AMERICAN MATHEMATICAL SOCIETY

CURRENT EVENTS BULLETIN

Friday, January 8, 2016, 1:00 PM to 5:00 PM

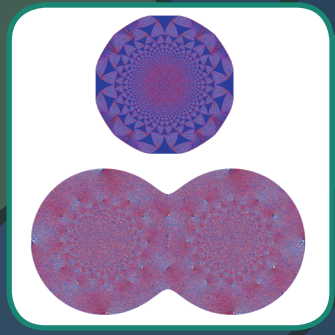
Room 4C-3 Washington State Convention Center

Joint Mathematics Meetings, Seattle, WA

1:00 PM Carina Curto, *Pennsylvania State University*

What can topology tell us about the neural code?

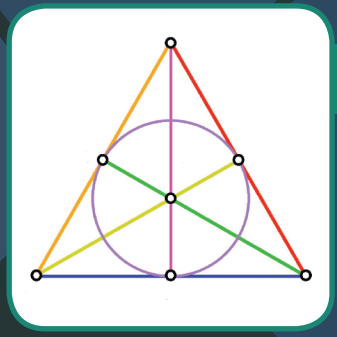
Surprising new applications of what used to be thought of as "pure" mathematics.



2:00 PM Lionel Levine, *Cornell University,* and Yuval Peres, *Microsoft Research and University of California, Berkeley*

Laplacian growth, sandpiles and scaling limits

Striking large-scale structure arising from simple cellular automata.



3:00 PM Timothy Gowers, *Cambridge University*

Probabilistic combinatorics and the recent work of Peter Keevash

The major existence conjecture for combinatorial designs has been proven!

4:00 PM Amie Wilkinson, *University of Chicago*

What are Lyapunov exponents, and why are they interesting?

A basic tool in understanding the predictability of physical systems, explained.

Organized by David Eisenbud,
Mathematical Sciences Research Institute

AMS
AMERICAN MATHEMATICAL SOCIETY

Introduction to the Current Events Bulletin

Will the Riemann Hypothesis be proved this week? What is the Geometric Langlands Conjecture about? How could you best exploit a stream of data flowing by too fast to capture? I think we mathematicians are provoked to ask such questions by our sense that underneath the vastness of mathematics is a fundamental unity allowing us to look into many different corners -- though we couldn't possibly work in all of them. I love the idea of having an expert explain such things to me in a brief, accessible way. And I, like most of us, love common-room gossip.

The Current Events Bulletin Session at the Joint Mathematics Meetings, begun in 2003, is an event where the speakers do not report on their own work, but survey some of the most interesting current developments in mathematics, pure and applied. The wonderful tradition of the Bourbaki Seminar is an inspiration, but we aim for more accessible treatments and a wider range of subjects. I've been the organizer of these sessions since they started, but a varying, broadly constituted advisory committee helps select the topics and speakers. Excellence in exposition is a prime consideration.

A written exposition greatly increases the number of people who can enjoy the product of the sessions, so speakers are asked to do the hard work of producing such articles. These are made into a booklet distributed at the meeting. Speakers are then invited to submit papers based on them to the *Bulletin of the AMS*, and this has led to many fine publications.

I hope you'll enjoy the papers produced from these sessions, but there's nothing like being at the talks -- don't miss them!

David Eisenbud, Organizer
Mathematical Sciences Research Institute
de@msri.org

Color graphics: Any graphics created in color will be rendered in grayscale for the printed version. Color graphics will be available in the online version of the 2016 *Current Events Bulletin*.

For PDF files of talks given in prior years, see
http://www.ams.org/ams/current-events-bulletin.html.
The list of speakers/titles from prior years may be found at the end of this booklet.

WHAT CAN TOPOLOGY TELL US ABOUT THE NEURAL CODE?

CARINA CURTO

Abstract. Neuroscience is undergoing a period of rapid experimental progress and expansion. New mathematical tools, previously unknown in the neuroscience community, are now being used to tackle fundamental questions and analyze emerging data sets. Consistent with this trend, the last decade has seen an uptick in the use of topological ideas and methods in neuroscience. In this talk I will survey recent applications of topology in neuroscience, and explain why topology is an especially natural tool for understanding neural codes.

1. Introduction

Applications of topology to scientific domains outside of pure mathematics are becoming increasingly common. Neuroscience, a field undergoing a golden age of progress in its own right, is no exception. The first reason for this is perhaps obvious – at least to anyone familiar with topological data analysis. Like other areas of biology, neuroscience is generating a lot of new data, and some of these data can be better understood with the help of topological methods. A second reason is that a significant portion of neuroscience research involves studying networks, and networks are particularly amenable to topological tools. Although my talk will touch on a variety of such applications, most of my attention will be devoted to a third reason – namely, that many interesting problems in neuroscience contain topological questions in disguise. This is especially true when it comes to understanding *neural codes*, and questions such as: how do the collective activities of neurons represent information about the outside world?

I will begin this talk with some well-known examples of neural codes, and then use them to illustrate how topological ideas naturally arise in this context. Next, I'll take a brief detour to describe other uses of topology in neuroscience. Finally, I will return to neural codes and explain why topological methods are helpful for studying their intrinsic properties. Taken together, these developments suggest that topology is not only useful for analyzing neuroscience data, but may also play a fundamental role in the theory of how the brain works.

Received by the editors November 10, 2015.

2. Neurons: nodes in a network or autonomous sensors?

It has been known for more than a century, since the time of Golgi and Ramon y Cajal, that the neurons in our brains are connected to each other in vast, intricate networks. Neurons are electrically active cells. They communicate with each other by firing action potentials (spikes) – tiny messages that are only received by neighboring (synaptically-connected) neurons in the network. Suppose we were eavesdropping on a single neuron, carefully recording its electrical activity at each point in time. What governs the neuron's behavior? The obvious answer: it's the network, of course! If we could monitor the activity of all the other neurons, and we knew exactly the pattern of connections between them, and were blessed with an excellent model describing all relevant dynamics, then (maybe?) we would be able to predict when our neuron will fire. If this seems hopeless now, imagine how unpredictable the activity of a single neuron in a large cortical network must have seemed in the 1950s, when Hodgkin and Huxley had just finished working out the complex nonlinear dynamics of action potentials for a simple, isolated cell [30].

And yet, around 1959, a miracle happened. It started when Hubel and Wiesel inserted a microelectrode into the primary visual cortex of an anesthetized cat, and eavesdropped on a single neuron. They could neither monitor nor control the activity of any other neurons in the network – they could only listen to one neuron at a time. What they *could* control was the visual stimulus. In an attempt to get the neuron to fire, they projected black and white patterns on a screen in front of the open-eyed cat. Remarkably, they found that the neuron they were listening to fired rapidly when the screen showed a black bar at a certain angle – say, 45°. Other neurons responded to different angles. It was as though each neuron was a sensor for a particular feature of the visual scene. Its activity could be predicted without knowing anything about the network, but by simply looking *outside* the cat's brain – at the stimulus on the screen.

Hubel and Wiesel had discovered orientation-tuned neurons [19], whose collective activity comprises a *neural code* for angles in the visual field (see Figure 1B). Although they inhabit a large, densely-connected cortical network, these neurons do not behave as unpredictable units governed by complicated dynamics. Instead, they appear to be responding directly to stimuli in the outside world. Their activity has *meaning*.

A decade later, O'Keefe made a similar discovery, this time involving neurons in a different area of the brain – the hippocampus. Unlike the visual cortex, there is no obvious sensory pathway to the hippocampus. This made it all the more mysterious when O'Keefe reported that his neurons were responding selectively to different locations in the animal's physical environment [26]. These neurons, dubbed *place cells*, act as position sensors in space. When an animal is exploring a particular environment, a place cell increases its firing rate as the animal passes through its corresponding

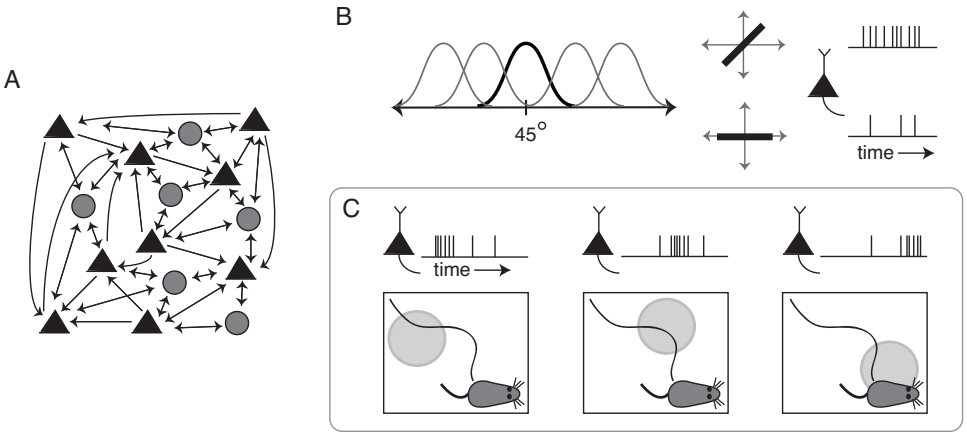


FIGURE 1. The neural network and neural coding pictures. (A) Pyramidal neurons (triangles) are embedded in a recurrent network together with inhibitory interneurons (circles). (B) An orientation-tuned neuron in primary visual cortex with a preferred angle of 45°. The neuron fires many spikes in response to a bar at a 45° angle in the animal's visual field, but few spikes in response to a horizontal bar. (C) Place cells in the hippocampus fire when the animal passes through the corresponding place field. The activity of three different neurons is shown (top), while the animal traces a trajectory starting at the top left corner of its environment (bottom). Each neuron's activity is highest when the animal passes through the corresponding place field (shaded disc).

place field – that is, the localized region to which the neuron preferentially responds (see Figure 1C).

Like Hubel and Wiesel, who received a Nobel prize for their work in 1981 [1], O'Keefe's discovery of place cells had an enormous impact in neuroscience. In 2014, he shared the Nobel prize with Edvard and May-Britt Moser [5], former postdocs of his who went on to discover an even stranger class of neurons that encode position, in a neighboring area of hippocampus called the entorhinal cortex. These neurons, called *grid cells*, display periodic place fields that are arranged in a hexagonal lattice. We'll come back to grid cells in the next section.

So, are neurons nodes in a network? or autonomous sensors of the outside world? Both pictures are valid, and yet they lead to very different models of neural behavior. Neural network theory deals with the first picture, and seeks to understand how the activity of neurons emerges from properties of the network. In contrast, neural coding theory often treats the network as a black box, focusing instead on the relationship between neural activity and external stimuli. Many of the most interesting problems in neuroscience are about *understanding the neural code*. This includes, but is not limited to, figuring out the basic principles by which neural activity represents sensory

inputs to the eyes, nose, ears, whiskers, and tongue. Because of the discoveries of Hubel and Wiesel, O'Keefe, and many others, we often know more about the coding properties of single neurons than we do about the networks to which they belong. But many open questions remain. And topology, as it turns out, is a natural tool for understanding the neural code.

3. Topology of hippocampal place cell codes

The term *hippocampal place cell code* refers to the neural code used by place cells in the hippocampus to encode the animal's position in space. Most of the research about place cells, including O'Keefe's original discovery, has been performed in rodents (typically rats), and the experiments typically involve an animal moving around in a restricted environment (see Figure 1C). It was immediately understood that a population of place cells, each having a different place field, could collectivity encode the animal's position in space [27], even though for a long time electrophysiologists could only monitor one neuron at a time. When simultaneous recordings of place cells became possible, it was shown via statistical inference (using previously measured place fields) that the animal's position could indeed be inferred from population place cell activity [3]. Figure 2 shows four place fields corresponding to simultaneously recorded place cells in area CA1 of rat hippocampus.

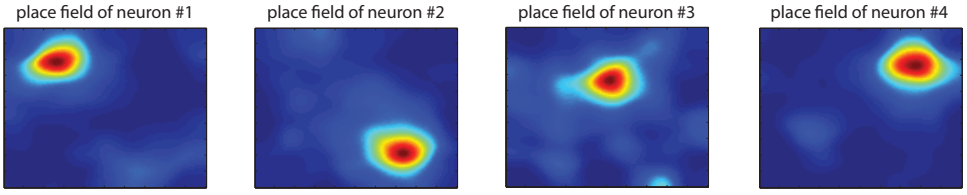


Figure 2. Place fields for four place cells, recorded while a rat explored a 2-dimensional square box environment. Place fields were computed from data provided by the Pastalkova lab.

The role of topology in place cell codes begins with a simple observation, which is perhaps obvious to anyone familiar with both place fields in neuroscience and elementary topology. First, let's recall the standard definitions of an open cover and a good cover.

Definition 3.1. Let X be a topological space. A collection of open sets, $\mathcal{U} = \{U_1, \ldots, U_n\}$, is an *open cover* of X if $X = \bigcup_{i=1}^{n} U_i$. We say that $\mathcal{U}$ is a *good cover* if every non-empty intersection $\bigcap_{i\in\sigma} U_i$, for $\sigma \subseteq \{1, \ldots, n\}$, is contractible.

Next, observe that a collection of place fields in a particular environment looks strikingly like an open cover, with each U_i corresponding a place field. Figure 3 displays three different environments, typical of what is used in hippocampal experiments with rodents, together with schematic arrangements of place fields in each.

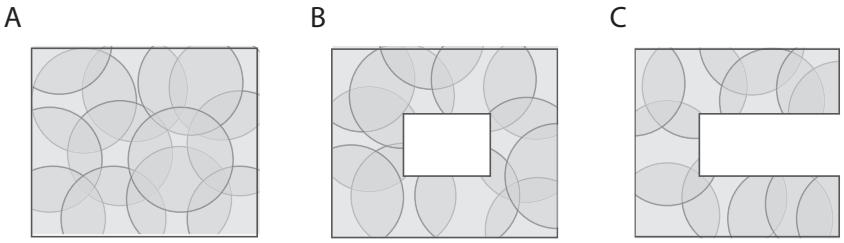


FIGURE 3. Three environments for a rat: (A) A square box environment, also known as an "open field"; (B) an environment with a hole or obstacle in the center; and (C) a maze with two arms. Each environment displays a collection of place fields (shaded discs) that fully cover the underlying space.

Moreover, since place fields are approximately convex (see Figure 2) it is not unreasonable to assume that they form a good cover of the underlying space. This means the Nerve Lemma applies. Recall the notion of the *nerve*[1] of a cover:

$$\mathcal{N}(\mathcal{U}) \stackrel{\text{def}}{=} \{\sigma \subset [n] \mid \bigcap_{i\in\sigma} U_i \neq \emptyset\},$$

where $[n] = \{1, \ldots, n\}$. Clearly, if $\sigma \in \mathcal{N}(\mathcal{U})$ and $\tau \subset \sigma$, then $\tau \in \mathcal{N}(\mathcal{U})$. This property shows that $\mathcal{N}(\mathcal{U})$ is an abstract *simplicial complex* on the vertex set $[n]$ – that is, it is a set of subsets of $[n]$ that is closed under taking further subsets. If X is a sufficiently "nice" topological space, then the following well-known lemma holds.

Lemma 3.2 (Nerve Lemma)**.** *Let $\mathcal{U}$ be a good cover of X. Then $\mathcal{N}(\mathcal{U})$ is homotopy-equivalent to X. In particular, $\mathcal{N}(\mathcal{U})$ and X have exactly the same homology groups.*

It is important to note that the Nerve Lemma fails if the good cover assumption does not hold. Figure 4A depicts a good cover of an annulus by three open sets. The corresponding nerve (right) exhibits the topology of a circle, which is indeed homotopy-equivalent to the covered space. In Figure 4B, however, the cover is *not* good, because the intersection $U_1 \cap U_2$ consists of two disconnected components, and is thus not contractible. Here the nerve (right) is homotopy-equivalent to a point, in contradiction to the topology of the covered annulus.

The wonderful thing about the Nerve Lemma, when interpreted in the context of hippocampal place cells, is that $\mathcal{N}(\mathcal{U})$ can be inferred from the activity of place cells alone – without actually knowing the place fields $\{U_i\}$. This is because the concurrent activity of a group of place cells, indexed

[1]Note that the name "nerve" here predated any connection to neuroscience!

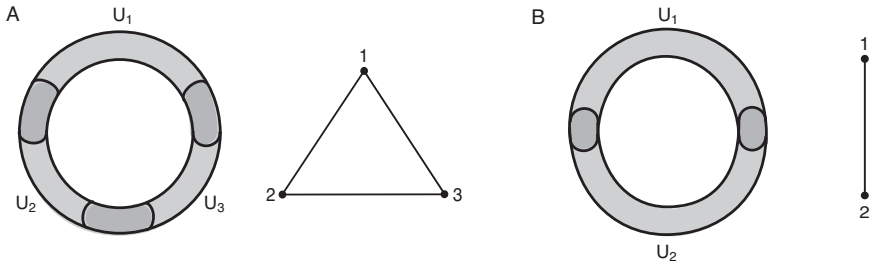


FIGURE 4. Good and bad covers. (A) A good cover $\mathcal{U} = \{U_1, U_2, U_3\}$ of an annulus (left), and the corresponding nerve $\mathcal{N}(\mathcal{U})$ (right). (B) A "bad" cover of the annulus (left), and the corresponding nerve (right). Only the nerve of the good cover accurately reflects the topology of the annulus.

by $\sigma \subset [n]$, indicates that the corresponding place fields have a non-empty intersection: $\bigcap_{i\in\sigma} U_i \neq \emptyset$. In other words, if we were eavesdropping on the activity of a population of place cells as the animal fully explored its environment, then by finding which subsets of neurons co-fire (see Figure 5) we could in principle estimate $\mathcal{N}(\mathcal{U})$, even if the place fields themselves were unknown. Lemma 3.2 tells us that the homology of the simplicial complex $\mathcal{N}(\mathcal{U})$ precisely matches the homology of the environment X. The place cell code thus naturally reflects the topology of the represented space.[2]

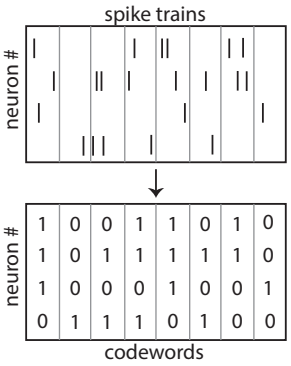


FIGURE 5. By binning spike trains for a population of simultaneously-recorded neurons, one can infer subsets of neurons that co-fire. If these neurons were place cells, then the first codeword 1110 indicates that $U_1 \cap U_2 \cap U_3 \neq \emptyset$, while the third codeword 0101 tells us $U_2 \cap U_4 \neq \emptyset$.

These and related observations have led some researchers to speculate that the hippocampal place cell code is fundamentally topological in nature

[2]In particular, place cell activity from the environment in Figure 3B could be used to detect the non-trivial first homology group of the underlying space, and thus distinguish this environment from that of Figure 3A or 3C.

[12, 6], while others (including this author) have argued that considerable geometric information is also present and can be extracted using topological methods [9, 18]. In order to disambiguate topological and geometric features, Dabaghian et. al. performed an elegant experiment using linear tracks with flexible joints [11]. This allowed them to alter geometric features of the environment, while preserving the topological structure as reflected by the animal's place fields. They found that place fields recorded from an animal running along the morphing track moved together with the track, preserving the relative sequence of locations despite changes in angles and movement direction. In other words, the place fields respected topological aspects of the environment more than metric features [11].

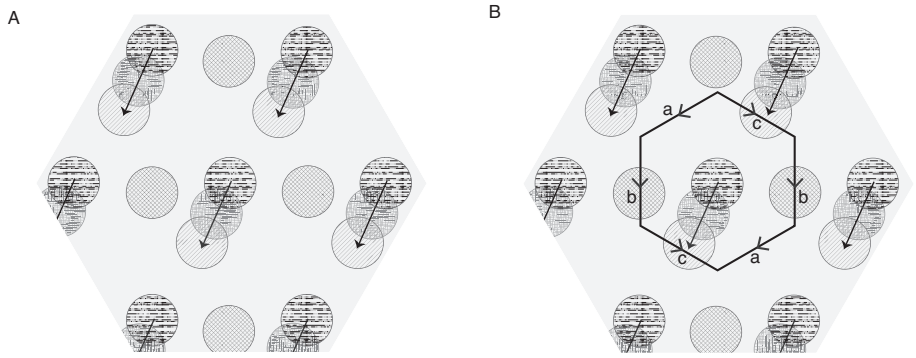


FIGURE 6. Firing fields for grid cells. (A) Firing fields for four entorhinal grid cells. Each grid field forms a hexagonal grid in the animal's two-dimensional environment, and each field thus has multiple disconnected regions. (B) A hexagonal fundamental domain contains just one disc-like region per grid cell. Pairs of edges with the same label (a, b, or c) are identified, with orientations specified by the arrows.

What about the entorhinal grid cells? These neurons have firing fields with multiple disconnected components, forming a hexagonal grid (see Figure 6A). This means that grid fields violate the good cover assumption of the Nerve Lemma – if we consider them as an open cover for the entire 2-dimensional environment. If, instead, we restrict attention to a fundamental domain for these firing fields, as illustrated in Figure 6B, then each grid field has just one (convex) component, and the Nerve Lemma applies. From the spiking activity of grid cells we could thus infer the topology of this fundamental domain. The reader familiar with the topological classification of surfaces may recognize that this hexagonal domain, with the identification of opposite edges, is precisely a torus. To see this, first identify the edges labeled "a" to get a cylinder. Next, observe that the boundary circles on each end of the cylinder consist of the edges "b" and "c", but with a 180°

twist between the two ends. By twisting the cylinder, the two ends can be made to match so that the "b" and "c" edges get identified. This indicates that the space represented by grid cells is not the full environment, but a torus.

4. Topology in neuroscience: a bird's-eye view

The examples from the previous section are by no means the only way that topology is being used in neuroscience. Before plunging into further details about what topology can tell us about neural codes, we now pause for a moment to acknowledge some other interesting applications. The main thing they all have in common is their recency. This is no doubt due to the rise of computational and applied algebraic topology, a relatively new development in applied math that was highlighted in the Current Events Bulletin nearly a decade ago [14].

Roughly speaking, the uses of topology in neuroscience can be categorized into three (overlapping) themes: (i) "traditional" topological data analysis applied to neuroscience; (ii) an upgrade to network science; and (iii) understanding the neural code. Here we briefly summarize work belonging to (i) and (ii). In the next section we'll return to (iii), which is the main focus of this talk.

4.1. "Traditional" TDA applied to neuroscience data sets.

The earliest and most familiar applications of topological data analysis (TDA) focused on the problem of estimating the "shape" of point-cloud data. This kind of data set is simply a collection of points, $x_1, \ldots, x_\ell \in \mathbb{R}^n$, where n is the dimensionality of the data. A question one could ask is: do these points appear to have been sampled from a lower-dimensional manifold, such as a torus or a sphere? The strategy is to consider open balls $B_\varepsilon(x_i)$ of radius ε around each data point, and then to construct a simplicial complex $\mathcal{K}_\varepsilon$ that captures information about how the balls intersect. This simplicial complex can either be the Cech complex (i.e., the nerve of the open cover defined by the balls), or the Vietoris-Rips complex (i.e., the clique complex of the graph obtained from pairwise intersections of the balls). By varying ε, one obtains a sequence of nested simplicial complexes $\{\mathcal{K}_\varepsilon\}$ together with natural inclusion maps. Persistent homology tracks homology cycles across these simplicial complexes, and allows one to determine whether there were homology classes that "persisted" for a long time. For example, if the data points were sampled from a 3-sphere, one would see a persistent 3-cycle.

There are many excellent reviews of persistent homology, including [14], so I will not go into further details here. Instead, it is interesting to note that one of the early applications of these techniques was in neuroscience, to analyze population activity in primary visual cortex [31]. Here it was found that the topological structure of activity patterns is similar between spontaneous and evoked activity, and consistent with the topology of a two-sphere. Moreover, the results of this analysis were interpreted in the context

of neural coding, making this work exemplary of both themes (i) and (iii). Another application of persistent homology to point cloud data in neuroscience was the analysis of the spatial structure of afferent neuron terminals in crickets [4]. Again, the results were interpreted in terms of the coding properties of the corresponding neurons, which are sensitive to air motion detected by thin mechanosensory hairs on the cricket. Finally, it is worth mentioning that these types of analyses are not confined to neural activity. For example, in [2] the statistics of persistent cycles were used to study brain artery trees.

4.2. **An upgrade to network science.** There are many ways of constructing networks in neuroscience, but the basic model that has been used for all of them is the graph. The vertices of a graph can represent neurons, cell types, brain regions, or fMRI voxels, while the edges reflect interactions between these units. Often, the graph is weighted and the edge weights correspond to correlations between adjacent nodes. For example, one can model a functional brain network from fMRI data as a weighted graph where the edge weights correspond to activity correlations between pairs of voxels. At the other extreme, a network where the vertices correspond to neurons could have edge weights that reflect either pairwise correlations in neural activity, or synaptic connections.

Network science is a relatively young discipline that focuses on analyzing networks, primarily using tools derived from graph theory. The results of a particular analysis could range from determining the structure of a network to identifying important subgraphs and/or graph-theoretic statistics (the distribution of in-degree or out-degree across nodes, number of cycles, etc.) that carry meaning for the network at hand. Sometimes, graph-theoretic features do not carry obvious meaning, but are nevertheless useful for distinguishing networks that belong to distinct classes. For example, a feature could be characteristic of functional brain networks derived from a subgroup of subjects, distinguishing them from a "control" group. In this way graph features may be a useful diagnostic tool for distinguishing diseased states, pharmacologically-induced states, cognitive abilities, or uncovering systematic differences based on gender or age.

