Mathematisches Forschungsinstitut Oberwolfach

Report No. 29/2008

Computational Algebraic Topology

Organised by
Gunnar Carlsson, Stanford
Dmitry Kozlov, Bremen/Zürich

June 29th – July 5th, 2008

Abstract. Computational algebraic topology is a dynamic field of mathematics, which has close connections to the classical algebraic topology, combinatorial algebraic topology, theory of algorithms, as well as an abundance of applications. The purpose of the workshop was to bring together the leading figures in the subject to foster interaction and better understanding of the current momentum.

Mathematics Subject Classification (2000): 55xx.

Introduction by the Organisers

The workshop was conducted jointly with a workshop in statistical learning theory. There was substantial interaction between the two groups, both formally in terms of talks attended by members of both groups, as well as via informal discussions. The intellectual themes which were presented during the workshop are described below.

Sensor nets and engineering applications:

In the opening talk R. Ghrist spoke about the topology necessary to develop methods for determining intruders have entered a net of sensors, and for counting their number. Ghrist, jointly with V. de Silva and Y. Baryshnikov, has developed techniques based directly on homological calculations as well as on integrals over Euler characteristics which hold promise for implementable algorithms. In order for such algorithms to be maximally useful, one must develop error insensitive methods, which will require more probabilistic methods to be included within the algebraic topological framework.

Combinatorial applications:

Several presentations at the workshop elaborated on the subject of combinatorial algebraic topology. D. Kozlov has given a survey talk, which has set the accents on the subject, tying together structures, methods, and applications, as these are present at the current state of the development. Talks by R. Jardine and M. Raussen concerned the combinatorial and computational aspects of homotopy theory, finding applications of such abstract notions as Quillen's closed model category. K. Knudson gave an interesting account of connections between persistent homology and discrete morse theory. Finally, the talk of E. Babson dealt with more probabilistic aspects and served as a bridge to the presentations of M. Kahle and P. Bubenik.

Dynamical systems:

K. Mischaikow and S. Day spoke about the use of algebraic topology to understand the qualitative structure of dynamical systems. Mischaikow introduced his paradigm of building databases of dynamical systems based on choices of parameter values. His methods permit the construction of partitions of parameter space within which the qualitative structure remains the same. In addition, Conley index methods, or rather their computational versions, are used to prove the existence of fixed points, recurrent points, and invariant subsets within a given region in a spatial domain.

Data analysis:

G. Carlsson and V. de Silva spoke about applications of various kinds of diagrams to understand the qualitative geometric nature of data sets. For example, persistence diagrams allow one to recover Betti numbers of sublevel sets of a probability distribution, multidimensional persistence allows one to study sublevel sets of various functions as well, and the analysis of structure theorems for certain kinds of quivers permits one to extend the bootstrap methods to clustering, Betti numbers, as well as to perform dynamic clustering (i.e. clustering over time). There are now viable computational methods for all of these applications.

Probabilistic methods:

M. Kahle and P. Bubenik spoke about the beginnings of stochastic algebraic topology. Work at the level of zeroth Betti numbers has already been carried out by M. Penrose, under the heading of "geometric random graphs". What is now needed is an extension of this work to higher dimensional homology groups, as well as to the barcodes which arise in persistent homology. Ultimately, precise results along these lines will open up the possibility of direct evaluation of significance of various qualitative observations given a null hypothesis.

There were also several talks more centered at applications, such as vision recognition (J. Giesen) and material science (R. MacPherson).

All things considered, the workshop was a great success in terms of scientific interaction, both within this group, as well as with the researchers in statistical learning theory, as was witnessed by many involved discussions, which often lasted well into the late evenings.

Workshop: Computational Algebraic Topology

Table of Contents

Abstracts

Fundamental Groups of random 2-complexes

Eric Babson
(joint work with Christopher Hoffman and Matthew Kahle)

The random graph $G(n,p)$ is the probability space of all graphs on vertex set $[n] := \{1, 2, \ldots, n\}$ with each edge inserted independently with probability p. One frequently considers p a function of n and asks whether a typical graph in $G(n,p)$ is likely to have a given monotone property as $n \to \infty$. We say $G(n,p)$ **almost always** has property $\mathcal{P}$ if $\lim_{n\to\infty} \mathbf{P}(G(n,p) \in \mathcal{P}) = 1$, which we sometimes abbreviate as 'a.a.' A famous result of Erdős and Rényi [Erdős and Rényi, 1959] is that $p = \log n/n$ is the threshold function for the connectivity of the random graph. More precisely, they showed the following.

Theorem 1. [Erdős and Rényi, 1959] *Let $\omega(n) \to \infty$ as $n \to \infty$. If $p = (\log n - \omega(n))/n$ then $G(n,p)$ is almost always disconnected, and if $p = (\log n + \omega(n))/n$, $G(n,p)$ is almost always connected.*

Nathan Linial and Roy Meshulam recently exhibited a 2-dimensional homological analogue of Theorem 1. They defined the random 2-dimensional simplicial complex $Y(n,p)$ to be the probability space of simplicial complexes on vertex set $[n]$ and edge set $\binom{[n]}{2}$, with each 2-face appearing independently with probability p.

Theorem 2. [Linial and Meshulam, 2005a] *Let $\omega(n) \to \infty$ as $n \to \infty$. If $p = (2\log n - \omega(n))/n$ then almost always $H_1(Y, \mathbb{F}_2) \neq 0$, and if $p = (2\log n + \omega(n))/n$ then $H_1(Y, \mathbb{F}_2) = 0$ almost always.*

Meshulam and Wallach later extended this result to $H_1(Y, \mathbb{F}_q)$ for any prime q and proved analogous results for random higher dimensional simplicial complexes [Meshulam and Wallach, 2006].

In this article we address the analogous question for $\pi_1(Y(n,p))$. We show that the threshold for vanishing of $\pi_1(Y(n,p))$ is approximately $p = n^{-1/2}$, in contrast to the Linial-Meshulam threshold for homology of roughly $p = n^{-1}$.

Theorem 3. *If*

$$p \geq \left(\frac{3\log n + \omega(n)}{n}\right)^{1/2}$$

where $\omega(n) \to \infty$, then $\pi_1(Y(n,p)) = \{1\}$ a.a.

The proof of Theorem 3 is based on showing that once every pairwise intersections of vertex links is connected, every triangle bounds an embedded disk. Then every cycle can be factored as a product of triangles, since the underlying graph is complete, and by simplicial approximation, every loop is homotopy equivalent to a product of cycles.

Our main result shows that the exponent $1/2$ in Theorem 3 is best possible.

Theorem 4. *For any* $\epsilon > 0$ *if*

$$p \leq \frac{n^{-\epsilon}}{n^{1/2}}$$

then $\pi_1(Y(n,p)) \neq \{1\}$ *and is hyperbolic a.a.*

Our proof of Theorem 4 is geometric in spirit and relies on general notions of negative curvature due to Gromov. In particular, we show that when p is in this range, $Y(n,p)$ is hyperbolic in the sense of satisfying a linear isoperimetric inequality.

References

[Bollobas, 2001] Bollobás B., Random Graphs, Second Edition, Cambridge University Press, 2001.

[Bollobás and Thomason, 1985] Bollobás B. and Thomason A.G., Random graphs of small order. In Random Graphs, Annals of Discr. Math, pp. 47-97, 1985.

[Björner, 2003] Björner A., Nerves, fibers, and homotopy groups, J. Combinatorial Theory, Ser. A, 102, 88–93, 2003.

[Bridson, 1999] Bridson, M., Non-positive curvature in group theory. (English summary) Groups St. Andrews 1997 in Bath, I, 124–175, London Math. Soc. Lecture Note Ser., 260, Cambridge Univ. Press, Cambridge, 1999.

[Erdős and Rényi, 1959] Erdős P., and Rényi A., On random graphs I, Publ. Math. Debrecen **6**, 290-297, 1959.

[Gromov, 1987] Gromov, M., Hyperbolic groups. Essays in group theory, 75–263, Math. Sci. Res. Inst. Publ., 8, Springer, New York, 1987.

[Gromov, 1993] Gromov, M. Asymptotic invariants of infinite groups. Geometric group theory, Vol. 2 (Sussex, 1991), 1–295, London Math. Soc. Lecture Note Ser., 182, Cambridge Univ. Press, Cambridge, 1993.

[Hatcher] Hatcher, A., Algebraic Topology, http://www.math.cornell.edu/ hatcher/AT/ATpage.html

[Linial and Meshulam, 2005a] Linial N. and Meshulam R., Homological connectivity of random 2-complexes, to appear in Combinatorica.

[Meshulam and Wallach, 2006] R. Meshulam, R. and Wallach. N. Homological connectivity of random k-dimensional complexes. math.CO/0609773 available at http://front.math.ucdavis.edu/math.CO/0609773

[Ollivier, 2005] Ollivier Y. A January 2005 Invitation to Random Groups.

[Ollivier, 2006] Ollivier Y. Sharp phase transition theorems for hyperbolicity of random groups. http://front.math.ucdavis.edu/math.GR/0301187

[Papasoglu, 1996] An algorithm detecting hyperbolicity. in G. Baumslag (ed.) et al., Geometric and computational perspectives on infinite groups DIMACS Ser. Discrete Math. Theoret. Comput. Sci., **25**, (1996) 193–200.

[Zùk, 2003] Property (T) and Kazhdan constants for discrete groups, GAFA, Geom. Funct. Anal. 13 (2003), **3**, 643–670.

Statistical persistent homology

Peter Bubenik
(joint work with Gunnar Carlsson, Peter T. Kim, Zhiming Luo)

We are interested in the following problem. We have a sample of points $X_1, \ldots, X_N$ from some manifold, which we assume was obtained, with errors, from some probability density. We would like to use the sample to recover, as best as possible, the topology of the underlying density.

First we need to say what we mean by the topology of the underlying density. Let $(\mathbb{M}, g)$ be a compact, connected Riemannian manifold, with a dominating measure ν (the invariant measure which in local coordinates is $\sqrt{|g|}dx_1...dx_d$). Let $f : \mathbb{M} \to [0, \infty]$ such that $\int_{\mathbb{M}} f d\nu = 1$. This probability density gives an increasing filtration of $\mathbb{M}$ by sublevel sets

$$\mathbb{M}_{f \leq r} = \{x \in \mathbb{M} \mid f(x) \leq r\}.$$

This induces an increasing filtration on $C_*(\mathbb{M})$, the *Morse filtration*: $\mathcal{F}_r(C_*(\mathbb{M})) = C_*(\mathbb{M}_{f \leq r})$, from which we can calculate the *persistent homology* [3, 4]. This is what we earlier referred to as the topology of f. Of course, the study of other invariants on $\mathbb{M}_{f \leq r}$ is possible.

In order that we may do statistics, we need to assume that $f = f_\theta$ belongs to a family of densities

$$\{f_\theta \mid \theta \in \Theta\}$$

where θ is the *parameter* and Θ is the *parameter space* which can be finite dimensional (the parametric case), or, infinite dimensional (the nonparametric case).

Our goal is to find an estimate $\hat{\theta}$ of θ so that the persistent homology of $f_{\hat{\theta}}$ is close to the persistent homology of f_θ.

For a basic example, consider the following unimodal densities on the unit sphere S^{p-1}, called von Mises-Fisher densities:

$$f_{\mu,\kappa}(x) = c_\kappa \exp\{\kappa x^t \mu\},$$

where $\mu \in S^{p-1}$, $\kappa \in [0, \infty)$, and c_κ is a normalizing constant. Using the Morse filtration on $C_*(S^{p-1})$ with $f_{\mu,\kappa}$, we see that the only non-trivial persistent homology occurs in degrees 0 and $p-1$. Here we encode the persistent homology in a set of intervals called a *barcode*:

$$\beta_0(f_{\mu,\kappa}) = \{[c_\kappa e^{-\kappa}, \infty]\} \text{ and } \beta_{p-1}(f_{\mu,\kappa}) = \{[c_\kappa e^{\kappa}, \infty]\}.$$

Let $X = (X_1, X_2, \ldots X_n)$ be a sample on S^{p-1} according to the von Mises-Fisher density $f_{\mu,\kappa}$. Let $\bar{X} = \frac{1}{n}\sum X_i$ be the sample mean. The maximum likelihood estimators are given by:

$$\hat{\mu} = \frac{\bar{X}}{||\bar{X}||} \text{ and } \hat{\kappa} = A_p^{-1}(||\bar{X}||), \text{ where } A_p(\lambda) = \frac{I_{\frac{p}{2}}(\lambda)}{I_{\frac{p-1}{2}}(\lambda)},$$

with I_ν denoting the the modified Bessel function of the first kind and order ν. Using the estimator $\hat{\kappa}$ we obtain the following estimates for the persistent homology

of the Morse filtration in degrees 0 and $p-1$:

$$\beta_0(f_{\hat{\mu},\hat{\kappa}}) = \{[c_{\hat{\kappa}}e^{-\hat{\kappa}}, \infty]\} \text{ and } \beta_{p-1}(f_{\hat{\mu},\hat{\kappa}}) = \{[c_{\hat{\kappa}}e^{-\hat{\kappa}}, \infty]\}.$$

The large sample asymptotics are given by

$$\sqrt{n}(\hat{\kappa} - \kappa) \to_d N(0, \tfrac{1}{A_p'(\kappa)}) \quad \text{as} \quad n \to \infty,$$

where N denotes the normal distribution. To measure how close our estimators $\beta_i(f_{\hat{\kappa}})$ to the true parameter $\beta_i(f_\kappa)$, we use the least symmetric difference metric on the space of barcodes. We obtain [1]

$$\mathbb{E}(\mathcal{D}(\beta_i(f_{\hat{\mu},\hat{\kappa}}), \beta_i(f_{\mu,\kappa}))) \leq \frac{c_\kappa}{\sqrt{n}}.$$

Now let us consider the more interesting nonparametric case. For simplicity, take $f : \mathbb{M} \to \mathbb{R}$ to be a Morse function with distinct critical values. Here we encode the persistent homology using the *persistence diagram* [2]. Let $D_p(f)$ denote the persistence diagram for the degree p persistent homology of $C_*(\mathbb{M})$ filtered using the sublevel sets of f. Let d_B denote the *bottleneck distance* [2] between persistence diagrams.

The bridge between topology and statistics is provide by the Stability Theorem [2]. Take f to be an unknown function and $\hat{f}$ to its statistical estimator. Then, by the Stability Theorem,

$$d_B(D_p(f), D_p(\hat{f})) \leq \|f - \hat{f}\|_\infty.$$

We would like to be able to solve the following *Data Analysis Problem*. Let $X_1, \ldots, X_N$ be data sampled (with noise) from an unknown probability density f on a compact manifold $\mathbb{M}$. Find an estimator $\hat{f}$ that minimizes $\|f - \hat{f}\|_\infty$ and calculate the asymptotics as $N \to \infty$.

For now, we consider the closely related *Gaussian White Noise model*, which is easier to work with. It remains to show that the two problems are equivalent. Let $\mathbb{M}$ be a compact, connected m-dimensional Riemannian manifold with Riemannian metric $\rho(\cdot,\cdot)$ and volume element $d\omega$. Consider the Hölder class of functions on $\mathbb{M}$,

$$\Lambda(\beta, L) = \left\{ f : \mathbb{M} \to \mathbb{R} \;\mid\; |f(x) - f(y)| \leq L\rho(x,y)^\beta, x, y \in \mathbb{M} \right\},$$

where $0 < \beta \leq 1$.

Our result is the following. In the white–noise model,

$$\mathbb{E}d_B(D_p(\hat{f}_\varepsilon), D_p(f)) \leq C\psi_\varepsilon,$$

as $\varepsilon \to 0$, where $\psi_\varepsilon = \varepsilon^2 \log(\varepsilon^{-2})$ and $C = L^{m/(2\beta+m)} \left(\frac{(\beta+m)m^2}{\mathrm{vol}(S^{m-1})\beta^2} \right)^{\frac{\beta}{(2\beta+m)}}$.

