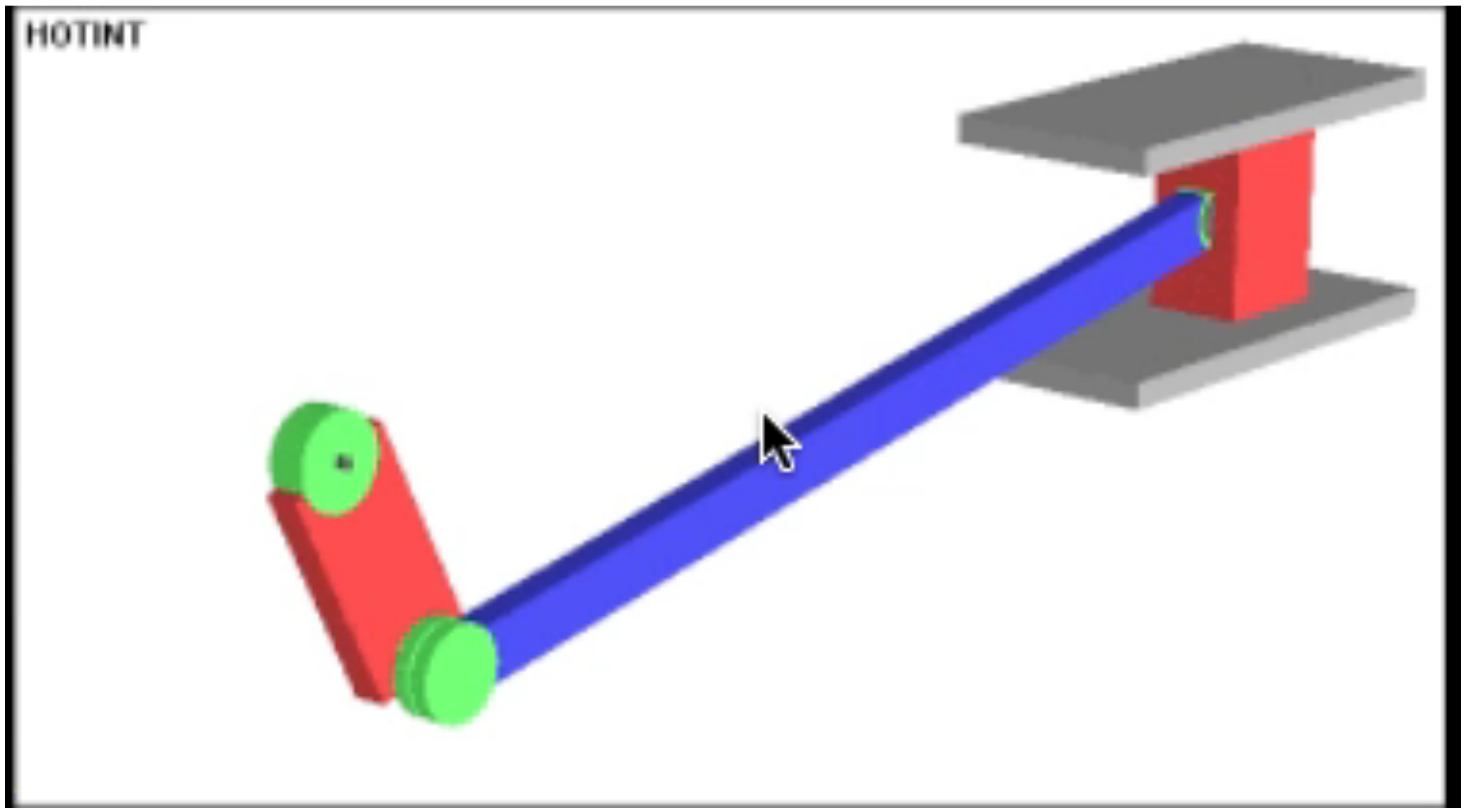
Intrinsic dimension

Yoav Freund
UCSD

Intrinsic and extrinsic dimensions



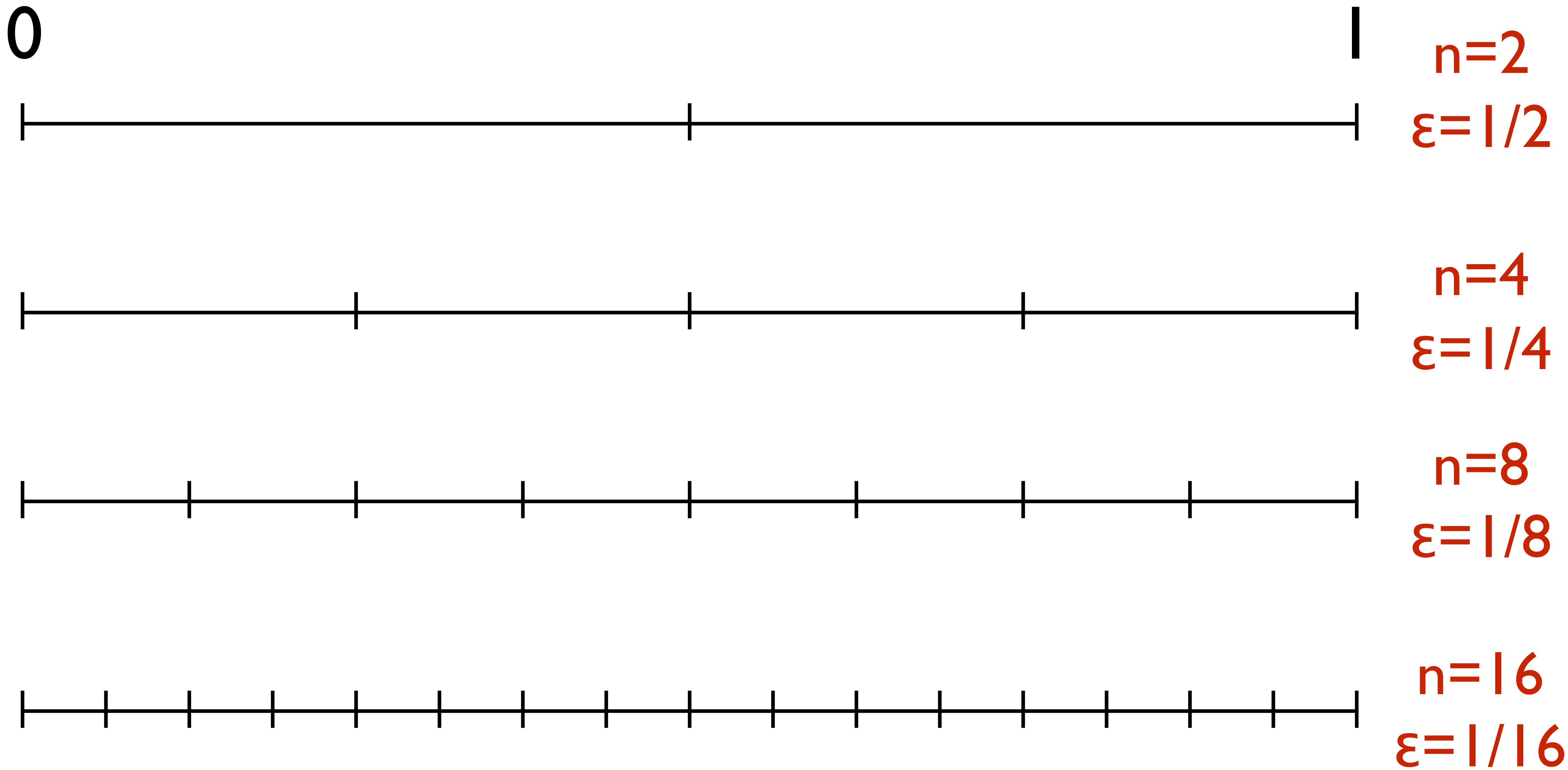
Dimension ~ number of degrees of freedom

- **Extrinsic:** Dimension as a video frame: 600x400
- **Intrinsic:** Dimension as a mechanical system: 1

Intrinsic dimension of a data cloud

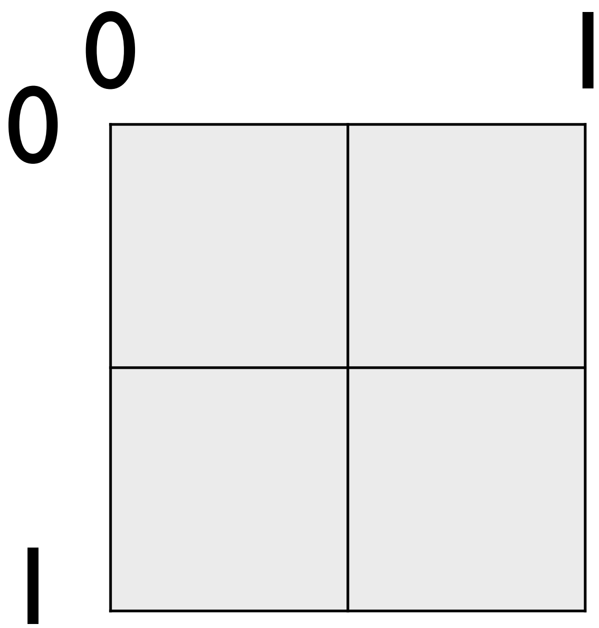
- Suppose we have a uniform distribution over some domain.
- We partition it into n cells.
- The "Diameter" ε of the partition is the maximal distance between two points belonging to the same cell.
- As n increases, ε decreases, but at what rate?
- Lets look at some simple examples.

