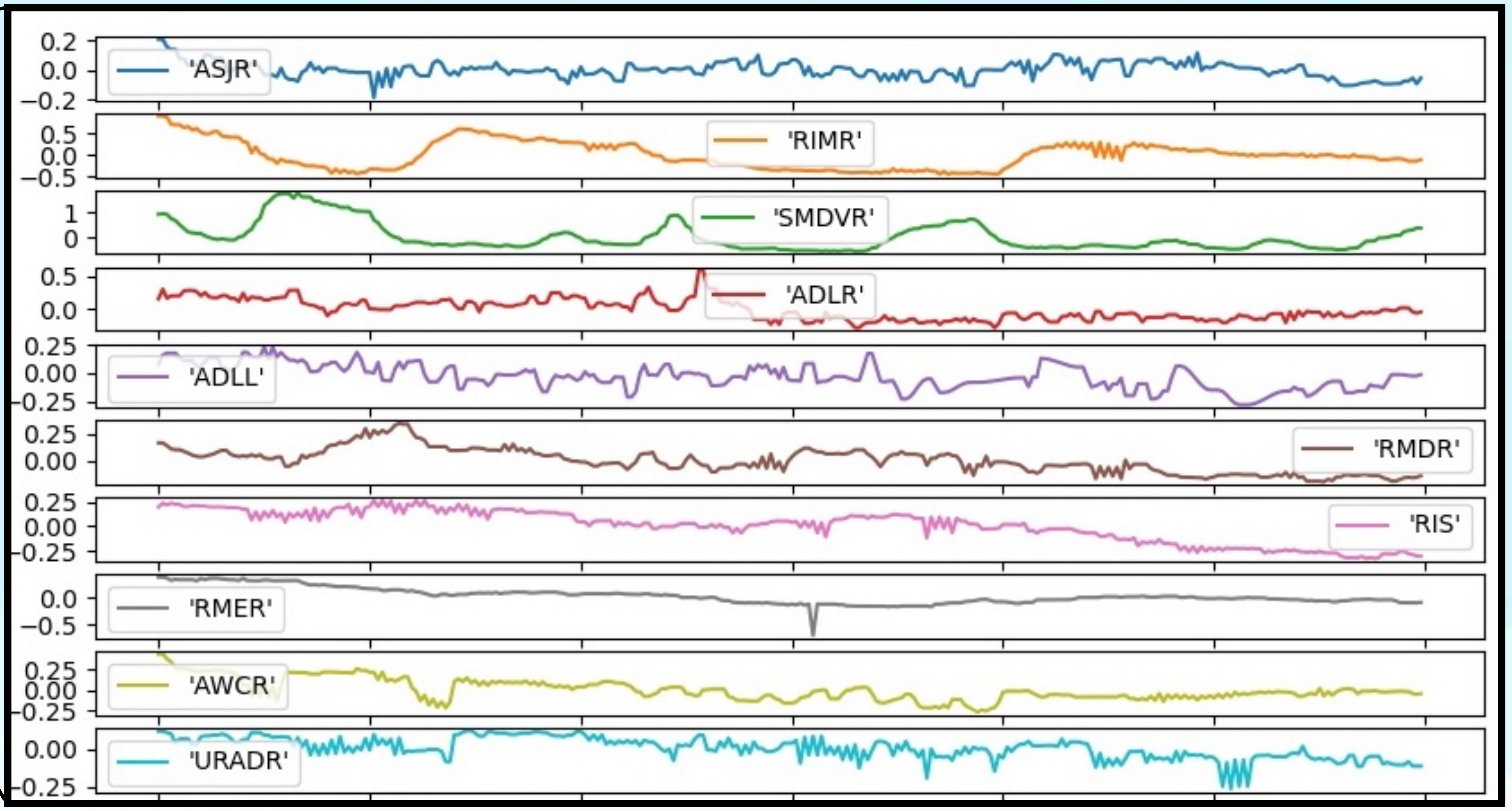
Theorem: $v \circ X$ satisfies a Law of Large Number and Central Limit Theorem

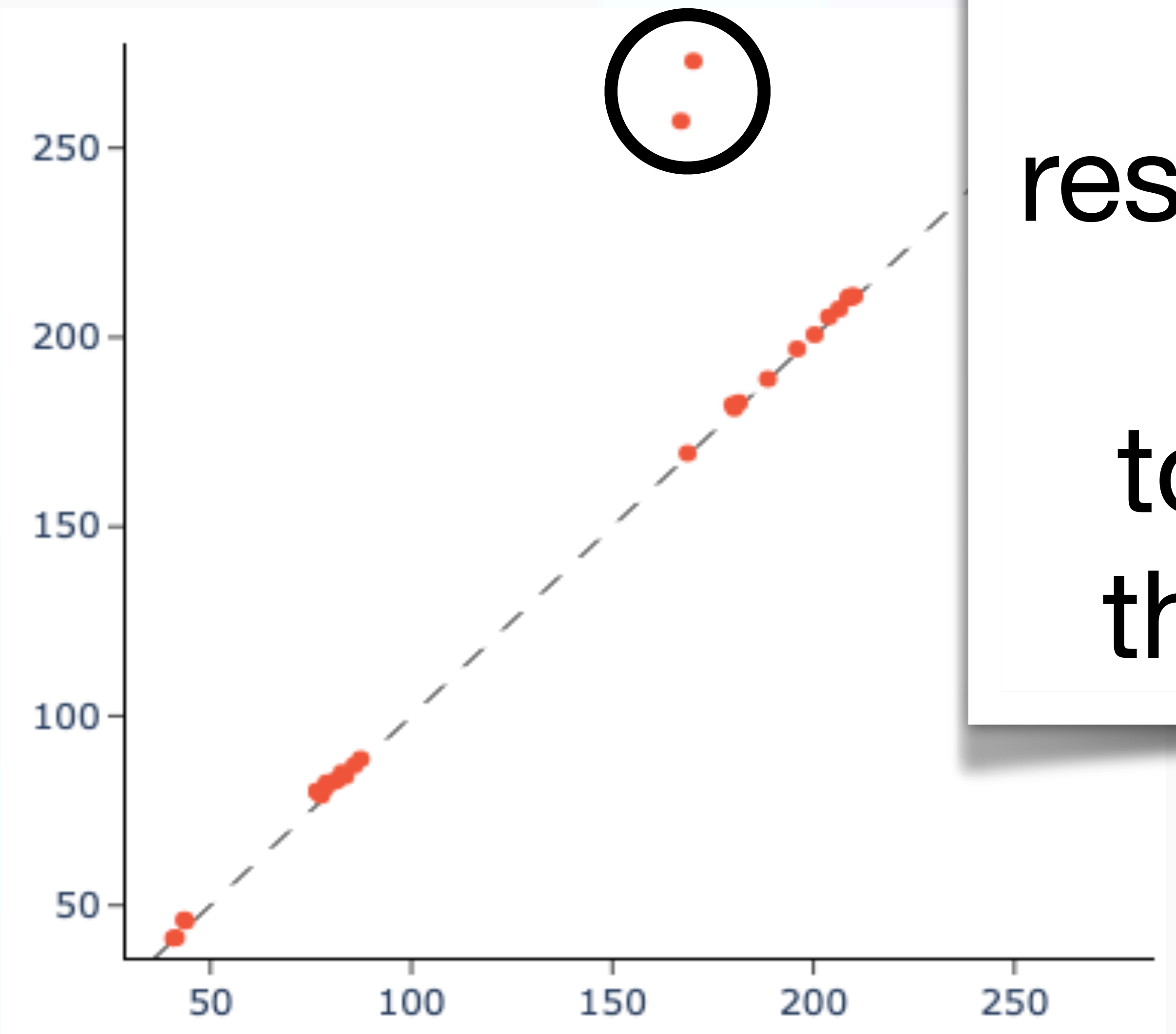
Corollary: The sample mean gradient path converges to the mean gradient path