References

[1] Peter Bubenik and Peter T. Kim. A statistical approach to persistent homology. *Homology, Homotopy and Applications*, 9(2):337–362, 2007.

[2] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. In *SCG '05: Proceedings of the twenty-first annual symposium on Computational geometry*, pages 263–271, New York, NY, USA, 2005. ACM Press.

[3] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. *Discrete Comput. Geom.*, 28(4):511–533, 2002. Discrete and computational geometry and graph drawing (Columbia, SC, 2001).
[4] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete Comput. Geom.*, 33(2):249–274, 2005.

Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition

Gunnar Carlsson

It is well understood that the analysis of high dimensional data sets of various kinds is a task which is of crucial importance to modern sciences and engineering. Data is being produced at a rate which is much greater than our capacity to analyze it, and furthermore traditional methods for analysis are often inadequate. One particularly useful set of methods which has been developed are the *geometric* ones, i.e. methods which rely on the introduction of a metric, or notion of distance, on the set of data points. The distance is often reflecting an intuitive notion of similarity, in that points which are a small distance apart are regarded as similiar. Methods based on these ideas are the multidimensional scaling and locally linear embedding methods, which find low dimensional visual representations of the data in question, and R. Coifman's notions of *diffusion distance* and *diffusion mapping*. Inducing a metric can then permit the introduction of geometric methods from pure mathematics which can be quite useful in making sense of the data. It is then interesting to ask what methods from geometry are particularly well adapted to data geometries. In considering this question, it is useful to keep in mind the following two points about distances frequently used in studying data.

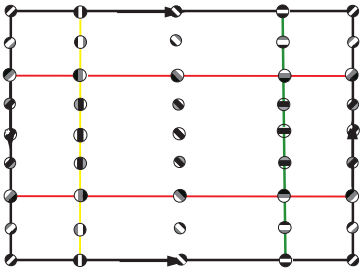
- Although short distances reflect a notion of similarity well, it is often the case that long distances are not particularly well related to the problem in question. For example, when comparing two genomic sequences, they are defined to be close if one can be obtained from the other by a small number of substitutions. On the other hand, knowing that two sequences are related by a particular large number of substitutions is not so useful.
- Quantitative comparison of short distances is often not so significant either.

In the above image, two comparisons of pairs of nearby points which are far apart within the space does not, for the most part, suggest that one pair is reliably more similar than the other.

These observations suggest that topological methods should be useful in data analysis, since topology studies qualitative properties of metric spaces by ignoring quantitative comparisons between distances, and instead studies only the notion

of a point being infinitely close to a subset. In this talk, we described three specific ways in which topological methods can be imported into the analysis of data.

(1) Homological signatures can be inferred using the method of persistent homology [4] and [7]. This technique was used to study particular examples, one from natural image statistics [3] and one from neuroscience [6]. The method involves the study and classification of a family of diagrams of vector spaces, called persistence diagrams. The method can be used to produce a map describing the topological type of various subsets of data sets, perhaps those of high density, as determined by a choice of density estimator and a threshold for density. Below is a map obtained for the case of Mumford's data set of high contrast 3×3 image patches in natural images.



It shows how representations of patches can be mapped out to lie on a Klein bottle. The rectangular picture corresponds to the usual identification space model of a Klein bottle.

(2) Simplicial constructions from topology can be used to devise methods for visualization which in many situations are more flexible and sensitive than standard methods which are more strongly metric dependent. This methodology, whose software realization is called Mapper, was introduced in [5], and was used in [1] to understand some questions in RNA folding. Here is a picture which is output from Mapper.

The loop present in the middle of this diagram turns out to yield potentially different folding mechanisms. This example illustrates the fact that this method is capable of finding small features within a larger data set. Sometimes these features will turn out to be artifacts, but this can only be determined by returning to examination of the data.

(3) Topological methods can be introduced using different diagrams, so-called zig-zag diagrams. These diagrams turn out to be interpretable as representations of quivers, in fact of quivers of finite representation type. They can shed light on the stability and consistency of qualitative invariants (such as the cluster decompositions) and the presence of loops, as well as

provide a useful summary for the understanding of dynamical clustering. This is discussed in [2].

References

[1] Bowman, G. R.; Huang, X.; Yao, Y.; Sun, J.; Carlsson, G.; Guibas, L. J.; Pande, V. S., *Structural insight into RNA hairpin folding intermediates* J. Am. Chem. Soc., 2008.

[2] G. Carlsson, *Topology and data*, to appear, Bulletin of American Mathematical Society.

[3] G. Carlsson, T. Ishkhanov, V. de Silva, and A. Zomorodian, *On the local behavior of spaces of natural images*, International Journal of Computer Vision, (76), 1, 2008, pp. 1-12.

[4] H. Edelsbrunner, D. Letscher, and A. Zomorodian, *Topological persistence and simplification*, Discrete and Computational Geometry **28**, 2002, 511-533

[5] G. Singh, F. Memoli and G. Carlsson, *Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition*, Point Based Graphics 2007, Prague, September 2007

[6] G. Singh, F. Memoli, T. Ishkhanov, G. Carlsson, G. Sapiro and D. Ringach, *Topological Structure of Population Activity in Primary Visual Cortex*, Journal of Vision, vol. 8, 8, 2008, pp. 1-18.

[7] A. Zomorodian and G. Carlsson, *Computing persistent homology*, Discrete and Computational Geometry, **33** (2), 2005, pp. 247-274

Zig-zag persistence

Vin de Silva

(joint work with Gunnar Carlsson)

1. Discrete invariants

It is well understood how to assign discrete algebraic invariants to continuous topological spaces; this is the art of classical algebraic topology. These invariants capture and quantify the intuitive topological properties of common objects: the hole 'in' a bagel; the cavity inside a tennis ball; the one-sided nature of a Möbius band; the triplewise but not pairwise linking of the Borromean rings.

It is less well understood how to assign algebraic invariants to *statistical* spaces; by which I mean finite samples taken from a probability distribution, perhaps concentrated on a simplicial complex or submanifold of some Euclidean space. Consider the simplest possible topological invariant, b_0, which counts the connected components of a topological space. The corresponding statistical quantity is the number of *clusters* of a sampled distribution. This is much more slippery than b_0; in fact there exists a huge literature on clustering in statistics and machine learning, and even an impossibility theorem, due to Kleinberg [1].

Why are connected components easy, and clusters difficult? A simple explanation is that the class Top of continuous topological spaces is *discrete*, whereas the class Stat of discrete statistical spaces is *continuous* in character. (One can evolve continuously from a data set with two separate clusters to a data set which has only one cluster; whereas to reach a surface of genus 2 from a surface of genus 1 requires a quantum leap.) Thus, discrete invariants (such as b_0) occur naturally

on TOP, but are necessarily discontinuous when defined on STAT. Such invariants are unstable.

2. Persistent homology

The theory of persistence provides a mechanism for overcoming these difficulties. We wish to assign homological invariants (which are discrete) to statistical spaces (which vary continuously). The way around this is to decorate the homological invariants with a continuous parameter. For instance, let $X \subset \mathbb{R}^n$ be a finite set. Replacing each point by a ball of radius ϵ, we can define $X^\epsilon = \bigcup_{x \in X} B_x(\epsilon)$ and homology groups $H^\epsilon = H_*(X^\epsilon)$. Whenever $\epsilon \leq \delta$ there is an inclusion $X^\epsilon \subseteq X^\delta$, and hence an induced map $H^\epsilon \to H^\delta$. Thus we have a direct system $\{H^\epsilon\}$ which we can regard as the primary homological invariant of the point cloud.

The advantage of this framework is that homological cycles (or *features*) come labelled with a real-valued lifetime: given $\alpha \in H^\epsilon$ its remaining lifetime is defined as $\inf\{\delta \geq \epsilon \mid \alpha \in \ker(H^\epsilon \to H^\delta)\}$. A feature which persists over a long interval $[\epsilon, \delta)$ is by default taken more seriously than a feature which perishes quickly. Short-lived features are more likely to be artifacts of the construction or of noise in the data. If X is a topologically ambiguous configuration, then this may be reflected by the presence of features whose lifetime is not clearly short or long.

Two ingredients make these ideas effective: a concise description of the structure of the directed system $\{H^\epsilon\}$, and an algorithm for computing this description. Such a description and algorithm first appeared in [4], with a more general algebraic treatment in [5]. The compact description comes from the fact that $\{H^\epsilon\}$ decomposes as a direct sum of one-dimensional summands, each comprising a single cycle α_i which appears at time b_i (birth) and disappears at time d_i (death). Moreover, the summands are canonical, in the sense the the collection of pairs (b_i, d_i) is independent of the splitting. This description requires that we work with field coefficients, and we also assume that the vector spaces H_ϵ are finite dimensional (to avoid obvious pathology).

The multiset of pairs (b_i, d_i) can be regarded as a collection of half-open intervals $[b_i, d_i)$ (the barcode description) or else as a subset of $\mathbb{R}^2$ lying above the diagonal $x_1 = x_2$ (the persistence diagram). It is now known that this object is stable under perturbations of the underlying construction under very mild hypotheses [6].

The complete protocol for topological measurement can be summarised as follows. Given a statistical space X, represent the spaces X^ϵ as simplicial complexes $\mathrm{Cech}^\epsilon(X)$ which are the nerves of the covering $X^\epsilon = \bigcup_x B_x(\epsilon)$; then apply the persistence algorithm to $\{\mathrm{Cech}^\epsilon(X)\}$ to obtain a barcode description of the family $\{H^\epsilon\}$. There are several alternatives to the Cech construction (such as Rips complexes, witness complexes [7]) but all have the same qualitative properties for the purposes of this discussion.

The idea of using persistence in statistical topology appears to have occurred independently several times. Two early references are [2] and [3]. In the last few years there has been considerable work in this area. One recent example is a response [8] to Kleinberg's impossibility theorem.

3. Zig-zag persistence

Persistent homology as described in [4, 5] works only for nested (or at least direct) systems of spaces $\{X^\epsilon\}$. We propose a method for extending this paradigm to more general 1-parameter families of spaces $\{X^t\}$ derived from statistical data. Our main technical support comes from the theory of quivers. We note in passing that 2-parameter persistence is known to be a hard problem [9], even in the simplest case of a bifiltration: there is provably no concise description of a system of vector spaces of the form $\{H^{\delta,\epsilon}\}$.

Consider a sequence of topological spaces $X^0, \ldots, X^n$ and maps $p_1, \ldots, p_n$. Each $p_i = f_i$ or g_i takes one of two possible forms: $f_i : X^{i-1} \to X^i$ (forwards) or $g_i : X^i \to X^{i-1}$ (backwards). For instance if every p_i maps forwards then we recover ordinary persistence, discretized at the values $\epsilon = 0, \ldots, n$.

Applying the homology functor (with field coefficients) produces a diagram of vector spaces $\{V^i = H_*(X^i)\}$. In the language of quiver theory, this is precisely a quiver representation of type A_n. As with ordinary persistence, such representations have a canonical decomposition into summands. Each interval $[i, j] \subset \{0, \ldots, n\}$ corresponds to a summand type $I_{i,j}$ which has one-dimensional components over the indices $i, \ldots, j$ connected by isomorphisms, and which is zero elsewhere. This structure theorem can be proved directly, but can also be seen from the higher vantage points of Gabriel's theorem (which classifies quivers with a well-behaved decomposition theory) or Kac's theorem (which asserts that the decomposition theory is in some sense independent of the directions of the maps p_i). See [10] for a general introduction to quivers.

Because the decomposition is canonical (up to isomorphism), the summands of an A_n-diagram of vector spaces can be calculated explicitly using greedy induction along the index i. The situation is more involved when the vector spaces are presented homologically, but the principle remains the same.

Now consider a 1-parameter family of finite sets $X_t \subset \mathbb{R}^m$. To define persistence over $\{X_t\}$, we discretize t to a finite set of parameter values $0, \ldots, n$ and construct spaces $Y_i = (X_i)^\epsilon$ for a fixed value of ϵ. There are no natural maps between the Y_i, so we introduce interpolating spaces $Z_i = Y_{i-1} \cap Y_i$. The inclusion maps $Z_i \to Y_{i-1}$, $Z_i \subset Y_i$ lead to a diagram with alternating arrows (or *zig-zag*)

$$Y_0 \leftarrow Z_1 \to Y_1 \leftarrow \cdots \to Y_{n-1} \leftarrow Z_n \to Y_n.$$

Applying the homology functor gives a quiver representation which can be decomposed into summands, giving us the desired barcode description.

When the Y_i are represented as Čech complexes $\mathrm{Cech}^\epsilon(X_i)$, each Z_i is most naturally represented as a bisimplicial complex $\mathrm{Cech}_2^\epsilon(X_{i-1}, X_i)$ contained in the Cartesian product $\mathrm{Cech}^\epsilon(X_{i-1}) \times \mathrm{Cech}^\epsilon(X_i)$. There are analogous bicomplexes for the Rips and witness constructions.

References

[1] J. Kleinberg, *An impossibility theorem for clustering*, Advances in Neural Information Processing Systems (NIPS) **15** (2002).

[2] P. Frosini, *Discrete computation of size functions*, Journal of Combinatorics, Information & System Sciences, **17** 3-4 (1992), 232–250.
[3] V. Robins, *Toward computing homology from finite approximations*, Topology Proceedings **24** (1999), 503–532.
[4] H. Edelsbrunner, D. Letscher, A. Zomorodian, *Topological persistence and simplification*, Discrete and Computational Geometry **28** (2002), 511–533.
[5] A. Zomorodian, G. Carlsson, *Computing persistent homology*, Discrete and Computational Geometry **33** (2005), 249–274.
[6] D. Cohen-Steiner, H. Edelsbrunner, J. Harer, *Stability of persistence diagrams*, Discrete and Computational Geometry **37** (2007), 103–120.
[7] V. de Silva, G. Carlsson, *Topological estimation using witness complexes*, Eurographics Symposium on Point-Based Graphics (2004).
[8] G. Carlsson, F. Memoli, *Persistent clustering and a theorem of J. Kleinberg*, `http://comptop.stanford.edu/preprints/` (unpublished manuscript).
[9] G. Carlsson, A. Zomorodian, *The theory of multidimensional persistence*, Proceedings 23rd ACM Symposium on Computational Geometry, (2007).
[10] H. Derksen, J. Weyman, *Quiver representations*, Notices of the AMS **52** (2005), 200–206.

Symbolic Dynamics via Computational Topology

Sarah Day

With recent advances in computing power, numerical studies of nonlinear dynamical systems have become increasingly more popular. However, errors inherent to such studies may obscure the dynamics or, at the very least, raise doubts about the existence of numerically observed structures. Furthermore, unstable behavior, an intrinsic element of complicated systems, may be difficult to track even with very careful numerical work. I will discuss techniques based on computational algebraic topology, and in particular Conley index theory, which allow for the rigorous detection of dynamical structures of various stability types.

I will focus on the application of these techniques to discrete-time dynamical systems given as continuous maps $f : X \to X$. The phase space X may be finite or infinite dimensional, so long as it may be discretized in a natural way. Two examples are 2-dimensional Euclidean space $\mathbb{R}^2$ and the infinite-dimensional function space L^2. The maps themselves may also take many forms, but obtaining rigorous verification of the results requires the computation of bounds for truncation and other errors inherent to the numerical study. These errors reflect a necessary loss of information resulting from viewing the system at a fixed resolution required for a numerical study. The well-studied Hénon map and the Kot-Schaffer integrod-ifference population model from ecology are two examples of discrete models for which discretization and the corresponding error bounds may be computed.

Sample results for the Hénon and Kot-Schaffer models illustrate the use of the topological tools. Our goal is to produce a semi-conjugacy between the studied system and a constructed symbolic dynamical system. The symbolic system then serves as a (partial) catalogue of important structures called *invariant sets*. In joint work with O. Junge and K. Mischaikow on the Kot-Schaffer model ([1]), we used the techniques to locate (up to machine precision) and rigorously verify the existence of unstable periodic orbits, connecting orbits, and more complicated

invariant sets–one with positive topological entropy (one measurement of complexity) and another described by chaotic symbolic dynamics. In more recent work with R. Frongillo and R. Treviño on the Hénon map ([3]), we automated both the initial numerical work and the processing of the Conley index information in order to construct a semi-conjugate symbolic dynamical system on many symbols, yielding a high lower bound on the topological entropy.

