Persistent Homology of Data, Groups, Function Spaces, and Landscapes.

Shmuel Weinberger
Department of Mathematics
University of Chicago

William Benter Lecture
City University of Hong Kong
Liu Bie Ju Centre for Mathematical Sciences

May 12, 2010

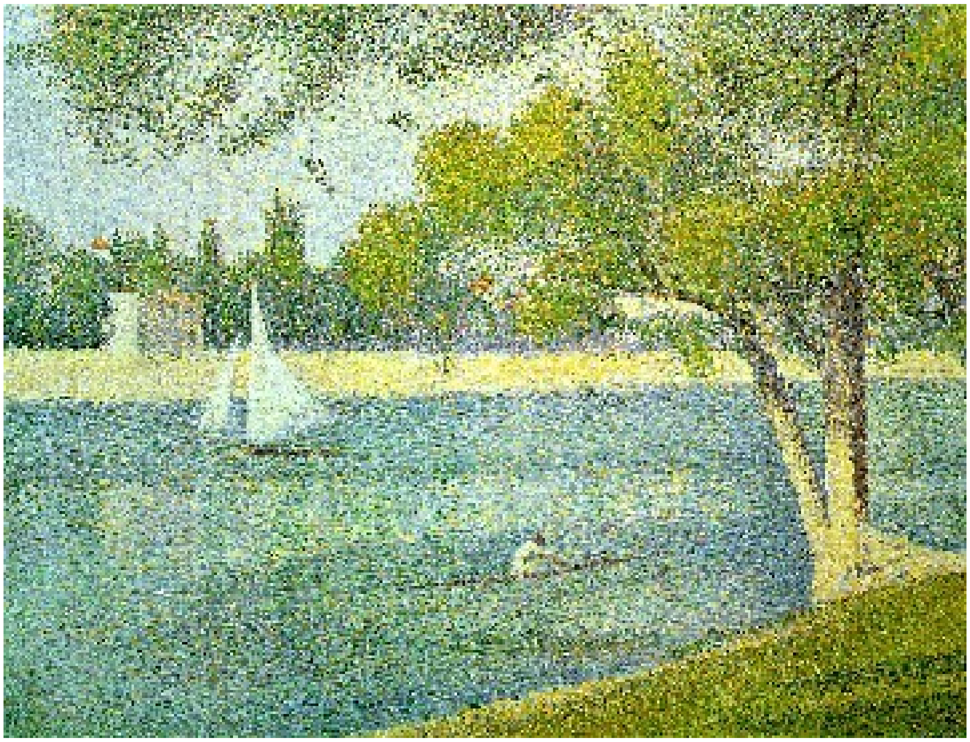
Outline:

I. Statements of Problems.

II. Persistent Homology, and Stability theorems

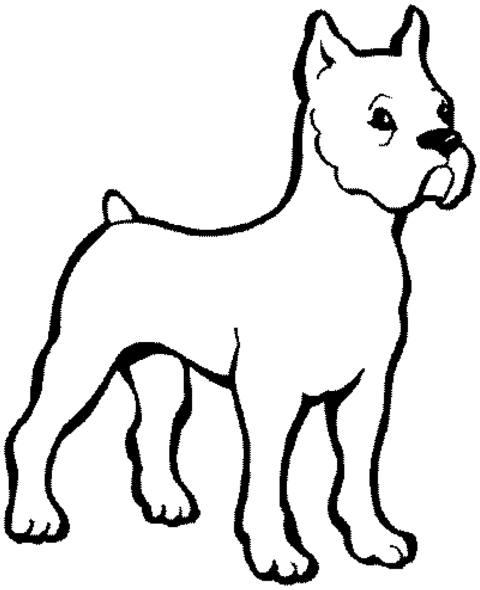
III. Applications.

IV. Further Directions.

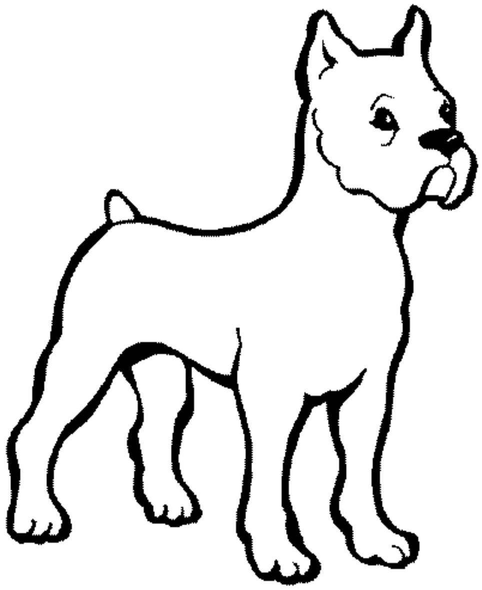


How do we interpret the dots of this painting as the picture of a boat and a canoe and a tree?

Cat vs. Dog



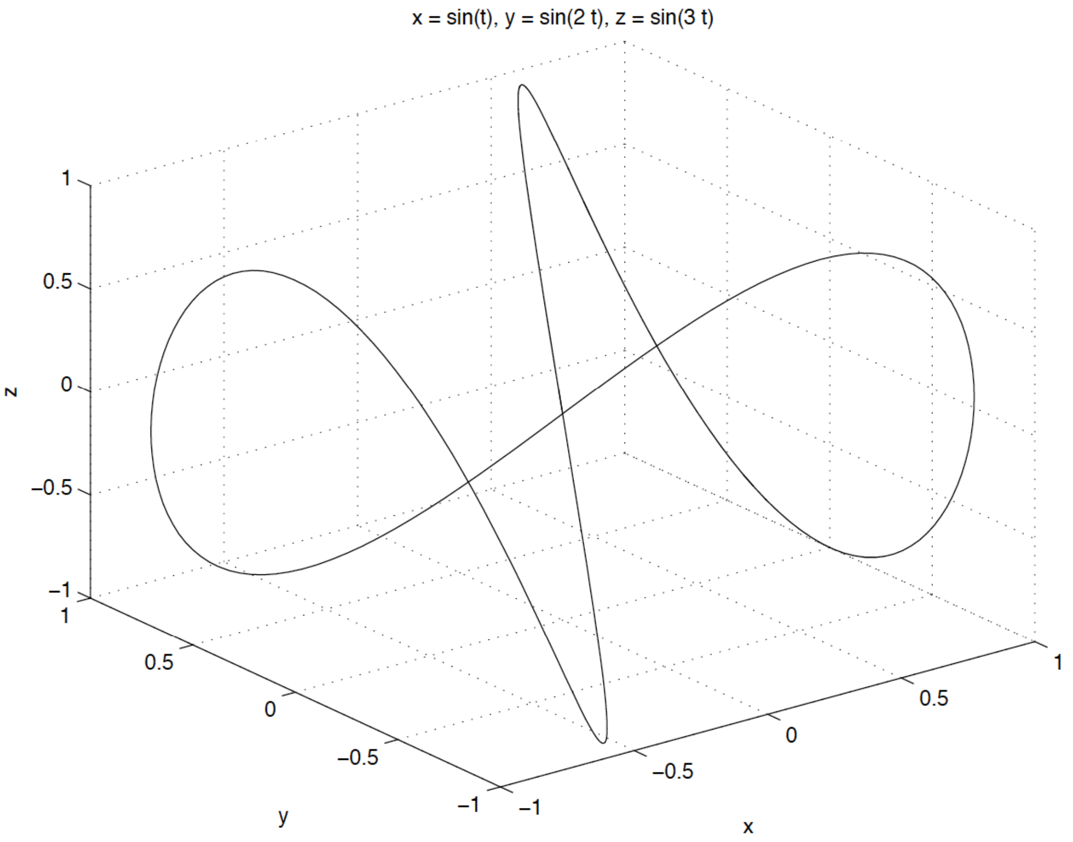
Cat vs. Dog



TWO RATHER DIFFERENT ASPECTS OF THE PROBLEM:

1. Pattern recognition
2. Concept formation and clustering in a Hilbert Space.

Observe Data. When can you hope to learn about it?