Application to Biology

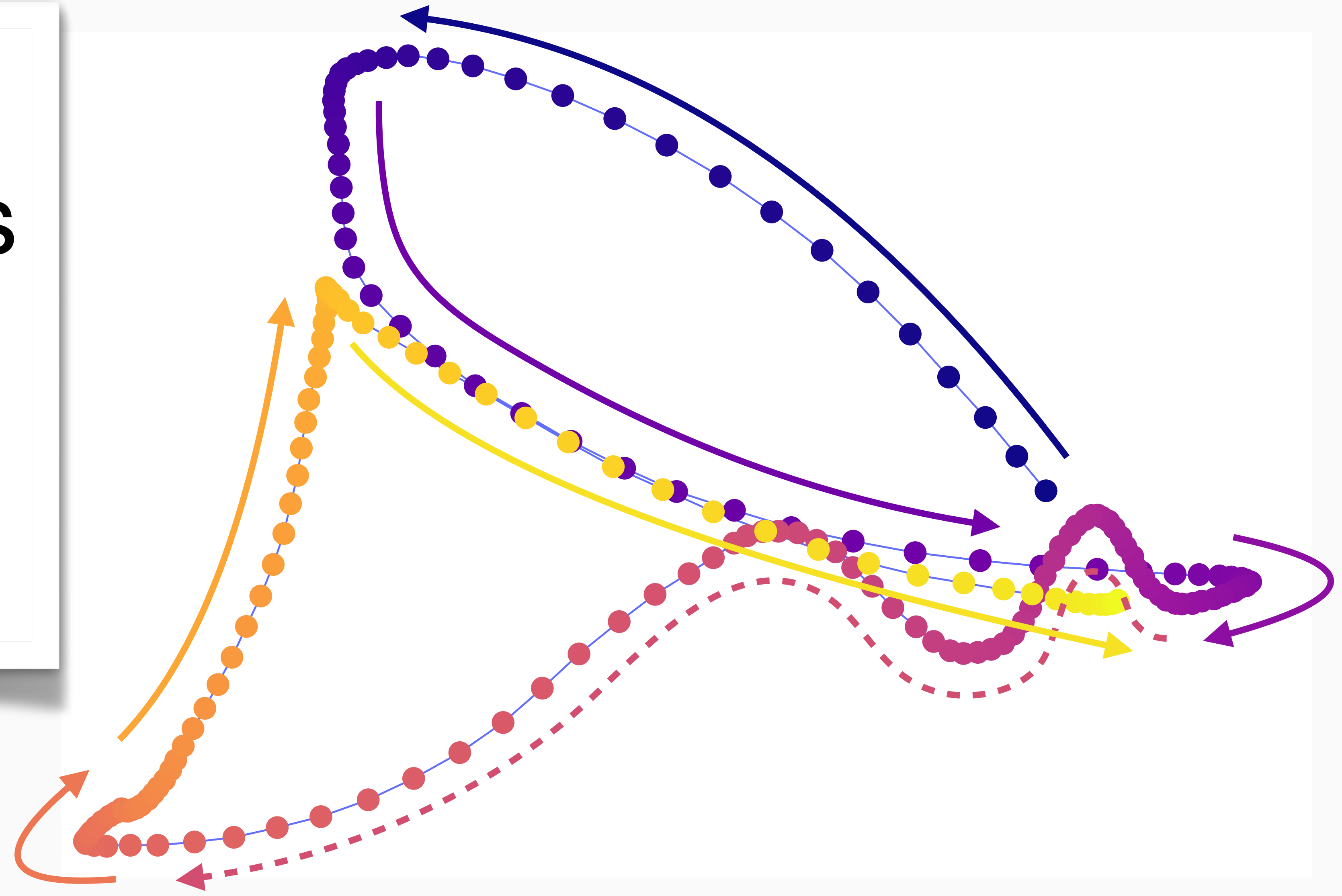
Returning to *C. elegans*…



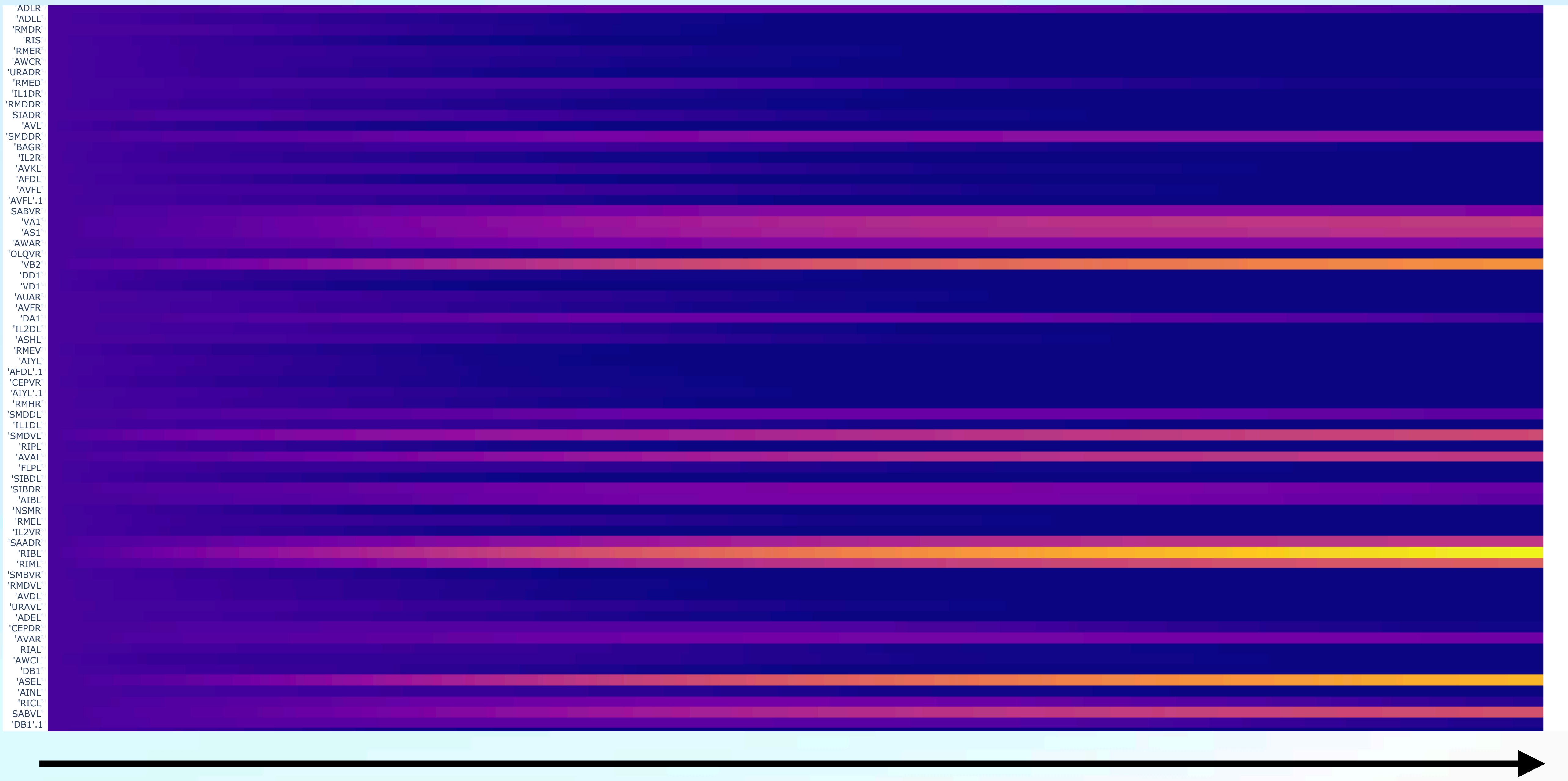