A line segment



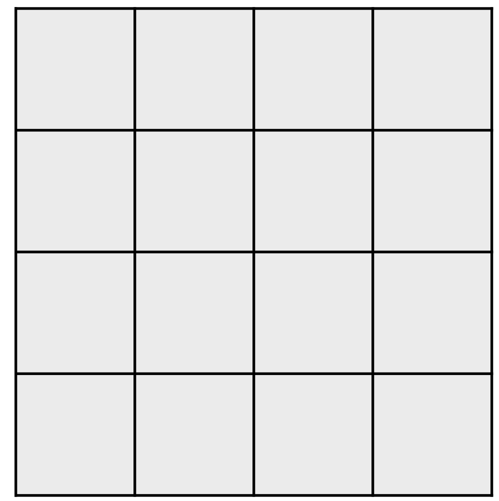
General rule: $\varepsilon = 1/n$

A 2-D set



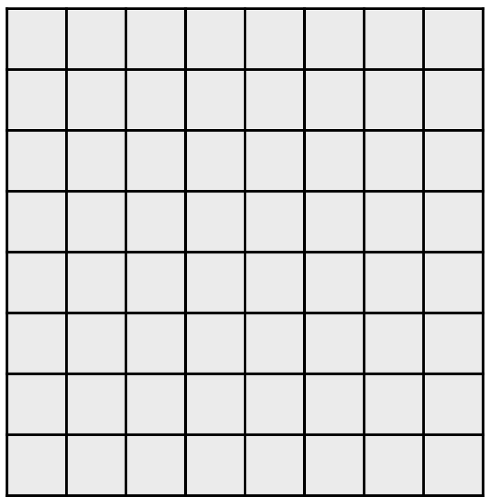
$n = 4$

$$\epsilon = \frac{\sqrt{2}}{2} = \frac{1}{\sqrt{2}}$$



$n = 16$

$$\epsilon = \frac{\sqrt{2}}{4} = \frac{1}{2\sqrt{2}}$$



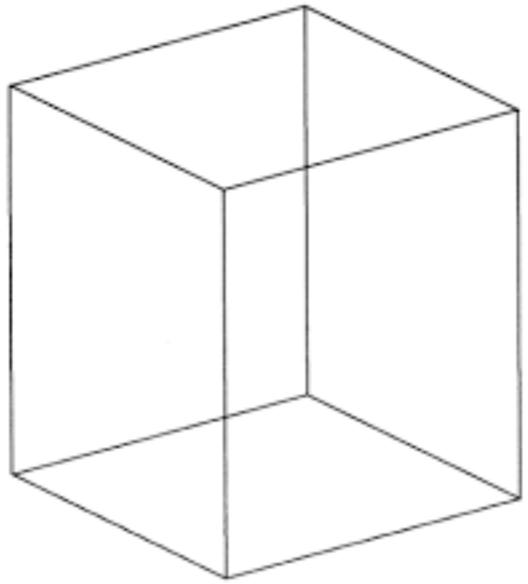
$n = 64$

$$\epsilon = \frac{\sqrt{2}}{8} = \frac{1}{4\sqrt{2}}$$

general formula $\epsilon = \sqrt{\frac{2}{n}}$

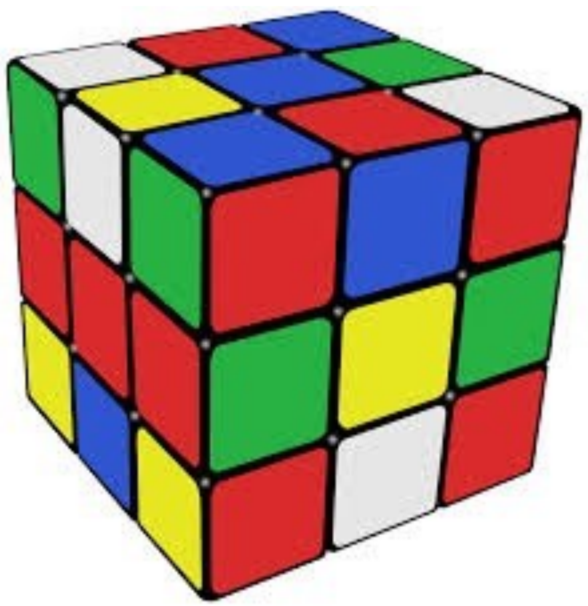
or $n = \frac{2}{\epsilon^2}$

A 3d set



$n = 1$

$\epsilon = \sqrt{3}$



$n = 27$

$$\epsilon = \frac{\sqrt{3}}{3}$$

$n = 125$

$$\epsilon = \frac{\sqrt{3}}{5}$$

general formula $\epsilon = \dfrac{\sqrt{3}}{\sqrt[3]{n}}$

or $n = \dfrac{3\sqrt{3}}{\epsilon^3}$

General dependence of number of elements on diameter

$\epsilon =$ max diameter

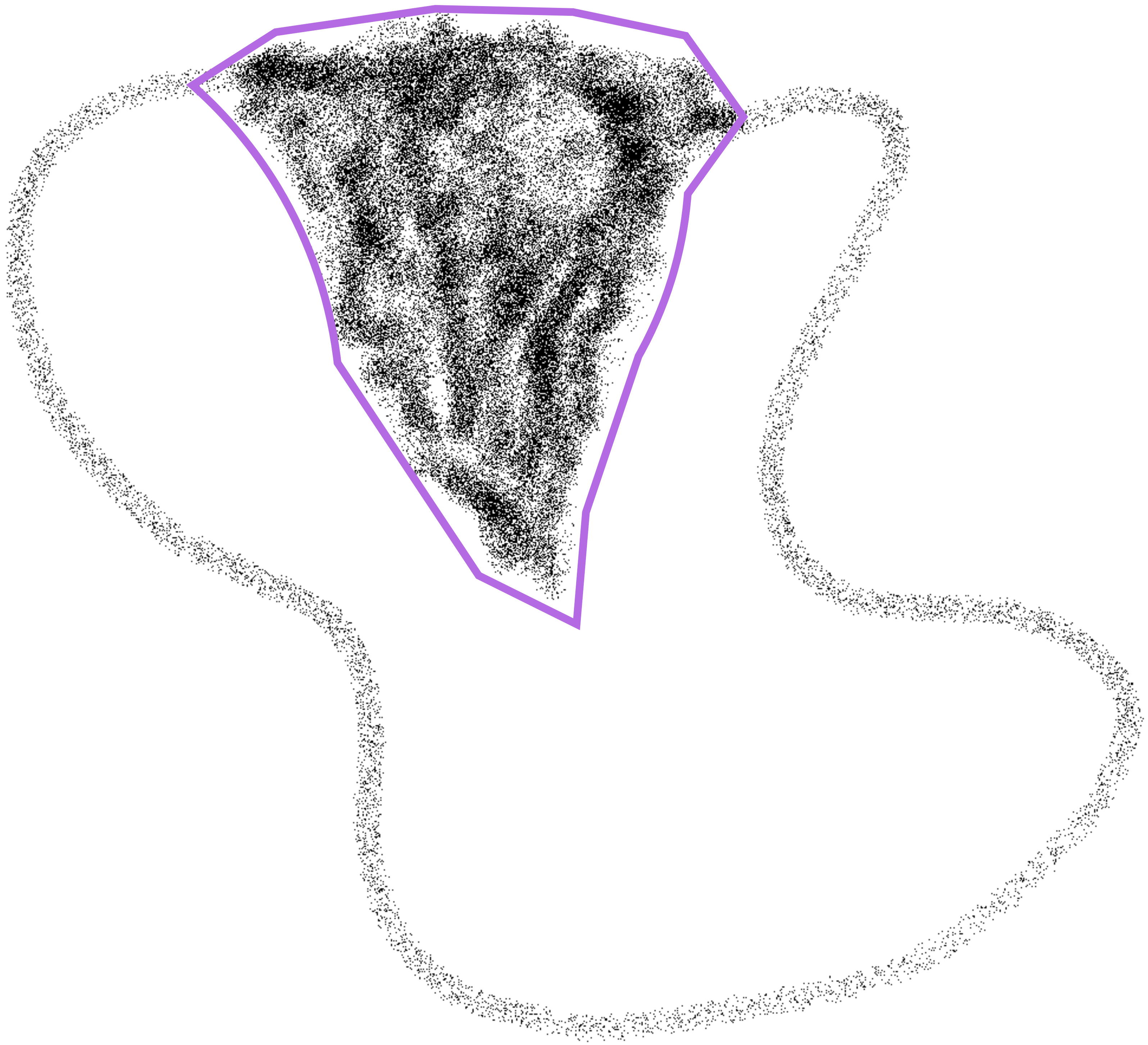
$n =$ number of cells

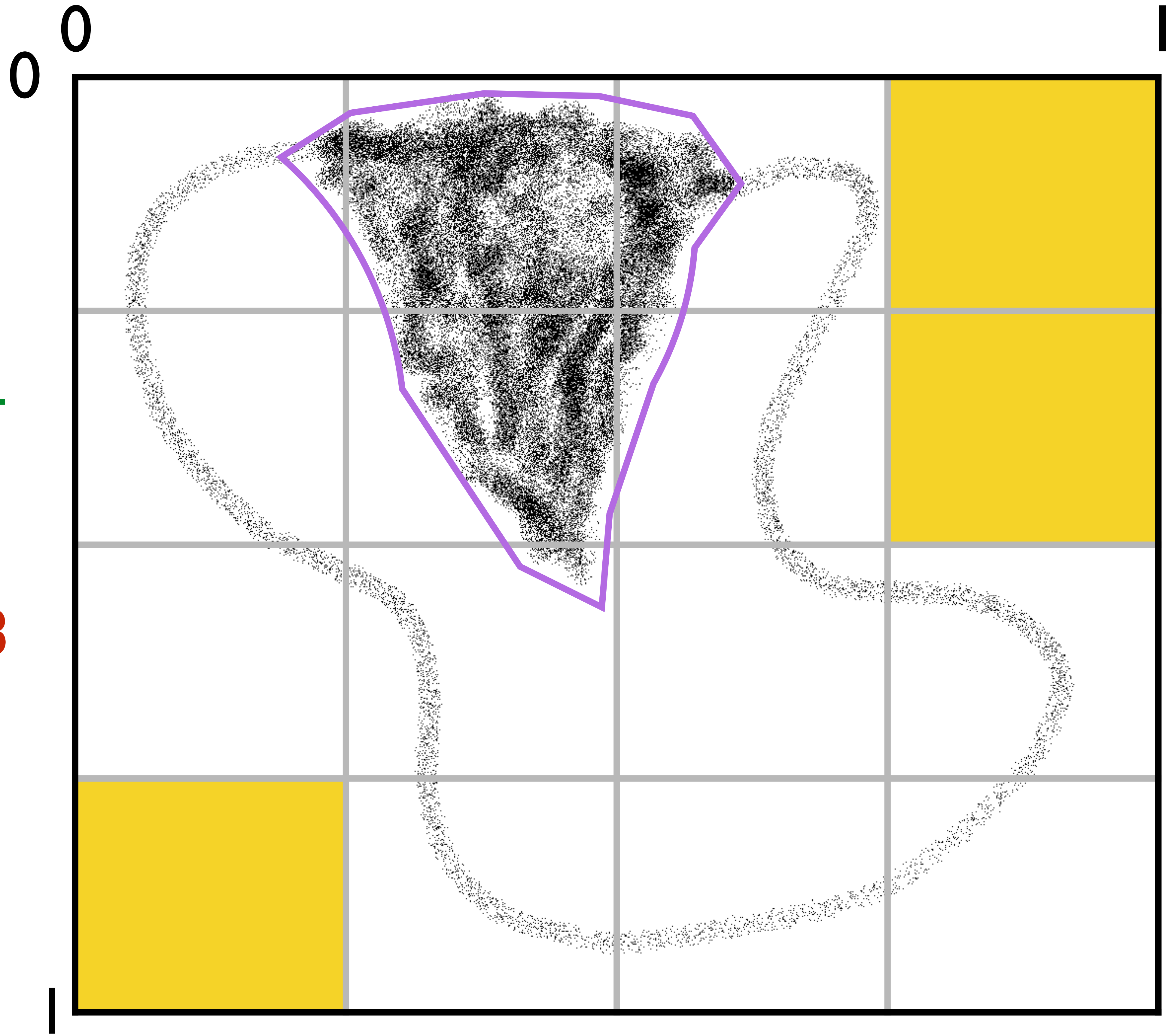
$d =$ dimension of space

General Formula: $n = \dfrac{C}{\epsilon^{d}}$

Alternatively: $\log n = \log C + d \log \dfrac{1}{\epsilon}$

We can use the last equation to **define** the dimension of a dataset





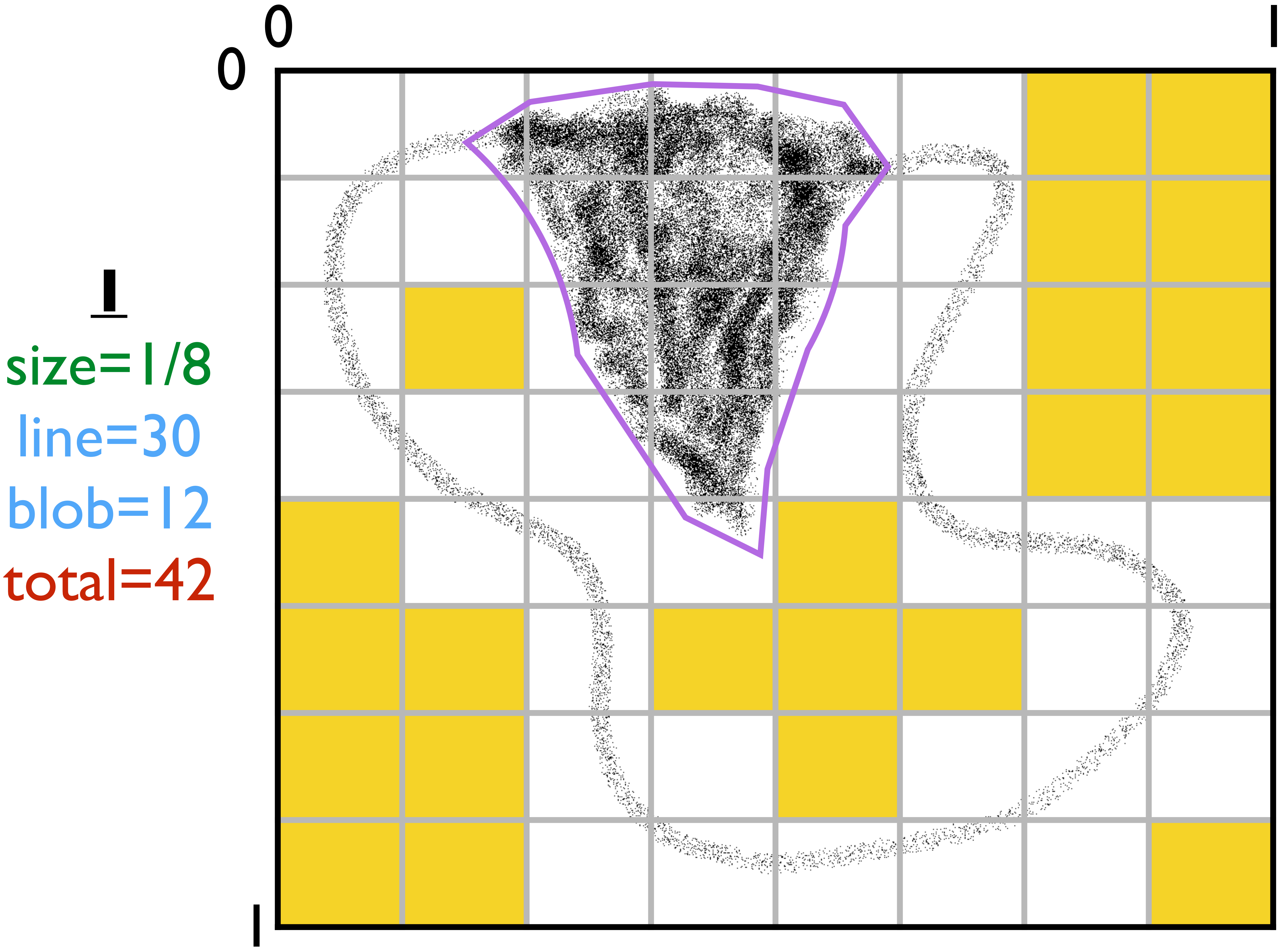
0

size=1/4

line=7

blob=6

total=13


0
0
1
1
1
size=1/8
line=30
blob=12
total=42

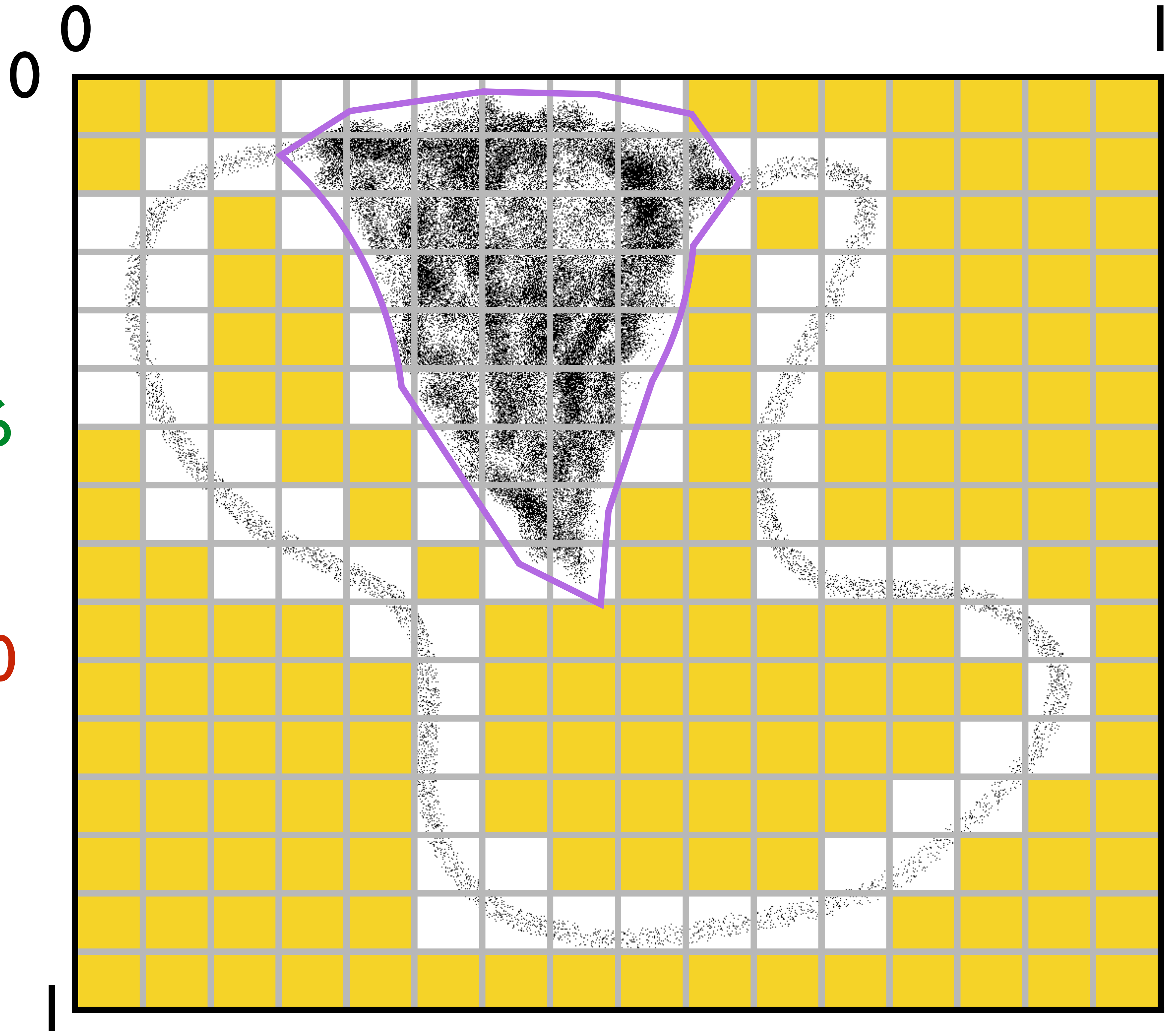
2

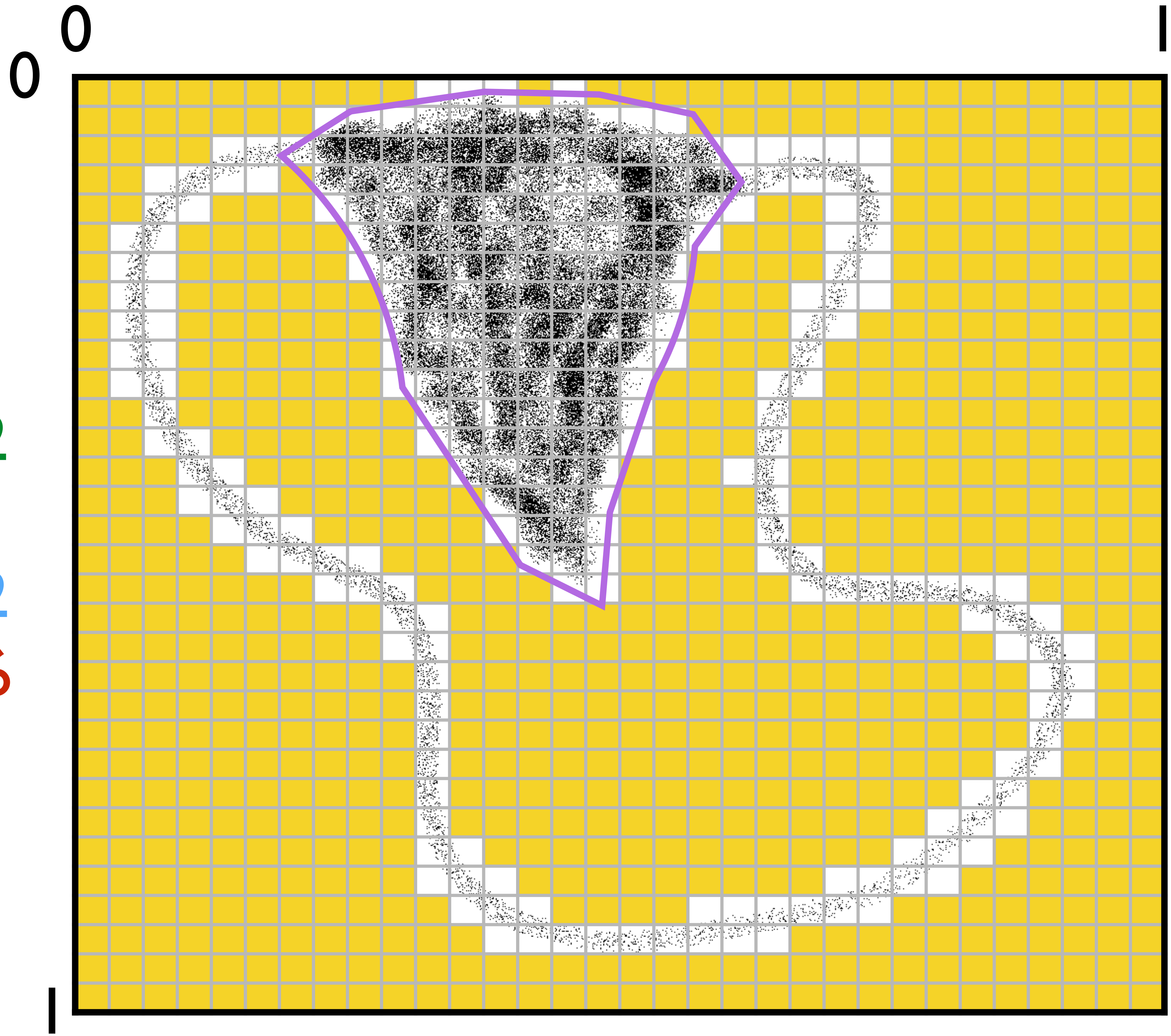
size=1/16

line=56

blob=44

total=100





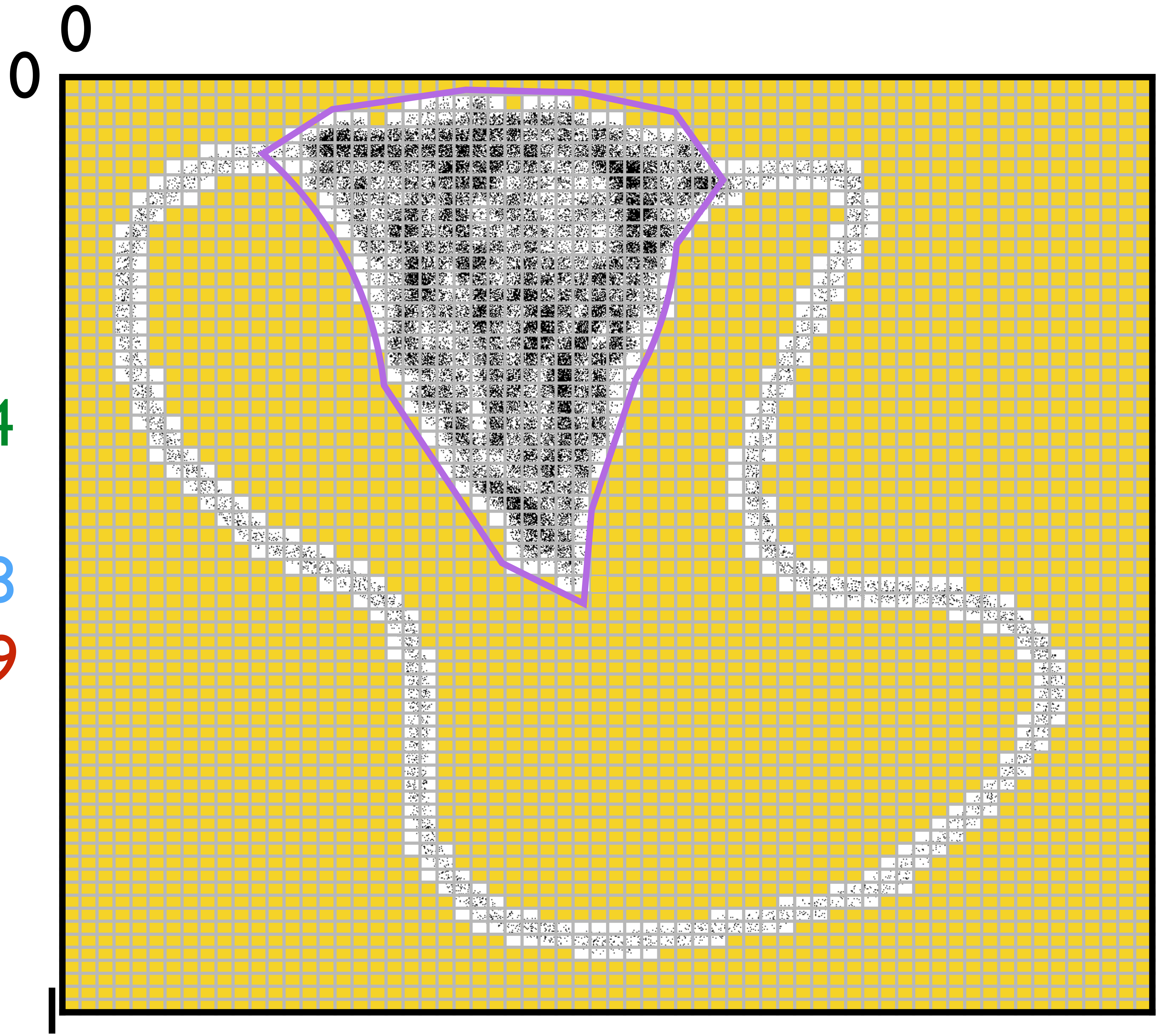
3

size=1/32

line=134

blob=142

total=276



4

size=1/64

line=281

blob=598

total=879

Estimating intrinsic dimension

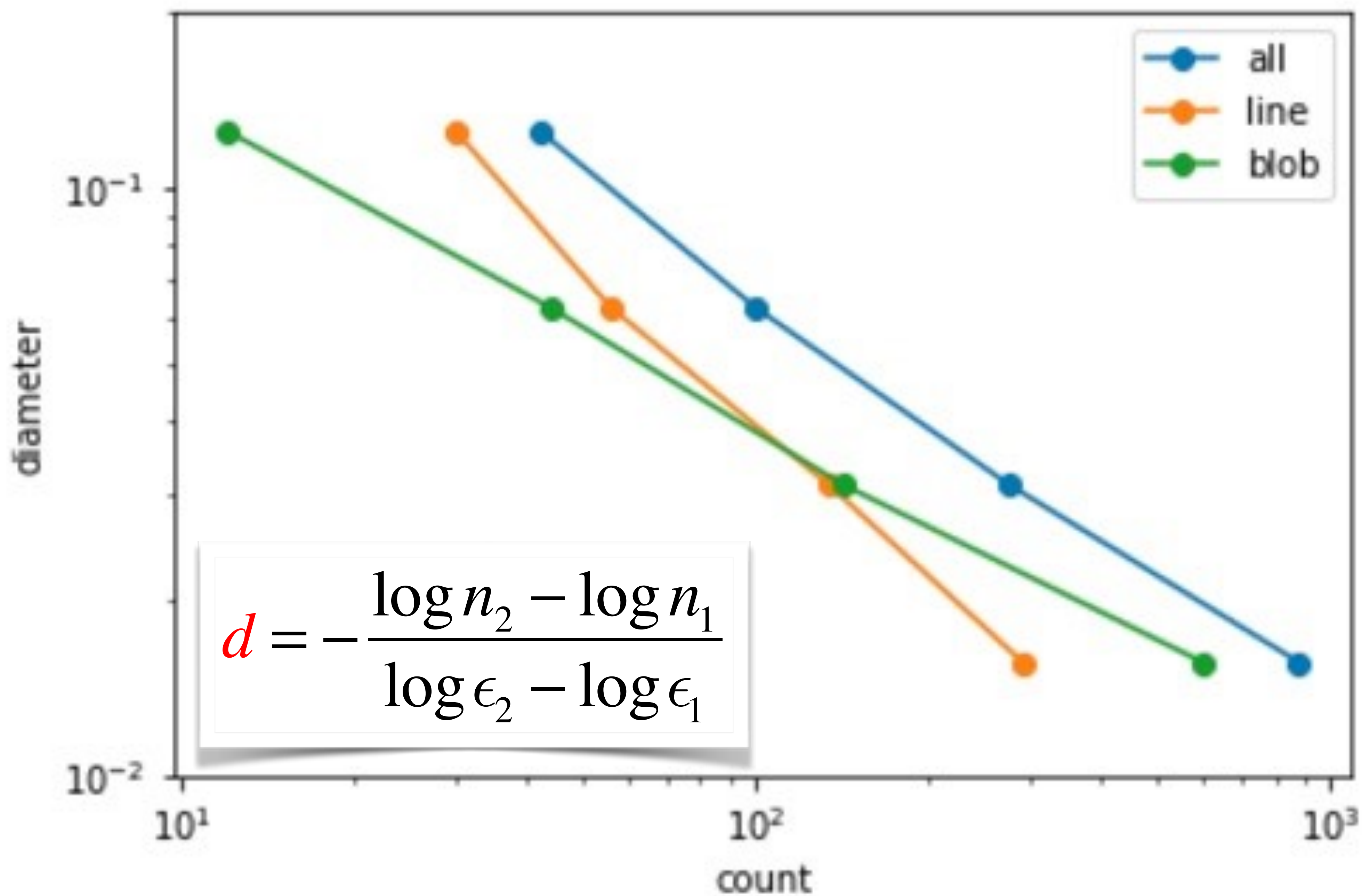
$$\log n = \log C + d \log \frac{1}{\epsilon}$$

Two Scales: $(n_1, \epsilon_1), (n_2, \epsilon_2)$; $\quad n_1 < n_2, \quad \epsilon_1 > \epsilon_2$

$$\log \frac{n_2}{n_1} = d \log \frac{\epsilon_1}{\epsilon_2}$$

$$d = \frac{\log n_2 - \log n_1}{\log \epsilon_1 - \log \epsilon_2}$$

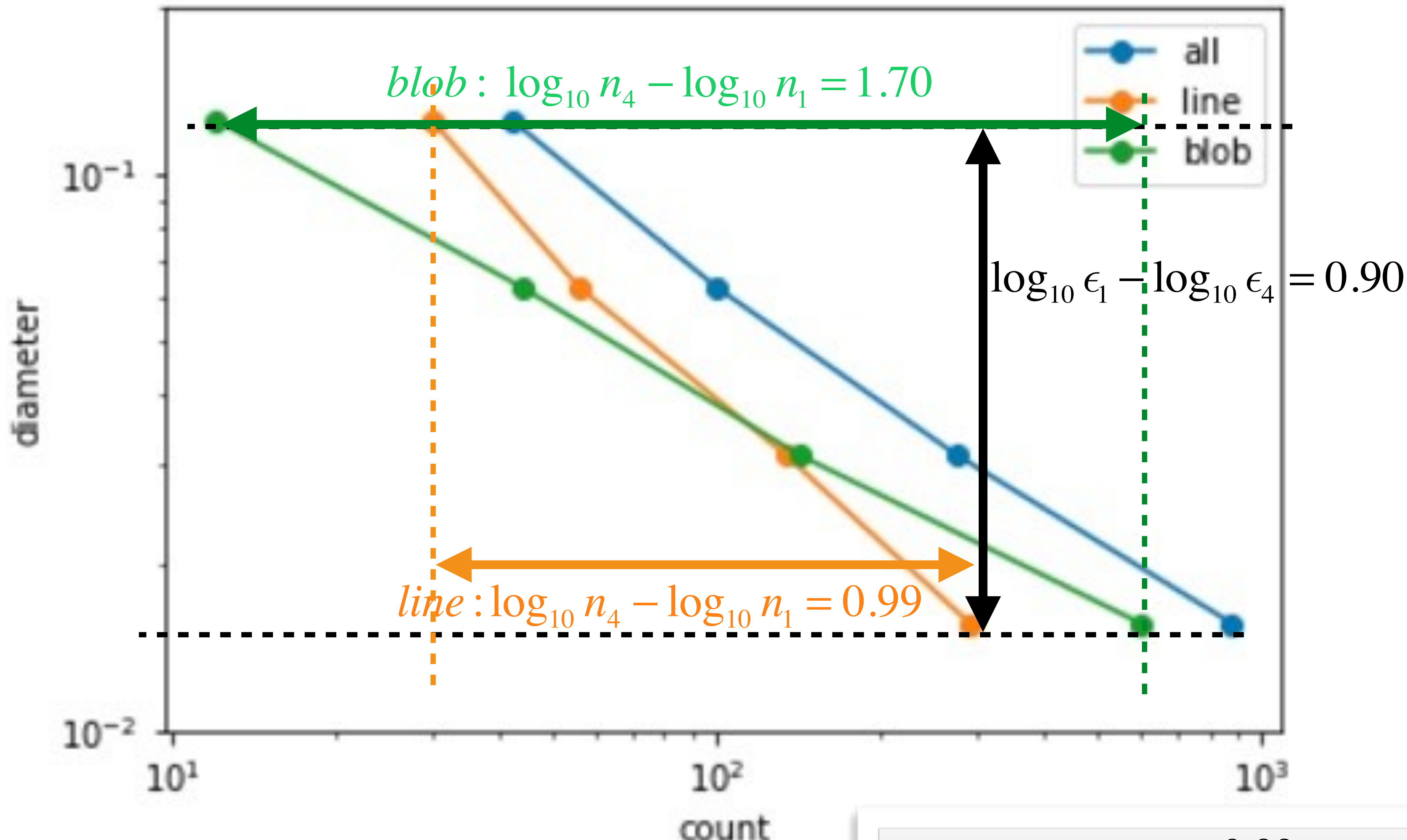
Estimating the dimension



$$d = -\frac{\log n_2 - \log n_1}{\log \epsilon_2 - \log \epsilon_1}$$

Steeper decline = lower dimension

Estimating the dimension



$$d = -\frac{\log n_2 - \log n_1}{\log \epsilon_2 - \log \epsilon_1}$$

$$\text{dimension of line} = \frac{0.99}{0.90} = 1.10 \approx 1$$

$$\text{dimension of blob} = \frac{1.70}{0.90} = 1.89 \approx 2$$

Estimating using kmeans++

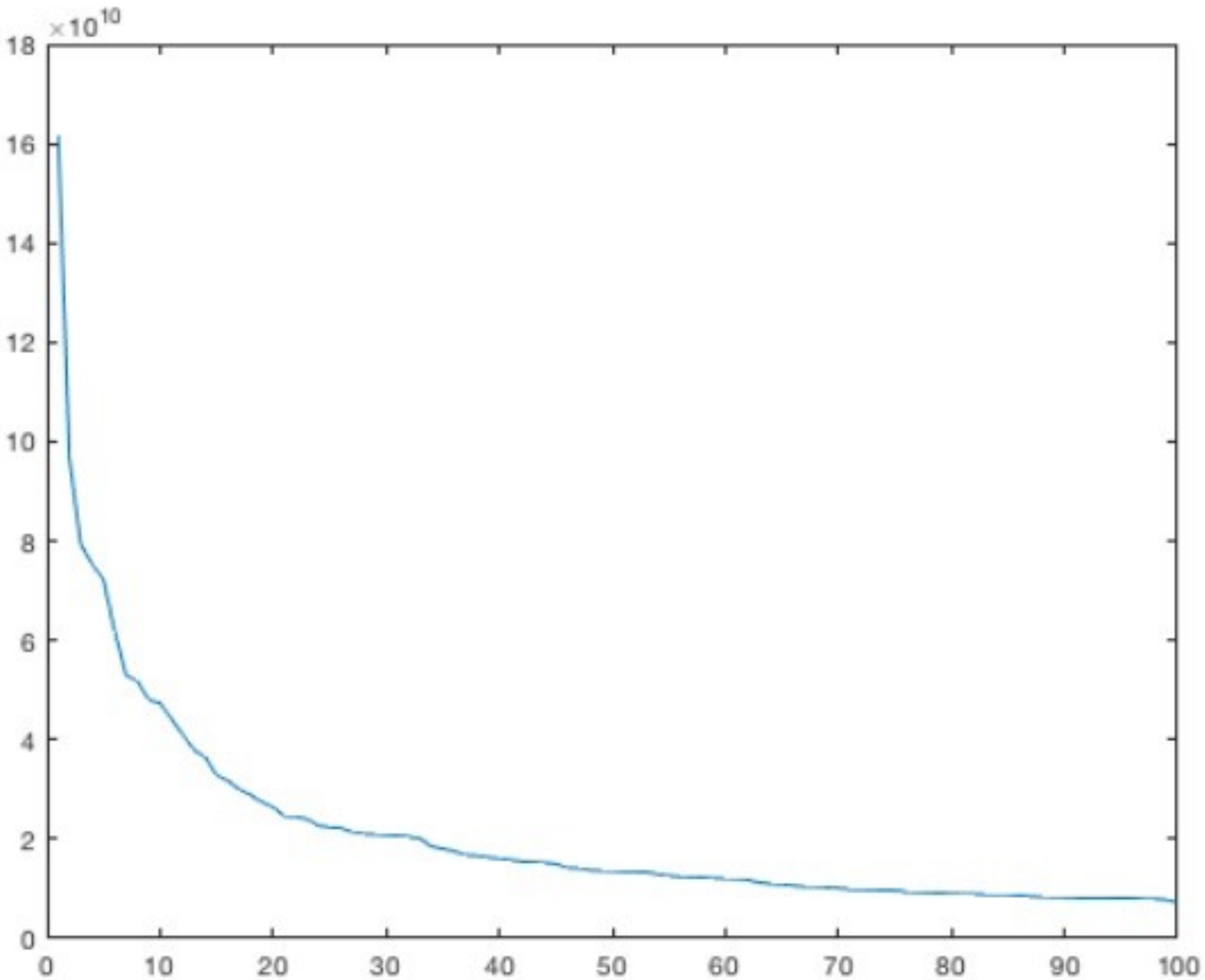
- Add representatives using the K-means++ rule.
- After adding a representative, estimate the average square distance.

$$d = \frac{\log n_2 - \log n_1}{\log\sqrt{\epsilon_1} - \log\sqrt{\epsilon_2}} = 2\frac{\log n_2 - \log n_1}{\log\epsilon_1 - \log\epsilon_2}$$

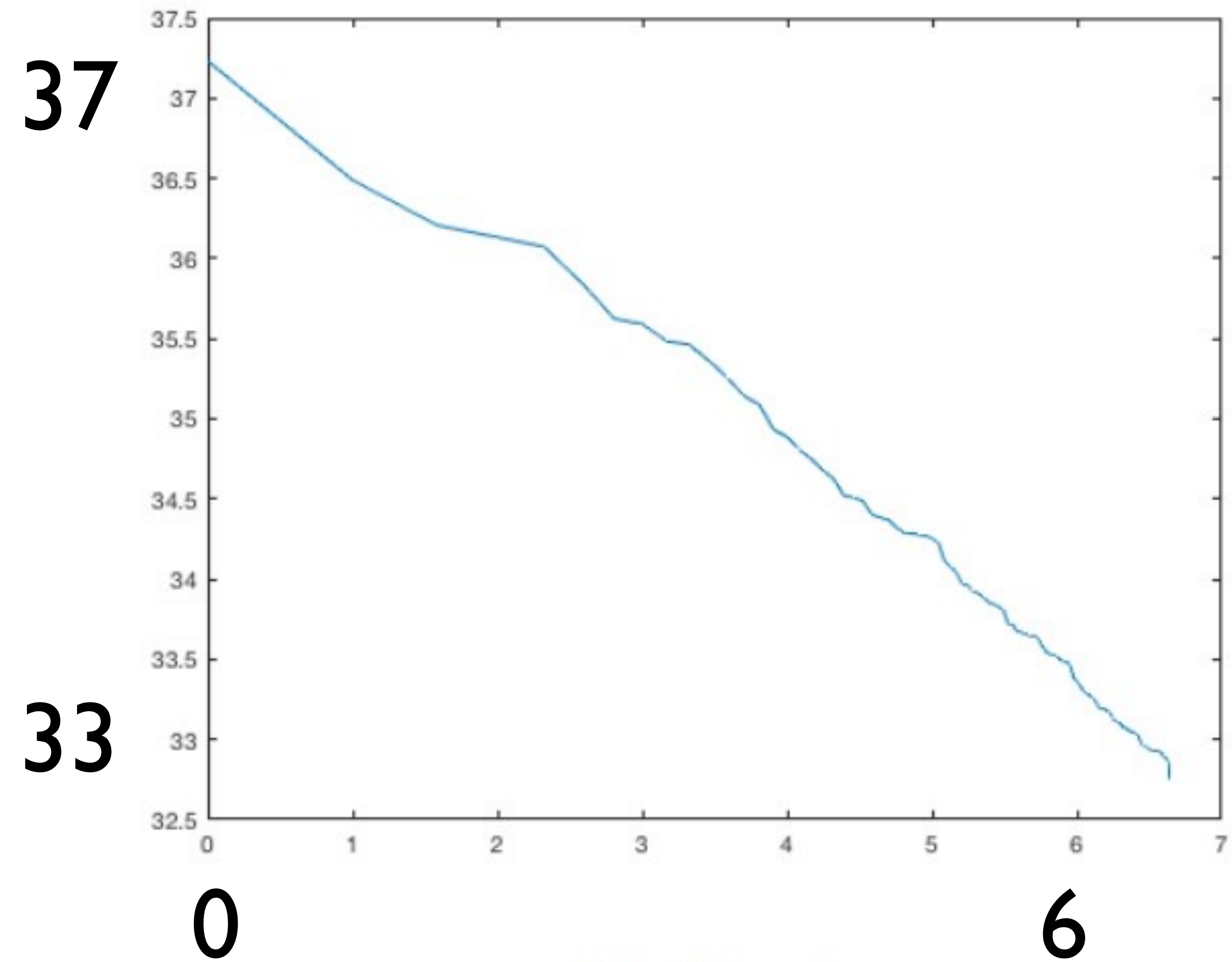
rotating hand

http://vasc.ri.cmu.edu/idb/html/motion/hand/index.html

Rotating hand dimension estimation



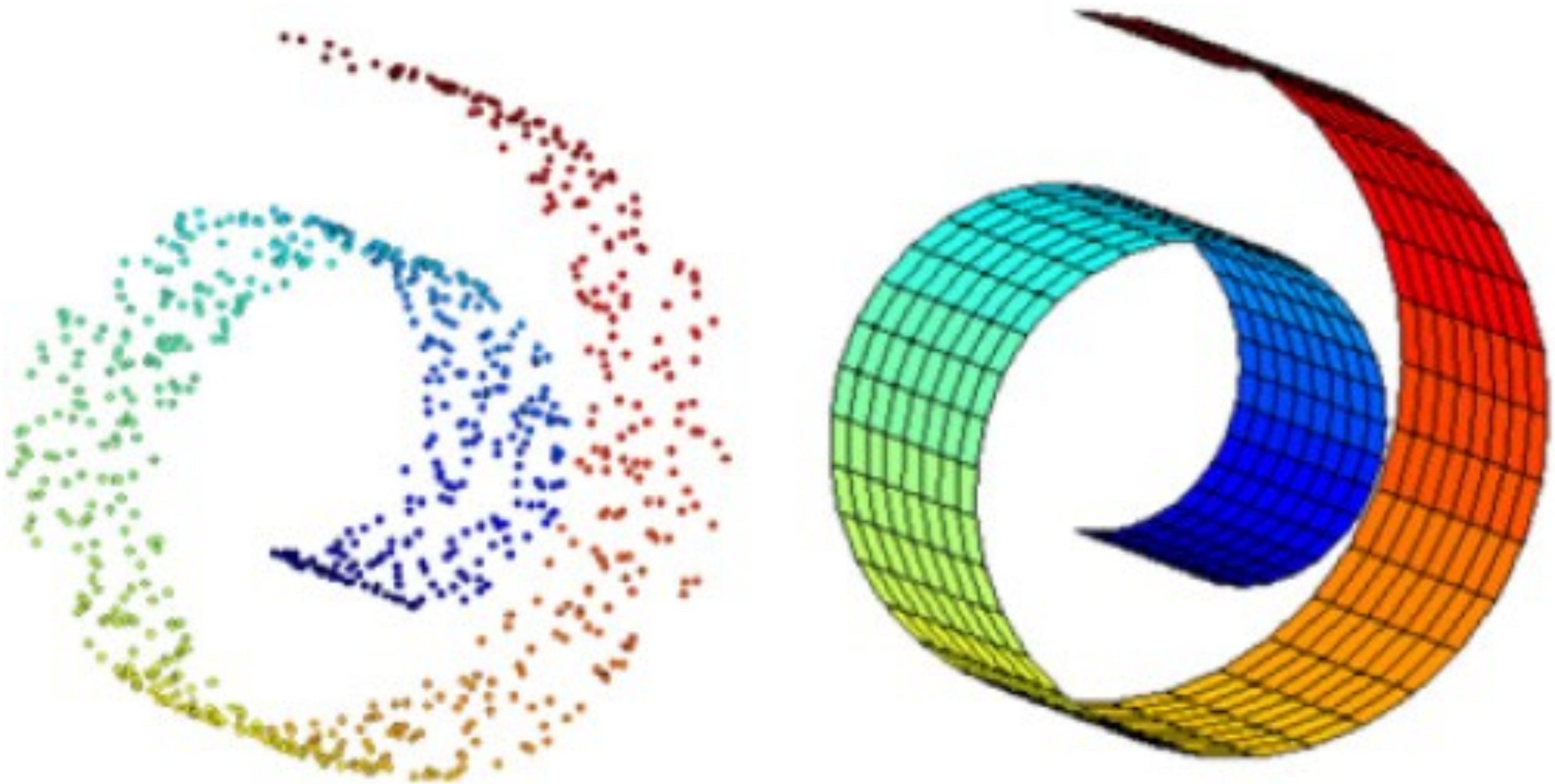
(c) Hand



(d) Hand

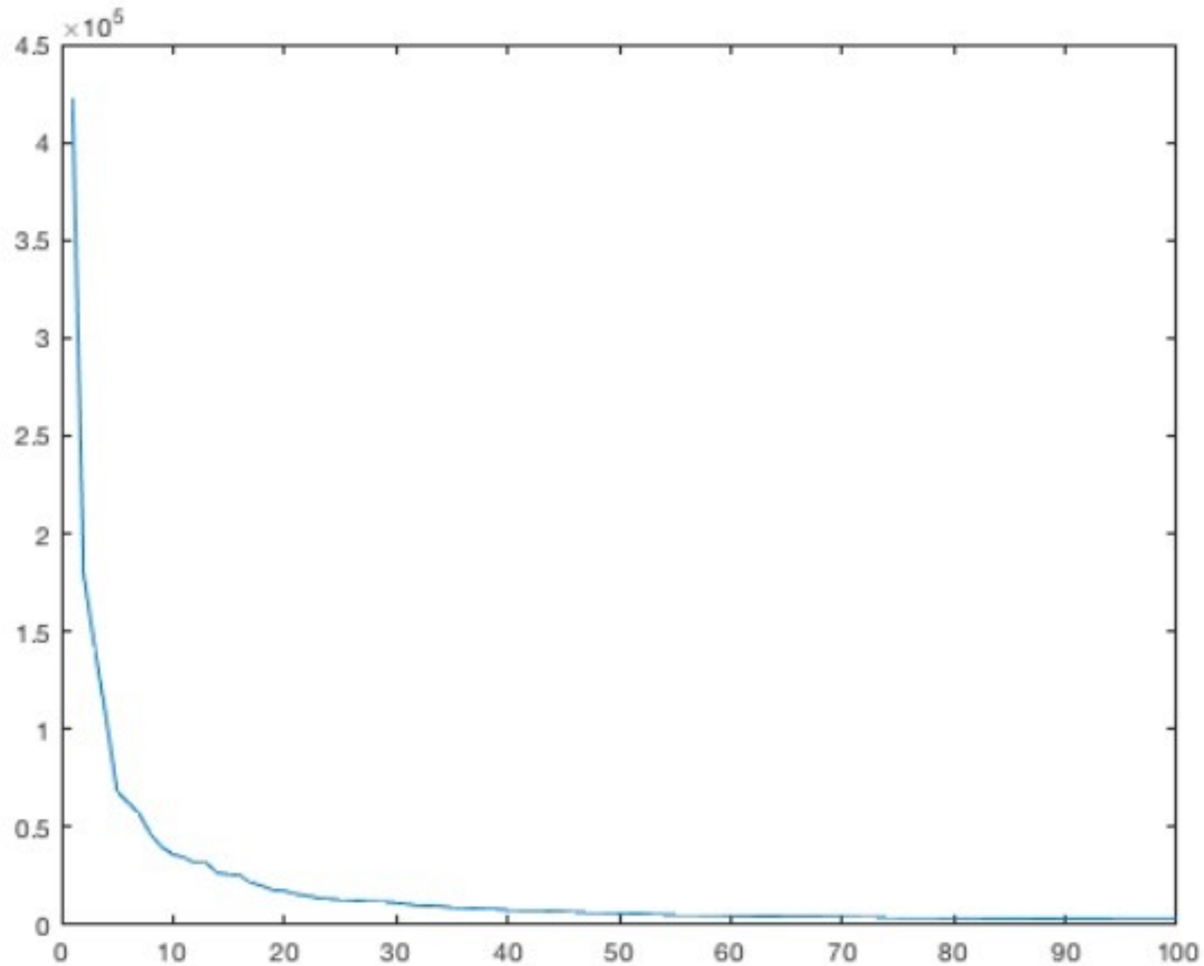
2*6/4 = 3

Swiss Roll



(a) (b)

