AMERICAN MATHEMATICAL SOCIETY

Current Events Bulletin

Friday, January 5, 2024
2:00 p.m.–6:00 p.m.

Moscone North/South, Moscone Room 205 | Joint Mathematics Meetings, San Francisco, CA

2:00 p.m. | **Will Perkins**
Georgia Tech

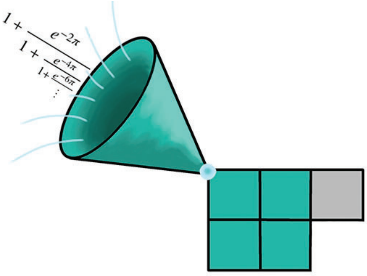
Searching for (sharp) thresholds in random structures: where are we now?

Phase transitions, hard computational problems, and the emergence of intricate structures in random graphs—how are these phenomena connected and how can we understand them?

3:00 p.m. | **Hussein Mourtada**
Université Paris Cité

Hilbert meets Ramanujan: singularity theory and integer partitions

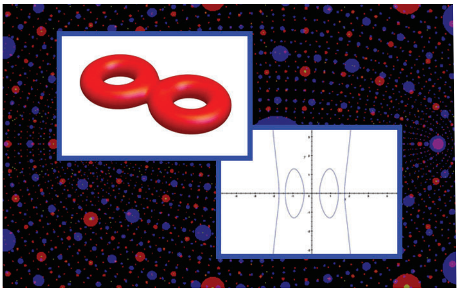
What can singularities of algebraic varieties say about the various decompositions of a positive integer into a sum of positive integers?


$1+\cfrac{e^{-2\pi}}{1+\cfrac{e^{-4\pi}}{1+\cfrac{e^{-6\pi}}{\ddots}}}$

4:00 p.m. | **Holly Krieger**
University of Cambridge

Uniformity when arithmetic meets geometry

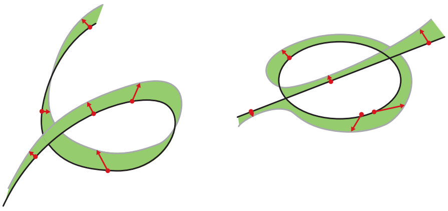
Understanding how algebra and geometry provide uniform control over the number of rational points on a curve.



5:00 p.m. | **Ravi Vakil**
Stanford University

Passing a curve through n points—solution of a 100-year-old problem

When can you string a curve through a number of points in space? How two young researchers finally settled an ancient problem.



This lecture is supported by the Bose, Datta, Mukhopadhyay, and Sarkar Fund.

Organized by **David Eisenbud**, *University of California, Berkeley*

AMS American Mathematical Society

jmm 2024 Joint Mathematics Meetings
San Francisco • January 3–6

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

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.

Thank you

The AMS is deeply grateful for the support of the Bose, Datta, Mukhopadhyay and Sarkar Fund, which sponsors one lecture at the Current Events Bulletin annually.

The Fund's purpose is to bring appreciation for mathematics to a broader audience. It was established in 2021 and made possible by the generosity of Dr. Salilesh Mukhopadhyay.

Searching for (sharp) thresholds in random structures: where are we now?

Will Perkins[*]

November 10, 2023

Abstract

We survey the current state of affairs in the study of thresholds and sharp thresholds in random structures on the occasion of the recent proof of the Kahn–Kalai Conjecture by Park and Pham and the fairly recent proof of the satisfiability conjecture for large k by Ding, Sly, and Sun. Random discrete structures appear as fundamental objects of study in many scientific and mathematical fields including statistical physics, combinatorics, algorithms and complexity, social choice theory, coding theory, and statistics. While the models and properties of interest in these fields vary widely, much progress has been made through the development of general tools applicable to large families of models and properties all at once. Historically these tools originated to solve or make progress on specific, difficult conjectures in the areas mentioned above. We will survey recent progress on some of these hard problems and describe some challenges for the future.

1 Introduction

Randomness is a powerful tool for algorithm design, scientific discovery, modeling the world, engineering, and mathematical proof. To use randomness effectively and confidently, we would like to accurately understand the properties of a typical random object or structure. Perhaps most familiar is the use of the Central Limit Theorem to understand statistical significance and margin of error in scientific studies and political polling.

In modern applications in algorithms, physics, social science, and other fields the random objects of interest can be very large and complex (think large computer networks, social networks, neural networks, or models of many interacting particles). We would like to be able to understand with as much accuracy as possible, what properties random structures typically possess, and how these typical properties change as underlying parameters change.

Even for the simplest class of random structures (those in which elements of a structured set are included independently with the same probability) this question can be extremely challenging, depending on the complexity of the property of interest. Over the past 40 or more years, specific problems have driven different fields (including probabilistic combinatorics, algorithms, and statistical physics) to develop powerful tools for investigating this type of problem. Some of these tools are designed for specific settings while others are very general.

Recently major progress on both general and specific problems has been made: Park and Pham proved the 'Kahn–Kalai Conjecture' [111] and Ding, Sly, and Sun proved the 'Satisfiability Conjecture' for large k [45]. The aim of this survey will be to put these exciting developments in a shared context and give an idea of some of the remaining and pressing challenges in the general area of thresholds in

[*]School of Computer Science, Georgia Institute of Technology, Atlanta, GA, USA

random structures. A key theme will be the way in which probability, combinatorics, and algorithms interact in these questions.

1.1 Random structures, monotone properties, and thresholds

The main setting for this survey will be p-biased product probability measures on the discrete cube $\{0,1\}^N$; that is, associating to a vector $x \in \{0,1\}^N$ the set $S_x = \{i \in [N] : x_i = 1\}$, probability measures of the form

$$\mu_p(S) = p^{|S|}(1-p)^{N-|S|} .$$

Typically we will think of N being large, and p small, possibly depending on N. This is simply a random subset with elements chosen independently with the same probability. Often the set of coordinates $[N] = \{1, \ldots, N\}$ will have additional structure; the coordinates might represent vertices or edges of some graph, or elements of an ordered set, or many other possibilities. Then we can think of the random set $S \sim \mu_p$ as being a *random structure.*

We will see numerous examples shortly, but perhaps the first example to have in mind is flipping a sequence of N independent coins each with probability p of landing heads. The structure here is the ordering of coordinates first to last.

A *property* $\mathcal{F}$ is simply an event in this probability space, $\mathcal{F} \subseteq \{0,1\}^N$. A property $\mathcal{F}$ is non-trivial if $\mathcal{F} \neq \emptyset$, $\mathcal{F} \neq \{0,1\}^N$. In our example, non-trivial properties include 'Flipping all heads'; 'Flipping at least $N/3$ heads'; 'Flipping an odd number of heads', 'Flipping a tail on the 7th and 11th flips', and so on. A property is *monotone* if it is closed under changing 0's to 1's; that is, if $x \in \mathcal{F}$ and $y \geq x$ coordinate-wise, then y must also be in $\mathcal{F}$. In the examples above about coin flips, all properties are monotone (or have a monotone complement) except the property of flipping an odd number of heads.

For a non-trivial monotone property, the probability that $\mathcal{F}$ holds under μ_p increases from 0 to 1 as p increases from 0 to 1, and we will be interested in how rapidly it does so. Take the property of flipping at least $N/3$ heads. By using a normal approximation to a binomial, we see that when $p = \frac{1}{3} + \frac{c\sqrt{2}}{3\sqrt{N}}$, $\mu_p(\mathcal{F})$ is approximately equal to the probability that a standard normal is at least $-c$; and so as p passes $1/3$ in an interval of length on the order $N^{-1/2}$, $\mu_p(\mathcal{F})$ jumps rapidly from near 0 to near 1. The rest of this survey is devoted to understanding similar phenomena for much more complex properties of random structures.

1.2 Examples

Here we give some examples from combinatorics, probability, statistical physics, computer science, and social choice theory. These examples have driven considerable interest in thresholds in random structures from different fields and their study from these different perspectives has also given the field shared language and intuition.

1.2.1 Probabilistic combinatorics

A central example comes from combinatorics. Here we take $n \in \mathbb{N}$ and let $N = \binom{n}{2}$, with coordinates representing edges of the complete graph K_n on n vertices. Then the elements of $\{0,1\}^N$ are in correspondence with (labeled) graphs on n vertices. A sample from μ_p is a random graph on n vertices in which each possible edge is included independently with probability p; the model is known as the *Erdős-Rényi random graph* [64, 50, 74, 58]. See two samples from $G(n,p)$ in Figures 1 and 2. Despite the simplicity of its definition, the Erdős-Rényi random graph exhibits a wide range of fascinating behaviors, and has been studied from many angles: as an interesting random structure in its own right, as a source of examples and counterexamples in graph theory, as a source of conjectured hard instances

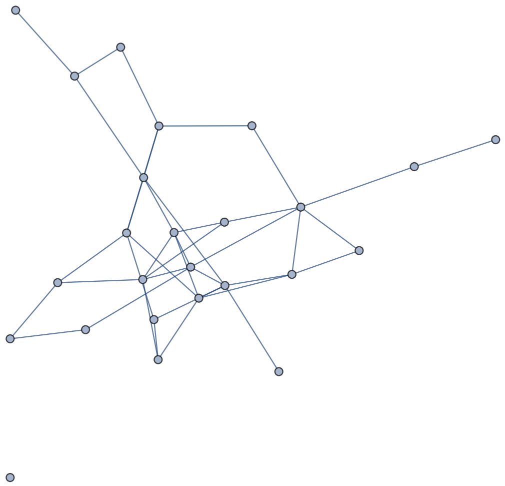


Figure 1: A realization of the random graph $G(25, 1/8)$

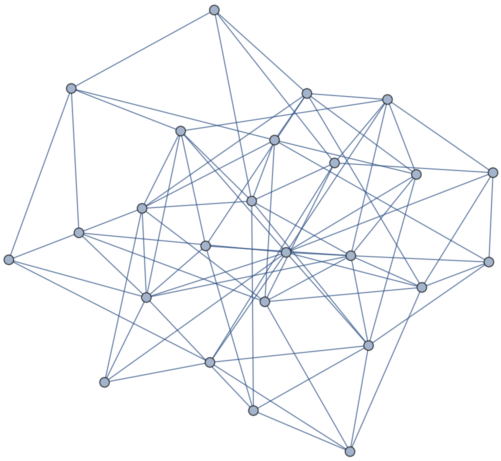


Figure 2: A realization of the random graph $G(25, 1/4)$

of combinatorial optimization problems, and as a host graph for random processes in probability and statistical physics.

1.2.2 Computer science

In computer science, a computational decision problem can be represented by a property of the discrete cube; or equivalently, *boolean function* $f : \{0, 1\}^N \to \{0, 1\}$, where the input x is encoded as a binary string and $f(x) = 1$ if x is a YES instance and 0 otherwise. A graph-theoretic decision problem corresponds directly to a property of graphs, as above. The focus in computer science is on the computational resources (time, space, etc.) required to compute a given boolean function f, either in the worst case over inputs, or, as we discuss more below, for typical or random instances x, such as those drawn according to μ_p.

1.2.3 Statistical physics

Much of the language and intuition around thresholds and sharp thresholds comes from physical systems that exhibit sudden changes in qualitative behavior as some parameter is changed in a small way. This is the phenomenon of a *phase transition*. The phase transitions of water from liquid to solid as temperature drops below 32° F or from liquid to gas as temperature rises past 100° F are familiar to everyone.

Mathematically, phase transitions are non-analytic points (or discontinuities in functions or their derivatives) of observables of infinite systems as some parameter varies. This is not quite the setting described above of sequences of finite random structures with parameters that vanish as the system size grows, but there is much to be gained by pushing this analogy in both directions; in Section 4 we describe the impact of the phase transition perspective on computer science, while [46] takes the analogy in the other direction.

The statistical physics model perhaps most related to μ_p is that of *percolation* [85, 67, 25]. Here the setting is typically an infinite graph, say a lattice like $\mathbb{Z}^d$ or the hexagonal lattice. In site (resp. bond) percolation each vertex (resp. edge) is declared 'open' independently with probability p; the main question is about the existence or non-existence of an infinite, connected 'open' component. See

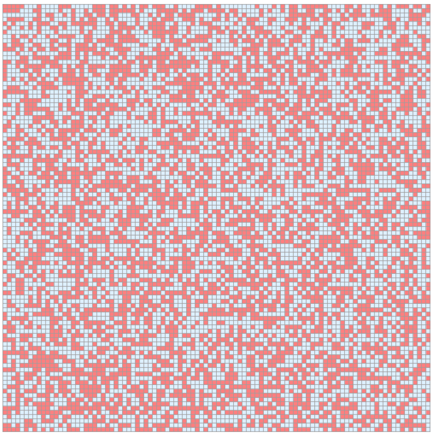


Figure 3: Site percolation on a 100×100 grid with $p = 1/2$

Figure 3 for a depiction of site percolation on a finite portion of $\mathbb{Z}^2$.

1.2.4 Social choice theory

Social choice theory is the study of collective decision making, including the study of voting mechanisms. We can view an election between two candidates with N voters as a boolean function $f : \{0,1\}^N \to \{0,1\}$, where the candidates are labeled 0 and 1, and a vector $x \in \{0,1\}^N$ is the list of the votes of the N voters. The function f is a rule for determining the winner of the election given the votes. A monotone function f can be thought of as a rule with the sensible property that if a voter changes their vote from 0 to 1 the outcome of the election cannot change from 1 to 0. Social choice theory lends a lot of evocative terminology to the study of boolean functions and thresholds. A 'dictator' boolean function is one that depends only on one coordinate; a 'junta' is one that depends only on a small number of coordinates. The majority function (i.e. equal representation democracy) has some extremal properties with applications in approximation algorithms ('Majority is stablest' [86, 108]), while the 'tribes' function, a hierarchical majority function from [21], inspired conjectures on general properties of boolean functions that have proved essential to the study of thresholds.

1.3 Outline

In Section 2 we define critical probabilities, thresholds, sharp and coarse thresholds, and scaling windows, then illustrate these notions with three examples from the study of random graphs. In the next three sections, we present some problems and conjectures that have driven progress on thresholds in different fields. In Section 3 we present the first of these main motivating questions, the question of thresholds for spanning structures in random graphs and hypergraphs. We discuss perfect matchings in hypergraphs, Latin squares, and the Kahn–Kalai Conjecture. In Section 4 we discuss the random k-SAT model, the satisfiability conjecture, and statistical physics predictions for thresholds in random computational problems. In Section 5 we describe the stochastic block model and the predicted information-theoretic and algorithmic thresholds in the model.

We conclude in Section 6 with some questions and open problems.

1.4 Further reading

This survey is not meant to be exhaustive at all, but rather aims to describe some recent results and open problems in different areas, all around the topic of thresholds and sharp thresholds in random structures. This area is very fortunate to have several excellent books and surveys devoted to the topic from different perspectives. Some of these sources are listed below with a few remarks.

Friedgut's 2005 survey 'Hunting for sharp thresholds' [55], following the developments in [54, 7, 57], is devoted to the question of which properties have sharp thresholds (defined below), criteria for proving sharpness of a threshold, and some intuition and meta conjectures on the topic.

Kalai and Safra's 2006 survey [81] takes a broad look at threshold phenomena in computer science, mathematics, and social choice theory and explains how notions arising in the latter (of 'influence' or 'pivotality' of coordinates) along with Fourier analysis of boolean functions can be used to study the sharpness or coarseness of thresholds. O'Donnell's textbook on discrete Fourier analysis of boolean functions [109] is a great reference to learn about these tools.

From a different perspective, Duminil-Copin's 2019 survey [46] describes how general tools from the study of sharp thresholds of boolean functions can be used to study phase transitions in classical statistical physics models like percolation on infinite graphs.

Rao's recent survey [115] describes the 2021 breakthrough of Alweiss, Lovett, Wu, and Zhang on sunflowers [14], the method of which led to the proof of the fractional Kahn–Kalai conjecture by Frankston, Kahn, Narayanan, and Park [53]. Even more recently Park has written an expository article [110] explaining the intuition behind the Kahn–Kalai conjecture and the consequences of its proof by Park and Pham [111].

Finally, in Sections 4 and 5 on random computational problems and statistical inference we discuss areas that have been shaped to a great extent by the field of statistical physics through questions, conjectures, and methods. The textbook of Mezard and Montanari [98] is a great resource for understanding these methods, while the surveys of Zdeborová and Krzakala [129], Moore [104], and Abbe [1] give an account of developments in this area in the last decade.

2 General questions, notions, and phenomena

In what follows, the perspective will be asymptotic as some underlying parameter, usually N or n, tends to infinity. We use standard asymptotic notation: $O(\cdot)$, $o(\cdot)$, $\Omega(\cdot)$, $\omega(\cdot)$, $\Theta(\cdot)$ to compare growth rates of functions. We use a subscript, e.g. $O_\varepsilon(\cdot)$, to indicate the implied constant may depend on ε. When discussing probabilities we say an event A holds 'with high probability' or 'whp' as $N \to \infty$ if $\lim_{N\to\infty} \Pr(A) = 1$; that is, $\Pr(A) = 1 - o(1)$. Here we implicitly consider a sequence of probability spaces.

For a non-trivial monotone property $\mathcal{F}$, the probability $\mu_p(\mathcal{F})$, as a function of p, is a strictly increasing function that starts with $\mu_0(\mathcal{F}) = 0$ and $\mu_1(\mathcal{F}) = 1$ (in fact it is a polynomial in p). The basic question about a monotone property of a random structure is how $\mu_p(\mathcal{F})$ increases from 0 to 1 as p increases. More generally we will think of a sequence of random structures indexed by N along with a sequence of monotone properties, and ask how $\mu_p(\mathcal{F}_N)$ increases from 0 to 1 as a function of N. From here on, we will write $\mathcal{F}$ and μ_p even though both implicitly depend on N.

Because $\mu_p(\mathcal{F})$ is strictly increasing, we can define the *critical probability* of $\mathcal{F}$, $p_c(\mathcal{F})$, as

$$p_c(\mathcal{F}) = \{p : \mu_p(\mathcal{F}) = 1/2\} .$$

The first task in studying a particular property $\mathcal{F}$ is to identify, at least approximately, the critical probability.

Task 1. *For a given monotone property $\mathcal{F}$, determine $p_c(\mathcal{F})$ asymptotically as $N \to \infty$, or at least determine the asymptotic order of $p_c(\mathcal{F})$.*

Obtaining the asymptotics of p_c means finding some $f(n)$ so that $p_c(\mathcal{F}) = (1+o(1))f(n)$; finding the asymptotic order means finding $f(n)$ so that $p_c(\mathcal{F}) = \Theta(f(n))$.

In the context of random graphs, this task was put forward in the original paper of Erdős and Rényi [50] and has been a central topic in probabilistic combinatorics since. In Section 3 we discuss one of the most challenging classes of problems for this task: that of determining the asymptotic order of p_c for the existence of different kinds of spanning subgraphs (including perfect matchings for random graphs and hypergraphs, H-factors of random graphs, and other combinatorial designs (including Latin squares) in random structures). We will discuss the recent breakthrough work on the Kahn–Kalai conjecture in [53, 111] that has led to a new powerful tool for determining the asymptotic order of p_c.

After Task 1, the most fundamental question about monotone properties is how quickly $\mu_p(\mathcal{F})$ increases from near 0 to near 1. To quantify this, we define the *width of the scaling window* of a monotone property. Let

$$T_\varepsilon(\mathcal{F}) = \{p_{1-\varepsilon} - p_\varepsilon : \mu_{p_{1-\varepsilon}}(\mathcal{F}) = 1-\varepsilon \text{ and } \mu_{p_\varepsilon}(\mathcal{F}) = \varepsilon\}.$$

In words T_ε is the amount p has to increase for the probability of $\mathcal{F}$ to increase from ε to $1-\varepsilon$.

Erdős and Rényi [50] defined a *threshold function* for a monotone property as follows: $p_{\mathrm{t}}(n)$ is a threshold function for a monotone property $\mathcal{F}$ if the following hold:

1. If $p = \omega(p_{\mathrm{t}}(n))$, then $\mu_p(\mathcal{F}) = 1 - o(1)$.
2. If $p = o(p_{\mathrm{t}}(n))$, then $\mu_p(\mathcal{F}) = o(1)$.

Equivalently, $\mathcal{F}$ has a threshold function (and we may take the function to be $p_c(\mathcal{F})$) if for every small $\varepsilon > 0$, $T_\varepsilon(\mathcal{F}) = O_\varepsilon(p_c(\mathcal{F}))$.

The following result of Bollobas and Thomason justifies the abstract study of thresholds in monotone properties and the specific definition of $p_c(\mathcal{F})$ above.

Theorem 1 (Bollobas–Thomason [27])**.** *Every non-trivial monotone property has a threshold function; moreover one can take $p_c(\mathcal{F})$ to be this threshold function.*

The original proof of this theorem uses the Kruskal–Katona theorem, but a simple modern proof is as follows. Let K be a large positive integer. If $\mu_p(\mathcal{F}) = \varepsilon$, then $\mu_{Kp}(\mathcal{F}) \geq 1 - (1-\varepsilon)^K$ by superimposing K independent copies of the random structure and applying monotonicity; taking $K \approx \log(1/\varepsilon)/\varepsilon$ proves the theorem.

Erdős and Rényi further classified thresholds as *sharp* or *coarse* according to how $T_\varepsilon(\mathcal{F})$ compares to $p_c(\mathcal{F})$. We say $\mathcal{F}$ has a *sharp threshold* if for every $\varepsilon > 0$, $T_\varepsilon(\mathcal{F}) = o_\varepsilon(p_c(\mathcal{F}))$; otherwise $\mathcal{F}$ has a coarse threshold. Equivalently, $\mathcal{F}$ has a sharp threshold if for every $\varepsilon > 0$,

1. If $p \geq (1+\varepsilon)p_c(\mathcal{F})$, then $\mu_p(\mathcal{F}) = 1 - o(1)$.
2. If $p \leq (1-\varepsilon)p_c(\mathcal{F})$, then $\mu_p(\mathcal{F}) = o(1)$.

Task 2. *Determine if a monotone property $\mathcal{F}$ has a sharp or coarse threshold.*

In fact, in very general settings Task 2 has been solved, with the machinery of discrete Fourier analysis. Roughly, as Friedgut describes in [55] and Kalai and Safra write in [81], we should expect a sharp threshold unless there is reason to expect a coarse threshold; and the reason to expect a coarse threshold is if $\mathcal{F}$ is essentially determined by a small number of coordinates or by the presence or absence of a small substructure. This is made rigorous in increasing generality in Friedgut's theorem on sharp thresholds [54], Bourgain's theorem in the appendix of [54], and Hatami's result in [68]. We describe in Section 4 the random k-SAT model, a main motivation for Friedgut's important result.

A final task in understanding a particular property $\mathcal{F}$ with a sharp threshold is getting more precise bounds on the scaling window.

Task 3. *Determine the asymptotic order of the width of the scaling window of a monotone property $\mathcal{F}$.*

Unlike with Tasks 1 and 2, as of yet there is no general principle in determining the magnitude of $T_\varepsilon(\mathcal{F})$ beyond distinguishing sharp from coarse: the width of the scaling window really seems to depend on the particular details of the property $\mathcal{F}$ in question, and being able to determine the asymptotic order of the width indicates a near-complete understanding of the property.

2.1 Examples

We give three examples to illustrate these different tasks. The examples are the properties in the random graph $G(n,p)$ of *containing a triangle*, *being connected*, and *being 3-colorable*; call these properties $\mathcal{F}_{K_3}$, $\mathcal{F}_{\text{connected}}$, $\mathcal{F}_{3-\text{col}}$ respectively.

To understand the probability that $G(n,p)$ contains a triangle, we can use the first- and second-moment methods (see e.g. [12] for an exposition). Let X be the number of triangles. Then $\mathbb{E}X = \binom{n}{3}p^3$ and $\text{var}(X) = (1+o(1))\binom{n}{3}p^3$. Using this we can determine that $p_c(\mathcal{F}_{K_3}) = \Theta(1/n)$. Markov's Inequality yields:

$$\Pr(\mathcal{F}_{K_3}) = \Pr(X \geq 1) \leq \mathbb{E}X = (1+o(1))\frac{n^3p^3}{6} = o(1)$$

when $p = o(1/n)$. Chebyshev's Inequality yields

$$\Pr(\overline{\mathcal{F}_{K_3}}) = \Pr(X = 0) \leq \Pr(|X - \mathbb{E}X| \geq \mathbb{E}X) \leq \frac{\text{var}(X)}{(\mathbb{E}X)^2} = (1+o(1))\frac{6}{n^3p^3} = o(1)$$

when $p = \omega(1/n)$. This pair of facts tells us that $p = c/n$ (for any constant $c > 0$) is a threshold function for $\mathcal{F}_{K_3}$. In fact much more is known: when $p = c/n$, the distribution of X converges to a Poisson$(c^3/6)$ random variable. Thus the scaling window is of length $\Theta(1/n)$ and the threshold is coarse; see Figure 4.