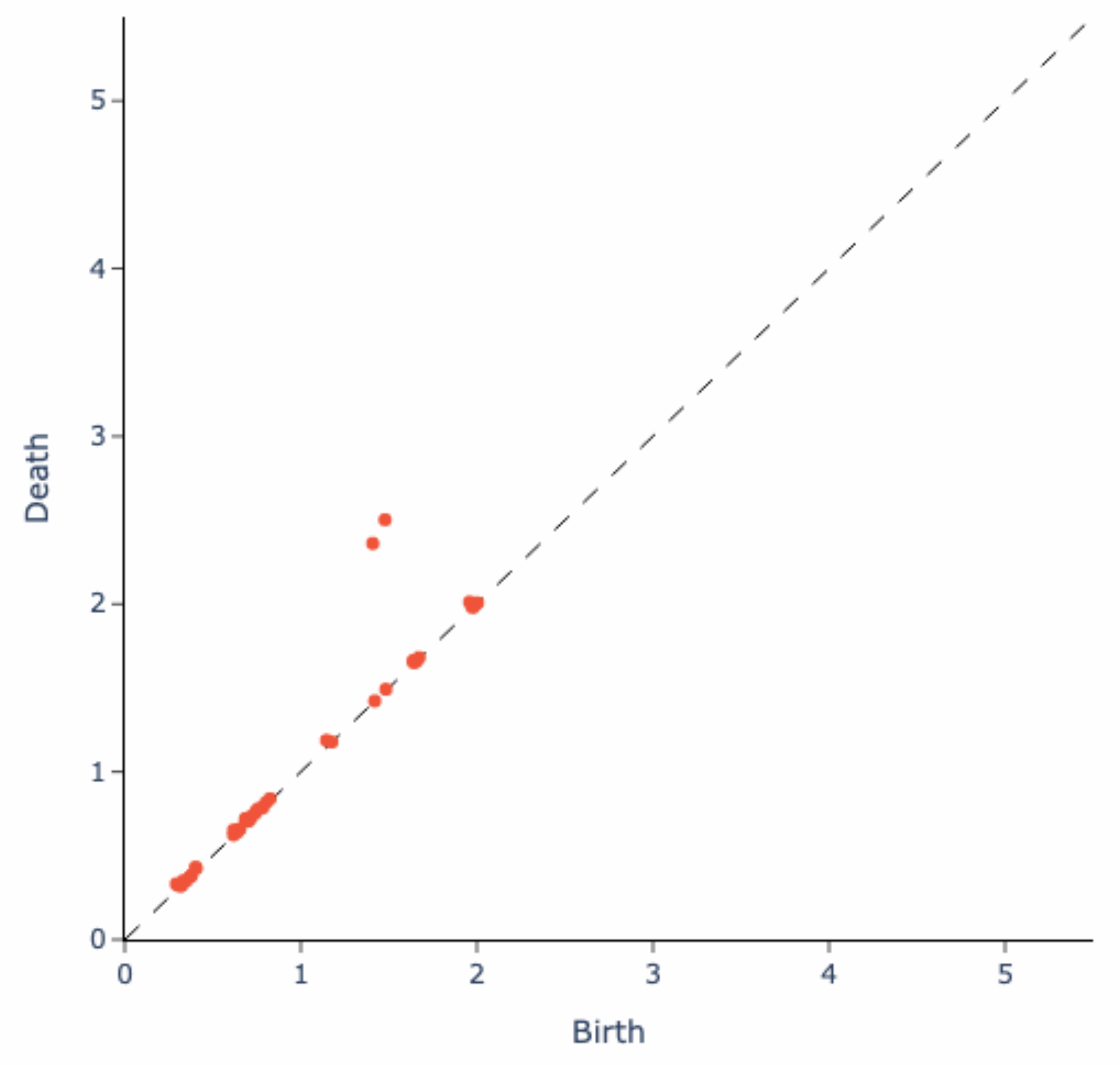
Goal:
rescale the neuron activations
on the time series
to increase the lifetimes of
these two dominant cycles