The recent emergence of topological methods in network science stems from the following "upgrade" to the network model: instead of a graph, one considers a simplicial complex. Sometimes this simplicial complex reflects higher-order interactions that are obtained from the data, and sometimes it is just the *clique complex* of the graph G:

$$X(G) = \{\sigma \subset [n] \mid (ij) \in G \text{ for all } i, j \in \sigma\}.$$

In other words, the higher-order simplices correspond to cliques (all-to-all connected subgraphs) of G. Figure 7A shows a graph (top) and the corresponding clique complex (bottom), with shaded simplices corresponding to two 3-cliques and a 4-clique. The clique complex fills in many of the 1-cycles

in the original graph, but some 1-cycles remain (see the gold 4-gon), and higher-dimensional cycles may emerge. Computing homology groups for the clique complex is then a natural way to detect topological features that are determined by the graph. In the case of a weighted graph, one can obtain a sequence of clique complexes by considering a related sequence of simple graphs, where each graph is obtained from the previous one by adding the edge corresponding to the next-highest weight (see Figure 7B). The corresponding sequence of clique complexes, $\{X(G_i)\}$, can then be analyzed using persistent homology. Other methods for obtaining a sequence of simplicial complexes from a network are also possible, and may reflect additional aspects of the data such as the temporal evolution of the network.

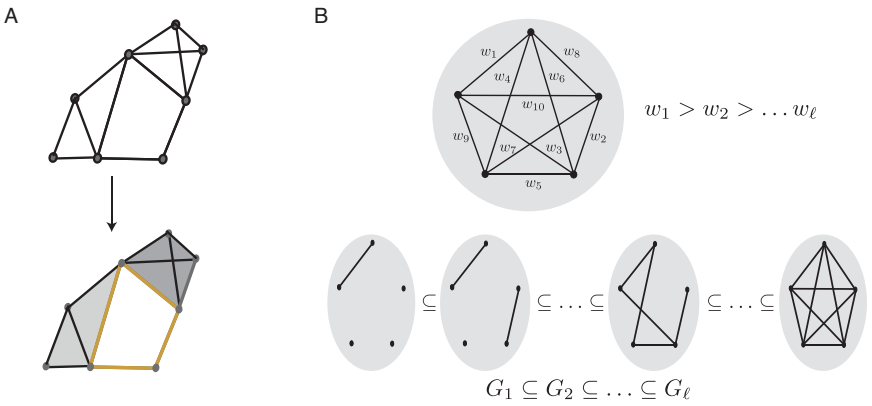


FIGURE 7. Network science models: from graphs to clique complexes and filtrations.

For a more thorough survey of topological methods in network science, I recommend the forthcoming review article [15]. Here I will only mention that topological network analyses have already been used in a variety of neuroscience applications, many of them medically-motivated: fMRI networks in patients with ADHD [13]; FDG-PET based networks in children with autism and ADHD [23]; morphological networks in deaf adults [22]; metabolic connectivity in epileptic rats [7]; and functional EEG connections in depressed mice [21]. Other applications to fMRI data include human brain networks during learning [33] and drug-induced states [28]. At a finer scale, recordings of neural activity can also give rise to functional connectivity networks among neurons (which are not the same as the neural networks defined by synaptic connections). These networks have also been analyzed with topological methods [29, 18, 32].

5. The code of an open cover

We now return to neural codes. We have already seen how the hippocampal place cell code reflects the topology of the underlying space, via the nerve $\mathcal{N}(\mathcal{U})$ of a place field cover. In this section, we will associate a binary

code to an open cover. This notion is closer in spirit to a combinatorial neural code (see Figure 5), and carries more detailed information than the nerve. In the next section, we'll see how topology is being used to determine intrinsic features of neural codes, such as convexity and dimension.

First, a few definitions. A *binary pattern* on n neurons is a string of 0s and 1s, with a 1 for each active neuron and a 0 denoting silence; equivalently, it is a subset of (active) neurons $\sigma \subset [n]$. (Recall that $[n] = \{1, \ldots, n\}$.) We use both notations interchangeably. For example, 10110 and $\sigma = \{1, 3, 4\}$ refer to the same pattern, or codeword, on $n = 5$ neurons. A *combinatorial neural code* on n neurons is a collection of binary patterns $\mathcal{C} \subset 2^{[n]}$. In other words, it is a binary code of length n, where we interpret each binary digit as the "on" or "off" state of a neuron. The *simplicial complex of a code*, $\Delta(\mathcal{C})$, is the smallest abstract simplicial complex on $[n]$ that contains all elements of $\mathcal{C}$. In keeping with the hippocampal place cell example, we are interested in codes that correspond to open covers of some topological space.

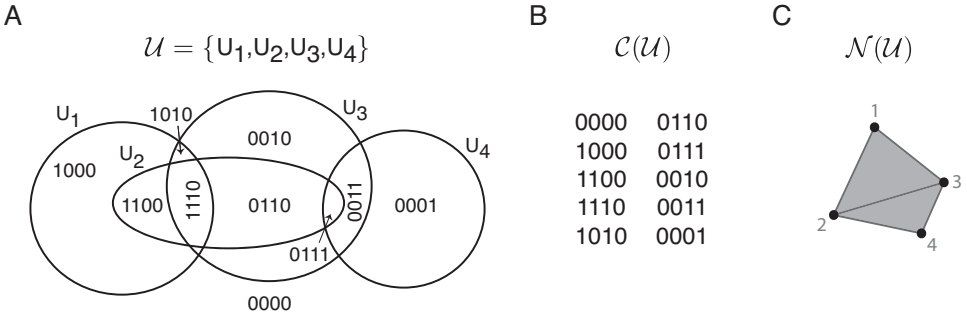


FIGURE 8. Codes and nerves of open covers. (A) An open cover $\mathcal{U}$, with each region carved out by the cover labeled by its corresponding codeword. (B) The code $\mathcal{C}(\mathcal{U})$. (C) The nerve $\mathcal{N}(\mathcal{U})$.

Definition 5.1. Given an open cover $\mathcal{U}$, the *code of the cover* is the combinatorial neural code

$$\mathcal{C}(\mathcal{U}) \stackrel{\text{def}}{=} \{\sigma \subset [n] \mid \bigcap_{i \in \sigma} U_i \setminus \bigcup_{j \in [n] \setminus \sigma} U_j \neq \emptyset\}.$$

Each codeword in $\mathcal{C}(\mathcal{U})$ corresponds to a region that is defined by the intersections of the open sets in $\mathcal{U}$ (Figure 8A). Note that the code $\mathcal{C}(\mathcal{U})$ is not the same as the nerve $\mathcal{N}(\mathcal{U})$. Figures 8B and 8C display the code and the nerve of the open cover in Figure 8A. While the nerve encodes which subsets of the U_is have non-empty intersections, the code also carries information about set containments. For example, the fact that $U_2 \subseteq U_1 \cup U_3$ can be inferred from $\mathcal{C}(\mathcal{U})$ because each codeword of the form $*1**$ has an additional 1 in position 1 or 3, indicating that if neuron 2 is firing then so is neuron 1 or 3. Similarly, the fact that $U_2 \cap U_4 \subseteq U_3$ can be inferred from the

code because any word of the form $*1*1$ necessarily has a 1 in position 3 as well. These containment relationships go beyond simple intersection data, and cannot be obtained from the nerve $\mathcal{N}(\mathcal{U})$. On the other hand, the nerve can easily be recovered from the code since $\mathcal{N}(\mathcal{U})$ is the smallest simplicial complex that contains it – that is,

$$\mathcal{N}(\mathcal{U}) = \Delta(\mathcal{C}(\mathcal{U})).$$

$\mathcal{C}(\mathcal{U})$ thus carries more detailed information than what is available in $\mathcal{N}(\mathcal{U})$. The combinatorial data in $\mathcal{C}(\mathcal{U})$ can also be encoded algebraically via the *neural ideal* [10], much as simplicial complexes are algebraically encoded by Stanley-Reisner ideals [25].

It is easy to see that any binary code, $\mathcal{C} \subseteq \{0,1\}^n$, can be realized as the code of an open cover.[3] It is not true, however, that any code can arise from a good cover or a *convex cover* – that is, an open cover consisting of convex sets. The following lemma illustrates the simplest example of what can go wrong.

Lemma 5.2. *Let $\mathcal{C} \subset \{0,1\}^3$ be a code that contains the codewords* 110 *and* 101*, but does not contain* 100 *and* 111*. Then $\mathcal{C}$ is not the code of a good or convex cover.*

The proof is very simple. Suppose $\mathcal{U} = \{U_1, U_2, U_3\}$ is a cover such that $\mathcal{C} = \mathcal{C}(\mathcal{U})$. Because neuron 2 or 3 is "on" in any codeword for which neuron 1 is "on," we must have that $U_1 \subset U_2 \cup U_3$. Moreover, we see from the code that $U_1 \cap U_2 \neq \emptyset$ and $U_1 \cap U_3 \neq \emptyset$, while $U_1 \cap U_2 \cap U_3 = \emptyset$. This means we can write U_1 as a disjoint union of two non-empty sets: $U_1 = (U_1 \cap U_2) \cup (U_1 \cap U_3)$. U_1 is thus disconnected, and hence $\mathcal{U}$ can be neither a good nor convex cover.

6. Using topology to study intrinsic properties of neural codes

In our previous examples from neuroscience, the place cell and grid cell codes can be thought of as arising from convex sets covering an underlying space. Because the spatial correlates of these neurons are already known, it is not difficult to infer what space is being represented by these codes. What could we say if we were given just a code, $\mathcal{C} \subset \{0,1\}^n$, without *a priori* knowledge of what the neurons were encoding? Could we tell whether such a code can be realized via a convex cover?

[3]For example, if the size of the code is $|\mathcal{C}| = \ell$, we could choose disjoint open intervals $B_1, \ldots, B_\ell \subset \mathbb{R}$, one for each codeword, and define the open sets $U_1, \ldots, U_n$ such that U_i is the union of all open intervals B_j corresponding to codewords in which neuron i is "on" (that is, there is a 1 in position i of the codeword). Such a cover, however, consists of highly disconnected sets and its properties reflect very little of the underlying space – in particular, the good cover assumption of the Nerve Lemma is violated.

6.1. **What can go wrong.** As seen in Lemma 5.2, not all codes can arise from convex covers. Moreover, the problem that prevents the code in Lemma 5.2 from being convex is topological in nature. Specifically, what happens in the example of Lemma 5.2 is that the code dictates there must be a set containment,

$$U_\sigma \subseteq \bigcup_{j\in\tau} U_j,$$

where $U_\sigma = \bigcap_{i\in\sigma} U_i$, but the nerve of the resulting cover of U_σ by the sets $\{U_\sigma \cap U_j\}_{j\in\tau}$ is not contractible. This leads to a contradiction if the sets U_i are all assumed to be convex, because the sets $\{U_\sigma \cap U_j\}_{j\in\tau}$ are then also convex and thus form a good cover of U_σ. Since U_σ itself is convex, and the Nerve Lemma holds, it follows that $\mathcal{N}(\{U_\sigma \cap U_j\}_{j\in\tau})$ must be contractible, contradicting the data of the code.

These observations lead to the notion of a *local obstruction* to convexity [16], which captures the topological problem that arises if certain codes are assumed to have convex covers. The proof of the following lemma is essentially the argument outlined above.

Lemma 6.1 ([16])**.** *If $\mathcal{C}$ can be realized by a convex cover, then $\mathcal{C}$ has no local obstructions.*

The idea of using local obstructions to determine whether or not a neural code has a convex realization has been recently followed up in a series of papers [8, 24, 17]. In particular, local obstructions have been characterized in terms of links, $\mathrm{Lk}_\Delta(\sigma)$, corresponding to "missing" codewords that are not in the code, but are elements of the simplicial complex of the code.

Theorem 6.2 ([8])**.** *Let $\mathcal{C}$ be a neural code, and let $\Delta = \Delta(\mathcal{C})$. Then $\mathcal{C}$ has no local obstructions if and only if $\mathrm{Lk}_\Delta(\sigma)$ is contractible for all $\sigma \in \Delta \setminus \mathcal{C}$.*

It was believed, until very recently, that the converse of Lemma 6.1 might also be true. However, in [24] the following counterexample was discovered, showing that this is not the case. Here the term *convex code* refers to a code that can arise from a convex open cover.

Example 6.3 ([24])**.** The code $\mathcal{C} = \{2345, 123, 134, 145, 13, 14, 23, 34, 45, 3, 4\}$ is not a convex code, despite the fact that it has no local obstructions.

That this code has no local obstructions can be easily seen using Theorem 6.2. The fact that there is no convex open cover, however, relies on convexity arguments that are not obviously topological. Moreover, this code does have a good cover [24], suggesting the existence of a new class of obstructions to convexity which may or may not be topological in nature.

6.2. **What can go right.** Finally, it has been shown that several classes of neural codes are guaranteed to have convex realizations. *Intersection-complete codes* satisfy the property that for any $\sigma, \tau \in \mathcal{C}$ we also have $\sigma \cap \tau \in \mathcal{C}$. These codes (and some generalizations) were shown constructively to have

convex covers in [17]. Additional classes of codes with convex realizations have been described in [8].

Despite these developments, a complete characterization of convex codes is still lacking. Finding the minimum dimension needed for a convex realization is also an open question.

7. Codes from networks

We end by coming back to the beginning. Even if neural codes give us the illusion that neurons in cortical and hippocampal areas are directly sensing the outside world, we know that of course they are not. Their activity patterns are shaped by the networks in which they reside. What can we learn about the architecture of a network by studying its neural code? This question requires an improved understanding of neural networks, not just neural codes. While many candidate architectures have been proposed to explain, say, orientation-tuning in visual cortex, the interplay of neural network theory and neural coding is still in early stages of development.

Perhaps the simplest example of how the structure of a network can constrain the neural code is the case of simple feedforward networks. These networks have a single input layer of neurons, and a single output layer. The resulting codes are derived from hyperplane arrangements in the positive orthant of $\mathbb{R}^k$, where k is the number of neurons in the input layer and each hyperplane corresponds to a neuron in the output layer (see Figure 9). Every codeword in a *feedforward code* corresponds to a chamber in such a hyperplane arrangement.

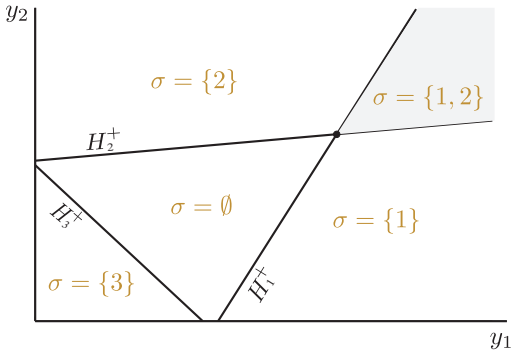


Figure 9. A hyperplane arrangement in the positive orthant, and the corresponding feedforward code.

It is not difficult to see from this picture that all feedforward codes are realizable by convex covers – specifically, they arise from overlapping halfspaces [16]. On the other hand, not every convex code is the code of a feedforward network [20]. Moreover, the discrepancy between feedforward codes and convex codes is not due to restrictions on their simplicial complexes. As was shown in [16], every simplicial complex can arise as $\Delta(\mathcal{C})$ for

a feedforward code. As with convex codes, a complete characterization of feedforward codes is still unknown. It seems clear, however, that topological tools will play an essential role.

8. Acknowledgments

I would like to thank Chad Giusti for his help in compiling a list of references for topology in neuroscience. I am especially grateful to Katie Morrison for her generous help with the figures.

References

1. *Physiology or Medicine 1981-Press Release*, 2014, *Nobelprize.org* Nobel Media AB. http://www.nobelprize.org/nobel_prizes/medicine/laureates/1981/press.html.
2. Paul Bendich, JS Marron, Ezra Miller, Alex Pieloch, and Sean Skwerer, *Persistent homology analysis of brain artery trees*, Ann. Appl. Stat. **to appear** (2014).
3. E.N. Brown, L.M. Frank, D. Tang, M.C. Quirk, and M.A. Wilson, *A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells*, J. Neurosci. **18** (1998), 7411–7425.
4. Jacob Brown and Tomás Gedeon, *Structure of the afferent terminals in terminal ganglion of a cricket and persistent homology.*, PLoS ONE **7** (2012), no. 5.
5. N. Burgess, *The 2014 Nobel Prize in Physiology or Medicine: A Spatial Model for Cognitive Neuroscience*, Neuron **84** (2014), no. 6, 1120–1125.
6. Zhe Chen, Stephen N Gomperts, Jun Yamamoto, and Matthew A Wilson, *Neural representation of spatial topology in the rodent hippocampus*, Neural Comput. **26** (2014), no. 1, 1–39.
7. Hongyoon Choi, Yu Kyeong Kim, Hyejin Kang, Hyekyoung Lee, Hyung-Jun Im, E Edmund Kim, June-Key Chung, Dong Soo Lee, et al., *Abnormal metabolic connectivity in the pilocarpine-induced epilepsy rat model: a multiscale network analysis based on persistent homology*, NeuroImage **99** (2014), 226–236.
8. C. Curto, E. Gross, J. Jeffries, K. Morrison, M. Omar, Z. Rosen, A. Shiu, and N. Youngs, *What makes a neural code convex?*, Available online at http://arxiv.org/abs/1508.00150.
9. C. Curto and V. Itskov, *Cell groups reveal structure of stimulus space*, PLoS Comp. Bio. **4** (2008), no. 10, e1000205.
10. C. Curto, V. Itskov, A. Veliz-Cuba, and N. Youngs, *The neural ring: an algebraic tool for analyzing the intrinsic structure of neural codes*, Bull. Math. Biol. **75** (2013), no. 9, 1571–1611.
11. Yuri Dabaghian, Vicky L Brandt, and Loren M Frank, *Reconceiving the hippocampal map as a topological template*, Elife **3** (2014), e03476.
12. Yuri Dabaghian, Facundo Mémoli, L Frank, and Gunnar Carlsson, *A topological paradigm for hippocampal spatial map formation using persistent homology*, PLoS Comp. Bio. **8** (2012), no. 8, e1002581.
13. Steven P Ellis and Arno Klein, *Describing high-order statistical dependence using "concurrence topology," with application to functional mri brain data*, H. H. A. **16** (2014), no. 1.
14. R. Ghrist, *Barcodes: the persistent topology of data*, Bull. Amer. Math. **45** (2008), 61–75.
15. C. Giusti, R. Ghrist, and D.S. Bassett, *Two's company, three (or more) is a simplex: Algebraic-topological tools for understanding higher-order structure in neural data*, Submitted.
16. C. Giusti and V. Itskov, *A no-go theorem for one-layer feedforward networks*, Neural Comput. **26** (2014), no. 11, 2527–2540.

17. C. Giusti, V. Itskov, and W. Kronholm, *On convex codes and intersection violators*, In preparation.
18. Chad Giusti, Eva Pastalkova, Carina Curto, and Vladimir Itskov, *Clique topology reveals intrinsic geometric structure in neural correlations*, Proceedings of the National Academy of Sciences (2015).
19. D. H. Hubel and T. N. Wiesel, *Receptive fields of single neurons in the cat's striate cortex*, J. Physiol. **148** (1959), no. 3, 574–591.
20. V. Itskov, Personal communication, 2015.
21. Arshi Khalid, Byung Sun Kim, Moo K Chung, Jong Chul Ye, and Daejong Jeon, *Tracing the evolution of multi-scale functional networks in a mouse model of depression using persistent brain network homology*, Neuroimage **101** (2014), 351–363.
22. Eunkyung Kim, Hyejin Kang, Hyekyoung Lee, Hyo-Jeong Lee, Myung-Whan Suh, Jae-Jin Song, Seung-Ha Oh, and Dong Soo Lee, *Morphological brain network assessed using graph theory and network filtration in deaf adults*, Hear. Res. **315** (2014), 88–98.
23. Hyekyoung Lee, Moo K Chung, Hyejin Kang, Bung-Nyun Kim, and Dong Soo Lee, *Discriminative persistent homology of brain networks*, Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on, IEEE, 2011, pp. 841–844.
24. C. Lienkaemper, A. Shiu, and Z. Woodstock, *Obstructions to convexity in neural codes*, Available online at http://arxiv.org/abs/1509.03328.
25. E. Miller and B. Sturmfels, *Combinatorial commutative algebra*, Graduate Texts in Mathematics, vol. 227, Springer-Verlag, New York, 2005. MR MR2110098 (2006d:13001)
26. J. O'Keefe and J. Dostrovsky, *The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat*, Brain Res. **34** (1971), no. 1, 171–175.
27. J. O'Keefe and L. Nadel, *The hippocampus as a cognitive map*, Clarendon Press Oxford, 1978.
28. G Petri, P Expert, F Turkheimer, R Carhart-Harris, D Nutt, PJ Hellyer, and Francesco Vaccarino, *Homological scaffolds of brain functional networks*, J. Roy. Soc. Int. **11** (2014), no. 101, 20140873.
29. Virginia Pirino, Eva Riccomagno, Sergio Martinoia, and Paolo Massobrio, *A topological study of repetitive co-activation networks in in vitro cortical assemblies.*, Phys. Bio. **12** (2014), no. 1, 016007–016007.
30. J. Rinzel, *Discussion: Electrical excitability of cells, theory and experiment: Review of the hodgkin-huxley foundation and update*, Bull. Math. Biol. **52** (1990), no. 1/2, 5–23.
31. Gurjeet Singh, Facundo Memoli, Tigran Ishkhanov, Guillermo Sapiro, Gunnar Carlsson, and Dario L Ringach, *Topological analysis of population activity in visual cortex*, J. Vis. **8** (2008), no. 8, 11.
32. Gard Spreemann, Benjamin Dunn, Magnus Bakke Botnan, and Nils A Baas, *Using persistent homology to reveal hidden information in neural data*, arXiv:1510.06629 [q-bio.NC] (2015).
33. Bernadette Stolz, *Computational topology in neuroscience*, Master's thesis, University of Oxford, 2014.

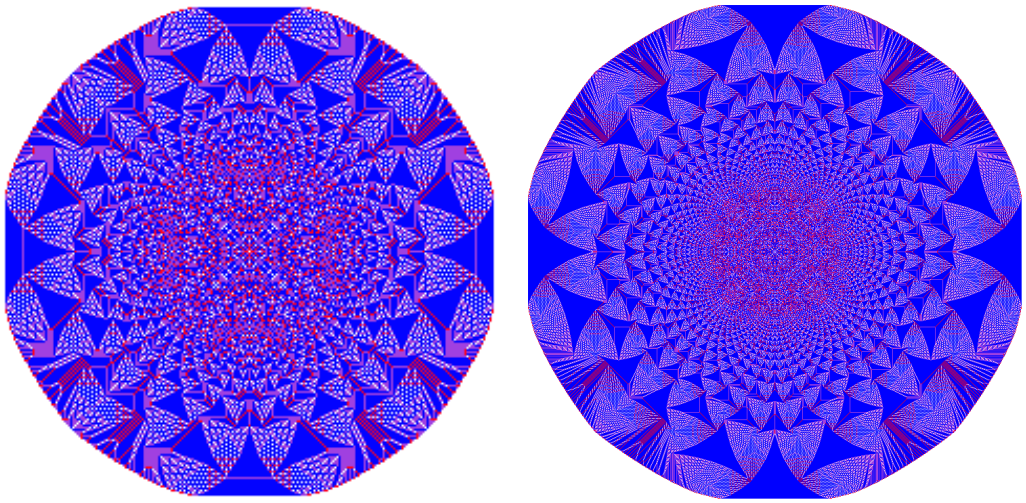
Department of Mathematics, The Pennsylvania State University
E-mail address: ccurto@psu.edu

LAPLACIAN GROWTH, SANDPILES AND SCALING LIMITS

LIONEL LEVINE AND YUVAL PERES

1. The abelian sandpile model

Start with n particles at the origin in the square grid $\mathbb{Z}^2$, and let them spread out according to the following rule: whenever any site in $\mathbb{Z}^2$ has 4 or more particles, it gives one particle to each of its 4 nearest neighbors (North, East, South and West). The final configuration of particles does not depend on the order in which these moves are performed (which explains the term "abelian"; see Lemma 1.1 below).



$n = 10^5$ $n = 10^6$

FIGURE 1. Sandpiles in $\mathbb{Z}^2$ formed by stabilizing 10^5 and 10^6 particles at the origin. Each pixel is colored according to the number of sand grains that stabilize there (white 0, red 1, purple 2, blue 3). The two images have been scaled to have the same diameter.

This model was invented in 1987 by the physicists Bak, Tang and Wiesenfeld [7]. While defined by a simple local rule, it produces self-similar global patterns that call for an explanation. Dhar [15] extended the model to any base graph and discovered the abelian property. The abelian sandpile was independently discovered by combinatorialists [10], who called it *chip-firing*. Indeed, in the last two

Date: November 11, 2015.
2010 *Mathematics Subject Classification.*

decades the subject has been enriched by an exhilirating interaction of numerous areas of mathematics, including statistical physics, combinatorics, free boundary PDE, probability, potential theory, number theory and group theory. More on this below. There are also connections to algebraic geometry [49, 8, 60], commutative algebra [52, 53] and computational complexity [55, 6, 12]. For software for experimenting with sandpiles, see [61].

