Topological Data Analytics

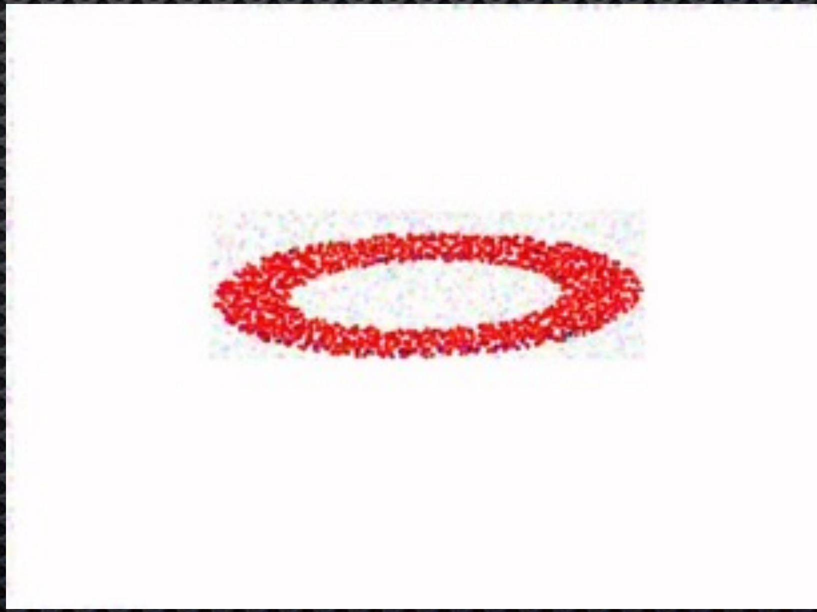
John Baillieul
Intelligent Mechatronics Lab
Boston University
johnb@bu.edu

10-th Summer School on Geometry,
Mechanics and Control June, 2016

Why do we need TDA?

Data analytics largely rely on linear methods, . . .

but, not all data is linear.

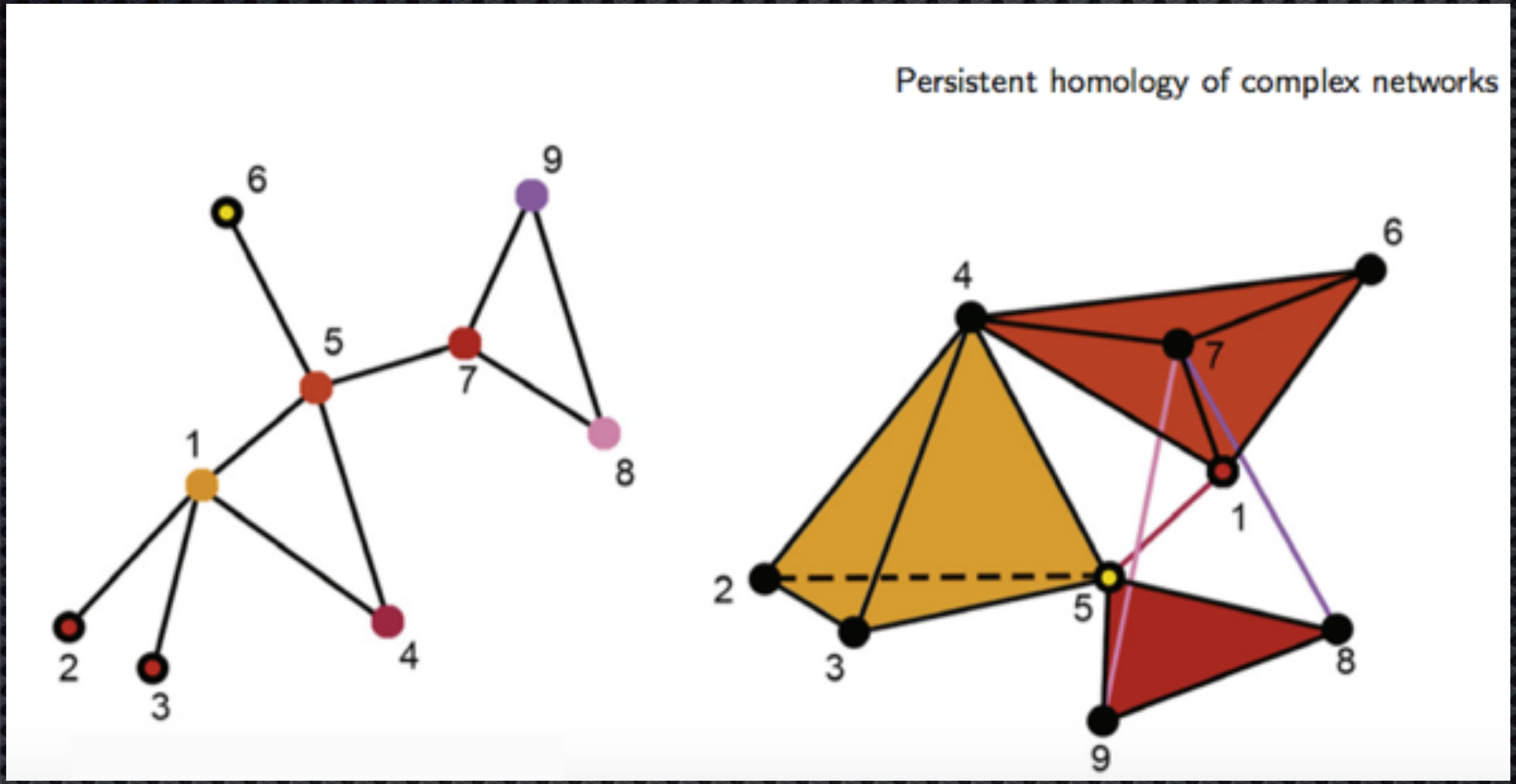


Talk Outline

- Inroduction to topological persistence
- A topological approach to saliency
- An information-theoretic approach to saliency

Topological Persistence

How can we quantify feature *significance*?



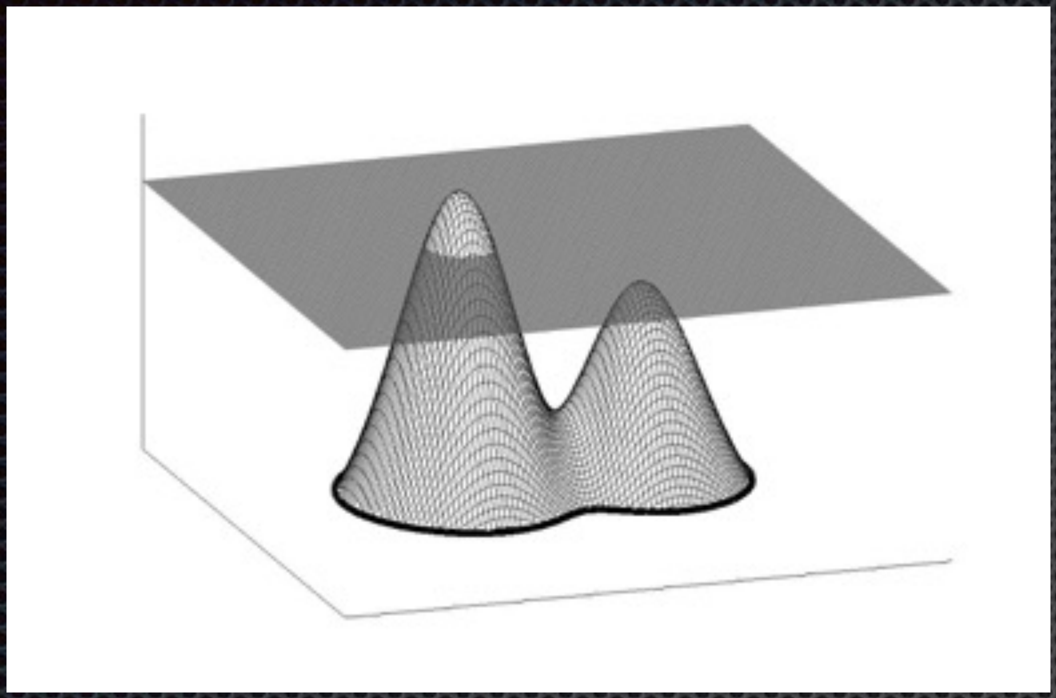
The starting point is an evolution a simplicial complex in which simplices are added in sequence:

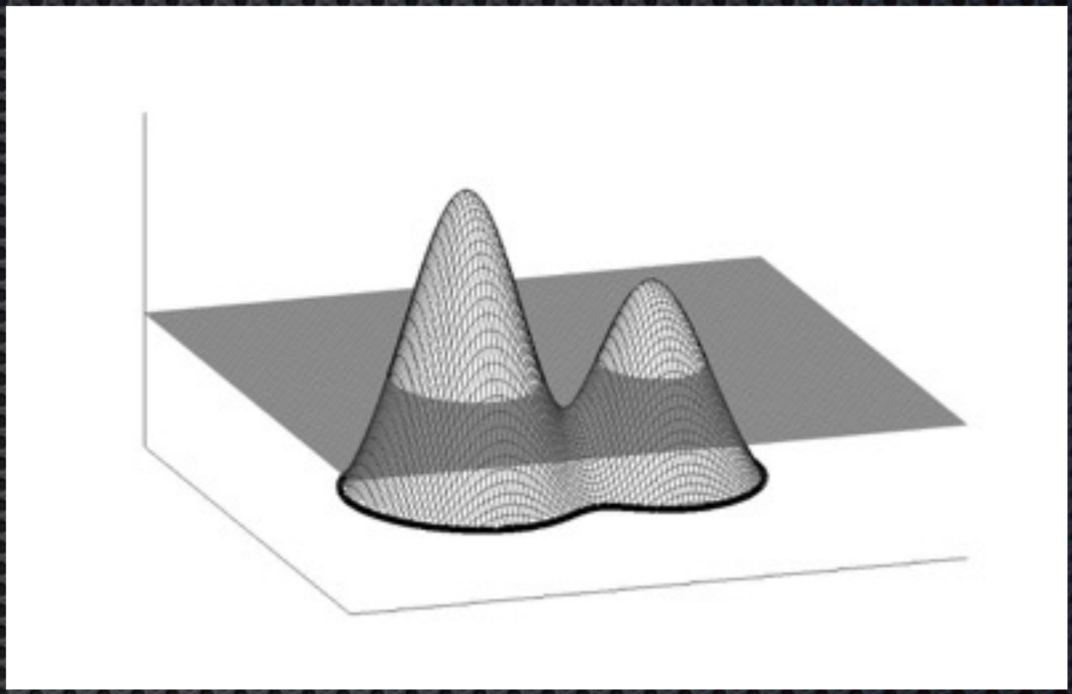
$$K_0 \subset K_1 \subset \cdots \subset K_n = K$$

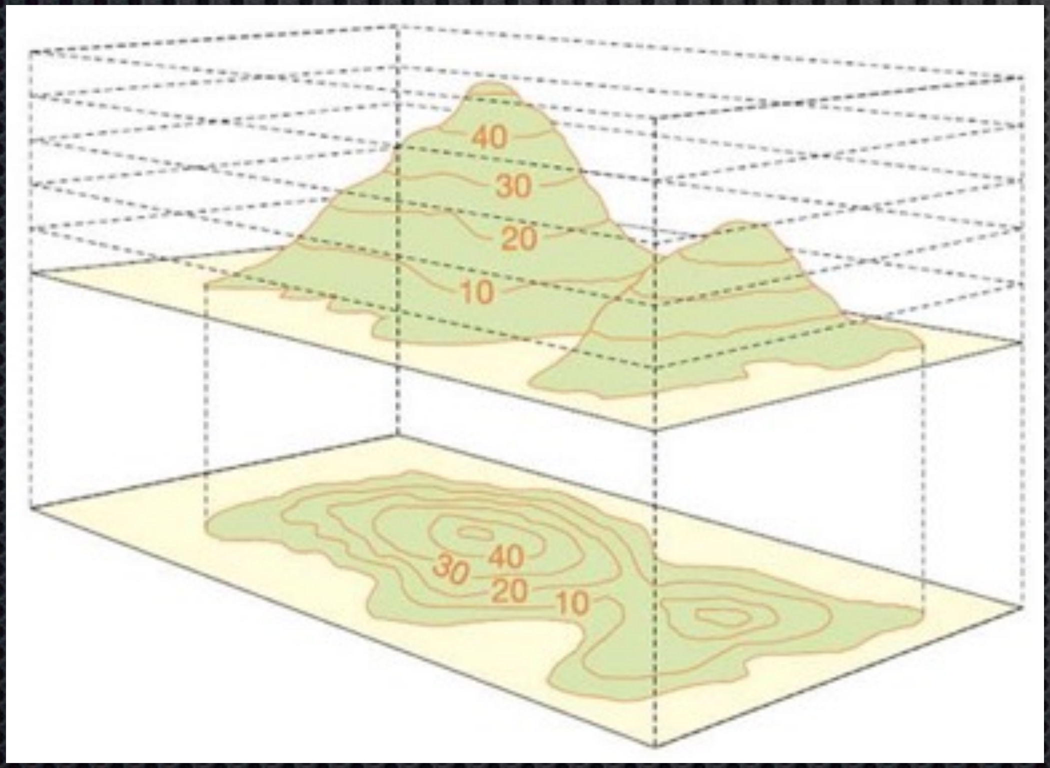
This is called a *filtration*. The idea of *persistence* is to keep track of how many steps occur between the step when a topological feature appears in the filtration and the when it is annihilated. The persistence parameter take discrete values in this case.

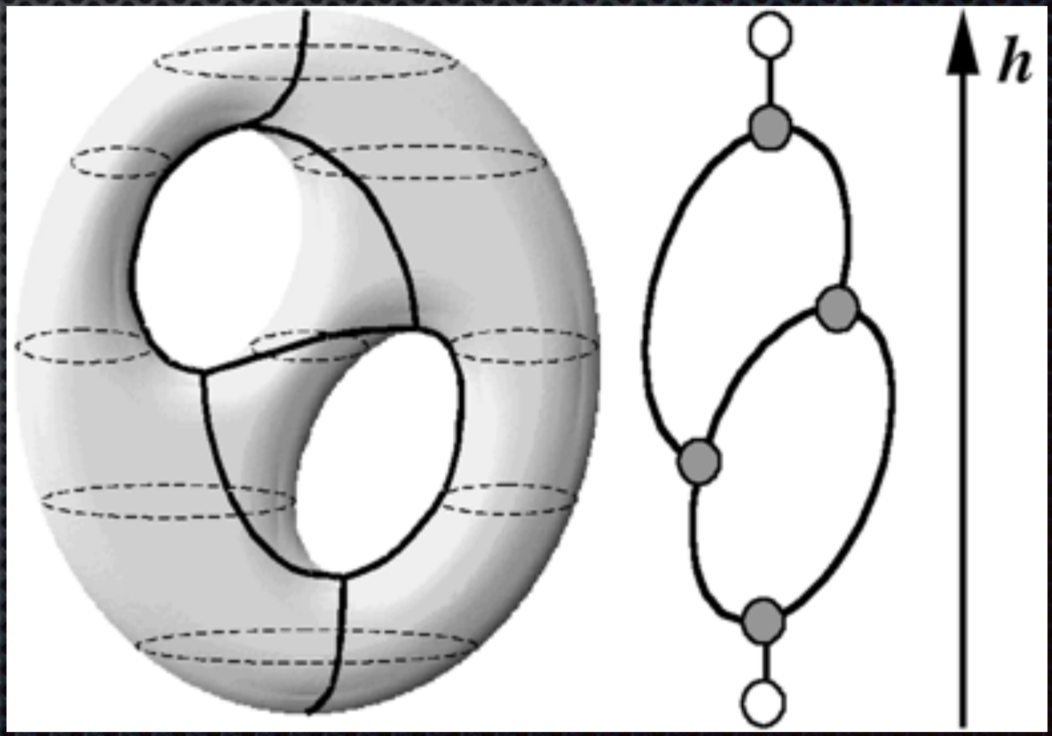
Continuous parameterizations are also studied—e.g. *Rips-Vietoris complexes*

