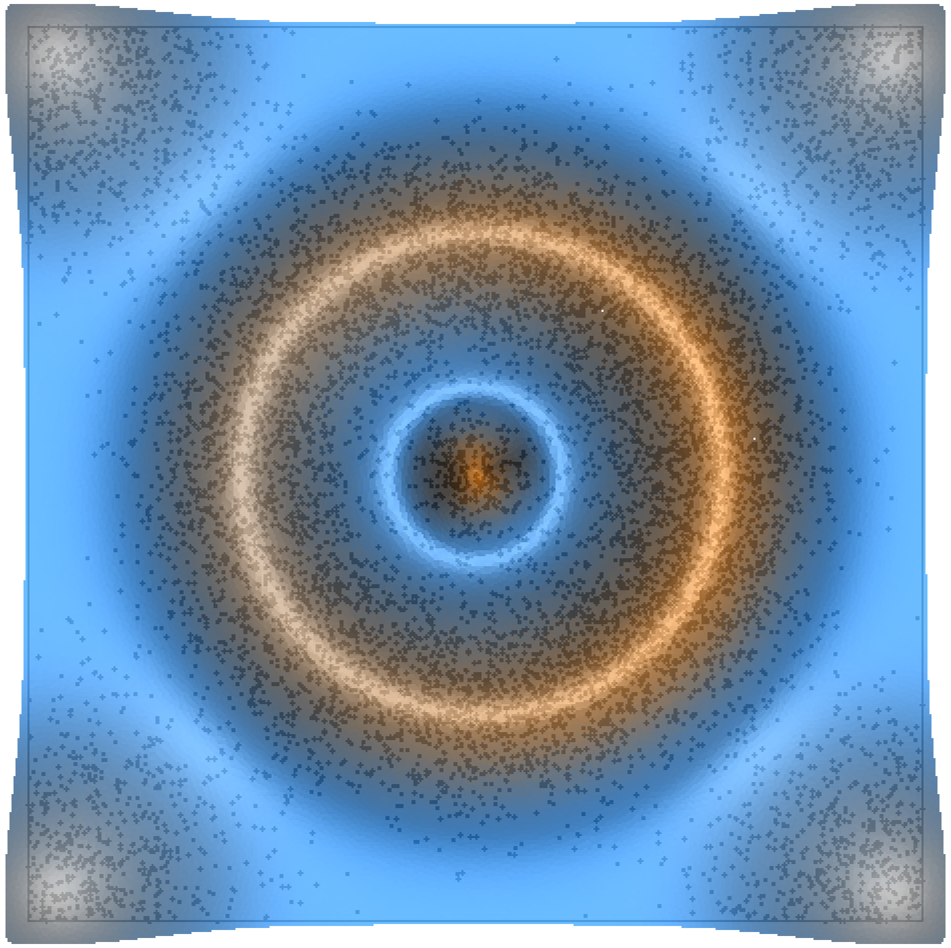
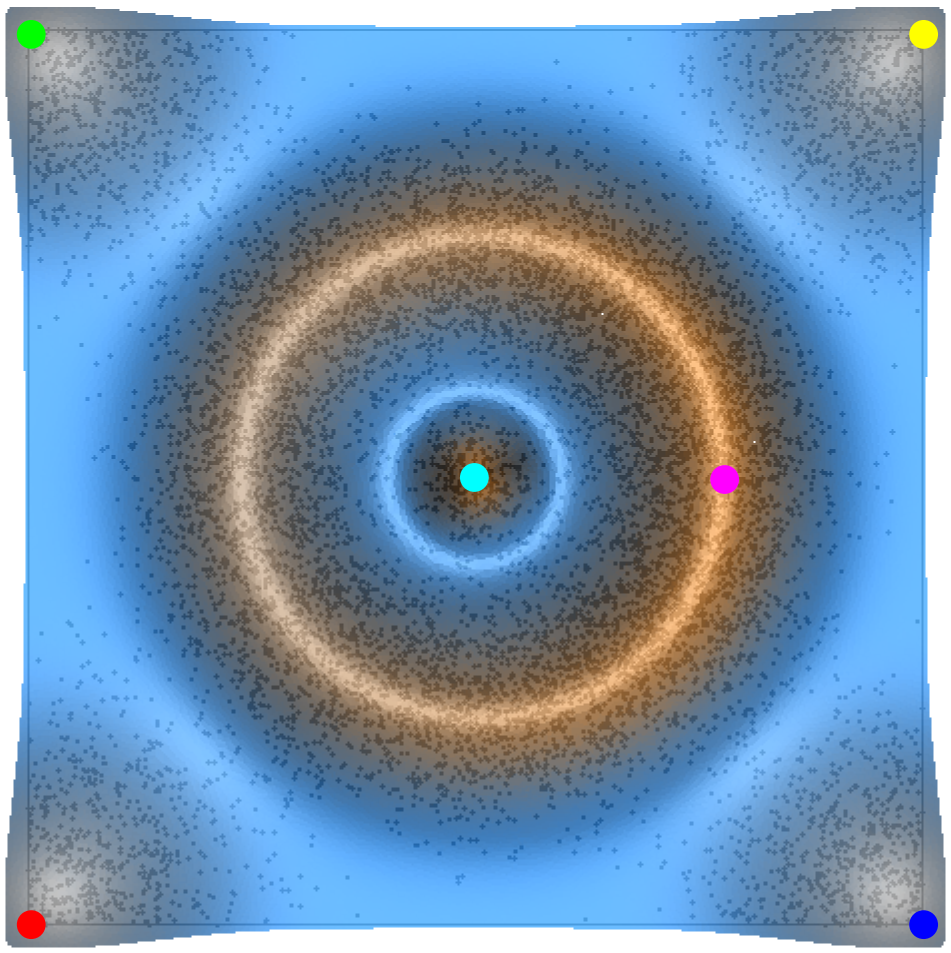
Back to the Mode-Seeking Paradigm

- Assume the data points are sampled from some unknown probability distribution
- Partition the data according to the basins of attraction of the peaks of the density



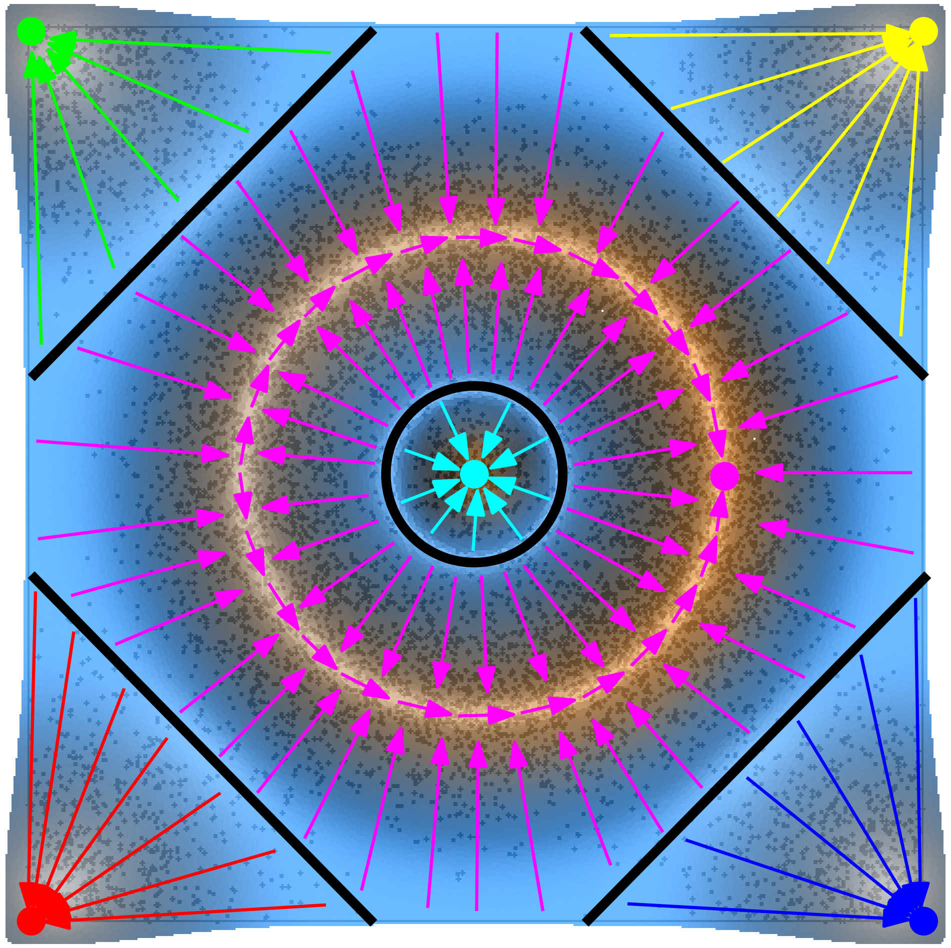
Back to the Mode-Seeking Paradigm

- Assume the data points are sampled from some unknown probability distribution
- Partition the data according to the basins of attraction of the peaks of the density



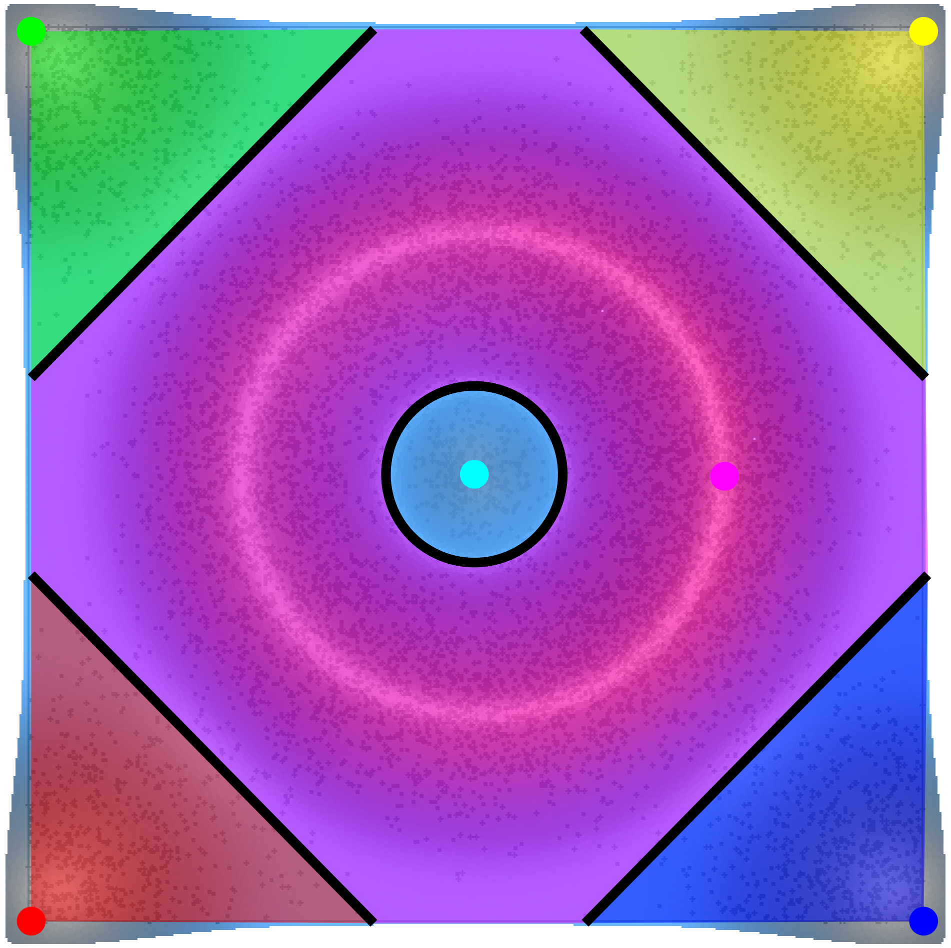
Back to the Mode-Seeking Paradigm

- Assume the data points are sampled from some unknown probability distribution
- Partition the data according to the basins of attraction of the peaks of the density



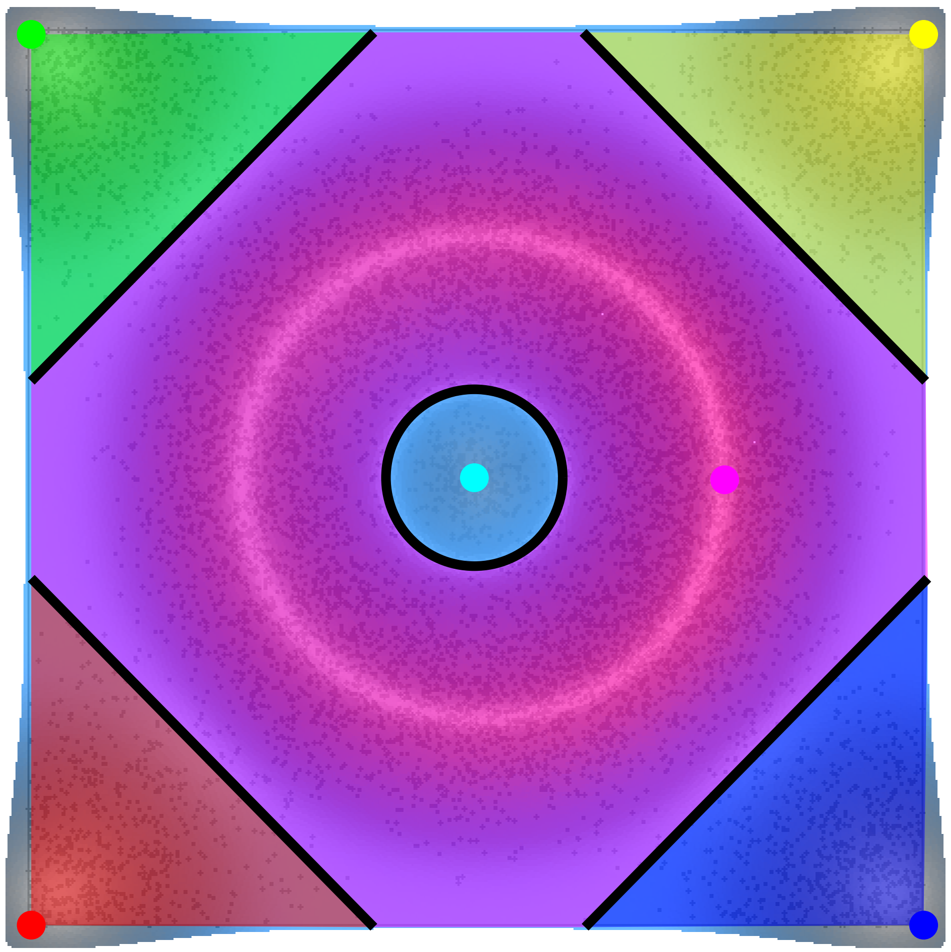
Back to the Mode-Seeking Paradigm

- Assume the data points are sampled from some unknown probability distribution
- Partition the data according to the basins of attraction of the peaks of the density



Back to the Mode-Seeking Paradigm

- Assume the data points are sampled from some $\boxed{\text{unknown}}$ probability distribution
- Partition the data according to the basins of attraction of the peaks of the density

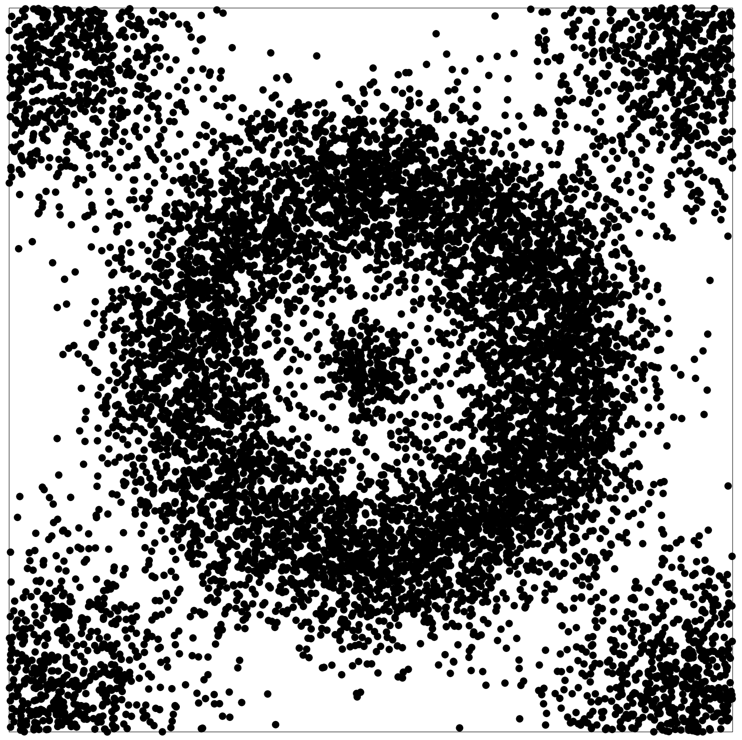


Hill-Climbing Schemes

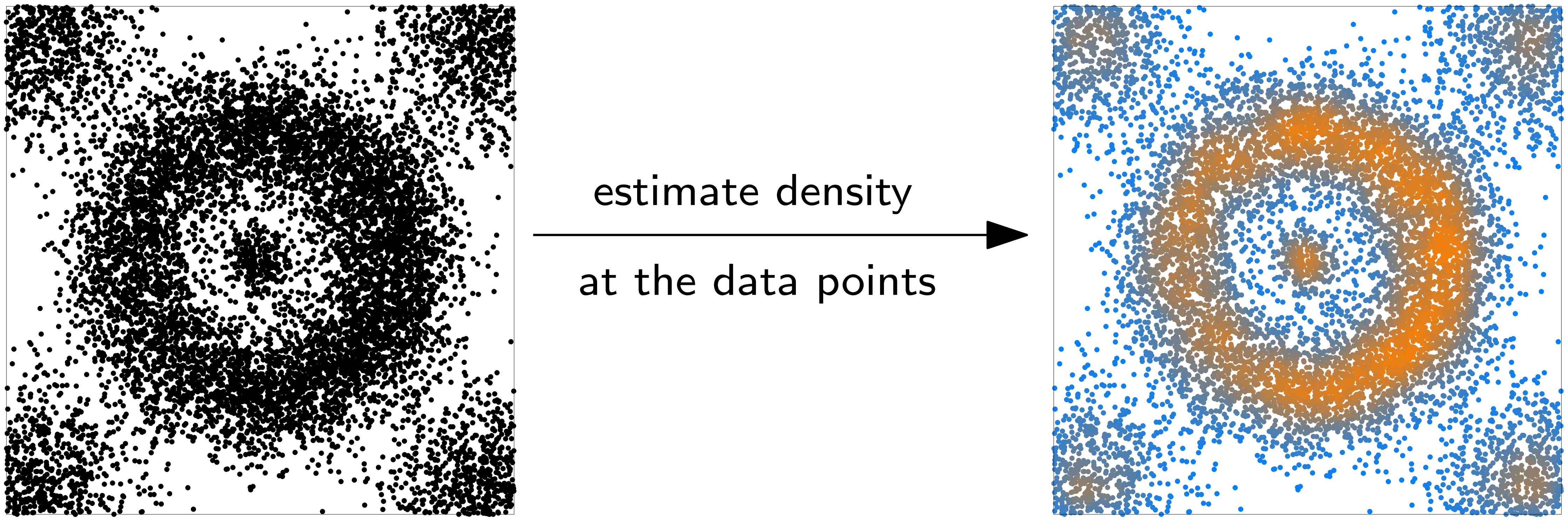
- **Iterative**, e.g. D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(5):603619, May 2002.

- **Non-iterative**, e.g. W. L. Koontz, P. M. Narendra, and K. Fukunaga. A graph-theoretic approach to nonparametric cluster analysis. *IEEE Trans. on Computers*, 24:936944, September 1976.

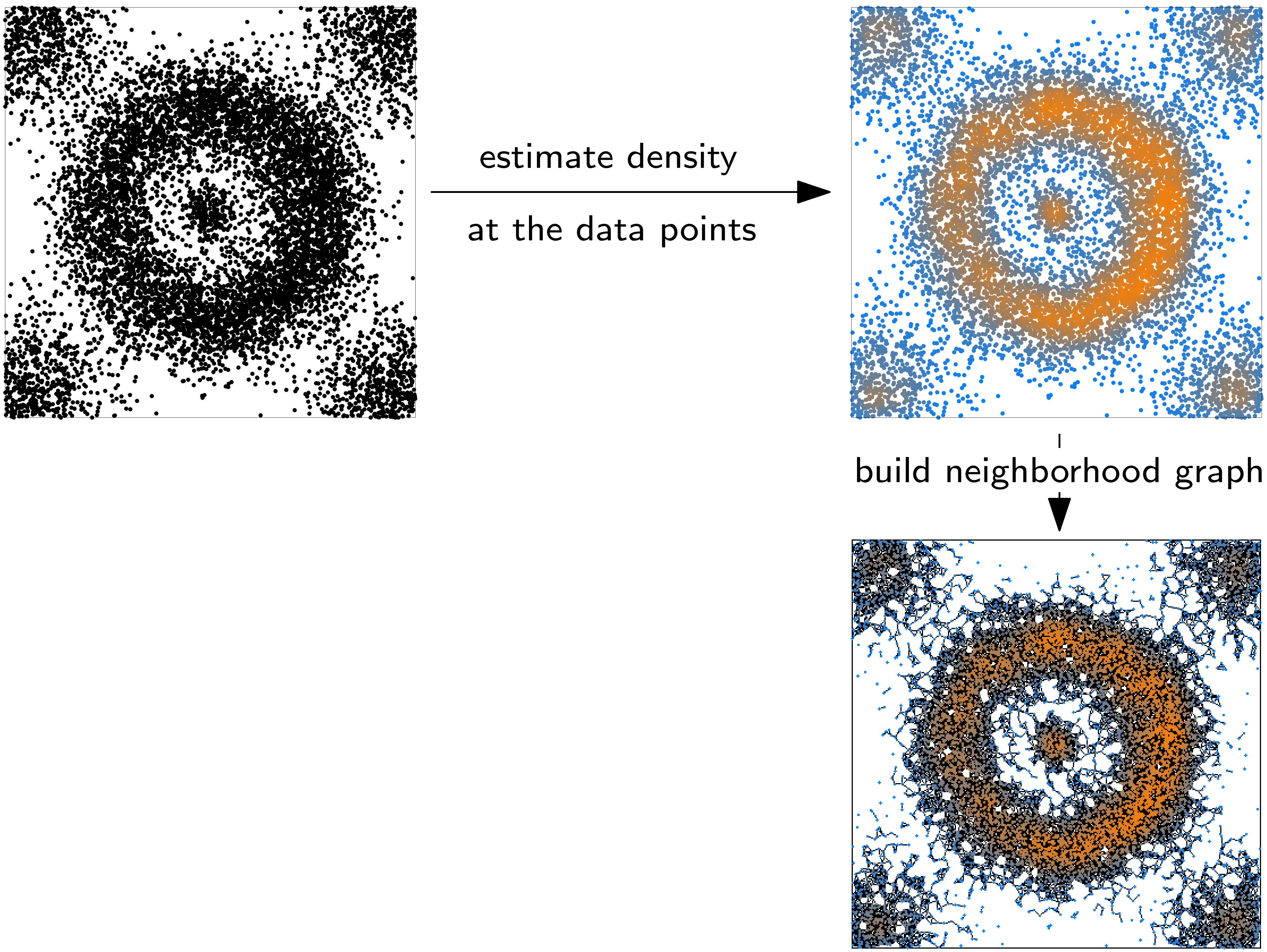
[Koontz, Narendra, Fukunaga'76] in a Nutshell



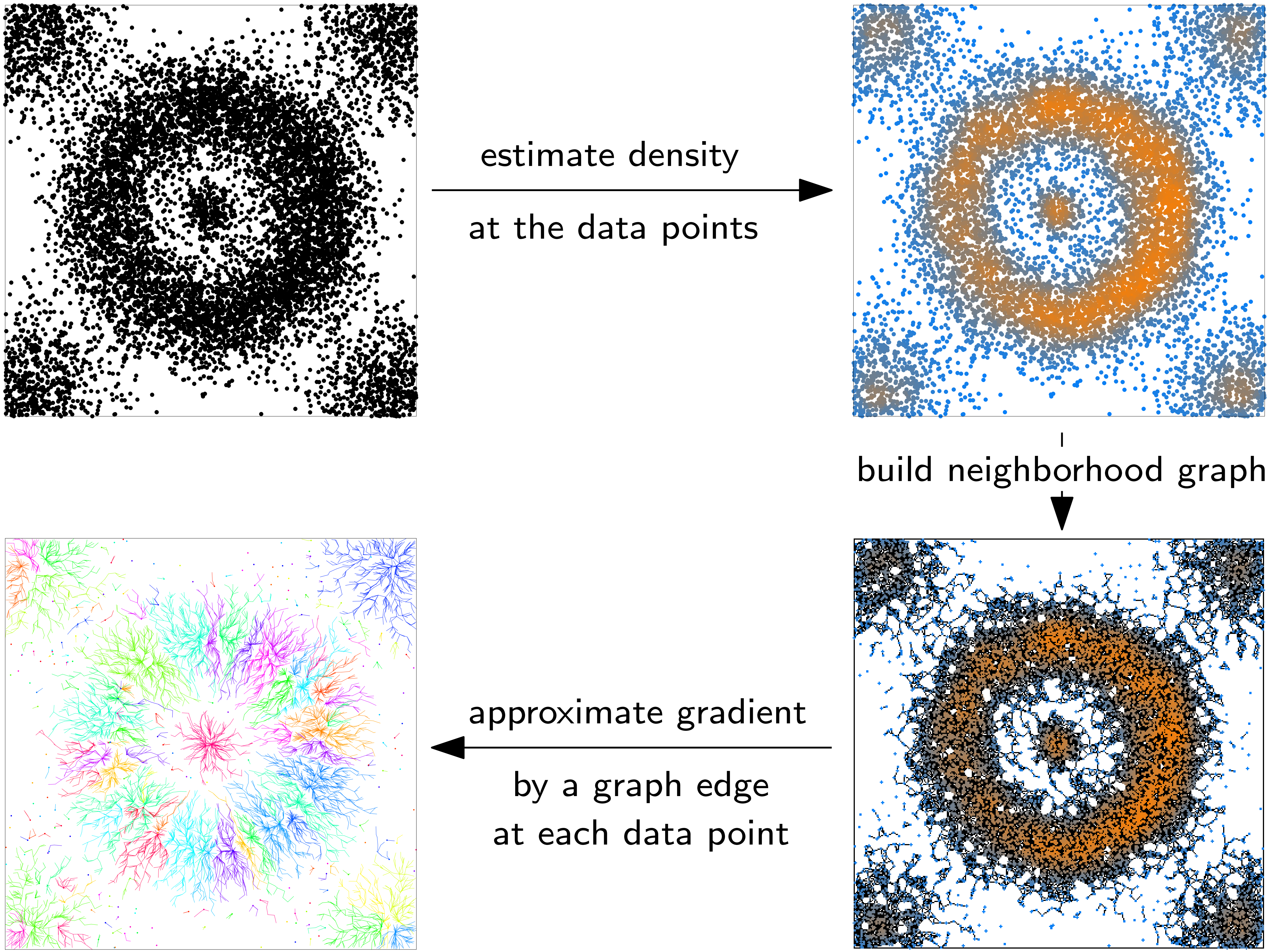
[Koontz, Narendra, Fukunaga'76] in a Nutshell



[Koontz, Narendra, Fukunaga'76] in a Nutshell



[Koontz, Narendra, Fukunaga'76] in a Nutshell



Pseudo-code:

Input: neighborhood graph G with n vertices, n-dimensional vector $\hat{f}$ (density estimator)

Sort the vertex indices $\{1, 2, \cdots, n\}$ so that $\hat{f}(1) \geq \hat{f}(2) \geq \cdots \geq \hat{f}(n)$;
Initialize a union-find data structure (disjoint-set forest) $\mathcal{U}$ and two vectors g, r of size n;

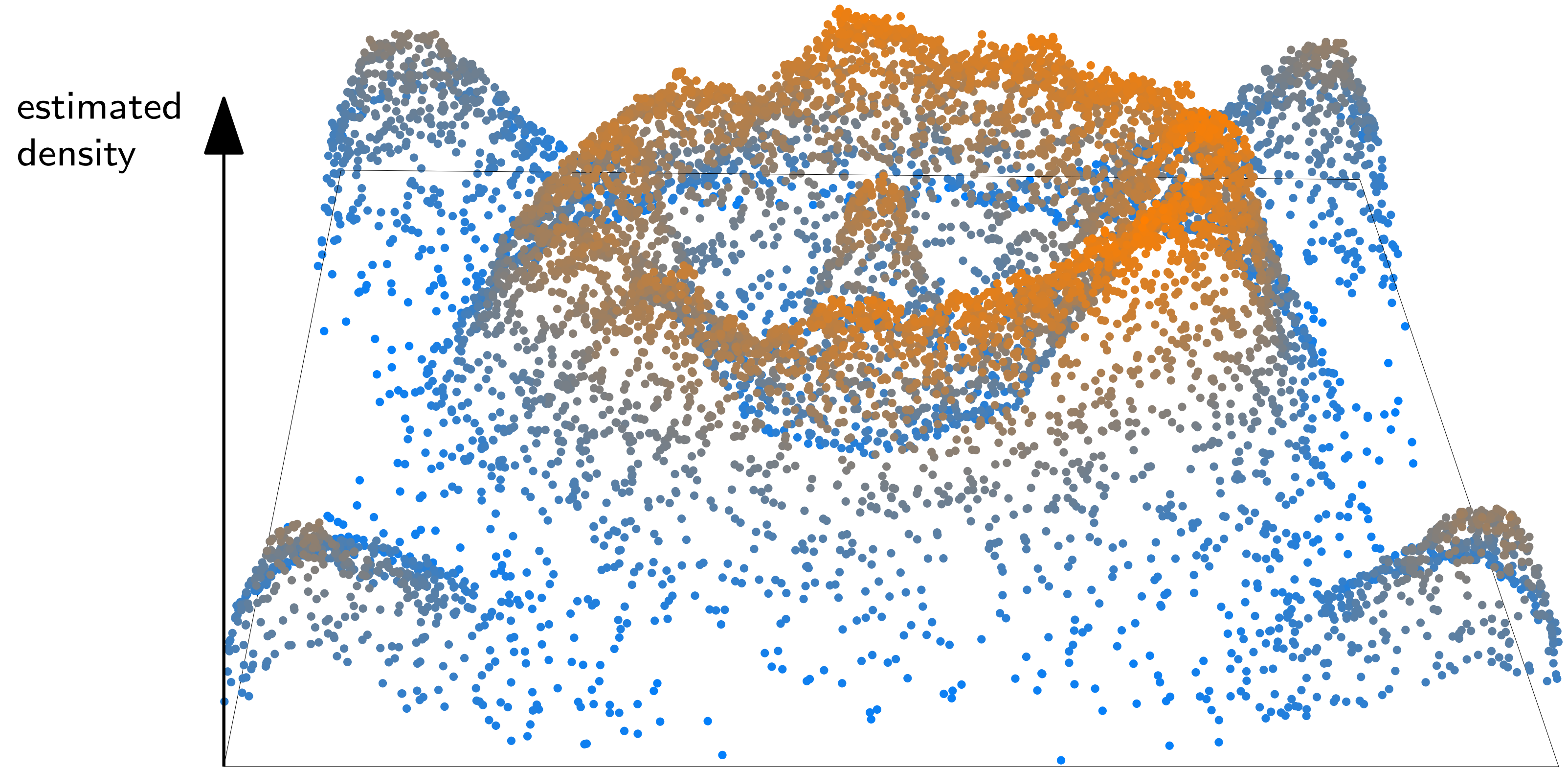
for $i = 1$ to n **do**
    Let $\mathcal{N}$ be the set of neighbors of i in G that have indices lower than i;
    if $\mathcal{N} = \emptyset$ *// vertex i is a peak of $\hat{f}$ within G*
        Create a new entry e in $\mathcal{U}$ and attach vertex i to it;
        $r(e) \leftarrow i$ *// $r(e)$ stores the root vertex associated with the entry e*
    else *// vertex i is not a peak of $\hat{f}$ within G*
        $g(i) \leftarrow \mathrm{argmax}_{j \in \mathcal{N}} \hat{f}(j)$ *// $g(i)$ stores the approximate gradient at vertex i*
        $e_i \leftarrow \mathcal{U}.\mathtt{find}(g(i))$;
        Attach vertex i to the entry e_i;

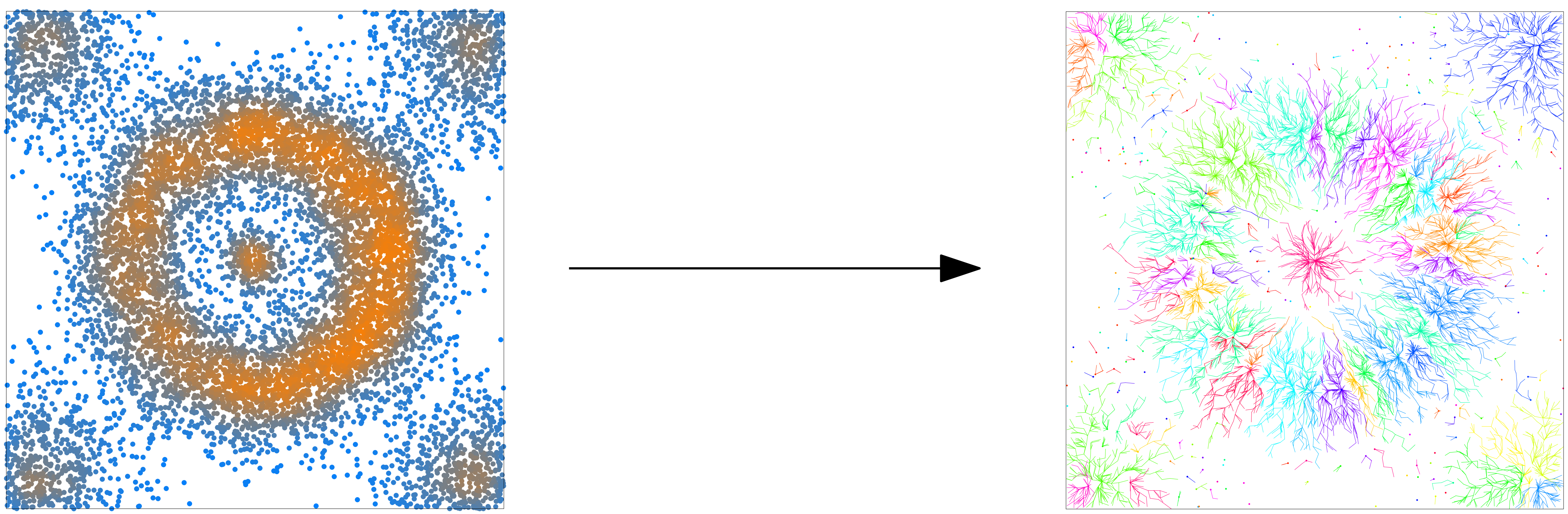
graph-based
hill-climbing
(1976)

Output: the collection of entries e in $\mathcal{U}$

Why things are likely to go ill

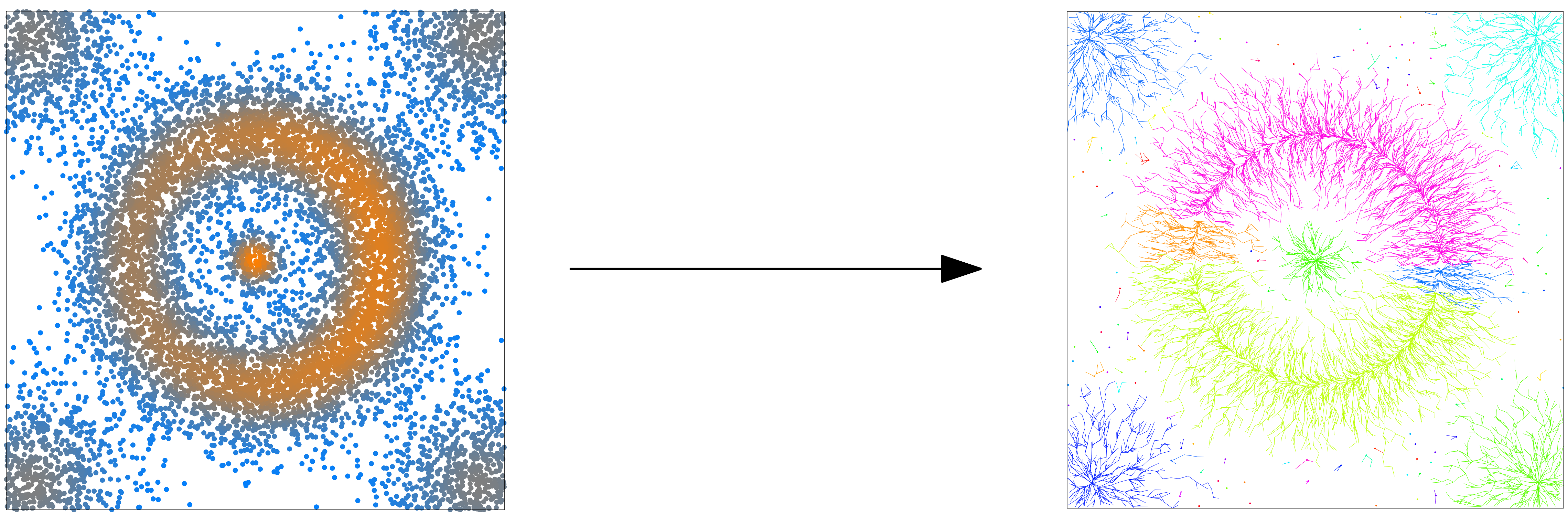
- Noisy estimator





Why things are likely to go ill

- Noisy estimator
- Neighborhood graph



Why things are likely to go ill

- Noisy estimator
- Neighborhood graph

Solutions:

1. **Be proactive**: smooth-out estimator before clustering, a la Mean-Shift

   $\rightarrow$ how much smoothing is needed?

