ICM 2018 LECTURE SAMPLER

From left to right: Gil Kalai, Sanjeev Arora, Catherine Goldstein, Sylvia Serfaty, Gregory F. Lawler.

In this sampler, the speakers below have kindly provided introductions to their Plenary Lectures for the International Congress of Mathematicians 2018 (ICM 2018) taking place August 1–9 in Rio de Janeiro.

ICM 2018
Rio de Janeiro | August 1–9

Three Puzzles on Mathematics, Computation, and Games
Gil Kalai
page 782

Toward Mathematical Understanding of Deep Learning
Sanjeev Arora
page 785

On the History of Mathematical Concepts
Catherine Goldstein
page 786

Systems with Coulomb Interaction
Sylvia Serfaty
page 787

Loops and Loop-Erased Random Walk
Gregory F. Lawler
page 788

Gil Kalai

Three Puzzles on Mathematics, Computation, and Games

The theory of computing and computer science as a whole are precious resources for mathematicians.

They bring new questions, new profound ideas, and new perspectives on classical mathematical objects, and serve as new areas for applications of mathematics and of mathematical reasoning.

In my lecture I will talk about three mathematical puzzles involving mathematics and computation (and, at times, other fields) that have preoccupied me over the years.

Puzzle 1: What is the Explanation for the Success of the Simplex Algorithm?

Linear programming is the problem of maximizing a linear function ϕ subject to a system of linear inequalities. The set of solutions for the linear inequalities is a convex polytope P (which can be unbounded). The simplex algorithm was developed by George Danzig. Geometrically it can be described by moving from one vertex to a neighboring vertex so as to improve the value of the objective function.

The simplex algorithm is one of the most successful mathematical algorithms. The explanation of this success is an applied, vaguely stated question, which is connected with computers. The problem has strong relations to the study of convex polytopes, which fascinated mathematicians from ancient times, and which served as a starting point for my own research.

Two important elements in studying the mathematics of linear programming are abstractions and reductions. Various abstract forms of linear programming, convex polytopes and other objects play an important role, and there are very clever and important reductions from one set of problems and objects to others.

If I needed to choose the single most important mathematical explanation for the success of the simplex algorithm, my choice would point to a theorem about another algorithm. I would choose Khachiyan's 1979 theorem asserting that there is a polynomial-time algorithm for linear programming. (Or briefly $LP \in \mathbf{P}$.) Khachiyan's theorem refers to the ellipsoid method, and the answer is given in the language of computational complexity, a language that was not at all available when the question was originally raised. A more recent explanation is the

Gil Kalai is professor of mathematics, Hebrew University of Jerusalem and Yale University. He is supported in part by ERC advanced grant 320924.

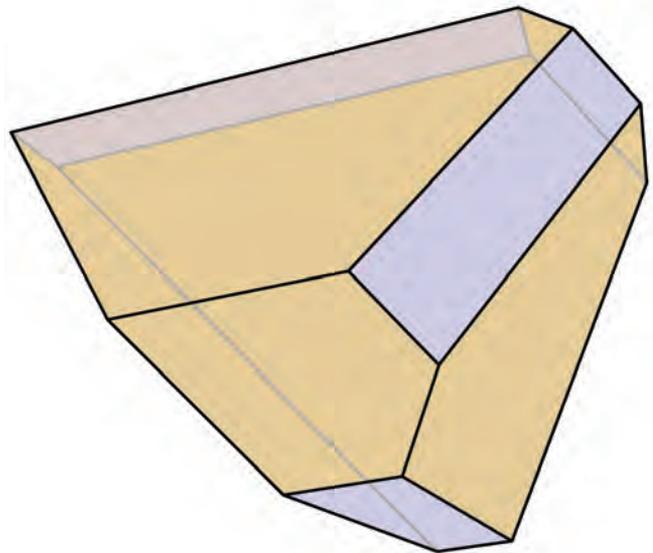*His email address is* kalai@math.huji.ac.il.

Figure 1. The associahedron (Stasheff polytope).

work of Spielman and Teng asserting that a certain pivot rule requires an expected polynomial number of steps for a random perturbation of every linear programming problem.