Topological Persistence



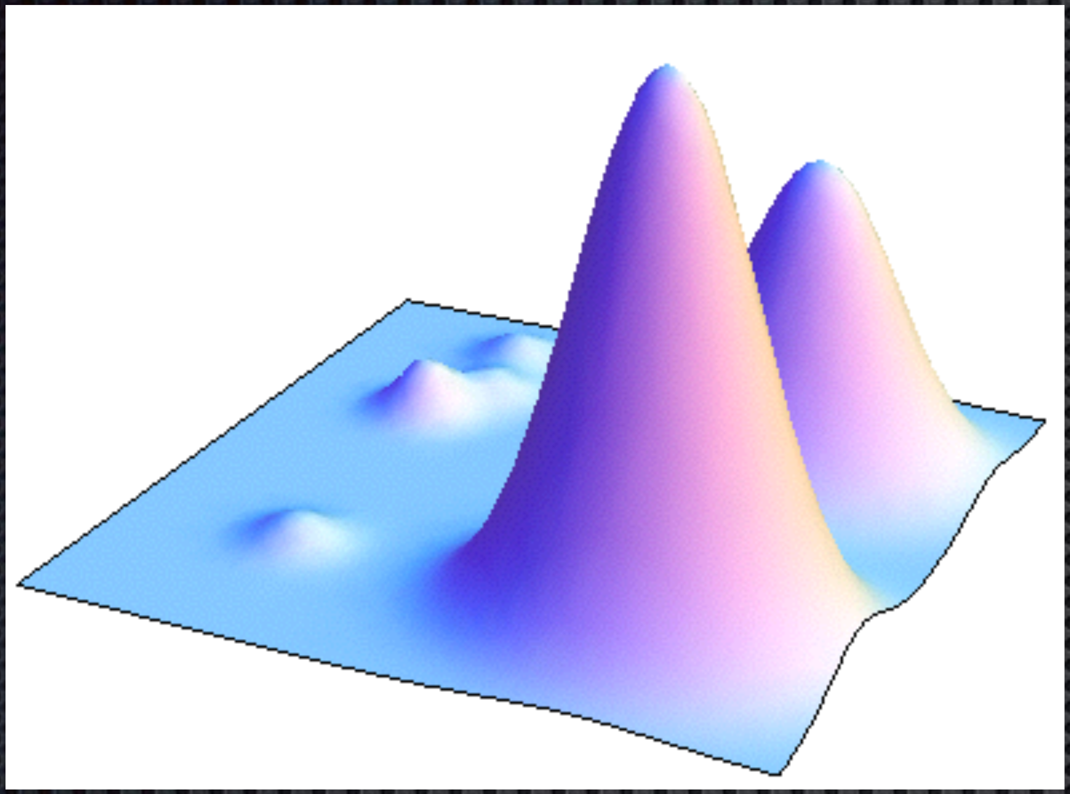






Continuous parameterizations are also studied—e.g. *height map complexes*

Random fields as information channels - diversity and noise



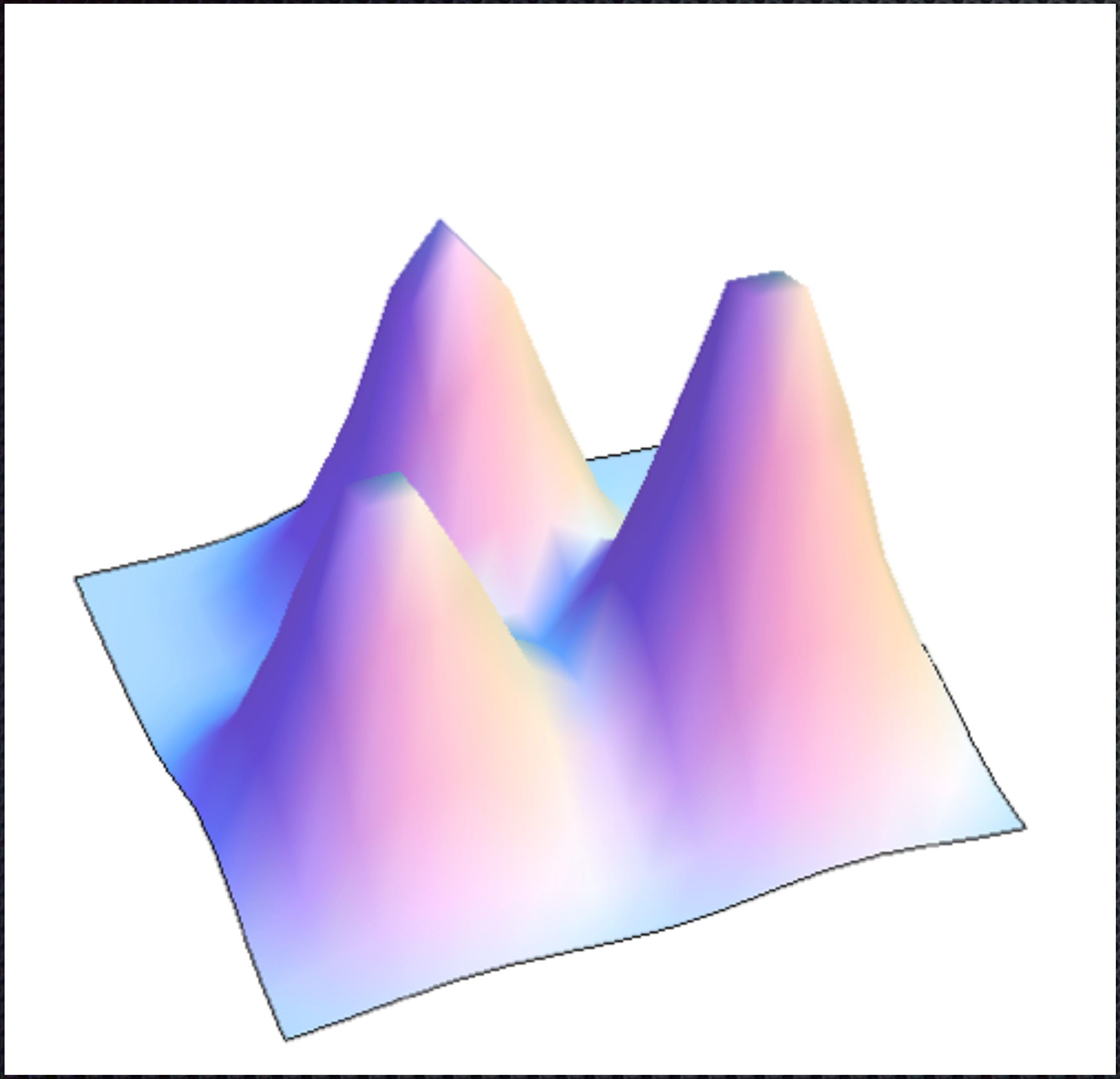
Some peaks are more significant than others.

How can we quantify feature *significance*?

How do we respect diversity and reject noise?

1. Height.
2. *Topological persistence*.
3. *Information utility*.

Random fields as information channels - diversity and noise



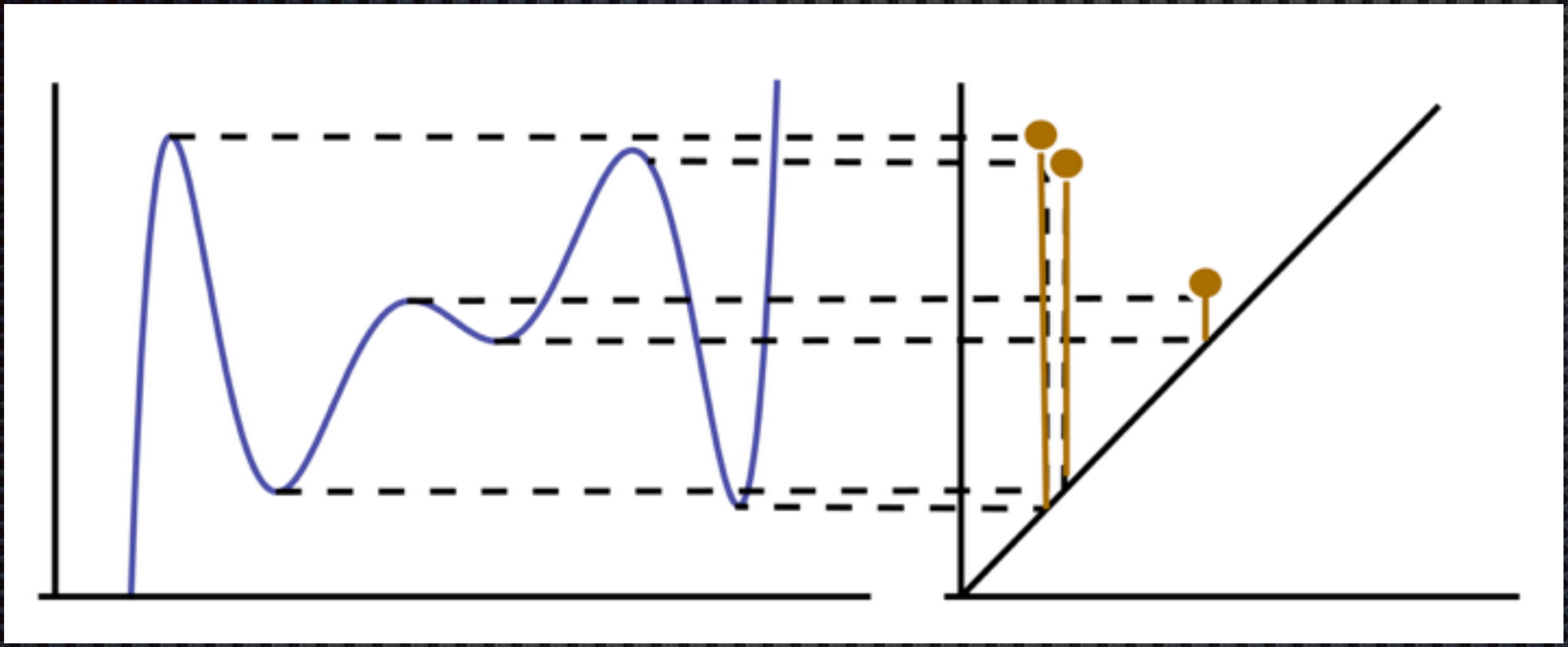
As the dimension of X increases, the topology of the sublevel sets

$$\mathbb{R}_t = f^{-1}(-\infty, t]$$

becomes more complex.

Random fields as information channels - diversity and noise

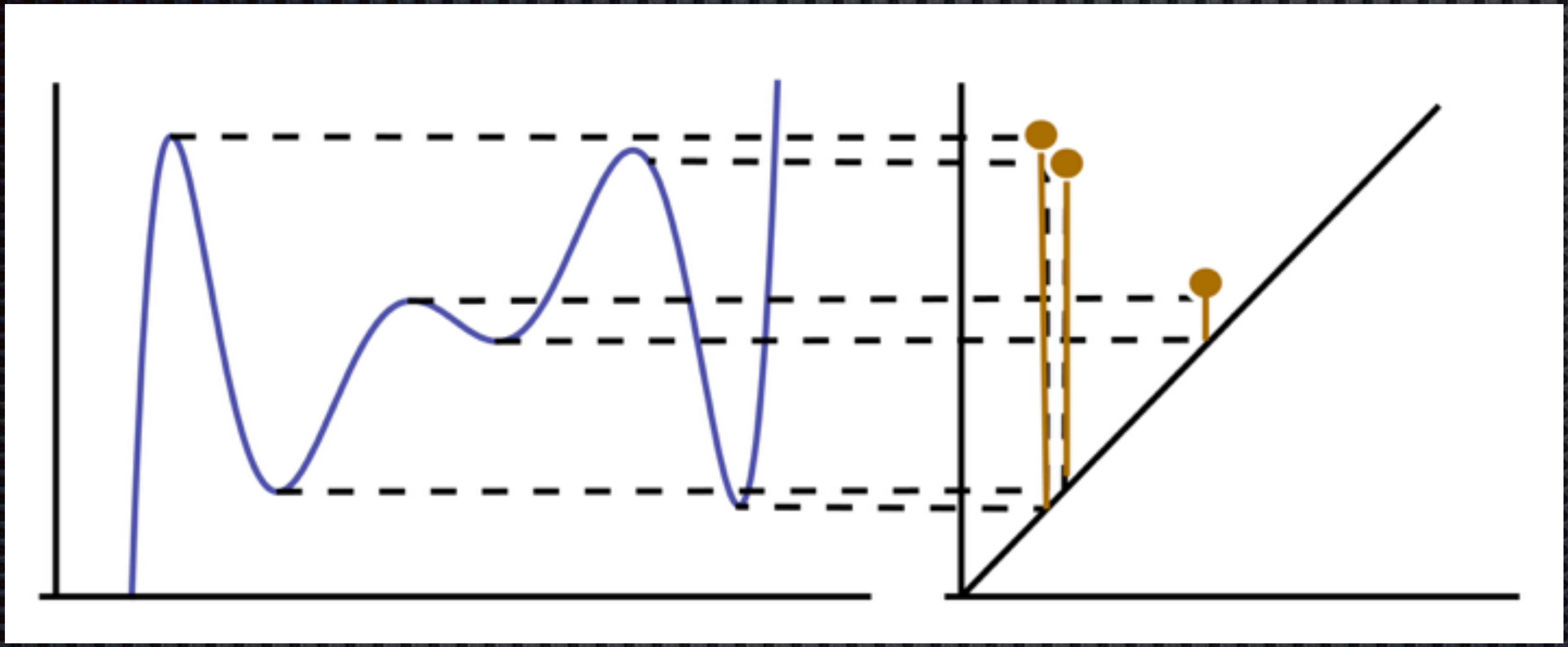
How can we quantify feature *significance*?



1. Height of critical values.
2. *Topological persistence*.
3. *Information utility*.

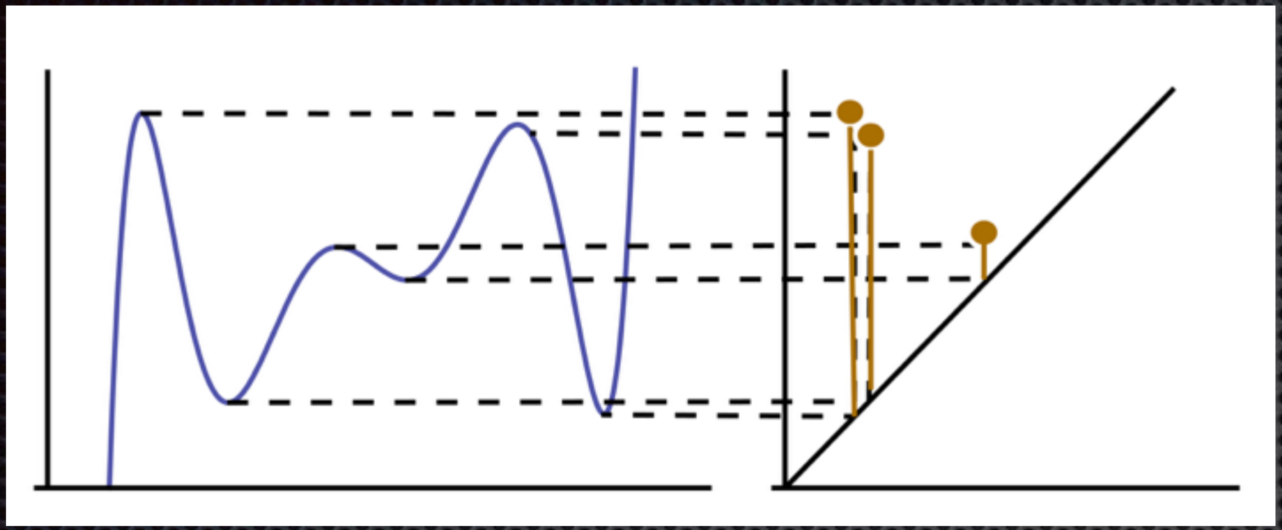
Random fields as information channels - diversity and noise

How can we quantify feature *significance*?



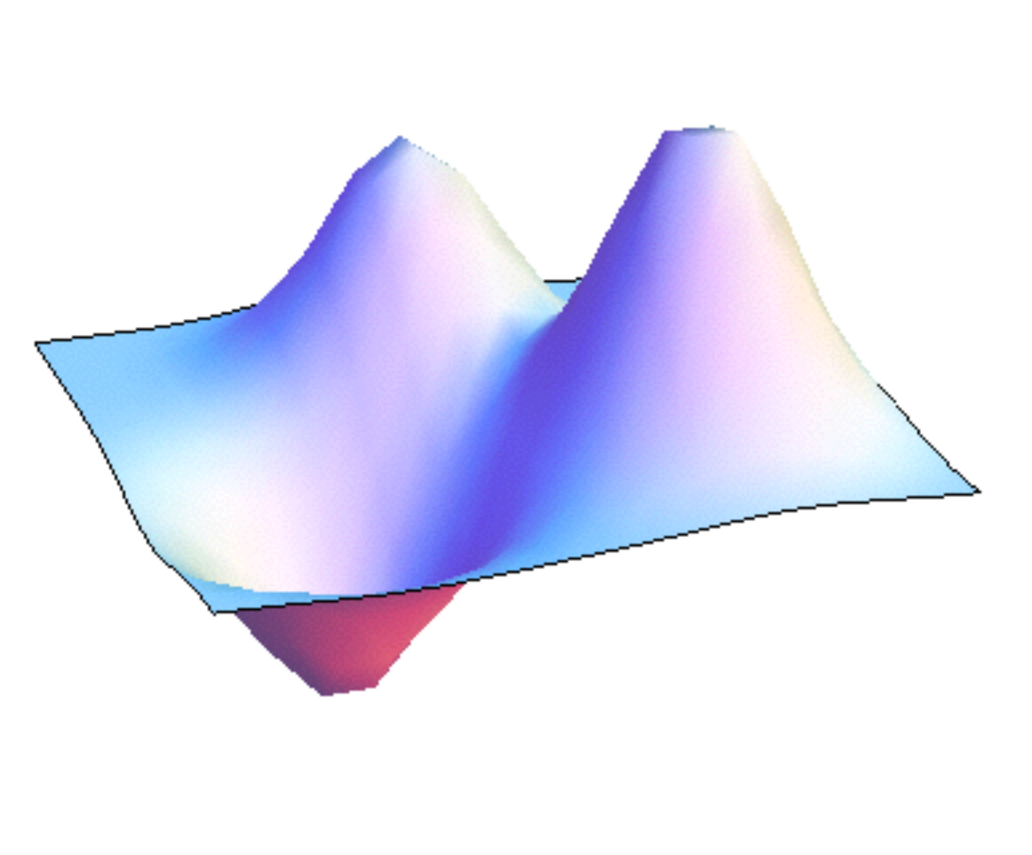
Topological persistence. is defined in terms of the topology of sublevel sets $\mathbb{R}_t = f^{-1}(-\infty, t]$.

Random fields as information channels - diversity and noise



The *persistence diagram* is defined in terms of a threshold set: $I\alpha = \{x \in I \mid f(x) \leq \alpha\}$. Here we are interested in the number of connected components of $I\alpha$. As α increases through a local min., a connect component of $I\alpha$ is *born*. When α increases through a local max., two connected components merge, and we say that the one that has persisted for a smaller range of α *dies*.

The Differential Topology of Scalar Fields



Let regular values x_j

$$a = x_0 < x_1 < \cdots < x_m = b.$$

bracket the m ciritical values of

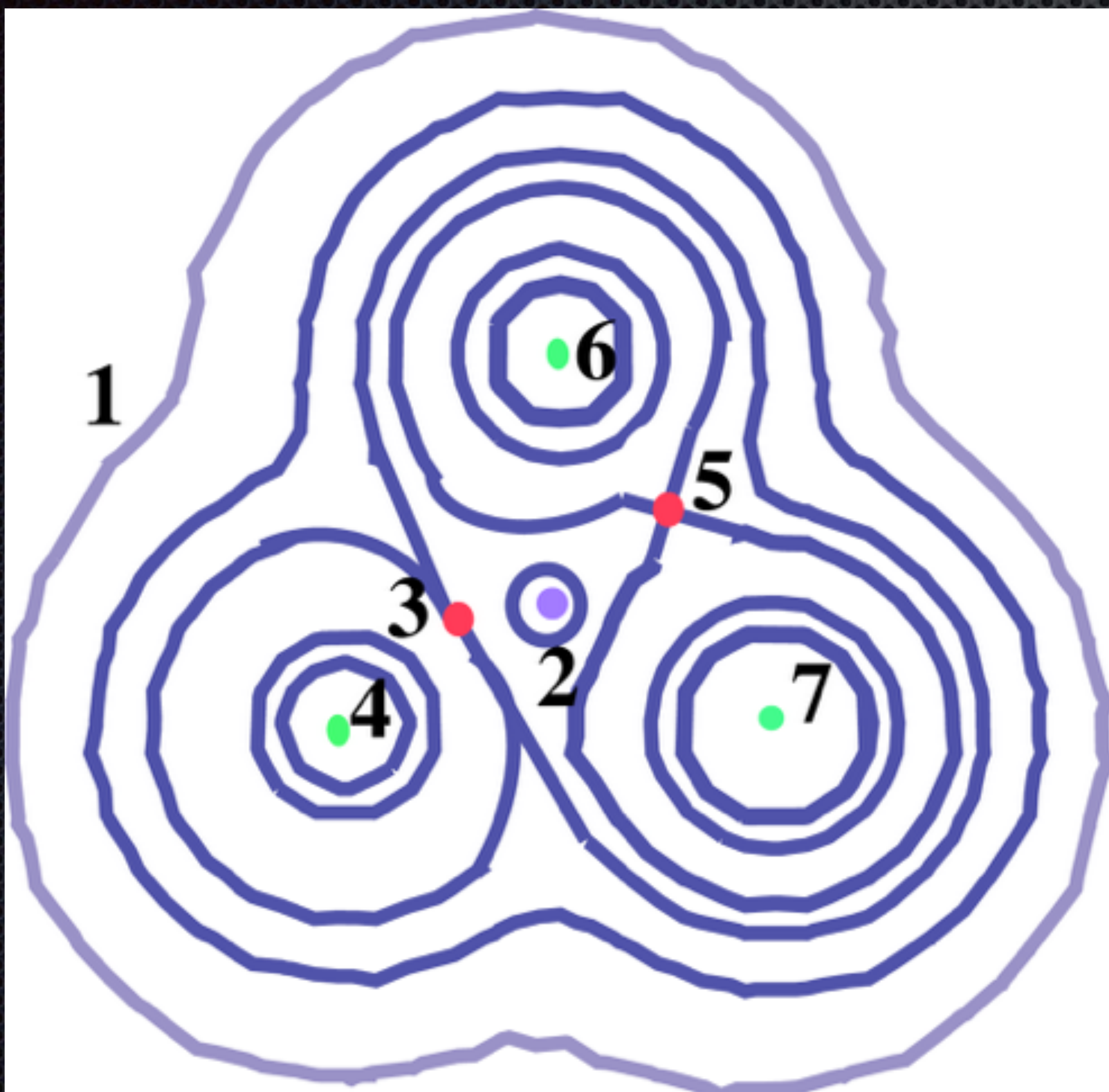
$$f : X \to \mathbb{R}.$$

Case dim X=1, persistance tracks $\beta_0(\mathbb{R}_{x_j}) = rank\ \mathcal{H}_0.$

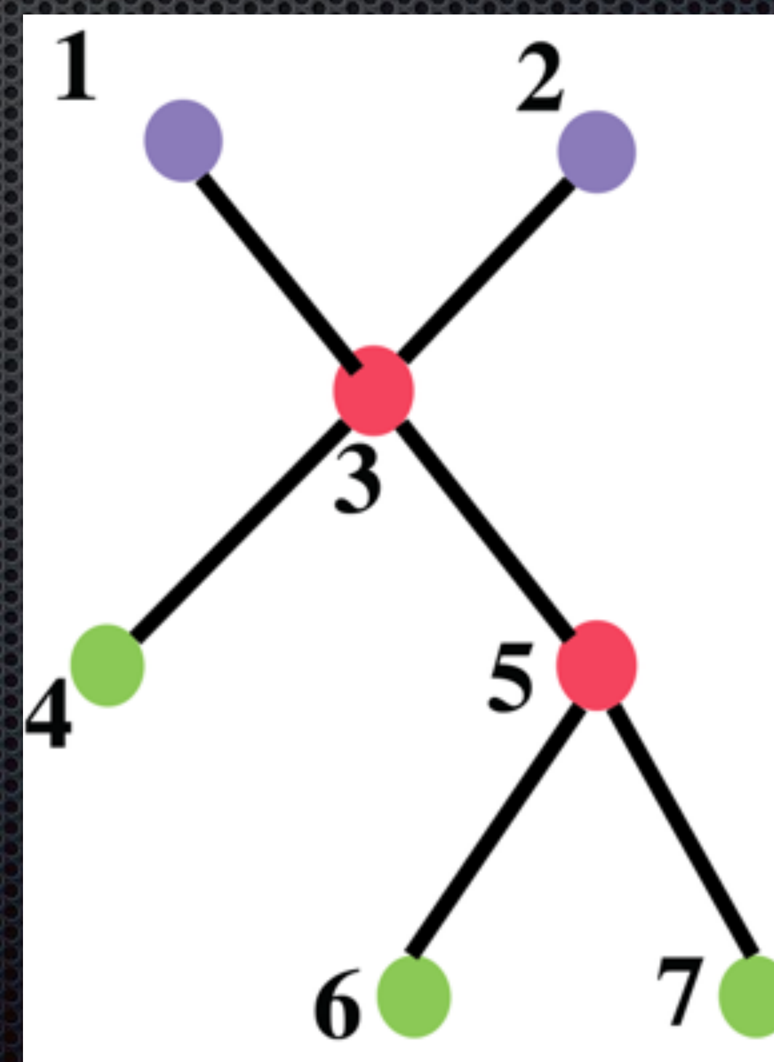
Case dim X>1, persistance tracks $\beta_p(\mathbb{R}_{x_j}) = rank\ \mathcal{H}_p.$

There is a persistence diagram for each p.