   $\rightarrow$ does not solve the neighborhood graph issue

2. **Be reactive**: merge clusters after clustering, to regain some stability

   $\rightarrow$ repeat mode-seeking until convergence (Medoid-Shift [SKK'07])

   $\rightarrow$ use topological persistence to guide a single-pass merging step

Why things are likely to go ill

- Noisy estimator
- Neighborhood graph

Solutions:

1. **Be proactive**: smooth-out estimator before clustering, a la Mean-Shift

   $\rightarrow$ how much smoothing is needed?

   $\rightarrow$ does not solve the neighborhood graph issue

2. **Be reactive**: merge clusters after clustering, to regain some stability

   $\rightarrow$ repeat mode-seeking until convergence (Medoid-Shift [SKK'07])

   $\rightarrow$ use topological persistence to guide a single-pass merging step

Enter Topological Persistence...

Topological Persistence (in a nutshell)

X topological space

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

$\downarrow$

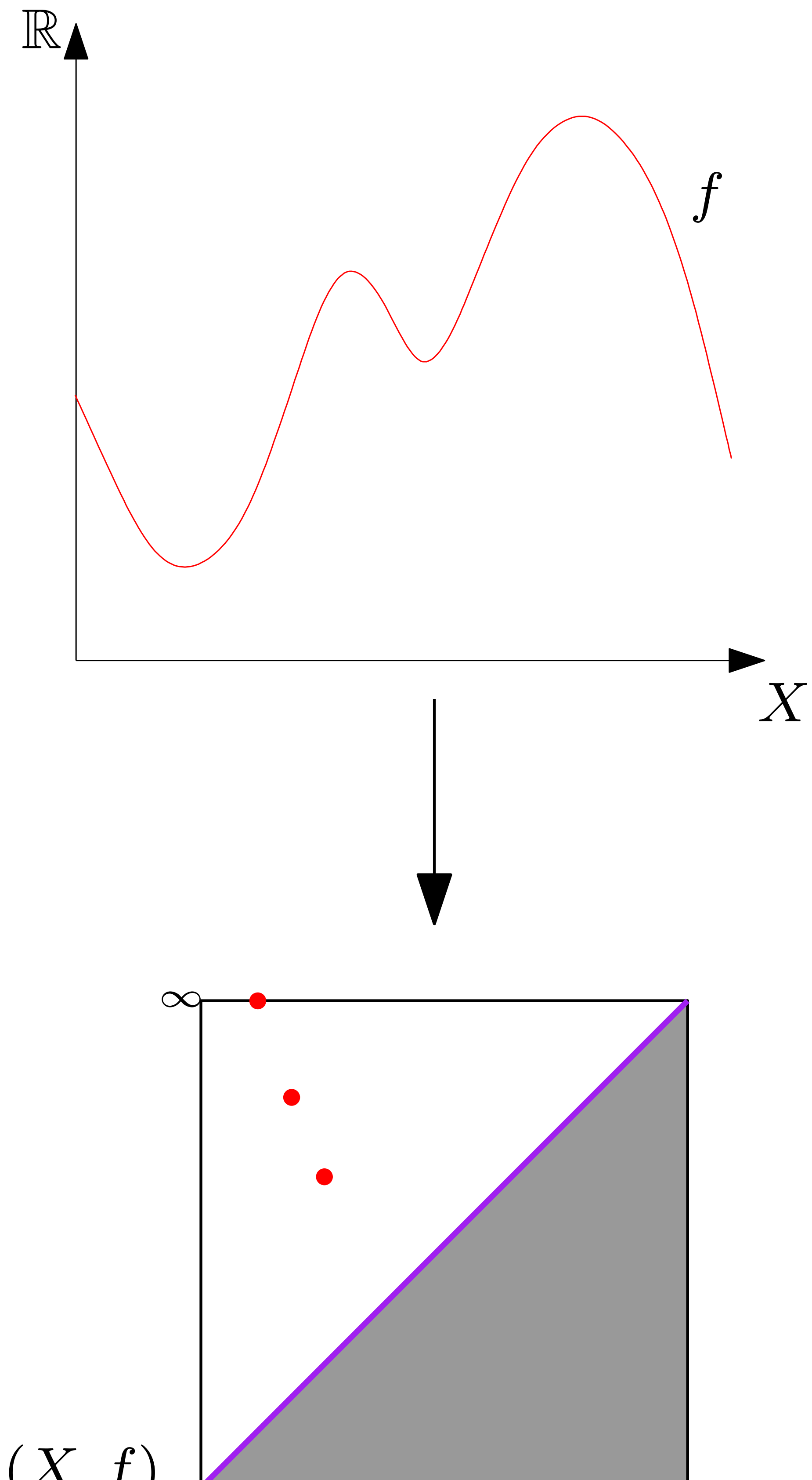
persistence

$\downarrow$

$\mathrm{Dg}\, f$

signature: *persistence diagram*

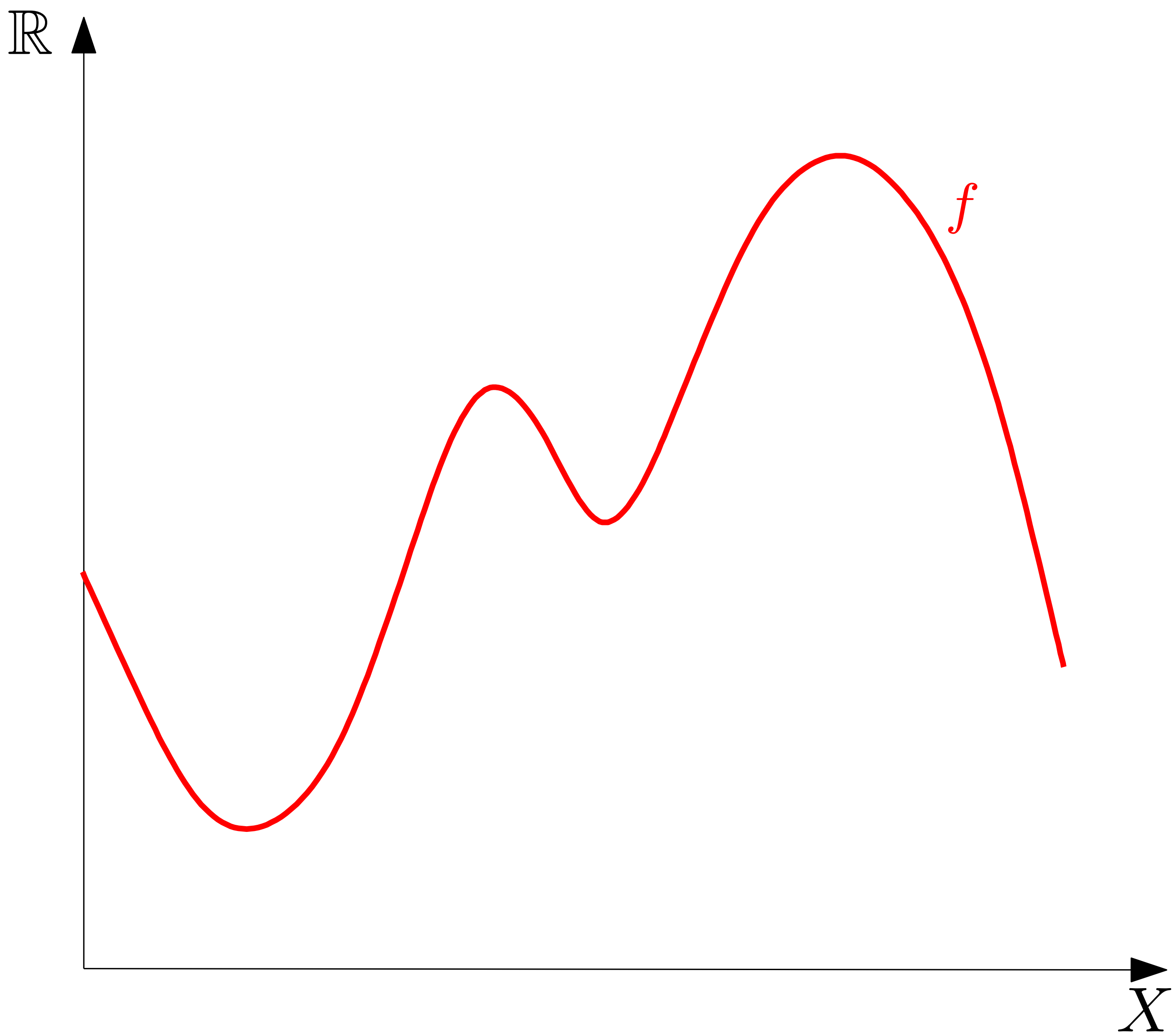
encodes the topological structure of the pair (X, f)



Topological Persistence (in a nutshell)

Inside the black box:

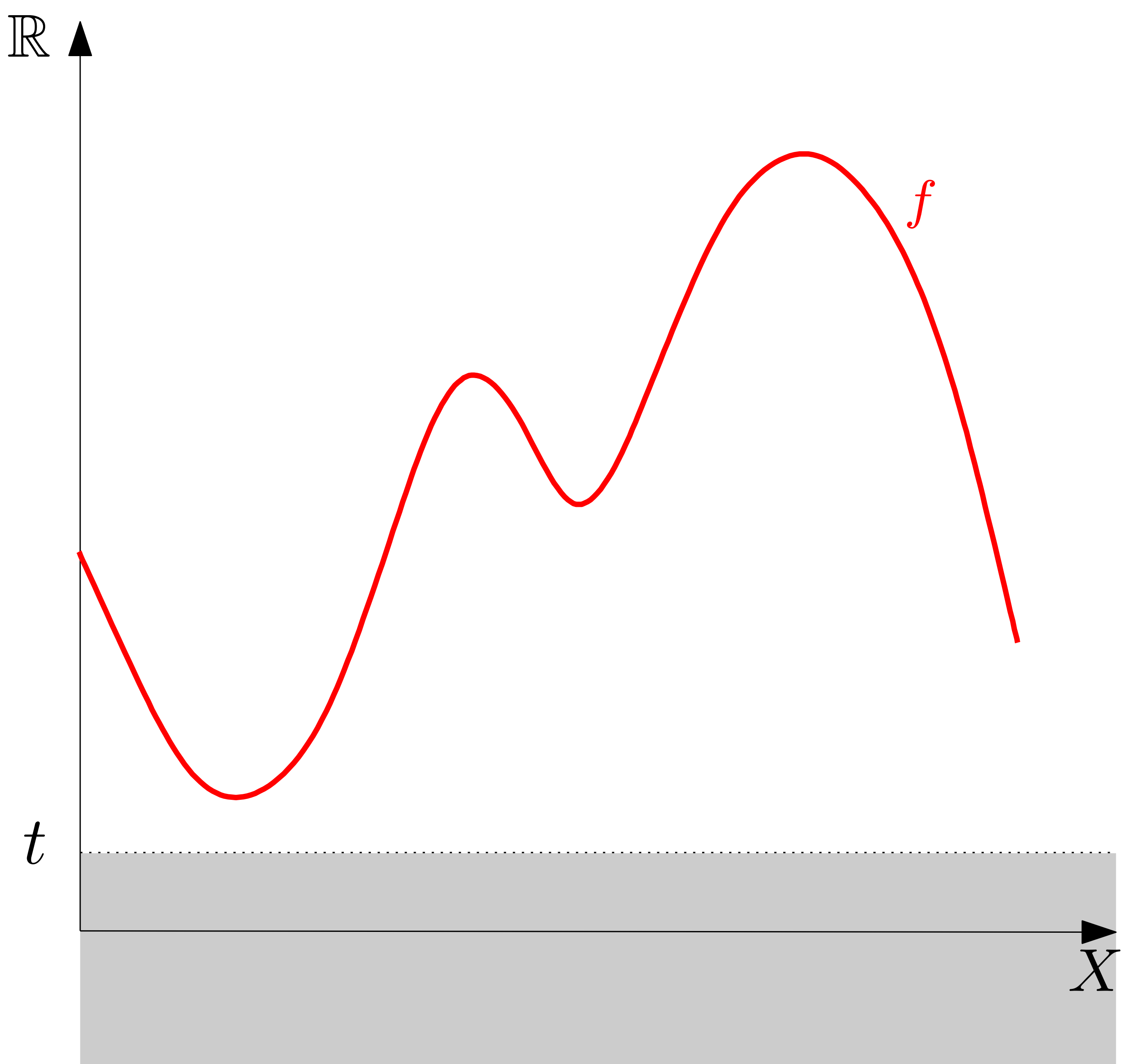
- Nested family (*filtration*) of sublevel-sets $f^{-1}((-\infty, t])$ for t ranging from $-\infty$ to $+\infty$
- Track the evolution of the topology throughout the family



Topological Persistence (in a nutshell)

Inside the black box:

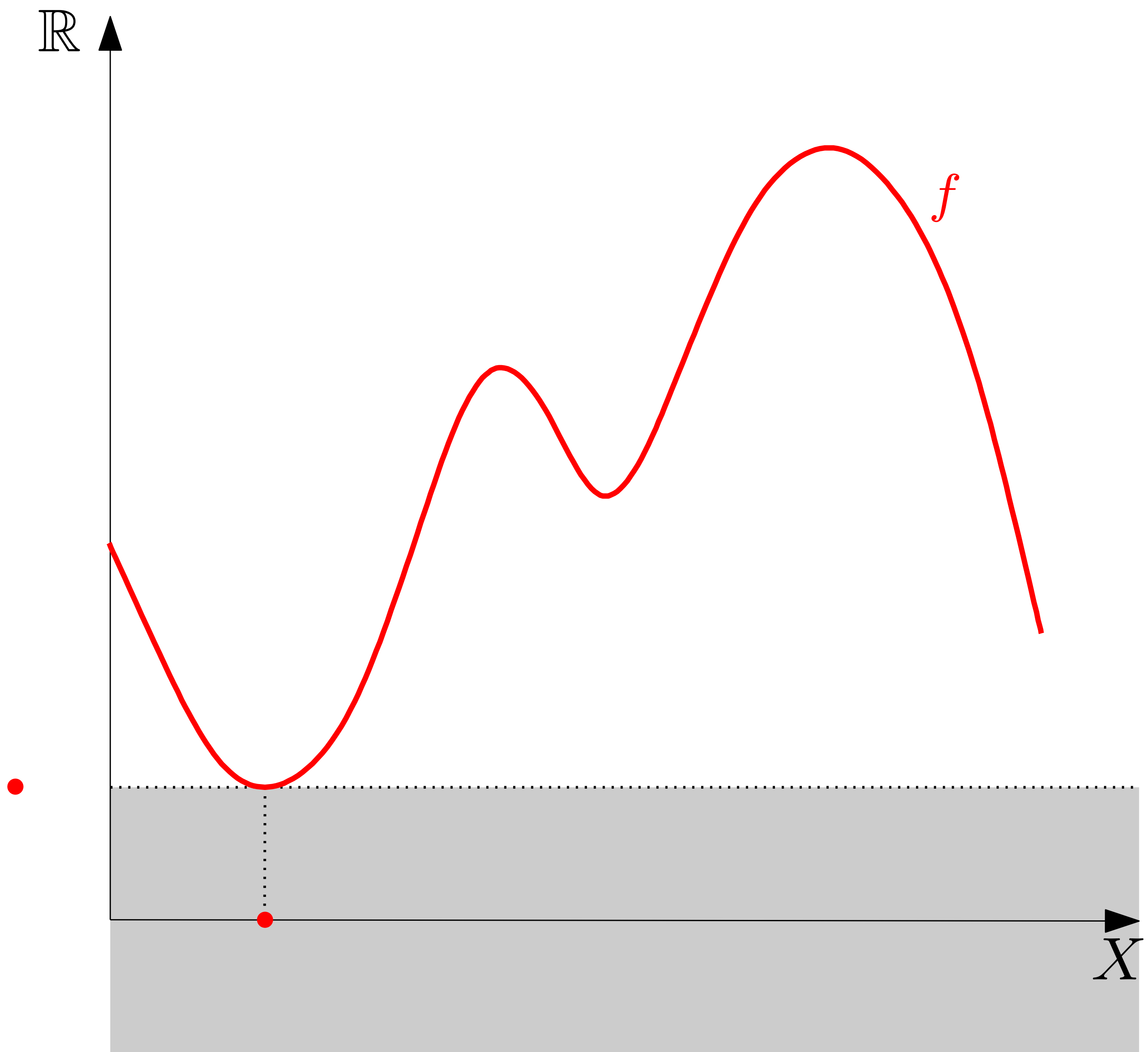
- Nested family (*filtration*) of sublevel-sets $f^{-1}((-\infty, t])$ for t ranging from $-\infty$ to $+\infty$
- Track the evolution of the topology throughout the family



Topological Persistence (in a nutshell)

Inside the black box:

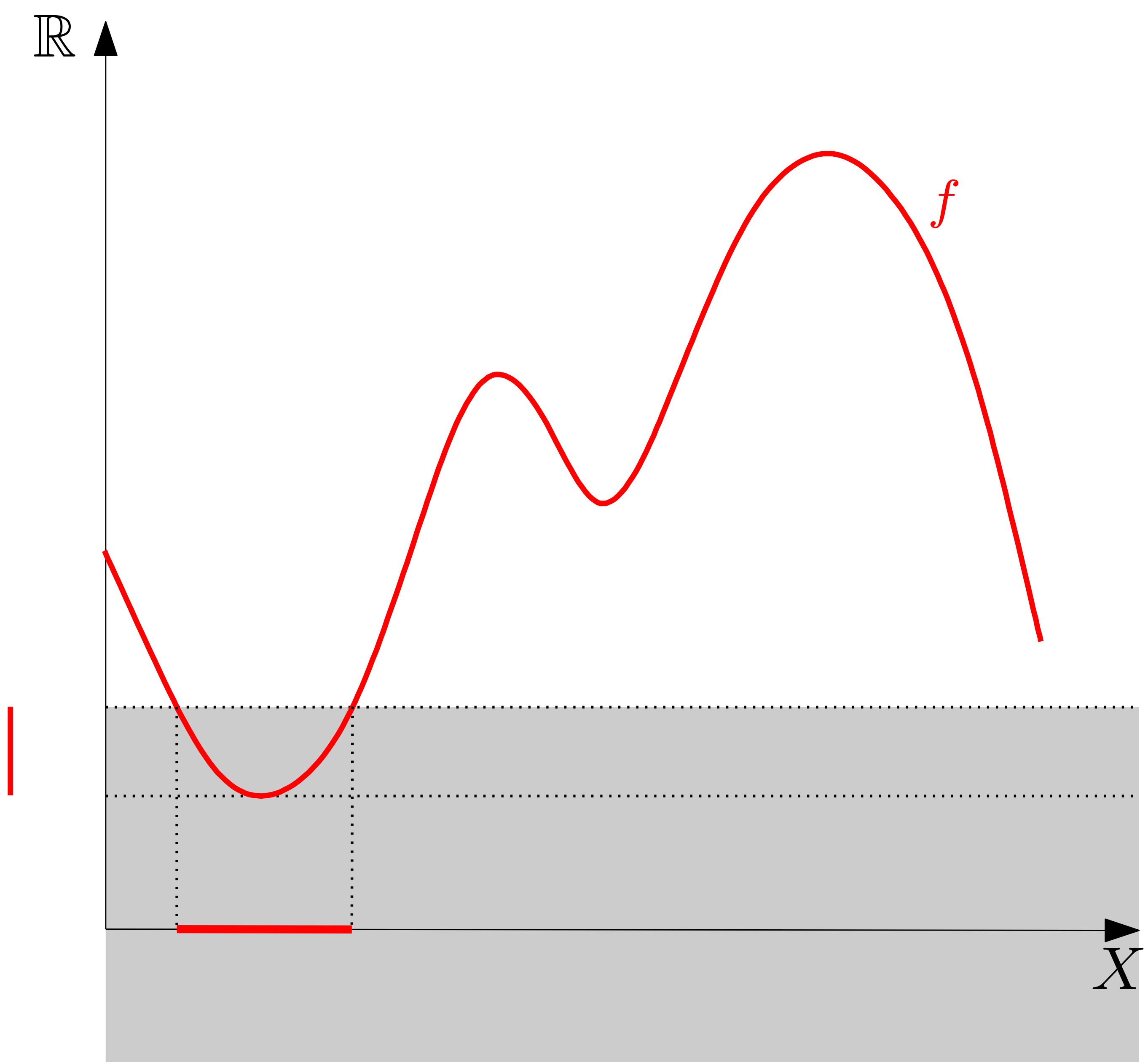
- Nested family (*filtration*) of sublevel-sets $f^{-1}((-\infty, t])$ for t ranging from $-\infty$ to $+\infty$
- Track the evolution of the topology throughout the family



Topological Persistence (in a nutshell)

Inside the black box:

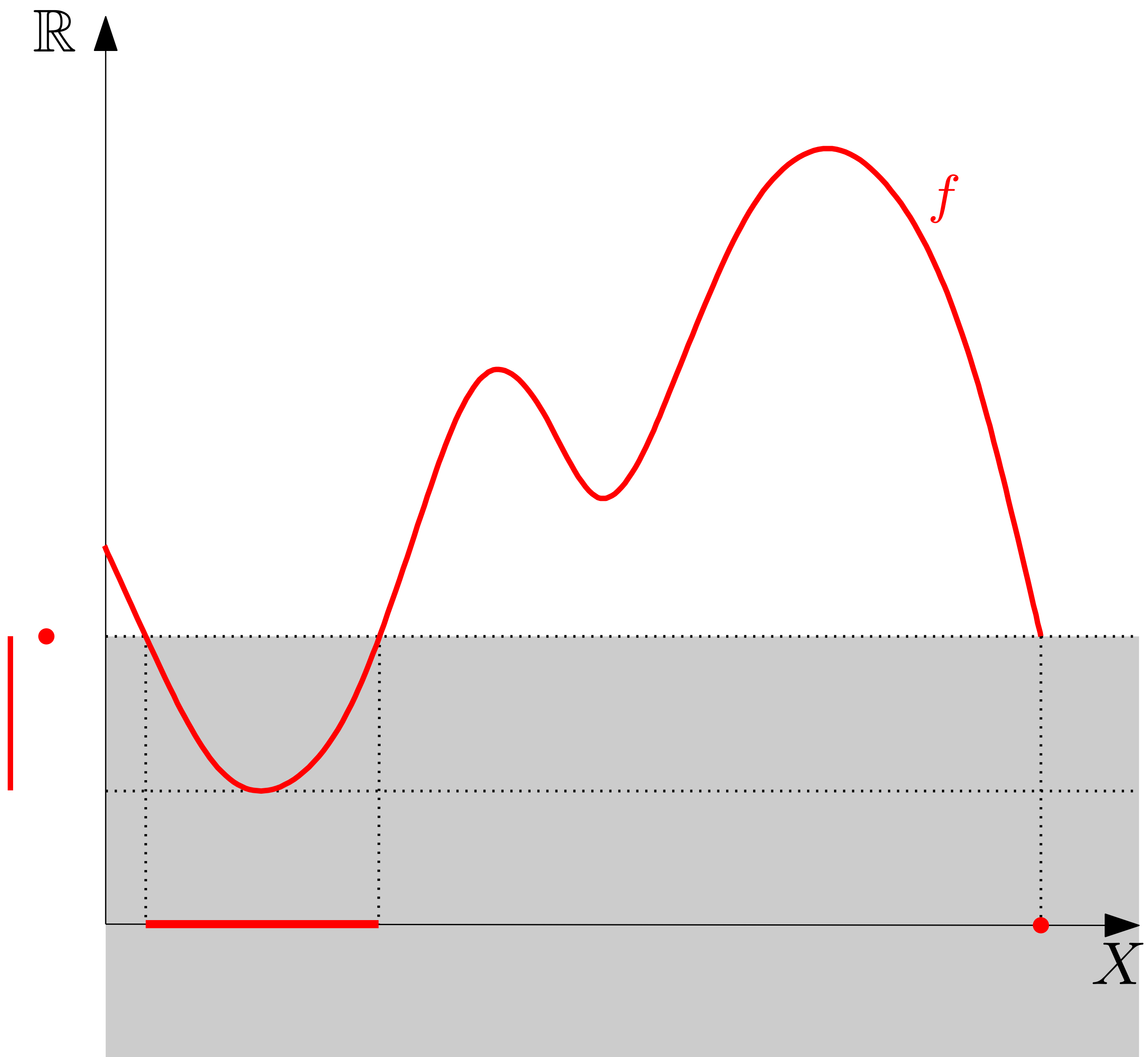
- Nested family (*filtration*) of sublevel-sets $f^{-1}((-\infty, t])$ for t ranging from $-\infty$ to $+\infty$
- Track the evolution of the topology throughout the family



Topological Persistence (in a nutshell)

Inside the black box:

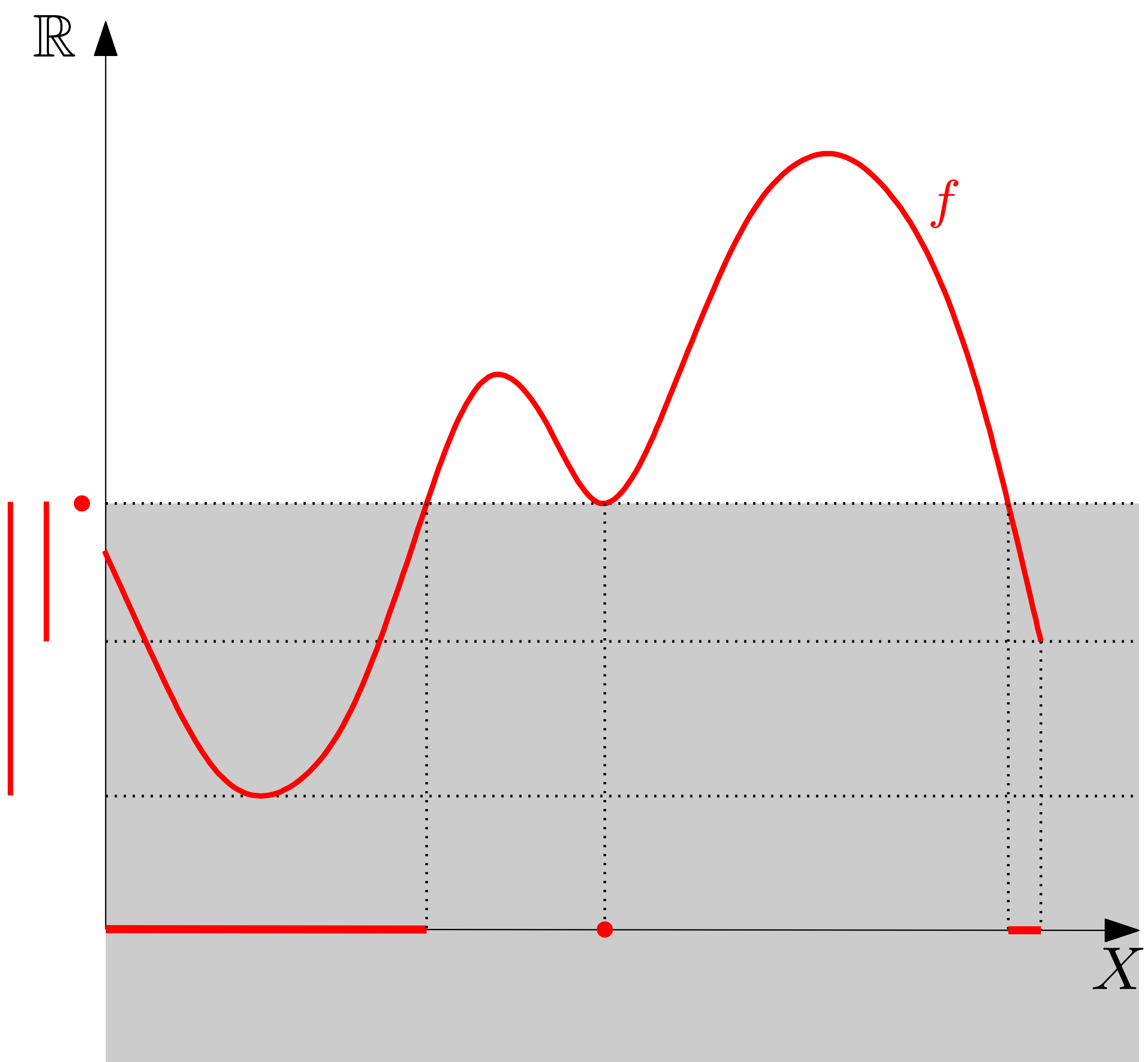
- Nested family (*filtration*) of sublevel-sets $f^{-1}((-\infty, t])$ for t ranging from $-\infty$ to $+\infty$
- Track the evolution of the topology throughout the family



Topological Persistence (in a nutshell)

Inside the black box:

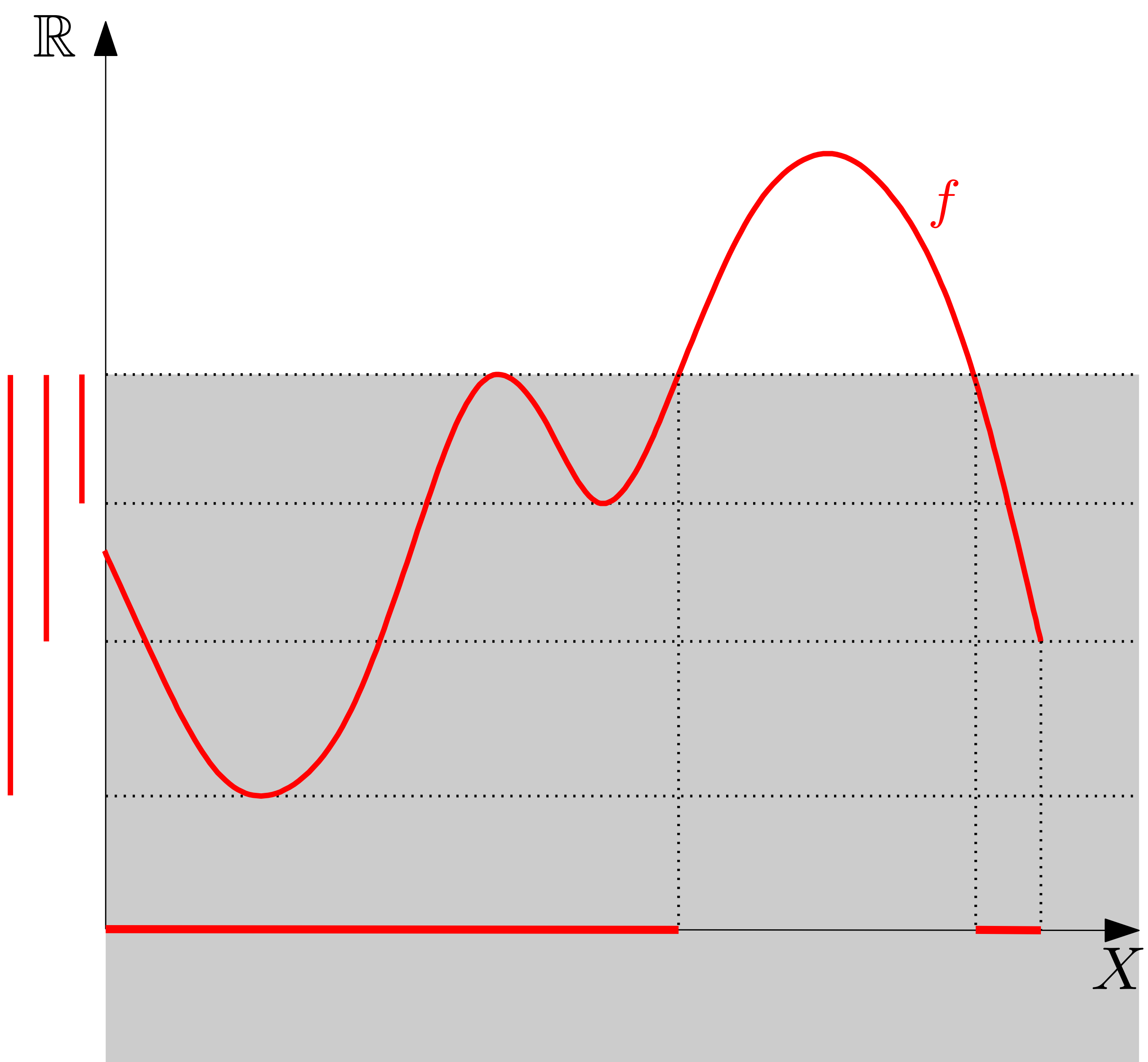
- Nested family (*filtration*) of sublevel-sets $f^{-1}((-\infty, t])$ for t ranging from $-\infty$ to $+\infty$
- Track the evolution of the topology throughout the family



Topological Persistence (in a nutshell)

Inside the black box:

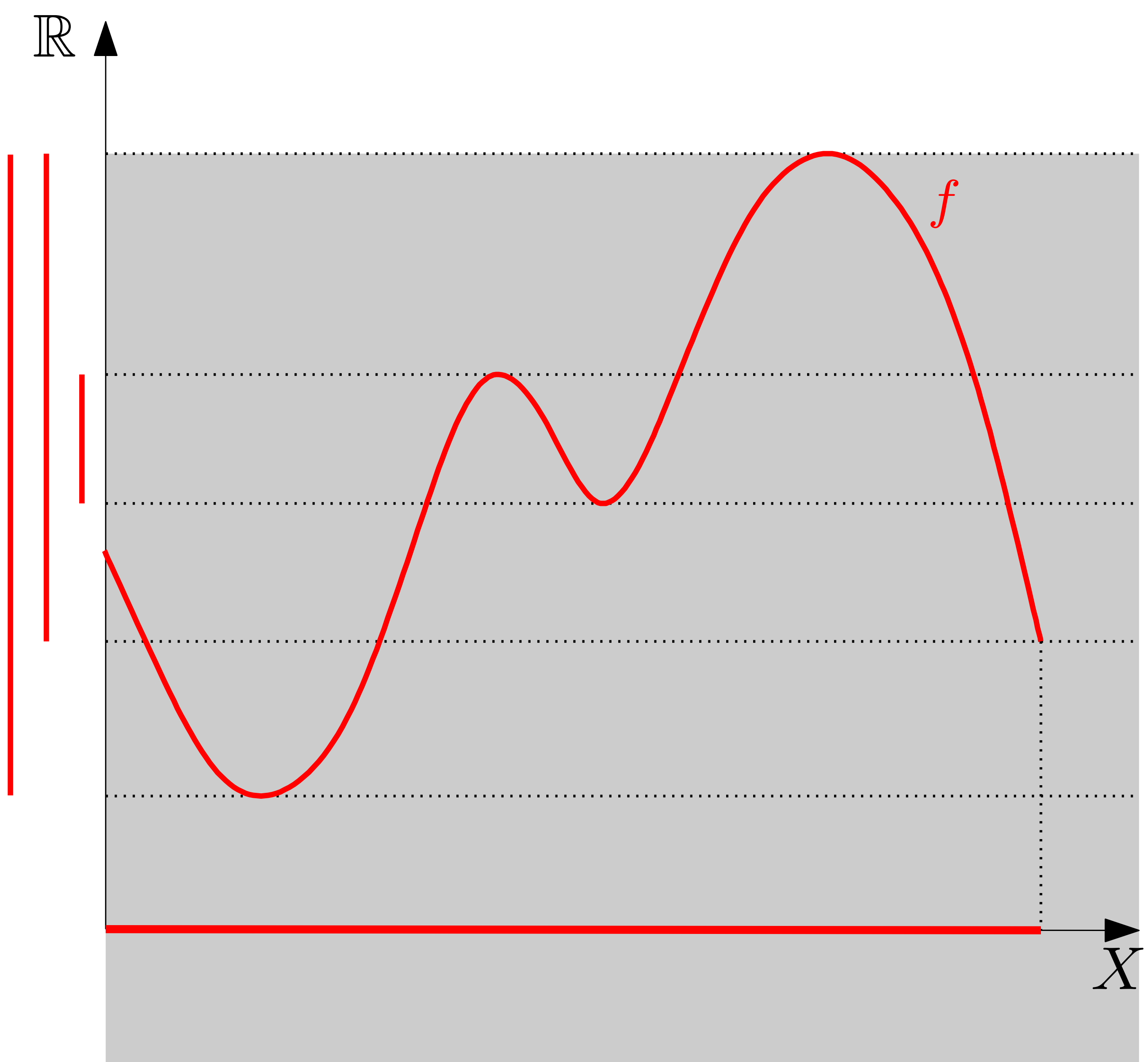
- Nested family (*filtration*) of sublevel-sets $f^{-1}((-\infty, t])$ for t ranging from $-\infty$ to $+\infty$
- Track the evolution of the topology throughout the family