In this work, the constructed symbolic dynamical system, or *subshift of finite type*, is specified by a directed graph. Verifying that all allowed periodic symbol sequences (cycles in the directed graph) correspond to orbits in the original system requires both computation and interpretation of the Conley index. Continuity and compactness arguments allow us to extend the result from all cycles in the graph to all infinite paths, thereby proving the semi-conjugacy to the constructed symbolic system. Since the directed graphs we are interested in often contain infinitely many cycles, there may be an infinite list of indices to check. Obtaining the results in [3] relied on developing and automating a procedure for reducing this infinite set of computations to a finite list.

Related topics and possible extensions include:

(1) *Detecting spurious solutions* – studying the relationship between the numerics/simulations alone and the topological information given by the Conley index.

(2) *Optimization* of the algorithms for automated construction of subshifts of finite type on many symbols and topological entropy bounds.

(3) *Applications* to broader classes and families of discrete dynamical systems and interpretation of the results.

References

[1] S. Day, O. Junge, and K. Mischaikow, *A rigorous numerical method for the global analysis of infinite-dimensional discrete dynamical systems*, SIAM J. Appl. Dyn. Syst., **3**(2) (2004), 117–160.

[2] S. Day, O. Junge, and K. Mischaikow, *Towards automated chaos verification*, Proceeding fo Equadiff 2003, (2005), 157–162.

[3] S. Day, R. Frongillo, and R. Treviño, *Algorithms for rigorous entropy bounds and symbolic dynamics*, to appear in SIAM J. Appl. Dyn. Syst., (2008).

Euler Calculus

Robert Ghrist

(joint work with Yuliy Baryshnikov)

Consider the problem of using a network of sensors to count targets. For reasons of power conservation and simplicity in design, sensors are assumed to be minimal: targets cannot be localized or identified, but merely counted. This presents the

problem of aggregating all the redundant data over the network in order to return an accurate count.

We formalize the problem as follows. Fix an o-minimal structure on subsets of Euclidean space [10] (e.g., one may restrict to piecewise-linear, or semi-algebraic, or subanalytic sets): these are fixed as the "tame" sets in what follows. Assume that W is the **target space**, where a finite number of targets T_α reside. Assume a dense collection of counting sensors, parametrized as a **sensor space** X. The **sensing relation** is the subset $S \subset W \times X$ given by $S = \{(w, x) : \text{a target at } w \text{ is seen by a sensor at } x\}$. The sensor field on X returns a **counting function** $h : X \to \mathbb{N}$ given by $h(x) =$ the number of targets seen by sensor x.

Each target T_α activates some collection of sensors $U_\alpha \subset X$ — the fiber of S over T_α under projection $W \times X \mapsto W$. Call this fiber the **target support**. The following theorem gives a means of enumerating targets with minimal assumptions on the shape or size of the target supports.

Theorem 1 (Baryshnikov-Ghrist)**.** *Given $h : X \to \mathbb{N}$ the counting function for $\{U_\alpha\}$ a collection of tame target supports in X with uniform Euler characteristic: $\chi(U_\alpha) = N \neq 0$ for all α. Then*

$$\#\alpha = \frac{1}{N} \int_X h \, d\chi. \tag{1}$$

The theorem and proof uses a calculus based on Euler characteristic χ viewed as a measure $d\chi$. The history of this calculus is convolved: it appears in inchoate form in the works of Blaschke [2], Hadwiger [6], Groemer [5], and Rota [7], all of which view the Euler characteristic as a scale-invariant valuation on polyhedra. In the 1980s, Viro [11] (based on MacPherson's work on sheaves) and Schapira [8] (based on Kashiwara's work on sheaves) independently formulated a rich integral calculus based on Euler characteristic. Most recently, the Euler calculus has been rediscovered as a foreshadowing of **motivic integration** [3, 4].

A few definitions suffice to set up the calculus. By the o-minimal Triangulation Theorem [10], any tame set admits a partition into the homeomorphic image of a (not necessarily closed) simplicial complex. The **geometric Euler characteristic** of a tame space is defined as

$$\chi(A) = \sum_\sigma (-1)^{\dim \sigma} = \sum_{k=0}^{\infty} (-1)^k \dim H_k^{BM}(A), \tag{2}$$

where σ denotes the simplices in a triangulation of A and H_k^{BM} denotes the **Borel-Moore homology** of A. The geometric Euler characteristic is a homeomorphism invariant.

Let $CF(X)$ denote the class of **constructible functions** on X: functions $h : X \to \mathbb{Z}$ of compact support whose level sets are tame. The **Euler integral** is the pushforward $\int_X \cdot \, d\chi : CF(X) \to CF(\{pt\}) \cong \mathbb{Z}$ induced by the trivial map $X \to \{pt\}$. It has an explicit construction: any $h \in CF(X)$ can be expressed as $h = \sum_\alpha c_\alpha \mathbb{1}_{\sigma_\alpha}$ for $\{\sigma_\alpha\}$ a collection of (sets homeomorphic to open) simplices and $\{c_\alpha\}$ coefficients. Then $\int_X h \, d\chi = \sum_\alpha (-1)^{\dim \sigma_\alpha} c_\alpha$.

As this integral operator is expressed as a pushforward, it respects all the gluing and restriction operations implicit in sheaves of constructible functions: it really is an integration theory. To wit, since $\chi(A \cup B) = \chi(A) + \chi(B) - \chi(A \cap B)$, one has that $\int_X h \, d\chi$ is invariant under how h is decomposed [11, 8]. Much more is knowable and known, including a Fubini theorem, convolutions, integral transforms, and duality [9, 3].

Given this calculus, the proof of Theorem 1 is trivial.

Proof.

$$\int_X h \, d\chi = \int_X \left(\sum_\alpha \mathbb{1}_{U_\alpha} \right) d\chi = \sum_\alpha \int_X \mathbb{1}_{U_\alpha} \, d\chi = \sum_\alpha \chi(U_\alpha) = N \, \#\alpha. \tag{3}$$

□

Unfortunately, sensors do not yet come in the continua Theorem 1 assumes: they come in discrete networks. Intuition from numerical analysis suggests that, given a discrete sampling of an integrand h, one should use the piecewise-linear interpolant h_{PL} as a good approximation to h. However, $\int_X \cdot \, d\chi$ does not take real-valued functions as an argument. Rota [7] and those following define integration for $\mathbb{R}$-valued integrands; however, in that theory, integrals of continuous functions always vanish. We propose the following definition.

Given a tame mapping $h : X \to \mathbb{R}$ (the graph of h is a tame set), define

$$\int_X h \lfloor d\chi \rfloor = \lim_{n \to \infty} \frac{1}{n} \int_X \lfloor nh \rfloor d\chi. \tag{4}$$

This limit exists, is well-defined, and gives a plethora of computational formulæ, including some deep relations to Morse theory — the integral of a Morse function is an alternating sum of critical values of h, graded by the Morse index. Unfortunately, the integral operator is no longer linear. In recompense, we do have the following:

Theorem 2 (Baryshnikov-Ghrist)**.** *Let $h \in CF(\mathbb{R}^n)$ be a constructible upper semi-continuous function whose chambers are all codimension-0 in $\mathbb{R}^n$. Then, for a sufficiently fine and regular triangulation of $\mathbb{R}^n$, the PL interpolation h_{PL} of h over the vertex set of the triangulation satisfies*

$$\int_{\mathbb{R}^n} h_{PL} \, d\chi = \int_{\mathbb{R}^n} h \, d\chi. \tag{5}$$

This allows one to begin the process of developing numerical analysis, signal processing, filters, and expectations for the topological integration theory of Euler calculus.

References

[1] Y. Baryshnikov and R. Ghrist, "Target enumeration via Euler characteristic integrals," to appear, *SIAM J. Appl. Math.*, preprint 2008.

[2] W. Blaschke, *Vorlesungen über Integralgeometrie*, Berlin, 1955.

[3] R. Cluckers and M. Edmundo, "Integration of positive constructible functions against Euler characteristic and dimension," *J. Pure Appl. Algebra*, 208(2), 2007, 691 - 698.
[4] R. Cluckers and F. Loeser, "Constructible motivic functions and motivic integration," Inventiones Mathematicae, 173(1), 2008, 23–121.
[5] H. Groemer, "Minkowski addition and mixed volumes," *Geom. Dedicata* 6, 1977, 141–163.
[6] H. Hadwiger, "Integralsätze im Konvexring," *Abh. Math. Sem. Hamburg*, 20, 1956, 136–154.
[7] G.-C. Rota, "On the combinatorics of the Euler characteristic," Studies in Pure Mathematics, Academic Press, London, 1971, 221–233.
[8] P. Schapira, "Operations on constructible functions," *J. Pure Appl. Algebra* 72, 1991, 83–93.
[9] P. Schapira, "Tomography of constructible functions," in proceedings of *11th Intl. Symp. on Applied Algebra, Algebraic Algorithms and Error-Correcting Codes*, 1995, 427–435.
[10] L. Van den Dries, *Tame Topology and O-Minimal Structures*, Cambridge University Press, 1998.
[11] O. Viro, "Some integral calculus based on Euler characteristic," Lecture Notes in Math., vol. 1346, Springer-Verlag, 1988, 127–138.

Surface Reconstruction

Joachim Giesen

(joint work with Tamal K. Dey, Bardia Sadri, Edgar A. Ramos, Madhusudan Manjunath)

Given a smooth closed surface Σ embedded into $\mathbb{R}^3$. Let $P \subset \Sigma$ be a finite sampling of Σ. The *surface reconstruction problem* asks to reconstruct Σ from P, i.e., to compute a piecewise linear surface Σ' interpolating P that is isotopic to Σ and geometrically close.

There are *sampling conditions* on P such that the surface reconstruction problem can be solved. A popular condition introduced by Amenta, Bern and Eppstein [1] is based on the medial axis of Σ. The *medial axis* $M(\Sigma)$ is the closure of the set of all maximal, empty, open balls in $\mathbb{R}^3$. Empty means that the ball does not intersect Σ, and maximal means the ball is not contained in another empty, open ball. The *local feature size* is the function $f : \Sigma \to \mathbb{R}_+, x \mapsto d(x, M(\Sigma))$, and P is an *ε-sampling* of Σ if for all $x \in \Sigma$ there is a $p \in P$ with $d(x, p) \leq \varepsilon f(x)$. In the following we require P to be an ε-sampling with sufficiently small $\varepsilon > 0$.

One way to get a isotopic reconstruction of Σ from P is based on the approximation of the distance function $d_\Sigma : \mathbb{R}^3 \to \mathbb{R}_+, x \mapsto d(x, \Sigma)$ by $d_P : \mathbb{R}^3 \to \mathbb{R}_+, x \mapsto d(x, P)$. The gradient of d_Σ at $x \in \mathbb{R}^3$ is defined as as follows: let $N(x) = \{y \in \Sigma \,|\, d(x, y) = d(x, \Sigma)\}$ and $B(x)$ be the smallest enclosing ball of $N(x)$ with center $c \in \mathbb{R}^3$ and radius $r \in \mathbb{R}_+$. If $x \notin \mathrm{conv}(N(x))$, then $\nabla_{d_\Sigma}(x) = \frac{x-c}{d_\Sigma(x)}$, where $\mathrm{conv}(N(x))$ is the convex hull of $N(x)$. If $x \in \mathrm{conv}(N(x))$, then x is called a *critical point* of d_Σ. The gradient of d_Σ can be approximated by the gradient ∇_{d_P} of d_P. It can be shown [3] that the critical points of d_P are either very close to Σ (in terms of ε) or close to $M(\Sigma)$. The two types of critical points can be separated algorithmically [3]. The *stable manifold* of a critical point c is the closure of all points that flow into c under the gradient flow of ∇_{d_P}. The union of all stable manifolds of critical points close to Σ essentially provide an isotopic reconstruction of Σ if $\epsilon > 0$ is small enough.

So far guarantees known for smoothly embedded surfaces cannot be extended to surfaces with sharp edges and corners. Empirically one observes that isotopic reconstruction is often possible in practice and that making a computed isotopic piecewise linear approximation Σ' to Σ tighter gives geometrically better results. A surface is called *tightly embedded* if any hyperplane cuts it into at most two pieces. Tight surfaces minimize the total absolute Gaussian curvature for the topological type of Σ. When aiming for a tighter surface this observation leads to the problem to minimize the total absolute Gaussian curvature of Σ' while keeping the topological type and the vertex set P. For a variant of this problem NP-hardness has been shown [2]. A probably weaker goal is to minimize the absolute Gaussian curvature locally, i.e., to move a vertex $p \in P$ such that absolute Gaussian curvature of Σ' at p is minimized. Note that the Gaussian and absolute Gaussian curvature of a piecewise linear surface Σ' is concentrated in the vertices of Σ'. The problem to minimize the absolute Gaussian curvature locally can be abstracted as follows: given a piecewise linear embedding C of $\mathbb{S}^1$ into $\mathbb{R}^3$ such that the vertices of C have rational coordinates. Find a point $q \in \mathbb{R}^3 \setminus C$ such that piecewise linear surface that contains one triangle with vertex q for every edge in C has minimum absolute Gaussian curvature. We have shown [4] that in general q is not constructible, i.e., there exists no finite sequence of expressions $(E_1, \ldots, E_l)$ starting with the coordinates of the vertices of C, where E_i is obtained from previous expressions by addition, subtraction, multiplication, division or taking of k'th roots. We implicitly construct a polygon C whose vertices have rational coordinates and for which a unique minimizer q exists, and show that the coordinates of q are not constructible.

References

[1] Nina Amenta, Marshall W. Bern, and David Eppstein. The crust and the beta-skeleton: Combinatorial curve reconstruction. *Graphical Models and Image Processing*, 60(2):125–135, 1998.

[2] Maike Buchin and Joachim Giesen. Minimizing the total absolute gaussian curvature in a terrain is hard. In *CCCG*, pages 195–198, 2005.

[3] Tamal K. Dey, Joachim Giesen, Edgar A. Ramos, and Bardia Sadri. Critical points of distance to an epsilon-sampling of a surface and flow-complex-based surface reconstruction. *Int. J. Comput. Geometry Appl.*, 18(1/2):29–61, 2008.

[4] Joachim Giesen and Madhusudan Manjunath. Minimizing absolute gaussian curvature locally: Algebraic hardness and approximation in value. 2008.

Local persistence: homotopy theory of filtrations

J.F. Jardine

Suppose that a, b are real numbers, and that $a < b$. A filtration of a simplicial set X is a simplicial set-valued functor

$$F_\bullet(X) : [a, b] \to s\mathbf{Set}$$

defined on the poset $[a, b]$, with $t \mapsto F_t(X)$ for $a \leq t \leq b$, such that

- $F_t(X) = \varinjlim_{s<t} F_s(X)$ for all $t \in [a,b]$,
- $F(b) = X$.

Examples include the Čech and Rips complex constructions, defined on an interval $[a,b]$ where a is a small positive number and b is large. In persistence applications, $F_a(X)$ is a discrete set of points, $F_b(X) = X$ is a big contractible space, and the vertex functor $F_t(X)_0 = X_0$ is constant.

The idea of "persistent homotopy theory" is to find homotopical features of the filtration steps $F_t(X)$ that survive (or are defined) along subintervals of $[a,b]$. Persistent homology theory concerns the functors $t \mapsto H_*(F_t(X))$.

For a filtration $F_\bullet(X)$, if $I \subset [a,b]$ is a subinterval, say that a persistent homology class on I in $H_n(F_\bullet(X))$ is a family of classes $\alpha_t \in H_n(F_t(X))$ such that $\alpha_t \mapsto \alpha_{t'}$ under the map

$$H_n(F_t(X)) \to H_n(F_{t'}(X))$$

for all $t \leq t'$ in I. In other words, a persistent homology class on I is an element of

$$\varprojlim_{t \in I} H_n(F_t(X)).$$

The assignment

$$I \mapsto \varprojlim_{t \in I} H_n(F_t(X))$$

defines a contravariant functor on the category $\mathrm{Int}[a,b]$ of open subintervals of $[a,b]$, with inclusions. One similarly defines persistent classes in $h(F_\bullet(X))$ on I for all homotopy functors h, such as homotopy groups.