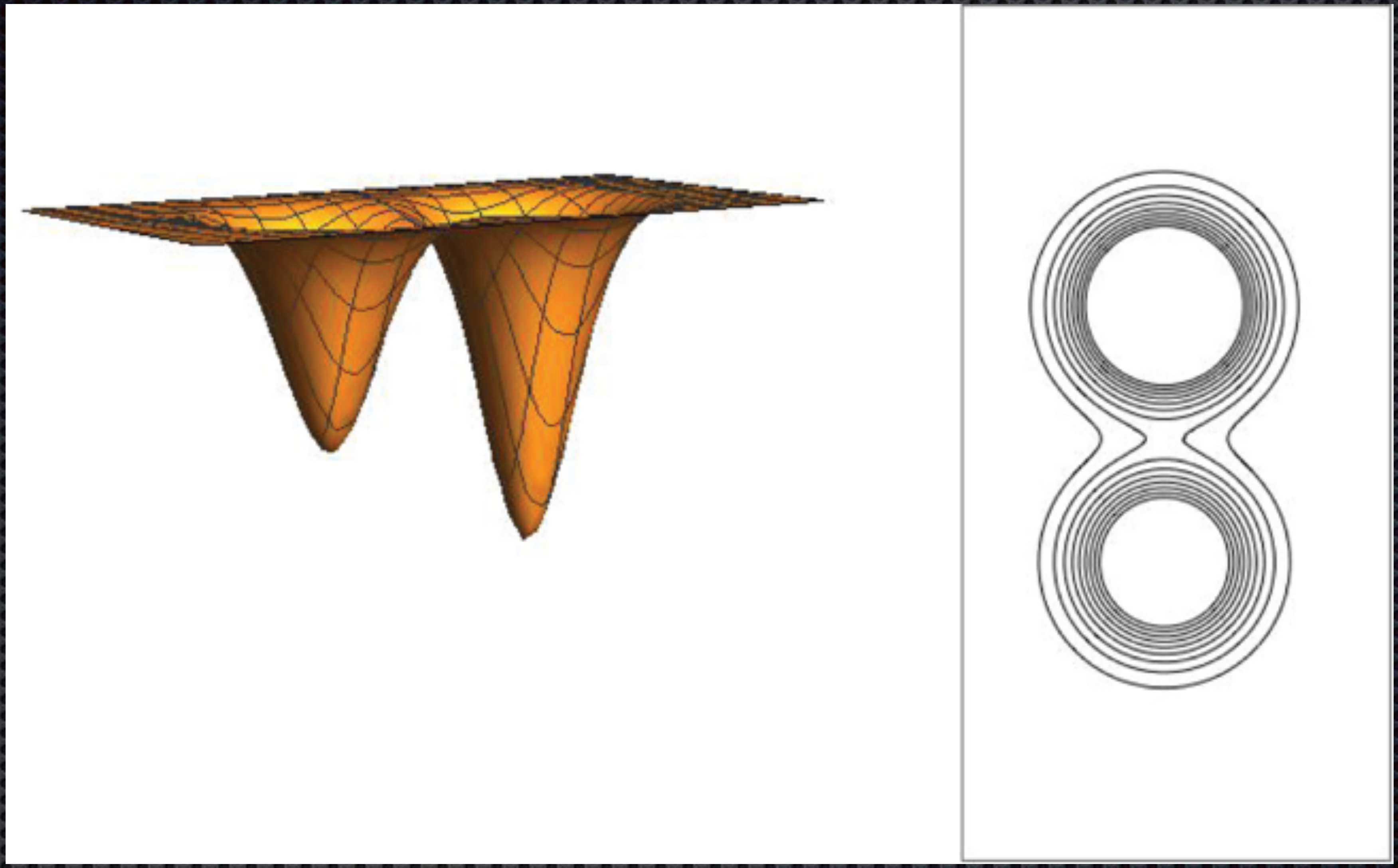
Morse-Smale fields as information channels - diversity and noise



Topological persistence looks at the topology of sublevel sets.



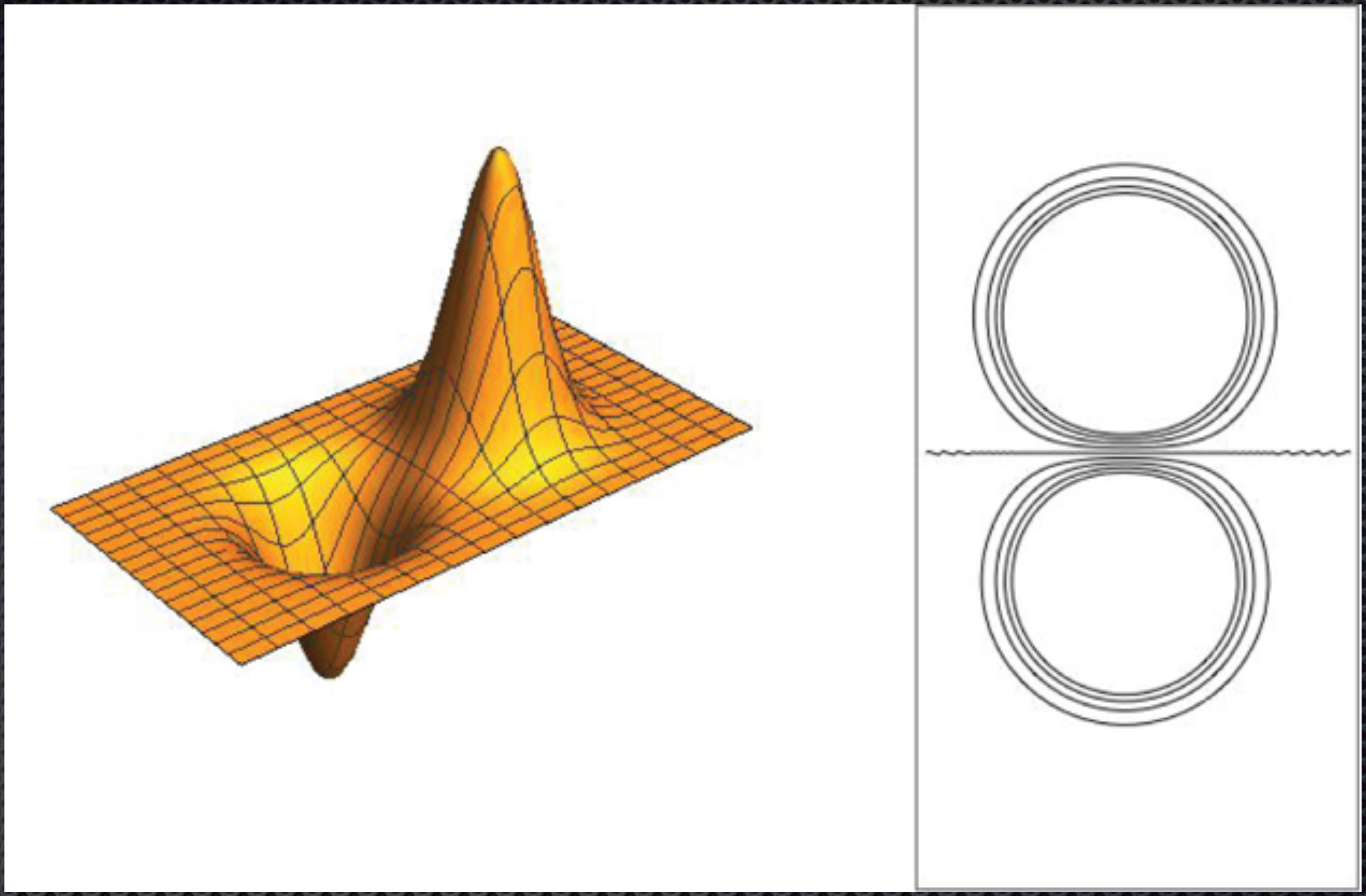
Topology of Critical Level Sets



A minimum is paired with and destroyed by a "negative saddle":

β_0 — —

Topology of Critical Level Sets

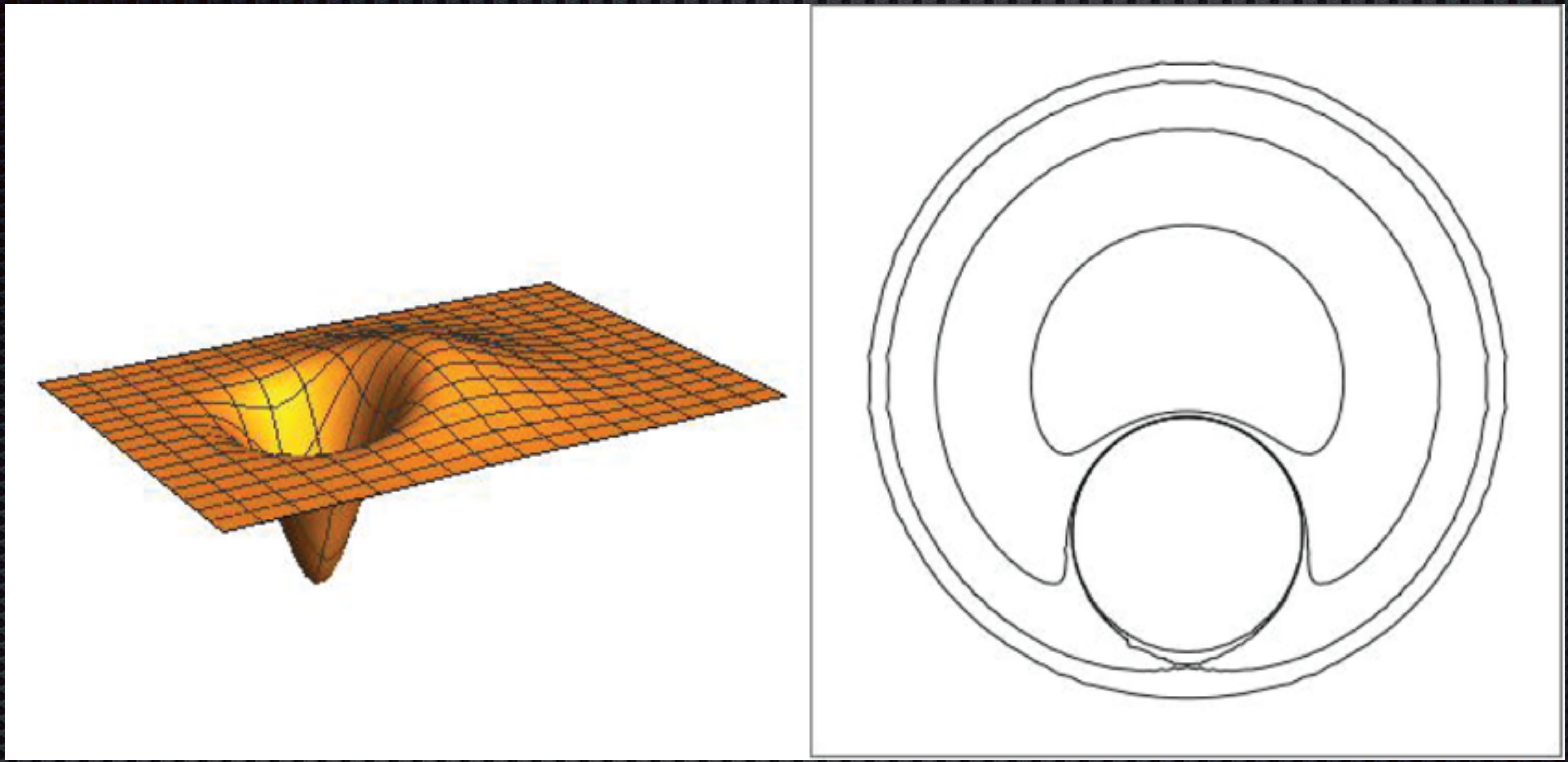


A minimum can also be paired with a "mixed saddle":

$\beta_0 - -$

$\beta_1 + +$

Topology of Critical Level Sets

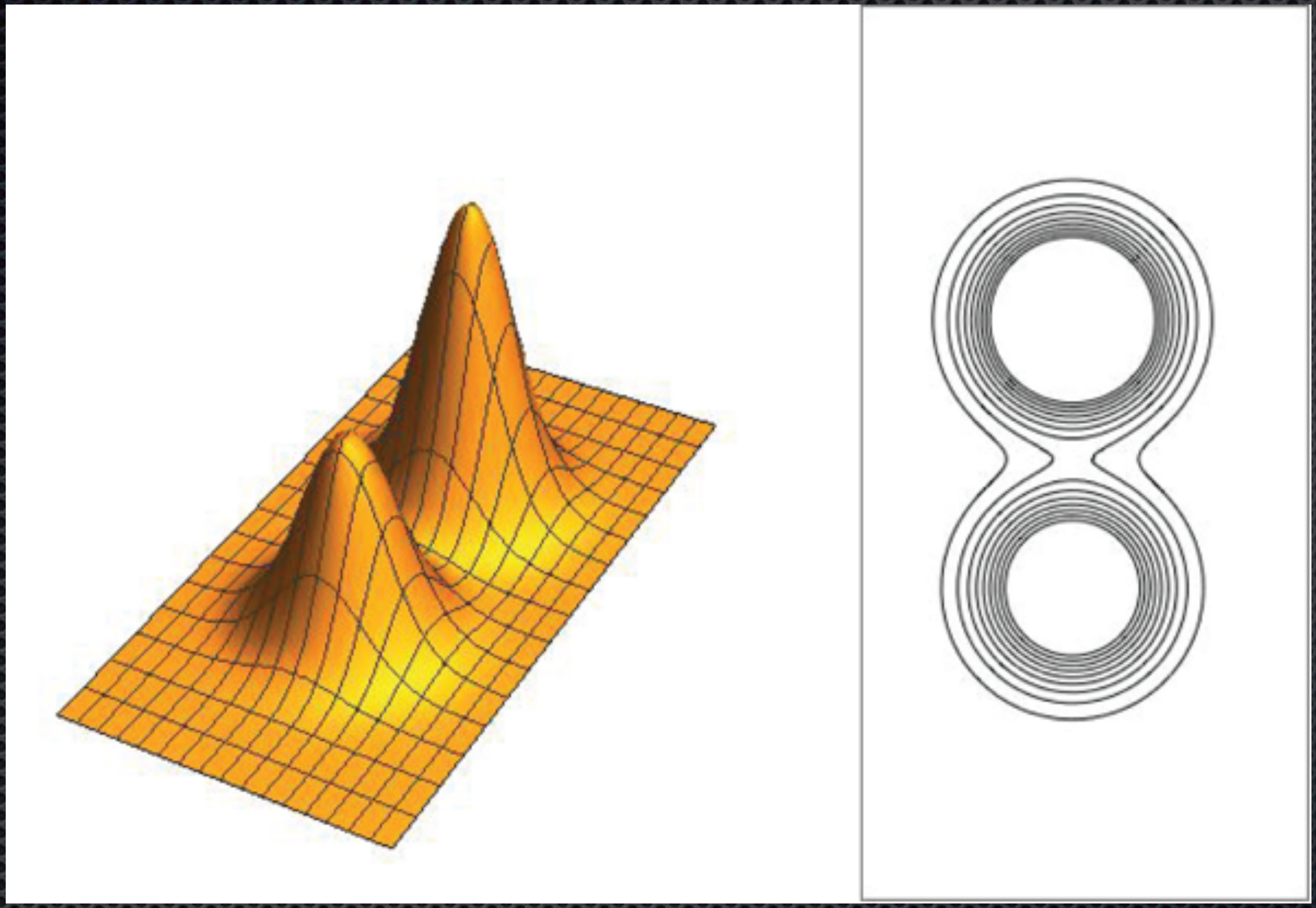


A minimum can also be paired with a "mixed saddle":

