THE CITY UNIVERSITY OF NEW YORK
College of Staten Island

Introduction to Topological Data Analysis

Mikael Vejdemo-Johansson

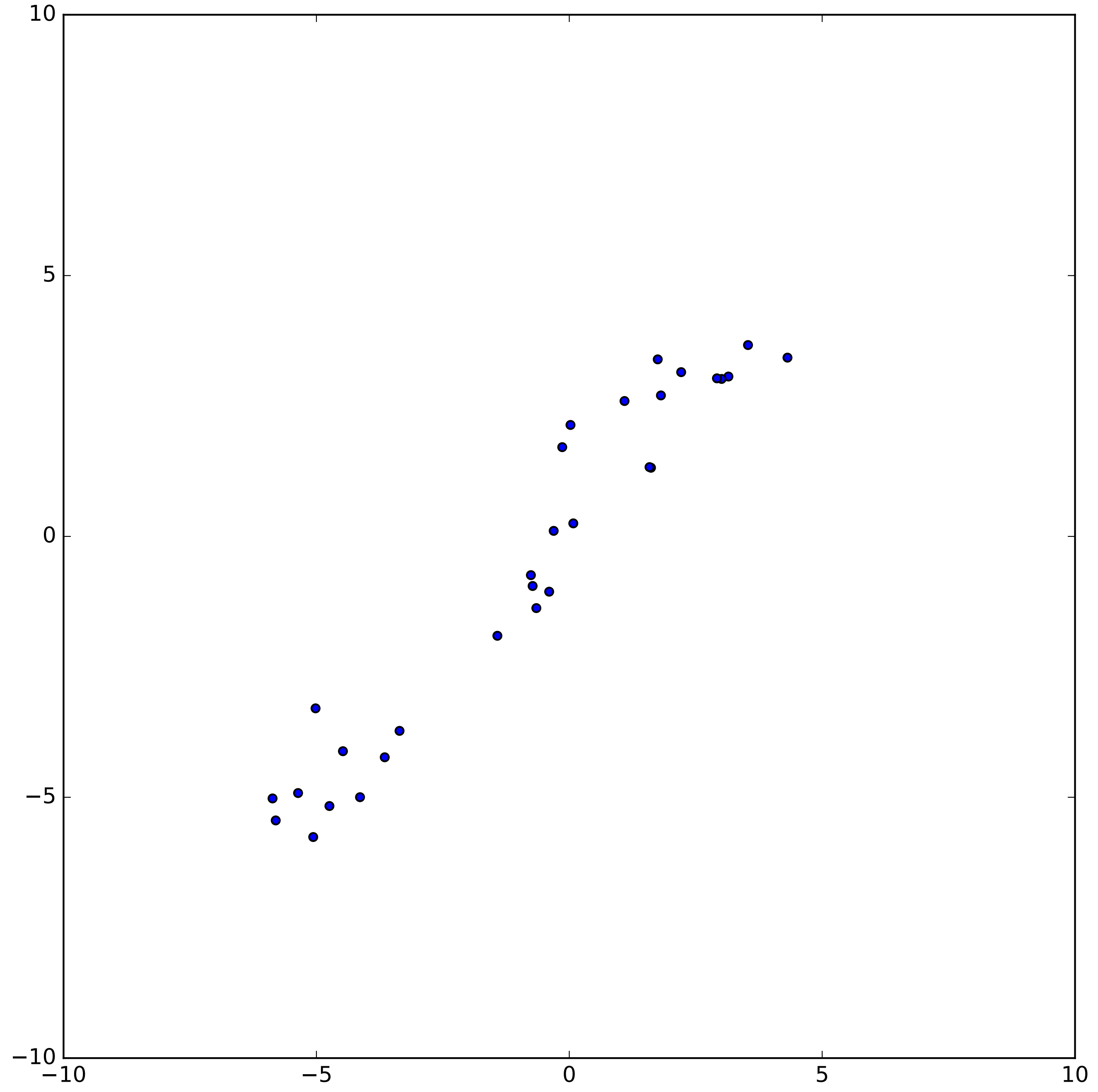
Outline

- Data has shape
- Homology: linear algebra measures shape
- Persistence: squinting with mathematics
- Cohomology
- Applications

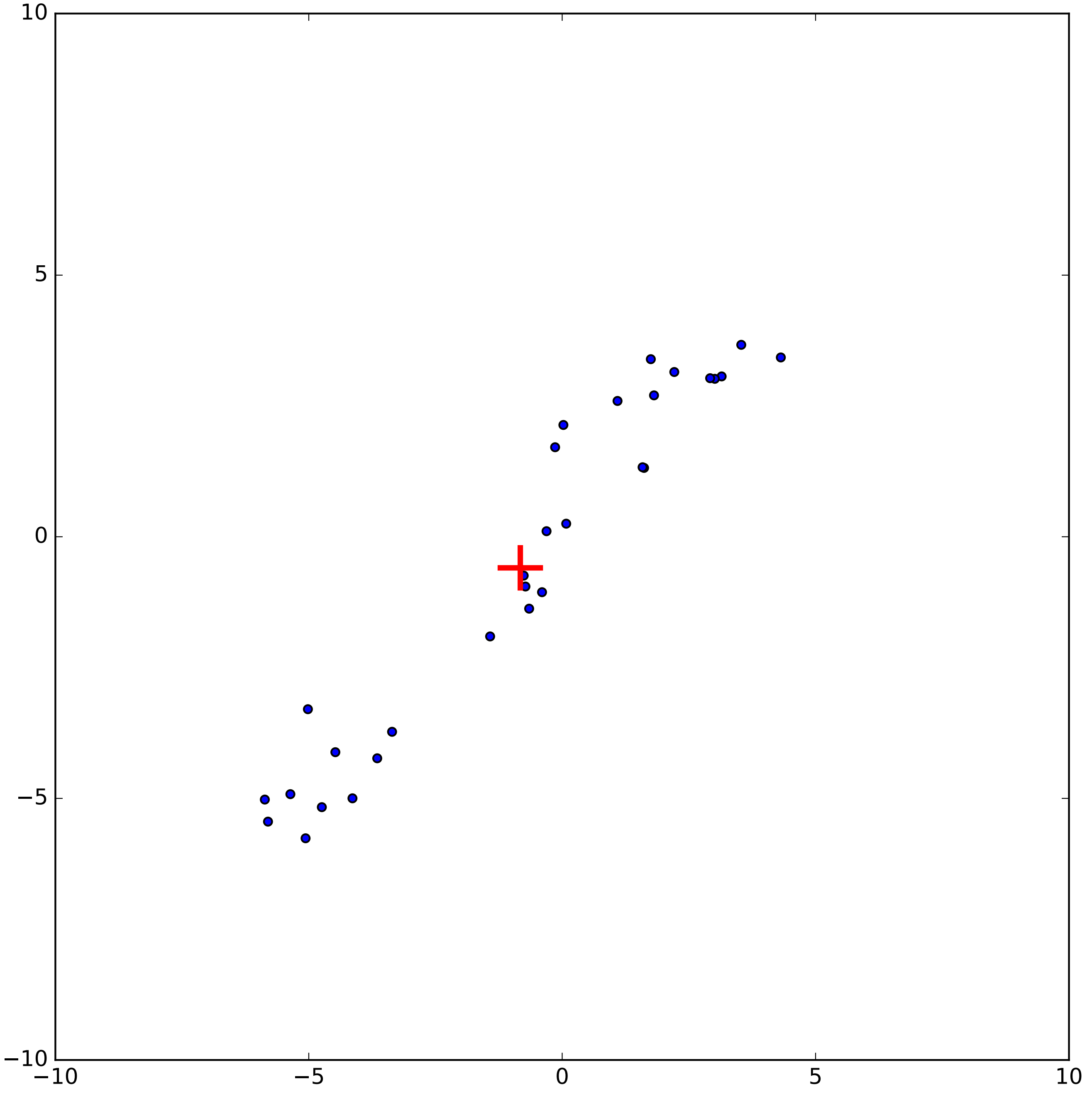
Outline

- **Data has shape**
- Homology: linear algebra measures shape
- Persistence: squinting with mathematics
- Cohomology
- Applications