Observe that there could be a non-zero element $\alpha \in \varprojlim_{t \in I} H_n(F_t(X))$ in the kernel of the map

$$\varprojlim_{t \in I} H_n(F_t(X)) \to H_n(F_s(X))$$

for some $s \in I$, in which case α would be persistently 0 on some open subinterval of I.

There are various ways to study filtrations $F_\bullet(X)$, or more generally to study functors $Y : I \to s\mathbf{Set}$ as homotopy theoretic objects, where I is some interval. Here are two related constructions:

1) The category $s\mathbf{Set}^I$ consists of all functors $Y : I \to s\mathbf{Set}$ taking values in simplicial sets and all natural transformations between them.

The injective model structure on $s\mathbf{Set}^I$ has for weak equivalences (respectively cofibrations) all maps $Y \to Z$ for which all maps $Y(t) \to Z(t)$, $t \in I$, are weak equivalences (respectively inclusions) of simplicial sets. Injective fibrations are defined by a right lifting property.

Application: to construct the homotopy inverse limit $\underleftarrow{\mathrm{holim}}_I X$ of an I-diagram $X : I \to s\mathbf{Set}$, take an injective fibrant model $j : X \to Y$ (ie. trivial cofibration such that Y is injective fibrant) and set

$$\underleftarrow{\mathrm{holim}}_I X = \varprojlim_I Y.$$

2) The open subintervals of $[a,b]$ are the subintervals of the form $[a,c)$, (d,e) and $(f,b]$, along with the empty set, and these are the objects of the poset $\mathrm{Int}[a,b]$ of open subintervals, where the morphisms are inclusions. Covering families of open subintervals are defined in the usual way, and we get a Grothendieck topology on $\mathrm{Int}[a,b]$. The corresponding sheaf category on $\mathrm{Int}[a,b]$ is equivalent to the standard sheaf category for the topological space $[a,b]$

Suppose that $X : [a,b] \to \mathbf{Set}$ is a set-valued functor defined on the poset $[a,b]$. Suppose that $I \subset [a,b]$ is an open subinterval. Define

$$\varprojlim X(I) = \varprojlim X|_I,$$

where $X|_I$ is the composite functor

$$I \subset [a,b] \xrightarrow{X} \mathbf{Set}.$$

Then the assignment $I \mapsto \varprojlim X(I)$ defines a presheaf $\varprojlim X$ on $\mathrm{Int}[a,b]$, and we have the following:

- $\varprojlim X$ is a sheaf on $\mathrm{Int}[a,b]$.
- Let $(\varprojlim X)_t$ be the stalk of the sheaf $\varprojlim X$ at $t \in [a,b]$. Then we have the following:
  1) $(\varprojlim X)_t \xrightarrow{\cong} \varinjlim_{s<t} X(t)$ if $t \in (a,b]$, and
  2) $(\varprojlim X)_a \cong X(a)$.

There is a model structure on the category $s\mathbf{Pre}(\mathrm{Int}[a,b])$ of simplicial presheaves on $\mathrm{Int}[a,b]$, for which a cofibration is an inclusion of simplicial presheaves, and a weak equivalence is a map $X \to Y$ which induces weak equivalences $X_x \to Y_x$ of simplicial sets in all stalks. Fibrations for this theory are called global (or injective) fibrations, and the weak equivalences are called local weak equivalences.

We have the following:

- If $Y : [a,b] \to s\mathbf{Set}$ is a functor, then $\varprojlim Y$ is a simplicial sheaf on $\mathrm{Int}[a,b]$.
- A weak equivalence (or cofibration) $Y \to Z$ in $s\mathbf{Set}^{[a,b]}$ induces a local weak equivalence (or cofibration)

  $$\varprojlim Y \to \varprojlim Z$$

  of simplicial sheaves on $\mathrm{Int}[a,b]$.
- If $F_\bullet(X) : [a,b] \to s\mathbf{Set}$ is a filtration of a space X, then

  $$(\varprojlim F_\bullet X)_t \simeq F_t(X).$$

Suppose that $Y = F_\bullet(X) : [a,b] \to s\mathbf{Set}$ is a filtration of a space X, and let $j : Y \to Z$ be a functorial injective fibrant model in $s\mathbf{Set}^{[a,b]}$. Then the induced map

$$j_* : \varprojlim\ Y \to \varprojlim\ Z$$

is a local weak equivalence of simplicial presheaves on $\mathrm{Int}[a,b]$, and the restricted map

$$j|_I : X|_I \to Z|_I$$

is an injective fibrant model for the restriction of X along the inclusion $I \subset [a,b]$, for each open subinterval of $[a.b]$.

There is a spectral sequence

$$E_2^{p,q} = \varprojlim{}_I^p(\pi_q Y) = \varprojlim{}_I^p(\pi_q Z) \Rightarrow \pi_{q-p}(\varprojlim{}_I Z),$$

involving the higher derived functors $\varprojlim_I^p$ of inverse limit.

The simplicial presheaf $\varprojlim Z$ on $\mathrm{Int}[a,b]$ may satisfy descent (there are good reasons for expecting this), but in general it has a functorial injective fibrant model

$$i : \varprojlim Z \to W$$

in the model structure for simplicial presheaves on $\mathrm{Int}[a,b]$. The map i induces weak equivalences $F_t(X) = Y(t) \xrightarrow{\simeq} W_t$ for all $t \in [a,b]$.

Finally, there is a descent spectral sequence with

$$E_2^{p,q} = H^p([a,b], \tilde{\pi}_q(\varprojlim Y)) \Rightarrow \pi_{q-p} W([a,b]).$$

where the E_2-term is composed of sheaf cohomology groups for the space $[a,b]$ with coefficients in homotopy group sheaves. This spectral sequence is global sections of a presheaf of spectral sequences on $\mathrm{Int}[a,b]$, for the homotopy groups of the spaces $W(I)$.

References

[1] .G. Goerss and J.F. Jardine, *Simplicial Homotopy Theory*, Progress in Mathematics **174**, Birkhäuser Verlag, Basel (1999).

[2] .F. Jardine, *Simplicial presheaves*, J. Pure Appl. Algebra **47**(1) (1987), 35–87.

Geometric random complexes

Matthew Kahle

(joint work with Gunnar Carlsson, Persi Diaconis)

There has recently been exciting activity in understanding the topology of point cloud data. However, there is a feeling that we are still in need of detailed null hypotheses to carefully compare with real world situations.

One step in this direction was some recent work by Steve Smale, et. al. on understanding what happens when we build the Čech complex on points uniformly sampled from a compact manifold embedded in Euclidean space. Their main result is to estimate the number of epsilon-balls necessary to take before one completely covers the manifold, and hence can compute the homology of the manifold by working with the nerve. This was further extended to the situation where the points are assumed to be sampled uniformly, but each with an unknown Gaussian error term. This second analysis is more difficult, since, in particular, the distribution does not have compact support.

A difference in our approach here is that we consider all possible values of epsilon, and not only the ones which are large enough that we can completely recover some underlying space. In practice, it is clear that one needs a robust,

multi-scale approach to understand, for example, the homology of point cloud data. For this reason, there is much current research centered around persistent homology. This is part of our motivation to study even small values of epsilon – to get a clearer picture of the "signature" of random points.

The main problem we study is the following. Take n points, sampled independently and identically distributed according to a standard normal distribution on $\mathbb{R}^d$. Connect pairs of points by an edge of their distance is less than $r(n)$, and build the Vietoris-Rips complex of this graph. The basic question is what we can say about the topology of such a complex as n approaches infinity. For example, what can one say about the expected size of the Betti numbers, etc.?

There are four main regimes of interest. These regimes have already been identified by random geometric graph theorists, particularly by Matthew Penrose in his book "Random geometric graphs." From our point of view here, only the 1-skeleton of our complex has been studied in detail, but we rely heavily on the detailed calculations that have already been done for subgraph counts, etc., in geometric random graphs.

Sparse regime: When the radius is small enough, the underlying graph consists of almost as many connected components as vertices. The f-vector is decreasing, and the dimension is finite, but increasing. Subgraphs of arbitrary large order start appearing. In particular, subgraphs representing the boundary of the $k+1$-dimensional cross-polytope appear for every k, and as isolated components. These represent nontrivial k-dimensional homology classes, and in fact they dominate k-dimensional homology.

What is perhaps surprising is that homology appears in arbitrarily high dimension, even when $d = 2$. Basically, the dimension of k-dimensional homology is growing as a power of n, and the exponent approaches one as the radius grows toward the critical regime.

Because we have good control over the structure of these complexes, and by using Poisson and normal approximation theorems for subgraph counts, we can prove strong laws of large numbers for the Betti numbers.

Another upshot is that although we show Betti numbers can grow arbitrarily large, we know exactly what the classes look like, so it is easy to see that the persistent homology is extremely small by factoring through the Čech complex.

Critical regime: This is sometimes called the thermodynamic limit, since it is where percolation appears. In this regime, we are able to continue the argument from above, and the expectation for every Betti number is growing linearly with n.

It is known that a giant component appears in this range. It would be interesting to know if there are higher dimensional analogues of the giant – it is plausible that they each have their own threshold here, even when $d = 2$.

Dense regime: Just as the Betti numbers increased as the radius approached the critical regime, now they start to decrease. In this regime, there is a spatial phase transition between "dense" and "sparse" regions. In the dense region, a discrete Morse theory argument shows that the complex is contractible. In the sparse

region, there are many isolated components, and we can repeat the argument from the sparse regime. A little care is required at the boundary between these regions.

Connected regime: Eventually the radius $r(n)$ is large enough that every vertex has at least one edge. In this same range, we see that the Čech complex is homotopy equivalent to a star-shaped domain in the plane. Similarly, one can use Morse theory to show that the Rips complex is contractible.

This seems to give a relatively complete picture of the rate of growth of the Betti numbers of Gaussian random complexes. Many of the techniques used here (and even some of the results, as stated) are applicable to any density function, and not just normally distributed points.

One notable feature of these results is that they actually tell us quite a lot about persistent homology. First of all, they give strong evidence that persistence should give us a much better picture than homology on its own, since the persistence bars should all be quite short. This is intuitively clear, but we are not aware of any results in that direction. Second, the results show that the most natural thing to check when measuring the length of a persistence bar is not the difference between the two endpoints, but the ratio. This might be the most interesting aspect of this research in terms of applying the ideas to real world implementations of computational homology.

Persistent homology and discrete Morse theory

Kevin P. Knudson

Let K be a finite simplicial complex equipped with a filtration

$$\emptyset = K_{-1} \subset K_0 \subset K_1 \subset \cdots \subset K_n = K.$$

We assume that each subcomplex K_{i+1} is obtained from K_i by attaching a single simplex: $K_{i+1} = K_i \cup \sigma_{i+1}$. In this talk we explore the relationship between the persistent homology pairing of Edelsbrunner–Letscher–Zomorodian [1] and Forman's discrete Morse theory [2]. In what follows, we use the notation $\alpha < \beta$ to indicate that α is a codimension-one face of β.

The basic idea behind persistence is the following. As simplices enter the complex K homology classes are created or destroyed. If a class α is born at filtration level i and dies at level j, we say the *persistence* of α is $j - i - 1$. This concept is useful in topological data analysis as homology classes that persist for long intervals may be considered to be real features of a point cloud while those that live for short intervals may be regarded as noise.

The persistence pairing is defined using homology with $\mathbb{Z}/2$-coefficients. Say the simplex σ has dimension k. When σ enters the complex, one of two events occurs. Either σ creates a new k-cycle, or it destroys a $(k-1)$-dimensional homology class. In the first case, we call σ *positive* and in the second case it is called *negative*. This allows one to pair the simplices of K in the following way. Given a simplex τ, consider the list of positive simplices in $\partial\tau$. Say σ is the youngest simplex in $\partial\tau$ in

the sense that it entered the filtration at the latest stage. If σ is not already paired with one of its cofaces, then we pair σ and τ. If σ is paired with some τ', then we may write $\partial\tau$ in terms of $\partial\tau'$ and some other positive simplices, and we then pair τ with the youngest in this list of positive simplices, unless it is already paired. We iterate this procedure until an unpaired positive simplex is found (which must occur since we always search backwards in the filtration). A complete description of the algorithm may be found in [1]. Denote the resulting collection of pairs $\{\sigma, \tau\}$ by P. Note that it is possible that σ is not a face of τ in such a pair.

Forman's discrete Morse theory is meant to mimic the smooth theory on an arbitrary simplicial complex (or more generally a regular cell complex). The most convenient formulation of this theory is in terms of discrete vector fields on the complex. A *discrete vector field* V on K is a collection of pairs of simplices $\{\alpha < \beta\}$ such that each simplex of K is in at most one pair in V. A V-path is a collection of simplices

$$\alpha_0 < \beta_0 > \alpha_1 < \beta_1 > \cdots > \alpha_r < \beta_r > \alpha_{r+1},$$

where each $\{\alpha_i < \beta_i\} \in V$. Such a path is called *closed* if $\alpha_{r+1} = \alpha_0$. The vector field V is a *gradient* if there are no closed V-paths. A simplex in K which does not appear in any pair in V is called *critical*, while those which are paired are called *regular*. The usual theorems of smooth Morse theory have their analogues in the discrete case; in particular, if V is a discrete gradient on K, then K has the homotopy type of a CW-complex with one cell of dimension p for each critical simplex of dimension p.

Now suppose we have a filtered complex as above, and we produce a persistence pairing P. For the pairs $\{\sigma, \tau\}$ in P, there are two possibilities: either $\sigma < \tau$ or σ is not a face of τ. Define a discrete vector field V_P on K by

$$\{\sigma, \tau\} \in V_P \Leftrightarrow \sigma < \tau \text{ and } \{\sigma, \tau\} \in P.$$

If a simplex is unpaired by P or if a simplex is paired with another simplex not containing it as a face, then the simplex is critical for V_P.

Theorem 1. *V_P is a gradient vector field on K.*

Note that simplices not included in P correspond to homology classes in the complex K. If two non-adjacent simplices are a pair in P, then the persistence of the associated homology class is probably somewhat large, and we may want to view them as defining some topologically interesting feature. There will be gradient paths between them, however, and we expect the following to be true.

Conjecture 2. *If $\{\sigma, \tau\} \in P$ are such that σ is not a face of τ, then there is a single gradient path from τ to σ in the vector field V_P.*

Since the cells σ and τ do not correspond to homology classes in K, the existence of a single gradient path joining them implies that we may cancel the cells (see [2]), thereby simplifying the vector field. In general, though, this procedure is computationally expensive, which we should expect in light of the result of Joswig and Pfetsch [3] which asserts that the construction of optimal discrete gradients

is NP-complete (the persistence pairing runs in polynomial time, so cancellation must be complicated).

Now, suppose V is a discrete gradient on K. A useful way to visualize this is to modify the Hasse diagram of K in following way. Recall that the Hasse diagram of K is the directed graph whose vertices correspond to the simplices of K, and there is an arrow from β to α if $\alpha < \beta$. We modify the Hasse diagram by reversing the arrows for every $\{\alpha < \beta\} \in V$. Since V is a gradient, the resulting directed graph G has no directed loops. There is therefore an integer-valued function f on the vertices of G which is strictly decreasing along every directed path. We may assume that the absolute minimum of f is 0. Given such an f, we define a filtration of K by setting K_i to be the subcomplex generated by $f^{-1}(-\infty, i]$. This may lead to multiple simplices entering at one time, but we may order the simplices in such a way that we may still talk about the persistence pairing associated to this filtration. We therefore obtain a persistence pairing $P_{\{V,f\}}$. Of course, this also depends on the choice of the function f, but we expect the following to hold.

