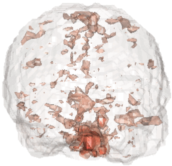
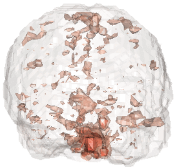




Leonhard Euler
1707 – 1783

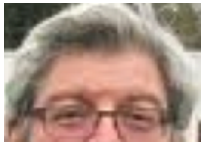
Pondering Persistence
and
Extolling Euler

Robert Adler

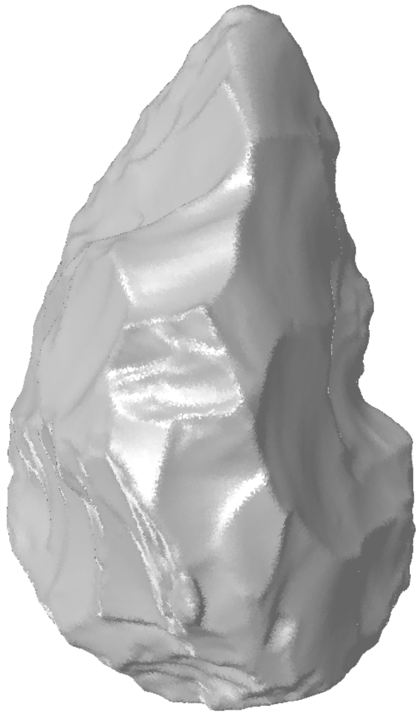
Electrical Engineering

Technion – Israel Institute of Technology

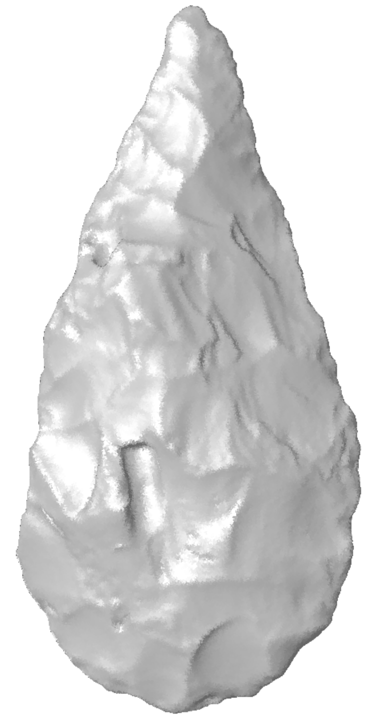
and a whole bunch of other people



A prehistoric problem

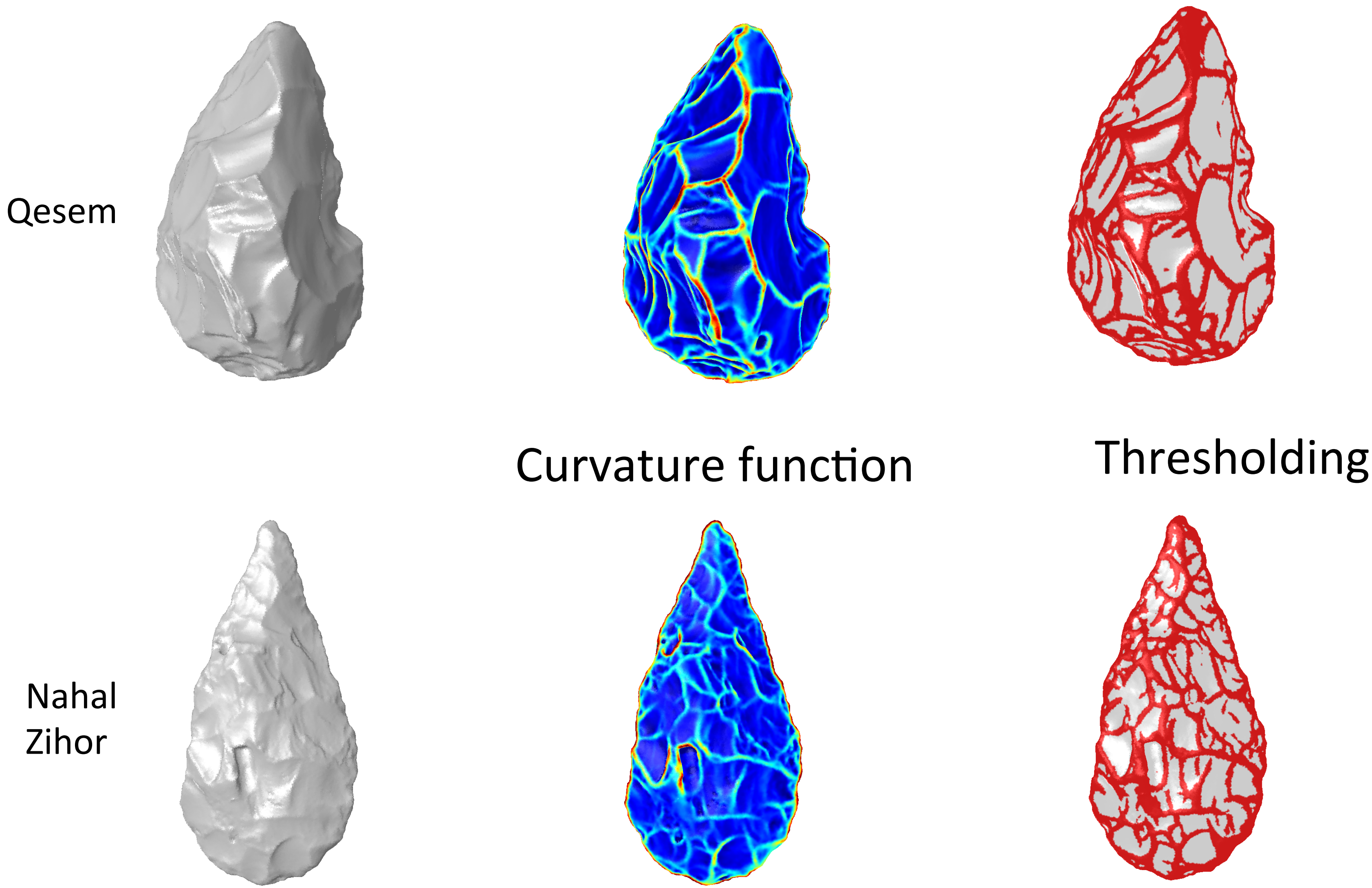


Qesem



Nahal Zihor

Thresholding via the curvature function



How should we analyze data?