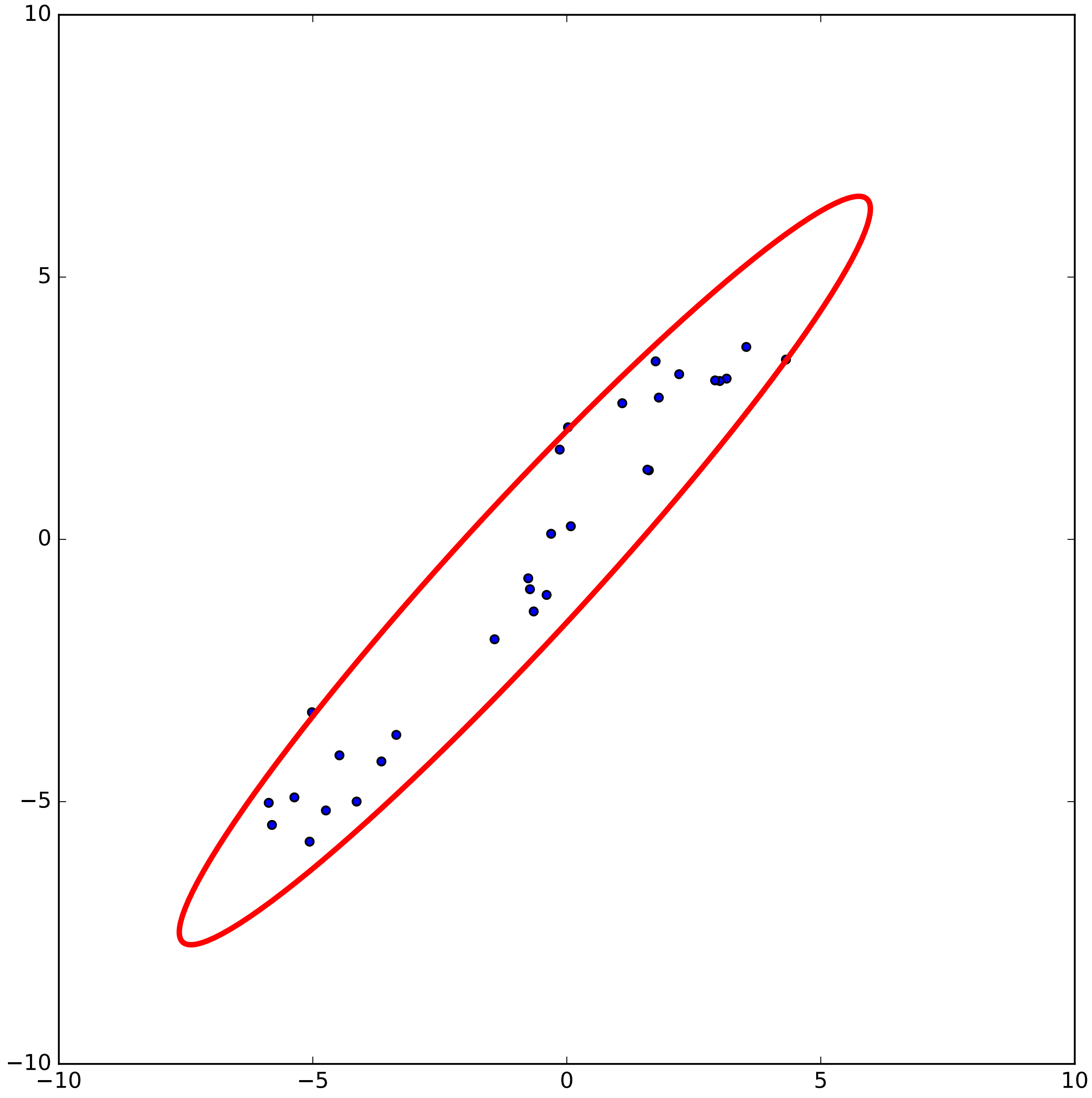
Data has shape



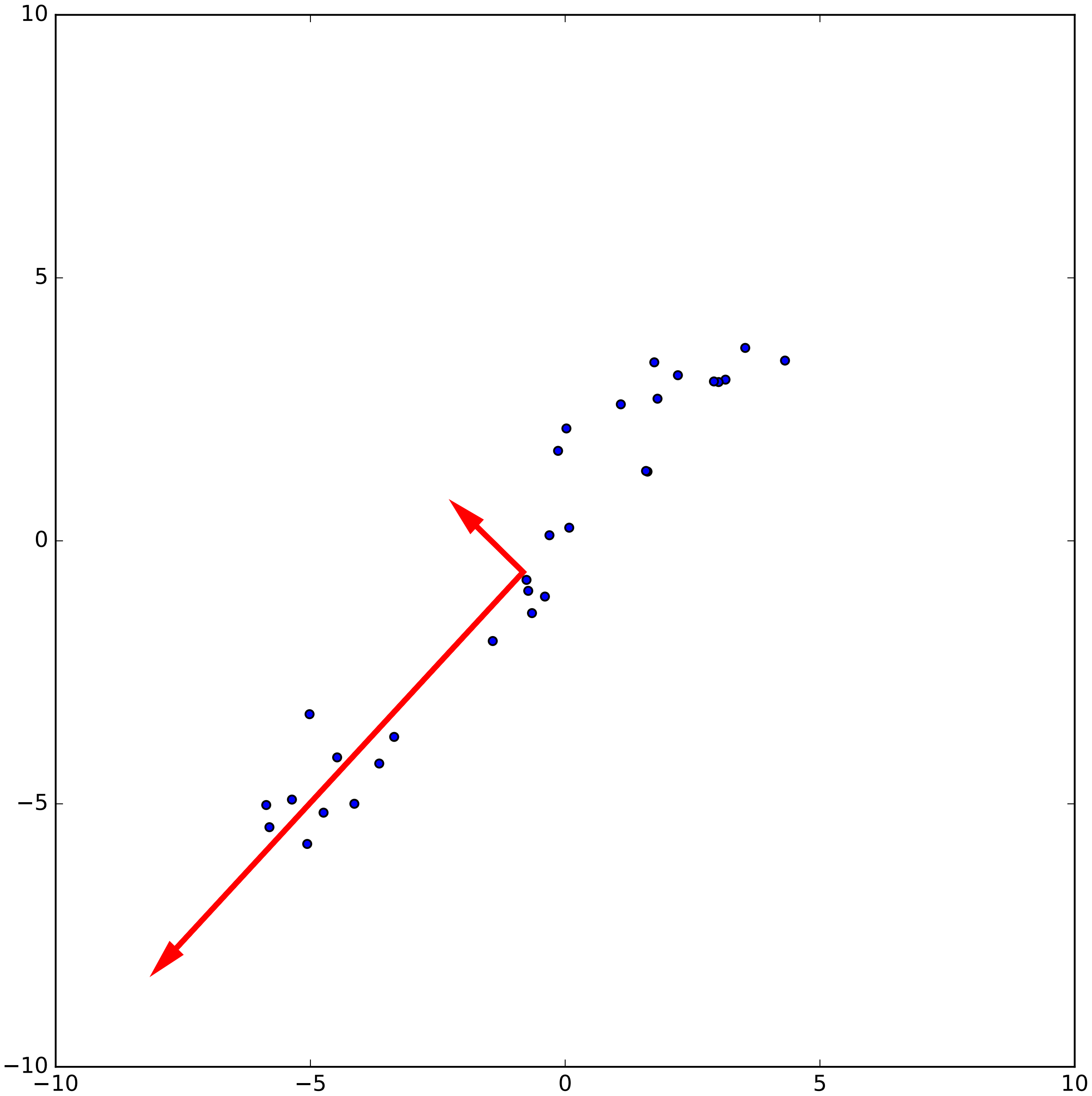
Data has shape



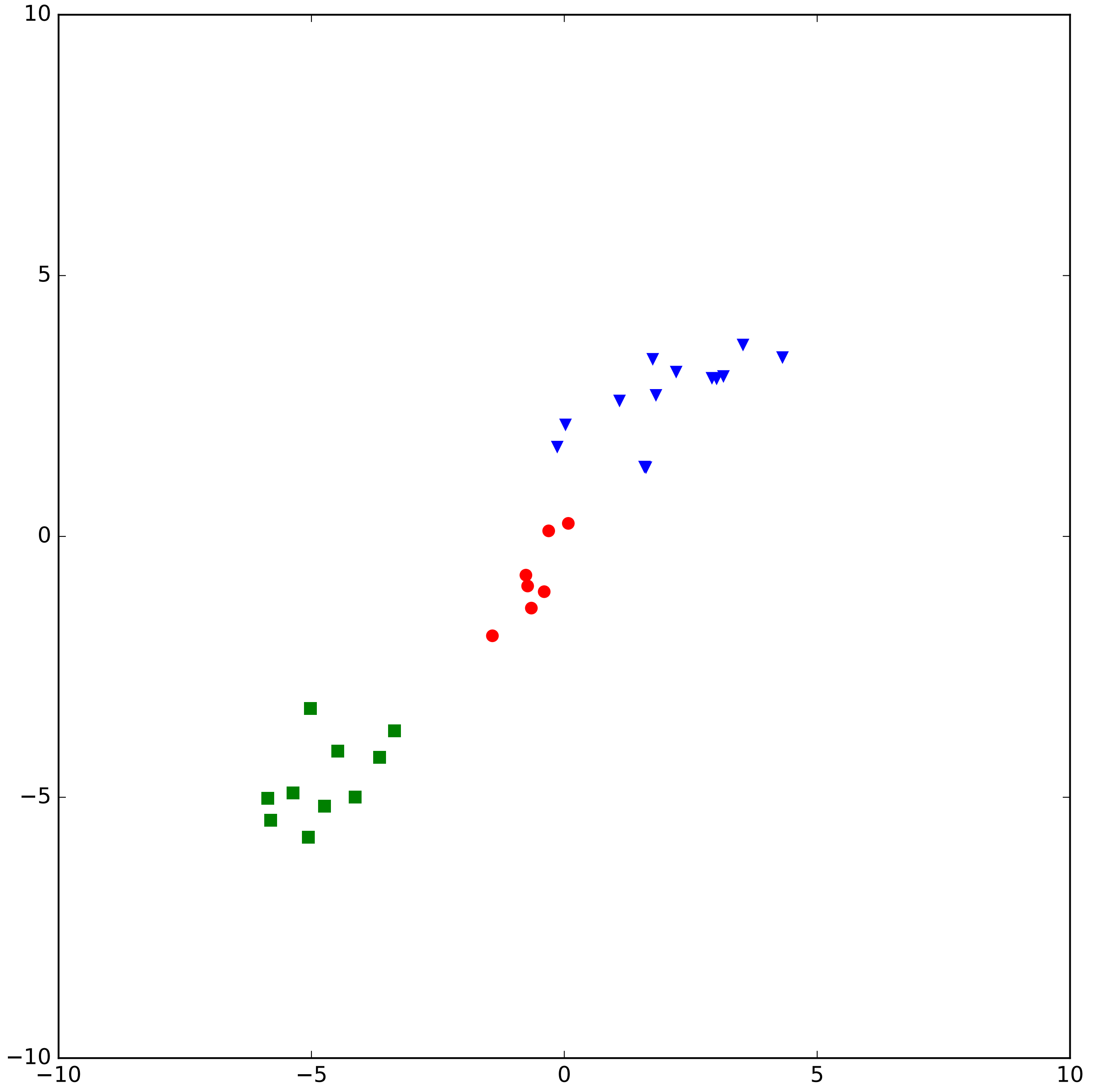
Data has shape



Data has shape

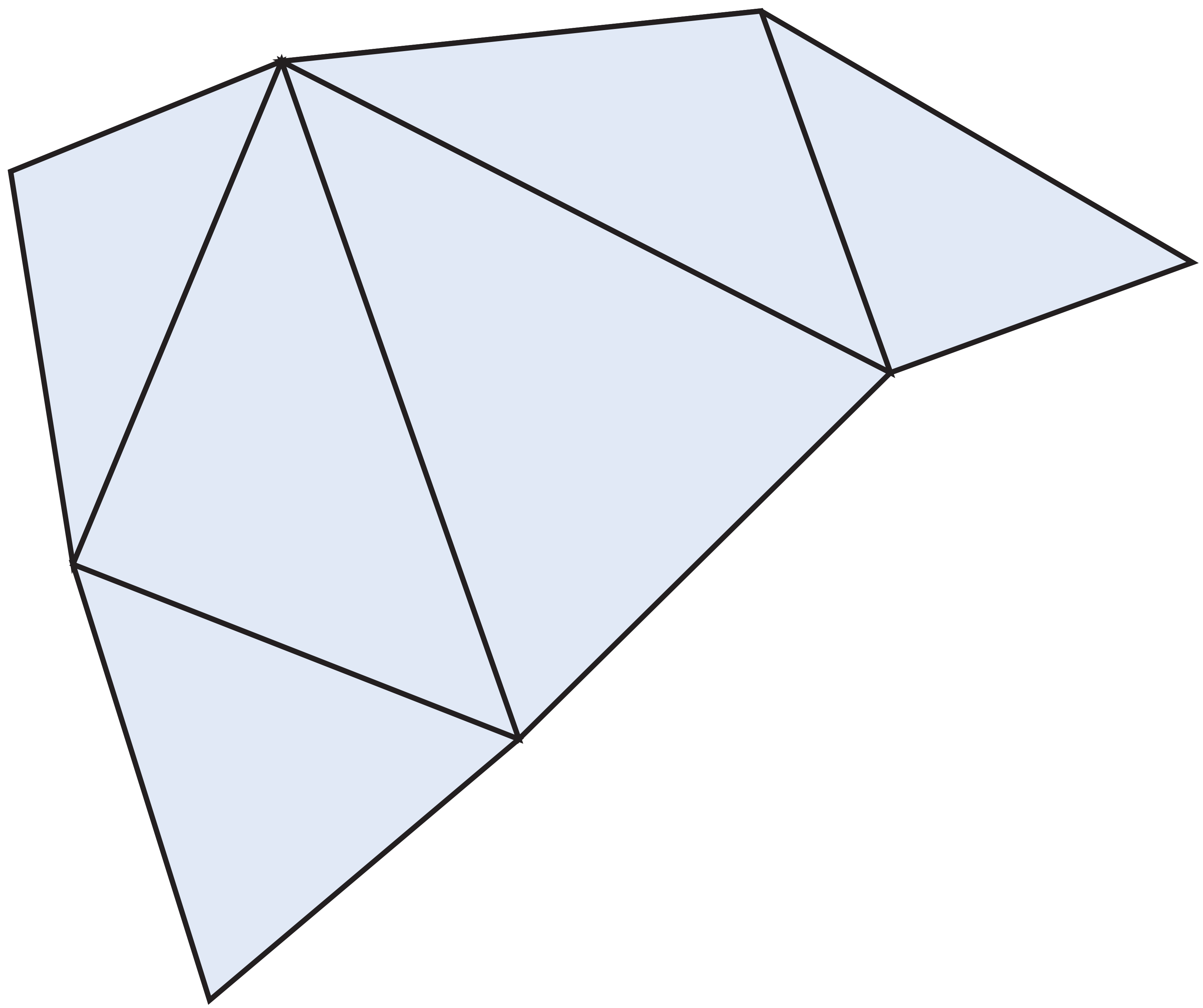


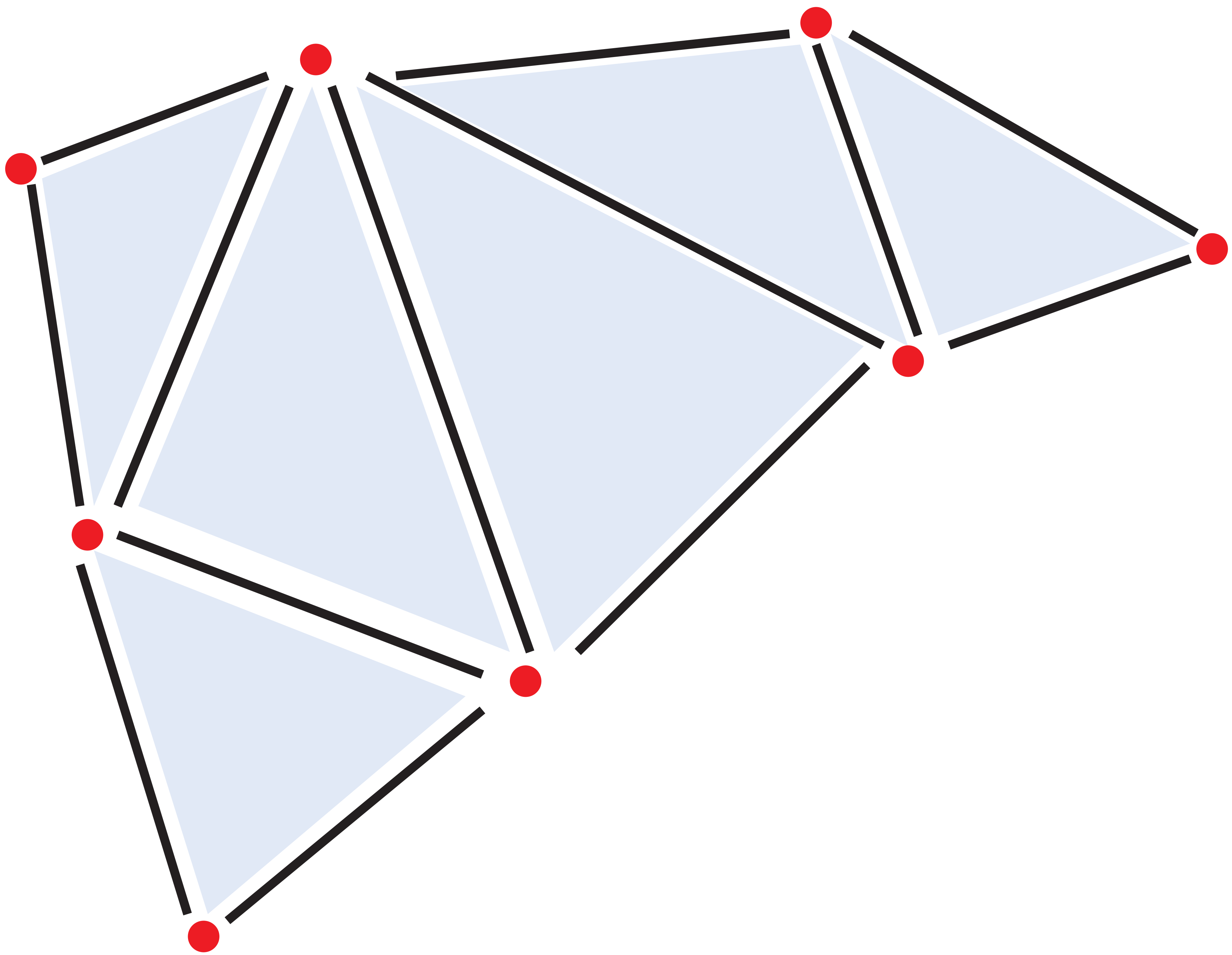
Data has shape



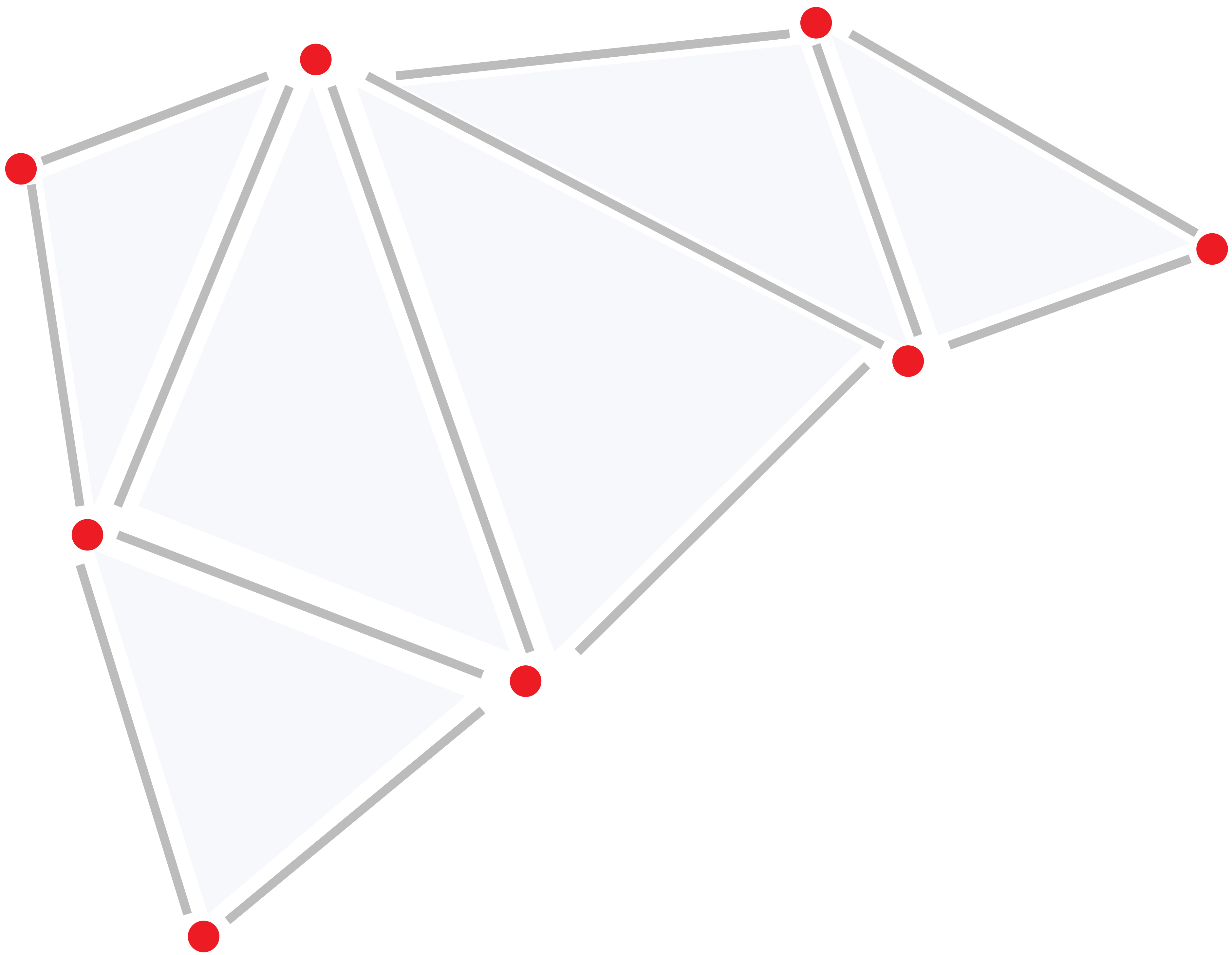
Outline

- Data has shape
- **Homology: linear algebra measures shape**
- Persistence: squinting with mathematics
- Cohomology
- Applications

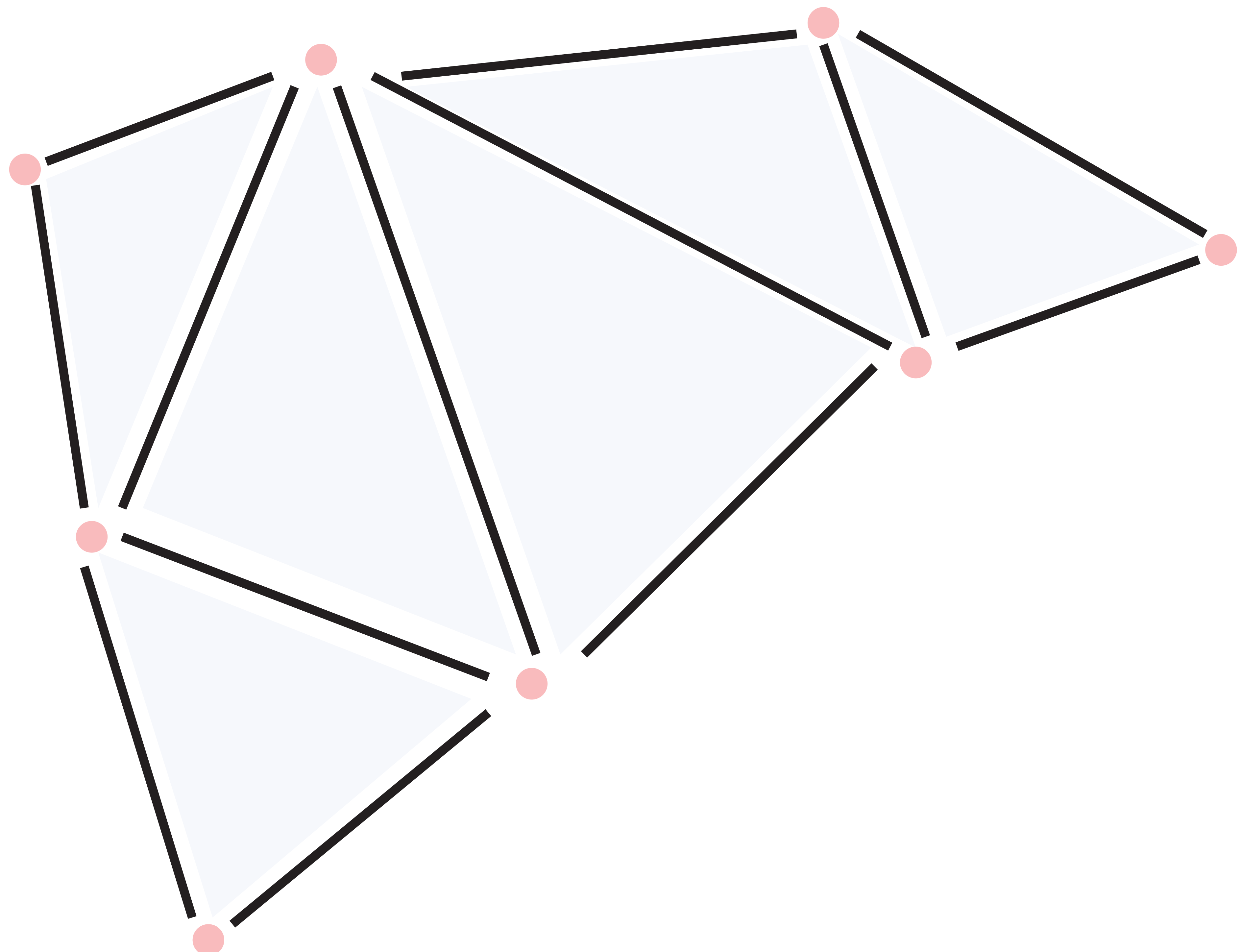




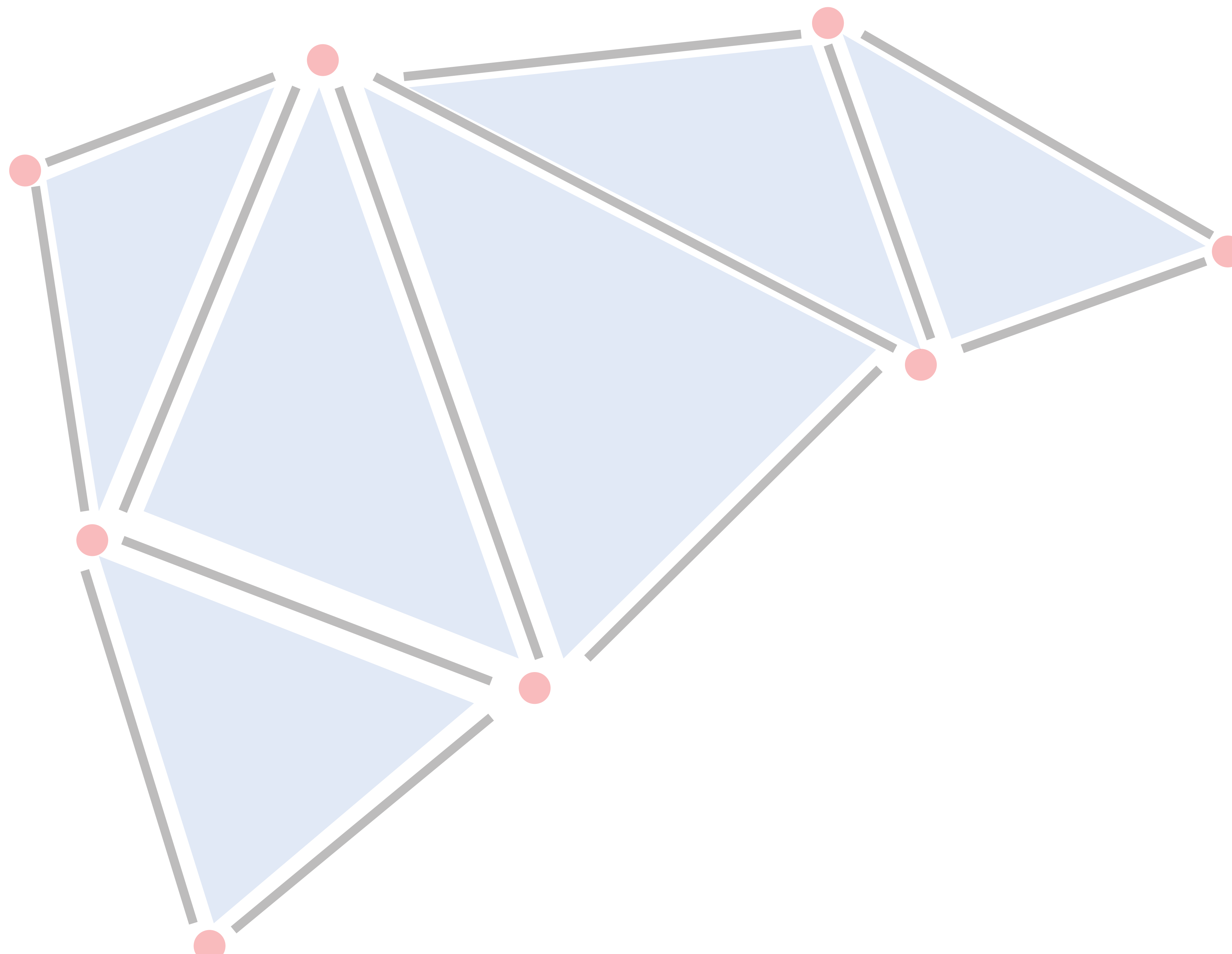
C_0



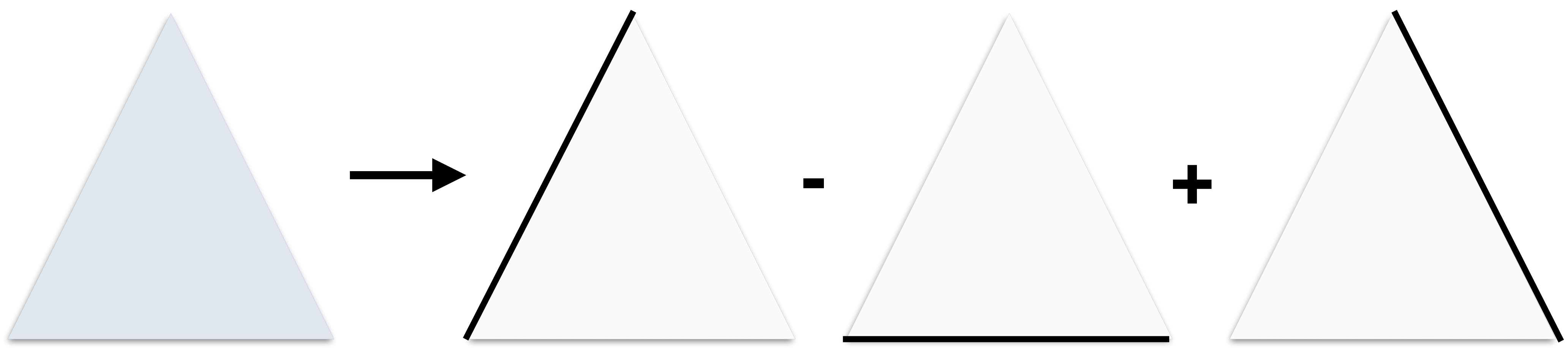
C_1



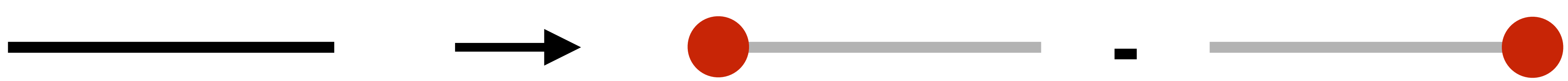
C_2



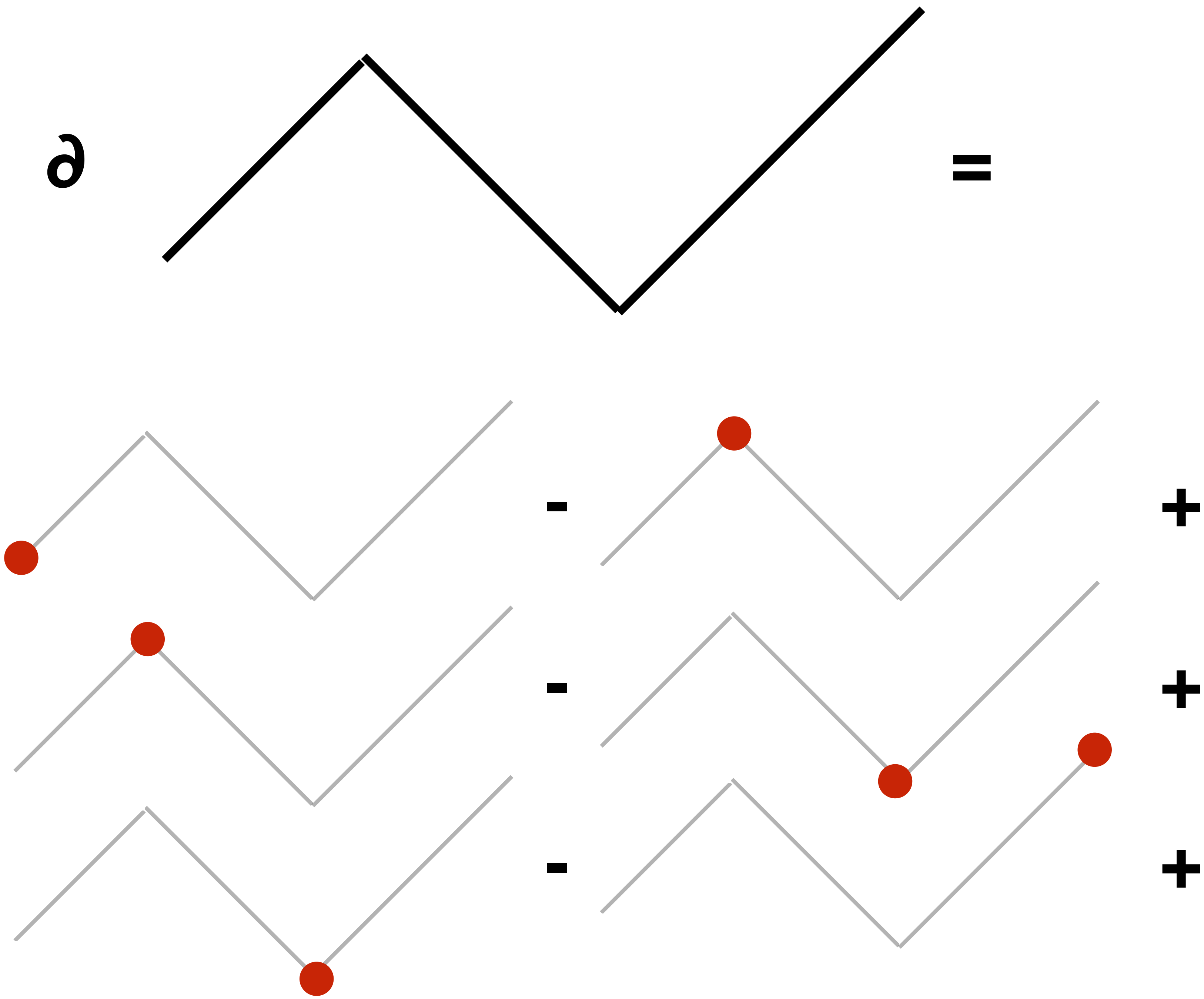
Linear maps $\partial: C_2 \longrightarrow C_1$