$\beta_0 - -$

$\beta_1 + +$

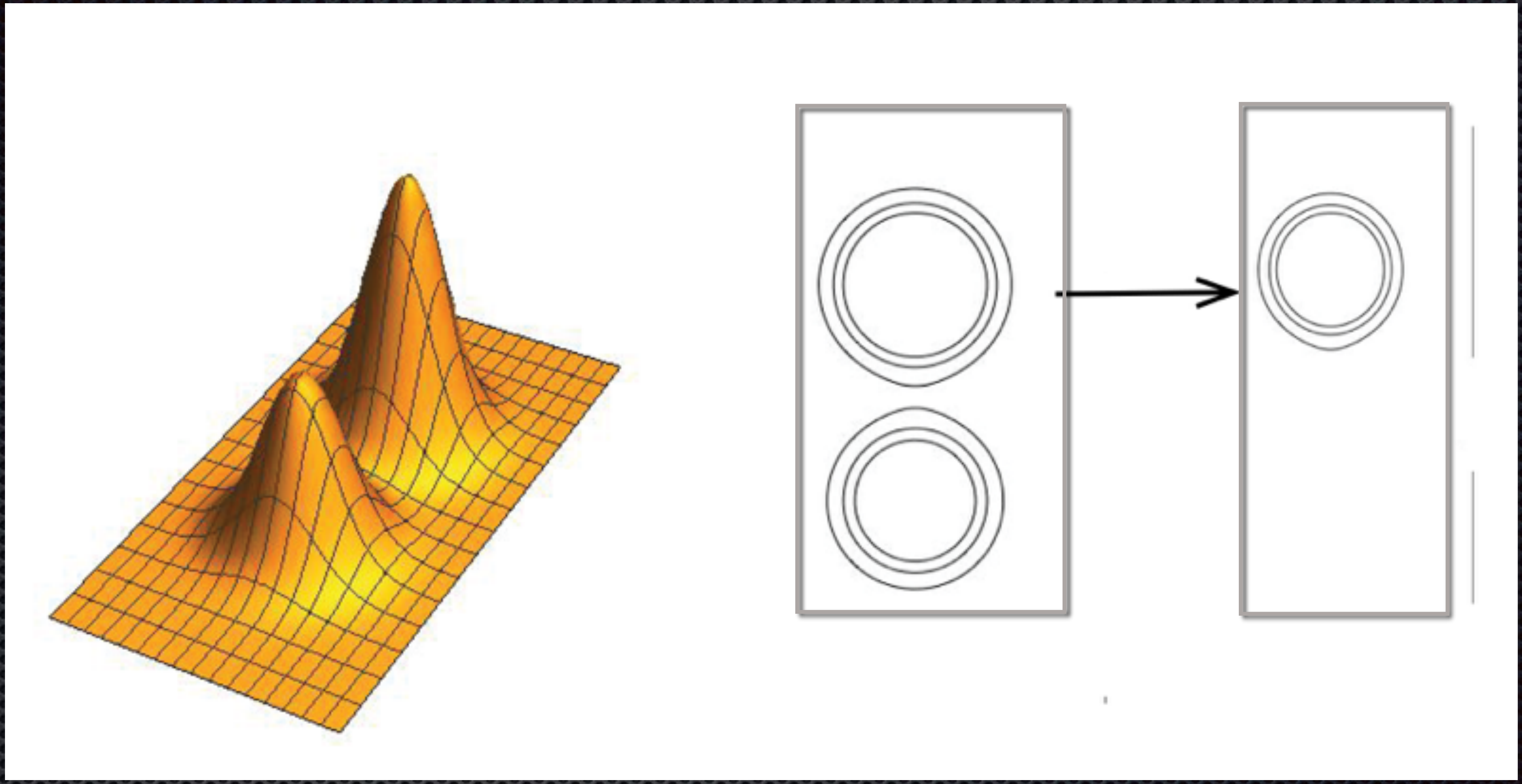
Topology of Critical Level Sets



When t^* is a "positive saddle", $f^{-1}[-\infty, t]$ topology changes:

$\beta_1 + +$

Topology of Critical Level Sets

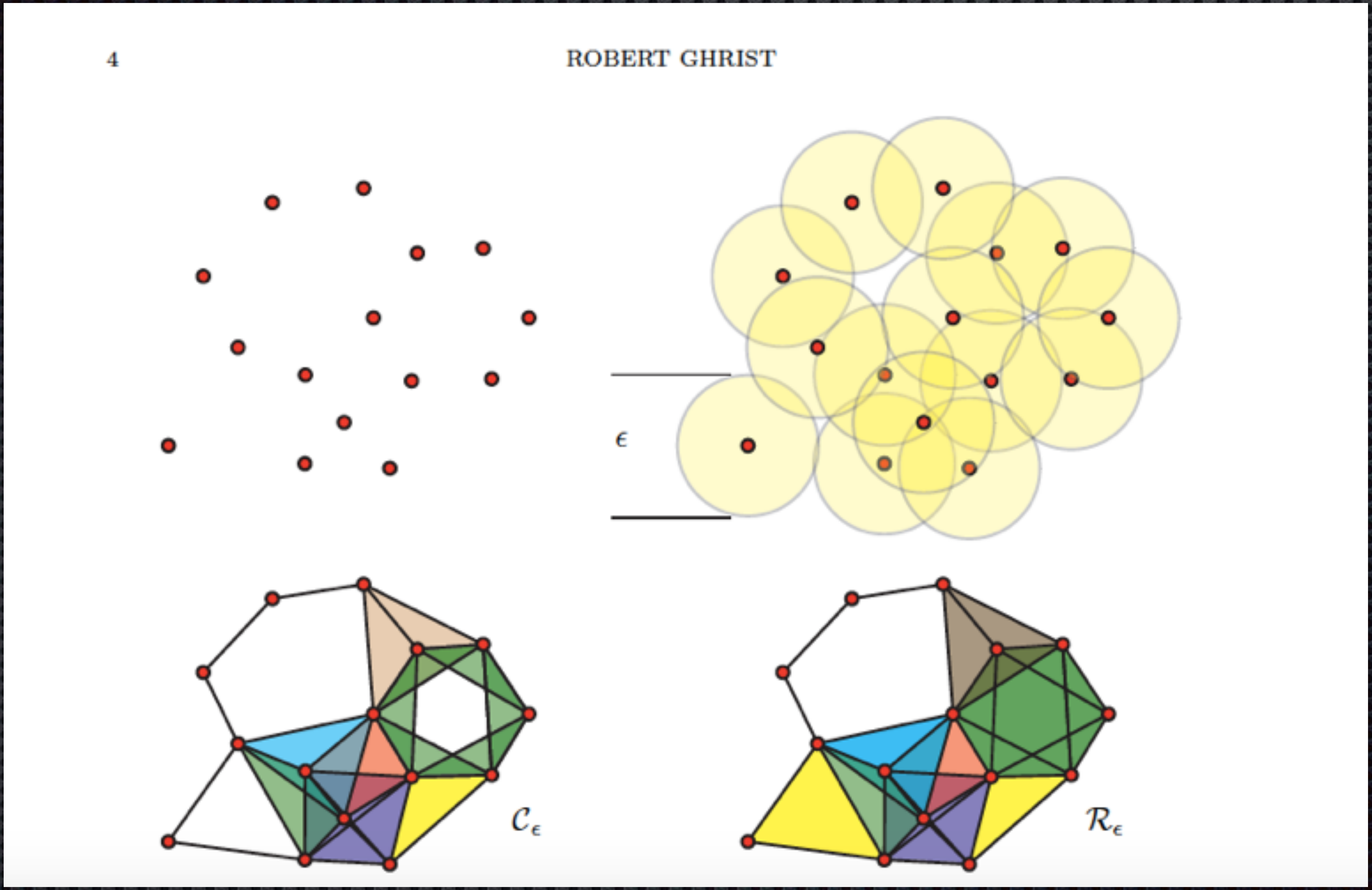


A positive saddle is paired with and eventually destroyed by a local max:

$\beta_1 - -$

Topological Persistence

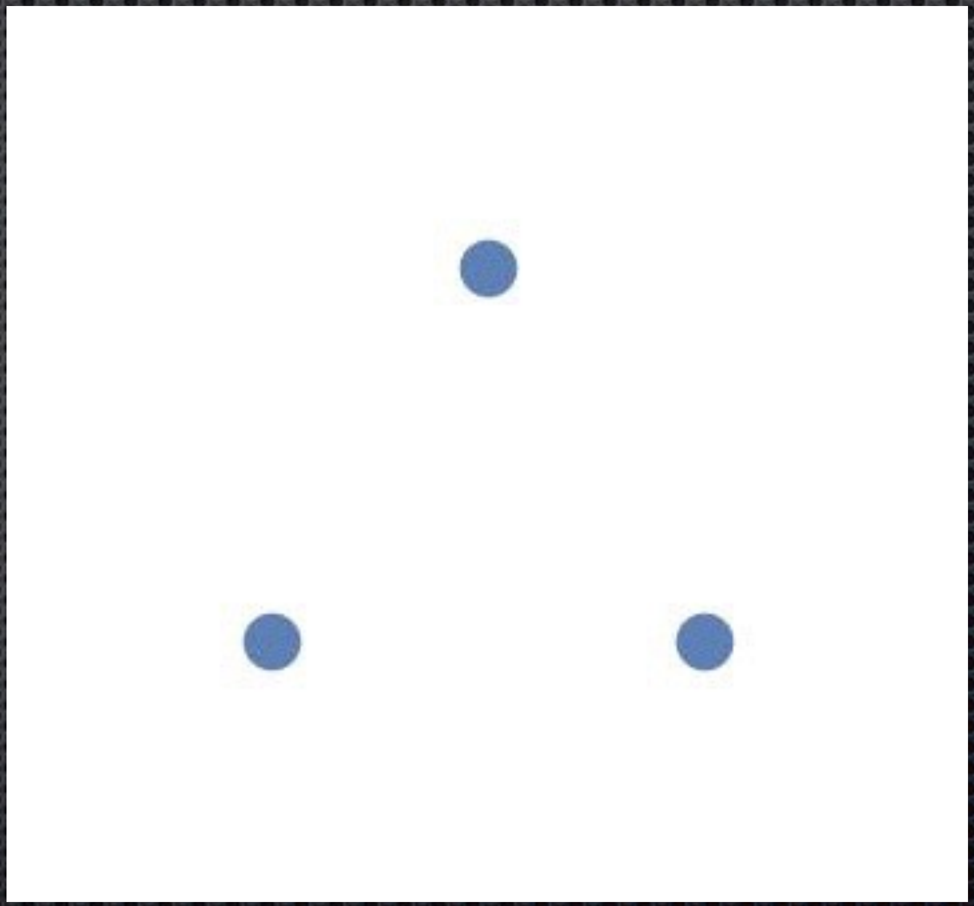
Filtrations via the *Vietoris-Rips complex*:



Topological Persistence

Filtrations via the *Vietoris-Rips complex* and the *Cech* complex:

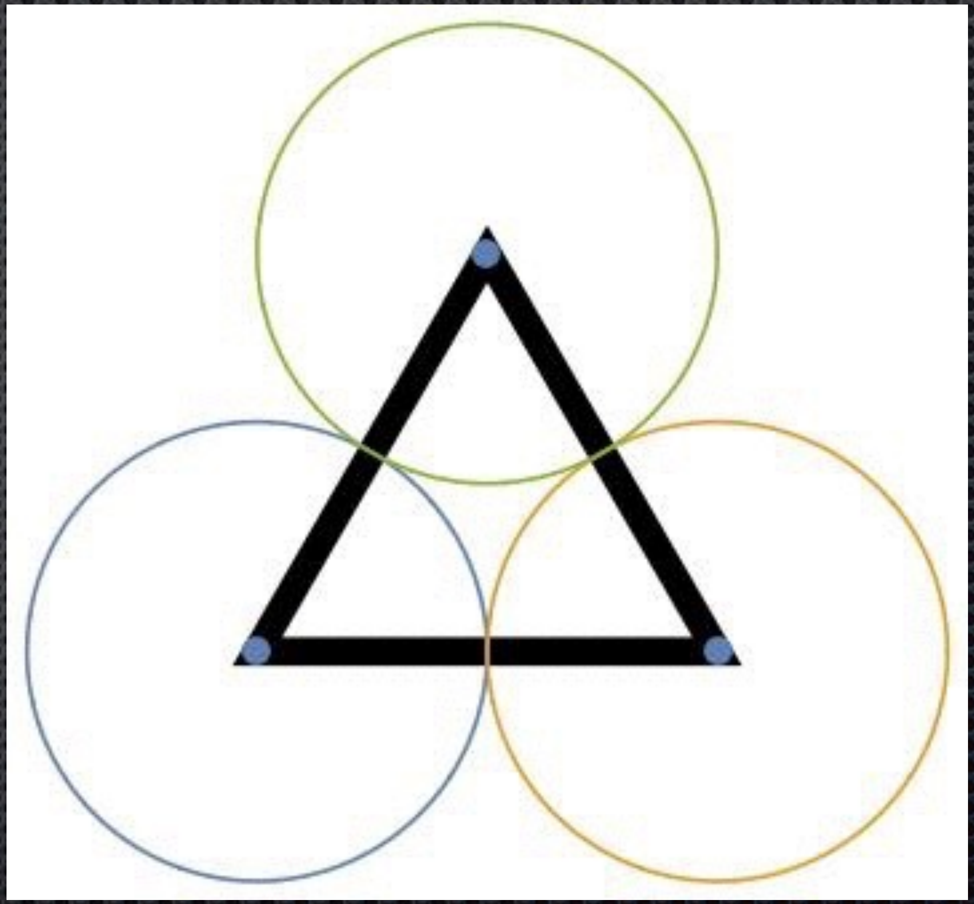
The Cech complex is the abstract simplicial complex whose k-simplices are determined by unordered (k + 1)-tuples of points whose closed r/2-ball neighborhoods have a point of common intersection.



Topological Persistence

Filtrations via the *Vietoris-Rips complex* and the *Cech* complex:

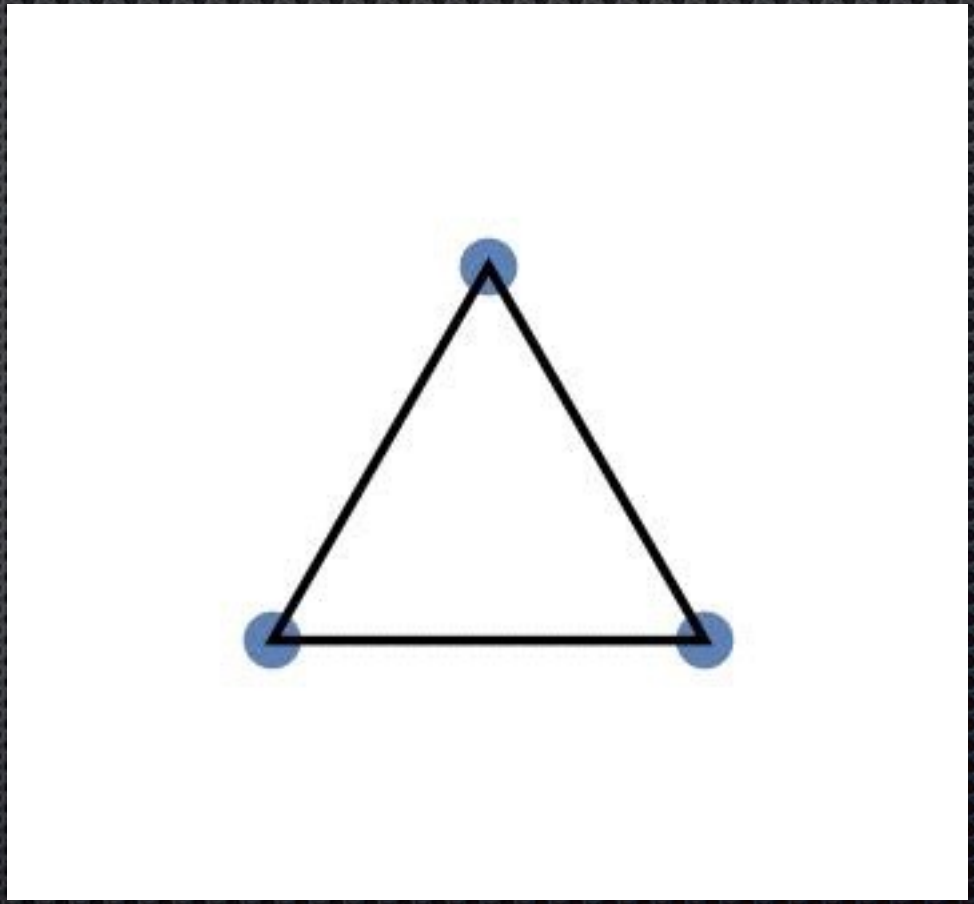
The Cech complex is the abstract simplicial complex whose k-simplices are determined by unordered (k + 1)-tuples of points whose closed r/2-ball neighborhoods have a point of common intersection.



Topological Persistence

Filtrations via the *Vietoris-Rips complex* and the *Cech* complex:

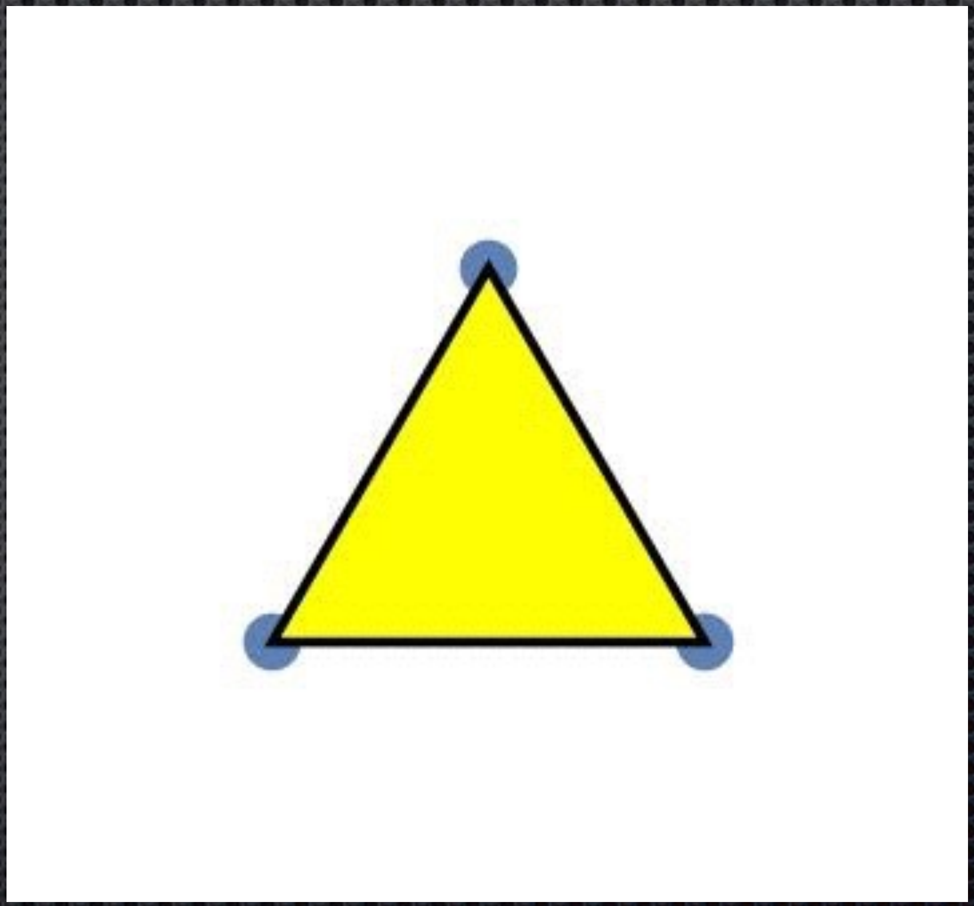
The Cech complex is the abstract simplicial complex whose k-simplices are determined by unordered (k + 1)-tuples of points whose closed r/2-ball neighborhoods have a point of common intersection.