Let $G = (V, E)$ be a locally finite connected graph. A **sandpile** on G is a function $s : V \to \mathbb{Z}$. We think of a positive value $s(x) > 0$ as a number of sand grains (or "particles") at vertex x, and negative value as a hole that can be filled by particles. Vertex x is **unstable** if $s(x) \geq \deg(x)$, the number of edges incident to x. **Toppling** x is the operation of sending $\deg(x)$ particles away from x, one along each incident edge. We say that a sequence of vertices $\mathbf{x} = (x_1, \ldots, x_m)$ is **legal** for s if $s_i(x_i) \geq \deg(x_i)$ for all $i = 1, \ldots, m$, where s_i is the sandpile obtained by toppling $x_1, \ldots, x_{i-1}$; we say that $\mathbf{x}$ is **stabilizing** for s if $s_m \leq \deg -1$. (All inequalities between functions are pointwise.)

Lemma 1.1. *Let $s : V \to \mathbb{Z}$ be a sandpile, and suppose there exists a sequence $\mathbf{y} = (y_1, \ldots, y_n)$ that is stabilizing for s.*

(i) Any legal sequence $\mathbf{x} = (x_1, \ldots, x_m)$ for s is a permutation of a subsequence of $\mathbf{y}$.

(ii) There exists a legal stabilizing sequence for s.

(iii) Any two legal stabilizing sequences for s are permutations of each other.

Proof. Since $\mathbf{x}$ is legal for s we have $s(x_1) \geq \deg(x_1)$. Since $\mathbf{y}$ is stabilizing for s it follows that $y_i = x_1$ for some i. Toppling x_1 yields a new sandpile s'. Removing x_1 from $\mathbf{x}$ and y_i from $\mathbf{y}$ yields shorter legal and stabilizing sequences for s', so (i) follows by induction.

Let $\mathbf{x}$ be a legal sequence of maximal length, which is finite by (i). Such $\mathbf{x}$ must be stabilizing, which proves (ii).

Statement (iii) is immediate from (i). □

We say that s **stabilizes** if there is a sequence that is stabilizing for s. If s stabilizes, we define its **odometer** as the function on vertices

$$u(x) = \text{number of occurences of } x \text{ in any legal stabilizing sequence for } s.$$

The **stabilization** $\hat{s}$ of s is the result of toppling a legal stabilizing sequence for s. The odometer determines the stabilization, since

$$\hat{s} = s + \Delta u \tag{1}$$

where Δ is the **graph Laplacian**

$$\Delta u(x) = \sum_{y \sim x} (u(y) - u(x)). \tag{2}$$

Here the sum is over vertices y that are neighbors of x.

By Lemma 1.1(iii), both the odometer u and the stabilization $\hat{s}$ depend only on s, and not on the choice of legal stabilizing sequence, which is one reason the model is called **abelian** (another is the role played by an abelian group; see Section 7).

What does a very large sandpile look like? The similarity of the two sandpiles in Figure 1 suggests that some kind of limit exists as we take the number of particles $n \to \infty$ while "zooming out" so that each square of the grid has area $1/n$. The first step toward making this rigorous is to reformulate Lemma 1.1 in terms of the Laplacian as follows.

Least Action Principle. *If there exists $w : V \to \mathbb{N}$ such that*

$$s + \Delta w \leq \deg -1 \tag{3}$$

then s stabilizes, and $w \geq u$ where u is the odometer of s. Thus,

$$u(x) = \inf\{w(x) \,|\, w : V \to \mathbb{N} \text{ satisfies (3)}\}. \tag{4}$$

Proof. If such w exists, then any sequence $\mathbf{y}$ such that $w(x) = \#\{i : \mathbf{y}_i = x\}$ for all x is stabilizing for s. The odometer is defined as $u(x) = \#\{i : \mathbf{x}_i = x\}$ for a *legal* stabilizing sequence $\mathbf{x}$, so $w \geq u$ by part (i) of Lemma 1.1. The last line now follows from (1). □

The Least Action Principle expresses the odometer as the solution to a variational problem (4). In the next section we will see that the same problem, without the integrality constraint on w, arises from a variant of the sandpile which will be easier to analyze.

2. Relaxing Integrality: The Divisible Sandpile

Let $\mathbb{Z}^d$ be the set of points with integer coordinates in d-dimensional Euclidean space $\mathbb{R}^d$, and let $\mathbf{e}_1, \ldots, \mathbf{e}_d$ be its standard basis vectors. We view $\mathbb{Z}^d$ as a graph in which points x and y are adjacent if and only if $x - y = \pm\mathbf{e}_i$ for some i. For example, when $d = 1$ this graph is an infinite path, and when $d = 2$ it is an infinite square grid.

In the divisible sandpile model, each point $x \in \mathbb{Z}^d$ has a continuous amount of mass $\sigma(x) \in \mathbb{R}_{\geq 0}$ instead of a discrete number of particles. Start with mass m at the origin and zero elsewhere. At each time step, choose a site $x \in \mathbb{Z}^d$ with mass $\sigma(x) > 1$ where σ is the current configuration, and distribute the excess mass $\sigma(x) - 1$ equally among the $2d$ neighbors of x. We call this a *toppling*. Suppose that these choices are sufficiently **thorough** in the sense that whenever a site attains mass > 1, it is eventually chosen for toppling at some later time. Then we have the following version of the abelian property.

Lemma 2.1. *For any initial $\sigma_0 : \mathbb{Z}^d \to \mathbb{R}$ with finite total mass, and any thorough sequence of topplings, the mass function converges pointwise to a function $\sigma_\infty : \mathbb{Z}^d \to \mathbb{R}$ satisfying $0 \leq \sigma_\infty \leq 1$. Any site z satisfying $\sigma_0(z) < \sigma_\infty(z) < 1$ has a neighboring site y satisfying $\sigma_\infty(y) = 1$.*

Proof. Let $u_k(x)$ be the total amount of mass emitted from x during the first k topplings, and let $\sigma_k = \sigma_0 + \Delta u_k$ be the resulting mass configuration. Since u_k is increasing in k, we have $u_k \uparrow u_\infty$ for some $u_\infty : V \to [0, \infty]$. To rule out the value

∞, consider the *quadratic weight*

$$Q(\sigma_k) := \sum_{x\in\mathbb{Z}^d} (\sigma_k(x) - \sigma_0(x))|x|^2 = \sum_{x\in\mathbb{Z}^d} u_k(x).$$

To see the second equality, note that Q increases by h every time we topple mass h. The set $\{\sigma_k \geq 1\}$ is connected and contains $\mathbf{0}$, and has cardinality bounded by the total mass of σ_0, so it is bounded. Moreover, every site z with $\sigma_k(z) > \sigma_0(z)$ has a neighbor y with $\sigma_k(y) \geq 1$. Hence $\sup_k Q(\sigma_k) < \infty$, which shows that u_∞ is bounded.

Finally, $\sigma_\infty := \lim \sigma_k = \lim(\sigma_0 + \Delta u_k) = \sigma_0 + \Delta u_\infty$. By thoroughness, for each $x \in \mathbb{Z}^d$ we have $\sigma_k(x) \leq 1$ for infinitely many k, so $\sigma_\infty \leq 1$. □

The picture is thus of a set of "filled" sites ($\sigma_\infty(z) = 1$) bordered by a strip of partially filled sites ($\sigma_0(z) < \sigma_\infty(z) < 1$). Every partially filled site has a filled neighbor, so the thickness of this border strip is only one lattice spacing. Think of pouring maple syrup over a waffle: most squares receiving syrup fill up completely and then begin spilling over into neighboring squares. On the boundary of the region of filled squares is a strip of squares that fill up only partially (Figure 3).

The limit u_∞ is called the **odometer** of σ_0. The preceding proof did not show that u_∞ and σ_∞ are independent of the thorough toppling sequence. This is a consequence of the next result.

Least Action Principle For The Divisible Sandpile. *For any $\sigma_0 : \mathbb{Z}^2 \to [0,\infty)$ with finite total mass, and any $w : V \to [0,\infty)$ such that*

$$\sigma + \frac{1}{2d}\Delta w \leq 1 \tag{5}$$

we have $w \geq u_\infty$ for any thorough toppling sequence. Thus,

$$u_\infty(x) = \inf\{w(x) \,:\, w : V \to [0,\infty) \text{ satisfies (5)}\}. \tag{6}$$

Proof. With the notation of the preceding proof, suppose for a contradiction that $u_k \not\leq w$ for some k. For the miminal such k, the functions u_k and u_{k-1} agree except at x_k, hence

$$1 = \Big(\sigma + \frac{1}{2d}\Delta u_k\Big)(x_k) < \Big(\sigma + \frac{1}{2d}\Delta w\Big)(x_k) \leq 1\,,$$

which yields the required contradiction. □

2.1. **The superharmonic tablecloth.** The variational problem (6) has an equivalent formulation:

Lemma 2.2. *Let $\gamma : \mathbb{Z}^d \to \mathbb{R}$ satisfy $\frac{1}{2d}\Delta\gamma = \sigma_0 - 1$. Then the odometer u of (6) is given by*

$$u = s - \gamma$$

where

$$s(x) = \inf\{f(x) \,|\, f \geq \gamma \text{ and } \Delta f \leq 0\}. \tag{7}$$

Proof. f is in the set on the right side of (7) if and only if $w := f - \gamma$ is in the set on the right side of (6). □

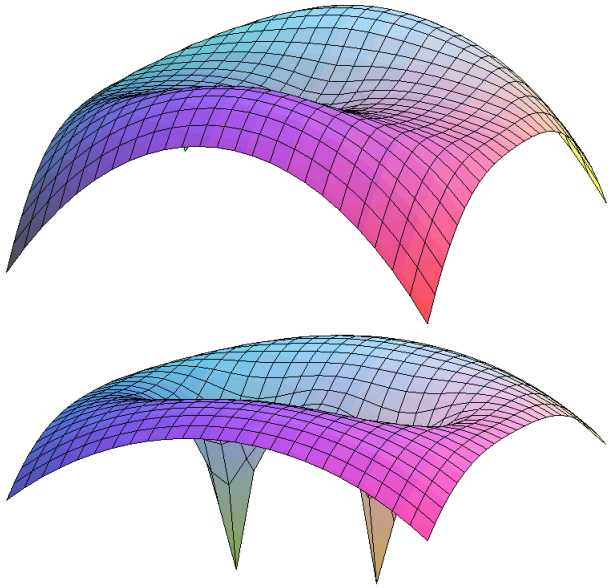


FIGURE 2. The obstacles γ corresponding to starting mass 1 on each of two overlapping disks (top) and mass 100 on each of two nonoverlapping disks.

The function γ is sometimes called the **obstacle**, and the minimizing function s in (7) called the **solution to the obstacle problem**. To explain this terminology, imagine the graph of γ as a fixed surface (for instance, the top of a table), and the graph of f as a surface that can vary (a tablecloth). The tablecloth is constrained to stay above the table ($f \geq \gamma$) and is further constrained to be **superharmonic** ($\Delta f \leq 0$), which in particular implies that f has no local minima. Depending on the shape of the table γ, these constraints may force the tablecloth to lie strictly above the table in some places.

The solution s is the lowest possible position of the tablecloth. The set where strict inequality holds

$$D := \{x \in \mathbb{Z}^d \,:\, s(x) > \gamma(x)\}.$$

is called the **noncoincidence set**. In terms of the divisible sandpile, the odometer function u is the gap $s - \gamma$ between tablecloth and table, and the set $\{u > 0\}$ of sites that topple is the noncoincidence set.

2.2. **Building the obstacle.** The reader ought now be wondering, given a configuration $\sigma_0 : \mathbb{Z}^d \to [0, \infty)$ of finite total mass, what the corresponding obstacle $\gamma : \mathbb{Z}^d \to \mathbb{R}$ looks like. The only requirement on γ is that it has a specified discrete Laplacian, namely

$$\frac{1}{2d}\Delta\gamma = \sigma_0 - 1.$$

Does such γ always exist?

Given a function $f : \mathbb{Z}^d \to \mathbb{R}$ we would like to construct a function F such that $\Delta F = f$. The most straightforward method is to assign arbitrary values for F on a pair of parallel hyperplanes, from which the relation $\Delta F = f$ determines the other values of F uniquely.

This method suffers from the drawback that the growth rate of F is hard to control. A better method uses what is called the **Green function** or **fundamental solution** for the discrete Laplacian Δ. This is a certain function $g : \mathbb{Z}^d \to \mathbb{R}$ whose discrete Laplacian is zero except at the origin.

$$\frac{1}{2d}\Delta g(x) = -\delta_{\mathbf{0}}(x) = \begin{cases} -1 & x = \mathbf{0} \\ 0 & x \neq \mathbf{0}. \end{cases} \tag{8}$$

If f has finite support, then we can construct F as a convolution

$$F(x) = -f * g := -\sum_{y \in \mathbb{Z}^d} f(y)g(x-y)$$

in which only finitely many terms are nonzero. (The condition that f has finite support can be relaxed to fast decay of $f(x)$ as $|x| \to \infty$, but we will not pursue this.) Then for all $x \in \mathbb{Z}^d$ we have

$$\Delta F(x) = \sum_{y \in \mathbb{Z}^d} f(y)\delta_{\mathbf{0}}(x-y) = f(x)$$

as desired. By controlling the growth rate of the Green function g, we can control the growth rate of F. The minus sign in equation (8) is a convention: as we will now see, with this sign convention g has a natural definition in terms of random walk.

Let $\xi_1, \xi_2, \ldots$ be a sequence of independent random variables each with the uniform distribution on the set $\mathcal{E} = \{\pm\mathbf{e}_1, \ldots, \pm\mathbf{e}_d\}$. For $x \in \mathbb{Z}^d$, the sequence

$$X_n = \xi_1 + \ldots + \xi_n, \qquad n \geq 0$$

is called **simple random walk** started from the origin in $\mathbb{Z}^d$: it is the location of a walker who has wandered from $\mathbf{0}$ by taking n independent random steps, choosing each of the $2d$ coordinate directions $\pm\mathbf{e}_i$ with equal probability $1/2d$ at each step.

In dimensions $d \geq 3$ the simple random walk is **transient**: its expected number of returns to the origin is finite. In these dimensions we define

$$g(x) := \sum_{n \geq 0} \mathbb{P}(X_n = x),$$

a function known as the **Green function** of $\mathbb{Z}^d$. It is the expected number of visits to x by a simple random walk started at the origin in $\mathbb{Z}^d$. The identity

$$-\frac{1}{2d}\Delta g = \delta_{\mathbf{0}} \tag{9}$$

is proved by conditioning on the first step X_1 of the walk:

$$g(x) = P(X_0 = x) + \sum_{n\geq 1}\sum_{e\in\mathcal{E}} P(X_n = x | X_1 = e)P(X_1 = e).$$
$$= \delta_{\mathbf{0}}(x) + \sum_{n\geq 1}\sum_{e\in\mathcal{E}} P(X_{n-1} = x - e)\frac{1}{2d}$$

Interchanging the order of summation, the second term on the right equals $\frac{1}{2d}\sum_{y\sim x} g(y)$, and (9) now follows by the definition of the Laplacian Δ.

The case $d = 2$ is more delicate because the simple random walk is **recurrent**: with probability 1 it visits x infinitely often, so the sum defining $g(x)$ diverges. In this case, g is defined instead as

$$g(x) = \sum_{n\geq 0} \left(\mathbb{P}(X_n = x) - \mathbb{P}(X_n = \mathbf{0})\right).$$

One can show that this sum converges and that the resulting function $g : \mathbb{Z}^2 \to \mathbb{R}$ satisfies (9); see [72]. The function $-g$ is called the **recurrent potential kernel** of $\mathbb{Z}^2$.

Convolving with the Green function enables us to construct functions on $\mathbb{Z}^d$ whose discrete Laplacian is any given function with finite support. But we want more: In Lemma 2.2 we seek a function γ satisfying $\Delta\gamma = \sigma - 1$, where σ has finite support. Fortunately, there is a very nice function whose discrete Laplacian is a constant function, namely the squared Euclidean norm

$$q(x) = |x|^2 := \sum_{i=1}^{d} x_i^2.$$

(In fact, we implicitly used the identity $\frac{1}{2d}\Delta q \equiv 1$ in the quadratic weight argument for Lemma 2.1.) We can therefore take as our obstacle the function

$$\gamma = -q - (g * \sigma). \tag{10}$$

In order to determine what happens when we drape a superharmonic tablecloth over this particular table γ, we should figure out what γ looks like! In particular, we would like to know the asymptotic order of the Green function $g(x)$ when x is far from the origin. It turns out [22, 38, 73] that

$$g(x) = (1 + O(|x|^{-2}))G(x)$$

where G is the spherically symmetric function

$$G(x) := \begin{cases} -\frac{2}{\pi}\log|x|, & d = 2; \\ a_d|x|^{2-d}, & d \geq 3. \end{cases} \tag{11}$$

(The constant $a_d = \frac{2}{(d-2)\omega_d}$ where ω_d is the volume of the unit ball in $\mathbb{R}^d$.) As we will now see, this estimate in combination with $-\frac{1}{2d}\Delta g = \delta_{\mathbf{0}}$ is a powerful package. We start by analyzing the initial condition $\sigma = m\delta_{\mathbf{0}}$ for large m.

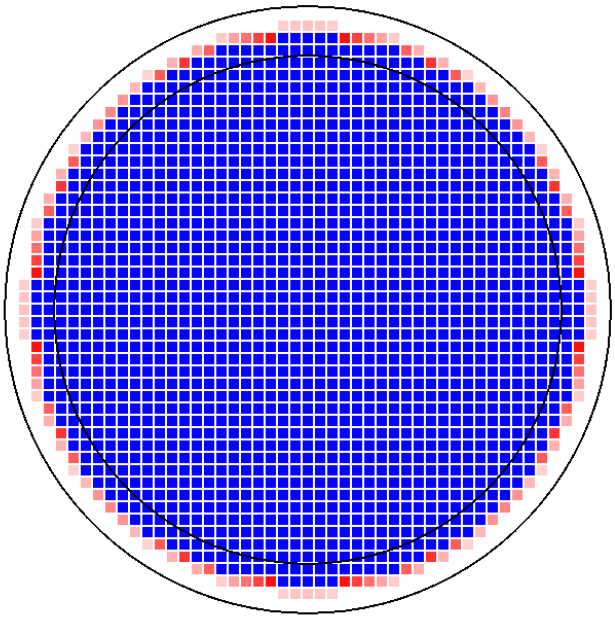


FIGURE 3. Divisible sandpile in $\mathbb{Z}^2$ started from mass $m = 1600$ at the origin. Each square is colored blue if it fills completely, red if it fills only partially. The black circles are centered at the origin, of radius $r \pm 2$ where $\pi r^2 = m$.

2.3. **Point sources.** Pour m grams of maple syrup into the center square of a very large waffle. Supposing each square can hold just 1 gram of syrup before it overflows, distributing the excess equally among the four neighboring squares, What is the shape of the resulting set of squares that fill up with syrup?

Figure 3 suggests the answer is "very close to a disk". Being mathematicians, we wish to quantify "very close", and why stop at two-dimensional waffles? Let $B(\mathbf{0}, r)$ be the Euclidean ball of radius r centered at the origin in $\mathbb{R}^d$.

Theorem 2.3. [46] *Let $D_m = \{\sigma_\infty = 1\}$ be the set of fully occupied sites for the divisible sandpile started from mass m at the origin in $\mathbb{Z}^d$. There is a constant $c = c(d)$, such that*

$$B(\mathbf{0}, r - c) \cap \mathbb{Z}^d \subset D_m \subset B(\mathbf{0}, r + c)$$

where r is such that $B(\mathbf{0}, r)$ has volume m. Moreover, the odometer u_∞ satisfies

$$u_\infty(x) = mg(x) + |x|^2 - mg(r\mathbf{e}_1) - r^2 + O(1) \tag{12}$$

for all $x \in B(\mathbf{0}, r + c) \cap \mathbb{Z}^d$, where the constant in the O depends only on d.

The idea of the proof is to use Lemma 2.2 to write the odometer function as

$$u_\infty = s - \gamma$$

for an obstacle γ with discrete Laplacian $\frac{1}{2d}\Delta\gamma = m\delta_{\mathbf{0}} - 1$. What does such an obstacle look like?

Recalling that the Euclidean norm $|x|^2$ and the discrete Green function g have discrete Laplacians 1 and $-\delta_{\mathbf{0}}$, respectively, a natural choice of obstacle is

$$\gamma(x) = -|x|^2 - mg(x). \tag{13}$$

The claim of (12) is that $u(x)$ is within an additive constant of $\gamma(r\mathbf{e}_1) - \gamma(x)$. To prove this one uses two properties of γ: it is nearly spherically symmetric (because g is!) and it is maximized near $|x| = r$. From these properties one deduces that s is nearly a constant function, and that $\{s > \gamma\}$ is nearly the ball $B(\mathbf{0}, r) \cap \mathbb{Z}^d$.

The Euclidean ball as a limit shape is an example of **universality**: Although our topplings took place on the cubic lattice $\mathbb{Z}^d$, if we take the total mass $m \to \infty$ while zooming out so that the cubes of the lattice become infinitely small, the divisible sandpile assumes a perfectly spherical limit shape. Figure 1 strongly suggests that the abelian sandpile, with its indivisible grains of sand, does *not* enjoy such universality. However, discrete particles are not incompatible with universality, as the next two examples show.

3. Internal DLA

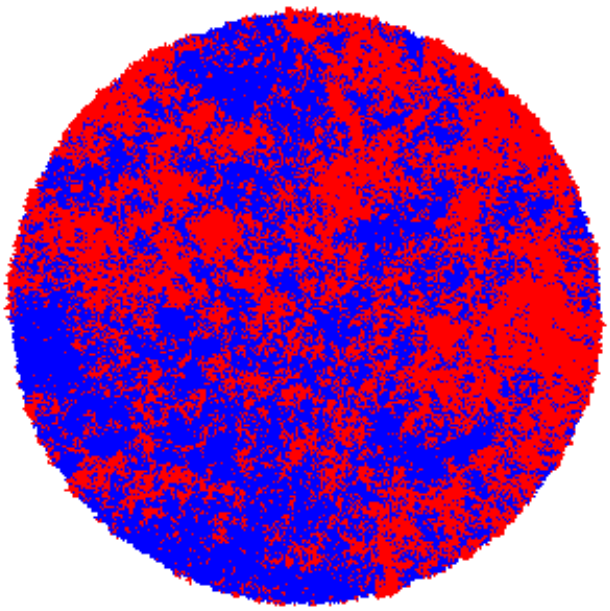


FIGURE 4. An internal DLA cluster in $\mathbb{Z}^2$. The colors indicate whether a point was added to the cluster earlier or later than expected: the random site $x(j)$ where the j-th particle stops is colored red if $\pi|x(j)|^2 > j$, blue otherwise.

Let $m \geq 1$ be an integer. Starting with m particles at the origin in the d-dimensional integer lattice $\mathbb{Z}^d$, let each particle in turn perform a simple random walk until reaching an unoccupied site; that is, the particle repeatedly jumps to

an nearest neighbor chosen independently and uniformly at random, until it lands on a site containing no other particles.

This procedure, known as *internal DLA*, was proposed by Meakin and Deutch [54] and independently by Diaconis and Fulton [18]. It produces a random set A_m of m occupied sites in $\mathbb{Z}^d$. This random set is close to a ball, in the following sense. Let r be such that the Euclidean ball $B(\mathbf{0}, r)$ of radius r has volume m. Lawler, Bramson and Griffeath [42] proved that for any $\epsilon > 0$, with probability 1 it holds that

$$B(\mathbf{0}, (1-\epsilon)r) \cap \mathbb{Z}^d \subset A_m \subset B(\mathbf{0}, (1+\epsilon)r) \qquad \text{for all sufficiently large } m.$$

A sequence of improvements followed, showing that the fluctuations of A_m around $B(\mathbf{0}, r)$ are logarithmic in r [40, 2, 3, 4, 31, 32, 33].

4. Rotor-routing: derandomized random walk

In a **rotor-router walk** on a graph, the successive exits from each vertex follow a prescribed periodic sequence. Walks of this type were studied in [75] as a model of mobile agents exploring a territory, and in [65] as a model of self-organized criticality. Propp [67] proposed rotor walk as a derandomization of random walk, a perspective explored in [14, 28].

In the case of the square grid $\mathbb{Z}^2$, each site has a *rotor* pointing North, East, South or West. A particle starts at the origin; during each time step, the rotor at the particle's current location rotates 90 degrees clockwise, and the particle takes a step in the direction of the newly rotated rotor.

In **rotor aggregation**, we start with n particles at the origin; each particle in turn performs rotor-router walk until it reaches a site not occupied by any other particles. Importantly, we do not reset the rotors between walks! Let R_n denote the resulting region of n occupied sites in $\mathbb{Z}^2$. For example, if all rotors initially point north, the sequence will begin $R_1 = \{\mathbf{0}\}$, $R_2 = \{\mathbf{0}, \mathbf{e}_1\}$, $R_3 = \{\mathbf{0}, \mathbf{e}_1, -\mathbf{e}_2\}$. The region R_{10^6} is pictured in Figure 5. The limiting shape is again a Euclidean ball [46].