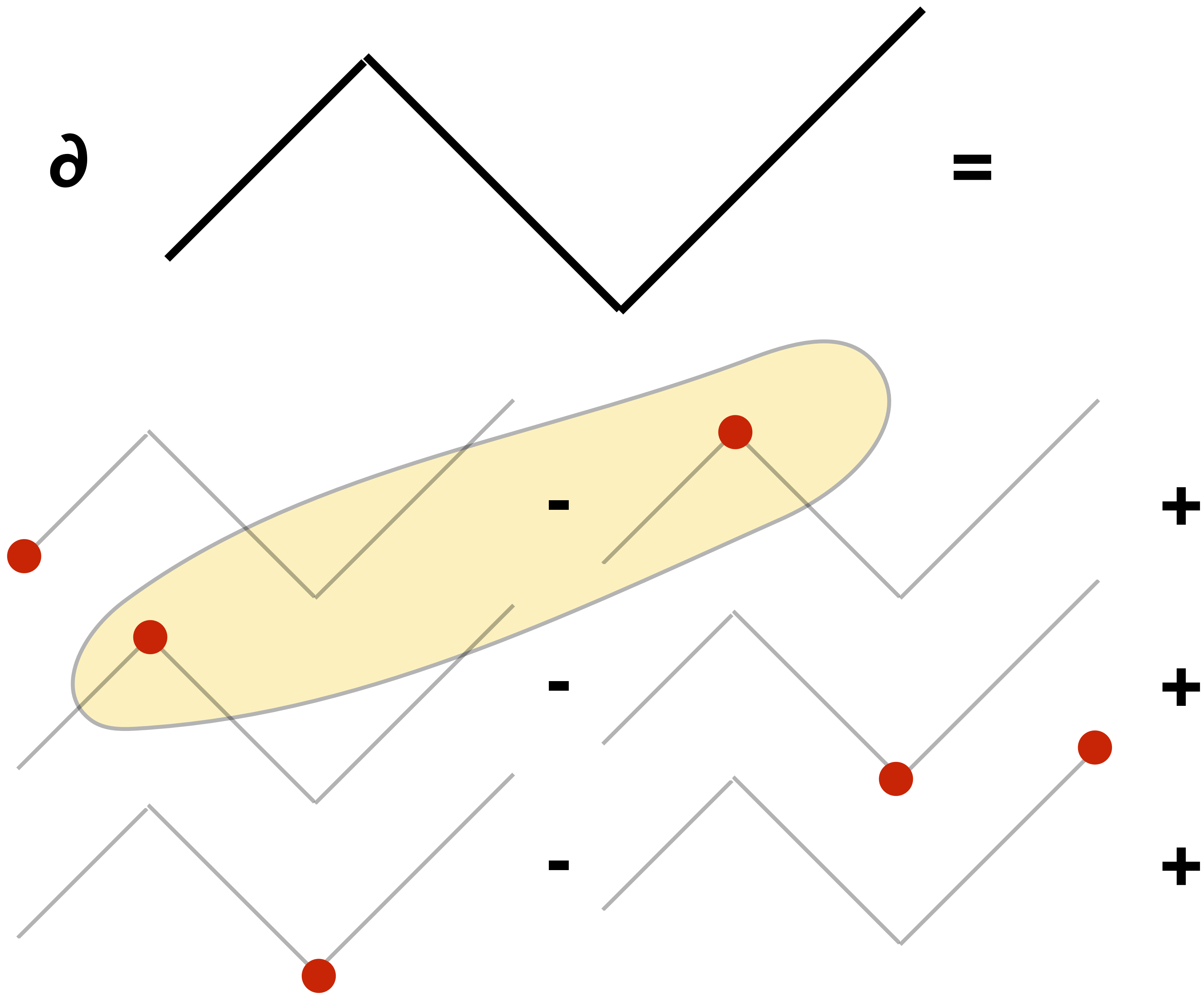
$\partial: C_1 \longrightarrow C_0$



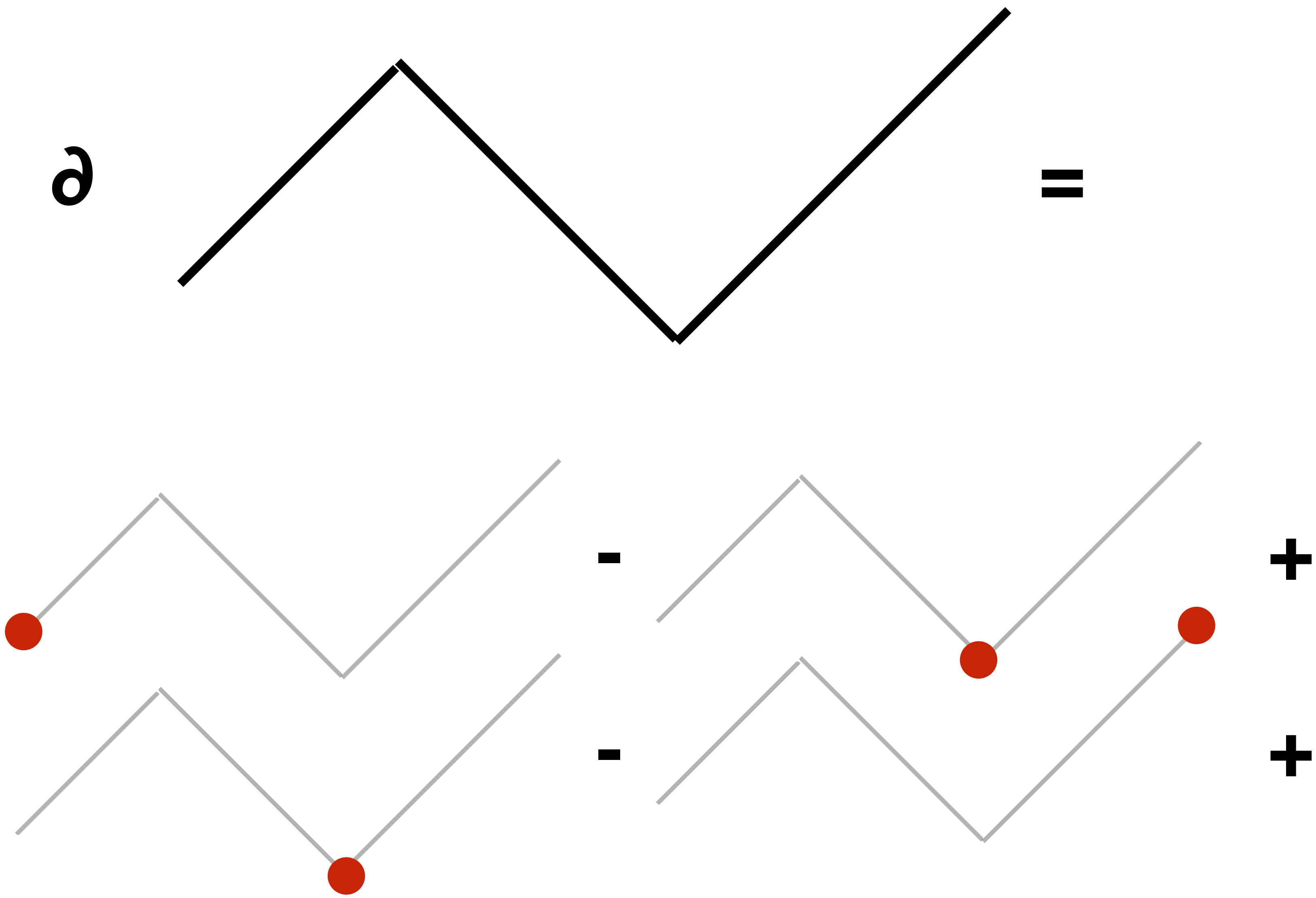
Boundary of paths telescope



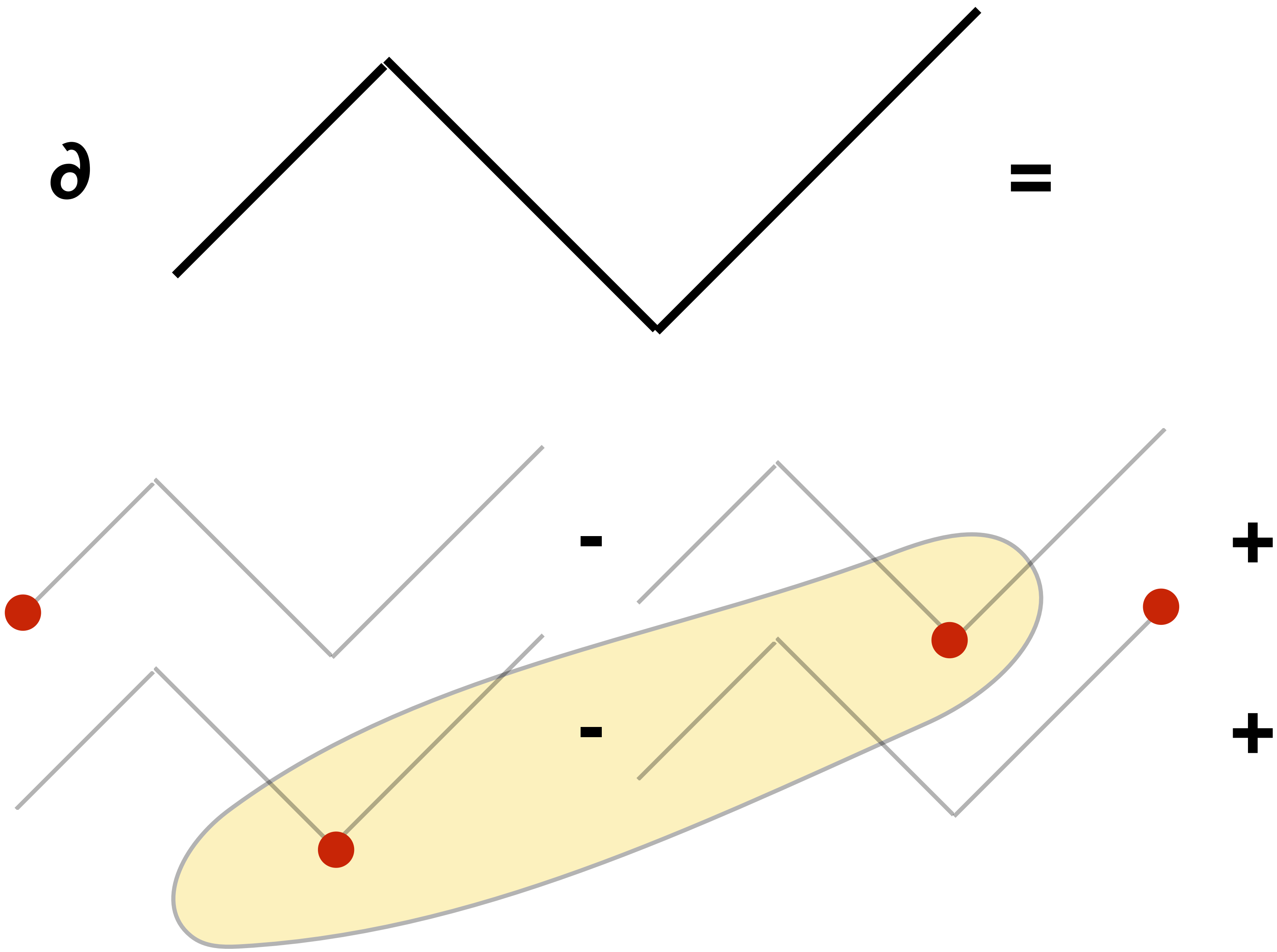
Boundary of paths telescope



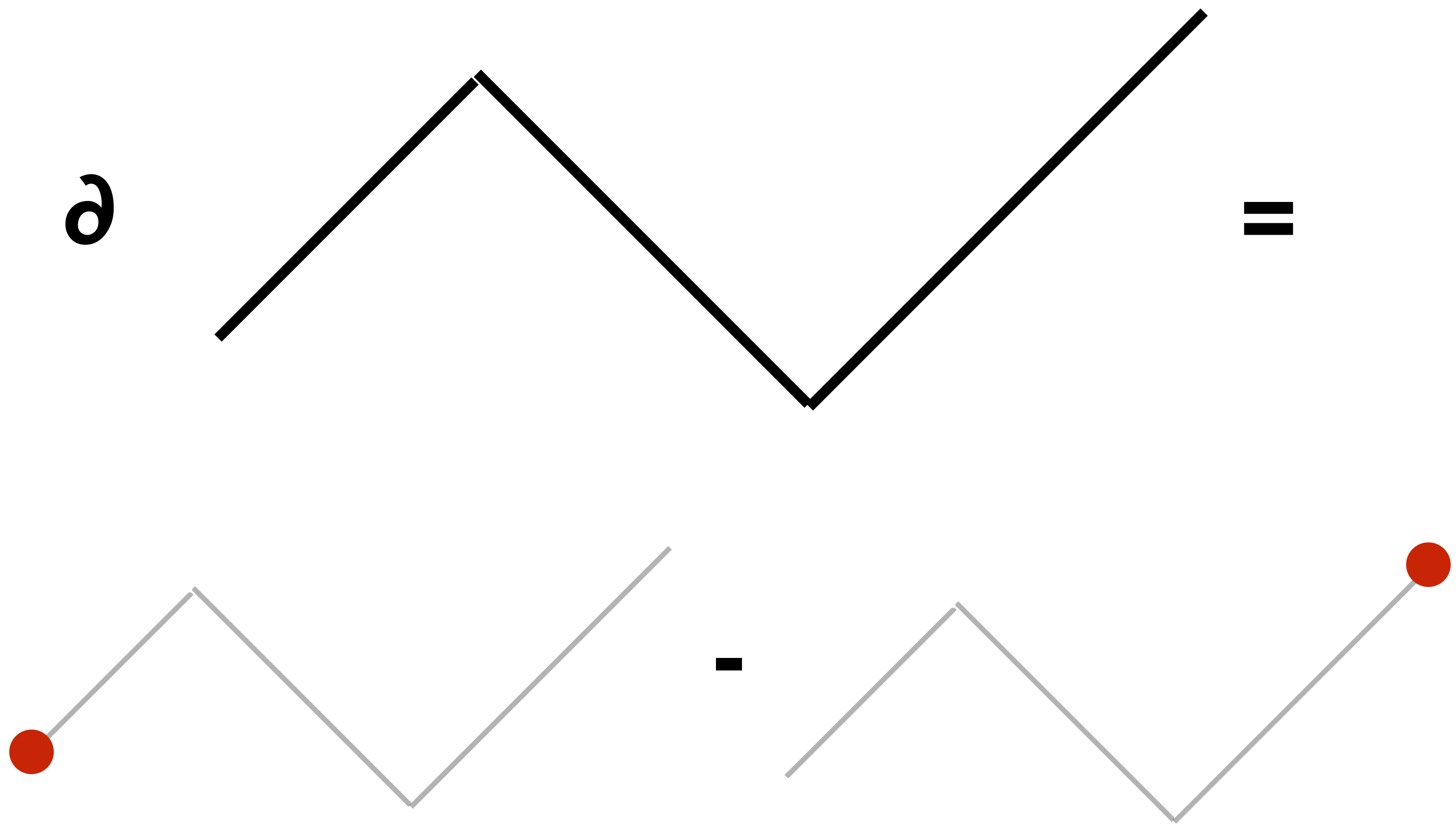
Boundary of paths telescope



Boundary of paths telescope



Boundary of paths telescope



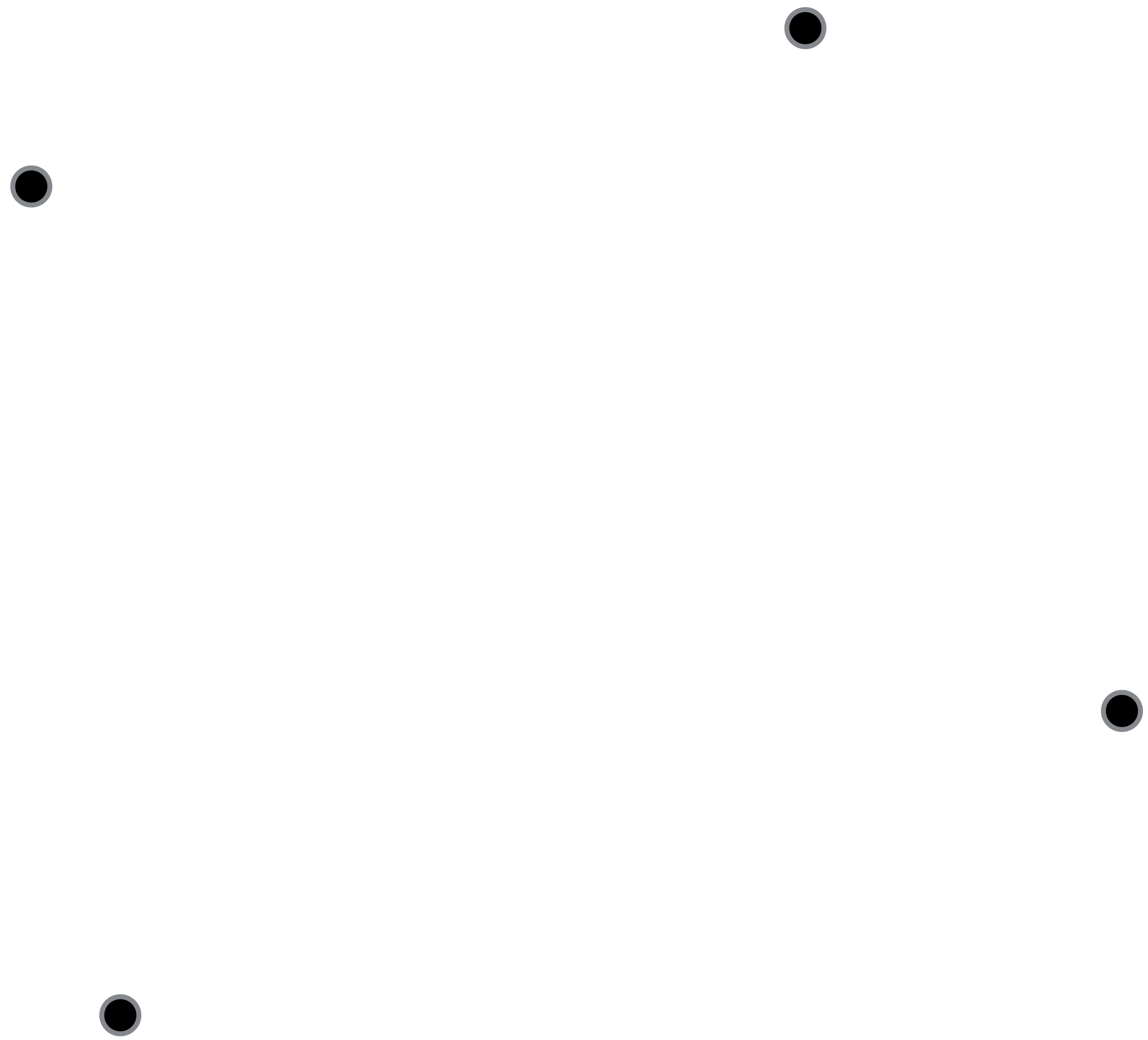
Note: if the path z is a cycle, endpoints coincide, so $\partial z=0$

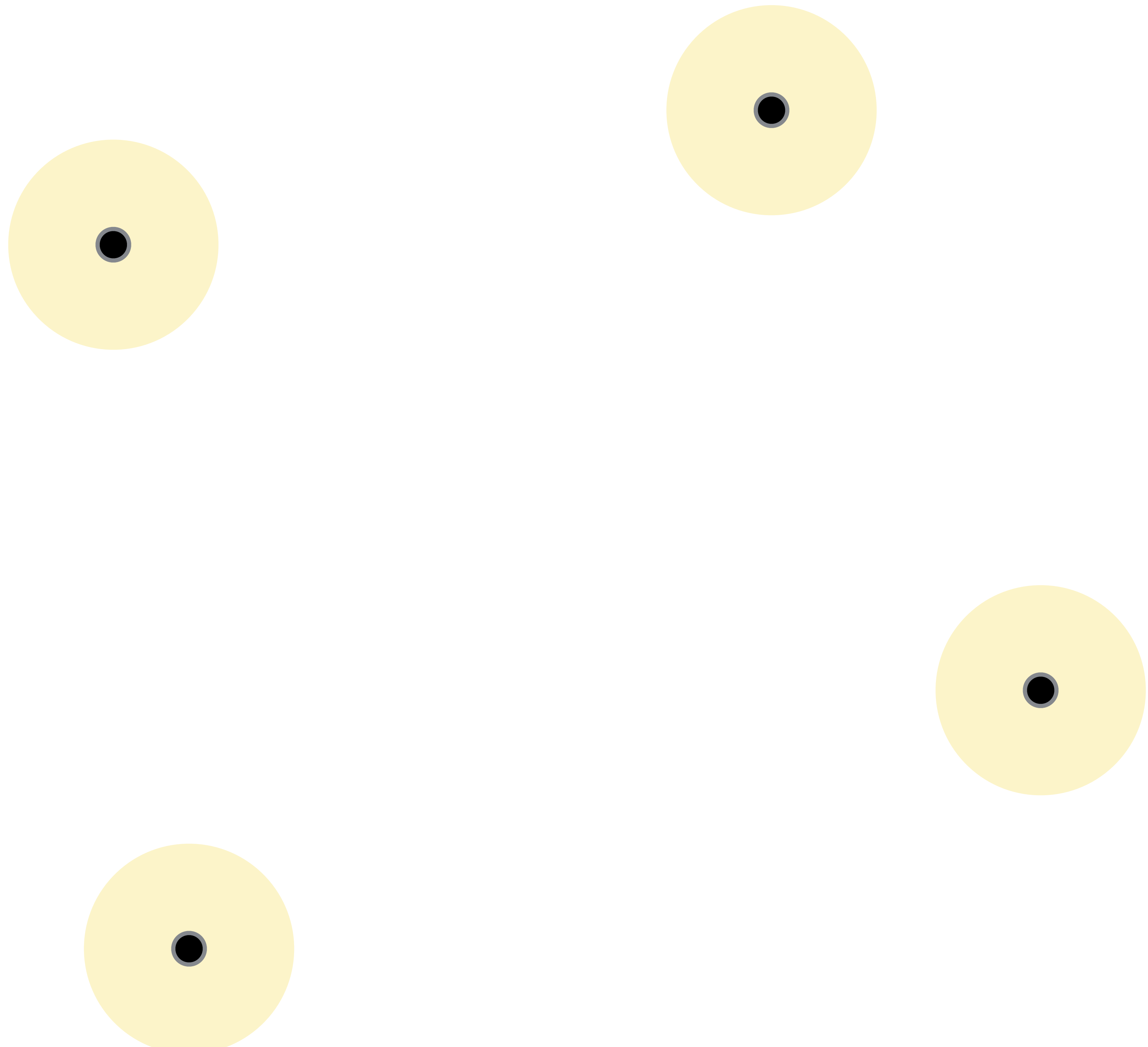
Definitions

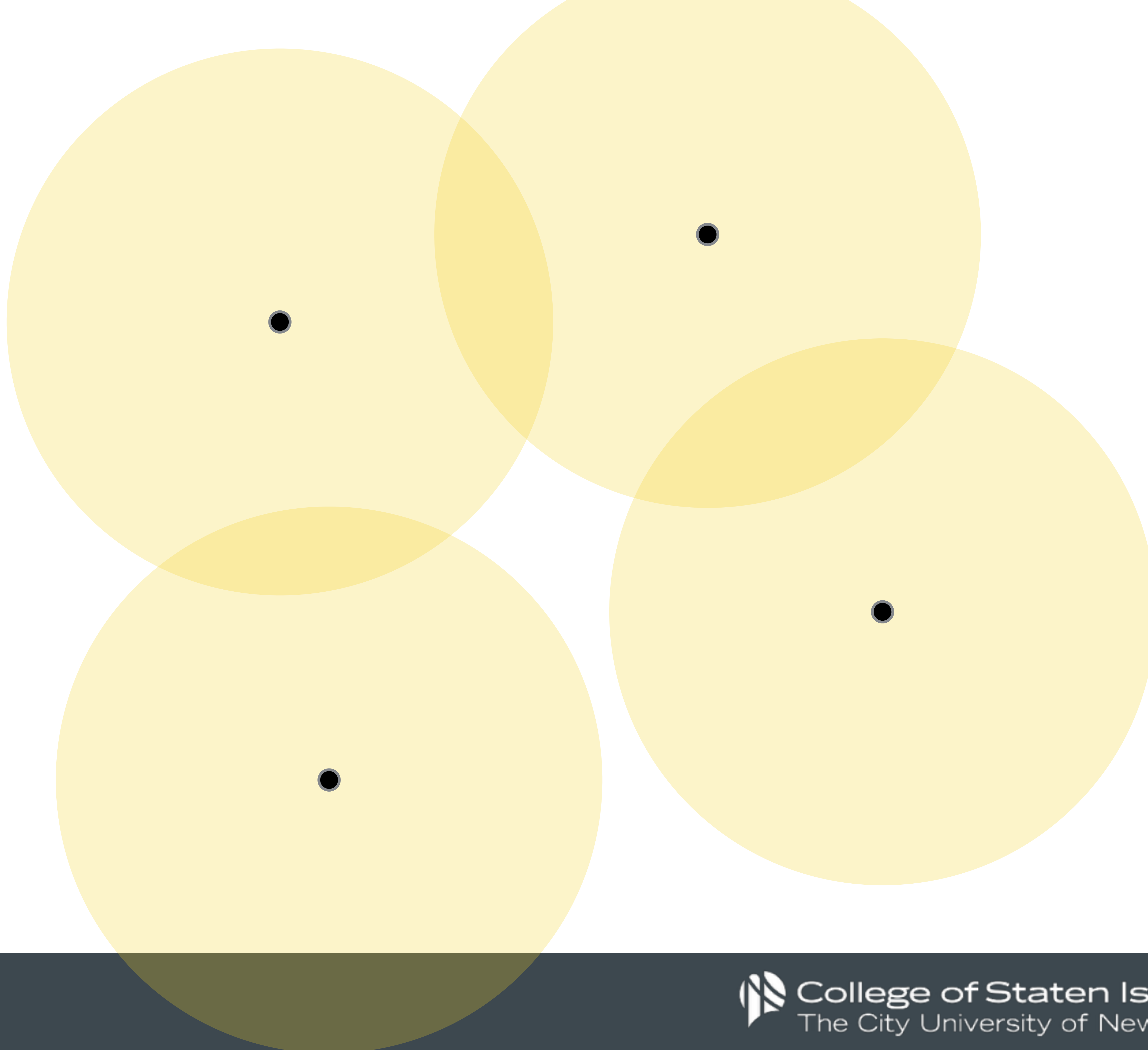
- A **chain** is a linear combination of **simplices**
- A **cycle** is an element of ker ∂
  Something that looks like a closed path
- A **boundary** is an element of img ∂
  Something that should look like a closed path
- **Homology** is the quotient vector space
  ker ∂ / img ∂
  Essential (non-obvious) cycles

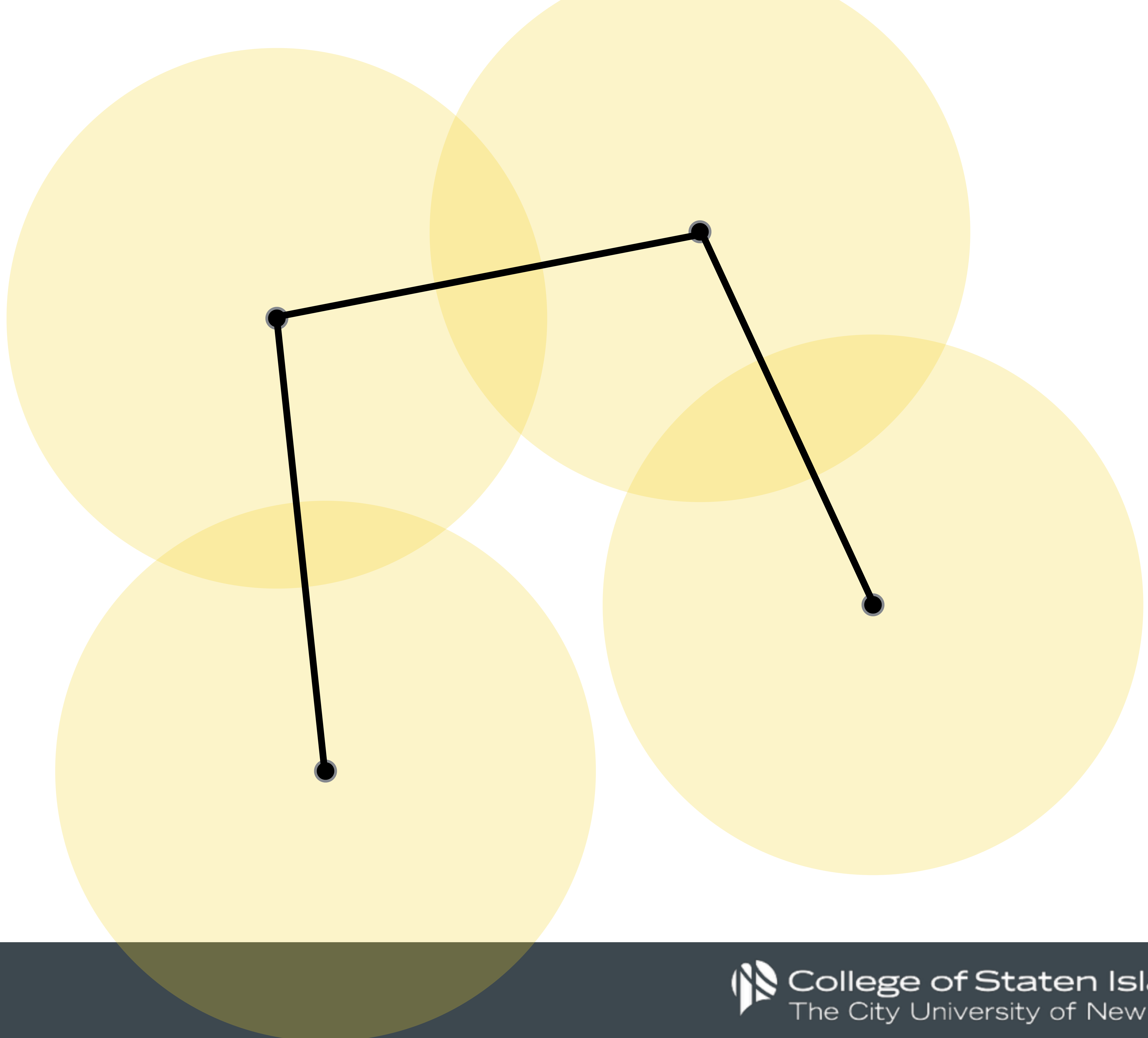
Outline

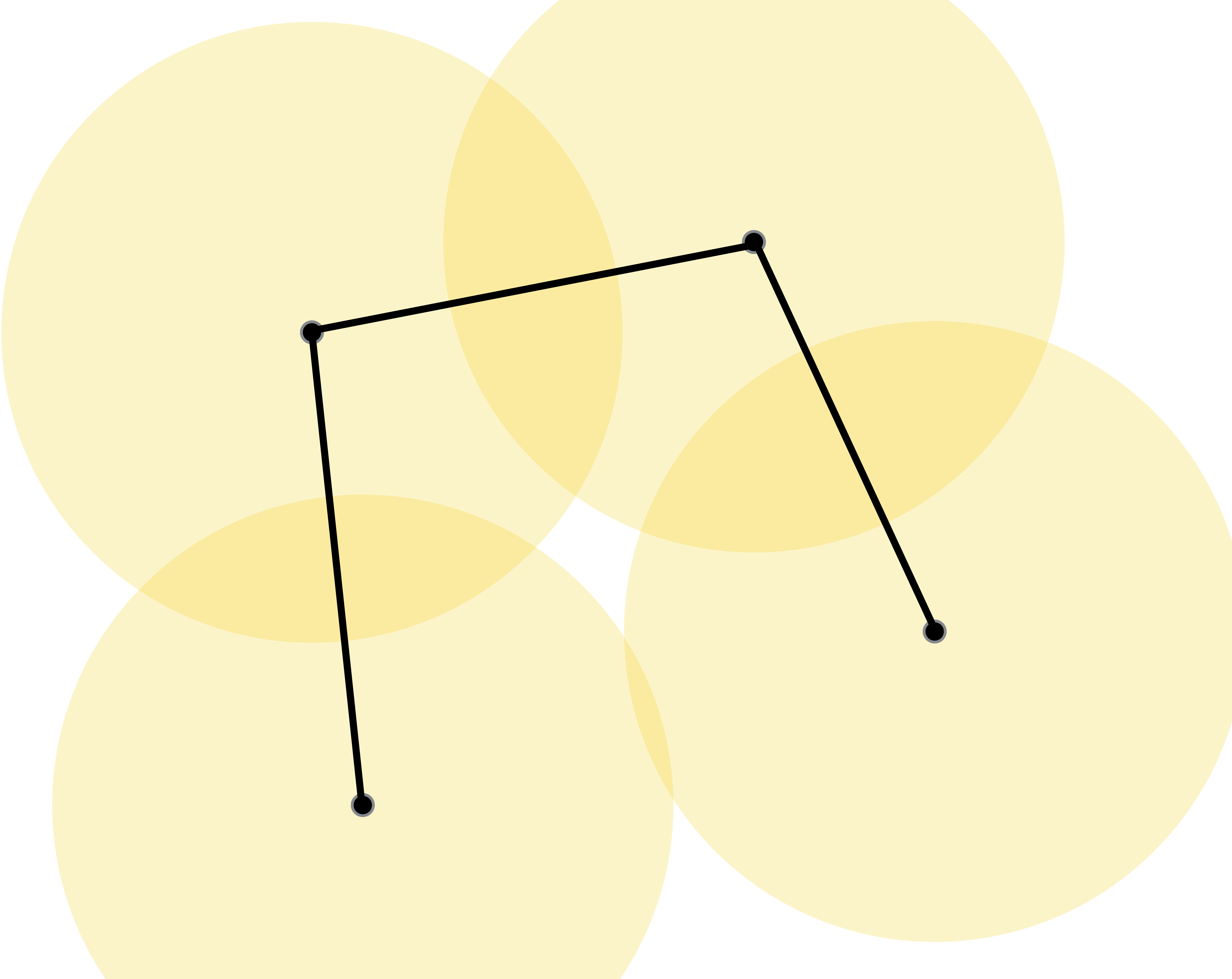
- Data has shape
- Homology: linear algebra measures shape
- **Persistence: squinting with mathematics**
- Cohomology
- Applications