Conjecture 3. *Suppose P is a persistence pairing on K. Then with the correct choice of f, we have $P_{\{V_P,f\}} = P$.*

Of course, we cannot expect to have $V_{P_{\{V,f\}}} = V$ for an arbitrary gradient V. Indeed, the easiest example of this is $V = \emptyset$, in which every simplex is critical. Since any associated persistence pairing $P_{\{V,f\}}$ will contain a pair of adjacent simplices (in particular an edge and one of its vertices), $V_{P_{\{V,f\}}} \neq \emptyset$.

References

[1] H. Edelsbrunner, D. Letscher, A. Zomorodian, *Topological persistence and simplification*, Discrete Comput. Geom. **28** (2002), 511–533.

[2] R. Forman, *Morse theory for cell complexes*, Adv. Math. **134** (1998), 90–145.

[3] M. Joswig, M. Pfetsch, *Computing optimal Morse matchings*, SIAM J. Discrete Math. **20** (2006), 11–25.

Combinatorial Algebraic Topology: Structures, Methods, and Applications

Dmitry Kozlov

Combinatorial Algebraic Topology is a contemporary field of mathematics which concerns itself with computing invariants in Algebraic Topology for **combinatorial cell complexes** by **combinatorial means.**

Here, by combinatorial cell complexes we mean CW complexes which are combinatorial *locally*, i.e., having simple cell attachments, for example - simplicial or prodsimplicial complexes, as well as *globally*, which means that their cells are indexed by various combinatorial objects, such as graphs, partitions, permutations, various combinations and taking boundary is described by a combinatorial rule for the indexing objects, for example - removing vertices in graphs, merging blocks in partitions, relabeling, etc.

We also say that the computation is done by combinatorial means if it is mostly done by using matchings, orderings, labelings, et cetera, to simplify or to completely eliminate algebraic computations and topological deformations.

We would like to distinguish the following aspects of this theory:

- discrete structures are derived as models for topological spaces,
- there exist many constructions of cell complexes with discrete structures as input,
- using standard Algebraic Topology tools uncovers ties between discrete structures,
- properties of discrete structures are distinguished by how well they behave with respect to Algebraic Topology.

There are also numerous applications stemming from the fact that the study of discrete structures yields information about the topology of the original spaces and vice versa. The main reference for this field is the recent monograph by the speaker, [4].

The talk was divided into three parts, which were called *structures*, *methods*, and *applications* respectively, which contained an example or two of each.

I. Structures

1. Acyclic categories.

A category is called **acyclic** if every non-identity morphism

- is not a loop
- has no inverse.

To an acyclic category C we associate a regular trisp $\Delta(C)$.

The **category of intervals** $I(C)$ (aka *factorization category, edgewise subdivision*) of an acyclic category C is defined by:

- the set of objects of $I(C)$ is the set of morphisms of C;
- for two objects m_1 and m_2 of $I(C)$, the set of morphisms from m_1 to m_2 in $I(C)$ is indexed by all pairs (α, β) such that $\alpha, \beta \in \mathcal{M}(C)$ and $m_2 = \beta \circ m_1 \circ \alpha$,

where the composition rule in $I(C)$ is the standard trapezoidal combination rule.

Theorem. *For any acyclic category C, the regular trisp $\Delta(I(C))$ is a subdivision of the regular trisp $\Delta(C)$, i.e., we have*

$$\Delta(I(C)) \rightsquigarrow \Delta(C).$$

2. Witness complex.

Definition. *The abstract simplicial complex $W(A, B)$, called the* **witness complex***, consists of all subsets $\sigma \subseteq B$ such that for every $\tau \subseteq \sigma$ there exists a point w in A that "witnesses" τ in the following sense: every point in τ is closer to w than every point in $B \setminus \tau$.*

The points in A are called *landmark points*, and the points in B are called *data points*. Usually one assumes that there are many more data points than landmark points.

Reformulation in the language of Combinatorial Algebraic Topology.

Every landmark point x induces an order on the data points: sort with respect to their distances to x. Every such ordering can be visualized as a path in the Hasse diagram of the Boolean lattice: start from the point nearest to the chosen landmark point, proceed to the union of the two closest ones, then to the union of the three closest ones, and so on. The witness complex $W(A, B)$ is the maximal abstract simplicial complex whose face poset is contained in the union of these paths.

II. Methods: Poset maps with small fibers

Definition. *A partial matching M on the Hasse diagram of a poset P is called* **acyclic** *if there does not exist a cycle*

$$b_1 \succ d(b_1) \prec b_2 \succ d(b_2) \prec \cdots \prec b_n \succ d(b_n) \prec b_1,$$

where $n \geq 2$, all $b_i \in P$ are distinct, and $(a_i, b_i) \in M$ for all i.

The unmatched elements are called **critical**, their set is denoted by $C(P, M)$.

Proposition. *A polyhedral complex Δ is collapsible if and only if the poset $\mathcal{F}(\Delta) \cup \{\hat{0}\}$ allows a complete acyclic matching.*

Definition. *A poset map $\varphi : P \to Q$ is said to have* **small fibers** *if for any $q \in Q$, the fiber $\varphi^{-1}(q)$ is one of the following:*

empty - a single element - two comparable elements.

The fibers of cardinality 2 yield a partial matching $M(\varphi)$.

Theorem. (DK, 2007) (Acyclic matchings via poset maps with small fibers) *For any poset map with small fibers $\varphi : P \to Q$, the partial matching $M(\varphi)$ is acyclic. Conversely, any acyclic matching on P can be represented as $M(\varphi)$ for some poset map with small fibers φ.*

Definition. *A poset map with small fibers $\varphi : P \to Q$ is called a* **collapsing order** *if φ is surjective as a set map, and Q is a chain.*

Definition. *For an acyclic matching M on a poset P we define the poset $U(P, M)$ as follows:*

- *the set of elements of $U(P, M)$ is $M \cup C(P, M)$;*
- *the partial order is the transitive closure of the elementary relations given by $S_1 \leq_U S_2$, for $S_1, S_2 \in U(P, M)$, if and only if $x \leq y$, for some $x \in S_1$, $y \in S_2$.*

Theorem. (DK, 2007) (Universality of $U(P, M)$)
For any acyclic matching M on a poset P, we have:

(1) *the partial order on $U(P, M)$ is well-defined;*
(2) *the induced quotient map $q : P \to U(P, M)$ is a poset map with small fibers;*
(3) *the linear extensions of $U(P, M)$ are in 1-to-1 correspondence with collapsing orders for M;*
this correspondence is given by the composition of the quotient map q with a linear extension map.

Definition. *A* **poset fibration** *(aka Grothendieck construction) is a pair $(B, \mathcal{F})$, where*

- *B is a poset - the* **base** *of the fibration;*
- *$\mathcal{F} = \{F_x\}_{x \in B}$ is a collection of posets, indexed by the elements of B - the* **fibers**.

The **total space** is the poset $E(B, \mathcal{F}) = \cup_{x \in B} F_x$, with the order relation: $\alpha \geq \beta$ in $E(B, \mathcal{F})$ if either $\alpha, \beta \in F_x$, and $\alpha \geq \beta$ in F_x, for some $x \in B$, or $\alpha \in F_x$, $\beta \in F_y$, and $x > y$ in B.

Theorem. (DK, 2007) (Decomposition theorem)
For an arbitrary poset fibration $(B, \mathcal{F})$, where $\mathcal{F} = \{F_x\}_{x \in B}$, and an arbitrary poset P, there is a 1-to-1 correspondence between

- *poset maps $\varphi : P \to E(B, \mathcal{F})$;*
- *pairs $(\psi, \{g_x\}_{x \in B})$, where ψ and g_x are poset maps $\psi : P \to B$ and $g_x : \psi^{-1}(x) \to F_x$, for each $x \in B$.*

Under this bijection, the fibers of φ are the same as the fibers of the maps g_x.

Theorem. (DK, 2007)(Patchwork theorem)
Assume that $\varphi : P \to Q$ is a poset map, and that we have acyclic matchings on subposets $\varphi^{-1}(q)$, for all $q \in Q$. Then the union of these matchings is itself an acyclic matching on P.

III. Applications: Complexes of graph homomorphisms

The applications which we have presented in this talk went along the following logical scheme.

$$\begin{array}{ccc} \text{graph } G & \longrightarrow & \text{topological space } X(G) \\ & & \downarrow \\ \begin{array}{c}\text{combinatorial}\\ \text{properties of } G\end{array} & \longleftarrow & \begin{array}{c}\text{topological}\\ \text{properties of } X(G)\end{array} \end{array}$$

The main idea was to introduce a cell complex related to the set of graph colorings of a fixed graph. Recall that for two graphs T and G, a **graph homomorphism** from T to G is a map $\phi : V(T) \to V(G)$, such that for every edge (x, y) in T the pair $(\phi(x), \phi(y))$ is an edge in G. It is an easy observation that G is n-colorable if and only if there exists a graph homomorphism $\phi : G \to K_n$.

Furthermore, the composition of two homomorphisms $\phi_1 : G_1 \to G_2$ and is again a homomorphism $\phi_2 \circ \phi_1 : G_1 \to G_3$, hence we obtain the category **Graphs** with graphs as objects and graph homomorphisms as morphisms.

A cell in $\texttt{Hom}\,(T, K_n)$ is an assignment of subsets of $[n]$ to vertices of T, such that an arbitrary choice of one color per list yields an admissible coloring of T.

Generalizing we replace K_n with an arbitrary graph G, and see that

- the vertices of G are the colors;
- homomorphisms $T \to G$ replace the valid colorings.

Hom-complexes enjoy several nice properties:

(1) The cells in $\texttt{Hom}\,(T, G)$ are products of simplices: every cell η is a product of $|V(T)|$ simplices of dimension $|\eta(x)|-1$ for $x \in V(T)$. Hence, $\texttt{Hom}\,(T, G)$ is a **prodsimplicial** complex.

(2) $\texttt{Hom}\,(T, -)$ is a covariant and $\texttt{Hom}\,(-, G)$ is a contravariant functor from **Graphs** to **Top**. In particular, the group $\mathcal{A}ut(T) \times \mathcal{A}ut(G)$ acts on $\texttt{Hom}\,(T, G)$. When G is loopfree and $\phi \in \mathcal{A}ut(T)$ flips an edge, the induced map $\phi_G \circlearrowright \texttt{Hom}\,(T, G)$ is fixed point free.

Introducing characteristic classes as obstructions to graph colorings.

Recall that a CW complex X is called a $\mathbb{Z}_2$**-space**, if $\mathbb{Z}_2$ freely acts on X. In this case there exists a continuous $\mathbb{Z}_2$-map $\psi : X \to S^\infty_a$. The induced quotient map $\phi : X/\mathbb{Z}_2 \to \mathbb{R}P^\infty$ is up to homotopy independent of the choice of ψ. This induces a map of algebras

$$\phi^* : H^*(\mathbb{R}P^\infty; \mathbb{Z}_2) \to H^*(X/\mathbb{Z}_2; \mathbb{Z}_2),$$

which is independent of the choice of ψ. Let z be the generator of $H^1(\mathbb{R}P^\infty; \mathbb{Z}_2)$. Then

$$\boldsymbol{w_1(X)} := \phi^*(z)$$

is called the **Stiefel-Whitney class**. These classes are functorial in the following sense: if $\psi : X \to Y$ is a $\mathbb{Z}_2$-map, and $\widetilde{\psi} : X/\mathbb{Z}_2 \to Y/\mathbb{Z}_2$ is the induced map between the quotient spaces, then

$$\widetilde{\psi}^*(w_1(Y)) = w_1(X)\,.$$

Stiefel-Whitney characteristic classes were recently applied by Eric Babson and the speaker to settle the following conjecture (more generally

$$w_1^{n-2}(\texttt{Hom}\,(C_{2r+1}, K_n)) = 0$$

holds).

Lovász Conjecture:

$$\texttt{Hom}\,(C_{2r+1}, G) \quad \text{is} \quad k\text{-connected} \quad \Longrightarrow \quad \chi(G) \geq k+4\,.$$

See [1] for details. Roughly, the proof is based on spectral sequence computations coupled with using topological obstructions as follows.

Assume that $\chi(G) \leq k+3$, that is $\exists\, \varphi : G \to K_{k+3}$. The functoriality of the $\texttt{Hom}$-construction yields a $\mathbb{Z}_2$-map

$$\psi : \texttt{Hom}\,(C_{2r+1}, G) \to \texttt{Hom}\,(C_{2r+1}, K_{k+3}).$$

On the other hand, when $\texttt{Hom}\,(C_{2r+1}, G)$ is k-connected, there exists a $\mathbb{Z}_2$-map $\rho : S_a^{k+1} \to \texttt{Hom}\,(C_{2r+1}, G)$. For the characteristic classes we get

$$0 = w_1^{k+1}(\texttt{Hom}\,(C_{2r+1}, K_{k+3})) \longrightarrow w_1^{k+1}(\texttt{Hom}\,(C_{2r+1}, G)) \longrightarrow w_1^{k+1}(S_a^{k+1}) \neq 0,$$

which yields a contradiction.

Further spectral sequence computations were also used by the author to derive the following result.

Theorem. (DK, 2005) *For integers m, n, such that $m \geq 5$, $n \geq 4$, we have*

$$\widetilde{H}^*(\texttt{Hom}\,(C_m, K_n); \mathbb{Z}) = \left(\bigoplus_{t=1}^{\lfloor (m-2)/3 \rfloor} A_{t,m,n} \right) \oplus B_{m,n},$$

where

$$A_{t,m,n} = \begin{cases} \mathbb{Z}(tn-3t) \oplus \mathbb{Z}(tn-3t+1), & \text{if } n \text{ or } m+t \text{ odd,} \\ \mathbb{Z}_2(tn-3t+1), & \text{if } n \text{ and } m+t \text{ even,} \end{cases}$$

and

$$B_{m,n} = \begin{cases} \mathbb{Z}^{2^n-3}(nk-m), & \text{if } m = 3k, \\ \mathbb{Z}(nk-m+2), & \text{if } m = 3k+1, \\ \mathbb{Z}(nk-m), & \text{if } m = 3k-1. \end{cases}$$

So, for example, we have

- $\widetilde{H}^*(\texttt{Hom}\,(C_6, K_4); \mathbb{Z}) = A_{1,6,4} \oplus B_{6,4} = \mathbb{Z}(1) \oplus \mathbb{Z}(2) \oplus \mathbb{Z}^{13}(2) = \mathbb{Z}(1) \oplus \mathbb{Z}^{14}(2)$;
- $\widetilde{H}^*(\texttt{Hom}\,(C_8, K_6); \mathbb{Z}) = A_{1,8,6} \oplus A_{2,8,6} \oplus B_{8,6} =$
  $\mathbb{Z}(3) \oplus \mathbb{Z}(4) \oplus \mathbb{Z}_2(7) \oplus \mathbb{Z}(10)$.

The author has also developed the following homology tests for graph colorings. Recall that the **Stiefel-Whitney height** of a non-empty $\mathbb{Z}_2$-space X is the maximal $\texttt{h}(X)$, such that $w_1^{\texttt{h}(X)}(X) \neq 0$. The following is true in general.

Theorem. *Let X be a non-empty $\mathbb{Z}_2$-space of finite Stiefel-Whitney height, then $\widetilde{H}^{\texttt{h}(X)}(X; \mathbb{Z}_2) \neq 0$.*

Hence we get

Theorem. (DK, 2006) *Let T be a graph with a $\mathbb{Z}_2$-action flipping an edge, such that*
(1) T is a Stiefel-Whitney test graph: $\texttt{h}(\texttt{Hom}\,(T, K_n)) = n - \chi(T)$,
(2) $\widetilde{H}_i(\texttt{Hom}\,(T, G); \mathbb{Z}_2) = 0$, for $i \leq d$,

then $\chi(G) \geq d + 1 + \chi(T)$.

References

[1] E. Babson, D.N. Kozlov: *Proof of the Lovász Conjecture*, `Annals of Mathematics`, **165** (2007), 965–1007.

[2] E. Babson, D.N. Kozlov, *Topological obstructions to graph colorings*, Electron. Res. Announc. Amer. Math. Soc. **9** (2003), 61–68.