This doesn't look like it's near any lower dimensional linear subspace so the usual statistical methods, e.g. PCA don't directly apply.

KEY PROBLEMS

1. Clustering.
2. Dimensionality.
3. Entropy for time series.

KEY PROBLEMS

1. Clustering.
2. Dimensionality.
3. Entropy for time series.

ALL OF THESE ARE RELATED TO HOMOLOGY.

(WE WILL REVIEW HOMOLOGY LATER)

KEY PROBLEMS

1. Clustering.
2. Dimensionality.
3. Entropy for time series.

ALL OF THESE ARE RELATED TO HOMOLOGY.

AND WE MUST ALSO DISCUSS A TOOL, Persistent homology, FOR INFERRING HOMOLOGY OF A SPACE FROM ITS SAMPLES.

PROBLEM (Furstenberg, from the 1950's): Can a lattice in SLn(R) also be a lattice in SLm(R) if $n \neq m$

PROBLEM (Furstenberg, from the 1950's): Can a lattice in SLn(R) also be a lattice in SLm(R) if $n \neq m$

Let's consider by analogy lattices in $\mathbf{R}^n$ versus ones in $\mathbf{R}^m$.

PROBLEM (Furstenberg, from the 1950's): Can a lattice in SLn(R) also be a lattice in SLm(R) if $n \neq m$?

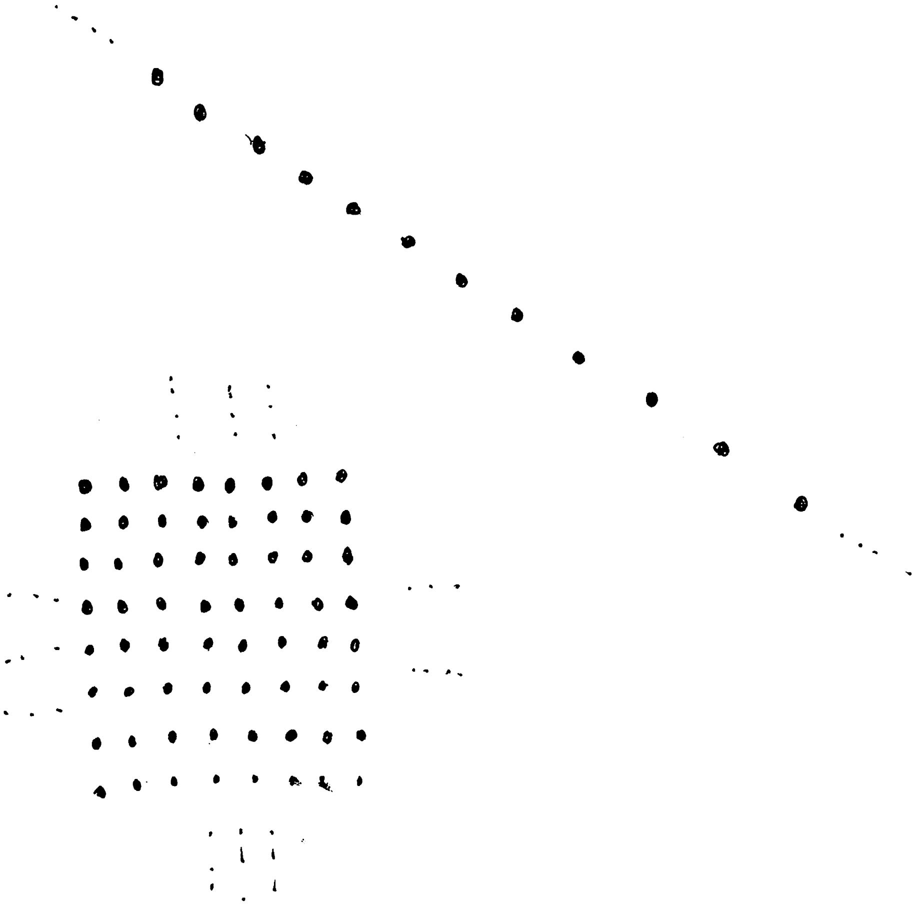
Let's consider by analogy lattices in $\mathbf{R}^n$ versus ones in $\mathbf{R}^m$.

(And, let's not use algebraic ideas like abelian groups, vector spaces, (algebraic) dimension etc.)

PROBLEM (Furstenberg, from the 1950's): Can a lattice in $SLn(R)$ also be a lattice in $SLm(R)$ if $n \neq m$?

Let's consider by analogy lattices in $\mathbf{R}^n$ versus ones in $\mathbf{R}^m$.

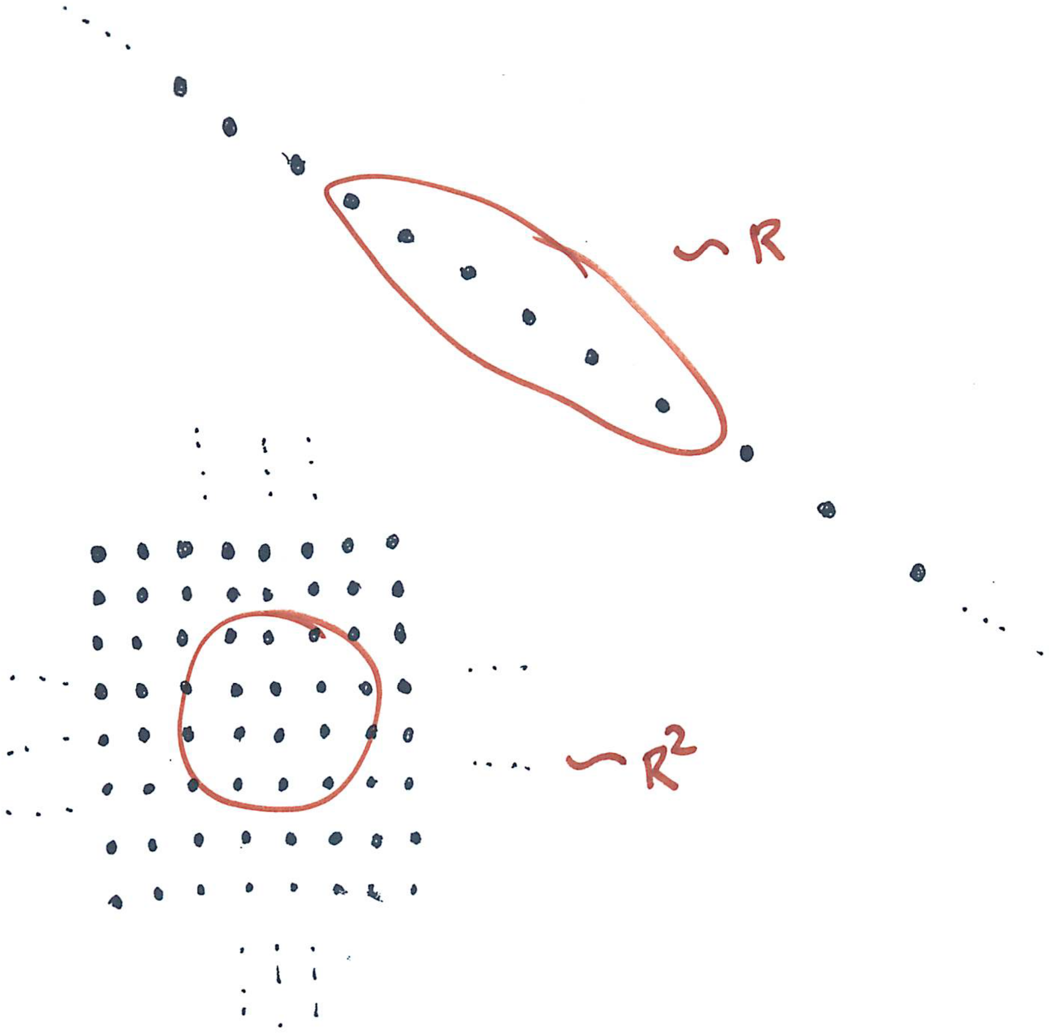
(And, let's not use algebraic ideas like abelian groups, vector spaces, (algebraic) dimension etc.)