In the lecture I will concentrate on the study of diameter of graphs of polytopes and the discovery of randomized subexponential variants of the simplex algorithm, and mention recent advances: The disproof of the Hirsch conjecture by Santos and the connection between linear programming and stochastic games leading to subexponential lower bounds proved by Friedman, Hansen, and Zwick, for certain randomized pivot rules for the simplex algorithm.

Puzzle 2: What are Methods of Election that are Immune to Errors in the Counting of Votes?

The second puzzle can be seen in the context of understanding and planning of electoral methods. We all remember the sight of vote recounts in Florida in the 2000 US presidential election.

Is the American electoral system, based on electoral votes, inherently more susceptible to mistakes than the majority system? And what is the most stable method? Together with Itai Benjamini and Oded Schramm I investigated these and similar problems. We asked the following question: given that there are two candidates and each voter chooses at random and with equal probability (independently) between them, what is the stability of the outcome, when in the vote-counting process one percent of the votes is counted incorrectly? The mathematical jargon for these errors is "noise." We defined a measure

Figure 2.The US 2000 presidential elections raised questions about whether the Electoral College system is more susceptible to mistakes than popular vote.

of noise sensitivity of electoral methods and found that weighted majority methods are immune to noise: when the probability of error is small, the chances that the elections' outcome will be affected diminish. We also showed that every stable to noise method is close in some mathematical sense to a weighted majority method. In later work, O'Donnell, Oleszkiewicz, and Mossel showed that the majority system is most stable to noise among all non-dictatorial methods.

Our work was published in 1999, a year before the question appeared in the headlines in the US presidential election, and it did not even deal with the subject of elections.[1] We were interested in understanding the problem of planar percolation, a mathematical model derived from statistical physics. In our article we showed that if we adopt an electoral system based on the model of percolation, this method will be very sensitive to noise. This insight is of no use at all in planning good electoral methods, but it makes it possible to understand interesting phenomena in the study of percolation.

After the US presidential election in 2000 we tried to understand the relevance of our model and the concepts of stability and noise in real-life elections: is the measure for noise stability that we proposed relevant, even though

the basic assumption that each voter randomly votes with equal probability for one of the candidates is far from realistic? The attempt to link mathematical models to questions about elections (and, more generally, to social science) is fascinating and complicated, and a true pioneer in this study was the Marquis de Condorcet, a mathematician and philosopher, a democrat, a human rights advocate, and a feminist who lived in France in the eighteenth century. One of Condorcet's findings, often referred to as Condorcet's paradox, is that when there are three candidates, the majority rule can sometimes lead to cyclic outcomes, and it turns out that the probability for cyclic outcomes depends on the stability to noise of the voting system. We will discuss the connection of noise stability to computational complexity and mainly to the important theory of hardness of approximations and PCP.

Puzzle 3: Are Quantum Computers Possible?[2]

A quantum computer is a hypothetical physical device that exploits quantum phenomena such as interference and entanglement in order to enhance computing power. The study of quantum computation combines fascinating physics, mathematics, and computer science. In the 1990s, Peter Shor discovered that quantum computers would make it possible to perform certain computational tasks hundreds of orders of magnitude faster than ordinary computers and, in particular, would break most of today's encryption methods. At that time, the first doubts about the model were raised: quantum systems are of a noisy and unstable nature. Peter Shor himself found a key to a possible solution to the problem of noise: quantum error-correcting codes and quantum fault-tolerance.

In the mid-1990s, three groups of researchers studied the noisy quantum computer model and showed that noisy quantum computers still make it possible to perform all miracles of universal quantum computing, as long as engineers succeeded in lowering the noise level below a certain threshold.

One possibility, which reflects widespread opinion, is that the construction of quantum computers is possible, that the remaining challenge is essentially of an engineering nature, and that such computers will be built in the coming decades. Moreover, people expect to build in the next few years quantum codes of the quality required for quantum fault-tolerance, and to demonstrate the concept of "quantum computational supremacy" on quantum computers with fifty qubits. A second possibility, which expresses my position, is that it will not be possible to construct quantum codes that are required for quantum computation, nor will it be possible to demonstrate quantum computational superiority in other quantum systems. Let me explain why.