Topological Persistence (in a nutshell)

Inside the black box:

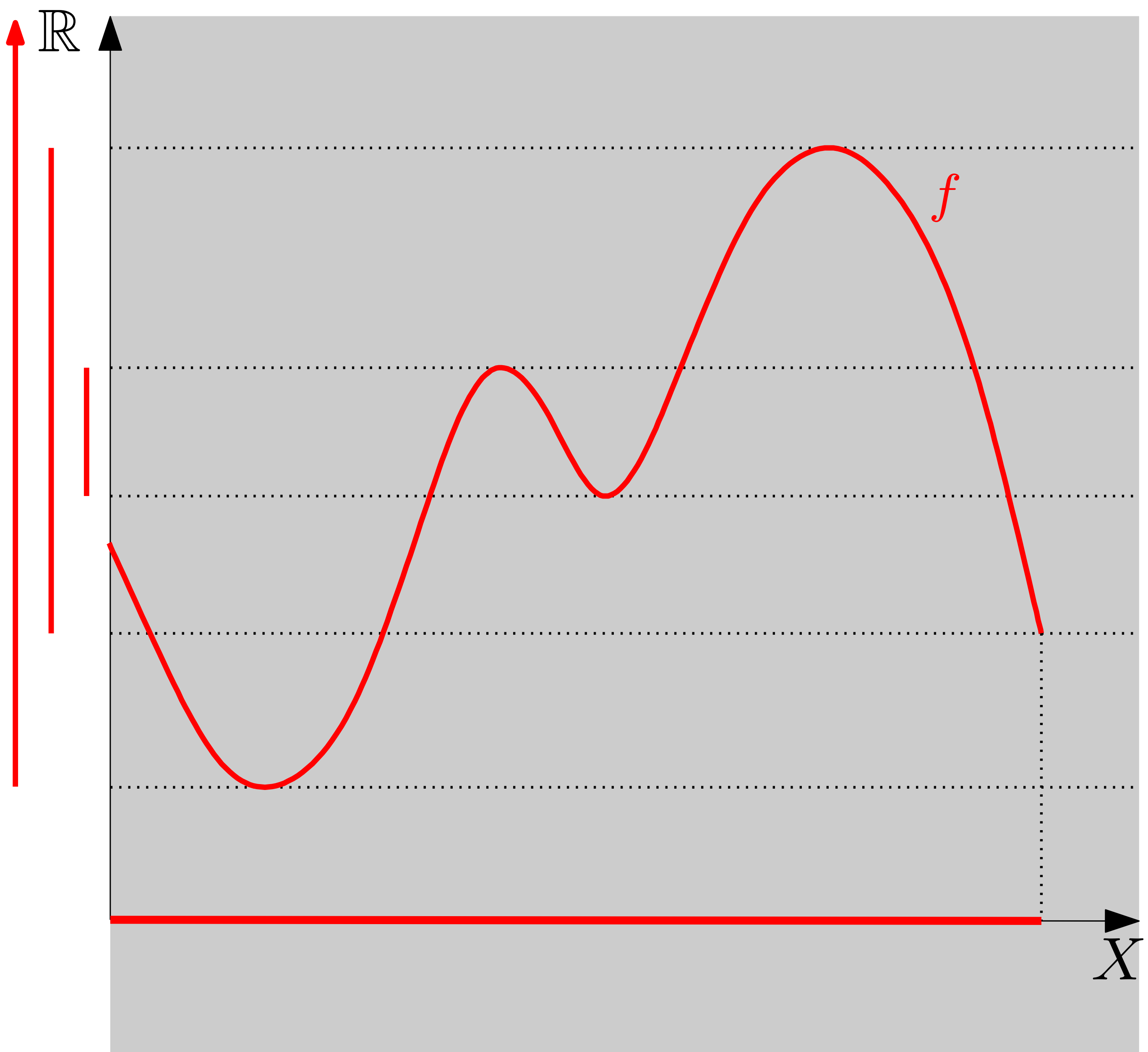
- Nested family (*filtration*) of sublevel-sets $f^{-1}((-\infty, t])$ for t ranging from $-\infty$ to $+\infty$
- Track the evolution of the topology throughout the family



Topological Persistence (in a nutshell)

Inside the black box:

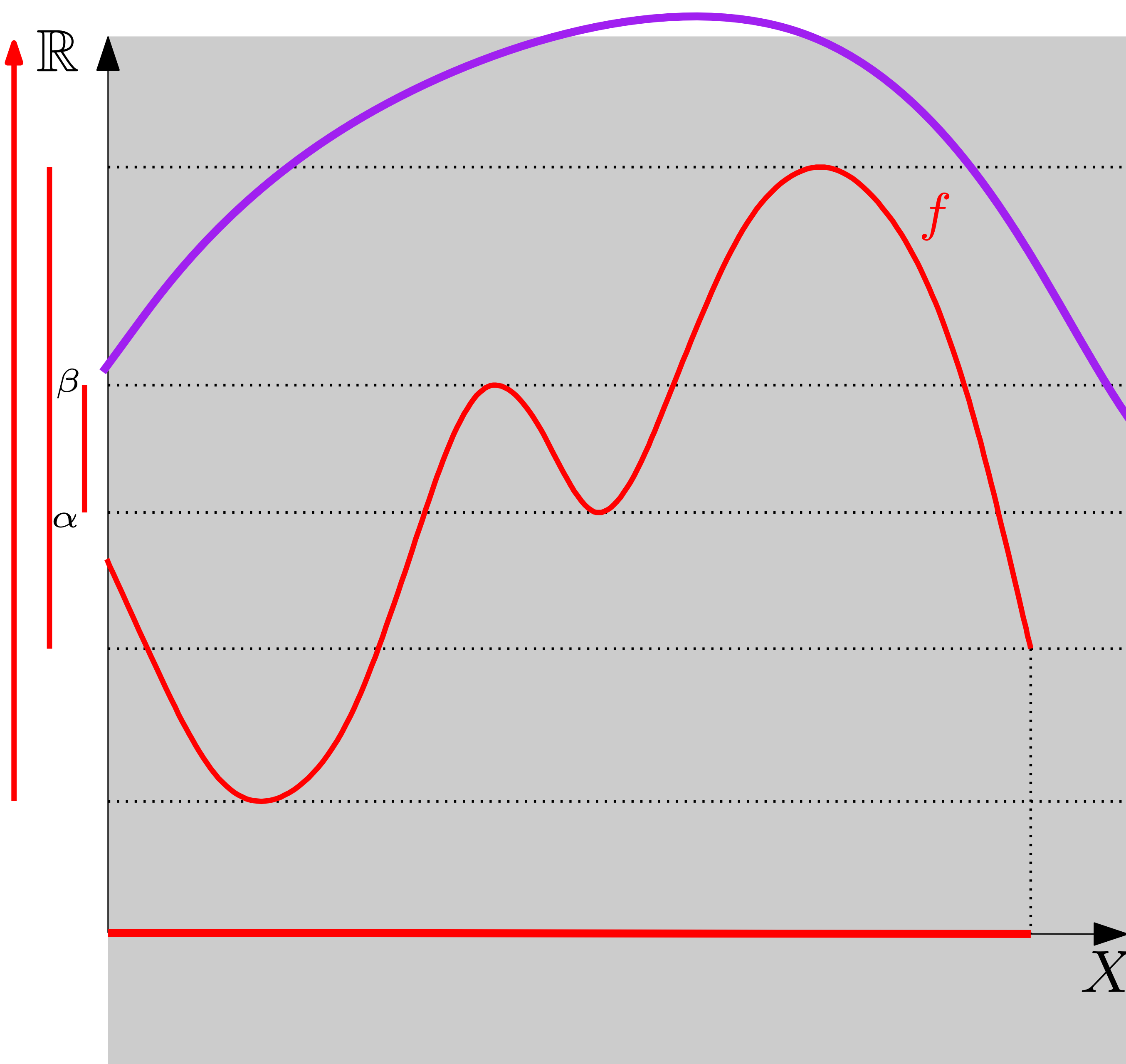
- Nested family (*filtration*) of sublevel-sets $f^{-1}((-\infty, t])$ for t ranging from $-\infty$ to $+\infty$
- Track the evolution of the topology throughout the family
- Finite set of intervals (barcode) encodes births/deaths of topological features



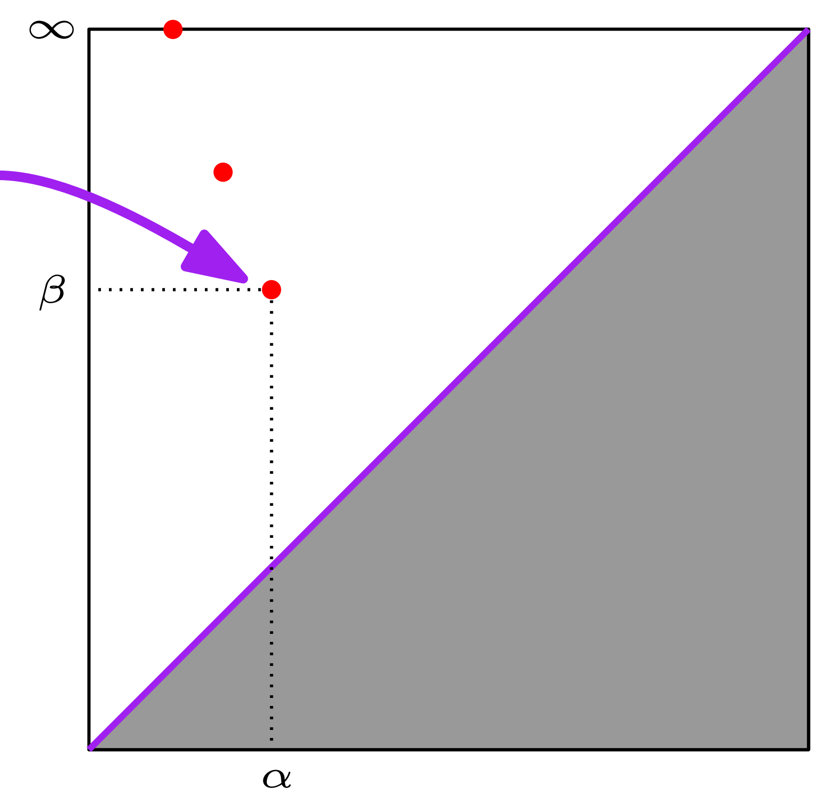
Topological Persistence (in a nutshell)

Inside the black box:

- Nested family (*filtration*) of sublevel-sets $f^{-1}((-\infty, t])$ for t ranging from $-\infty$ to $+\infty$
- Track the evolution of the topology throughout the family
- Finite set of intervals (barcode) encodes births/deaths of topological features



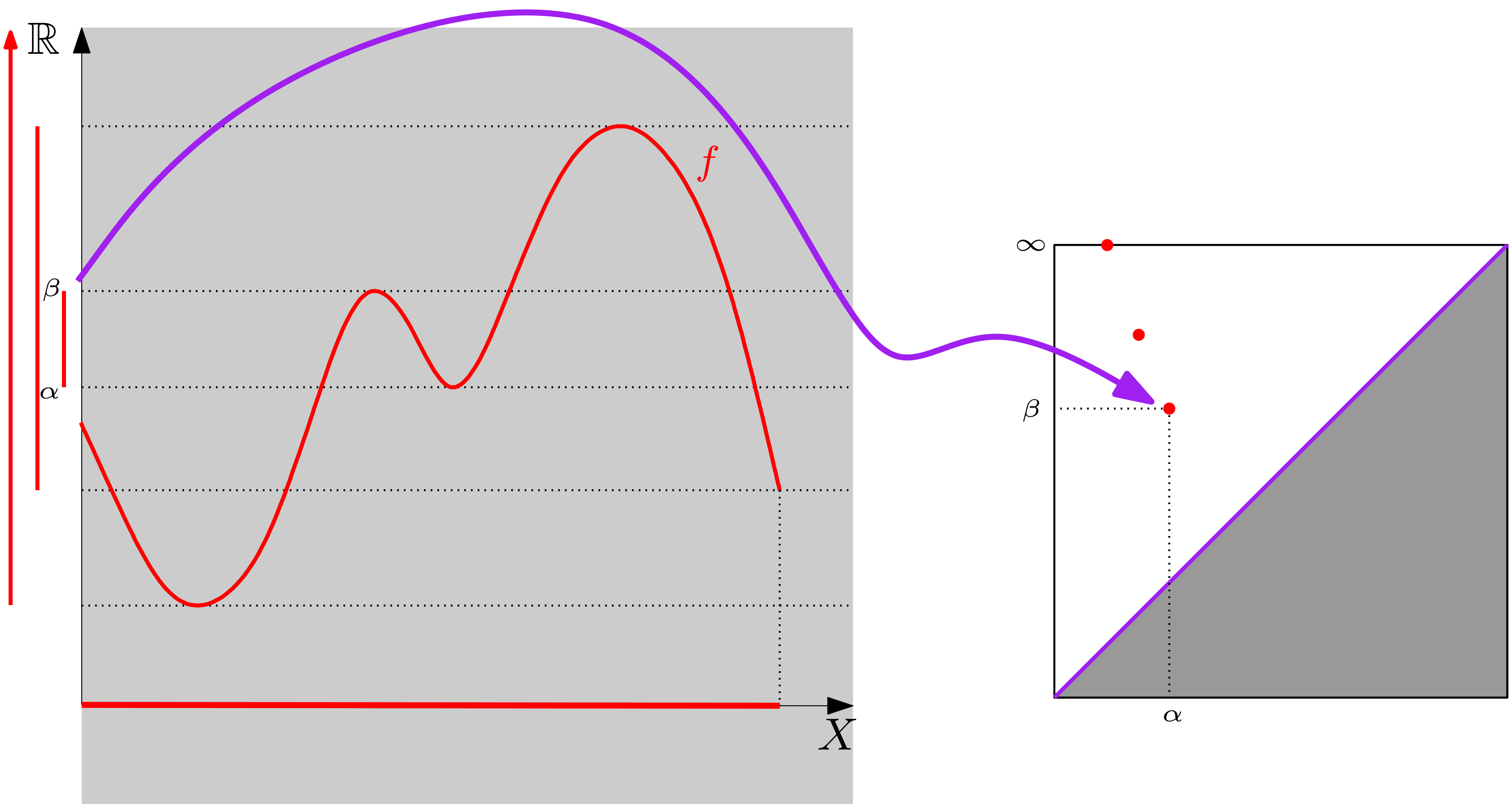
- Alternate representation as a multiset of points in the plane (*diagram*).



Topological Persistence (in a nutshell)

Algorithm:

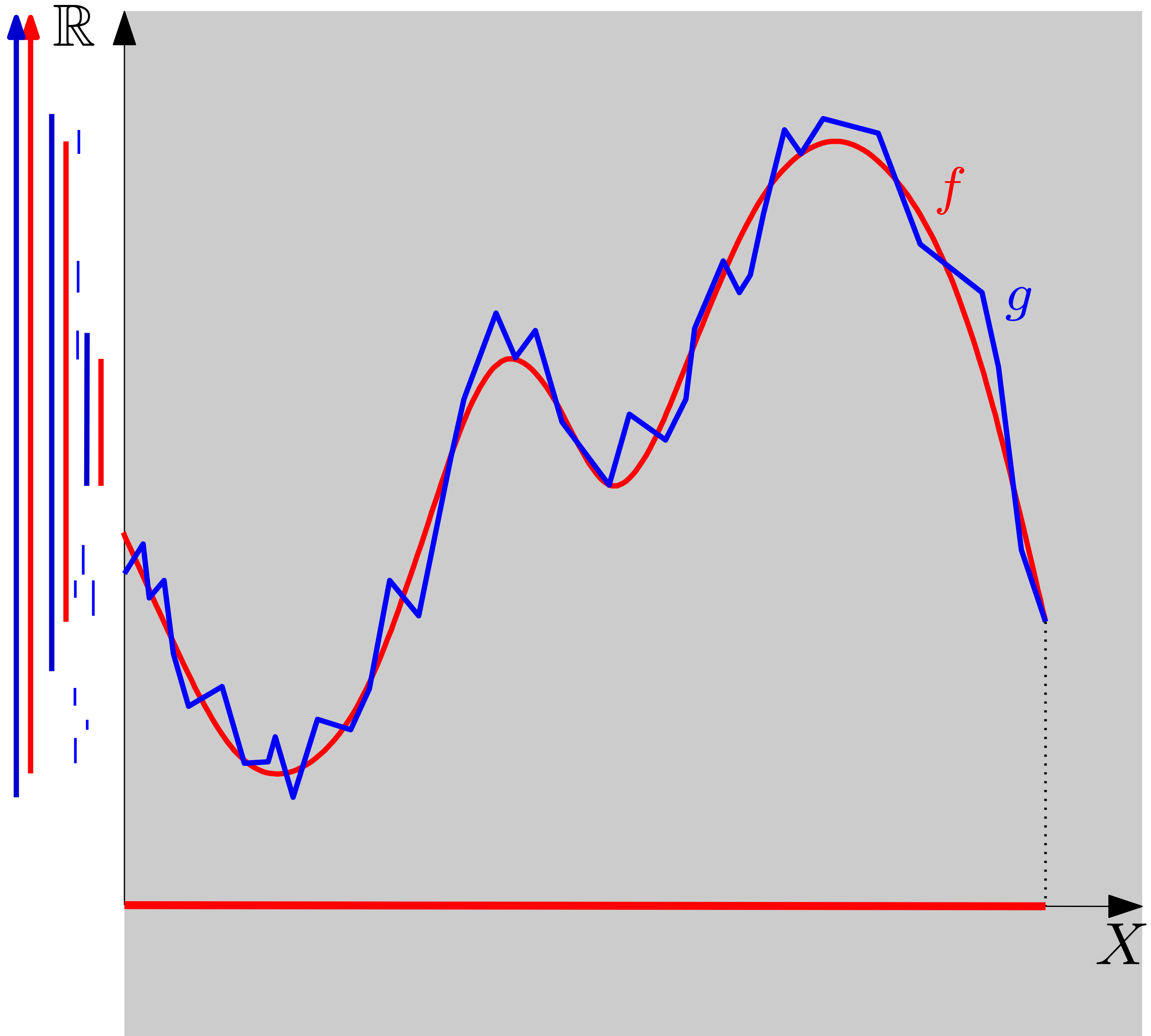
- input: graph $G = (V, E)$ + map $f : V \sqcup E \to \mathbb{R}$
- procedure: scan graph by increasing f-values, update CCs by union-find



Topological Persistence (in a nutshell)

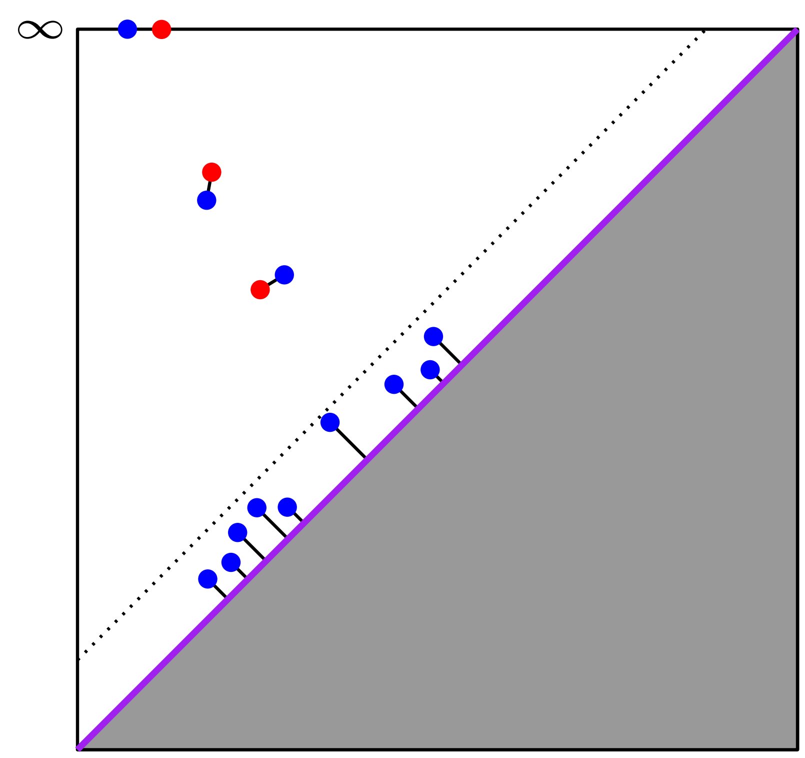
Inside the black box:

- Nested family (*filtration*) of sublevel-sets $f^{-1}((-\infty, t])$ for t ranging from $-\infty$ to $+\infty$
- Track the evolution of the topology throughout the family
- Finite set of intervals (barcode) encodes births/deaths of topological features



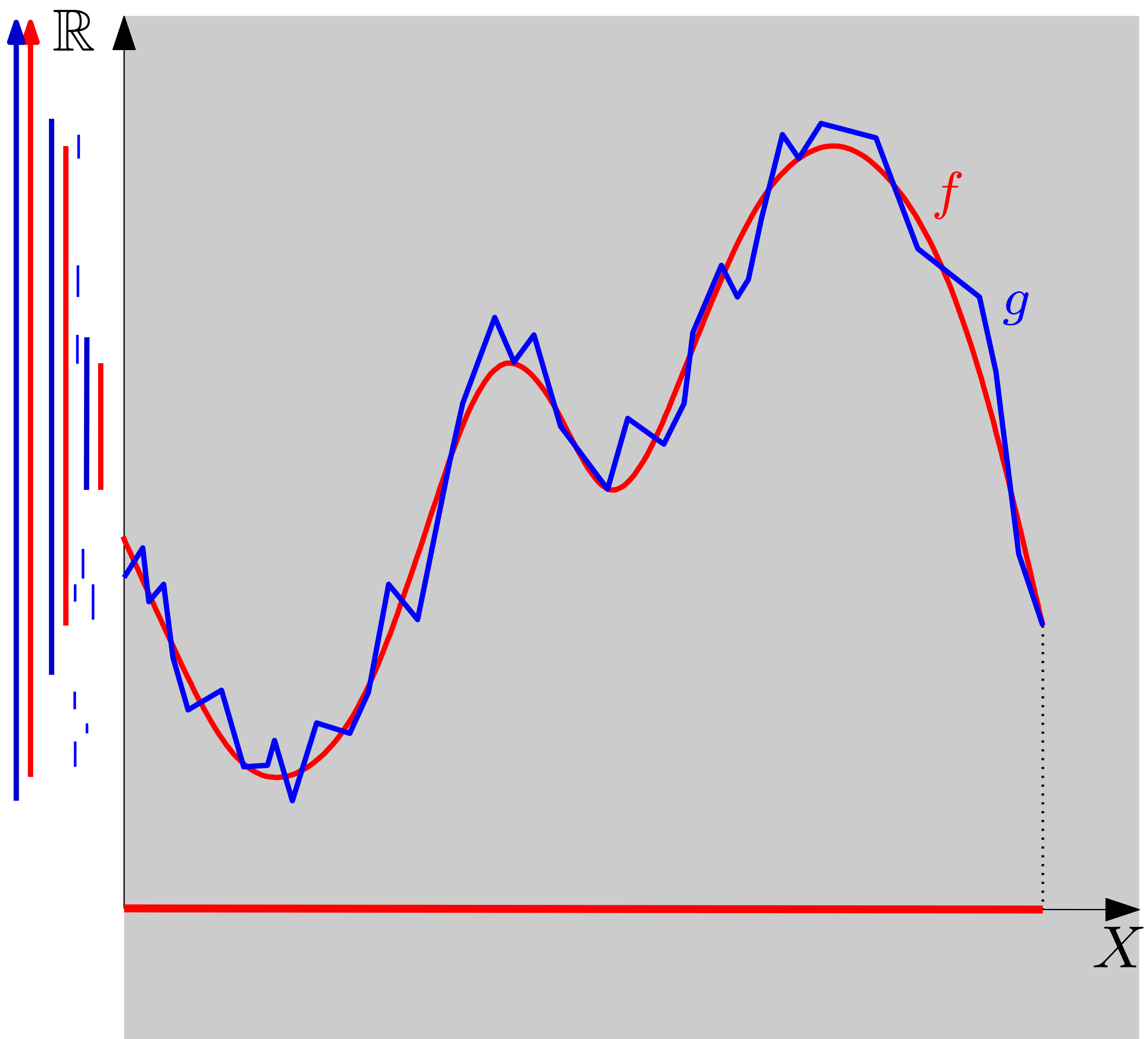
- Alternate representation as a multiset of points in the plane (*diagram*).

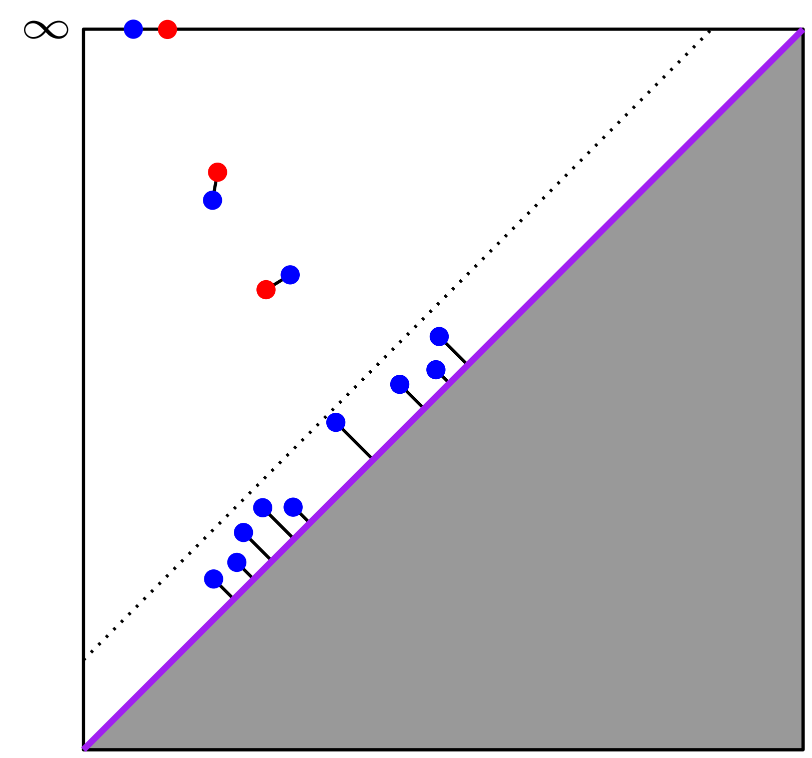
What if f is slightly perturbed?



Topological Persistence (in a nutshell)

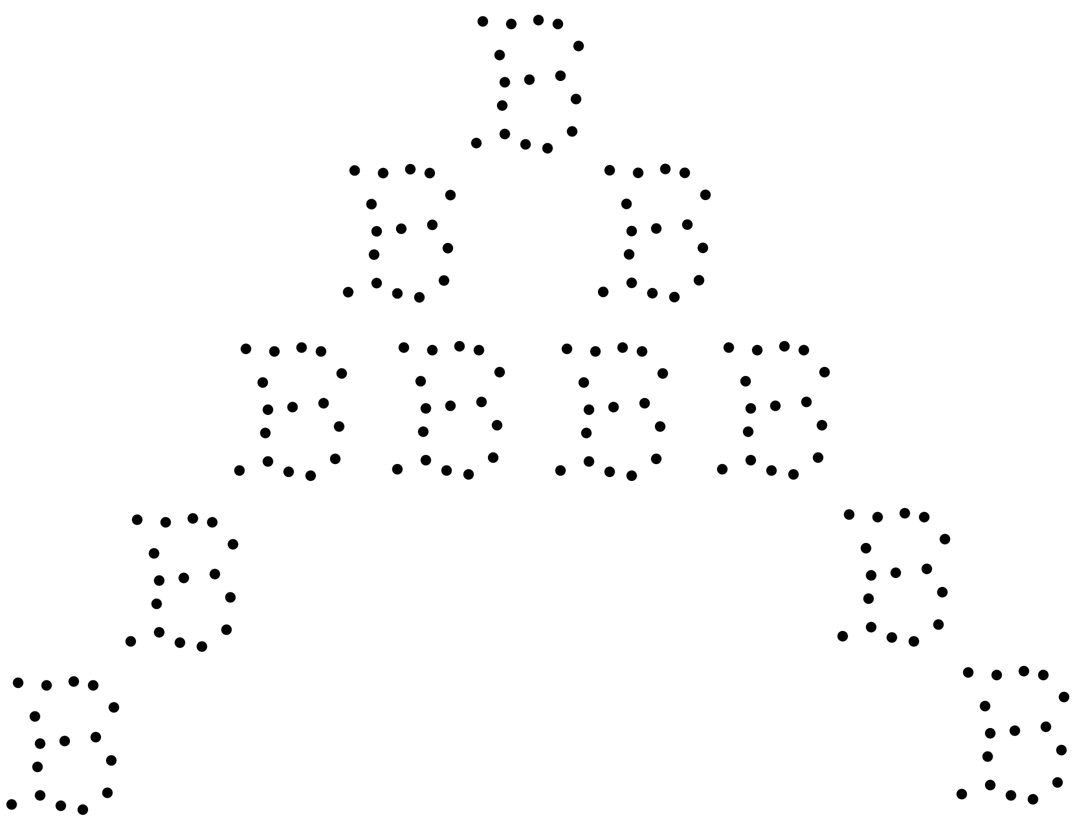
Theorem (Stability): [Cohen-Steiner et al. 2005, Chazal, O. et al. 2009]
For any *tame* functions $f, g : \mathbb{X} \to \mathbb{R}$, $\mathrm{d}_B^\infty(\mathrm{Dg}\, f, \mathrm{Dg}\, g) \leq \|f - g\|_\infty$.

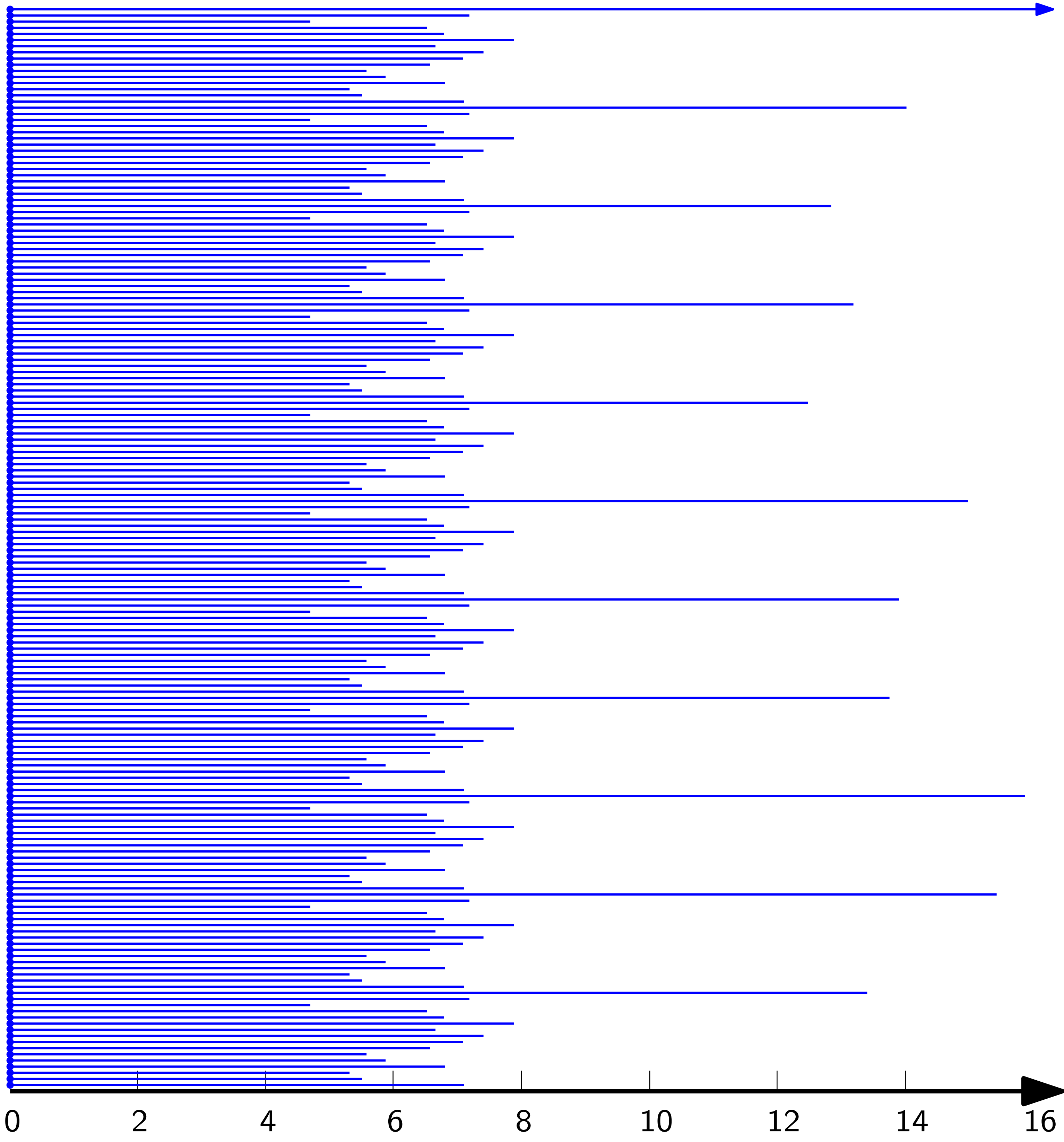




Example: Distance Function

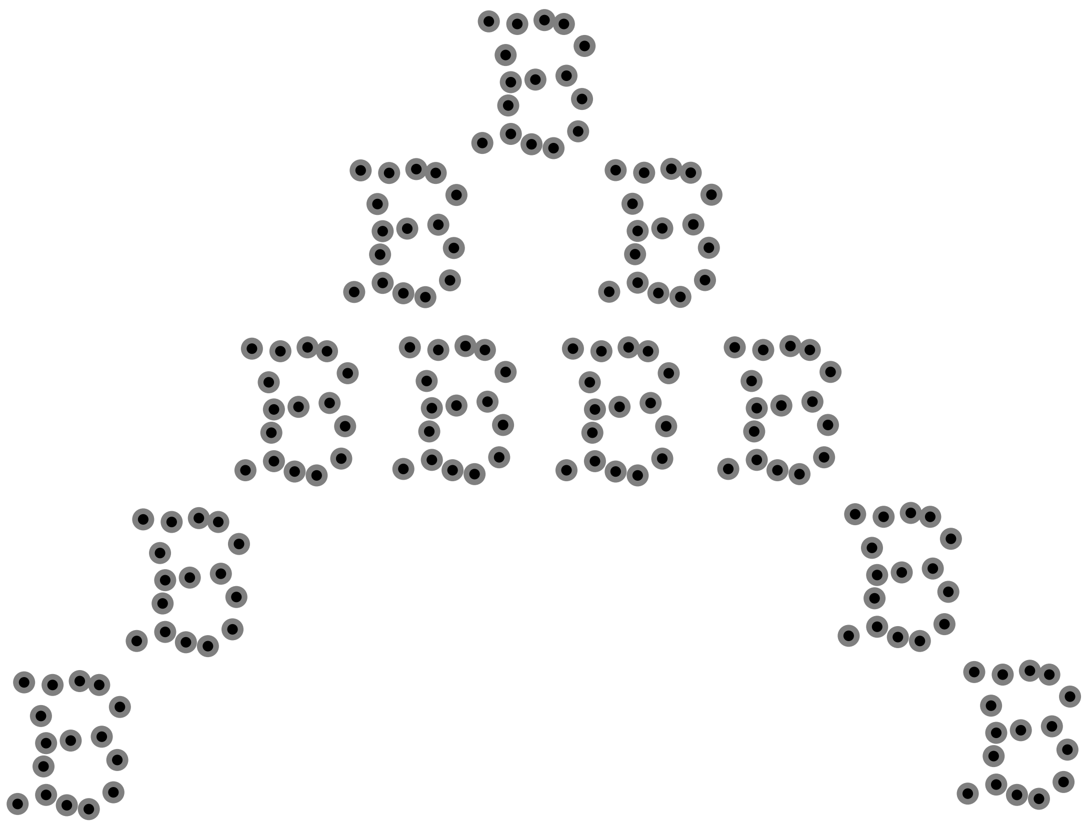
$$f_P : \quad \mathbb{R}^2 \to \mathbb{R}$$
$$x \mapsto \min_{p \in P} \|x - p\|_2$$