PROBLEM (Furstenberg, from the 1950's): Can a lattice in $SL_n(R)$ also be a lattice in $SL_m(R)$ if $n \neq m$?

Let's consider by analogy lattices in $\mathbf{R}^n$ versus ones in $\mathbf{R}^m$.

(And, let's not use algebraic ideas like abelian groups, vector spaces, (algebraic) dimension etc.)



For lattices in $\mathbf{R}^n$, the **volume growth** is like $(\text{radius})^n$, so the growth rate distinguishes in this case.

For lattices in $\mathbf{R}^n$, the **<u>volume growth</u>** is like $(\text{radius})^n$, so the growth rate distinguishes in this case.

For lattices in $SL_n(R)$ the growth rate is EXPONENTIAL for every $n>1$. So we need a new idea.

For lattices in $\mathbf{R}^n$, the **volume growth** is like $(\text{radius})^n$, so the growth rate distinguishes in this case.

For lattices in SLn(R) the growth rate is EXPONENTIAL for every $n>1$. So we need a new idea.

But not completely new…

Volume growth (of a lattice)

= Hausdorff dimension of the

enveloping space.

Recall: Hausdorff dimension essentially measures how many balls of radius R does it take to cover the ball of radius 2R. It should be $2^{\text{dimension}}$.

Since Hausdorff dimension doesn't work, we can try **topological dimension** instead. (The topological dimension is always at most the Hausdorff dimension.)

Since Hausdorff dimension doesn't work, we can try **topological dimension** instead. (The topological dimension is always at most the Hausdorff dimension.)

Just like in sampling: We need a method, intrinsic to a sampled set, do determine a topological feature of a space.

Since Hausdorff dimension doesn't work, we can try **topological dimension** instead.

Just like in sampling: We need a method, intrinsic to a sampled set, do determine a topological feature of a space.

We will do so, later, using persistent homology.

A Third Example in Riemannian Geometry.

We will prove (and generalize):

Theorem: If M is a compact Riemannian manifold whose fundamental group has unsolvable word problem, then M has infinitely many closed contractible geodesics.

A Third Example in Riemannian Geometry.

We will prove (and generalize):

Theorem (Gromov 1970's): If M is a compact Riemannian manifold whose fundamental group has unsolvable word problem, then M has infinitely many closed contractible geodesics.

Where's the sampling? The inference of an enveloping structure from a substructure?

A Third Example in Riemannian Geometry.

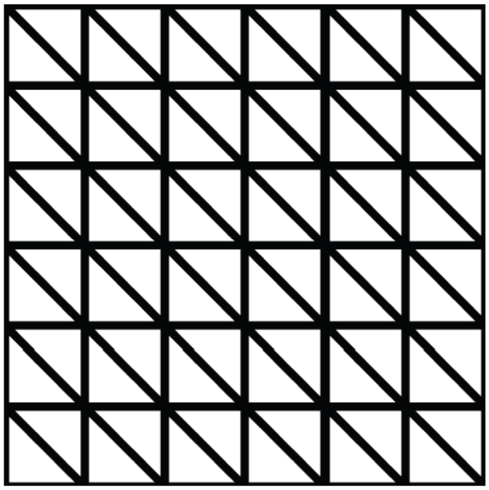
We will prove (and generalize):

Theorem (Gromov 1970's): If M is a compact Riemannian manifold whose fundamental group has unsolvable word problem, then M has infinitely many closed contractible geodesics.

Where's the sampling? The inference of an enveloping structure from a substructure?

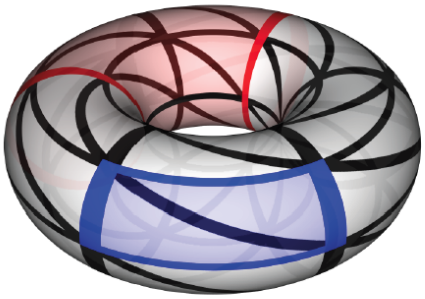
We will see later...

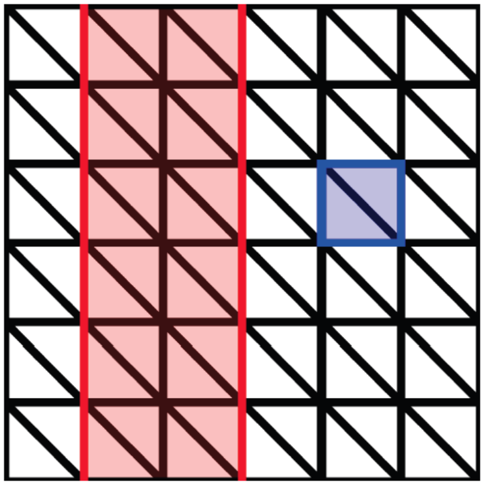
Homology



Homology $= \{$no boundary$\}$ mod boundaries

Homology





$$\text{Homology} = \{\text{no boundary}\} \bmod \text{boundaries}$$

Basic facts about homology:

1. It is well defined (i.e. independent of triangulation – although it can be computed from a triangulation).

2. It only depends on the homotopy type (=deformation type) of the space.

3. $H_0(X)$ measures how many components X has.

4. $H_1(X)$ is a commutative measure of whether X is simply connected (or whether irrotational vector fields on X are necessarily gradient).

5. The dimension of X (if $< \infty$) = sup $\{k \mid H_{k+1}(U) = 0$ for all open $U \subset X\}$.

Basic facts about homology:

1. It is well defined (i.e. independent of triangulation – although it can be computed from a triangulation).

So the construction is intrinsic.