Gunnar Carlsson

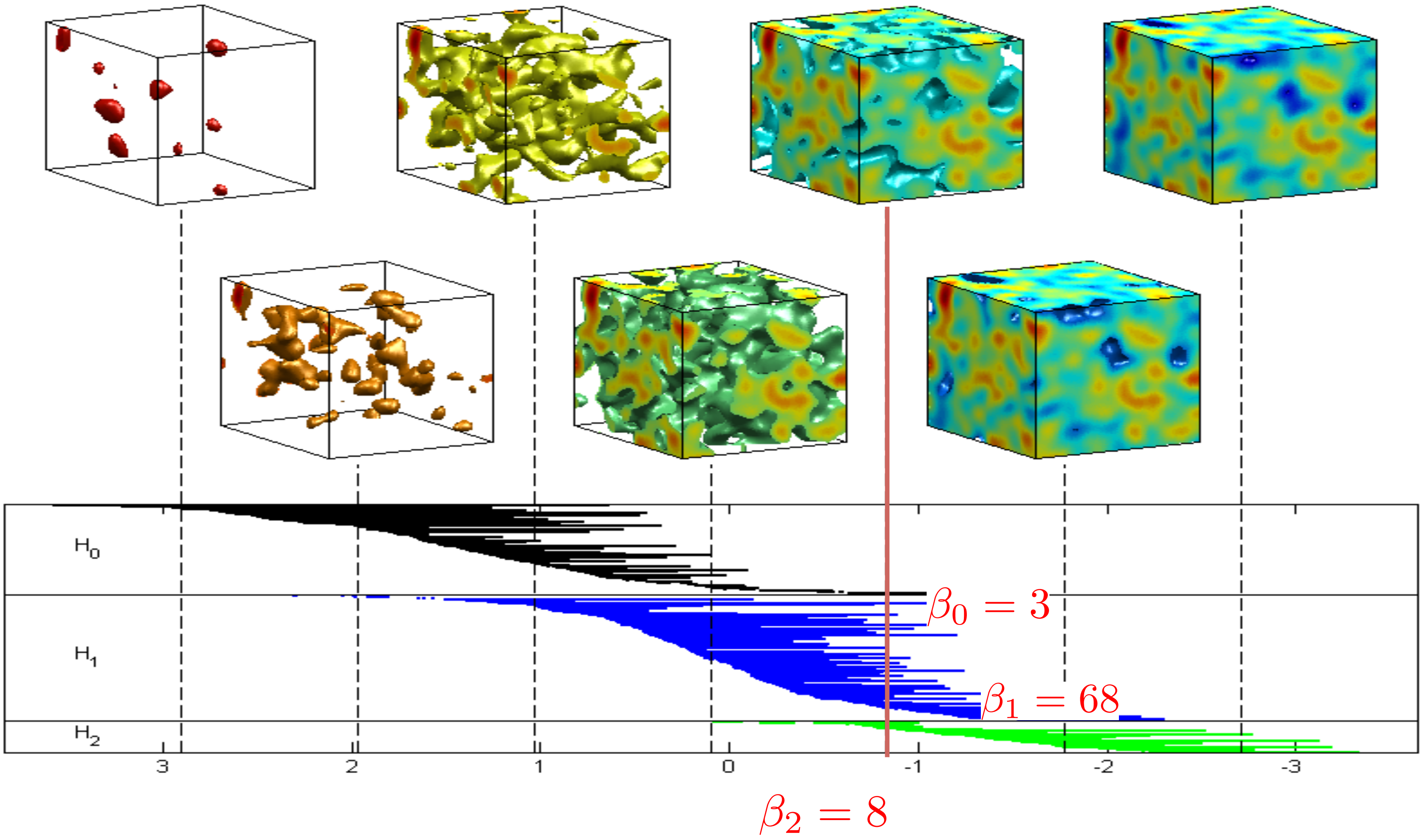
TOPOLOGISTS DO IT WITH PERSISTENCE

John Harer

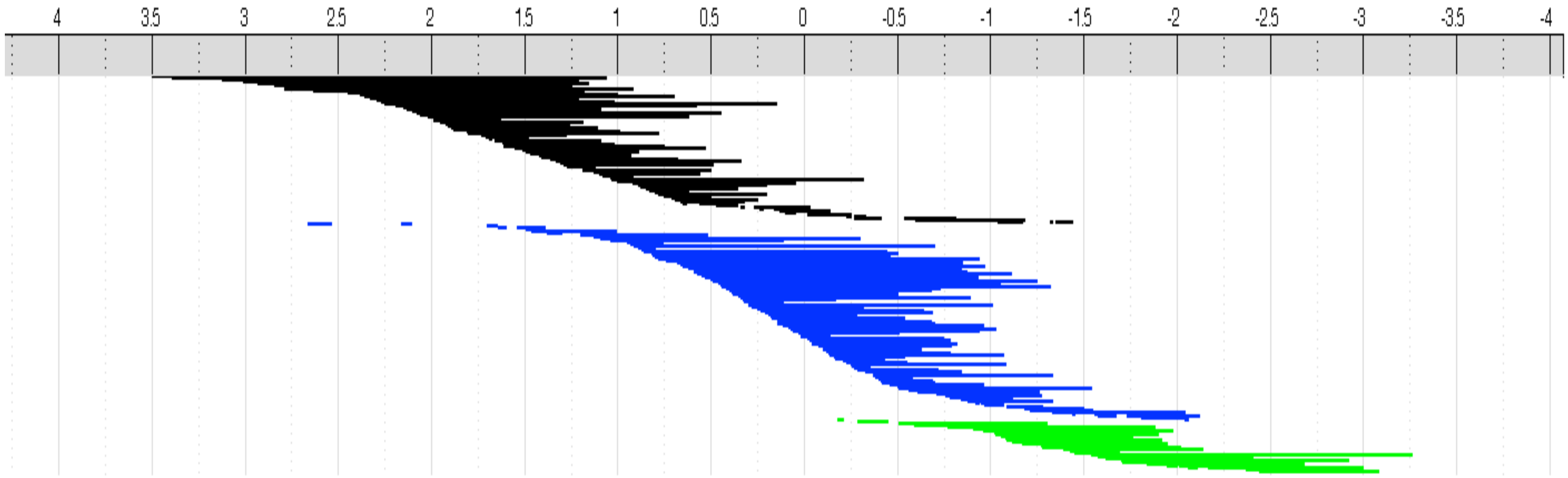
Herbert Edelsbrunner

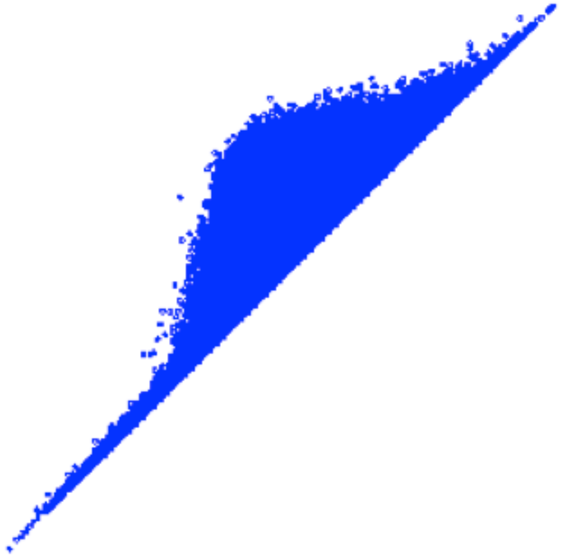
Robert Ghrist

Persistence for dummies

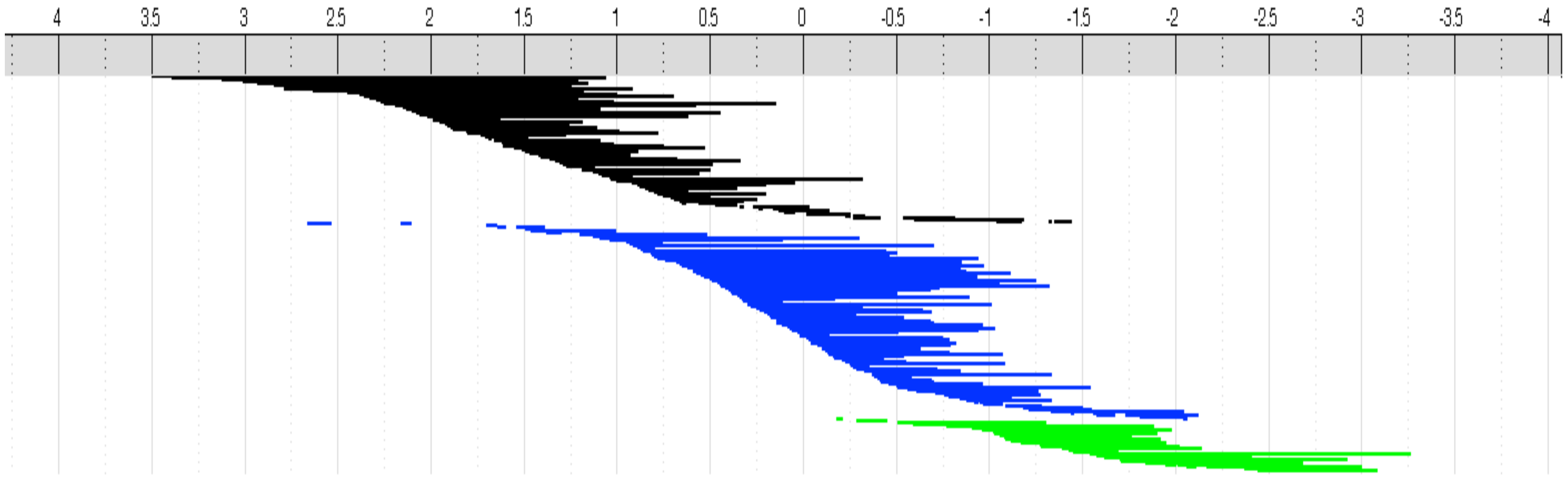


Persistence diagrams

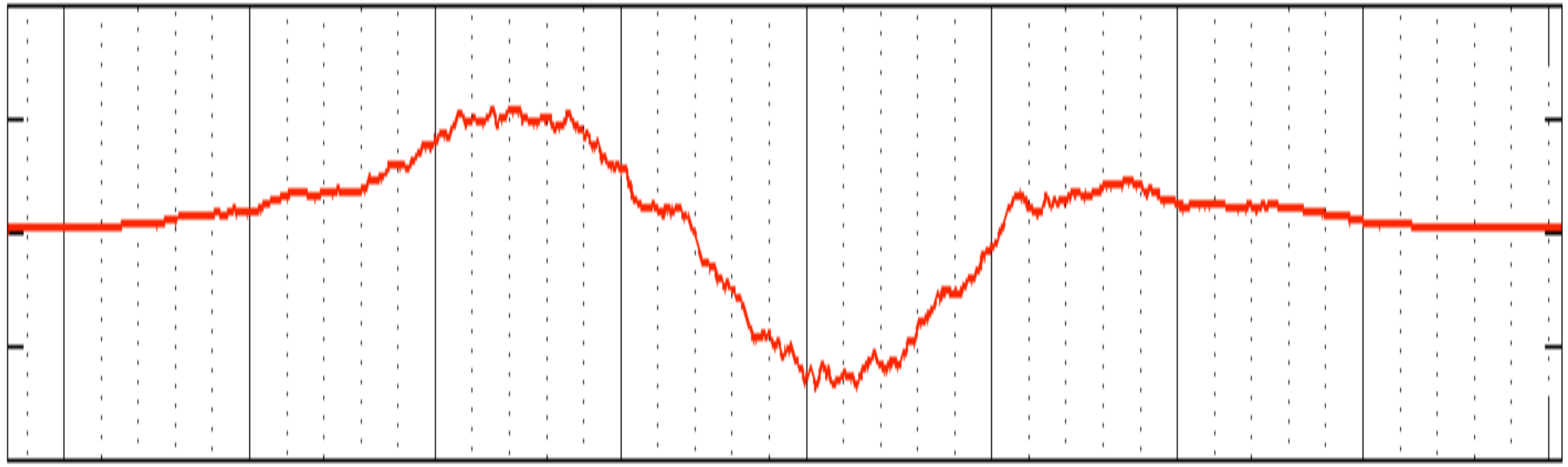




From persistence to the Euler curve



$$\chi_u = \sum_{k=0}^{\dim} (-1)^k \beta_k(u)$$



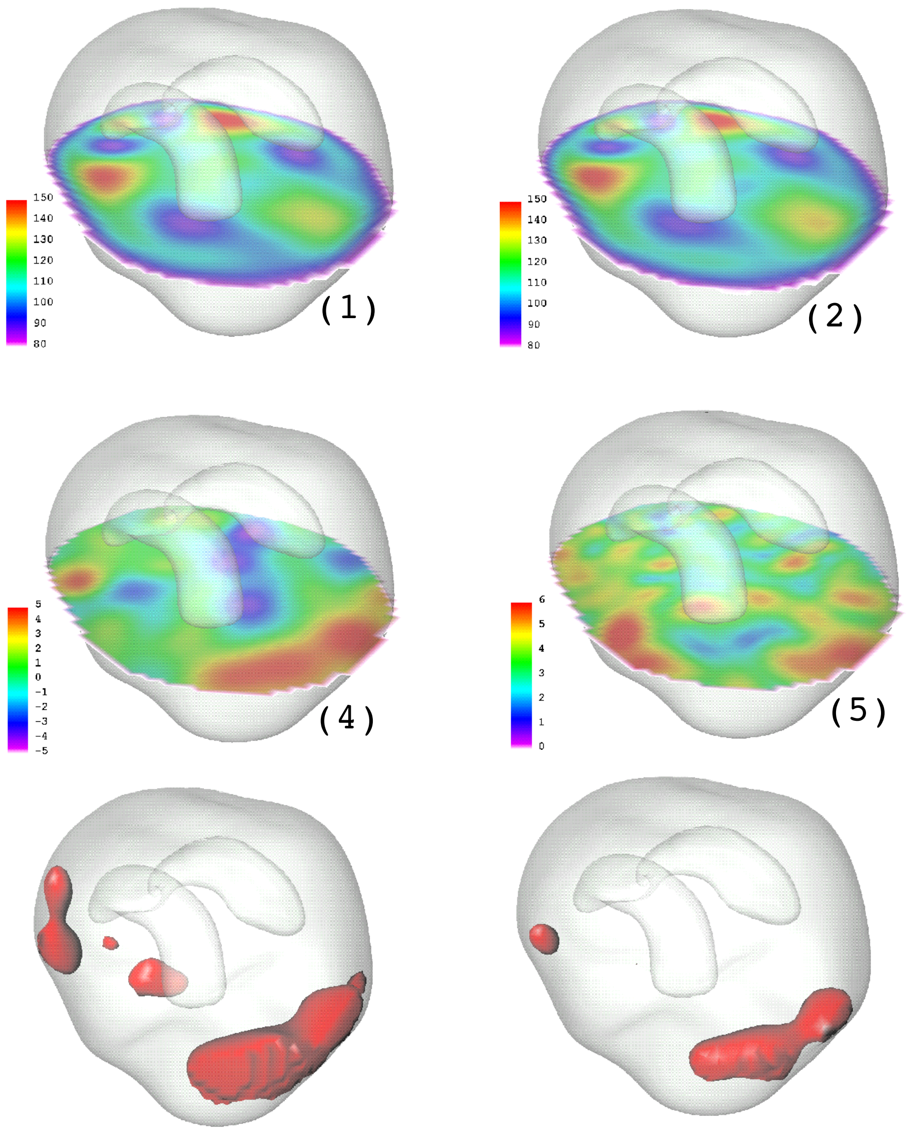
Keith Worsley (1951-2009)