Next we turn to the property $\mathcal{F}_{\text{connected}}$. Being connected is a more 'global' property than that of containing a triangle, but there is an easy lower bound on $p_c(\mathcal{F}_{\text{connected}})$: if G has an isolated vertex (and $n > 1$) then G must be disconnected. Or, taking complements, $\mathcal{F}_{\text{connected}} \subseteq \mathcal{F}_{\text{no isolated vertex}}$. As with triangles it is straightforward to understand the threshold for containing isolated vertices using the method of moments. Let Y be the number of isolated vertices; then $\mathbb{E}Y = n(1-p)^{n-1}$. As with triangles, when $\mathbb{E}Y$ tends to a positive constant, the distribution of Y converges to a Poisson random variable; when $\mathbb{E}Y \to 0$ whp $Y = 0$; and when $\mathbb{E}Y \to \infty$, whp $Y \geq 1$. This tells us the exact scaling window for $\mathcal{F}_{\text{no isolated vertex}}$: if $p = \frac{\log n + c}{n}$ with $c \in \mathbb{R}$ constant, then $\Pr(\mathcal{F}_{\text{no isolated vertex}}) = (1+o(1))e^{-e^{-c}}$, see Figure 5. In this case the width of the scaling window is $\Theta(1/n) = o(p_c)$ and so the threshold is sharp.

It turns out that a first-moment argument on connected components of size ≥ 2 from Erdős and Rényi [49] shows that in fact $\mathcal{F}_{\text{no isolated vertex}}$ approximates $\mathcal{F}_{\text{connected}}$ very well indeed around the threshold. An effective way to state the approximation is as a *hitting time result*, which we now describe.

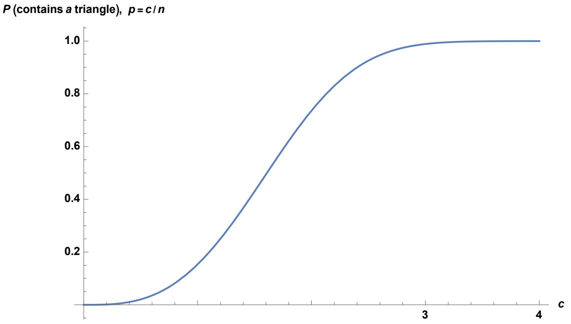


Figure 4: Scaling window of $\mathcal{F}_{K_3}$

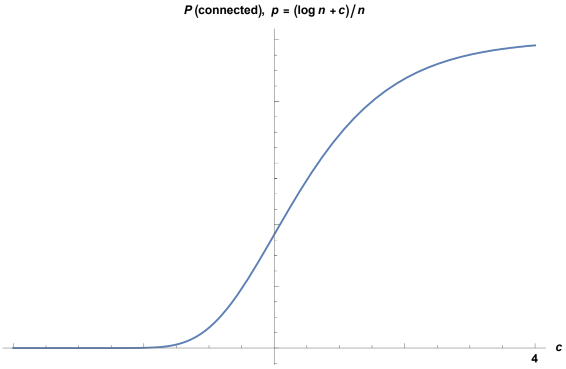


Figure 5: Scaling window of $\mathcal{F}_{\text{connected}}$

We can couple the random graphs $G(n,p)$ for all $p \in [0,1]$ simultaneously by drawing iid Uniform$[0,1]$ random variables U_{ij}, $1 \le i < j \le n$. Then we form $G(n,p)$ by including $(ij) \in E$ if and only if $U_{ij} \le p$. Under this coupling $G(n,p')$ is a subgraph of $G(n,p)$ if $p' \le p$. We can then define the 'random graph process' as the discrete time process $G_0, G_1, \dots, G_{\binom{n}{2}}$ where at each step a uniformly random edge (not already present) is added to the graph. This process is what results from raising p from 0 to 1 and marking the appearance of new edges in the coupling.

Theorem 2 (Hitting time for connectivity [49])**.** *Whp over the random graph process we have*

$$\min_t \{G_t \textit{ is connected}\} = \min_t \{G_t \textit{ has no isolated vertices}\} .$$

That is, the edge that touches the last isolated vertex in the process also connects the graph, with high probability over the process. This implies that the threshold and scaling widow are the same for $\mathcal{F}_{\text{connected}}$ and $\mathcal{F}_{\text{no isolated vertex}}$.

We will see below in Section 3 further examples of hitting time results along the same lines for more complicated properties than connectivity.

Finally, consider a property for which we do not yet have a full understanding. Let $\mathcal{F}_{3-\text{col}}$ be the property that G has a proper 3-coloring of its vertices (no monochromatic edges). The complement of this property is a non-trivial monotone property.

One can show that if $p = c/n$, $c < 1$, then whp $G(n,p)$ is 3-colorable by showing that whp all connected components are trees or unicyclic (and thus 3-colorable). For an upper bound on p_c one can use the first moment method on Z, the number of 3-colorings of G. As a first attempt we can bound $\mathbb{E}Z \le 3^n (1-p)^{\frac{n^2}{6}}$ using the fact that a balanced partition of n vertices has the fewest potential monochromatic edges. This tells us that $G(n, c/n)$ is not 3-colorable whp when $c > 6 \log 3 \approx 6.5917$. However, we can do something a little bit more clever: whp $G(n, c/n)$ has $\frac{cn}{2} + o(n)$ edges and so we can condition on the number of edges m, and compute $\mathbb{E}Z$ in the random graph G_m from the random graph process. This gives the bound

$$\mathbb{E}_{G_m} Z \le 3^n \left(1 - \frac{1}{3}\right)^m$$

which tends to 0 when $m \sim cn/2$, $c > \frac{2 \log 3}{\log 3 - \log 2} \approx 5.419$. We will see more of this idea of combining the moment method with conditioning on typical events in Section 4.

Together these facts tell us that $p_c(\mathcal{F}_{3-\text{col}}) = \Theta(1/n)$ and the width of the scaling window is $O(1/n)$, uniformly over ε. What are the asymptotics of p_c? Is there some d_3 so that $G(n, d/n)$ is 3-colorable whp when $d < d_3$ and whp not 3-colorable when $d > d_3$? In fact this is the only remaining open problem from the original Erdős and Rényi paper on random graphs [50]. In Section 4 we describe predictions for this problem made by statistical physicists.

3 Perfect matchings, Latin squares, Kahn–Kalai, and spread distributions

From the beginning of the study of random graphs, a key question has been 'For a given subgraph H, when in the evolution of the random graph does a copy of H first appear?' That is, what is the threshold for the property $\mathcal{F}_H$ of containing a copy of H.

For fixed size H, the answer is now known completely [23], though more delicate in general than the case of $H = K_3$ above. A threshold function for $\mathcal{F}_H$ is the *expectation threshold* for the densest subgraph H' of H (maximizing the ratio of edges to vertices over all induced subgraphs of H); the value $p_E(H')$ so that the expected number of copies of H' in $G(n, p_E(H'))$ is 1. For $p = o(p_E(H'))$, the expected number of copies of H' tends to 0, and so whp $G(n,p)$ has no copies of H' and thus no copies of H. When the expected number of copies of H' for each $H' \subseteq H$ tends to infinity, it has been showed that whp a copy of H exists.

It can be much more difficult to determine the threshold for the appearance of subgraphs whose size grows with n; in particular, for *spanning subgraphs* that include each of the n vertices. Two classical examples are those of H being a perfect matching (with n even) and H being a Hamilton cycle. As with connectivity, there are natural lower bounds for $p_c(\mathcal{F}_H)$ in these cases, coming from local obstructions. If G has an isolated vertex, it cannot have a perfect matching; if G has a vertex of degree at most 1 it cannot have a Hamilton cycle. In both cases these obstructions determine the thresholds, even in the strong form of hitting-time results [51, 90, 26] that give a complete understanding of the scaling windows.

Other spanning subgraph problems remained open for many years. One famous such problem is Shamir's Problem [119] (mentioned first in print by Erdős in [48]): the problem of determining the threshold for a random k-uniform hypergraph to contain a perfect matching.

A k-uniform hypergraph $G = (V, E)$ is a set of vertices V along with a collection E of k-sets of V (the case $k = 2$ is a usual graph). The random k-uniform hypergraph $G^{(k)}(n,p)$ is a k-uniform hypergraph on n vertices in which each possible k-set is an edge independently with probability p. A perfect matching in a k-uniform hypergraph on n vertices is a set of n/k vertex-disjoint hyperedges that (necessarily) cover all n vertices; n must be divisible by k for a perfect matching to be possible. For $k = 2$ (a graph) this coincides with the usual definition of a perfect matching.

As with perfect matchings in graphs, the natural obstruction to a perfect matching in a hypergraph is a vertex not appearing in any hyperedges; this immediately yields a lower bound of $\Omega(\log n \cdot n^{1-k})$ for the threshold for perfect matchings.

After many years of effort, Johansson, Kahn, and Vu proved that the lower bound is tight up to constants.

Theorem 3 (Johansson–Kahn–Vu [76]). *Let $\mathcal{F}$ be the property of $G^{(k)}(n,p)$ containing a perfect matching when k divides n. Then*

$$p_c(\mathcal{F}) = \Theta_k(\log n \cdot n^{1-k}) .$$

Much more recently Kahn proved much finer results: first-order asymptotics of $p_c(\mathcal{F})$ in [78] and the definitive hitting-time result [77] (see also the related [69, 116, 70]).

3.1 The Kahn–Kalai conjectures

Motivated by challenges like Shamir's Problem, Kahn and Kalai [79] made two bold conjectures about thresholds. These conjectures arise from the fact that it is often straightforward to give a lower bound for $p_c(\mathcal{F})$; that is, to show that when p is small enough $\mu_p(\mathcal{F}) = o(1)$. A matching or near matching upper bound is often much more difficult. This is especially apparent in the examples above (connectivity, perfect matchings, etc.) in which there is a trivial local obstruction to some property. Kahn and Kalai's two conjectures both have a similar flavor: 'The best possible easy lower bound on p_c is not too far from the truth.' We present the two conjectures in the opposite order they appear in [79] (but in the order they arrived at them).

Their second conjecture is about the property $\mathcal{F}_H$ of containing an isomorphic copy of some subgraph H in a random graph (or hypergraph). We have seen several examples of this for different H, including a triangle, a perfect matching, a Hamilton cycle, and a hypergraph perfect matching. As we saw for fixed-size subgraphs, it can be crucial to consider the rarest (in terms of expected number of copies) subgraphs of H. This motivates the following definition of the *subgraph expectation threshold* of H:

$$p_E(H) = \min\{p : \mathbb{E}_{G(n,p)} X_{H'} \geq 1 \, \forall \text{ subgraphs } H' \subseteq H \} \,.$$

Clearly $p_E(H)$ bounds below the threshold of $\mathcal{F}_H$; if $p = o(p_E(H))$ then there is some subgraph $H' \subseteq H$ that whp does not appear in $G(n,p)$. Kahn and Kalai conjecture this cannot be too far from the truth.

Conjecture 4 (Kahn and Kalai [79])**.** *There exists an absolute constant K so that for all graphs H,*

$$p_c(\mathcal{F}_H) \leq K \cdot p_E(H) \cdot \log |V(H)| \,.$$

For H of fixed size, Conjecture 4 asserts that $p_e(H)$ is a threshold for $\mathcal{F}_H$ and this is known to be true as described above. The real content of the conjecture is for H that grow with n. In particular Conjecture 4 is tight (up to the constant K) for perfect matchings in graphs and hypergraphs.

Their first conjecture was in the more general setting of monotone properties of $\{0,1\}^N$. To state it we must define a more abstract notion of an expectation threshold (following [123, 124, 79]). A property $\mathcal{F} \subseteq \{0,1\}^N$ is *p-small* if there exists a 'cover' $\mathcal{G} \subseteq \{0,1\}^N$ so that:

1. $\forall T \in \mathcal{F} \; \exists S \in \mathcal{G}, S \subseteq T$.
2. $\sum_{S \in \mathcal{G}} p^{|S|} \leq \frac{1}{2}$.

To give some intuition for this definition, if H is a graph and $H' \subseteq H$, then $\mathcal{F}_{H'}$ is a cover of $\mathcal{F}_H$, satisfying the first condition above.

The *expectation threshold* of $\mathcal{F}$, $q(\mathcal{F})$, is the largest p for which $\mathcal{F}$ is p-small. Finally let $L(\mathcal{F})$ be the maximum size of a minimal element of $\mathcal{F}$.

Notice that $q(\mathcal{F})$ is a lower bound for $p_c(\mathcal{F})$. Let $q = q(\mathcal{F})$. Then

$$\begin{aligned}\mu_q(\mathcal{F}) &\leq \sum_{S \in \mathcal{G}} \sum_{T \in \mathcal{F}, T \supseteq S} \mu_q(T) \\ &\leq \sum_{S \in \mathcal{G}} \sum_{T \supseteq S} \mu_q(T) \\ &= \sum_{S \in \mathcal{G}} q^{|S|} \leq \frac{1}{2} \,.\end{aligned}$$

Kahn and Kalai conjectured that in the abstract setting this cannot be far from the truth for p_c. Park and Pham then proved this.

Theorem 5 (Kahn–Kalai Conjecture, now Park–Pham Theorem [111]). *There exists an absolute constant K so that for every monotone $\mathcal{F}$,*

$$p_c(\mathcal{F}) \leq K \cdot q(\mathcal{F}) \cdot \log L(\mathcal{F}) .$$

Theorem 5 gives a powerful method for bounding p_c from above: find a good (or the best) $\mathcal{G}$ with the properties above, and this will determine p_c up to a $\log N$ factor or better. Looking ahead, we can characterize $q(\mathcal{F})$ via the following integer optimization problem over the variables $g(x)$, $x \in \{0,1\}^N$ which we interpret as the indicator vector of a potential cover $\mathcal{G}$ witnessing $\mathcal{F}$ being p-small.

$$\begin{aligned} V(\mathcal{F},p) = \min & \sum_{x \in \{0,1\}^N} p^{|x|} g(x) \\ & \text{subject to} \\ & g(x) \in \{0,1\} \text{ for all } x \in \{0,1\}^N \\ & \sum_{S \subseteq T} g(x_S) \geq 1 \text{ for all } T \in \mathcal{F} \end{aligned}$$

In particular, if $V(\mathcal{F},p) \geq 1/2$, then $q(\mathcal{F}) \leq p$, and by Theorem 5, $p_c(\mathcal{F}) \leq K \cdot p \cdot \log L(\mathcal{F})$. However, solving an integer program over 2^N variables (for all large N) seems somewhat formidable.

3.2 Duality and spread distributions

Before Park and Pham proved Theorem 5, Frankston, Kahn, Narayanan, and Park [53] (following the breakthrough of Alweiss, Lovett, Wu, and Zhang on the sunflower conjecture [14]) proved a related but weaker conjecture of Talagrand from [125], dubbed the 'Fractional Kahn–Kalai Conjecture'. To state this, we first relax the integrality constraint above and define a linear program (in 2^N variables)

$$\begin{aligned} V_f(\mathcal{F},p) = \min & \sum_{x \in \{0,1\}^N} p^{|x|} g(x) \\ & \text{subject to} \\ & g(x) \in [0,1] \text{ for all } x \in \{0,1\}^N \\ & \sum_{S \subseteq T} g(x_S) \geq 1 \text{ for all } T \in \mathcal{F} . \end{aligned}$$

Then define $q_f(\mathcal{F})$, the *fractional expectation threshold* of $\mathcal{F}$, to be the largest p for which $V_f(\mathcal{F},p) \leq 1/2$. Again it is easy to see that $p_c(\mathcal{F}) \geq q_f(\mathcal{F})$, with the same proof as above.

Moreover, we have the relations

$$q(\mathcal{F}) \leq q_f(\mathcal{F}) \leq p_c(\mathcal{F}) \leq K \cdot q(\mathcal{F}) \cdot \log L(\mathcal{F}) \leq K \cdot q_f(\mathcal{F}) \cdot \log L(\mathcal{F}) ,$$

where the second-to-last inequality is Theorem 5; the weaker inequality $p_c(\mathcal{F}) \leq K \cdot q_f(\mathcal{F}) \cdot \log L(\mathcal{F})$ is the main result of [53].

Theorem 6 (Frankston, Kahn, Narayanan, and Park [53]). *There exists an absolute constant K so that for every monotone $\mathcal{F}$,*

$$p_c(\mathcal{F}) \leq K \cdot q_f(\mathcal{F}) \cdot \log L(\mathcal{F}) .$$

To bound p_c with Theorem 6 it suffices to lower bound $V_f(\mathcal{F},p)$: if $V_f(\mathcal{F},p) \geq 1/2$ then $p_c(\mathcal{F}) \leq K \cdot p \cdot \log L(\mathcal{F})$. The nice thing about the linear programing formulation is that we can use duality to give a lower bound. We can write the dual as a linear program with variables $\nu(T)$ for each $T \in \mathcal{F}$ and constraints for each $S \subseteq \{0,1\}^N$.

$$\begin{aligned} V_f(\mathcal{F}, p) = &\max \sum_{T \in \mathcal{F}} \nu(T) \\ &\text{subject to} \\ &\nu(T) \geq 0 \text{ for all } T \in \mathcal{F} \\ &\sum_{T \supseteq S} \nu(T) \leq p^{|S|} \text{ for all } S \subseteq \{0,1\}^N . \end{aligned}$$

We are interested in showing $V_f(\mathcal{F}, p) \geq 1/2$ to bound p_c and via duality this can be accomplished by exhibiting a good $\nu(\cdot)$. A slightly more elegant formulation is due to Talagrand [125], who made the following definition of a *spread probability distribution*.

Definition 7. *Let $\mathcal{F} \subseteq \{0,1\}^N$. A probability measure ν supported on $\mathcal{F}$ is p-spread if for all $S \subseteq \{0,1\}^N$,*

$$\sum_{T \supseteq S} \nu(T) \leq 2p^{|S|} .$$

Putting all of the above together we obtain a very useful theorem.

Theorem 8. *There is an absolute constant K so that the following is true. Let $\mathcal{F}$ be a monotone property that supports a p-spread probability measure ν. Then*

$$p_c(\mathcal{F}) \leq K \cdot p \cdot \log L(\mathcal{F}) .$$

Theorem 8 follows from Theorem 6, duality, and the extra factor 2 allowing us to take ν to be a probability measure.

3.3 Applications

Theorem 8 is a brand-new tool for proving upper bounds on $p_c(\mathcal{F})$ and we will recount here some spectacular applications. But first as a warm-up we will see how it can be used to establish $p_c(\mathcal{F}_{\text{perfect matching}}) = \Theta\left(\frac{\log n}{n}\right)$.

Let n be even, and let ν be the uniform distribution on perfect matchings of the complete graph K_n. Let A be some set of edges of K_n and let $M \sim \nu$. We want to bound the probability that $A \subseteq M$. First note that if A is not a matching then this probability is 0. Now suppose A is a matching of size k. Then, letting $\text{pm}(G)$ denote the number of perfect matchings of G and $(a)_b = \frac{a!}{(a-b)!}$,

$$\Pr(A \subseteq M) = \frac{\text{pm}(K_{n-2k})}{\text{pm}(K_n)} = \frac{(n-2k)!}{2^{n/2-k}(n/2-k)!} \frac{2^{n/2}(n/2)!}{n!} = \frac{2^k (n/2)_k}{(n)_{2k}} \leq \left(\frac{e}{n}\right)^k ,$$

and so ν is $\frac{e}{n}$-spread. Applying Theorem 8 gives the result, though note that this is weaker than the sharp threshold obtained in [51]. Here spreadness is proved by counting, and it helps a lot that we have an explicit formula for $\text{pm}(K_n)$ to use.

From Theorem 8 and similar counting arguments (see e.g. [53]), one can also rather easily obtain the asymptotic order of the threshold for perfect matchings in hypergraphs and K_r-factors as well as the threshold for containing any bounded-degree spanning tree, previously known through the long proofs of Johansson, Kahn, and Vu [76] and Montgomery [103] respectively.

Perhaps even more exciting are applications which before the theorem were completely out of reach but which now can be approached through finding spread distributions. We describe one application here but see also [114, 84, 15] for other applications of either Theorem 8 or the notion of spread distributions.

A *Latin square* of order n is an $n \times n$ matrix with entries in $\{1, \ldots, n\}$ in which no row or column contains repeated entries. A Latin square of order 8 is given in Table 1. One can ask about the existence of Latin squares of order n, the number of Latin squares of order n, or the existence of Latin squares with additional properties. Johansson [75] asked a very natural threshold question about Latin squares: if each position (ij) in an $n \times n$ matrix is assigned a list L_{ij} of allowed symbols in $\{1, \ldots, n\}$ by including each element in L_{ij} independently with probability p, what is the threshold in p for the existence of a Latin square in which each entry appears in the corresponding list L_{ij}? The lower bound comes from a simple local obstruction: for $p \leq (1-\varepsilon)\frac{\log n}{n}$ there will be empty lists whp. A natural conjecture given the results and conjectures on perfect matchings and factors in random graphs is that this local obstruction determines the threshold: its asymptotic order, first-order asymptotics, and even in the form of a hitting time. This conjecture is stated explicitly in [94], and see also the closely related conjectures in [32, 121, 82, 118].

1	3	4	8	2	6	7	5
6	1	2	5	7	4	8	3
5	2	3	1	8	7	4	6
4	6	5	3	1	8	2	7
2	4	6	7	3	5	1	8
7	8	1	2	6	3	5	4
8	5	7	6	4	1	3	2
3	7	8	4	5	2	6	1

Table 1: A Latin square of order 8

Conjecture 9. *The property $\mathcal{F}_{Latin}$ has a sharp threshold at $p = \frac{\log n}{n}$*

Before Theorem 8, there was really no effective way to to prove upper bounds on $p_c(\mathcal{F}_{\text{Latin}})$: the best known result was that $p_c \leq 1 - \delta$ for some fixed $\delta > 0$ [16].

Using Theorem 8 and finding a sufficiently spread distribution on Latin squares, Sah, Sawhney and Simkin proved Conjecture 9 up to subpolynomial factors: $p_c = n^{-1+o(1)}$ [118]. Intriguingly, this bound already established that the threshold is sharp without determining the threshold, by applying Friedgut's theorem discussed in the next section. Following this, Kang, Kelly, Kühn, Methuku, and Osthus [82] proved $p_c = O(\log^2 n/n)$, within a factor $\log n$ of the conjecture. Then very recently Jain and Pham, and independently Keevash, established the correct order of the threshold.

Theorem 10 (Jain–Pham [73]; Keevash [83])**.**

$$p_c(\mathcal{F}_{Latin}) = \Theta\left(\frac{\log n}{n}\right) .$$

Using Theorem 8, it suffices to construct an $O(1/n)$-spread distribution on Latin squares. It is natural to expect that (as in the case of perfect matchings) the *uniform distribution* is $O(1/n)$ spread, but unlike perfect matchings it is very challenging to enumerate Latin squares. So instead, the above authors have constructed non-uniform spread distributions using sophisticated tools from probabilistic and extremal combinatorics, namely iterative absorption (in [118, 82]), analyzing the Lovász Local Lemma probability distribution (in [73]), and analyzing a random greedy stochastic process (in [83]).

4 Random k-SAT and the satisfiability conjecture

The most important open problem in theoretical computer science (and one of the most important in all of mathematics) is the P vs NP question: can computational decision problems with polynomial-time checkable certificates be solved in polynomial time?

Most computer scientists and mathematicians believe that P$\neq$NP and that many hard computational problems exist; these include classic problems like MAX-CUT, Max Independent Set, traveling salesman, boolean satisfiability, graph coloring, among many others. It is believed that these problem are computationally intractable in the worst-case over instances. On the other hand, large instances of these problems are solved every day (see for example a survey on the success of SAT solvers in [61] or exact solutions to very large instances of the traveling salesman problem in e.g. [17, 33]). What can explain this discrepancy? One possibility is that 'typical' instances of certain NP-hard problems are tractable while hard instances are exceptional.

This is one motivation for the study of *average-case complexity*: the computational tractability or intractability of random instances of computational problems. Average-case complexity is a huge topic with many fascinating facets and many mysteries (see e.g. [22, 126]). Here we will deal will a small slice of the topic.

Recall that a boolean CNF formula is the AND of OR's of literals (boolean variables and their negations). A k-CNF formula is the AND of clauses of k literals each. For example, the following is a small 3-SAT formula:

$$(x_7 \vee \overline{x}_2 \vee \overline{x}_3) \wedge (\overline{x}_1 \vee x_8 \vee \overline{x}_4) \wedge (x_2 \vee x_4 \vee x_1)\,.$$

The k-SAT problem is to find an assignment of True and False to the boolean variables so that the given k-CNF formula evaluates to true; if such an assignment exists the formula is satisfiable and unsatisfiable otherwise.

An empirical observation about algorithms brought the study of random computational problems to the attention of computer scientists, statistical physicists, and probabilists. In 1996, Mitchell, Selman, and Levesque [120] (working on problems in Artificial Intelligence and computational deductive reasoning) generated uniformly random 3-SAT instances on n variables with m constraints, for fairly large n and different values of m. They ran standard heuristic SAT solving and SAT refutation algorithms and observed the following: the running time required to find a solution (or find a proof that none existed) showed a sharp peak (as a function of m) right around the point at which an estimate of the probability of such a random instance being satisfiable made a sharp decrease from near 1 to near 0.

These two empirical observations – that random k-SAT exhibits a sharp threshold and that instances near the threshold are computationally hard – set off an explosion of work on random k-SAT and related models (random graph coloring, k-NAE-SAT, k-XOR-SAT, etc.) in many different directions.

We can model random k-SAT in the setting of this paper. For a given n and p, let $F_k(n,p)$ be a random k-SAT formula generated by including each of the $\binom{2n}{k}$ possible k-clauses independently with probability p. The property of being satisfiable is a non-trivial property with a monotone complement (and thus has a threshold function).