2. It only depends on the homotopy type (=deformation type) of the space.

This is the key to avoiding "overfitting".

3. $H_0(X)$ measures how many components X has.

So we can solve "clustering" problems.

4. $H_1(X)$ is a commutative measure of whether X is simply connected (or whether irrotational vector fields on X are necessarily gradient).

In general, the k-th homology of a space only depends in its k dimensional aspects.

5. The dimension of X (if $< \infty$) = sup $\{k \mid H_{k+1}(U) = 0$ for all open $U \subset X\}$.

So we will be able to use homology to decide problems of dimensionality (especially relevant to the group theory example).

Persistent Homology

Definition: Suppose that we have $X = X_r$ a nested sequence of spaces (satisfying mild technical conditions) then we define persistent homology $PH_k(X)$ by the formula:

$$PH_k(X) = \Pi\ H_k(X_r).$$

Persistent Homology

Definition: Suppose that we have $X = X_r$ a nested sequence of spaces (satisfying mild technical conditions) then we define persistent homology $PH_k(X)$ by the formula:

$$PH_k(X) = \Pi\ H_k(X_r).$$

What kind of object is the right hand side?

Persistent Homology

Definition: Suppose that we have $X = X_r$ a nested sequence of spaces (satisfying mild technical conditions) then we define persistent homology $PH_k(X)$ by the formula:

$$PH_k(X) = \Pi\ H_k(X_r).$$

What kind of object is the right hand side?

If we use coefficients in a field, then we can think of it as a collection of intervals in the real line, parametrizing the r's from which a particular homology class through its lifetime, till it dies.

Persistent Homology

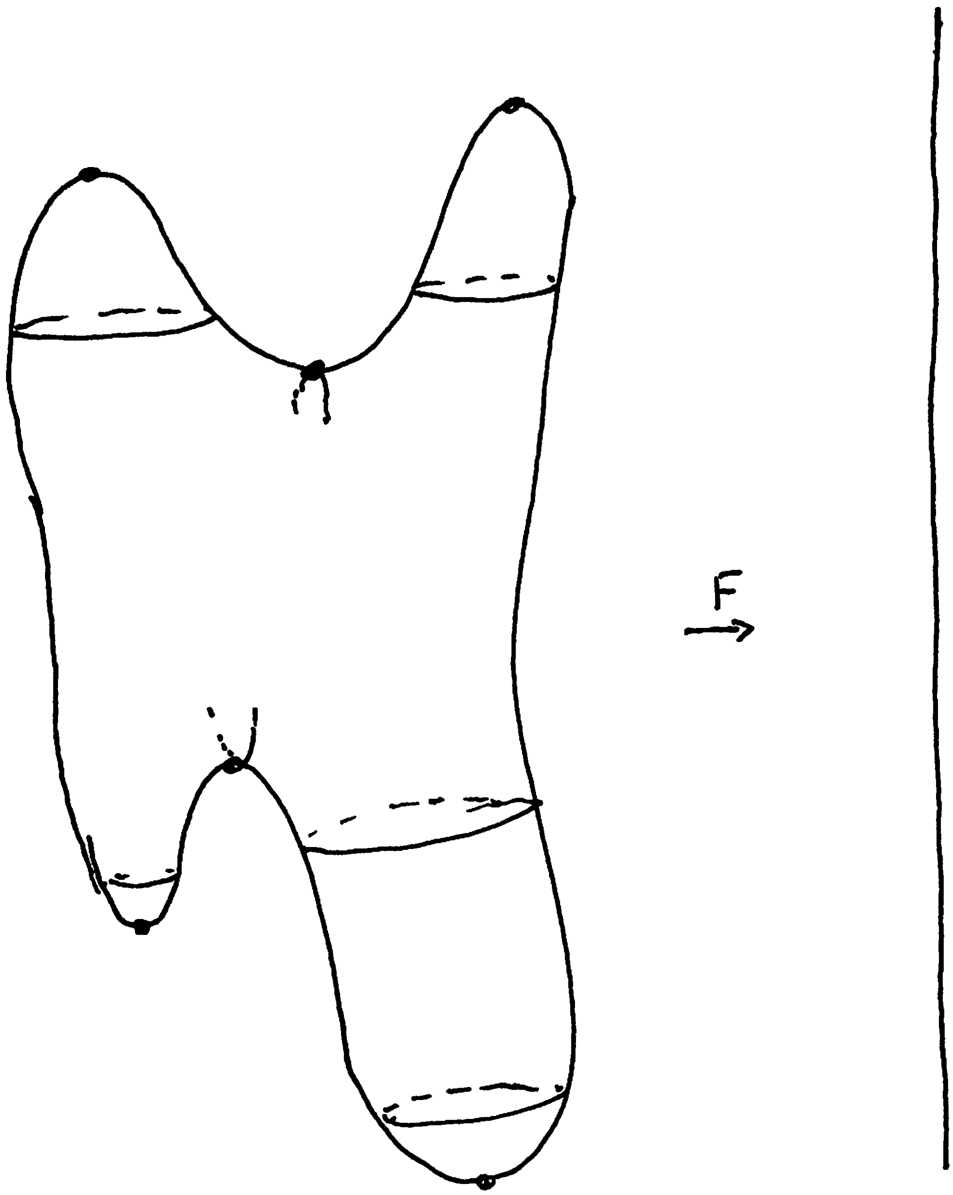
Definition: Suppose that we have $X = X_r$ a nested sequence of spaces (satisfying mild technical conditions) then we define persistent homology $PH_k(X)$ by the formula:

$$PH_k(X) = \Pi\ H_k(X_r).$$

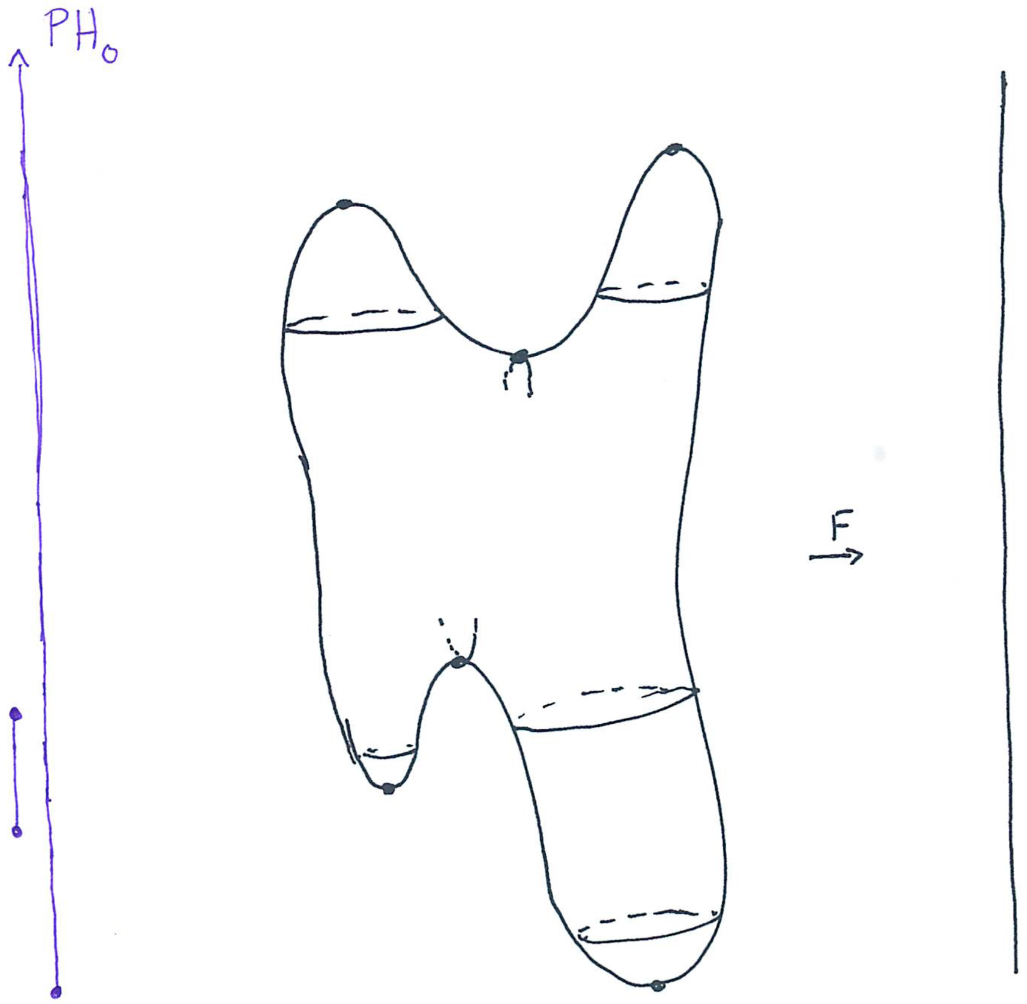
What kind of object is the right hand side?