Topological Persistence

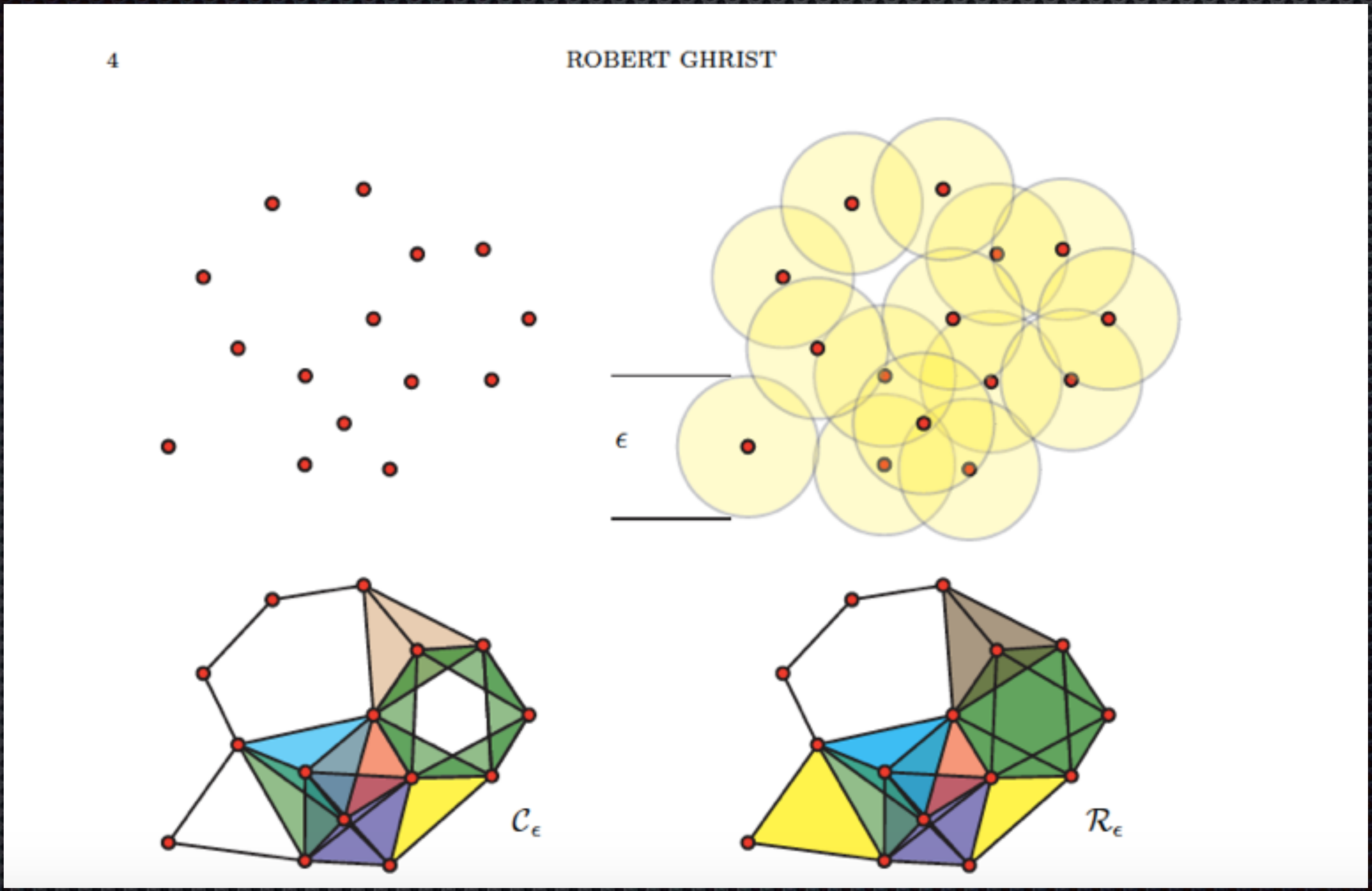
Filtrations via the *Vietoris-Rips complex* and the *Cech* complex:

The Rips complex is the abstract simplicial complex whose k-simplices correspond to unordered (k + 1)-tuples of points which are pairwise within distance r.



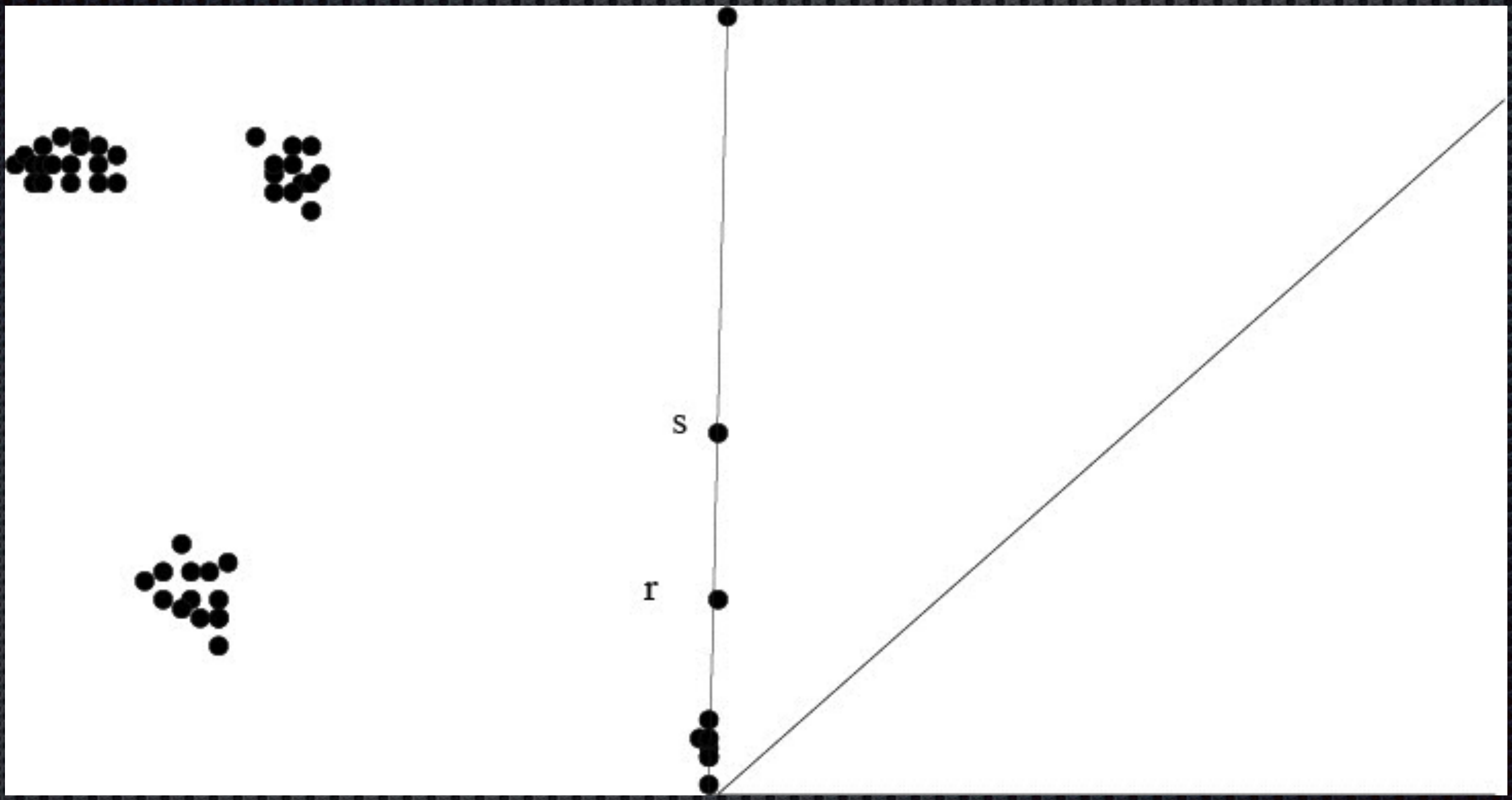
Topological Persistence

Filtrations via the *Vietoris-Rips complex* and the *Cech* complex:



Topological Persistence

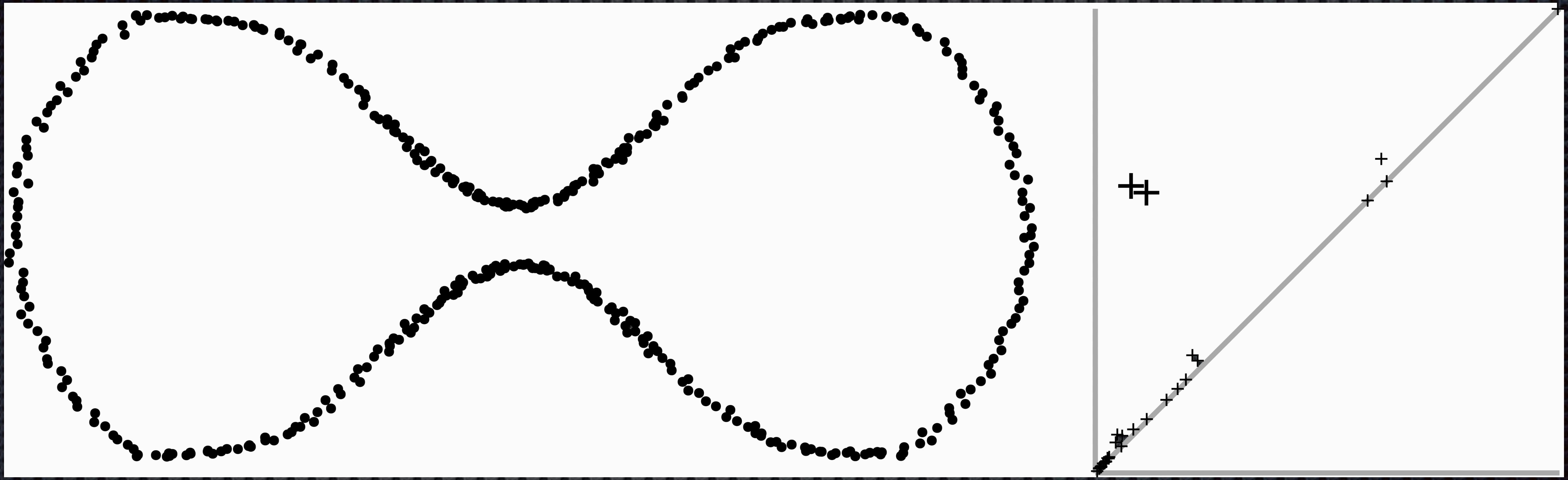
dgm_0 keeps track of Betti number β_0 as a function of the Rips/Čech parameter.



The point *r* is half the distance between the top two clusters, *s* is half the distance between the top right and bottom cluster. There is only "death."

Topological Persistence

dgm_1 keeps track of Betti number β_1 as a function of the Rips/Čech parameter.



Here there is birth and death.

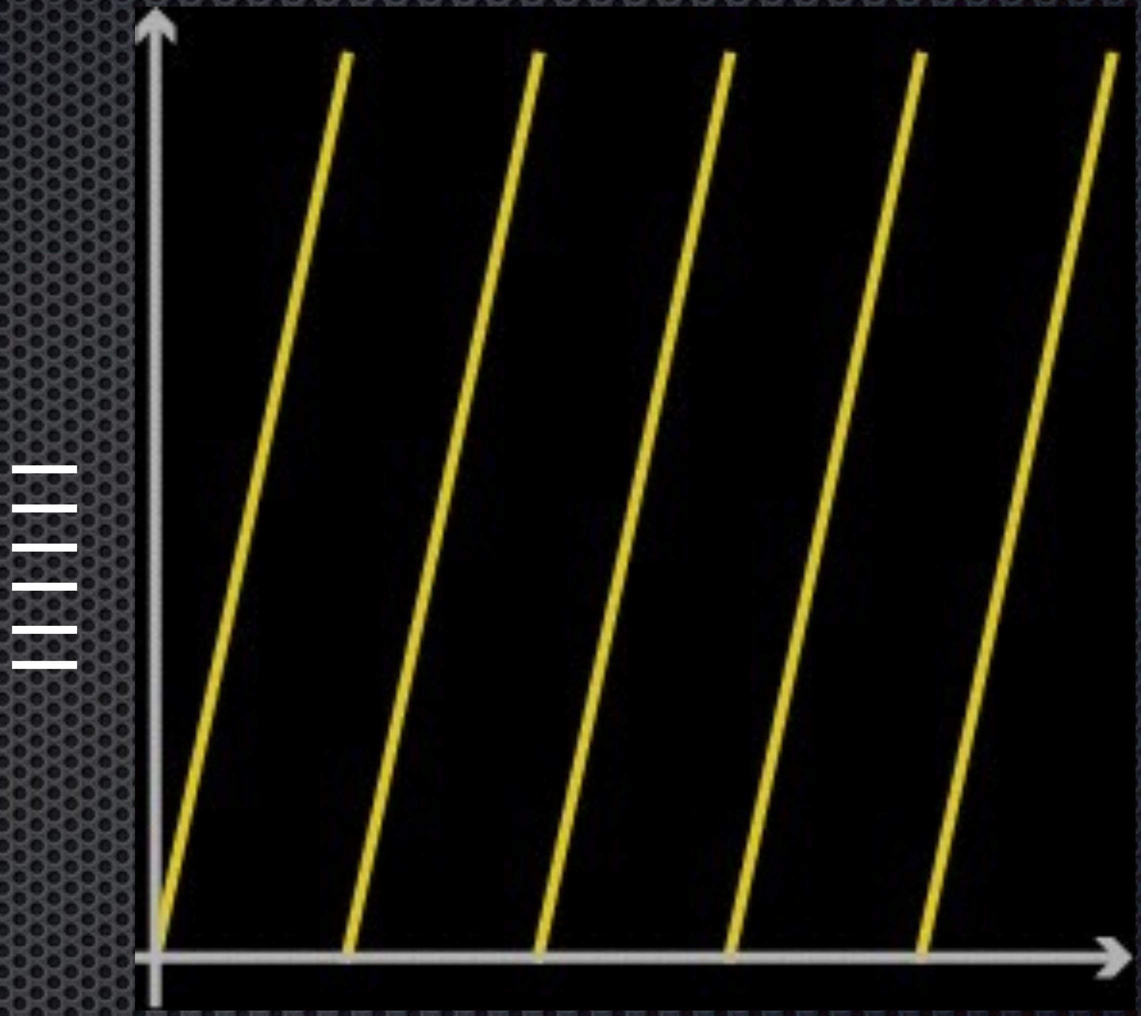
A Sidebar on Entropy

Consider the comb function:

$$f(x) = kx \pmod 1$$

In the equation $y = f(x)$

how much uncertainty do you have about *x* if you know *y*?



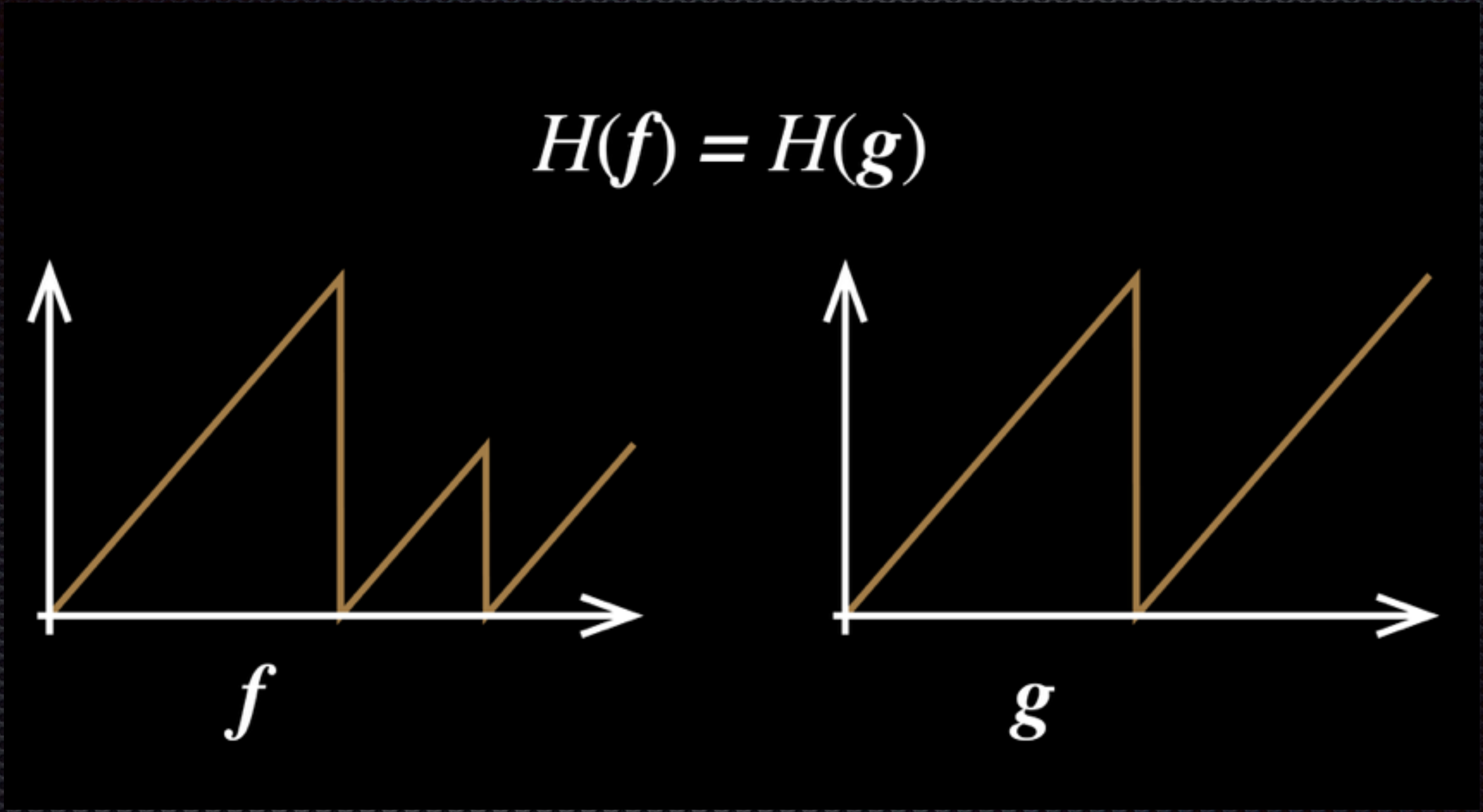
Consider a uniform partition into n subintervals of the range.

$$-\sum_{j=1}^{n} \frac{\mu([y_{k-1}, y_k])}{n} \log_2 \left(\frac{\mu([y_{k-1}, y_k])}{n} \right) = -\sum_{j=1}^{n} \frac{1}{n} \log_2 \frac{1}{n} = \log_2 n$$

Corresponding entropy in the domain:

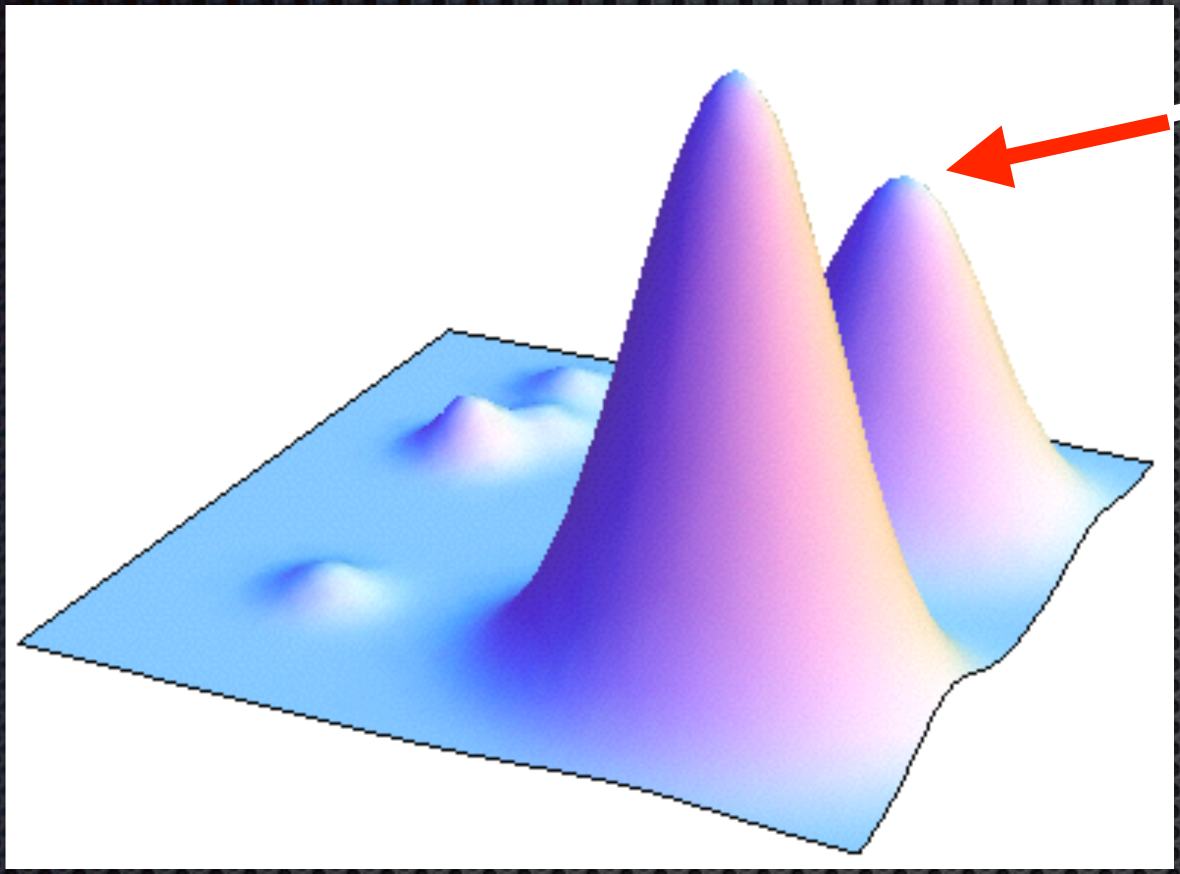
$$-\sum_{i=1}^{k} \sum_{j=1}^{n} \frac{1}{kn} \log_2 \frac{1}{kn} = \log_2 k + \log_2 n$$

A Sidebar on Entropy



Entropy does not capture the intuitive notion of saliency.