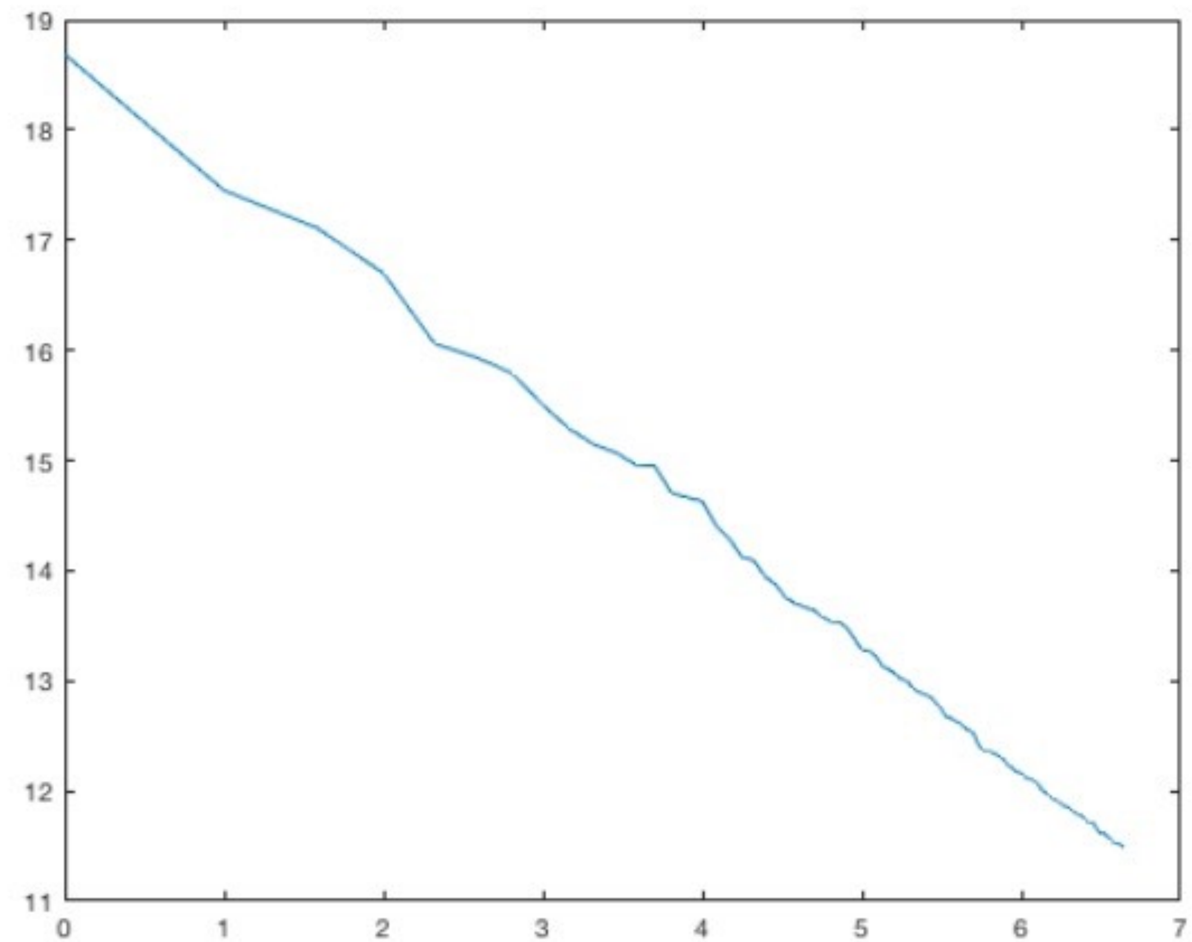
Swiss Roll dimension estimation



(c) Swissroll

18.5

11.5



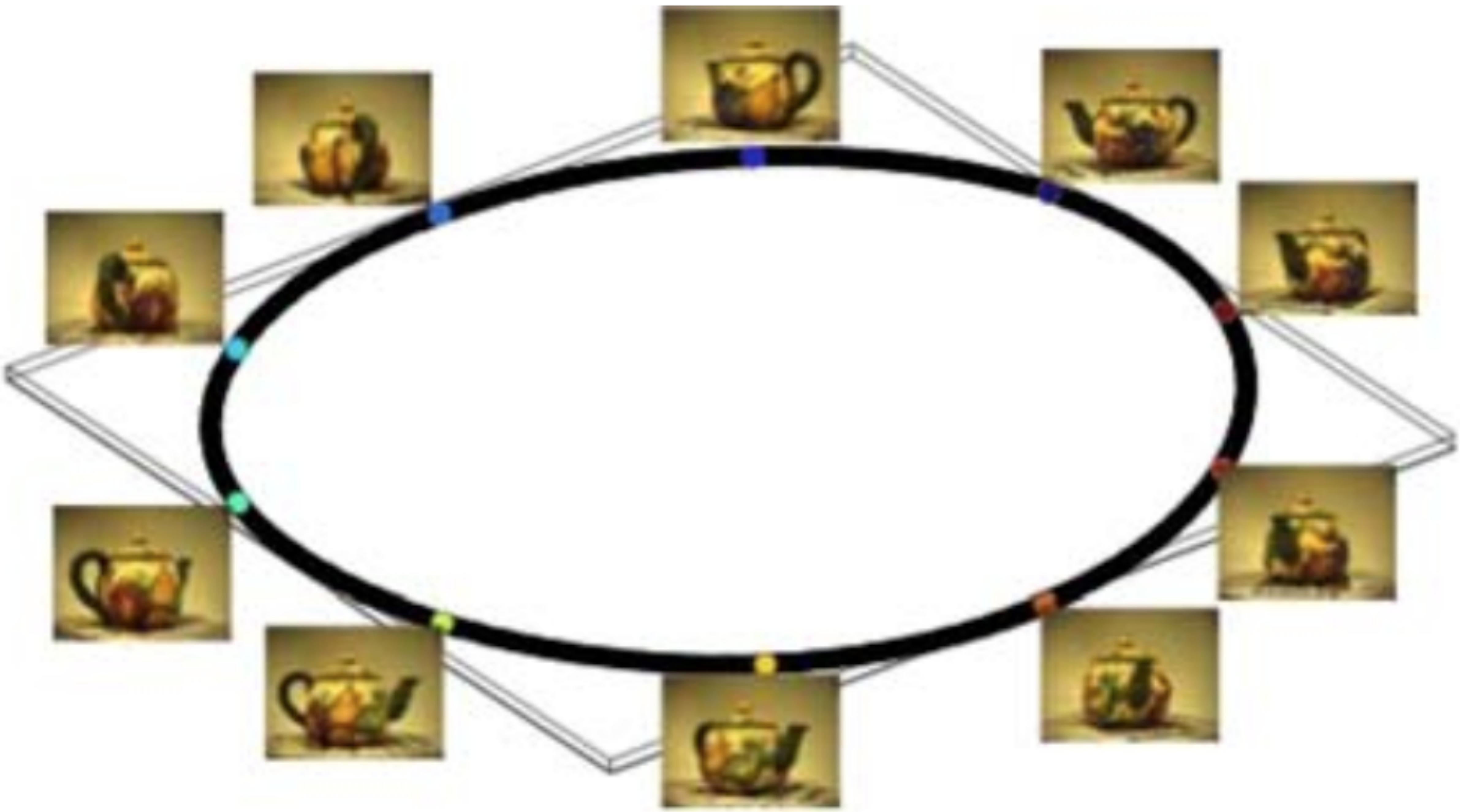
0

6.5

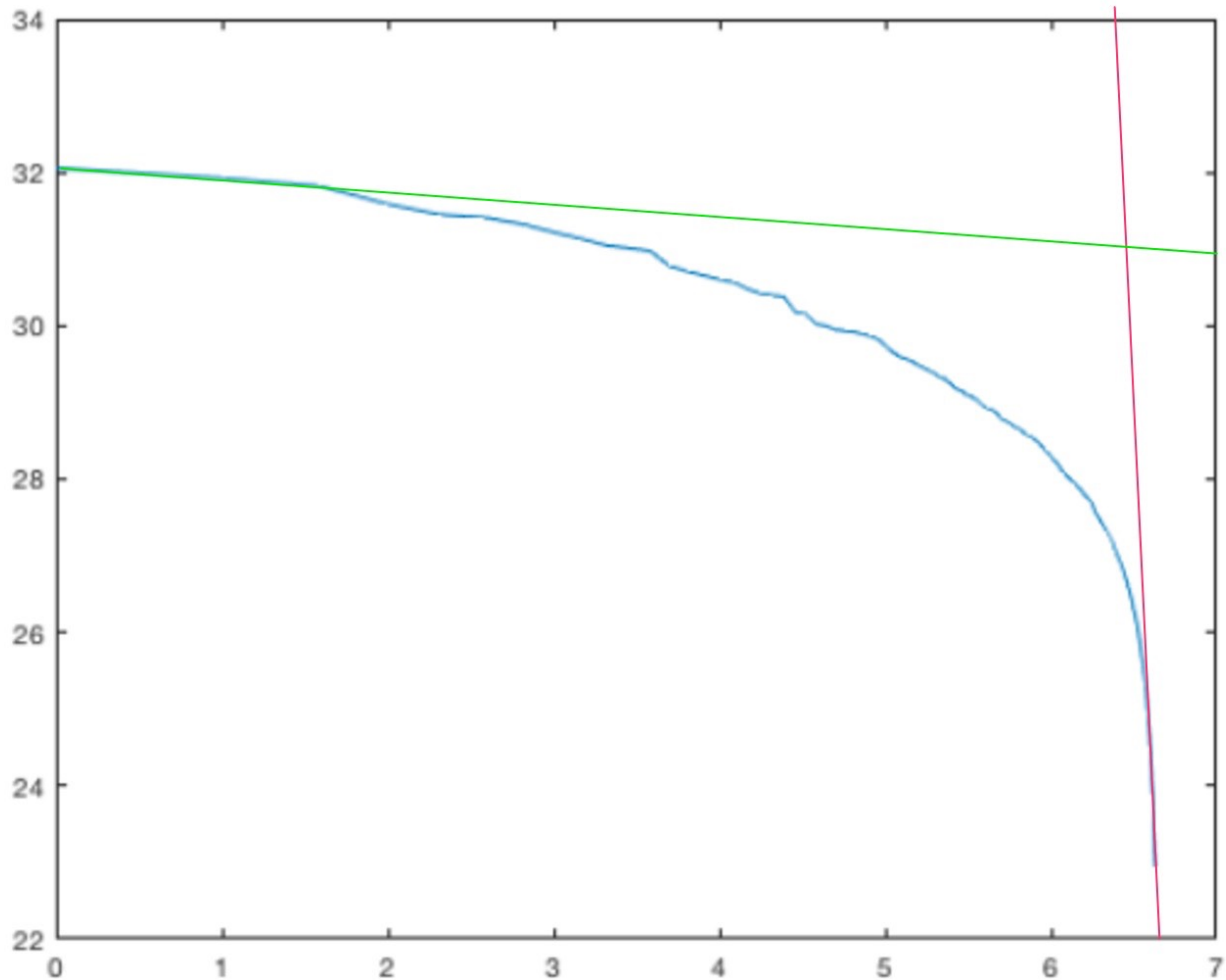
(d) Swissroll

2*7/6.5 ~ 2

The turning tea-pot



Tea-pot dimension estimation



https://arxiv.org/abs/1702.08638

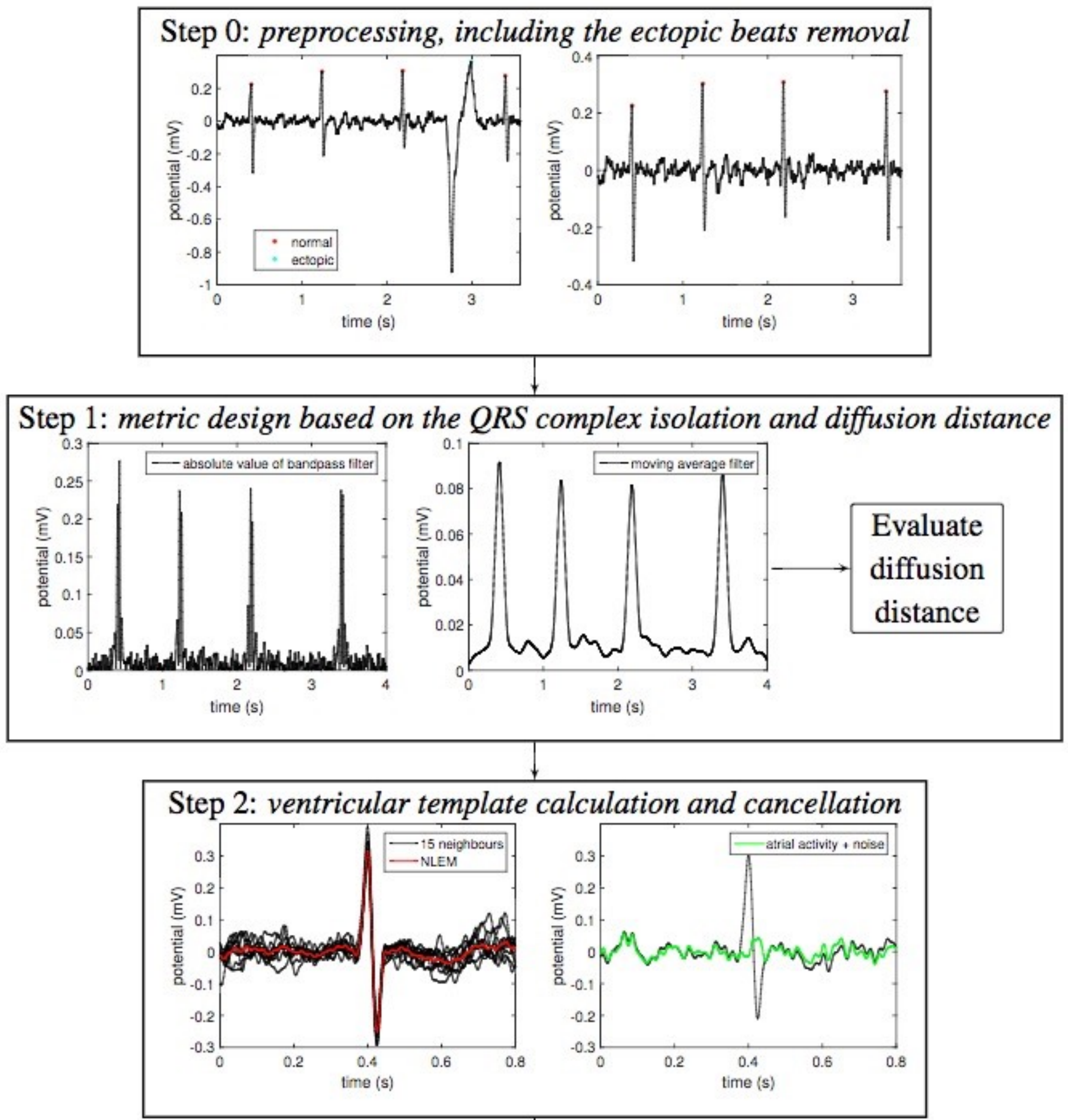
Single-lead f-wave extraction using diffusion geometry

John Malik[1*], Neil Reed[1*], Chun-Li Wang[2,3†], Hau-tieng Wu[1,4†]

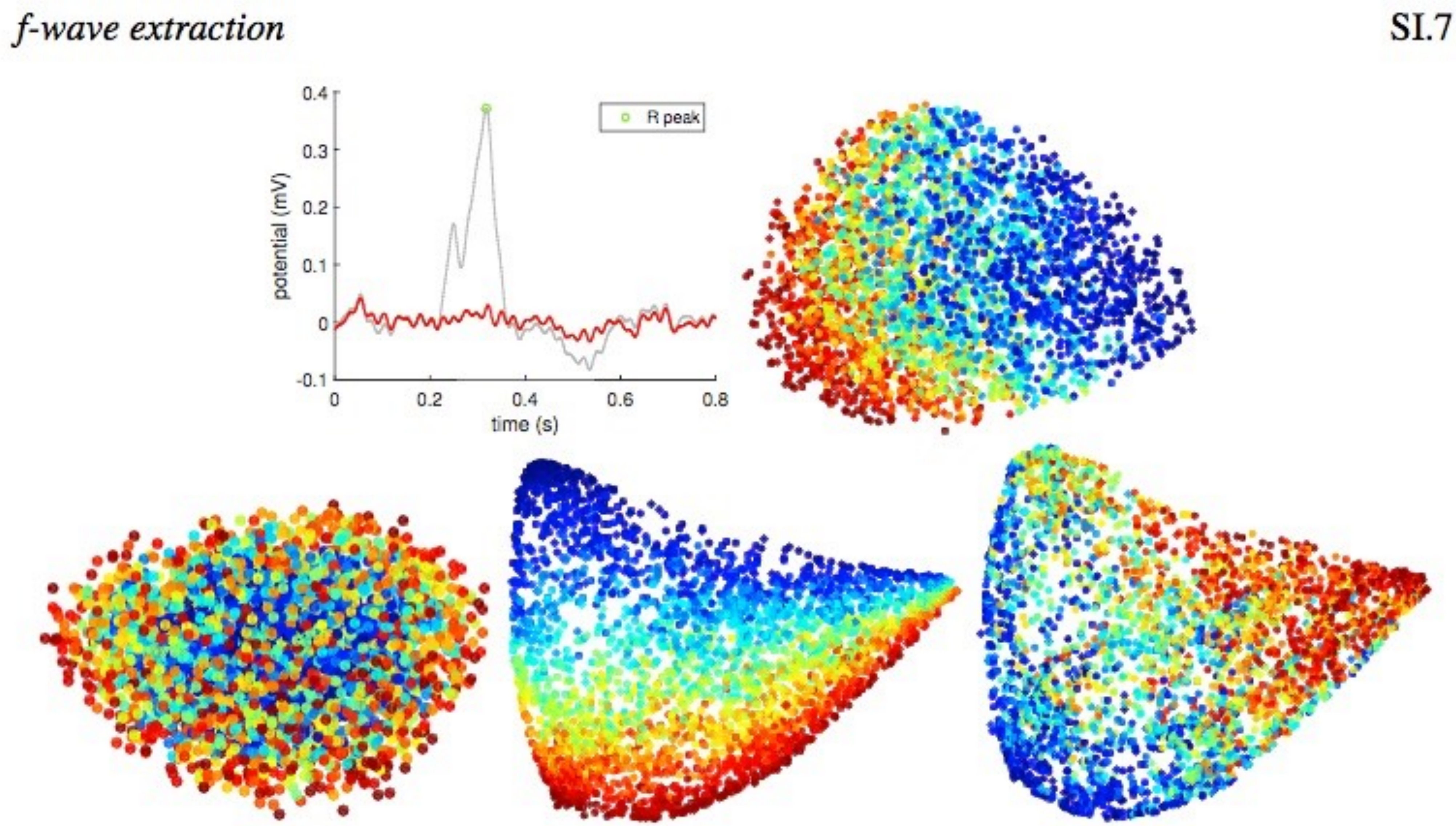
[1] Department of Mathematics, University of Toronto, Toronto, Ontario, Canada
[2] Cardiovascular Division, Department of Internal Medicine, Chang Gung Memorial Hospital, Linkou Medical Center, Taoyuan, Taiwan
[3] College of Medicine, Chang Gung University, Taoyuan, Taiwan
[4] Mathematics Division, National Center for Theoretical Sciences, Taipei, Taiwan
*: these two authors contribute equally to this work. †: co-correspondence.

Atrial fibrillation (Af) is the most commonly sustained arrhythmia encountered in clinical practice and continues to receive considerable research interest. Interventions such as rhythm or rate control improve Af-related symptoms and may preserve cardiac function. However, current Af management guidelines provide no treatment recommendations that take the various mechanisms and patterns of Af into account [25, 21] and therefore tests are developed that quantify Af and guide its management. The fibrillation wave (f-wave) related analysis of the surface ECG or long-term Holter monitoring for Af patients is undoubtedly one of the most challenging questions encountered in the clinical practice [36, 3]; for example, what is the mechanism underlying the initiation, termination, and maintenance of paroxysmal Af [23], and what is the outcome of Af treatment [29]? A summary of the available information on

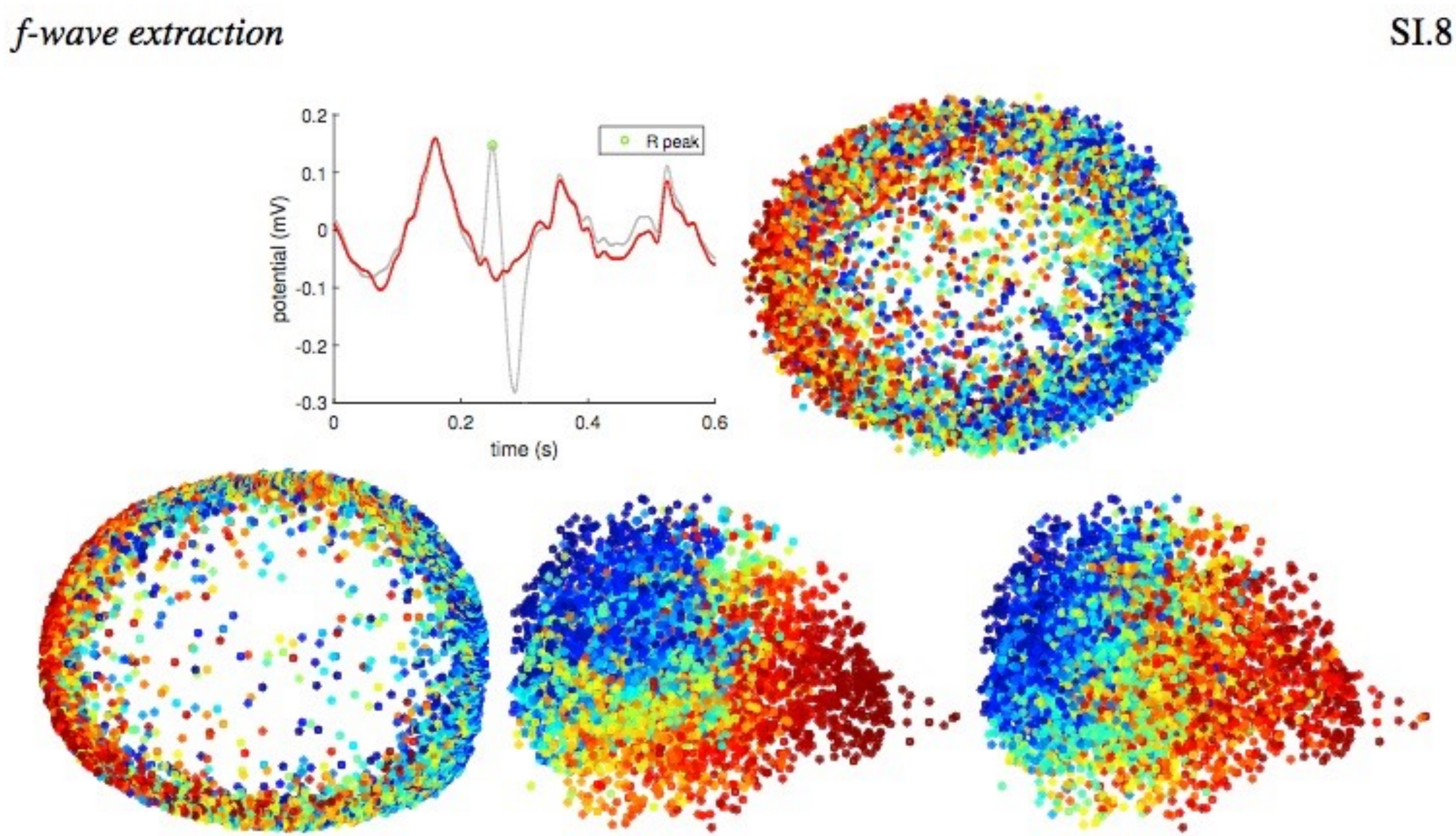
Signal Processing



Normal Heart



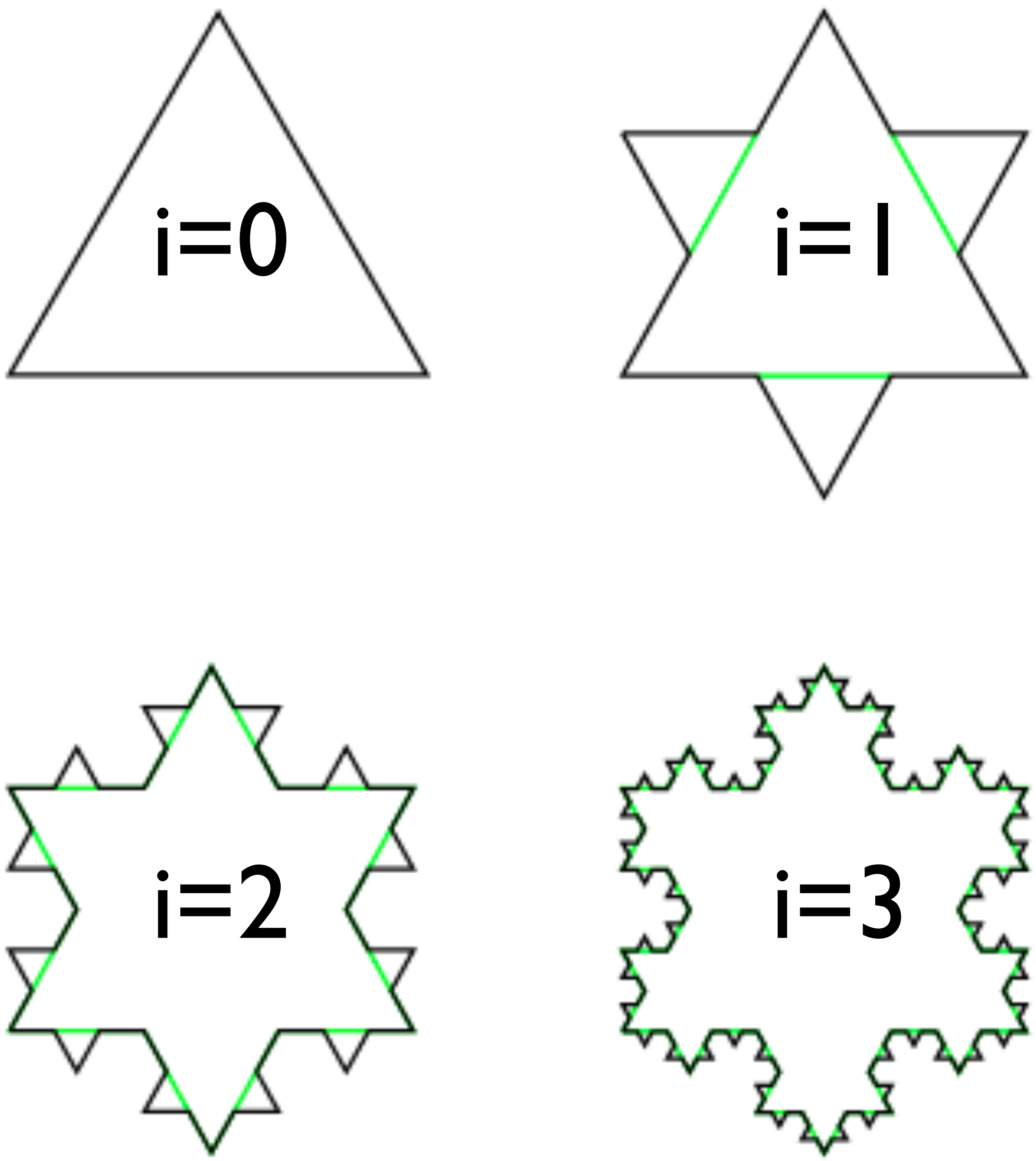
Anomalous Heart



Integer and fractional dimensions

- We saw dimensions 1,2,3,….
- can there be fractional dimensions?

Koch Snowflake



$$\epsilon_i = \frac{1}{3^i}$$

$$n_i = 3 \times 4^i$$

$$n_i = 3 \times \left(\frac{1}{\epsilon_i} \right)^{\boxed{\frac{\log 4}{\log 3}}}$$

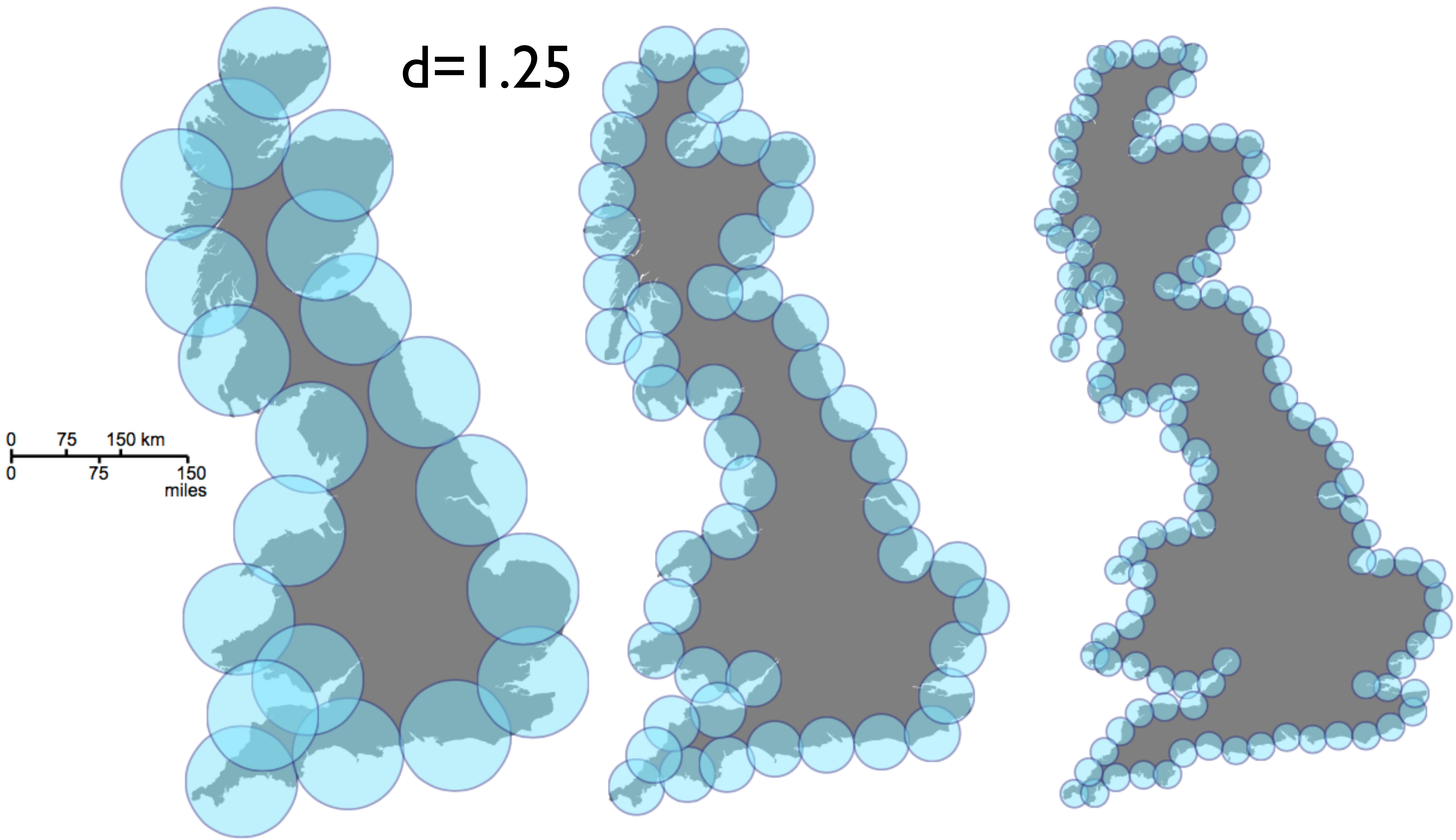
=dimension
=1.26

Snowflake corresponds to $i \to \infty$

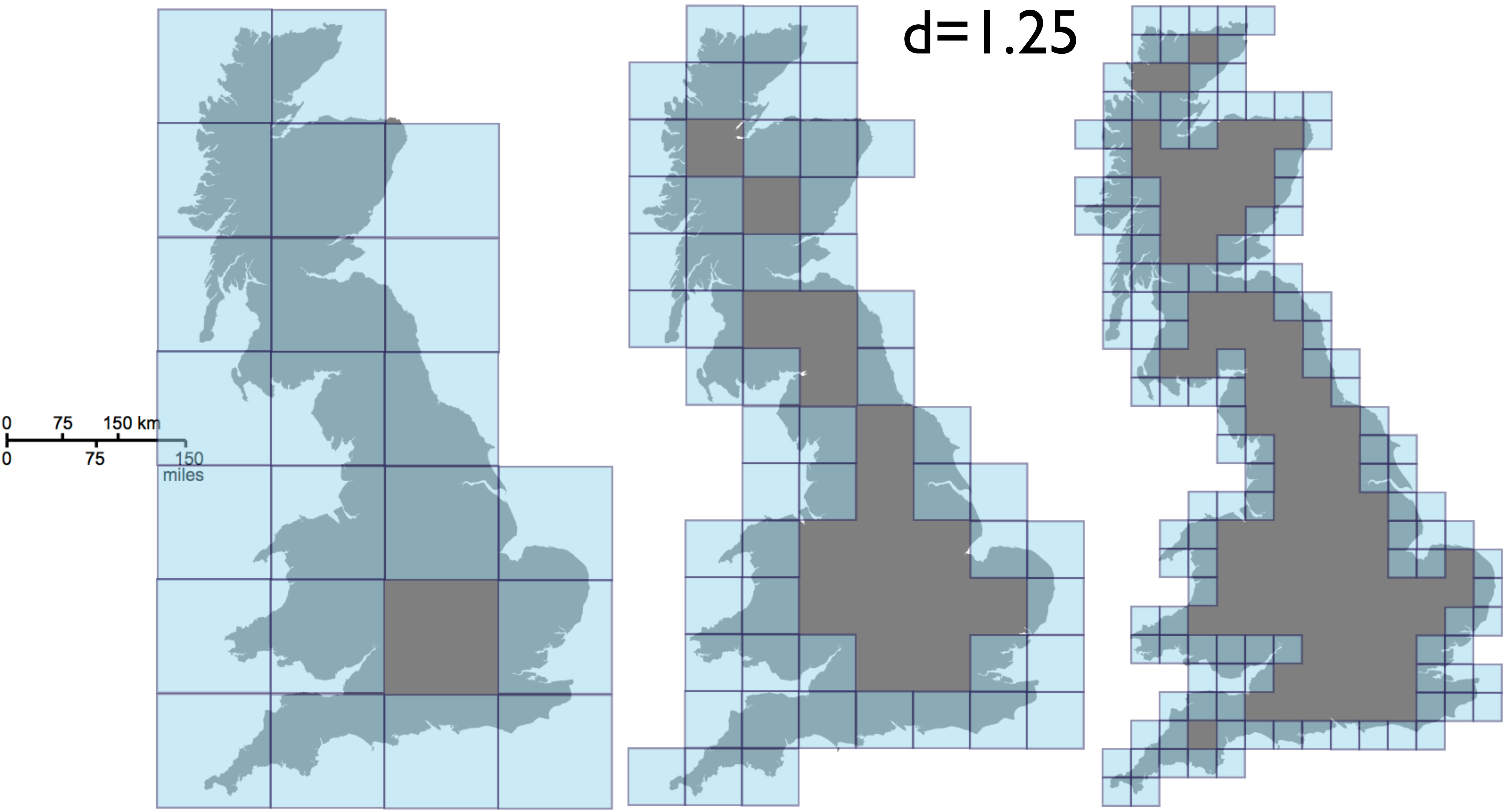
Variations on a theme

- Partition count can be defined in many ways
  - Housdorff dimension: max distance between 2 points in the same cell
  - VQ: Average distance to representative.
  - Epsilon-cover: all points are at a distance of at most epsilon from a representative.
  - One can use grids, circles, triangles,line segments ….
- **In most cases they all converge to the same number!**

How many **balls** of radius r it takes to cover the British coastline?



How many **squares** of size $1/2^i$ it takes to cover the British coastline?