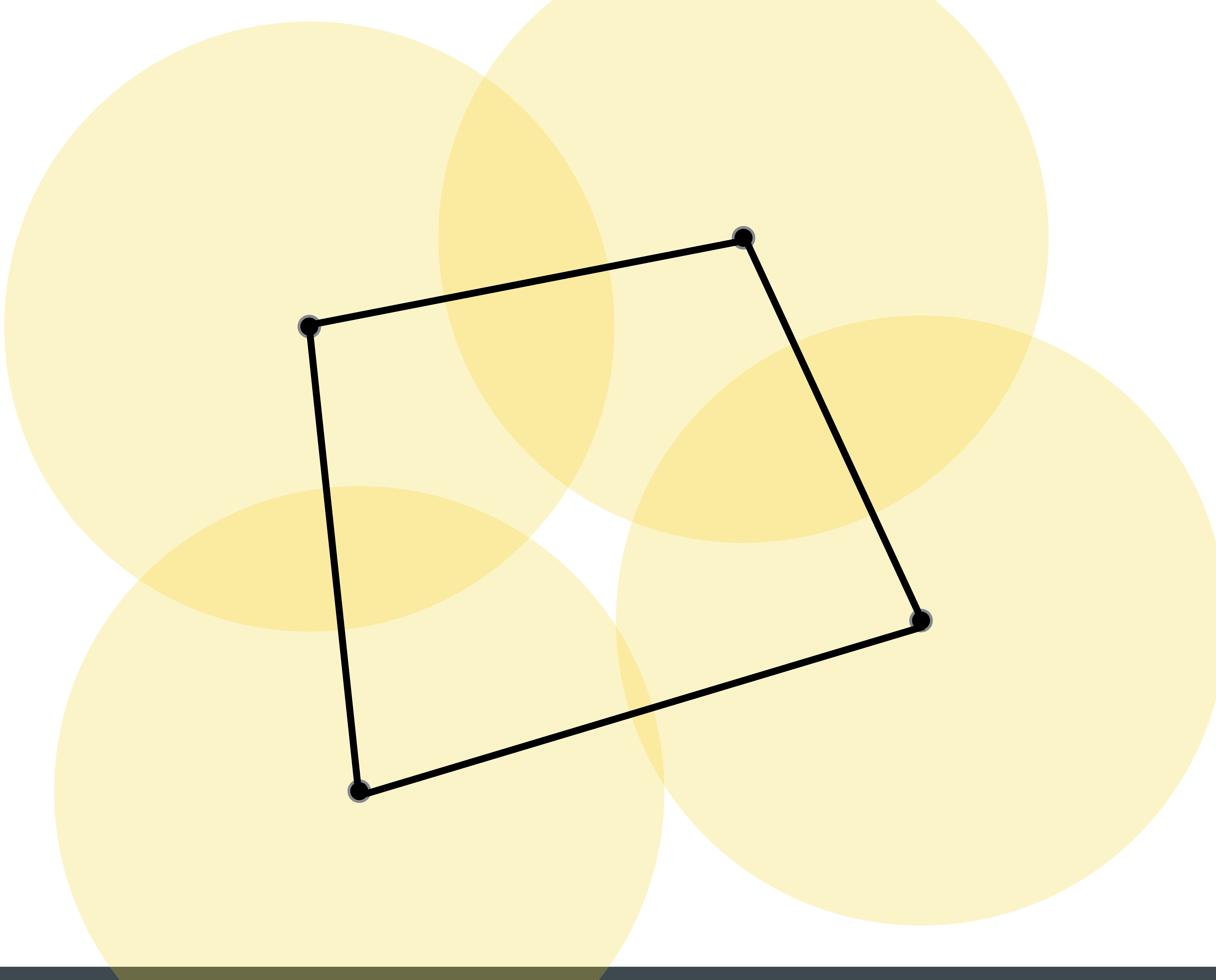


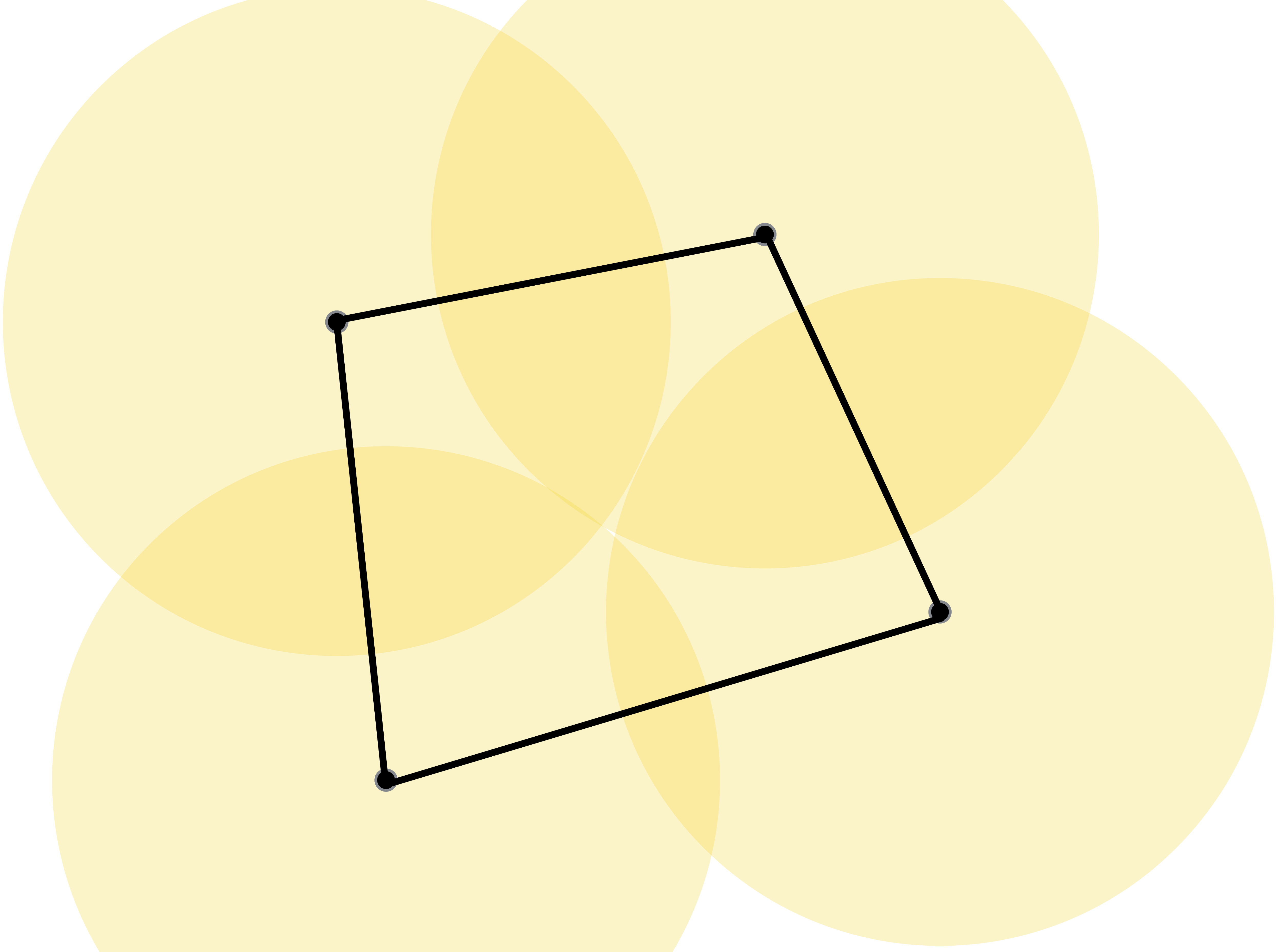


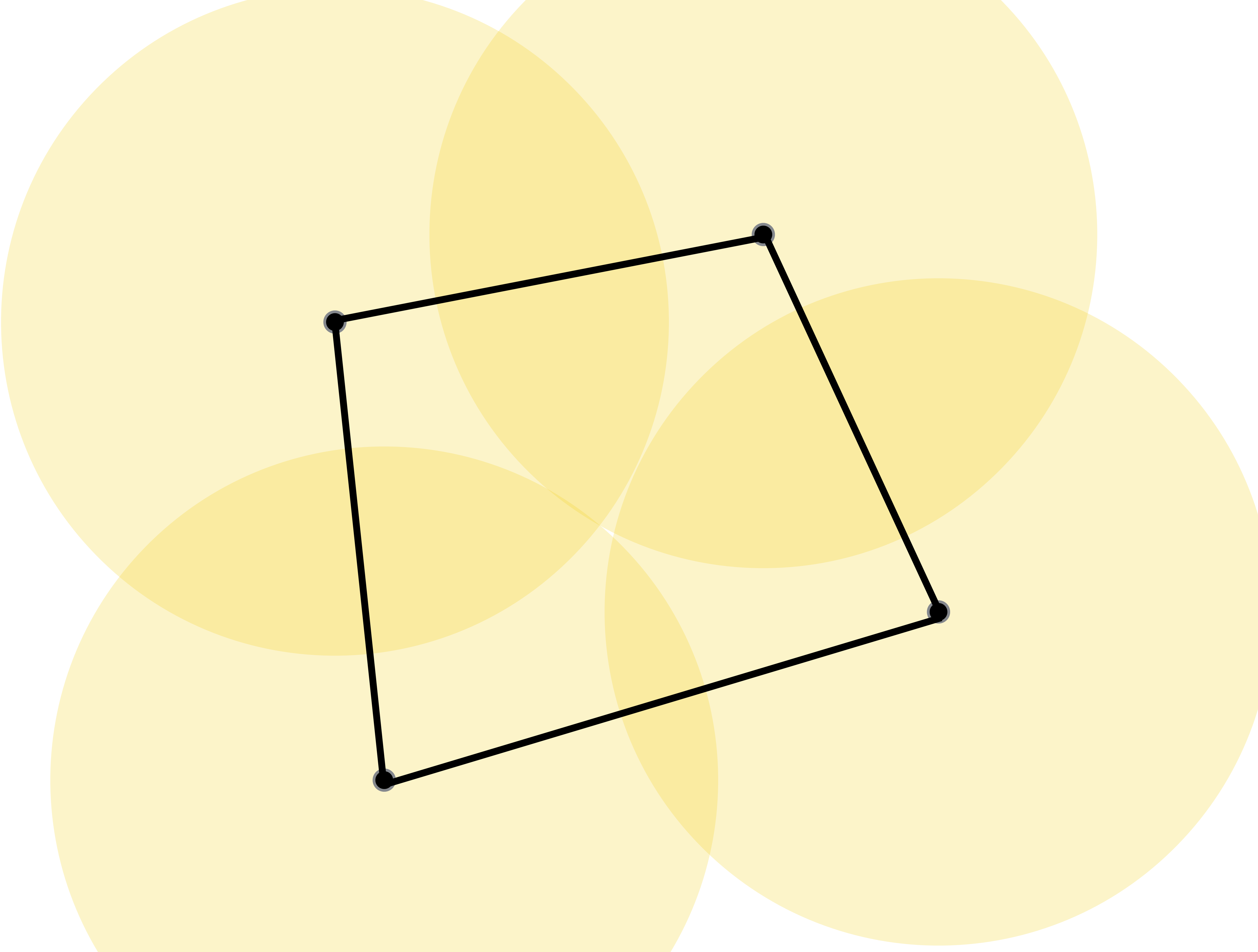


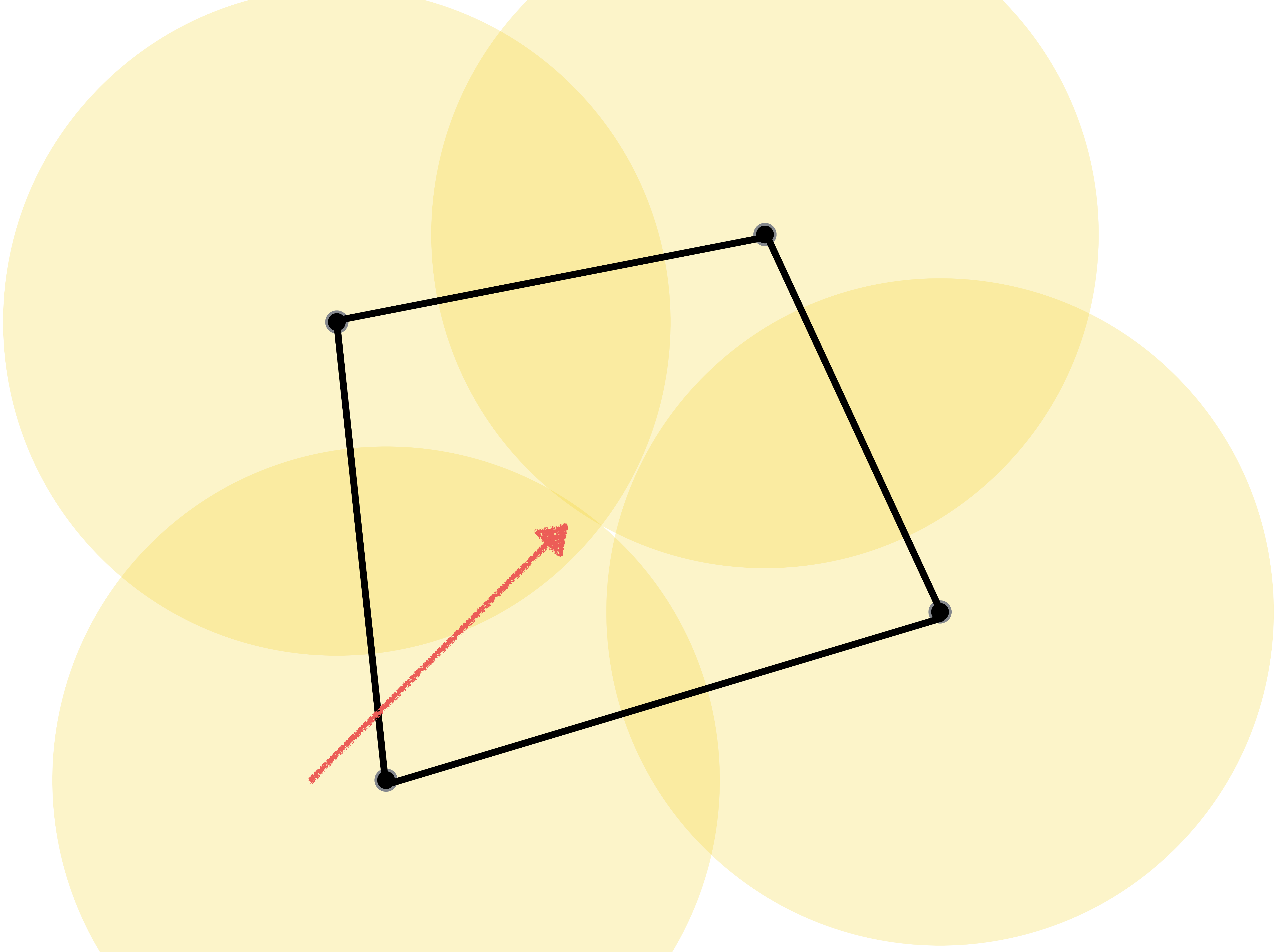


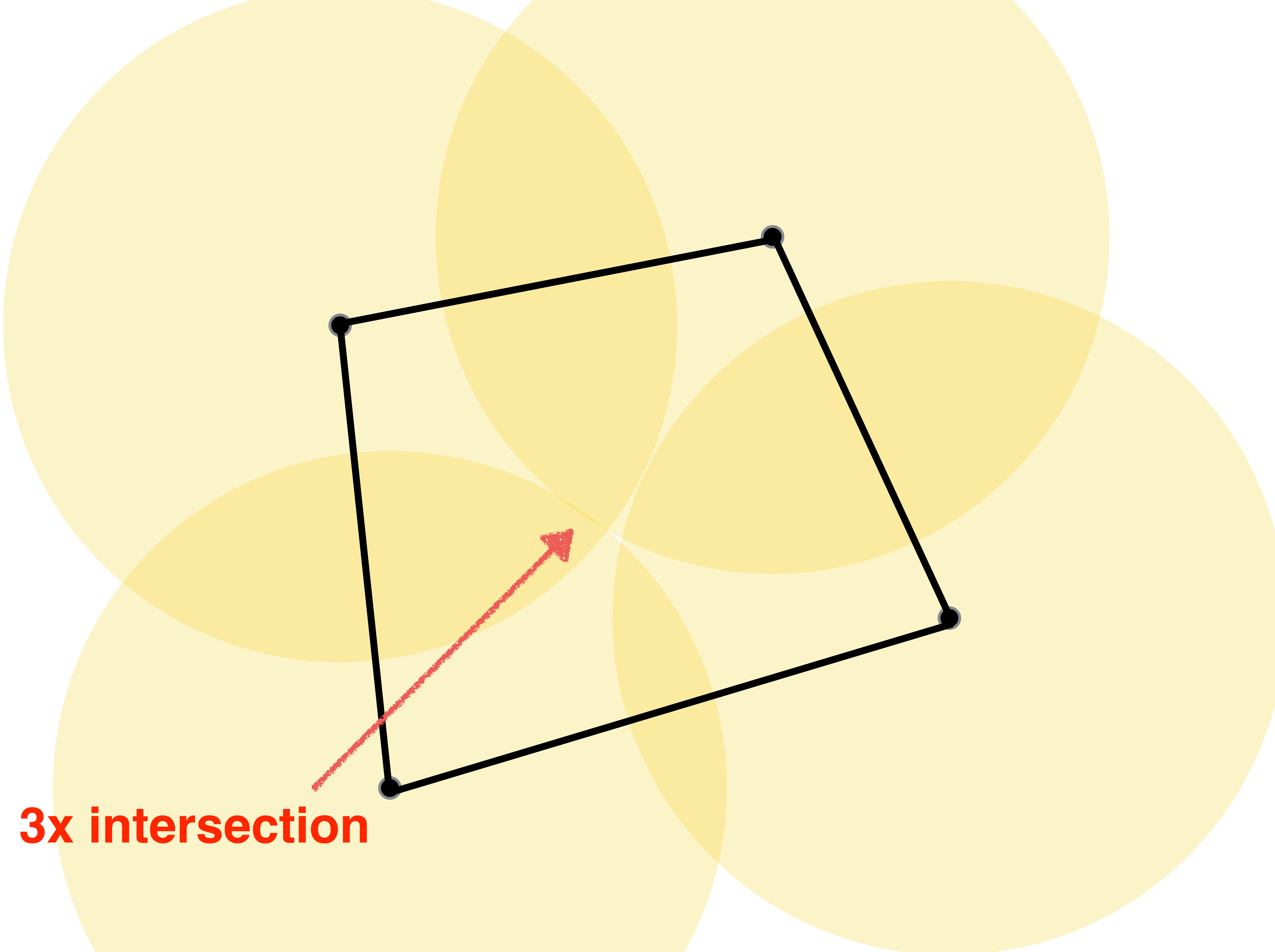


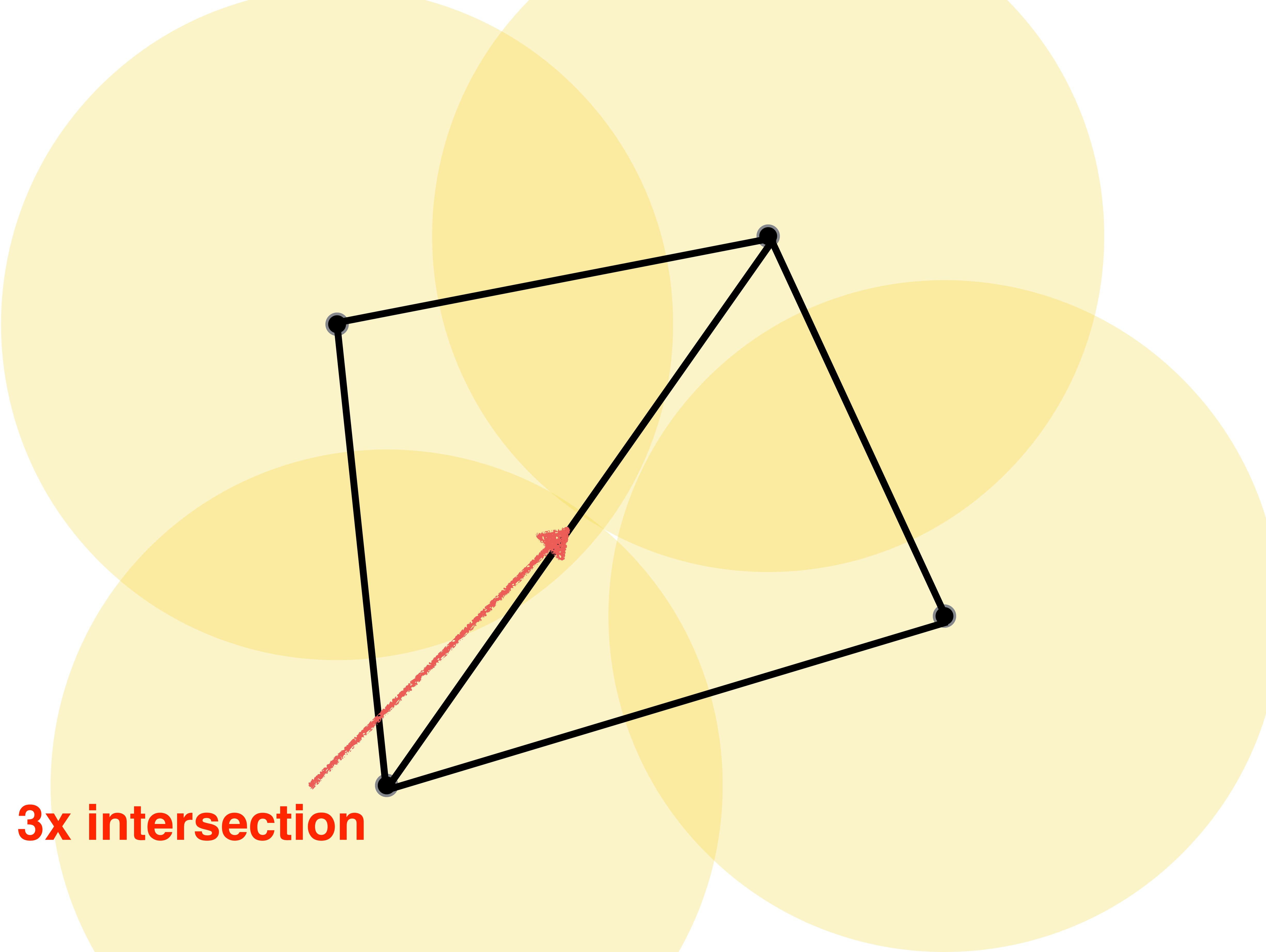


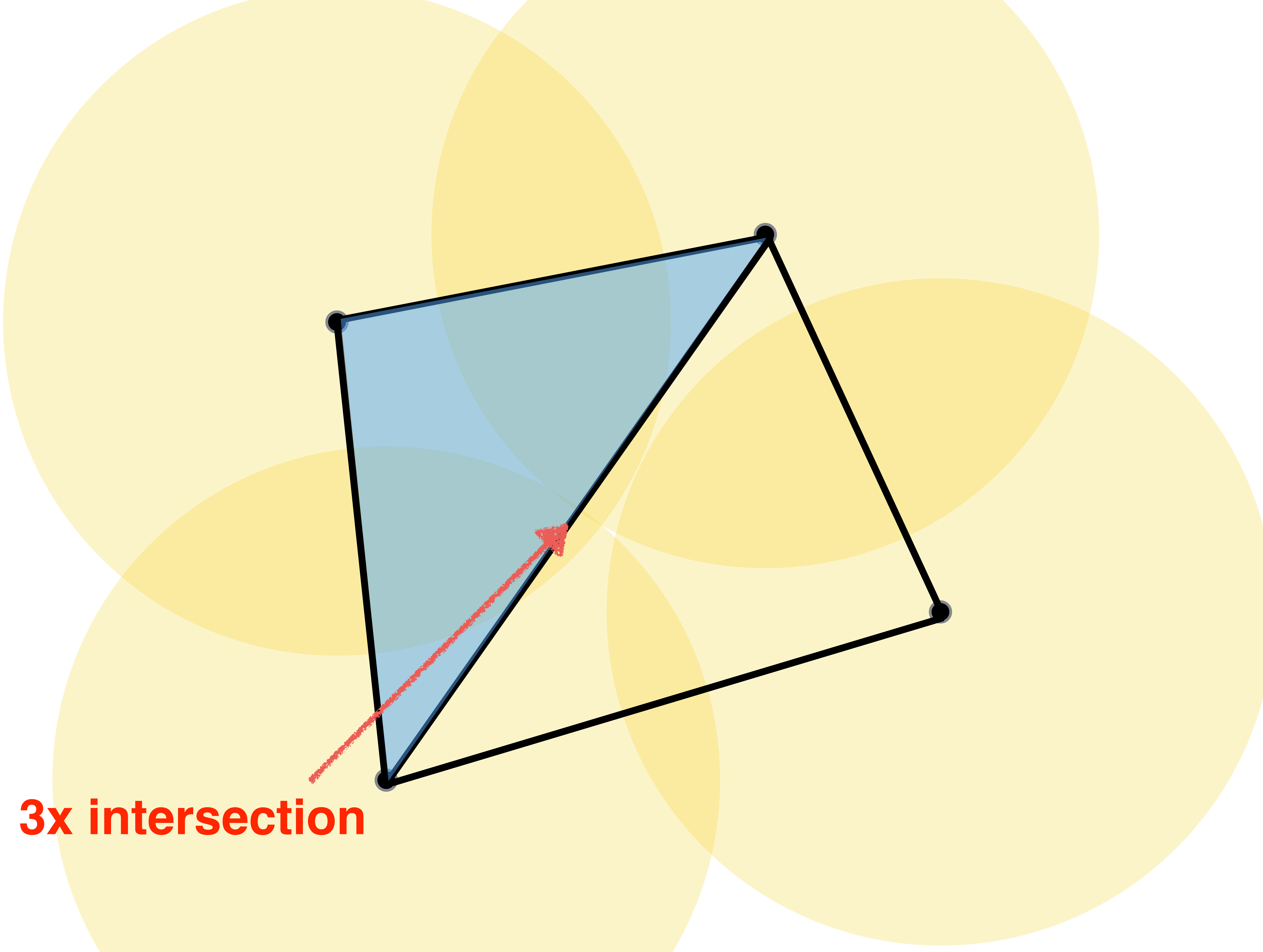


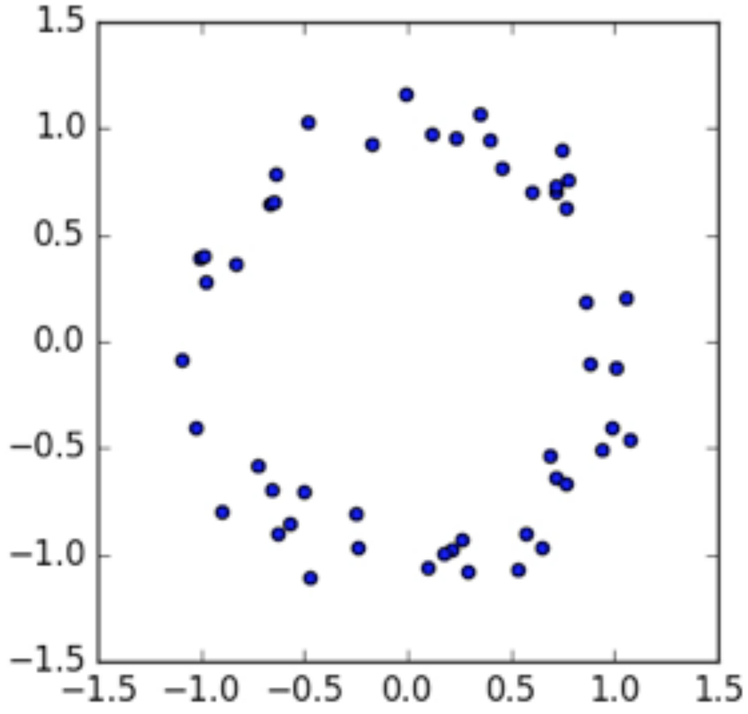


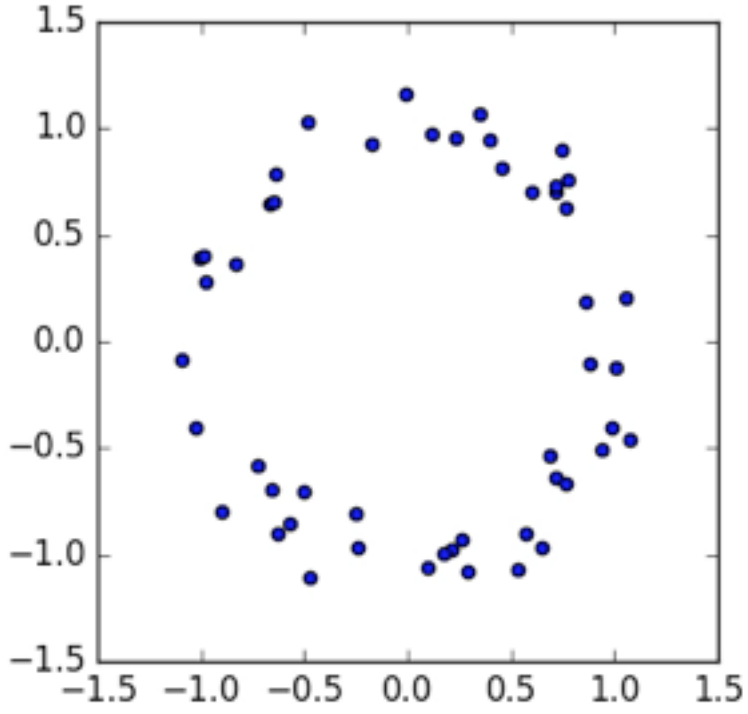


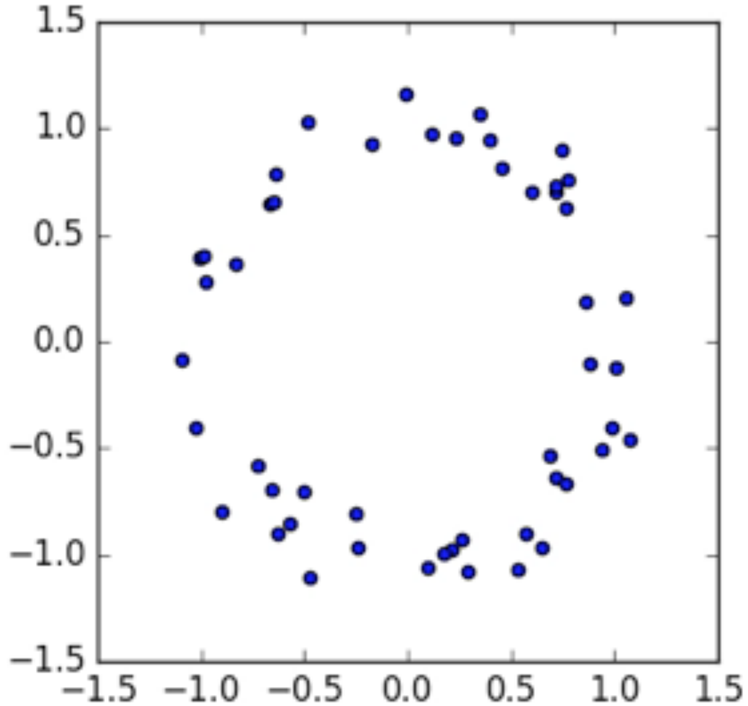

3x intersection


3x intersection

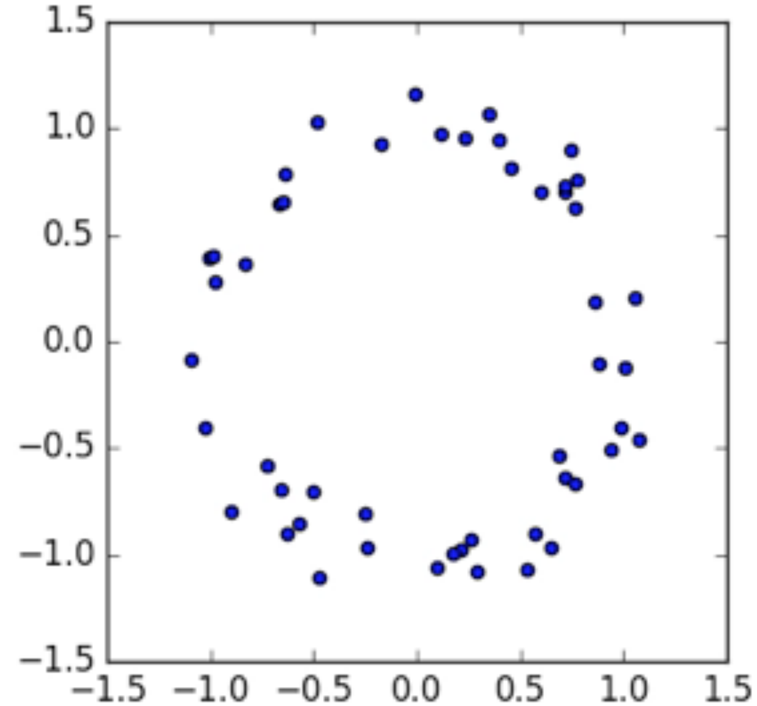

3x intersection


1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.5 −1.0 −0.5 0.0 0.5 1.0 1.5

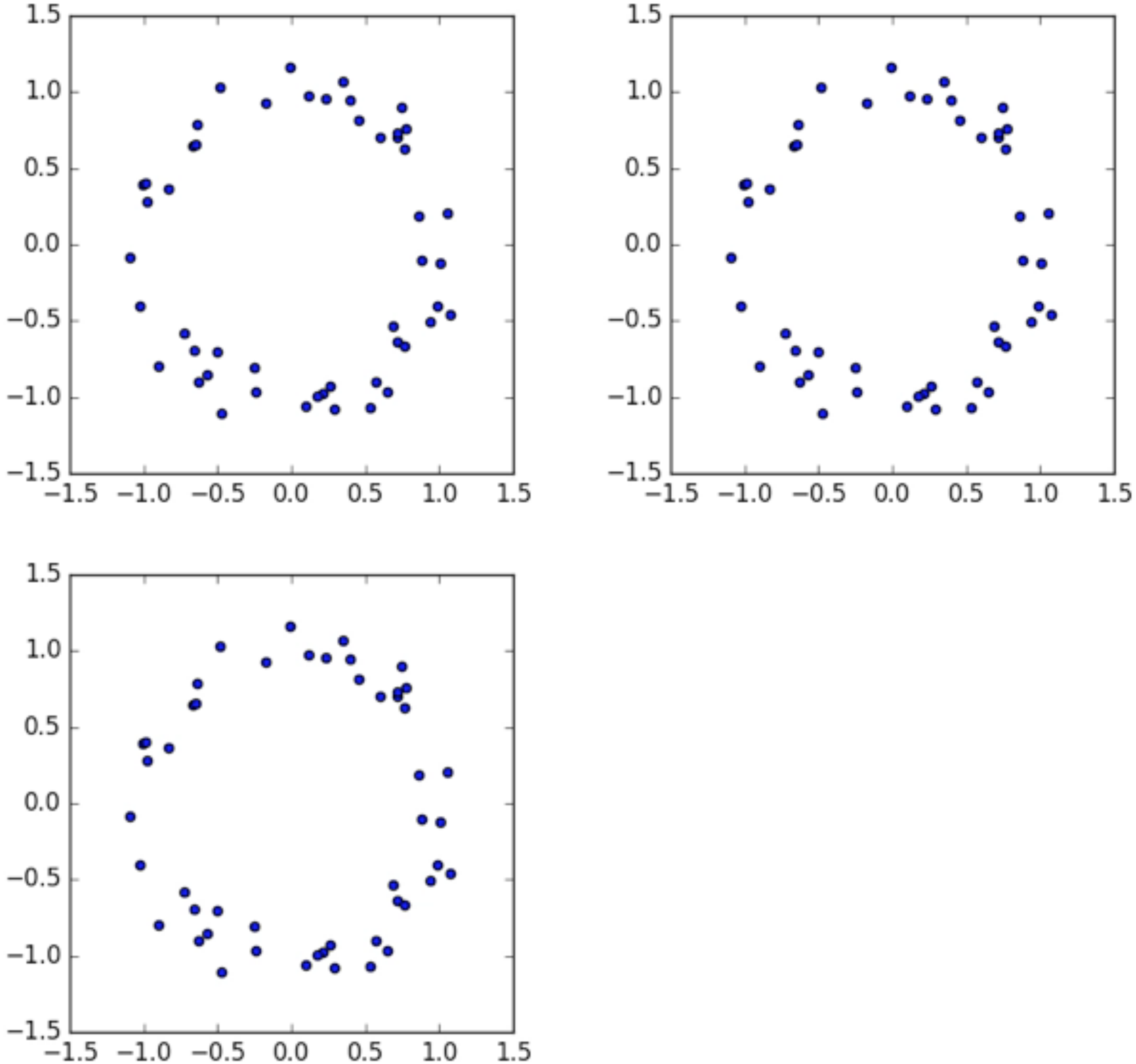

1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.5 −1.0 −0.5 0.0 0.5 1.0 1.5

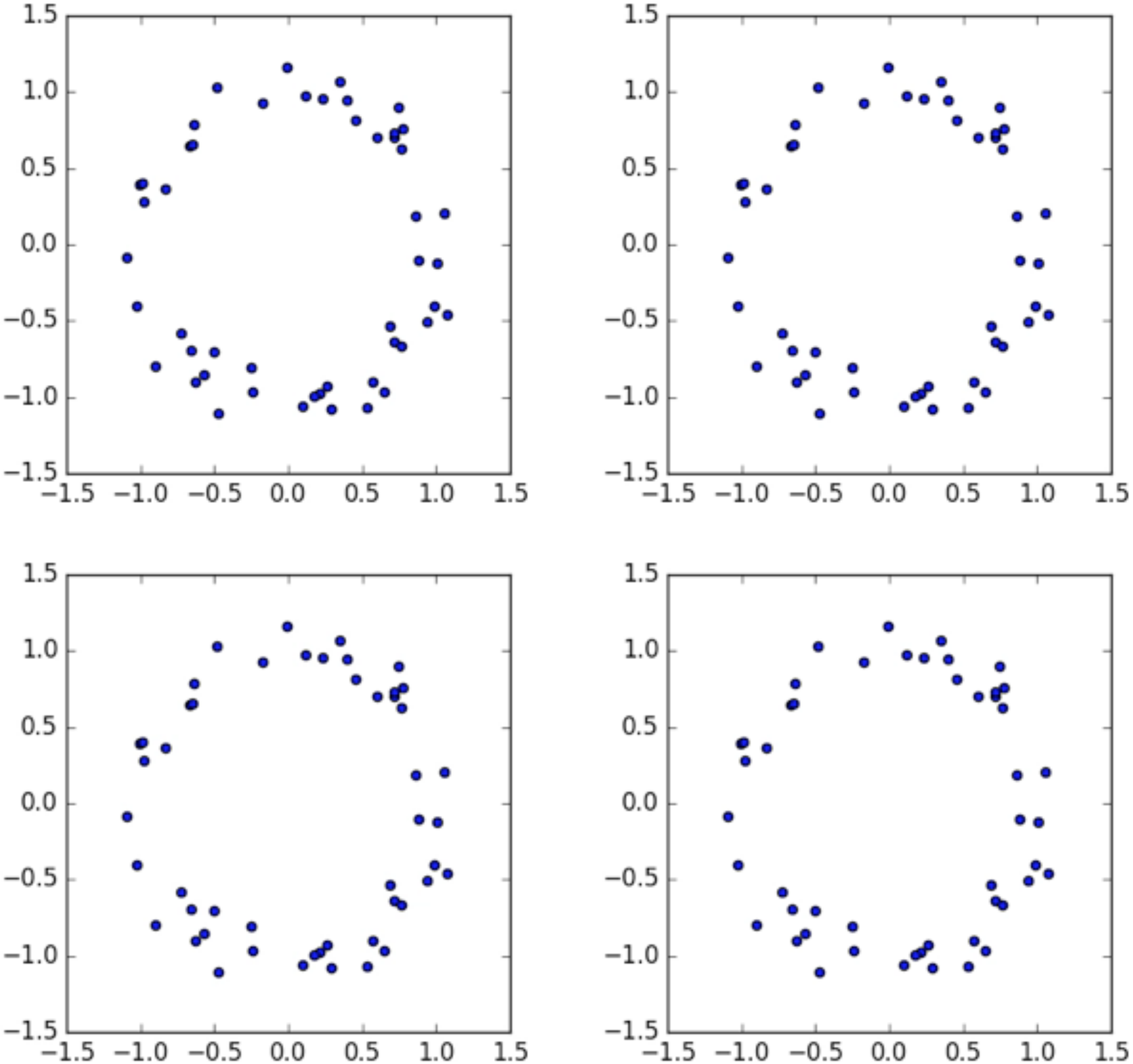


1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.5 −1.0 −0.5 0.0 0.5 1.0 1.5

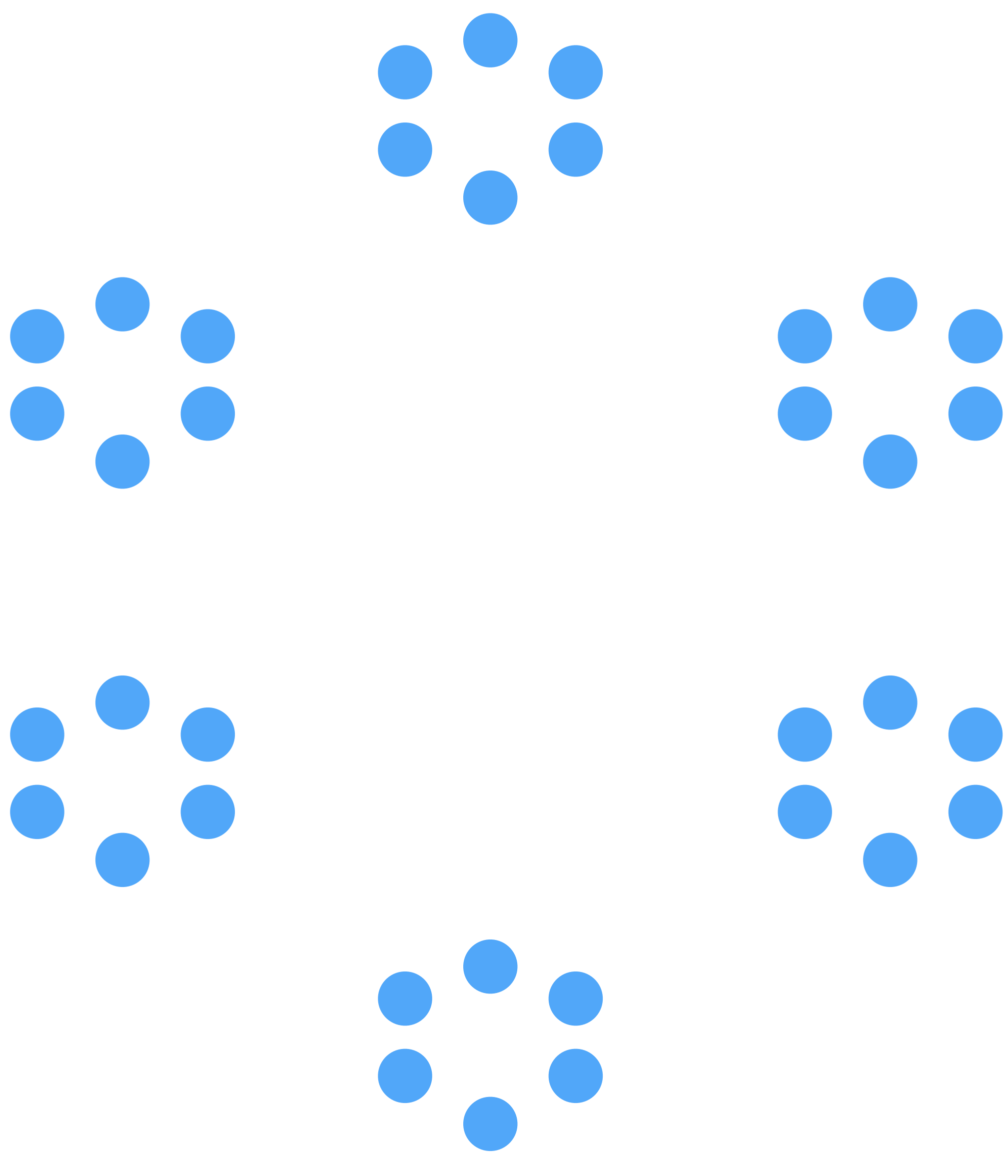


1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.5 −1.0 −0.5 0.0 0.5 1.0 1.5


1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.5 −1.0 −0.5 0.0 0.5 1.0 1.5