My analysis is based on the same model of noise that led researchers in the 1990s to optimism about quantum computation, and it points to the need for different analyses

---

[1]*Here one can witness a secret of our trade. Mathematicians recycle their models, and the same model can be used for very different purposes.*

[2]*See Kalai's article on "The Quantum Computer Puzzle" in the May 2016 Notices.*
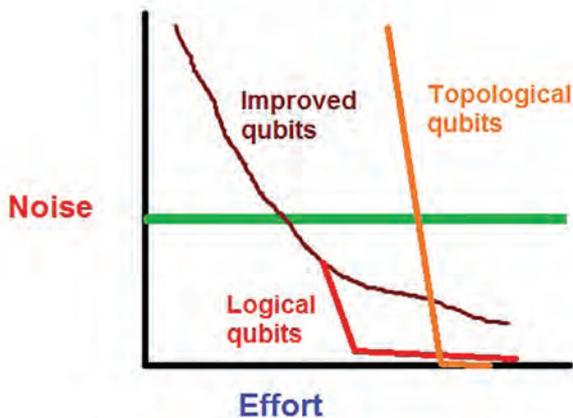
Figure 3. It is commonly believed that by putting more effort into creating qubits the noise level can be pushed down to as close to zero as we want. Once the noise level is small enough and crosses the green threshold line, quantum error correction allows logical qubits to reduce the noise even further with a small amount of additional effort. Very high quality topological qubits are also expected.
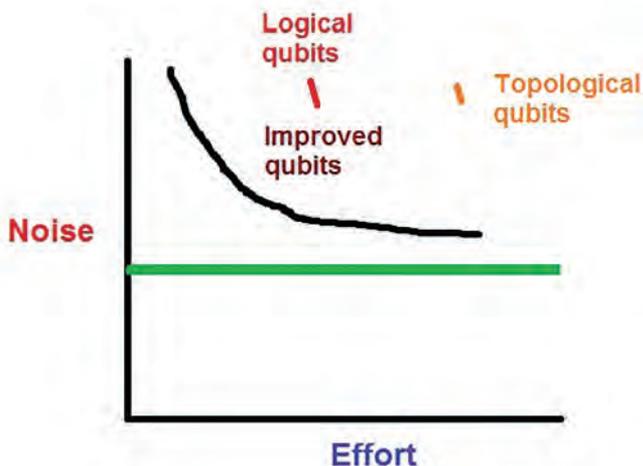

Figure 4. My analysis gives good reasons to expect that we will not be able to reach the green line and that all attempts for good quality logical and topological qubits will fail.

on different scales. My analysis shows that noisy quantum computers on the small scale (a few dozen qubits) express such a primitive computational power that it will not allow the creation of quantum codes that are required as building blocks for quantum computers on a higher scale. A reader may ask: how does the noisy quantum world allow classical information and classical computation? The difference between classical and quantum

The connection and tension between the pure and the applied, between models and reality, and between foundations and engineering.

information is related to puzzle 2. Encoding using repetitions and decoding using the majority rule (or related techniques) enable classical information and calculation: the majority method corrects errors and is stable to noise, and therefore can be implemented in a very primitive computational system. This enables, on larger scales, classical information and computation.

Conclusion

The lecture deals with three fascinating puzzles on mathematics and computation telling a story of pure and applied mathematics, theoretical computer science, physics and social sciences, as well as games of various kinds.

All our three puzzles involve the study of beautiful underlying mathematical objects, and deep mathematical methods, and also demonstrate the connection and tension between the pure and the applied, between models and reality, and between foundations and engineering.

Image Credits

Figure 1 by Bill Casselman.
Figure 2 by Neta Kalai.
Figures 3 and 4 by Gil Kalai.
Author photo courtesy of Matas Šileikis.

Gil Kalai

ABOUT THE AUTHOR

Many members of **Gil Kalai's** family are artists, and sometimes he thinks about mathematics as a form of art.