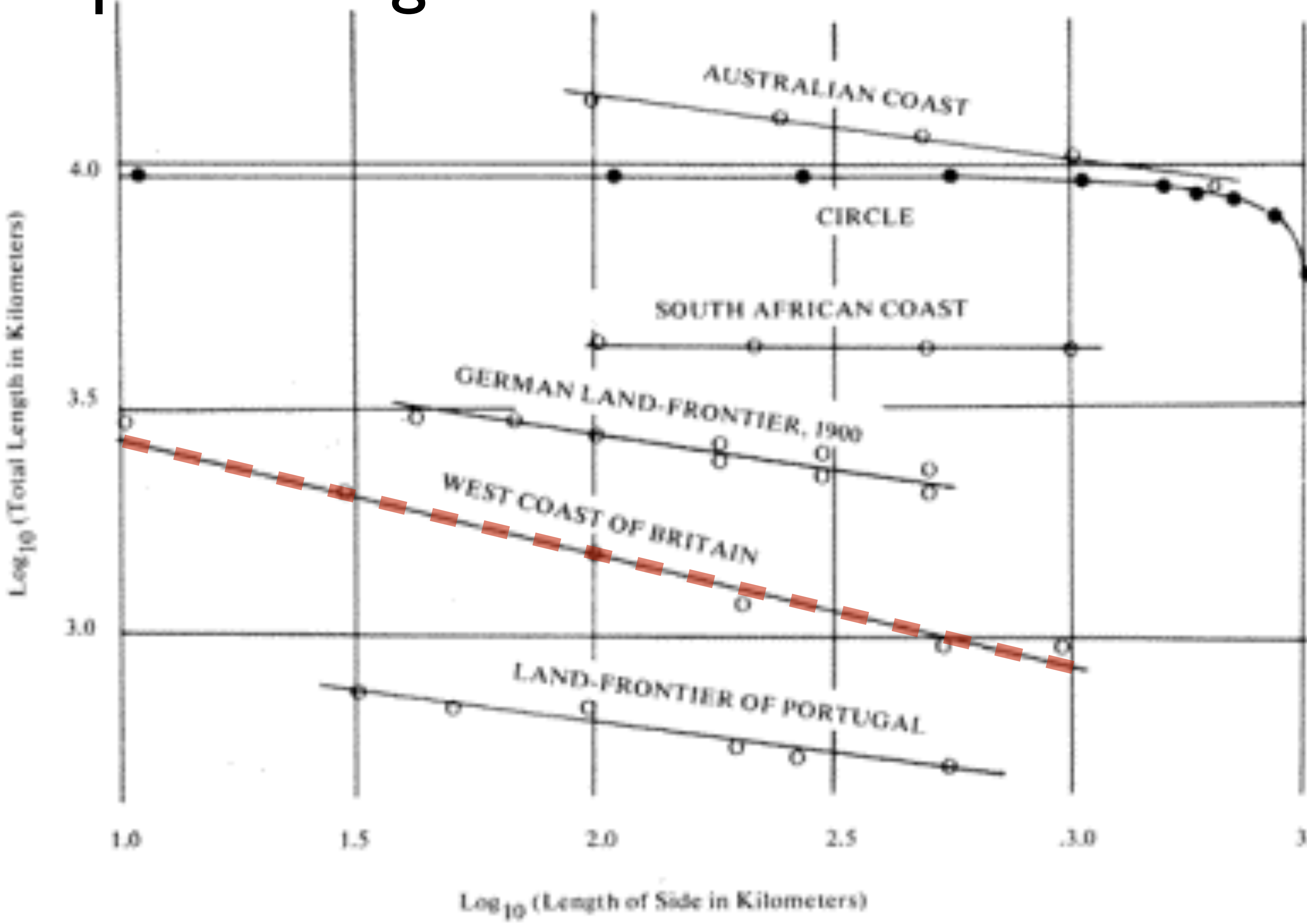


Using line segments: how many **line segments** of length $1/2^i$ it takes to trace the British coastline?

d=1.25

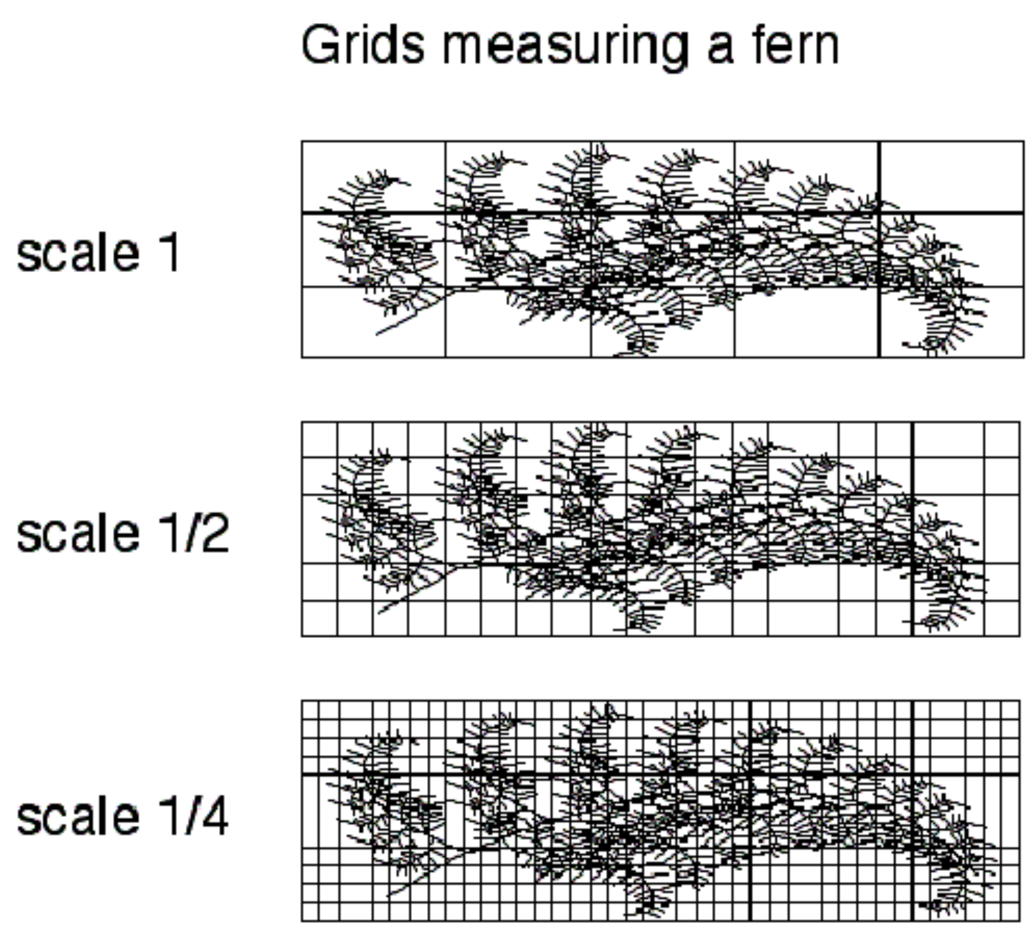
A comparative study of coastlines

Only slopes are significant!



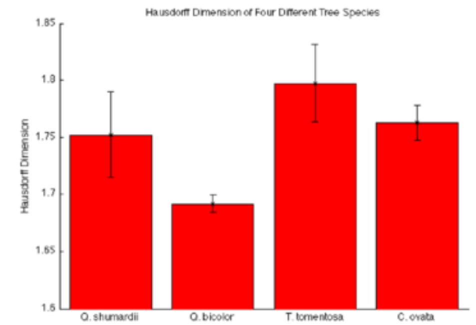
Plants

d btwn 1 and 2

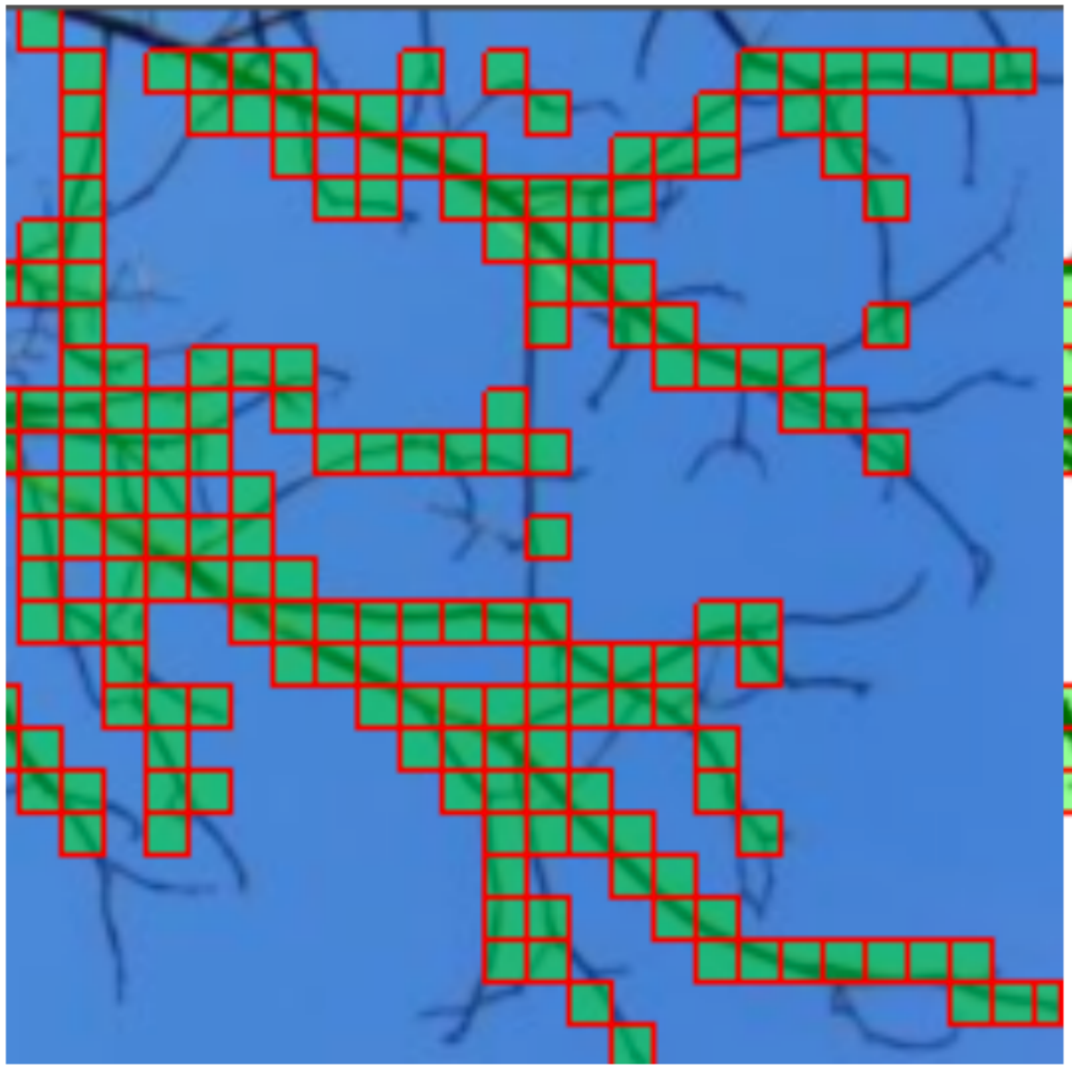


Dimensions for different tree types
Boccio and Bastian 2011

http://www.andreasbastian.com/fractal/fractal.html

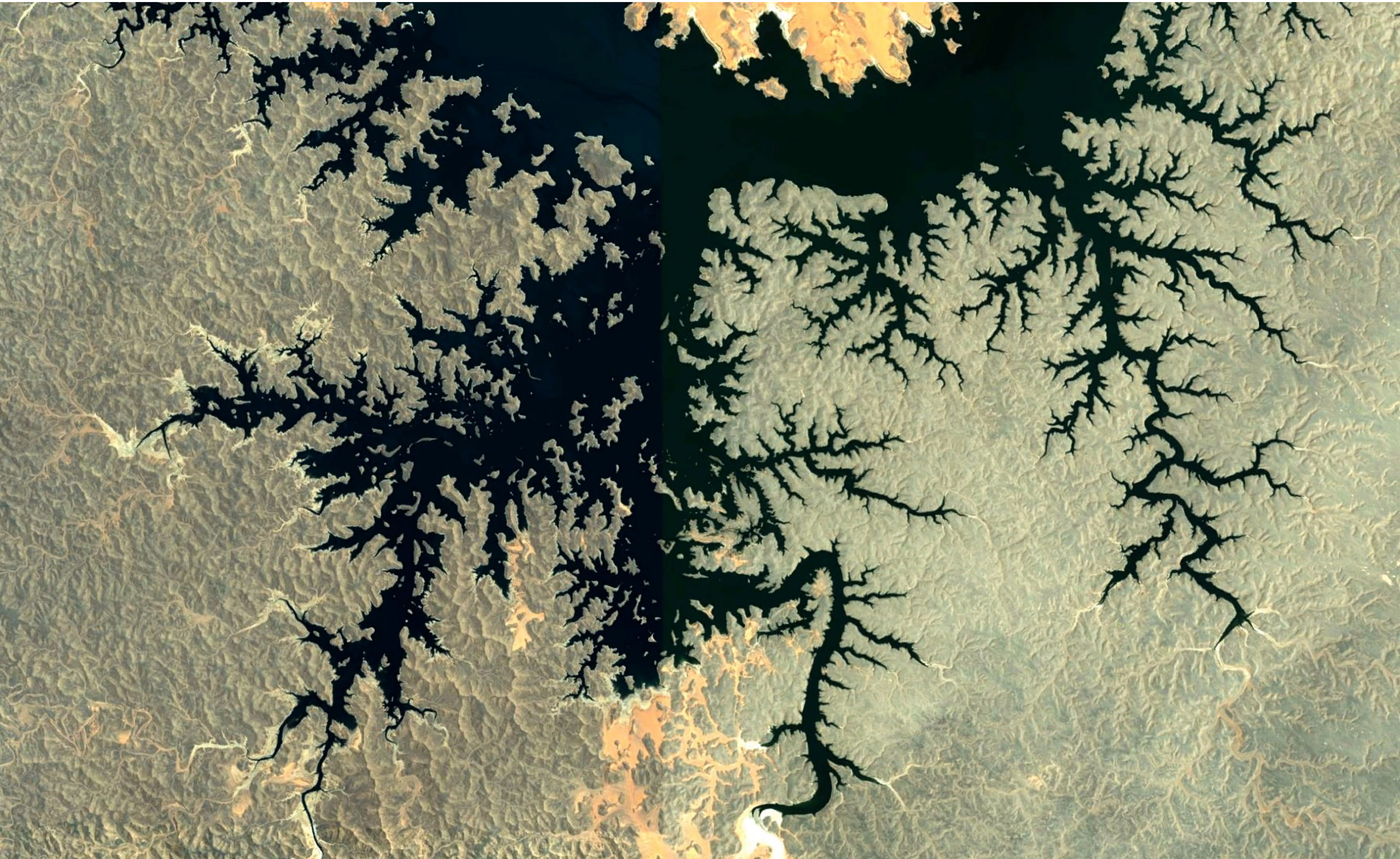


Original (color)

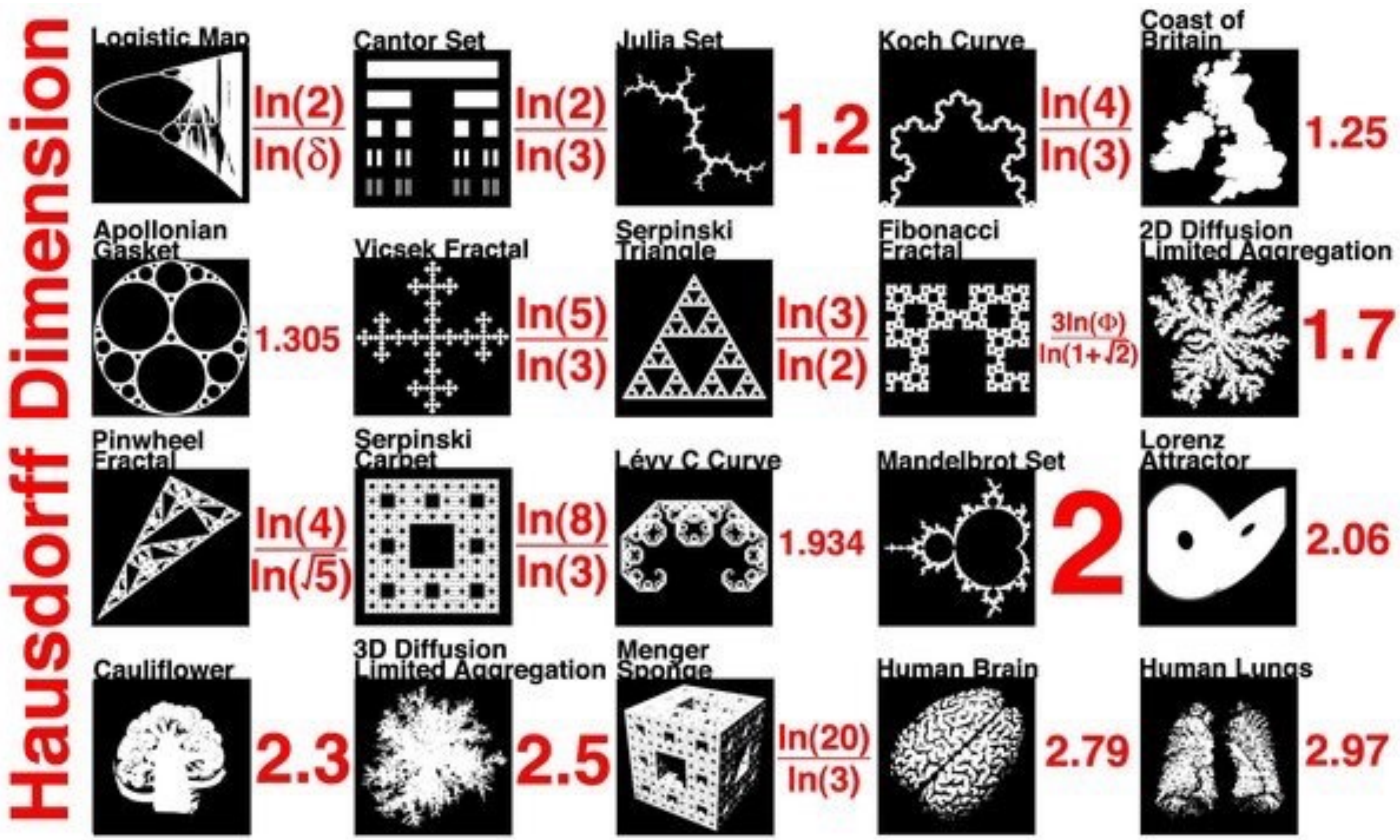


preprocessed

The nile from the air.



More examples



Examples of objects with
different Hausdorff Dimension:
http://en.wikipedia.org/wiki/List_of_fractals_by_Hausdorff_dimension

Application to gesture recognition

Facial Motion Capture - Avatar

Motion Capture - Avatar

- Intrinsic dimension=number of degree of freedom < number of muscles in the human face: around 23.
- 23 markers suffice to capture all expressions!

Degrees of freedom of facial movements in face-to-face conversational speech

Gérard Bailly, Frédéric Elisei, Pierre Badin, Christophe Savariaux 2007

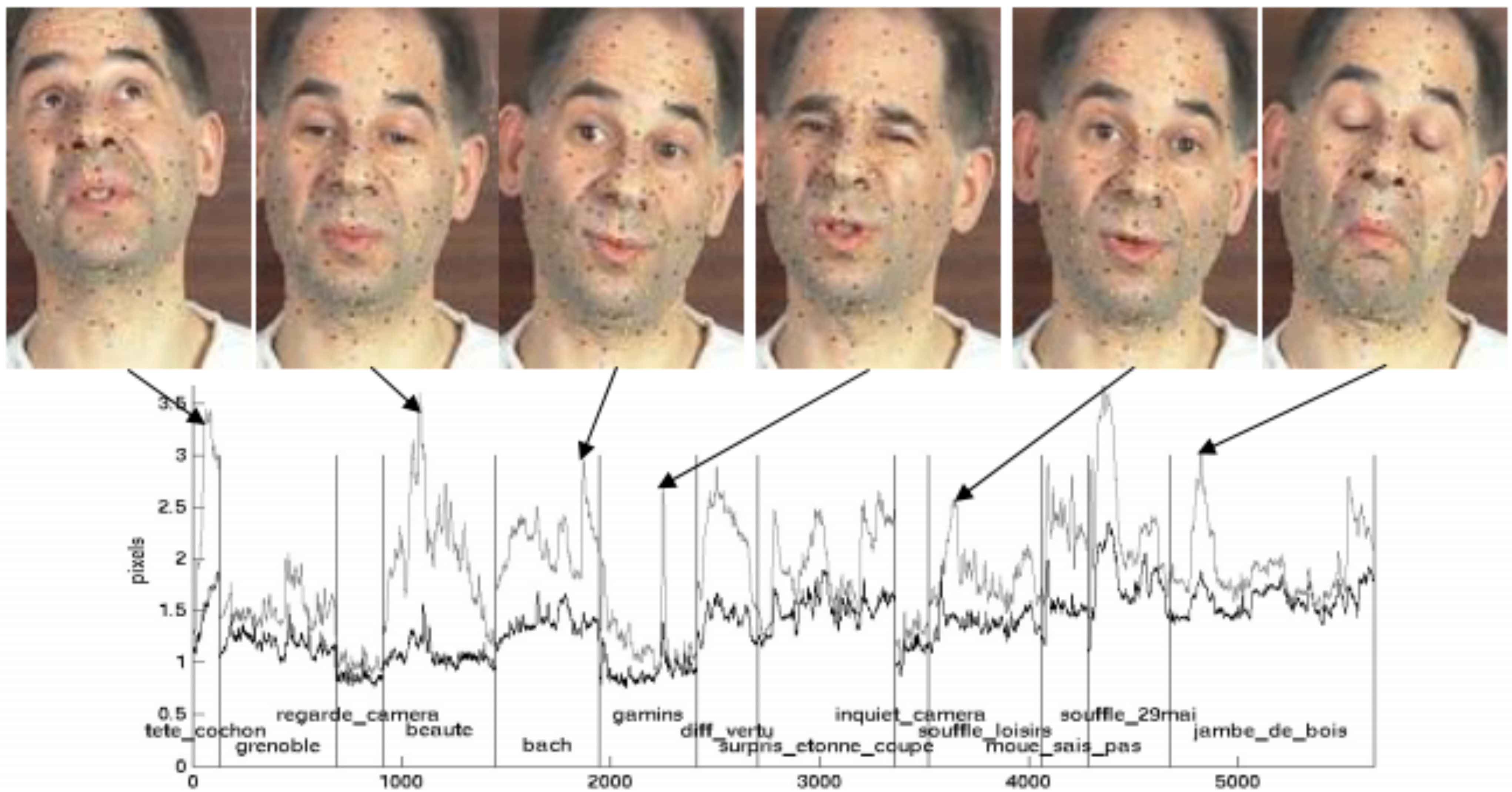
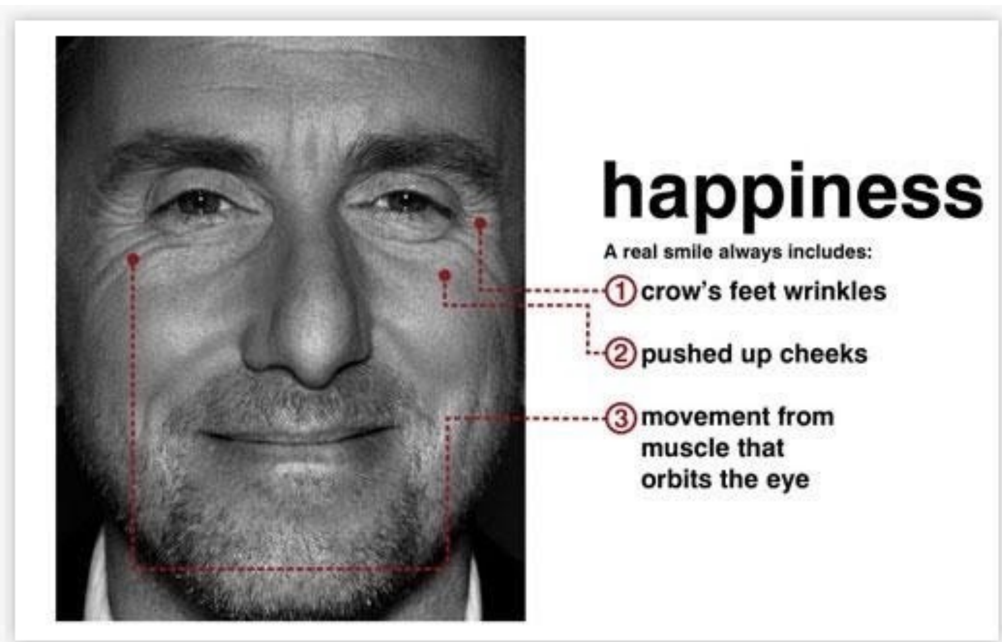


Figure 1: Comparing prediction errors of facial shapes using a model built using 52 speech visemes (light gray) with one incorporating 102 additional expressemes (dark gray), for a series of selected video sequences. The mean error lowers from 1.7 to 1.3 pixels. Frames shown at the top are generating the most important prediction errors of the speech-only model.

- Viseme ~ simple model using 11 DoF (Degrees of freedom)
- expresseme ~ Using additional codewords to detect extremal expressions
- Goal of work: complement speech signal to improve language recognition.

Emotions and facial expressions



Human facial expressions are universal, not learned

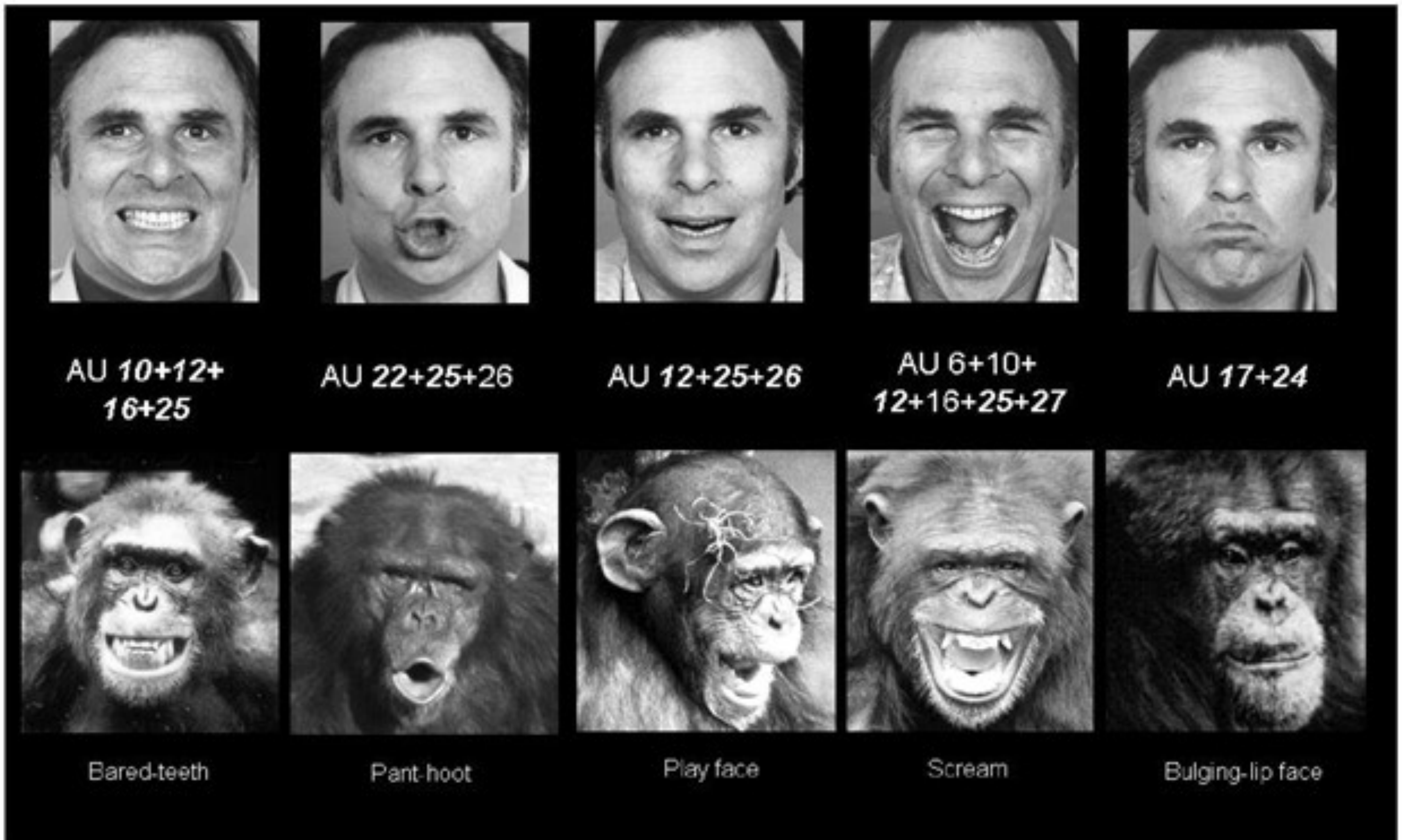
Paul Ekman / 1963 / New Guinea

(a) show me what your face would look like if you were about to fight.

(b) show me what your face would look like if you learned your child had died.

(c) show me what your face would look like if you met friends.