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.5 −1.0 −0.5 0.0 0.5 1.0 1.5
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.5 −1.0 −0.5 0.0 0.5 1.0 1.5


1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.5 −1.0 −0.5 0.0 0.5 1.0 1.5
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.5 −1.0 −0.5 0.0 0.5 1.0 1.5
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.5 −1.0 −0.5 0.0 0.5 1.0 1.5
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.5 −1.0 −0.5 0.0 0.5 1.0 1.5

What parameter to pick?



Persistent homology

- Look at **all** parameters at once
- Scale-independent
- Compressed data summary

- As the parameter grows, complexes include
  Inclusions create linear maps on homology
- Track homology classes as the parameter grows
  Homology is born, lives, and dies.

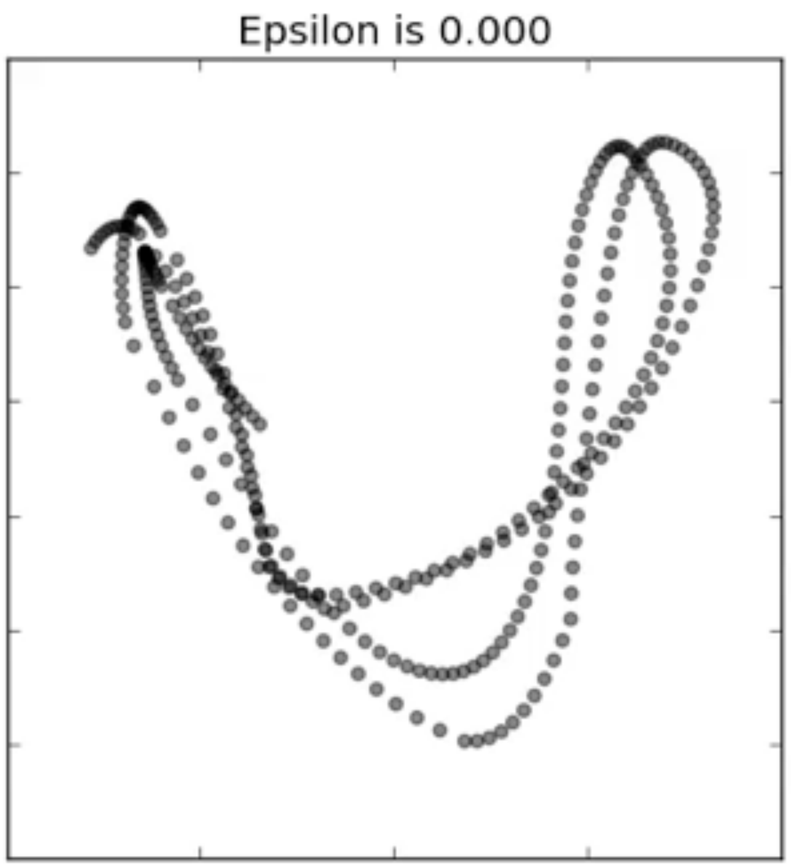
Persistent homology

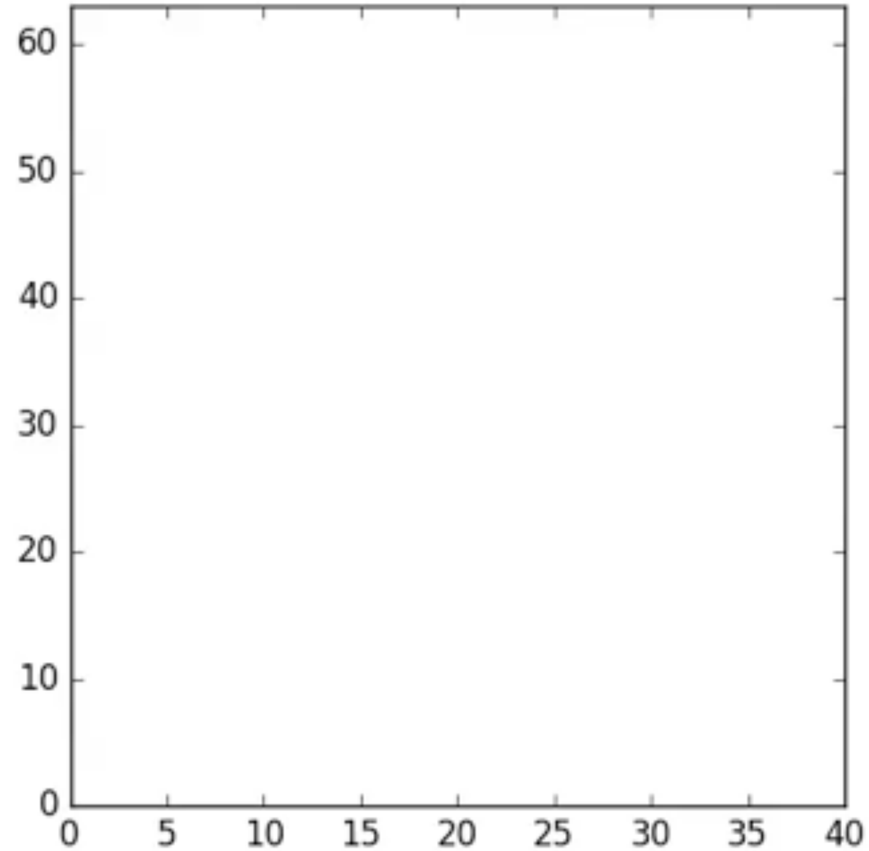
- Category theory: Homology is a **functor**.
  Sequence of spaces $X_1 \to X_2 \to X_3 \to X_4$.