Sanjeev Arora

Toward Mathematical Understanding of Deep Learning

Deep learning, the modern name for learning using neural networks, has in recent years become the dominant paradigm of machine learning. Performance on tasks such as image recognition and game-playing now exceeds human performance.
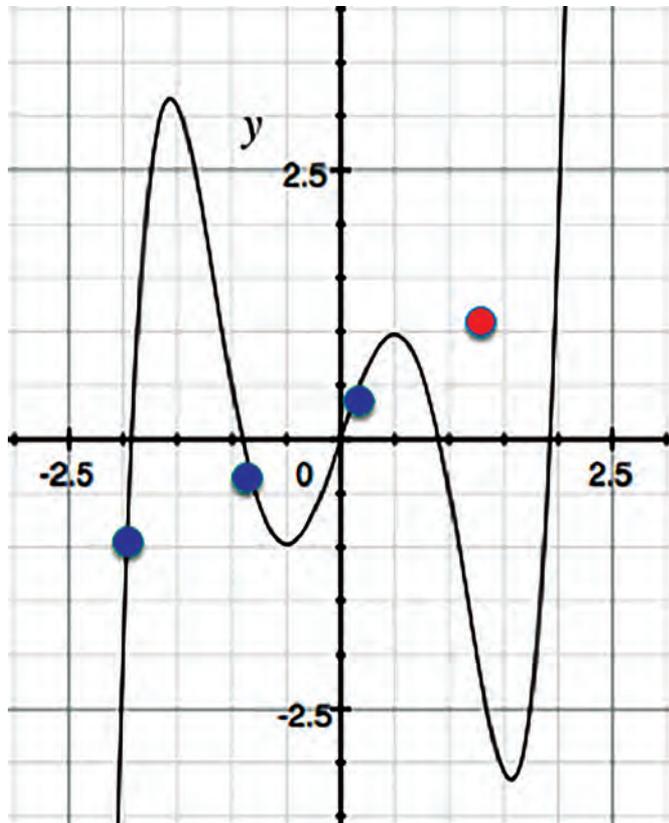

Figure 1. If one tries to fit a degree 5 polynomial to three blue data points, then there is no reason such a fitted polynomial will predict the value of some previously unseen fourth red point. This classic intuition suggests that deep nets with too many parameters would overfit to the data, but in practice they do not.

At a basic level deep learning—like many other machine learning algorithms—can be seen as fitting a model to a data. This is quite analogous to classical statistical techniques such as regression, but with far more trainable parameters and more complex

Sanjeev Arora is Charles Fitzmorris Professor of computer science at Princeton University and visiting professor in mathematics at the Institute for Advanced Study. His email address is arora@princeton.edu.

For permission to reprint this article, please contact: reprint-permission@ams.org.

DOI: http://dx.doi.org/10.1090/noti1698

optimization. For example, vision tasks may involve models with tens of millions of parameters and nonconvex optimization objective.

Mathematical understanding of this field is still in its infancy, and my talk will be a survey of the nascent efforts to develop such understanding. Some of the new mathematical results concern properties of optimization algorithms: when/how do they find reasonable solutions? Other results concern generalization: why do the trained nets perform well on unseen data? This is a nontrivial question because usually the net has far more parameters than the number of training examples. Remember the overfitting issue in modeling, as in Figure 1. If one tries to fit a degree 5 polynomial to three data points, then there is no reason such a fitted polynomial will predict the value of some previously unseen fourth point. Thus clearly the optimization algorithm somehow implicitly cuts down the effective number of parameters in the net, but it is unclear how this happens. Recent results begin to cast some light on this mystery.

Next, there is the question of what kinds of functionalities can or cannot be expressed by deep nets of a certain size and depth. This includes the ability of deep nets to model interesting distributions of images, using ideas such as dueling deep nets (GANs).

The talk will be largely self-contained. I have also prepared a short article for the ICM proceedings, which surveys the main conceptual frameworks of machine learning. I will also put that article on arXiv and on my webpage.

Image Credits

Figure 1 by Editors.
Author photo courtesy of Sanjeev Arora.

ABOUT THE AUTHOR

Sanjeev Arora has lived in Princeton, New Jersey, for twenty-four years. His interests outside work include listening to classical music and jazz, cooking, travel, and photography.

Sanjeev Arora

Catherine Goldstein

On the History of Mathematical Concepts

A key issue in my field, history of mathematics, is to describe and understand the development of mathematics. The focus might be on the shaping of concepts, of results, of applications, but also on ways and practices of thinking.
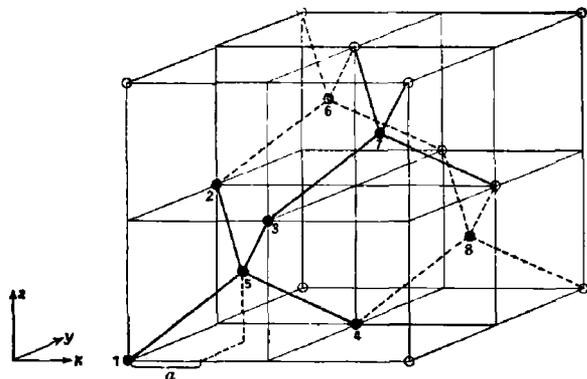

Fig. 1.