[3] E. Babson, D.N. Kozlov, *Complexes of graph homomorphisms*, Israel J. Math. **152** (2006), 285–312.

[4] D.N. Kozlov: *Combinatorial Algebraic Topology*, 407 pp., Springer-Verlag, 2008.

[5] D.N. Kozlov: *Chromatic numbers, morphism complexes, and Stiefel-Whitney characteristic classes*, `IAS/Park City Mathematics Series` **14**, Amer. Math. Soc., Providence, RI, 2007; Institute for Advanced Study, Princeton, NJ, 249–316.

[6] D.N. Kozlov: *Cobounding odd cycle colorings*, Electron. Res. Announc. Amer. Math. Soc. **12** (2006), 53–55.

[7] D.N. Kozlov: *Homology tests for graph colorings*, in: Contemporary Mathematics **423**, Amer. Math. Soc., Providence, RI, 2007.

[8] D.N. Kozlov: *Cohomology of colorings of cycles*, American J. Math., to appear. `arXiv:math.AT/0507117`

The Geometry of Materials

Robert MacPherson

1. Cell Complexes Arising from Materials.

The structure of two different sorts of materials leads to a cell decomposition of a region of 3-space:

- Metals and ceramics are generally divided into individual crystals, typically of size $\sim 10^{-5}$ meters, which are the cells of a cell complex.
- Foams are a cell complex whose cells are the individual bubbles.

Both of these situations have interesting two dimensional analogues, which give cell decompositions of a region of 2-space:

- In very thin sheets of metals and ceramics, the grains are essentially two dimensional.
- Bubbles trapped between two parallel sheets of glass are essentially two dimensional.

The structures of these cell complexes are key to understanding many of the properties of the material. The cell complexes are very large and complicated. New ideas in topology are probably needed to understand them adequately.

2. Time Evolution.

In all of these situations, the cell complex evolves over time. In the case of metals and ceramics, individual atoms jump from being part of one crystal to being part of a neighboring crystal. In the case of foams, air molecules diffuse through the liquid separating one bubble from the next one. As a result of this time evolution, some cells become smaller and disappear, so the average cell size becomes larger. This process is sometimes called coarsening.

There is a standard mathematical model for this time evolution. In the case of a metal or ceramic, this model can be informally summarized by saying that the cell decomposition evolves according to $-C$ grad $\mathcal{A}$ where C is a constant depending on the material and the temperature, and $\mathcal{A}$ is the total area (or length in 2 dimensions) of all the grain boundaries separating adjacent grains. The idea is that the grain boundaries have an energy cost proportional to their area, which is decreased over time. Made mathematically rigorous, this model stipulates that the grain boundaries move by mean curvature flow (curvature flow in 2 dimensions), and that adjacent grain boundaries meet at 120° angles. Like most mathematical models, this one is an approximation. More accurate models would involve the complication of an energy cost that varies with the orientation of the boundary. However, this model is already very accurate for many materials. There are interesting mathematical questions that are not understood even for this for this model.

Foams evolve according to a different model which we won't go in to here. However, the results stated below hold for foams as well.

3. Growth of an Individual Cell.

A fundamental result in the field is the formula for the growth of an individual grain, due to von Neumann (1952) and Mullins (1956).

The von Neumann - Mullins Relation for Grains in Two Dimensions. Assuming the model for grain evolution described above, rate of change of the area A of an individual grain is given by

$$\frac{dA}{dt} = K(n-6)$$

where n is the number of 0-cells around the grain, K is a constant depending on the material and the temperature, and t is time.

In joint work with David Srolovitz, we generalized von Neumann - Mullins result to higher dimensions. First restate the two dimensional von Neumann-Mullins relation as follows:

Restatement of von Neumann - Mullins. The rate of change of the area of a grain G in two dimensions is

$$\frac{dA}{dt} = K\left(\mathcal{L}_0(G_0) - 6\mathcal{L}_0(G)\right)$$

where $\mathcal{L}_0$ denotes the Euler characteristic and G_0 is the union of the 0-dimensional cells around G.

This is the same formula, because $\mathcal{L}_0(G_0) = n$ and $\mathcal{L}_0(G)$ is 1 since G is a cell. (Actually, we've already gained something by this restatement. It is possible to set up conditions where an individual grain has holes in it. Then, $\mathcal{L}_0(G)$ is no longer 1. The restated formula is the correct one, in this situation.)

The three dimensional analogue is then is the following:

Three dimensional von Neumann - Mullins relation. ***(joint with D. Srolovitz)*** The rate of change of the volume V of a grain G in three dimensions is

$$\frac{dV}{dt} = K\left(\mathcal{L}_1(G_1) - 6\mathcal{L}_1(G)\right)$$

where $\mathcal{L}_1$ denotes the *mean width* and G_1 is the union of the 1-dimensional cells around G. (The quantity $\mathcal{L}_1(G_1)$ is just the sum of the lengths of the 1-dimensional cells in G_1.)

The mean width $\mathcal{L}_1(G)$ a measure of length associated to G. It is one of a series of measures from integral geometry that go by various names: "quermassintegrals', "intrinsic volumes", "Minkowski functionals", or "Hadwiger measures". The initial one, $\mathcal{L}_0(G)$ is the Euler characteristic, and $\mathcal{L}_i(G)$ scales as i dimensional area. These measures have an extensive literature of characterizations, formal properties, and algorithms for computation that we won't go in to here.

4. Universality of the Long Time Limit.

Almost all materials scientists believe the following *universality conjecture*: For generic initial conditions, the cell complex of grains evolves toward a universal (random) cell complex, defined up to a scaling factor. This universal cell complex should be independent of the material. If a cell complex is already universal, then with further time evolution, it stays universal, but scaled in accord with the increase of the average grain size. This universality conjecture is certainly not proved. I don't even know a mathematically precise formulation. However there is a lot of evidence for it as a phenomenon - both from numerical simulations and from physical experiments.

Assuming the conjecture, the universal cell decomposition by grains is a very important object. All real metals and ceramics are likely to be in this universal state, because early in their formation the temperature was high so the evolution was fast. However, very little is known about it.

Enumerative Combinatorics of Cell Complexes. The universality conjecture implies that quantities associated to the cell complex should have a universal distributions. The subject of enumerative combinatorics of cell complexes studies several such quantities obtained by counting faces of various dimensions (the "f-vectors"). These quantities are subject to identities called the Dehn-Sommerville

relations. Using these identities we find that for a two dimensional complex of grains

$$\langle \text{The number of vertices on a grain} \rangle = 6$$

where $\langle \text{X} \rangle$ means the expected value of X.

In three dimensions, the situation is more interesting. There is one free parameter F.

$$\langle \text{The number of faces on a grain} \rangle = F$$

Then Dehn-Sommerville relations yield:

$$\langle \text{The number of edges on a grain} \rangle = 3F - 6$$

$$\langle \text{The number of vertices on a grain} \rangle = 2F - 4$$

$$\langle \text{The number of edges on a face} \rangle = 6 - \frac{12}{F}$$

What can we say about this quantity F for the universal cell decomposition by grains? We have one rigorous result:

Theorem.

$$F = \frac{13.39}{1 + 2.13C}$$

where C is the average of the integral of the Gauss curvature over a face of the cell complex. Here, the precise values of the constants are:

$$2.13 = 1 + \frac{1}{3\cos^{-1}(1/3) - \pi}$$

$$13.39 = 2 + \frac{2\pi}{3\cos^{-1}(1/3) - \pi}$$

I call 13.39 this Coxeter's number, because Coxeter gave a plausibility argument that F should be this number. One can think of our theorem as a rigorous version of Coxeter's argument. In fact, simulations and experiments show that F is between 13.5 and 14 for the universal cell complex in three dimensions, so our theorem can be read backwards to say that the average Gauss curvature over a face is negative.

Other Questions. One contribution that pure mathematics could make to the study of universal cell complexes is to identify what the interesting invariants to study are (just as the mean width, the Gauss curvature, and the invariants from enumerative combinatorics mentioned above came from pure mathematics). This raises new mathematical questions, such as the following:

- Are there interesting (quantum) invariants of cell decompositions of a manifold with boring global topology, such as a ball in three space? (This would be analogous to the phenomenon that there are very interesting invariants of a one manifold embedded in a three manifold with boring topology.)
- Is there a hierarchy of invariants in the style of enumerative combinatorics of which the f-vectors discussed above are the leading item in the hierarchy? (For example, counting how many 3-cells have n 2-cells as faces might be in the next order of the hierarchy.)
- What are interesting metric invariants of such cell complexes? Obvious examples are the volumes of 3-cells, the areas of 2-cells, etc. However there should be more interesting metric invariants, perhaps using the ideas of persistent homology.

Stability of Clustering Methods

Facundo Mémoli
(joint work with Gunnar Carlsson)

Despite being one of the most commonly used tools for unsupervised exploratory data analysis and despite its extensive literature, very little is known about the theoretical foundations of clustering methods.

Standard clustering methods take as input a finite metric space (X, d) and output a partition of X. Kleinberg [2] discussed this situation in an axiomatic way and identified a set of reasonable properties of standard clustering schemes, namely, scale invariance, richness and consistency. He then proved, in the same spirit of Arrow's impossibility theorem, that no clustering scheme satisfying these conditions simultaneously can exist.

Datasets can exhibit multiscale structure and this can render standard clustering algorithms inapplicable in certain situations, see Figure 1. This motivates the use of *Hierarchical clustering methods.*

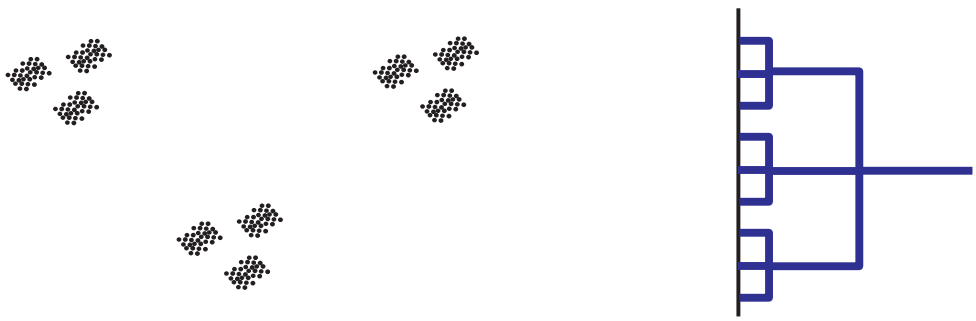


Figure 1. Dataset with multiscale structure.

In a similar spirit to Kleinberg's theorem we prove that in the context of hierarchical methods, one obtains uniqueness instead of non-existence.

1. Hierarchical clustering: formulation

Given a finite metric space (X, d), a hierarchical clustering method f returns a nested family of partitions, or **dendrogram** (a.k.a. persistent set) of X:

$$f(X, d) \in \mathcal{D}(X)$$

where $\mathcal{D}(X) = \{(X, \theta) |\, \theta : [0, \infty) \to \mathcal{P}(X)\}$ s.t.

(1) $\theta(0) = \{\{x_1\}, \ldots, \{x_n\}\}$.
(2) There exists t_0 s.t. $\theta(t)$ is the *single block partition* for all $t \geq t_0$.
(3) If $r \leq s$ then $\theta(r)$ *refines* $\theta(s)$.
(4) For all r there exists $\varepsilon > 0$ s.t. $\theta(r) = \theta(t)$ for $t \in [r, r + \varepsilon]$.

We represent dendrograms (= rooted trees) as *ultrametric* spaces: a metric space (X, u) is an ultrametric space if and only if for all $x, x', x'' \in X$,

$$\max(u(x, x'), u(x', x'')) \geq u(x, x'').$$

For $n \in \mathbb{N}$ let $\mathcal{X}_n$ (resp. $\mathcal{U}_n$) denote the set of all metric spaces (resp. ultra-metric spaces) with n points. Let $\mathcal{X} = \sqcup_{n \geq 1} \mathcal{X}_n$ denote set of all finite metric spaces and $\mathcal{U} = \sqcup_{n \geq 1} \mathcal{U}_n$ all finite ultrametric spaces. Then, a hierarchical clustering method can be regarded as a map

$$T : \mathcal{X} \to \mathcal{U}$$

s.t. $\mathcal{X}_n \ni (X, d) \mapsto (X, u) \in \mathcal{U}_n$, $n \in \mathbb{N}$.

There is a canonical construction: Let $T^* : \mathcal{X} \to \mathcal{U}$ be given by $(X, d) \mapsto (X, u^*)$ where

$$u^*(x, x') := \min \left\{ \max_{i=0,\ldots,\ell-1} d(x_i, x_{i+1}) |\, x = x_0, \ldots, x_\ell = x' \right\}.$$

This construction yields exactly *single linkage clustering*, [1].

For $X \in \mathcal{X}$ let $\mathrm{sep}(X, d) := \min_{x \neq x'} d(x, x')$.

Theorem 1. *Let T be a clustering method s.t.*

(1) $T\left(\{p, q\}, \left(\begin{smallmatrix} 0 & \delta \\ \delta & 0 \end{smallmatrix}\right)\right) = \left(\{p, q\}, \left(\begin{smallmatrix} 0 & \delta \\ \delta & 0 \end{smallmatrix}\right)\right)$ *for all* $\delta > 0$.
(2) *Whenever $X, Y \in \mathcal{X}$ and $\phi : X \to Y$ are such that for all $x, x' \in X$ we have $d_X(x, x') \geq d_Y(\phi(x), \phi(x'))$, then it also holds*

$$u_X(x, x') \geq u_Y(\phi(x), \phi(x'))$$

for all $x, x' \in X$, where $T(X, d_X) = (X, u_X)$ and $T(Y, d_Y) = (Y, u_Y)$.
(3) *For all $(X, d) \in \mathcal{X}$,*

$$u(x, x') \geq sep(X, d) \text{ for all } x \neq x' \in X$$

where $T(X, d) = (X, u)$.

Then $T = T^*$.

1.1. **Metric stability of T^*.** We also obtain the Proposition and Theorem below asserting metric stability and asymptotic consistency of the method T^*. We use the notion of Gromov-Hausdorff distance between metric spaces, [3].

The Gromov-Hausdorff distance $d_{\mathcal{GH}}(X,Y)$ between compact metric spaces (X,d_X) and (Y,d_Y) is defined to be the infimal $\varepsilon > 0$ s.t. there exists a metric d on $X \sqcup Y$ with $d_{|_{X\times X}} = d_X$ and $d_{|_{Y\times Y}} = d_Y$ for which the Hausdorff distance between X and Y (as subsets of $(X \sqcup Y, d)$) is less than ε.

Proposition 2. *For any finite metric spaces (X,d_X) and (Y,d_Y)*

$$d_{\mathcal{GH}}((X,d_X),(Y,d_Y)) \geq d_{\mathcal{GH}}(T^*(X,d_X),T^*(Y,d_Y)).$$

Fix a finite set X. For a symmetric function $W : X \times X \to \mathbb{R}^+$ let $\mathcal{L}(W)$ denote the maximal metric on X less than of equal to W, i.e.

$$\mathcal{L}(W)(x,x') = \min\left\{ \sum_{i=0}^{m} W(x_i,x_{i+1}) |\, x = x_0,\ldots,x_m = x \right\}$$

for $x, x' \in X$.

Theorem 3. *Assume (Z,d_Z) is a compact metric space. Let X and X' be any two finite sets of points sampled from Z and $r,r' > 0$ such that $Z \subset \cup_{x\in X} B(x,r)$ and $Z \subset \cup_{x'\in X'} B(x',r')$. Let $d_X = d_{Z|_{X\times X}}$ and $d_{X'} = d_{Z|_{X'\times X'}}$. Let $T^*(X,d_X) = (X,u_X)$ and $T^*(X',d_{X'}) = (X',u_{X'})$. Then,*

1. *(Finite Stability) $d_{\mathcal{GH}}((X,u_X),(X',u_{X'})) \leq (r+r')$.*
2. *(Convergence/consistency) Assume in addition that $Z = \sqcup_{\alpha\in A} Z_\alpha$ where A is a finite index set and Z_α are compact, disjoint and path-connected sets. Let (A,d_A) be the finite metric space with underlying set A and metric given by $d_A := \mathcal{L}(W)$ where*

$$W(\alpha,\alpha') := \min_{z\in Z_\alpha, z'\in Z_{\alpha'}} d_Z(z,z') \quad \text{for } \alpha,\alpha' \in A.$$

Let $T^*(A,d_A) = (A,u_A)$. Then, as $r \to 0$ one has

$$d_{\mathcal{GH}}((X,u_X),(A,u_A)) \to 0.$$

An extended version of this work has appeared in [4].