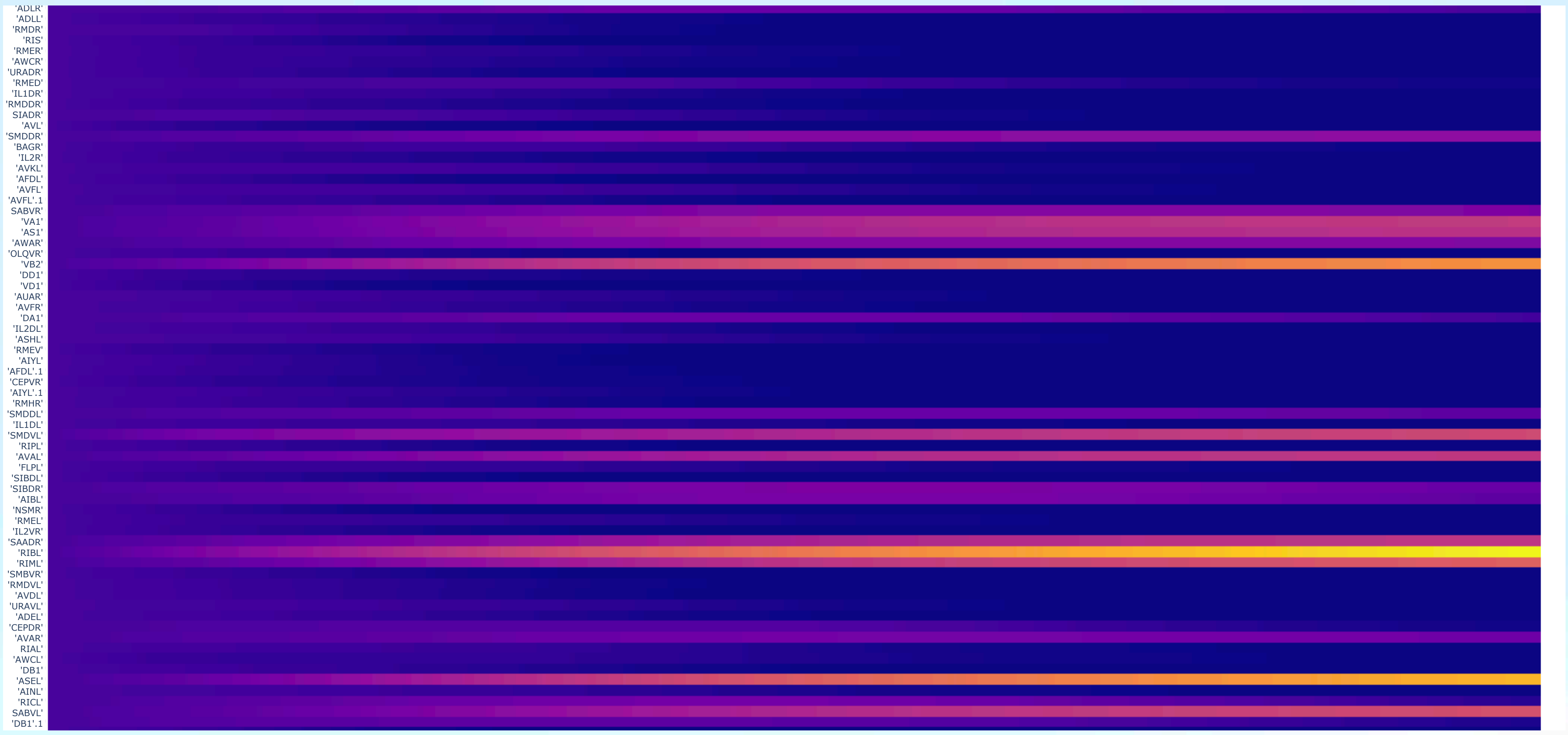
mean path through the simplex



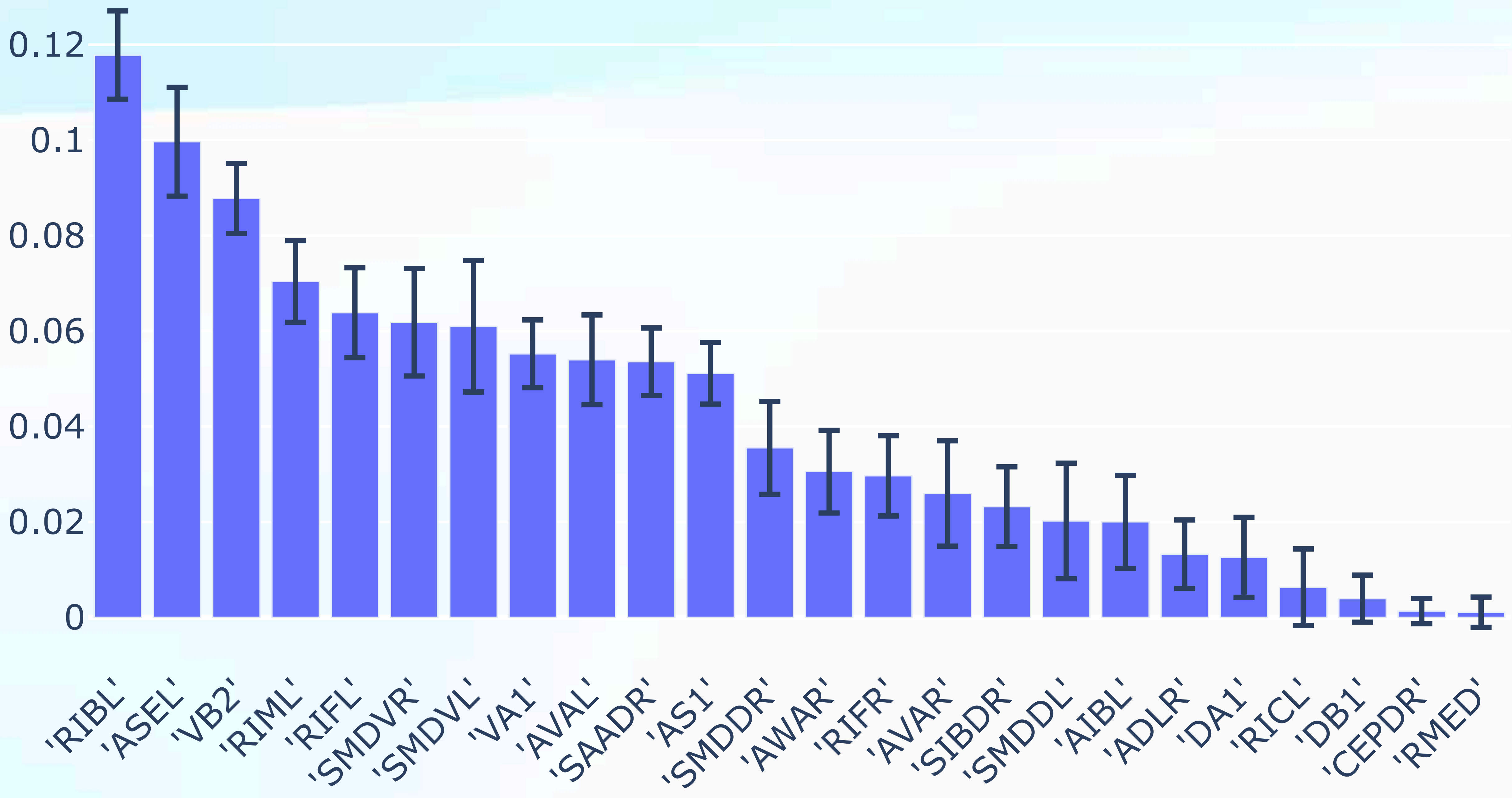
24 neurons with mean scores $> 10^{-3}$



mean path through the simplex



24 neurons with mean scores $> 10^{-3}$



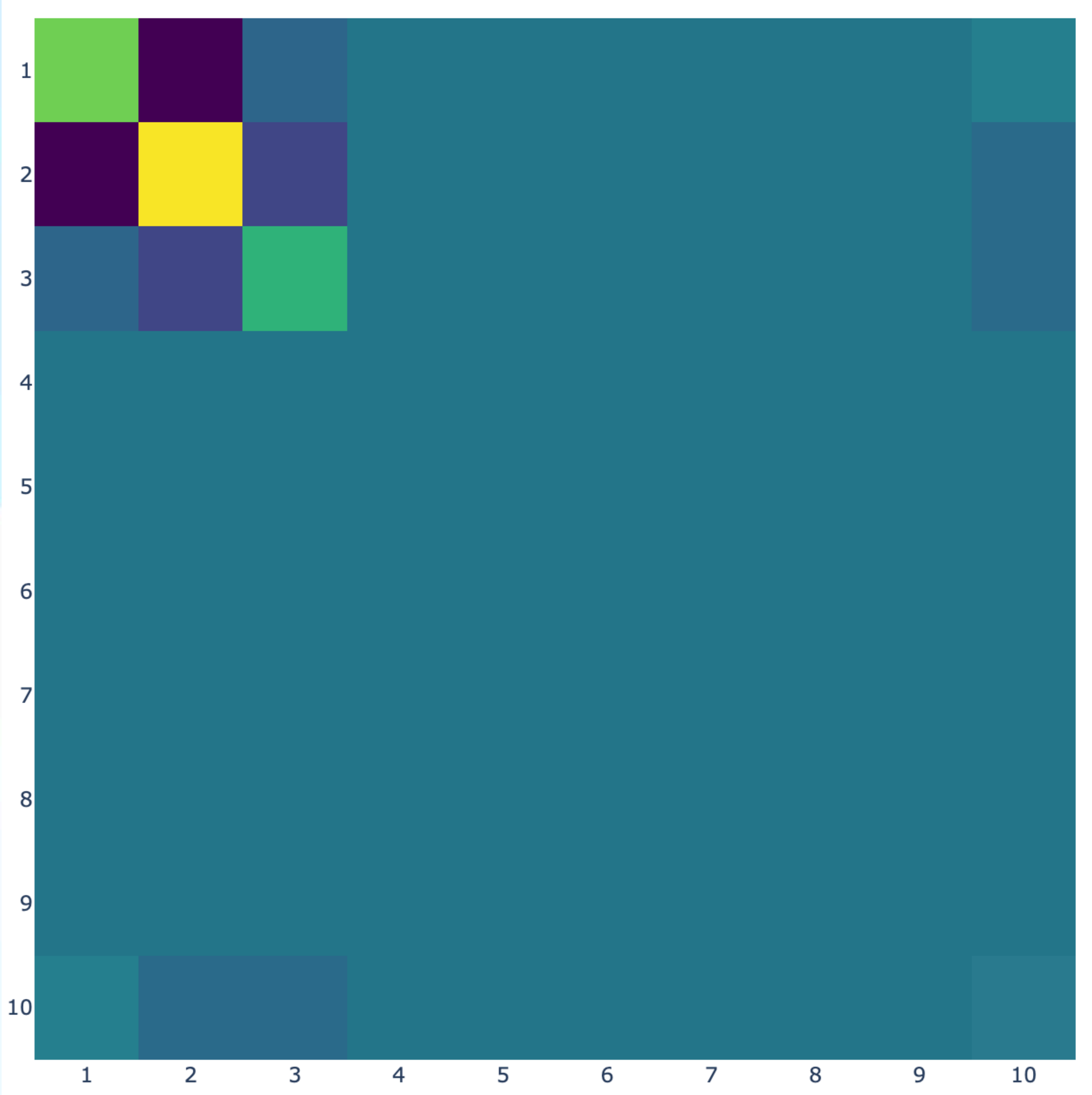
- RIBL = head interneuron known to integrate inputs from sensory modalities and contribute to locomotor regulation
- ASEL = a chemosensory neuron
- VB2 and VA1 = motor neurons responsible for backward and forward locomotion, respectively
- RIML, RIFL = motor coordination and neuromodulatory pathways
- SMDVR, SMDVL = motor neurons involved in head movement