Human/ape facial expressions



https://www.youtube.com/watch?v=R6galodflTQ

Different notions of dimension and low-D embeddings

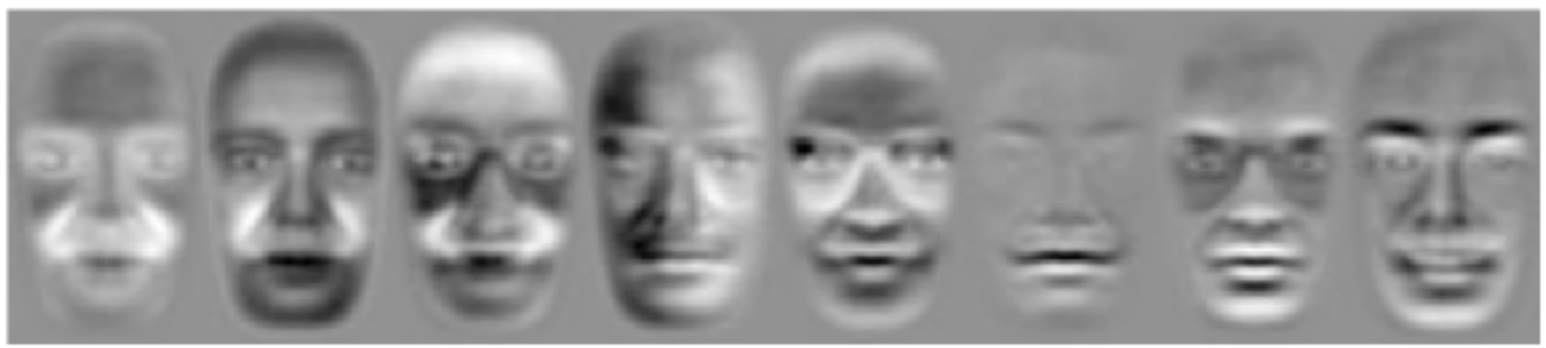
- PCA (Linear dimension)
- Locally near Embedding
- Differential Geometry
- Doubling / Haussdorf dimension
- RP-trees

Eigen-Faces

Beyond Eigenfaces: Probabilistic Matching for Face Recognition

Baback Moghaddam
Mitsubishi Electric Research Laboratory

Wasiuddin Wahid and Alex Pentland
MIT Media Laboratory

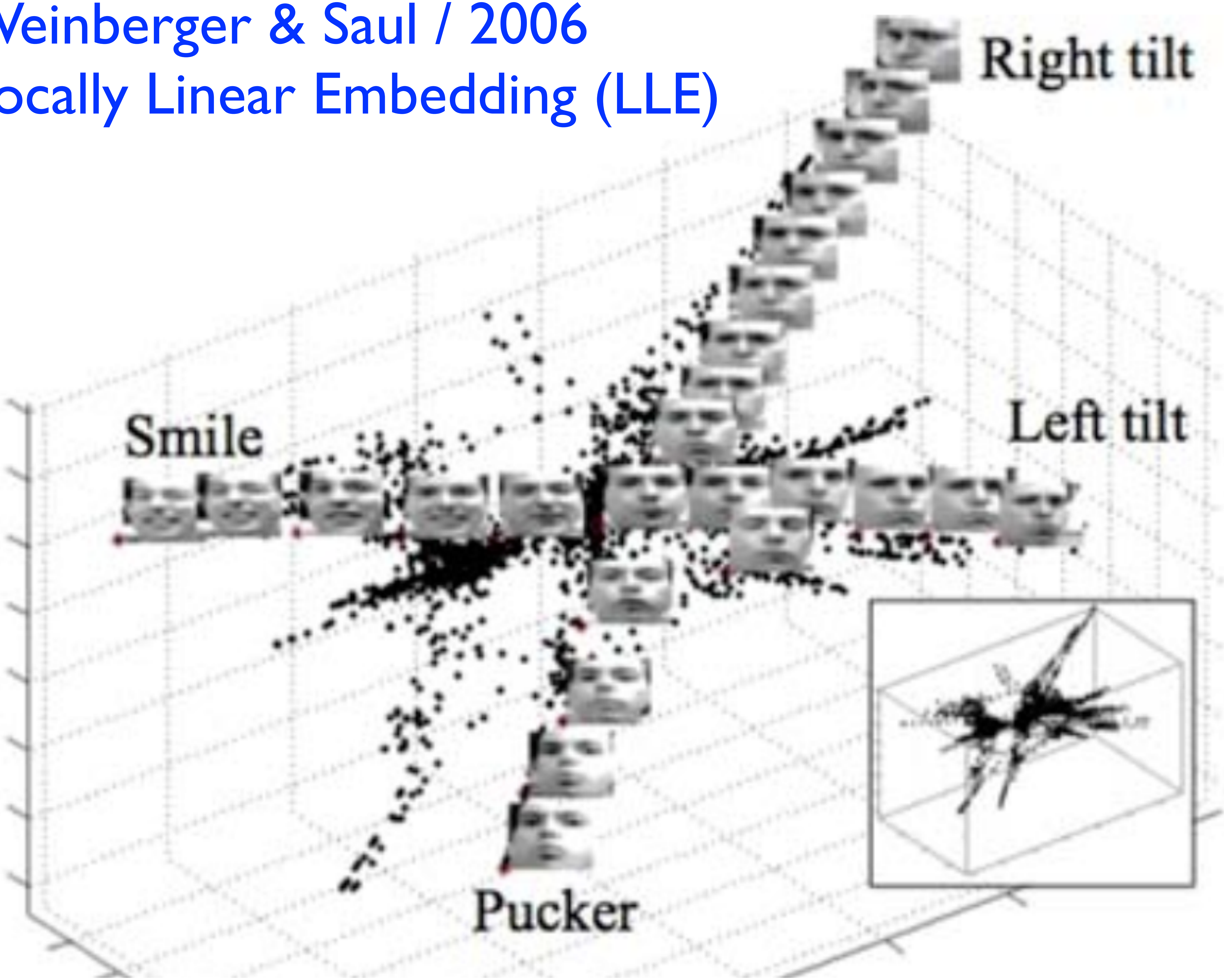


(a)

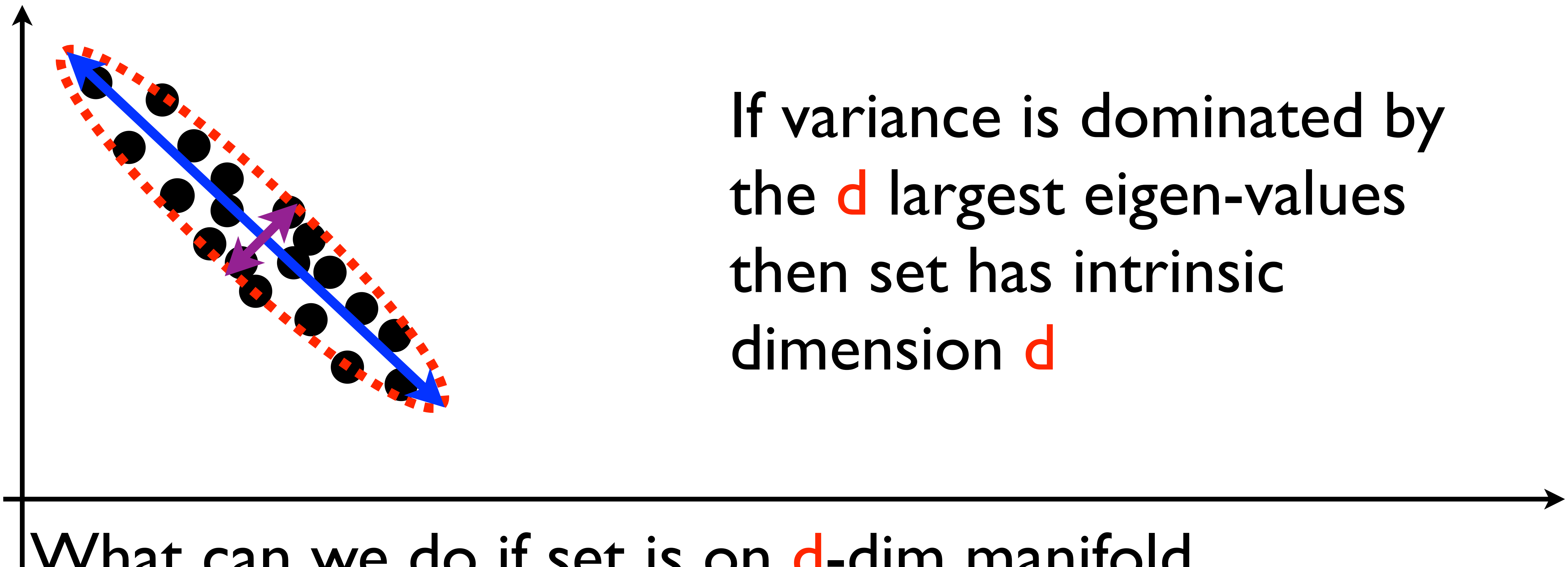
(b)

Figure 6: "Dual" Eigenfaces: (a) Intrapersonal, (b) Extrapersonal

Weinberger & Saul / 2006
Locally Linear Embedding (LLE)



PCA



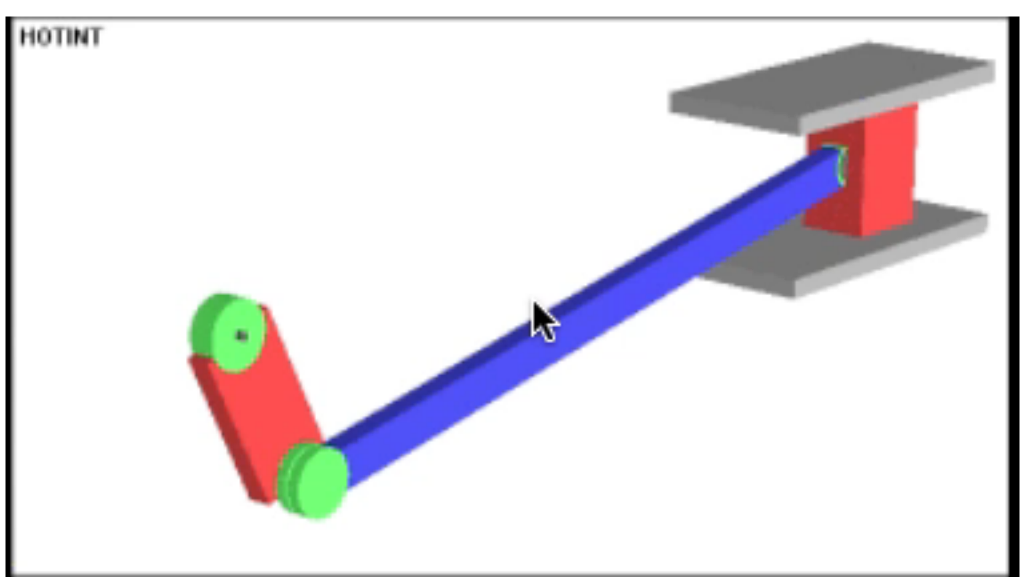
If variance is dominated by the d largest eigen-values then set has intrinsic dimension d

What can we do if set is on d-dim manifold that is not affine?

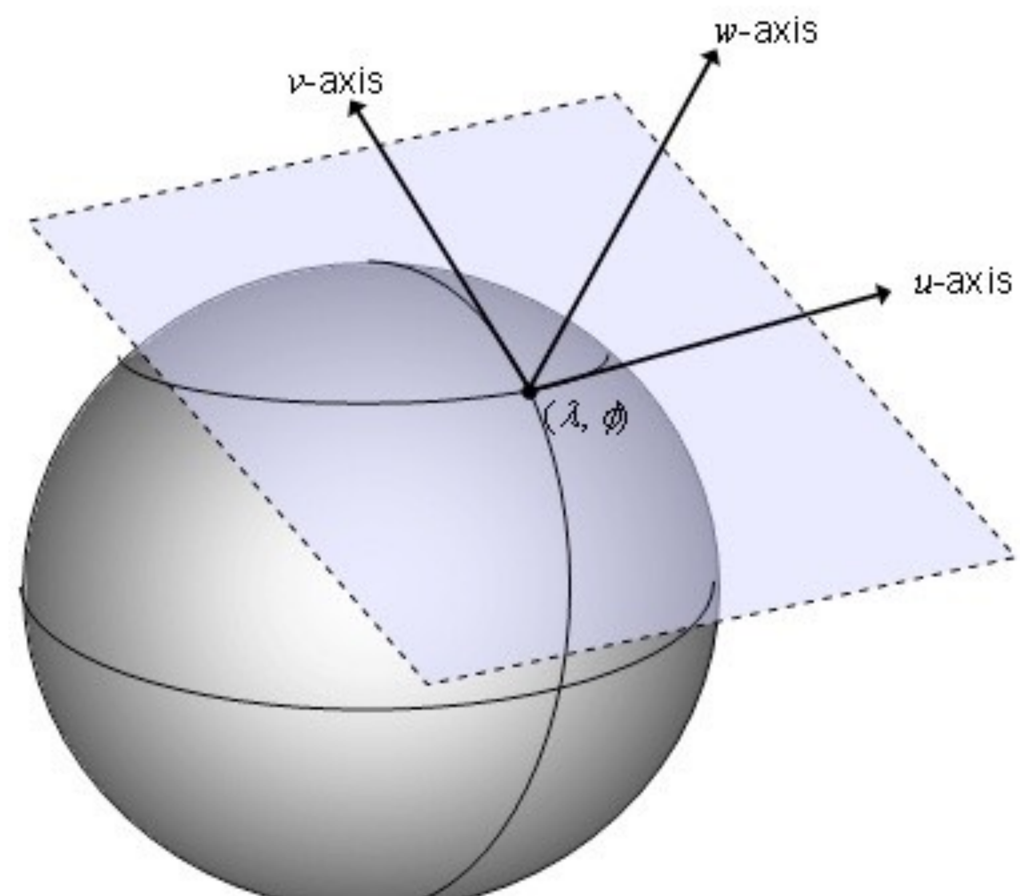
Partition space into small regions in which the set is approximately affine.

Manifold dimension

- Differentiable manifold dimension: dimension of local tangent space.
- local, infinitesimally small regions. Requires smoothness. Hard to use for sampled data.



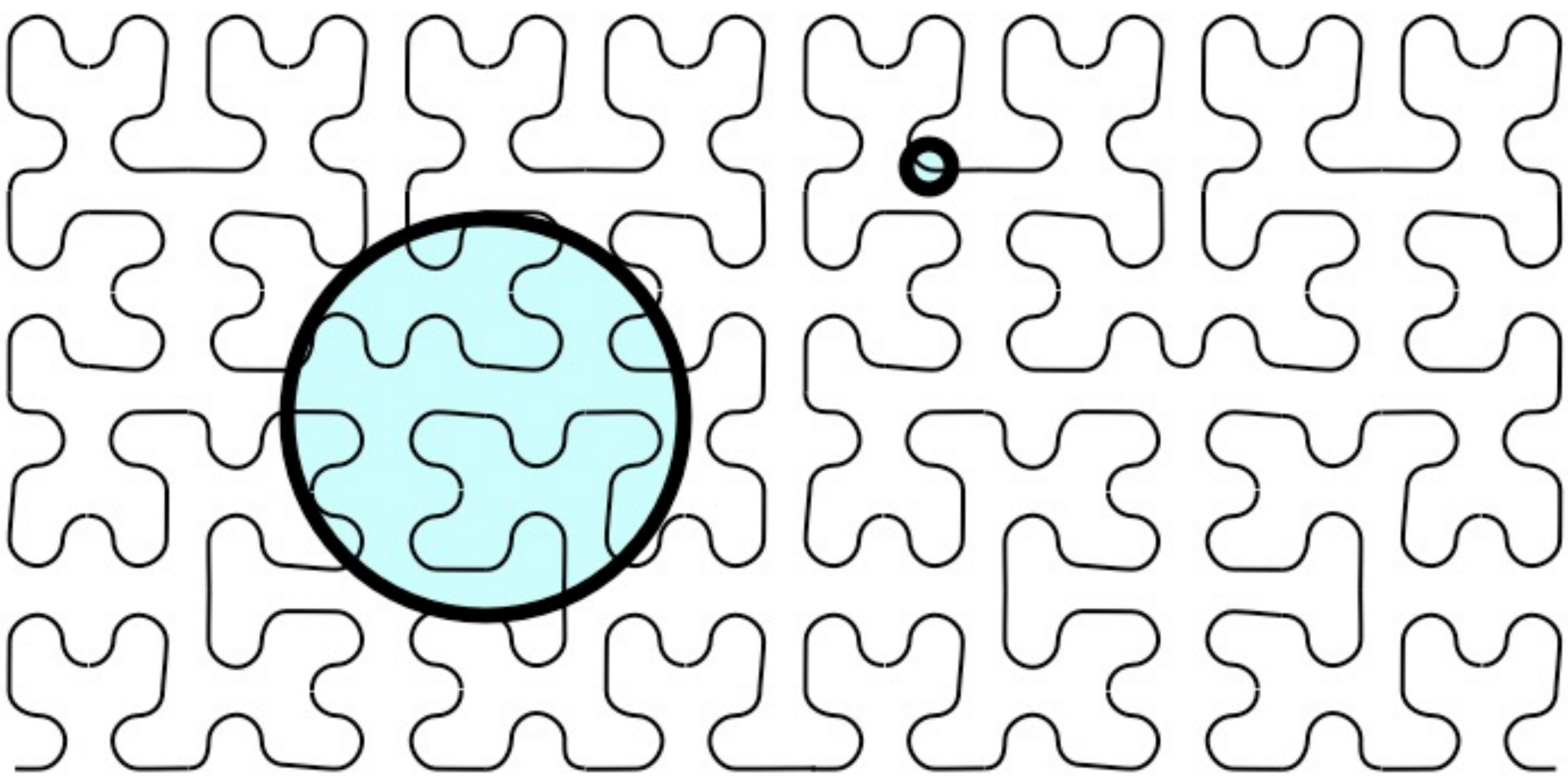
http://mathinsight.org/dynamical_system_idea



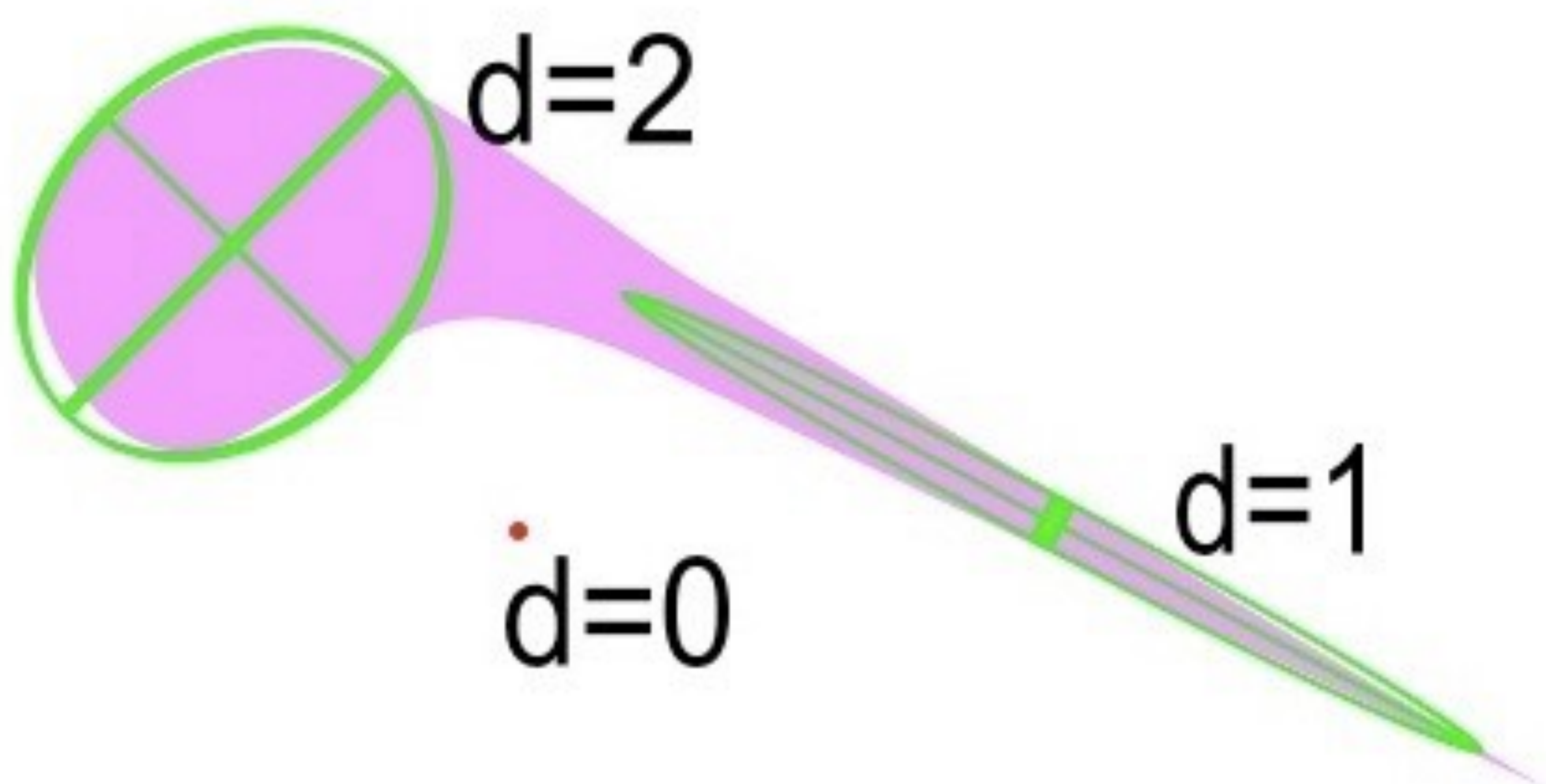
Doubling dimension

- Similar to Hausdorff dimension.
- Doubling dimension of set S is d if:
  - For any ball B of radius r
  - Intersection of set S and B can be covered by at most 2^d balls of radius $r/2$.
- Global, all scales, does not require smoothness.
- More general than manifold dimension.

Dimension can depend on scale



Dimension can depend on location

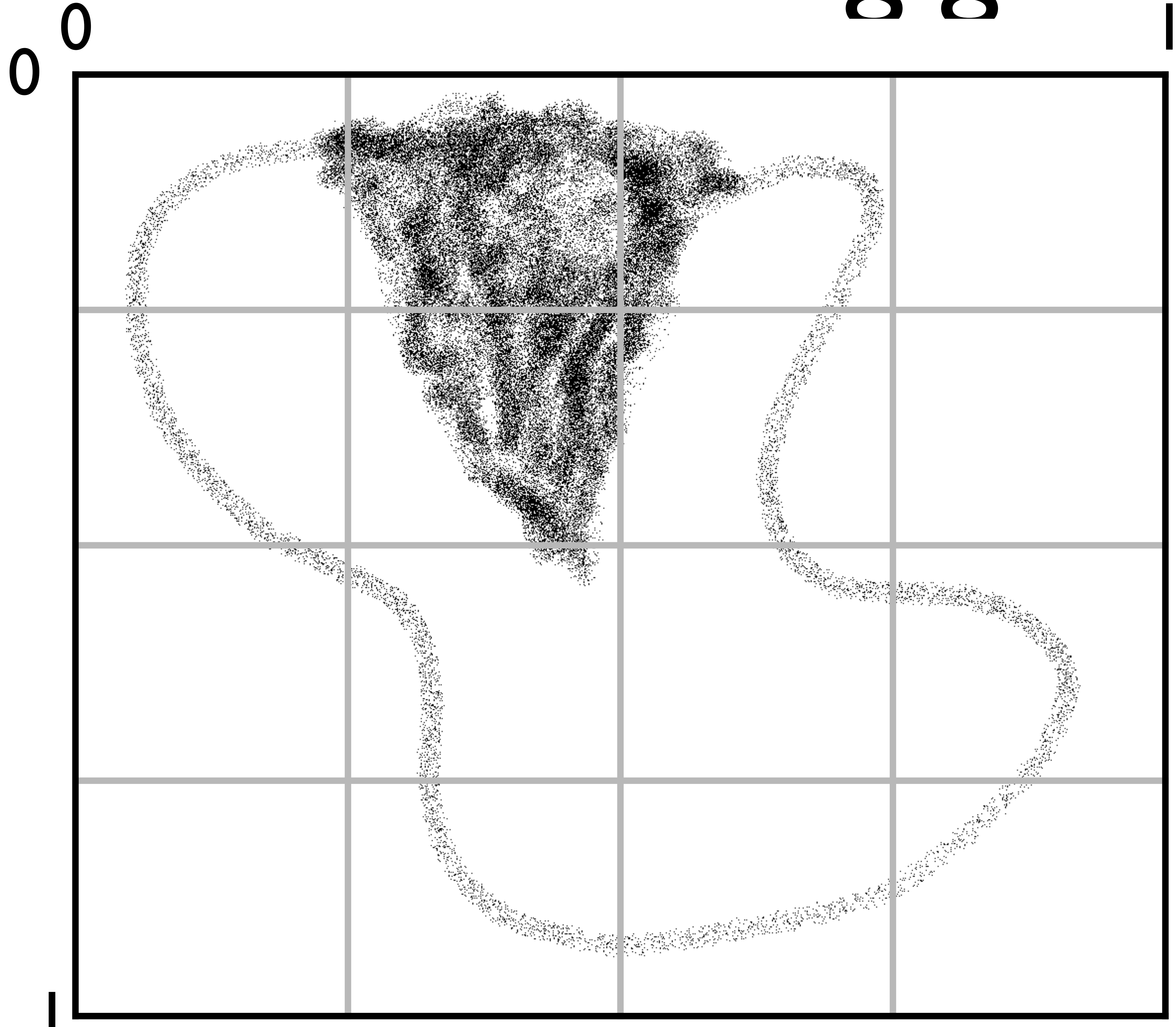


Haussdorff vs. PCA

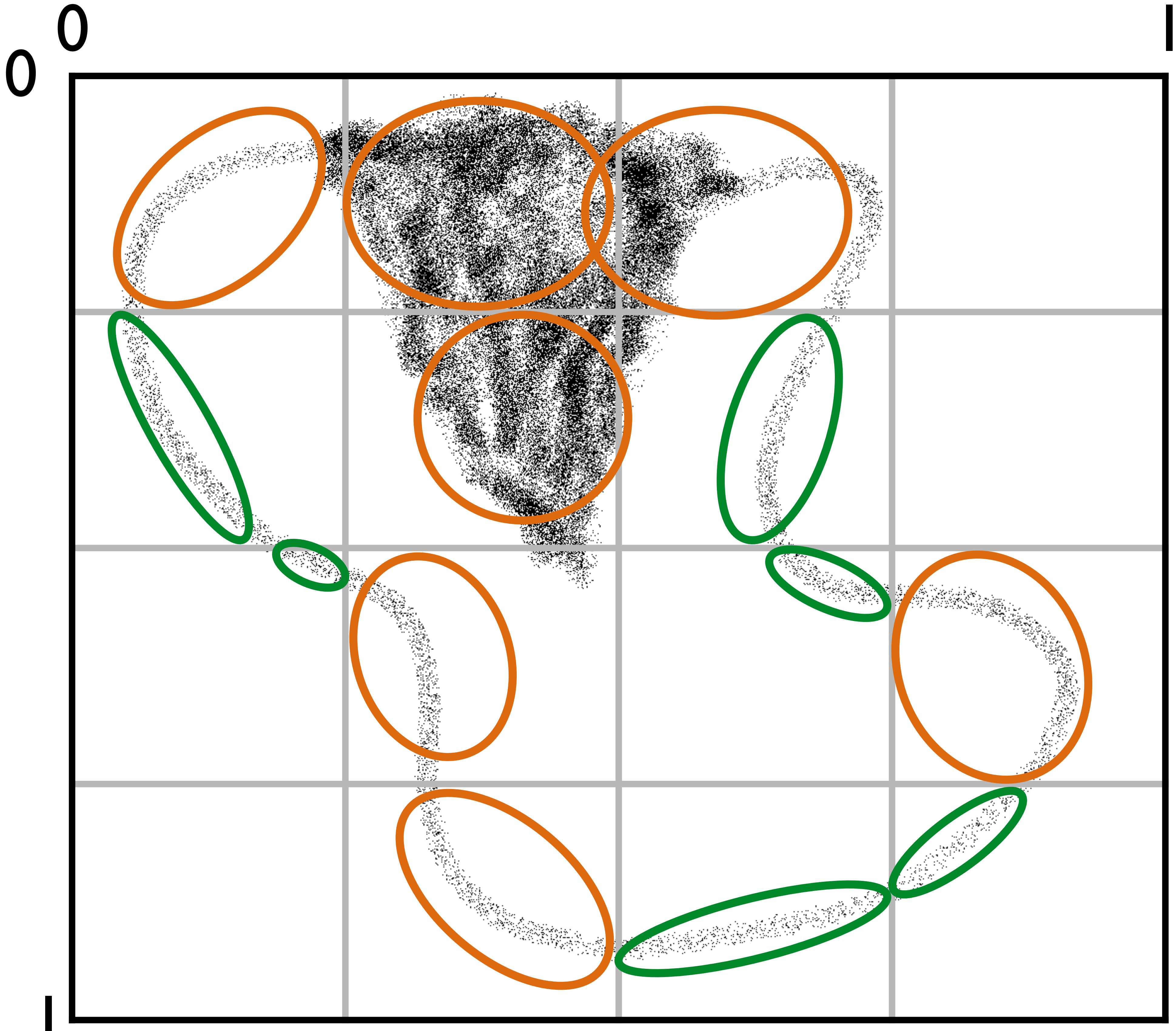
- With PCA we can find a low dimensional representation (eigen-vectors explaining 90% of variance). But only for a linear mapping.
- With Hausdorff dimension we can identify arbitrary low dimensional structure, but there is no coordinate system.
- Can we combine the two?

Data

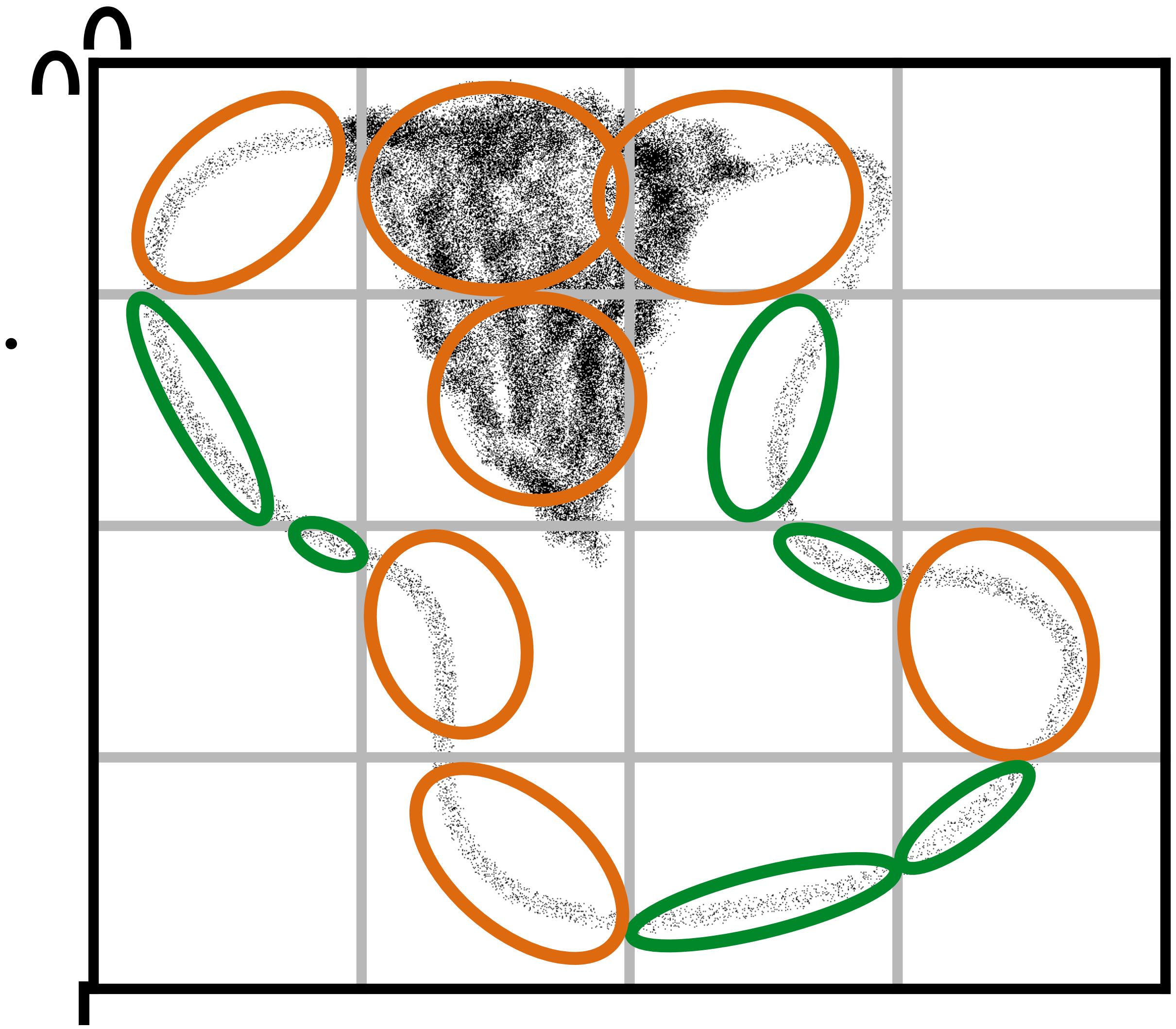
Partition using grid



PCA in each cell



- Green ellipses: First eigenvector explains > X% of variance in cell. - we are done.
- Orange ellipses: First eigenvector explains < X% of variance in cell - subdivide cell.
- In high dimensions data can be divided very unequally among the cells. -> leads to non-uniform accuracy.
- We need a better way to divide cells.