References

[1] A. Jain and R.Dubes, *Algorithms for Clustering Data*, Prentice-Hall, Englewood Cliffs, NJ, 1988. xiv+320 pp. ISBN: 0-13-022278-X

[2] J. Kleinberg, *An impossibility theorem for clustering*, Advances in Neural Information Processing Systems (NIPS) 15, 2002.

[3] D. Burago, Yu. Burago, and S. Ivanov, *A course in Metric Geometry*, AMS Graduate Studies in Mathematics Volume 33.

[4] G. Carlsson and F. Mémoli, *Persistent Clustering and a theorem of J. Kleinberg*, `comptop.stanford.edu/preprints`.

Building a database for the global dynamics of multi-parameter systems

Konstantin Mischaikow
(joint work with Z. Arai, W. Kalies, H. Kokubu, H. Oka, and P. Pilarczyk)

The global dynamics of a nonlinear system can exhibit structures at all spatial scales, for example, the fractal structures associated to chaotic dynamics. The same phenomenon can occur with respect to the parameters, that is, global dynamical structures can change on Cantor sets in parameter space. From the point of view of scientific computation, only a finite amount of information can be computed, and therefore, any computational characterization of global dynamics of a multiparameter system can be expected to represent a dramatic reduction of information. Nevertheless, the computation of global dynamical information is an important problem for applications, which leads to the questions of how to characterize global dynamical structures and how to identify changes in these structures in practice. The fact that this is a nontrivial task has been made clear by the work of the dynamical systems community over the last century.

Identifying and classifying the qualitative properties of models over a wide ranges of parameter values is of fundamental importance in many disciplines. For biologists the question of the preservation of qualitative structures over large ranges of parameter values is of considerable interest. This question is in some sense contrary to the classical topic of singularity or bifurcation theory, where the focus is on understanding how the dynamics changes. The fact that most topics of interest in systems biology are dynamic in nature suggests the need for a comprehensive, yet efficient method for cataloging the global dynamics of nonlinear systems. In other words, a method is desired which computationally constructs a database of global dynamical behavior of a specific system over a range of parameters.

The starting point for our computational methods is Conley's topological approach to dynamics. We shall argue that Conley theory provides an appropriate theoretical base for designing algorithms in computational dynamics and demonstrate that these ideas can be used to design an efficient computational framework for constructing databases of global dynamics of specific systems over multiple parameters. While the methods we propose are general, we illustrate them using a two-dimensional version of an overcompensatory Leslie population model $g\colon \mathbb{R}^2 \times \mathbb{R}^4 \to \mathbb{R}^2$ given by

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mapsto \begin{bmatrix} (\theta_1 x_1 + \theta_2 x_2) e^{-c(x_1+x_2)} \\ p x_1 \end{bmatrix},$$

where the fertility rates decay exponentially with population size.

The database takes the form of a continuation graph. The nodes contain the information about the continuation classes and the edges indicate the connectivity (in parameter space) of the continuation classes. The continuation classes are associated to regions in parameter space and consist of acyclic directed graphs, called Conley-Morse graphs. Each node of a Conley-Morse graph is associated

with a Morse set and its Conley index. The continuation graph provides a compact easily queriable description of the global dynamics over the entire range of parameter spaces.

Trace spaces: Organization, Calculation, Applications

Martin Raussen

A topological approach to the study of concurrency phenomena in theoretical Computer Science has led to the definition of Higher Dimensional Automata: The main interest, both mathematically and in concurrency, is in the study of *directed* (d)-paths in a pre-cubical complex (and more generally, a d-space) up to *directed* homotopy (d-homotopy). These d-paths can, in general, not be reversed; hence algebraic topological invariants of the spaces of d-paths in a d-space may vary with the selection of start and end points. For technical reasons, it is often advisable to divide out (weakly) increasing reparametrizations; the quotient objects are the so-called traces – whence the title of the talk.

In a given d-space X, let $\vec{T}(X)(x,y)$ denote the space of all d-paths starting at x and ending at y with the CO-topology. If X is a pre-cubical complex, we show that this space is metrizable, separable, locally contractible, locally compact and that it has the homotopy type of a CW-complex.

Apart from the number of components of such a trace space that often can be calculated using a Seifert-van Kampen type theorem due to Marco Grandis, not much has been known in general about the topology of d-path or trace spaces. We use a version of the Vietoris-Begle theorem due to S. Smale to arrive at additional information, e.g.:

Decomposition: Suppose given a collection of disjoint "layers" $L_1, \ldots, L_{n-1} \subset X$ in a d-space X with additional properties (satisfied by the geometric realization of a pre-cubical complex). If the set of layers is *unavoidable* from x to y – in the sense that every d-path can be decomposed into d-paths that start in one layer, end in another without touching any of these or other layers in between – then the space $\vec{T}(X)(x,y)$ of traces in X from x to y is (at least weakly) homotopy equivalent to a union of fibered product of spaces $\vec{T}(X)(L_i, L_{i+1})$.

Reachable sequences: If, moreover, $\vec{T}(X)(x_i, x_{i+1})$ is either empty or (weakly) contractible for every $x_i \in L_i, x_j \in L_j$ belonging to subsequent layers, then $\vec{T}(X)(x,y)$ is weakly homotopy equivalent to the subspace of "mutually reachable" sequences $(x_1, \ldots, x_{n-1}) \in X^{n-1}$ connecting points in these layers. In particular, for a nice space X, the infinite dimensional trace spaces can often be identified with a subspace of a finite dimensional space.

Piecewise linear traces: If X is the geometric realization of a precubical complex, one may consider the *piecewise linear* (geodesic) d-paths in X

that are linear on each cell (including its boundary). The space of piecewise linear traces $\vec{T}_l(X)(x,y)$ is (at least weakly) homotopy equivalent to the space of all traces $\vec{T}(X)(x,y)$. It can be subdivided into "cube paths" containing all PL-d-paths that are contained in a given sequence of subsequent cubes. Such a cube path turns out to be a product of simplices, and $\vec{T}_l(X)$ comes thus equipped with a prodsimplicial combinatorial structure. Since cube paths can be ordered by their length in "rounds", (higher) connectivity computations as in the work of Herlihy et al. seem promising along these lines.

Moreover, the following topics have been dealt with:

Arc length: In a pre-cubical complex, there is a natural notion of arc-length (l_1) and natural parametrization; this makes it possible to represent every trace by a canonical d-path. It is not difficult to see that arc length is preserved under a d-homotopy.

Variation of invariants: A categorical organisation of the algebraic topological invariants of trace spaces with varying end points giving rise to a decomposition of the parameter space, that has to be viewed as a subspace of $X \times X$.

References

[1] M. Grandis, *Directed homotopy theory I. The fundamental category*, Cahiers Topologie Géom. Différentielle Catég. **44** (2003), 281-316.

[2] L. Fajstrup, E. Goubault, E. Haucourt, M. Raussen, *Components of the fundamental category*, Appl. Categ. Struct. **12** (2004), 91-108.

[3] L. Fajstrup, E. Goubault, M. Raussen, *Algebraic Topology and Concurrency*, Theor. Comput. Sci. **357** (2006), 241-278.

[4] U. Fahrenberg, M. Raussen, *Reparametrizations of continuous paths*, J. Homotopy Relat. Struct. **2** (2007), 93-117.

[5] M. Raussen, *Invariants of directed spaces*, Appl. Categ. Struct. **15** (2007), 355-386.

Computational Algebraic Topology à la Eilenberg-Mac Lane

Pedro Real

By Computational Algebraic Topology we mean an area of Computational Mathematics dealing with the effective and efficient construction of algebraic topological invariants for discrete or continuous objects. Eilenberg and Mac Lane in the sixties of the last century develop a framework for Algebraic Topology mainly based on the notions of simplicial sets and chain homotopy equivalences. This setting which gives preference to chain homotopy operators in combinatorial ambiance is suitable for developing computational algebraic topological work related to $A(\infty)$-algebra cohomology computation. As example of this modus operandi, we show an application in Digital Volume Processing. More concretely, a binary digital (voxel-based) volume is seen from an geometric-algebraic point of view as a special chain homotopy operator acting on every cell of the polyhedral continuous analogous associated to the volume. Discrete Morse Theory and Homological Perturbation

Theory can be integrated to this framework in a straightforward and algorithmic manner.

Local Scale Selection for Large High Dimensional Data

Gurjeet Singh
(joint work with Gunnar Carlsson, Facundo Mémoli)

We introduced a method called `Mapper` (in [3]) for the qualitative analysis, simplification and visualization of high dimensional datasets with respect to *filter functions* defined on the dataset. In many cases, data coming from real applications is massive and it is not possible to visualize and discern structure even in low dimensional projections. The method can be used to reduce high dimensional datasets into simplicial complexes with far fewer points which can capture topological and geometric information at a specified resolution. Our construction provides a coordinatization of the data by providing a discrete and combinatorial object, a simplicial complex, to which the dataset maps and which can represent the dataset in a useful way. We propose an extension to `Mapper` in the form of a method that is able to extract different scale dependent views of the data. This method is a scale detection procedure in which we compute a graph that encodes all the multiscale structure present in the data as viewed w.r.t a function prescribed on the data.

In the simplest case, `Mapper` begins with a sample from a metric space X, and a continuous real valued function $f : X \to \mathbb{R}$, to produce a graph. This function can be a function which reflects geometric properties of the dataset, such as the result of a density estimator, or can be a user defined function, which reflects properties of the data being studied. In the first case, one is attempting to obtain information about the qualitative properties of the dataset itself, and in the second case one is trying to understand how these properties interact with interesting functions on the dataset. The basic idea behind `Mapper` can be referred to as *partial clustering*, in that a key step is to apply standard clustering algorithms to subsets of the original dataset, and then to understand the interaction of the partial clusters thus formed with each other. That is, if U and V are subsets of the dataset, and $U \cap V$ is non-empty, then the clusters obtained from U and V respectively may have non-empty intersections, and these intersections are used to construct a simplicial complex.

In more detail, assume $f : X \to \mathbb{R}$ is given together with an open covering $\{U_\alpha\}_{\alpha \in A}$ of its range, for a finite index set A. Then $\{f^{-1}(U_\alpha)\}_{\alpha \in A}$ forms an open covering of X. For each α we consider the decomposition of $f^{-1}(U_\alpha)$ into its path connected components: $f^{-1}(U_\alpha) = \cup_{i=1}^{j_\alpha} V(\alpha, i)$ where j_α is the number of connected components of U_α. Let $W = \{W_\beta\}_{\beta \in B}$ be the set of all $V(\alpha, i)$ for all $1 \leq i \leq j_\alpha$ and $\alpha \in A$. Note that $|B| = \sum_{\alpha \in A} j_\alpha$. Next we construct the *nerve* $\mathcal{N}(W)$ of the covering W, i.e. the simplicial complex with vertex set B and where a family $\{\beta_0, \ldots, \beta_k\}$ spans a k-simplex in $\mathcal{N}(W)$ if and only if $W_{\beta_0} \cap W_{\beta_1} \cap \ldots \cap W_{\beta_k} \neq \emptyset$. The practical counterpart of the operation of *decomposing a set*

[htbp]

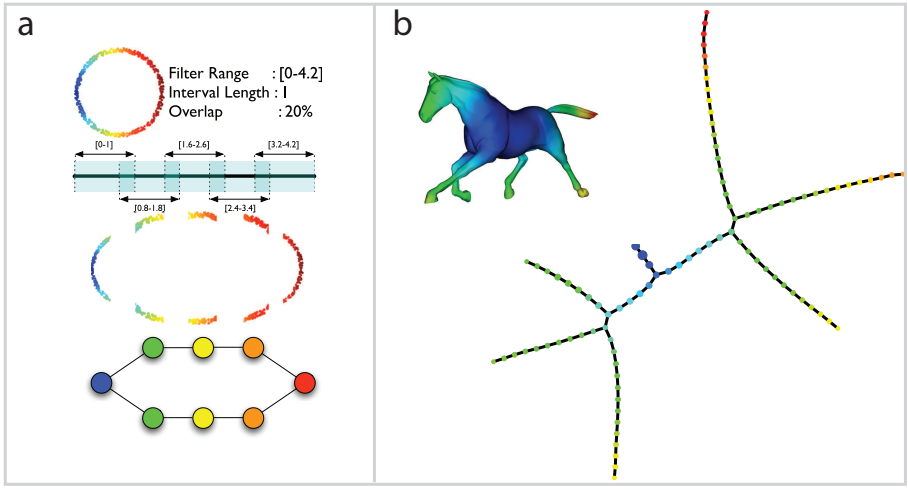


FIGURE 1. (a) The data is sampled from a noisy circle, and the function used is $f(x) = ||x - p||_2$, where p is the leftmost point in the data. The dataset is shown on the top left, colored by the value of the function. We divide the range of the function into 5 equal intervals and a 20% overlap between subsequent intervals. For each interval we compute the clustering of the points lying in all the preimages of the intervals. Finally we connect the clusters with edges whenever they have non empty intersection. At the bottom is the simplicial complex which we recover; its vertices are colored by the average function value. (b) Application of `Mapper` to shape simplification. The horse shape is colored with the function f which we choose to be the average geodesic distance to all other points on the shape (blue is low and red is high). The output of `Mapper` is a graph, its vertices being colored by the average function value of the points in the corresponding cluster.

into its path connected components is of course clustering. We show two practical examples of application of `Mapper` in the Figure 1.

Instead of using a function $f : X \to \mathbb{R}$, one can use a function $f : X \to Z$, where Z is any topological space e.g. Z could be $\mathbb{R}^2$ or S^1. In the first case, one produces a two dimensional simplicial complex, together with a natural map from the dataset to it. In the second case, one constructs a graph with a map from the graph to a circle. Roughly, our construction amounts to a generalization of the *Reeb graph* (see [1]) associated with the filter function.

In the practical implementation of the procedure one must deal with some clustering algorithm. Our procedure is not tied to any particular choice. In our experiments, however, we choose to work with single linkage clustering. Regardless of the clustering algorithm one picks, it is always necessary to estimate certain parameters (thresholds) that ultimately determine the number of clusters recovered.

In the case of using a hierarchical clustering procedure and a function $f : X \to \mathbb{R}$, we propose a procedure that detects spatial coherence of the scales in the partial datasets $f^{-1}(U_\alpha)$ to obtain a coherent choice of parameters. This procedure is called the *scale space procedure*. Roughly, we construct a *scale space graph*, which encodes all the scale dependent information in the dataset relative to the function f. This graph is a layered, directed and planar graph. Refer to Figure 2 for an example.