5. Multiple sources; Quadrature domains

The Euclidean ball as a limiting shape is not too hard to guess. But what if the particles start at two different points of $\mathbb{Z}^d$? For example, fix an integer $r \geq 1$ and a positive real number a, and start with $m = \lfloor \omega_d (ar)^d \rfloor$ particles at each of $r\mathbf{e}_1$ and $-r\mathbf{e}_1$. Alternately release a particle from $r\mathbf{e}_1$ and let it perform simple random walk until it finds an unoccupied site, and then release a particle from $-r\mathbf{e}_1$ and let it perform simple random walk until it finds an unoccupied site. The result is a random set $A_{m,m}$ consisting of $2m$ occupied sites in $\mathbb{Z}^d$.

If $a < 1$, then the distance between the source points $\pm r\mathbf{e}_1$ is so large compared to the number of particles that with high probability, the particles starting at $r\mathbf{e}_1$ do not interact with those starting at $-r\mathbf{e}_1$. In this case $A_{m,m}$ is a disjoint union of two ball-shaped clusters each of size m. On the other hand, if $a \gg 1$, so that the two source points are very close together relative to the number of particles released, then the cluster $A_{m,m}$ will look like a single ball of size $2m$. In between

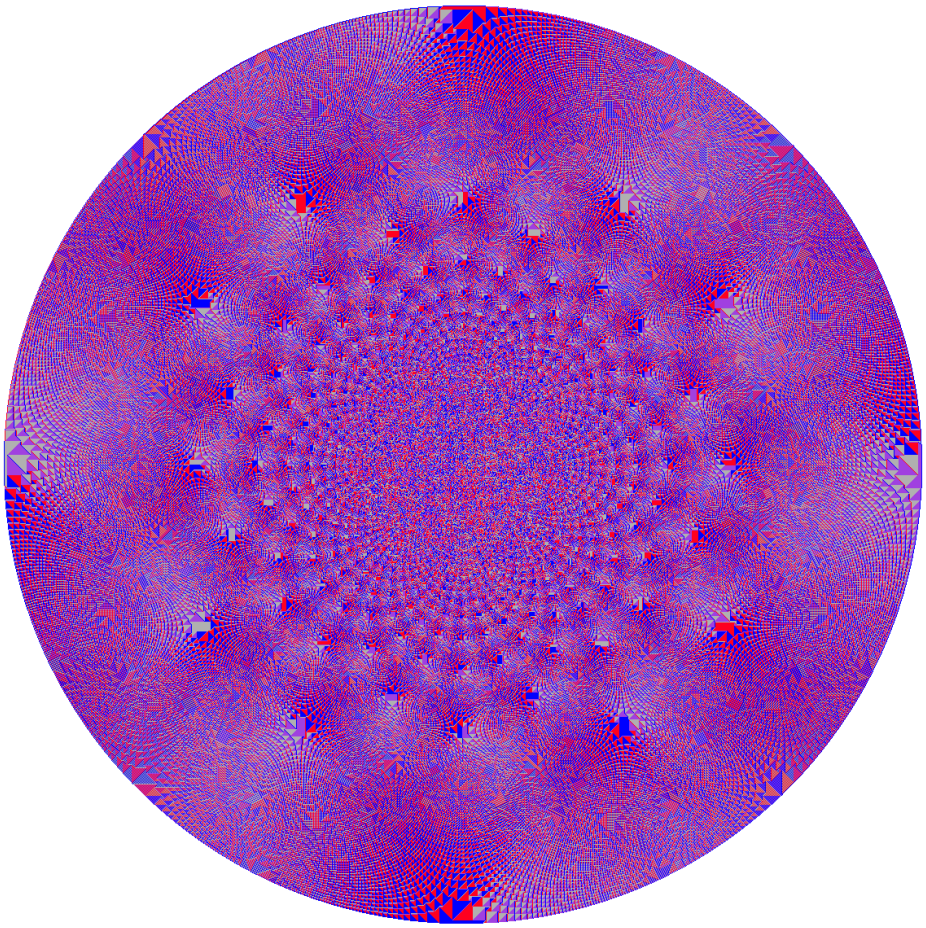


FIGURE 5. Rotor-router aggregate of one million particles started at the origin in $\mathbb{Z}^2$, with all rotors initially pointing North. Each site is colored according to the final direction of its rotor (North, East, South or West).

these extreme cases there is a more interesting behavior, described by the following theorem.

Theorem 5.1. [47] *There exists a deterministic domain $D \subset \mathbb{R}^d$ such that with probability* 1

$$\frac{1}{r} A_{m,m} \to D \tag{14}$$

as $r \to \infty$.

The precise meaning of the convergence of domains in (14) is the following: given $D_r \subset \frac{1}{r}\mathbb{Z}^d$ and $\Omega \subset \mathbb{R}^d$, we write $D_r \to \Omega$ if for all $\epsilon > 0$ we have

$$\Omega_\epsilon \cap \frac{1}{r}\mathbb{Z}^d \subset D_r \subset \Omega^\epsilon \tag{15}$$

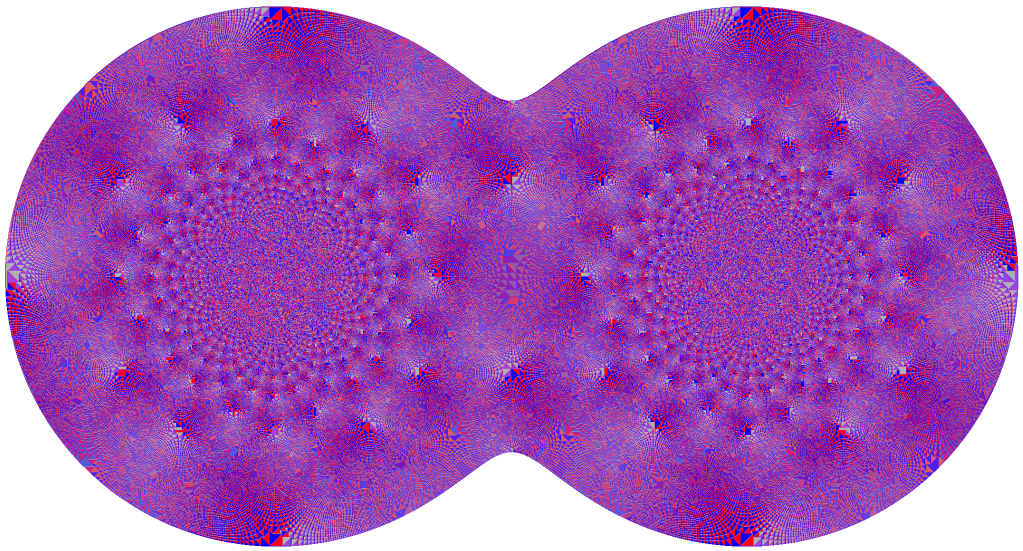


FIGURE 6. Rotor-router aggregation started from two point sources in $\mathbb{Z}^2$. Its scaling limit is a two-point quadrature domain in $\mathbb{R}^2$, satisfying (16).

for all sufficiently large r, where

$$\Omega_\epsilon = \{x \in \Omega \mid \overline{\mathrm{B}}(x, \epsilon) \subset \Omega\}$$

and

$$\Omega^\epsilon = \{x \in \mathbb{R}^d \mid \mathrm{B}(x, \epsilon) \not\subset \Omega^c\}$$

are the inner and outer ϵ-neighborhoods of D.

The limiting domain D is called a **quadrature domain** because it satisfies

$$\int_D h \, \mathrm{d}x = h(-\mathbf{e}_1) + h(\mathbf{e}_1) \tag{16}$$

for all integrable harmonic functions h on D, whre $\mathrm{d}x$ is Lebesgue measure on $\mathbb{R}^d$. This identity is analogous to the mean value property $\int_B h \, \mathrm{d}x = h(\mathbf{0})$ for integrable harmonic functions on the ball B of unit volume centered at the origin.

In dimension $d = 2$, the domain D has a much more explicit description: Its boundary in $\mathbb{R}^2$ is the quartic curve

$$\left(x^2 + y^2\right)^2 - 2a^2\left(x^2 + y^2\right) - 2(x^2 - y^2) = 0. \tag{17}$$

When $a = 1$, the curve (17) becomes

$$(x^2 + y^2 - 2x)(x^2 + y^2 + 2x) = 0$$

which describes the union of two unit circles centered at $\pm\mathbf{e}_1$ and tangent at the origin. This case corresponds to two clusters that just barely interact, whose interaction is small enough that we do not see it in the limit. When $a \gg 1$, the

term $2(x^2 - y^2)$ is much smaller than the others, so the curve (17) approaches the circle

$$x^2 + y^2 - 2a^2 = 0.$$

This case corresponds to releasing so many particles that the effect of releasing them alternately at $\pm re_1$ is nearly the same as releasing them all at the origin.

Theorem 5.1 extends to the case of any k point sources in $\mathbb{R}^d$ as follows.

Theorem 5.2. [47] *Fix $x_1, \dots, x_k \in \mathbb{R}^d$ and $\lambda_1, \dots, \lambda_k > 0$. Let $x_i^{\vdots}$ be a closest site to x_i in the lattice $\frac{1}{n}\mathbb{Z}^d$, and let*

$$D_n = \{\textit{occupied sites for the divisible sandpile}\}$$
$$R_n = \{\textit{occupied sites for rotor aggregation}\}$$
$$I_n = \{\textit{occupied sites for internal DLA}\}$$

started in each case from $\lfloor \lambda_i n^d \rfloor$ particles at each site $x_i^{\vdots}$ in $\frac{1}{n}\mathbb{Z}^d$.

Then there is a deterministic set $D \subset \mathbb{R}^d$ such that

$$D_n, R_n, I_n \to D$$

where the convergence is in the sense of (15); the convergence for R_n holds for any initial setting of the rotors; and the convergence for I_n is with probability 1.

The limiting set D is called a k-point quadrature domain. It is characterized up to measure zero by the inequalities

$$\int_D h \, \mathrm{d}x \leq \sum_{i=1}^{k} \lambda_i h(x_i)$$

for all integrable superharmonic functions h on D, where $\mathrm{d}x$ is Lebesgue measure on $\mathbb{R}^d$. The subject of quadrature domains in the plane begins with Aharonov and Shapiro [1] and was developed by Gustafsson [24], Sakai [69, 70] and others. The boundary of a quadrature domain for k point sources in the plane lies on an algebraic curve of degree $2k$. In dimensions $d \geq 3$, it is not known whether the boundary of D is an algebraic surface!

6. Scaling limit of the abelian sandpile on $\mathbb{Z}^2$

Now that we have seen an example of a universal scaling limit, let us return to our very first example, the abelian sandpile with discrete particles.

Take as our underlying graph the square grid $\mathbb{Z}^2$, start with n particles at the origin and stabilize. The resulting configuration of sand appears to be *non-circular* (Figure 1)—so we do not the scaling limit to be universal like the one in Theorem 5.2. In a breakthrough work [58], Pegden and Smart proved existence of its scaling limit as $n \to \infty$. To state their result, let

$$s_n = n\delta_{\mathbf{0}} + \Delta u_n$$

be the sandpile formed from n particles at the origin in $\mathbb{Z}^d$, and consider the rescaled sandpile

$$\bar{s}_n(x) = s_n(n^{1/d}x).$$

Theorem 6.1. [58] *There is a function $s : \mathbb{R}^d \to \mathbb{R}$ such that $\bar{s}_n \to s$ weakly-$*$ in $L^\infty(\mathbb{R}^d)$.*

The weak-$*$ convergence of $\bar{s}_n$ in L^∞ means that for every ball $B(x,r)$, the average of s_n over $\mathbb{Z}^d \cap n^{1/d}B(x,r)$ tends as $n \to \infty$ to the average of s over $B(x,r)$.

The limiting sandpile s is lattice dependent. Examining the proof in [58] reveals that the lattice dependence enters in the following way. Each real symmetric $d \times d$ matrix A defines a quadratic function $q_A(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T A\mathbf{x}$ and an associated sandpile $s_A : \mathbb{Z}^d \to \mathbb{Z}$

$$s_A = \Delta \lceil q_A \rceil \ .$$

For each matrix A, the sandpile s_A either **stabilizes locally** (that is, every site of $\mathbb{Z}^d$ topples finitely often) or fails to stabilize (in which case every site topples infinitely often). The set of **allowed Hessians** $\Gamma(\mathbb{Z}^d)$ is defined as the closure (with respect to the Euclidean norm $\|A\|_2^2 = \mathrm{Tr}(A^T A)$) of the set of matrices A such that s_A stabilizes locally.

One can convert the Least Action Principle into an obstacle problem analogous to Lemma 2.2 with an additional integrality constraint. The limit of these discrete obstacle problems on $\frac{1}{n}\mathbb{Z}^d$ as $n \to \infty$ is the following variational problem on $\mathbb{R}^d$.

Limit of the least action principle.

$$u = \inf\left\{w \in C(\mathbb{R}^d) \mid w \geq -G \text{ and } D^2(w+G) \in \Gamma(\mathbb{Z}^d)\right\}. \tag{18}$$

Here G is the fundamental solution of the Laplacian in $\mathbb{R}^d$. The infimum is pointwise, and the minimizer u is related to the the sandpile odometers u_n by

$$\lim_{n\to\infty} \frac{1}{n} u_n(n^{1/2}x) = u(x) + G(x).$$

The Hessian constraint in (18) is interpreted **in the sense of viscosity**:

$$D^2\varphi(x) \in \Gamma(\mathbb{Z}^d)$$

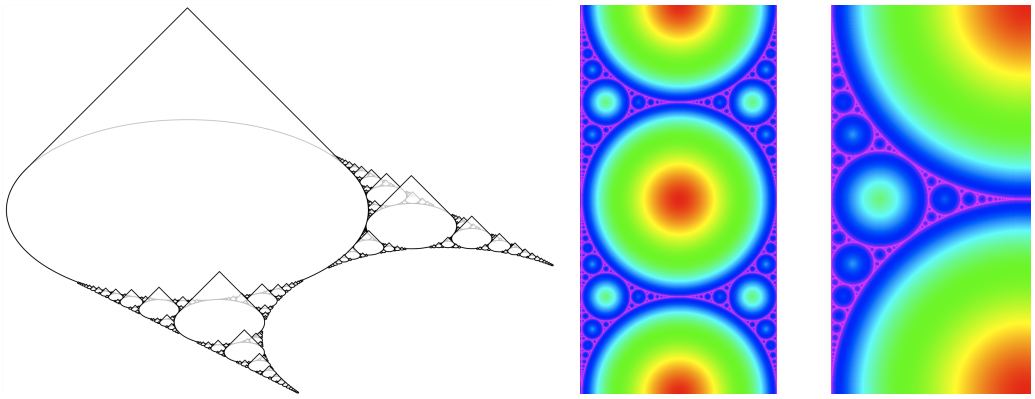
whenever φ is a C^∞ function touching $w + G$ from below at x (that is, $\varphi(x) = w(x) + G(x)$ and $\varphi - (w + G)$ has a local maximum at x).

The obstacle G in (18) is a spherically symmetric function on $\mathbb{R}^d$, so the lattice-dependence arises solely from $\Gamma(\mathbb{Z}^d)$. Put another way, the set $\Gamma(\mathbb{Z}^d)$ is a way of quantifying *which features of the lattice $\mathbb{Z}^d$ are still detectable in the limit* of sandpiles as the lattice spacing shrinks to zero.

An explicit description of $\Gamma(\mathbb{Z}^2)$ appears in [45] (see Figure 7), and explicit fractal solutions of the **sandpile PDE**

$$D^2 u \in \partial\Gamma(\mathbb{Z}^2)$$

are constructed in [44]. See [59] for images of $\Gamma(L)$ for some other two-dimensional lattices L.



(a) (b) (c)

FIGURE 7. (a) According to the main theorem of [45], the set of allowed Hessians $\Gamma(\mathbb{Z}^2)$ is the union of slope 1 cones based at the circles of an **Apollonian circle packing** in the plane of 2×2 real symmetric matrices of trace 2. (b) The same set viewed from above: Color of point (a, b) indicates the largest c such that $\left[\begin{smallmatrix} c-a & b \\ b & c+a \end{smallmatrix}\right] \in \Gamma(\mathbb{Z}^2)$. The rectangle shown, $0 \le a \le 2$, $0 \le b \le 4$ extends periodically to the entire plane. (c) Close-up of the lower left corner $0 \le a \le 1$, $0 \le b \le 2$.

7. The sandpile group of a finite graph

Let $G = (V, E)$ be a finite connected graph and fix a **sink** vertex $z \in V$. A **stable** sandpile is now a map $s : V \setminus \{z\} \to \mathbb{N}$ satisfying $s(x) < \deg(x)$ for all $x \in V \setminus \{z\}$. As before, sites x with $s(x) \ge \deg(x)$ topple by sending one particle along each edge incident to x, but now particles falling into the sink disappear.

We define a Markov chain on the set of stable sandpiles as follows: at each time step, add one sand grain at a vertex of $V \setminus \{z\}$ selected uniformly at random, and then perform all possible topplings until the sandpile is stable. Recall that a state s in a finite Markov chain is called **recurrent** if whenever s' is reachable from s then also s is reachable from s'. Dhar [15] observed that the operation a_x of adding one particle at vertex x and then stabilizing is a permutation of the set $\mathrm{Rec}(G, z)$ of recurrent sandpiles. These permutations obey the relations

$$a_x a_y = a_y a_x \quad \text{and} \quad a_x^{\deg(x)} = \prod_{u \sim x} a_u$$

for all $x, y \in V \setminus \{z\}$. The subgroup $K(G, z)$ of the permutation group $\mathrm{Sym}(\mathrm{Rec}(G, z))$ generated by $\{a_x\}_{x \ne z}$ is called the **sandpile group** of G. Although the set $\mathrm{Rec}(G, z)$ depends on the choice of sink vertex, the sandpile groups for different choices of sink are isomorphic (see, e.g., [27, 29]).

The sandpile group $K(G, z)$ has a free transitive action on $\mathrm{Rec}(G, z)$, so $\#K(G, z) = \#\mathrm{Rec}(G, z)$. One can use rotor-routing to define a free transitive action of $K(G, z)$ on the set of spanning trees of G [27]. In particular, the number of spanning trees

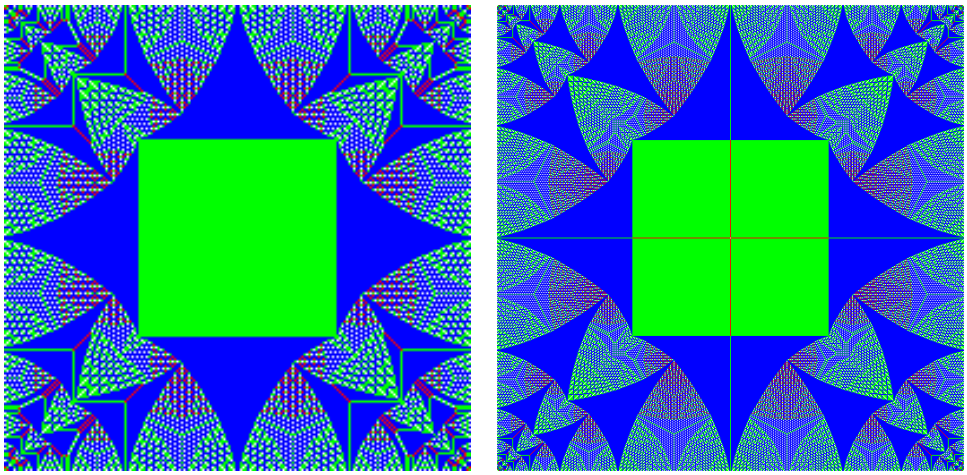


FIGURE 8. Identity elements of the sandpile group $\mathrm{Rec}([0,n]^2, z)$ of the $n \times n$ grid graph with sink at the wired boundary (i.e., all boundary vertices are identified to a single vertex z), for $n = 198$ (left) and $n = 521$.

also equals $\#K(G,z)$. The most important bijection between recurrent sandpiles and spanning trees uses Dhar's burning algorithm [15, 51].

A group operation $\oplus$ can also be defined directly on $\mathrm{Rec}(G,z)$, namely $s \oplus s'$ is the stabilization of $s + s'$. Then $s \mapsto \prod_x a_x^{s(x)}$ defines an isomorphism from $(\mathrm{Rec}(G,z), \oplus)$ to the sandpile group.

8. Loop erasures, Tutte polynomial, Unicycles

Fix an integer $d \geq 2$. The **looping constant** $\xi = \xi(\mathbb{Z}^d)$ is defined as the expected number of neighbors of the origin on the infinite loop-erased random walk in $\mathbb{Z}^d$. In dimensions $d \geq 3$, this walk can be defined by erasing cycles from the simple random walk in chronological order. In dimension 2, one first defines the loop erasure of the simple random walk stopped on exiting the box $[-n,n]^2$ and shows that the resulting measures converge weakly [39, 41].

A **unicycle** is a connected graph with the same number of edges as vertices. Such a graph has exactly one cycle (Figure 9). If G is a finite (multi)graph, a *spanning subgraph* of G is a graph containing all of the vertices of G and a subset of the edges. A **uniform spanning unicycle** (USU) of G is a spanning subgraph of G which is a unicycle, selected uniformly at random.

An **exhaustion** of $\mathbb{Z}^d$ is a sequence $V_1 \subset V_2 \subset \cdots$ of finite subsets such that $\bigcup_{n \geq 1} V_n = \mathbb{Z}^d$. Let G_n be the multigraph obtained from $\mathbb{Z}^d$ by collapsing V_n^c to a single vertex z_n, and removing self-loops at z_n. We do not collapse edges, so G_n may have edges of multiplicity greater than one incident to z_n. Theorem 8.1, below, gives a numerical relationship between the looping constant ξ and the **mean**

unicycle length

$$\lambda_n = \mathbb{E}\left[\text{length of the unique cycle in a USU of } G_n\right].$$

as well as the **mean sandpile height**

$$\zeta_n = \mathbb{E}\left[\text{number of particles at } \mathbf{0} \text{ in a uniformly random recurrent sandpile on } V_n\right].$$

To define the last quantity of interest, recall that the **Tutte polynomial** of a finite (multi)graph $G = (V, E)$ is the two-variable polynomial

$$T(x, y) = \sum_{A \subset E} (x-1)^{c(A)-1}(y-1)^{c(A)+\#A-n}$$

where $c(A)$ is the number of connected components of the spanning subgraph (V, A). Let $T_n(x, y)$ be the Tutte polynomial of G_n. The **Tutte slope** is the ratio

$$\tau_n = \frac{\frac{\partial T_n}{\partial y}(1,1)}{(\#V_n)T_n(1,1)}.$$

A combinatorial interpretation of τ_n is the number of spanning unicycles of G_n divided by the number of rooted spanning trees of G_n.

For a finite set $V \subset \mathbb{Z}^d$, write ∂V for the set of sites in V^c adjacent to V.

Theorem 8.1. [48] *Let $\{V_n\}_{n\geq 1}$ be an exhaustion of $\mathbb{Z}^d$ such that $V_1 = \{\mathbf{0}\}$, $\#V_n = n$, and $\#(\partial V_n)/n \to 0$. Let $\tau_n, \zeta_n, \lambda_n$ be the Tutte slope, sandpile mean height and mean unicycle length in V_n. Then the following limits exist:*

$$\tau = \lim_{n\to\infty} \tau_n, \qquad \zeta = \lim_{n\to\infty} \zeta_n, \qquad \lambda = \lim_{n\to\infty} \lambda_n.$$

Their values are given in terms of the looping constant $\xi = \xi(\mathbb{Z}^d)$ by

$$\tau = \frac{\xi - 1}{2}, \qquad \zeta = d + \frac{\xi - 1}{2}, \qquad \lambda = \frac{2d-2}{\xi - 1}. \tag{19}$$

The two-dimensional case is of particular interest, because the quantities $\xi, \tau, \zeta, \lambda$ rather intriguingly come out to be rational numbers.

Corollary 8.2. *In the case $d = 2$, we have* [37, 66, 13]

$$\xi = \frac{5}{4} \quad \textit{and} \quad \zeta = \frac{17}{8}.$$

Hence by Theorem 8.1,

$$\tau = \frac{1}{8} \quad \textit{and} \quad \lambda = 8.$$

The value $\zeta(\mathbb{Z}^2) = \frac{17}{8}$ was conjectured by Grassberger (see [16]). Poghosyan and Priezzhev [62] observed the equivalence of this conjecture with $\xi(\mathbb{Z}^2) = \frac{5}{4}$, and shortly thereafter three proofs [66, 37, 13] appeared.

The proof that $\zeta(\mathbb{Z}^2) = \frac{17}{8}$ by Kenyon and Wilson [37] uses the theory of vector bundle Laplacians [36], while the proof by Poghosyan, Priezzhev and Ruelle [66] uses monomer-dimer calculations. Earlier, Jeng, Piroux and Ruelle [30] had reduced the computation of $\zeta(\mathbb{Z}^2)$ to evaluation of a certain multiple integral which

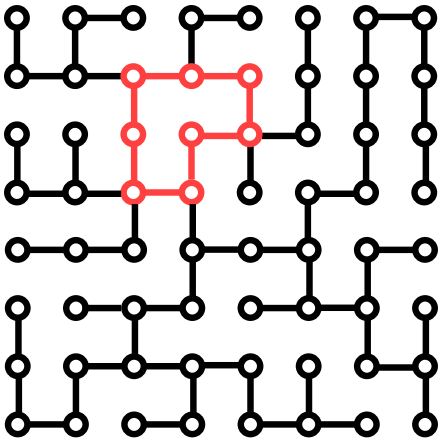


FIGURE 9. A spanning unicycle of the 8×8 square grid. The unique cycle is shown in red.

they evaluated numerically as 0.5 ± 10^{-12}. This integral was proved to equal $\frac{1}{2}$ by Caracciolo and Sportiello [13], thus providing another proof.