Local covariance dimension

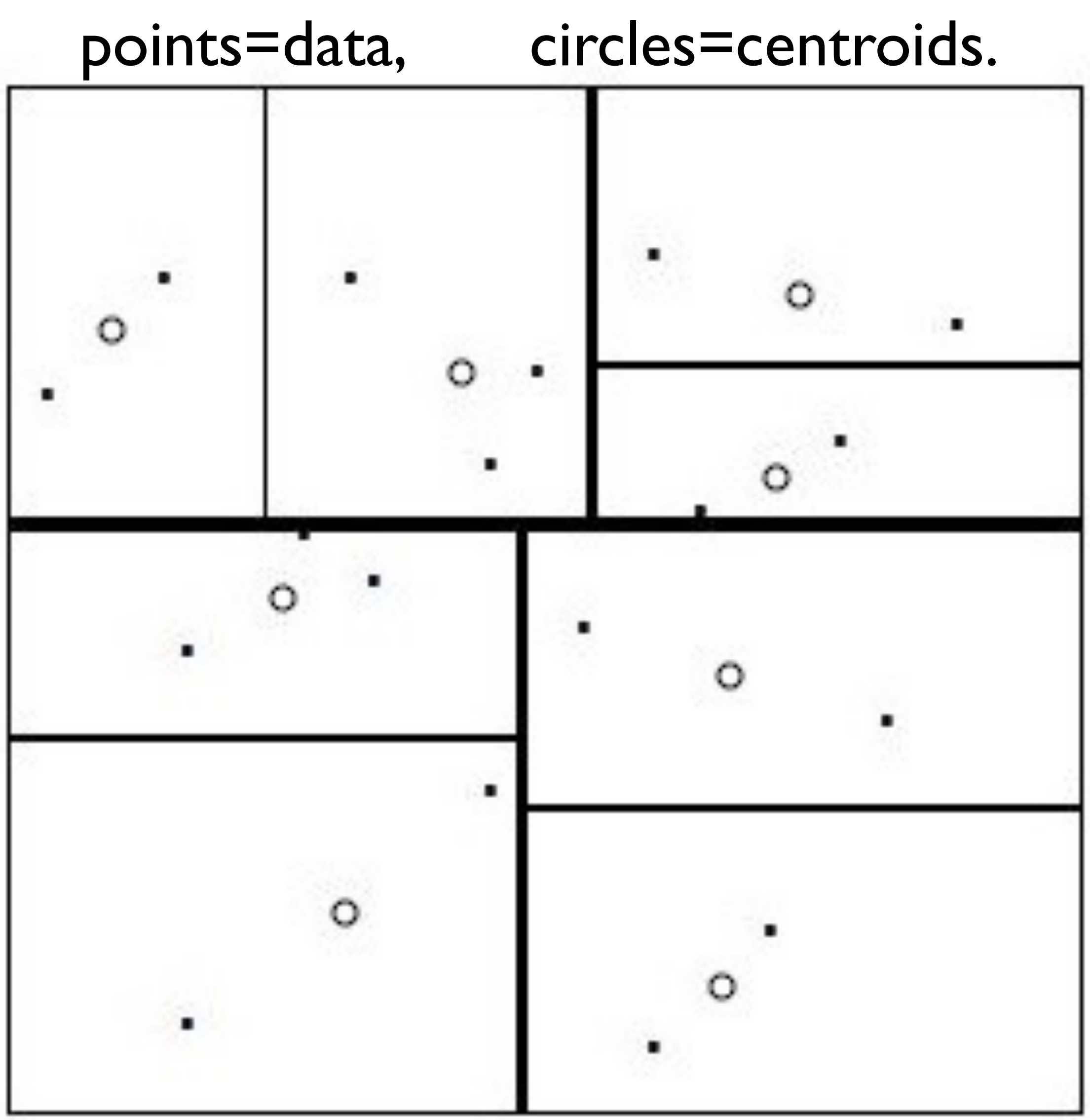
- $S\{x_i\}_{i=1}^N$ is a finite set in R^D (a sample).
- Mean vector: $\mu = \frac{1}{N}\sum_{i=1}^N x_i$. Assume wlog $\mu = 0$
- Covariance matrix: $C = \frac{1}{N}\sum_{i=1}^N x_i^T x_i$
- $\{v_i\}_{i=1}^D$ are eigen-vectors of C with eigen-values $\sigma_1^2 \geq \sigma_2^2 \geq \ldots \geq \sigma_D^2$
- S has covariance dimension (d, ϵ) if

$$\sum_{i=1}^{d} \sigma_i^2 \geq (1-\epsilon) \sum_{i=1}^{D} \sigma_i^2$$

- S has local covariance dimension (d, ϵ) in the ball $B(x, r)$ if $S \cap B(x, r)$ has covariance dimension (d, ϵ).

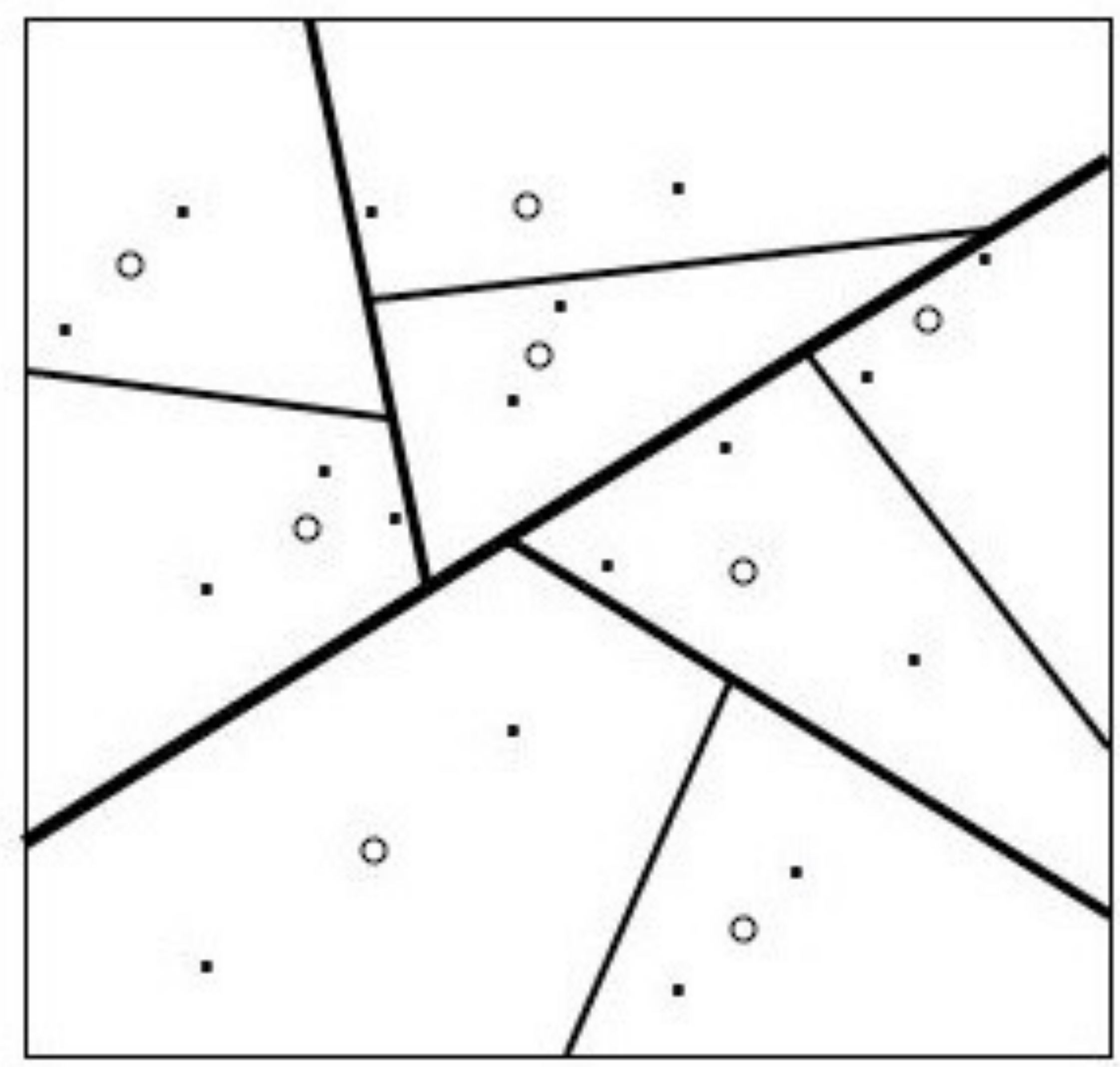
Balanced space partitioning using KD-Trees

- Goal: partition space into regions with similar number of examples in each.
- KD-trees:
  - Choose a coordinate at random.
  - Partition the data at the median.
  - repeat for leaves.
- Works well for low-dimensional spaces.
- Works poorly for data with low intrinsic dimension embedded in a high dimensional space.
- If dimension is D, then D levels are required to half the max-diameter of the cells.
- D=20 -> 2^{20} >1,000,000 cells to reduce the diameter from 1 to 1/2.

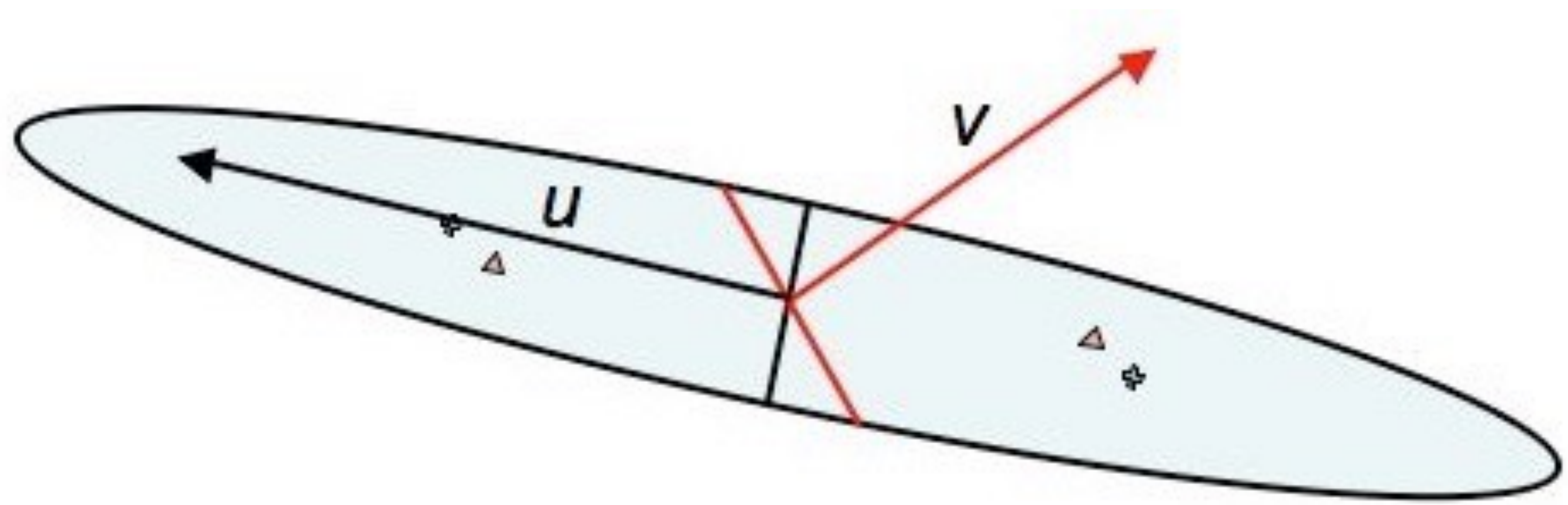


Random-Projection trees

- Goal: partition space into regions with similar number of examples in each.
    - AND create shallow trees if data has low intrinsic dimension.
- RP-trees:
    - Choose a **direction** uniformly at random.
    - Partition the data at the median.
    - repeat for leaves.
- Works well for datasets with low covariance dimension. Even if embedded in a high dimensional space.
- If covariance dimension is d, then d levels are required to half the max-diameter of the cells.



Splitting a set with low covariance dimension

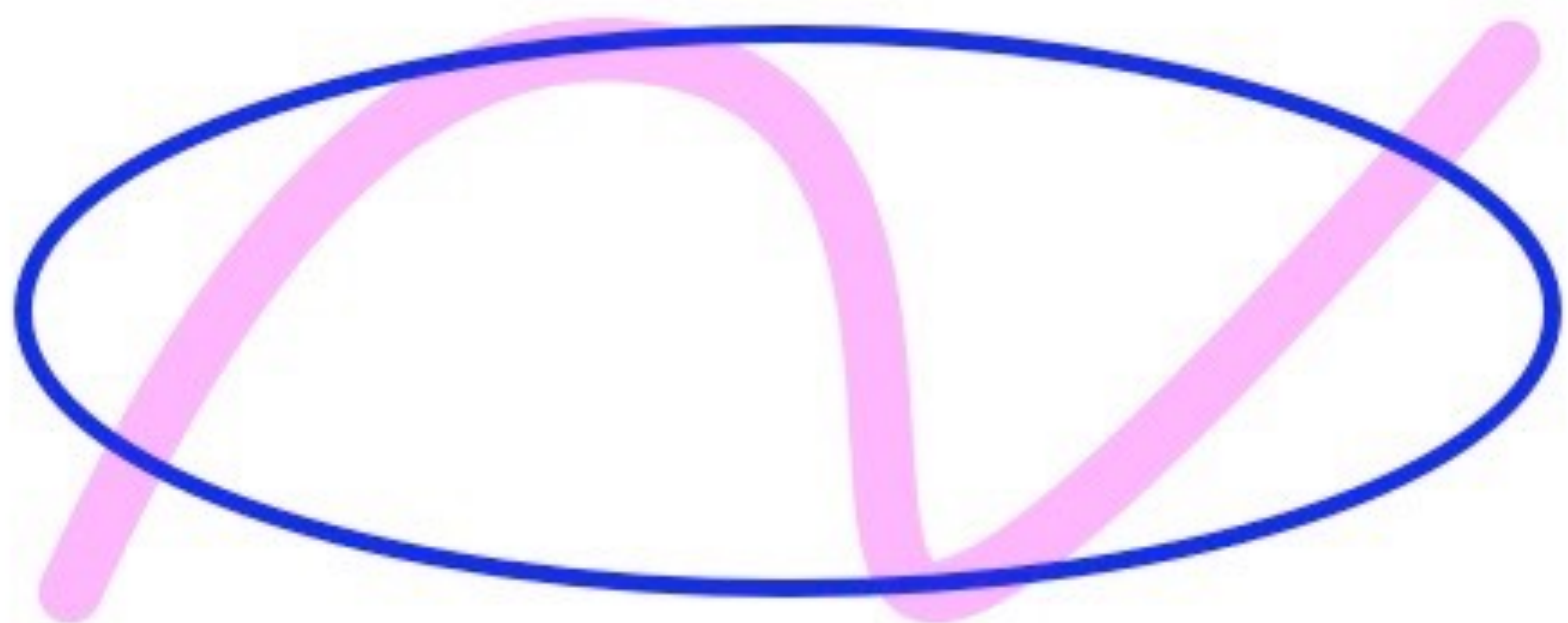


- “optimal” split - orthogonal to largest eigen-vector.
- Split on random direction - almost optimal with constant probability.

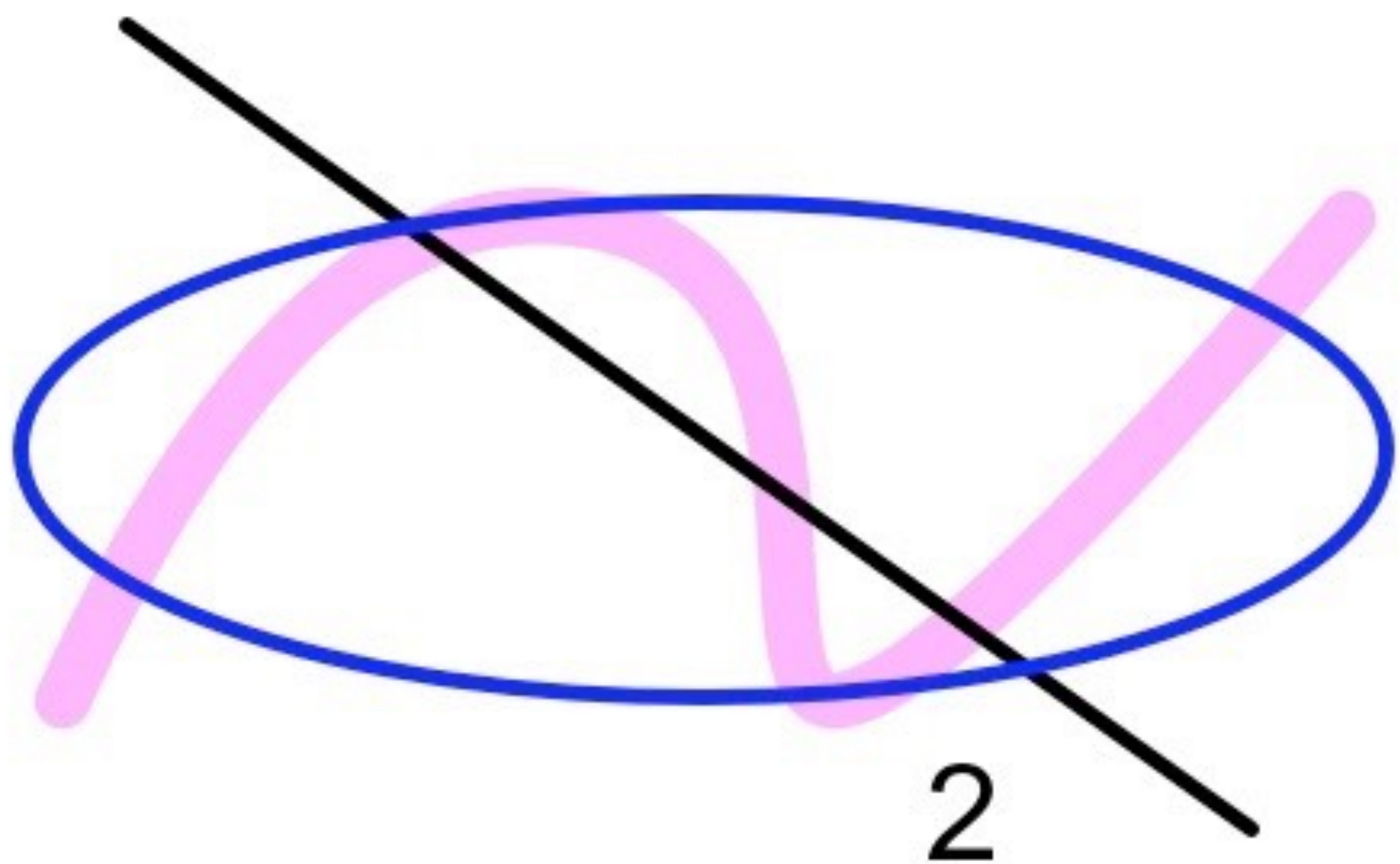
theoretical properties of RP-trees.

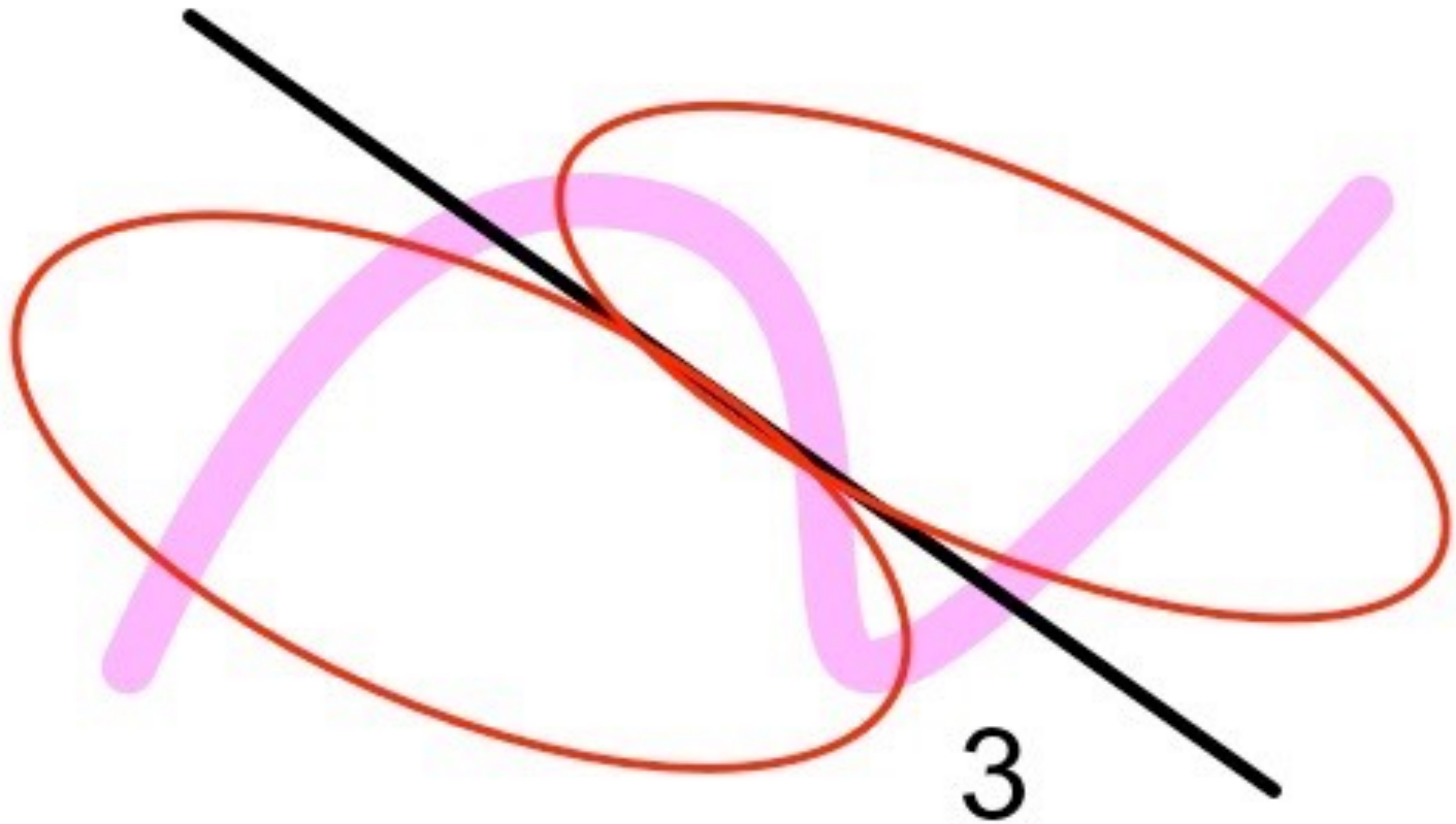
Dasgupta & Freund, STOC08

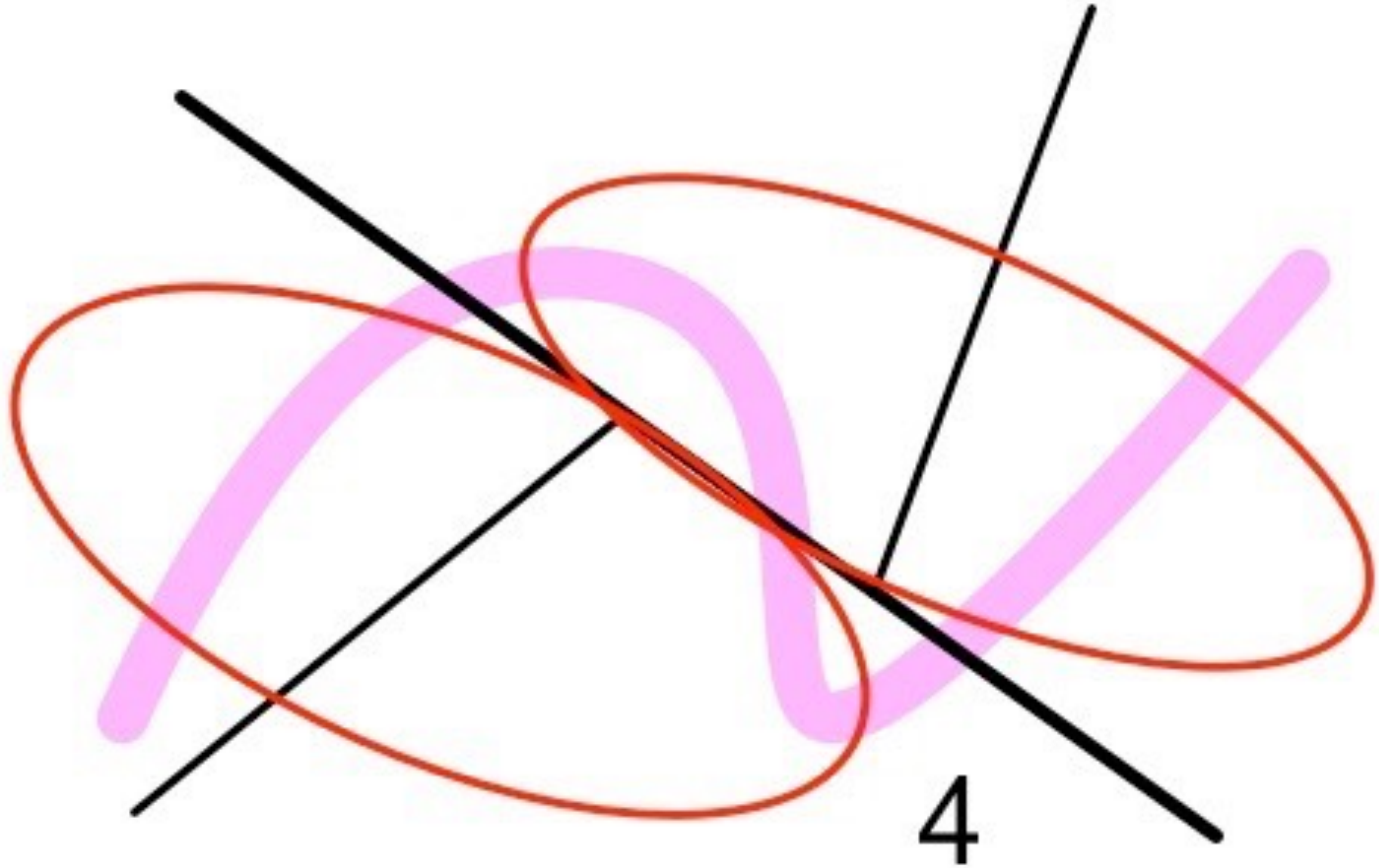
- Space: R^D
- Measure of progress: average cell diameter
- Tree-structured VQ: average diameter halved every D tree levels
- Data of intrinsic dimension d<<D
- RP-tree: average diameter halved every d tree levels (with constant probability)