One long-standing conjecture is that the random k-SAT model exhibits a sharp threshold.

Conjecture 11 (Satisfiability Conjecture)**.** *For each $k \geq 2$, there exists $c_k > 0$ so that for every $\varepsilon > 0$, the following hold:*

- *If $p \leq (1-\varepsilon)c_k n^{-1/(k-1)}$, then whp $F_k(n,p)$ is satisfiable.*
- *If $p \geq (1+\varepsilon)c_k n^{-1/(k-1)}$, then whp $F_k(n,p)$ is unsatisfiable.*

Conjecture 11 was proved early on for the special case of random 2-SAT [34]; and in fact the precise scaling window was determined [24]. The reason for this is that determining satisfiability of a 2-SAT formula can be reduced to determining the strongly connected components of the implication graph on literals; this gives a linear-time algorithm for 2-SAT in the worst case and gives a strong analogy between the random 2-SAT threshold and the emergence of a giant component in a random graph. The case $k \geq 3$ is fundamentally different.

Random k-SAT is an example of a *random constraint satisfaction problem* (random CSP); there is an initial set of possible solutions (here all possible assignments to n boolean variables) and a set of random constraints is selected, each of which rules out some solutions (here each clause rules out a 2^{1-k} fraction of all possible solutions). Random graph coloring is another random CSP: the q^n colorings of n vertices are the possible solutions; each edge rules out $\frac{1}{q}$-fraction of the solutions.

How can one locate the satisfiability threshold in a random CSP? As with many of the examples above, a starting point is to try the first- and second-moment methods on the random variable Z that counts the number of solutions. If p is large enough that $\mathbb{E}Z \to 0$, we know $p_c \leq p$; conversely if p is such that $\mathbb{E}Z \to \infty$ and, say, $\mathbb{E}(Z^2) \leq C(\mathbb{E}Z)^2$, then we know $p_c \geq p$.

This approach, combined with smart ideas of conditioning, replacing Z with a related random variable, and solving difficult optimization problems to bound $\mathbb{E}(Z^2)$, has been able to pin down the satisfiability threshold in many random CSP's to within a small constant factor (e.g. [89, 9, 11, 10]). These probabilistic methods face a barrier, however, and cannot address Conjecture 11 or determine p_c. Instead, new tools and ideas from mathematics and statistical physics have been brought to bear on the problem.

4.1 Friedgut's Theorem

Major progress towards Conjecture 11 was made by Friedgut in [54] who proved an n-dependent sharp threshold.

Theorem 12 (Friedgut). *For each $k \geq 3$ there is a function $c_k(n)$ bounded above and below by constants so that for every $\varepsilon > 0$ the following hold:*

- *If $p \leq (1-\varepsilon)c_k(n) \cdot n^{-1/(k-1)}$, then whp $F_k(n,p)$ is satisfiable.*
- *If $p \geq (1+\varepsilon)c_k(n) \cdot n^{-1/(k-1)}$, then whp $F_k(n,p)$ is unsatisfiable.*

This theorem shows that the scaling window of random k-SAT is $o(p_c)$ but it leaves open the possibility that the critical density c_k fluctuates as $n \to \infty$.

Though motivated by the Satisfiability Conjecture, Friedgut proved much more in [54]: he gave a very general characterization of what monotone properties of random graphs can have a coarse threshold: those that are well approximated by the property of containing some bounded-size subgraph from a bounded-size list of subgraphs. All other properties have a sharp threshold (in the non-uniform sense of Theorem 12).

Returning to the example from Section 2.1, Achlioptas and Friedgut [7] use [54] to show that for $q \geq 3$ the property of $G(n,p)$ being q-colorable exhibits a sharp threshold, in the sense of Theorem 12. It remains open to show that there exists constants d_q so that a sharp threshold for q-colorability occurs at d_q/n.

4.2 The cavity method and the structure of solution spaces

Statistical physicists soon turned their attention to random k-SAT and related models using tools and intuition from the study of spin glasses [99, 101, 100, 97, 92, 102]. The replica and cavity methods

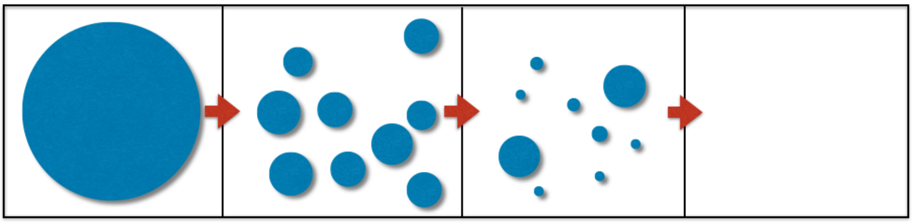


Figure 6: Cartoon of the evolution of the solution space for random k-SAT or random graph coloring as the edge density increases.

are powerful analytic tools for analyzing disordered (random) systems, based on assumptions that to a large extent have not been proved mathematically. Under these assumptions, these methods are capable of making very detailed predictions for thresholds and phase transitions in a broad class of random structures, *Gibbs measures on random graphs*.

Applying these methods to random CSP's involves shifting perspective and instead of asking whether or not there is a solution, asking 'what does the uniform distribution on solutions (if they exist) look like?' Then the main properties of interest are *correlations* and *overlaps*. Correlations are measured, for instance, by the covariance in the indicator random variables that two variables take the value True (or two vertices take the color Red). Overlap is the random variable measuring the fraction of coordinates on which two independent samples from the uniform distribution on solutions agree.

Figure 6 depicts a cartoon of how the space of solutions in, say, random k-SAT or random graph q-coloring, changes (or is predicted to change) as the density of constraints increases. Three distinct thresholds are pictured. At low densities, solutions lie in one large component, connected under single variable changes. Then after the *shattering threshold* solutions break apart into exponentially many clusters of roughly equal exponential size, each separated by linear Hamming distance ($\Theta(n)$ variables must be changed to move between clusters). Next, at the *condensation threshold* a constant number of large clusters contain almost all solutions, while there are exponentially many smaller clusters. Finally, after the *satisfiability threshold*, no solutions remain.

The condensation threshold is critical for correlations and overlap. In the random graph coloring model, below the condensation threshold average correlations between vertices vanish as $n \to \infty$, and the overlap concentrates on $1/q$ (as it would if the graph were empty and we were sampling uniformly from $[q]^n$). Above the threshold, average correlations are bounded away from 0 and overlap concentrates on two points, $1/q$ and some $\eta > 1/q$, with the second corresponding to the case in which both samples come from the same dominant cluster. In the language of the cavity method, the model is *replica symmetric* below the condensation threshold and exhibits *replica symmetry breaking* above the threshold.

In terms of identifying the satisfiability threshold, condensation is important because it presents a fundamental obstacle to applying the second-moment method to the random variable Z counting the number of solutions: no amount of conditioning can reduce the variance sufficiently and so another approach is required (see the discussion in [39, 38, 37, 20]).

4.3 Proof of the satisfiability conjecture for large k

How can one locate the satisfiability threshold in light of condensation blocking the second-moment method? One very successful solution is to use the cavity method predictions to select a different random variable on which to perform the second-moment method. In particular, the predicted '1-RSB' behavior of both random k-SAT and random graph coloring suggests that while the uniform distribution over solutions exhibits replica symmetry breaking above the condensation threshold, the uniform distribution on *clusters* does not.

The innovation of Coja-Oghlan and Panagiotou in [39] was to design a random variable Z_β counting the number of solutions with βn 'blocked' variables that is a stand-in for counting the number of clusters of a given (exponential) size. The same authors apply this idea to random k-SAT in [37], additionally conditioning on typical vertex degrees. This kind of random variable was also used in [44, 43] to determine thresholds and limiting values of optimization problems in random regular models (in which each vertex or variable has the same edge or constraint degree).

In [45], Ding, Sly, and Sun used this cavity-method inspired second-moment argument along with conditioning on the typical empirical distribution of arbitrary depth neighborhoods of variables to resolve Conjecture 11 for k large enough.

Theorem 13 (Ding, Sly, Sun 11)**.** *There exists $k_0 > 0$ so that for each $k \geq k_0$ there exists c_k so that the following hold.*

- *If $m \leq (1-\varepsilon)c_k n$, then whp $F_k(n,m)$ is satisfiable.*
- *If $m \geq (1+\varepsilon)c_k n$, then whp $F_k(n,m)$ is unsatisfiable.*

The proof is a mathematical tour-de-force, employing several sophisticated tools (for both the upper and lower bounds) and performing very challenging probabilistic and combinatorial calculations. Notably, to prove Conjecture 11 in this case they establish the exact threshold as predicted by the 1-RSB cavity method.

One could hope to apply the same strategy to random graph coloring, but this seems hopelessly complex: having $q > 2$ possibilities for each variable makes everything more complicated, and the proof in [45] is already over 300 pages.

4.4 Algorithmic thresholds

What about the other question asked by Mitchell, Selman, and Levesque – is it true that random k-SAT instances near the satisfiability threshold are computationally intractable?

There was some hope among statistical physicists that *survey propagation*, a message-passing algorithm based on the 1-RSB predictions for random k-SAT and other models, might provide an efficient search algorithm for instances near the satisfiability threshold [28]. However (at least for large k) this was disproved [71, 29].

Instead, evidence of computational hardness based on the structure of the solution space has emerged. Achlioptas and Coja-Oghlan proved the shattering of the solution space in [6] and observed that the threshold for shattering approximately coincides with the density above which no efficient search algorithms are known (though see below for a caveat in linking the two). While it did not establish a direct link between solution space structure and algorithms, this paper was innovative both in its techniques (using the *planted model*, described more below) and in making a conceptual link between the two.

A more direct link with algorithms is the Overlap Gap Property (OGP) pioneered by Gamarnik and Sudan [60, 59]. The basic OGP states that there is an interval (a, b) so that whp over an instance of a random computational problem, no pair of solutions have their normalized overlap in (a, b). This can be proved using a first-moment method. Amazingly, this simple property then implies that entire classes of algorithms (local algorithms, low-degree algorithms) cannot find a solution whp. For random k-SAT the OGP perspective has been applied to determine approximately the low-degree algorithmic threshold of the problem [29].

4.5 Scaling windows

With Friedgut's theorem establishing a sharp threshold in the random k-SAT model, one could hope to say something quantitatively stronger about scaling windows in this and related models. Friedgut's theorem (or Bourgain's more abstract theorem) gives bounds on the k-SAT scaling window of the form $T_\varepsilon = O(p_c/\log\log n)$, just barely enough for a sharp threshold. Surprisingly, Abbe and Montanari [4] show that even a very mild improvement to this bound would resolve Conjecture 11

Proposition 14 ([4]). *If for any fixed $\delta > 0$, the scaling window for random k-SAT satisfies $T_\varepsilon = O(p_c/\log^{1+\delta} n)$, then Conjecture 11 holds.*

The statistical physics cavity method does not give much guidance on what to expect for scaling windows. In 2002, Wilson [127] used basic probabilistic arguments to prove generic polynomial lower bounds on the scaling window for random k-SAT and other models, thus disproving a number of conjectures from statistical physics. Good upper bounds for random k-SAT or random graph q-coloring are completely lacking currently.

In a different random CSP, however, scaling windows have recently been pinned down precisely. The *binary perceptron* [40] arose in the 1960's as a toy model of a neural network then attracted attention in statistical physics via the work of Gardner and others [62, 63, 91]. A symmetric variant was recently introduced in [18] and has some remarkably nice properties from a mathematical point of view. For one, the plain first- and second-moment methods (along with some concentration arguments) suffice to pin down precisely the satisfiability threshold in the model [18, 113, 3]. Closely related is the fact that the solution space looks very different than the cartoon in Figure 6: for all constraint densities below the satisfiability threshold, whp over the instance, almost all solutions are isolated, at linear Hamming distance from the nearest other solution. The corresponding cartoon would be a sprinkling of points.

The nature of the solution space in this model raises a lot of questions: if almost all solutions are isolated should the search problem be hard at all densities? In fact there are efficient search algorithms at low densities [87] even when almost all solutions are isolated; these algorithms in fact find solutions lying in very rare clusters (with maximum possible diameter) [2], as predicted by physicists working on machine learning problems [19]. This indicates that the cartoon in Figure 6 is not really relevant to algorithms: the cartoon depicts properties of typical solutions, while efficient algorithms may indeed find rare solutions. The structural properties like the OGP that apply to *all* solutions (or tuples of solutions) are more algorithmically relevant.

Finally, in a recent breakthrough Altschuler proved something remarkable about the scaling window: the scaling window, measured in number of constraints, is of width $O(\log n)$ (while the critical number of constraints is $O(n)$) [13]. Contrast this to what is known via abstract results like Friedgut's theorem which would give a bound of $O(n/\log\log n)$; and to what is known for models like random k-SAT and random graph coloring which is only what the abstract results give. Even more recently Sah and Sawhney determined the scaling window completely, showing it is of width $O(1)$ and giving the limiting probability of satisfiability inside the window [117].

5 Statistical inference and the stochastic block model

A fundamental statistical question can be phrased as 'Under what circumstances can a signal be recovered from a noisy observation of that signal?' Or - 'Can we distinguish a signal from noise?'

As the study of statistics has evolved in the age of fast computers and massive data sets, these same questions remain fundamental but the data sets of interest are now very high dimensional and the question of efficient computation becomes paramount. Just as we want to know whether 'typical' instances of 3-SAT are algorithmically tractable we also want to know if 'typical' statistical inference problems are algorithmically tractable.

A very useful framework for the rigorous study of these questions is the *teacher–student* framework (see e.g., [129]). In this framework, a teacher describes a generative probabilistic model of data to a student. The model takes as an input a 'ground truth' and adds to this noise of some form. The teacher chooses a ground truth from some known prior distribution, generates the data from the model, and presents the data to the student. The student's task is to recover the ground truth from the data, using knowledge of the generative model and prior distribution.

Perhaps the most studied teacher–student model in the *stochastic block model*, a toy model for the statistical and machine learning task of clustering: partitioning a set of data points into subsets with similar characteristics. The stochastic block model deals with a simple specialization of clustering: community detection in which the the data is a graph and the task is to partition the vertex set into subsets of vertices with similar connectivity structure; this could mean finding a partition in which most edges lie within parts, or finding a partition in which most edges cross the parts.

Formally, a symmetric version of the stochastic block model is defined as follows. Fix an integer $q \geq 2$, and $p_{\text{in}}, p_{\text{out}} \in [0,1]$, $p \neq q$, and let n denote the number of vertices of the graph.

- Choose a partition $\sigma \in [q]^n$ of $[n]$ into q parts uniformly at random.
- For each pair $i, j \in [n]$, $i \neq j$, include the edge (i, j) with probability p_{in} if $\sigma(i) = \sigma(j)$ and with probability p_{out} if $\sigma(i) \neq \sigma(j)$, all edges independently of the others.
- Call the resulting graph $\hat{G}$.

If $p_{\text{in}} > p_{\text{out}}$, then, on average, more edges will be drawn between vertices with the same label (we say the model is *assortative*); if $p_{\text{in}} > p_{\text{out}}$, the opposite is true (the model is *disassortative*). If $p_{\text{in}} = p_{\text{out}}$ the model is simply $G(n, p)$.

The inference task is to *recover* the partition σ given the graph $\hat{G}$. Recovery can be defined in different ways: exact recovery (up to a permutation of the q labels); recovery of almost all the labels up to permutation; or recovery of a partition σ' that (after a permutation) agrees with σ on $\frac{1}{q} + \varepsilon$ fraction of vertices, for some small $\varepsilon > 0$; that is, just a tiny bit better than random guessing. We focus here on the last notion, called *weak recovery.*

Though we have left the setting of monotone properties of the hypercube under μ_p, we have retained some of the essential features: this model has independent edges and an important monotonicity property. If we fix the ratio $\frac{p_{\text{in}}}{p_{\text{out}}} =: \alpha$ and set $p_{\text{in}} = p\alpha$ and $p_{\text{out}} = p$ then the weak recovery problem only becomes *easier* as p increases: we have more data (observed edges) from which to deduce the signal (the underlying partition).

5.1 Sharp thresholds for inference

Though the stochastic block model was defined independently in different fields in the 1980's [72, 31, 47], mathematical interest in the model exploded in 2011 when Decelle, Krzakala, Moore, and Zdeborová [41] used the cavity method to make a series of beautiful predictions about sharp thresholds in the model.

We focus on their conjectures for thresholds for weak recovery when $p_{\text{in}}, p_{\text{out}} = O(1/n)$ and so the random graph has constant average degree.

There is an *information theoretic threshold* at d_{inf} if

1. when $d < d_{\text{inf}}$, $p_{\text{in}} = \alpha \frac{d}{n}$, $p_{\text{out}} = \frac{d}{n}$, there is no algorithm (efficient or not) that whp finds a partition σ' that agrees with σ on $\frac{1}{q} + \varepsilon$ fraction of vertices (after a permutation of the parts) for any fixed $\varepsilon > 0$.

2. when $d > d_{\text{inf}}$, $p_{\text{in}} = \alpha\frac{d}{n}$, $p_{\text{out}} = \frac{d}{n}$, there is $\varepsilon > 0$ and an algorithm (perhaps inefficient) that whp finds a partition σ' that agrees with σ on $\frac{1}{q} + \varepsilon$ fraction of vertices (after a permutation of the parts).

There is an *algorithmic threshold* at d_{alg} if

1. when $d < d_{\text{alg}}$, $p_{\text{in}} = \alpha\frac{d}{n}$, $p_{\text{out}} = \frac{d}{n}$, there is no polynomial-time algorithm that whp finds a partition σ' that agrees with σ on $\frac{1}{q} + \varepsilon$ fraction of vertices (after a permutation of the parts) for any fixed $\varepsilon > 0$.
2. when $d > d_{\text{alg}}$, $p_{\text{in}} = \alpha\frac{d}{n}$, $p_{\text{out}} = \frac{d}{n}$, there is $\varepsilon > 0$ and a polynomial-time algorithm that whp finds a partition σ' that agrees with σ on $\frac{1}{q} + \varepsilon$ fraction of vertices (after a permutation of the parts).

For the case $q = 2, 3$ they conjectured that for any α, $d_{\text{inf}} = d_{\text{alg}}$, but for $q \geq 5$ they conjectured a gap: $d_{\text{inf}} < d_{\text{alg}}$ (now known as a statistical–computational gap, and the subject of great recent interest in statistics and computer science). The $q = 2$ conjecture was proved in celebrated works of Mossel–Neeman–Sly [105, 106] and Massoulié [95]

The positive side of their general conjecture about d_{alg}, that weak recovery is possible above the 'Kestum–Stigum threshold', was proved by Abbe and Sandon [5]. The information–theoretic threshold conjecture in the disassortative case ($\alpha < 1$) was proved in [36] by implementing a rigorous version of the cavity method, but the assortative case remains open in general. The surveys [104, 1] contain detailed accounts of further developments.

5.2 The planted model

To wrap things up, we discuss a concept with connections to all of the developments recounted in Sections 3, 4, and 5.

The stochastic block model is an example of a *planted model.* A solution (in this case a partition into q parts) is chosen, then a random instance is drawn consistent with this *planted solution.* The inference task above is to recover information about the planted solution given the random instance. An even easier computation task is to distinguish the planted model from the purely random model (in this case the Erdős-Rényi random graph $G(n, p)$).

It is straightforward to devise planted models for random k-SAT and random graph coloring: pick a solution σ uniformly at random and choose independent constraints or edges among those that are satisfied by σ.

Achlioptas and Coja-Oghlan [6] show that if the number of solutions in the random model is sufficiently concentrated, then high probability results about the planted solution in the planted model can be transferred to high probability statements about uniformly random solutions in the random model. This is the key to establishing results on the structure of the solution space (and is used in the symmetric perceptron results [113, 3] as well).

Returning again to random graph coloring, the planted model is exactly the extreme case ($\alpha = 0$) of the disassortative stochastic block model. In [20, 36, 35] the condensation threshold for random graph q-coloring is determined precisely, in terms of a solution to a variational problem arising from the cavity method. This bound is the best known lower bound on the q-colorability threshold for $q \geq 4$ (the best bound for $q = 3$ is in [8]). In [36], a precise connection between condensation thresholds and information theoretic thresholds is made: the two thresholds coincide for a large class of models exhibiting some symmetry and convexity; these include random graph coloring, the anti-ferromagnetic Potts model, k-NAE-SAT. They do not include random k-SAT (asymmetry) or the assortative stochastic block model (lacking the needed convexity). Overcoming these technical obstacles and being able to identify condensation thresholds in asymmetric models is a major challenge and could eventually lead to a generic approach to Conjecture 11. See related discussion in [128, 88].

To conclude, somewhat surprisingly the planted model is closely related to the developments recounted in Section 3 as well. Mossel, Niles-Weed, Sun and Zadik [107] recently showed that a well-chosen planted model and a second-moment argument can be used to prove the 'spread lemma', the key technical ingredient in the proof of improved sunflower bounds [14] and the fractional Kahn–Kalai conjecture [53].

6 Conclusions and questions

As we saw in Section 3, the fractional Kahn–Kalai theorem has sufficed for all known applications, but one can ask if there are applications that require Theorem 5.

Question 1. *Are there threshold applications that need the full power of Theorem 5 rather than the fractional version of Theorem 6? Or is there a constant K so that $q_f(\mathcal{F}) \leq Kq(\mathcal{F})$?*

The second is conjectured in the affirmative by Talagrand in [125].

As we have seen, one effective way of determining sharp thresholds and scaling windows is to prove a hitting time result, relating a complex property to a simple property.

Question 2. *Is there a general way to find first-order asymptotics of p_c for the property of $G(n,p)$ containing a subgraph H when there is not a natural hitting-time conjecture?*

A nice conjecture for one such property is by Kahn, Narayanan, and Park on the threshold for the existence of the square of a Hamilton cycles in $G(n,p)$; that is, the existence of a cyclic ordering of the n vertices so that there is an edge between every pair of neighbors and second neighbors.

Conjecture 15 (Kahn, Narayanan, and Park [80])**.** *For the property $\mathcal{F}$ of $G(n,p)$ containing the square of a Hamilton cycle,*

$$p_c(\mathcal{F}) = (1+o(1))\sqrt{\frac{e}{n}}\,.$$

In [80] they prove $p_c = \Theta(n^{-1/2})$ using a finer understanding of the spreadness of the uniform distribution on squares of Hamilton cycles; a direct application of Theorem 8 would have lost a $\log n$ factor. See also [122, 52] for generalizations.

Question 3. *Are there any general-purpose tools for bounding the width of the scaling window of a monotone property with a sharp threshold, beyond the bounds given by Friedgut's theorem? Can Conjecture 11 be proved in a generic way, without the need to precisely identify the threshold?*

Random geometric graphs

We conclude the survey by stating one of the author's favorite open problems on thresholds in random structures.

In this survey we have discussed the Erdős-Rényi random graph in depth and another model of a random graph in the stochastic block model; both models have the important property that edges are independent. There are many other important random graph models that do not have this property. Here we discuss one, the *random geometric graph* [65, 112], in which edges between randomly placed points are determined geometrically.

In particular, let $\mathbb{S}^{d-1}$ be the unit sphere in d-dimensions[1]. The random graph $G_d(n,p) = (V,E)$ is formed as follows:

- Let $V = [n]$
- Select n points $x_1, \ldots, x_n$ independently and uniformly from $\mathbb{S}^{d-1}$.
- Let $E = \{(i,j) : x_i \cdot x_j \geq \tau_p\}$ where τ_p is chosen so that $\Pr[x_i \cdot x_j \geq \tau_p] = p$.

In particular, like the Erdős-Rényi random graph $G(n,p)$, $G_d(n,p)$ is a random graph on n vertices with edge probability p; but now the edges are not independent.

Thresholds in a random geometric graphs have been studied extensively; see the textbook of Penrose [112] for results on connectivity, subgraph containment, existence of a giant component. More recently, following the influential paper of Devroye, György, Lugosi, and Udina [42], thresholds in $G_d(n,p)$ when $d = d(n) \to \infty$ have been studied, along with a different kind of threshold: for a given $p(n)$ for which values of $d(n)$ are the two random graphs $G(n,p)$ and $G_d(n,p)$ statistically distinguishable? [30, 93].

Just as in $G(n,p)$, a general study of monotone properties is possible in $G_d(n,p)$. In particular, let $\mathcal{F}$ be a non-trivial monotone property (where as always monotone pertains to adding edges). Then again the probability that $\mathcal{F}$ holds in $G_d(n,p)$ is a strictly increasing function of p, and so $p_c(\mathcal{F})$ can be uniquely defined, and all of the same questions from Section 1.1 can be asked in this setting too.

Perhaps surprisingly then, the analogue of Theorem 1, the Bollobas-Thomason theorem, that lays the foundation for the general study of thresholds of monotone properties in random graphs, is not known in general for random geometric graphs. In the special case of $d = 1$ (points on a circle), McColm proved that every monotone property has a threshold [96]. Moreover, Goel, Rai, and Krishnamachari [66] proved that the scaling window of every monotone property is bounded by a function (depending on d) that vanishes as $n \to \infty$, the analogue of the result of Friedgut and Kalai [56] for $G(n,p)$. The general statement, however, remains open.

Conjecture 16. *Every non-trivial monotone property has a threshold in $G_d(n,p)$.*

Acknowledgements

The author thanks Vishesh Jain, Tom Kelly, and Richard Montgomery for very helpful discussions while preparing this survey.