Figure 1. Hermitian forms appear in a 1914 paper by Max Born on the elasticity theory of the diamond, long before their entry with the same person into quantum theory.

One main advance of the last decades is a much more acute awareness of the richness of mathematical activities and of the various layers of these activities. In a mathematical result many aspects interfere. Who were the mathematicians involved, perhaps academic mathematicians, or engineers, or scribes, or lawyers, or even amateurs? What was their training? What were their sources? What were their mathematical priorities, whether specific applications or effective results or englobing conceptualizations? What were their arguments and proofs? What were their ways of transmitting their findings? Recent focus has been on collective issues such as the history of mathematical journals or societies, but also on the ways figures and formulas intervene in mathematical texts and how they help or hamper the creation and diffusion of mathematical ideas. Such a focus favors interactions between historians of mathematics and general historians, sociologists, linguists, and other specialists of the human sciences. The understanding of the cultural variety encapsulated in a single mathematical text or concept has just begun to be explored.

On the other hand, historians of mathematics have been much more reflexive about our own terminology and arguments. What do we mean precisely by such things as a "discipline" or a "mathematical school" or even the "reception" of a result or domain? This is particularly important if we want to write a long-term history of the development of a concept or a theorem or a method, using our new understanding of how a piece of mathematics, born in a particular environment, has been partially taken over, partially rediscovered, or contextualized afresh. The textual, technical, cultural, social, economical, and logical strata all have and their different scales and rhythms of change to be taken into account if we want to understand how mathematical constructs evolve.

In my ICM talk, I intend to explain and illustrate this concretely, using the simple, but historically intriguing example of Hermitian forms, arising from the classical arithmetical question concerning the decomposition of any integer as a sum of four squares and becoming within a few decades a common tool in geometry, algebra, and physics.

Image Credits

Figure 1 courtesy of M. Born, Zur Raumgittertheorie des Diamanten, *Annalen der Physik*, vol. 349, Issue 12, pp. 605–642. Copyright ©1914 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

Author photo courtesy of Vincent Hoel, Paris Photo Studio.

Catherine Goldstein

ABOUT THE AUTHOR

Catherine Goldstein is particularly interested in the history of number theory, in women in mathematics, and in the role of observation for mathematicians. She has recently coedited a book on mathematics and WWI (*A War of Guns and Mathematics*, AMS, 2014) and another on the history of combinatorics and Delannoy numbers (*Les travaux combinatoires en France (1870–1914) et leur actualité*, PULIM, 2017).

Catherine Goldstein is director of research at the CNRS, Institut de mathématiques de Jussieu-Paris Gauche, Sorbonne Université, Univ Paris Diderot. Her email address is `catherine.goldstein@imj-prg.fr`*.*

Sylvia Serfaty

Systems with Coulomb Interaction

What do vortices in superconductors, eigenvalues of random matrices, and optimal interpolation points on spheres have in common? They all consist of points that repel one another via a logarithmic interaction, i.e. have an energy of the form

$$(0.1) \qquad \mathcal{H}_N(x_1, \ldots, x_N) = \sum_{1 \leq i \neq j \leq N} \mathsf{g}(x_i - x_j),$$

where $\mathsf{g}(x) = -\log|x|$. It is natural to generalize this to the Coulomb interaction $\mathsf{g}(x) = |x|^{2-d}$ in dimension $d \geq 3$ and to add a possible confinement energy $\sum_{i=1}^N V(x_i)$, where V grows sufficiently fast at infinity. One is interested in understanding the large N behavior of such systems. This can be in the setting of minimizers of such energies, of the time-evolutions of such systems, but also of states with temperature. When the temperature is $\frac{1}{\beta}$, the probability density of observing the system in the configuration $(x_1, \ldots, x_N)$ is given by the Gibbs measure

$$(0.2) \qquad d\{_{N,\beta}(x_1, \ldots, x_N) = \frac{1}{Z_{N,\beta}} e^{-\beta \mathcal{H}_N(x_1, \ldots, x_N)} dx_1 \ldots dx_N,$$

where $Z_{N,\beta}$ is a normalizing constant to make the total probability 1. Motivation comes for instance from statistical mechanics, random matrix theory, quantum mechanics, condensed matter physics, fluid mechanics, approximation theory, and even geometry.