Because it leads to the idea of

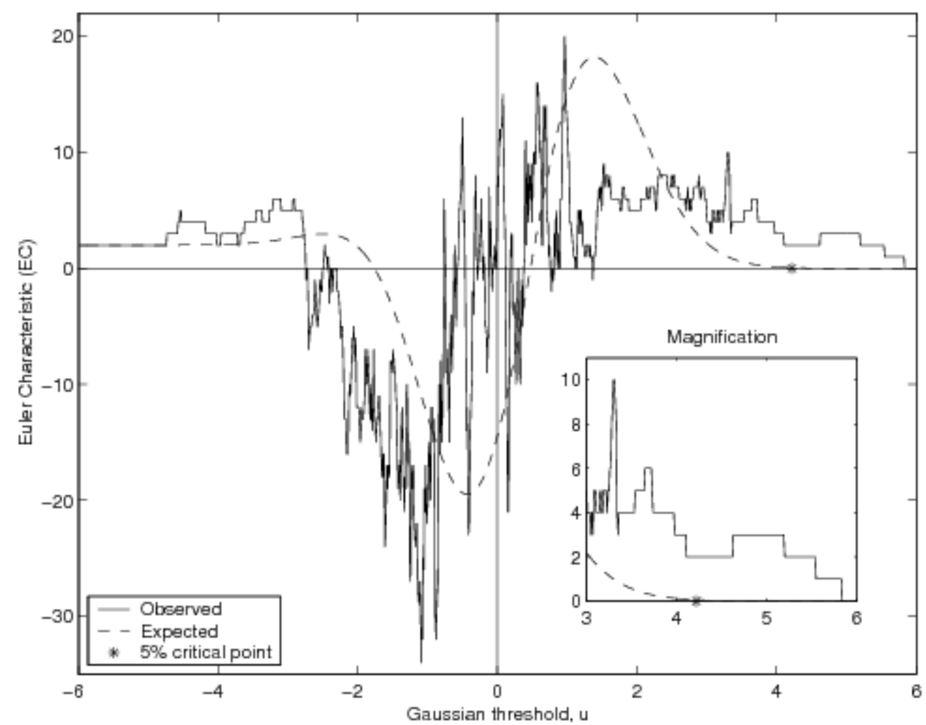
TOPOLOGICAL INFERENCE

Worsley, Friston et al, late 1990s

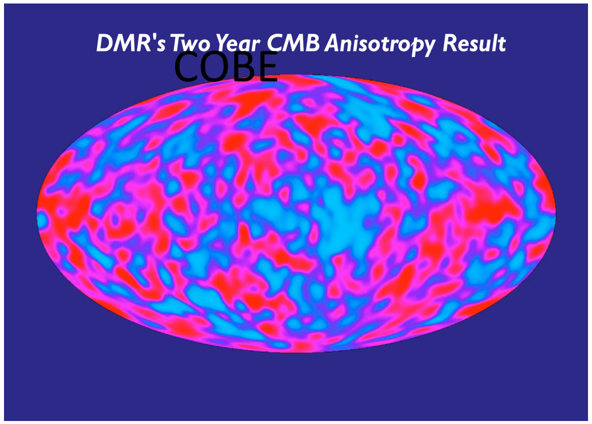
Brain activity via the EC curve

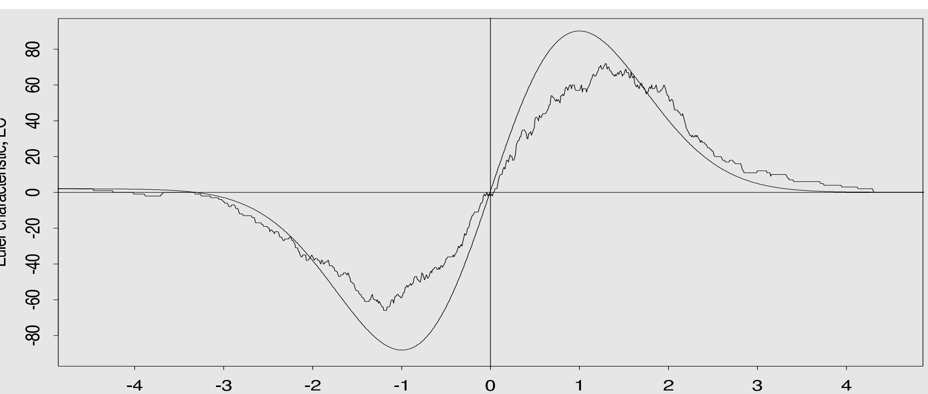


(1) $B = \text{Before}$
(2) $A = \text{After}$
(3) $D = A - B$
(4) σ_D^2
(5) D/σ_D

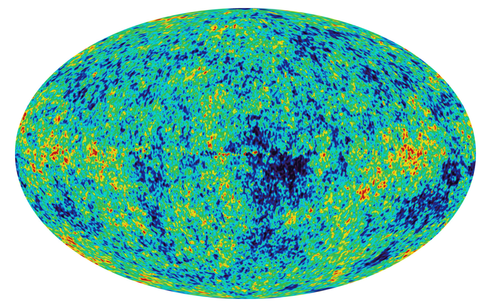


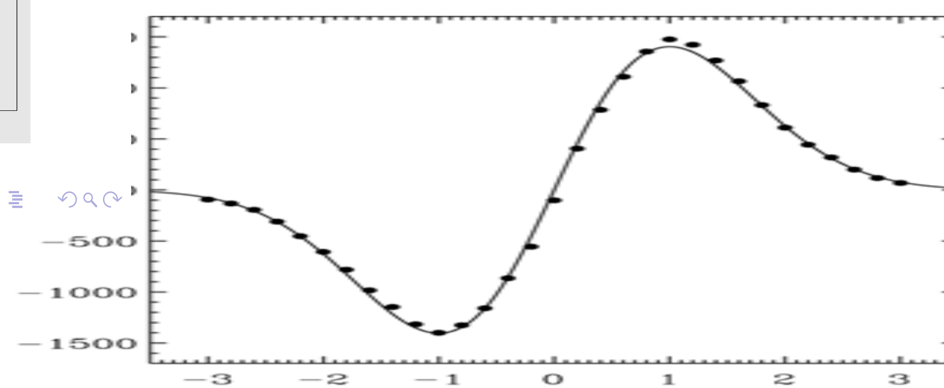
CMB and Euler characteristic curves





WMAP

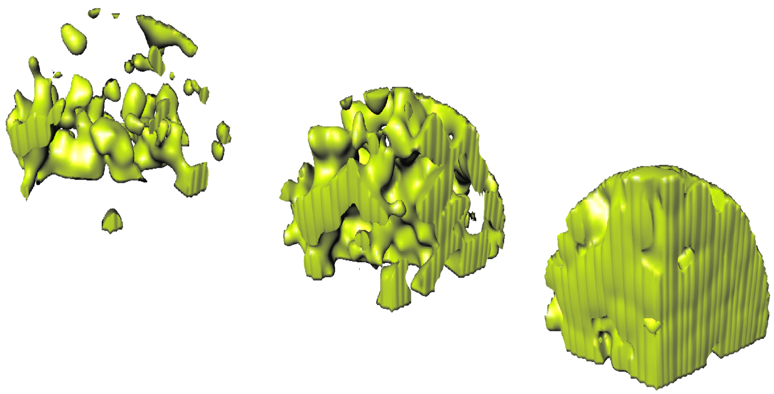


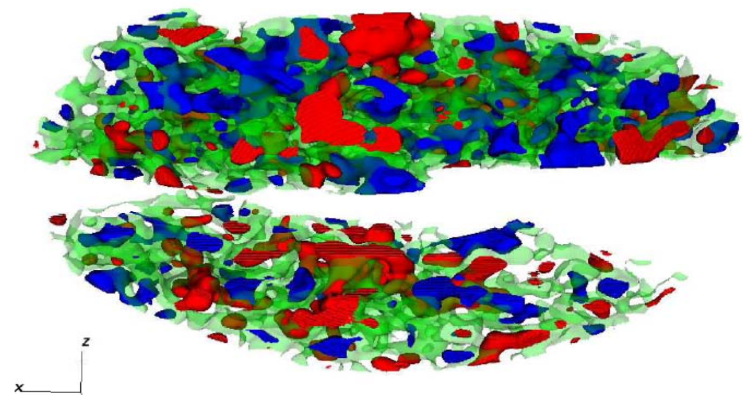


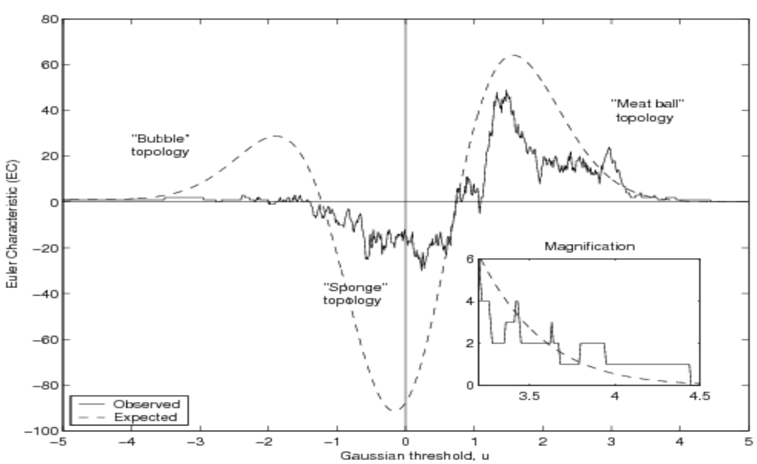
CfA, SDSS and Euler curves

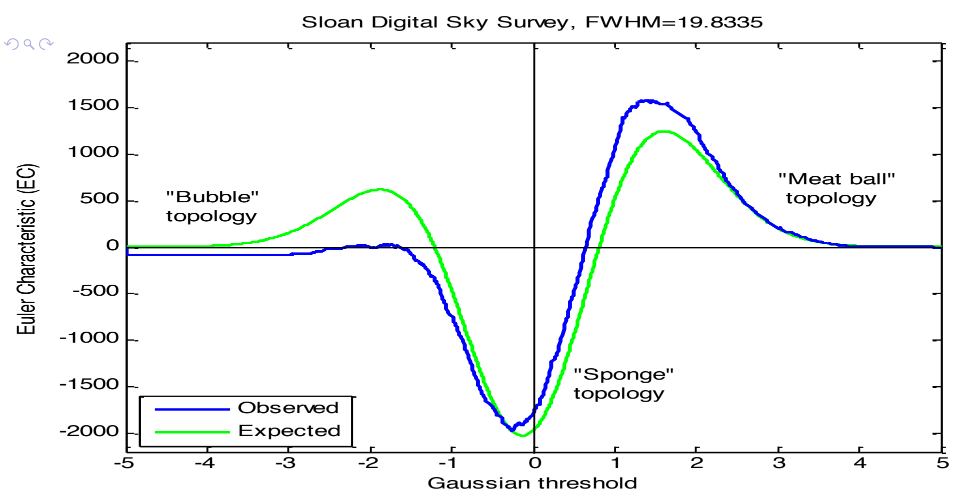
Center for Astrophysics (CfA) survey

10,506 galaxies in the cone-shaped survey region, which extends out to 135 megaparsecs in the northern hemisphere, with the earth at the apex of the cone.