References

[1] E. Abbe. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, 18(1):6446–6531, 2017.

[2] E. Abbe, S. Li, and A. Sly. Binary perceptron: efficient algorithms can find solutions in a rare well-connected cluster. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 860–873, 2022.

[3] E. Abbe, S. Li, and A. Sly. Proof of the contiguity conjecture and lognormal limit for the symmetric perceptron. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 327–338. IEEE, 2022.

[1] We alternatively could choose the underlying space to be the d-dimensional unit torus, $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$, the results and questions below would still apply.

[4] E. Abbe and A. Montanari. On the concentration of the number of solutions of random satisfiability formulas. *Random Structures & Algorithms*, 45(3):362–382, 2014.

[5] E. Abbe and C. Sandon. Proof of the achievability conjectures for the general stochastic block model. *Communications on Pure and Applied Mathematics*, 71(7):1334–1406, 2018.

[6] D. Achlioptas and A. Coja-Oghlan. Algorithmic barriers from phase transitions. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 793–802. IEEE, 2008.

[7] D. Achlioptas and E. Friedgut. A sharp threshold for k-colorability. *Random Structures & Algorithms*, 14(1):63–70, 1999.

[8] D. Achlioptas and C. Moore. Almost all graphs with average degree 4 are 3-colorable. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 199–208, 2002.

[9] D. Achlioptas and C. Moore. The asymptotic order of the random k-SAT threshold. In *The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings.*, pages 779–788. IEEE, 2002.

[10] D. Achlioptas, A. Naor, and Y. Peres. Rigorous location of phase transitions in hard optimization problems. *Nature*, 435(7043):759–764, 2005.

[11] D. Achlioptas and Y. Peres. The threshold for random k-SAT is $2^k \ln 2 - O(k)$. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 223–231, 2003.

[12] N. Alon and J. H. Spencer. *The probabilistic method.* John Wiley & Sons, 2016.

[13] D. J. Altschuler. Critical window of the symmetric perceptron. *Electronic Journal of Probability*, 28:1–28, 2023.

[14] R. Alweiss, S. Lovett, K. Wu, and J. Zhang. Improved bounds for the sunflower lemma. *Annals of Mathematics*, 194(3):795–815, 2021.

[15] M. Anastos and D. Chakraborti. Robust Hamiltonicity in families of Dirac graphs. *arXiv preprint arXiv:2309.12607*, 2023.

[16] L. J. Andrén, C. J. Casselgren, and L.-D. Öhman. Avoiding arrays of odd order by Latin squares. *Combinatorics, Probability and Computing*, 22(2):184–212, 2013.

[17] D. Applegate, R. Bixby, V. Chvátal, and W. Cook. On the solution of traveling salesman problems. *Documenta Mathematica*, pages 645–656, 1998.

[18] B. Aubin, W. Perkins, and L. Zdeborová. Storage capacity in symmetric binary perceptrons. *Journal of Physics A: Mathematical and Theoretical*, 2019.

[19] C. Baldassi, A. Ingrosso, C. Lucibello, L. Saglietti, and R. Zecchina. Subdominant dense clusters allow for simple learning and high computational performance in neural networks with discrete synapses. *Physical Review Letters*, 115(12):128101, 2015.

[20] V. Bapst, A. Coja-Oghlan, S. Hetterich, F. Raßmann, and D. Vilenchik. The condensation phase transition in random graph coloring. *Communications in Mathematical Physics*, 341:543–606, 2016.

[21] M. Ben-Or and N. Linial. Collective coin flipping, robust voting schemes and minima of Banzhaf values. In *26th Annual Symposium on Foundations of Computer Science (sfcs 1985)*, pages 408–416. IEEE, 1985.

[22] A. Bogdanov, L. Trevisan, et al. Average-case complexity. *Foundations and Trends® in Theoretical Computer Science*, 2(1):1–106, 2006.

[23] B. Bollobás. *Random graphs.* Springer, 1998.

[24] B. Bollobás, C. Borgs, J. T. Chayes, J. H. Kim, and D. B. Wilson. The scaling window of the 2-SAT transition. *Random Structures & Algorithms*, 18(3):201–256, 2001.

[25] B. Bollobás and O. Riordan. *Percolation.* Cambridge University Press, 2006.

[26] B. Bollobás and A. Thomason. Random graphs of small order. In *North-Holland Mathematics Studies*, volume 118, pages 47–97. Elsevier, 1985.

[27] B. Bollobás and A. G. Thomason. Threshold functions. *Combinatorica*, 7(1):35–38, 1987.

[28] A. Braunstein, M. Mézard, and R. Zecchina. Survey propagation: An algorithm for satisfiability. *Random Structures & Algorithms*, 27(2):201–226, 2005.

[29] G. Bresler and B. Huang. The algorithmic phase transition of random k-SAT for low degree polynomials. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 298–309. IEEE, 2022.

[30] S. Bubeck, J. Ding, R. Eldan, and M. Z. Rácz. Testing for high-dimensional geometry in random graphs. *Random Structures & Algorithms*, 49(3):503–532, 2016.

[31] T. N. Bui, S. Chaudhuri, F. T. Leighton, and M. Sipser. Graph bisection algorithms with good average case behavior. *Combinatorica*, 7:171–191, 1987.

[32] C. J. Casselgren and R. Häggkvist. Coloring complete and complete bipartite graphs from random lists. *Graphs and Combinatorics*, 32:533–542, 2016.

[33] V. Chvátal, W. Cook, G. B. Dantzig, D. R. Fulkerson, and S. M. Johnson. Solution of a large-scale traveling-salesman problem. *50 Years of Integer Programming 1958-2008: From the Early Years to the State-of-the-Art*, pages 7–28, 2010.

[34] V. Chvátal and B. Reed. Mick gets some (the odds are on his side) (satisfiability). In *Proceedings., 33rd Annual Symposium on Foundations of Computer Science*, pages 620–627. IEEE Computer Society, 1992.

[35] A. Coja-Oghlan, C. Efthymiou, N. Jaafari, M. Kang, and T. Kapetanopoulos. Charting the replica symmetric phase. *Communications in Mathematical Physics*, 359(2):603–698, 2018.

[36] A. Coja-Oghlan, F. Krzakala, W. Perkins, and L. Zdeborová. Information-theoretic thresholds from the cavity method. *Advances in Mathematics*, 333:694–795, 2018.

[37] A. Coja-Oghlan and K. Panagiotou. The asymptotic k-sat threshold. *Advances in Mathematics*, 100(288):985–1068, 2016.

[38] A. Coja-Oghlan and L. Zdeborová. The condensation transition in random hypergraph 2-coloring. In *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*, pages 241–250. SIAM, 2012.

[39] A. Coja-Oglan and K. Panagiotou. Catching the k-NAESAT threshold. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 899–908, 2012.

[40] T. M. Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transactions on Electronic Computers*, (3):326–334, 1965.

[41] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical review E*, 84(6):066106, 2011.

[42] L. Devroye, A. György, G. Lugosi, and F. Udina. High-dimensional random geometric graphs and their clique number. *Electronic Journal of Probability*, 16:2481–2508, 2011.

[43] J. Ding, A. Sly, and N. Sun. Maximum independent sets on random regular graphs. *Acta Math*, 217:263–340, 2016.

[44] J. Ding, A. Sly, and N. Sun. Satisfiability threshold for random regular NAE-SAT. *Communications in Mathematical Physics*, 341(2), 2016.

[45] J. Ding, A. Sly, and N. Sun. Proof of the satisfiability conjecture for large k. *Annals of Mathematics*, 196(1):1–388, 2022.

[46] H. Duminil-Copin. Sharp threshold phenomena in statistical physics. *Japanese Journal of Mathematics*, 14:1–25, 2019.

[47] M. E. Dyer and A. M. Frieze. The solution of some random NP-hard problems in polynomial expected time. *Journal of Algorithms*, 10(4):451–489, 1989.

[48] P. Erdős. On the combinatorial problems which I would most like to see solved. *Combinatorica*, 1(1):25–42, 1981.

[49] P. Erdős and A. Rényi. On random graphs I. *Publ. math. debrecen*, 6(290-297):18, 1959.

[50] P. Erdős and A. Rényi. On the evolution of random graphs. *Publ. math. inst. hung. acad. sci*, 5(1):17–60, 1960.

[51] P. Erdős and A. Rényi. On the existence of a factor of degree one of a connected random graph. *Acta Math. Acad. Sci. Hungar*, 17:359–368, 1966.

[52] A. Espuny Diaz and Y. Person. Spanning f-cycles in random graphs. *Combinatorics, Probability and Computing*, 32(5):833–850, 2023.

[53] K. Frankston, J. Kahn, B. Narayanan, and J. Park. Thresholds versus fractional expectation-thresholds. *Annals of Mathematics*, 194(2):475–495, 2021.

[54] E. Friedgut. Sharp thresholds of graph properties, and the k-SAT problem. *Journal of the American mathematical Society*, 12(4):1017–1054. Appendix by J. Bourgain, 1999.

[55] E. Friedgut. Hunting for sharp thresholds. *Random Structures & Algorithms*, 26(1-2):37–51, 2005.

[56] E. Friedgut and G. Kalai. Every monotone graph property has a sharp threshold. *Proceedings of the American mathematical Society*, 124(10):2993–3002, 1996.

[57] E. Friedgut and M. Krivelevich. Sharp thresholds for certain Ramsey properties of random graphs. *Random Structures & Algorithms*, 17(1):1–19, 2000.

[58] A. Frieze and M. Karoński. *Introduction to random graphs.* Cambridge University Press, 2016.

[59] D. Gamarnik. The overlap gap property: A topological barrier to optimizing over random structures. *Proceedings of the National Academy of Sciences*, 118(41):e2108492118, 2021.

[60] D. Gamarnik and M. Sudan. Limits of local algorithms over sparse random graphs. *The Annals of Probability*, pages 2353–2376, 2017.

[61] V. Ganesh and M. Y. Vardi. On the unreasonable effectiveness of SAT solvers. In T. Roughgarden, editor, *Beyond the Worst-Case Analysis of Algorithms*, pages 547–566. Cambridge University Press, 2020.

[62] E. Gardner. Maximum storage capacity in neural networks. *Europhysics Letters*, 4(4):481, 1987.

[63] E. Gardner and B. Derrida. Optimal storage properties of neural network models. *Journal of Physics A: Mathematical and general*, 21(1):271, 1988.

[64] E. N. Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.

[65] E. N. Gilbert. Random plane networks. *Journal of the society for industrial and applied mathematics*, 9(4):533–543, 1961.

[66] A. Goel, S. Rai, and B. Krishnamachari. Monotone properties of random geometric graphs have sharp thresholds. *Annals of Applied Probability*, pages 2535–2552, 2005.

[67] G. Grimmett and G. Grimmett. *What is percolation?* Springer, 1999.

[68] H. Hatami. A structure theorem for Boolean functions with small total influences. *Annals of Mathematics*, pages 509–533, 2012.

[69] A. Heckel. Random triangles in random graphs. *Random Structures & Algorithms*, 59(4):616–621, 2021.

[70] A. Heckel, M. Kaufmann, N. Müller, and M. Pasch. The hitting time of clique factors. *arXiv preprint arXiv:2302.08340*, 2023.

[71] S. Hetterich. Analysing survey propagation guided decimation on random formulas. In *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2016.

[72] P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

[73] V. Jain and H. T. Pham. Optimal thresholds for Latin squares, Steiner triple systems, and edge colorings. *arXiv preprint arXiv:2212.06109*, 2022.

[74] S. Janson, T. Luczak, and A. Rucinski. *Random graphs*. John Wiley & Sons, 2011.

[75] A. Johansson. Triangle factors in random graphs, 2006.

[76] A. Johansson, J. Kahn, and V. Vu. Factors in random graphs. *Random Structures & Algorithms*, 33(1):1–28, 2008.

[77] J. Kahn. Hitting times for Shamir's problem. *Transactions of the American Mathematical Society*, 375(1):627–668, 2022.

[78] J. Kahn. Asymptotics for Shamir's problem. *Advances in Mathematics*, 422:109019, 2023.

[79] J. Kahn and G. Kalai. Thresholds and expectation thresholds. *Combinatorics, Probability and Computing*, 16(3):495–502, 2007.

[80] J. Kahn, B. Narayanan, and J. Park. The threshold for the square of a Hamilton cycle. *Proceedings of the American Mathematical Society*, 149(8):3201–3208, 2021.

[81] G. Kalai and S. Safra. Threshold phenomena and influence with some perspectives from mathematics, computer science, and economics. *Computational complexity and statistical physics*, page 25, 2006.

[82] D. Y. Kang, T. Kelly, D. Kühn, A. Methuku, and D. Osthus. Thresholds for Latin squares and Steiner triple systems: Bounds within a logarithmic factor. *Transactions of the American Mathematical Society*, 2023.

[83] P. Keevash. The optimal edge-colouring threshold. *arXiv preprint arXiv:2212.04397*, 2022.

[84] T. Kelly, A. Müyesser, and A. Pokrovskiy. Optimal spread for spanning subgraphs of Dirac hypergraphs. *arXiv preprint arXiv:2308.08535*, 2023.

[85] H. Kesten. *Percolation theory for mathematicians*, volume 2. Springer, 1982.

[86] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM Journal on Computing*, 37(1):319–357, 2007.

[87] J. H. Kim and J. R. Roche. Covering cubes by random half cubes, with applications to binary neural networks. *Journal of Computer and System Sciences*, 56(2):223–252, 1998.

[88] A. Kireeva and J.-C. Mourrat. Breakdown of a concavity property of mutual information for non-Gaussian channels. *arXiv preprint arXiv:2304.05129*, 2023.

[89] L. M. Kirousis, E. Kranakis, D. Krizanc, and Y. C. Stamatiou. Approximating the unsatisfiability threshold of random formulas. *Random Structures & Algorithms*, 12(3):253–269, 1998.

[90] J. Komlós and E. Szemerédi. Limit distribution for the existence of Hamiltonian cycles in a random graph. *Discrete mathematics*, 43(1):55–63, 1983.

[91] W. Krauth and M. Mézard. Storage capacity of memory networks with binary couplings. *Journal de Physique*, 50(20):3057–3066, 1989.

[92] F. Krzakała, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová. Gibbs states and the set of solutions of random constraint satisfaction problems. *Proceedings of the National Academy of Sciences*, 104(25):10318–10323, 2007.

[93] S. Liu, S. Mohanty, T. Schramm, and E. Yang. Testing thresholds for high-dimensional sparse random geometric graphs. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 672–677, 2022.

[94] Z. Luria and M. Simkin. On the threshold problem for Latin boxes. *Random Structures & Algorithms*, 55(4):926–949, 2019.

[95] L. Massoulié. Community detection thresholds and the weak ramanujan property. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 694–703, 2014.

[96] G. L. McColm. Threshold functions for random graphs on a line segment. *Combinatorics, Probability and Computing*, 13(3):373–387, 2004.

[97] S. Mertens, M. Mézard, and R. Zecchina. Threshold values of random k-SAT from the cavity method. *Random Structures & Algorithms*, 28(3):340–373, 2006.

[98] M. Mezard and A. Montanari. *Information, physics, and computation.* Oxford University Press, 2009.

[99] M. Mézard, G. Parisi, and M. A. Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company, 1987.

[100] M. Mézard, G. Parisi, and R. Zecchina. Analytic and algorithmic solution of random satisfiability problems. *Science*, 297(5582):812–815, 2002.

[101] R. Monasson and R. Zecchina. Statistical mechanics of the random k-satisfiability model. *Physical Review E*, 56(2):1357, 1997.

[102] A. Montanari, F. Ricci-Tersenghi, and G. Semerjian. Clusters of solutions and replica symmetry breaking in random k-satisfiability. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(04):P04004, 2008.

[103] R. Montgomery. Spanning trees in random graphs. *Advances in Mathematics*, 356:106793, 2019.

[104] C. Moore. The computer science and physics of community detection: Landscapes, phase transitions, and hardness. *Bulletin of EATCS*, 1(121), 2017.

[105] E. Mossel, J. Neeman, and A. Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162:431–461, 2015.

[106] E. Mossel, J. Neeman, and A. Sly. A proof of the block model threshold conjecture. *Combinatorica*, 38(3):665–708, 2018.

[107] E. Mossel, J. Niles-Weed, N. Sun, and I. Zadik. A second moment proof of the spread lemma. *arXiv preprint arXiv:2209.11347*, 2022.

[108] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, pages 21–30. IEEE, 2005.

[109] R. O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[110] J. Park. Threshold phenomena for random discrete structures. *arXiv preprint arXiv:2306.13823*, 2023.

[111] J. Park and H. Pham. A proof of the Kahn–Kalai conjecture. *Journal of the American Mathematical Society*, 2023.

[112] M. Penrose. *Random geometric graphs*, volume 5. Oxford University Press, 2003.

[113] W. Perkins and C. Xu. Frozen 1-rsb structure of the symmetric Ising perceptron. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1579–1588, 2021.

[114] C. Pohoata, L. Sauermann, and D. Zakharov. Sharp bounds for rainbow matchings in hypergraphs. *arXiv preprint arXiv:2212.07580*, 2022.

[115] A. Rao. Sunflowers: from soil to oil. *Bulletin of the American Mathematical Society*, 60(1):29–38, 2023.

[116] O. Riordan. Random cliques in random graphs and sharp thresholds for F-factors. *Random Structures & Algorithms*, 61(4):619–637, 2022.

[117] A. Sah and M. Sawhney. Distribution of the threshold for the symmetric perceptron. *arXiv preprint arXiv:2301.10701*, 2023.

[118] A. Sah, M. Sawhney, and M. Simkin. Threshold for Steiner triple systems. *Geometric and Functional Analysis*, pages 1–32, 2023.

[119] J. Schmidt and E. Shamir. A threshold for perfect matchings in random d-pure hypergraphs. *Discrete mathematics*, 45(2-3):287–295, 1983.

[120] B. Selman, D. G. Mitchell, and H. J. Levesque. Generating hard satisfiability problems. *Artificial intelligence*, 81(1-2):17–29, 1996.

[121] M. Simkin. $(n, k, k-1)$-Steiner systems in random hypergraphs. *arXiv preprint arXiv:1711.01975*, 2017.

[122] S. Spiro. A smoother notion of spread hypergraphs. *Combinatorics, Probability and Computing*, 32(5):809–818, 2023.

[123] M. Talagrand. Are all sets of positive measure essentially convex? In *Geometric Aspects of Functional Analysis: Israel Seminar (GAFA) 1992–94*, pages 295–310. Springer, 1995.

[124] M. Talagrand. Selector processes on classes of sets. *Probability theory and related fields*, 135:471–486, 2006.

[125] M. Talagrand. Are many small sets explicitly small? In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 13–36, 2010.

[126] A. Wigderson. P, NP and mathematics–a computational complexity perspective. In *Proceedings of the ICM*, volume 6, pages 665–712, 2006.

[127] D. B. Wilson. On the critical exponents of random k-SAT. *Random Structures & Algorithms*, 21(2):182–195, 2002.

[128] Q. Yu and Y. Polyanskiy. Ising model on locally tree-like graphs: Uniqueness of solutions to cavity equations. *IEEE Transactions on Information Theory*, 2023.

[129] L. Zdeborová and F. Krzakala. Statistical physics of inference: Thresholds and algorithms. *Advances in Physics*, 65(5):453–552, 2016.

HILBERT MEETS RAMANUJAN: SINGULARITY THEORY AND INTEGER PARTITIONS

HUSSEIN MOURTADA

Abstract. What can singularities of algebraic varieties say about the decompositions of a positive integer into a sum of positive integers ?

1. Introduction

In his first letter to Hardy, dated 16 January 1913 ([18], p. 29) Ramanujan stated the formulas

$$1+\cfrac{e^{-2\pi}}{1+\cfrac{e^{-4\pi}}{1+\cfrac{e^{-6\pi}}{\vdots}}}=\left(\sqrt{\frac{5+\sqrt{5}}{2}}-\frac{1+\sqrt{5}}{2}\right)e^{\frac{2\pi}{5}} \tag{1.1}$$

$$1+\cfrac{e^{-\pi}}{1+\cfrac{e^{-2\pi}}{1+\cfrac{e^{-3\pi}}{\vdots}}}=\left(\sqrt{\frac{5-\sqrt{5}}{2}}-\frac{1-\sqrt{5}}{2}\right)e^{\frac{\pi}{5}} \tag{1.2}$$

about which Hardy writes in the article "The Indian Mathematician Ramanujan" ([34], p. 144):

> *"[These formulas] defeated me completely. I had never seen anything in the least like them before. A single look at them is enough to show that they could only be written down by a mathematician of the highest class. They must be true because, if they were not true, no one would have had the imagination to invent them."*

This article is not exactly about these formulas, but about some identities which are at the heart of their proofs; this allows the author to enjoy writing them and probably the reader (who already knew them or not yet) to enjoy the scene. According to [17], the first proof of these formulas was given by Watson [56]; following [9], let us see how partitions, via the Rogers-Ramaunujan identities, play a fundamental role in the proof. Consider the $q-$difference equation

$$F(x)=F(xq)+xqF(xq^2), \tag{1.3}$$

where $q\in\mathbf{C}^*$ and $F(x)=\sum a_n(q)x^n$ is an analytic function satisfying $F(0)=1$.

2020 *Mathematics Subject Classification.* 14B05,13P10,11P84,11P81,05A17,05A19,13D40.

Let $c(x,q) := \frac{F(x)}{F(xq)}$; we have

$$c(x,q) = 1 + \frac{xq}{c(xq,q)} = 1 + \frac{xq}{1 + \frac{xq^2}{c(xq^2,q)}}.$$

Iterating this last identity, we find that the left member of the identity (1.1) is equal to $c(1, e^{-2\pi})$ and that the left member of the identity (1.2) is equal to $c(1, e^{-\pi})$. Now, if we plug $F(x) = \sum a_n(q)x^n$ in the equation (1.3), by comparing the coefficients of x^n on both sides, we get

$$a_n(q) = \frac{q^{n^2}}{(q)_n} = \frac{q^{n^2}}{(1-q)(1-q^2)\cdots(1-q^n)}.$$

This gives the left equalities in the following two identities:

$$F(1) = 1 + \sum_{n\geq 1} \frac{q^{n^2}}{(q)_n} \qquad = \prod_{i\equiv\ 1,4\ (\mathrm{mod}\ 5)} \frac{1}{1-q^i}. \tag{1.4}$$

$$F(q) = 1 + \sum_{n\geq 1} \frac{q^{n^2+n}}{(q)_n} \qquad = \prod_{i\equiv\ 2,3\ (\mathrm{mod}\ 5)} \frac{1}{1-q^i}. \tag{1.5}$$

The equalities on the right in (1.4) and (1.5) are two miracles, which are central in this article. They allow us to represent $c(1,q)$ as an infinite product and we may then deduce Ramanujan's continued fraction (1.1),(1.2) by an appeal to the theory of elliptic theta functions.

The "miracles" in (1.4) and (1.5) are called the Rogers-Ramanujan identities; it is magic how they appear "in many different domains": statistical mechanics, combinatorics and number theory, representation theory, probability theory and in Algebraic Geometry and Commutative Algebra; see [12, 16, 20, 22, 30, 33, 29]. Here we will concentrate on the Algebro-Geometric side of the story. But at first, since we have stated the Rogers-Ramanujan identities in terms of $q-$series, let us explain why these are partition identities.

Definition 1.1. A partition of a positive integer n is a decreasing sequence $\lambda = (\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r)$ such that $\lambda_1 + \cdots + \lambda_r = n$. The λ_i's are called the parts of λ and r is its size.

For instance, 4 has 5 partitions:

$$4, 3+1, 2+2, 2+1+1, 1+1+1+1. \tag{1.6}$$

The combinatorial version of Rogers-Ramanujan identities in terms of integer partitions is due to MacMahon [40] and Schur [53].

Theorem 1.2 (Rogers–Ramanujan identities, combinatorial version)**.** *Let n be a nonnegative integer and set $i \in \{1;2\}$. Denote by $T_{2,i}(n)$ the number of partitions of n such that the difference between two consecutive parts is at least 2 and the part 1 appears at most $i-1$ times. Let $E_{2,i}(n)$ be the number of partitions of n into parts congruent to $\pm 2 + i \mod 5$. Then we have*

$$T_{2,i}(n) = E_{2,i}(n).$$

For example, the partitions of 4 (see 1.6) which are counted by $T_{2,2}(4)$ are 4 and $3+1$; those which are counted by $E_{2,2}(4)$ are 4 and $1+1+1+1$. In particular we have $T_{2,2}(4) = E_{2,2}(4) = 2$, and the theorem says that this is the case for every positive integer n. The relation between the identities (1.4) and (1.5) and theorem 1.2 is that one can prove that the left member of (1.4) (respectively (1.5)) is the generating series of the sequence $T_{2,2}(n)$ (respectively $T_{2,1}(n)$) and it is not a difficult exercise to see that the right member of (1.4) (respectively (1.5)) is the generating series of the sequence $E_{2,2}(n)$ (respectively $E_{2,1}(n)$). Recall here that the generating series of a sequence of integer numbers $(a_n)_{n\in\mathbf{Z}_{\geq 0}}$ is by definition

$$\sum_{n\in\mathbf{Z}_{\geq 0}} a_n q^n.$$

The other important object (with integer partitions) for this article is the arc space, coming from algebraic geometry. Let $X \subset \mathbf{C}^e$ be an algebraic variety: *i.e.*, X is the zero locus in $\mathbf{C}^e$ of a set of polynomials in e variables with coefficients in the field $\mathbf{C}$ of complex numbers. The arc space X_∞ of X is a space which parametrizes the arcs (germs of formal curves) which are traced on X; so a point of X_∞ corresponds to an arc on X. As we will see, this is also an "algebraic variety" (or a scheme) which often is of infinite dimension. Arc spaces (and their finite dimensional approximations) play an important role in singularity theory, for instance via the Nash problem [49], motivic integration [25, 23], birational geometry [47] or equisingularity [45, 46, 37, 38].