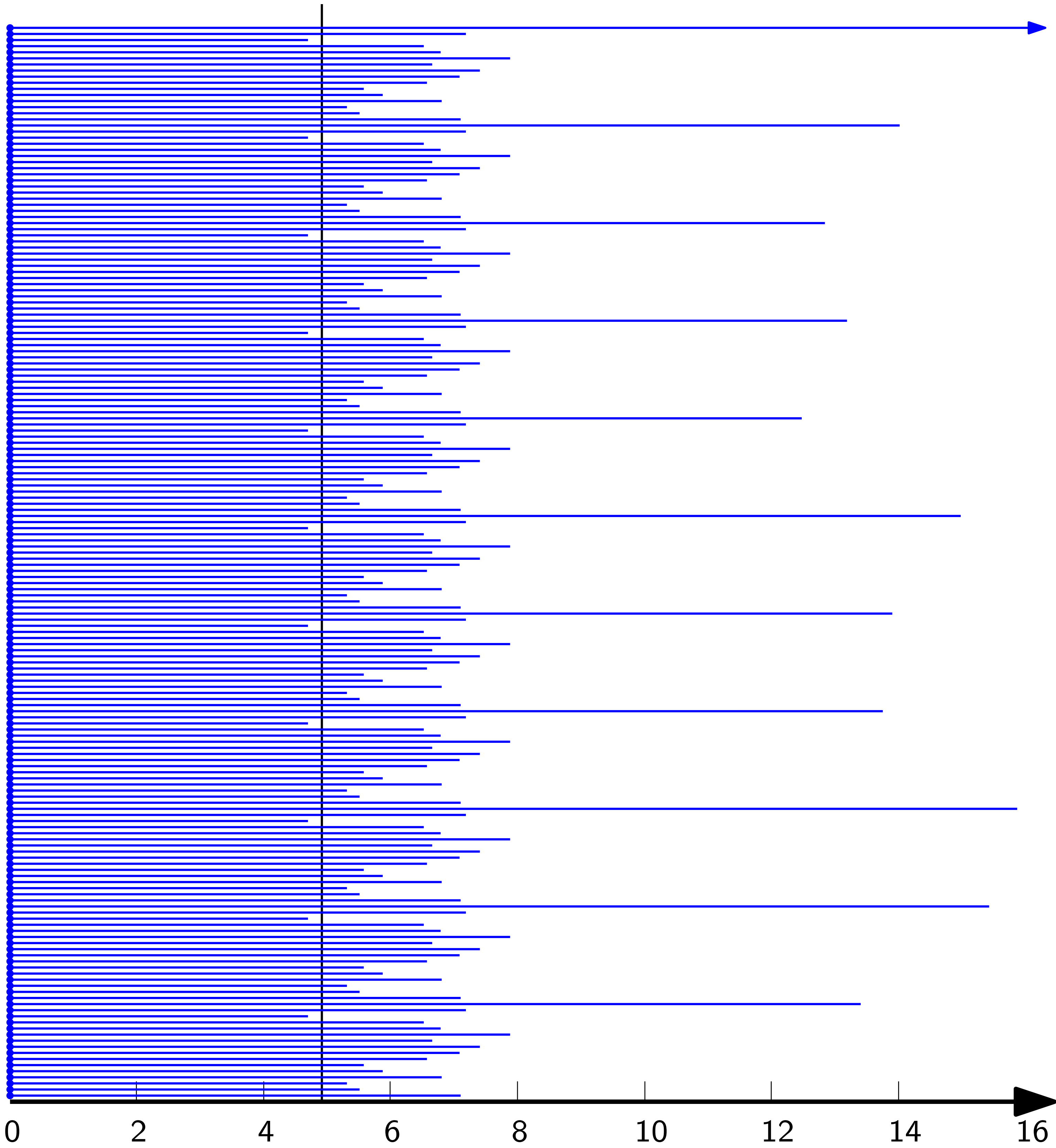




Example: Distance Function

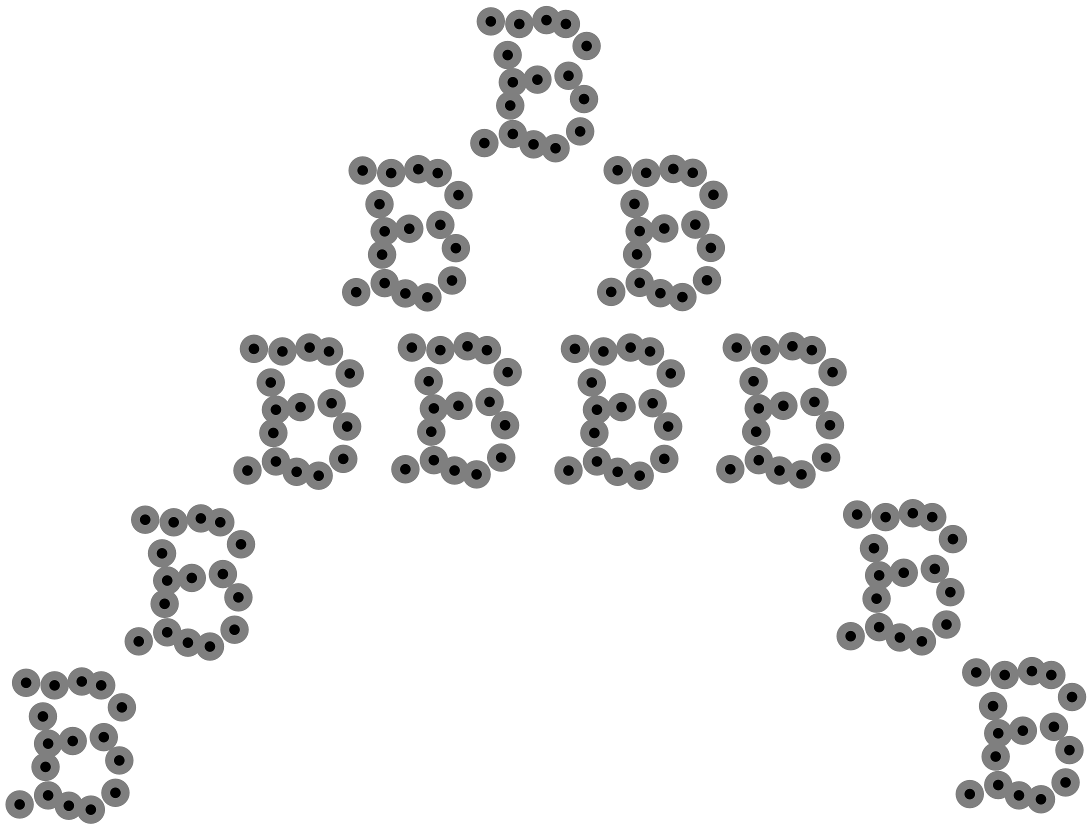
$$f_P : \quad \mathbb{R}^2 \to \mathbb{R}$$
$$x \mapsto \min_{p \in P} \|x - p\|_2$$

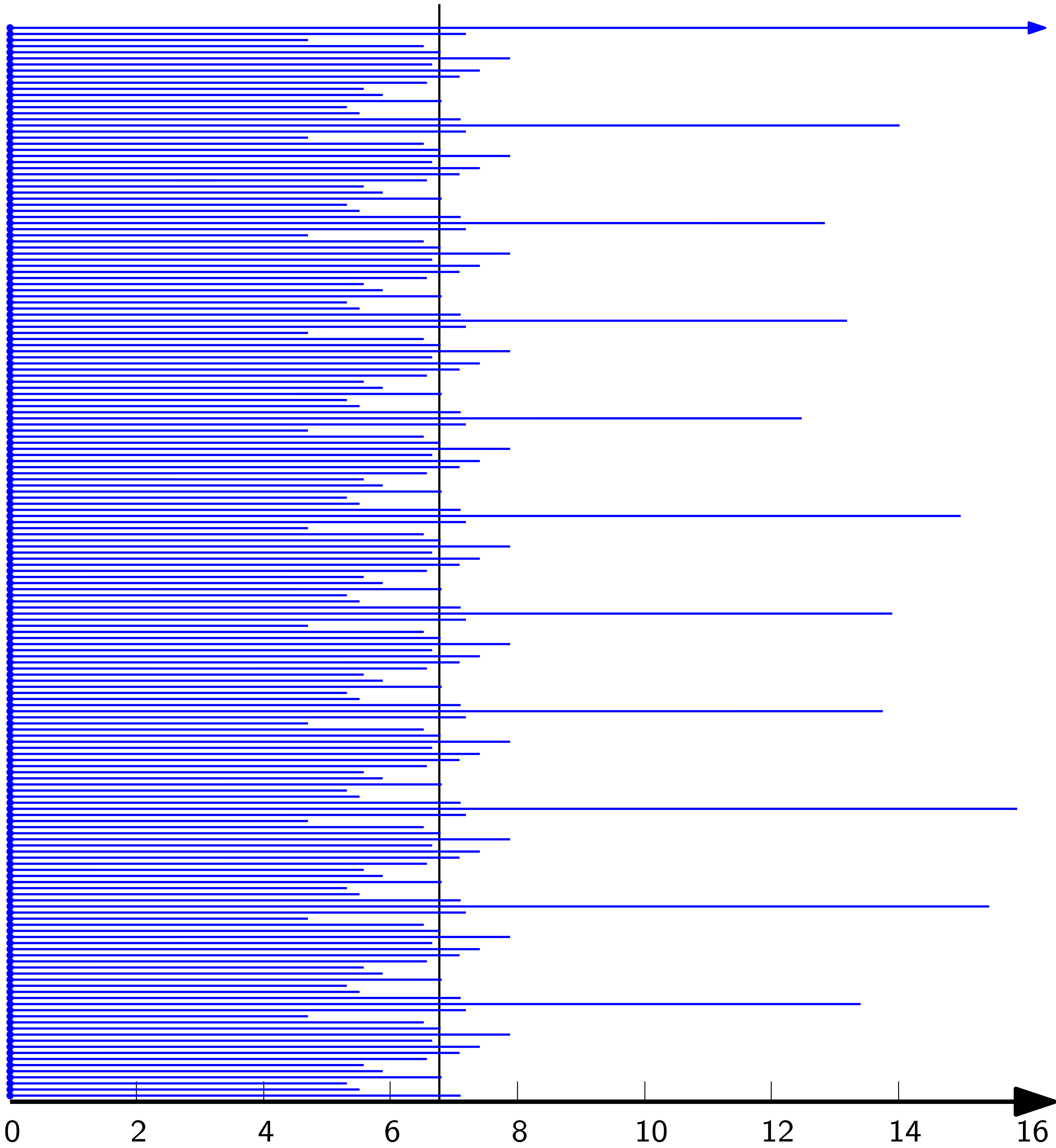




Example: Distance Function

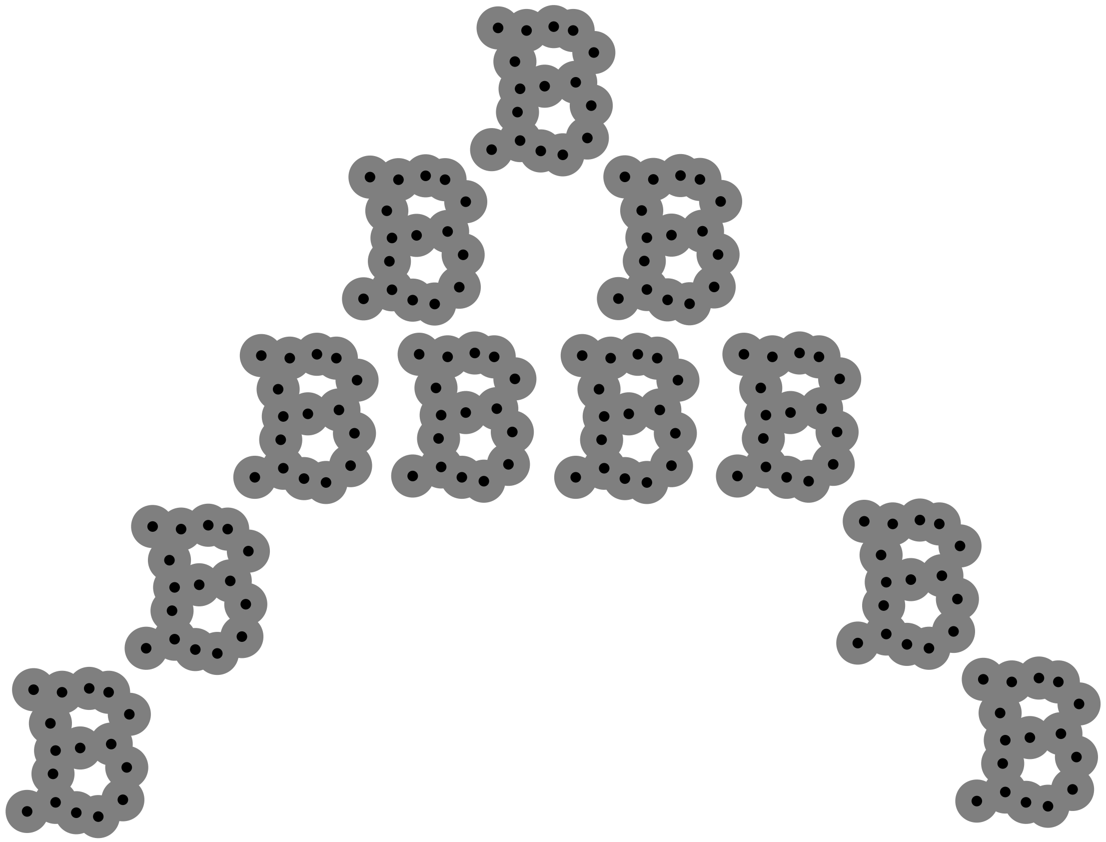
$$f_P : \quad \mathbb{R}^2 \to \mathbb{R}$$
$$x \mapsto \min_{p \in P} \|x - p\|_2$$

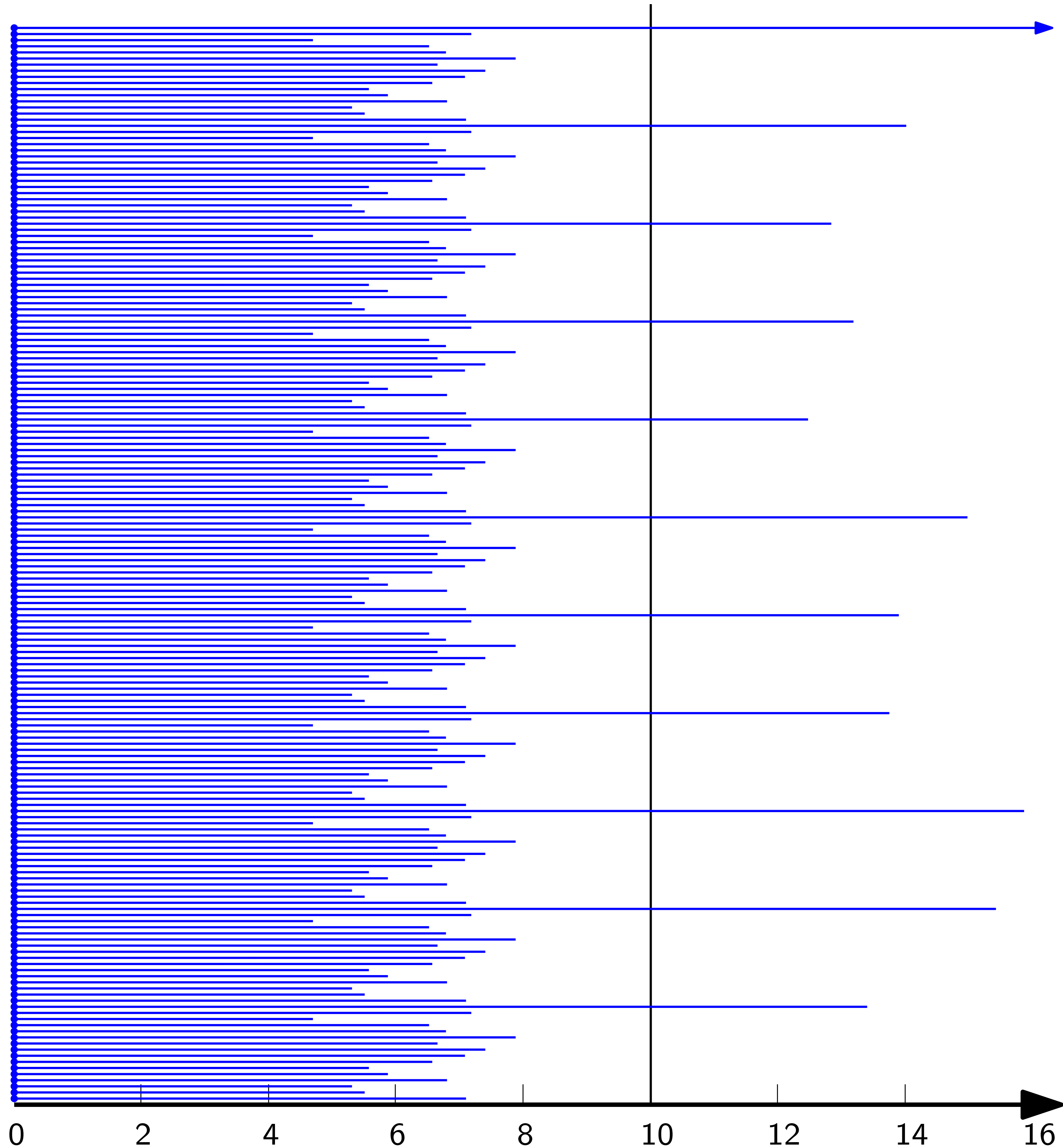




Example: Distance Function

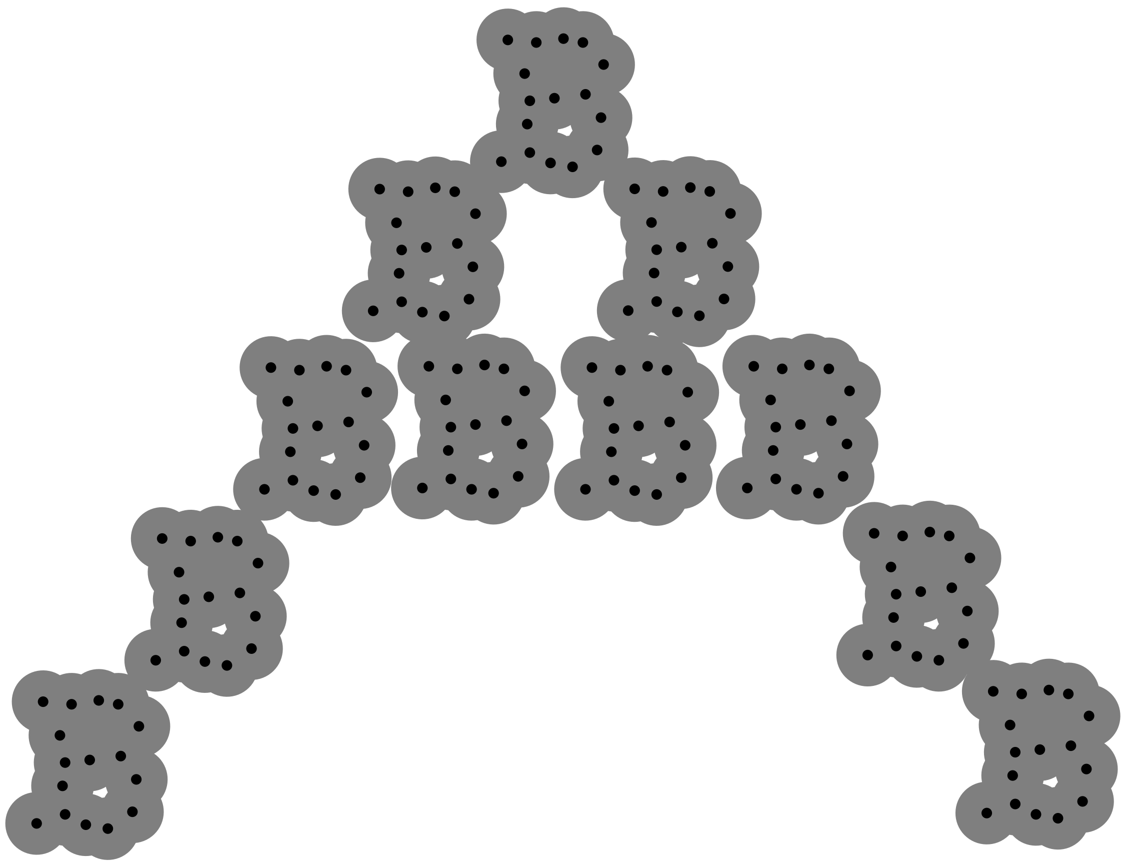
$$f_P : \quad \mathbb{R}^2 \to \mathbb{R}$$
$$x \mapsto \min_{p \in P} \|x - p\|_2$$

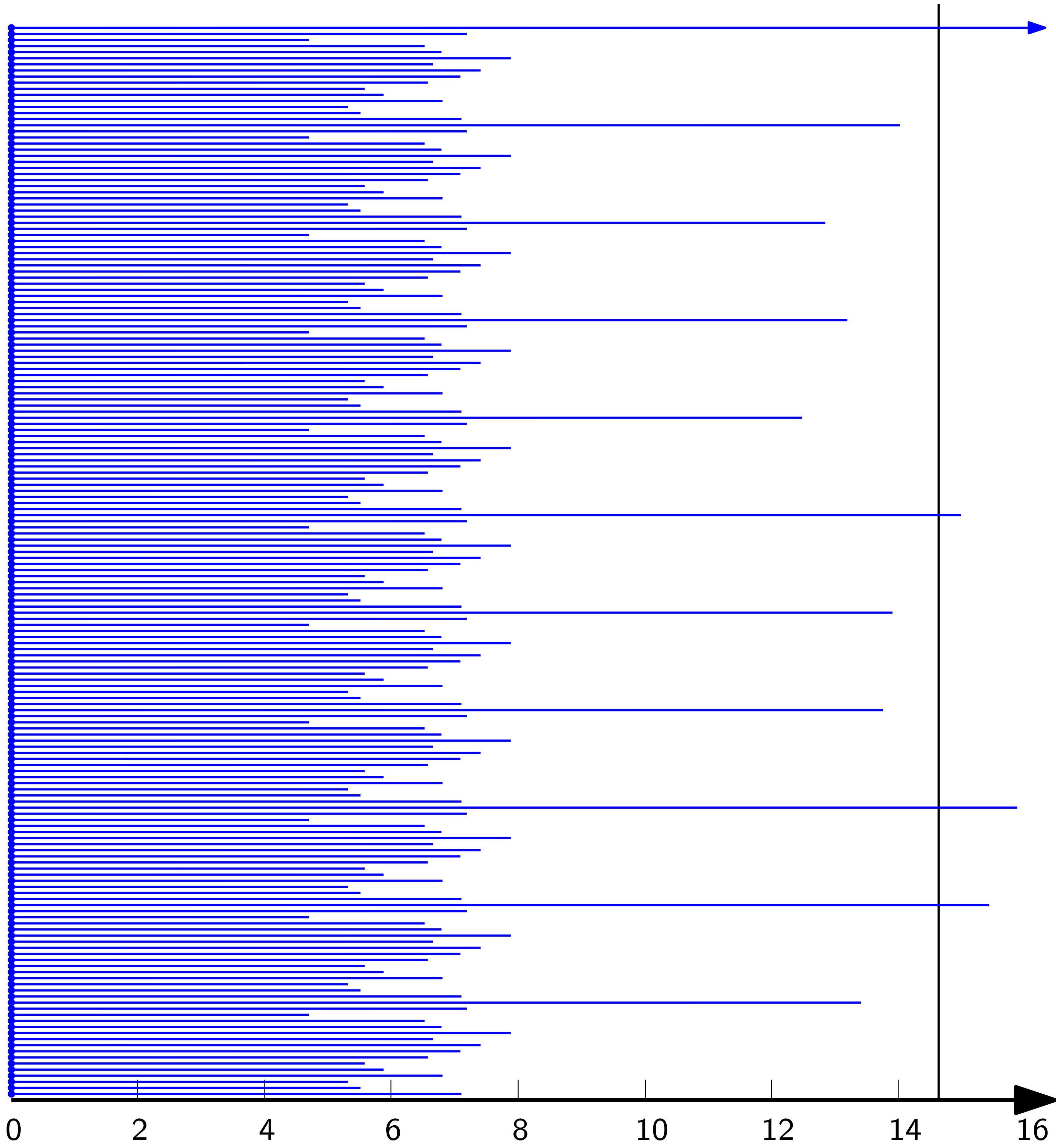




Example: Distance Function

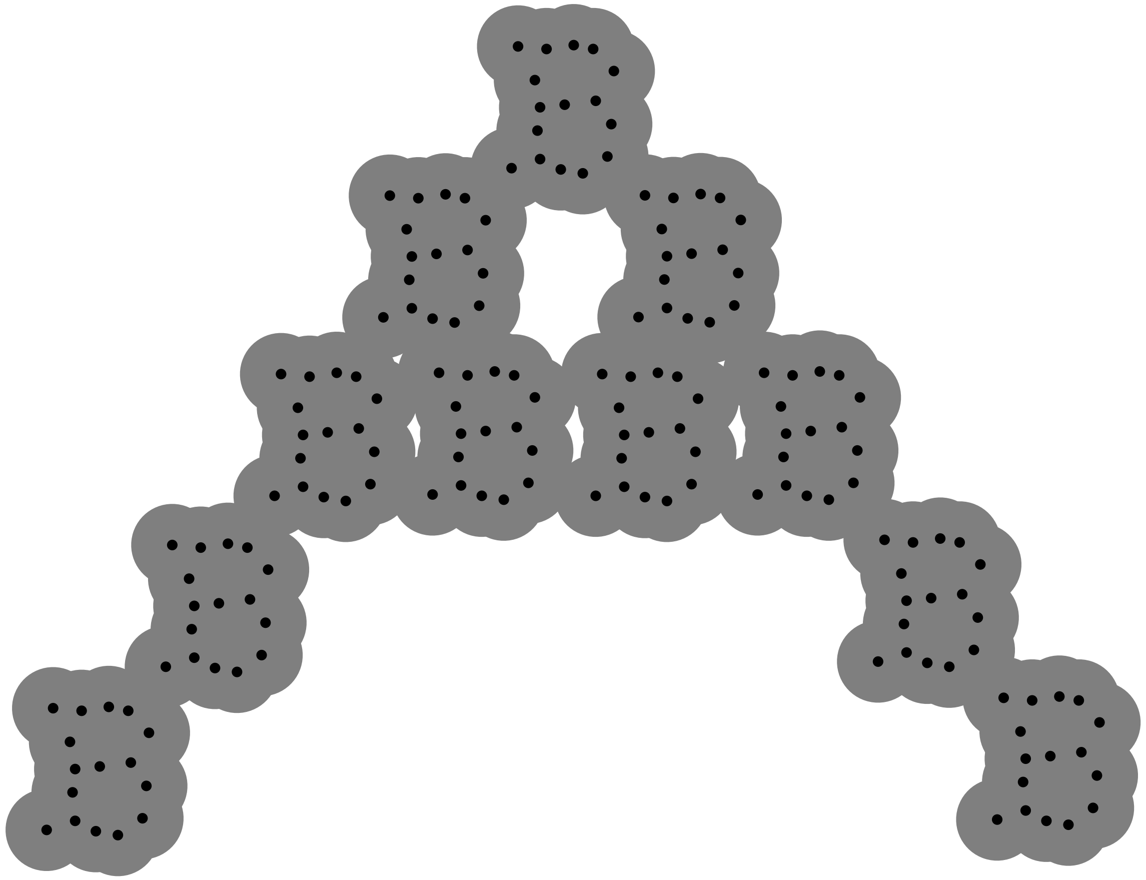
$$f_P: \quad \mathbb{R}^2 \to \mathbb{R}$$
$$x \mapsto \min_{p \in P} \|x - p\|_2$$

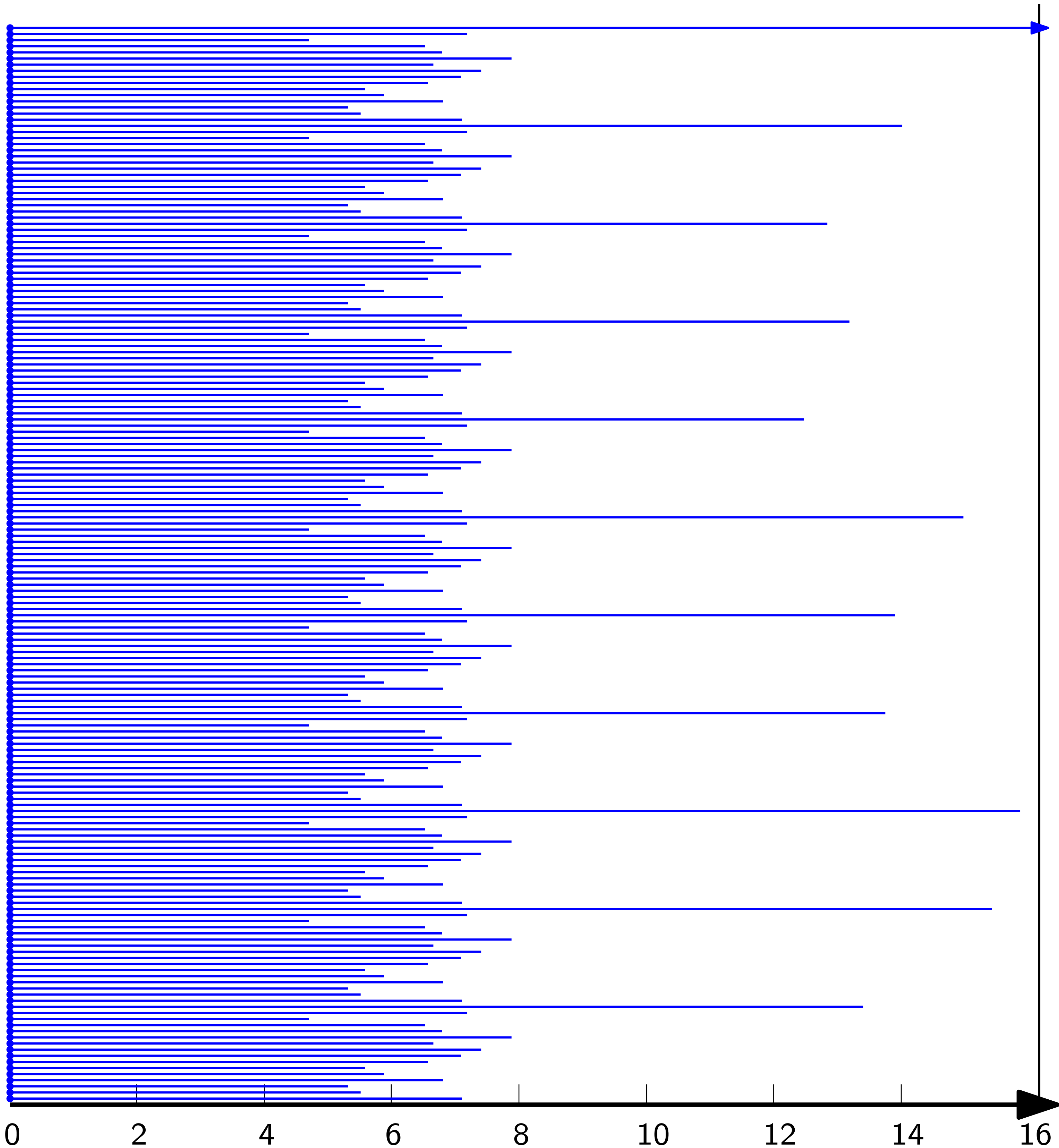




Example: Distance Function

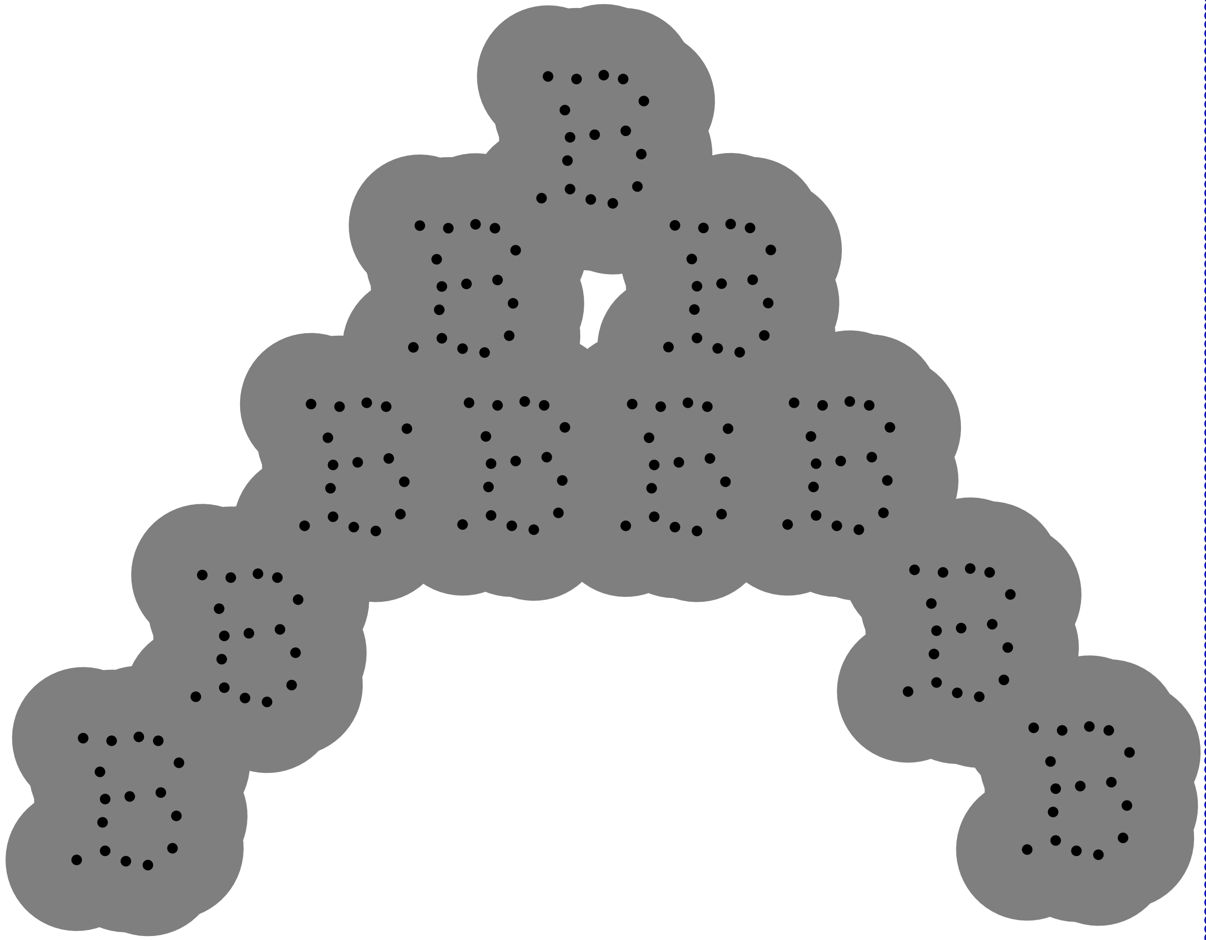
$$f_P: \quad \mathbb{R}^2 \to \mathbb{R}$$
$$x \mapsto \min_{p \in P} \|x - p\|_2$$