Simulations of realizations of (0.2) for the logarithmic interaction in two dimensions with a quadratic confining potential are shown in Figure 1.
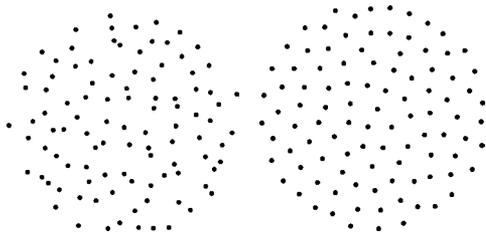
Figure 1. Configurations of $N = 100$ points tend to be less ordered at higher temperatures ($1/\beta = 1/5$ on the left) than at lower temperatures ($1/\beta = 1/400$ on the right). Simulation with $V(x) = |x|^2$ by Thomas Leblé.

They seem to indicate that the macroscopic distribution of the particles does not depend much on the temperature, while the microscopic patterns strongly depend on it. On the left, the temperature is larger and the system seems quite disordered, while on the right, the temperature is much smaller and the system seems more ordered at the microscopic level. Can this be proven rigorously? Is there a phase transition happening between a "liquid state" and a "solid state"? Is there a limiting pattern that the

Sylvia Serfaty is Silver Professor of Mathematics at the Courant Institute of Mathematical Sciences. Her email address is `serfaty@cims.nyu.edu`.
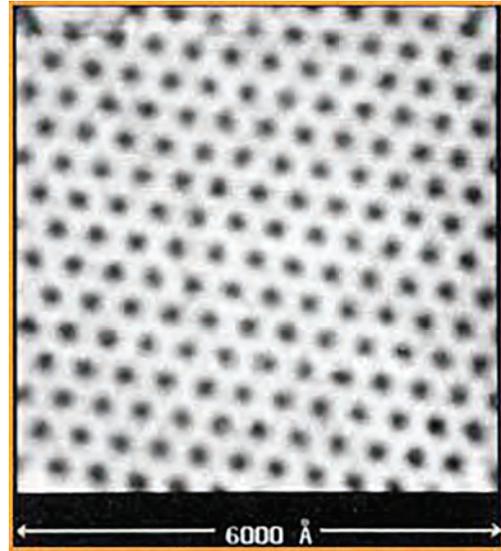
Figure 2. Vortices in superconductors form perfect triangular Abrikosov lattices, which are also optimal packings.

points want to form as temperature goes to 0? Is that the same as the patterns that energy minimizers would form?

A hint is provided by the behavior of vortices in superconductors, which can be directly connected to minimizers of (0.1): they form perfect triangular lattices, called Abrikosov lattices in physics, as in Figure 2. How can one mathematically explain the formation of these triangular lattices? The triangular lattice is also the solution of the best-packing problem in two dimensions and of other minimization questions related to modular functions. A 2007 conjecture of H. Cohn and A. Kumar [3] says that the triangular lattice should be *universally minimizing* in the sense that it should minimize all interactions with kernels which are *completely monotone* (i.e. satisfy $(-1)^k \mathsf{g}^{(k)} \geq 0$), indicating that this is part of a broader phenomenon. One can also wonder about higher dimensions: what would play the analogous role of the triangular lattice? Many of these questions are actually still open, but certain qualitative properties of these large N limit configurations can be rigorously established.

[3]*H. Cohn, A. Kumar, Universally optimal distribution of points on spheres.* J. Amer. Math. Soc. **20** *(2007), no. 1, 99–148, Conjecture 9.4.*

Image Credits

ABOUT THE AUTHOR

Sylvia Serfaty is interested in developing analysis tools to understand problems from physics. She has extensively studied the Ginzburg–Landau model of superconductivity and this has led her to questions on the statistical mechanics of Coulomb-type systems. She has been splitting her career between Paris and New York.