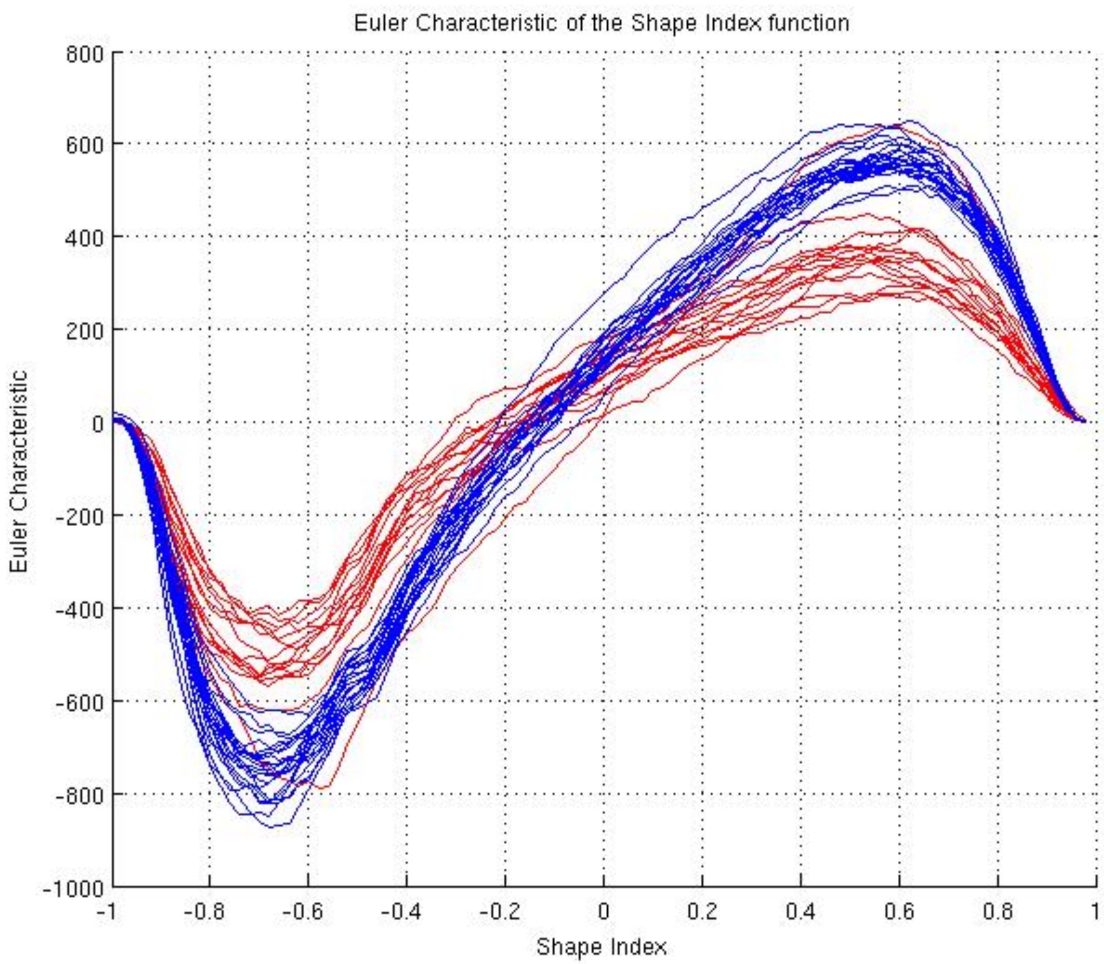








Qesem and Nahal Zihor

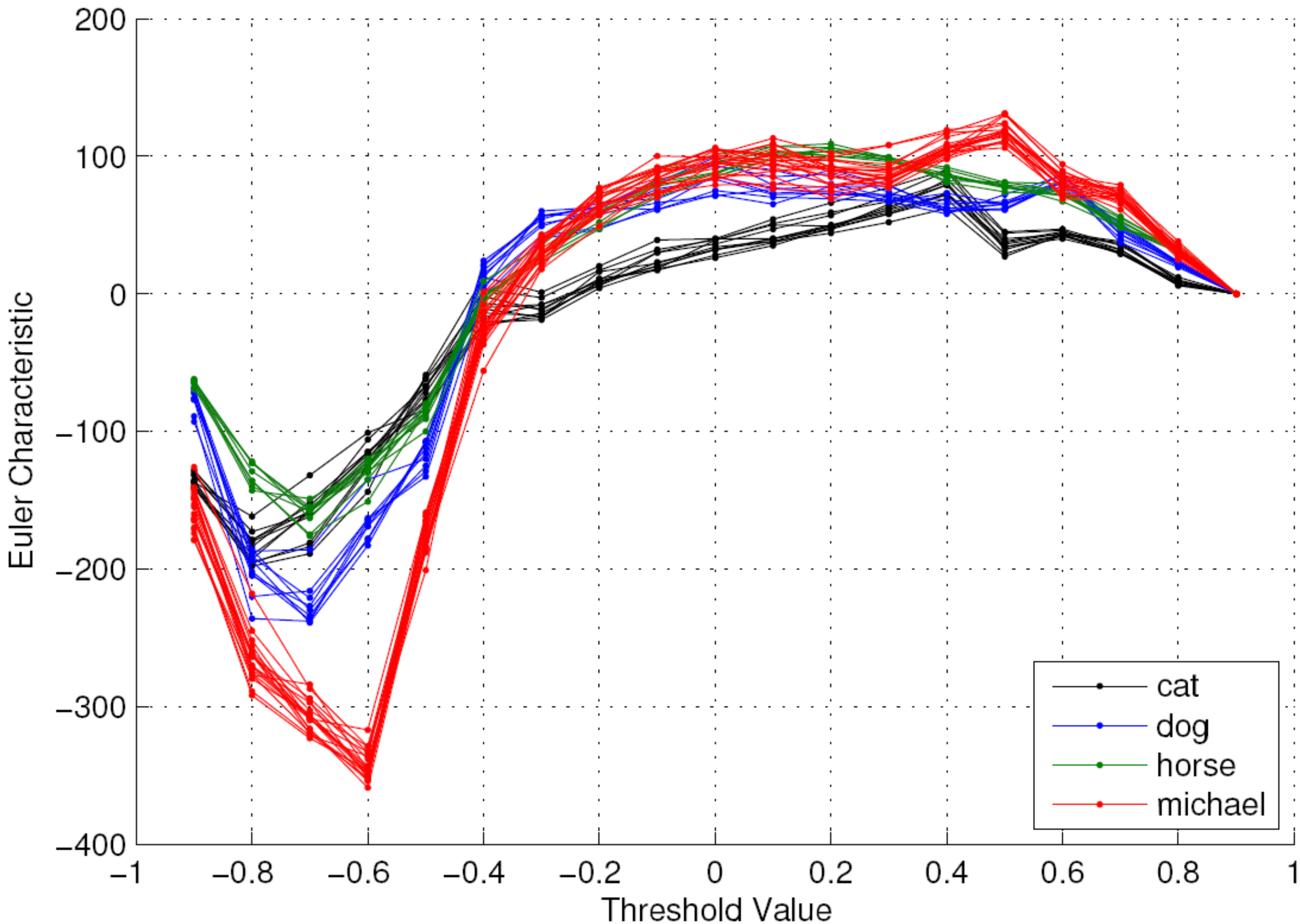


SVM classification rates:

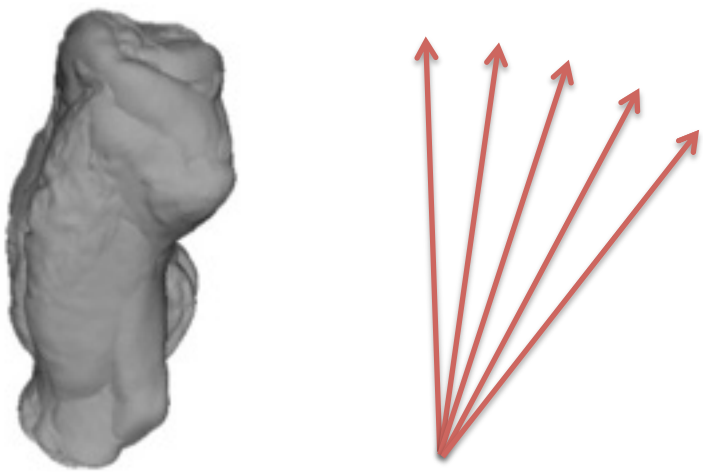
Curvature 93.7%
Shape index: 96.8%
Distance from center 52.2%
Normal distance: 70%

Eitan Richardson
Mike Werman

Cats, dogs, horses, and Michael



Euler Characteristic ---- Transform



Theorem (Turner, Mukherjee, Boyer)

Set $n = 2$ or 3. Let $\mathcal{M}$ be the space of (finite) piecewise linear simplicial complexes in R^n. Then the Euler characteristic transform

$$\begin{aligned} \text{ECT} : \mathcal{M} &\rightarrow C(S^{n-1}, \mathcal{D}) \\ M &\mapsto \text{ECT}(M) : v \mapsto X(M, v)) \end{aligned}$$

is injective.

Euler characteristics in random systems

The wonderful world of geotopology

SIMPLICIAL TOPOLOGY

Simplices, complexes,
cycles, numbers of simplices,
Betti numbers

$$\sum_k (-1)^k \#\{k\text{-dimensional simplices}\}$$

INTEGRAL GEOMETRY

Convexity, convex ring
kinematic formulae
Minkowski functionals

$$\mathcal{M}_k(M) = c_{dk} \int_{\mathrm{Graff}(d,d-k)} \chi(M \cap V)\, d\mu^d_{d-k}(V)$$

$$\chi$$

$$\sum_k (-1)^k \beta_k$$

$$\sum_k (-1)^k \#\{\text{critical points of index } k\}$$

$$\int_M \mathrm{Tr}(R^{m/2}) \mathrm{Vol}_g$$

ALGEBRAIC TOPOLOGY

Homology, homotopy,
dimensions of groups,
Betti numbers, persistence

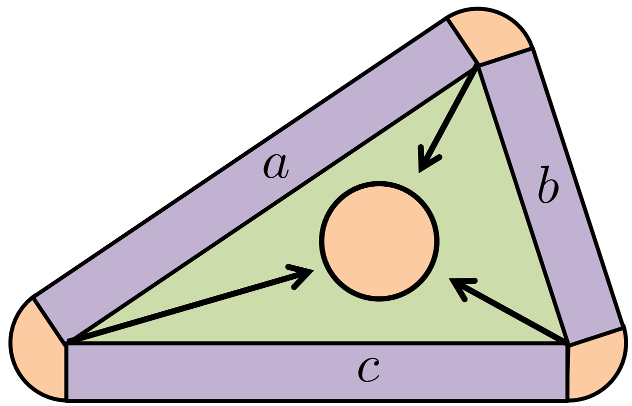
DIFFERENTIAL TOPOLOGY

Curvature, forms, Betti numbers,
Morse theory, integration,
Lipschitz-Killing curvatures

Some geometry - The tube around a triangle

$M = \text{Triangle}$

$\text{Tube}(M, \rho)$



$$\text{Vol}\,(\text{Tube}(M,\rho)) = \text{Vol}(M) + (a+b+c)\rho + \pi\rho^2$$

$$\Rightarrow \mathcal{L}_0(M) = 1 = \chi(M)$$

$$\mathcal{L}_1(M) = (a+b+c)/2$$

$$\mathcal{L}_2(M) = \text{Vol}(M)$$

Lipschitz-Killing curvatures/Minkowski F'nls

Theorem (Weyl's Tube Formula)

Let M be a 'nice' stratified space embedded in $\mathbb{R}^l$, then there exist numbers $\mathcal{L}_i(M)$ such that

$$\mathrm{Vol}\left(\mathrm{Tube}(M,\rho)\right) = \sum_{i=0}^{\dim M} \rho^{l-i}\omega_{l-i}\mathcal{L}_i(M)$$

Where:

- $\mathrm{Tube}(M,\rho) = \left\{t \in \mathbb{R}^l : \mathrm{dist}(t,M) \leq \rho\right\}$
- ω_n - volume of n-dimensional unit ball

The coefficients $\mathcal{L}_i(M)$ are called the **Lipschitz-Killing curvatures** of M

A 1-slide course on Gaussian random fields

- On a topological space *M*

  (maybe a Riemannian manifold)

- choose a set $\{\varphi_k\}$ of dense functions on *M*

  (e.g. Eigenfunctions of the Laplacian)

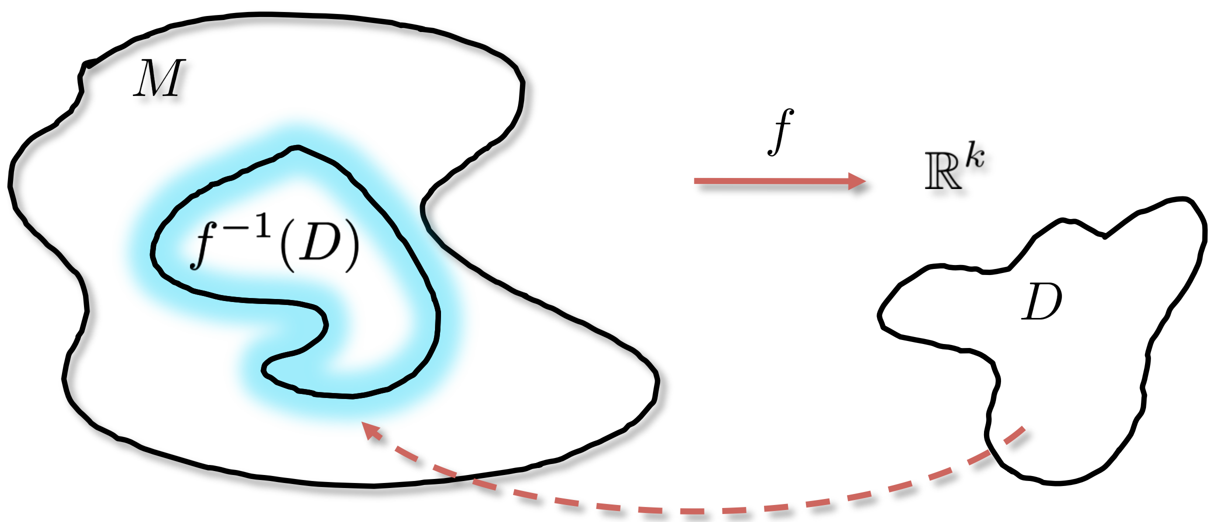
- Then take independent Gaussian random variables, $\xi_1, \xi_2, \ldots$

$$\mathcal{P}\left\{(\xi_{j_1}, \ldots, \xi_{j_k}) \;\in\; A\right\} \;=\; \frac{1}{(2\pi)^{k/2}} \int_A e^{-\|x\|^2/2}\, dx$$

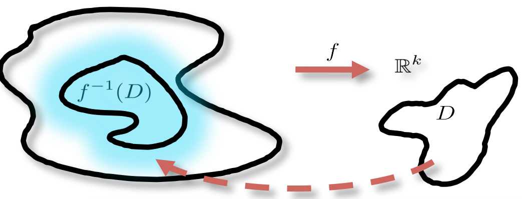
- and define a Gaussian random field:

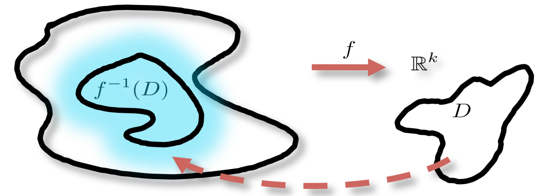
$$f(p) \;= \sum_j a_j \xi_j \varphi_j(p)$$