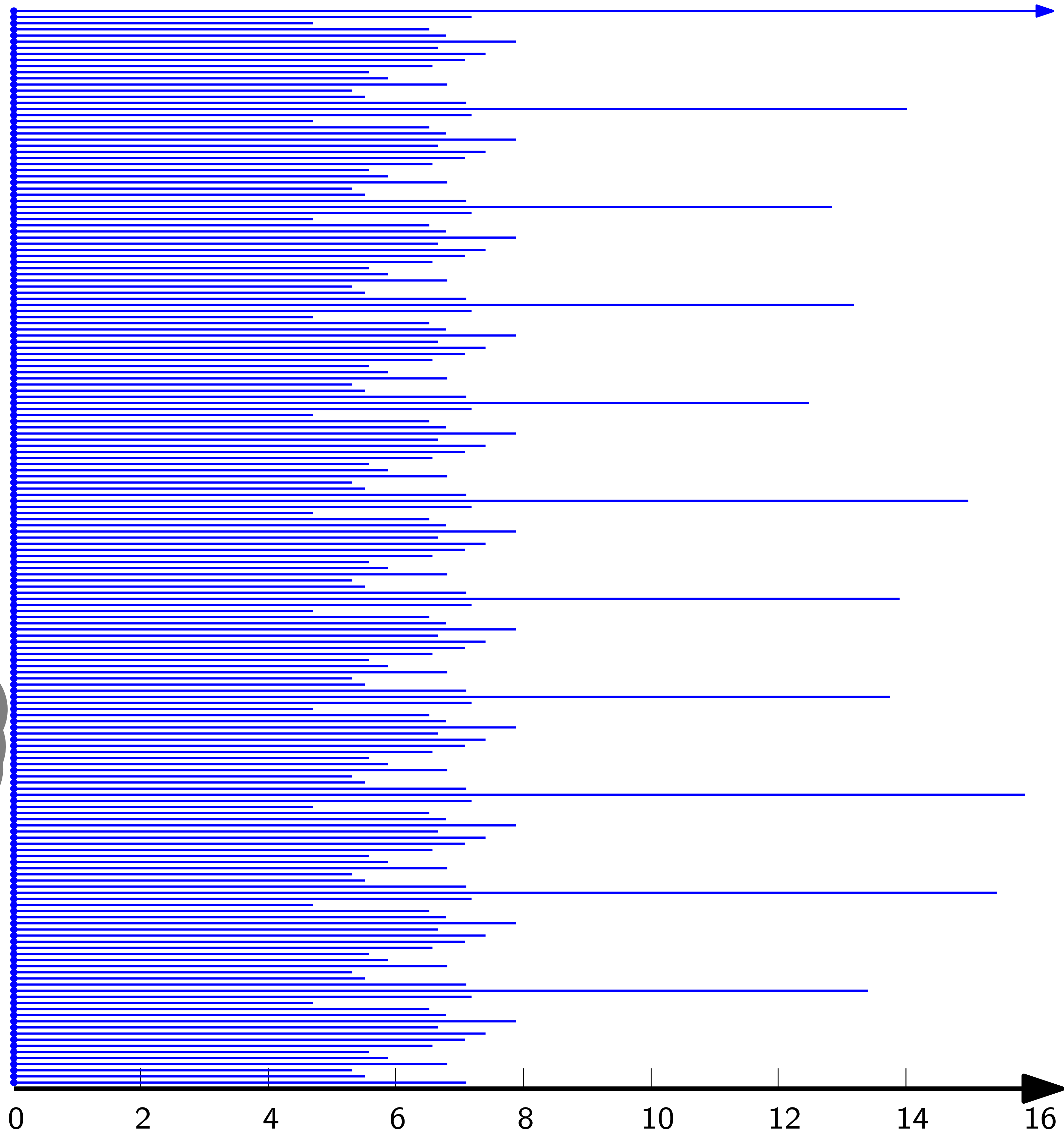




Example: Distance Function

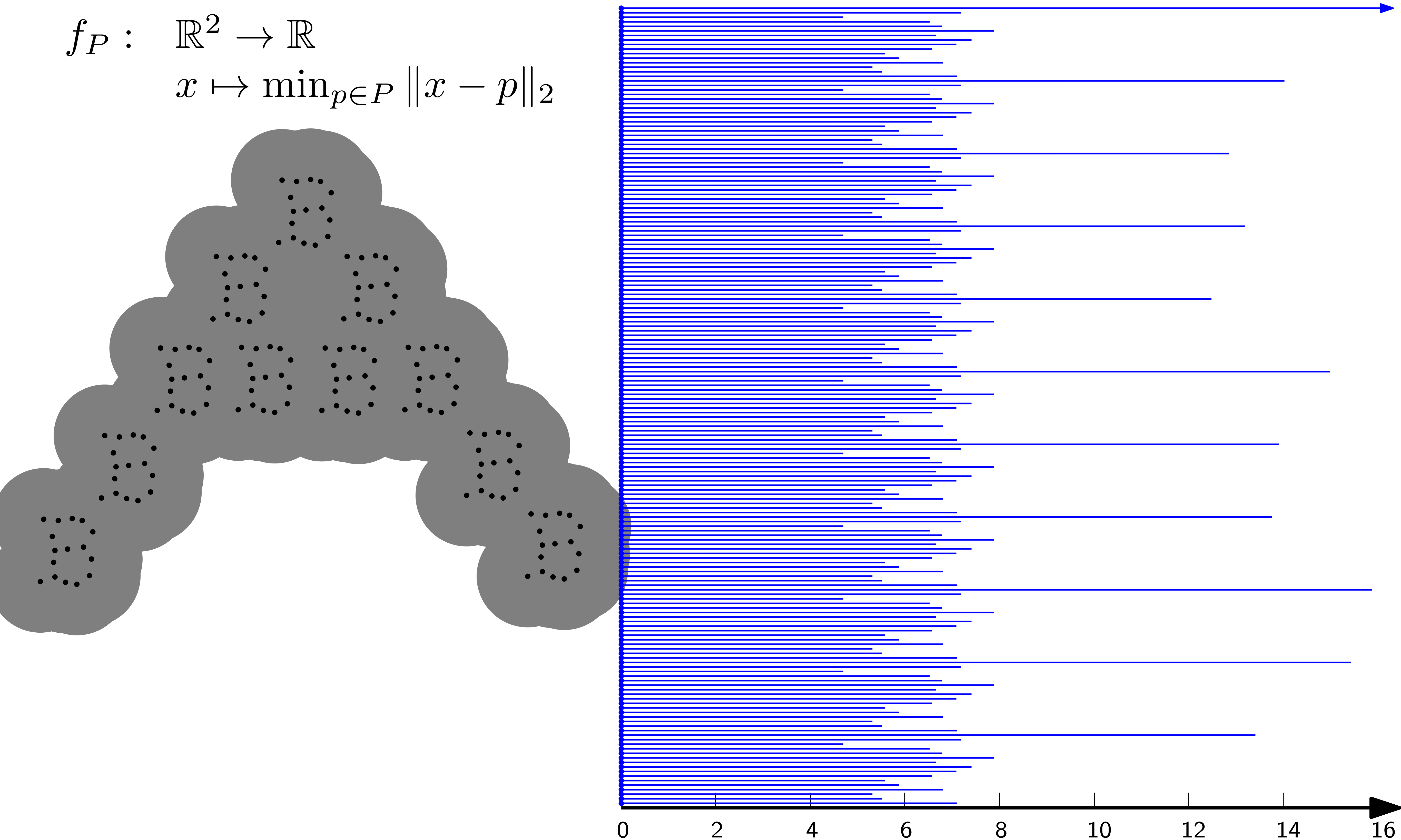
$$\begin{aligned} f_P : \quad & \mathbb{R}^2 \to \mathbb{R} \\ & x \mapsto \min_{p \in P} \|x - p\|_2 \end{aligned}$$





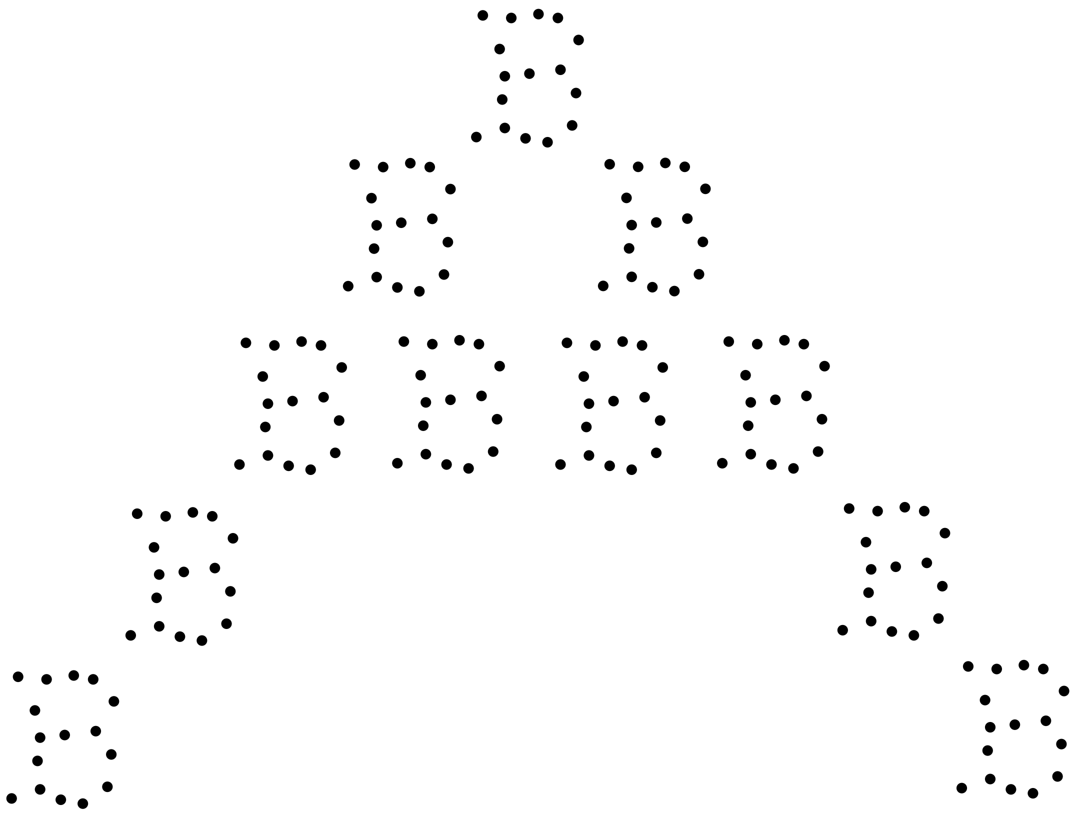
Example: Distance Function

$$f_P: \quad \mathbb{R}^2 \to \mathbb{R}$$
$$x \mapsto \min_{p \in P} \|x - p\|_2$$

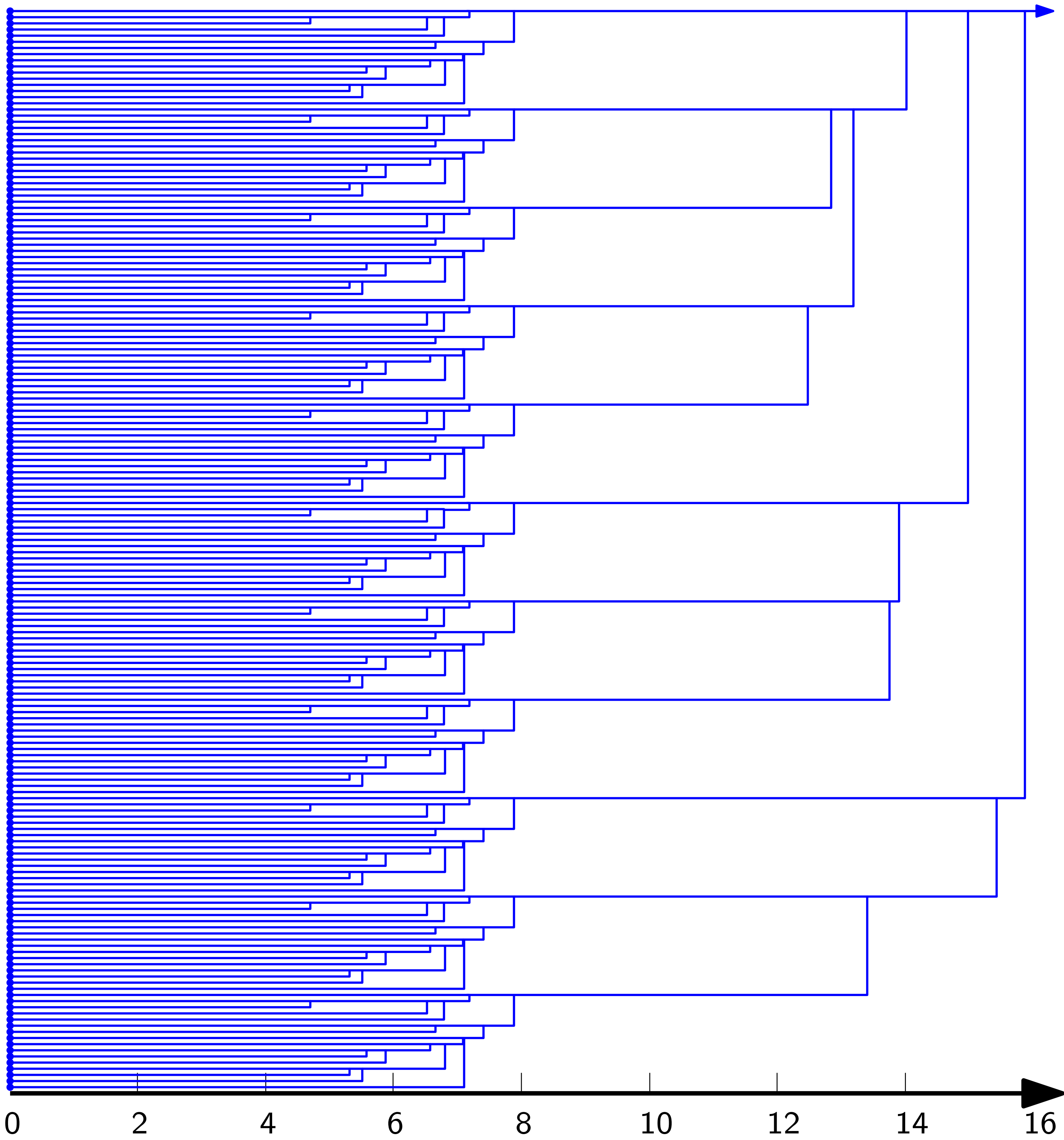


Example: Distance Function

$$f_P : \quad \mathbb{R}^2 \to \mathbb{R}$$
$$x \mapsto \min_{p \in P} \|x - p\|_2$$

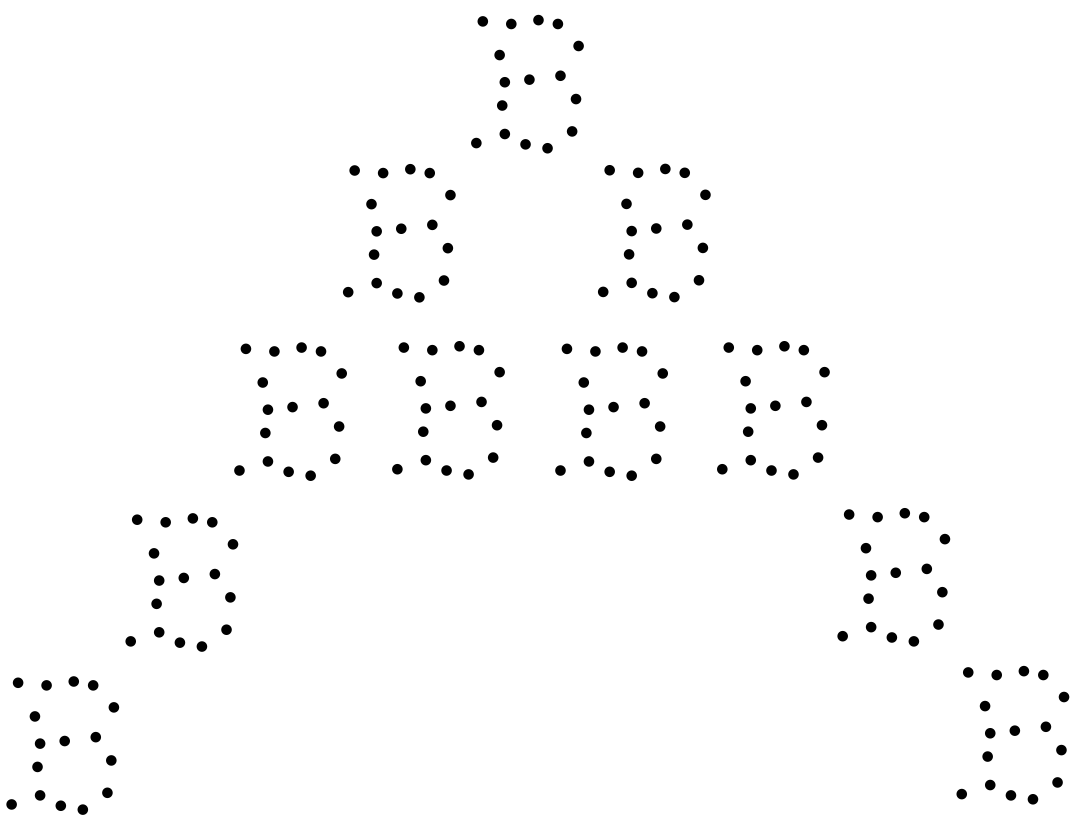


barcode $\to$ merge tree

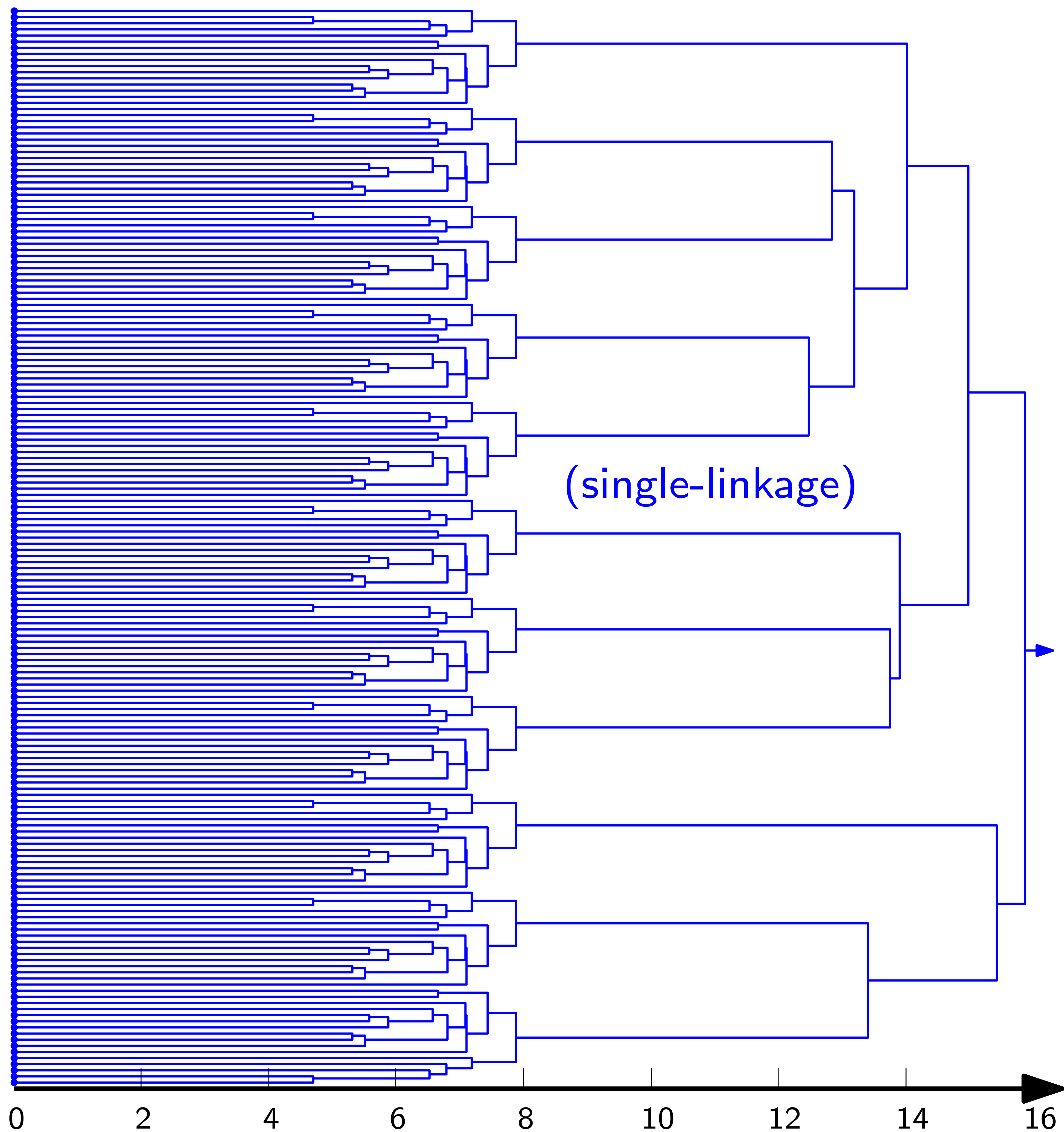


Example: Distance Function

$$f_P: \quad \mathbb{R}^2 \to \mathbb{R}$$
$$x \mapsto \min_{p \in P} \|x - p\|_2$$



barcode $\to$ merge tree $\to$ dendrogram



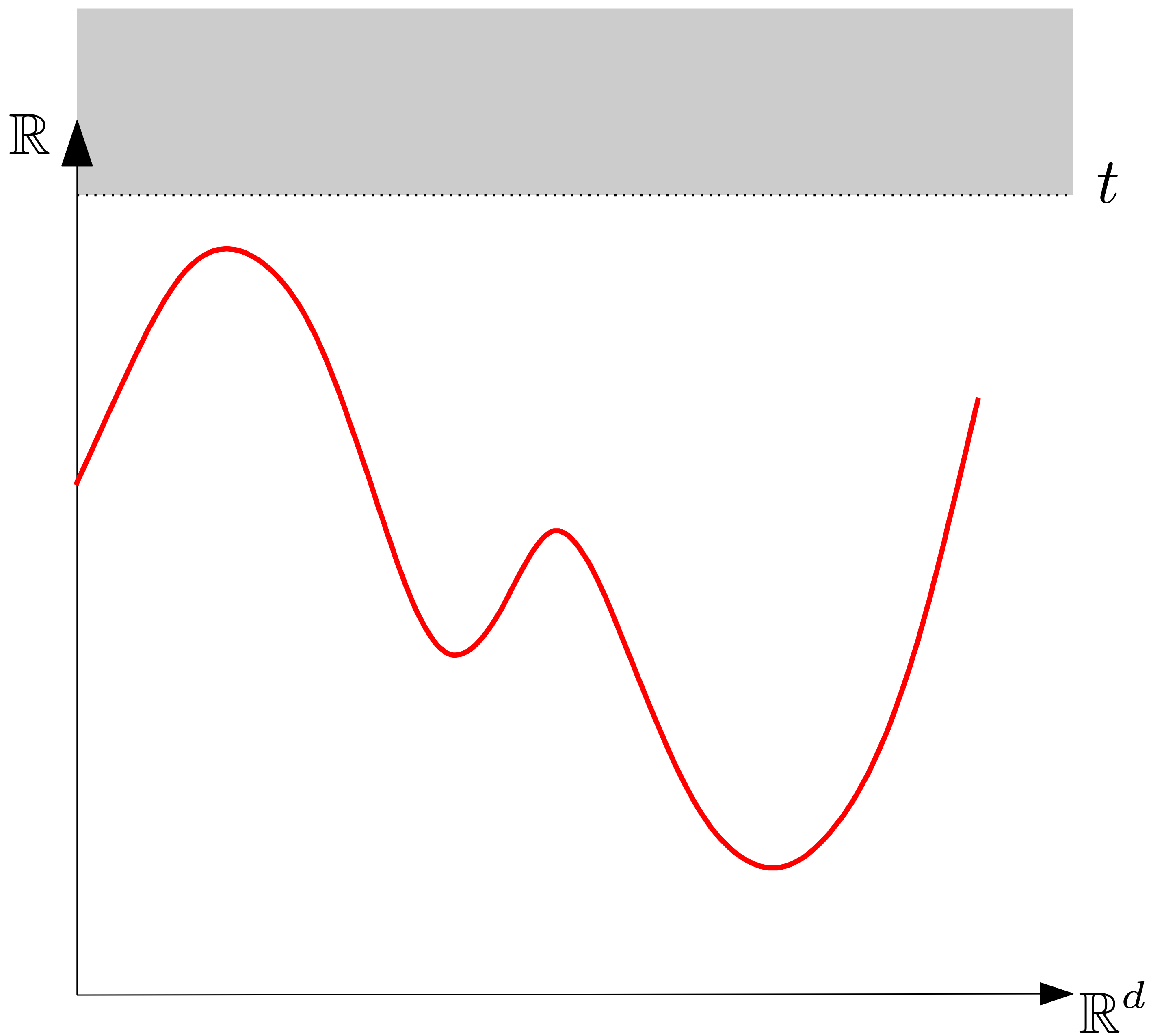
Back to Mode Seeking

(use density estimator instead of distance function)

Persistence for Mode Seeking

Given a probability density f:

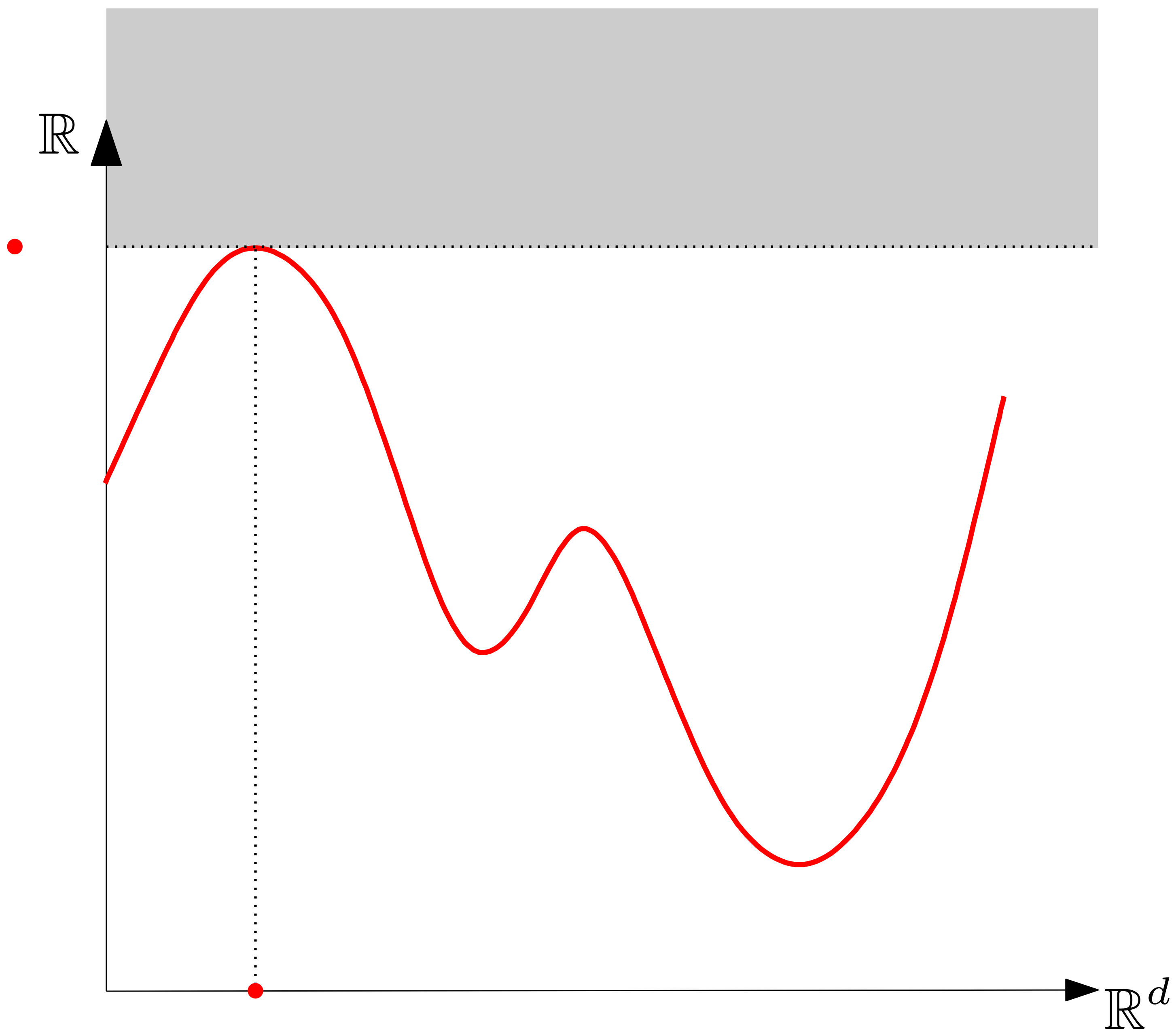
- Nested family (filtration) of **superlevel-sets** $f^{-1}([t, +\infty))$ for t from $+\infty$ to $-\infty$.
- Track evolution of connected components throughout the family.



Persistence for Mode Seeking

Given a probability density f:

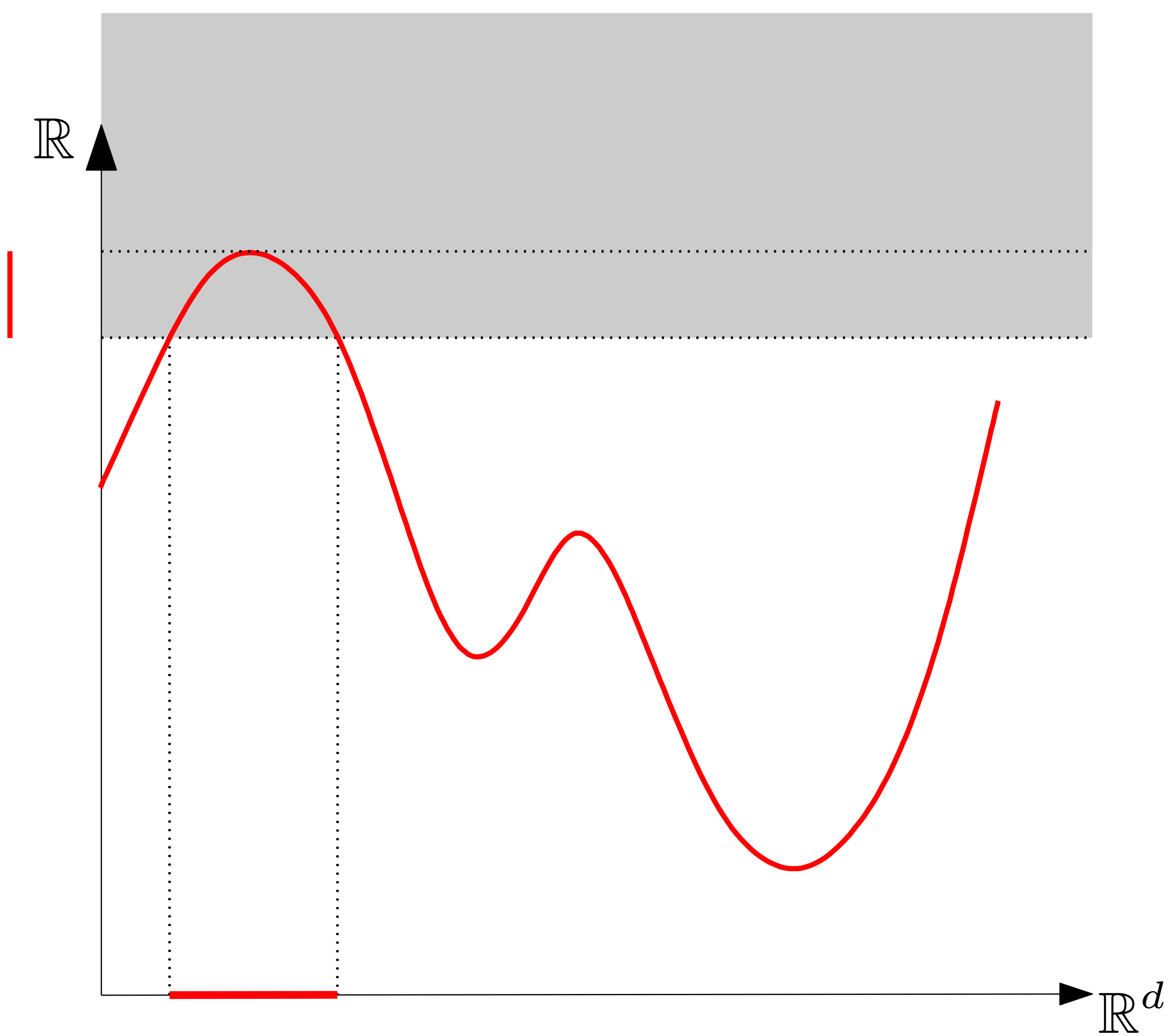
- Nested family (filtration) of **superlevel-sets** $f^{-1}([t, +\infty))$ for t from $+\infty$ to $-\infty$.
- Track evolution of connected components throughout the family.



Persistence for Mode Seeking

Given a probability density f:

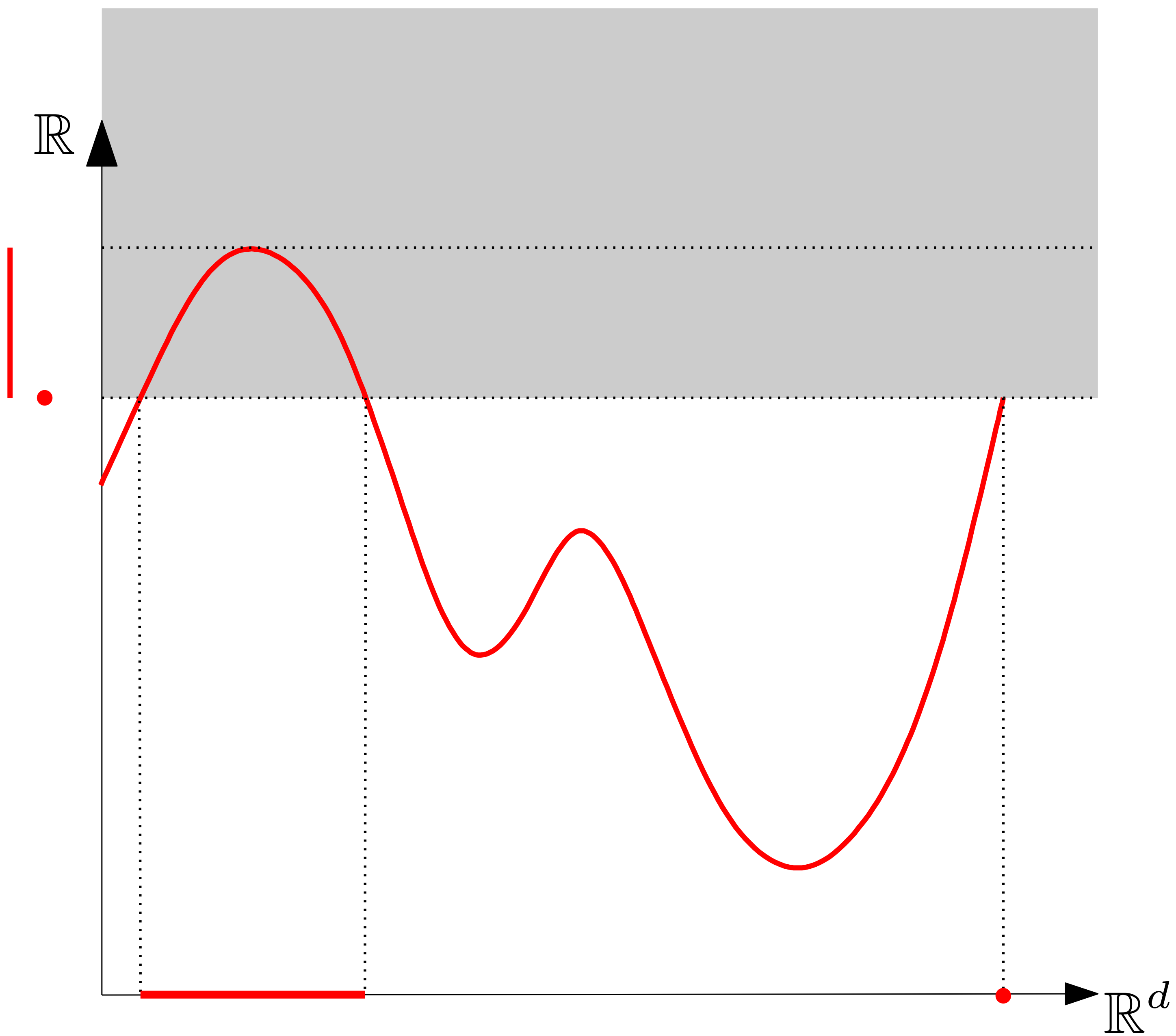
- Nested family (filtration) of **superlevel-sets** $f^{-1}([t, +\infty))$ for t from $+\infty$ to $-\infty$.
- Track evolution of connected components throughout the family.