Sylvia Serfaty

Gregory F. Lawler

Loops and Loop-Erased Random Walk

Random fractal structures arise in statistical mechanical models viewed at or near a phase transition. A challenge for mathematics is to describe the random curves $\gamma : [0, \infty) \to \mathbb{R}^d$ that arise in the continuum limit. Usual calculus defines smooth curves in terms of derivatives. At the other extreme, "(completely) random continuous" motion is modeled by random walk or Brownian motion, as in Figure 1. One can combine completely random and smooth motions using stochastic calculus. The curves we are studying are more complicated than these because of their strong interaction with the past. For example, many of the curves are forbidden to visit any place that has already been visited.
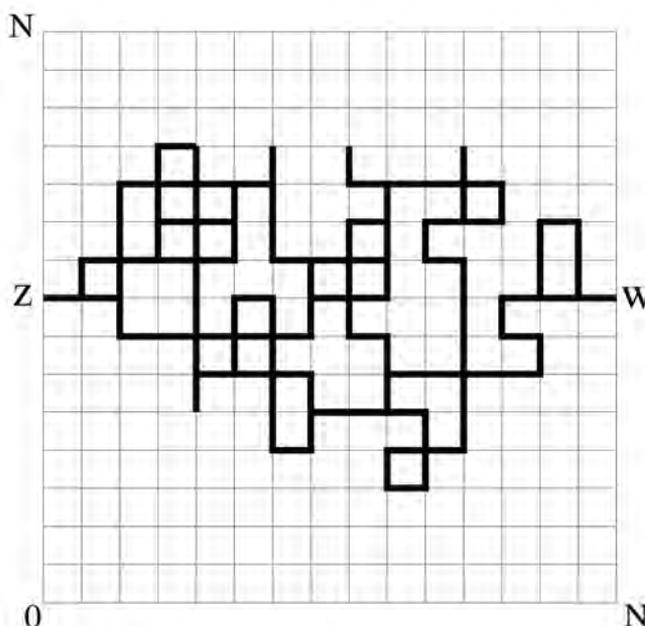

Figure 1. A random walk from z to w in the square typically has self-intersections and loops.

To try to define such curves, one often discretizes by giving a precise definition on each discrete scale with the hope of being able to take a limit. For curves with self repulsion, the most interesting dimensions are $d = 2, 3$. For $d = 1$ "self-avoidance" means that curves can only go in a single direction, and for $d \geq 4$ it turns out that completely random paths do not want to have self-intersections. Roughly speaking, this is because the path of a random walk has "fractal dimension two." This follows from the central limit theorem, the most fundamental fact in probability.

There have been many major advances in the last twenty years towards the rigorous understanding of two-dimensional critical

Gregory F. Lawler is the George Wells Beadle Distinguished Service Professor in mathematics and statistics at the University of Chicago. His email address is lawler@math.uchicago.edu.
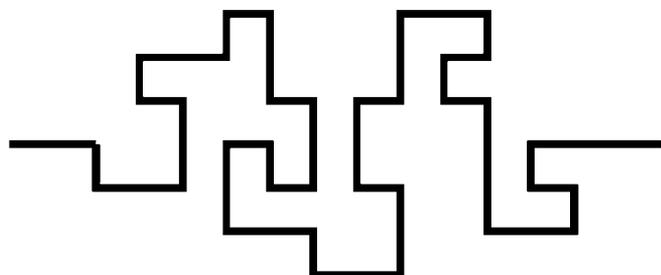
Figure 2. The loop-erased random walk (LERW) obtained from Figure 1 by erasing loops chronologically.

systems, taking advantage of a nonrigorous prediction from theoretical physics that such systems would exhibit conformal invariance in the limit. The situation in three dimensions is much more open.

I will focus on a particular example, the *loop-erased random walk* (LERW). One obtains the process by starting with a usual random walk as in Figure 1 and erasing the loops in a chronological order. Such paths also arise as paths of a spanning tree of the lattice chosen at random. Figures 1 and 2 show a two-dimensional simple random walk and the LERW obtained after erasing the loops.