For more details: arXiv:2310.17494v3

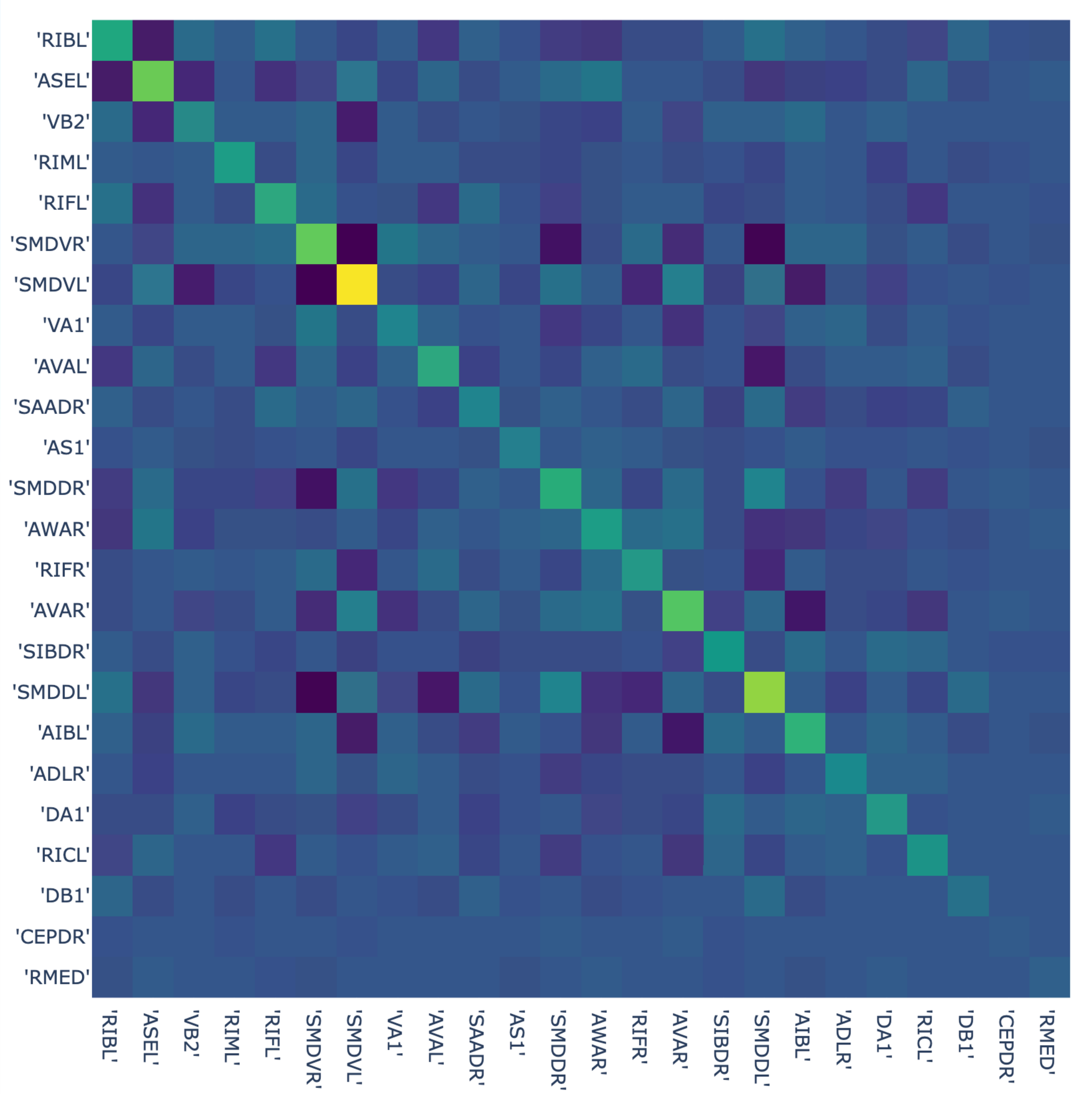
Extra slides

Covariance matrices