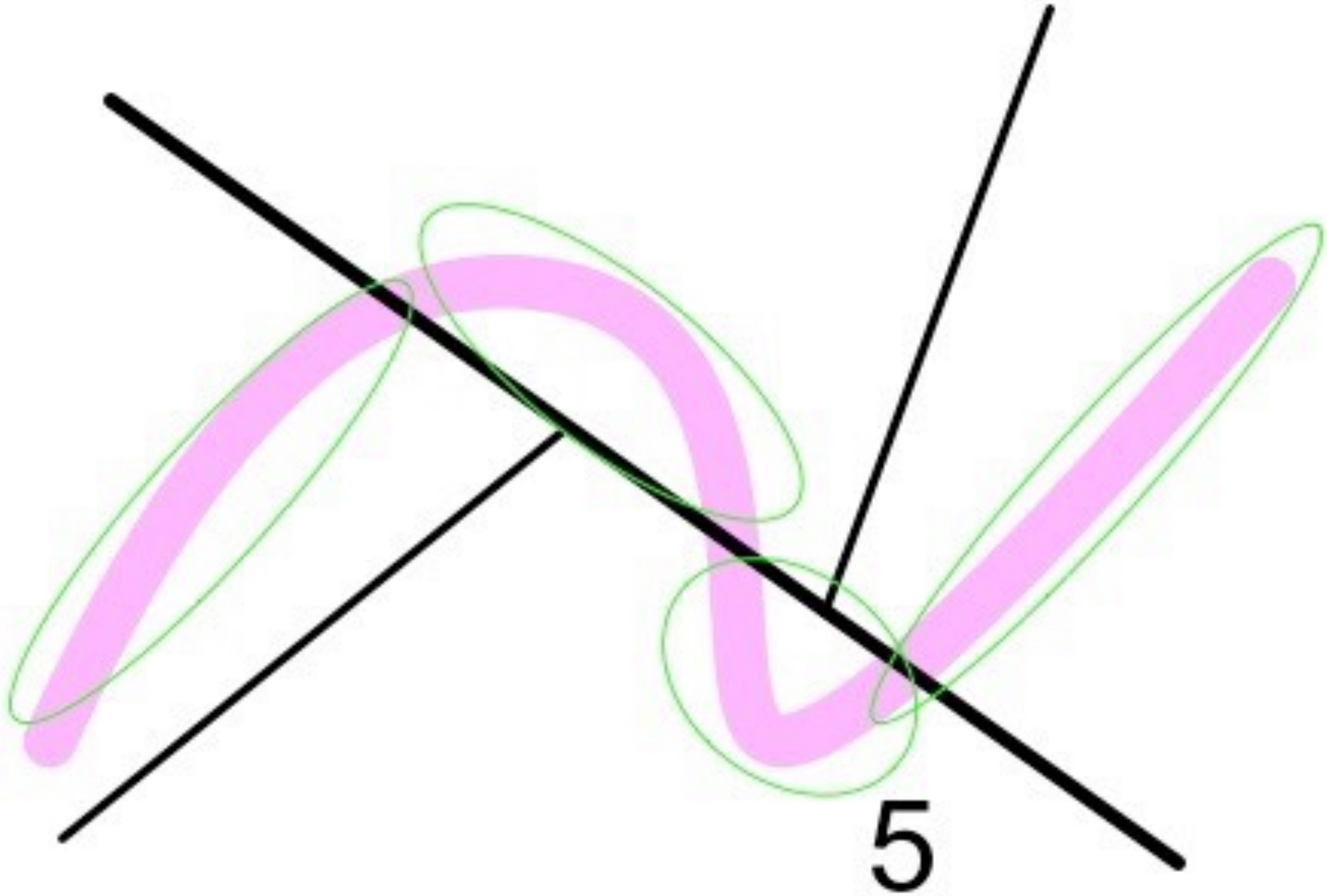


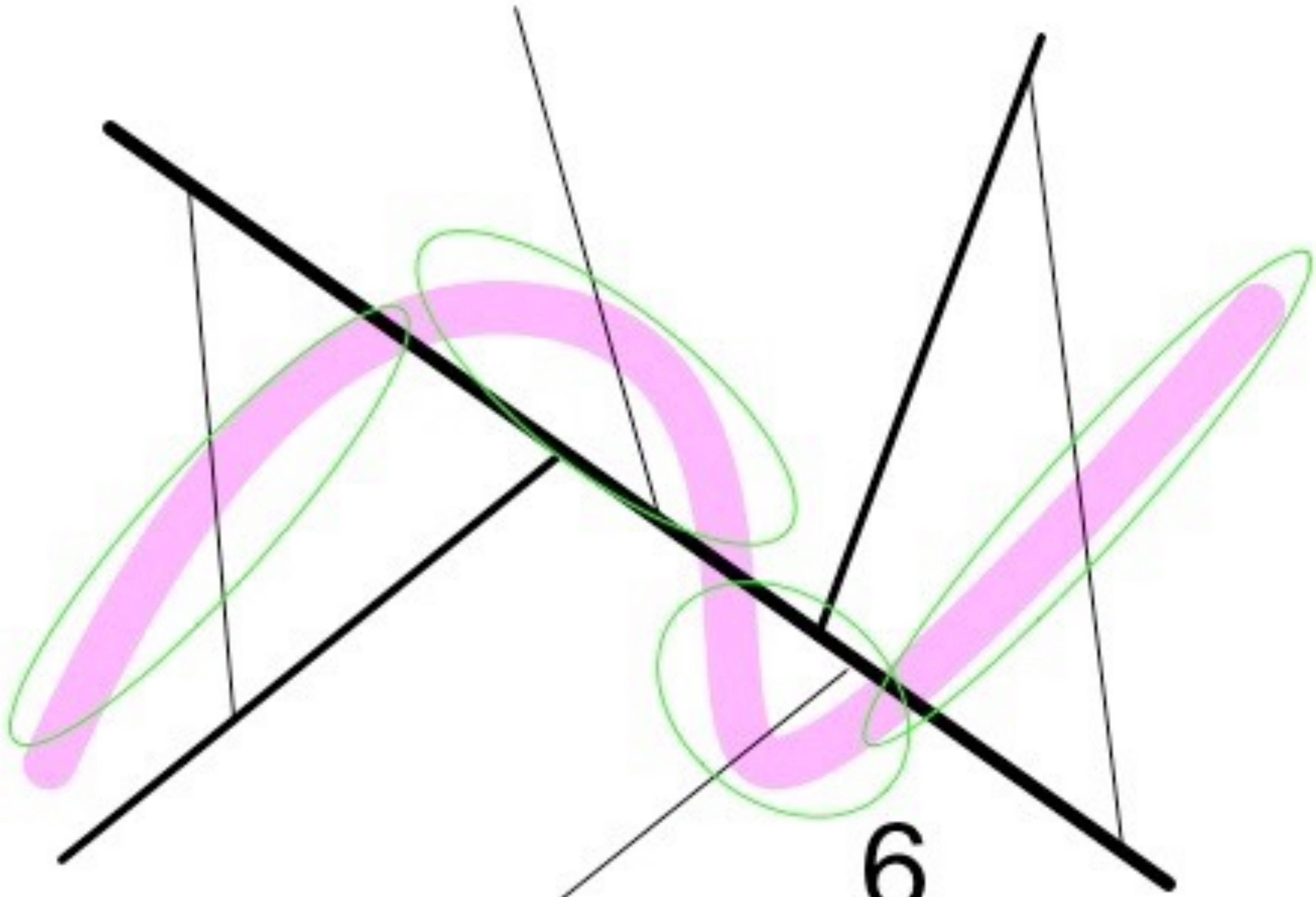
1


2


3

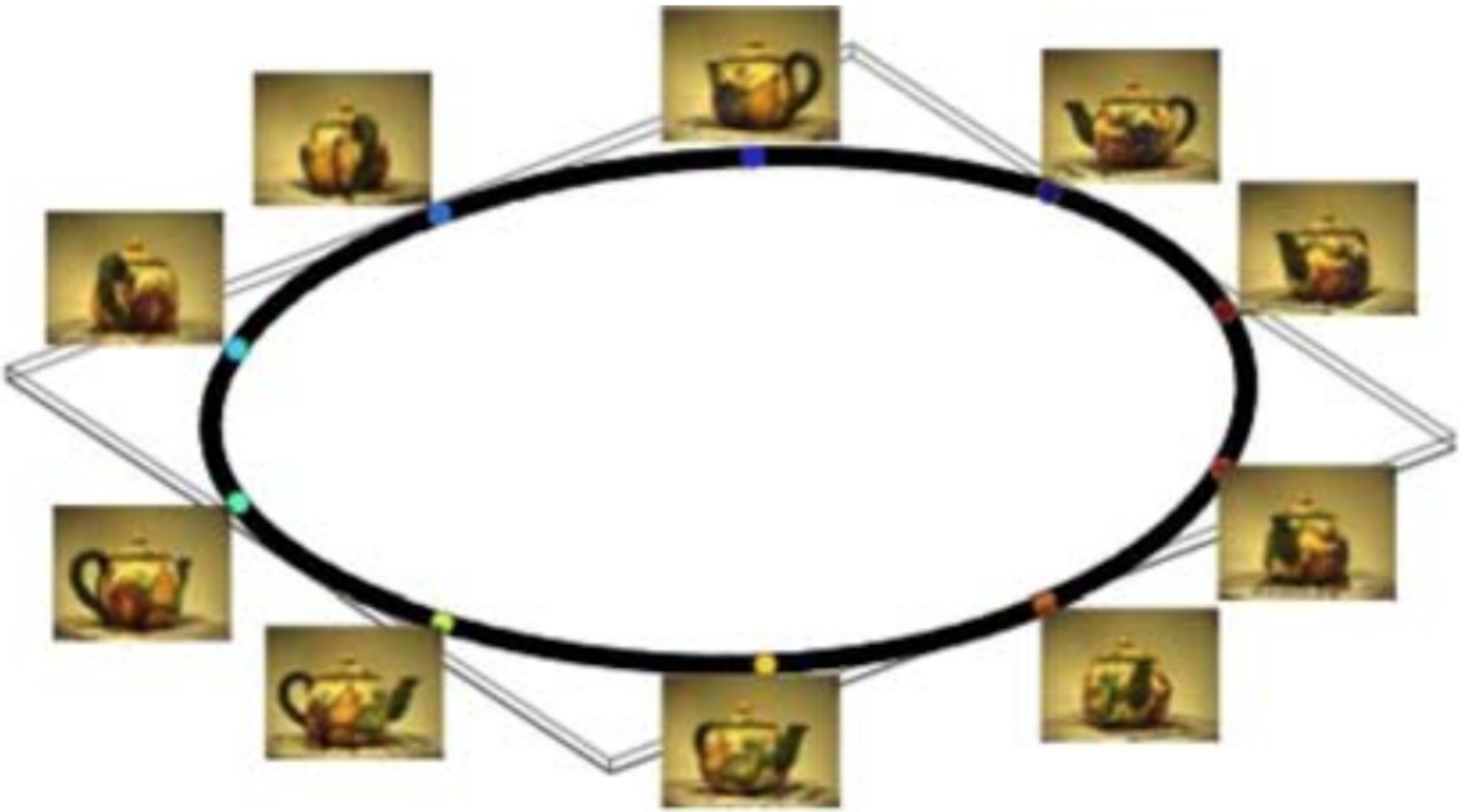

4


5



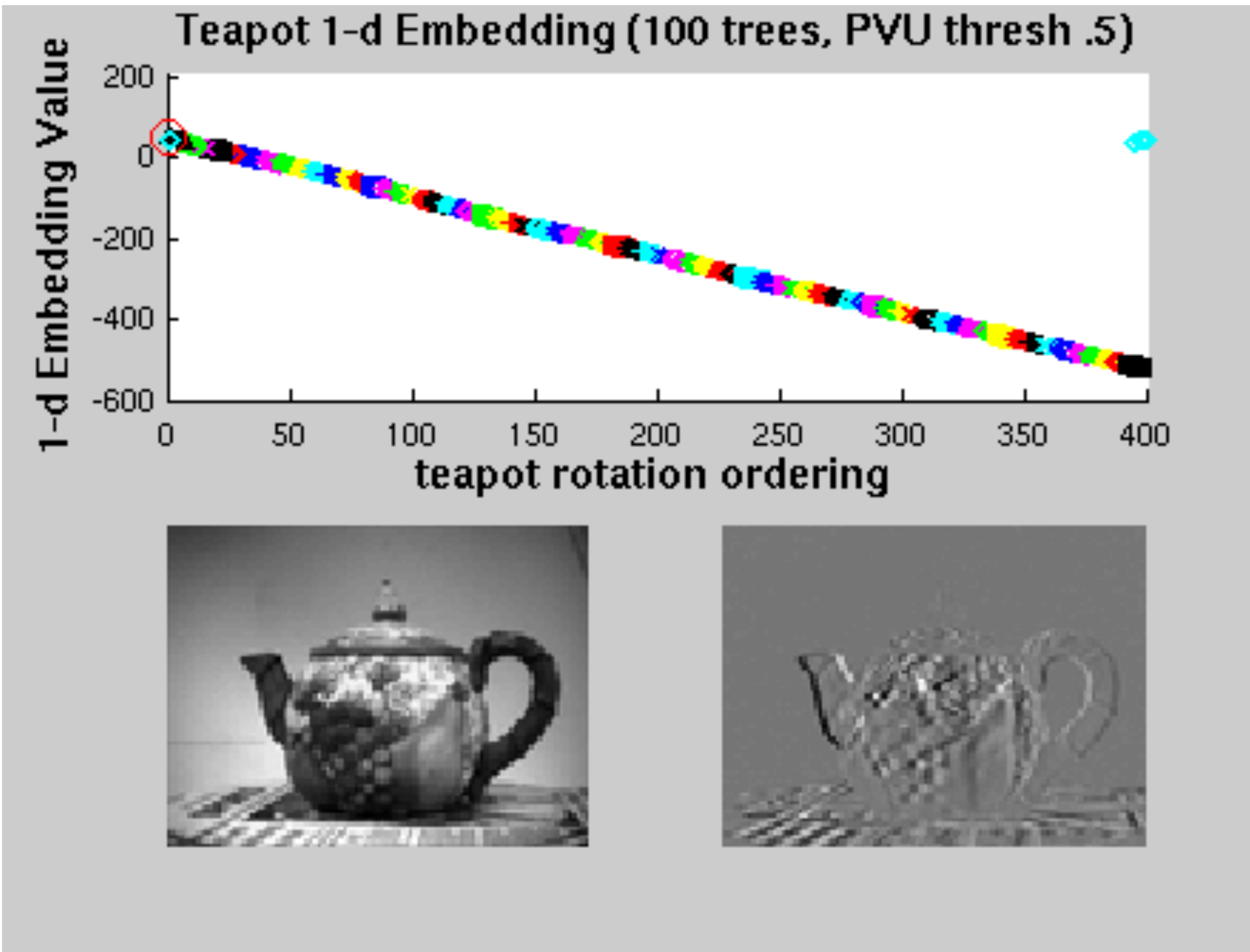
6

The turning tea-pot



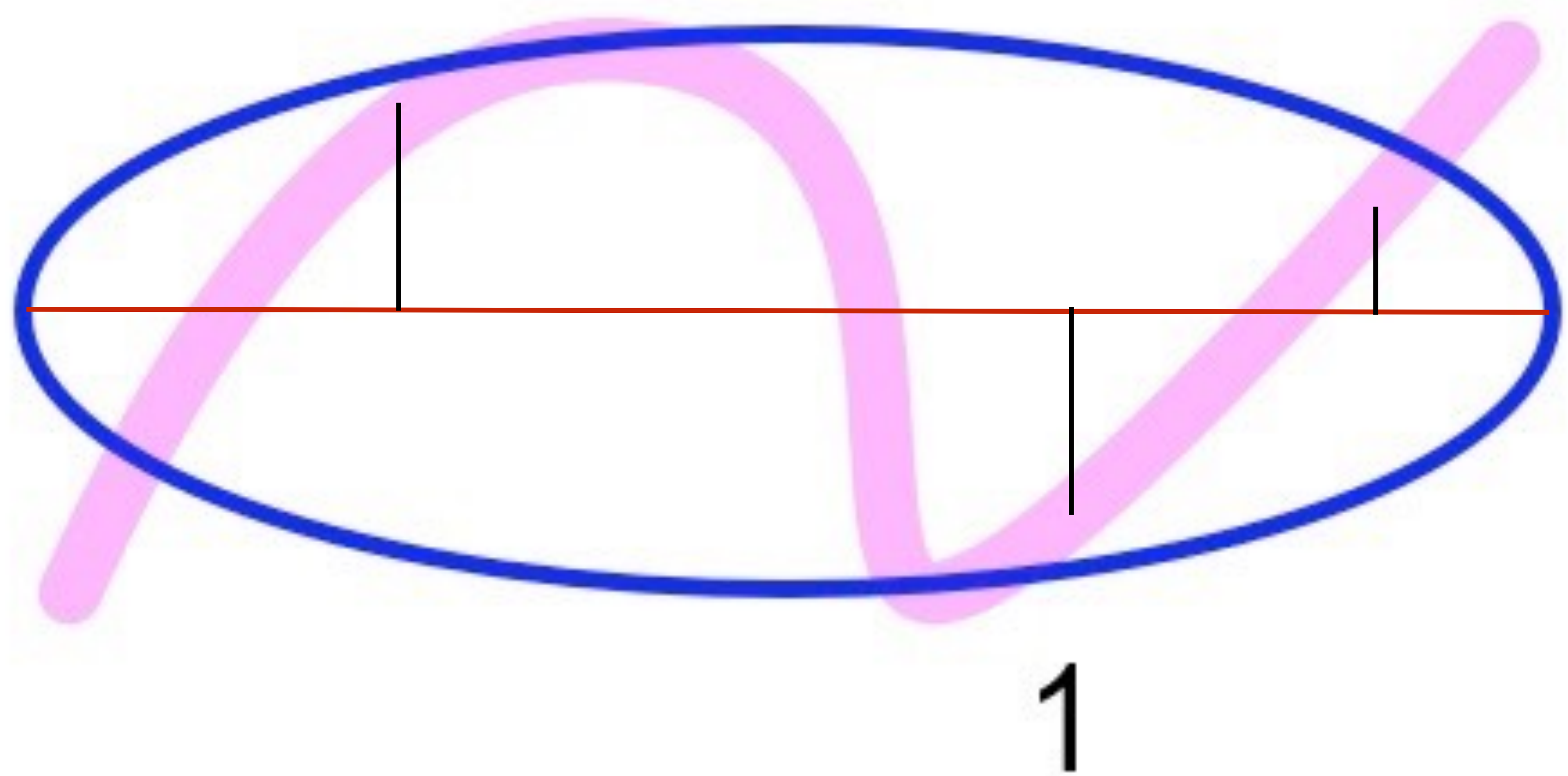
Charting turning teapot manifold

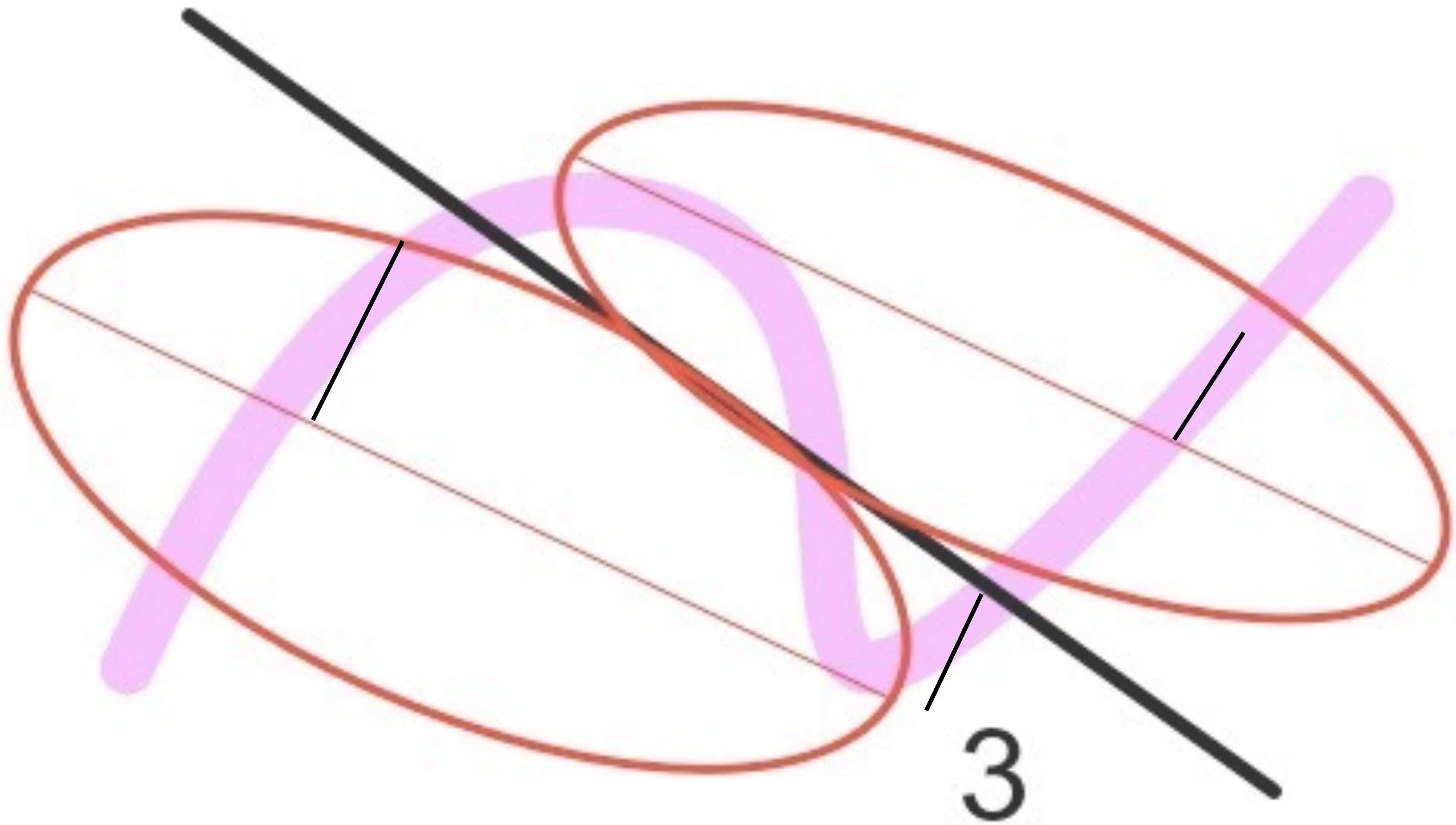
Problem: put an unordered set of images in order of rotational angle.

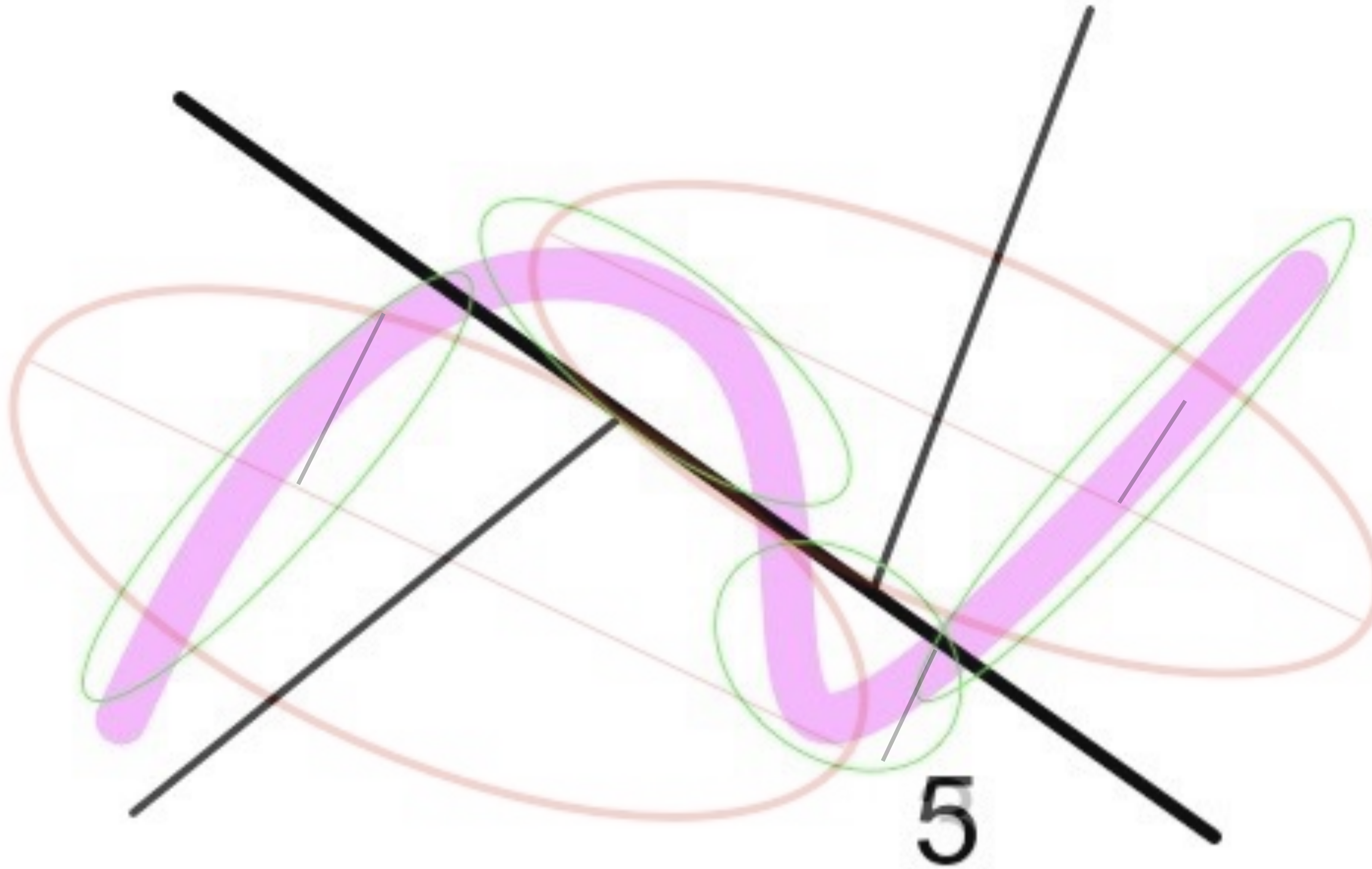


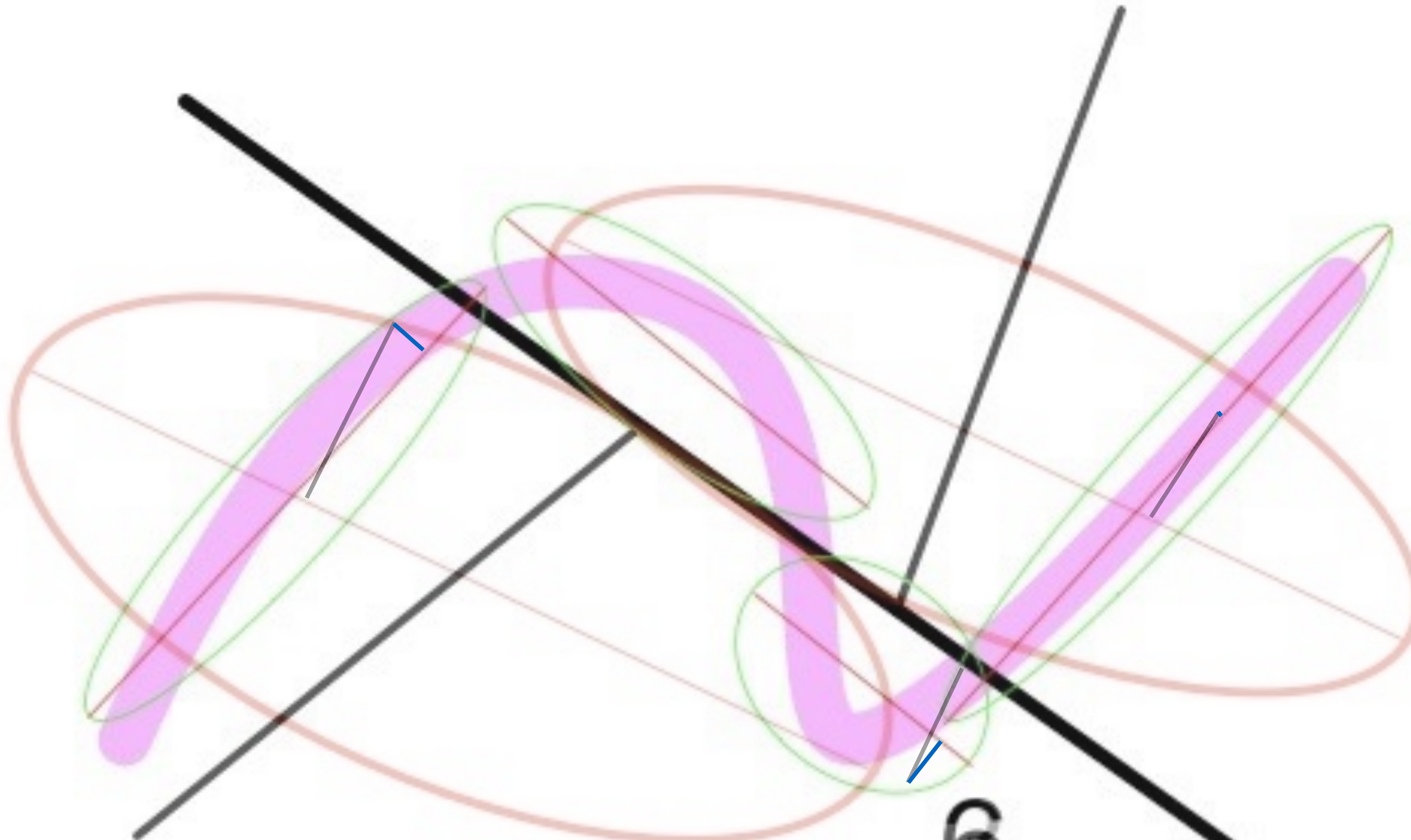
Using RP-trees to represent high-dimensional data

- Goal: map each data point to a localized PCA projection.
- Identify the sufficiently linear pieces. (percent variance explained)
- Combine representations from different nodes along the path.