How does one construct a limit path from the LERW ω? We start by determining the fractal dimension $\alpha > 1$ which roughly states that if the random walk is of diameter N then the number of steps of the LERW is of order N^{α}. We then set $\gamma^{(N)}(t) = N^{-1}\omega(tN^{\alpha})$, and hope to obtain a limit. Or in the language of nonstandard analysis one can take N^{-1} to be infinitesimal and consider the macroscopic behavior of the paths. The scaling rule is key: the increment in space is the increment in time raised to the reciprocal of the fractal dimension. For $d = 2$, F. Viklund and I [2] completed the final part of this program for LERW. The scaling exponent $\alpha = 5/4$ had been predicted by theoretical physics and was first established rigorously by Kenyon in 2000. Convergence to a limit process, which is now called the *Schramm-Loewner evolution (SLE)* with parameter 2, was proved by O. Schramm, W. Werner, and me in 2004, but that was for a curve with a singular parametrization.

Conformal invariance is a key aspect of the analysis. The correct parametrization of the continuous curve is by (5/4)-Minkowski content, a fractal analogue of arc length. At the moment, LERW is the only process for which the complete result is known but there are many other processes for which we understand the "up to parametrization" limit.

The fractal curves can be considered as an observable of a more complicated random field. Much of the most exciting work
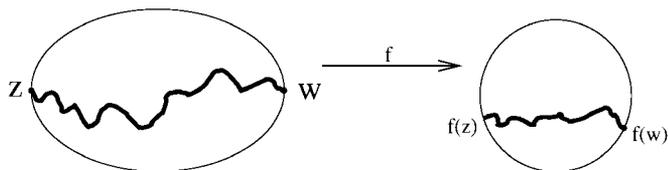
recently has taken this perspective. For our example, we can consider not only the path obtained by erasing loops but also the "soup" of loops that have been erased in forming the path. These loops have a continuum limit known as the *Brownian loop soup*; it exists in all dimensions, but has been particularly useful in two dimensions, where it exhibits conformal invariance. It is a Poisson point process of loops. At time t, the intensity of loops is t. The parameter $\mathbf{c} := -2t$ is the same as an algebraic quantity the *central charge* of conformal field theory. The loop-erased walk and the uniform spanning tree have $\mathbf{c} = -2$ and correspond to intensity 1. Different central charges and intensities correspond to models in different *universality classes*.

Suppose we restrict to an $N \times N$ lattice square and view the loops inside the square. For each vertex we can consider n_x the number of times the vertex has been visited. Then $n_x + 1$ is (very close to) the norm squared of a *complex Gaussian free field* [3]. In other words, it looks like $|Z_x|^2$ where $Z_x = (X_x + iY_x)/\sqrt{2}$ and X, Y are independent centered joint normal random variables with $\mathbb{E}[X_x X_y] = \mathbb{E}[Y_x Y_y] = G(x, y)$. Here $G(x, y)$ is the expected number of visits to y by a random walk starting at x before it leaves the square. This Gaussian free field has become one of the most important building blocks for constructing conformally invariant models. One particular construction goes under the name *(Liouville) quantum gravity* [1]. This is a random planar measure that assigns measure proportional to $|e^{\lambda Z}|$ for some λ to each site.

While I will be discussing some of the exciting work that has been done in two dimensions, I will also emphasize the important and harder case of three dimensions, a challenge for future researchers. Here the powerful techniques of conformal invariance are not available, but perhaps some ideas will carry through. The first big challenge is to find a nice way to describe continuous random fractal curves with self-repulsion.

References

[1] B. DUPLANTIER, J. MILLER, and S. SHEFFIELD, Liouville quantum gravity as a mating of trees, arXiv:1409.7055. MR3729063

[2] G. LAWLER and F. VIKLUND, Convergence of loop-erased random walk in the natural parametrization, arXiv:1603:05203.

[3] Y. LE JAN, Markov Paths, Loops, and Fields, *Lecture Notes in Mathematics*, (2011), **2026**, Springer-Verlag. MR2815763

Image Credits

Article figures by Greg Lawler.

Author photo courtesy of University of Chicago.

Figure 3. We consider measures on curves that are invariant under conformal transformations.

ABOUT THE AUTHOR

When not teaching or pursuing his research in random walks with strong interactions, **Greg Lawler** plays guitar and recently has been advising the tournament bridge world on the use of statistics to determine cheating.

Greg Lawler