All three proofs involve powers of $1/\pi$ which ultimately cancel out. For $i = 0, 1, 2, 3$ let p_i be the probability that a uniform recurrent sandpile in $\mathbb{Z}^2$ has exactly i grains of sand at the origin. The proof of the distribution

$$\begin{aligned} p_0 &= \frac{2}{\pi^2} - \frac{4}{\pi^3} \\ p_1 &= \frac{1}{4} - \frac{1}{2\pi} - \frac{3}{\pi^2} + \frac{12}{\pi^3} \\ p_2 &= \frac{3}{8} + \frac{1}{\pi} - \frac{12}{\pi^3} \\ p_3 &= \frac{3}{8} - \frac{1}{2\pi} + \frac{1}{\pi^2} + \frac{4}{\pi^3} \end{aligned}$$

is completed in [66, 37], following work of [51, 64, 30]. In particular, $\zeta(\mathbb{Z}^2) = p_1 + 2p_2 + 3p_3 = \frac{17}{8}$.

Kassel and Wilson [35] give a new and simpler method for computing $\zeta(\mathbb{Z}^2)$, relying on planar duality, which also extends to other lattices. For a survey of their approach, see [34].

The objects we study on finite subgraphs of $\mathbb{Z}^d$ also have "infinite-volume limits" defined on $\mathbb{Z}^d$ itself: Lawler [39] defined the infinite loop-erased random walk, Pemantle [57] defined the uniform spanning tree in $\mathbb{Z}^d$, and Athreya and Járai [5] defined the infinite-volume stationary measure for sandpiles in $\mathbb{Z}^d$. The latter limit uses the burning bijection of Majumdar and Dhar [51] and the one-ended property of the trees in the uniform spanning forest [57, 9]. As for the Tutte polynomial, the limit

$$t(x, y) := \lim_{n \to \infty} \frac{1}{n} \log T_n(x, y)$$

can be expressed in terms of the pressure of the Fortuin-Kasteleyn random cluster model. By a theorem of Grimmett (see [23, Theorem 4.58]) this limit exists for all real $x, y > 1$. Theorem 8.1 concerns the behavior of this limit as $(x, y) \to (1, 1)$; indeed, another expression for the Tutte slope is

$$\tau_n = \frac{\partial}{\partial y} \left[\frac{1}{n} \log T_n(x, y) \right] \Big|_{x=y=1}.$$

9. Open problems

We conclude by highlighting a few of the key open problems in this area.

(1) Suppose $s(x)_{x \in \mathbb{Z}^2}$ are independent and identically distributed random variables taking values in $\{0, 1, 2, 3, 4\}$. Viewing s as a sandpile, the event that every site of $\mathbb{Z}^2$ topples infinitely often is invariant under translation, so it has probability 0 or 1. We do not know of an algorithm to decide whether this probability is 0 or 1! See [19].

(2) Does the rotor-router walk in $\mathbb{Z}^2$ with random initial rotors (independently North, East, South, or West, each with probability $\frac{1}{4}$) return to the origin with probability 1? The number of sites visited by such a walk in n steps is predicted to be of order $n^{2/3}$ [63]. For a lower bound of that order, see [21]. As noted there, even an upper bound of $o(n)$ would imply recurrence, which is not known!

(3) Recall that the weak-$*$ convergence in Theorem 6.1 means that the average height of the sandpile s_n in any ball B converges as the lattice spacing shrinks to zero. A natural refinement would be to show that for any ball B and any integer j, the fraction of sites in B with j particles converges. Understanding the scaling limit of the sandpile identity elements (Figure 8) is another appealing problem.

(4) By [45], the maximal elements of $\Gamma(\mathbb{Z}^2)$ correspond to the circles in the Apollonian band packing of $\mathbb{R}^2$. Because the radius and the coordinates of the center of each such circle are rational numbers, each maximal element of $\Gamma(\mathbb{Z}^2)$ is a matrix with rational entries. Describe the maximal elements of $\Gamma(\mathbb{Z}^d)$ for $d \geq 3$. Are they isolated? Do they have rational entries?

References

[1] D. Aharonov and H. S. Shapiro, Domains on which analytic functions satisfy quadrature identities, *J. Analyse Math.* **30** (1976), 39–73.

[2] Amine Asselah and Alexandre Gaudillière, From logarithmic to subdiffusive polynomial fluctuations for internal DLA and related growth models, *The Annals of Probability* 41.3A (2013): 1115–1159.

[3] Amine Asselah and Alexandre Gaudillière, Sublogarithmic fluctuations for internal DLA. *The Annals of Probability* 41.3A (2013): 1160–1179.

[4] Amine Asselah and Alexandre Gaudillière, Lower bounds on fluctuations for internal DLA, *Probability Theory and Related Fields* 158.1-2 (2014):39–53.

[5] Siva R. Athreya and Antal A. Járai, Infinite volume limit for the stationary distribution of Abelian sandpile models, *Communications in mathematical physics* 249.1 (2004): 197–213.

[6] László Babai, Lecture notes on the sandpile model. http://people.cs.uchicago.edu/~laci/REU05/#problem

[7] P. Bak, C. Tang and K. Wiesenfeld, Self-organized criticality: an explanation of the $1/f$ noise, *Phys. Rev. Lett.* **59**, no. 4 (1987), 381–384.

[8] Matthew Baker and Serguei Norine, Riemann-Roch and Abel-Jacobi theory on a finite graph, *Adv. Math.* **215**(2):766–788, 2007.

[9] Itai Benjamini, Russell Lyons, Yuval Peres and Oded Schramm, (2001). Special invited paper: uniform spanning forests. *Annals of probability* **29**: 1–65.

[10] Anders Björner, László Lovász and Peter Shor, Chip-firing games on graphs, *European J. Combin.* **12**(4):283–291, 1991.

[11] Luis A. Caffarelli, The obstacle problem revisited, *J. Fourier Anal. Appl.* **4** (1998), no. 4-5, 383–402.

[12] Hannah Cairns, Some halting problems for abelian sandpiles are undecidable in dimension three. arXiv:1508.00161

[13] Sergio Caracciolo and Andrea Sportiello, Exact integration of height probabilities in the Abelian Sandpile model, Journal of Statistical Mechanics: Theory and Experiment 2012.09 (2012): P09013.

[14] Joshua N. Cooper and Joel Spencer, Simulating a random walk with constant error, *Combin. Probab. Comput.* **15** (2006) 815–822. arXiv:math.CO/0402323.

[15] Deepak Dhar (1990): Self-organized critical state of sandpile automaton models, *Phys. Rev. Lett.* **64**, 1613.

[16] Deepak Dhar (2006): Theoretical studies of self-organized criticality, *Physica A* **369**: 29–70. See also arXiv:cond-mat/9909009

[17] Deepak Dhar, Tridib Sadhu and Samarth Chandra (2009): Pattern formation in growing sandpiles, *Europhys. Lett.* **85**, 48002. arXiv:0808.1732

[18] Persi Diaconis and William Fulton (1991): A growth model, a game, an algebra, Lagrange inversion, and characteristic classes, *Rend. Sem. Mat. Univ. Pol. Torino* **49**, no. 1, 95–119.

[19] Anne Fey-den Boer, Ronald Meestera and Frank Redig (2009): Stabilizability and percolation in the infinite volume sandpile model, *Ann. Probab.* **37**, no. 2, 654–675. arXiv:0710.0939

[20] Anne Fey, Lionel Levine and Yuval Peres, Growth rates and explosions in sandpiles, *J. Stat. Phys.* **138**: 143–159, 2010. arXiv:0901.3805

[21] Laura Florescu, Lionel Levine and Yuval Peres, The range of a rotor walk, 2014. arXiv:1408.5533

[22] Yasunari Fukai and Kôhei Uchiyama, Potential kernel for two-dimensional random walk. *Ann. Probab.* **24** (1996), no. 4, 1979–1992.

[23] Geoffrey R. Grimmett, The random-cluster model, Springer, 2006.

[24] Björn Gustafsson, Quadrature Identities and the Schottky double, *Acta Appl. Math.* **1** (1983), 209–240.

[25] B. Gustafsson and H. S. Shapiro, "What is a quadrature domain?" in Quadrature Domains and Their Applications, *Operator Theory: Advances and Applications* **156** (2005), 1–25.

[26] Lester L. Helms, *Potential Theory*, Springer, 2009.

[27] Alexander E. Holroyd, Lionel Levine, Karola Mészáros, Yuval Peres, James Propp and David B. Wilson, Chip-firing and rotor-routing on directed graphs, *In and out of equilibrium 2*, 331–364, Progr. Probab. **60**, Birkhäuser, 2008. arXiv:0801.3306

[28] Alexander E. Holroyd and James G. Propp, Rotor walks and Markov chains, *Algorithmic Probability and Combinatorics*, American Mathematical Society, 2010. arXiv:0904.4507

[29] Antal A. Járai, Sandpile models, 2014. arXiv:1401.0354

[30] Monwhea Jeng, Geoffroy Piroux and Philippe Ruelle (2006): Height variables in the Abelian sandpile model: scaling fields and correlations, *J. Stat. Mech.* **61** p. 15.

[31] David Jerison, Lionel Levine and Scott Sheffield, Logarithmic fluctuations for internal DLA *Journal of the American Mathematical Society* 25.1 (2012): 271–301. arXiv:1010.2483

[32] David Jerison, Lionel Levine and Scott Sheffield, Internal DLA in higher dimensions. *Electron. J. Probab.* 18.98 (2013): 1–14. arXiv:1012.3453

[33] David Jerison, Lionel Levine and Scott Sheffield, Internal DLA and the Gaussian free field. *Duke Mathematical Journal* 163.2 (2014): 267–308. arXiv:1101.0596

[34] Adrien Kassel, Learning about critical phenomena from scribbles and sandpiles, ESAIM: Proceedings and Surveys 51 (2015): 60–73.
[35] Adrien Kassel and David B. Wilson, Looping rate and sandpile density of planar graphs, 2014. arXiv:1402.4169
[36] Richard Kenyon, Spanning forests and the vector bundle Laplacian, The Annals of Probability (2011): 1983–2017. arXiv:1001.4028
[37] Richard Kenyon and David Wilson, Spanning trees of graphs on surfaces and the intensity of loop-erased random walk on planar graphs, *Journal of the American Mathematical Society* (2014). arXiv:1107.3377
[38] Gady Kozma and Ehud Schreiber, An asymptotic expansion for the discrete harmonic potential, *Electronic J. Probab.* **9**, no. 1, 1–17, 2004. arXiv:math/0212156
[39] Gregory F. Lawler, A self-avoiding walk, *Duke Math. J.* Volume 47, Number 3 (1980), 655–693.
[40] Gregory F. Lawler, Subdiffusive fluctuations for internal diffusion limited aggregation, *Ann. Probab.* **23** (1995) no. 1, 71–86.
[41] Gregory F. Lawler, *Intersections of Random Walks*, Birkhäuser, 1996.
[42] Gregory F. Lawler, Maury Bramson and David Griffeath, Internal diffusion limited aggregation, *Ann. Probab.* **20**, no. 4 (1992), 2117–2140.
[43] Gregory F. Lawler and Vlada Limic, *Random walk: a modern introduction*. Vol. 123. Cambridge University Press, 2010. http://www.math.uchicago.edu/~lawler/srwbook.pdf
[44] Lionel Levine, Wesley Pegden and Charles K. Smart, Apollonian structure in the Abelian sandpile, *Geometric and Functional Analysis*, to appear. arXiv:1208.4839
[45] Lionel Levine, Wesley Pegden and Charles K. Smart, The Apollonian structure of integer superharmonic matrices, 2013. arXiv:1309.3267
[46] Lionel Levine and Yuval Peres, Strong spherical asymptotics for rotor-router aggregation and the divisible sandpile, *Potential Anal.* **30** (2009), 1–27. arXiv:0704.0688
[47] Lionel Levine and Yuval Peres, Scaling limits for internal aggregation models with multiple sources, *J. Analyse Math.* 111.1 (2010): 151–219. arXiv:0712.3378
[48] Lionel Levine and Yuval Peres, The looping constant of $\mathbb{Z}^d$, *Random Structures & Algorithms* 45.1 (2014): 1–13.
[49] Dino J. Lorenzini, Arithmetical graphs, *Math. Ann.* **285**(3):481–501, 1989.
[50] László Lovász, Discrete analytic functions: an exposition, *Surveys in differential geometry* **IX**:241–273, 2004.
[51] S.N. Majumdar and Deepak Dhar, Height correlations in the Abelian sandpile model, *J. Phys. A: Math. Gen.* 24.7 (1991): L357.
[52] Madhusudan Manjunath and Bernd Sturmfels, Monomials, binomials and Riemann-Roch, *Journal of Algebraic Combinatorics* 37.4 (2013): 737–756.
[53] Fatemeh Mohammadi and Farbod Shokrieh, Divisors on graphs, connected flags, and syzygies, *International Mathematics Research Notices* 2014, no. 24: 6839–6905. arXiv:1210.6622
[54] Paul Meakin and J.M. Deutch (1986), The formation of surfaces by diffusion-limited annihilation, *J. Chem. Phys.* **85**: 2320.
[55] Cristopher Moore and Martin Nilsson. The computational complexity of sandpiles. *J. Stat. Phys.* **96**:205–224, 1999.
[56] Srdjan Ostojic, Patterns formed by addition of grains to only one site of an abelian sandpile, *Physica A* **318**:187–199, 2003.
[57] Robin Pemantle, Choosing a spanning tree for the integer lattice uniformly, *The Annals of Probability* (1991): 1559–1574.
[58] Wesley Pegden and Charles K. Smart, Convergence of the Abelian sandpile, *Duke Mathematical Journal* 162.4 (2013): 627–642.
[59] Wesley Pegden, Sandpile galleries. http://www.math.cmu.edu/~wes/sandgallery.html
[60] David Perkinson, Jacob Perlman and John Wilmes, Primer for the algebraic geometry of sandpiles. In *Tropical and non-Archimedean geometry* (2013), American Mathematical Society, pages 211–256. arXiv:1112.6163

[61] David Perkinson and Bryan Head, SandpilesApp. http://people.reed.edu/~davidp/sand/program/program.html
[62] V.S. Poghosyan and V.B. Priezzhev, The problem of predecessors on spanning trees, *Acta Polytechnica* Vol. 51 No. 1 (2011) arXiv:1010.5415
[63] A.M. Povolotsky, V.B. Priezzhev, and R.R. Shcherbakov, Dynamics of Eulerian walkers, *Physical Review E* 58.5 (1998): 5449.
[64] V. B. Priezzhev (1994): Structure of two dimensional sandpile. I. Height probabilities, *J. Stat. Phys.* **74**, 955-979.
[65] V. B. Priezzhev, Deepak Dhar, Abhishek Dhar and Supriya Krishnamurthy, Eulerian walkers as a model of self-organised criticality, *Phys. Rev. Lett.* **77**:5079–5082, 1996. arXiv:cond-mat/9611019
[66] V.S. Poghosyan, V.B. Priezzhev and P. Ruelle, Return probability for the loop-erased random walk and mean height in the Abelian sandpile model: a proof. *Journal of Statistical Mechanics: Theory and Experiment* 2011(10), P10004.
[67] James Propp, Random walk and random aggregation, derandomized, 2003. http://research.microsoft.com/apps/video/default.aspx?id=104906.
[68] S. Richardson, Hele-Shaw flows with a free boundary produced by the injection of fluid into a narrow channel, *J. Fluid Mech.* **56** (1972), 609–618.
[69] Makoto Sakai, *Quadrature Domains*, Lect. Notes Math. **934**, Springer, 1982.
[70] Makoto Sakai, Solutions to the obstacle problem as Green potentials, J. Analyse Math. **44** (1984/85), 97–116.
[71] Stanislav Smirnov, Discrete complex analysis and probability, *Proceedings of the International Congress of Mathematicians, Hyderabad, India*, 2010. arXiv:1009.6077
[72] Frank Spitzer, *Principles of Random Walk*, Springer, 1976.
[73] Kôhei Uchiyama, Green's functions for random walks on $\mathbb{Z}^N$, *Proc. London Math. Soc.* **77** (1998), no. 1, 215–240.
[74] Alexandr Nikolaevich Varchenko and Pavel I. Etingof, *Why the Boundary of a Round Drop Becomes a Curve of Order Four*, AMS University Lecture Series, vol. 3, 1992.
[75] Israel A. Wagner, Michael Lindenbaum and Alfred M. Bruckstein, Smell as a computational resource — a lesson we can learn from the ant, *4th Israeli Symposium on Theory of Computing and Systems*, pages 219–230, 1996.

PROBABILISTIC COMBINATORICS AND THE RECENT WORK OF PETER KEEVASH

W.T. GOWERS

1. The probabilistic method

A *graph* is a collection of points, or *vertices*, some of which are joined together by *edges*. Graphs can be used to model a wide range of phenomena: whenever you have a set of objects such that any two may or may not be related in a certain way, then you have a graph. For example, the vertices could represent towns and the edges could represent roads from one town to another, or the vertices could represent people, with an edge joining two people if they know each other, or the vertices could represent countries, with an edge joining two countries if they border each other, or the vertices could represent websites, with edges representing links from one site to another (in this last case the edges have directions – a link is *from* one website *to* another – so the resulting mathematical structure is called a *directed* graph).

Mathematically, a graph is a very simple object. It can be defined formally as simply a set V of vertices and a set E of unordered pairs of vertices from V. For example, if we take $V = \{1, 2, 3, 4\}$ and $E = \{12, 23, 34, 14\}$ (using ab as shorthand for $\{a, b\}$), then we obtain a graph known as the *4-cycle*.

Despite the simplicity and apparent lack of structure of a general graph, there turn out to be all sorts of interesting questions one can ask about them. Here is a classic example.

Question 1.1. *Does there exist a triangle-free graph with chromatic number 2016?*

Let me quickly explain what the various terms in the question mean. A *triangle* in a graph is, as one would expect, a triple of vertices x, y, z such that all of xy, yz and xz are edges. A graph is called *triangle free* if it contains no triangles. A *proper colouring* of a graph is a way of assigning colours to its vertices such that no two vertices of the same colour are joined by an edge. The *chromatic number* of a graph is the smallest number of colours you need in a proper colouring. (One of the most famous results of graph theory, the four-colour theorem, asserts that any graph that you can draw in the plane without any of its edges crossing each other has chromatic number at most 4. This can be interpreted as saying that if you want to colour the countries in a map in such a way that no two adjacent countries have the same colour, then four colours will suffice.)

How might one answer Question 1.1? The obvious thing would be to sit down and try to construct one. But this seems to be rather hard. How can we ensure that plenty of colours are needed? The only method that is immediately apparent is to have a large set of vertices that are all joined to each other by edges, since then all those vertices have to have different colours. But of course, if we do that, then we massively violate the other main constraint, that the graph should contain no triangles.

That suggests that perhaps one can always properly colour a triangle-free graph with a small number of colours. So how might we do that? Perhaps we could list the vertices in some order and colour them in that order, using new colours only when forced to do so. We could colour the first vertex red, then the second blue (if it is joined to the first). The third would be joined to at most one of the first two, so could be coloured either red or blue. And so on.

If one keeps going like this, one soon discovers that the number of colours one needs can grow without limit. Unfortunately, this does not solve the problem: it merely shows that choosing the colours in a greedy way does not work.

Although it is not the only way of solving the problem, there is an extraordinarily simple and powerful idea that does the job. It is to take an appropriate *random* graph. (This hint does not entirely spoil the problem, since it is still a very nice challenge to come up with an explicit construction of a graph that works.)

Here is how the argument works. Let p be a probability that we will choose later. Now let G be a random graph where we join a vertex x to a vertex y with probability p, making all choices independently.

If there are n vertices, then the expected number of triangles in the graph is at most $p^3\binom{n}{3}$. This means that it is possible to make the graph triangle free by deleting at most $p^3\binom{n}{3}$ edges.

We now make two simple observations. Recall that an *independent set* in a graph is a set of vertices, no two of which are joined by an edge. The very simple observation is that in a proper colouring, the vertices of any one colour must form an independent set, so if a graph has n vertices and chromatic number k, then it must have an independent set of size at least n/k. Therefore, to prove that the chromatic number of a graph is large, it is sufficient to find a small upper bound on the *independence number* of the graph – that is, the size of its largest independent set.

The second observation is that if every set of m vertices contains more than $p^3\binom{n}{3}$ edges, then we can remove an arbitrary set of $p^3\binom{n}{3}$ edges and will be left with a graph that has independence number greater than m. So our task is reduced to proving that every reasonably large set of vertices includes a reasonably large number of edges.

At this point I will not give full details. I will just say that the probability $q(m,r)$ that some given set of m vertices spans fewer than r edges is given by the binomial distribution $B(\binom{m}{2},p)$, for which we have standard estimates. So to ensure that with probability greater than $1/2$ every set of m vertices spans at least r edges, one needs to choose the parameters in such a way that $\binom{n}{m}q(m,r)<1/2$. This is not hard to do, and it leads to a proof that there exists a triangle-free graph with chromatic number proportional to $\sqrt{n}/\log n$. Finally, choosing n large enough, we obtain a triangle-free graph with chromatic number at least 2016. (If we want, we can remove edges from it until it has chromatic number exactly 2016, thereby answering Question 1.1 exactly as it was asked.)

2. When is the probabilistic method appropriate?

This is not an easy question to answer with complete precision, but there are some useful rough guidelines that one can give. Take the proof I have just sketched. After the two simple observations our aim became to find a triangle-free graph with

the property that for any set of m vertices there would be at least r edges between those vertices. This task has the following four properties.

- Graphs are very structureless objects, in the sense that a general graph has very few constraints that it needs to satisfy.
- It seems to be hard to define such a graph explicitly.
- We want to scatter the edges around very evenly.
- We want to do that very efficiently (so that we don't end up choosing so many edges that they bunch together to form triangles).

In such a situation, choosing edges at random makes very good sense: somehow, it allows us to do a lot of tasks in parallel, each with a very high chance of success.

Contrast that with a task such as solving the famous Burnside problem: does there exist a prime p and an infinite but finitely generated group such that every element has order p? It would be ludicrous to suggest picking a random group as a possible solution. That is partly because for an object to be a group, it has to satisfy a set of axioms that impose significantly greater constraints than those of a graph, where one simply chooses an arbitrary set of edges. Indeed, it is far from obvious what a good model for a random group should be. (There are in fact some very interesting models introduced by Gromov where one chooses a set of generators and a random set of relations of a certain length, but this model of random groups is of no help for the Burnside problem.) It is also because for the Burnside problem there is an obvious construction: choose the free group $G(k,p)$ on k generators subject to the relations that every element has order p. This is a well-defined group, and if such groups can be infinite, then this one must be. So the problem boils down to showing that some $G(k,p)$ is infinite.

There are also some intermediate problems – that is, problems that seem to have too many constraints for probabilistic methods to be appropriate, but too few for there to be obvious explicit constructions. An example of such a problem is the following beautiful question. Recall that a *Hamilton cycle* in a graph is a cycle that visits every vertex. That is, it is a sequence $v_1, v_2, \dots, v_n$ of vertices, where each vertex is included exactly once, such that $v_1v_2, v_2v_3, \dots, v_{n-1}v_n$ and v_nv_1 are all edges.

Question 2.1. *Let $n = 2m+1$ be an odd integer greater than or equal to 3 and define a graph G_n as follows. Its vertices are all subsets of $\{1,2,\dots,n\}$ of size m or $m+1$, and the edges are all pairs AB such that $|A| = m$, $|B| = m+1$, and $A \subset B$. Does every G_n contain a Hamilton cycle?*

The graph G_n is called the *middle layers* graph in the *discrete cube*.

If one starts trying to build a Hamilton cycle in G_n, one runs into the problem of having too much choice, and no obvious way of making it. (A natural thing to try to do is find some sort of inductive construction, but a lot of people have tried very hard to do this, with no success – a natural pattern just doesn't seem to emerge after the first few small cases.)

So there are not enough constraints to force one's hand and in that way lead one to a solution. At first this sounds like just the kind of situation for which the probabilistic method was designed: why not start at a set and then keep randomly choosing neighbouring vertices that have not yet been visited? Of course, that cannot be the whole story, since there is no guarantee that one will not get stuck

at some point, but perhaps one can combine this basic idea with a little bit of backtracking here and there and end up proving the existence of a Hamilton cycle.

Unfortunately, this does not seem to work either. After a while, the constraints start to bite, and one gets sufficiently stuck (at least potentially) that even being allowed a little local tinkering does not allow one to proceed.

So here we have a situation that is difficult because it is somehow "intermediate" between highly structured, where one has few options, so constructions, if they exist, are in a certain sense easier to find, and highly unstructured, where one has so many options that making random choices does not violate the constraints.

This well-known problem appears to have been recently solved by Torsten Mütze (`http://arxiv.org/abs/1404.4442`), though I do not know to what extent his argument has been thoroughly checked.