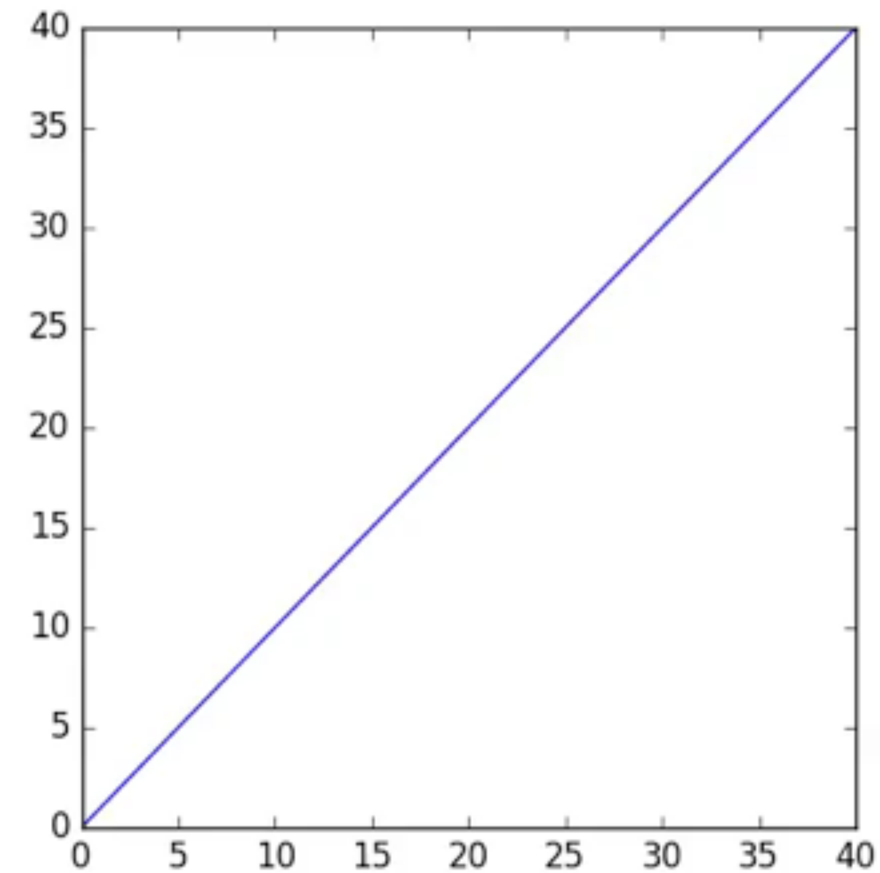
  Maps to sequence of vector spaces **with induced linear maps**:
  $H_k X_1 \to H_k X_2 \to H_k X_3 \to H_k X_4$.

- Module theory (over $\mathbb{k}[t]$ or over an A_n quiver or…):
  Sequence of vector spaces and linear maps decomposes into direct sum of **interval modules**.
  These are defined by **birth index** and **death index**.

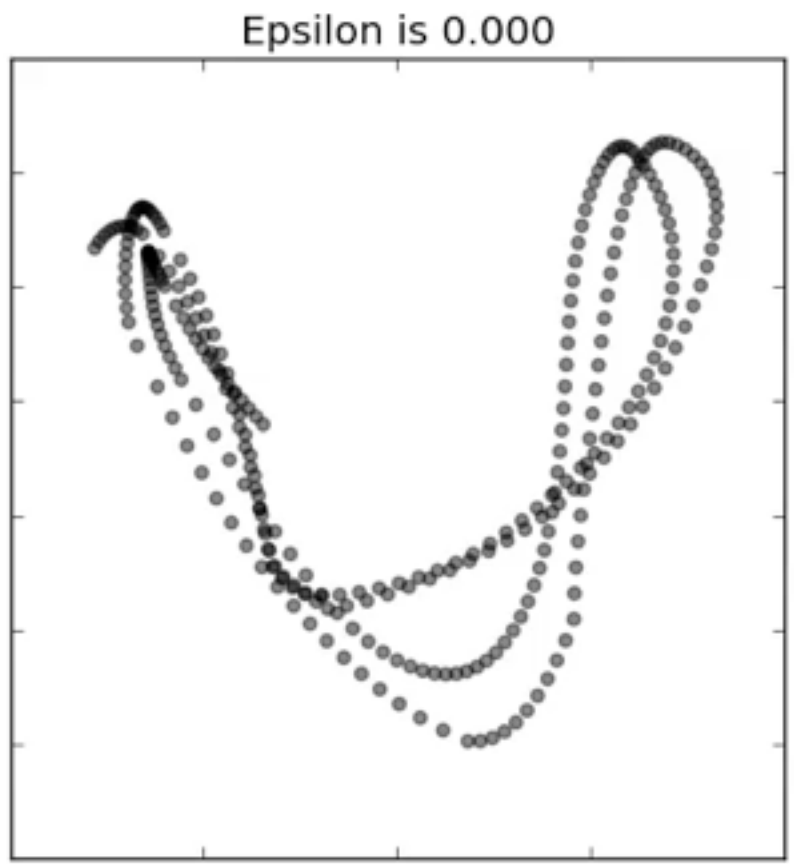
Persistent barcodes
Persistent diagrams

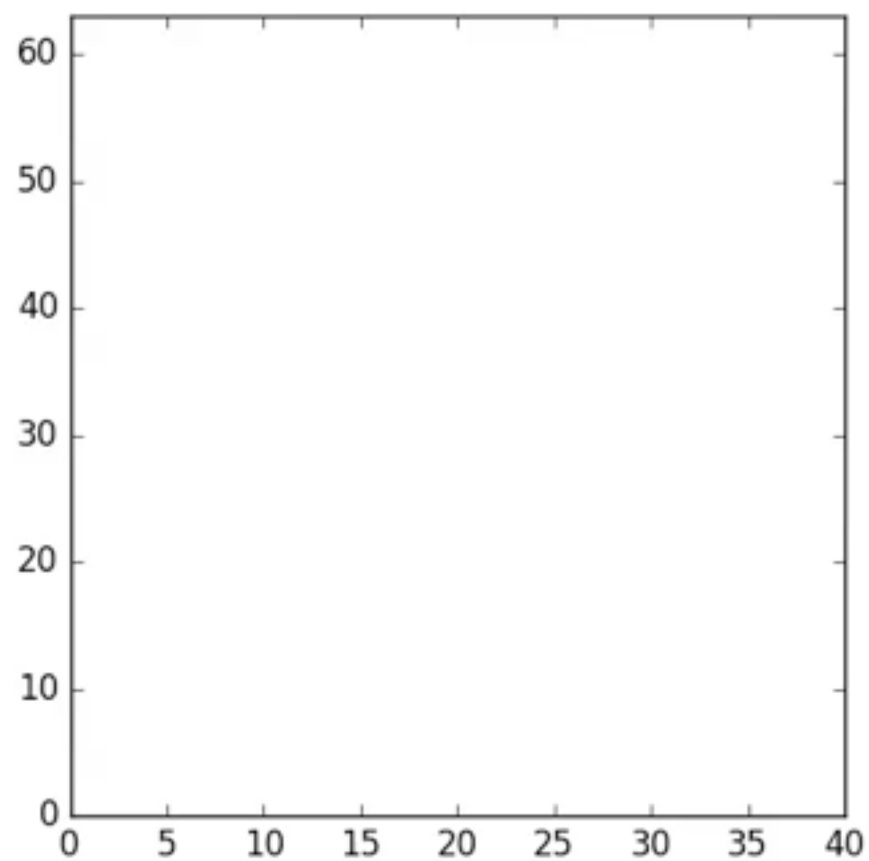


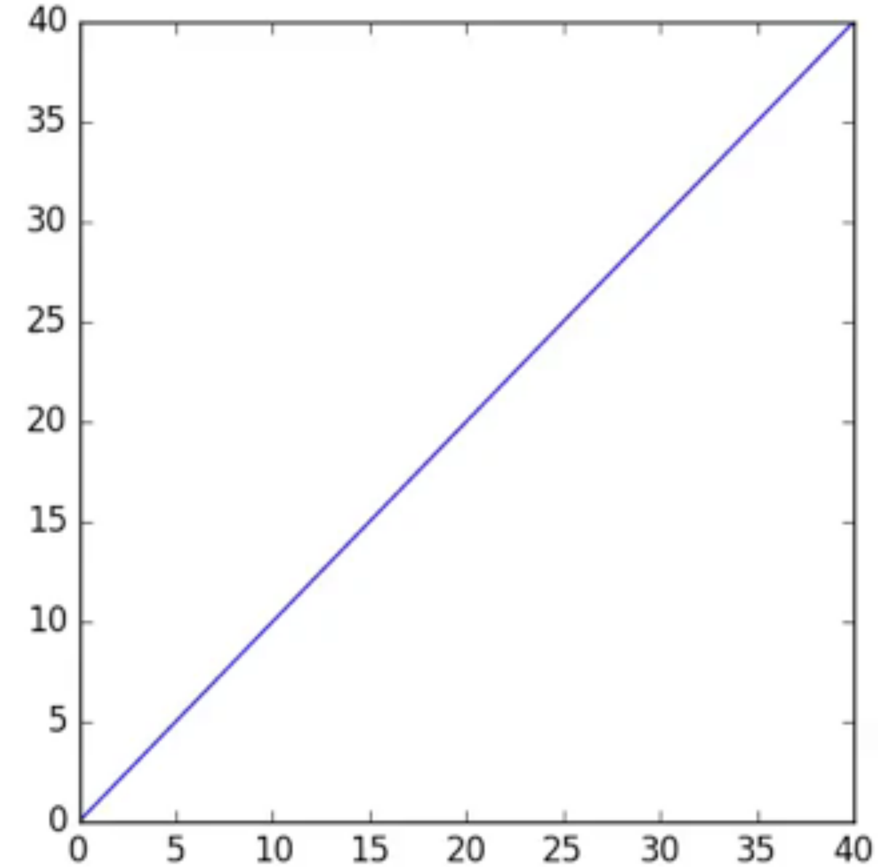




Persistent barcodes
Persistent diagrams







Outline

- Data has shape
- Homology: linear algebra measures shape
- Persistence: squinting with mathematics
- **Cohomology**
- Applications

Cohomology: dualize everything

- Instead of chains C_n, use cochains $C^n = \mathrm{Hom}(C_n, \mathbb{k})$
- Instead of boundary ∂ use coboundary $\delta = \partial^T$
- Cocycles: follow a generalized path invariance:
  $f(x,y) + f(y,z) = f(x,z)$ for any paths $x \to y$, $x \to z$, $y \to z$
- Coboundaries: path invariance follows from a generalized potential function construction:
  $f = g(y) - g(x)$

1-Cohomology is Circle-valued functions

- Fun fact:
  $H^1(X; \mathbb{Z})$ is bijective with homotopy equivalence classes of functions $X \to S^1$

- Construction for $[f] \in H^1(X; \mathbb{Z})$:

  - Send vertices to 0

  - Use f(e) as a wrapping number: wrap e around the circle f(e) times

  - All higher simplices work out bc path invariance

1-Cohomology is Circle-valued functions

- Construction adapts to data:

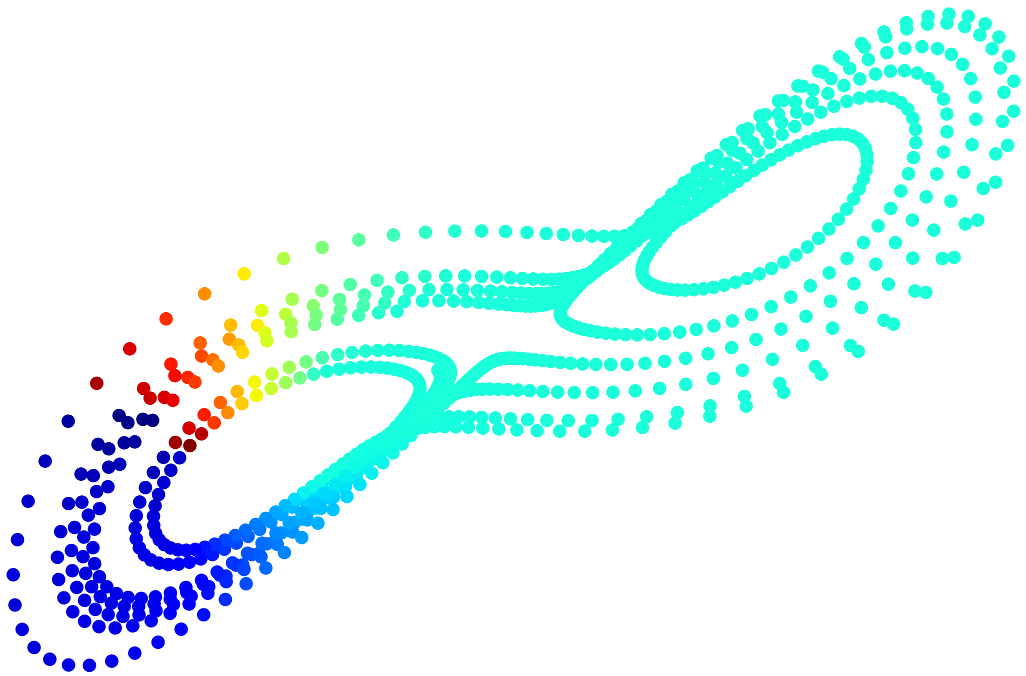
  Pretend [f] is instead from $H^1(X; \mathbb{R})$.

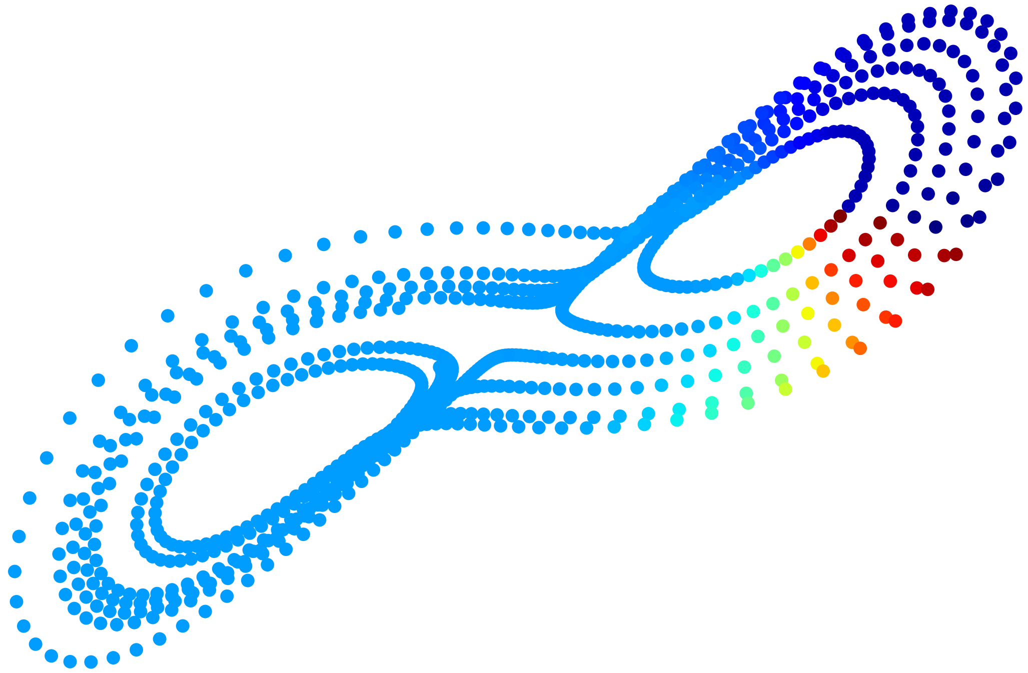
  $\arg\min_g | f - \delta g |_2 \bmod 1.0$
  is a smooth function $C_0 \to S^1$.

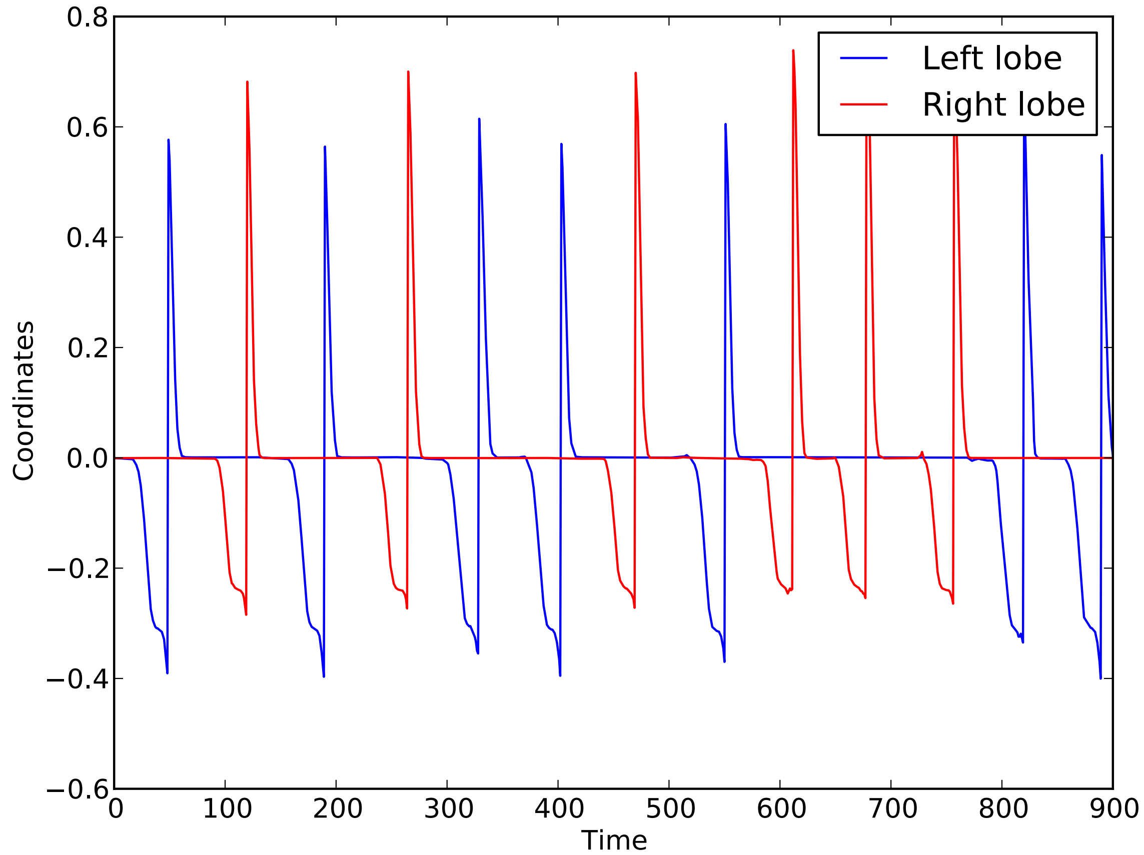
- Can compute intrinsic phase variables from geometry of time series

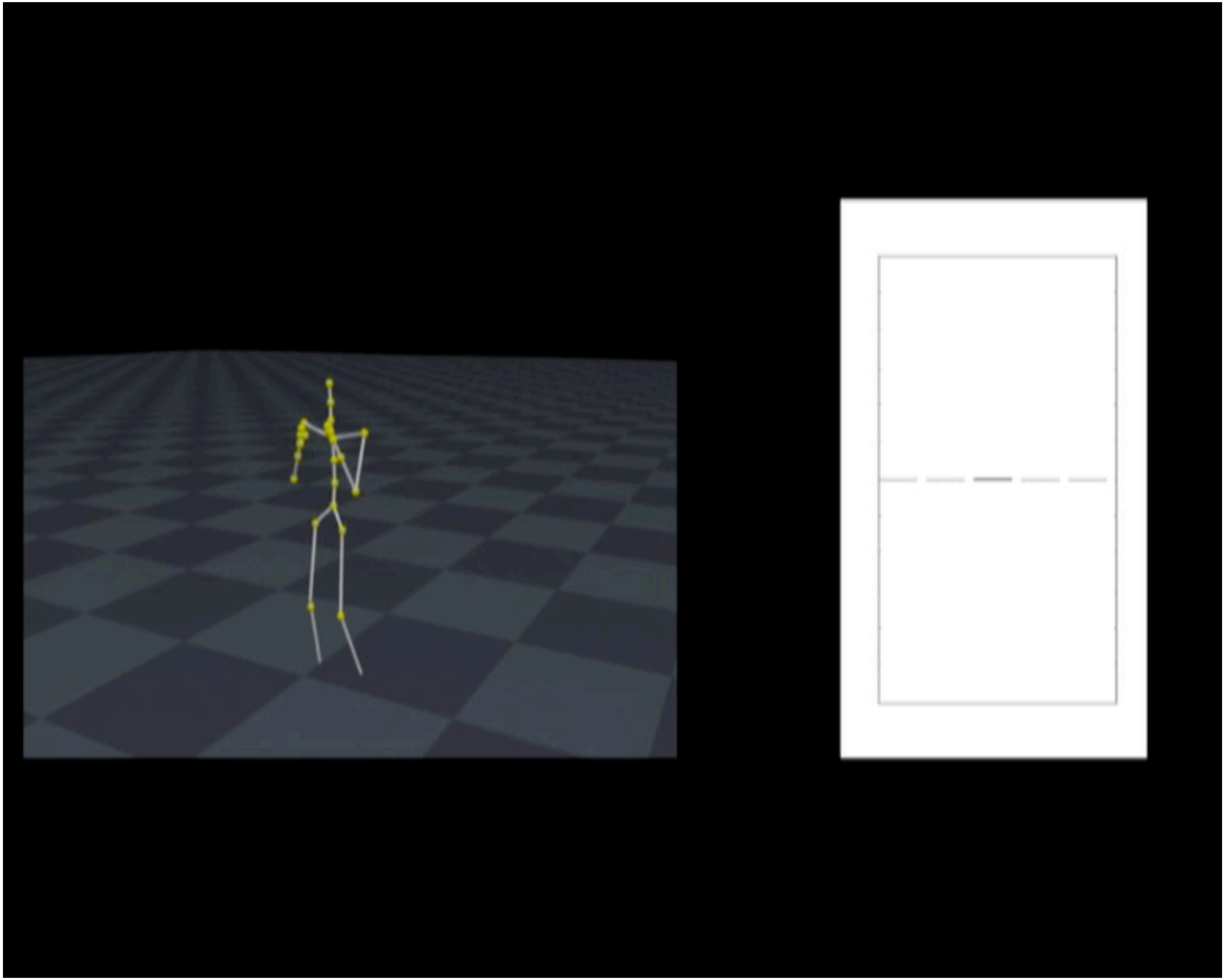
Outline

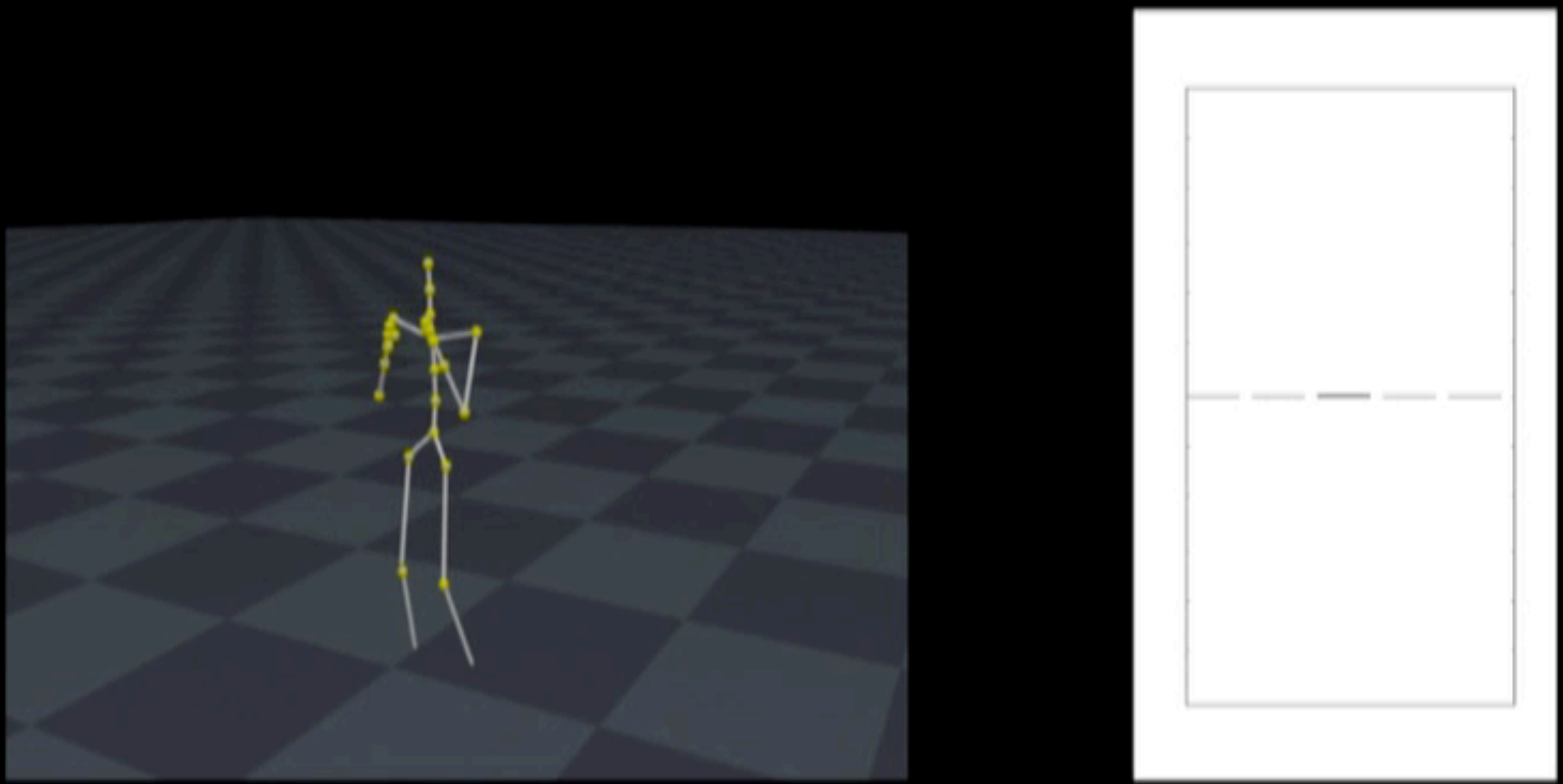
- Data has shape
- Homology: linear algebra measures shape
- Persistence: squinting with mathematics
- Cohomology
- **Applications**