Persistence for Mode Seeking

Given a probability density f:

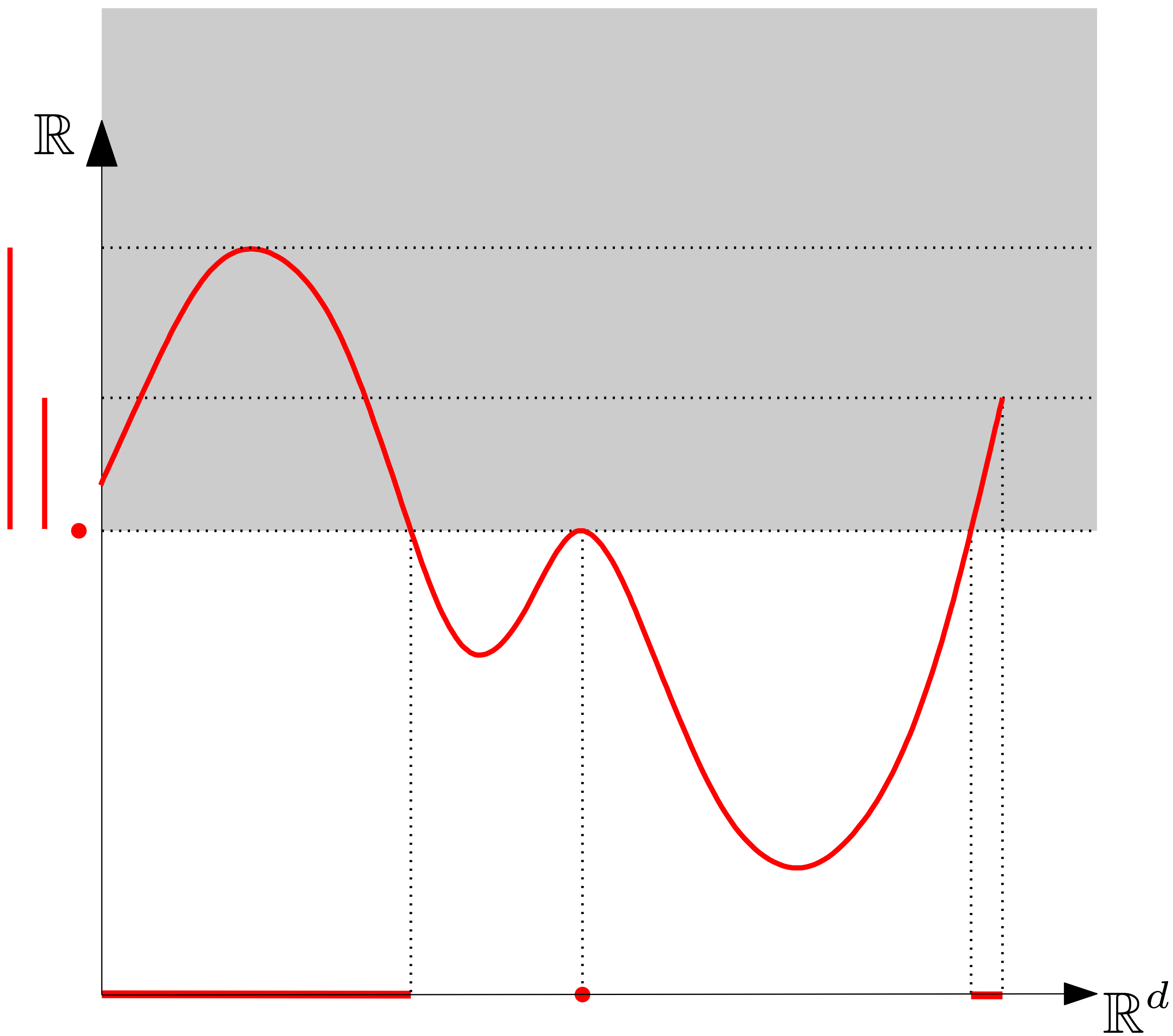
- Nested family (filtration) of **superlevel-sets** $f^{-1}([t, +\infty))$ for t from $+\infty$ to $-\infty$.
- Track evolution of connected components throughout the family.



Persistence for Mode Seeking

Given a probability density f:

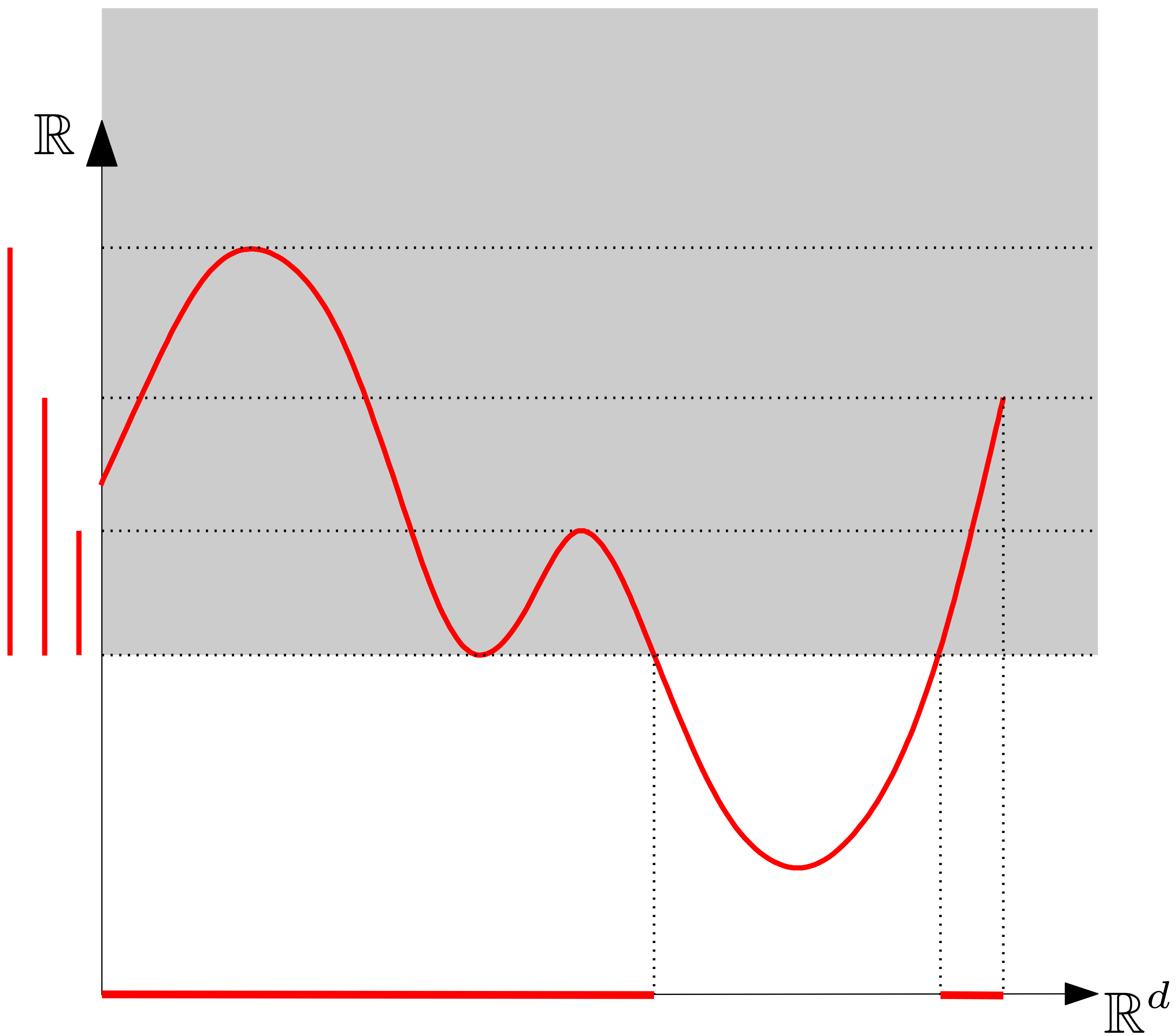
- Nested family (filtration) of **superlevel-sets** $f^{-1}([t, +\infty))$ for t from $+\infty$ to $-\infty$.
- Track evolution of connected components throughout the family.



Persistence for Mode Seeking

Given a probability density f:

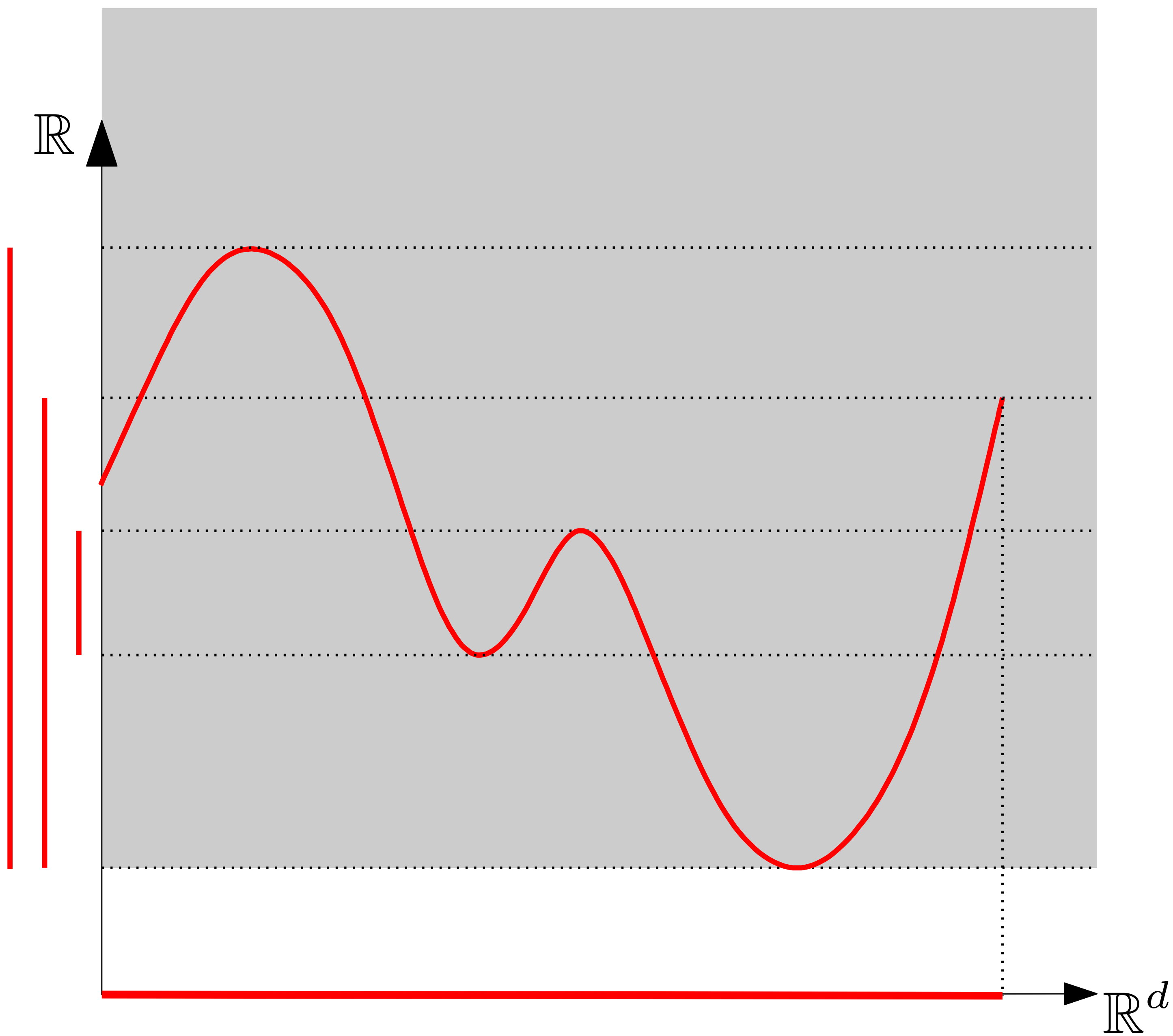
- Nested family (filtration) of **superlevel-sets** $f^{-1}([t, +\infty))$ for t from $+\infty$ to $-\infty$.
- Track evolution of connected components throughout the family.



Persistence for Mode Seeking

Given a probability density f:

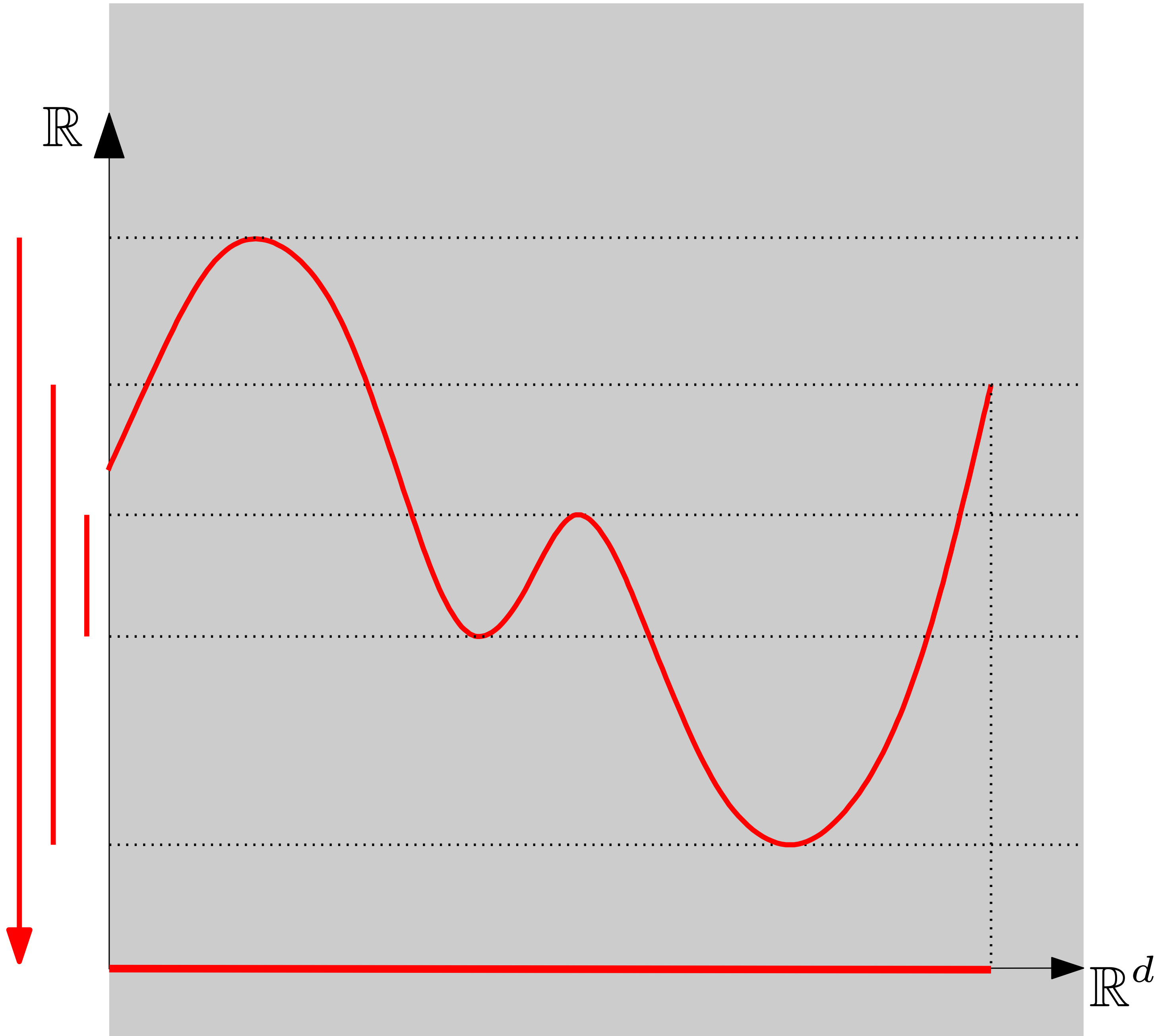
- Nested family (filtration) of **superlevel-sets** $f^{-1}([t, +\infty))$ for t from $+\infty$ to $-\infty$.
- Track evolution of connected components throughout the family.



Persistence for Mode Seeking

Given a probability density f:

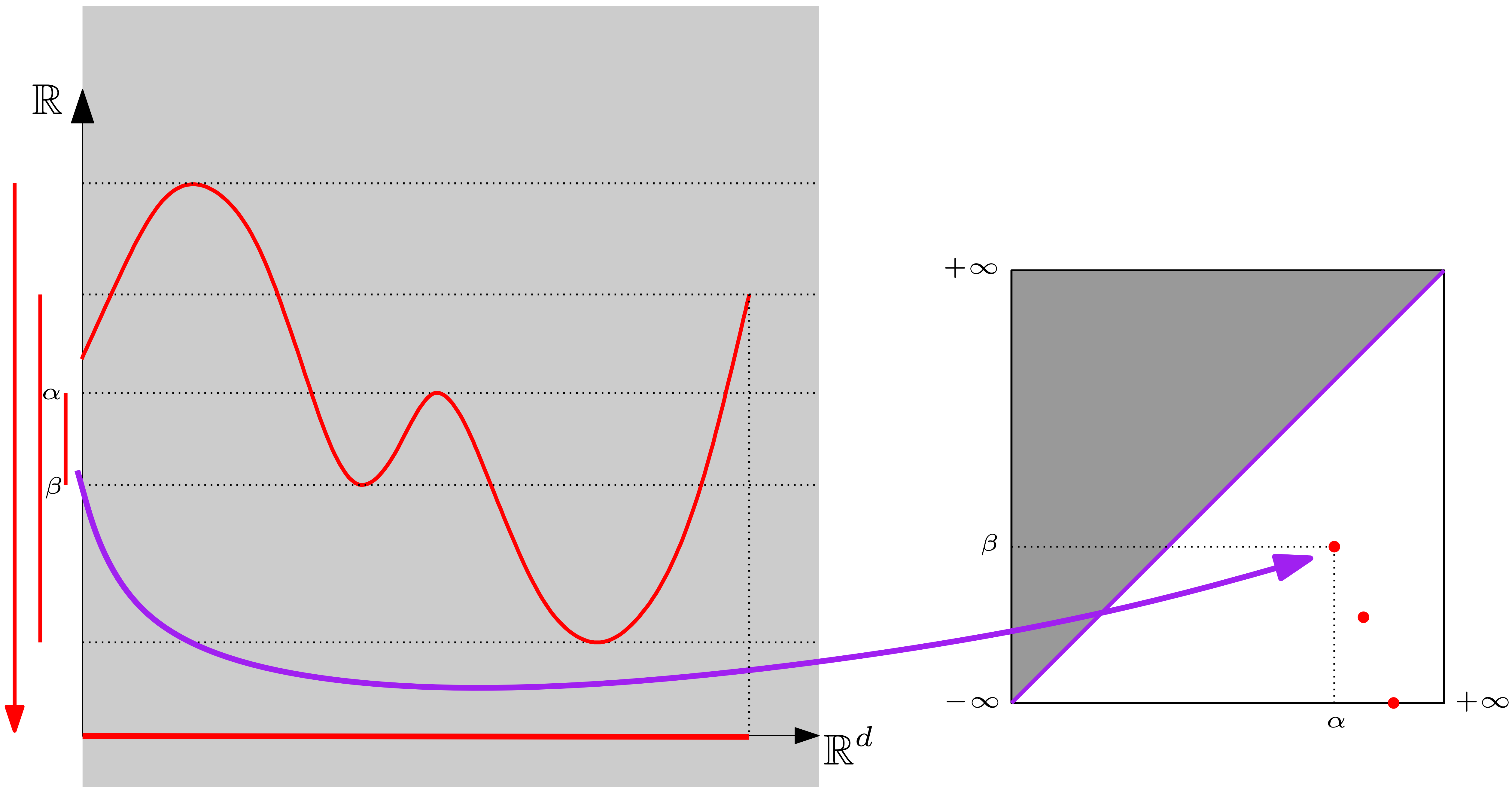
- Nested family (filtration) of **superlevel-sets** $f^{-1}([t, +\infty))$ for t from $+\infty$ to $-\infty$.
- Track evolution of connected components throughout the family.
- Finite set of intervals (barcode) encodes births/deaths of topological features.



Persistence for Mode Seeking

Given a probability density f:

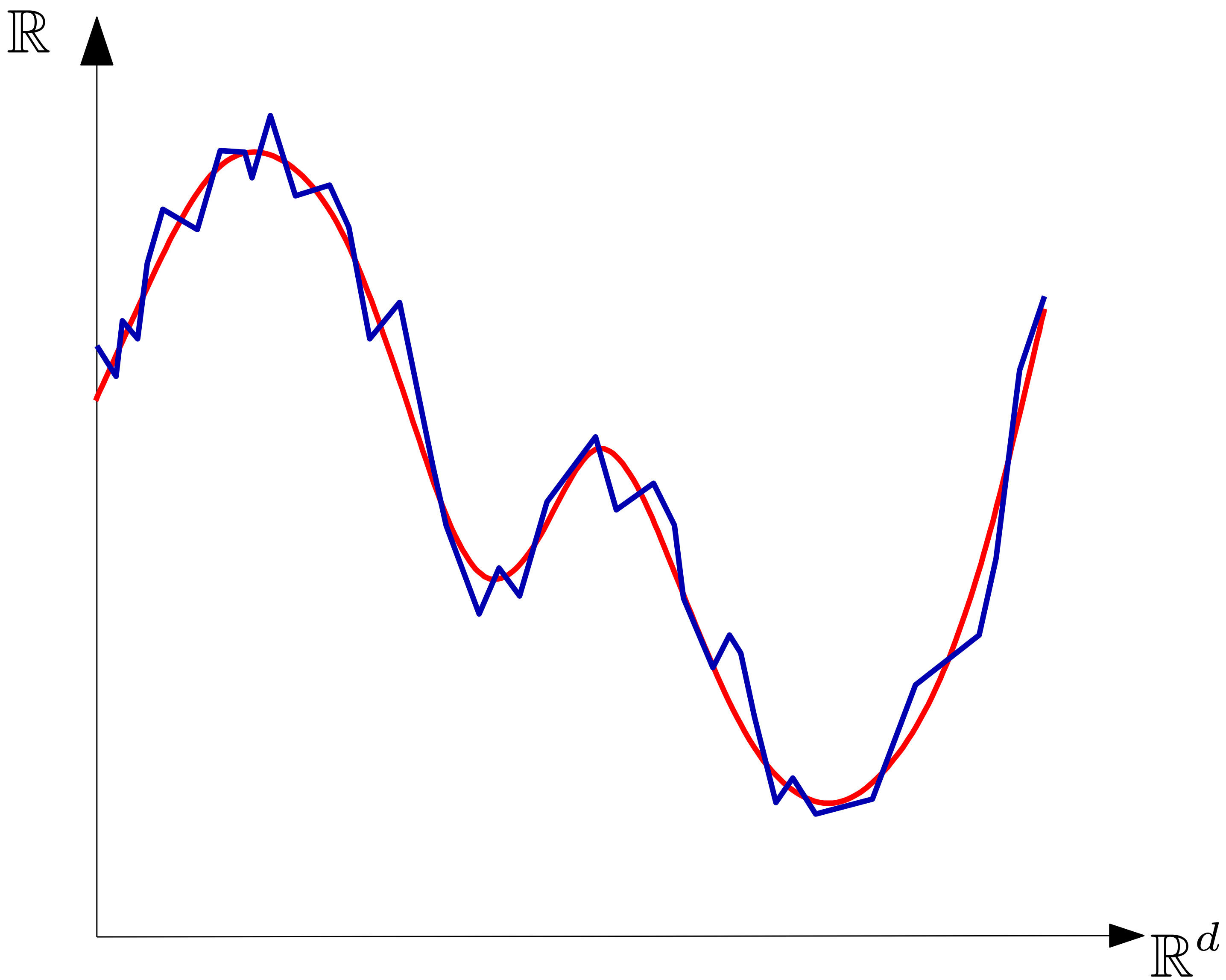
- Nested family (filtration) of **superlevel-sets** $f^{-1}([t, +\infty))$ for t from $+\infty$ to $-\infty$.
- Track evolution of connected components throughout the family.
- Finite set of intervals (barcode) encodes births/deaths of topological features.

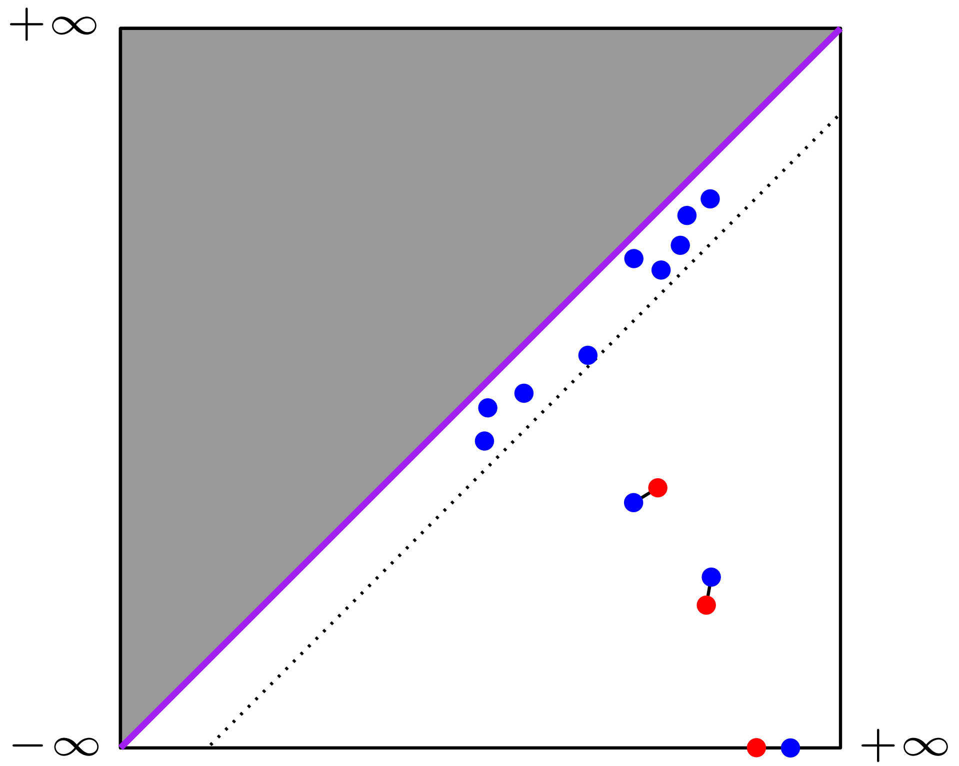


Persistence for Mode Seeking

Given an estimator $\hat{f}$:

Stability Theorem $\Rightarrow \mathrm{d}_B^\infty(\mathrm{Dg}\, f, \mathrm{Dg}\, \hat{f}) \leq \|f - \hat{f}\|_\infty$.

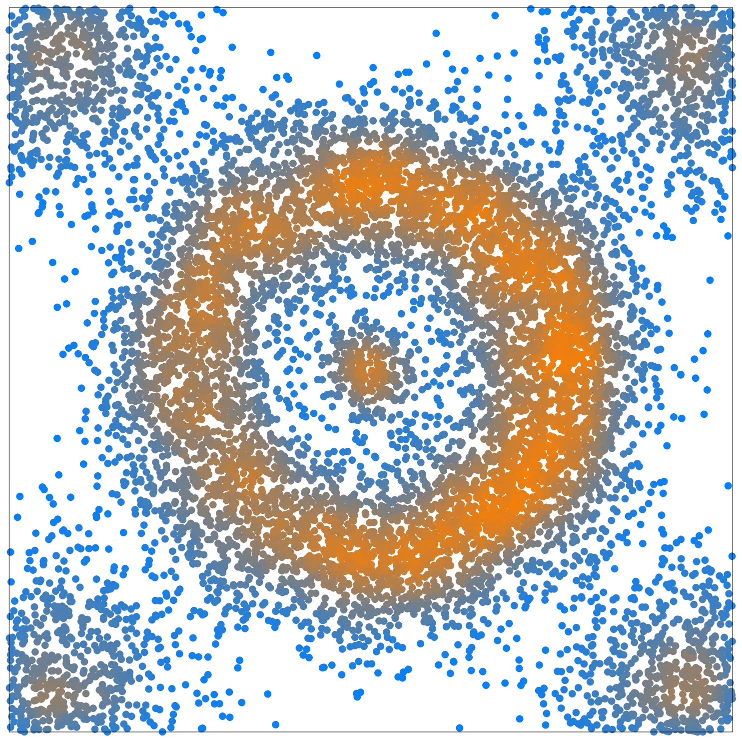




More precisely...

- Density estimator $\hat{f}$ defines an order on the point cloud

  (sort data points by **decreasing** estimated density values)



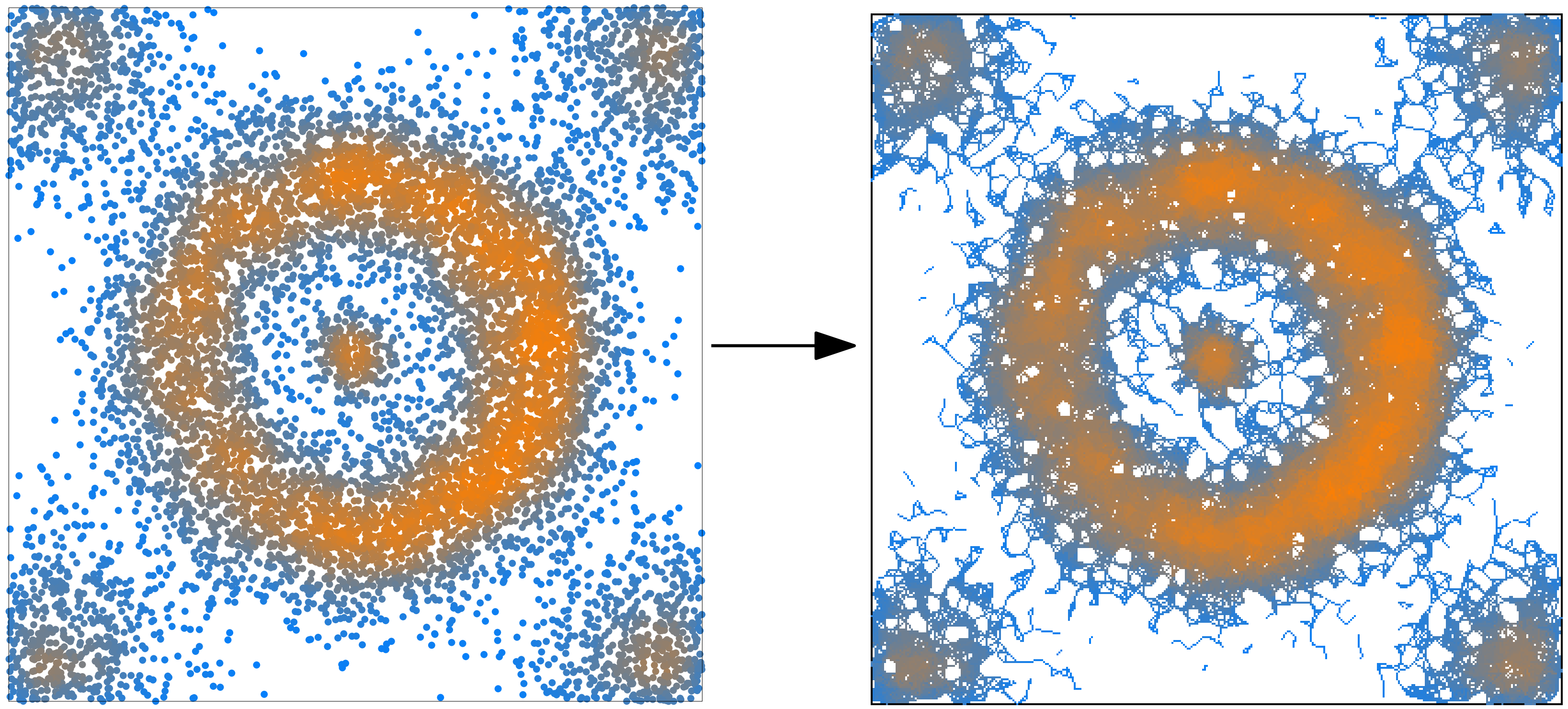
More precisely...

- Density estimator $\hat{f}$ defines an order on the point cloud

  (sort data points by **decreasing** estimated density values)

- Extend order to the graph edges $\rightarrow$ *upper-star filtration*

  $(\hat{f}([u,v]) = \min\{\hat{f}(u),\ \hat{f}(v)\})$



More precisely...