Gaussian and Gaussian-related excursions



The Gaussian Kinematic Formula





Theorem

Let $M \subset \mathbb{R}^N$ and $D \subset \mathbb{R}^d$ be nice stratified spaces. Let $f = (f_1, \ldots, f_d) : M \to \mathbb{R}^d$ be a d-dimensional Gaussian field, with *iid* components all having zero mean, unit variance and a nice covariance function. Then,

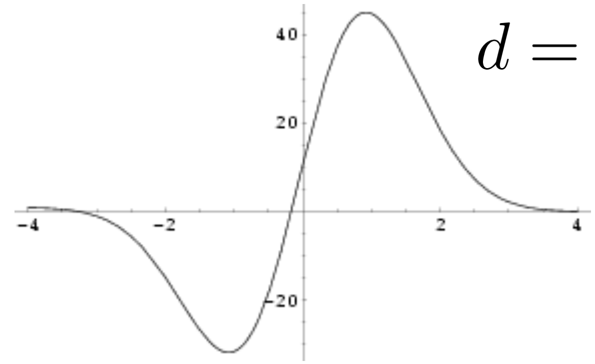
$$\mathbb{E}\left\{\chi(f^{-1}(D))\right\} = \sum_{j=0}^{\dim M} (2\pi)^{-j/2} \mathcal{L}_j(M) \mathcal{M}_j(D)$$

$$\mathbb{E}\left\{\mathcal{L}_i(f^{-1}(D))\right\} = \sum_{j=0}^{\dim M - i} \begin{bmatrix} i+j \\ j \end{bmatrix} (2\pi)^{-j/2} \mathcal{L}_{i+j}(M) \mathcal{M}_j(D)$$

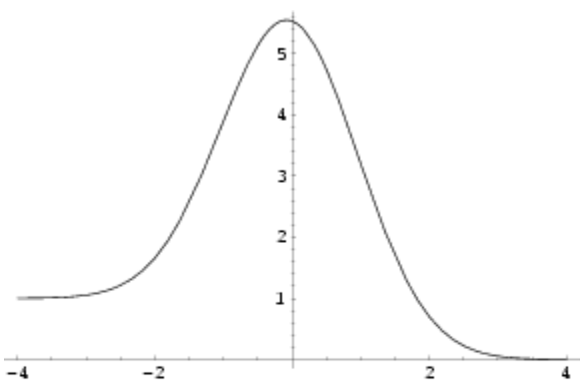
Mean Euler characteristics – Gaussian case

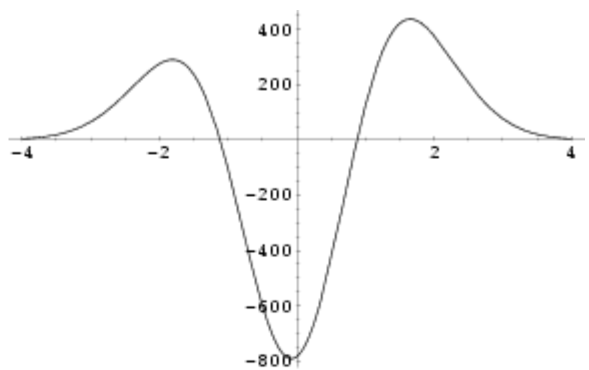
$$E\left\{\chi\left([0,T]^2 \cap f^{-1}[u,\infty)\right)\right\}$$

$$= \left[\frac{T^2\lambda}{(2\pi)^{3/2}}u + \frac{2T\lambda^{1/2}}{2\pi}\right]e^{-u^2/2} + \Psi(u)$$



$d = 2$





$d = 1$

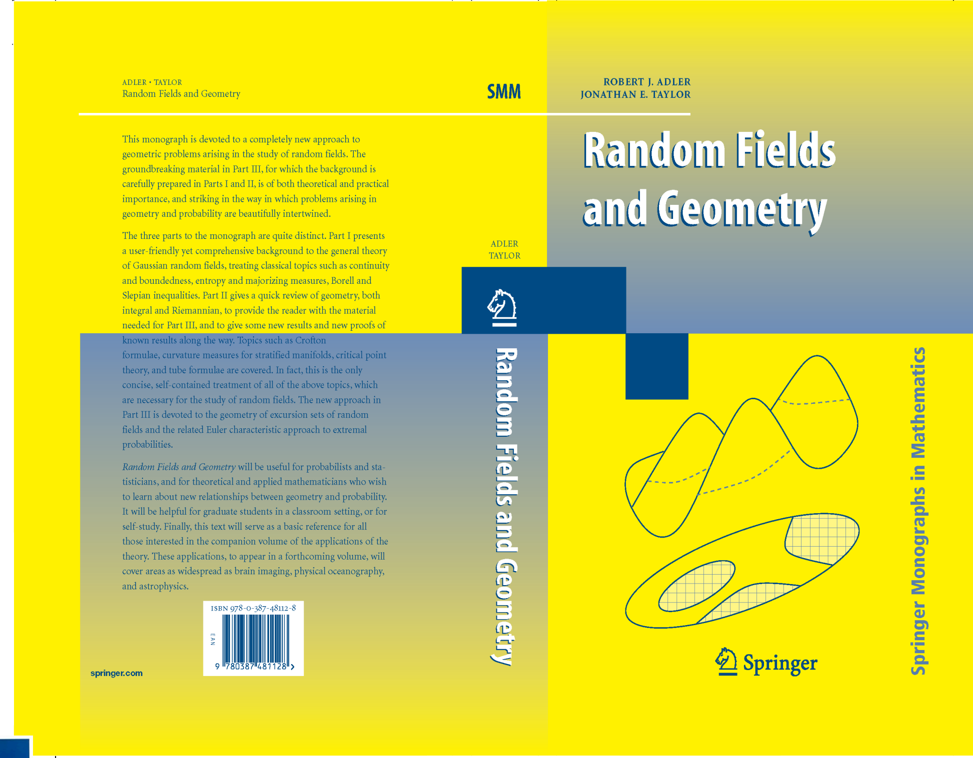
$d = 3$

The Euler characteristic heuristic

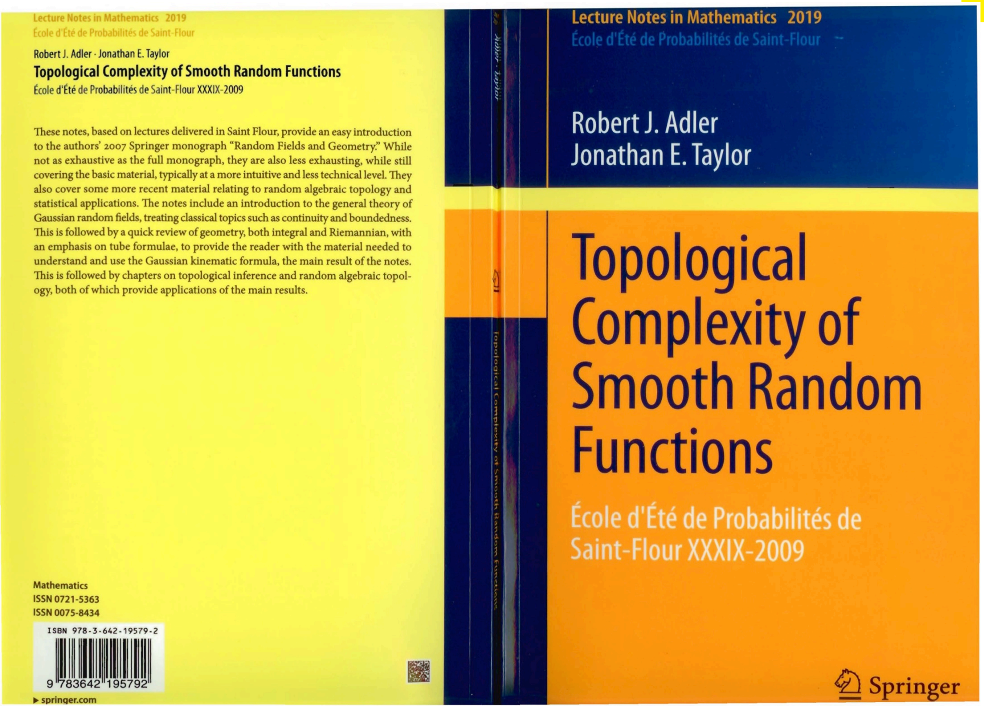
$$\mathbb{P}\left\{ \sup_M f \ \geq u \right\} \sim \mathbb{E}\{\chi(A_u(f,M))\}$$

$$\sim u^{\dim M - 1} e^{-u^2/2}$$

$$\liminf_{u\to\infty} u^{-2} \log|\mathbb{P} - \mathbb{E}| \geq \frac{1}{2} + \frac{1}{2\sigma^2(f)}$$



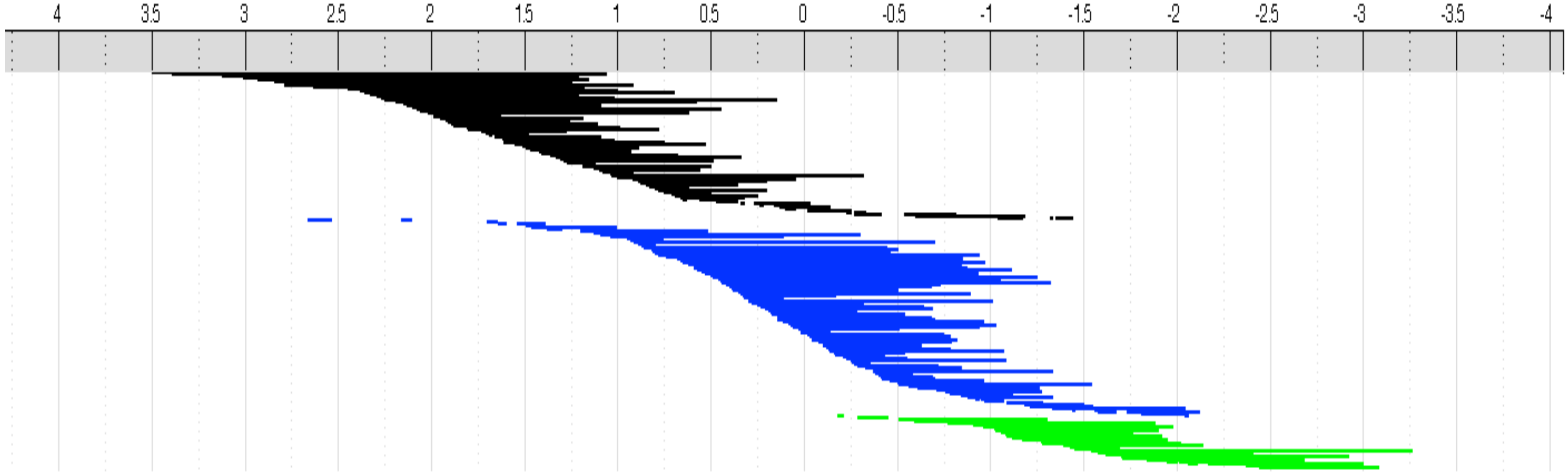
ADLER · TAYLOR
Random Fields and Geometry
SMM
This monograph is devoted to a completely new approach to geometric problems arising in the study of random fields. The groundbreaking material in Part III, for which the background is carefully prepared in Parts I and II, is of both theoretical and practical importance, and striking in the way in which problems arising in geometry and probability are beautifully intertwined.
The three parts to the monograph are quite distinct. Part I presents a user-friendly yet comprehensive background to the general theory of Gaussian random fields, treating classical topics such as continuity and boundedness, entropy and majorizing measures, Borell and Slepian inequalities. Part II gives a quick review of geometry, both integral and Riemannian, to provide the reader with the material needed for Part III, and to give some new results and new proofs of known results along the way. Topics such as Crofton formulae, curvature measures for stratified manifolds, critical point theory, and tube formulae are covered. In fact, this is the only concise, self-contained treatment of all of the above topics, which are necessary for the study of random fields. The new approach in Part III is devoted to the geometry of excursion sets of random fields and the related Euler characteristic approach to extremal probabilities.
Random Fields and Geometry will be useful for probabilists and statisticians, and for theoretical and applied mathematicians who wish to learn about new relationships between geometry and probability. It will be helpful for graduate students in a classroom setting, or for self-study. Finally, this text will serve as a basic reference for all those interested in the companion volume of the applications of the theory. These applications, to appear in a forthcoming volume, will cover areas as widespread as brain imaging, physical oceanography, and astrophysics.
ISBN 978-0-387-48112-8
springer.com
ADLER
TAYLOR
Random Fields and Geometry
ROBERT J. ADLER
JONATHAN E. TAYLOR
Random Fields and Geometry
Springer
Springer Monographs in Mathematics