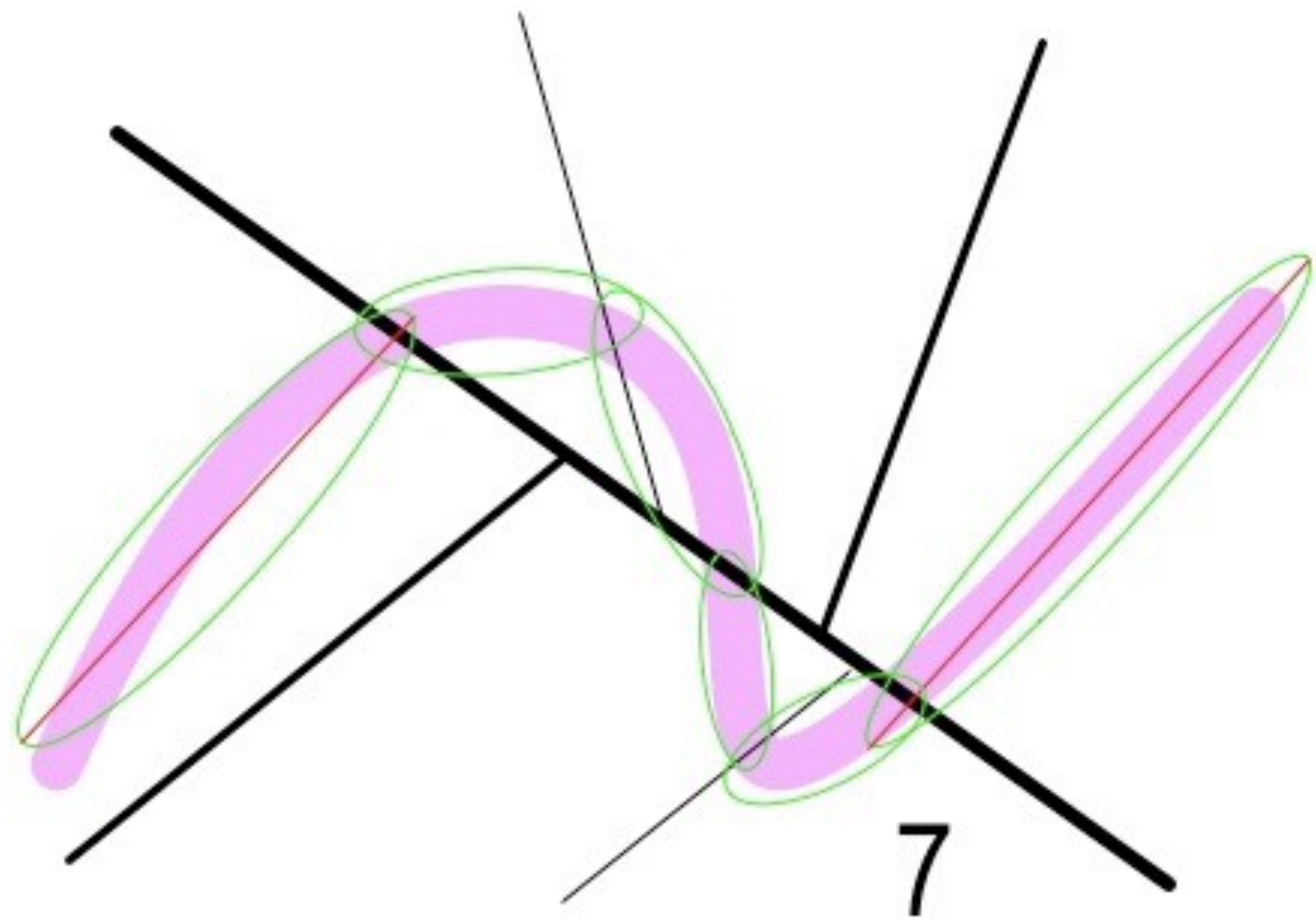

1

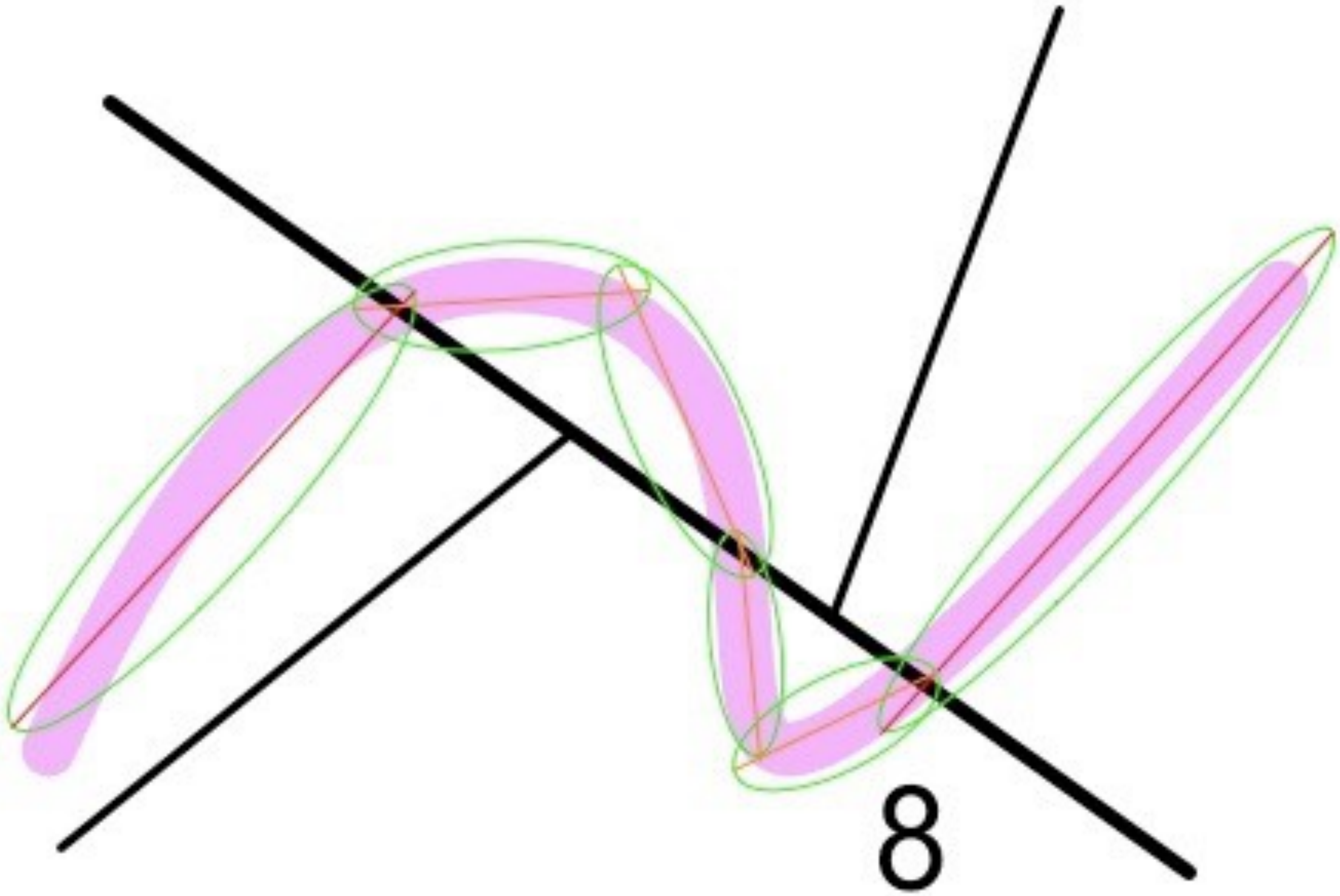

3


5



6

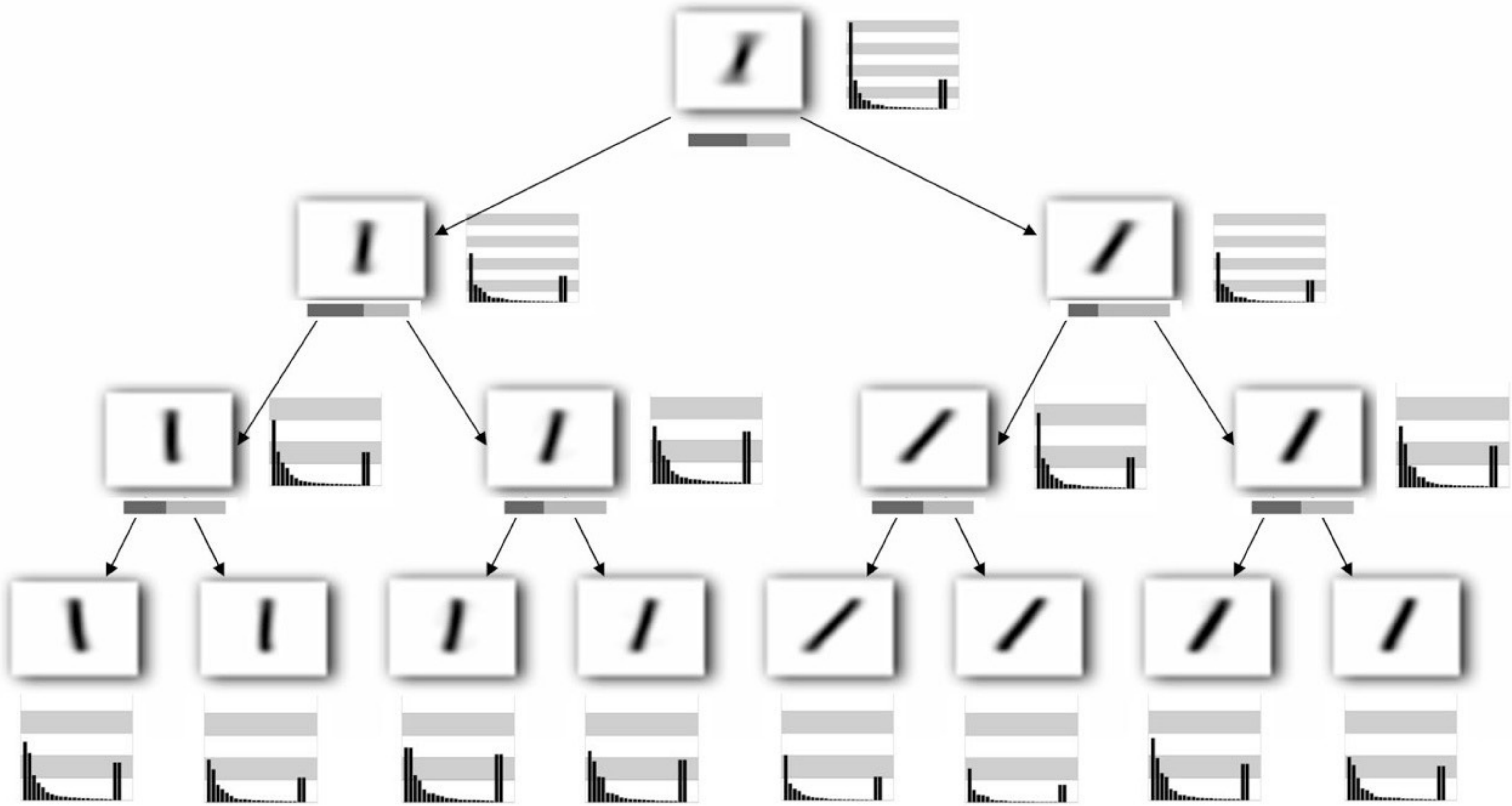

7


8

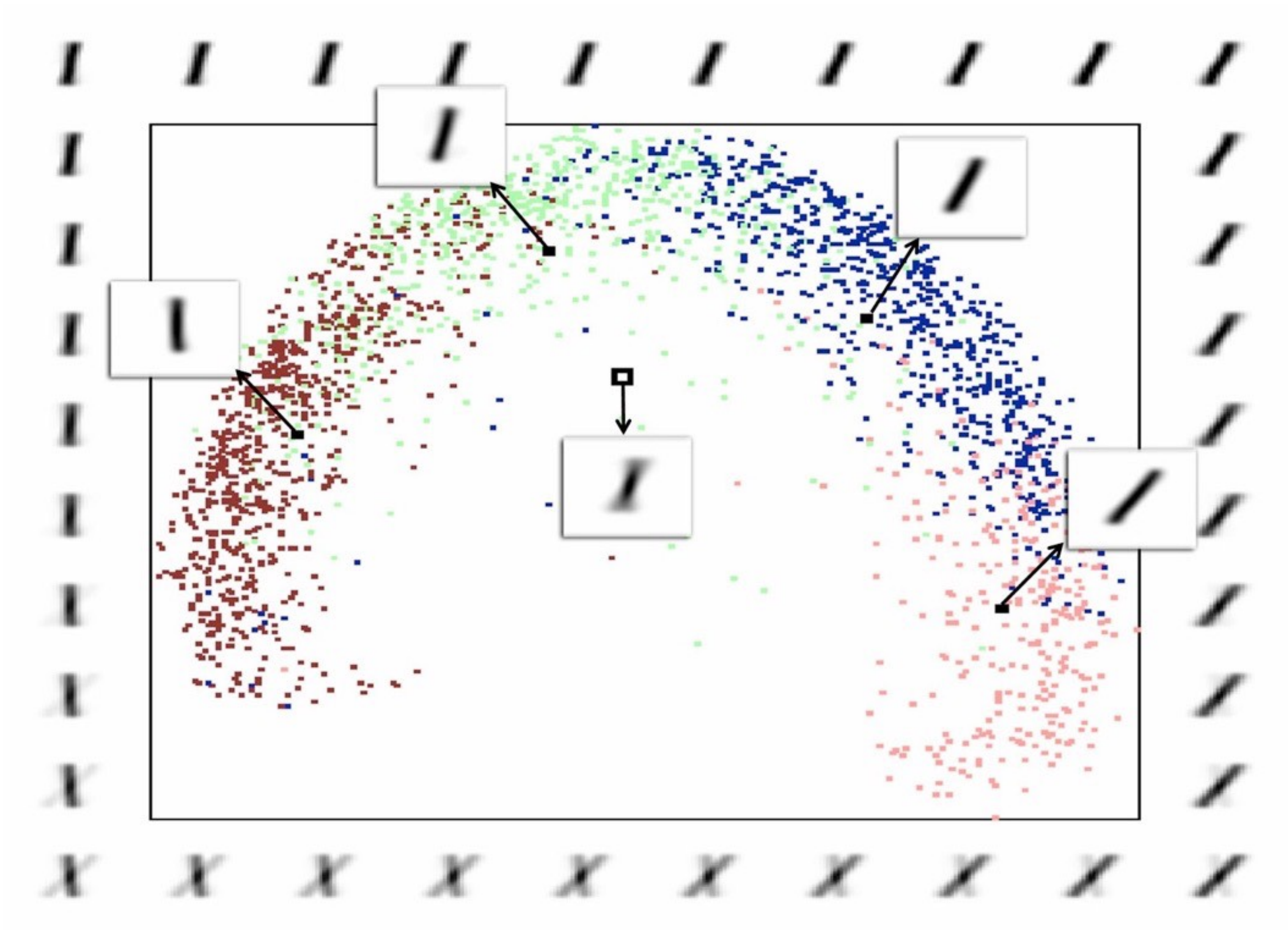
Modeling the manifold of handwritten digits

- Using the MNIST digit dataset.
- We use RP-trees to model one digit at a time.
- Can be a useful pre-processing step for digit recognition.

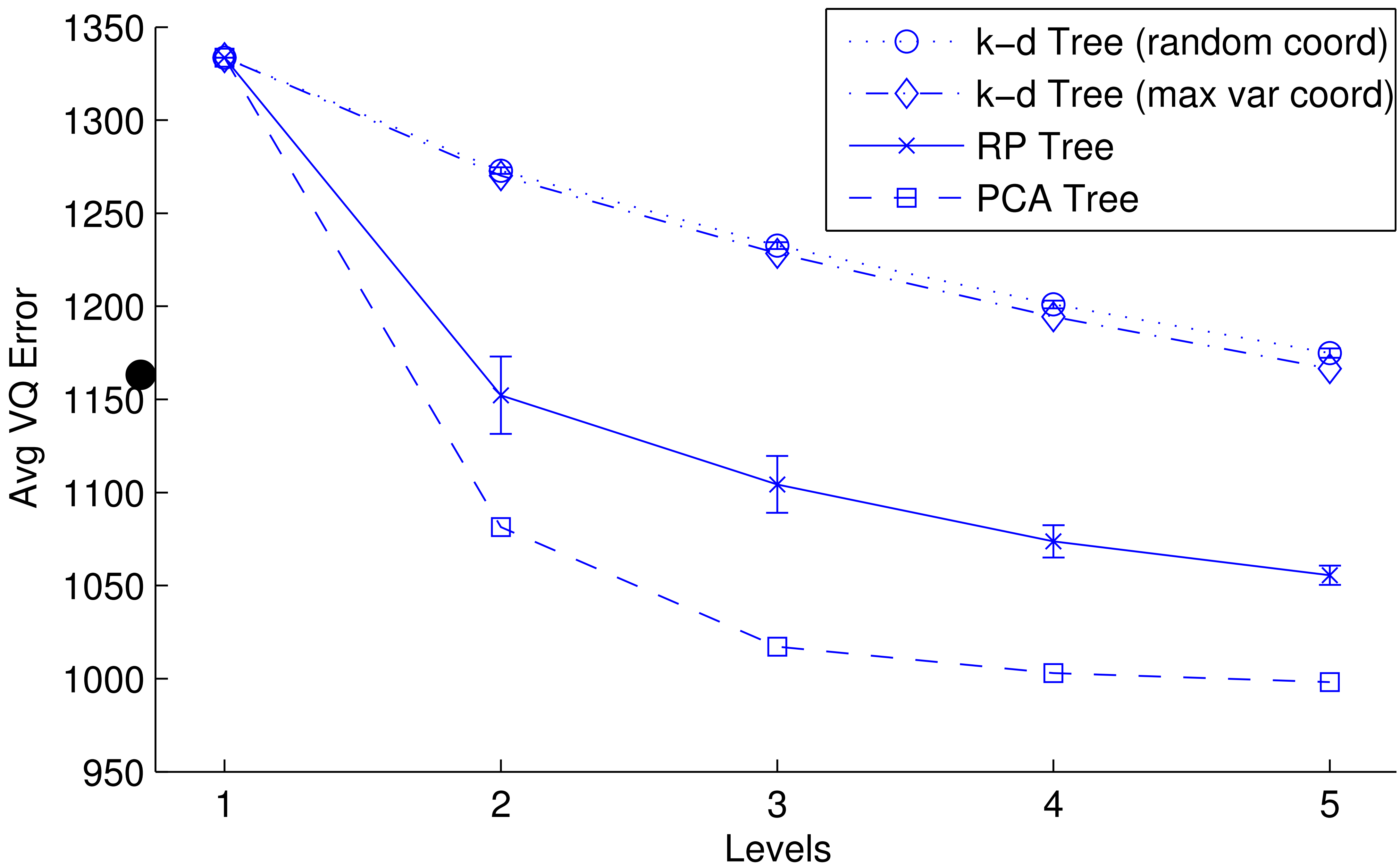
RP-tree for the digit 1



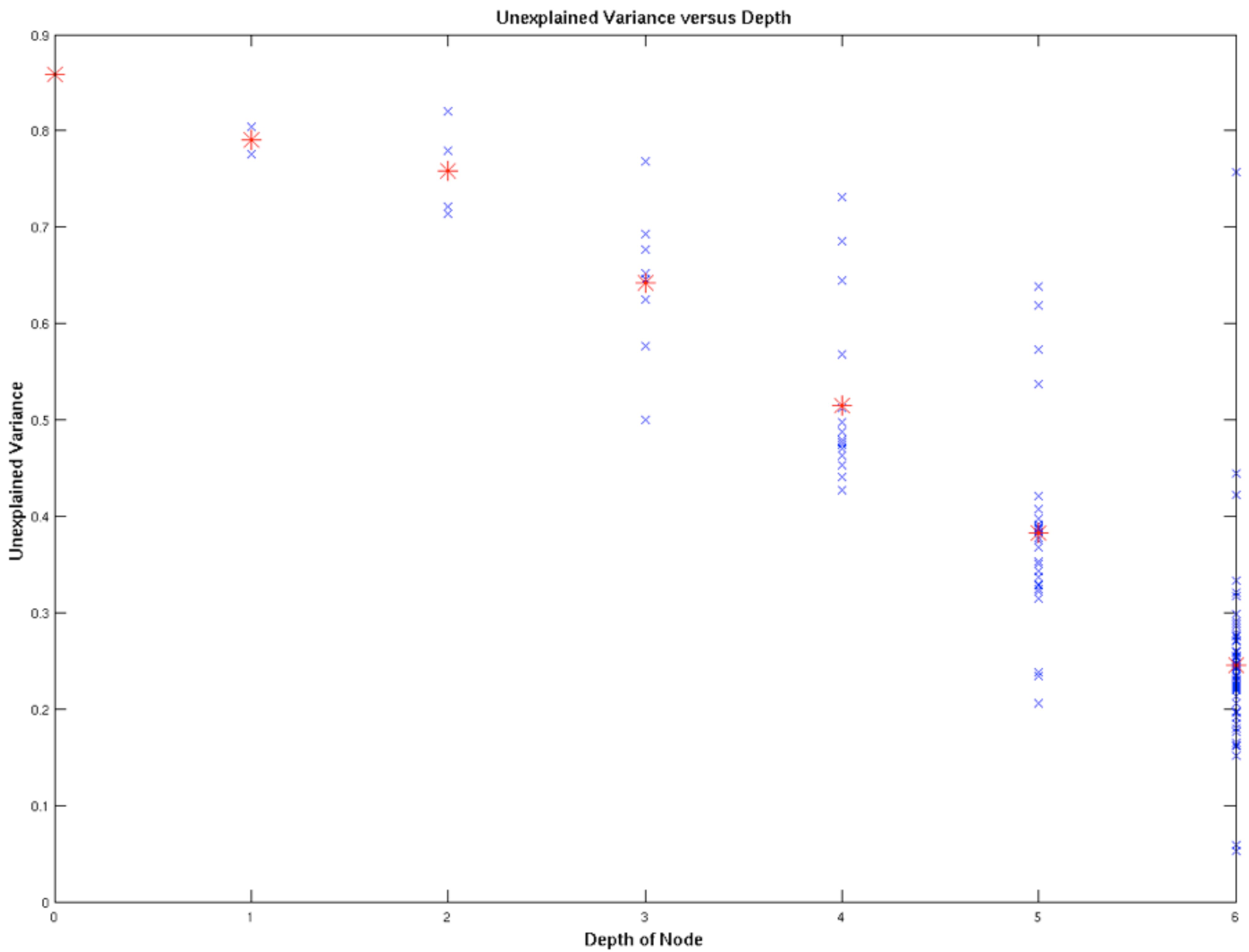
2d distribution of 1



KD-tree vs. RP-tree performance



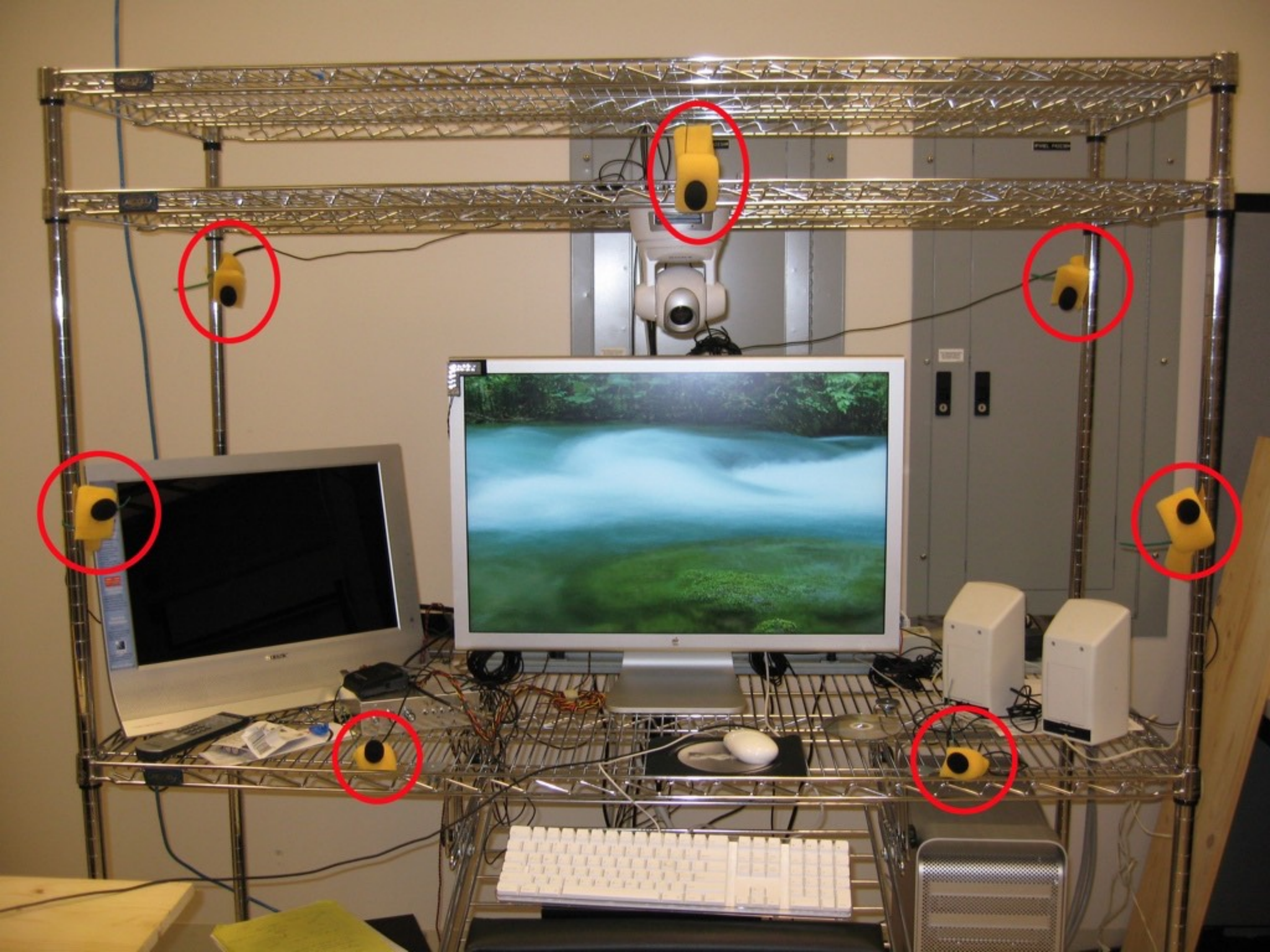
Unexplained variance vs. tree depth



Another Application of RP trees

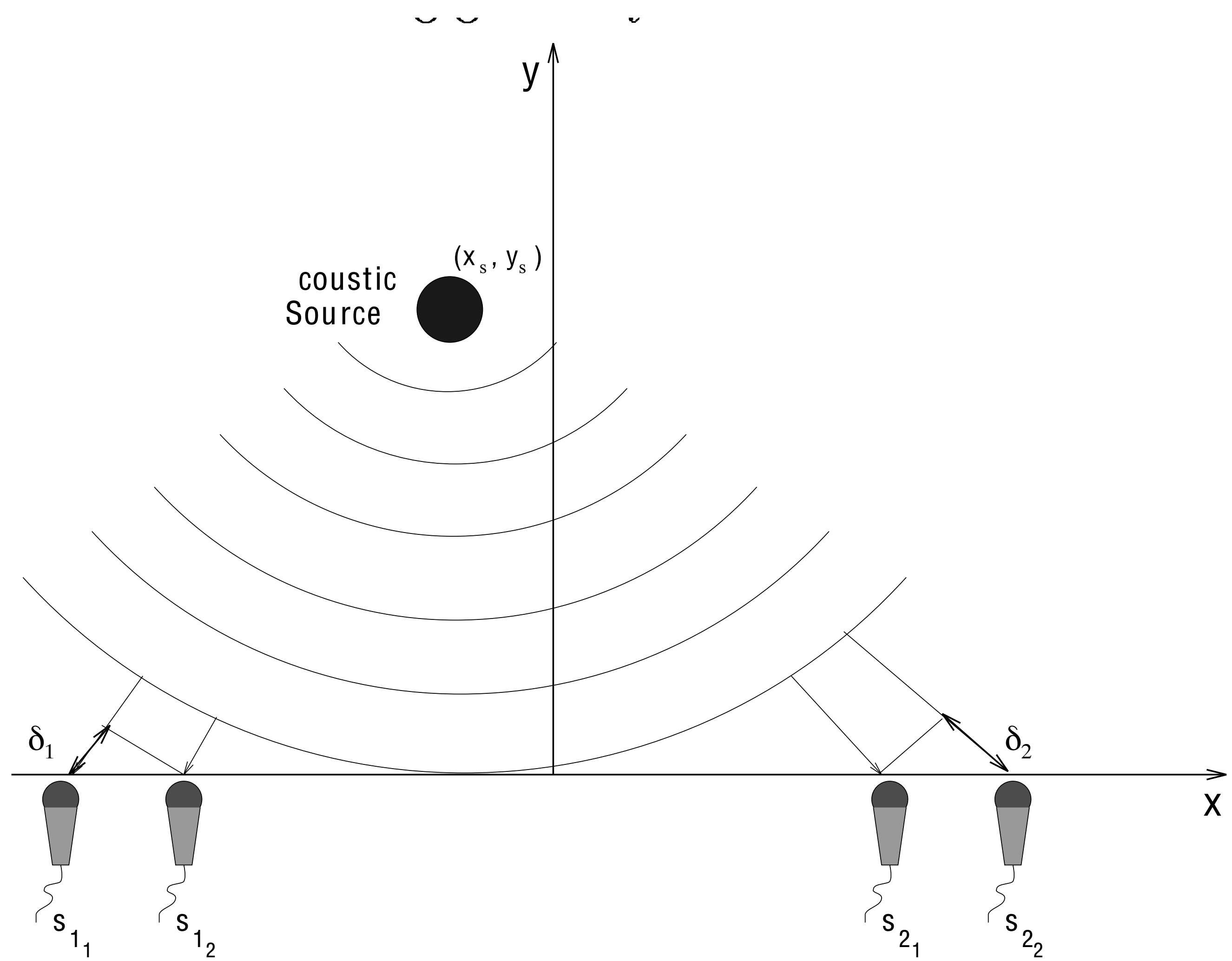
Automatic Cameraman

- Controlling a PTZ camera using audio triangulation
- Learning low dimensional manifolds from sampled data.
- http://www.cse.ucsd.edu/~yfreund/cameraman/index.html



Beamforming basics

- Arrays allow us to *FOCUS* on a source...these techniques are called *beamformers*.
- The signal arrives with a delay $\mathbf{\Delta}_{ij}$ between microphones i and j.



Calibration process

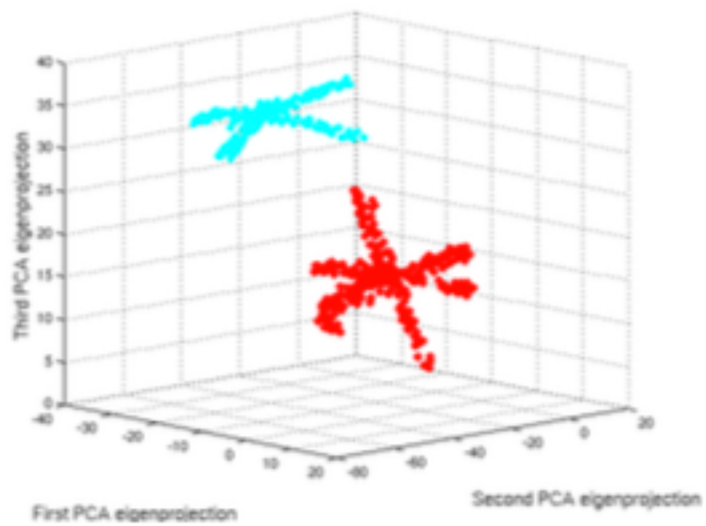
- Goal: map measured delays to pan-tilt direction of camera.
- Training data:
  - High-correlation delay for each microphone pair (21)
  - Camera pan+tilt (2)

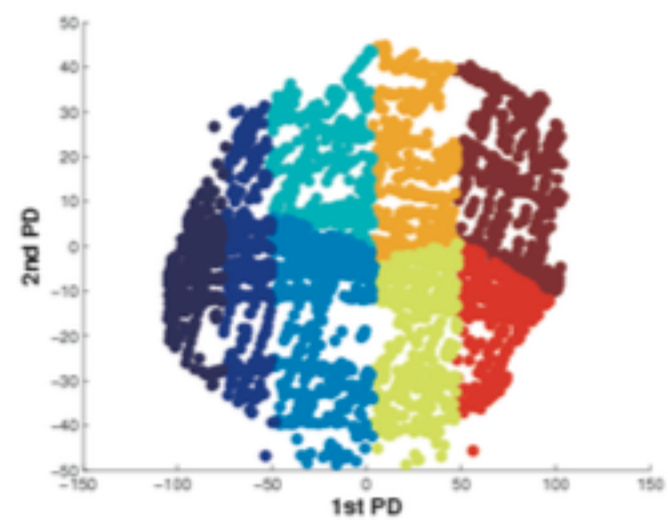
delay 1,2	delay 1,3	delay 2,3	.	.	.	pan	tilt
9±2	35±1	?				77±2	31±2
13±2	30±2	50±20				80±2	33±2

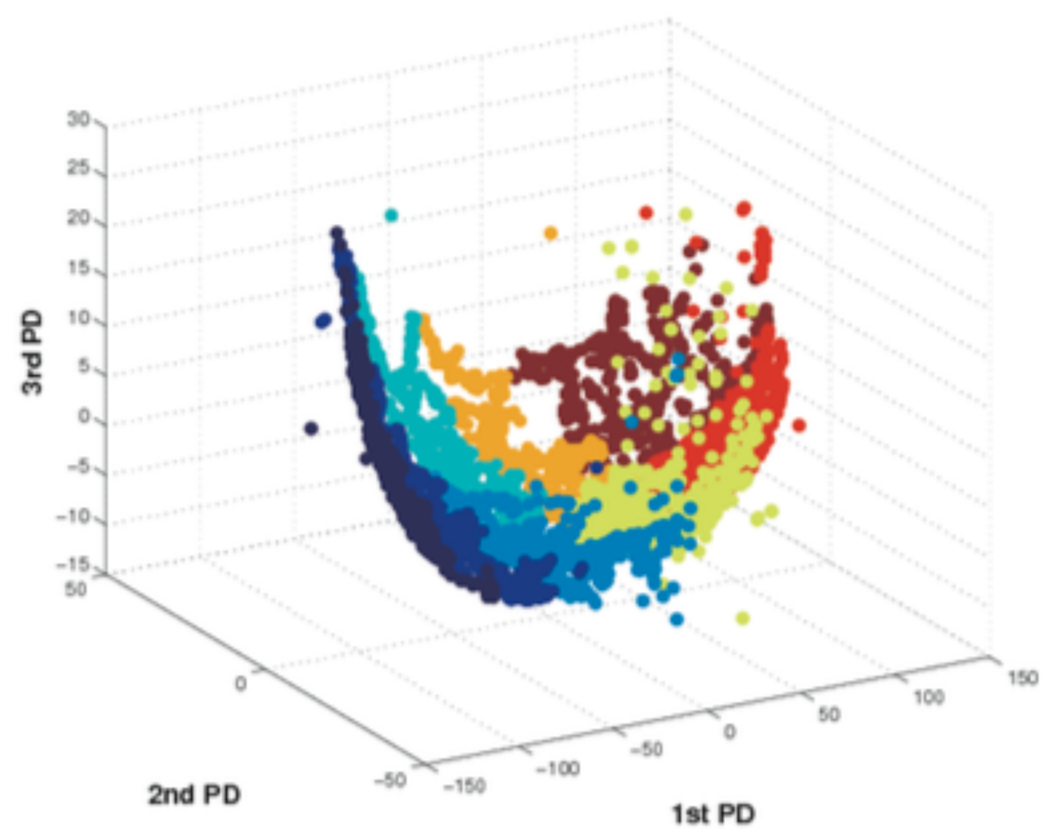
The delay manifold

- 7 microphones
- 21 microphone pairs
- 2 camera coordinates: pan,tilt
- Together: 23 dimensional space
- Data lies (close to) a smooth 3 dimensional manifold.
- If we can learn manifold from data
  we can map delay vector to (pan,tilt)

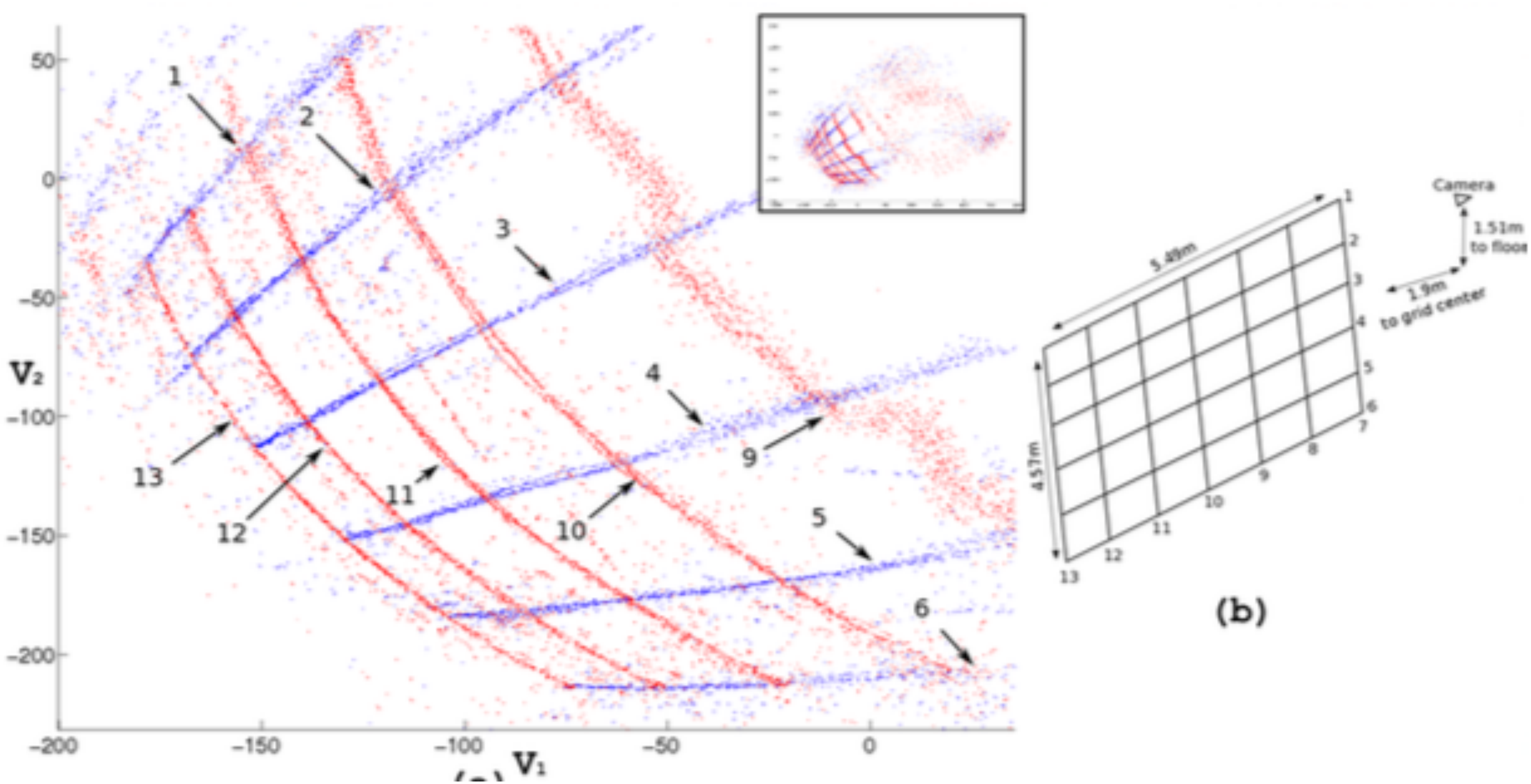
Delay manifold for laboratory setup







Mapping of Hallway using top 2 eigenvectors For one node of RP-tree.



Summary I

- Dimensionality reduction / Lossy compression are methods for reducing data without losing much of the information.
- PCA is the most popular method, but it can only find linear mappings. We say that PCA find a k-dimensional representation if >X% of the variance is explained by the top k eigen-vectors. Equivalently, the top k eigen-values sum to >X% of the total variance.
- PCA dimension is a global concept.

An old video

- https://www.youtube.com/watch?v=rrOy6LpL940

Summary 2

- Vector quantization is generic but it only finds a partition,
  not a mapping into new coordinates.
- Scaling dimension / Haussdorf dimension / Metric dimension: characterizes the rate of increase in the number of partition as the radius/diameter of the parts decreases.

$$\log n = \log C + d \log \frac{1}{\epsilon}$$

$$\log \frac{n_2}{n_1} = d \log \frac{\epsilon_1}{\epsilon_2}$$

$$d = \frac{\log \frac{n_2}{n_1}}{\log \frac{\epsilon_1}{\epsilon_2}}$$

Summary 3

- Low dimensional manifold: a subset of the space that is defined by a set of constraints.
- Not a statistical concept
- The local dimension of the manifold is defined by the tangent hyperplane at that point.
- Dimension is an infinitesimal concept.

Local covariance dimension

- A local but not an infinitesimal concept.
- Perform PCA on the data that is in a ball.
- RP-Trees - a space-partitioning data structure that performs well (as opposed to KD-trees) when the intrinsic dimension is low.

Future direction
learning piecewise-linear control

Tedrake et al. "Learning to walk in 20 minutes" Science 2004