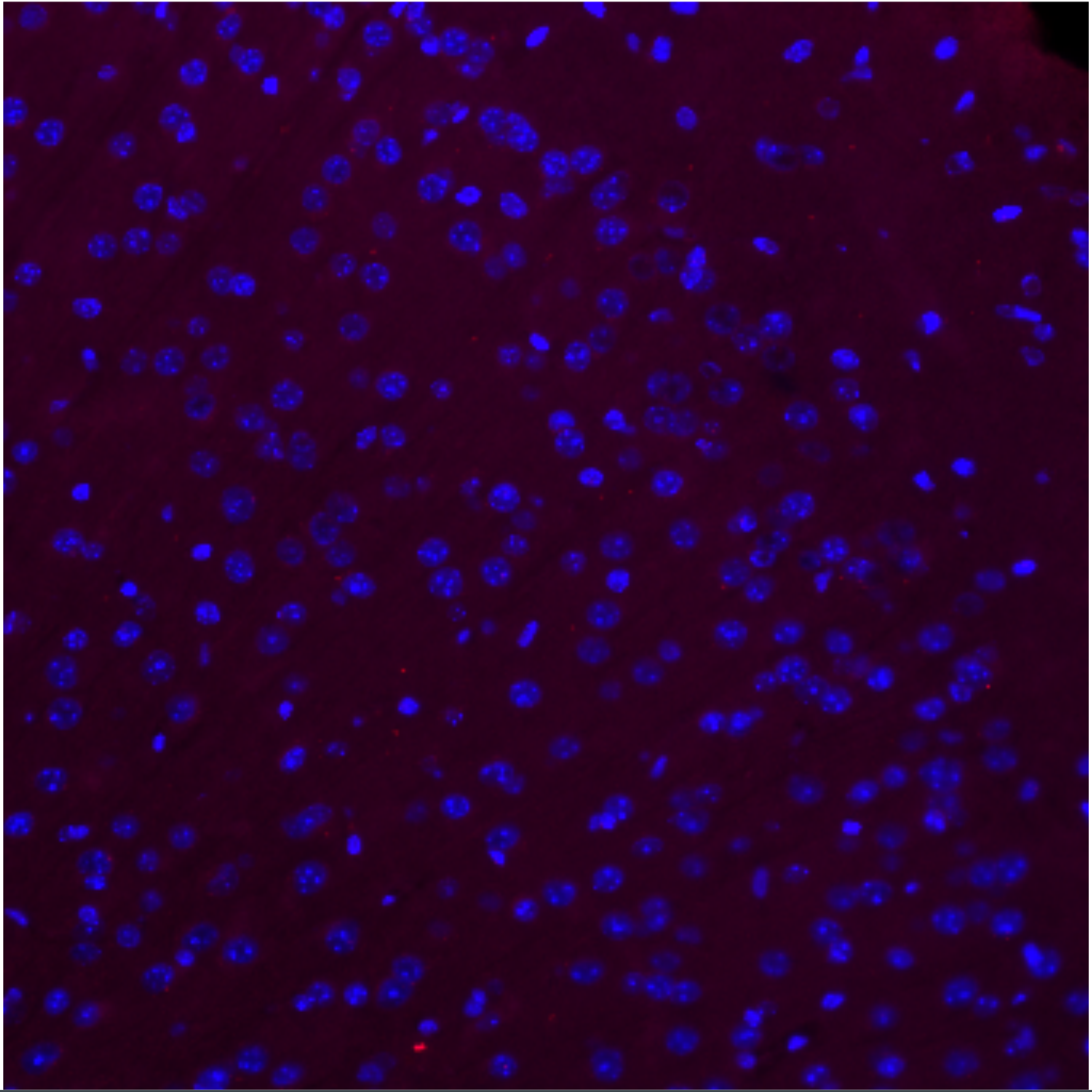


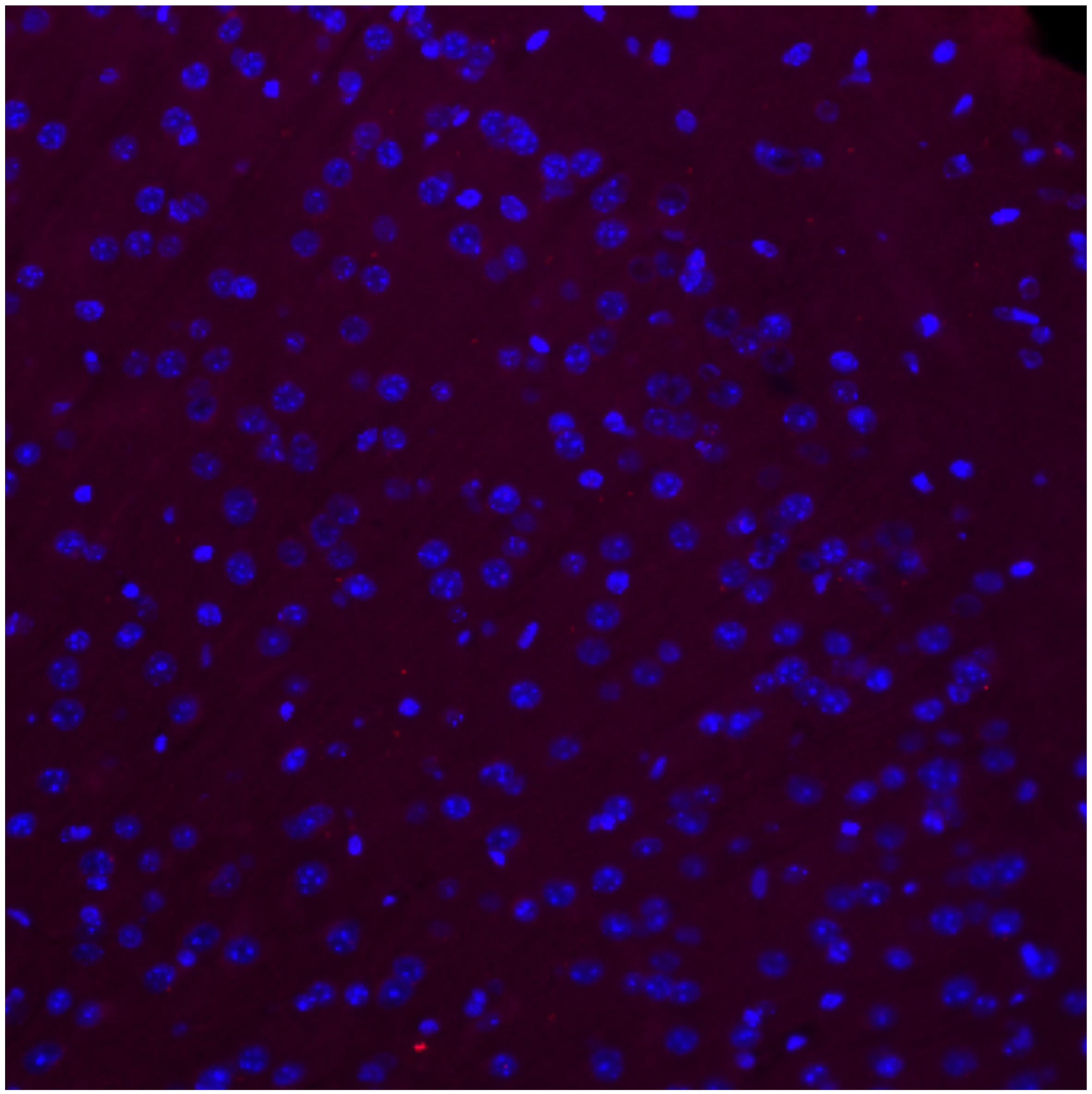


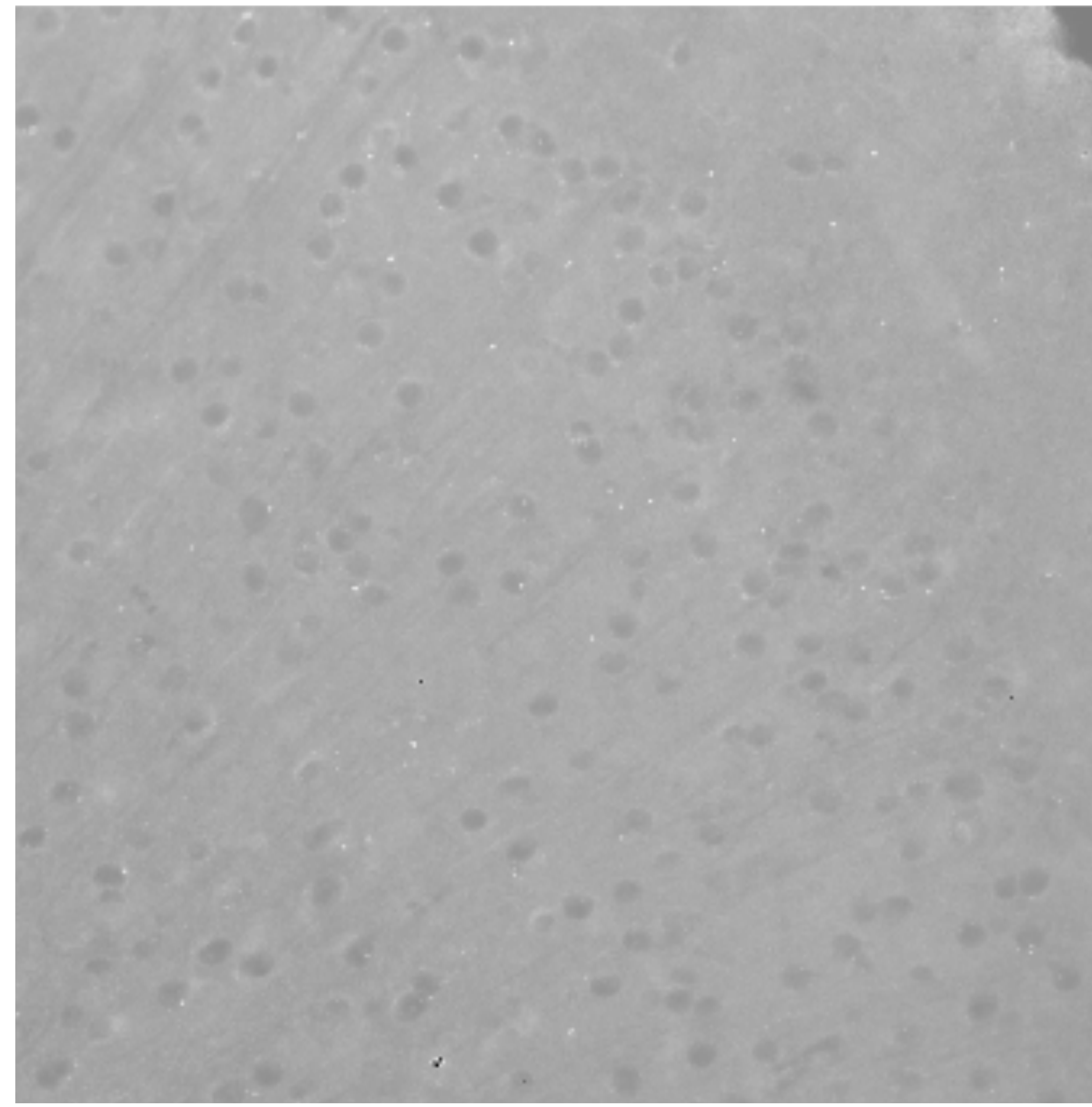

Left lobe
Right lobe
Coordinates
Time
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
0
100
200
300
400
500
600
700
800
900