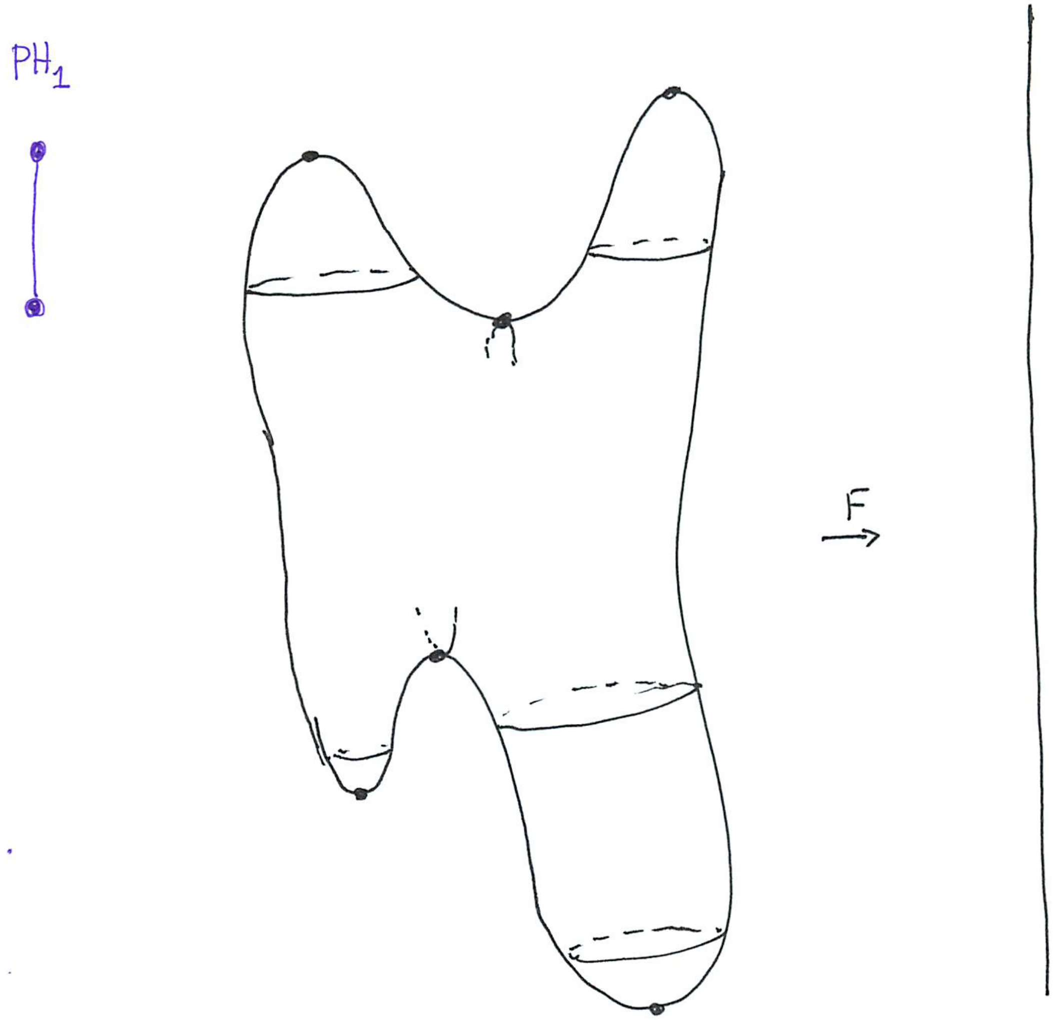
If we use coefficients in a field, then we can think of it as a collection of intervals in the real line, parametrizing the r's from which a particular homology class through its lifetime, till it dies.

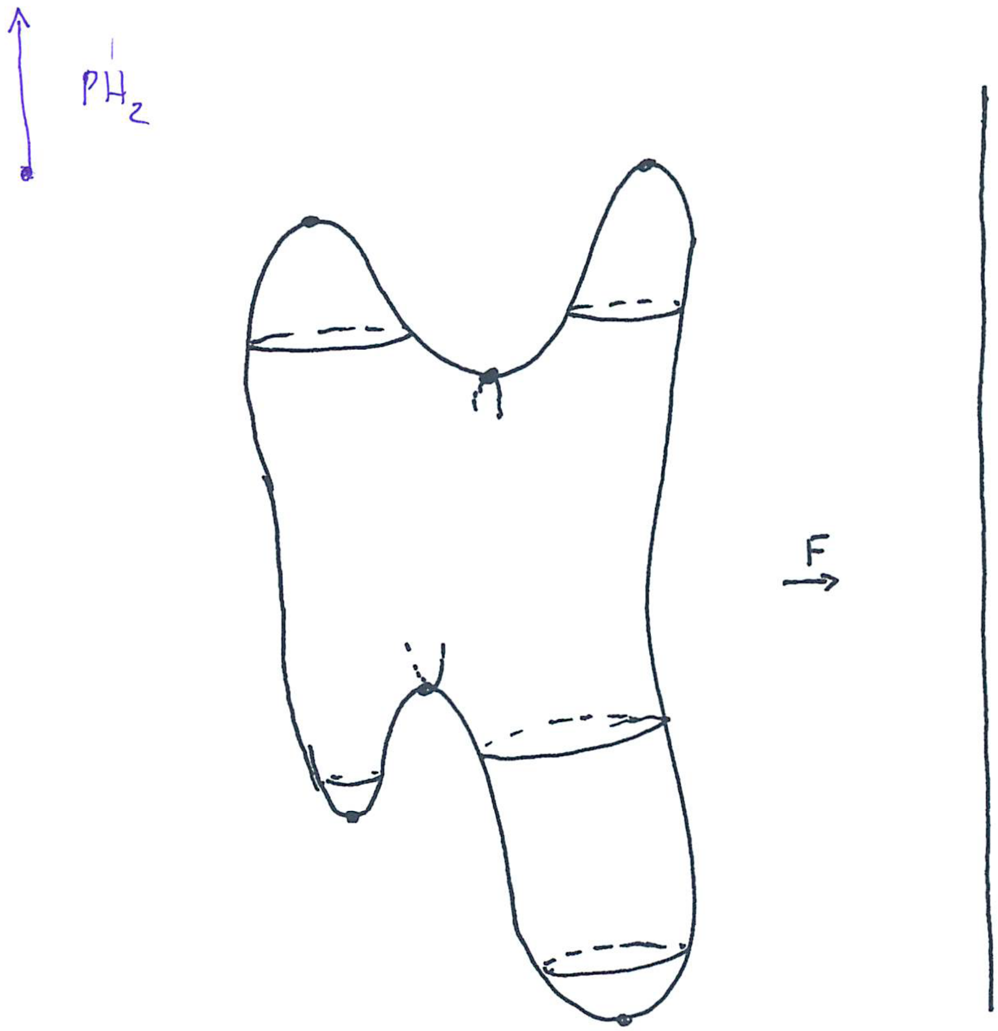
Here is an example:



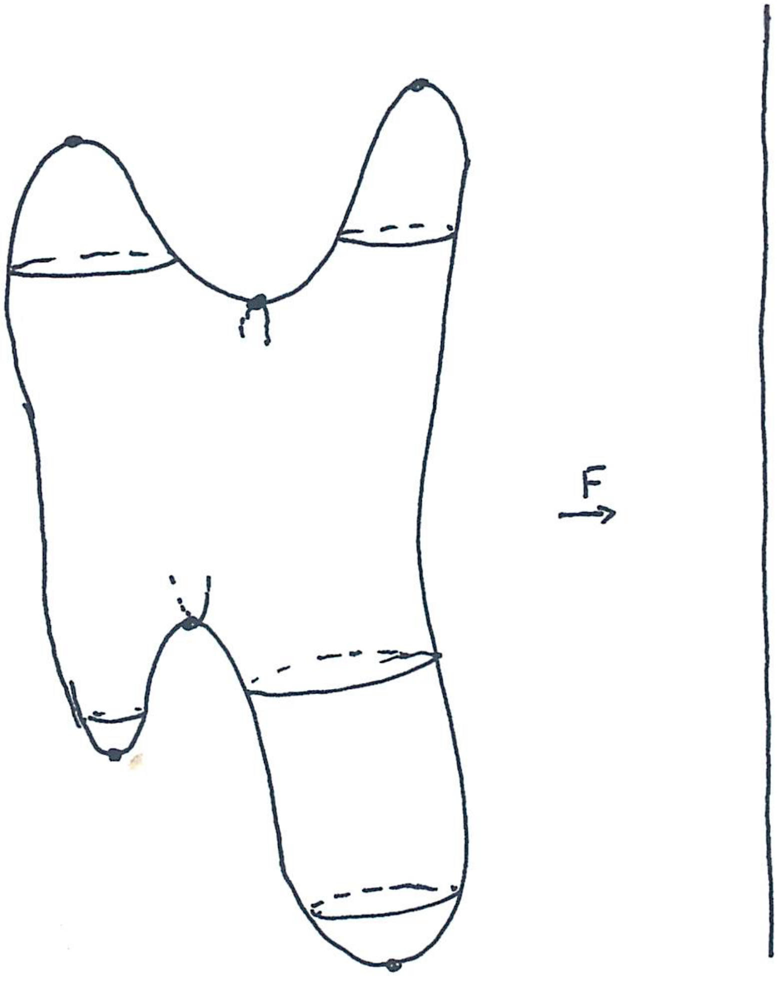
$X_r = F^{-1}(-\infty, r]$


PH_0
F

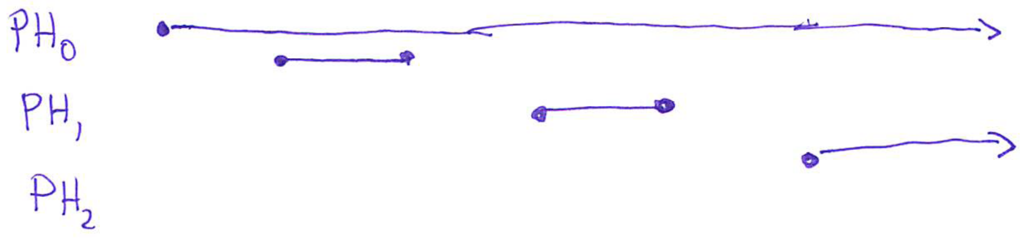

PH_1
F


PH_2
F

In Summary:



$$X_r = F^{-1}(-\infty, r]$$



Examples:

(1) Given X, use a function f: X → **R**.

$$X_r = \{ x \in X \mid f(x) \leq r\}$$

Examples:

(1) Given X, use a function f: $X \rightarrow \mathbf{R}$.

$X_r = \{ x \in X \mid f(x) \leq r\}$

(2) For $X \subset \mathbf{R}^n$, let

$X_r = \{ u \in \mathbf{R}^n \mid \exists\, x \in X$, such that $\|x-u\| \leq r\}$

Examples:

(1) Given X, use a function f: $X \to \mathbf{R}$.

$X_r = \{ x \in X \mid f(x) \le r\}$

(2) For $X \subset \mathbf{R}^n$, let

$X_r = \{ u \in \mathbf{R}^n \mid \exists\, x \in X$, such that $\|x-u\| \le r\}$

(3) Let (X,d) be a metric space. We can embed X in $L^\infty(X)$

by $x \to \inf(d(x, ?), 1)$. Now define

$X_r = \{ u \in L^\infty(X) \mid \exists\, x \in X$, such that $\|x-u\|_\infty \le r\}$.

Examples:

(1) Given X, use a function f: $X \to \mathbf{R}$.

$X_r = \{ x \in X \mid f(x) \leq r\}$

(2) For $X \subset \mathbf{R}^n$, let

$X_r = \{ u \in \mathbf{R}^n \mid \exists\, x \in X$, such that $\|x-u\| \leq r\}$

(3’) Let (X,d) be a metric space, * a base point. We can embed X in $L^\infty(X)$ by $x \to d(x, ?) - d(*,)$. Now define

$X_r = \{ u \in L^\infty(X) \mid \exists\, x \in X$, such that $\|x-u\|_\infty \leq r\}$.

(3) is sometimes better for “small scale” and (3’) is always better for large scale problems.

The key to applications of PH is that it has some stability properties.

The key to applications of PH is that it has some stability properties.

We will focus on example (1) since it is typical.

The key to applications of PH is that it has some stability properties.

We will focus on example (1) since it is typical.

In some sense: PH(f) is a continuous function of f.

The key to applications of PH is that it has some stability properties.

We will focus on example (1) since it is typical.

In some sense: PH(f) is a continuous function of f.

Note: This is a change of perspective from usual topology – where invariants are supposed to be “functorial”. Here they are “functional”.