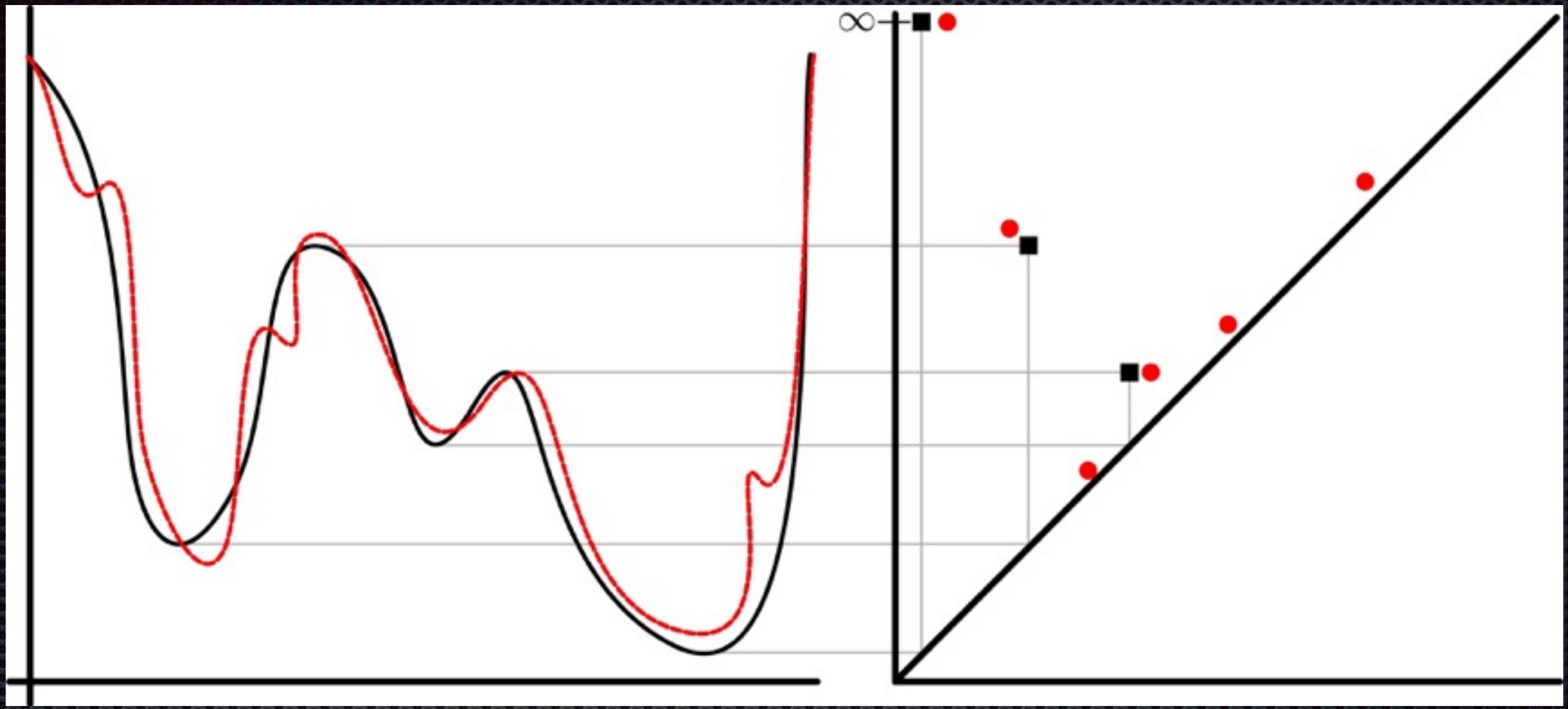
Random fields as information channels - diversity and noise



There is clearly some important diversity.

But it is not captured completely by entropy.

Distinguishing Noise from Features

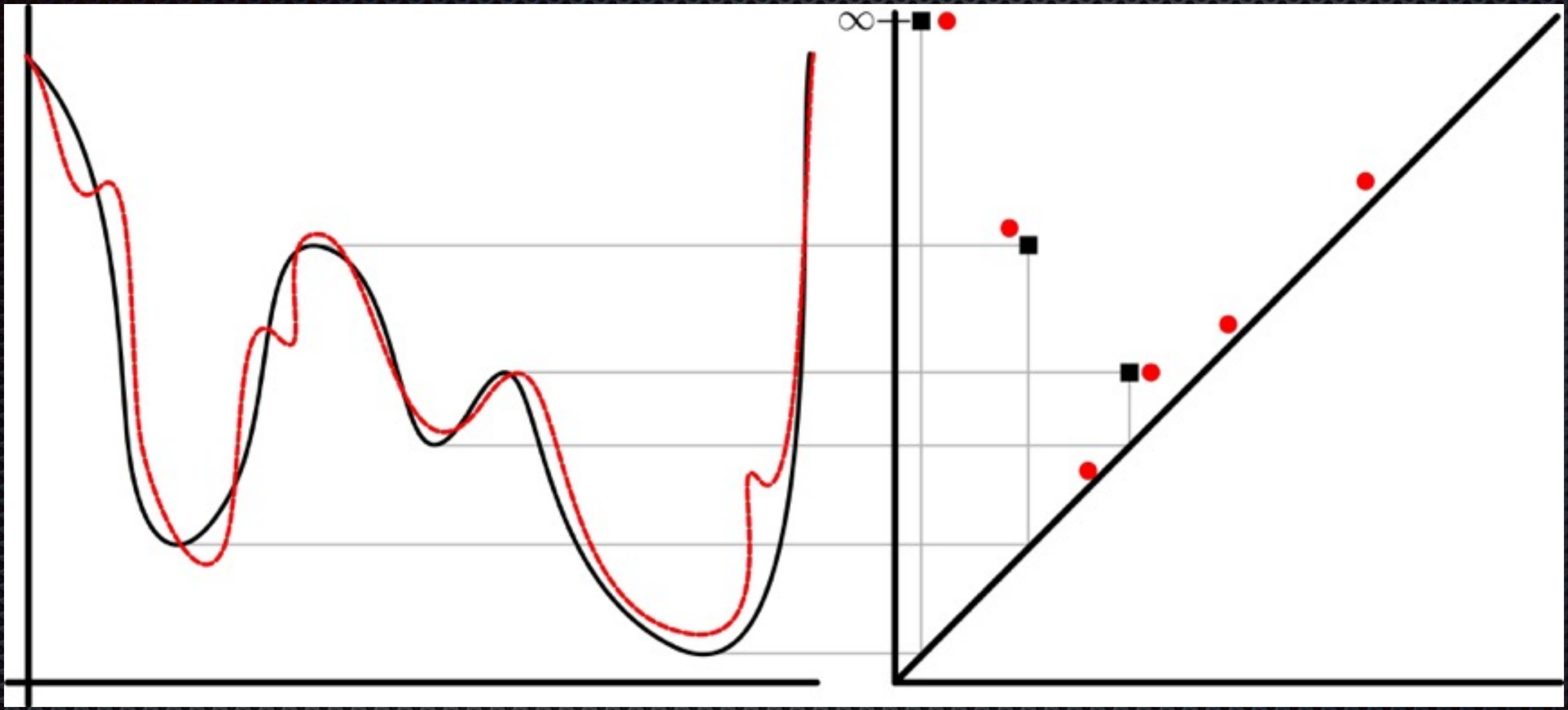


Bottleneck distance between persistence diagrams:

$$W_\infty(X,Y) = \inf_{\eta:X\to Y} \max\{\|x-\eta(x)\|_\infty\}$$

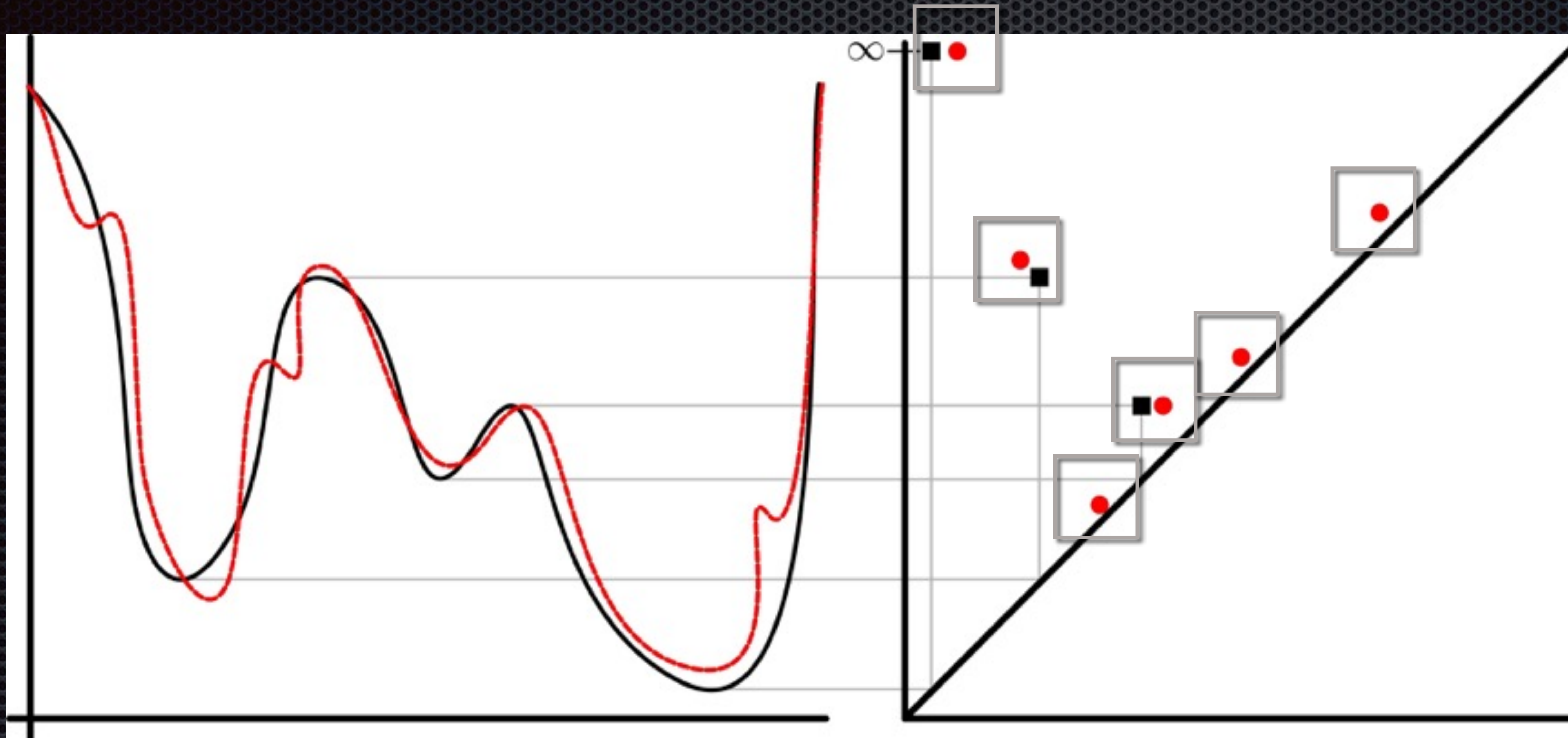
The $\inf$ is over all bijections between persistence diagrams.

Distinguishing Noise from Features



Theorem $W_\infty(X_f, Y_g) \leq \|f - g\|_\infty.$

Distinguishing Noise from Features



Theorem $W_\infty(X_f, Y_g) \leq \|f - g\|_\infty.$

Topological Entropy

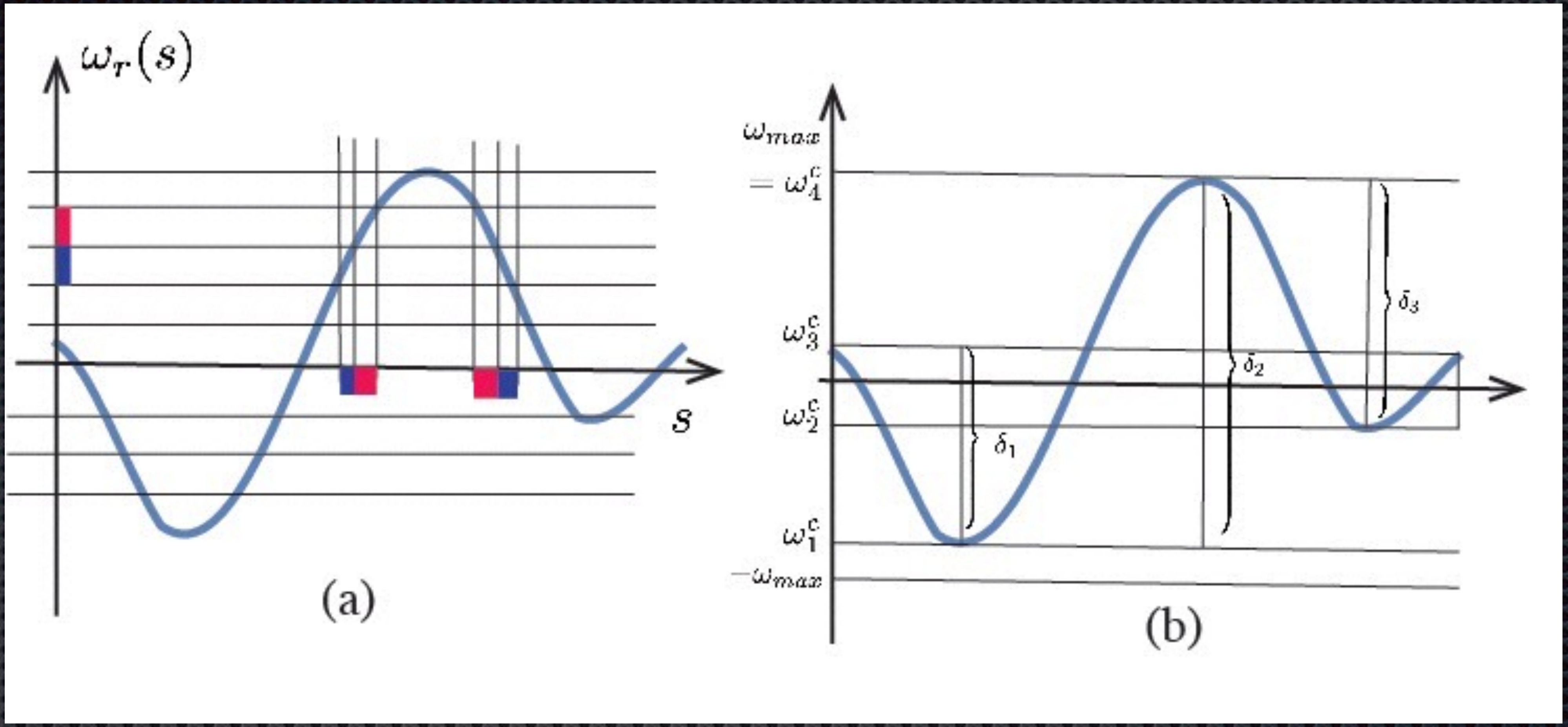
R. L. Adler, A. G. Konheim and M. H. McAndrew, "Topological entropy," *Trans. Amer. Math. Soc.* 114 (1965), 309-319. MR 30 #5291.

Let X be a compact topological space, and let $\mathcal{U}$ be an open cover.

The *entropy* $h(f,\mathcal{U})$ of a mapping $f : X \to X$ with respect to a cover $\mathcal{U}$ is defined as $\lim_{n\to\infty} H(\mathcal{U} \cup f^{-1}\mathcal{U} \cup \cdots \cup f^{-n+1}\mathcal{U})/n$, where H is the partition entropy defined as $H(\mathcal{A}) = \log N(\mathcal{A})$ for aany partition $\mathcal{A}$.

The *entropy* of f, $h(f)$ is the supremum over all covers $\mathcal{U}$ of X.

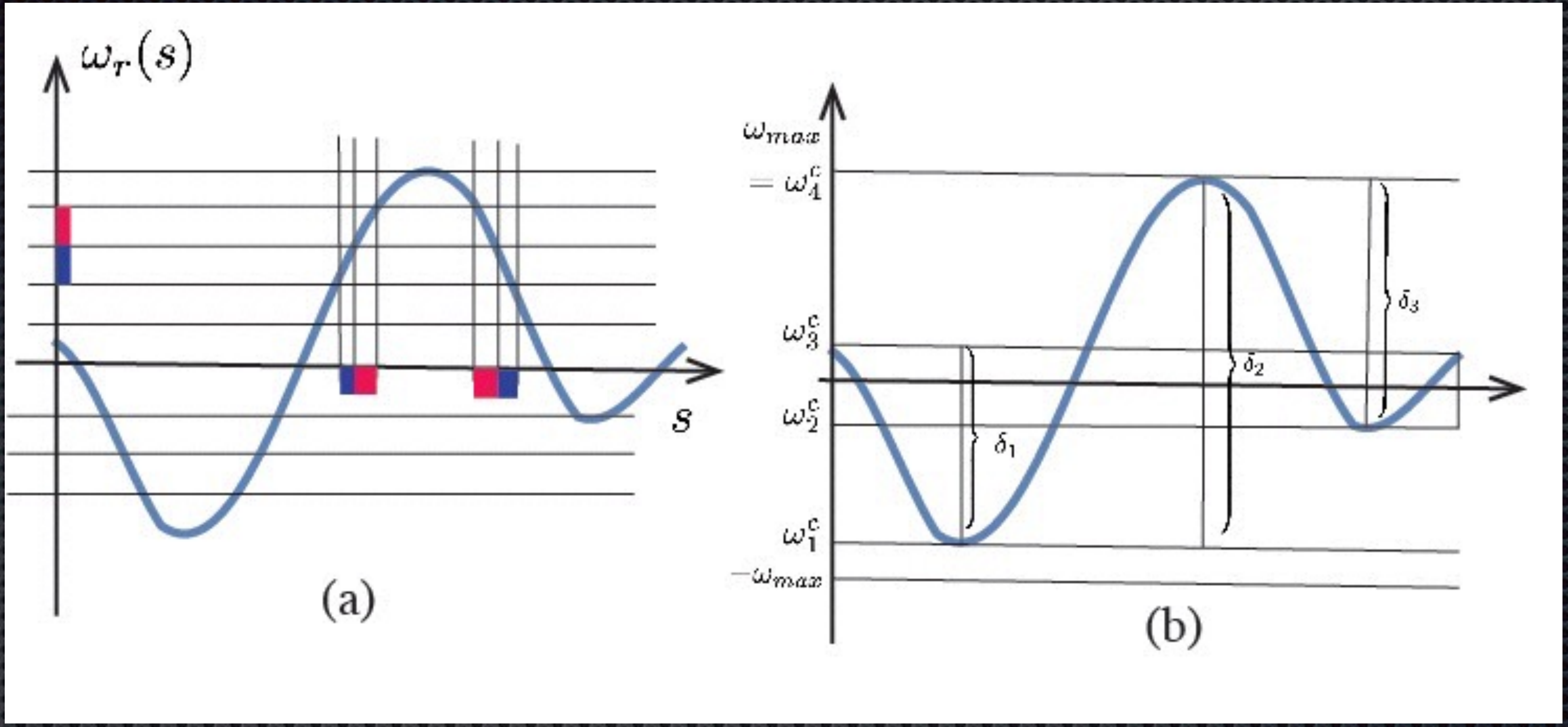
Topological Entropy of Piecewise Monotonic Interval Maps



The vertical axis parameter is ω_r, and the horizontal axis parameter is *s*.