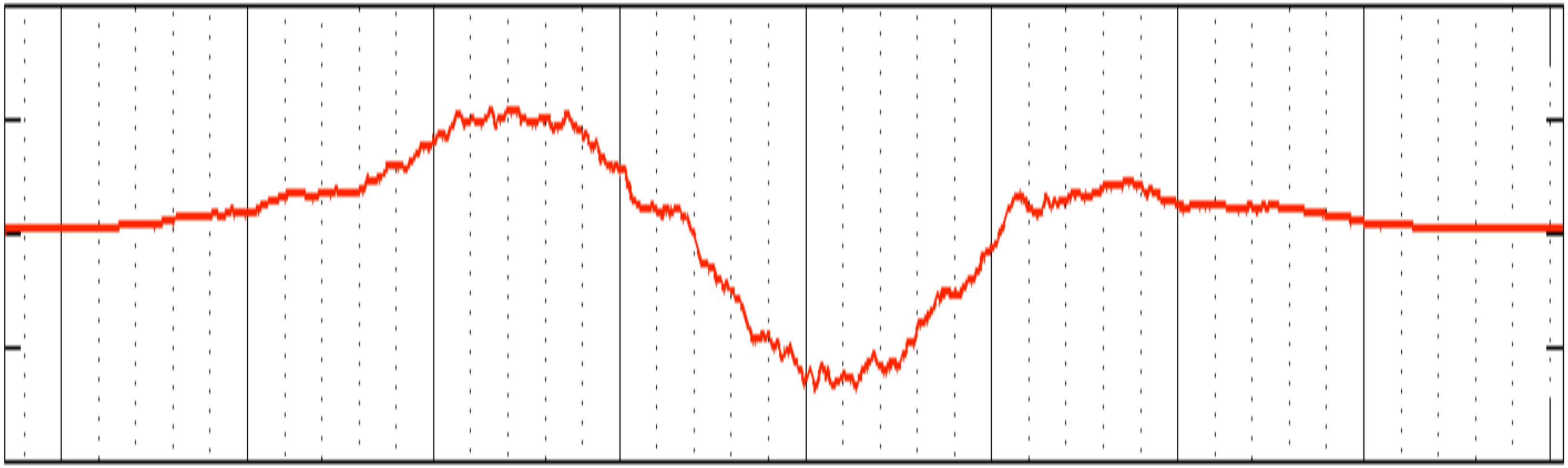
Lecture Notes in Mathematics 2019
École d'Été de Probabilités de Saint-Flour
Robert J. Adler · Jonathan E. Taylor
Topological Complexity of Smooth Random Functions
École d'Été de Probabilités de Saint-Flour XXXIX-2009
These notes, based on lectures delivered in Saint Flour, provide an easy introduction to the authors' 2007 Springer monograph "Random Fields and Geometry." While not as exhaustive as the full monograph, they are also less exhausting, while still covering the basic material, typically at a more intuitive and less technical level. They also cover some more recent material relating to random algebraic topology and statistical applications. The notes include an introduction to the general theory of Gaussian random fields, treating classical topics such as continuity and boundedness. This is followed by a quick review of geometry, both integral and Riemannian, with an emphasis on tube formulae, to provide the reader with the material needed to understand and use the Gaussian kinematic formula, the main result of the notes. This is followed by chapters on topological inference and random algebraic topology, both of which provide applications of the main results.
Mathematics
ISSN 0721-5363
ISSN 0075-8434
ISBN 978-3-642-19579-2
springer.com
Lecture Notes in Mathematics 2019
École d'Été de Probabilités de Saint-Flour
Robert J. Adler
Jonathan E. Taylor
Topological Complexity of Smooth Random Functions
École d'Été de Probabilités de Saint-Flour XXXIX-2009
Springer

Euler characteristics and Betti numbers

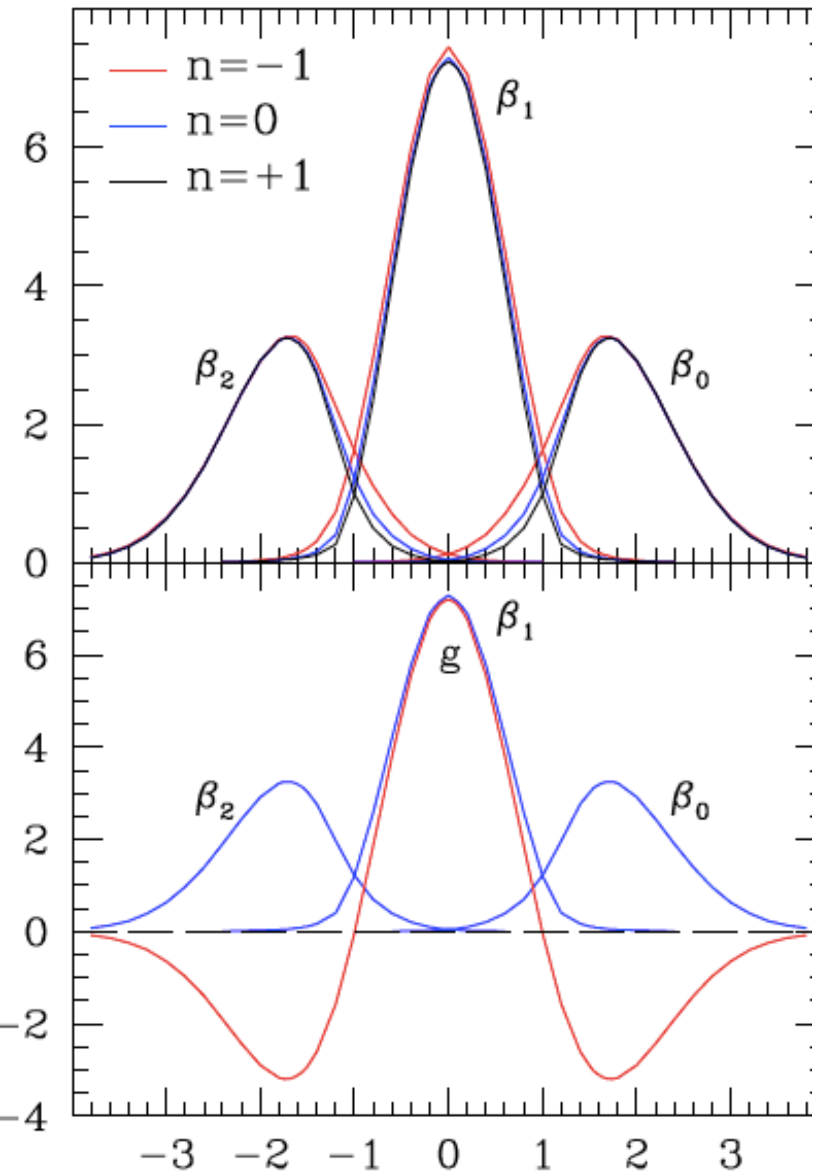
Persistence diagrams and EC curves



$$\chi_u = \sum_{k=0}^{\dim} (-1)^k \beta_k(u)$$



Betti numbers and EC curves for functions



BETTI NUMBERS OF GAUSSIAN FIELDS

Changbom Park[1], Pratyush Pranav[2], Pravabati Chingangbam[3], Rien van de Weygaert[2], Bernard Jones[2], Gert Vegter[4], Inkang Kim[5], Johan Hidding[2], and Wojciech A. Hellwing[6,7],

[1] School of Physics, Korea Institute for Advanced Study, Seoul 130-722, Korea

E-mail : cbp@kias.re.kr

[2] Kapteyn Astron. Inst., Univ. of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands

[3] Indian Institute of Astrophysics, Koramangala II Block, Bangalore 560 034, India

E-mail : pravabati@gmail.com

[4] Johann Bernoulli Inst. for Mathematics and Computer Science, Univ. of Groningen, P.O. Box 407, 9700 AK Groningen, The Netherlands

[5] School of Mathematics, Korea Institute for Advanced Study, Seoul 130-722, Korea

[6] Institute of Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, United Kingdom

[7] Interdisciplinary Centre for Mathematical and Computational Modeling (ICM), University of Warsaw, ul. Pawinskiego 5a, Warsaw, Poland

From point clouds to simplicial complexes

Definition 2.1 (The Čech complex). Let $\mathcal{P} = \{x_1, x_2, \ldots\}$ be a collection of points in a metric space X. Construct an abstract simplicial complex $C(\mathcal{P}, \varepsilon)$ in the following way:

1. The 0-simplices are the points in $\mathcal{P}$,
2. An n-simplex $[x_{i_0}, \ldots, x_{i_n}]$ is in $C(\mathcal{P}, \varepsilon)$ if $\bigcap_{k=0}^{n} B_\varepsilon(x_{i_k}) \neq \emptyset$,

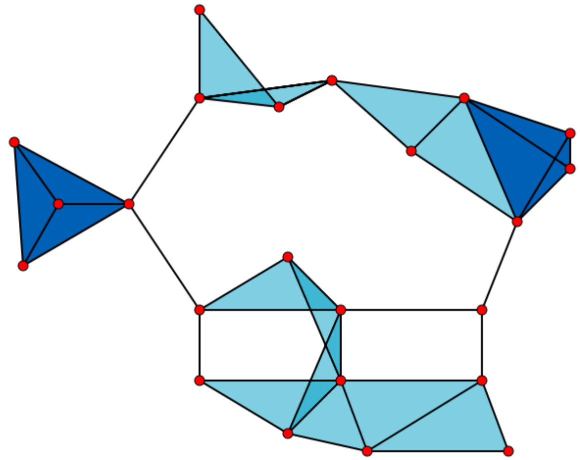
where $B_\varepsilon(x)$ is the ball of radius ε around x. The complex $C(\mathcal{P}, \varepsilon)$ is called the Čech complex attached to $\mathcal{P}$ and ε.

Definition 2.2 (The Vietoris-Rips complex). Let $\mathcal{P} = \{x_1, x_2, \ldots\}$ a collection of points in a metric space X. Construct an abstract simplicial complex $R(\mathcal{P}, \varepsilon)$ in the following way:

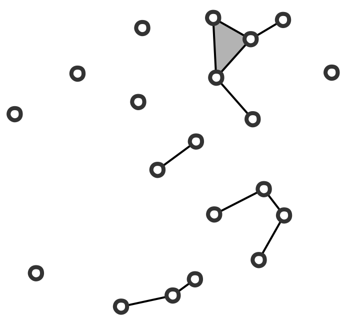
1. The 0-simplices are the points in $\mathcal{P}$.
2. An n-simplex $[x_{i_0}, \ldots, x_{i_n}]$ is in $R(\mathcal{P}, \varepsilon)$ if $B_\varepsilon(x_{i_k}) \cap B_\varepsilon(x_{i_m}) \neq \emptyset$ for every $0 \leq k < m \leq n$.

The complex $R(\mathcal{P}, \varepsilon)$ is called the Rips complex attached to $\mathcal{P}$ and ε.

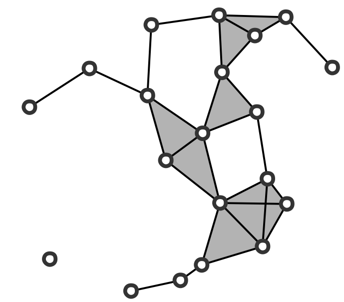
A Vietoris–Rips complex of a set of 23 points in the plane.
This complex has sets of up to four points:
1: the points themselves (shown as red circles),
2: pairs of points (black edges),
3: triples of points (pale blue triangles),
4: quadruples of points (dark blue tetrahedrons).