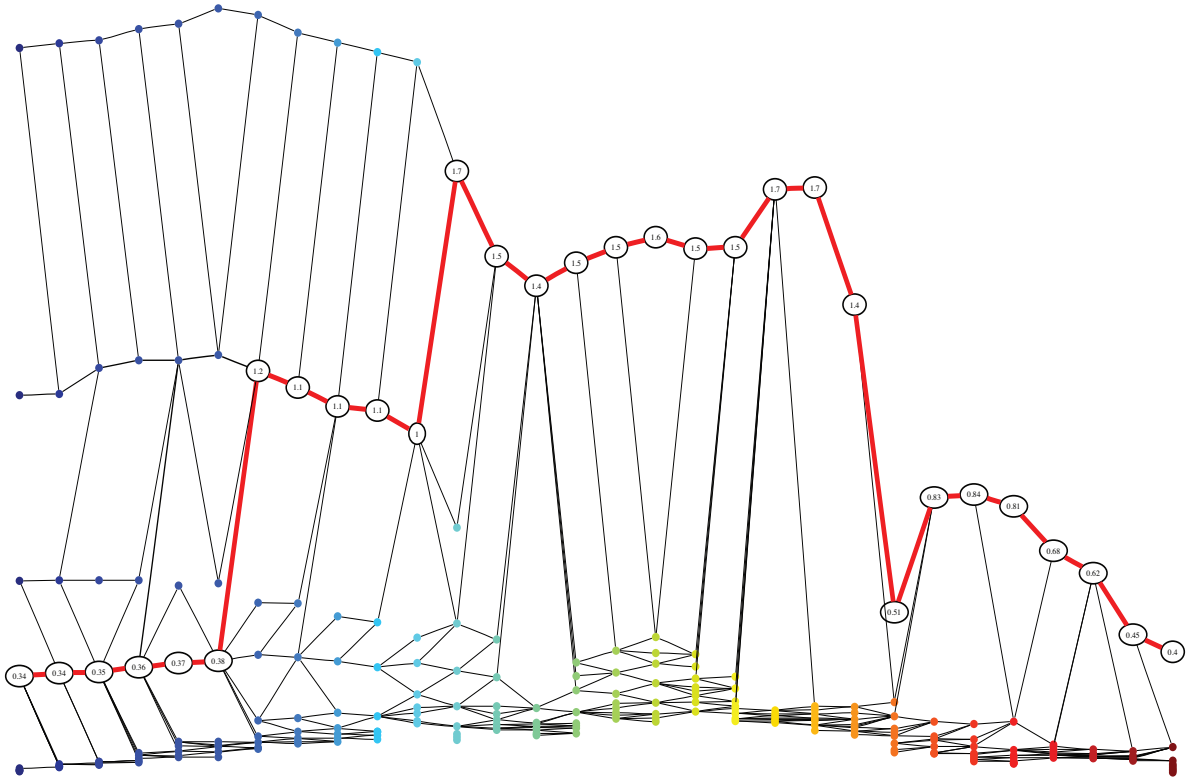


FIGURE 2. This is the scale space graph corresponding to the shape and the filter function shown in Figure 1(b). The filter was divided into 30 intervals; there is a layer of vertices corresponding to each interval. A path (like the red path shown here) picks a vertex from each layer which represents a threshold for each layer. A path in this graph which picks one vertex from each layer can be used to generate one `Mapper` output. The `Mapper` output shown in Figure 1(b) corresponds to the red path shown here. The set of all paths in this graph represent the set of all possible `Mapper` outputs. One can assign weights to the edges in this graph and search for *optimal* paths.

REFERENCES

[1] G. Reeb, *Sur les points singuliers d'une forme de Pfaff complètement intégrable ou d'une fonction numérique.* C.R. Acad. Sci. Paris, 222:847-849, 1946.

[2] A. Hatcher, *Algebraic topology.* Cambridge University Press, Cambridge, 2002.
[3] G.Singh, F. Mémoli, G. Carlsson, *Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition.* Point Based Graphics, 91-100, 2007.

(Multidimensional) Persistence

Afra Zomorodian
(joint work with Gunnar Carlsson and Gurjeet Singh)

Persistence refers to one type of multi-scale analysis: the technique of identifying features by analyzing geometric *histories* of data. The key notion is that features *persist* through history while noise is short-lived. In this talk, we review two theories of persistence: persistent homology [9] and multidimensional persistence [2]. We also discuss algorithms for both theories, including recent work [1].

We model scientific data as a set of d-dimensional points S augmented with $n-1$ real-valued functions $f_j : S \to \mathbb{R}$, $1 \leq j \leq n-1$, defined at the samples, for $n > 1$. Our representation for such data is a multifiltered complex. A topological space X is *multifiltered* if we are given a family of subspaces $\{X_u\}_u$, where $u \in \mathbb{N}^n$ and $X_u \subseteq X$ such that for $u, w_1, w_2, v \in \mathbb{N}^n$, the diagrams

$$\begin{array}{ccc} X_u & \longrightarrow & X_{w_1} \\ \downarrow & & \downarrow \\ X_{w_2} & \longrightarrow & X_v \end{array} \tag{1}$$

commute whenever $u \leq w_1, w_2 \leq v$. We call the family of subspaces $\{X_u\}_u$ a *multifiltration.*

To generate multifiltrations, we first capture the multi-scale connectivity of S with a one-parameter family K_ϵ of cell complexes. There are a variety of techniques for constructing such complexes, such as the *Čech*, *Rips-Vietoris* [5], and *witness* [4] complexes. These techniques give a *scale function* $\epsilon : K \to \mathbb{R}$, where K is the complex at maximum scale $\epsilon_{\max}$, and $\epsilon(\sigma)$ is when cell σ enters K. Extending the functions f_j piecewise-linearly over K, we may define $F : K \to \mathbb{R}^n$, where $F(\sigma) = (f_1(\sigma), f_2(\sigma), \ldots, f_{n-1}(\sigma), \epsilon(\sigma))$. We filter K via the *excursion sets* $\{K_u\}_u$ of F: $K_u = \{\sigma \in K \mid F(\sigma) \leq u \in \mathbb{R}^n\}$. Excursion sets are also called *sublevel sets* and are inspired by Morse theory. These complexes form an n-dimensional multifiltration [2]. Note that cell σ enters K_u at $u = F(\sigma)$ and will remain in the complex for all $u \geq F(\sigma)$. We call this property 1-critical.

When S does not come with functions f_j, we simply have a one-dimensional multifiltration or a *filtration.* The theory of *persistent homology* captures the homology of a filtration. In any dimension, this homology corresponds to a graded $R[t]$-module, where $R[t]$ is the ring of polynomials with indeterminate t over ring R. Over fields k, $k[t]$ is a *principal ideal domain (PID)*, so a consequence of the

standard structure theorem for graded $k[t]$-modules gives the full classification:

$$\bigoplus_{i=1}^{n} \Sigma^{\alpha_i} F[t] \;\oplus\; \bigoplus_{j=1}^{m} \Sigma^{\gamma_j} F[t]/(t^{n_j}), \tag{2}$$

where Σ^{α} denotes an α-shift upward in grading. The classification gives us n half-infinite intervals $[\alpha^i, \infty)$ and m finite intervals $[\gamma_j, \gamma_j + n_j)$. The multiset of $n+m$ intervals is a complete discrete invariant. We call this multiset the *persistence barcode*.

This algebraic framework also enables us to derive an algorithm for persistent homology directly from the standard *reduction algorithm* in algebraic topology [6]. The *persistence algorithm* has complexity $\Theta(n^3)$, although it has shown a linear running time in practice. While initially defined for simplicial complexes, the theory extends to singular homology, and the algorithm extends to a broad class of filtered cell complexes. A *generic* implementation of this algorithm is part of PLEX, a library of MATLAB routines developed at Stanford University for manipulating simplicial objects [8].

The theory of *multidimensional persistence* captures the homology of a multifiltration. In any dimension, this homology corresponds to an n-graded A_n-module M, where $A_n = k[x_1, \ldots, x_n]$ is the n-graded module of polynomials with n indeterminates over a field k. Unlike its one-dimensional counterpart, A_n is not a PID and A_n-modules have no structure theorem. Nevertheless, we establish a full classification of this structure in terms of three invariants. The first invariant, $\xi_0(M)$ is the multiset of generators for the free approximation of M. The second invariant, $\xi_1(M)$ is the multiset of generators for the *free hull* of M. These invariants have intuitive meaning as analogs of the left and right endpoints of the intervals in a barcode, respectively. Unfortunately, we show that there is no way to *match* these endpoints consistently as the remaining invariant corresponds to the set of orbits of an algebraic group action on an algebraic variety. Unfortunately, such as set is not, in general, an algebraic variety. The number of orbits may be uncountable, giving us a *continuous* invariant.

Multidimensional persistence shows that no complete discrete invariant exists for parametrized data. To capture persistence information, we have proposed a new discrete invariant by extending the group definition of persistent homology. For each pair $u, v \in \mathbb{N}^n$ with $u \leq v$, $X_u \subseteq X_v$ by definition, so $X_u \hookrightarrow X_v$. This inclusion, in turn, induces an injection $\iota_i(u, v)$ at the ith homology level $H_i(X_u) \to H_i(X_v)$ that maps a homology class in X_u to the one that contains it in X_v. The *ith persistence module* is the image of ι_i for all pairs $u \leq v$. The *ith rank invariant* is $\rho_i(u, v) = \operatorname{rank} \iota_i(u, v)$. for all pairs $u \leq v \in \mathbb{N}^n$, where ι_i is the injection defined above. We prove that the rank invariant is equivalent to the persistent barcode in the one-dimensional case, so it is complete when it can be [2]. Unlike the barcode, the rank invariant extends to higher dimensions as an incomplete but useful invariant.

For an n-dimensional multifiltration of m cells, a naive computation of the rank invariant gives an $O(m^{2n+3})$ time algorithm. To store the rank invariant, we also

require $\Theta(m^{2n})$ space. Neither is feasible. We have recently utilized algorithms from computational algebraic geometry for direct computation of multidimensional persistence [3]. For 1-critical multifiltrations, the boundary operator is a linear homomorphisms between free n-graded free chain modules over A_n and may be written as a matrix. To compute homology, we have three tasks, all of which may be translated into problems within algebraic geometry.

(1) Compute the boundary module ($\operatorname{im}\partial_{i+1}$): Intuitively, this is analogous to computing the *range space* of the matrix for ∂_{i+1}. In this case, this is the *submodule membership problem (SMP)*. This problem is solved by computing a *Gröbner* basis for the submodule using the *Buchberger* algorithm. We may then check membership using the division algorithm for multivariate polynomials.
(2) Compute the cycle module ($\ker\partial_i$): As before, this is analogous to the *null space* of the boundary operator's matrix. With multivariate polynomial entries, the problem is equivalent to computing the *syzygy submodule* [3]. This problem is solved using *Schreyer's algorithm*.
(3) Compute the quotient H_i: All we need to do is test whether the generators of the syzygy submodule (cycles) are in the boundary submodule. This is an instance of the SMP problem as in our first task.

While the above algorithms exist, they are not widely used for analysis due to their complexity. The SMP problem is a generalization of the *polynomial ideal membership problem (PIMP)* at the ring level, and PIMP is already EXPSPACE-complete, requiring exponential space and time [7]. However, our current work shows that in our setting, the algorithms are poly-time, requiring $O(n^7)$ time and $O(n^4)$ space. This significant reduction is due to the additional structure coming from the multigrading as the matrix entries are homogeneous monomials. This time and space bound still imply that computation is out of reach for large datasets with our current algorithms.

Open questions include efficient algorithms for computing the rank invariant, defining alternate discrete invariants that contain persistence information, and new theories of persistence for other classes of data, such as dynamic data, as discussed by Vin de Silva in this workshop.

References

[1] G. Carlsson, G. Singh, and A. Zomorodian. Computing multidimensional persistence, 2008. Manuscript.
[2] G. Carlsson and A. Zomorodian. The theory of multidimensional persistence. In *Proc. ACM Symposium on Computational Geometry*, pages 184–193, 2007. (Invited to Discrete & Computational Geometry).
[3] D. A. Cox, J. Little, and D. O'Shea. *Using algebraic geometry*, volume 185 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 2005.
[4] V. de Silva and G. Carlsson. Topological estimation using witness complexes. In *Proc. IEEE Eurographics Symposium on Point-Based Graphics*, pages 157–166, 2004.
[5] M. Gromov. Hyperbolic groups. In S. Gersten, editor, *Essays in Group Theory*, pages 75–263. Springer-Verlag, New York, NY, 1987.
[6] A. Hatcher. *Algebraic Topology*. Cambridge University Press, New York, NY, 2002.

[7] E. W. Mayr. Some complexity results for polynomial ideals. *Journal of Complexity*, 13(3):303–325, 1997.
[8] PLEX: Simplicial complexes in MATLAB. `http://math.stanford.edu/comptop/programs`.
[9] A. Zomorodian and G. Carlsson. Computing persistent homology. *Discrete & Computational Geometry*, 33(2):249–274, 2005.

Reporters: Gunnar Carlsson, Dmitry N. Kozlov

Participants

Prof. Dr. Eric Babson
Department of Mathematics
University of California, Davis
One Shields Avenue
Davis CA 95616-8633
USA

Prof. Dr. Peter Bubenik
Dept. of Mathematics
Cleveland State University
2121 Euclid Avenue
Cleveland, OH 44115
USA

Prof. Dr. Gunnar Carlsson
Department of Mathematics
Stanford University
Stanford, CA 94305-2125
USA

Peter Csorba
Department of Mathematics and
Computer Science
Eindhoven University of Technology
Postbus 513
NL-5600 MB Eindhoven

Prof. Dr. Sarah Day
Dept. of Mathematics
College of William and Mary
P.O. Box 8795
Williamsburg, VA 23187-8795
USA

Prof. Dr. Vin De Silva
Department of Mathematics
Pomona College
610 N. College Ave.
Claremont, CA 91711-6348
USA

Emanuele Delucchi
Dept. of Mathematical Sciences
State University of New York
at Binghamton
Binghamton, NY 13902-6000
USA

Prof. Dr. Eva Maria Feichtner
Universität Bremen
FB3 Mathematik und Informatik
28344 Bremen

Prof. Dr.
Dmitry N. Feichtner-Kozlov
Universität Bremen
FB3 Mathematik und Informatik
28344 Bremen

Prof. Dr. Robert Ghrist
Dept. of Mathematics, University of
Illinois at Urbana-Champaign
273 Altgeld Hall MC-382
1409 West Green Street
Urbana, IL 61801-2975
USA

Dr. Joachim Giesen
Max-Planck-Institut für Informatik
Geb. 46
Stuhlsatzenhausweg 85
66123 Saarbrücken

Dr. Tigran Ishkanov
Department of Mathematics
Stanford University
Stanford, CA 94305-2125
USA

Prof. Dr. John Frederick Jardine
Department of Mathematics
The University of Western Ontario
London ONT N6A 5B7
CANADA

Dr. Matthew K. Kahle
Department of Mathematics
Stanford University
Stanford, CA 94305-2125
USA

Prof. Dr. Kevin P. Knudson
Dept. of Math. and Statistics
Mississippi State University
P.O.Drawer MA
Mississippi State, MS 39762
USA

Juliane Lehmann
Fachbereich 3
Mathematik und Informatik
Universität Bremen
Bibliothekstr.1
28359 Bremen

Prof. Dr. Robert D. MacPherson
School of Mathematics
Institute for Advanced Study
1 Einstein Drive
Princeton, NJ 08540
USA

Dr. Facundo Memoli
Department of Mathematics
Stanford University
Stanford, CA 94305-2125
USA

Prof. Dr. Konstantin Mischaikow
Department of Mathematics
Rutgers University
Hill Center, Busch Campus
110 Frelinghuysen Road
Piscataway, NJ 08854-8019
USA

Dr. Martin Raußen
Department of Mathematical Sciences
Aalborg University
Fredrik Bajers Vej 7G
DK-9220 Aalborg Ost

Prof. Dr. Pedro Real
Departamento de Matemática Aplicada I
ETS de Ingeniería Informática
Avenida Reina Mercedes sn
E-41012 Sevilla

Prof. Dr. Gurjeet Singh
Department of Mathematics
Stanford University
Stanford, CA 94305-2125
USA

Prof. Dr. Ross E. Staffeldt
New Mexico State University
College of Arts and Sciences
Department of Mathematical Sciences
Box 30001
Las Cruces NM 88003-0001
USA

Prof. Dr. Alek Vainshtein
Dept. of Mathematics & Computer Sciences
University of Haifa
Mount Carmel
Haifa 31905
ISRAEL

Dr. Mikael Vejdemo Johansson
Mathematisches Institut
Friedrich-Schiller-Universität Jena
07737 Jena

Prof. Dr. Afra Zomorodian
Dept. of Computer Science
Dartmouth College
Hanover, NH 03755
USA