3. The existence of designs

A *Steiner triple system* is a finite set X and a collection T of triples of elements of X with the property that every pair x, y of elements of X is contained in exactly one triple xyz from T. One can think of it as a set of triangles that partition the edges of the complete graph with vertex set X.

Do Steiner triple systems exist? Well, an obvious constraint is that the number of edges in the complete graph with vertex set X should be a multiple of 3. That is, if $|X| = n$, we need that $3|\binom{n}{2}$. This holds if and only if $n \equiv 0$ or 1 mod 3. A slightly less obvious constraint (until it is pointed out) is that n should be odd. That is because the triangles xyz that contain a given vertex x will partition the $n-1$ edges that meet x into sets of size 2. Thus, n must, for very simple reasons, be congruent to 1 or 3 mod 6.

When $n = 1$ the empty set is a trivial Steiner system. When $n = 3$ a single triangle does the job. The next case, when $n = 7$, is more interesting: there is a famous example known as the *Fano plane*. If we number its vertices from 1 to 7, then the triples can be given as follows: 123, 246, 257, 167, 347, 356, 145. A less mysterious definition is to take as vertices the set of all triples of 0s and 1s apart from the triple 000, and then to take as our triple system (which must consist of triples of these triples!) the set of all xyz such that $x + y + z = 000$, where the addition is mod-2 in each coordinate. For example, one of the elements of this triple system is the triple xyz with $x = 100, y = 101$ and $x = 001$.

Why does this work? Well it is clear that an edge xy can be contained in at most one triple from the system, since the only possibility is the unique triple xyz such that $x + y + z = 000$, which forces z to equal $-(x + y)$, which is the same as $x + y$ since addition is mod 2. The only thing that could conceivably go wrong is if z turned out to equal x or y, but if $x + y$ is equal to x or y, then one of x and y is 000, which is not allowed.

Note that this construction gives an infinite family of Steiner triple systems, since the same proof works if we take non-zero 01-sequences of any fixed length k. So there are Steiner triple systems for every n of the form $2^k - 1$. It turns out that there are constructions similar in spirit to this one (making use of objects called quasigroups) that prove that Steiner triple systems exist for every n that satisfies the obviously necessary divisibility conditions – that is, for every n congruent to 1 or 3 mod 6.

A *design* is a generalization of a Steiner triple system, where instead of aiming to include every set of size 2 in exactly one set (from a carefully chosen collection) of size 3, one wishes to include every set of size r in exactly one set of size s. People also consider a further generalization where one wishes for every set of size r to be contained in exactly λ sets of size s. When $\lambda = 1$, the case I shall concentrate on here, designs are called *Steiner systems*.

No sooner is this definition presented, than several obvious questions immediately arise. Here are five, in decreasing order of optimism. Let us say that a *Steiner* (n, r, s)*-system* is a collection Σ of subsets size r of a set X of size n such that every subset of X of size s is contained in exactly one set from Σ.

Question 3.1. *If obvious necessary divisibility conditions are satisfied, does it follow that a Steiner* (n, r, s)*-system exists?*

Question 3.2. *If obvious necessary divisibility conditions are satisfied and* n *is sufficiently large, does it follow that a Steiner* (n, r, s)*-system exists?*

Question 3.3. *Is it the case that for all* $s < r$ *there are infinitely many* n *such that a Steiner* (n, r, s)*-system exists?*

Question 3.4. *Do Steiner* (n, r, s)*-systems exist with* s *arbitrarily large?*

Question 3.5. *Do Steiner* (n, r, s)*-systems exist with* r *arbitrarily large?*

For the rest of this note I want to discuss a remarkable recent result of Peter Keevash, but to set the scene let me give an idea of what was known before his work.

First, it was known that Question 3.1 was too optimistic: there are triples (n, r, s) for which no Steiner (n, r, s)-system exists even though the existence of such a system is not ruled out on simple divisibility grounds. Second, the answer to Question 3.5 was known to be positive: in a famous sequence of papers in the 1970s, Richard Wilson proved that for any fixed r, a Steiner $(n, r, 2)$-system exists provided that n is sufficiently large and satisfies the obviously necessary divisibility conditions. (His proof also gives a similar result for all higher values of λ.)

However, the answers to the questions in between were not known. There have been many ingenious constructions of designs for specific values of n, r and s (and λ if one makes that extra generalization), and a few small pairs (r, s) for which Question 3.2 has a positive answer. But to give an idea of our level of ignorance, no Steiner (n, r, s)-systems at all were known with $s > 5$. Thus, even an answer to Question 3.4 would have been a remarkable achievement. But Keevash went much further than this: he proved that the answer to Question 3.2 is yes! This is about as non-incremental a result as one could imagine: going from a situation where it was a huge struggle to prove the existence of even one design when s was even slightly large, to proving that for each r and s the only obstacle to the existence of a Steiner (n, r, s)-system was the trivial divisibility condition, for all but at most a finite set of n.

4. Are probabilistic methods appropriate for the problem?

In the previous section we saw a simple algebraic construction of a family of Steiner triple systems. That might suggest that explicit constructions are the right way of tackling this problem, and indeed much research in design theory has concerned using algebra to create interesting examples of designs.

On the other hand, the constraints on a Steiner system are not all that strong. Even for a triple system, if one wants to include all pairs in exactly one triple, the constraints are very slack: there is no sense at all in which one's moves are forced, at least to start with.

In fact, the situation here is rather similar to the situation with the problem about Hamilton cycles in middle-layer graphs. Suppose one just chooses sets of size r more or less arbitrarily, but making sure that no two of them intersect in a set of size s or more. (This ensures that no set of size s is contained in more than two sets from our collection.) For a long time we will have no problem, but eventually we will start to find that there are sets of size s that we do not seem to be able to cover with a set of size r without that set overlapping too much with a set we have already chosen. So the problem is difficult for a similar reason: there is too much choice for algebraic constructions to be easy to discover, and too little choice for simple probabilistic arguments to work.

Interestingly, however, it turns out that more sophisticated probabilistic arguments enable one to prove the existence of collections of sets that are *almost* Steiner systems. (This can mean one of two things: either one asks for no set of size s to be covered more than once and almost all sets of size s to be covered, or one asks for all sets of size s to be covered, and almost no sets of size s to be covered more than once. It turns out that if you can achieve one, then you can achieve the other.) Even more interestingly, Keevash's proof uses an intriguing *mixture* of probabilistic and algebraic methods, thus having the best of both worlds, and reflecting the "intermediate" nature of the problem.

5. The Rödl nibble

Let us return to the simplest non-trivial case of our general problem, that of Steiner triple systems. I shall give some indication of how probabilistic methods can be used to create almost Steiner triple systems. The basic technique was invented by Vojta Rödl, though an important precursor to it appeared in a paper of Ajtai, Komlós and Szemerédi.

For this example, Rödl's technique, which was dubbed the Rödl nibble, works as follows. Let us discuss the version of the problem where we are trying to cover all edges at least once and almost no edges more than once.

Let G_0 be the complete graph on n vertices. We begin by taking a small "bite" out of G_0, by choosing triangles randomly with probability $\epsilon/(n-2)$ for some small ϵ and removing them. The expected number of triangles we choose that contain any given edge xy is then ϵ (since there are $n-2$ possible z and each xyz has a probability $\epsilon/(n-2)$ of being chosen). We can in fact say more than this: the distribution of the number of triangles we choose that contain an edge xy is Poisson with mean ϵ.

Once we have done this, we will typically have covered a fraction roughly equal to ϵ of all the edges of the graph. We now eliminate all the edges in all the triangles that we have chosen and call the resulting graph G_1.

We then repeat the process for G_1. That is, we choose each triangle in G_1 independently with probability p, for a suitably chosen p, and remove all the edges in all the triangles we have chosen, obtaining a new graph G_2. Note that none of the triangles we choose in this second bite share an edge with any of the triangles we chose in the first bite, since they are all triangles in G_1.

But what is a suitable probability p? There would not be a clear answer to this question were it not for the fact that G_1 has an extremely useful property that is crucial to Rödl's argument: with very high probability it is *quasirandom*. I will not say exactly what a quasirandom graph is here, but roughly speaking a quasirandom graph is one that behaves in many ways like a random graph of the same density. It is a remarkable and important fact that this can be defined in a precise and useful way – in fact, it can be defined in several ways that turn out to be equivalent, not always for trivial reasons.

Once we know that G_1 is quasirandom, we know that almost all edges will be contained in approximately the same number of triangles in G_1. If this number is t, then we can choose the probability to be ϵ/t.

The Rödl nibble continues this process, showing at each stage that the graph G_r that results after r bites have been taken is quasirandom. Eventually, one loses control of the quasirandomness of G_r, but by that time there are so few edges that one can simply include each remaining edge in a triangle and one will not have covered too many edges more than once.

Rödl used a generalization of this argument to prove the existence of almost Steiner (n, r, s)-systems for all sufficiently large n. (If you do not require an exact Steiner system, then the divisibility constraints no longer apply.)

6. Keevash's contribution

Before Peter Keevash's work, the received wisdom was that with probabilistic methods, results like that of Rödl were the best one could hope for. Indeed, one can almost prove it: the probabilistic methods are insensitive to the divisibility constraints, and we know that Steiner systems do not exist when the divisibility constraints are not satisfied. Thus, it is very hard to see how a probabilistic argument could be made to work without also proving a false result. This does not of course completely rule out using probabilistic methods – it just means that it is very hard to see how one could use them.

A natural if hopeless-looking preliminary idea is that one might try to use probabilistic methods to get almost all the way, and then some kind of clever local backtracking and adjustments right at the end to get from an almost Steiner system to an exact one. To oversimplify a lot, this is roughly what Keevash does. His proof is long and complicated, and there is no hope of explaining it in a short time to a general audience, so I will content myself with trying to explain roughly how algebra enters the picture.

Recall the example we saw earlier of a family of Steiner triple systems. More or less the same example (suitably generalized for the more general Steiner systems) plays an important role in Keevash's proof, but he uses the finite field $\mathbb{F}^{2^k}$ rather than just the Abelian group $\mathbb{F}_2^k$.

More precisely, he chooses k such that $2n \leq 2^k < 4n$, takes a random map $\phi : V(G) \to \mathbb{F}^{2^k}$, and lets T be the system of triangles xyz such that $\phi(x) + \phi(y) + \phi(z) = 0$. As before, no two triangles in T share an edge. He then defines G^* to be the graph obtained by taking all the edges in all the triangles of T. This graph has positive density but there are many edges not included.

Next, he uses the Rödl nibble to cover the complement of G^* as well as he can, but for the reasons discussed above, the best he can hope for is to cover *most* of the edges outside G^* in this way. Let H be the union of all these edges. So now he

has an almost Steiner triple system, one part of which covers G^* and one part of which covers H.

Why is he in any better a situation than one would be after simply applying the Rödl nibble? This is (as it must be) the key to his argument. Whereas after applying the Rödl nibble, there is almost nothing useful one can say about the small graph that is not yet covered, we are now in a situation where we have the union of a small graph (about which we can still say almost nothing) and the graph G^*, about which it turns out that we can say rather a lot, thanks to its algebraic definition.

What kind of thing would one like to say about G^*? To answer this question, let us think about what we plan to do next. Let E be the small graph that consists of the edges that we have not managed to cover. Obviously we would now like to cover E, but there is no guarantee that E contains any triangles, so we are going to have to be prepared to modify the system of triangles we have already chosen. We do not touch any of the triangles used to cover H, because we do not know enough about it, but we are happy to touch the triangles used to cover G^*, because those are described very explicitly. As Keevash puts it in a key phrase, G^* "carries a rich structure of possible local modifications". Roughly speaking, given an edge of E, there are rather a lot of ways in which one can try to contain it in a triangle with the other two edges in G^*, and once one does contain it, there are many ways in which one can "repair the damage", by removing the triangles used to cover those other two edges and in a clever way organize for the triangles covering G^* to be adjusted slightly so that the edges that have just been uncovered get covered up again.

One might ask why it is necessary for G^* to be defined algebraically. Would it not be possible just to take G^* as a random union of edge-disjoint triangles? In fact, why bother with G^* at all? Why not just do the local modifications in H, which will look pretty random? Would the randomness not make it highly likely that local adjustments of the kind one wants to make are possible?

The answer to this turns out to be no. The reason is that algebraic constructions have a lot of small structures that one does not get with random constructions. For example, in both a random construction and an algebraic construction there will be a large number of triples of triangles of the form $xy'z', x'yz', x'y'z$, but whereas in the random case the probability that xyz also belongs to the system of triangles is proportional to n^{-1} (because the number of triangles we choose is proportional to n^2 and the total number of triples of vertices is proportional to n^3, and because there is almost no correlation between choosing the first three triangles and the fourth), in the algebraic case it is 1, since in $\mathbb{F}^{2^k}$ the equations

$$x + y' + z' = x' + y + z' = x' + y' + z = 0$$

imply the equation $x+y+z = 0$ (as can be seen by adding together the three sums). Thus, an algebraically constructed triangle system is rich in configurations that look like four alternate faces of an octahedron, and in other small configurations of a similar kind, whereas a randomly constructed system is not.

Something like this idea works in general, but becomes much more complicated – and it is already complicated even in the case of triangles and edges. Keevash uses an algebraic construction to define what he calls a "template", proves that the template is rich in small configurations that can be used to make local adjustments, and then uses that richness to deal with the small collection of sets that is left over

after the Rödl nibble has done what it can. Part of the reason that the proof is complicated is that it uses induction on s, so he has to consider simplicial complexes and not just collections of sets of a given size. Thus, the paper is a technical tour de force based on a beautiful underlying idea, and it solves one of the oldest problems in combinatorics.

7. Further reading

These notes are not meant as a formal document, and therefore I have not given a detailed bibliography. However, I should obviously give a reference to Keevash's original paper, or rather preprint, since it is not yet published. It is called "The existence of designs," and is available at `http://arxiv.org/abs/1401.3665`. It contains references to many important earlier papers in the subject, some of which I have alluded to above.

Gil Kalai presented Keevash's work to the Bourbaki Seminar. His write-up, which goes into more detail than I have (though nothing like full detail), can be found here: `http://www.bourbaki.ens.fr/TEXTES/1100.pdf`.

Finally, an excellent introduction to the probabilistic method is the book The Probabilistic Method, by Noga Alon and Joel Spencer: `http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470170204.html`. Alternatively, it is easy to find many good treatments of the method online.

What are Lyapunov exponents, and why are they interesting?

Amie Wilkinson

Introduction

Taking as inspiration the Fields Medal work of Artur Avila, I'd like to introduce you to Lyapunov exponents. My plan is to show how Lyapunov exponents play a key role in three areas in which Avila's results lie: smooth ergodic theory, billiards and translation surfaces, and the spectral theory of 1-dimensional Schrödinger operators.

But first, what are Lyapunov exponents? Let's begin by viewing them in one of their natural habitats. The barycentric subdivision of a triangle is a collection of 6 smaller triangles obtained by joining the midpoints of the sides to opposite vertices. Here's what happens when you start with an equilateral triangle and repeatedly barycentrically subdivide:

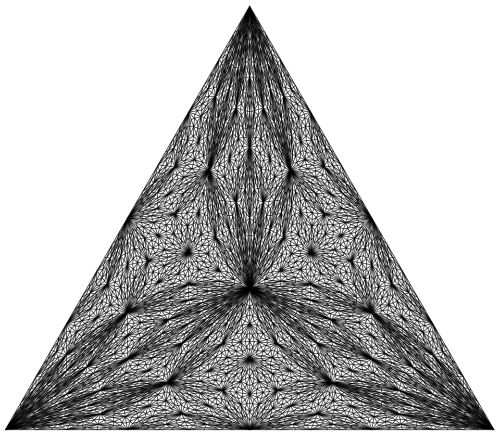


Figure 1: Iterated barycentric subdivision, from [McM].

As the subdivision gets successively finer, notice that many of the triangles produced by subdivision get increasingly eccentric and needle-like. We can measure the skinniness of a triangle T via the aspect ratio $\alpha(T) = \text{area}(T)/L(T)^2$, where $L(T)$ is the length of the long side. Suppose we label the triangles in a possible subdivision 1 through 6, roll a six-sided die and at each stage choose a triangle to subdivide. The sequence of triangles $T_1 \supset T_2 \supset \dots$ obtained have aspect ratios $\alpha_1, \alpha_2, \dots$, where $\alpha_n = \alpha(T_n)$.

Theorem 0.1. *There exists a real number $\chi < 0$ such that almost surely:*

$$\lim_{n\to\infty} \frac{1}{n} \log \alpha_n = \chi.$$

In other words, there is a universal constant $\chi < 0$, such that if triangles are chosen successively by a random coin toss, then with probability 1, their aspect ratios will tend to 0 at an exponential rate governed by $\exp(n\chi)$. This magical number χ is a Lyapunov exponent. For more details, see [BBC] and [McM].

0.1 Cocycles, hyperbolicity and exponents

Formally, Lyapunov exponents are quantities associated to a cocycle over a measure-preserving dynamical system. A measure-preserving dynamical system is a triple (Ω, μ, f), where (Ω, μ) is a probability space, and $f\colon \Omega \to \Omega$ is a map preserving the measure μ, in the sense that $\mu(f^{-1}(X)) = \mu(X)$ for every measurable $X \subset \Omega$.

Here is a short list of examples of measure preserving systems that also turns out to be quite useful for our purposes.

- *Rotations on the circle.* On the circle $\Omega = \mathbb{R}/\mathbb{Z}$, let $f_\alpha(x) = x + \alpha \pmod 1$, where $\alpha \in \mathbb{R}$ is fixed. This preserves the Lebesgue-Haar measure on μ on the circle, which assigns to an interval I its length $|I|$.

- *Toral automorphisms.* Let $\Omega = \mathbb{T}^2 := \mathbb{R}^2/\mathbb{Z}^2$, the 2-torus. Let $A \in SL(2, \mathbb{Z})$, for example $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. Then A acts linearly on the plane by multiplication and preserves the lattice $\mathbb{Z}^2$, by virtue of having integer entries and determinant 1. It therefore induces a map $f_A\colon \mathbb{T}^2 \to \mathbb{T}^2$ of the 2-torus, a group automorphism. The area μ is preserved because $\det(A) = 1$.

- *The Bernoulli shift.* Let $\Omega = \{1, \dots, k\}^{\mathbb{N}}$ be the set of all infinite, one sided strings $\omega = (\omega_1, \omega_2, \cdots)$ on the alphabet $\{1, \dots, k\}$. The shift map $\sigma \colon \Omega \to \Omega$ is defined by $\sigma(\omega)_k = \omega_{k+1}$. Any probability vector $p = (p_1, \dots, p_k)$ (i.e. with $p_i \in [0, 1]$, and $\sum_i p_i = 1$) defines a product measure $\mu = p^{\mathbb{N}}$ on Ω. It is easy to see that the shift σ preserves μ.

A measurable map $A \colon \Omega \to M_{d\times d}$ into the space $d \times d$ matrices (real or complex) is called a cocycle over f. For each $n > 0$, and $\omega \in \Omega$, we write

$$A^{(n)}(\omega) = A(f^{n-1}(\omega)A(f^{n-2}(\omega) \cdots A(f(\omega))A(\omega),$$

where f^n denotes the n-fold composition of f with itself. For $n = 0$, we set $A^{(n)}(\omega) = I$, and if f is invertible, we also define, for $n \geq 1$:

$$A^{(-n)}(\omega) = (A^{(n)}(f^{-n}(\omega)))^{-1} = A^{-1}(f^{(-n+1)}(\omega)) \cdots A^{-1}(\omega).$$

Using the language of cocycles, we can encode the behavior of a random product of matrices. Let $\{A_1, \dots, A_k\} \subset M_{d\times d}$ be a finite collection of matrices. Suppose we take a k-sided (Dungeons and Dragons) die and roll it repeatedly. If the die comes up with the number j, we choose the matrix A_j,

Figure 2: Four-sided Dungeons and Dragons die.

thus creating a sequence $A_{\omega_1}, A_{\omega_2}, \dots$, where $\omega = (\omega_1, \omega_2, \dots) \in \{1, \dots, k\}^{\mathbb{N}}$. This process can be packaged in a cocycle A over a measure preserving system (Ω, μ, σ) by setting $\Omega = \{1, \dots, k\}^{\mathbb{N}}$, $\mu = (p_1, \dots, p_k)^{\mathbb{N}}$, σ to be the shift map, and $A(\omega) = A_{\omega_1}$, where p_j is the probability that the die shows j on a roll. Then $A_n(\omega)$ is simply the product of the first n matrices produced by this process.[1]

Another important class of cocycle is the *derivative cocycle*. Let $f \colon M \to M$ be a C^1 map on a compact d-manifold M preserving a probability measure μ. Suppose for simplicity that the tangent bundle is trivial: $TM =$

[1] More generally, suppose that η is a probability measure on the set of matrices $M_{d\times d}$. The space $\Omega = M_{d\times d}^{\mathbb{N}}$ of sequences $(M_1, M_2, \dots)$ carries the product (Bernoulli) measure $\eta^{\mathbb{N}}$ which is invariant under the shift map σ, where as above $\sigma(M_1, M_2, \dots) = (M_2, M_3, \dots)$. There is a natural cocycle $A \colon \Omega \to M_{d\times d}$ given by $A((M_1, M_2, \dots)) = M_1$. The matrices $A^{(n)}(\omega)$, for $\omega \in \Omega$ are just n-fold random products of matrices chosen independently with respect to the measure η.

$M \times \mathbb{R}^d$. Then for each $x \in M$, the derivative $D_x f \colon T_x M \to T_{fx} M$ can be written as a matrix $D_x f \in M_{d\times d}$.[2] The Chain Rule implies that if $A = Df$ is a derivative cocycle, then $D_x f^n = A^{(n)}(x)$. A simple example of the derivative cocycle is provided by the toral automorphism $f_A \colon \mathbb{T}^2 \to \mathbb{T}^2$ described above. Conveniently, the tangent bundle to $\mathbb{T}^2$ is trivial, and the derivative cocycle is the constant cocycle $D_x f_A = A$.

Before defining Lyapunov exponents, we mention an important concept called uniform hyperbolicity. A continuous cocycle A over a homeomorphism $f \colon \Omega \to \Omega$ of a compact metric space Ω is *uniformly hyperbolic* if there exists $n \geq 1$, and for every $\omega \in \Omega$, there is a continuous splitting $\mathbb{R}^d = E^u(\omega) \oplus E^s(\omega)$ such that, for every $\omega \in \Omega$:

- $A(\omega)E^u(\omega) = E^u(f(\omega))$, and $A(\omega)E^s(\omega) = E^s(f(\omega))$,
- $v \in E^u(\omega) \setminus \{0\} \implies \|A^{(n)}(\omega)v\| \geq 2\|v\|$, and
- $v \in E^s(\omega) \setminus \{0\} \implies \|A^{(-n)}(\omega)v\| \geq 2\|v\|$.

Notice that measure plays no role in the definition of uniform hyperbolicity. It is a topological property of the cocycle. Hyperbolicity is an *open* property of both the cocycle A and the dynamics f: that is, if A is uniformly hyperbolic over f, and we make a uniformly small perturbations to both A and f, then new cocycle will also be uniformly hyperbolic over the new homeomorphism.

A diffeomorphism $f \colon M \to M$ whose derivative cocycle is uniformly hyperbolic is called *Anosov*.[3] Anosov diffeomorphisms remain Anosov when the dynamics are perturbed in a C^1 way, by the openness of uniform hyperbolicity of cocycles.

The toral automorphism $f_A \colon \mathbb{T}^2 \to \mathbb{T}^2$, with $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ is Anosov; since the derivative cocycle is constant, the splitting $\mathbb{R}^2 = E^u(x) \oplus E^s(x)$, for $x \in \mathbb{T}^2$ does not depend on x: $E^u(x)$ is the expanding eigenspace for A corresponding to the larger eigenvalue $\lambda = (3 + \sqrt{5})/2 > 1$, and $E^s(x)$ is the contracting eigenspace for A corresponding to the smaller eigenvalue $\lambda^{-1} = (3 - \sqrt{5})/2 < 1$. In this example, we can choose $n = 2$.

[2] The case where TM is not trivializable is easily treated: either one trivializes TM over a full measure subset of M, or one expands the definition of cocycle to include bundle morphisms over f.

[3] Again, one needs to modify this definition when the tangent bundle TM is nontrivial. The splitting of R^d in the definition is replaced by a splitting $T_x M = E^u(x) \oplus E^s(x)$ into subspaces.

A real number χ is a *Lyapunov exponent for the cocycle A over (Ω, μ, f) at $\omega \in \Omega$* if there exists a nonzero vector $v \in \mathbb{R}^d$, such that

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^{(n)}(\omega)v\| = \chi. \tag{1}$$

Oseledec proved in 1968 [Os] that for any cocycle A over a measure preserving system (Ω, μ, f) satisfying the integrability condition $\log^+ \|A\| \in L^1(\Omega, \mu)$, for μ-almost every $\omega \in \Omega$ and for every nonzero $v \in \mathbb{R}^d$ the limit in (1) exists. It is not hard to see that this limit can assume at most d distinct values $\chi_1(\omega) > \chi_2(\omega) > \cdots > \chi_{k(\omega)}(\omega)$, where $k(\omega) \leq d$.