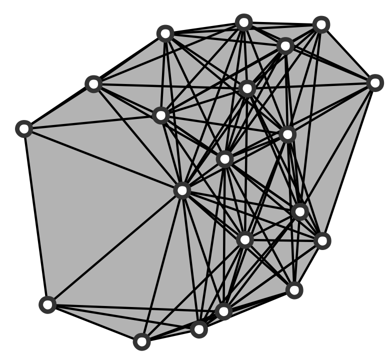
Phase transitions/persistence for complexes



Dust, or subcritical phase



Thermodynamic, or critical, or percolation phase



Connectivity, or supercritical phase

Cech complex on 1000 points in $[0,1]^3$

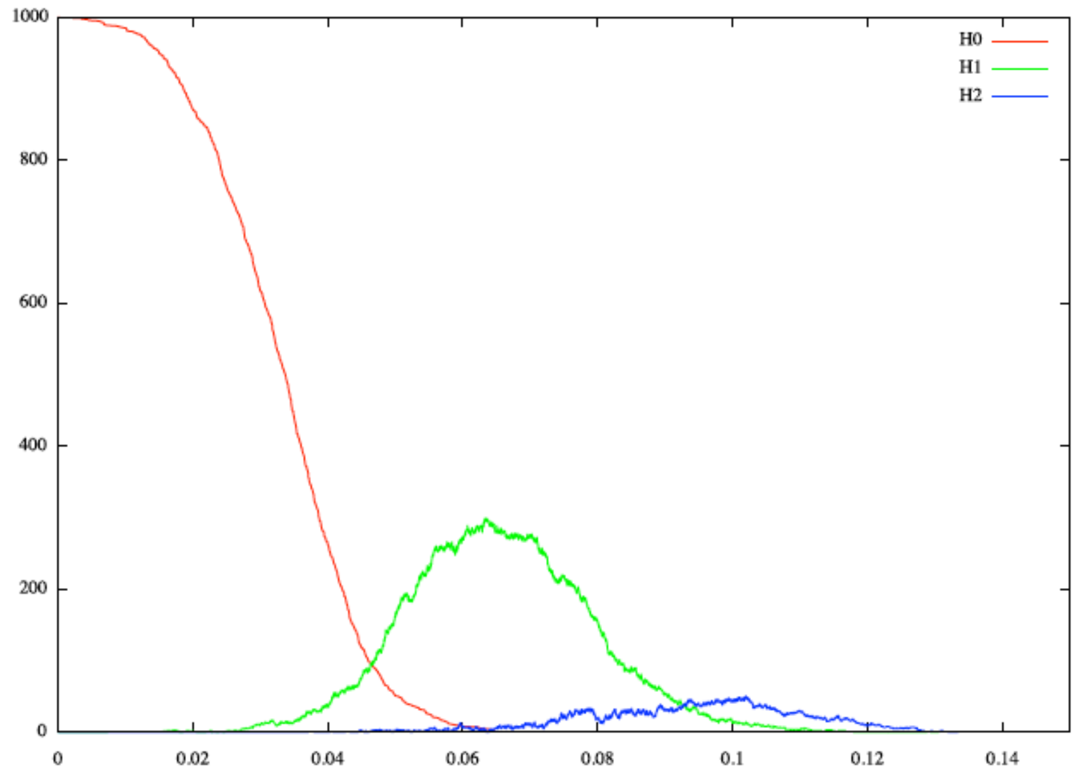
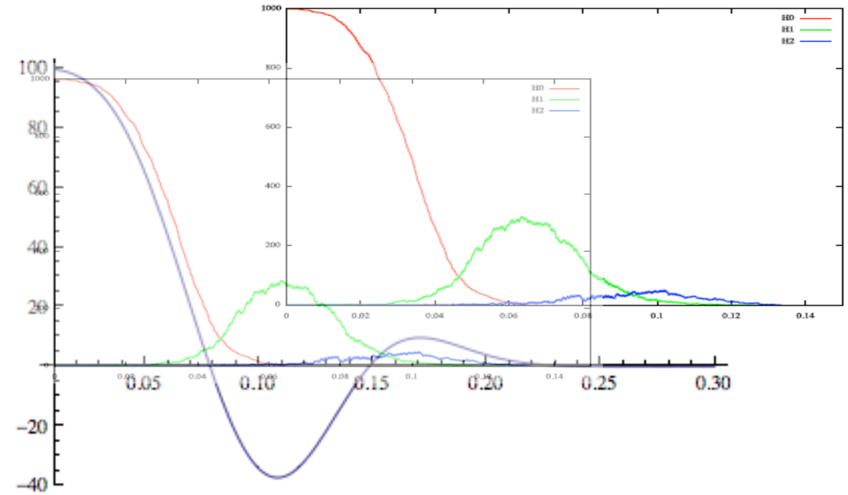
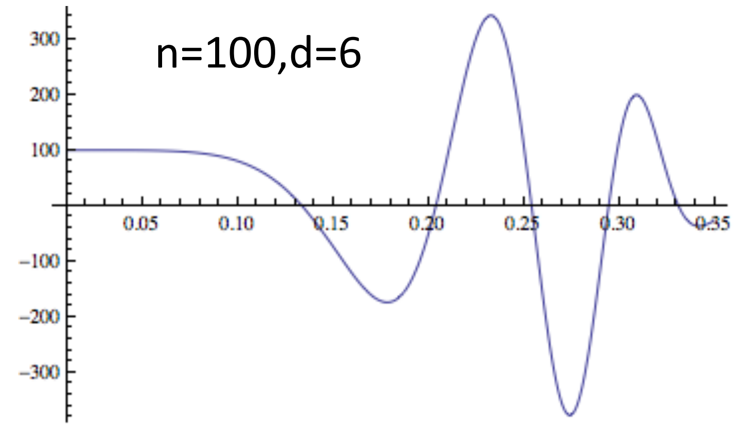


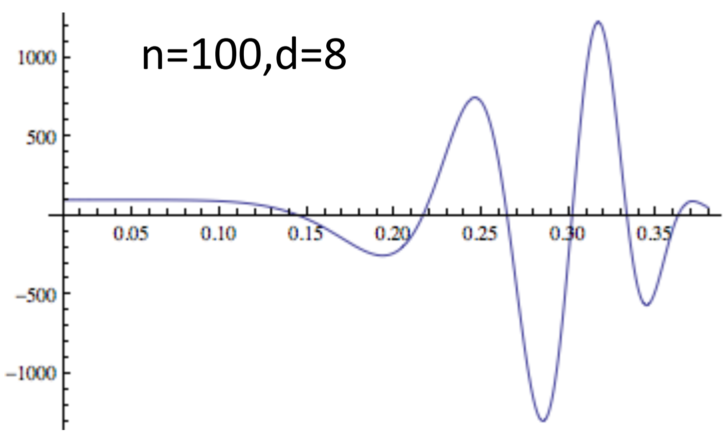
FIGURE 2. The Betti numbers of a random Čech complex on $n = 1000$ points versus the radius r. Computation and image courtesy of Dmitriy Morozov and his Dionysus library [39].

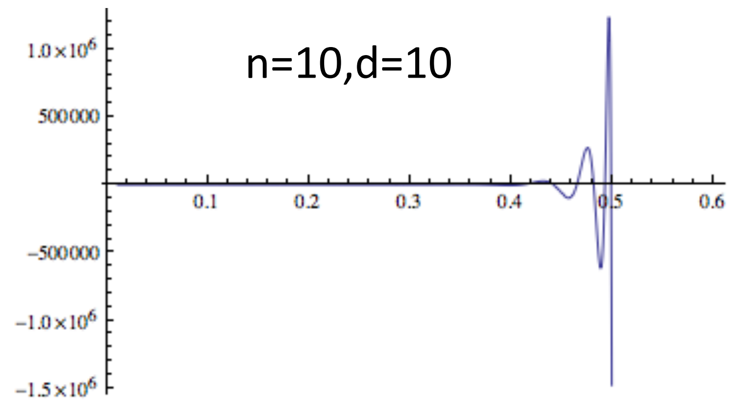
Matt K and Omer B

E(EC) for Cech, n random points in $[0,1]^d$





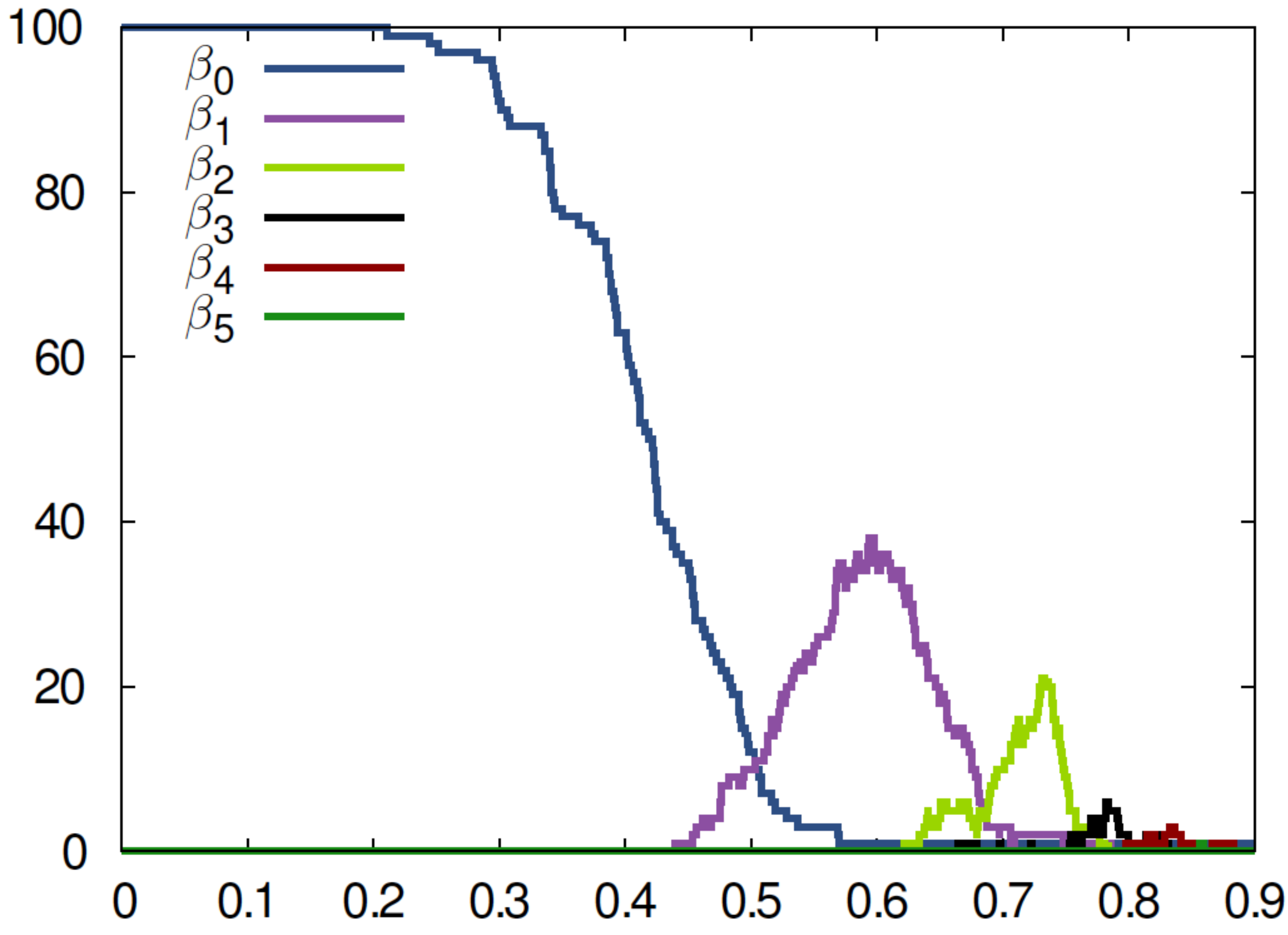




$$\mathbb{E}\{\chi\} = \sum_{k=0}^{n} (-1)^{k+1} \binom{n}{k} k^d (2r)^{d(k-1)}$$

Decreusefond, Ferraz, Randriam, Vergne, by counting simplices

Rips complex on 100 points in $[0,1]^6$



Matt K

Boolean models

Components of a Boolean model:

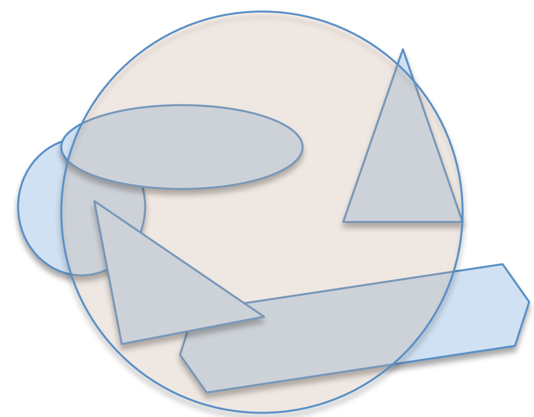
1: Poisson point process on *R^d,* *{x_i, i=1,2,..}*