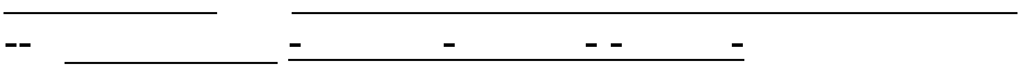
Stability theorem. (Cohen-Steiner, Edelsbrunner, Harer).

If f, g: X → **R** are functions, then

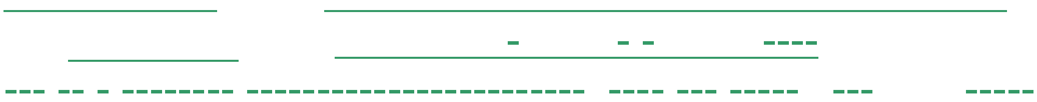
$$d_{\mathrm{Bottleneck}}(PH(f), PH(g)) \leq \| f - g \|_\infty$$

Technical issue:
In this theorem we should allow arbitrary numbers of 0-length homology intervals.

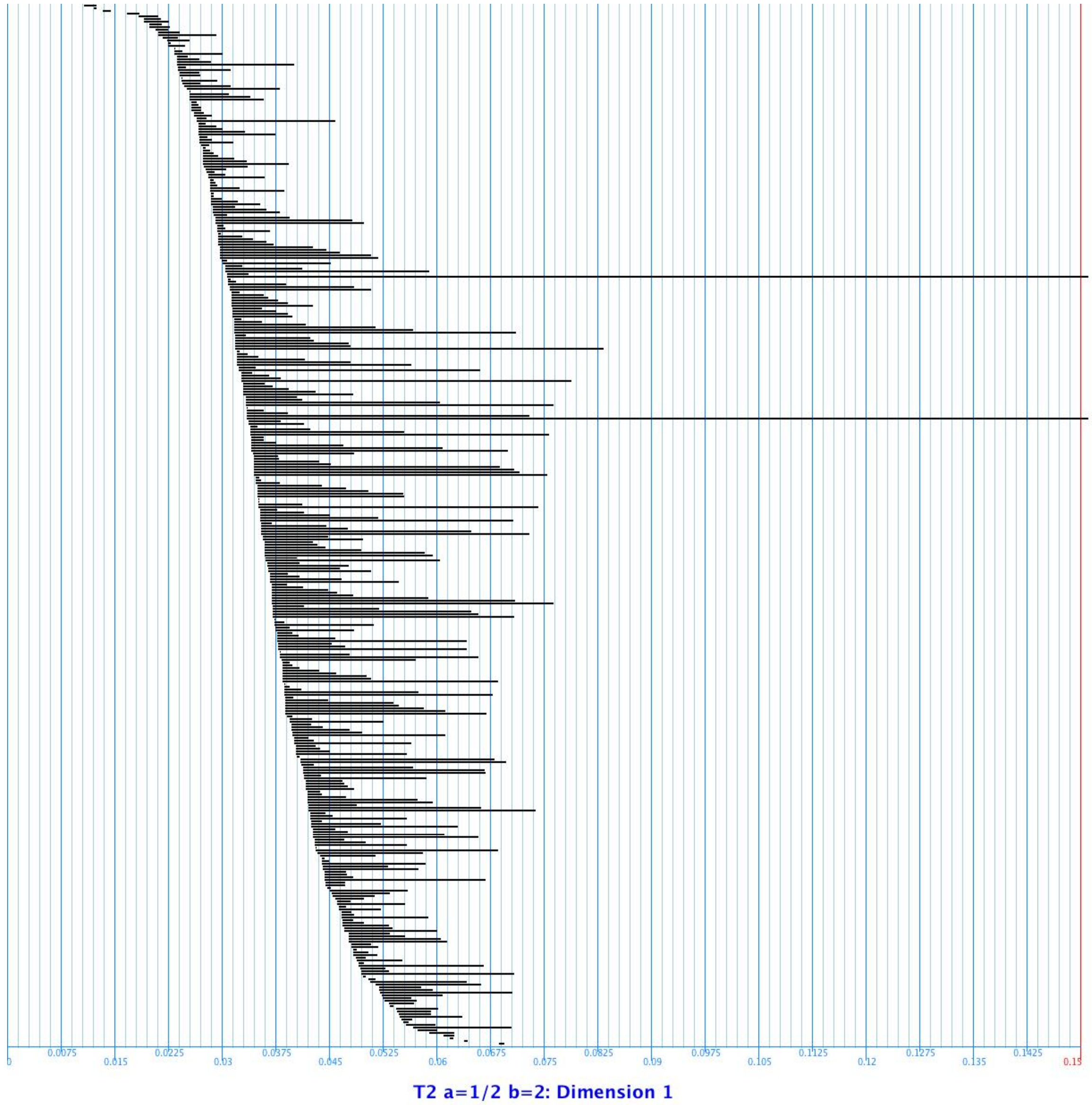
Example of “bottleneck distance”:



is close to



because the “long intervals” are placed with close start- and end-points.



A more typical barcode taken from a computer experiment by Steve Ferry.

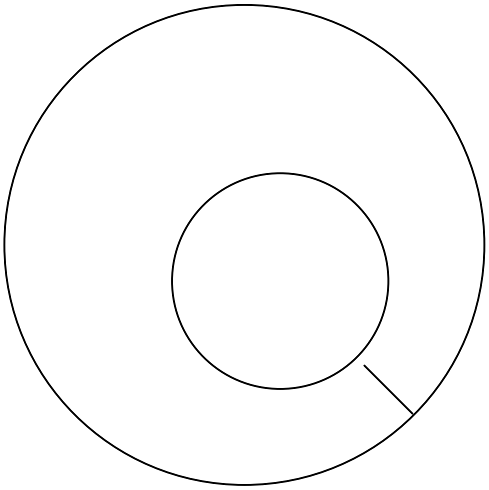
Application to sampling.

Hypotheses:

1. M^n is a compact smooth submanifold of $\mathbf{R}^{\mathbf{d}}$.
2. We are given $\tau > 0$ that is a condition number if $m_i \in$ M and $\upsilon_i \in TM_{m_i}$ with $\| \upsilon_i \| < \tau$ then

$$m_1 + \upsilon_1 = m_2 + \upsilon_2 \Rightarrow m_1 = m_2 \text{ and } \upsilon_1 = \upsilon_2.$$

The line segment below has length 2τ.



Theorem (Niyogi-Smale-Weinberger): Suppose that M is as above, and that one knows (and upper bound on) vol(M) or diam(M). Then it is possible to calculate a lower bound on the probability that for a sample S = { m_i i=1….N} chosen uniformly from M, that one has an isomorphism between $H_*(M)$ and the intervals of size > [ε/4, ε] for $PH_*(S)$ for any $\varepsilon < \tau$.

Theorem (Niyogi-Smale-Weinberger): Suppose that M is as above, and that one knows (and upper bound on) vol(M) or diam(M). Then it is possible to calculate a lower bound on the probability that for a sample S = { m_i i=1….N} chosen uniformly from M, that one has an isomorphism between $H_*(M)$ and the intervals of size > [ε/4, ε] for $PH_*(S)$ for any ε <τ.