This article tells, on the one hand, about a link between arc spaces and partition identities and on the other hand how this link allows one to discover and prove new partition identities. In the second section, we will introduce the arc space and the arc HP-series (the arc Hilbert-Poincaré series) which is an invariant of singularities of algebraic varieties; we will also show how to compute this series in some examples. The third section reveals the relation between the arc HP-series and Rogers-Ramanujan identities: differential algebra and Groebner basis theory play an important role here. The fourth section shows how one can guess and prove new partition identities using the link between arc spaces and integer partitions. The last section is about research directions which are related to the subject of this article but which have not been treated here. The article is meant to be self contained.

Aknowledgements. The author would like to thank several colleagues and friends with whom he discussed at a moment or another about the subject of this paper, in particular: P. Afsharijoo, L. Boccadifuoco, C. Bruscheck, S. Corm, S. Corteel, J. Dousse, M. Hajli, H. Hauser, F. Jouhet, M. Lejeune-Jalabert, Z. Mohsen, A. Rangachev, J. Schepers, B. Teissier.

2. The Arc Hilbert-Poincaré series

Let $\mathbf{C}$ be the field of complex numbers (any other field of characteristic zero would be good for this paper). Let $X \subset \mathbf{C}^3$ be an affine algebraic variety; the story is absolutely the same if we replace the 3 in $\mathbf{C}^3$ by an integer number e, modulo more notations; actually later we will consider examples where e (the embedding dimension) is $1, 2$ or 3. For the scope of this paper, we can consider X to be a hypersurface defined by a polynomial $f \in \mathcal{R} = \mathbf{C}[x,y,z]$, i.e.

$$X = \{(a_1,a_2,a_3) \in \mathbf{C}^3 \mid f(a_1,a_2,a_3) = 0\}.$$

Again, not much related to what we will tell changes if we replace the ideal generated by (f) by an ideal generated by a finite number of polynomials. We will also write

$$X = \mathrm{Spec}\frac{\mathbf{C}[x,y,z]}{(f)} = \mathrm{Spec}\frac{\mathcal{R}}{(f)}; \tag{2.1}$$

this latter notation emphasizes, as in modern Algebraic Geometry, on the fact the ring of polynomial functions defined on X with value in $\mathbb{C}$ is given by

$$\mathcal{O}_X = \frac{\mathbf{C}[x,y,z]}{(f)}.$$

For instance, the polynomial function defined by f (or any polynomial in the ideal (f) generated by f) is zero in $\mathcal{O}_X$; this meets the fact that for any $(a_1,a_2,a_3) \in X, f(a_1,a_2,a_3) = 0$. Moreover, the use of the notation Spec allows us to distinguish the variety (or scheme) defined by f from the one defined by f^2 (even though the underlying geometric object is the same); one can think of Spec$\mathcal{R}/(f^2)$ as a kind of thickening of Spec$\mathcal{R}/(f)$, since we have more polynomial functions on it, f for instance is not zero in $\mathcal{R}/(f^2)$.

An arc γ on X is defined by a string of power series

$$\gamma(t) = (x(t), y(t), z(t))$$

such that $f(\gamma(t)) = f(x(t),y(t),z(t)) = 0$. This latter equality says that the arc γ which was originally defined as an arc on $\mathbf{C}^3$ is an arc on X. Let us write

$$x(t) = \sum_{i\geq 0} x_i t^i, y(t) = \sum_{i\geq 0} y_i t^i, z(t) = \sum_{i\geq 0} z_i t^i, \tag{2.2}$$

and expand $f(\gamma(t)) =$

$$f(\sum_{i\geq 0} x_i t^i, \sum_{i\geq 0} y_i t^i, \sum_{i\geq 0} z_i t^i) = \sum_{j\geq 0} F_j(x_0,y_0,z_0,\ldots,x_j,y_j,z_j)t^j. \tag{2.3}$$

The data of an arc is then equivalent to the data of the coefficients

$$x_i, y_i, z_i, i \in \mathbf{Z}_{\geq 0},$$

which satisfy the equations $F_j(x_0,y_0,z_0,\ldots,x_j,y_j,z_j) = 0$ for every $j \in \mathbf{Z}_{\geq 0}$. Hence the arc space which is the space of all arcs on X is the algebraic variety X_∞ which is defined in an infinite dimensional affine space (whose coordinates are $x_i, y_i, z_i, i \in \mathbf{Z}_{\geq 0}$) by the polynomials $F_j, j \in \mathbf{Z}_{\geq 0}$. In other terms $X_\infty = \mathrm{Spec}\mathcal{O}_{X_\infty}$ where

$$\mathcal{O}_{X_\infty} = \frac{\mathbb{C}[x_i,y_i,z_i, i \in \mathbf{Z}_{\geq 0}]}{(F_j, j \in \mathbf{Z}_{\geq 0})}.$$

Giving the variables x_i, y_i and z_i the weight i, the polynomials F_j are weighted-homogeneous of degree j : Indeed, if we replace in the equation (2.3) the variables x_i, y_i, z_i by $\lambda^i x_i, \lambda^i y_i, \lambda^i z_i$, it becomes

$$f(\sum_{i\geq 0} \lambda^i x_i t^i, \sum_{i\geq 0} \lambda^i y_i t^i, \sum_{i\geq 0} \lambda^i z_i t^i) = \sum_{j\geq 0} F_j(\lambda^0 x_0, \lambda^0 y_0, \lambda^0 z_0, \ldots, \lambda^j x_j, \lambda^j y_j, \lambda^j z_j)t^j;$$

At the same time, noticing that $\lambda^i t^i = (\lambda t)^i$ we can write the equation as follows

$$f(\sum_{i\geq 0} x_i(\lambda t)^i, \sum_{i\geq 0} y_i(\lambda t)^i, \sum_{i\geq 0} z_i(\lambda t)^i) = \sum_{j\geq 0} F_j(x_0, y_0, z_0, \ldots, x_j, y_j, z_j)(\lambda t)^j;$$

hence, by collecting the coefficients of t^j in both forms of the equation, we have

$$F_j(\lambda^0 x_0, \lambda^0 y_0, \lambda^0 z_0, \ldots, \lambda^j x_j, \lambda^j y_j, \lambda^j z_j) = \lambda^j F_j(x_0, y_0, z_0, \ldots, x_j, y_j, z_j).$$

This gives $\mathcal{O}_{X_\infty}$ a structure of a grading ring, *i.e.*, we have a decomposition

$$\mathcal{O}_{X_\infty} = \bigoplus_{j\geq 0} \mathcal{O}_{X_\infty, j},$$

as a direct sum of subgroups $\mathcal{O}_{X_\infty,j}$ such that the product of an element in $\mathcal{O}_{X_\infty,j}$ with an element in $\mathcal{O}_{X_\infty,j'}$ is an element in $\mathcal{O}_{X_\infty,j+j'}$. The fact that the F_j are weighted-homogeneous is essential, otherwise, we can have two polynomials in $\mathbb{C}[x_i, y_i, z_i, i \in \mathbf{Z}_{\geq 0}]$ which are of different weights but whose images in $\mathcal{O}_{X_\infty}$ are equal. Still, $\mathcal{O}_{X_\infty}$ is not yet our favorite geometric object.

One notices that the data of a morphism of affine algebraic varieties $\phi : X \longrightarrow Y$ (a morphism which is defined by polynomial functions) is equivalent to the data of ring homomorphism $\phi^* : \mathcal{O}_Y \longrightarrow \mathcal{O}_X$ which to a polynomial function h on X, *i.e.* $h \in \mathcal{O}_Y$, associates $\phi^*(h) = \phi \circ h$. Hence the natural ring morphism given by $\mathcal{O}_X \longrightarrow \mathcal{O}_{X_\infty}$ which sends x, y, z respectively to x_0, y_0, z_0, defines a morphism

$$\psi_X : X_\infty \longrightarrow X.$$

We sometimes omit X in the notation ψ_X when X is clear from the context. This is the morphism which to an arc $\gamma(t) = (x(t), y(t), z(t)) \in X_\infty$ associates $\gamma(0) \in X$, the center of γ. Let us assume that the origin $O = (0, 0, 0) \in X$ (by a change of variable any point $x \in X$ can be considered to be the origin). We are interested in the fiber $\psi^{-1}(O)$ of ψ above O. We have $\psi^{-1}(O) = \text{Spec}\mathcal{A}_\infty$, where

$$\mathcal{A}_\infty = \frac{\mathbb{C}[x_i, y_i, z_i, i \in \mathbf{Z}_{\geq 1}]}{(f_j, j \in \mathbf{Z}_{\geq 1})};$$

the f_j's are obtained from the $F_j's$ by substituting x_0, y_0, z_0 by 0. Hence the f_i's are again weighted-homogeneous when giving x_i, y_i and $z_i, i \in \mathbf{Z}_{>0}$ the weight i and $\mathcal{A}_\infty$ inherits a graded structure $\mathcal{A}_\infty = \bigoplus_{j\geq 0} \mathcal{A}_{\infty,j}$. We are now ready to define our invariant, the arc HP-series.

Definition 2.1. The arc HP-series of X at O is defined by

$$\text{AHP}_{X,O}(q) := \sum_{j\in \mathbf{Z}_{\geq 0}} \dim_{\mathbf{C}} \mathcal{A}_{\infty,j} q^j.$$

Remark 2.2. The reason why we considered the arcs with center at a point (*i.e.* $\psi^{-1}(O)$) and not X_∞ is that the dimension over $\mathbf{C}$ of $\mathcal{O}_{X_\infty,0}$ (the homogeneous component of weight 0) is not finite ($\mathcal{O}_{X_\infty,0}$ is actually isomorphic to $\mathcal{O}_X$). Of course, one could consider the dimension over a generic point of an irreducible component of X, but in that case this series is much less interesting as it will be apparent later.

Example 2.3. The most basic example is the case where $X = \text{Spec}\mathbf{C}[y] = \mathbf{A}^1$ is the affine line and O is the origin. Following the explanation above, we have

$$\mathcal{A}_\infty = \mathbf{C}[y_j, j \in \mathbf{Z}_{>0}],$$

with the graded structure induced from giving y_j the weight j for every $j \in \mathbf{Z}_{>0}$. In particular, $\mathcal{A}_{\infty,j}$ is generated, as a vector space over $\mathbf{C}$ by the monomials

$$y_{j_1}y_{j_2}\cdots y_{j_r}$$

where $j_1 + j_2 + \cdots + j_r = j$ and where we can assume $j_1 \geq j_2 \geq \cdots \geq j_r$. These generators are in bijection with the partitions of j, simply by associating with the monomial $y_{j_1}y_{j_2}\cdots y_{j_r}$ the partition $j = j_1 + j_2 + \cdots + j_r$. Let us use the usual notation $p(j)$ to denote the number of partitions of j, where by convention $p(0) = 1$. We then have

$$\mathrm{AHP}_{\mathbf{A}^1,O}(q) = \sum_{j\in\mathbf{Z}_{\geq 0}} p(j)q^j = \prod_{j\in\mathbf{Z}_{>0}} \frac{1}{1-q^j}$$

The equality to the right in the above equation is a formula which is due to Euler; one can prove it simply by substituting in the product

$$\frac{1}{1-q^j} = (1 + q^j + q^{2j} + \cdots).$$

and then by expanding the product using the usual product of power series. A similar computation gives us

$$\mathrm{AHP}_{\mathbf{A}^d,O}(q) = \prod_{j\in\mathbf{Z}_{>0}} \frac{1}{(1-q^j)^d}$$

Example 2.3 actually allows us to compute the arc HP-series in many examples. To see that, let us use a slightly fancier definition of an arc γ on a variety X : an arc γ on X is a morphism

$$\gamma : \mathrm{Spec}\mathbf{C}[[t]] \longrightarrow X.$$

Here, $\mathbf{C}[[t]]$ is the ring of power series with coefficients in $\mathbf{C}$. One can see it as the completion of the local ring of the affine line at the origin, as follows: the local ring of the affine line $\mathbf{A}^1 = \mathrm{Spec}\mathbf{C}[t]$ is the ring $\mathbf{C}[t]_{(t)}$ which is obtained from $\mathbf{C}[t]$ by inverting all the polynomials $h \in \mathbf{C}[t]$ whose values at the origin O is not 0. This is a local ring with a unique maximal ideal (t); the powers $(t)^n$ of this maximal ideal gives a basis of a topology on $\mathbf{C}[t]_{(t)}$. The completion $\widehat{\mathbf{C}[t]_{(t)}}$ of $\mathbf{C}[t]_{(t)}$ with respect to this topology is $\mathbf{C}[[t]]$. One moral of the story is that $\mathrm{Spec}\mathbf{C}[[t]]$ can be thought as a formal neighborhood of the origin in the affine line $\mathbf{A}^1$, hence the intuition that the image of γ is a germ of a formal curve on X. Now, if we are interested only in the arcs centered at the origin $O \in X$, then such an arc γ corresponds to a morphism $\gamma : \mathcal{O}_{X,0} \longrightarrow \mathbf{C}[[t]]$. Since $\mathbf{C}[[t]]$ is complete the universal property of completeness tells us that γ factors through a morphism $\hat{\gamma} : \widehat{\mathcal{O}_{X,0}} \longrightarrow \mathbf{C}[[t]]$. So, if we assume that the variety X is non-singular at O (for a hypersurface this is equivalent to say that the partial derivatives at O are not all zero) then by Cohen structure theorem ([27], section 7.4), the completion $\widehat{\mathcal{O}_{X,0}}$ is isomorphic to $\mathbb{C}[[y_1, \ldots, y_d]]$, d being the dimension of X at O. It follows that the data of any γ is equivalent to the data of a morphism $\widehat{\gamma}* : \mathbf{C}[[y_1, \ldots, y_d]] \longrightarrow \mathbf{C}[[t]]$ and that $\Psi_X^{-1}(O)$ is isomorphic to $\Psi_{\mathbf{A}^d}^{-1}(O)$. We conclude from example 2.3 the computation of $AHP_{X,O}$; Moreover, one can show that if X is singular at O, $AHP_{X,O} \neq AHP_{\mathbf{A}^d,O}$.

Proposition 2.4. Let X be an algebraic variety and consider a point $O \in X$. We have that X is non-singular at O if and only if

$$\mathrm{AHP}_{X,O}(q)=\prod_{j\in\mathbf{Z}_{>0}}\frac{1}{(1-q^j)^d}$$

Proposition 2.4 tells us that the arc HP-series is an invariant of singularities since it detects singular points from non-singular ones. It also tells us that this series contains more information at singular closed points (the case that we are considering); for instance if X is irreducible, it is non-singular at its generic point and its arc HP-series (where dimensions are considered over the residue field of the generic point) is equal to the series in the proposition; see [44] section 9 for a comparison of the information contained in this invariant with more classical invariants of singularities.

In general, it is quite difficult to compute this series, essentially because the homological complexity of the jet schemes (the finite dimensional approximation of the arc space); for instance even for curves singularities [45], the jet schemes have a lot of irreducible components and they are very far from being equidimensional. We will actually use jet schemes to show how to compute HP-series for some "simple" singularities. For $m\in\mathbf{Z}_{\geq 0}$, an $m-$jet α on X is a morphism $\alpha:\mathrm{Spec}\mathbf{C}[t]/(t^{m+1})\longrightarrow X$. Following the same reasoning that we made to represent the arc space, we find that for an X like in (2.1) the $m-$th jet scheme of X is

$$X_m=\mathrm{Spec}\mathcal{O}_{X_m}=\mathrm{Spec}\frac{\mathbb{C}[x_i,y_i,z_i,i=0,\dots,m]}{(F_j,j=0,\dots,m)}.$$

Again, for the same reason as in the arc space case we have a natural morphism $\pi_m:X_m\longrightarrow X$ (again here, when it is clear from the context, we neglect the mentioning of X in the notation π_m) and we have $\pi^{-1}(O)=\mathrm{Spec}\mathcal{A}_m$ where

$$\mathcal{A}_m=\frac{\mathbb{C}[x_i,y_i,z_i,i=1,\dots,m]}{(f_1,\dots,f_m)}\tag{2.4}$$

We are ready to determine the arc HP-series for rational double point surface singularities. These latter are somehow ubiquitous in singularity theory and in algebraic geometry [26]. For instance these are the only locally complete intersection rational surface singularities. Embedded in $\mathbf{C}^3$, They are defined via the equations:

$$A_n,\ n\in\mathbb{N}:xy-z^{n+1}=0.$$
$$D_n,\ n\in\mathbb{N},n\geq 4:z^2-x(y^2+x^{n-2})=0.$$
$$E_6:z^2+y^3+x^4=0^{\cdot}$$
$$E_7:x^2+y^3+yz^3=0.$$
$$E_8:z^2+y^3+x^5=0^{\cdot}$$

The following theorem was first proved in [43]; we give here a proof following [22].

Theorem 2.5. *Let X be surface having a rational double point singularity at O. We have*

$$\mathrm{AHP}_{X,O}(q)=\frac{1}{(1-q)^3}\prod_{j\geq 2}\frac{1}{(1-q^j)^2}$$

Proof. We will prove that $\pi_m^{-1}(O) \subset \mathrm{Spec}\mathbf{C}[x_i, y_i, z_i, i = 0, \ldots, m]$ is a complete intersection (*i.e.*, the codimension of all its irreducible components is equal to the number of its defining equations); the result will then follow from [54], knowing that the weight of f_j is j for $j = 1, \ldots, m$, and that by definition the Hilbert-Poincaré series of $\mathcal{A}_m$ is equal to the Hilbert-Poincaré series of $\mathcal{A}_m$ modulo (q^{m+1}). Notice that embedded in $\mathbf{A}^3_m := \mathrm{Spec}\mathbf{C}[x_i, y_i, z_i, i = 0, \ldots, m]$, $\pi_m^{-1}(O)$ is defined by the ideal $(x_0, y_0, z_0, f_2, \ldots, f_m)$; *i.e.* by the equations given by all the generators of the ideal equal to 0 (f_1 does not appear here because it is equal to 0 modulo (x_0, y_0, z_0)). So the codimension of $\pi_m^{-1}(O)$ in $\mathbf{A}^3_m$ is smaller than or equal to $m+2$, the number of equations. We also know that $\pi_m^{-1}(X_{reg})$ (X_{reg} being the non-singular locus of X) is irreducible of codimension $m+1$: indeed, one can see that the equations $F_j, j = 0, \ldots, m$ are linear outside $(x, y, z) = (0, 0, 0)$. If the codimension $\pi_m^{-1}(O)$ is smaller than or equal to $m+1$, then $\pi_m^{-1}(O)$ cannot be included in the Zariski closure $\overline{\pi_m^{-1}(X_{reg})}$ of $\pi_m^{-1}(X_{reg})$ since its dimension is then larger than or equal to the dimension of $\overline{\pi_m^{-1}(X_{reg})}$; the other inclusion is also impossible since $\pi_m^{-1}(O) \subset \{x_0 = 0\}$ while $\overline{\pi_m^{-1}(X_{reg})}$ is not. We deduce that if the codimension $\pi_m^{-1}(O)$ is smaller than or equal to $m+1$, then X_m has at least two irreducible components; this contradicts the fact that X_m is irreducible since X is locally complete intersection with rational singularities [48]. Hence we deduce that the codimension of any irreducible component of $\pi_m^{-1}(O)$ in $\mathbf{A}^3_m$ is exactly equal to $m+2$, the number of the defining equations.

□

There are several other instances where the arc HP-series can be determined, see [44].

3. The arc HP-series and the Rogers-Ramanujan identities

The first Rogers-Ramanujan identity comes into the picture when considering one of the most elementary singularities, the one defined by (x^2) in the line. More precisely, from [21], we have:

Theorem 3.1. *Let* $X = \mathrm{Spec}\mathbf{C}[y]/(y^2)$. *We have*

$$\mathrm{AHP}_{X,O}(q) = \prod_{i \equiv 1,4 \mod 5} \frac{1}{1-q^i}$$

Moreover, let $\mathcal{B}_\infty := \mathcal{A}_\infty/(y_1)$. Again $\mathcal{B}_\infty$ inherits from $\mathcal{A}_\infty$ a graded structure $\mathcal{B}_\infty = \oplus_{j \in \mathbf{Z}_{\geq 0}} \mathcal{B}_{\infty,j}$ and one can consider its Hilbert-Poincaré series

$$HP_{\mathcal{B}_\infty}(q) = \sum_{j \in \mathbf{Z}_{\geq 0}} \dim_{\mathbf{C}} \mathcal{B}_{\infty,j} q^j.$$

Theorem 3.2. *The Hilbert-Poincaré series of* $\mathcal{B}_\infty$ *is*

$$HP_{\mathcal{B}_\infty}(q) = \prod_{i \equiv 2,3 \mod 5} \frac{1}{1-q^i}$$

We will now give a proof of theorem 3.1. This proof reduces the computations of the arc HP-series via the theory of Groebner basis to the computation of a Hilbert-Poincaré series of a quotient of an infinite dimensional polynomial ring by a monomial ideal. To apply this theory, we use the differential structure of the arc space. Let us say two words about these two concepts, one about each.

Groebner bases. The polynomial ring with one variable, $\mathbf{C}[x]$, is Euclidean, *i.e.* one can apply the Euclidean algorithm which says that given $g, h \in \mathbf{C}[x]$, there exists a unique couple of polynomials (q, r) such that

$$g = hq + r,$$

and $0 \leq \deg(r) < deg(h)$; we have that $r = 0$ if and only if h divides g. This algorithm is very useful to detect whether an element g belongs to an ideal $I \subset \mathbf{C}[x]$: indeed, again thanks to the Euclidean algorithm, $\mathbf{C}[x]$ is principal, $I = (h)$ is generated by one element and $g \in I$ if and only if h divides g, equivalently if $r = 0$. In the polynomial ring $R = \mathbf{C}[x_1, \ldots, x_n]$ with several variables, the ideals are finitely generated (Hilbert Basis theorem) but not principal in general; hence the need of a division algorithm which allows to divide a polynomial by several other polynomials. For that, the degree (which does not define a total ordering of monomials, many monomials may have the same degree) is replaced by a monomial ordering that we denote by $\prec$: this is a total ordering on the monomials of R which satisfies that for monomials m_1, m_2, m_3 if $m_1 \prec m_2$ then $m_1 m_3 \prec m_2 m_3$. We also demand for $\prec$ to be a well ordering, *i.e.*, any set of monomials of R has a smallest element with respect to $\prec$. Unlike the monomial of highest degree, the initial monomial $\text{in}_\prec(h)$ of $h \in R$ with respect to $\prec$ is unique, this is the largest monomial in h with respect to $\prec$. One can then divide a polynomial h by an ordered set of polynomials $(h_1, \ldots, h_s)$, and the result is:

$$h = h_1 q_1 + \ldots h_s q_s + r, \tag{3.1}$$

where $q_1, \ldots, q_s, r \in R$ and there is no monomial appearing in r which is divisible by any of $\text{in}_\prec(h_i), i = 1, \ldots, s$. In general r depends on the order of the $s-$tuple $(h_1, \ldots, h_s)$ and the condition that $r = 0$ is not necessary for f to belong to the ideal generated by $(h_1, \ldots, h_s)$: for instance (see example 5 page 68 of [24]), the division of $x_1 x_2^2 - x_1$ by $(x_1 x_2 - 1, x_2^2 - 1)$ with respect to the lexicographical ordering, where we assume $y \prec x$, is given by:

$$x_1 x_2^2 - x_1 = x_2(x_1 x_2 - 1) + 0.(x_2^2 - 1) + (-x_1 + x_2).$$

The remainder $r = -x_1 + x_2 \neq 0$ but

$$x_1 x_2^2 - x_1 = x_1(x_2^2 - 1) \in (x_1 x_2 - 1, x_2^2 - 1).$$

To fix this problem, one should consider a special (with respect to the chosen monomial order $\prec$) basis $(g_1, \ldots, g_l)$ of the ideal $I = (h_1, \ldots, h_s)$ which satisfies that the initial ideal $\text{in}_\prec(I) := (\text{in}_\prec(h); h \in I)$ is given by

$$\text{in}_\prec(I) := (\text{in}_\prec(g_1), \ldots, \text{in}_\prec(g_l)).$$

Such a basis is called a **Groebner basis** and it ensures when dividing by $(g_1, \ldots, g_l)$ the uniqueness of the remainder r. one notices that in the example above, $I = (h_1, h_2)$ where $h_1 = x_1 x_2 - 1$ and $h_2 = x_2^2 - 1$, the basis (h_1, h_2) is not a Greobner basis (with respect to the lexicographical ordering), indeed:

$$\text{S}(h_1, h_2) := x_2 h_1 - x h_2 = x - y \in I. \tag{3.2}$$

We have $\text{in}_\prec(x - y) = x \notin (\text{in}_\prec(h_1), \text{in}_\prec(h_2)) = (x_1 x_2, x_2^2)$. But the basis $(h_1, h_2, h_3 = x - y)$ is a Groebner basis. The S-polynomial defined in equation (3.2) is made so that one can eliminate the initials of both h_1 and h_2 and search for other elements in the ideal which give new initials that do not belong to the ideal generated by the initials of the generators of the input basis. As one can guess, the S-polynomial is the right tool in general to find a Groebner bases by applying it recursively to all the couple of elements in the basis and by adding them (actually the remainder of their divisions by the basis) to the basis when they are useful. The fact that such an algorithm (the Buchberger algorithm) stops, as for the division algorithm, is related to the property that the monomial order is a well ordering. Now one important thing for us, is that for a graded ring which is the quotient of a polynomial ring R by a (weighted-)homogeneous ideal, the Hilbert-Poincaré series satisfies (see e.g, theorem 5.2.6 in [32])

$$HP_{R/I}(q) = HP_{R/\mathrm{in}_{\prec}(I)}(q). \tag{3.3}$$

Note that the equality (3.3) is somehow natural, since by the discussion above, if we take a Greobner basis $I = (g_1, \dots, g_l)$, any element in R is congruent by the division algorithm by $(g_1, \dots, g_l)$ to a unique element r (the remainder) whose terms are not divisible by any $\mathrm{in}_{\prec}(g_i)$, *i.e*, by terms whose image in $R/\mathrm{in}_{\prec}(I)$ is a basis over $\mathbf{C}$. For more about Greobner bases, the reader can consult *e.g* [24, 27, 32].