- Density estimator $\hat{f}$ defines an order on the point cloud

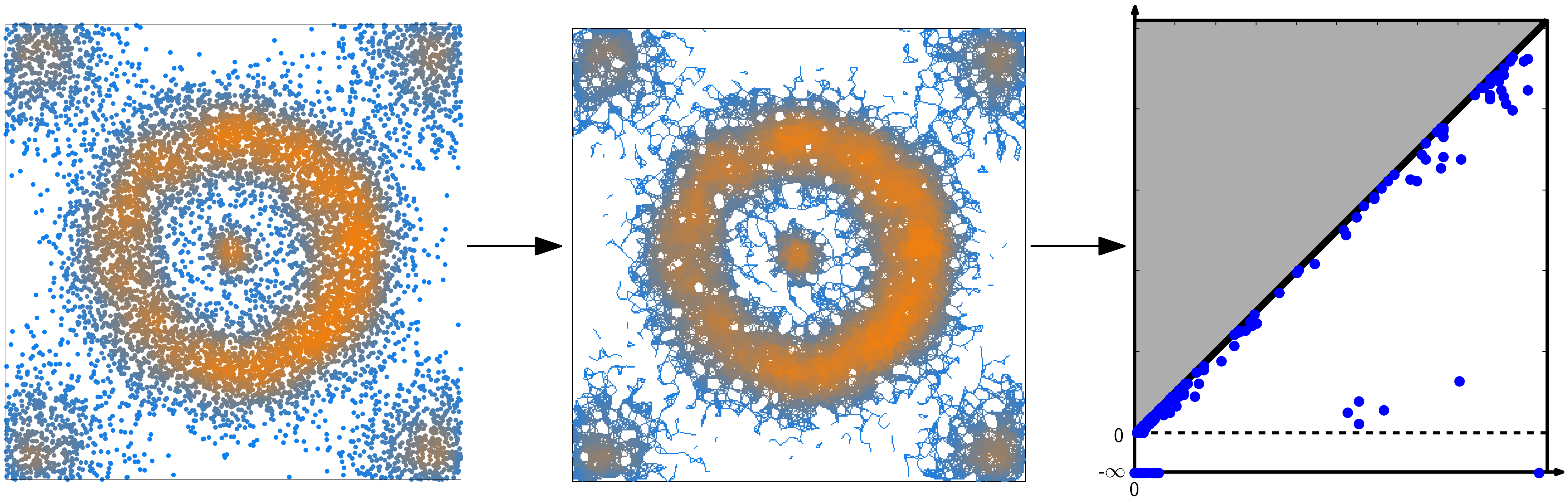
  (sort data points by **decreasing** estimated density values)

- Extend order to the graph edges $\rightarrow$ *upper-star filtration*

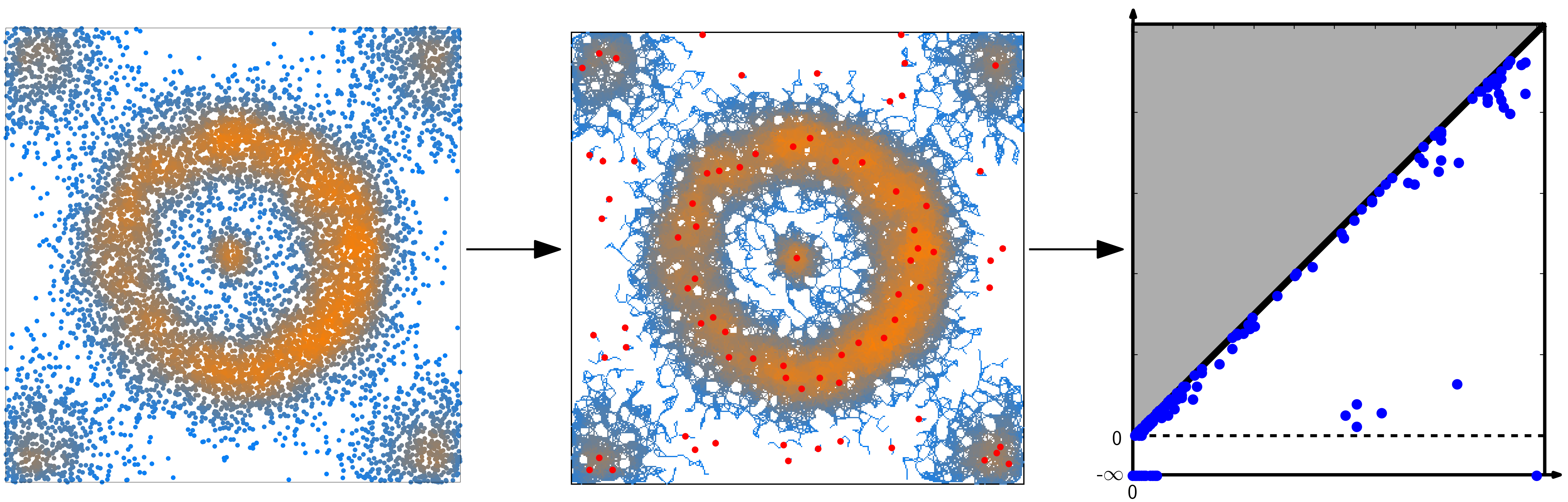
  $(\hat{f}([u,v]) = \min\{\hat{f}(u),\ \hat{f}(v)\})$

- Compute the 0-dimensional persistence diagram of this filtration

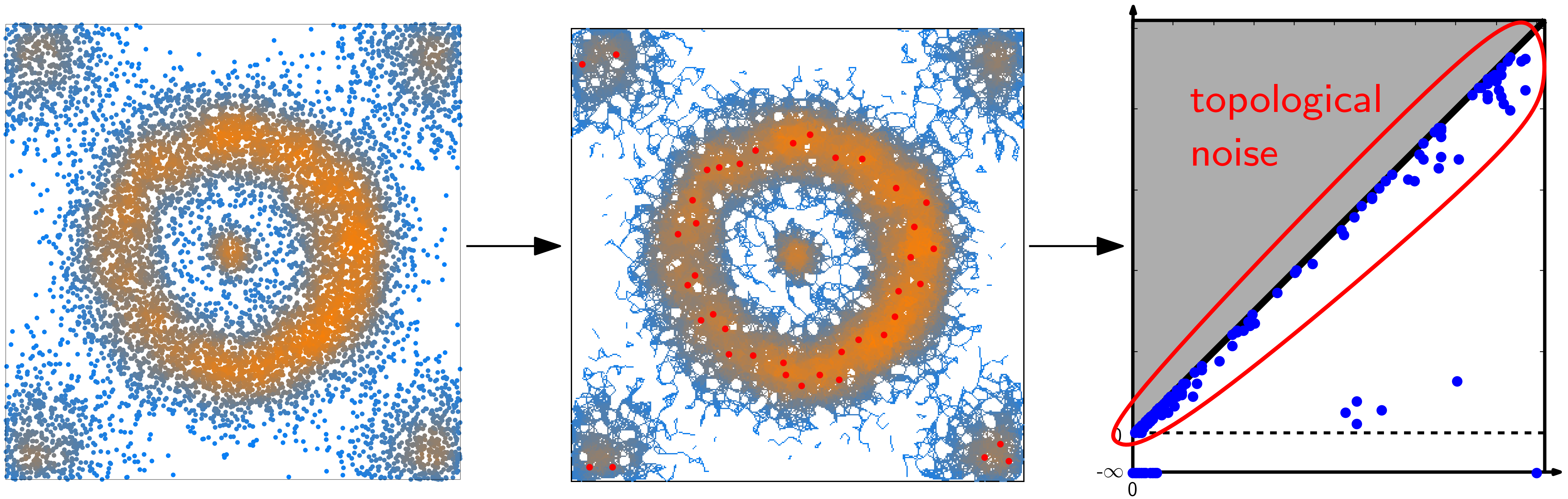
  (apply 0-dimensional persistence algorithm $\rightarrow$ union-find data structure)



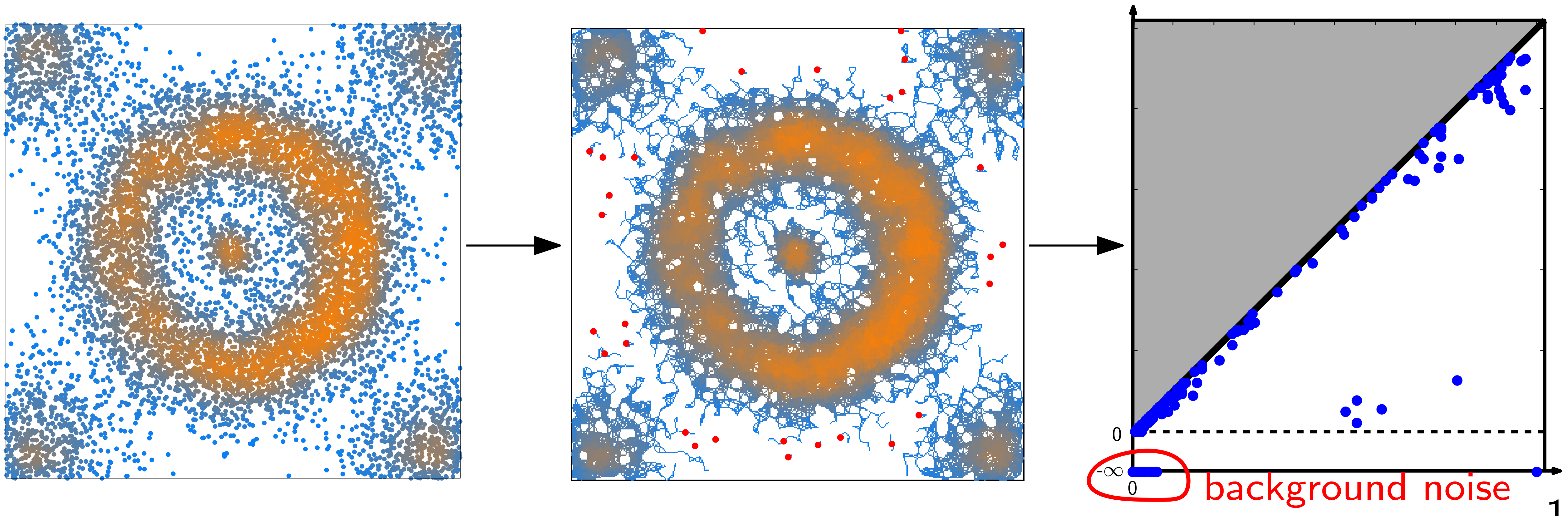
Estimating the Correct Number of Clusters



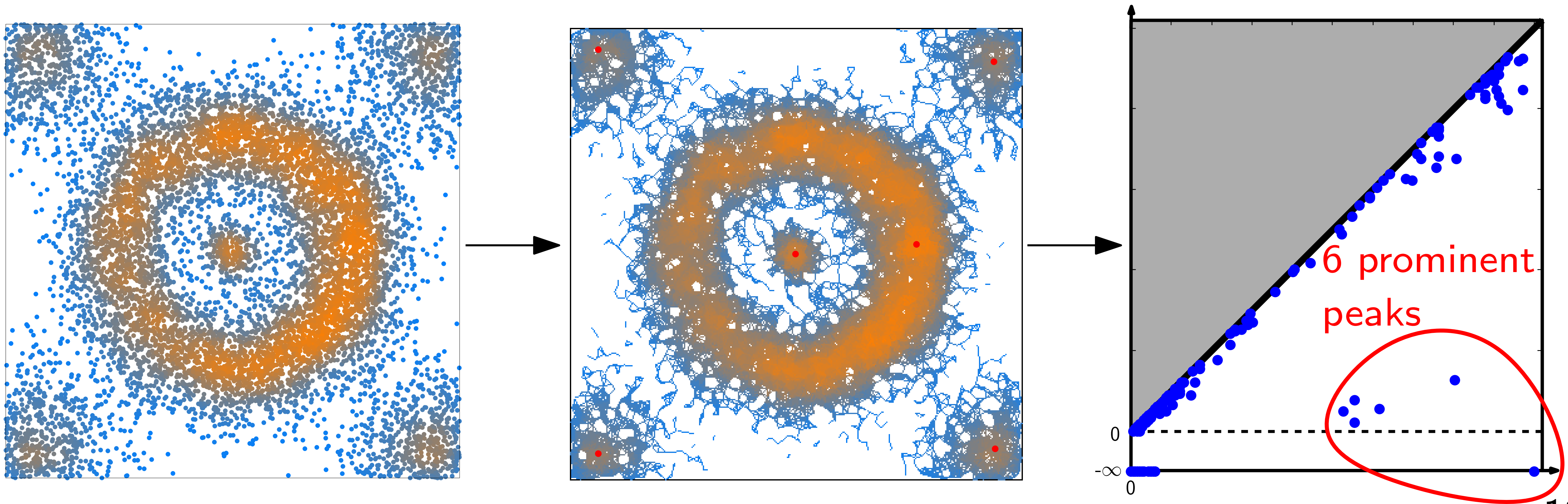
Estimating the Correct Number of Clusters



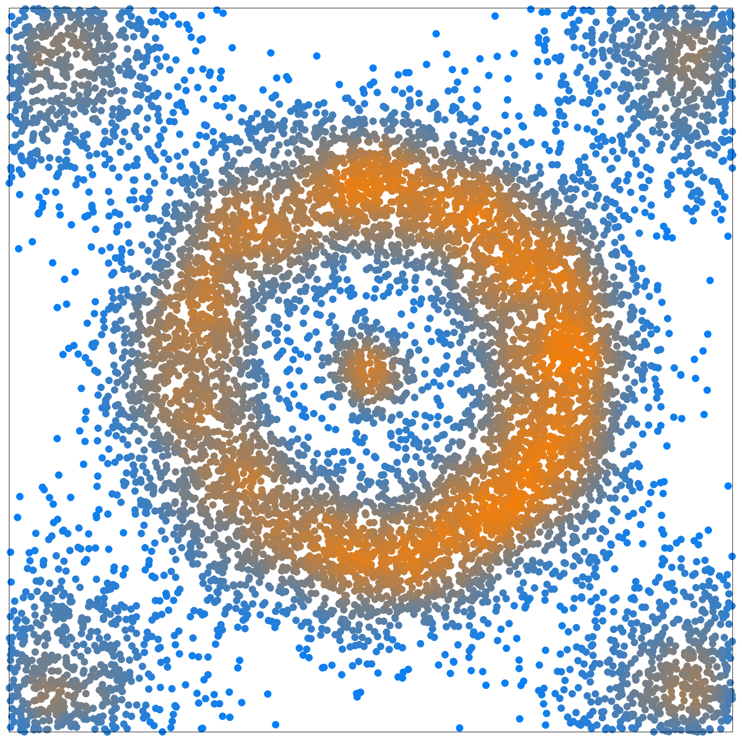
Estimating the Correct Number of Clusters

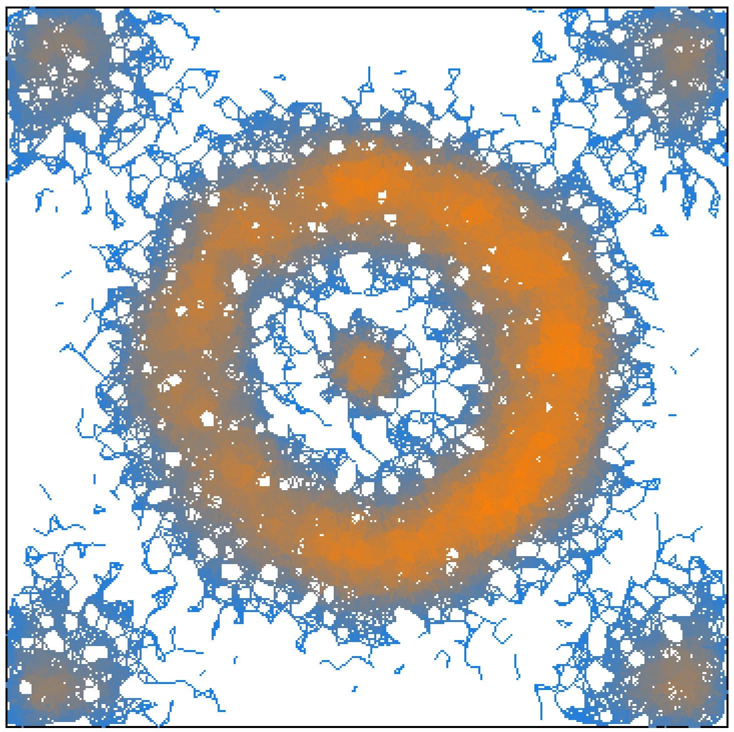


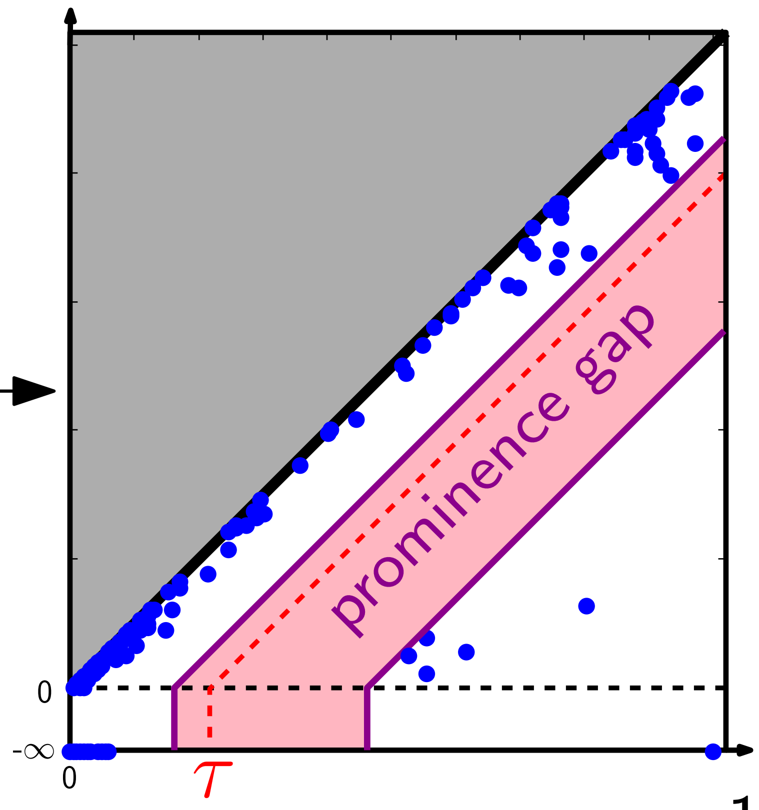
Estimating the Correct Number of Clusters



Estimating the Correct Number of Clusters

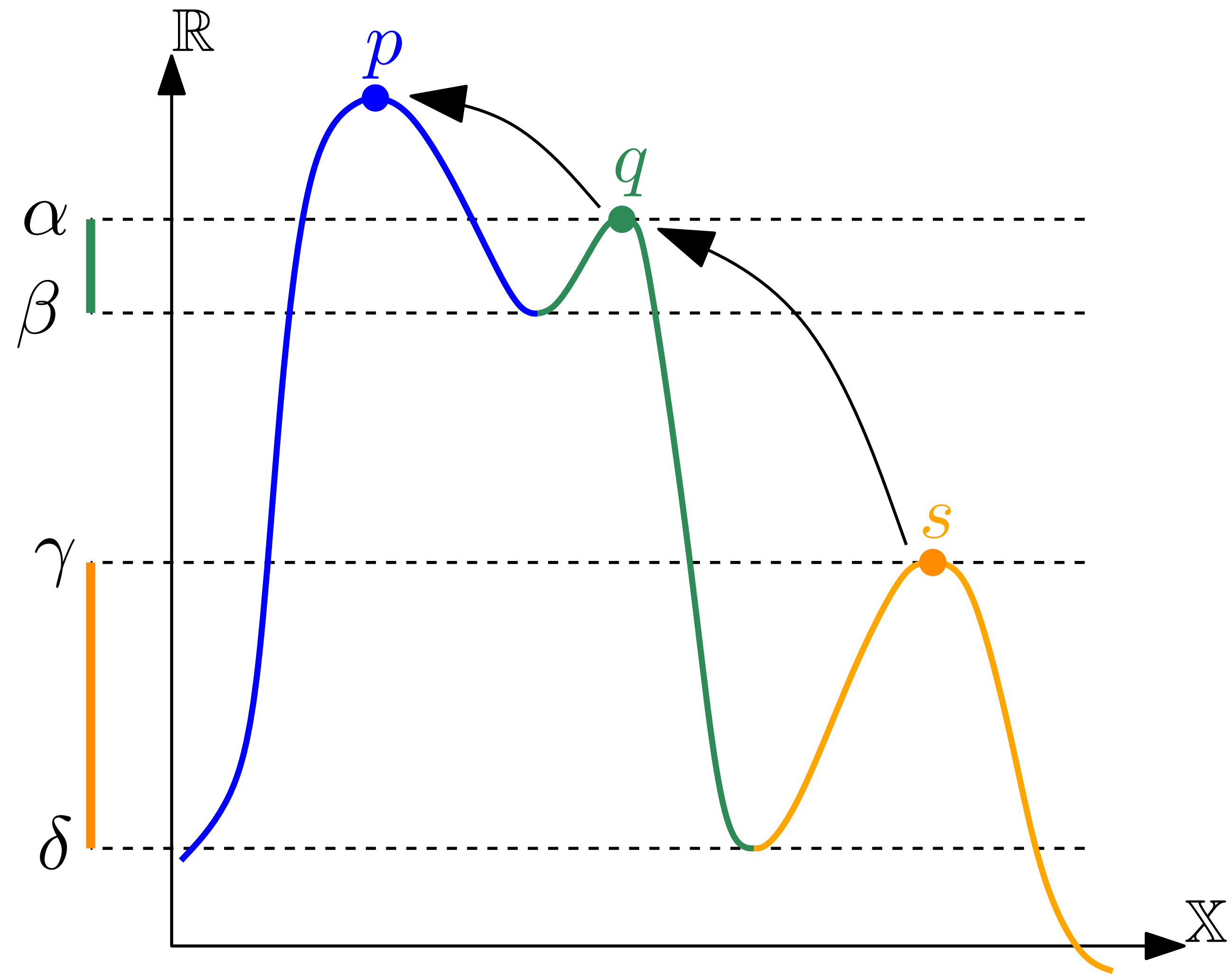






Merging Clusters

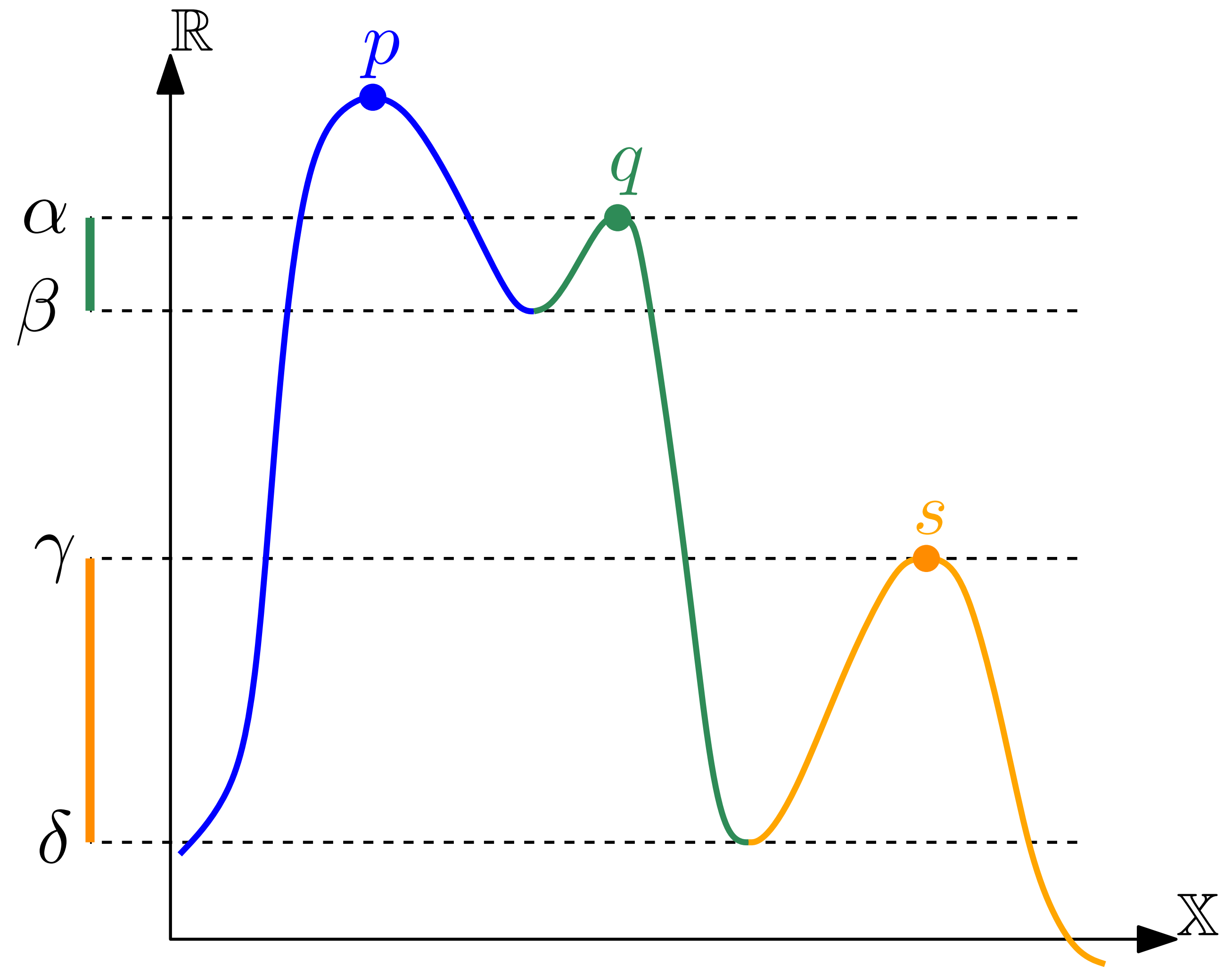
- degree-0 persistence algo. builds a hierarchy of the peaks of $\hat{f}$ (merge tree)
- merge clusters according to the hierarchy (merge each cluster into its parent)



Merging Clusters

- degree-0 persistence algo. builds a hierarchy of the peaks of $\hat{f}$ (merge tree)
- merge clusters according to the hierarchy (merge each cluster into its parent)
- given a fixed threshold $\tau \geq 0$, only merge those clusters of prominence $< \tau$

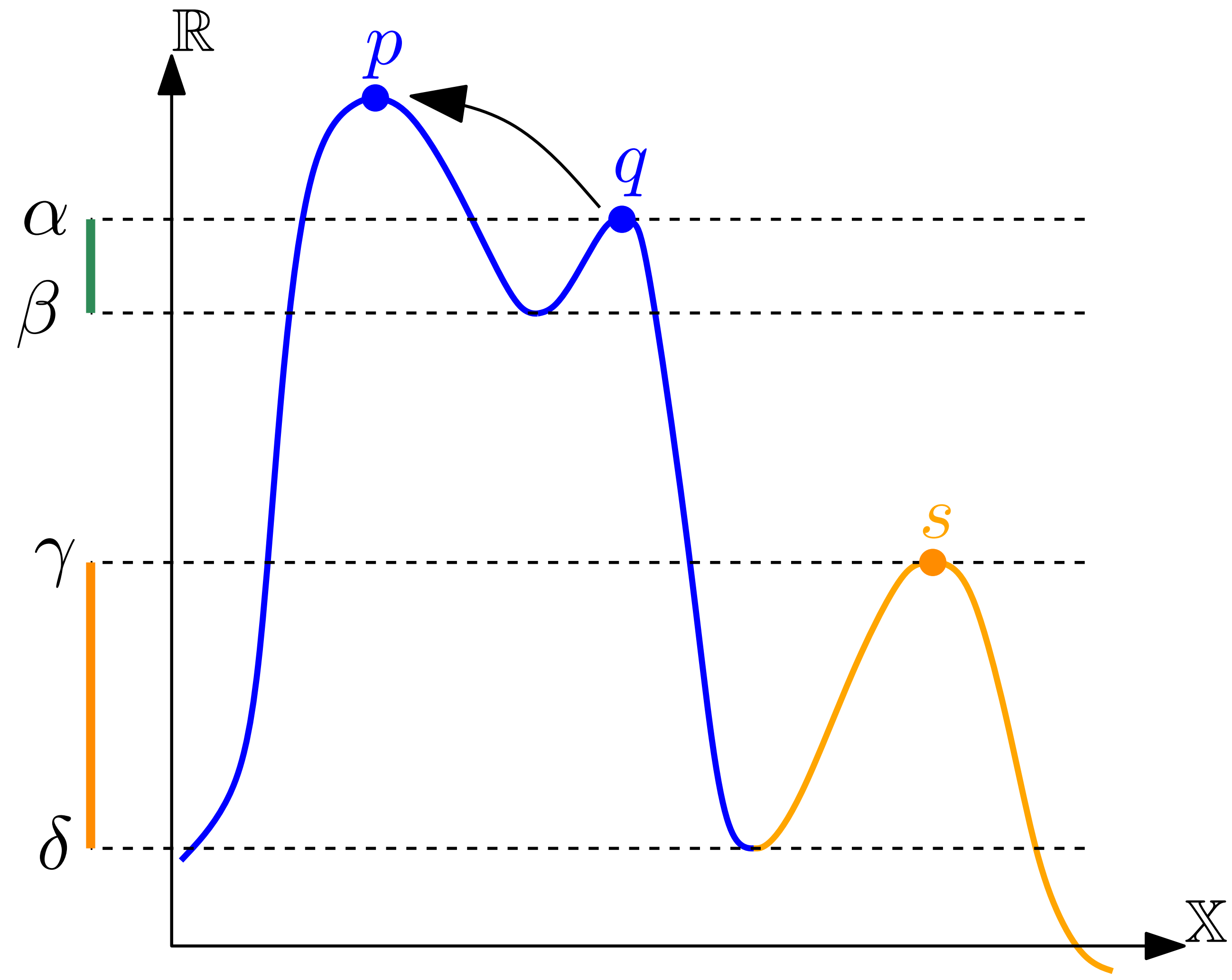
$$0 \leq \tau \leq \alpha - \beta$$



Merging Clusters

- degree-0 persistence algo. builds a hierarchy of the peaks of $\hat{f}$ (merge tree)
- merge clusters according to the hierarchy (merge each cluster into its parent)
- given a fixed threshold $\tau \geq 0$, only merge those clusters of prominence $< \tau$

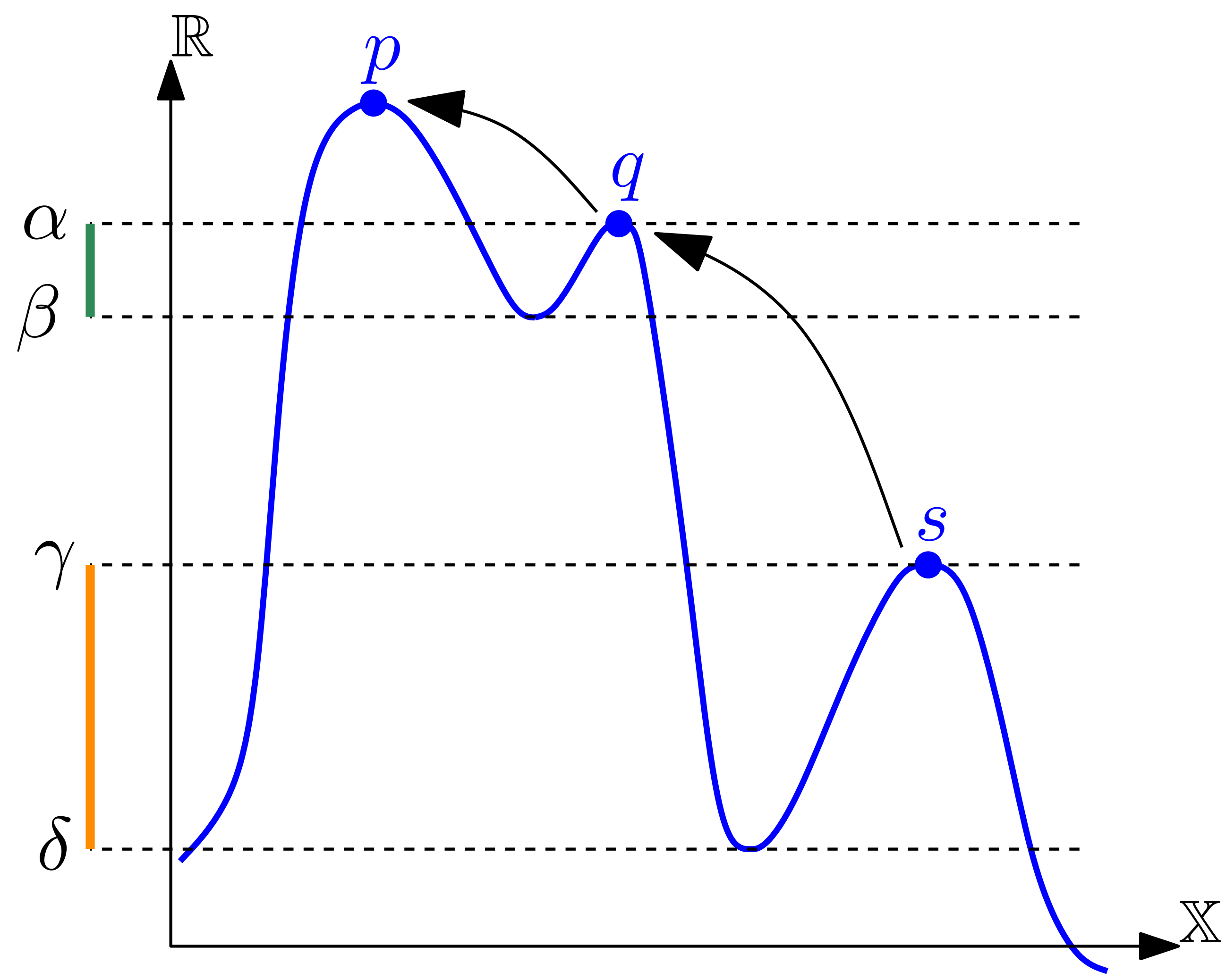
$$\alpha - \beta < \tau \leq \gamma - \delta$$



Merging Clusters

- degree-0 persistence algo. builds a hierarchy of the peaks of $\hat{f}$ (merge tree)
- merge clusters according to the hierarchy (merge each cluster into its parent)
- given a fixed threshold $\tau \geq 0$, only merge those clusters of prominence $< \tau$

$$\gamma - \delta < \tau \leq +\infty$$



Pseudo-code:

Input: simple graph G with n vertices, n-dimensional vector $\hat{f}$, real parameter $\tau \geq 0$.

Sort the vertex indices $\{1, 2, \cdots, n\}$ so that $\hat{f}(1) \geq \hat{f}(2) \geq \cdots \geq \hat{f}(n)$;
Initialize a union-find data structure $\mathcal{U}$ and two vectors g, r of size n;

for $i = 1$ to n **do**
  Let $\mathcal{N}$ be the set of neighbors of i in G that have indices lower than i;
  if $\mathcal{N} = \emptyset$ *// vertex i is a peak of $\hat{f}$ within G*
    Create a new entry e in $\mathcal{U}$ and attach vertex i to it;
    $r(e) \leftarrow i$ *// $r(e)$ stores the root vertex associated with the entry e*
  else *// vertex i is not a peak of $\hat{f}$ within G*
    $g(i) \leftarrow \text{argmax}_{j \in \mathcal{N}} \hat{f}(j)$ *// $g(i)$ stores the approximate gradient at vertex i*
    $e_i \leftarrow \mathcal{U}.\mathtt{find}(g(i))$;
    Attach vertex i to the entry e_i;

graph-based hill-climbing (1976)

    for $j \in \mathcal{N}$ **do**
      $e \leftarrow \mathcal{U}.\mathtt{find}(j)$;
      if $e \neq e_i$ and $\min\{\hat{f}(r(e)),\ \hat{f}(r(e_i))\} < \hat{f}(i) + \tau$
        $\mathcal{U}.\mathtt{union}(e,\ e_i)$;
        $r(e \cup e_i) \leftarrow \text{argmax}_{\{r(e),\ r(e_i)\}} \hat{f}$;
        $e_i \leftarrow e \cup e_i$;

cluster merges with persistence (2013)

Output: the collection of entries e of $\mathcal{U}$ such that $\hat{f}(r(e)) \geq \tau$.

Complexity of the Algorithm

Given a neighborhood graph with n vertices (with density values) and m edges:

1. the algorithm sorts the vertices by decreasing density values,

2. the algorithm makes a single pass through the vertex set, creating the spanning forest and merging clusters on the fly using a union-find data structure.