We say that a cocycle A over (Ω, μ, f) is *(measurably) hyperbolic* if for μ-a.e. ω, the exponents $\chi_j(\omega)$ are all nonzero. Since the role played by the measure is important in this definition, we sometimes say that μ is a hyperbolic measure for the cocycle A. Uniformly hyperbolic cocycles over a homeomorphism f are hyperbolic with respect to *any* f-invariant measure (exercise). On the other hand, in the nonuniform setting it is possible to be hyperbolic with respect one invariant measure, but not another.[4] For the toral automorphism f_A described above, the Lyaponov exponents with respect to any invariant measure, at any point, exist and equal $\pm \log(\lambda)$. This is a very special sitatuation.

Lyapunov exponents play an extensive role in the analysis of dynamical systems. Three areas that are touched especially deeply are smooth dynamics, billiards, and the spectral theory of 1-dimensional Schrödinger operators. What follows is a brief sampling of Avila's results in each of these areas. The last section is devoted to a discussion of some of the themes that arise in the study of Lyapunov exponents.

1 Ergodicity of "typical" diffeomorphisms

Smooth ergodic theory studies the dynamical properties of smooth maps from a statistical point of view. A natural object of study is a measure-preserving system (M, vol, f), where M is a smooth, compact manifold without boundary, vol is a probability measure on M equivalent to the volume, and $f\colon M \to M$ is a diffeomorphism preserving vol. Such a diffeomorphism is *ergodic* if its orbits are equidistributed, in the following sense: for almost

[4]The terminology is not consistent across fields. In smooth dynamics, a cocycle over a measurable system that is measurably hyperbolic is called nonuniformly hyperbolic, whether it is uniformly hyperbolic or not. In the spectral theory community, a cocycle is called nonuniformly hyperbolic if it is measurably hyperbolic but *not* uniformly hyperbolic.

every $x \in M$, and any continuous function $\phi\colon M \to \mathbb{R}$:

$$\lim_{n\to\infty} \frac{1}{n}\left(\phi(x) + \phi(f(x)) + \cdots \phi(f^{n-1}(x))\right) = \int_M \phi \, d\text{vol}.$$

An example of an ergodic diffeomorphism is the rotation f_α on $\mathbb{R}/\mathbb{Z}$, for α irrational. In fact this transformation has a stronger property called *unique ergodicity*, which means that the limit above exists for *every* $x \in \mathbb{R}/\mathbb{Z}$.[5] This is a highly degenerate example, as it is easily perturbed to be non-ergodic.

Another example of an ergodic diffeomorphism, in some sense at the opposite extreme of the rotations, is the automorphism f_A of the 2-torus induced by multiplication by the matrix $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. In spirit, this example is closely related to the Bernoulli shift, and in fact its orbits can be coded in such a way to produce a measure-preserving isomorphism with a Bernoulli shift. The reason this map is ergodic is uniform hyperbolicity: Anosov proved [An1] that any smooth uniformly hyperbolic, i.e. Anosov, diffeomorphism that preserves volume is ergodic.

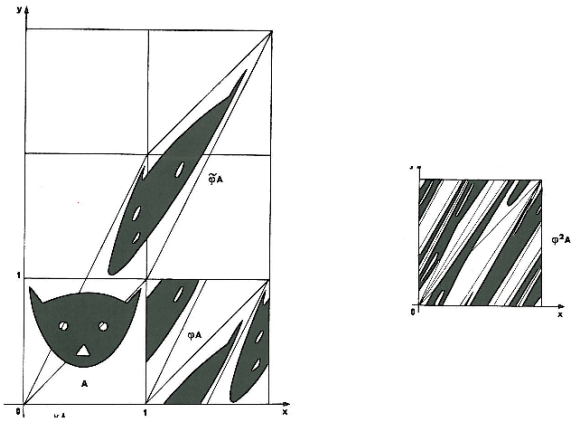


Figure 3: The action of f_A on a cat, from [AA].

Anosov's proof of ergodicity is involved, but viewing the action of f_A on a fundamental domain, one sees that f_A mixes up sets quite a bit. See Figure 3. This is an example of a *stably ergodic* diffeomorphism: since the

[5]This is a consequence of Weyl's equidistribution theorem and can be proved using elementary analysis. See, e.g. [He]. Note that f_α is definitely not ergodic when $\alpha = p/q$, for then $f_\alpha^q = id$.

Anosov property is a C^1-open one, the ergodicity cannot be destroyed by a small perturbation, in marked contrast with the irrational rotation f_α.

The question of whether ergodicity is a common property among diffeomorphisms is an old one, going back to Boltzmann's ergodic hypothesis of the late 19th Century, We can formalize the question by fixing a differentiability class $r \in [1, \infty]$ and considering the set $\mathrm{Diff}^r_{\mathrm{vol}}(M)$ of C^r, volume-preserving diffeomorphisms of M. This is a topological space in the C^r topology, and we say that a property holds *generically* in $\mathrm{Diff}^r_{\mathrm{vol}}(M)$ if it holds for all f in a countable intersection of open and dense subsets of $\mathrm{Diff}^r_{\mathrm{vol}}(M)$.[6]

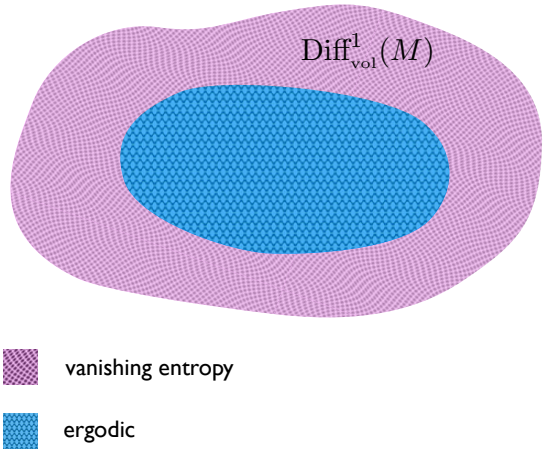


Figure 4: Generically, positive entropy implies ergodicity (and more).

Oxtoby and Ulam [OU] proved in 1939 that the generic volume-preserving *homeomorphism* of a compact manifold is ergodic. At the other extreme, KAM (Kolmogorov-Arnol'd-Moser) theory shows that ergodicity is *not* a dense property, let alone a generic one, in $\mathrm{Diff}^\infty_{\mathrm{vol}}(M)$, if $\dim(M) \geq 2$. The general question remains open for $r \in [1, \infty)$, but we now have a complete answer for any manifold when $r = 1$ under the assumption of *positive entropy*. Entropy is a numerical invariant attached to a measure preserving system that measures the complexity of orbits. The rotation f_α has entropy 0; the Anosov map f_A has positive entropy $\log(\lambda)$. By a theorem of Ruelle, positivity of entropy means that there is *some* positive volume subset of M on which the Lyapunov exponents are nonzero in *some* directions.

[6]Since $\mathrm{Diff}^r_{\mathrm{vol}}(M)$ is a Baire space, properties that hold generically hold for a dense set, and two properties that hold generically separately hold together generically.

Theorem 1.1 (Avila, Crovisier, Wilkinson [ACW]). *Generically in* $\mathrm{Diff}^1_{vol}(M)$, *positive entropy implies ergodicity, and moreover* measurable hyperbolicity.

See Figure 4. This result was proved in dimension 2 by Mañé-Bochi [Ma, Boc1] and dimension 3 by M.A. Rodriguez-Hertz [R]. Positive entropy is an *a priori* weak form of chaotic behavior that can be confined to an invariant set. Measurable hyperbolicity means that at almost every point *all* of the Lyapunov exponents of the derivative cocycle Df are nonzero. Conceptually, the proof divides into two parts:

1. C^1 generically, positive entropy implies nonuniform hyperbolicity. One needs to go from some nonzero exponents on some of the manifold to all nonzero exponents on almost all of the manifold. Since the cocycle and the dynamics are intertwined, carrying this out is a delicate matter. This relies on the relative flexibility of the C^1 topology.

2. C^1 generically, measurable hyperbolicity (with respect to volume) implies ergodicity. This has the flavor of arguments going back to E. Hopf (and later Anosov) which show that uniform hyperbolicity implies ergodicity. It builds on techniques later devloped by Pesin in the nonuniform setting [P].

2 Translation surfaces

A flat surface is a closed surface obtained by gluing together finitely many parallelograms in $\mathbb{R}^2$ along parallel edges:

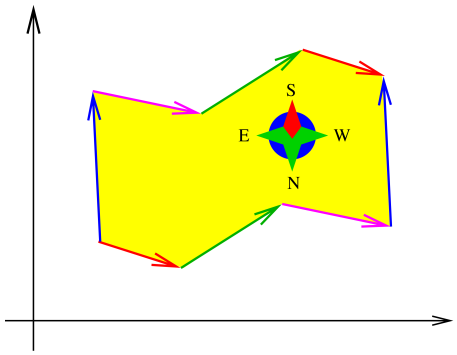


Figure 5: A flat surface with a distinguished "South," also known as a translation surface (courtesy Marcelo Viana).

Two flat surfaces are equivalent if one can be obtained from the other by cutting, translating, and rotating. A *translation* surface is a flat surface that comes equipped with a well-defined, distinguished vertical, "North" direction (or, "South" depending on your orientation). Two translation surfaces are equivalent if one can be obtained from the other by cutting and translating (but *not* rotating).

Fix a translation surface Σ of genus $g > 0$. If one picks an angle θ and a point x on Σ, and follows the corresponding straight ray through Σ, there are two possibilities: either it terminates in a corner, or it can be continued for all time. Clearly for any θ, and almost every starting point (with respect to area), the ray will continue forever. If it continues forever, either it returns to the initial point and direction and produces a closed curve, or it continues on a parallel course without returning. A version of the Pigeonhole Principle for area (Poincaré recurrence) implies that for almost every point and starting direction, the line will come back arbitrarily close to the starting point.

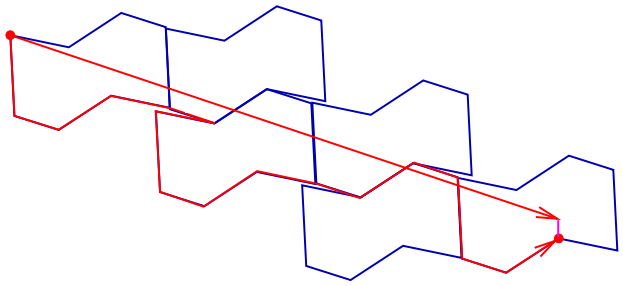


Figure 6: Closing up a ray that comes back close to itself (courtesy Marcelo Viana)

Kerckhoff-Masur-Smillie [KMS] proved more: for a fixed Σ, and almost every θ, the ray through any point x is dense in Σ, and in fact is equidistributed with respect to area. Such a direction θ is called *uniquely ergodic*, as it is uniquely ergodic in the same sense that f_α is, for irrational α. Suppose we start with a uniquely ergodic direction and wait for the successive times that this ray returns closer and closer to itself. This produces a sequence of closed curves γ_n which produces a sequence of cycles $[\gamma_n]$ in homology $H_1(\Sigma, \mathbb{Z}) \simeq \mathbb{Z}^{2g}$.

Unique ergodicity of the direction θ implies that there is a unique $c_1 \in$

$H_1(\Sigma, \mathbb{R})$ such that for any starting point x:

$$\lim_{n\to\infty} \frac{[\gamma_n]}{\ell(\gamma_n)} = c_1,$$

where $\ell(\gamma)$ denotes the length in Σ of the curve γ.

Theorem 2.1 (Forni, Avila-Viana, Zorich [Fo, AV2, Zo1, Zo2])**.** *Fix a topological surface S of genus $g \geq 1$, and let Σ be almost any translation surface modelled on S.* [7] *Then there exist real numbers $1 > \nu_2 > \ldots > \nu_g > 0$ and a sequence of of subspaces $L_1 \subset L_2 \subset \cdots L_g$ of $H_1(\Sigma, \mathbb{R})$ with $\dim(L_k) = k$ such that for almost every θ, for every x, and every γ in direction θ, the distance from $[\gamma]$ to L_g is bounded, and*

$$\limsup_{\ell(\gamma)\to\infty} \frac{\log \operatorname{dist}([\gamma], L_i)}{\log(\ell(\gamma))} = \nu_{i+1},$$

for all $i < g$.

This theorem gives precise information about the way the direction of $[\gamma_n]$ converges to its asymptotic cycle c_1: the convergence has a "directional nature" much in the way a vector $v \in \mathbb{R}^d$ converges to infinity under repeated application of a matrix

$$A = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & \lambda_2 & \cdots & * \\ 0 & \cdots & \cdots & * \\ 0 & 0 & \cdots & \lambda_d \end{pmatrix},$$

with $\lambda_1 > \lambda_2 > \cdots \lambda_d > 1$.

The numbers ν_i are the Lyapunov exponents of the *Kontsevich-Zorich* (KZ) cocycle over the so-called *Teichmüller flow*. The Teichmüller flow $\mathcal{F}_t$ acts on the moduli space $\mathcal{M}$ of translation surfaces (that is, translation surfaces modulo cutting and translation) by stretching in the East-West direction and contracting in the North-South direction. More precisely, if Σ is a translation surface, then $\mathcal{F}_t(\Sigma)$ is a new surface, obtained by transforming Σ by the linear map $\begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}$. Since a stretched surface can often

[7] "Almost any" means with respect to the Lebesgue measure on possible choices of lengths and directions for the sides of the pentagon. This statement can be made more precise in terms of Lebesgue measure restricted to various strata in the moduli space of translation surfaces.

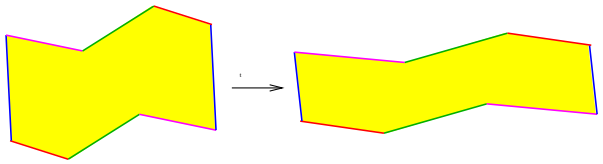


Figure 7: A local picture of the Teichmüller flow (courtesy Marcelo Viana).

be reassembled to obtain a more compact one, it is plausible that the Teichmüller flow has recurrent orbits (for example, periodic orbits). This is true and reflects the fact that the flow $\mathcal{F}_t$ preserves a natural volume that assigns finite measure to $\mathcal{M}$. The Kontsevich-Zorich cocycle takes values in the symplectic group $Sp(2g, \mathbb{R})$ and captures homological data about the cutting and translating equivalence on the surface.

Veech proved that $\nu_2 < 1$, Forni proved that $\nu_g > 0$, and Avila-Viana proved that the numbers $\nu_2, \nu_3, \ldots, \nu_{g-1}$ are all distinct. Zorich established the connection between exponents and the so-called deviation spectrum, which holds in greater generality. Many more things have been proved about

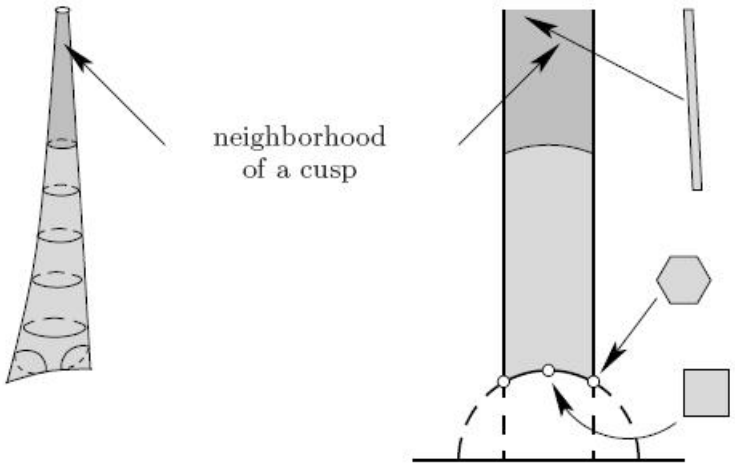


Figure 8: The moduli space of flat structures on the torus, a.k.a. the modular surface.

the Lyapunov exponents of the KZ cocycle, and some of their values have been caclulated which are (until recently, conjecturally) rational numbers! See [EKM, CE].

In the $g = 1$ case where Σ is a torus, this result has a simple explanation. The moduli space $\mathcal{M}$ is the set of all flat structures on the torus (up to homothety), equipped with a direction. This is the quotient $SL(2,\mathbb{R})/SL(2,\mathbb{Z})$, which is the unit tangent bundle of the modular surface $\mathbb{H}/SL(2,\mathbb{Z})$. The (continuous time) dynamical system on Ω is the flow $\mathcal{F}_t$ on Ω given by left multiplication by the matrix $\begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}$. The cocycle is, in essence, the derivative cocycle for this flow (transverse to the direction of the flow) This flow is *uniformly hyperbolic* (i.e. Anosov), and its exponents are $\log(e) = 1$ and $\log(e^{-1}) = -1$.

The proof for general translation surfaces is considerably more involved. We can nonetheless boil it down to some basic ideas.

1. The Teichmüller flow itself is nonuniformly hyperbolic with respect to a natural volume (Veech [Ve]), and can be coded in a way that the dynamics appear almost random.

2. Cocycles over perfectly random systems (for example i.i.d. sequences of matrices) have a tendency to have distinct, nonzero Lyapunov exponents. This was first proved by Furstenberg in the 2×2 case [F] and later by Gol'dsheid-Margulis [GM] and Guivarc'h-Raugi [GR].

3. Cocycles over systems that are nonrandom, but sufficiently hyperbolic and with good coding also tend to have distinct, nonzero Lyapunov exponents. This follows from series of results, beginning with Ledrappier in the 2×2 case [Le], and in increasing generality by Bonatti-Viana [BoVi], Viana [Vi], and Avila-Viana [AV1].

3 Hofstadter's butterfly

Pictured in Figure 9 is the spectrum of the operator $H_x^\alpha \colon \ell^2(\mathbb{Z}) \to \ell^2(\mathbb{Z})$ given by

$$[H_x^\alpha u](n) = u(n+1) + u(n-1) + 2\cos(2\pi(x + n\alpha))u(n),$$

where x is a fixed real number called the *phase*, and $\alpha \in [0, 1]$ is a parameter called the *frequency*. The vertical variable is α, and the horizontal variable is the spectral energy parameter E, which ranges in $[-4, 4]$. We can read off the spectrum of H_x^α by taking a horizontal slice at height α; the black region is the spectrum.

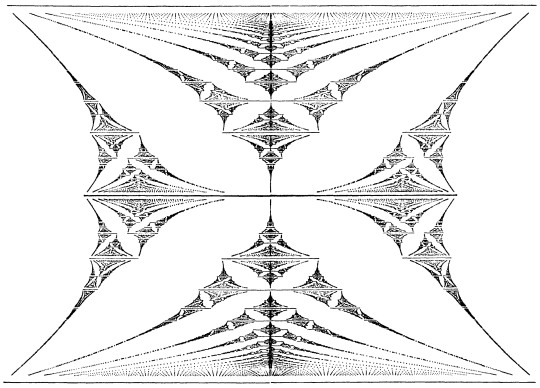


Figure 9: Hofstadter's butterfly, from [Ho].

In an influential 1976 paper, Douglas Hofstadter of *Gödel, Escher Bach* fame discovered this fractal picture while modelling the behavior of electrons in a crystal lattice under the force of a magnetic field [Ho]. This operator plays a central role in the Thouless et al. theory of the integer quantum Hall effect, and the butterfly has indeed appeared in von Klitzing's QHE experiments. Similar operators are used in modeling graphene and similar butterflies also appear in graphene related experiments.

Some properties of the butterfly have been established rigorously. For example, Avila and Krikorian proved:

Theorem 3.1 (Avila-Krikorian, [AK])**.** *For every irrational* $\alpha \in [0,1]$*, the* α*-horizontal slice of the butterfly has measure* 0*.*

Their proof complements and thus extends the earlier result of Last [La], who proved the same statement, but for a full measure set of α satisfying an arithmetic condition. In particular, we have:

Corollary 3.2. *The butterfly has measure* 0*.*

Some properties of the butterfly, for example its Hausdorff dimension, remain unknown.

The connection between the spectrum of this operator and cocycles is an interesting one. Recall the definition of the spectrum of H_x^α:

$$\sigma(H_x^\alpha) := \{E \in \mathbb{C} : H_x^\alpha - E \text{ is not invertible}\}.$$

The eigenvalues are those E so that the eigenvalue equation $H_x^\alpha u = Eu$ admits $\ell^2(\mathbb{Z})$ solutions.

The following simple observation is key. A sequence $(u_n : n \in \mathbb{Z}) \subset \mathbb{C}^{\mathbb{Z}}$ (not necessarily in $\ell^2(\mathbb{Z})$) solves $H_x^\alpha u = Eu$ if and only if

$$A_E(f_\alpha^n(x))\begin{pmatrix} u_n \\ u_{n-1} \end{pmatrix} = \begin{pmatrix} u_{n+1} \\ u_n \end{pmatrix},\ n \in \mathbb{Z},$$

where $f_\alpha \colon \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ is the translation mentioned above, and

$$A_E(x) = \begin{pmatrix} E - 2\cos(2\pi x) & -1 \\ 1 & 0 \end{pmatrix}, \tag{2}$$

which defines an $SL(2,\mathbb{R})$-cocycle, an example of a *Schrödinger cocycle*. Using the cocycle notation, we have

$$A_E^{(n)}(x)\begin{pmatrix} u_0 \\ u_{-1} \end{pmatrix} = \begin{pmatrix} u_n \\ u_{n-1} \end{pmatrix},\ n \in \mathbb{Z}.$$

Now let's connect the properties of this cocycle with the spectrum of H_x^α. Suppose for a moment that the cocycle A_E over f_α is uniformly hyperbolic, for some value of E. Then for every $x \in \mathbb{R}/\mathbb{Z}$ there is a splitting $\mathbb{R}^2 = E^u(x) \oplus E^s(x)$ invariant under cocycle, with vectors in $E^u(x)$ expanded under $A_E^{(mn)}(x)$, and vectors in $E^s(x)$ expanded under $A_E^{(-mn)}(x)$, both by a factor of 2^m. Thus no solution u to $H_x^\alpha u = Eu$ can be polynomially bounded simultaneously in both directions, which implies E is not an ℓ^2 eigenvalue of H_x^α. It turns out that the converse is also true, and moreover:

Theorem 3.3 (R. Johnson, [J])**.** *If α is irrational, then for every $x \in [0,1]$:*

$$\sigma(H_x^\alpha) = \{E : A_E \textit{ is not } \textit{uniformly hyperbolic over } f_\alpha\}. \tag{3}$$

For irrational α, we denote by Σ_α the spectrum of $\sigma(H_x^\alpha)$, which by Theorem 3.3 does not depend on x. Thus for irrational α, the set Σ_α is the α-horizontal slice of the butterfly.

The butterfly is therefore both a dynamical picture and a spectral one. On the one hand, it depicts the spectrum of a family of operators H_x^α parametrized by α, and on the other hand, it depicts, within a 2-parameter family of cocycles $\{(f_\alpha, A_E) : (E,\alpha) \in [-4,4] \times [0,1]\}$, the set of parameters corresponding to dynamics that are *not* uniformly hyperbolic.

Returning to spectral theory, let's continue to explore the relationship between spectrum and dynamics. If α is irrational, then f_α is ergodic, and Oseledec's theorem implies that the Lyapunov exponents for any cocycle over f_α take constant values over a full measure set. Thus the Lyapunov

exponents of A_E over f_α take two essential values, $\chi_E^+ \geq 0$, and χ_E^-; the fact that $\det(A_E) = 1$ implies that $\chi_E^- = -\chi_E^+ \leq 0$. Then either A_E is nonuniformly hyperbolic (if $\chi_E^+ > 0$), or the exponents of A_E vanish.

Thus for fixed α irrational, the spectrum Σ_α splits, from a dynamical point of view, into two (measurable) sets: the set of E for which A_E is nonununiformly hyperbolic, and the set of E for which the exponents of A_E vanish. On the other hand, spectral analysis gives us a different decomposition of the spectrum:

$$\sigma(H_x^\alpha) = \sigma_{ac}(H_x^\alpha) \cup \sigma_{sc}(H_x^\alpha) \cup \sigma_{pp}(H_x^\alpha)$$

where $\sigma_{ac}(H_x^\alpha)$ is the absolutely continuous spectrum, $\sigma_{pp}(H_x)$ is the pure point spectrum (i.e., the closure of the eigenvalues), and $\sigma_{sc}(H_x^\alpha)$ is the singular continuous spectrum. All three types of spectra have meaningful physical interpretations. While the spectrum $\sigma(H_x^\alpha)$ does not depend in x (since α is irrational), the decomposition into subspectra can depend on x.[8] It turns out that the absolutely continuous spectrum does not depend on x, so we can write $\Sigma_{ac,\alpha}$ for this common set.

The next deep relation between spectral theory and Lyapunov exponents is the following, which is due to Kotani:

Theorem 3.4 (Kotani, [Kot])**.** *Fix α irrational. Let $\mathcal{Z}$ be the set of E such that the Lyapunov exponents of A_E over f_α vanish. Let $\overline{\mathcal{Z}^{ess}}$ denote the essential closure of $\mathcal{Z}$, i.e. the closure of the Lebesgue density points of $\mathcal{Z}$. Then*

$$\Sigma_{ac} = \overline{\mathcal{Z}^{ess}}.$$

Thus Lyapunov exponents of the cocycle are closely related to the spectral type of the operators H_x. For instance, Theorem 3.4 implies that if A_E is nonuniformly hyperbolic over f_α for almost every $E \in \Sigma_\alpha$, then $\Sigma_{ac,\alpha}$ is empty: H_x^α has no absolutely continuous spectrum.