Differential structure on the arc space. The ring $\mathcal{O}_{X_\infty}$, where X is an affine variety, has a structure of a differential ring. Let us stick to the example of X in section 2 and to the notations there. The ring of global functions on $\mathbf{A}^3_\infty$ is

$$\mathcal{O}_{\mathbf{A}^3_\infty} = \mathbb{C}[x_i, y_i, z_i, i \in \mathbf{Z}_{\geq 0}].$$

We have a derivation D on $\mathcal{O}_{\mathbf{A}^3_\infty}$ defined by $D(x_i) = x_{i+1}, D(y_i) = y_{i+1}, D(z_i) = z_{i+1}$ for $i \in \mathbf{Z}_{\geq 0}$. If we replace in the equation (2.2) the variables x_i by $x_i/i!$ (where $j!$ is the factorial of j), and similarly for y_i and z_i, we find

$$f(\gamma(t)) = \sum_{j \geq 0} \frac{\mathcal{F}_j(x_0, y_0, z_0, \dots, x_j, y_j, z_j)}{j!} t^j. \tag{3.4}$$

where $\mathcal{F}_0 = f(x_0, y_0, z_0)$ and $\mathcal{F}_j$ is recursively defined by the identity $D(\mathcal{F}_j) = \mathcal{F}_{j+1}$; equation (3.4) follows from the fact that both sides are additive and multiplicative in f and that this equality is obviously true for $f = x, y$ or z. We obtain hence the desired differential structure which is induced by the derivation D on $\mathcal{O}_{X_\infty}$; this is because the rings

$$\frac{\mathbb{C}[x_i, y_i, z_i, i \in \mathbf{Z}_{\geq 0}]}{(F_j, j \in \mathbf{Z}_{\geq 0})} \text{ and } \frac{\mathbb{C}[x_i, y_i, z_i, i \in \mathbf{Z}_{\geq 0}]}{(\mathcal{F}_j, j \in \mathbf{Z}_{\geq 0})}$$

are isomorphic, the isomorphism being given by the change of variables expressed above. Fore more about differential algebras see [36, 52].

Proof. (of theorem 3.1) The ring of $\mathcal{A}_\infty$ of global functions on $\psi^{-1}(O)$ is (modulo an isomorphism) given by

$$\mathcal{A}_\infty = \frac{\mathbf{C}[y_j, j \in \mathbf{Z}_{>0}]}{[y_1^2]}$$

where

$$\begin{aligned}[y_1^2] =&(y_1^2,\\ &2y_1y_2,\\ &2y_2^2+2y_1y_3,\\ &6y_2y_3+2y_1y_4,\\ &6y_3^2+8y_2y_4+2y_1y_5,\ldots)\\ =&(f_2,f_3,\ldots)\end{aligned}$$

is the differential ideal generated by y_1^2 and all its iterated derivatives by the derivation D. For a general singularity $O \in X$, where X is affine, the ring $\mathcal{A}_\infty$ needs not be differential even if $\mathcal{O}_{X_\infty}$ is; in our case, this is true because one can construct an (non-homogeneous) isomorphism between $\mathcal{A}_\infty$ and $\mathcal{O}_{X_\infty}$. Now, when writing the generators f_i of the ideal, we ordered their terms by the weight (in f_i they are all of the same weight i) and by considering as smaller the monomials which make use of larger indices: for instance, y_3^2 is larger than y_2y_4 which is larger than y_1y_6; this order that we denote by $\prec$ sounds to us natural from a geometric point of view since y_3^2 says something about the third neighborhood while y_2y_4 concerns the fourth neighborhood ; so we want to see y_3^2 before y_2y_4. Now if we want to find a Groebner basis, we need to study the S$-$polynomial of the various couples of generators among the $f_i's$. If the the initial monomials of of f_i, f_j are coprime, then their S$-$polynomial will not "give" new initials (see *e.g.* proposition 1, page 106 [24]). So we need to consider the S$-$polynomials for the couples $(f_{2n}, f_{2n+1}), (f_{2n+1}, f_{2n+2})$ and (f_{2n+1}, f_{2n+3}). Let us study the first case, the other being similar. We have

$$\mathrm{S}(f_2, f_3) = 2y_2f_2 - y_1f_3 = 0. \tag{3.5}$$

Now deriving (3.5) iteratively $3n+4$ times, we obtain the equation

$$\sum_{j=1}^{3n-1} c_jy_jf_{3n+1-j} = 0, c_j \in \mathbb{C}. \tag{3.6}$$

Using the Leibniz formula, we find

$$c_n = 2C_{3(n-1)}^{n-2} - C_{3(n-1)}^{n-1}$$
$$c_{n+1} = 2C_{3(n-1)}^{n-1} - C_{3(n-1)}^{n}$$

where $C_n^k := \binom{n}{k}$ denotes the binomial coefficient. Let α_{2n} and α_{2n+1} be respectively the coefficients of y_n^2 in f_{2n} and of y_ny_{n+1} in f_{2n+1}. Since $f_{2n} = D^{2n-2}(f_2)$ and $f_{2n+1} = D^{2n-1}(f_2)$, again using the Leibniz formula we see that the coefficients α_n and α_{n+1} satisfies

$$\alpha_{2n} = C_{2(n-1)}^{n-1}$$
$$\alpha_{2n+1} = C_{2n}^{n}$$

Now, noticing that $\alpha_{2n}c_{n+1} = -\alpha_{2n+1}c_n$ we can rewrite the equation (3.6) as

$$\mathrm{S}(f_{2n}, f_{2n+1}) = \sum_{j=1,\cdots,3n-1; j\neq n,n+1} c_jy_jf_{3n+1-j}.$$

This latter formula says that $\mathrm{S}(f_{2n}, f_{2n+1}), n \geq 2$ does not give new initials (reduces to 0 modulo the basis $(f_2, f_3, \ldots)$, using the terms of [24]). Similarly, we can prove that the S-polynomials of the couples (f_{2n+1}, f_{2n+2}) and $(f_{2n+1}, f_{2n+3}), n \geq 1$ reduce to 0 modulo the basis $(f_2, f_3, \ldots)$ and by theorem 6 page 108 in [24], we deduce that $(f_2, f_3, \ldots)$ is a Groebner basis. Hence, since

$$\mathrm{in}_{\prec}(f_{2n}) = \alpha_{2n} y_n^2 \text{ and } \mathrm{in}_{\prec}(f_{2n+1}) = \alpha_{2n+1} y_n y_{n+1}$$

we have

$$\mathrm{in}_{\prec}([y_1^2]) = (y_n^2, y_n y_{n+1}, n \geq 1).$$

From the equality (3.3), we deduce that the arc HP-series of X at O is equal to the Hilbert-Poincaré series of

$$L := \frac{\mathbf{C}[y_j, j \in \mathbf{Z}_{>0}]}{(y_n^2, y_n y_{n+1}, n \geq 1)},$$

graded by giving the weight j to y_j. The $j-$th (wighted)-homogeneous component L_j of L is generated by the monomials

$$y_{j_1} y_{j_2} \cdots y_{j_s}$$

where $j_1 + j_2 + \cdots + j_s = j$ and where $y_{j_1} y_{j_2} \cdots y_{j_s}$ is not divisible by any monomial of the type y_n^2 or $y_n y_{n+1}$, this is equivalent to say that difference between two consecutive parts of the associated partition $j_1 + j_2 + \cdots + j_s$ of j is at least 2. Using theorem 1.2 and the identity (1.4) we obtain the form of the arc HP-series in the statement of the theorem.

□

Remark 3.3. (1) The fact that we derived (3.5) $3n+4$ times is not a trick, it is just that we know the weight of $\mathrm{S}(f_{2n}, f_{2n+1})$ and we derived enough times to reach this weight; deriving once make the weight grow of 1.

(2) It worth noticing, that the fact that we considering a non-finitely generated ideal in the above proof is a source of simplification : indeed, if we consider the finitely generated ideals $(f_2, f_3, \ldots, f_m), m \in \mathbf{Z}_{\geq 3})$, then the given basis is no longer a Groebner basis with respect to the considered monomial ordering; it is only when we let m goes to infinity that we have the miracle that the basis is a Groebner basis. This can for instance be seen in the equation 3.6, where some f_i's for $i > 2n+1$ may intervene.

The proof of theorem 3.2 follows the same ideas and computations in the proof of theorem 3.1. The proof above inspires the following approach (see [21]) towards the Rogers-Ramanujan identities. We begin by introducing some notations: Let $I_d = (y_n^2, y_n y_{n+1}, n \geq d)$,

$$L^{(d)} := \frac{\mathbf{C}[y_j, j \in \mathbf{Z}_{\geq d}]}{I_d},$$

graded as above and $h(d) = HP_{L^{(d)}}$. We have an exact sequence

$$0 \longrightarrow \frac{\mathbf{C}[y_j, j \in \mathbf{Z}_{\geq d}]}{(I_d : y_d)}[-d] \longrightarrow \frac{\mathbf{C}[y_j, j \in \mathbf{Z}_{\geq d}]}{I_d} \longrightarrow \frac{\mathbf{C}[y_j, j \in \mathbf{Z}_{\geq d}]}{(I_d, y_d)} \longrightarrow 0 \tag{3.7}$$

where the first non-zero morphism is the multiplication by y_d; the symbol $[-d]$ means that the graded structure is shifted by $-d$, so that the elements of weight 0 after adding the $[-d]$ correspond to those of weight $-d$ if we drop the $[-d]$, and the column ideal

$$(I_d : y_d) = \{h \in \mathbf{C}[y_j, j \in \mathbf{Z}_{\geq d}] \mid h \cdot y_d \in I_d\}.$$

The shift guarantees that all the morphisms are homogeneous (they send an element of a given weight to an element of the same weight) and hence we have exact sequences at the level of the graded components seen as $\mathbf{C}$-vector spaces. Noticing that

$$\frac{\mathbf{C}[y_j, j \in \mathbf{Z}_{\geq d}]}{(I_d : y_d)} = L^{(d+2)},$$

the rank theorem gives the following

$$h(d) = h(d+1) + q^d \cdot h(d+2),$$

and one deduces (see[21])

Proposition 3.4. The power series $h(1)$ satisfies

$$h(1) = A_d \cdot h(d) + B_{d+1} \cdot h(d+1);$$

for $A_i, B_i \in k[[q]]$ fulfilling the following recursion

$$\begin{array}{rcl} A_d & = & A_{d-1} + B_d \\ B_{d+1} & = & A_{d-1} \cdot q^{d-1} \end{array}$$

with initial conditions $A_1 = A_2 = 1$ and $B_2 = 0$,$B_3 = q$.

Since $ord_q B_d \geq d-2$, both $\lim A_d$ and $\lim B_d$ exist (limits with respect to the $q-$adic topology as sequence of power series), and they satisfy

$$\lim B_d = 0 \text{ and } h(1) = \lim A_d.$$

The recursion from Proposition 3.4 can be simplified to $h(1) = \lim A_d$ where A_d fulfills

$$A_d = A_{d-1} + q^{d-2} \cdot A_{d-2} \tag{3.8}$$

with initial conditions $A_1 = A_2 = 1$.

This last recursion is well-known from [13]. Its limit is the infinite product

$$\prod_{i=1,4 \text{mod } 5}^{\infty} \frac{1}{1-q^i},$$

i.e., the generating series of the number of partitions with parts equal to 1 or 4 modulo 5. The construction above gives the generating series G_d defined in [13] an interpretation as a Hilbert-Poincaré series of the quotients $\mathbf{C}[y_j, j \in \mathbf{Z}_{\geq d}]/I_d$. This immediately implies that the series G_d are of the form $G_d = 1 + \sum_{j \geq i} G_{dj} q^j$, the empirical hypothesis of [13].

4. Other Partition identities inspired by this viewpoint

An extension of Rogers-Ramanujan identities. In section 3, we showed that the arc HP-series for one of the simplest singularities is equal to the generating series of the number of partitions appearing in the Rogers-Ramanujan identities. At the heart of the proof, we find a computation of a Groebner basis of the ideal $[y_1^2]$, the defining ideal of the space of arcs centered at $O \in \text{Spec}\mathbf{C}[y]/(y^2)$; this Groebner basis is differentially finite,*i.e.*, it is built from a finite number of elements (here only one) and all their derivatives. The monomial order considered in section 3 is somehow "geometric" (chosen for geometric reasons), but one may also consider another monomial ordering $<$ for which the initial ideal $\mathrm{in}_<([y_1^2])$ of $[y_1^2]$ may vary but the Hilbert-Ponicar series of the quotient

$$HP_{\mathrm{Spec}\mathbf{C}[y_j,j\in\mathbf{Z}_{>0}]/\mathrm{in}_<([y_1^2])}=HP_{\mathrm{Spec}\mathbf{C}[y_j,j\in\mathbf{Z}_{>0}]/([y_1^2])}$$

will not vary, by equality (3.3). In [6], a Groebner basis computation with respect to a weighted lexicographical ordering was considered; but (see theorem 2.2 in [6]) such a basis cannot be differentially finite. This made it very difficult to actually compute a Groebner basis of $[y_1^2]$ with respect to this order; still, from the computation one can guess (without a proof) that the leading ideal should be

$$(y_k y_{i_1}\dots y_{i_k},\ \text{where } k\leq i_1\leq\cdots\leq i_k). \tag{4.1}$$

By playing this game with $[y_i^2]$ for $i\in\mathbf{Z}_{>0}$ and using iteratively exact sequences which are similar to (3.7), on can prove the following (Theorem 1.7 [6]):

Theorem 4.1. *Let $n\geq k$ be a positive integer. The number of partitions of n with parts larger than or equal to k and size less than or equal to (the smallest part minus $k-1$) is equal to the number of partitions of n with parts larger than or equal to k and such that the difference between two consecutive parts is at least* 2.

For $k=1$, theorem 4.1 says that:

For a positive integer $n\geq 1$, the number of partitions of n with size less than or equal to the smallest part is equal to the number of partitions of n such that the difference between two consecutive parts is at least 2; *this yields another member of Rogers-Ramanujan identities.*

Let us call $G_{2,2}(n)$ the number of partitions of n with size less than or equal to the smallest part. The partitions of 4 (see (1.6) which are counted by $G_{2,2}(4)$ are

$$4 \text{ and } 2+2.$$

In particular we have $T_{2,2}(4)=E_{2,2}(4)=G_{2,2}(4)=2$ (see theorem 1.2 for the notations), and theorem 4.1, for $k=1$, asserts that the equality

$$T_{2,2}(n)=E_{2,2}(n)=G_{2,2}(n)$$

is true for every n.

Remark 4.2. Recently, in [8], using new methods from differential algebra, the authors proved that the ideal appearing in (4.1) is actually the initial ideal of $[y_1^2]$ with respect the weighted lexicographical order. Still, until now we do not have a Groebner basis with respect to this order.

In [2], using similar ideas to those who led to theorem 4.1, the author proved another exciting extension to Rogers-Ramanujan identities, in which the parity (even odd) of the parts of a partition plays an important role.

Gordon's identities and their extensions. In the last section, we kept somewhat hidden the fact that there is a great generalization of theorem 3.1, proved in [22].

Theorem 4.3. *Let $n \in \mathbf{Z}_{\geq 2}$. For $X = Spec\frac{\boldsymbol{K}[y]}{(y^n)}$,*

$$AHP_{X,0}(q) = \prod_{i \not\equiv\ 0,n,n+1\ mod(2n+1)} \frac{1}{1-q^i}.$$

The proof follows the same strategy of the proof of theorem 3.1 but the differential calculus is much more involved. Another famous family of identities intervenes in the proof, Gordon's identities [31].

Theorem 4.4 (Gordon's identities)**.** *Let r and i be integers such that $r \geq 2$ and $1 \leq i \leq r$. Let $\mathcal{T}_{r,i}$ be the set of partitions $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_s)$ where $\lambda_j - \lambda_{j+r-1} \geq 2$ for all j, and at most $i-1$ of the parts λ_j are equal to 1. Let $\mathcal{E}_{r,i}$ be the set of partitions whose parts are not congruent to $0, \pm i \mod (2r+1)$. Let n be a nonnegative integer, and let $T_{r,i}(n)$ (respectively $E_{r,i}(n)$) denote the number of partitions of n which belong to $\mathcal{T}_{r,i}$ (respectively $\mathcal{E}_{r,i}$). Then we have*

$$T_{r,i}(n) = E_{r,i}(n).$$

Using ideas similar to those of section 3, in [3, 1], the author gave an alternative approach to Gordon's identities and conjectured a great generalization of theorem 4.1. This conjecture was proved recently in [5, 4]. Let us give the statement of this theorem: Given an integer $r \geq 2$, for $1 \leq i \leq r$, define the (i, ℓ)-new part of $\lambda = (\lambda_1, \ldots, \lambda_s)$ as follows:

$$p_{i,\ell}(\lambda) := \begin{cases} \lambda_s & \text{if } \ell = 1, \\ \lambda_{s-\sum_{j=1}^{\ell-1} p_{i,j}(\lambda)} & \text{if } 2 \leq \ell \leq i, \\ \lambda_{s+\ell-i-\sum_{j=1}^{\ell-1} p_{i,j}(\lambda)} & \text{if } i < \ell \leq r-1, \end{cases}$$

where $\lambda_j = 0$ for $j \leq 0$, and if $p_{i,\ell}(\lambda) = 0$ then $p_{i,j}(\lambda) = 0$ for $j > \ell$. We denote the number of all non-zero (i, ℓ)-new parts of λ by $N_{r,i}(\lambda)$.

Theorem 4.5. *Let $r \geq 2$ and $1 \leq i \leq r$ be two integers. Let $\mathcal{C}_{r,i}$ be the set of partitions of the form $\lambda = (\lambda_1, \ldots, \lambda_s)$, such that at most $i-1$ of the parts are equal to 1 and either $N_{r,i}(\lambda) < r-1$, or $N_{r,i}(\lambda) = r-1$ and $s \leq \sum_{j=1}^{r-1} p_{i,j}(\lambda) - (r-i)$. Let n be a nonnegative integer, and denote by $C_{r,i}(n)$ the number of partitions of n which belong to $\mathcal{C}_{r,i}$. Then we have*

$$C_{r,i}(n) = T_{r,i}(n) = E_{r,i}(n).$$

The proof uses on the one hand another classification theorem of the partitions in $\mathcal{C}_{r,i}$ in terms of a new type of Durfee dissection (inspired by [11], this is a classification in terms of Ferrers diagrams): the proof of this interpretation uses simple commutative algebra (another purely combinatorial proof of the same result is also given); On the other hand it uses Bailey lattices [15, 55], a very powerful tool for calculus with $q-$ *series.*

Another singularity and its associated family of partition identity. To have a taste of what kind of partition identities can come out of singularities in higher dimensions we give below a family of partition identities which is associated with the singularity at the origin of

$$Y = \text{Spec}\frac{\mathbf{C}[x,y]}{(xy)}.$$

Let us first introduce partitions with 2 colors. Consider that we have two copies of each positive integer m, one is blue and the other is red; we denote these copies by m_b and m_r. We define an order between the colored integers by $m_b > m_r$ (hence $m_b + m_r$ and $m_r + m_b$ are the same); if $m > k$, we set $m_c > k_{c'}$ for $c, c' \in \{b, r\}$.

An integer partition of a positive integer number n is a decreasing sequence (with respect to the order that we have just defined) of positive integers of one color or an other

$$\lambda = (\lambda_{1,c_1} \geq \lambda_{2,c_2} \geq \ldots \geq \lambda_{l,c_l}),$$

where $c_i \in \{b, r\}$ and such that $\lambda_{1,c_1} + \lambda_{2,c_2} + \cdots + \lambda_{l,c_l} = n$. For example, the two colors integer partitions of 2 are:

$$\begin{gathered} 2_b \\ 2_r \\ 1_b + 1_b \\ 1_r + 1_r \\ 1_b + 1_r. \end{gathered}$$

Our singularity Y, sometimes called the node singularity, is somehow related (but still very different in nature) to the singularity $X = \mathrm{Spec}\mathbf{C}[x]/(x^2)$, which led to the Rogers-Ramanujan identities: one can "put them" in a family

$$F : \mathrm{Spec}\frac{\mathbf{C}[x, y, t]}{(x(x - ty))} \longrightarrow \mathrm{Spec}\mathbf{C}[t].$$

The fibers over $t \neq 0$ are isomorphic to Y and the fiber above $t = 0$ is $X \times \mathbf{A}^1$. This can perhaps explain the small similarity of theorem 4.1 with the following theorem from [6]:

Theorem 4.6. *Let j be a positive integer number. The number of partitions of n with 2 colors (say blue and red) of $j, \ldots, 2j - 1$ and only the red color of any other positive integer larger than $2j$ is equal to the number of partitions n whose parts are larger than j and of two colors and such that the number of blue parts is strictly less than its smallest red part (if this latter exists) minus $(j - 1)$.*

5. Omissions

Many other research directions are directly related to the subjects of this article. I can mention the relation between Neighborly partitions, monomial ideals, graphs and hypergraphs [42, 7]; this subject which is a direct continuation of the story told in this article has led recently in [50] to a new proof of Rogers-Ramanujan identities. I can mention the relation with vertex operators and Virasoro Algebras [14, 28, 39]. And the reader possibly sees interactions with other research directions.

References

1. Afsharijoo, P. Looking for a new member of Gordon's identites. *Annals of Combinatorics 25* (2021), 543–571.
2. Afsharijoo, P., Even-odd partition identities of Rogers-Ramanujan type. *Ramanujan J..* **57**, 969-979 (2022), https://doi.org/10.1007/s11139-021-00470-3
3. Afsharijoo, P., Looking for a new version of Gordon's identities : from algebraic geometry to combinatorics through partitions. *PhD thesis, Université Paris Cité.* (2019)
4. Afsharijoo, P., Dousse, J., Jouhet, F., and Mourtada, H. New companions to Gordon identities from commutative algebra. *Sém. Lothar. Combin. 86B* (2022), Art. 48, 12.