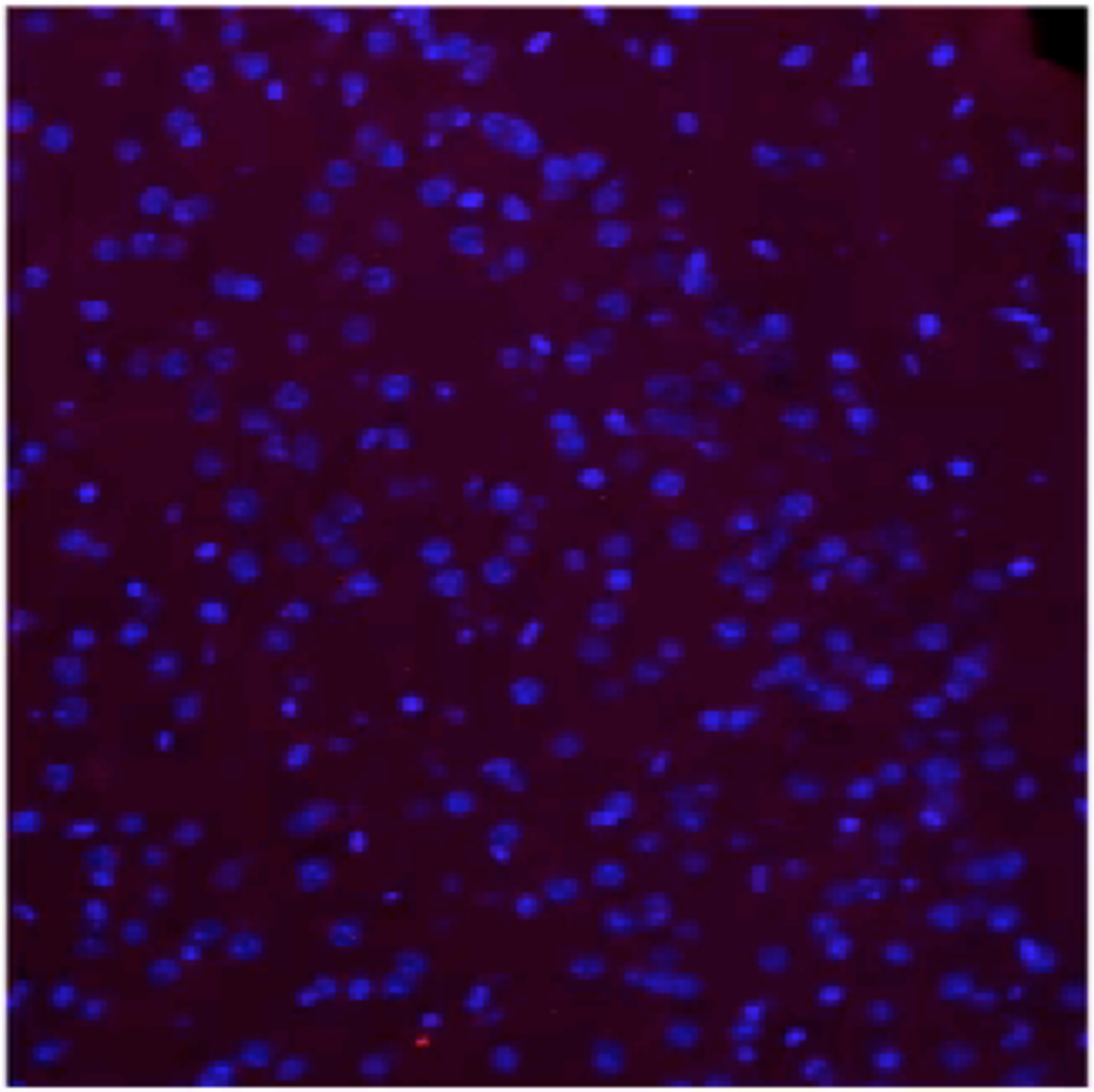


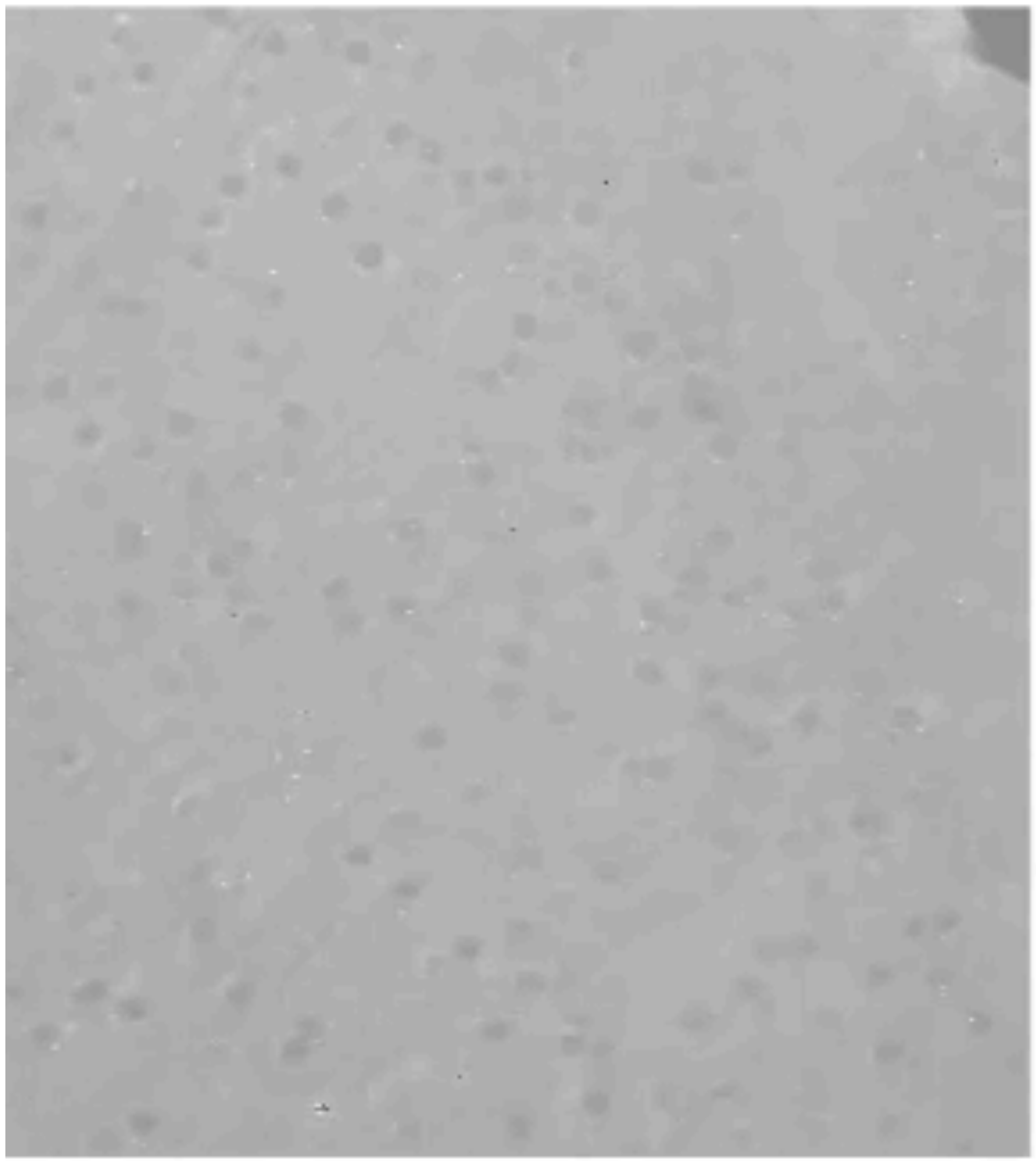


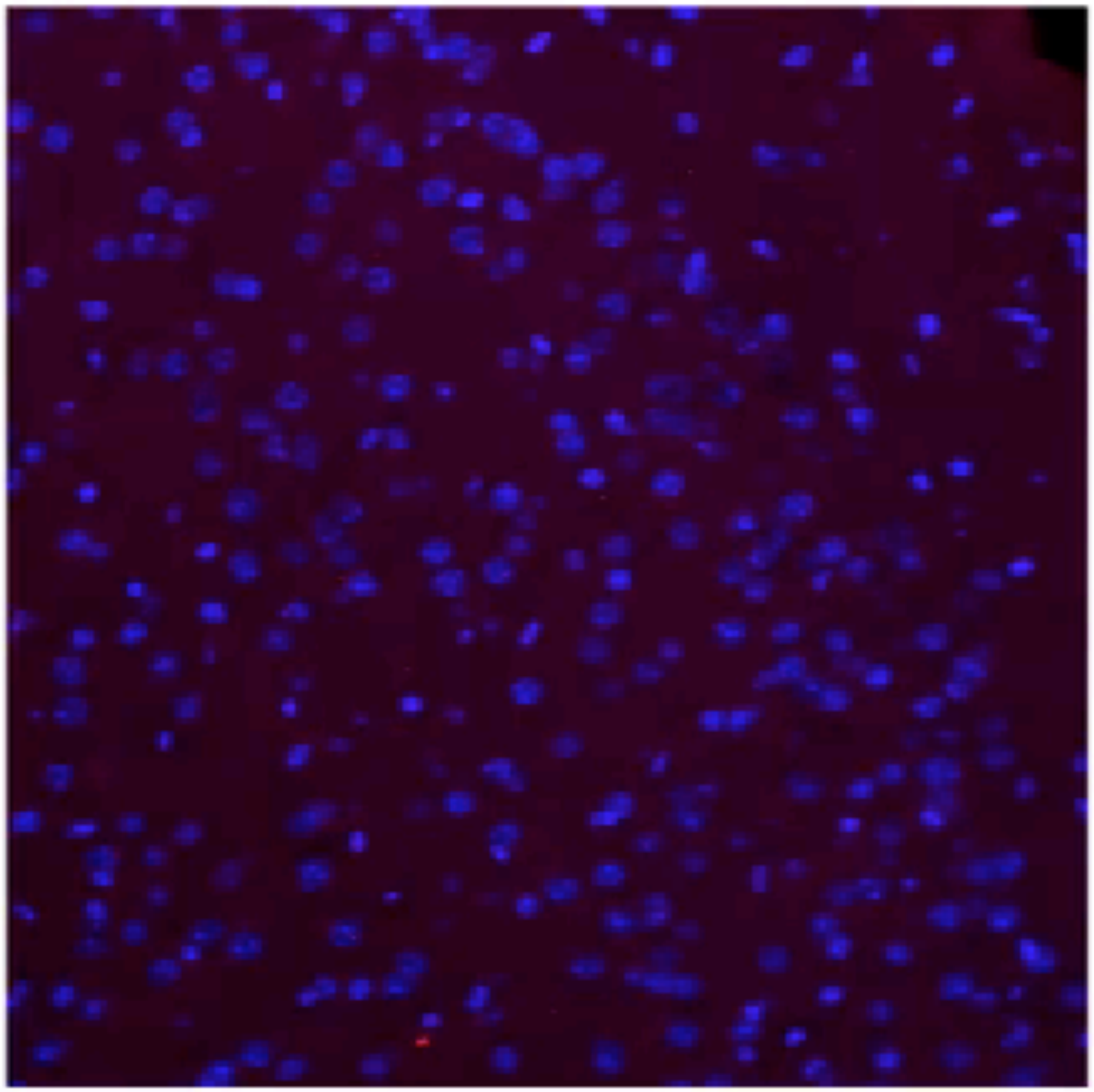


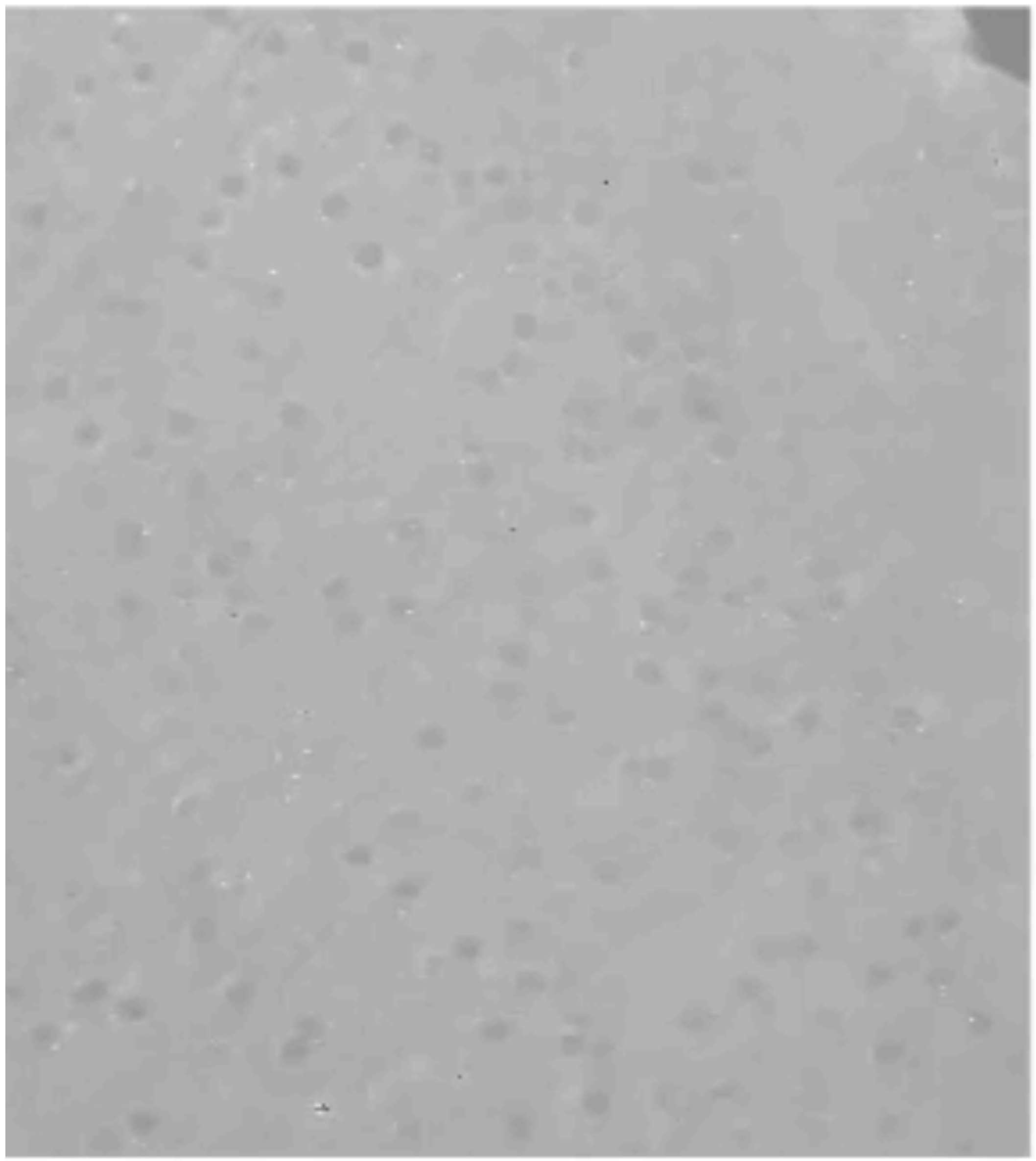


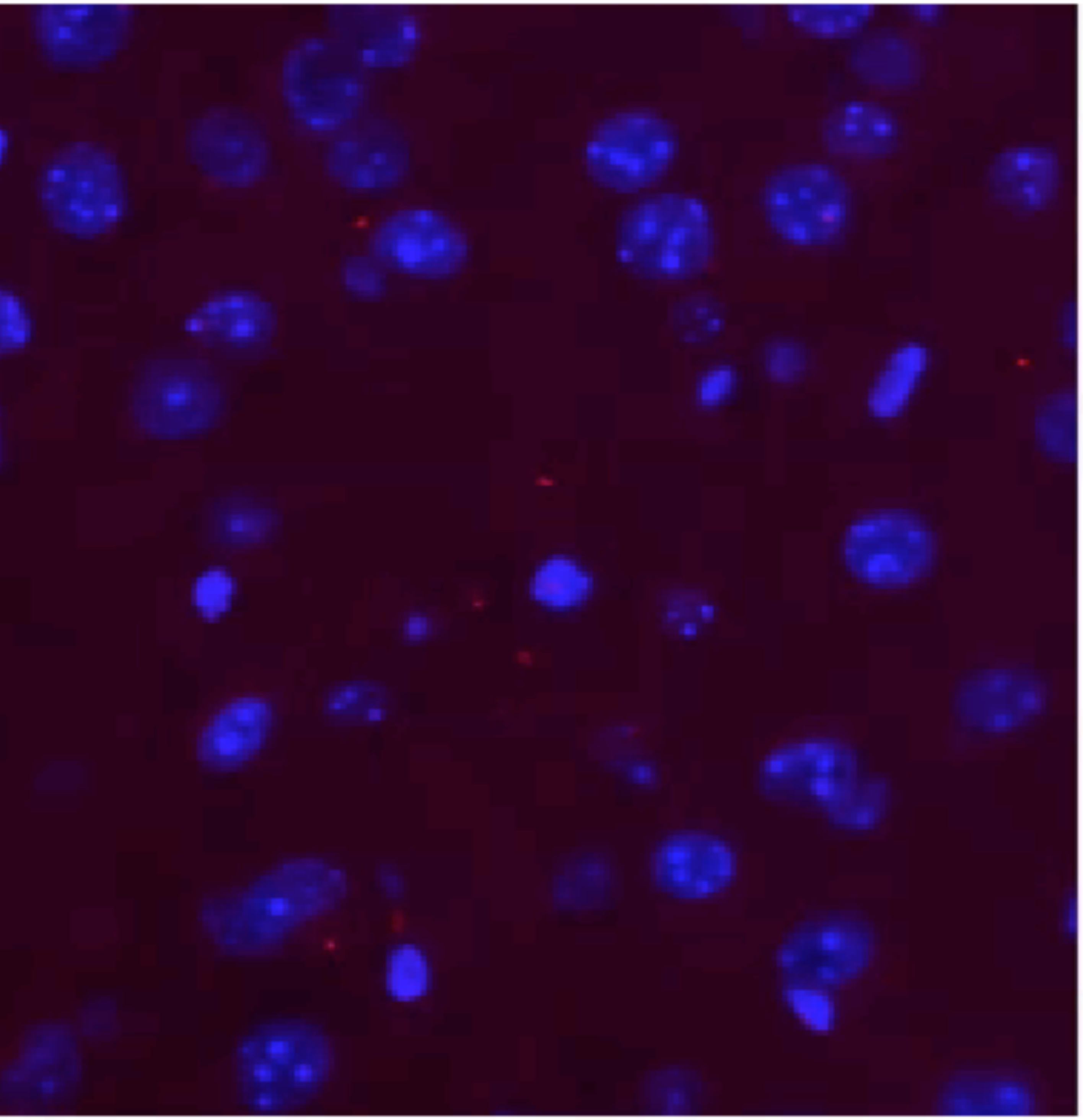


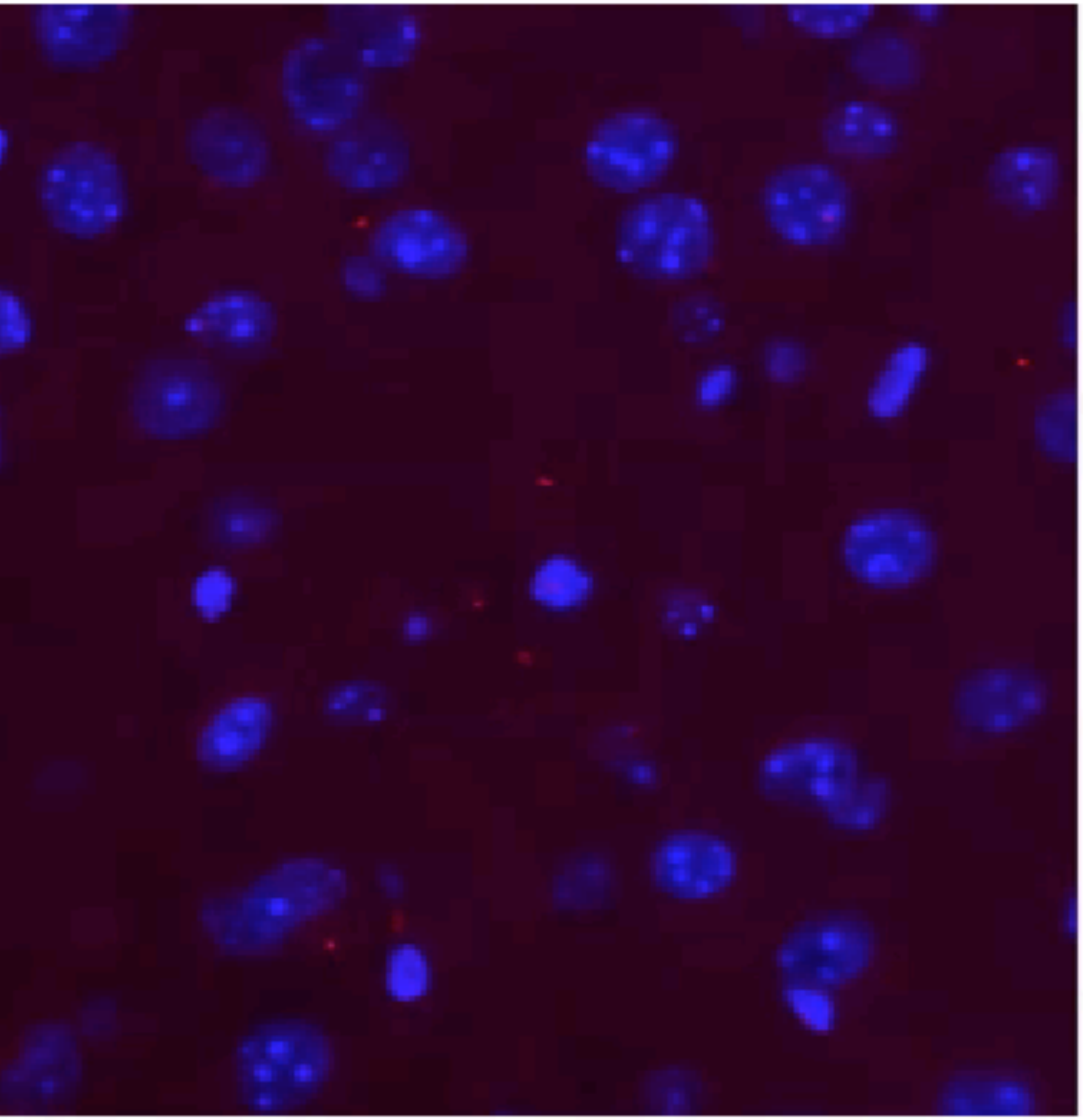


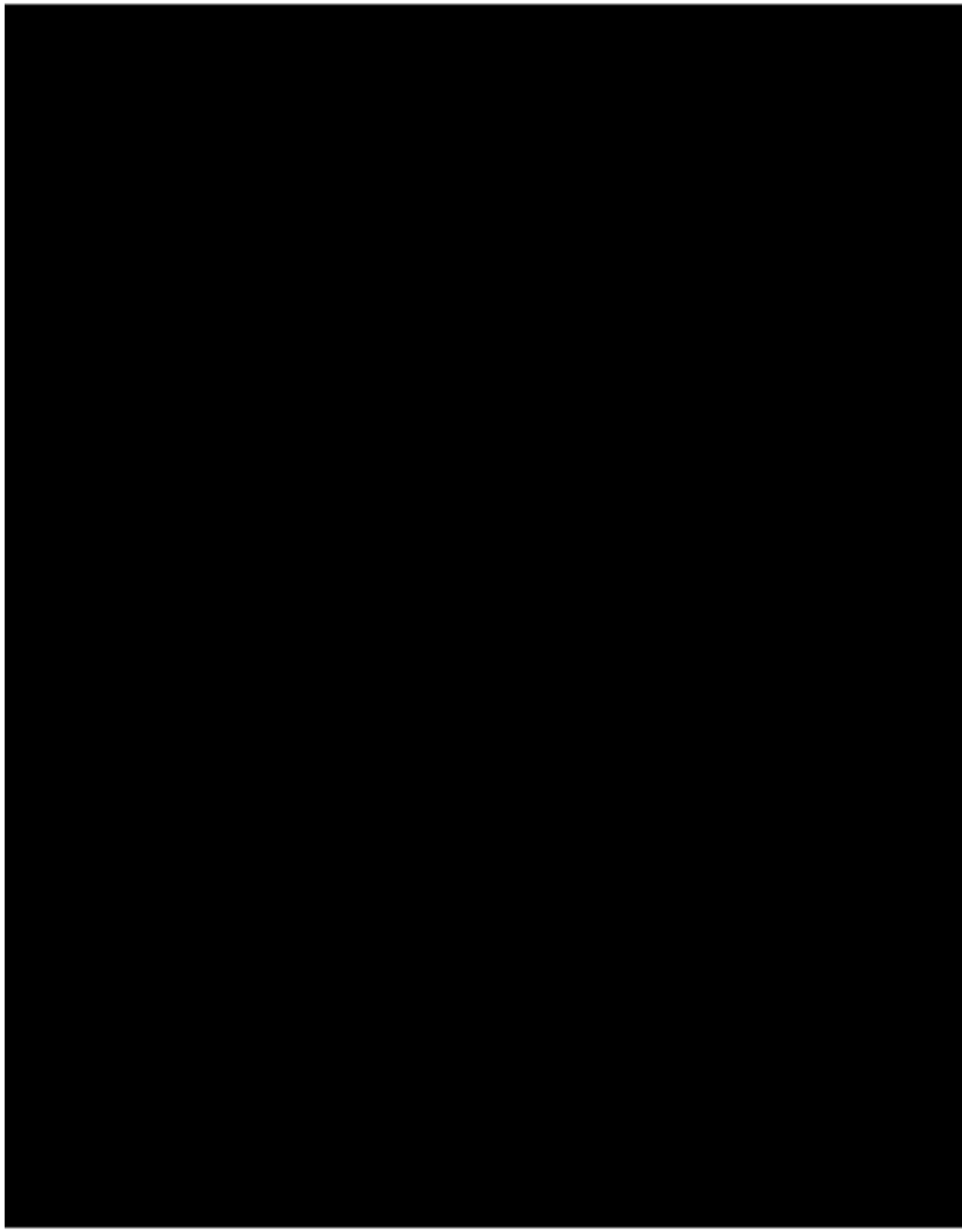


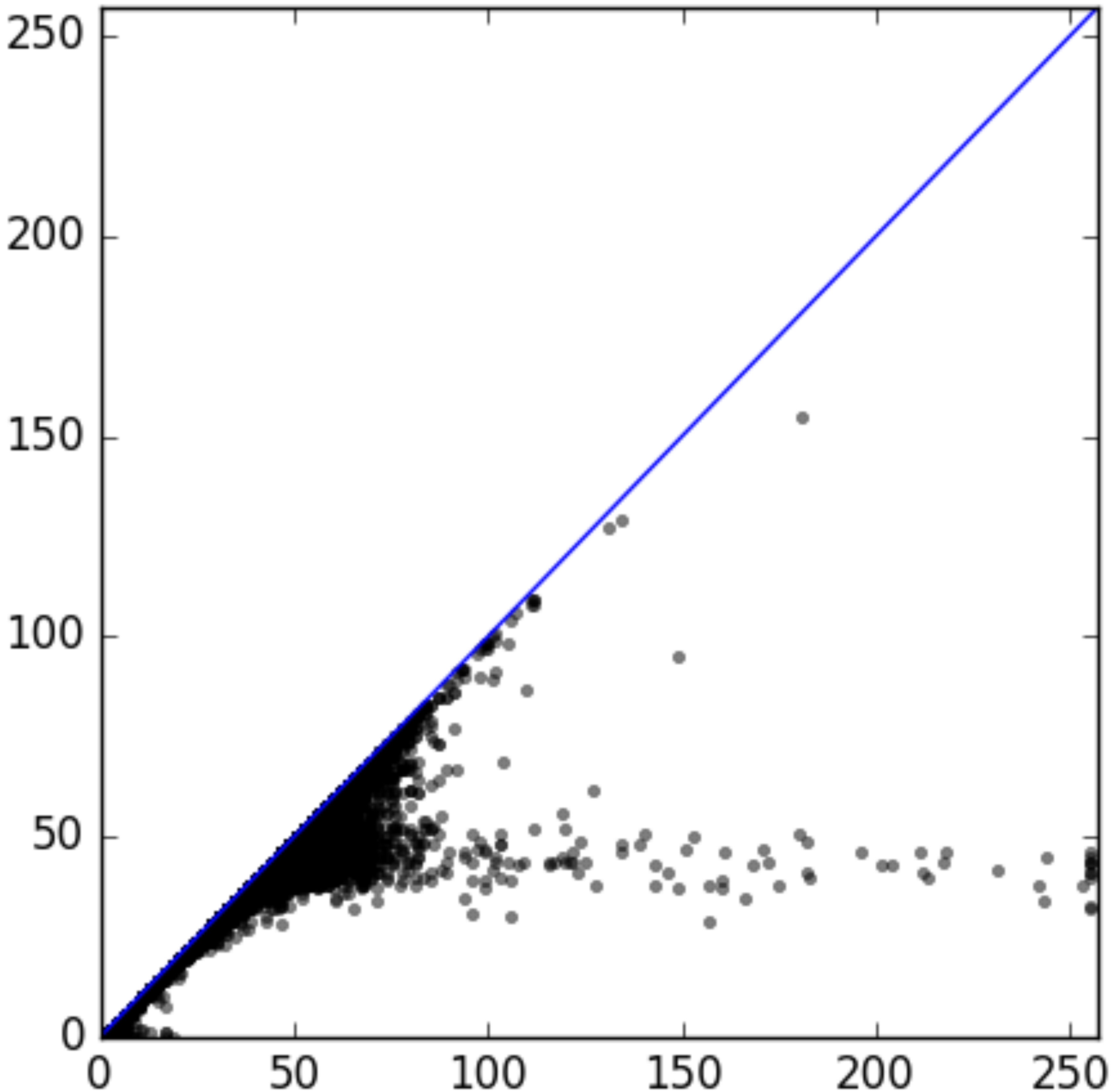


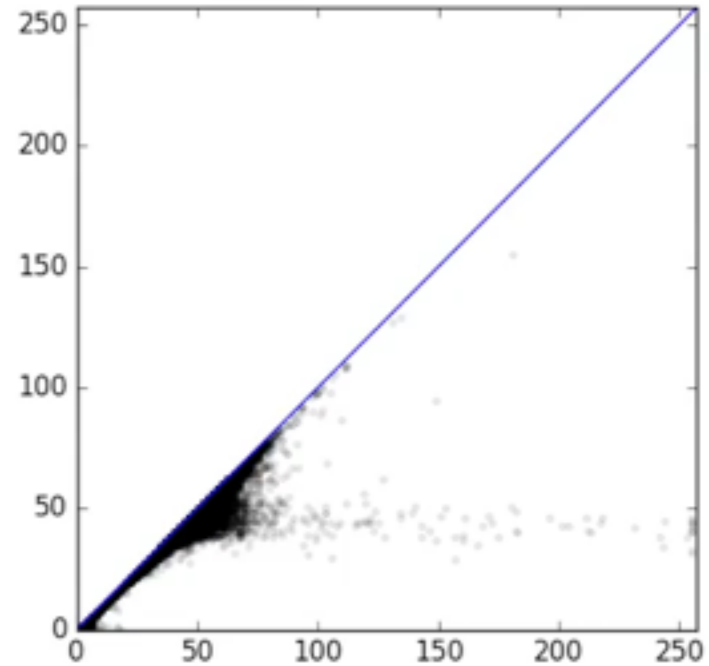


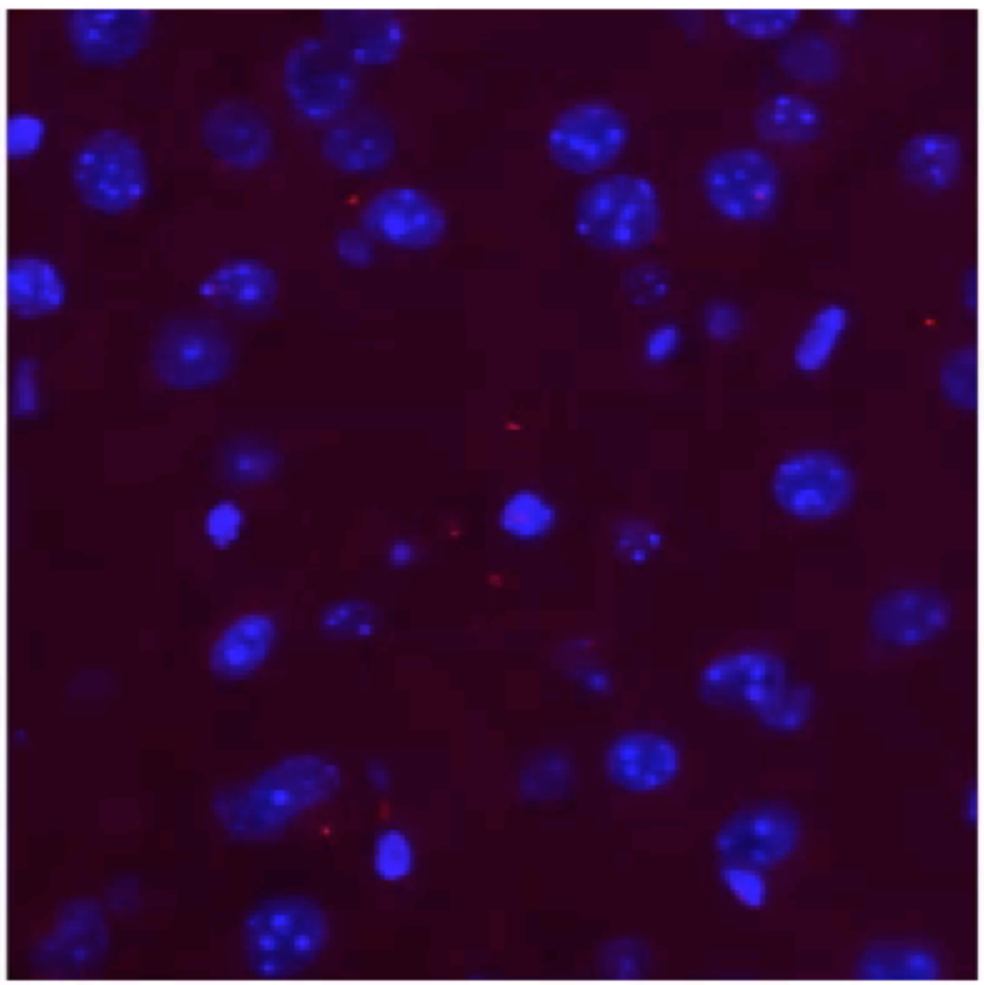


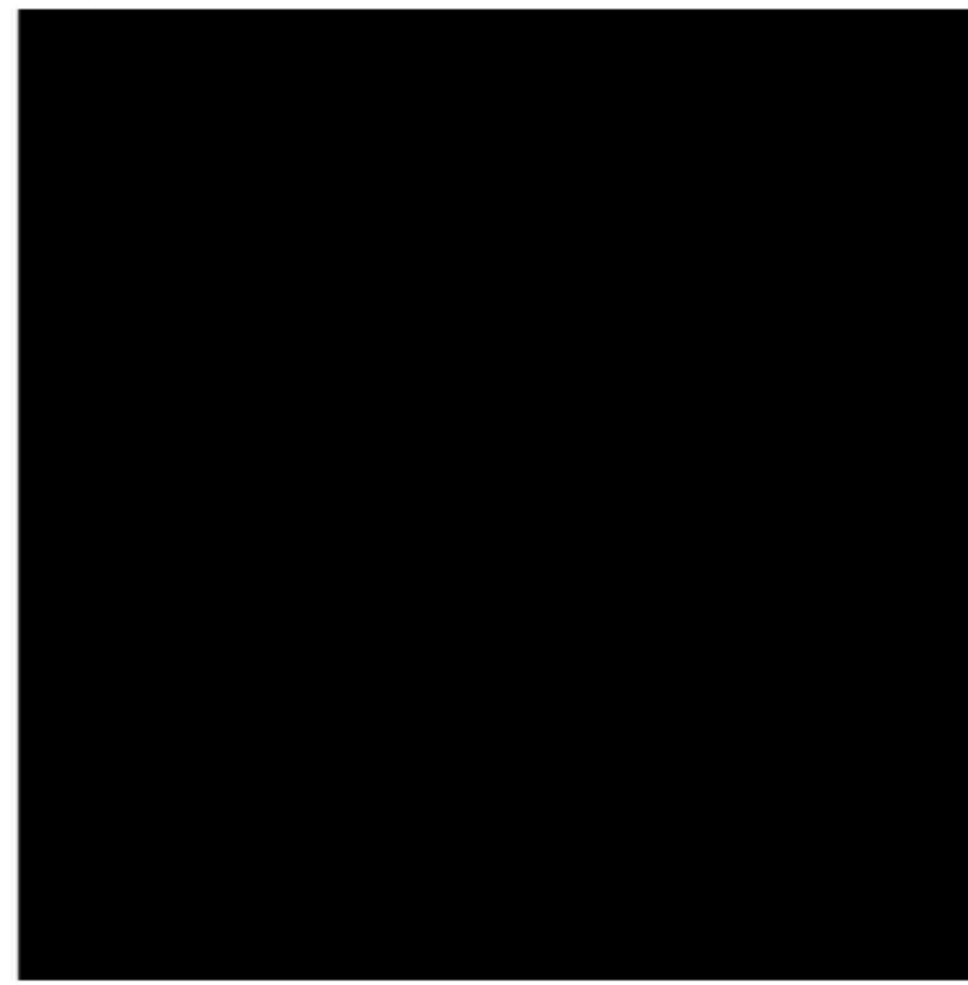


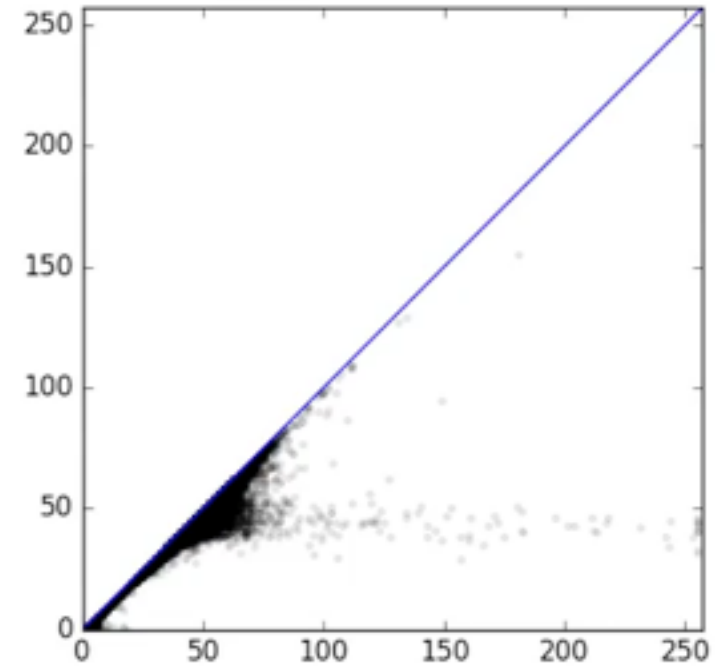



250
200
150
100
50
0
0
50
100
150
200
250


250
200
150
100
50
0
0
50
100
150
200
250






250
200
150
100
50
0
0
50
100
150
200
250