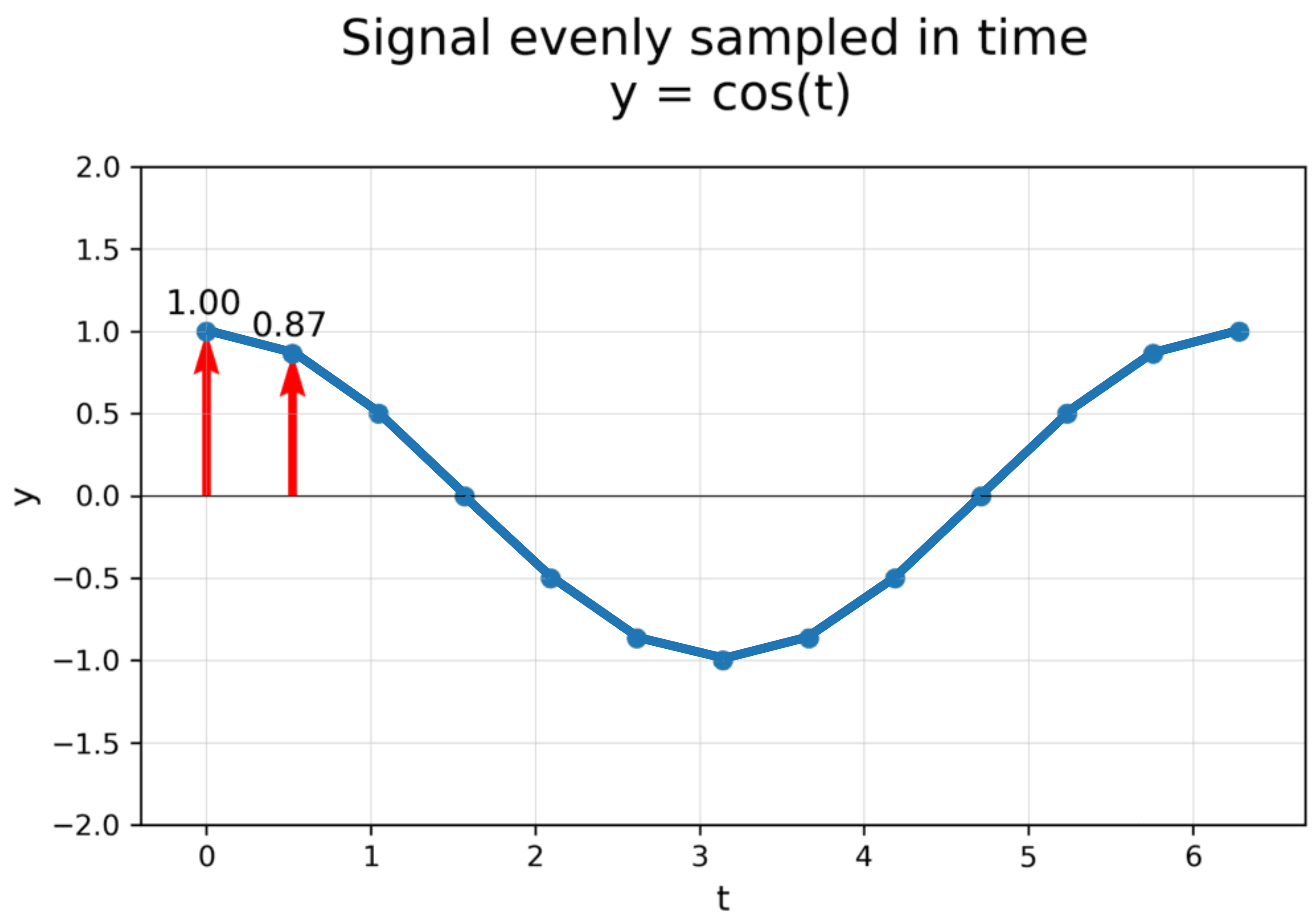
Synthetic example

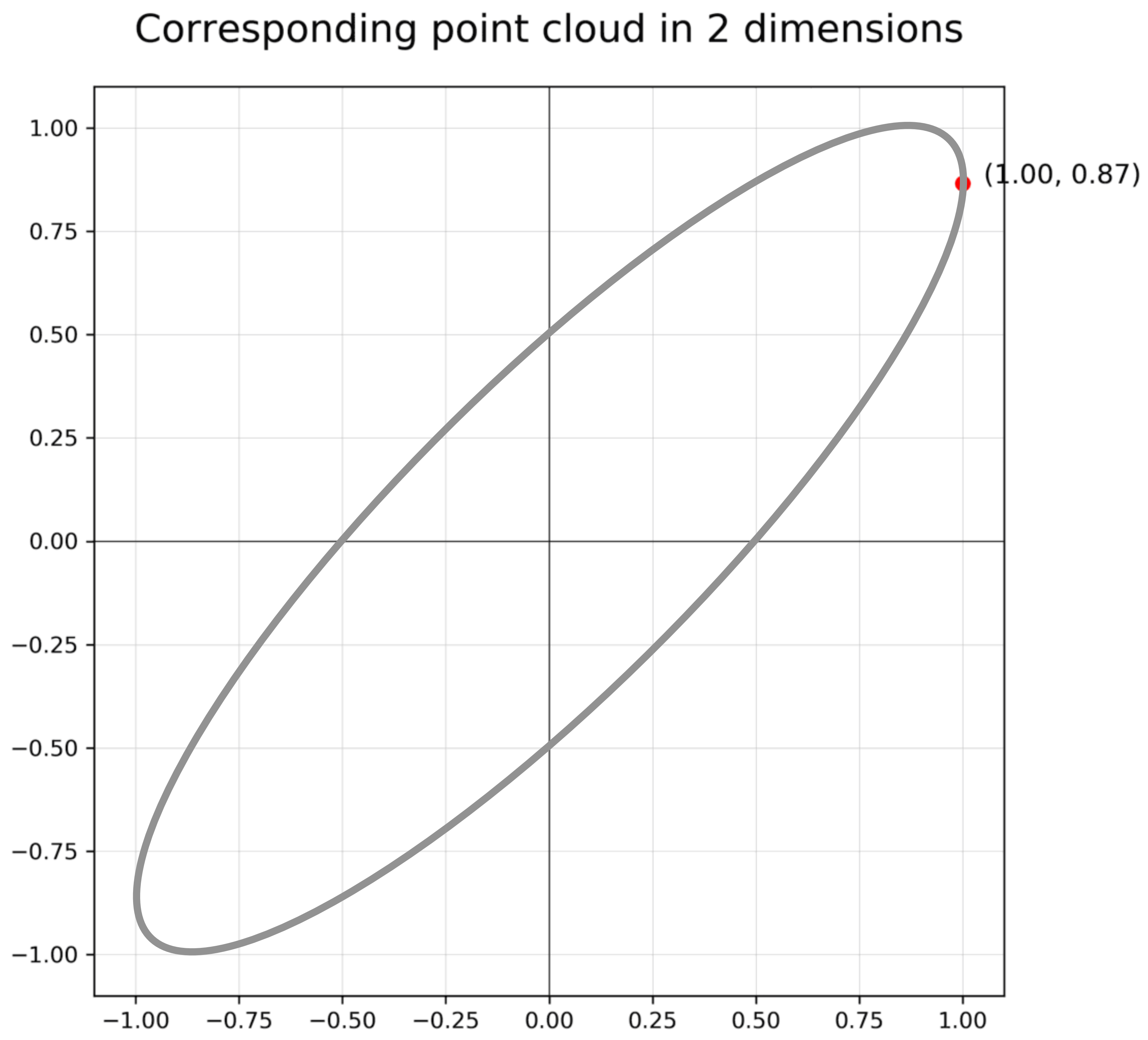


Biology example



Sliding Window Embedding (SWE)