5. Afsharijoo, P., Dousse, J., Jouhet, F., and Mourtada, H. New companions to the Andrews-Gordon identities motivated by commutative algebra. *Adv. Math. 417* (2023), Paper No. 108946, 40.
6. Afsharijoo, P., and Mourtada, H. Partition identities and application to infinite-dimensional Gröbner basis and vice versa. In *Arc schemes and singularities.* World Sci. Publ., Hackensack, NJ, [2020] ©2020, pp. 145–161.
7. Afsharijoo, P., and Mourtada, H. Neighborly partitions, hypergraphs and Gordon's identities *arXiv:2309.13334* (2023).
8. Ait El Manssour, R., and Pogudin, G. Multiplicity structure of the arc space of a fat point *Algebra and Number theory*, to appear.
9. Andrews, G. E. *The theory of partitions.* Cambridge Mathematical Library. Cambridge University Press, 1998.
10. Andrews, G. E. An analytic generalization of the Rogers-Ramanujan identities for odd moduli. *Proc. Nat. Acad. Sci. U.S.A. 71* (1974), 4082–4085.
11. Andrews, G., Partitions and Durfee dissection. *Amer. J. Math.*. **101**, 735-742 (1979), https://doi.org/10.2307/2373804
12. Andrews, G. E., *q-series: their development and application in analysis, number theory, combinatorics, physics and computer algebra*, CBMS Regional Conference Series in Mathematics, **66**, AMS, Providence, 1986.
13. Andrews, G. & Baxter, R., A motivated proof of the Rogers-Ramanujan identities. *Amer. Math. Monthly.* **96**, 401-409 (1989), https://doi.org/10.2307/2325145
14. Andrews, G., Ekeren, J. & Heluani, R., The singular support of the Ising model. *Int. Math. Res. Not. IMRN.*, 8800-8831 (2023), https://doi.org/10.1093/imrn/rnab328
15. Bailey W.N., *Identities of the Rogers-Ramanujan type*, Proc. London Math. Soc. (2) **50** (1949), 1–10.
16. R. J. Baxter, *Hard hexagons: exact solution*, J. Phys. A **13** (1980), no. 3, L6–L70.
17. Berndt, B. C., Chan, H. H., Huang, S.-S., Kang, S.-Y., Sohn, J., and Son, S. H. The Rogers-Ramanujan continued fraction. vol. 105. 1999, pp. 9–24. Continued fractions and geometric function theory (CONFUN) (Trondheim, 1997).
18. Berndt, B. C., Rankin, R.A. Ramanujan: Letters and Commentary. Amer. Math. Soc., Providence, 1995. London Math. Soc., London, 1995.
19. Bressoud, D., *A generalization of the Rogers–Ramanujan identities for all moduli*, J. Comb. Th. A **27** (1979), 64–68.
20. Bressoud, D., Ismail, M. and Stanton D., *Change of Base in Bailey Pairs*, The Ramanujan J. **4** (2000), 435–453.
21. Bruschek, C., Mourtada, H., and Schepers, J. Arc spaces and Rogers-Ramanujan identities. In *23rd International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC 2011)*, vol. AO of *Discrete Math. Theor. Comput. Sci. Proc.* Assoc. Discrete Math. Theor. Comput. Sci., Nancy, 2011, pp. 211–220.
22. Bruschek, C., Mourtada, H., and Schepers, J. Arc spaces and the Rogers-Ramanujan identities. *Ramanujan J. 30*, 1 (2013), 9–38.
23. Chambert-Loir, A., Nicaise, J. & Sebag, J., Motivic integration. (Birkhuser/Springer, New York,2018), https://doi.org/10.1007/978-1-4939-7887-8
24. Cox, D., Little, J. & O'Shea, D., Ideals, varieties, and algorithms. (Springer, New York,2015), https://doi.org/10.1007/978-0-387-35651-8
25. Denef, J. & Loeser, F., Germs of arcs on singular algebraic varieties and motivic integration. *Invent. Math.*. **135**, 201-232 (1999), https://doi.org/10.1007/s002220050284
26. Durfee, A., Fifteen characterizations of rational double points and simple critical points. *Enseign. Math. (2).* **25**, 131-163 (1979)
27. Eisenbud, D., Commutative algebra. (Springer-Verlag, New York,1995), https://doi.org/10.1007/978-1-4612-5350-1
28. Feigin, B. & Frenkel, E., Coinvariants of nilpotent subalgebras of the Virasoro algebra and partition identities. *I. M. Gelfand Seminar.* **16, Part 1** pp. 139-148 (1993)
29. Fulman, J., A probabilistic proof of the Rogers-Ramanujan identities. *Bull. London Math. Soc.*. **33**, 397-407 (2001), https://doi.org/10.1017/S0024609301008207
30. Garrett, K. Ismail, M. E. H. and Stanton D., *Variants of the Rogers-Ramanujan Identities*, Adv. in Appl. Math. **23** (1999), 274–299.

31. Gordon, B. A combinatorial generalizatin of the Rogers-Ramanujan identities. *Amer. J. Math. 83* (1961), 393–399.
32. Greuel, G. & Pfister, G. A, A Singular introduction to commutative algebra. (Springer, Berlin,2008), With contributions by Olaf Bachmann, Christoph Lossen and Hans Schnemann
33. M. J. Griffin, K. Ono, and S. O. Warnaar, *Framework of Rogers-Ramanujan identities and their arithmetic properties*, Duke Math. J. **165** (2016), no. 8, 147–1527.
34. Hardy, G. H. The Indian Mathematician Ramanujan. *Amer. Math. Monthly 44*, 3 (1937), 137–155.
35. Herzog, J., and Hibi, T. *Monomial ideals*, vol. 260 of *Graduate Texts in Mathematics.* Springer-Verlag London, Ltd., London, 2011.
36. Kolchin, E., Differential algebra and algebraic groups. (Academic Press, New York-London,1973)
37. Leyton-Alvarez, M., Deforming spaces of m-jets of hypersurfaces singularities. *J. Algebra.* **508** pp. 81-97 (2018), https://doi.org/10.1016/j.jalgebra.2018.04.014
38. Leyton-Alvarez, M., Mourtada, H. & Spivakovsky, M., Newton non-degenerate $\mu-$constant deformations admit simultaneous embedded resolutions. *Compos. Math..* **158**, 1268-1297 (2022), https://doi.org/10.1112/s0010437x22007576
39. Li, H., Some remarks on associated varieties of vertex operator superalgebras. *Eur. J. Math..* **7**, 1689-1728 (2021), https://doi.org/10.1007/s40879-021-00477-6
40. P. A. MacMahon, *Combinatory Analysis*, Volume 2, Cambridge University Press, Cambridge, 1916.
41. Miller, E., and Sturmfels, B. Combinatorial Commutative Algebra. *Graduate Texts in Mathematics (Springer-Verlag, New York) 227* (2005).
42. Mohsen, Z., and Mourtada, H. Neighborly partitions and the numerators of Rogers-Ramanujan identities. *Int. J. Number Theory 19*, 4 (2023), 859–872.
43. Mourtada, H. Jet schemes of rational double point singularities. In *Valuation theory in interaction*, EMS Ser. Congr. Rep. Eur. Math. Soc., Zürich, 2014, pp. 373–388.
44. Mourtada, H. Jet schemes and their applications in singularities, toric resolutions and integer partitions. In *Handbook of Geometry and Topology of Singularities IV*. Springer, Cham, 2023.
45. Mourtada, H., Jet schemes of complex plane branches and equisingularity. *Ann. Inst. Fourier (Grenoble).* **61**, 2313-2336 (2011), https://doi.org/10.5802/aif.2675
46. Mourtada, H., Veys, W. & Vos, L., The motivic Igusa zeta function of a space monomial curve with a plane semigroup. *Adv. Geom..* **21**, 417-442 (2021), https://doi.org/10.1515/advgeom-2021-0009
47. Mustata, M., Singularities of pairs via jet schemes. *J. Amer. Math. Soc..* **15**, 599-615 (2002), https://doi.org/10.1090/S0894-0347-02-00391-0
48. Mustata, M., Jet schemes of locally complete intersection canonical singularities. *Invent. Math..* **145**, 397-424 (2001), https://doi.org/10.1007/s002220100152, With an appendix by David Eisenbud and Edward Frenkel
49. Nash, J. Arc structure of singularities. *Duke Math. J..* **81** pp. 31-38 (1995), https://doi.org/10.1215/S0012-7094-95-08103-4, A celebration of John F. Nash, Jr.
50. O'Hara, K., and Stanton, D. Notes for neighborly partitions. *arXiv 2307.06786* (2023).
51. Peeva, I. Graded syzygies. *Algebra and Applications (Springer-Verlag, London) 14* (2011).
52. Ritt, J., Differential Algebra. (American Mathematical Society, New York,1950)
53. Schur, I, *Ein Beitrag zur additiven Zahlentheorie und zur Theorie der Kettenbrchen.* In: Gesammelte Abhandlungen. Band II. Springer-Verlag, Berlin-New York, (1973)
54. Stanley, R., Hilbert functions of graded algebras. *Advances In Math..* **28**, 57-83 (1978), https://doi.org/10.1016/0001-8708(78)90045-2
55. Warnaar, S. O., 50 *Years of Baileys lemma*, in Algebraic Combinatorics and Applications, pp. 333–347, A. Betten et al. eds. (Springer, Berlin, 2001).
56. Watson, G. N. Theorems Stated by Ramanujan (IX) : Two Continued Fractions. *J. London Math. Soc. 4*, 3 (1929), 231–237.

Université Paris Cité and Sorbonne Université, CNRS, IMJ-PRG, F-75013 Paris, France.
Email address: hussein.mourtada@imj-prg.fr

UNIFORMITY WHEN ARITHMETIC MEETS GEOMETRY

HOLLY KRIEGER

ABSTRACT. In 1983, Faltings proved Mordell's famous 1922 conjecture relating arithmetic to geometry: that for a polynomial equation $f(x,y) = 0$, if the topology of the set of solutions over the complex numbers is sufficiently complicated, then the set of solutions with rational numbers is finite.

Once we know a set is finite, natural questions arise: can we compute this finite set? How large can it be? What input data does its size depend on?

Recent works of Dimitrov-Gao-Habegger and Kühne have provided a strong and striking answer to this last question. More generally, they proved what is known as the uniform Mordell-Lang conjecture for curves embedded into their Jacobians, answering a question posed by Mazur in 1986. Here the word 'uniform' roughly means that the size bound depends only on the genus of the curve (a measure of the topological complexity of the solutions over $\mathbb{C}$) as well as a notion of algebraic complexity for the field in which we search for points satisfying the polynomial equation.

In this article, we will build up the tools to understand the uniform Mordell-Lang conjecture and survey the methods of Dimitrov-Gao-Habegger and Kühne that led to its resolution.

1. INTRODUCTION

From the fundamental starting point of the integers $\mathbb{Z}$ and their operations of addition and multiplication, we can build the rational numbers $\mathbb{Q}$ and polynomial functions $f(x_1, \ldots, x_n)$ with integer or rational coefficients. Diophantine problems are those that ask: what can we say about the integer or rational solutions to an equation of the form $f(x_1, \ldots, x_n) = 0$, or a system of such equations?

These questions are foundational, broad, and deep. From the non-existence of rational solutions to the equation $x^2 - 2 = 0$ we discover the irrational numbers. From integral solutions to equations of the form $x^2 - ny^2 = 1$ we encounter the continued fraction expansion of $\sqrt{n}$ and the more general theory of approximation of real numbers by rational numbers. Geometrically, the infinitude of rational solutions to the equation $x^2 + y^2 = 1$ yields the infinite collection of Pythagorean triples, and studying rational solutions of the equation $y^2 = x^3 - n^2x$ tells us whether a positive rational number n is the area of a right triangle with rational side lengths (this is known as the congruent number problem).

We will focus on the study of the set of rational solutions to equations of the form $f(x,y) = 0$, where f is a polynomial with integral (or rational, or algebraic) coefficients. The algebraic curve which is the set of complex solutions to $f(x,y) = 0$ (more precisely, a compact smooth model of the curve) is topologically a sphere (genus 0) or a sphere with g handles, in which case we say the curve has genus g.

2020 *Mathematics Subject Classification.* 14H40,11G50,11G10.

We are interested in rational points on curves here not only as a first case (dimension 1 being the simplest possible) but also because there is a fantastic and profitable trio of perspectives when investigating rational points on curves of genus $g \geq 1$: the arithmetic structure of the rational points, the analytic structure carried by the curve as a Riemann surface, and the algebraic structure obtained by embedding the curve in its Jacobian, which is both a group and an algebraic variety. The interplay between these three structures provides us with extraordinarily powerful tools to study the arithmetic and geometry of algebraic curves.

In this vein, Mordell conjectured [34], and Faltings proved [12], that a curve of genus $g \geq 2$ has only finitely many rational points; that is, the complex topology of the curve controls the arithmetic. The topology also controls the algebra: Raynaud's proof [36] of the Manin-Mumford conjecture established that a curve contains finitely many torsion points (for the group structure when embedded in its Jacobian) if the genus of the curve is greater than 1. These theorems are unified by a broader Mordell-Lang conjecture, which was settled by the work of Hindry [22] and Faltings [13].

Even before the finiteness of rational or torsion points on curves of genus $g \geq 2$ was established by Faltings and Raynaud, Mazur asked [31]

> is it reasonable to hope that the cardinality [...of the set of rational or torsion points...] admits an upper bound that depends only on the genus...?

An affirmative answer to this question became known as the *uniform Mordell-Lang conjecture*, which was recently resolved by the work of Dimitrov-Gao-Habegger [11] and Kühne [26]. The innovation leading to their solution relies on a characterization from each perspective - analytic, arithmetic, and geometric - of when a subvariety of a family of abelian varieties is *non-degenerate*, a property which allows one to control the number of rational points, torsion points, and other 'special' points on the subvariety. The theory of these non-degenerate subvarieties and its application to questions of points on curves in their Jacobians was built over a decade's work by all four authors [21] [19] [16] [15] [26] [10] [11].

Their breakthrough remarkably draws on tools from several very different approaches used for studying rational or torsion points on a curve: the theory of heights and divisors which allowed Vojta [41] to reprove Faltings theorem, the equidistribution theory used by Szpiro-Ullmo-Zhang [40] [43] to provide an alternative proof of the Manin-Mumford conjecture, and the application of abelian logarithms to study unlikely intersections in families initiated by Masser-Zannier [29]. Since their work, some of their tools used have been extended substantially; for example, work of Yuan-Zhang [42], DeMarco-Mavraki [9], and Gauthier-Vigny [20] pursue dynamical generalizations of the uniform Mordell-Lang conjecture by developing the theory of non-degenerate subvarieties in that setting, and Gao-Ge-Kühne [18] have extended the work of Dimitrov-Gao-Habegger and Kühne to prove a uniform Mordell-Lang type result for points on general subvarieties of abelian varieties.

In this article we will explain the uniform Mordell-Lang conjecture, survey a small sample of the preliminary work, and explain the basic approach of Dimitrov-Gao-Habegger and Kühne which resolved the conjecture. This paper is meant as a starting point for a reader unfamiliar with the theory; for further reading (or the more advanced reader) I strongly recommend the wonderful survey of Gao [17], to which much of the technical narrative of this survey is indebted.

Acknowledgements. Thanks to Laura DeMarco, Philipp Habegger, Ziyang Gao, Mattias Jonsson, Michael Stoll, and Jack Thorne for helpful conversations and correspondence related to the preparation of this article.

2. Rational points on curves: arithmetic and geometry

More than a century ago, Mordell [34] noted the depth and difficulty of the study of Diophantine problems.

> Mathematicians have been familiar with very few questions for so long a period with so little accomplished in the way of general results, as that of finding the rational solutions...

In this section we will study how the set of rational solutions to a polynomial equation of the form $f(x, y) = 0$, where f is a polynomial with rational coefficients, relates to the geometry of its complex solution set.

The polynomial f can be described by the information of its coefficients and the bidegrees of its monomial terms, and so we expect the set of rational solutions to depend in some way on this input information. However, this dependence can be subtle. For example, the equation $x^2+y^2 = 1$ has infinitely many rational solutions, arising from integer solutions to the Pythagorean equation $a^2 + b^2 = c^2$. However, the similar equation $x^2 + y^2 = -1$ has no rational (or even real) solutions, and as Fermat proved (as found in his notes after his death), the equation $x^4 + y^4 = 1$ can have no rational solutions when $xy \neq 0$.

While the degree of a polynomial is a natural quantity to work with, it is not quite the correct notion to be useful for studying solution sets. To see this, note that the set of solutions to $x^2 + y^2 = 1$ can be identified with the set of solutions to $x^{2024} + y^2 = x^{2022}$ via the map $(x, y) \mapsto (x, x^{1011}y)$; since this map is invertible away from the points $(0, \pm 1)$, the two solutions sets are bijectively identified, up to a finite number of points. The correct geometric perspective is that of the *genus* of the set of complex zeros of the polynomial.

2.1. Curves and genus.

If the partial derivatives of f do not simultaneously vanish at any points (x, y) satisfying $f(x, y) = 0$ (that is, the curve is smooth) and the set of complex solutions is connected, then the set of complex solutions has the structure of a 1-dimensional complex manifold; that is, a Riemann surface. In this exposition, a *curve* will refer to any compact Riemann surface; by a fundamental equivalence, these can be identified with smooth, irreducible, projective algebraic varieties of dimension 1.

The complex solutions to $f(x, y) = 0$ provide an example of a one-dimensional complex algebraic variety, though not one that is projective, or is necessarily smooth or irreducible. However, any irreducible variety of dimension 1 is birationally equivalent to a unique smooth projective curve, where a birational equivalence of varieties of dimension 1 is a map locally defined by rational functions which identifies the solution sets away from finitely many points, such as the example of $(x, y) \mapsto (x, x^{1011}y)$ given above.

The topological classification of compact Riemann surfaces tells us that any curve is homeomorphic to either the sphere or a g-holed torus for some $g \geq 1$. We define the *genus* of the curve to be 0 if the curve is a topological sphere, and g otherwise. As we will see, the genus provides a fundamental invariant for classifying the possibilities for rational points on a curve.

2.2. **Curves of small genus.** In the case of the genus 0 curve associated to $x^2 + y^2 = 1$, there is a simple geometric construction to understand why the set of rational solutions is infinite. Consider any line in the plane which contains the point $(1, 0)$; that is, a line of the form $x + \lambda y = 1$ for a real (or complex) slope λ. The intersection of the real circle $x^2 + y^2 = 1$ with this line is the set of points (x, y) so that $(1 - \lambda y)^2 + y^2 = 1$; solving for y and finding x with the line's equation, we obtain the points $(1, 0)$ and $(\frac{1-\lambda^2}{1+\lambda^2}, \frac{2\lambda}{1+\lambda^2})$. Thus the infinitude of rational points on the circle comes down to the infinitude of the set of rational slopes that we may choose for λ.

We can view this construction as an explicit presentation of the curve as a genus 0 curve; indeed, the map $\lambda \mapsto (\frac{1-\lambda^2}{1+\lambda^2}, \frac{2\lambda}{1+\lambda^2})$ provides a birational equivalence from $\mathbb{C}$ to the solution set of $x^2 + y^2 = 1$. For a general genus 0 curve, we may have infinitely many rational points as above, but we may have finitely many or none at all, as in the example $x^2 + y^2 = -1$, or the example $x^2 + y^2 = 3$, where we can see there are no rational solutions by working modulo 4 . If the defining polynomial f has rational coefficients, these 'local' obstructions to rational points are the only issue that arises in genus 0, and the dichotomy holds generally: either $f(x, y) = 0$ has infinitely many rational solutions, or none.

An even more interesting geometric construction to produce rational points arises in the setting of elliptic curves; that is, curves of genus 1. Any curve of genus 1 is birationally equivalent to one of the form $y^2 = x^3 + ax + b$, where a, b are complex numbers satisfying $4a^3 + 27b^2 \neq 0$ (this choice ensures that the curve is smooth). We can mimic the above construction as follows: if there is any one rational point P on a curve of this form (with coefficients in $\mathbb{Q}$), we consider the tangent line to the curve at P. Since the defining equation of the curve has degree 3 in x, we can see that this tangent line will intersect the curve at exactly one more point, whose coordinates will be rational functions of the coordinates of P and the coefficients a, b. In this way, we obtain an algebraic self-map of the elliptic curve which sends rational points to rational points, so we can use one rational point to produce another. Iterating this map, we can potentially obtain a large set of rational points on the elliptic curve; so really, this is a dynamical approach to propagating rational points. The set of rational points produced by this procedure is actually only potentially infinite, though, because we are not guaranteed at each step to produce *new* rational points. Indeed, the third point may turn out to be P itself, or more generally we may iterate this procedure only to end up in a finite cycle of points.

If we have two or more rational points on the curve, we can generalize this construction: given two rational points P and Q on the elliptic curve, the line through P and Q will intersect the curve at exactly one more point R, which will be rational if P and Q (and the defining coefficients a and b) are. The remarkable fact about elliptic curves is that this geometric construction is also algebraic! Explicitly, we can use it to define an abelian group law on the set of complex solutions to $y^2 = x^3 + ax + b$ by defining $P + Q$ to be the reflection of R about the x-axis, with R as described above. Notice that our procedure above with the tangent line was sending a point P of the curve to the point $-(P + P)$.

Since the sum of two rational points under this group law is again a rational point, the set of rational points on an elliptic curve defined by rational coefficients form a group. In the 1922 paper [34] quoted above, Mordell proved that this group is finitely generated. Several features of this result are ripe for generalization; there are higher dimensional analogues of elliptic curves known abelian varieties, and we may wish to consider fields other than $\mathbb{Q}$; for example, number fields, which are finite extensions of $\mathbb{Q}$. Indeed Mordell's result was extended to abelian varieties over number fields by Weil's doctoral thesis a few years later, and this generalization is now known as the *Mordell-Weil theorem.* Among other novel tools, Weil's work included the initial ideas for the height functions, which are fundamental to the modern study of arithmetic geometry (see subsection 5.1).

2.3. **Curves of higher genus.** Mordell concluded his 1922 paper with

> I might note that the preceding work suggests to me the truth of the following statements concerning indeterminate equations, none of which, however, I can prove [...] The equation
>
> $$ax^6 + bx^5y + \cdots + fxy^5 + gy^6 = z^2$$
>
> can be satisfied by only a finite number of *rational* values of x and y with the obvious extension to equations of higher degree [...] The same theorem holds for any homogeneous equation of genus greater than unity, say, $f(x, y, z) = 0$.

Mordell's implicit evidence for this conjecture, beyond the computational, primarily seemed to be the failure of any geometric or algebraic structure which might allow one to "boost" one rational point to infinitely many, such as the type described in the previous section for curves of genus 0 and 1. Let's consider the procedure we used in the genus 1 case to (potentially) obtain many rational points from a single point, and understand why it must fail in higher genus. On a concrete level, if we wish to obtain a unique further point from intersecting a line through a rational point P with a higher genus curve, we need the tangency of the tangent line at P to be of higher order; that is, P must be a flex point of the curve. But a plane curve which is not a line has only finitely many flex points, so we cannot hope to make this geometric procedure work in higher genus.

Independently of the particulars of the tangent line construction, though, any dynamical approach to propagating rational points must fail in higher genus. For we cannot hope to produce an infinite set of rational points by any algebraic self-map of the curve as we did in the elliptic curve case, since the theorem of Hurwitz [24] bounds the order of the group of holomorphic self-maps of a genus $g > 1$ compact Riemann surface by $84(g - 1)$ - a result which provides additional motivation in itself, perhaps, for Mordell's conjecture.

It turns out that this is not just a failure of innovation of methods to produce rational points on higher genus curves: Mordell's claim was proved by Faltings [12] and is now known as the Mordell conjecture or Faltings' theorem: a curve of genus $g \geq 2$ has at most finitely many rational points.

In light of Faltings' theorem, what can we say about the finite set of rational points on a higher genus curve? Here are a few well-studied examples.

Example 2.1. Fermat's Last Theorem asserts that the equation

$$x^n + y^n = z^n$$

has no solutions in positive integers x, y, z for any integer value $n \geq 2$. Rearranging a bit with negative signs we see that this is equivalent to the statement that rational solutions of $x^n + y^n = 1$ are a subset of $\{(\pm 1, 0), (0, \pm 1)\}$. For $n \geq 4$, the curve defined by $x^n + y^n = z^n$ has genus at least 2, and so Faltings' theorem tells us there are only finitely many rational points on this curve. Note the finiteness is a weaker statement that explicit computation of the solutions, now known by the proof of the (much!) stronger statement of Fermat's Last Theorem.

Example 2.2. It is easy to see that the set of rational points on a curve can be arbitrarily large; for example, if $r_1, \ldots, r_m$ are distinct rational numbers, then the set of zeroes of

$$f(x, y) = y^m - \prod_{i=1}^{m} (x - r_i)$$

contains the m rational points of the form $(r_i, 0), 1 \leq i \leq m$. However, notice that the genus in this curve is quadratic in m. Better constructions are known, primarily by improving interpolation by using curves with many symmetries; as Elkies pointed out, the construction of Mestre [32] yields for each $g \geq 2$ a curve over $\mathbb{Q}$ of genus g with at least $8g + 12$ rational points.

Example 2.3. Rational points of curves of genus 2 have been particularly well-studied, and while Mestre's work gives examples of genus 2 curves with at least 28 rational points, much larger sets are known. The current record is was found by Stoll, searching a family constructed by Elkies: the genus 2 curve

$$\begin{aligned} y^2 = 82342800x^6 &- 470135160x^5 + 52485681x^4 + \\ &+ 2396040466x^3 + 567207969x^2 - 985905640x + 247747600 \end{aligned}$$

has at least 642 rational points! See the list of these points at `https://www.mathe2.uni-bayreuth.de/stoll/recordcurve.html`.

2.4. **Working over number fields.** For expository purposes, we have largely constrained ourselves so far to considering points with coordinates in $\mathbb{Q}$ on polynomials defined over $\mathbb{Q}$. While reading this article, you can if you wish always restrict your attention to that setting. However, when one is in search of roots of polynomial equations, extending the field can always be done to provide more solutions, and so it is an important theoretical point to note that any *finite* extension of $\mathbb{Q}$ will not change the fundamental underlying finiteness provided by Faltings' theorem. Indeed, the arithmetic machinery brought to these questions all work in the broader context of number fields; that is, fields which are extensions of $\mathbb{Q}$ of finite degree.

The elements of these fields are all algebraic numbers; that is, they are roots of polynomials with integer coefficients. We write $\overline{\mathbb{Q}}$ for the minimal (infinite) extension of $\mathbb{Q}$ which contains all algebraic numbers.

We might, for example, be interested to know which points with coordinates in the Gaussian rationals $\mathbb{Q}(i)$ lie on the Fermat curve $x^n + y^n = 1$; or perhaps we wish to adjoin an nth root of unity ζ_n to $\mathbb{Q}$, in which case the Fermat curve gains points of the form $(\zeta_n^k, 0)$ and $(0, \zeta_n^k)$. Even more generally, we might also wish to choose coefficients for the polynomial f which lie in a number field, to enable studying rational points on curves like $x^n + \sqrt{2}y^n = 1$. In this setting, we often use the term 'rational points' or 'K-rational points' to denote the set of points with coordinates in the minimal field extension K of $\mathbb{Q}$ which contains all coefficients defining the polynomial $f(x, y)$. As noted above, although extending $\mathbb{Q}$ can provide some new points on a higher genus curve, a key feature of the connection between arithmetic and geometry is that the *finiteness* of rational points in higher genus is valid independent of the choice of this number field K; that is, Faltings' theorem holds in this generality.