2: Collection of (independent) random sets *{B_i, i=1,2,..}* in *R^d*

3: Or, non-random balls of fixed radius , *r.*

$$B = \bigcup_i \{x_i \oplus B_i\}$$

$$B_n = B \cap \text{Ball of radius } n$$



THEOREM: (Schneider and Weil)

Remember the nerve theorem

$$\lim_{n\to\infty} \frac{\mathbb{E}\{\mathcal{L}_k(B_n)\}}{n^d} = F_k\left(\lambda, \mathbb{E}\{\mathcal{L}_j(B_1)\}, \; j = 0, \ldots, d\right).$$

Theory: MK, EM, OB, YD, inter alia

Dust, subcritical

$r_n = o\left(n^{-1/d}\right)$

$$\lim_{n\to\infty} n^{-1}\mathbb{E}\left\{N_k(r_n)\right\} = 0 \quad ; \quad \mathbb{E}\left\{N_k(r_n)\right\} = \theta(n^{k+1}r_n^{dk})$$

$$\lim_{n\to\infty} n^{-1}\mathbb{E}\left\{\beta_k(r_n)\right\} = 0 \quad ; \quad \mathbb{E}\left\{\beta_k(r_n)\right\} = \theta(n^{k+2}r_n^{d(k+1)})$$

$$\lim_{n\to\infty} n^{-1}\mathbb{E}\left\{\chi_n(r_n)\right\} = 1$$

Critical, percolative, thermodynamic

$r_n = cn^{-1/d}$

$$\lim_{n\to\infty} n^{-1}\mathbb{E}\left\{N_k(r_n)\right\} = \frac{c^{dk}}{(k+1)!}C_k$$

$$\lim_{n\to\infty} n^{-1}\mathbb{E}\left\{\beta_k(r_n)\right\} = const > 0$$

$$\lim_{n\to\infty} n^{-1}\mathbb{E}\left\{\chi_n(r_n)\right\} = 1 + \sum_{k=1}^{d}\frac{(-c^d)^k}{(k+1)!}C_k$$

Dense, supercrictical

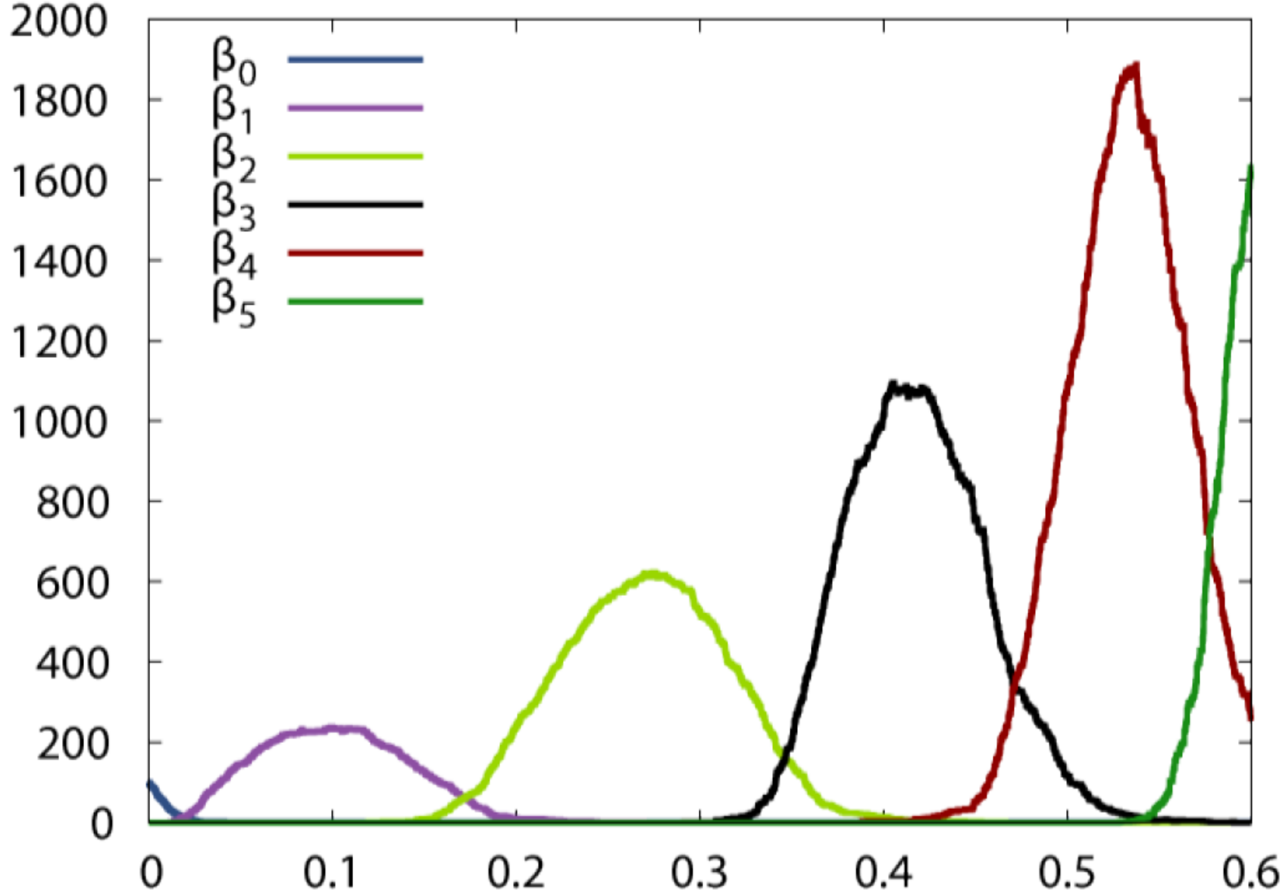
$r_n = \omega\left(n^{-1/d}\right)$

$$\lim_{n\to\infty} n^{-1}\mathbb{E}\left\{N_k(r_n)\right\} = \lim_{n\to\infty} n^{-1}\mathbb{E}\left\{N_k\right\}$$

$$\lim_{n\to\infty} n^{-1}\mathbb{E}\left\{\chi_n(r_n)\right\} = 0$$

Complexity of random smooth functions of many variables
[Antonio Auffinger, Gerard Ben Arous]

Erdos-Renyi flag complex on 100 vertices



Matt K, Elizabeth M, Afra Z, Gugan T.

The Euler Characteristic of a Barcode

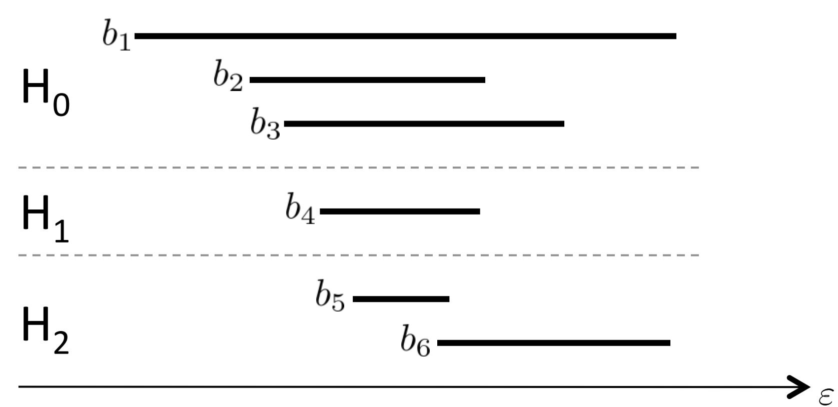
Definition (EC of a barcode)

The *Euler characteristic* of a barcode B_* with a finite number of bars and no bars of infinite length is

$$\chi(B_*) = \sum_{b_j \in B} (-1)^{\mu(b_j)} \ell(b_j).$$

Where:

- $\mu(b_j)$ - the degree of b_j
- $\ell(b_j)$ - the length of b_j



$$\chi(B_*) = l(b_1) + l(b_2) + l(b_3) - l(b_4) + l(b_5) + l(b_6)$$

Euler Integral and Persistent Homology

Proposition

Let $h : X \to \mathbb{R}$ be tame. Then

$$\chi(B_*(h, h_{\max})) = h_{\max}\,\chi(X) - \int_X h\lceil d\chi\rceil,$$

and in general

$$\chi(B_*(h, a)) = a\,\chi(X) - \int_X (G_a \circ h)\lceil d\chi\rceil.$$

Where:

- $B_*(h, a)$ is the persistent homology of h in the range $(-\infty, a]$
- $G_a(x) = \min(x, a)$

Expected EC of the Persistent Homology

Theorem (Bobrowski + Borman)

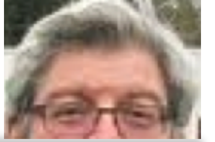
Let $f : M \to \mathbb{R}^k$ be a Gaussian random field satisfying the GKF conditions, $G : \mathbb{R}^k \to \mathbb{R}$ continuous and piecewise C^2, and $g = G \circ f$. Then

$$\mathbb{E}\left\{\chi(B_*(g, g_{\max}))\right\} = \chi(M)\left(\mathbb{E}\left\{g_{\max}\right\} - \mathbb{E}\left\{g\right\}\right) + \sum_{j=1}^{N} (2\pi)^{-j/2} \mathcal{L}_j(M) \int \mathcal{M}_j(D_u)\,du.$$

- For a real-valued Gaussian field - $\mathbb{E}\left\{\chi(B_*(f, f_{\max}))\right\} = \mathbb{E}\left\{f_{\max}\right\}\chi(M) + \frac{\mathcal{L}_1(M)}{\sqrt{2\pi}}$.
- Problem: $\mathbb{E}\left\{g_{\max}\right\} = ?$
- Workaround: Compute $\mathbb{E}\left\{\chi(B_*(g, a))\right\}$

In conclusion

Quantity of information in the EC curve



Applications of topological methods must either explain why the data should be suitable for those methods - e.g. why the complications that could arise, do not -- or they should be focused on invariants that can be measured.

Wise man has begun a study of such invariants, modeled on testability properties of graph properties. The simplest of these is the Euler characteristic divided by the volume – which is essentially (for Riemannian manifolds) the average (Pfaffian of the) curvature.

As an average, it is subject to sampling. Thus, a large submanifold in Euclidean space whose average curvature is large will surely have complicated topology, and discovering its topological properties will require enormous sampling and computational resources.

$$\chi(M) = \sum_{j=0}^{\dim M} (2\pi)^{-j/2} \sum_{m=0}^{\lfloor \frac{j}{2} \rfloor} C(N-j, j-2m) \frac{(-1)^m}{m!(j-2m)!}$$
$$\times \int_{\partial_j M} \int_{S(T_t \partial_j M^{\perp})} \mathrm{Tr}^{T_t \partial_j M}\left(R^m S^{j-2m}_{\nu_{N-j}}\right)$$
$$\times\, \alpha(\nu_{N-j}) \mathcal{H}_{N-j-1}(d\nu_{N-j}) \mathcal{H}_j(dt),$$

In conclusion: (sort of)

- Euler characteristics are easily computable via local computations on simplices.
- There exist many explicit formulae related to Euler characteristics of random systems, while for (eg) Betti numbers we only have asymptotics.
- Nature loves homological simplicity.
- When can persistence diagrams be replaced by Euler characteristic curves without a serious loss of useful information?

  For learning manifolds – probably never

  For comparing data sets – probably often

The preceding slides were brought to you by

European Research Council

Established by the European Commission

Uunderstanding Random Systems

URSAT

via Algebraic Topology

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

UNITED STATES AIR FORCE

SATA

COOPERATION

TOPOSYS

Doctoral and postdoctoral fellowships

at the

Technion

WE WANT YOU


THE
END

A little history

How Topologists Improve Citation Counts

John W. Tukey 1915 - 2000

- Princeton PhD with Solomon Lefshetz
- *Convergence and Uniformity in Topology*
  Princeton University Press 1940
  372 Google citations
- *An algorithm for the machine calculation of complex Fourier series*, with JW Cooley, 1965
  10,161 Google citations
- *Exploratory Data Analysis*, 1977
  12,115 Google citations.

40K on p1

Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise.

If we need a short suggestion of what exploratory data analysis is, I would suggest that
1: It is an attitude AND
2: A flexibility AND
3: Some graph paper (or transparencies, or both).