We remark that Theorems 3.3 and 3.4 hold for much broader classes of Schrödinger operators over ergodic measure preserving systems. For a short and self-contained proof of Theorem 3.3, see [Zh]. The spectral theory of one-dimensional Schrödinger operators is a rich subject, and we've only scratched the surface here; for further reading, see the recent surveys [JiM] and [D].

Avila's very recent work provides further fascinating connections of this type, linking analytic properties of Lyapunov exponents of general quasiperiodic operators with analytic potentials to the spectral decomposition.

[8] In fact, the decomposition is independent of a.e. x, just not all x.

4 To sum up

Let's pause for a minute to reflect.

As it turns out, all three sections are about families of dynamical systems. In Section 1, the family is the space of all volume preserving diffeomorphisms of a compact manifold M. This is an infinite dimensional, non-locally compact space, and we have thrown up our hands and depicted it in Figure 4 as a blob. Theorem 1.1 asserts that within the positive entropy systems (which turn out to be an open subset of the blob), measurable hyperbolicity (and ergodicity) is generic.

In Section 2, the moduli space $\mathcal{M}$ of directed flat surfaces can also be viewed as a space of dynamical systems, in particular the *billiard* flows on *rational* polygons, i.e., polygons whose corner angles are multiples of 2π. In a billiard system, one shoots a ball in a fixed direction and records the location of the bounces on the walls. By a process called unfolding, a billiard trajectory can be turned into a straight ray in a translation surface.[9] The process is illustrated here for the square torus billiard:

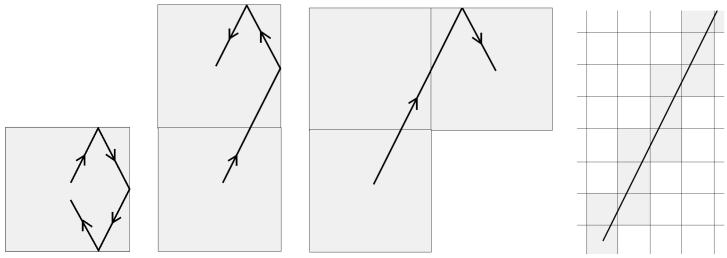


Figure 10: Unfolding billiards in a square to get lines in a torus (courtesy Diana Davis).

The moduli space $\mathcal{M}$ is not so easy to draw and not completely understood (except for $g = 1$). It is, however a manifold and carries some nice structures, which makes it easier to picture than $\mathrm{Diff}(M)$. Theorem 2.1 illustrates how dynamical properties of a *meta dynamical system*, i.e. the Teichmüller flow $\mathcal{F}_t \colon \mathcal{M} \to \mathcal{M}$ are tied to the dynamical properties of the elements of $\mathcal{M}$. For example, the Lyapunov exponents of the KZ cocycle

[9]Not every translation surface comes from a billiard, since the billiards have extra symmetries. But the space of billiards embeds inside the space of translation surfaces, and the Teichmüller flow preserves the set of billiards.

over $\mathcal{F}_t$ for a given billiard table with a given direction describe how well an infinite billiard ray can be approximated by closed, nearby billiard paths.

In Section 3, we saw how the spectral properties of a family of operators $\{H_x^\alpha : \alpha \in [0,1]\}$ are reflected in the dynamical properties of families of cocycles $\{(f_\alpha, A_E) : (E,\alpha) \in [-4,4]\times[0,1]\}$. Theorems about spectral properties thus have their dynamical counterparts. For example, Theorem 3.3 tells us that the butterfly is the complement of those parameter values where the cocycle (f_α, A_E) is uniformly hyperbolic. Since uniform hyperbolicity is an open property in both α and E, the complement of the butterfly is open. Corollary 3.2 tells us that the butterfly has measure 0. Thus the set of parameter values in the square that are hyperbolic form an open and dense, full-measure subset. In fact, work of Bourgain-Jitomirskaya [BoJi] implies that *the butterfly is precisely the set of parameter values* (E,α) *where the Lyapunov exponents of* (f_α, A_E) *vanish for some* x.[10] These results in some ways echo Theorem 1.1, within a very special family of dynamics.

The Hofstadter butterfly is just one instance of a low-dimensional family of dynamical systems containing very interesting dynamics and rich structure. A similar picture is seen in complex dynamics,[11] in the 1 (complex) parameter family of dynamical systems $\{p_c(z) = z^2 + c : c \in \mathbb{C}\}$. The Mandelbrot set consists of parameters c for which the map f_c has a connected Julia set J_c:

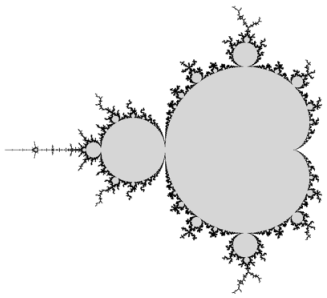


Figure 11: The Mandelbrot Set

It is conjectured that the set of parameters c such that p_c is uniformly hyperbolic on J_c is (open and) dense in the Mandelbrot set.

[10] which automatically means for all x in case of irrational α.

[11] Another field in which Avila has made significant contributions, which we have not touched upon here.

5 Themes

We end on a few themes that have come up in our discussion of exponents.

Nonvanishing exponents sometimes produce chaotic behavior. The bedrock result in this regard is Anosov's proof that smooth Anosov flows and diffeomorphisms are mixing (and in particular ergodic). Another notable result is Katok's proof [Ka] that measurable hyperbolicity of diffeomorphism produces exponential growth of periodic orbits.

Exponents can contain geometric information. We have not discussed it here, but there are delicate relationships between entropy, exponents and dimension [LY1, LY2].

Vanishing exponents sometimes present an exceptional situation that can be exploited. Both Furstenberg's theorem and Kotani theory are examples. Here's Furstenberg's criterion, presented in a special case:

Theorem 5.1 (Furstenberg, [F])**.** *Let $(A_1, \ldots, A_k) \subset SL(2, \mathbb{R})$, and let G be the smallest closed subgroup of $SL(2, \mathbb{R})$ containing $\{A_1, \ldots, A_k\}$. Assume that:*

1. *G is not compact.*
2. *There is no finite collection of lines $\emptyset \neq L \subset \mathbb{R}^2$ such that $M(L) = L$, for all $M \in G$.*

Then for any probability vector $p = (p_1, \ldots p_k)$ on $\{1, \ldots, k\}$ with $p_i > 0$, for all i, there exists $\chi^+(p) > 0$, such that for almost every $\omega \in \{1, \ldots, k\}^{\mathbb{N}}$ (with respect to the Bernoulli measure $p^{\mathbb{N}}$):

$$\lim_{n \to \infty} \frac{1}{n} \log \|A^{(n)}(\omega)\| = \chi^+.$$

One way to view what this result is saying is: if the exponent χ_+ vanishes, then the matrices either have a common eigenvector, or they generate a precompact group. Both possibilities are degenerate and are easily destroyed by perturbing the matrices. One proof of a generalization of this result [Le] exploits the connections between entropy, dimension and exponents alluded to before. This theorem can be used to prove our statement from the beginning about aspect ratios. See [McM] for details. See [BBC] for a related result.

Continuity and regularity of exponents is a delicate matter. There are still basic open questions here. Some of Avila's deepest results concern the dependence of Lyapunov exponents on parameters and dynamics, but this is the subject of a different talk.

Acknowledgments. Effusive thanks to Artur Avila, Svetlana Jitomirskaya, Curt McMullen, Zhenghe Zhang, and Anton Zorich for patiently explaining a lot of math to me, to Clark Butler and Jinxin Xue for catching many errors, and to Diana Davis, Carlos Matheus and Marcelo Viana for generously sharing their images.

References

[An1] D. Anosov, Geodesic flows on closed Riemannian manifolds of negative curvature. *Trudy Mat. Inst. Steklov.* **90** (1967).

[AA] Arnold, V. I.; Avez, A., *Ergodic problems of classical mechanics.* Translated from the French by A. Avez. W. A. Benjamin, Inc., New York-Amsterdam 1968.

[Av1] A. Avila, Global theory of one-frequency Schrödinger operators, *Acta Math.* **215** (2015), 1–54.

[Av2] A. Avila. KAM, Lyapunov exponents and spectral dichotomy for one-frequency Schrödinger operators. In preparation.

[Av3] A. Avila, On the regularization of conservative maps. *Acta Math.* **205** (2010), 5–18.

[AB] A. Avila, J. Bochi, Nonuniform hyperbolicity, global dominated splittings and generic properties of volume-preserving diffeomorphisms. *Trans. Amer. Math. Soc.* **364** (2012), no. 6, 2883–2907.

[ACW] A. Avila, S. Crovisier, A. Wilkinson, Diffeomorphisms with positive metric entropy, preprint.

[AJ] A. Avila, S. Jitomirskaya, The Ten Martini Problem, *Ann. Math.* **170** (2009), 303-342.

[AK] Avila, A.; Krikorian, R. Reducibility or non-uniform hyperbolicity for quasiperiodic Schrø"dinger cocycles. *Annals of Mathematics* **164** (2006), 911-940.

[AV1] Avila, Artur; Viana, Marcelo, Simplicity of Lyapunov spectra: a sufficient criterion. *Port. Math.* **64** (2007), 311-376.

[AV2] Avila, Artur; Viana, Marcelo, Simplicity of Lyapunov spectra: proof of the Zorich-Kontsevich conjecture. *Acta Math.* **198** (2007), no. 1, 1-56.

[BBC] I. Bárány, A. F. Beardon, and T. K. Carne. Barycentric subdivision of triangles and semigroups of Möbius maps. *Mathematika* **43** (1996), 165–171.

[Boc1] J. Bochi, Genericity of zero Lyapunov exponents. *Ergodic Theory Dynam. Systems* **22** (2002), no. 6, 1667–1696.

[Boc2] J. Bochi, C^1-generic symplectic diffeomorphisms: partial hyperbolicity and zero centre Lyapunov exponents. *J. Inst. Math. Jussieu* **9** (2010), no. 1, 49–93.

[BV1] J. Bochi, M. Viana, Lyapunov exponents: how frequently are dynamical systems hyperbolic? *Modern dynamical systems and applications* 271–297, Cambridge Univ. Press, Cambridge, 2004.

[BV2] J. Bochi, M. Viana, The Lyapunov exponents of generic volume-preserving and symplectic maps. *Ann. of Math.* **161** (2005), no. 3, 1423–1485.

[BC] C. Bonatti, S. Crovisier, Récurrence et généricité. *Invent. Math.* **158** (2004), 33–104.

[BGV] C. Bonatti, X. Gómez-Mont, M. Viana, Généricité d'exposants de Lyapunov non-nuls pour des produits déterministes de matrices. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **20** (2003), no. 4, 579–624.

[BoVi] C. Bonatti and M. Viana. Lyapunov exponents with multiplicity 1 for deterministic products of matrices. *Ergod. Th. & Dynam. Sys*, 24:1295–1330, 2004.

[BoGo] J. Bourgain and M. Goldstein, On nonperturbative localization with quasi-periodic potential, *Ann. of Math.* **152** (2000), 835-879.

[BoJi] Bourgain, J.; Jitomirskaya, S., Continuity of the Lyapunov exponent for quasiperiodic operators with analytic potential. *J. Statist. Phys.* **108** (2002), 1203-1218.

[CE] Chaika, Jon; Eskin, Alex, Every flat surface is Birkhoff and Oseledets generic in almost every direction. *J. Mod. Dyn.* **9** (2015), 123.

[D] D. Damanik, "Schrödinger Operators with Dynamically Defined Potentials: A Survey," to appear: *Erg. Th. Dyn. Syst.*

[EKM] Eskin, Alex; Kontsevich, Maxim; Zorich, Anton, Sum of Lyapunov exponents of the Hodge bundle with respect to the Teichmller geodesic flow. *Publ. Math. Inst. Hautes Études Sci.* **120** (2014), 207333.

[Fo] Forni, Giovanni, Deviation of ergodic averages for area-preserving flows on surfaces of higher genus. *Ann. of Math.* **155** (2002), 1-103.

[FK] Furstenberg, H., and H. Kesten, Products of random matrices. *Ann. Math. Statist.* **31** (1960) 457-469.

[F] Furstenberg, Harry, Noncommuting random products. *Trans. Amer. Math. Soc.* **108** (1963) 377428.

[GM] Goldsheĭd, I. Ya.; Margulis, G. A., Lyapunov exponents of a product of random matrices. (Russian) *Uspekhi Mat. Nauk 44* (1989), no. 5(269), 13–60; translation in Russian Math. Surveys 44 (1989), no. 5, 1171.

[GR] Guivarc'h, Yves ; Raugi, Albert. Propriétés de contraction d'un semi-groupe de matrices inversibles. Coefficients de Liapunoff d'un produit de matrices aléatoires indépendantes. *Israel J. Math.* **65** (1989), no. 2, 165196.

[He] Helson, Henry, Harmonic analysis. Second edition. Texts and Readings in Mathematics, **7**. Hindustan Book Agency, New Delhi, 2010.

[Ho] Hofstadter, Douglas, Energy levels and wavefunctions of Bloch electrons in rational and irrational magnetic fields. *Physical Review B* **14** (1976) 2239-2249.

[J] Johnson, R., Exponential dichotomy, rotation number, and linear differential operators with bounded coefficients, *J. Differential Equations* **61** (1986), 54–78.

[JiM] S. Jitomirskaya, C. A. Marx, "Dynamics and spectral theory of quasi-periodic Schrödinger-type operators," to appear: *Erg. Th. Dyn. Syst.*

[Ka] Katok, A., Lyapunov exponents, entropy and periodic orbits for diffeomorphisms. *Inst. Hautes Études Sci. Publ. Math.* **51** (1980), 137–173.

[KMS] Kerckhoff, Steven; Masur, Howard; Smillie, John, Ergodicity of billiard flows and quadratic differentials. *Ann. of Math.* **124** (1986), 293-311.

[Ko] A. Kolmogorov, Théorie générale des systèmes dynamiques et mécanique classique. *Proceedings of the International Congress of Mathematicians* (Amsterdam 1954) Vol. 1, 315–333

[Kot] S. Kotani, Ljapunov indices determine absolutely continuous spectra of stationary random one-dimensional Schrödinger operators, *Stochastic Analysis*(Katata/Kyoto, 1982), (North-Holland Math. Library 32, North-Holland, Amsterdam)(1984), 225-247.

[La] Y. Last, Zero measure spectrum for the almost Mathieu operator, *Comm. Math Phys.* **164**, 421432 (1994).

[Le] F. Ledrappier. Positivity of the exponent for stationary sequences of matrices. In *Lyapunov exponents (Bremen, 1984)*, volume 1186 of *Lect. Notes Math.*, pages 56–73. Springer-Verlag, 1986.

[LY1] F. Ledrappier, L.S. Young, The metric entropy of diffeomorphisms. I. Characterization of measures satisfying Pesin's entropy formula. *Ann. of Math.* **122** (1985), 509–539.

[LY2] F. Ledrappier, L.S. Young, The Metric Entropy of Diffeomorphisms. II. Relations between Entropy, Exponents and Dimension. *Ann. of Math.* **122** (1985), 540–574.

[Ma] R. Mañé, Oseledec's theorem from the generic viewpoint. *Proc. Int. Congress of Mathematicians* (Warszawa 1983) Vol. 2, 1259-76.

[McM] C. McMullen, "Barycentric subdivision, martingales and hyperbolic geometry", Preprint, 2011.

[Os] Oseledec, V. I., A multiplicative ergodic theorem. Characteristic Ljapunov, exponents of dynamical systems. (Russian) *Trudy Moskov. Mat. Obšč.* **19** (1968) 179-210.

[OU] J. Oxtoby, S. Ulam, Measure-preserving homeomorphisms and metrical transitivity. *Ann. of Math.* (2) **42** (1941), 874–920.

[P] Y. Pesin, Characteristic Ljapunov exponents, and smooth ergodic theory. *Uspehi Mat. Nauk* **32** (1977), no. 4 (196), 55-112, 287.

[R] M.A. Rodriguez-Hertz, Genericity of nonuniform hyperbolicity in dimension 3. *J. Mod. Dyn.* **6** (2012), no. 1, 121–138.

[SW] M. Shub, A. Wilkinson, Pathological foliations and removable zero exponents. *Invent. Math.* **139** (2000), no. 3, 495–508.

[Ve] W. A. Veech, The Teichmller geodesic flow, *Ann. of Math.* **124** (1986), 441530.

[Vi] M. Viana. Almost all cocycles over any hyperbolic system have nonvanishing Lyapunov exponents. *Ann. of Math.*, 167:643–680, 2008.

[Zh] Z. Zhang, Resolvent set of Schrödinger operators and uniform hyperbolicity, arXiv:1305.4226v2(2013).

[Zo1] A. Zorich, *How do the leaves of a closed 1-form wind around a surface*, "Pseudoperiodic Topology", V. Arnold, M. Kontsevich, A. Zorich (eds.), Translations of the AMS, Ser.2, vol. **197**, AMS, Providence, RI (1999), 135–178.

[Zo2] A. Zorich, *Asymptotic Flag of an Orientable Measured Foliation on a Surface*, dans "Geometric Study of Foliations", World Scientific Pb. Co., (1994), 479-498.

CURRENT EVENTS BULLETIN

Previous speakers and titles

For PDF files of talks, and links to Bulletin of the AMS articles, see
http://www.ams.org/ams/current-events-bulletin.html.

January 12, 2015 (San Antonio, TX)

Jared S. Weinstein, Boston University
Exploring the Galois group of the rational numbers: Recent breakthroughs.

Andrea R. Nahmod, University of Massachusetts, Amherst
The nonlinear Schrödinger equation on tori: Integrating harmonic analysis, geometry, and probability.

Mina Aganagic, University of California, Berkeley
String theory and math: Why this marriage may last.

Alex Wright, Stanford University
From rational billiards to dynamics on moduli spaces.

January 17, 2014 (Baltimore, MD)

Daniel Rothman, Massachusetts Institute of Technology
Earth's Carbon Cycle: A Mathematical Perspective

Karen Vogtmann, Cornell University
The geometry of Outer space

Yakov Eliashberg, Stanford University
Recent advances in symplectic flexibility

Andrew Granville, Université de Montréal
Infinitely many pairs of primes differ by no more than 70 million (and the bound's getting smaller every day)

January 11, 2013 (San Diego, CA)

Wei Ho, Columbia University
How many rational points does a random curve have?

Sam Payne, Yale University
Topology of nonarchimedean analytic spaces

Mladen Bestvina, University of Utah
Geometric group theory and 3-manifolds hand in hand: the fulfillment of Thurston's vision for three-manifolds

Lauren Williams, University of California, Berkeley
Cluster algebras

January 6, 2012 (Boston, MA)

Jeffrey Brock, Brown University
Assembling surfaces from random pants: the surface-subgroup and Ehrenpreis conjectures

Daniel Freed, University of Texas at Austin
The cobordism hypothesis: quantum field theory + homotopy invariance = higher algebra

Gigliola Staffilani, Massachusetts Institute of Technology
Dispersive equations and their role beyond PDE

Umesh Vazirani, University of California, Berkeley
How does quantum mechanics scale?

January 6, 2011 (New Orleans, LA)

Luca Trevisan, Stanford University
Khot's unique games conjecture: its consequences and the evidence for and against it

Thomas Scanlon, University of California, Berkeley
Counting special points: logic, Diophantine geometry and transcendence theory

Ulrike Tillmann, Oxford University
Spaces of graphs and surfaces

David Nadler, Northwestern University
The geometric nature of the Fundamental Lemma

January 15, 2010 (San Francisco, CA)

Ben Green, University of Cambridge
Approximate groups and their applications: work of Bourgain, Gamburd, Helfgott and Sarnak

David Wagner, University of Waterloo
Multivariate stable polynomials: theory and applications

Laura DeMarco, University of Illinois at Chicago
The conformal geometry of billiards

Michael Hopkins, Harvard University
On the Kervaire Invariant Problem

January 7, 2009 (Washington, DC)

Matthew James Emerton, Northwestern University
Topology, representation theory and arithmetic: Three-manifolds and the Langlands program

Olga Holtz, University of California, Berkeley
Compressive sensing: A paradigm shift in signal processing

Michael Hutchings, University of California, Berkeley
From Seiberg-Witten theory to closed orbits of vector fields: Taubes's proof of the Weinstein conjecture

Frank Sottile, Texas A & M University
Frontiers of reality in Schubert calculus

January 8, 2008 (San Diego, California)

Günther Uhlmann, University of Washington
Invisibility

Antonella Grassi, University of Pennsylvania
Birational Geometry: Old and New

Gregory F. Lawler, University of Chicago
Conformal Invariance and 2-d Statistical Physics

Terence C. Tao, University of California, Los Angeles
Why are Solitons Stable?

January 7, 2007 (New Orleans, Louisiana)

Robert Ghrist, University of Illinois, Urbana-Champaign
Barcodes: The persistent topology of data

Akshay Venkatesh, Courant Institute, New York University
Flows on the space of lattices: work of Einsiedler, Katok and Lindenstrauss

Izabella Laba, University of British Columbia
From harmonic analysis to arithmetic combinatorics

Barry Mazur, Harvard University
The structure of error terms in number theory and an introduction to the Sato-Tate Conjecture

January 14, 2006 (San Antonio, Texas)

Lauren Ancel Myers, University of Texas at Austin
Contact network epidemiology: Bond percolation applied to infectious disease prediction and control

Kannan Soundararajan, University of Michigan, Ann Arbor
Small gaps between prime numbers

Madhu Sudan, MIT
Probabilistically checkable proofs

Martin Golubitsky, University of Houston
Symmetry in neuroscience

January 7, 2005 (Atlanta, Georgia)

Bryna Kra, Northwestern University
The Green-Tao Theorem on primes in arithmetic progression: A dynamical point of view

Robert McEliece, California Institute of Technology
Achieving the Shannon Limit: A progress report

Dusa McDuff, SUNY at Stony Brook
Floer theory and low dimensional topology

Jerrold Marsden, Shane Ross, California Institute of Technology
New methods in celestial mechanics and mission design

László Lovász, Microsoft Corporation
Graph minors and the proof of Wagner's Conjecture

January 9, 2004 (Phoenix, Arizona)

Margaret H. Wright, Courant Institute of Mathematical Sciences, New York University
The interior-point revolution in optimization: History, recent developments and lasting consequences

Thomas C. Hales, University of Pittsburgh
What is motivic integration?

Andrew Granville, Université de Montréal
It is easy to determine whether or not a given integer is prime

John W. Morgan, Columbia University
Perelman's recent work on the classification of 3-manifolds

January 17, 2003 (Baltimore, Maryland)

Michael J. Hopkins, MIT
Homotopy theory of schemes

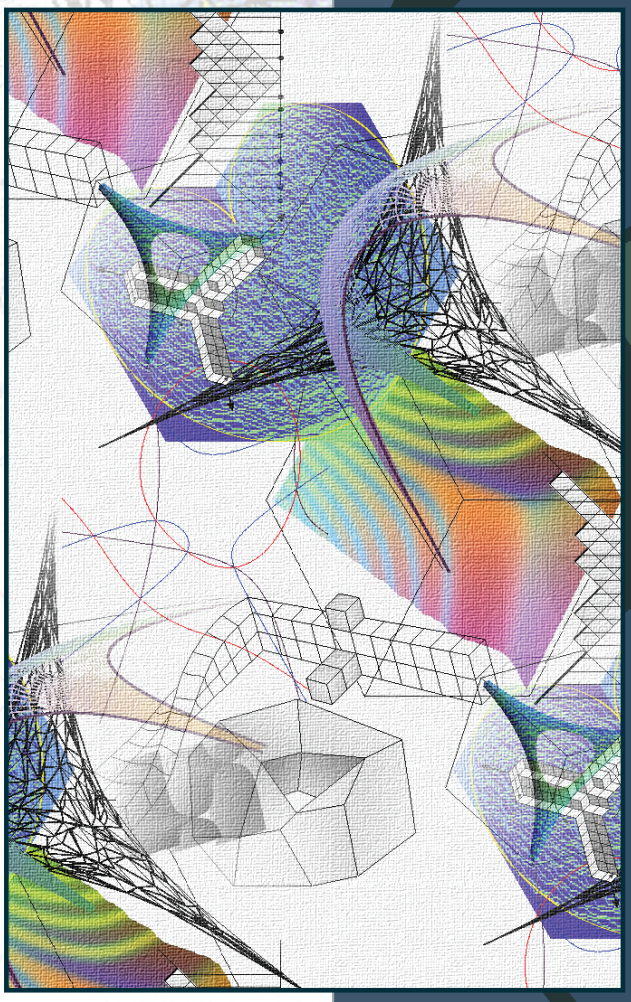
Ingrid Daubechies, Princeton University
Sublinear algorithms for sparse approximations with excellent odds

Edward Frenkel, University of California, Berkeley
Recent advances in the Langlands Program

Daniel Tataru, University of California, Berkeley
The wave maps equation

2016 CURRENT EVENTS BULLETIN

Committee



The back cover graphic is reprinted courtesy of Andrei Okounkov.

Cover graphic associated with Carina Curto's talk courtesy of Amanda Burnham.

Cover graphic associated with Amie Wilkinson's talk courtesy of Mike Field. Thorns, 1996. A symmetric chaotic attractor.

Cover graphic associated with Yuval Peres' talk courtesy of Itamar Landau and Lionel Levine.

Cover graphic associated with Timothy Gowers' talk courtesy of CMSC – University of Western Australia.