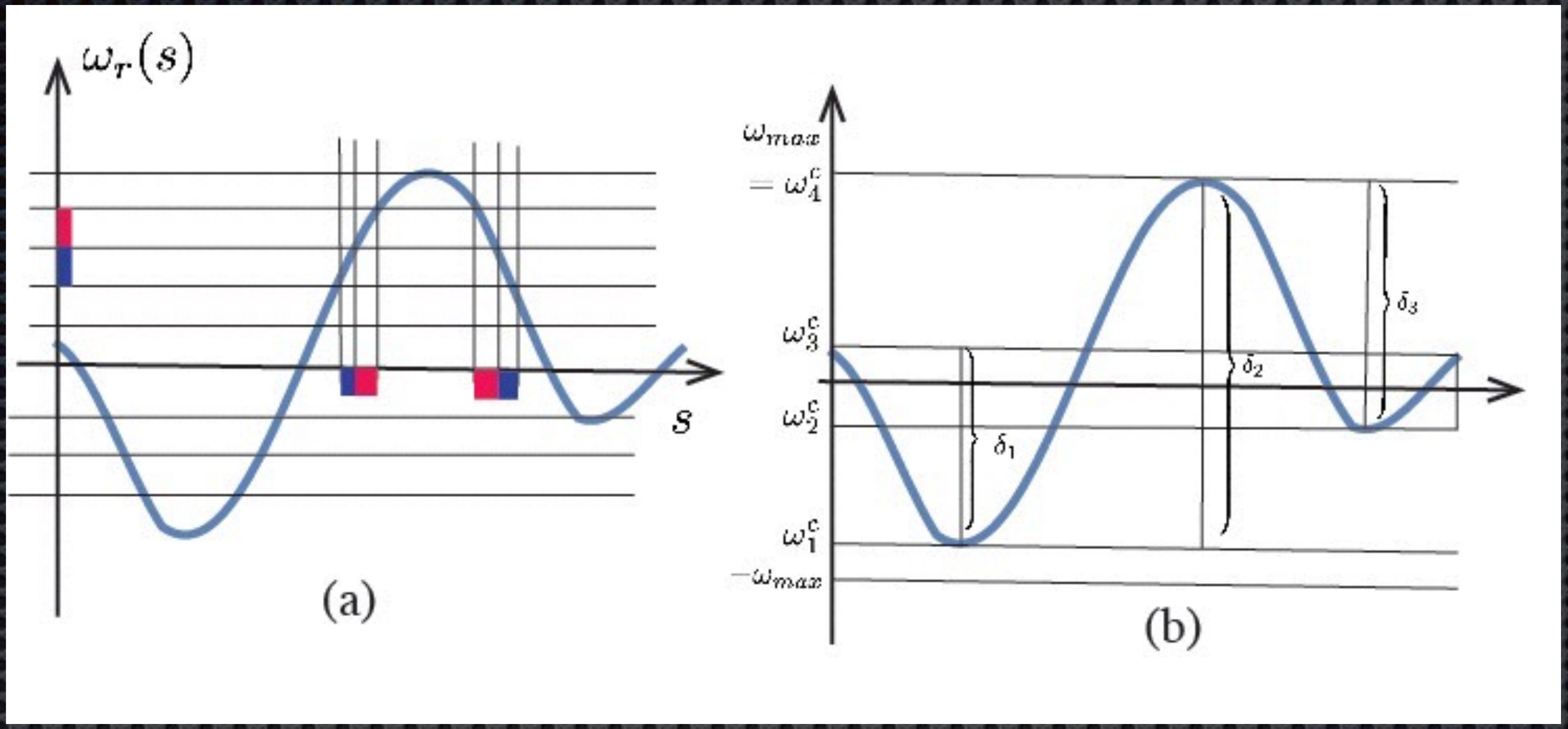
How much information does ω_r convey regarding *s*?

Topological Entropy of Piecewise Monotonic Interval Maps



We assume that ω_r takes values in a bounded range: $\omega_{\min} \leq \omega_r \leq \omega_{\max}$

Topological Entropy of Piecewise Monotonic Interval Maps



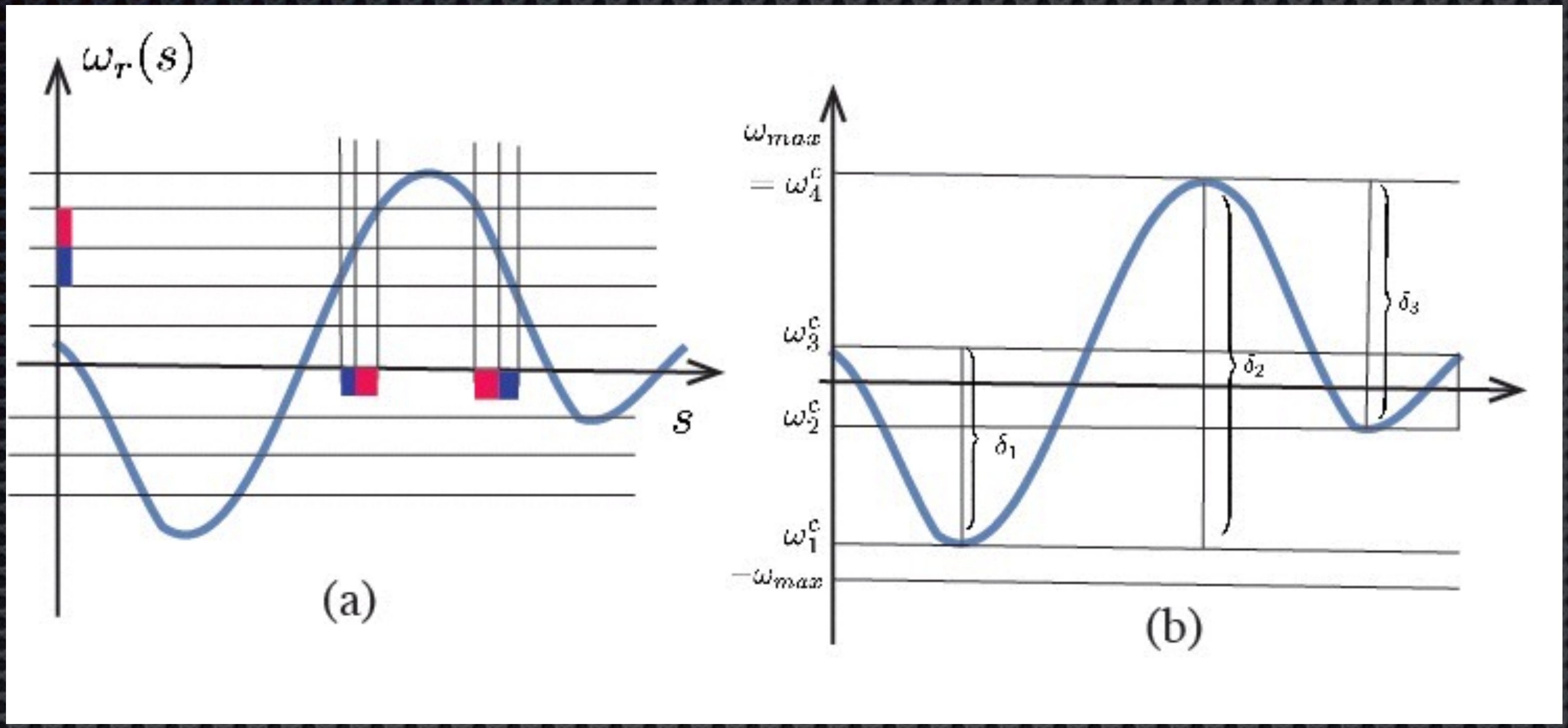
Partition the range into uniform subintervals.

$$(\omega_{max} - \omega_{min})/n \, : \, \omega_{min} = \omega_0 < \omega_1 < \cdots < \omega_n = \omega_{max}$$

This induces a partition in the domain.

$$\mathcal{V}_n = \bigcup_{k=1}^{n} cc\{\omega_r^{-1}([\omega_{k-1}, \omega_k]) = \{V_1, \ldots, V_N\}$$

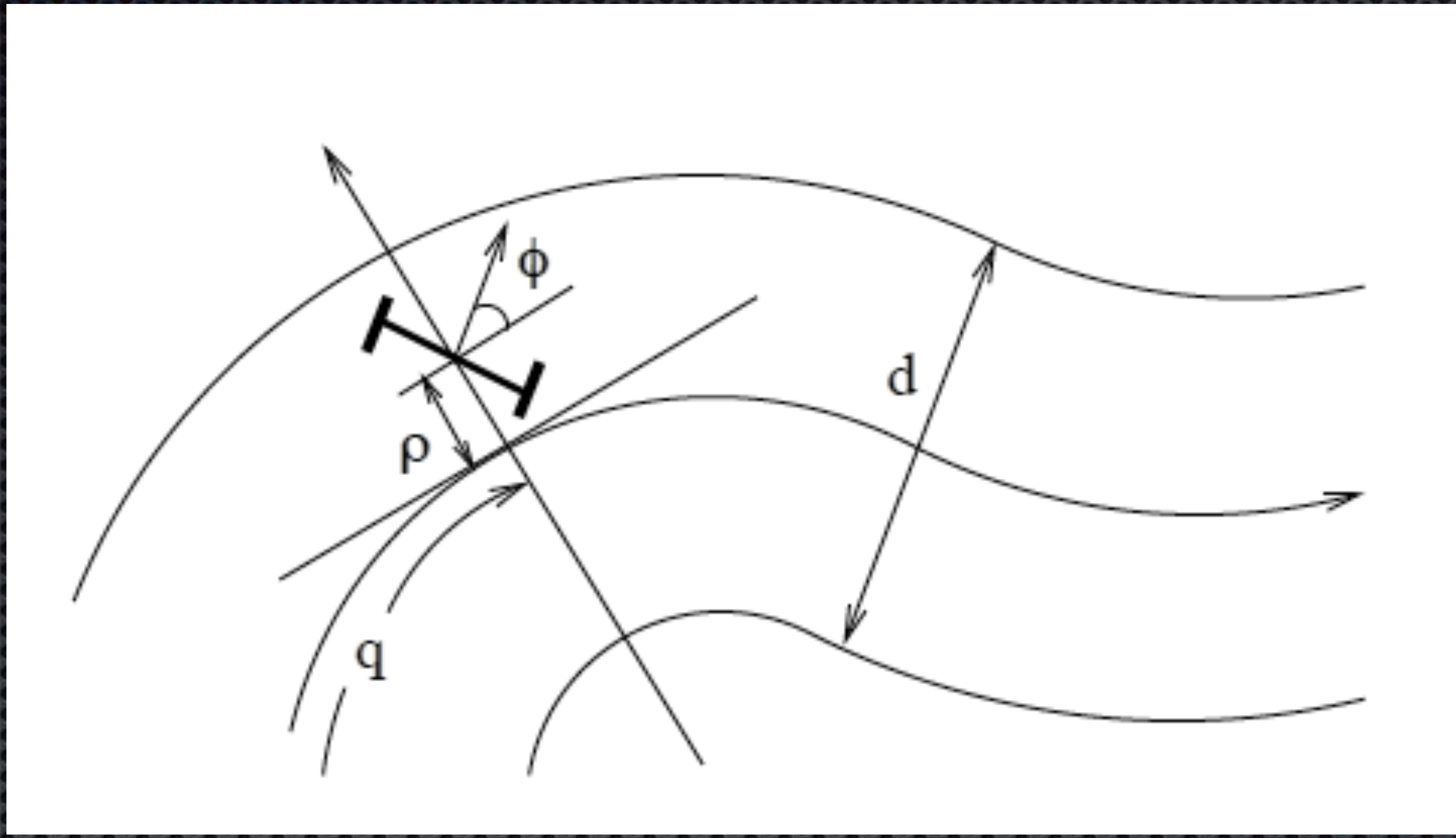
Topological Entropy of Piecewise Monotonic Interval Maps



The *range partition entropy* induced by the mapping is:

$$H(\mathcal{V}_n) = -\sum_{V_k \in \mathcal{V}_n} \frac{\mu(V_k)}{L} \log_2 \frac{\mu(V_k)}{L}, \text{ where } L = \sum_{V_k \in \mathcal{V}_n} \mu(V_k).$$

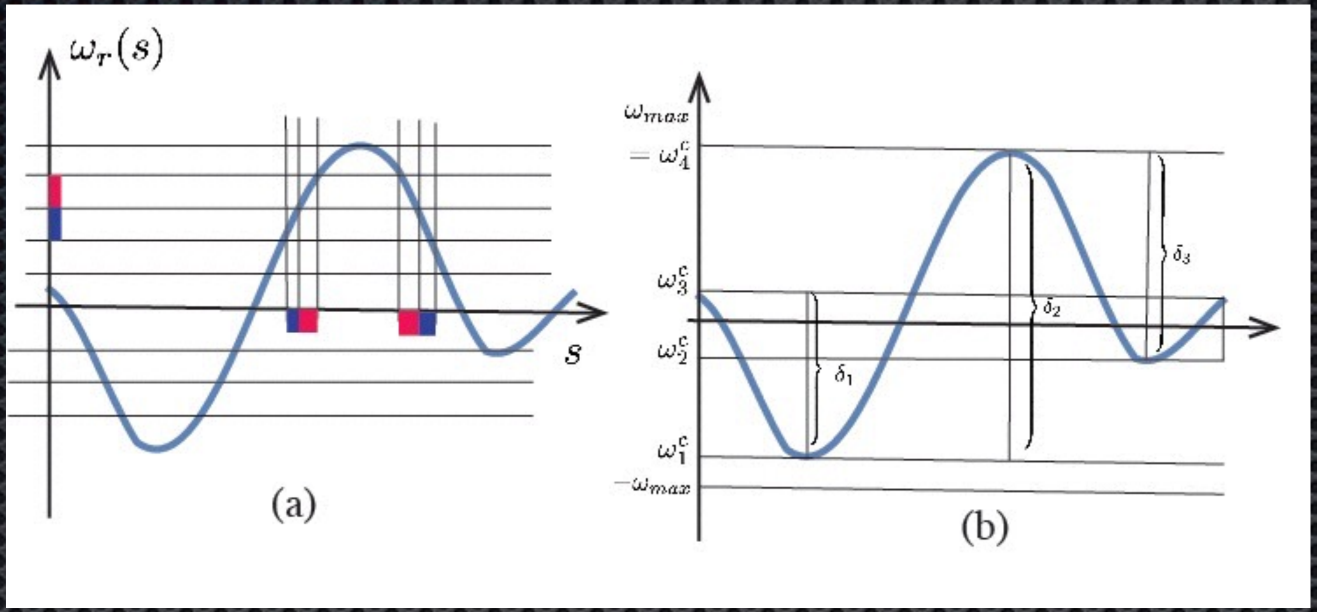
Topological Entropy — How do we distinguish significant features from noise?



$$\begin{cases} \frac{dx_r}{ds} = \cos\theta_r(s) \\ \frac{dy_r}{ds} = \sin\theta_r(s) \quad (A) \\ \frac{d\theta_r}{ds} = \omega_r(s) \end{cases}$$

$$\begin{aligned} \dot{q} &= \cos\phi \frac{1}{1+\omega_r(q)\rho} \\ \dot{\rho} &= -\sin\phi \qquad (B) \\ \dot{\phi} &= -\omega(t) - \cos\phi \frac{\omega_r(q)}{1+\omega_r(q)\rho}. \end{aligned}$$

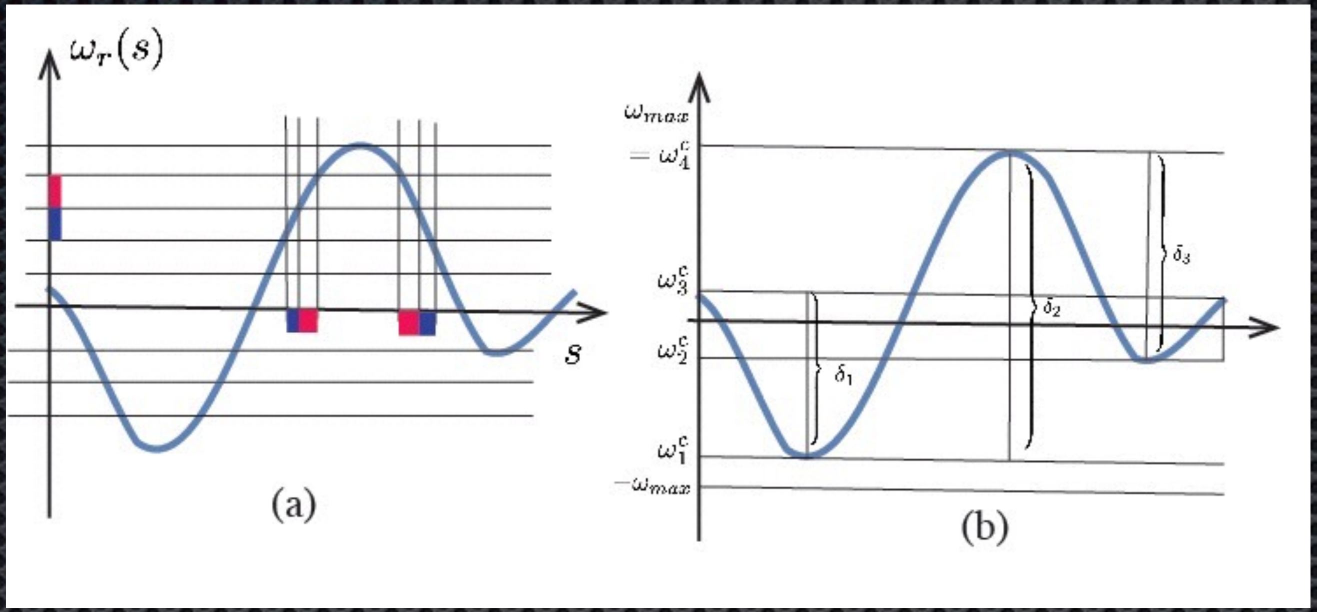
Topological Entropy — How do we distinguish significant features from noise?