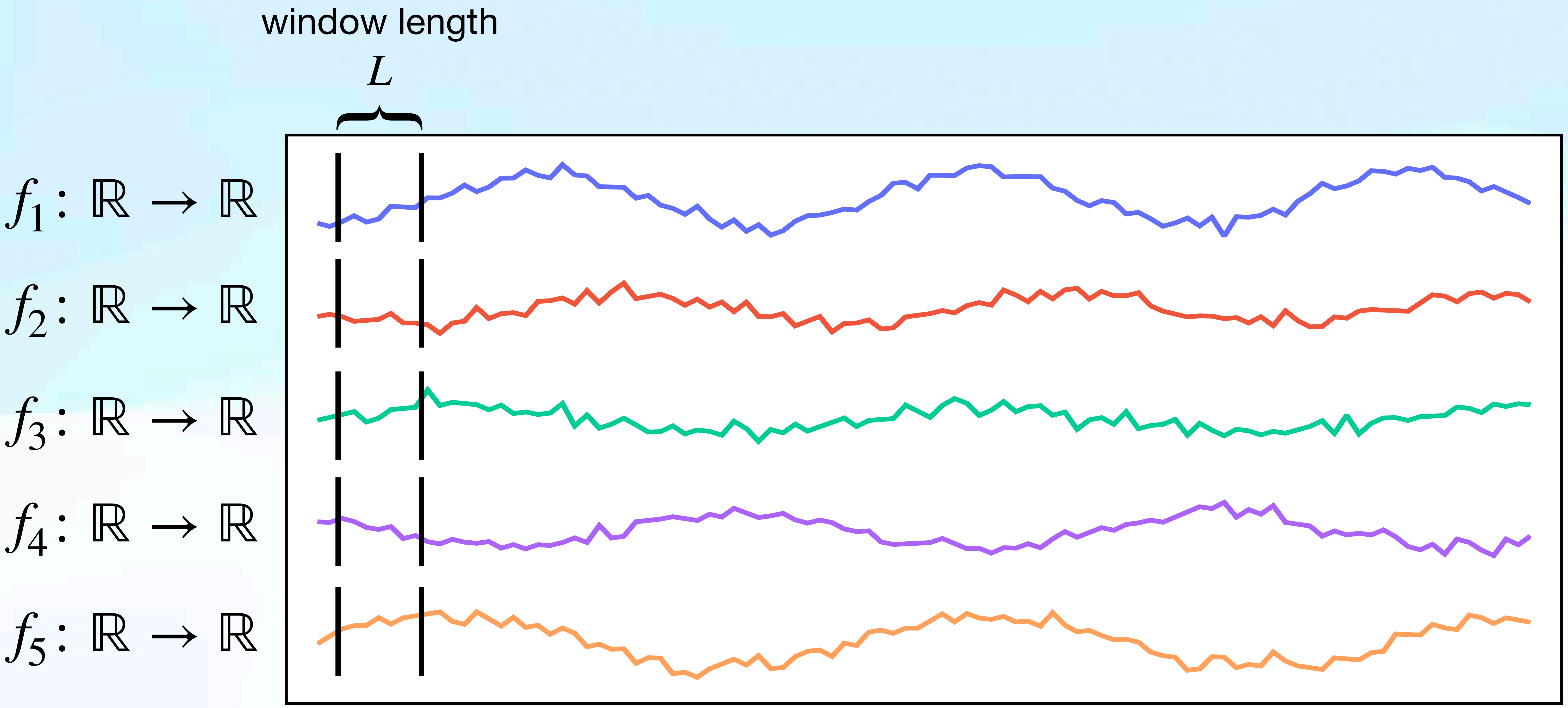




time series $\rightsquigarrow$ point cloud

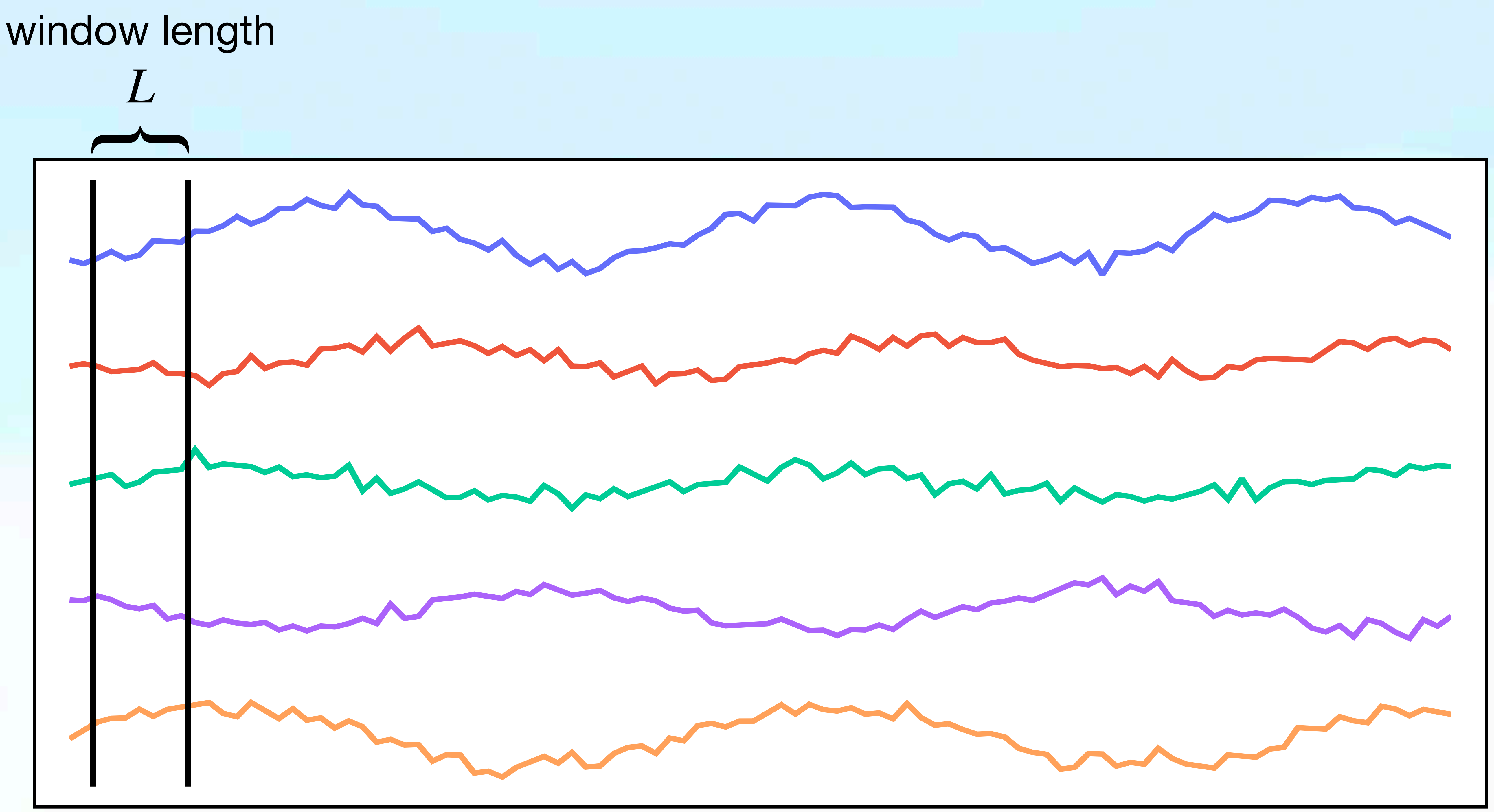
(quasi)periodicity $\rightsquigarrow$ topological loops

clarification:
“global” vs “component” embedding



$\Longrightarrow$ ~~five point clouds in $\mathbb{R}^L$~~

clarification:
“global” vs “component” embedding



$f\colon \mathbb{R} \to \mathbb{R}^5$

$\implies$ a single point cloud in $\mathbb{R}^{5 \cdot L}$