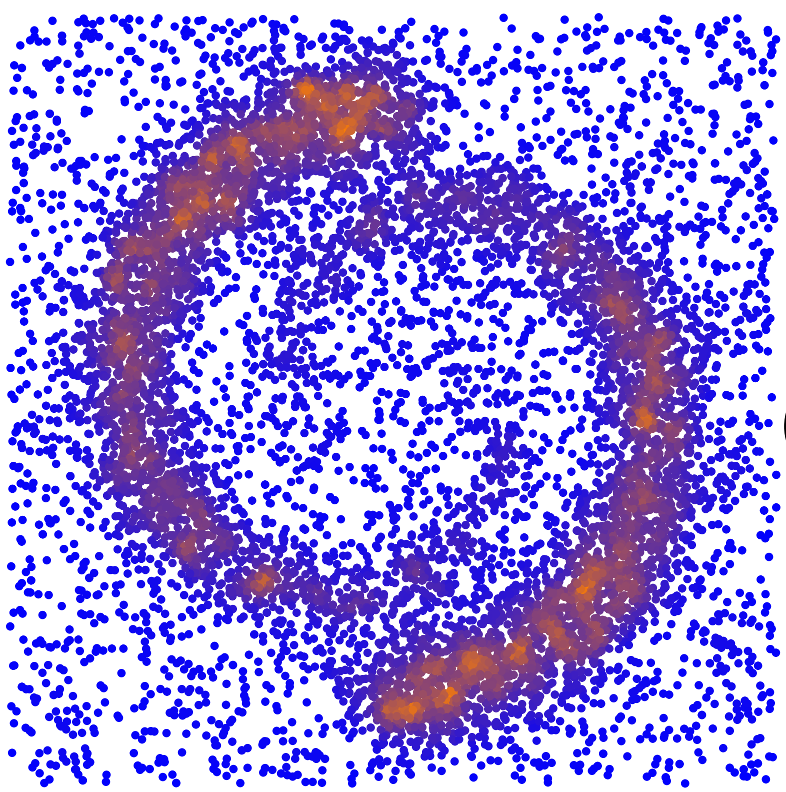
$\rightarrow$ Running time: $O(n \log n + (n+m)\alpha(n))$

$\rightarrow$ Space complexity: $O(n+m)$

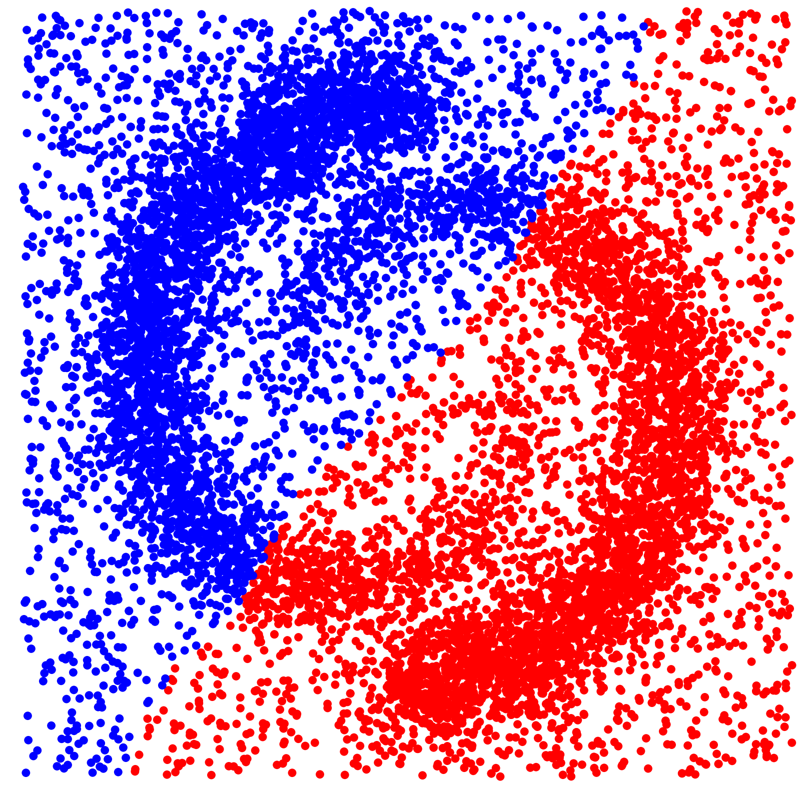
$\rightarrow$ Main memory usage: $O(n)$

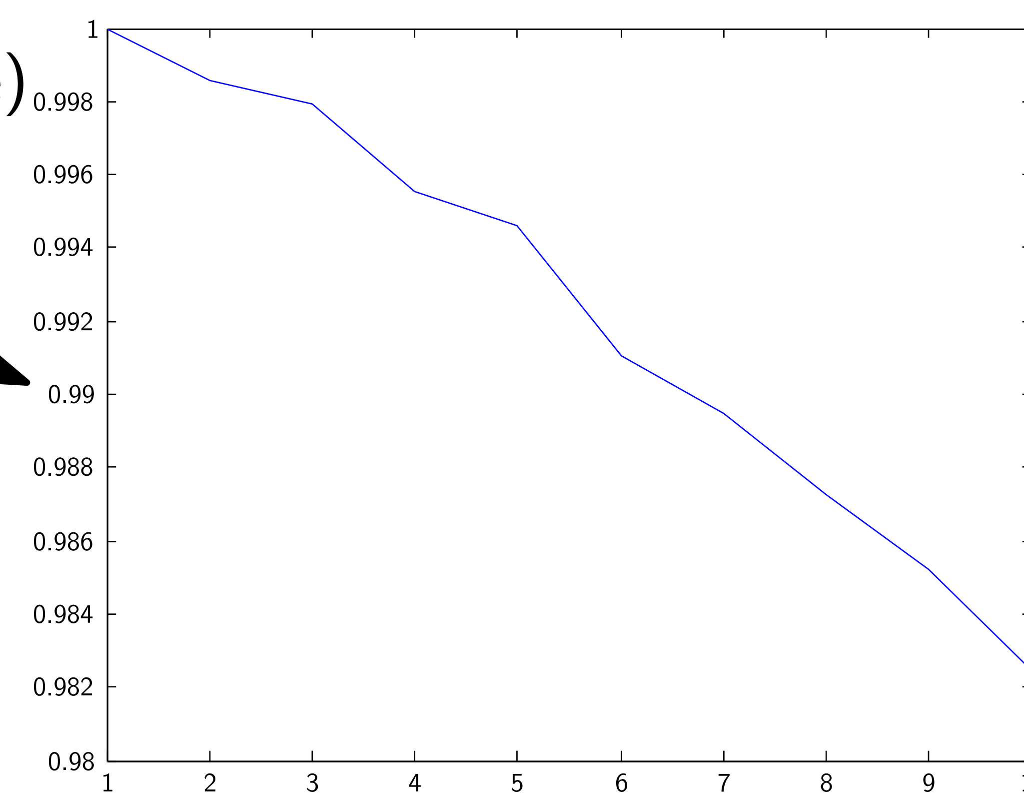
Experimental Results

Synthetic Data



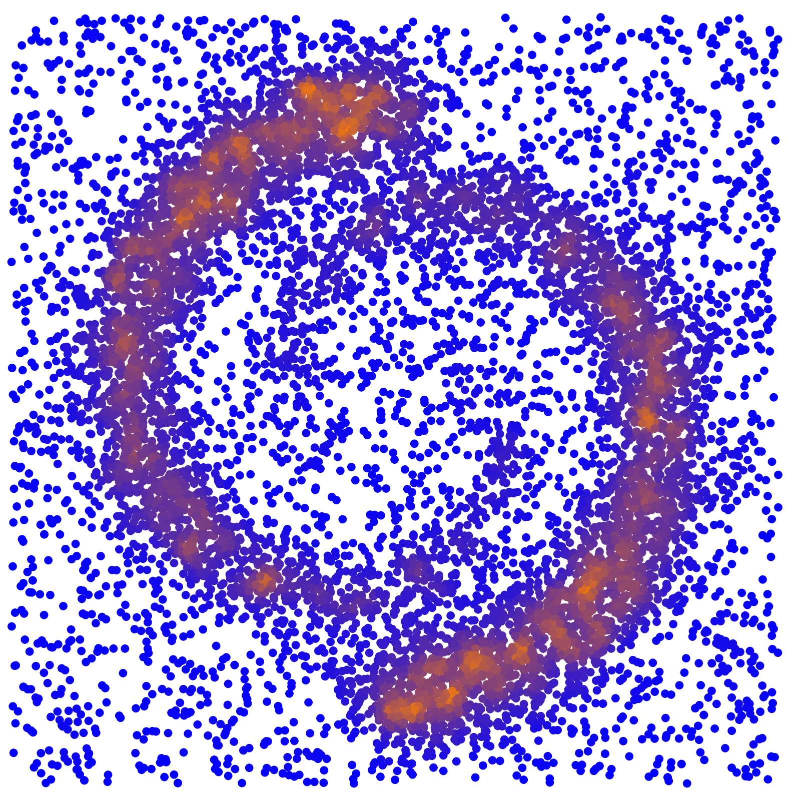
Spectral clustering
(k-means in eigenspace)





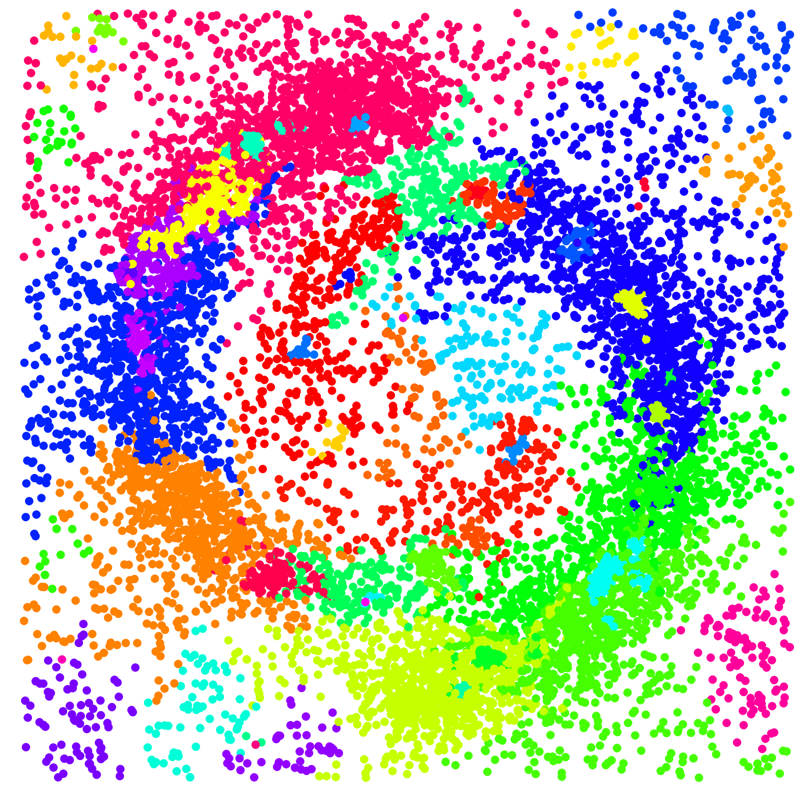
Experimental Results

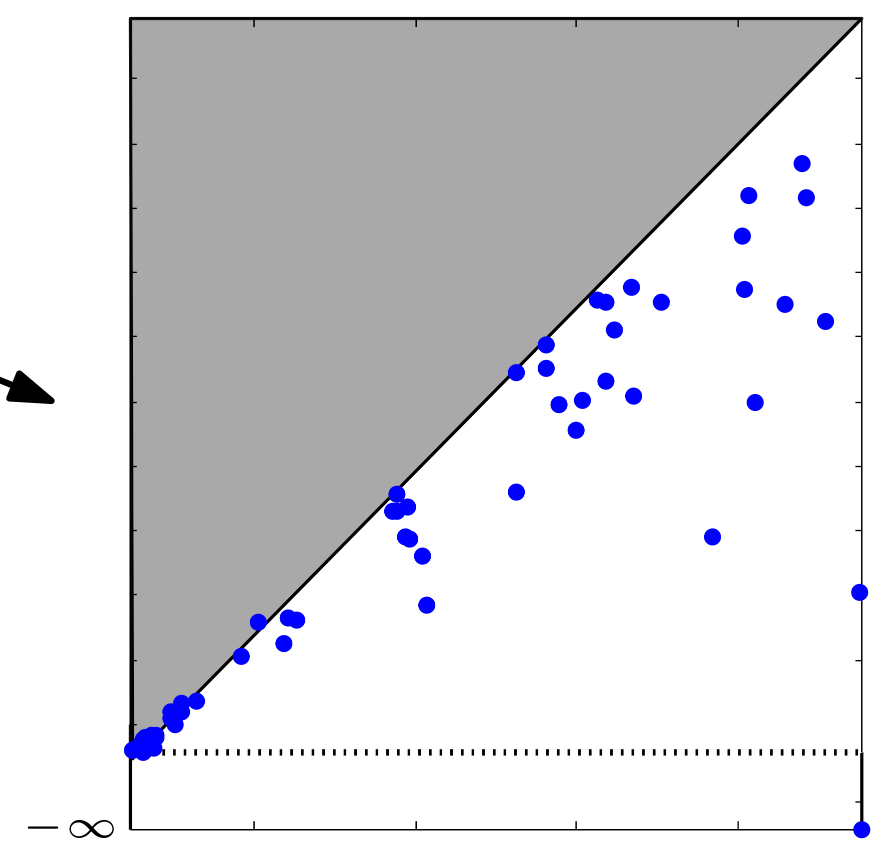
Synthetic Data



$\tau = 0$

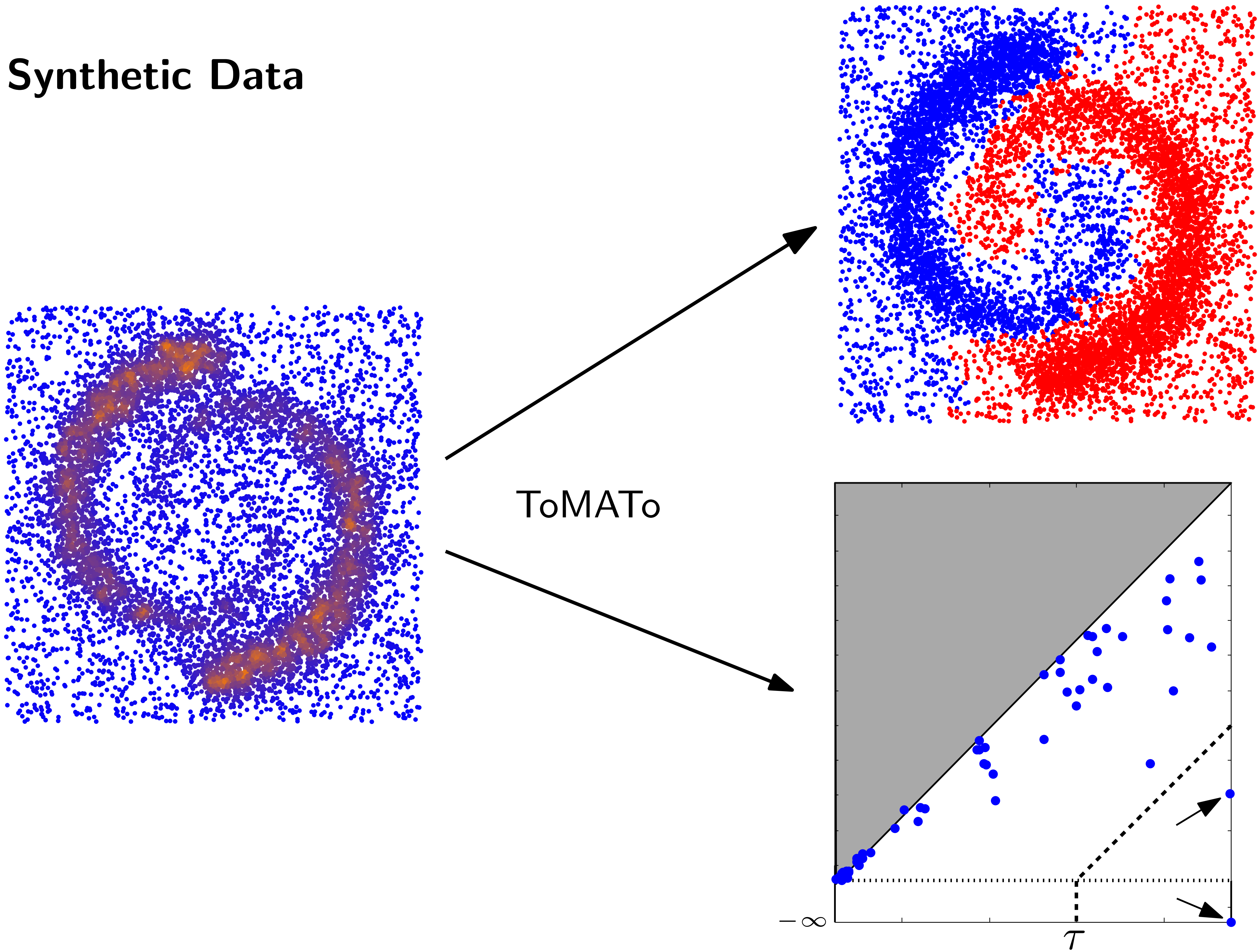
ToMATo





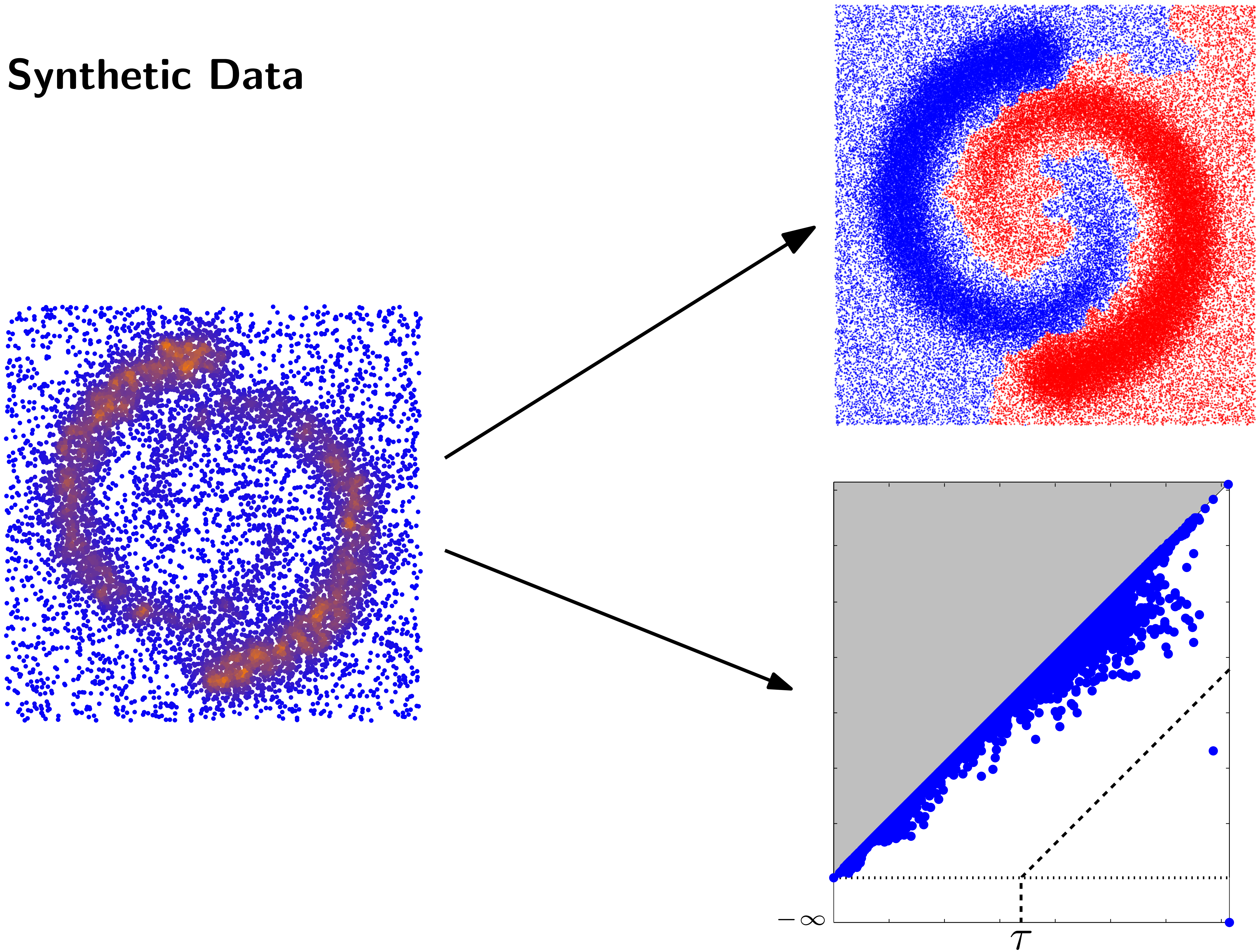
Experimental Results

Synthetic Data



Experimental Results

Synthetic Data

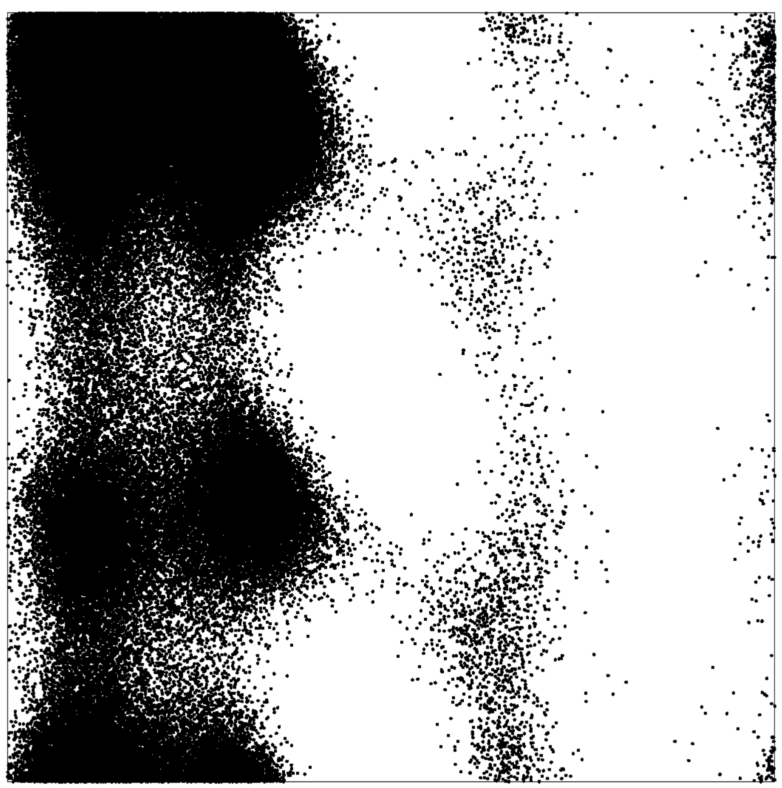


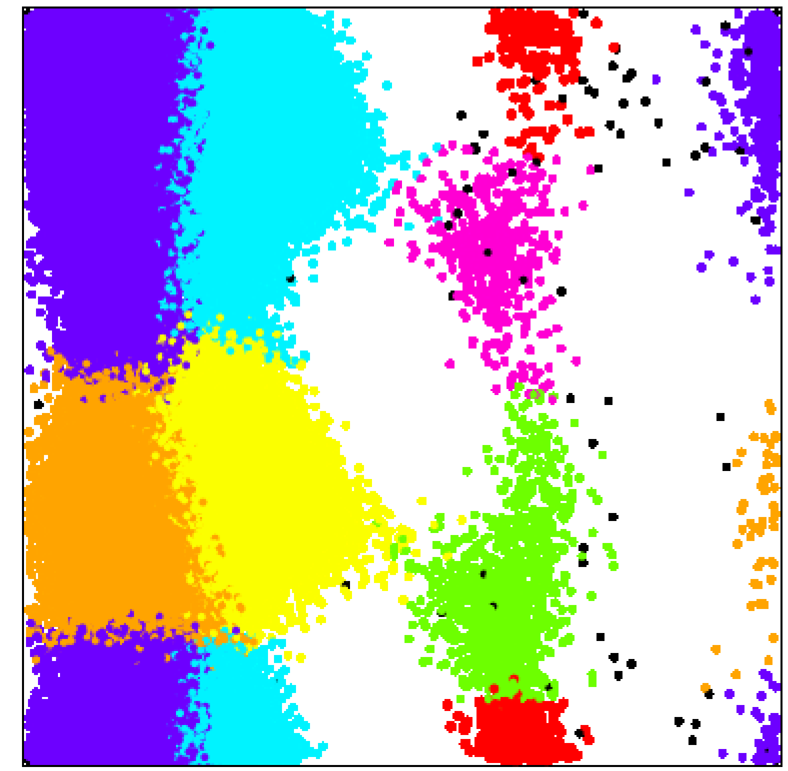
Experimental Results

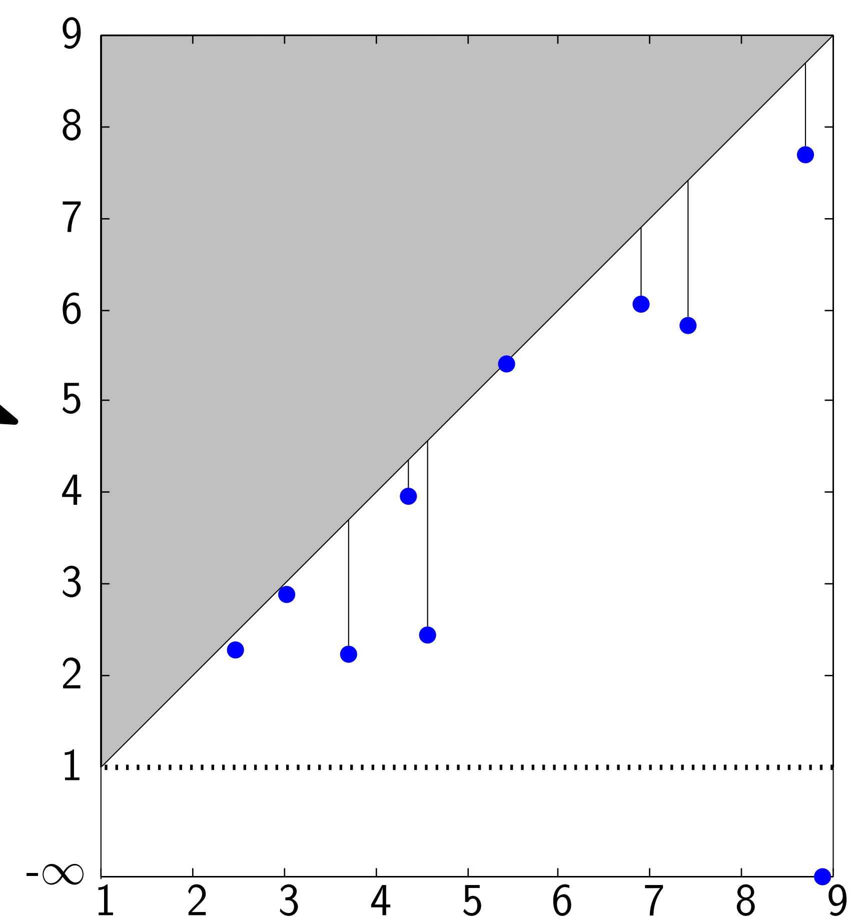
Biological Data

Alanine-Dipeptide conformations ($\mathbb{R}^{21}$)

RMSD distance (non-Euclidean)







Common belief: 6 metastable states

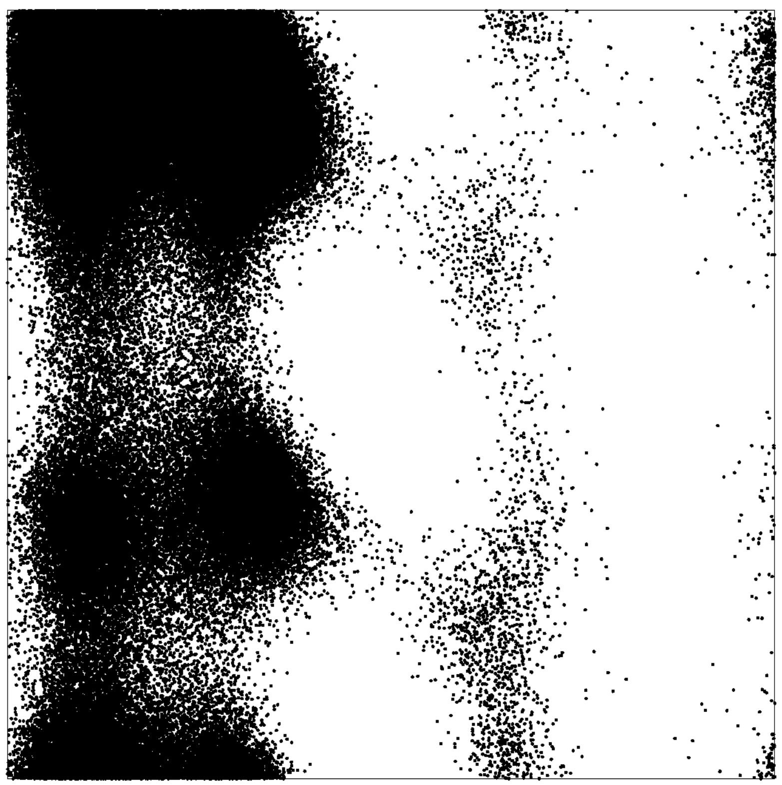
PD shows anywhere between 4 and 7 clusters

Experimental Results

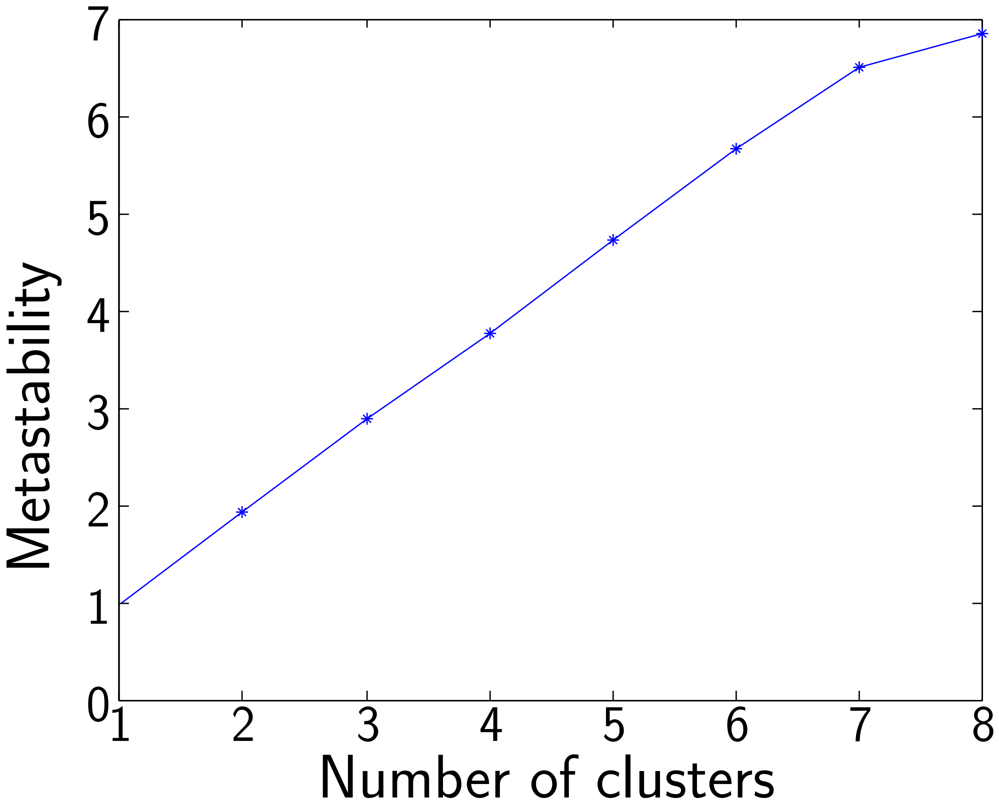
Biological Data

Alanine-Dipeptide conformations ($\mathbb{R}^{21}$)

RMSD distance (non-Euclidean)



Rank	Prominence	Metastability
1	$+\infty$	0.99982
2	3827	1.91865
3	1334	2.8813
4	557	3.76217
5	85	4.73838
6	32	5.65553
7	26	6.50757
8	7.2	6.8193
9	3.0	-
10	2.2	-



Common belief: 6 metastable states

PD shows anywhere between 4 and 7 clusters

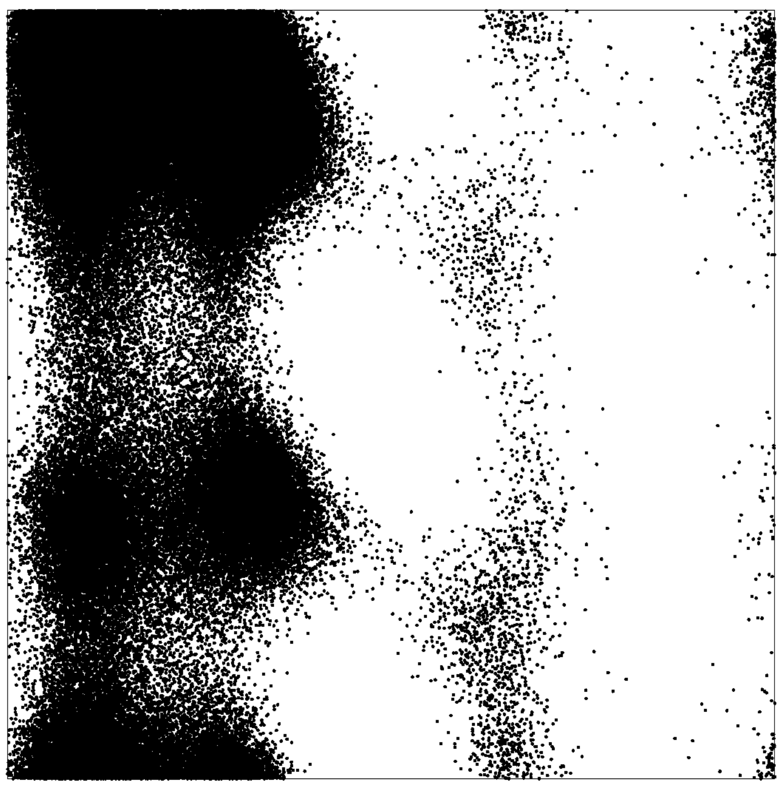
Measures of metastability confirm this insight

Experimental Results

Biological Data

Alanine-Dipeptide conformations ($\mathbb{R}^{21}$)

RMSD distance (non-Euclidean)



Note: Spectral Clustering takes a week of tweaking, while ToMATo runs out-of-the-box in a few minutes

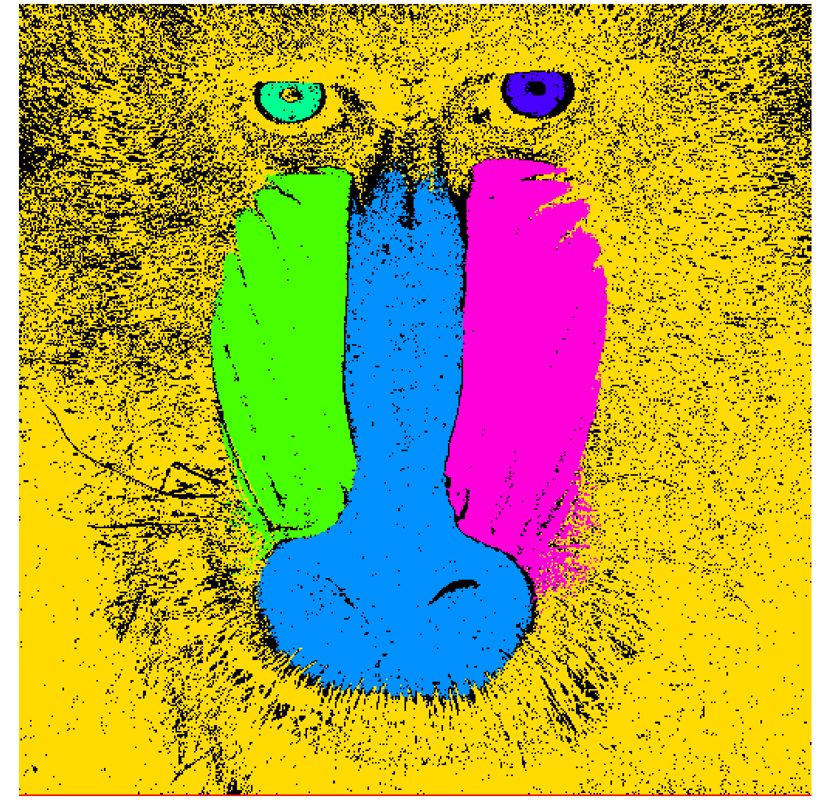
• Y. Yao, J. Sun, X. Huang, G. Bowman, G. Singh, M. Lesnick, L. Guibas, V. Pande, G. Carlsson, Topological methods for exploring low-density states in biomolecular folding pathways, *The Journal of Chemical Physics*, 2009.

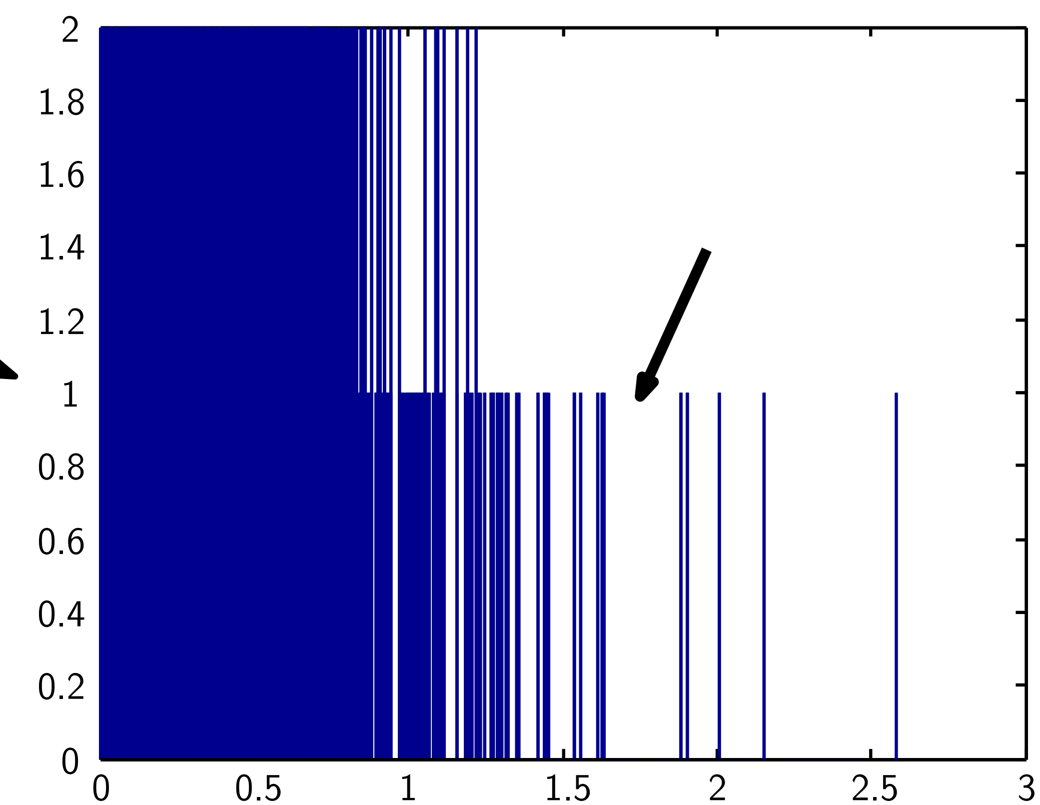
Experimental Results

Image Segmentation

Density is estimated in 3D color space (Luv)

Neighborhood graph is built in image domain





Distribution of prominences does not usually show a clear unique gap

Still, relationship between choice of τ and number of obtained clusters remains explicit