A partition that is (possibly much) coarser than the range partition is the *critical point partition*. Enumerate critical values: $Vr(\omega_r) = \{\omega_1^c < \cdots < \omega_\ell^c\}$

In terms of the partition $\mathcal{V}_{cr} = \bigcup cc\{\omega_r^{-1}(\Delta_k)\}$ (which is just the segments on which ω_r in monotone), we have another partition entropy: $H_{cr} = -\sum_{V \in \mathcal{V}_{cr}} \frac{\mu(V)}{L} \log_2 \frac{\mu(V)}{L}$

Topological Entropy — How do we distinguish significant features from noise?



$$\textbf{Theorem :} \quad \lim_{n\to\infty} \{H(\mathcal{V}_n) - \log_2 n\} \le H_{cr} + \sum_{V\in\mathcal{V}_{cr}} \frac{\mu(V)}{L} \log_2 \frac{\delta_k}{\omega_{max} - \omega_{min}}.$$

The δ_k s are the differences between successive critical values.

Topological Entropy — How do we distinquish significant features from noise?

In the case of dimensions 2 and higher, similar quantities are defined:

$$H(\mathcal{M}) + \sum_{i=1}^{n} \frac{\mu(M_i)}{\mu(X)} \log_2 \delta_i$$

This may be taken as a proxy fort *salient information*.

Information Based Image Segmentation

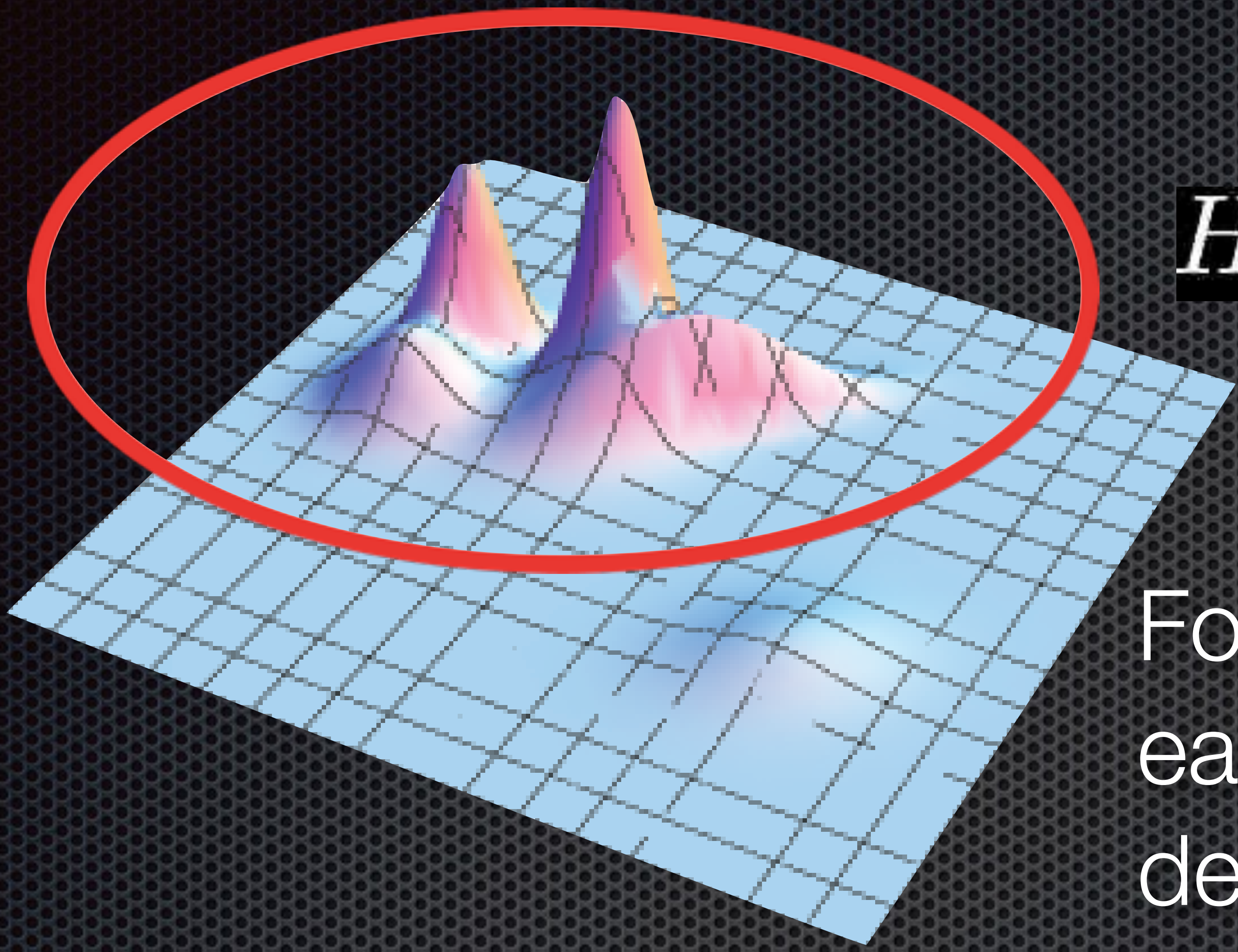
1. Reconnaissance of time-varying fields
2. Concepts of sensor fusion

Given multiple sensor fields $f_1, \ldots, f_N$ on a common domain X, partition X into sub-domains $Y_1, \ldots, Y_N$ such that on each sub-domain Y_j the *j*-th sensor field f_j is maximally informative.

Sensor-fusion: Form composite

$$f(x) = f_j(x) \text{ if } x \in Y_j$$

Entropy conditioned on a set



$$H(\alpha|Y) > H(\alpha)$$

For each sensor f_j and each subdomain Y, define associated critical set partition

$$\mathcal{M}_j = \mathcal{M}(f_j, Y) = cc(Y \backslash Cr(f_j, Y))$$

Concepts in sensor fusion

- Given a sensor field f_j and any subset compute

$$u(f_j, Y) = -\sum_{M_i \in \mathcal{M}_j} \frac{\mu(M_i \cap Y)}{\mu(Y)} \log_2 \frac{\mu(M_i \cap Y)}{\mu(Y)} + \sum_{M_i \in \mathcal{M}_j} \frac{\mu(M_i \cap Y)}{\mu(Y)} \log_2 \delta_{ij}$$

- Let $Y_j = \arg\max_{Y \subset X} u(f_j, Y)$
- Get a set of possibly overlapping subdomains

$$\{Y_1, \ldots, Y_N\}$$

Information utility in sensor-based imaging

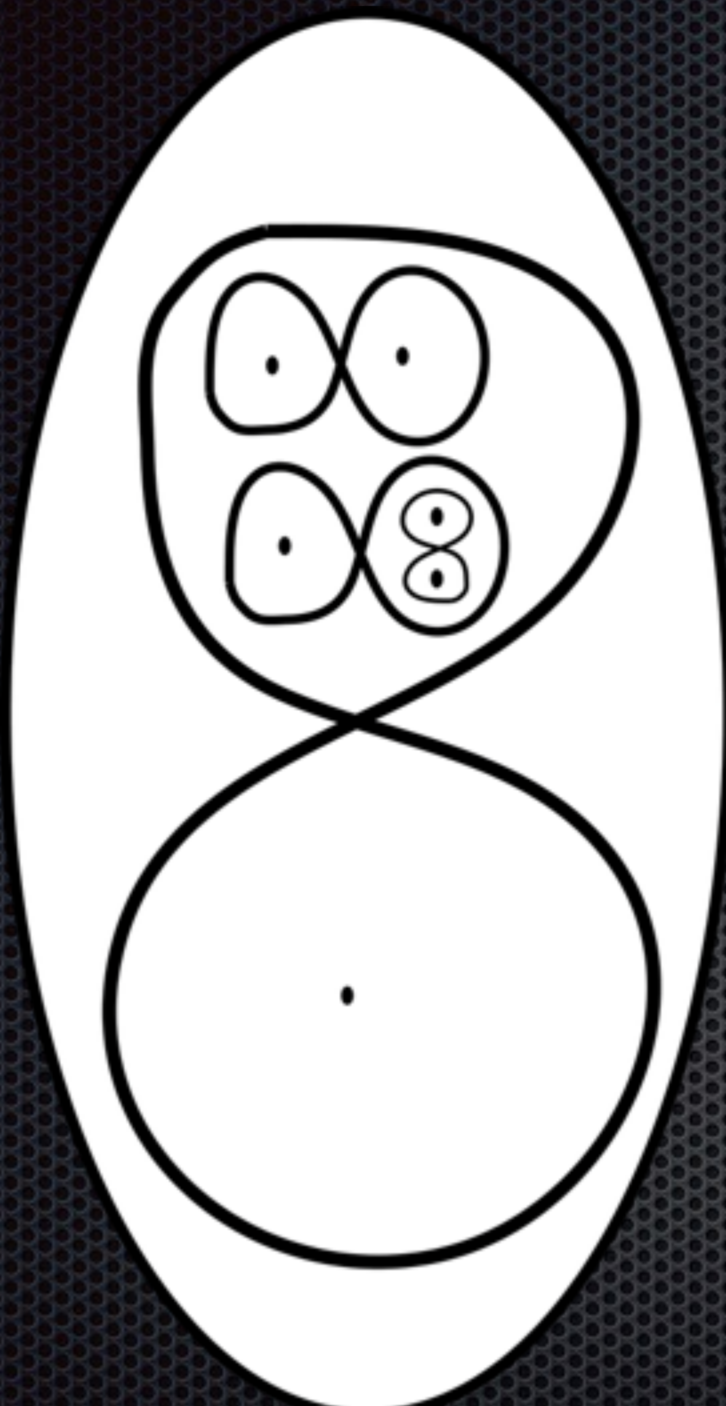
- Goal: Reconcile/merge field data from different sensors on nonempty overlaps $Y_{i_1} \cap \cdots \cap Y_{i_k} \neq \emptyset$
- Find partitions $\mathcal{Y} = \{\hat{Y}_1, \ldots, \hat{Y}_N\}$ that maximize the *information utility*

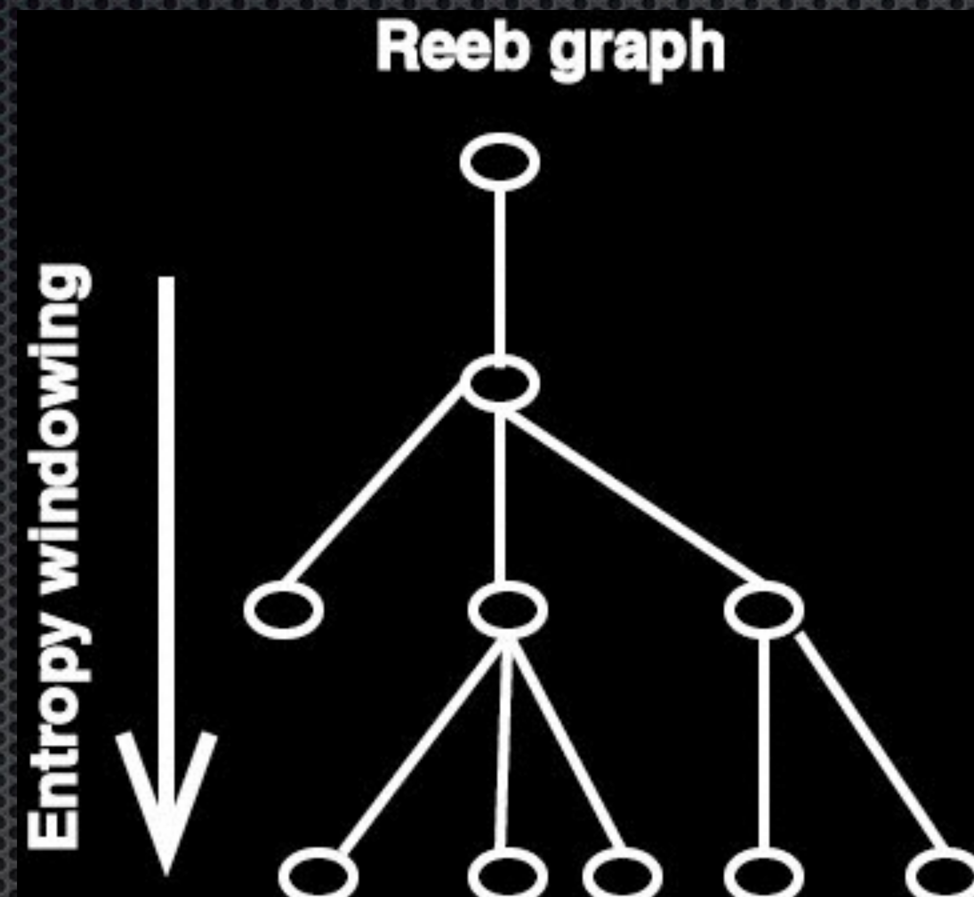
$$U(f_1, \cdots, f_N, \mathcal{Y}) = \sum_{j=1}^{N} \frac{\mu(\hat{Y}_j)}{\mu(X)} u(f_j | \hat{Y}_j).$$

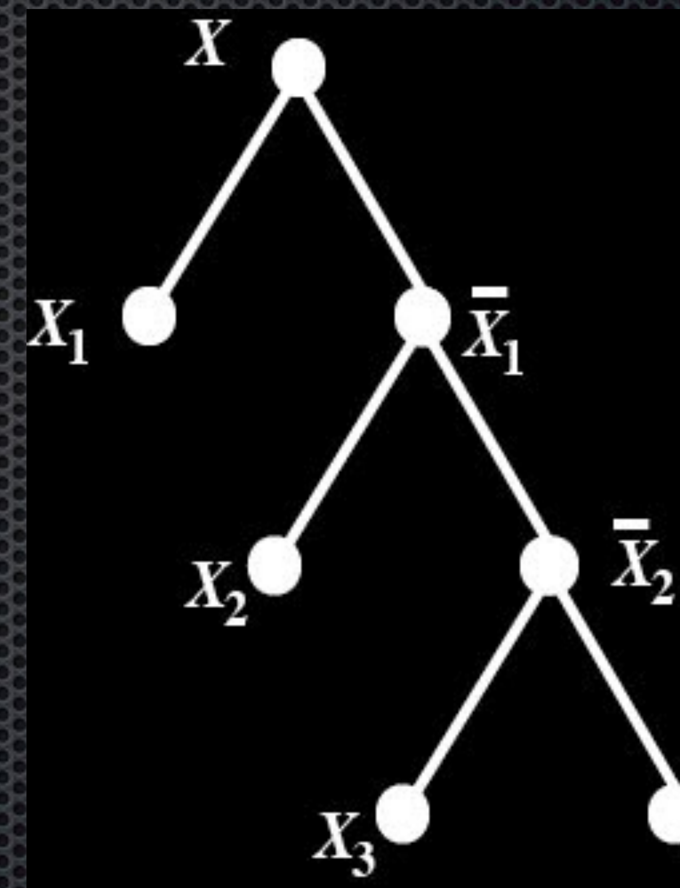
where

$$u(f|Y_{C_I}) = - \sum_{M_I^i \in \mathcal{M}_I} \frac{\mu(M_I^i \cap Y_{C_I})}{\mu(Y_{C_I})} \log_2 \frac{\mu(M_I^i \cap Y_{C_I})}{\mu(Y_{C_I})} + \sum_{M_I^i \in \mathcal{M}_I} \frac{\mu(M_I^i \cap Y_{C_I})}{\mu(Y_{C_I})} \log_2 \delta_{iI}$$

Information based segmentation using topological motifs







Maximum diversity decomposition

$$X_1, X_2, \ldots, X_k$$

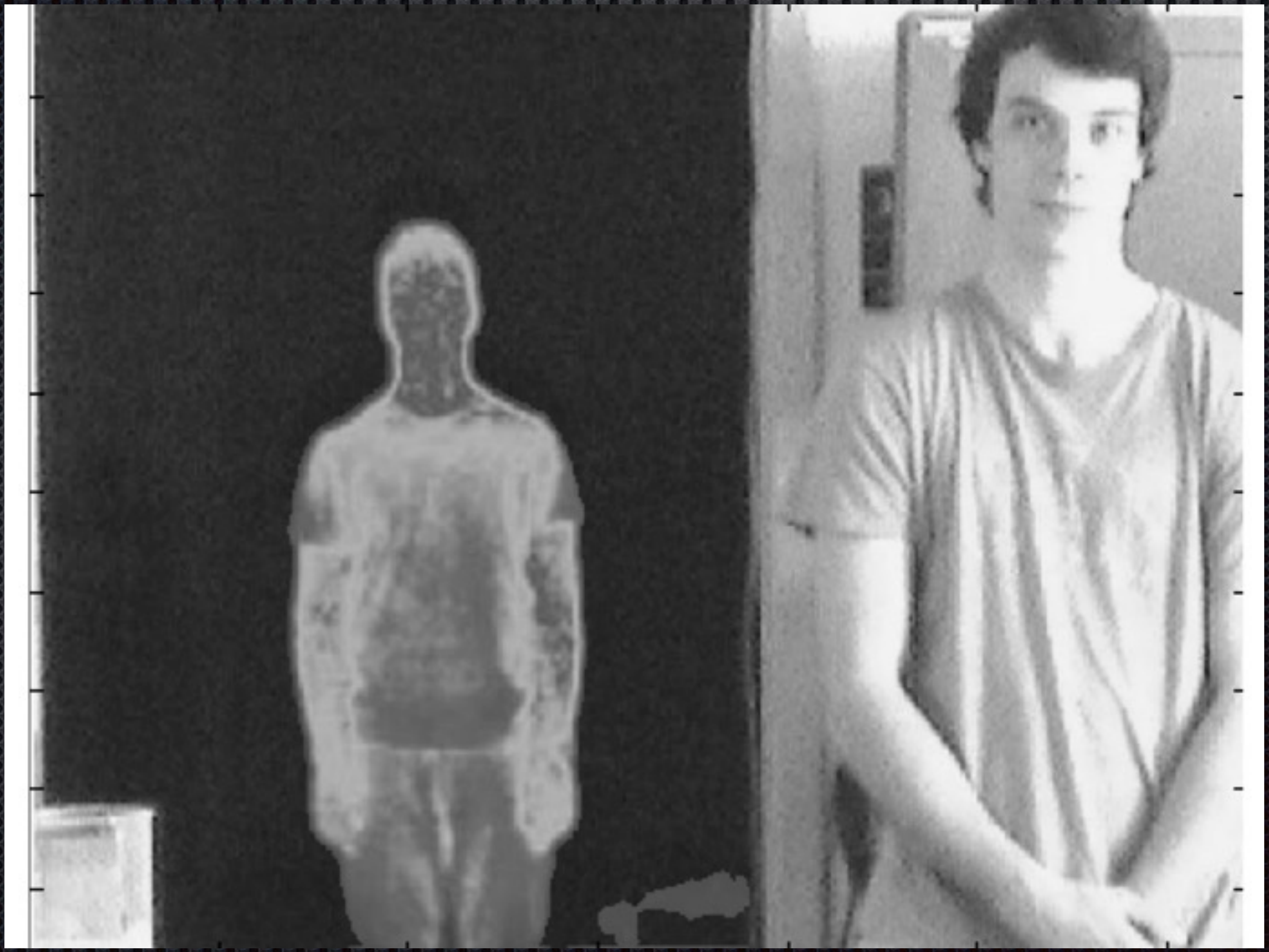
$$X = \cup_{j=1}^{k} X_j$$

Enhanced perception from sensor fusion

Visual spectrum

Infrared spectrum

Enhanced perception from sensor fusion



Conclusion

- Topological methods examine data relationships that may be missed by PCA and other essentially linear approaches to analytics.
- Such methods appear to be useful in data compression of multiband images.
- These methods also provide a baseline for studying human performance in directing reconnaissance (See tomorrow's talk.) and in studying visual cognitive styles.
- Current research is aimed at understanding how to extend this circle of ideas to time-varying fields and how to extend information theory to point cloud data sets.