Remarks:

1. This is a paraphrase of the theorem in [NSW], which gives a related statement even for integral homology.
2. τ incorporates 2 aspects:
    a. Local: measuring the second fundamental form of M.
    b. Global, e.g. measuring the separation between two parallel planes.
3. There is an algorithm for computing the persistence homology. We will discuss this a bit later.

4. The definition of PH for our purpose uses either of the inequivalent definitions (2) and (3). Using (2) one does not need positive length intervals, because of the following remark:
5. The paper actually gives a fixed scale where calculation is possible. As a result, the main result of [NSW] is rather stronger than the above formulation.
6. Related work was done by [Cohen-Steiner, Edelsbrunner, Harer], [Chazal-Liutier], [Chazal-Cohen-Steiner-Merigot].
7. That we can work at a fixed scale is useful for our approach to dealing with the problem of noise. [NSW2, to appear].
8. The use of the X_r of type (3) is
    a. Closely related, for samples, to Rips complexes, that have computational and theoretical advantages over the geometrically more natural Cech complexes, and
    b. Seems related to one of the ideas in the recent paper of [Bartholdi, Schick, Smale and Smale] on Hodge theory.
9. We will discuss more details of this at the end of the talk.

Example II. (Discrete groups).

Let X be a discrete metric space.

A finitely generated discrete group is made into a metric space using the word metric. (d(g,h) = smallest number of generators it takes to write $g^{-1}h$.)

The version of HP(X), where the filtration comes from type (3') can be concretely described as follows.

X_R is a simplicial complex, whose k simplices are k+1 tuples such that all pairwise distances are $\leq R$. Now we will let $R \to \infty$.

Proposition: If π is a discrete group acting properly discontinuously and cocompactly on a contractly polyhedron Z, then the limit as R → ∞ of $H_i^{lf}(\pi) = H_i^{lf}(P)$.

Thus the right hand side has a "coarse meaning". The infinitely long persistence intervals reflect something interesting about the geometry of the group.

For i=1, this tells us how many ends the group has (which equals the number of ends the universal cover of any compact space with that fundamental group has).

We can also consider the largest i for which this is non-zero. This is strong enough to distinguish many lattices from each other.

Corollary (Gersten, Block-Weinberger) For groups of finite type, cohomological dimension is a coarse quasi-isometry invariant.

In particular as $\mathrm{cd}(SL_n(\mathbf{Z})) = n(n-1)/2$, no lattices commensurable to $SL_n(\mathbf{Z})$ can be bi-Lipschitz to a lattice commensurable to $SL_m(\mathbf{Z})$, for $n \neq m$.

Example III: Closed geodesics.

We recall Gromov's theorem:

Theorem: If M is a compact Riemannian manifold whose fundamental group has unsolvable word problem, then M has infinitely many closed contractible geodesics.

Definition: We say that a manifold M has **property S** (Shrinking) if there is a constant C, such that any contractible curve of length L can be contracted through curves of length $\leq$ CL to one of length L/2.

Theorem: The question of whether a compact manifold M has **property S** only depends on $\pi_1 M$.

Theorem: The C implicit in property S is a function of the metric on M. It only depends on (inj, sup(|K|), vol(M)).

However C(inj, sup(|K|), vol(M)) cannot be bounded by any recursive (=computable) function of these arguments – even for metrics on the n-sphere, at least for n>4.

(For n=3 there is such a computable function. Indeed I believe that all compact 3-manifolds have property S as a consequence of Perelman's work.)

To understand these we need another notion, the Dehn function of a presentation of a group.

Definition: Let $\pi = < g_1 , g_2 \ldots., g_k \mid r_1 , r_2 , \ldots., r_m >$ be a finitely presented group.

$D_\pi(n) = \inf \{s \mid$ any word of length $\leq n$ is a product of at most s relations$\}$.

D(n) depends on the presentation, but its “growth rate” (e.g.polynomial, exponential, superexponential, computably bounded, etc.) does not.

D measures the following Riemannian property of manifolds with fundamental group π: What is the smallest area of all disks bounded by nullhomotopic curves of length $\leq L$? So for free abelian groups of rank > 1, D grows quadratically.

Remark: D is bounded by a computable function if and only if the fundamental group has a solvable word problem.

We now can assert our strengthening of Gromov’s theorem:

Theorem: If the Dehn function of π is super-exponential, then M does not have property S.

Simultaneous with proving this we give a characterization of Property S.

We still need one more idea:

Let M be a compact Riemannian manifold, and $\Lambda M = \{f: S^1 \to M\}$. We let $E: \Lambda M \to \mathbf{R}$ denote the energy

$$E(f) = \int \langle f'(t), f'(t) \rangle \, dt.$$

Proposition: Although the Energy of a curve depends on the Riemannian metric, the difference

$$\|\log E_1 - \log E_2\|_\infty \leq \sup |\log(\langle,\rangle_1 / \langle,\rangle_2)|$$

is bounded.

Hence:

Theorem: The “barcodes” PH(ΛM, log(E)) are well defined module “short intervals” of uniformly bounded size.

Property S $\Leftrightarrow$ $PH_0(\Lambda M ; \log(E))$ has arbitrarily long finite length intervals.

Note that the bottom of a $PH_0(\Lambda M ; \log(E))$ interval corresponds to a local minimum. The intervals in general all correspond exactly to various closed geodesics of various indices.

The rest of the theorem comes from an analysis of $PH_0(\Lambda K(\pi,1) ; \log(E))$.

This uses the combination of the Dehn function hypothesis and the topological entropy of ΛM.

Implicit in this are new types of algebraic topological invariants of finite complexes with variational meaning. We will later discuss some partial explorations of these.

IV. Further and future directions

Data Analysis.

1. What are the actual computational and sample complexities of these problems?
2. Are there topological features that are discoverable before the full homotopy type?
3. Can one use persistence homology at scales where the actual homology is not visible.
4. How does one measure the statistical significance of a persistence calculation of data?
5. What are the mechanisms for dealing with noise? (Cleaning, or kernel methods)
6. Extend the theory of PH for metric spaces to metric measure spaces.
7. What are the borders of well-posedness of these problems? Can complexity then be viewed as a measure of distance to the ill-posed set?

Geometric Group Theory and Large Scale Geometry.

1. Other functors, such as K-theory (applied to disprove Gromov's conjecture that uniformly contractible manifolds are hyperspherical)
2. Homotopy with coefficients can be used to produce barcodes. Ferry and I have studied this for [X: Y], Y simply connected and finite.

   This has many geometric applications, potentially, because of h-principles, surgery, cobordism….
3. It becomes necessary to develop new algebraic topology for this setting.
4. Dehn functions extend to other filling functions, and persistent homology has been varied into other coarse theories (e.g. uniformly finite, L^2, etc. that could have other applications).

5. Bounded propagation speed operators on metric spaces relates both to K-theory and to parallel processing.
6. Families of these can be applied to the Novikov conjecture (Ferry-W, Gromov-Lawson, Kasparov, Higson-Roe, Yu….) which gives information about compact manifolds, via the family of universal covers.

Landscapes:

(Epi)genetic & Economic.

Two mechanisms for the construction of “nontopological” critical points (and especially optima).

> Logical and computational complexity implies geometric complexity.
>
> Competition leads to computational complexity.
>
> Perturbation by random fields gives rise to these in a fashion, sometime computable by Rice-type formulae.