Theorem 2.4. *[Faltings] Let K be a number field, and C a curve of genus $g \geq 2$ defined over K. Then C contains finitely many K-rational points.*

This theorem was reproved using different techniques by Vojta in 1991 [41], simplified by Bombieri [1] and Faltings [13], and recently by a new approach of Lawrence and Venkatesh [28].

3. Torsion points on curves: algebra and geometry

We discussed in the last section why the geometric and dynamical approaches to propagating rational points fails in higher genus, but remember that in the case of genus 1, we also had algebraic structure; that is, the rational points (as well as the complex points) on an elliptic curve form an abelian group. This algebraic structure is an extremely powerful tool for the study of rational points on elliptic curves, and its analogue in higher genus - the Jacobian of a curve - will provide us with the same tools in that setting. In this section, all curves will have genus $g \geq 2$ unless stated otherwise.

3.1. Construction of the Jacobian.

While the complex points of a higher genus curve do not admit a (useful) group structure, we can embed the curve into an algebraic variety whose points form an abelian group; that is, an *abelian variety.* The downside of this is that application of the group law no longer necessarily maps two points on the curve to a sum which still lies on the curve, so we cannot use the group law to produce more rational points on the curve itself. Nonetheless, we can study the rational points in the larger variety - which itself exhibits this beautiful interaction between algebra, arithmetic, and geometry - to deduce useful information about the rational points on the curve.

This tool is technical but so important that we will describe it from several perspectives.

First, we will construct this embedding formally for any higher genus curve C. A *divisor* on C is a formal $\mathbb{Z}$-linear combinations of points on the curve; that is, sums of the form $\sum_{i=1}^m n_i P_i$ with $n_i \in \mathbb{Z}$ and P_i a complex point of the curve C for each i. Divisors on C with the addition operation form an abelian group which we'll denote by $\mathrm{Div}(C)$. The *degree* of a divisor is the integer $\sum_{i=1}^m n_i$, and the set $\mathrm{Div}^0(C)$ of divisors of degree 0 forms a subgroup of $\mathrm{Div}(C)$. An easy way to construct divisors of degree 0 is to consider the 'zeros and poles' divisor associated to an analytic function: if f is a nonconstant analytic map from C to the Riemann sphere, then the divisor obtained from f is

$$\mathrm{div}(f) := \sum_{x \in C} \mathrm{ord}_x(f) x,$$

where $\mathrm{ord}_x(f)$ is the order of vanishing of f at x if $f(x) = 0$, minus the order of the pole if $f(x) = \infty$, and zero otherwise. Since the degree of an analytic map between compact Riemann surfaces is well-defined, the degree of any zeros and poles divisor is 0.

We say that two divisors are *linearly equivalent* if their difference is of the form $\mathrm{div}(f)$ for some nonconstant analytic map f from C to the Riemann sphere. As a basic example, if $f(x, y) = 0$ describes a smooth plane curve C of degree d, then the divisor formed from intersecting C with a line - that is, the divisor which is the sum of the points in the intersection, counted with their multiplicities - is linearly equivalent to any other divisor from the sum of d collinear points of C.

Definition 3.1. The *Jacobian* of the curve C, denoted $\mathrm{Jac}(C)$, is the group of degree 0 divisors modulo linear equivalence.

To embed C into its Jacobian, we need only choose a basepoint $x_0 \in C$ to construct the *Abel-Jacobi map based at* x_0:

$$\Phi_{x_0}(x) := [x] - [x_0] \in \mathrm{Jac}(C),$$

sending a point to the linear equivalence class of the divisor $x - x_0$. By definition, $\mathrm{Jac}(C)$ is generated as a group by the image of C under any Abel-Jacobi map; in fact, $\mathrm{Jac}(C)$ is generated by elements of the form $[x] - [y]$, where $x, y \in C$. An Abel-Jacobi map is injective precisely when the genus of C is at least 1. In this way we see that any curve of genus $g \geq 1$ can be viewed as a subvariety of an abelian variety.

We can take an analytic perspective to define the Jacobian, if we prefer, and the flexibility of the two perspectives is key to the work of Dimitrov-Gao-Habegger and Kühne. Again starting with a curve C of genus $g \geq 2$, as C is a Riemann surface of genus g, C admits exactly g linearly independent holomorphic differential 1-forms, so we can fix a basis $\omega_1, \ldots, \omega_g$ of holomorphic differentials.

Since the first homology of C is of dimension $2g$, we may choose representatives $c_1, \ldots, c_{2g}$ generating the 1-dimensional homology of C, and define for each c_i the *period* $\omega(c_i)$ to be

$$\omega(c_i) := \left(\int_{c_i} \omega_1, \int_{c_i} \omega_2, \ldots, \int_{c_i} \omega_g \right),$$

and define the *period lattice* Ω in $\mathbb{C}^g$ to be the subgroup generated by the periods $\omega(c_i)$ for $1 \leq i \leq 2g$. One can show that the period lattice is discrete in $\mathbb{C}^g$ and a free abelian group on $2g$ generators, so the quotient $\mathbb{C}^g/\Omega$ is a complex torus of dimension g.

Recall that $\mathrm{Jac}(C)$ is generated by classes of the form $[x]-[y]$ where $x, y \in C$. Given any $x, y \in C$, let γ be a path from y to x, and define the vector

$$\omega(\gamma) := \left(\int_\gamma \omega_1, \int_\gamma \omega_2, \ldots, \int_\gamma \omega_g \right),$$

which lies in $\mathbb{C}^g$, and whose value modulo the period lattice Ω is independent of the choice of path γ from y to x. We can then define a map $\mathrm{Jac}(C) \to \mathbb{C}^g/\Omega$ on the generating set of classes of the form $[x]-[y]$ by sending the class of the divisor $[x]-[y]$ to $\omega(\gamma)$. A fundamental theorem is that this map provides a group isomorphism, and so we identify the Jacobian of C with the complex torus $\mathbb{C}^g/\Omega$ via this map, giving the analytic characterization of the Jacobian.

Note that this analytic process of constructing a period lattice varies holomorphically with C; that is, if we have a family of genus g curves which vary holomorphically, then for any given curve C_0 in the family and any choice $\omega(c_1), \ldots, \omega(c_{2g})$ of basis for the period lattice of the Jacobian for C_0, we may vary this basis holomorphically to provide a basis for curves which are holomorphic perturbations of C_0. This is the basis of the construction of the Betti map crucial to the work of Dimitrov-Gao-Habegger and Kühne, as described in subsection 6.4.

3.2. **Torsion points on curves.** We have been focused so far on the interaction between arithmetic (that is, rational points) and the geometry of curves, but now that we have a construction to embed a curve into an abelian variety, we can ask the natural question: how does the geometry of the curve impact the *algebraic* structure of the points on the curve embedded into this abelian variety? In particular, if C is embedded into its Jacobian $\mathrm{Jac}(C)$, how many torsion points of the group $\mathrm{Jac}(C)$ can the embedded curve contain?

This question was raised independently by Manin and Mumford and answered by Raynaud [36], now referred to as Raynaud's theorem or the Manin-Mumford conjecture.

Theorem 3.2. *[Raynaud] Let C be a curve of genus $g \geq 2$, and fix an embedding $C \hookrightarrow \mathrm{Jac}(C)$. Then the set of torsion points of* $\mathrm{Jac}(C)$ *in C is finite.*

Remark 3.3. Notice that we have not restricted ourselves to points in some fixed number field here; indeed, for any number field K the set of K-rational torsion points of $\mathrm{Jac}(C)$ is a finite set, so the question is not interesting unless we allow the field to vary. This is an easy consequence of basic height theory.

3.3. **The Mordell-Lang conjecture and unlikely intersections.** The question of torsion points on a curve in its Jacobian might seem very far from the question of rational points on a curve, but as Lang realized [27] well before either was proved, the Mordell conjecture and the Manin-Mumford conjecture can be unified into a single statement, now known as the *Mordell-Lang conjecture.* Given an abelian group A (with additive notation), the *rank* of A is the cardinality of a maximal linearly independent subset. For example, any finitely generated abelian group is isomorphic to $\mathbb{Z}^r \otimes T$ for a torsion subgroup T, and so a finitely generated group has finite rank r. However, an abelian group can have finite rank without being finitely generated, as is the case for the rational numbers $\mathbb{Q}$ under addition.

Lang made several conjectures of varying strength which generalized the Mordell conjecture; the form we state below was proved by the work of Faltings [13] and Hindry [22], generalizing Vojta's proof of Mordell.

Theorem 3.4. *Let C be a complex curve of genus $g \geq 2$, embedded into its Jacobian $\mathrm{Jac}(C)$. Then for any finite rank subgroup Γ of $\mathrm{Jac}(C)$, $C \cap \Gamma$ is finite.*

Let's understand why this is a unification of the Mordell and Manin-Mumford conjectures.

Proposition 3.5. *Theorem 3.4 implies Theorem 2.4 and Theorem 3.2.*

Proof. Let us first consider 2.4. Suppose C is a curve of genus $g \geq 2$ defined over a number field K. As noted in subsection 2.2, Mordell and Weil proved that the set of K-rational points of $\mathrm{Jac}(C)$ form a finitely generated group; in particular, a finite rank subgroup of $\mathrm{Jac}(C)$. By Theorem 3.4, we conclude that the embedded curve C contains only finitely many points of the finite rank subgroup of K-rational points in the Jacobian.

Now we turn to Theorem 3.2. By definition, the rank of the set of torsion points of any abelian group is 0, and so applying Theorem 3.4 we deduce that any curve of genus $g \geq 2$ embedded into its Jacobian contains only finitely many torsion points. □

4. Uniformity

Theorem 3.4 gives us a powerful finiteness result - on rational points, torsion points of the curve inside its Jacobian, and more - from the very basic information of the genus of a curve. Now that we know this set of rational points on a curve of genus at least two is finite, we want to know: what is this finite set?

Unfortunately, writing down a complete list of the rational points on a higher genus curve is a very difficult problem even for specific cases; even the most well-studied case of $x^n + y^n = 1$ took more than a decade after Faltings' theorem and a substantially different toolkit of techniques. In fact, the negative resolution of Hilbert's tenth problem provides some evidence that there can be no algorithm to decide existence or non-existence of *any* rational solution to a system of polynomial equations. While there is a wealth of interesting work in the direction of effective computation of rational points on a curve, we will address here a weaker question: given a curve of genus $g \geq 2$ over the rational numbers, how large can the set of rational points be?

4.1. Uniform Mordell-Lang.

We can fit this question into the Mordell-Lang framework:

Question 4.1. *Given a curve of genus $g \geq 2$ embedded into its Jacobian and a finite rank subgroup Γ of $\mathrm{Jac}(C)$, how large can $C \cap \Gamma$ be?*

We have already explored this question somewhat with Example 2.2, where we demonstrated that if we allow the genus of a curve to be large, then the curve can have a large number of rational points. It follows that any upper bound on the size of the set of rational points must have some dependence on the genus of the curve. Additionally, any upper bound must depend on the rank of the subgroup Γ; after all, we could simply expand Γ by adding elements of C, increasing both the rank of Γ and the number of points in the intersection $C \cap \Gamma$.

In 1986, Mazur asked if perhaps these were the *only* inputs needed to provide an upper bound on the size of $C \cap \Gamma$:

Question 4.2. *[uniform Mordell-Lang] Does there exist a bound $B = B(g, r)$ so that for every curve C of genus g embedded into its Jacobian, and every subgroup Γ of* $\mathrm{Jac}(C)$ *of rank r, we have*

$$|C \cap \Gamma| \leq B(g, r)?$$

The question is stated with deliberate ambiguity about the field over which we are working. Although Theorem 3.4 was stated over the complex numbers, one of the advantages of the uniformity question is that by a specialization argument of Masser [30], an affirmative answer for Question 4.2 over $\overline{\mathbb{Q}}$ suffices to deduce a positive answer to Question 4.2 for any field of characteristic 0.

4.2. **Previous work.** In the last few decades there have been a variety of tools used to chip away at Mazur's question, particularly in the situation when Γ is the set of rational points of the Jacobian (the setting of Faltings theorem) or when $\Gamma = \{0\}$ (the setting of Raynaud's theorem). We will mention only a few of these, though the literature is substantial.

Perhaps the most well-explored is the approach of Chabauty and Coleman; in 1941, Chabauty [3] proved a partial result towards the Mordell conjecture, proving finiteness of rational points for curves of genus $g \geq 2$ over $\mathbb{Q}$ for which the rational points of the Jacobian formed a group of rank at most $g - 1$ by a study of the p-adic Lie theory of the Jacobian. Coleman [4] made Chabauty's method explicit, but the bounds obtained depend on the geometry of curve when considered over finite fields $\mathbb{F}_p$ for a well-chosen prime p which depends on the curve, so are not uniform. In a more recent innovation, Stoll [39] and Katz-Rabinoff-Zureick-Brown [25] combined these methods with non-archimedean tools to prove very strong uniform results; in particular, they were able to remove the choice of a 'good' prime for the curve. These bounds were still subject to a condition on the rank of the group of rational points of the Jacobian; however, the method is appealing because it produces explicit and reasonably strong bounds. Katz-Rabinoff-Zureick-Brown also used these methods to make progress on the uniform Manin-Mumford question, obtaining an upper bound on torsion points on a curve defined over a number field K in its Jacobian: this bound depends on the genus and the extension degree $[K : \mathbb{Q}]$, so are not fully uniform.

Other progress has generally been limited to particular families of curves; based on the method of Vojta (see section 5), David-Philippon [6] and David-Nakamaye-Philippon [5] proved for some families of curves bounds of the form

$$|X(K)| \leq c(g, [K : \mathbb{Q}])^{1+\rho},$$

where K is a number field, $X(K)$ denotes the K-rational points on the curve, ρ is the rank of the K-rational points of the Jacobian of X, and c is a constant depending on the data of the genus g of X and the field extension degree of K over the rationals. In the direction of uniform Manin-Mumford, DeMarco-Krieger-Ye [7] provided a fully uniform bound for a family of genus two curves, using the quantitative equidistribution theory of Favre-Rivera-Letelier [14].

4.3. The result of Dimitrov-Gao-Habegger and Kühne.

In a striking series of innovations, the works of Dimitrov-Gao-Habegger [11] and of Kühne [26], in turn based on Habegger [21], Gao-Habegger [19], and Gao [15] [16], provide an affirmative answer to Mazur's uniform Mordell-Lang question:

Theorem 4.3. *[Dimitrov-Gao-Habegger and Kühne] Let $g \geq 2$ be an integer, and F a field of characteristic 0. Then there exists a constant $c(g)$ with the following property. Let C be a smooth curve of genus g defined over F, and let P_0 be an F-rational point of C, with associated Abel-Jacobi embedding $j : C \hookrightarrow \mathrm{Jac}(C)$. Let Γ be a subgroup of $\mathrm{Jac}(C)(F)$ of finite rank ρ. Then*

$$|j(C)(F) \cap \Gamma| \leq c(g)^{1+\rho}.$$

By taking F to be a number field and $\Gamma = \mathrm{Jac}(C)(F)$ we obtain a uniform bound (depending on g and ρ) on the number of F-rational points of C, proving a uniformity for the size of the finite sets asserted by Faltings' theorem. Taking $F = \mathbb{C}$ and Γ to be the torsion points of the Jacobian (a rank 0 subgroup, recall) we obtain a uniform version of Raynaud's theorem.

In structure, the work of Dimitrov-Gao-Habegger and Kühne is based on a proof of the Mordell conjecture given by Vojta in 1991 [41] and the quantitative version of that argument developed by Rémond [37]. The key essential issue in trying to understand how a curve in an abelian variety (such as its Jacobian) interacts with a subgroup Γ is how to *arithmetically* encode the information of the curve having higher genus. After all, there can be plenty of (genus 1) curves in an abelian variety with infinitely many rational points; for example, if E is an elliptic curve with an infinite set of rational points, and $A = E \times E$ is the product of E with itself, then A will contain many curves which have infinitely many rational points. In particular, any curve of the form $\{(P, [n]P) : P \in E\}$ with a fixed positive integer n will provide such a curve, as does a curve of the form $P \times E$ where P is any choice of torsion point of E.

In the case of a single abelian variety, this arithmetic encoding is done via the theory of heights associated to divisors. In Vojta's approach, the genus of the curve C is encoded in the properties of the so-called Vojta divisor on $C \times C$ and its associated height function; in the next section, we will outline some of the ideas of Vojta's argument. The primary innovation of the work of Dimitrov-Gao-Habegger and Kühne was the notion of *non-degenerate subvarieties* of an abelian scheme, along with a geometric description of these subvarieties, control over the arithmetic proximity of points on a non-degenerate subvariety, and an equidistribution result for points of small height on non-degenerate subvarieties.

5. Vojta's approach

To understand part of the outline of the Dimitrov-Gao-Habegger and Kühne arguments, we will first describe the structure of Vojta's approach to Faltings' theorem, as simplified and exposited by Bombieri.

5.1. Heights on projective space.

Heights on projective space. The fundamental tool to freely move between geometry and arithmetic is a *height function.* For details (and a much more carefully presented discussion) see Part B of [23].

A point in projective space $\mathbb{P}^n(\mathbb{C})$ of n dimensions is described as a line through the origin in $\mathbb{C}^{n+1}$; that is, a set of coordinates $[x_0, x_1, \ldots, x_n]$ defined up to multiplication by a non-zero complex constant. A $\mathbb{Q}$-rational point in projective space can then - after cancelling denominators - be represented by a point $x = [x_0, x_1, \ldots, x_n]$ for which $x_0, x_1, \ldots, x_n$ are integers with no common divisor. We define the *naive logarithmic height* of the point x to be

$$h(x) := \log \max\{|x_0|, \ldots, |x_n|\}.$$

This value $h(x)$ has the following arithmetic interpretation. While $\mathbb{Q}$ comes equipped with the absolute value inherited from its embedding into the complex numbers, it also admits *p-adic* absolute values, given for any prime p by

$$|q|_p = p^{-\mathrm{ord}_p(q)},$$

where $\mathrm{ord}_p(q)$ counts the number of factors of p that divide q (with negative sign if the factors appear in the denominator). The key fact unifying these absolute values on $\mathbb{Q}$ is known as the *product formula*, which states that for any non-zero $q \in \mathbb{Q}$ we have

$$\prod_{v \in M_{\mathbb{Q}}} |q|_v = 1,$$

where $M_{\mathbb{Q}}$ is the set consisting of the p-adic absolute values together with the 'usual' Euclidean absolute value. One can check that we may equivalently define

$$h(x) = \sum_{v \in M_{\mathbb{Q}}} \log \max\{|x_0|_v, \ldots, |x_n|_v\};$$

this is known as a *local decomposition* for the height, and by the product formula the sum will be independent of the choice of rational representative for the point x. Thus we understand the naive logarithmic height as providing a measure of *arithmetic* size, since it measures not only the usual absolute value but also the contributions from prime divisors of coordinates. The local decomposition definition of the height of a point can be extended to points in projective space over any number field, using Ostrowski's theorem to understand the embeddings of the field into the complex numbers as well as the analogues of the p-adic absolute values.

We obtain in this way a function $h : \mathbb{P}^n(\overline{\mathbb{Q}}) \to [0, \infty)$, where $\overline{\mathbb{Q}}$ denotes the algebraic closure of $\mathbb{Q}$, the minimal infinite extension containing every number field.

5.2. Heights from divisors.

Heights from divisors. The height function provides us with a description of arithmetic complexity for points in projective space and so for any algebraic variety which can be embedded into projective space as well; if we have a variety X and an embedding $\phi : X \hookrightarrow \mathbb{P}^n$, then we may define $h_\phi(x) := h(\phi(x))$ for any $x \in X(\overline{\mathbb{Q}})$. So how do we (potentially) find embeddings of a variety into projective space? The basic idea is the following: if we have a set $\{f_0, \ldots, f_k\}$ of rational functions taking $X \to \mathbb{P}^1$, then we can consider the map $X \to \mathbb{P}^k$ which (away from the poles and common zeros) is defined by

$$x \mapsto [f_0(x), f_1(x), \dots, f_k(x)].$$

This will by no means always give an embedding; we certainly need some basic requirements like an independence property for the maps and their vanishing loci, and for the set of rational functions on X to be large 'enough'. But the theory of divisors provides a natural setting where this idea can succeed to provide an embedding of X into projective space if such a map exists, and therefore allow X to inherit a height function from that embedding.

Recall from our construction in subsection 3.1 of the Jacobian of a curve the concept of a divisor on a curve: a formal sum of points (i.e. dimension 0 subvarieties) of the curve. For an arbitrary algebraic variety X of dimension d, we can mimic this construction, instead defining a divisor to be a formal sum of subvarieties of X with dimension $d-1$. As in the case of curves, we can define the divisor associated to a rational function on X, linear equivalence of divisors, and degree of a divisor. We can then associate to any divisor a set of rational functions as follows: let D be a divisor on X, and

$$L(D) = \{f \text{ a rational function on } X \mid D + \operatorname{div}(f) \text{ has non-negative coefficients}\}.$$

This is a finite-dimensional vector space, and for any basis $f_0, \dots, f_k$ of this space, we obtain a map $\phi_{L(D)} : X \to \mathbb{P}^k$. If this map is an embedding, the divisor is called *very ample*. If a divisor has a multiple which is very ample, we say the divisor is *ample*. We conclude that any ample divisor, if it exists, can be used (possibly after taking a multiple) to provide an embedding of the variety into projective space.

Let's illustrate this technique with an example. Let $E = \mathbb{C}/\Lambda$ be a an elliptic curve with point P corresponding to the equivalence class of 0. The theory of Weierstrass elliptic functions provides us with two linearly independent examples of rational maps on E; namely, the Weierstrass $\wp$ function and its derivative $\wp'$; these have poles of order two and three at the lattice points, respectively. One can show that the constant map is a basis for $L(P)$, $\{1, \wp\}$ is a basis for $L(2P)$, and $\{1, \wp, \wp'\}$ is a basis for $L(3P)$. The first two examples clearly do not give an embedding of E into projective space, but the map $E \to \mathbb{P}^2$ given by the basis $\{1, \wp, \wp'\}$ for $L(3P)$ is in fact an embedding, realizing E as a plane curve in Weierstrass form (up to a factor of 4). So $3P$ is very ample, and the embedding $E \hookrightarrow \mathbb{P}^2$ provides a height function on E, by composition with the height function on $\mathbb{P}^2$.

Finally we note that if a variety $X \hookrightarrow \mathbb{P}^k$ is already embedded into projective space, then the embedding corresponds to one coming from the divisor class associated to a *hyperplane*, which is the equivalence class containing divisors given by subvarieties of the form $L \cap X$ where L is the set of solutions to $\ell(P) = 0$ for a general linear form ℓ in the coordinates of $\mathbb{P}^k$. We can see this in the preceding example, as $3P$ is precisely the intersection of a Weierstrass elliptic curve with the tangent line at a flex point, and as noted in subsection 3.1, there is a single divisor class on E which contains the divisors formed from three collinear points, including the degenerate case of the flex point.

This discussion leads us to the *Weil height machine*, a construction which associates to any (smooth projective) variety V defined over a number field a map from divisors to height functions:

$$h : \mathrm{Div}(V) \to \{\text{functions } V(\overline{\mathbb{Q}}) \to \mathbb{R}\},$$

where a very ample divisor D is associated to a height function h_D from an embedding arising from a choice of basis of $L(D)$, and for the rest we use the fact that any divisor on a smooth projective variety can be written as a difference of two very ample divisors. There is clearly some ambiguity here, as we choose a basis for $L(D)$ and representations of D as a difference of very amples, and so to be precise we generally only have a height function well-defined up to a bounded constant; that is, if h_D is a choice of height function associated to a divisor D and $h : V(\overline{\mathbb{Q}}) \to \mathbb{R}$ is a function for which $|h_D(x) - h(x)|$ is uniformly bounded, then h is considered a height function for D as well.

In fact, this entire discussion goes through if we replace $\mathbb{Q}$ with an arbitrary number field, and the machine is normalized in the sense that the height associated to the hyperplane in projective space recovers the naive logarithmic height. We will not delve more deeply into the description of the Weil height machine, but we must note two important positivity features:

(1) if a divisor $D = n_1Y_1 + \cdots n_rY_r$ has non-negative coefficients $n_i \geq 0$ (we say D is *effective*), then there is a lower bound B so that $h_D(x) \geq B$ for all $x \notin Y_1 \cup \cdots \cup Y_r$, and

(2) if a divisor D on V is ample and C is any real constant, then for any number field K, the set
$$\{x \in V(K) \mid h_D(x) \leq C\}$$
is finite.

In this way, the geometric features of a divisor are translated into positivity properties for the associated height function, a feature which we will see has substantial arithmetic consequences.

5.3. **Heights on Jacobians.** The Jacobian of a curve C of genus $g \geq 2$ comes with a natural ample divisor: the *theta divisor*

$$\Theta := \{j(x_1) + \cdots + j(x_{g-1}) : x_1, \ldots, x_{g-1} \in C\},$$

where j is any choice of Abel-Jacobi embedding of the curve into its Jacobian. We will also consider

$$\Theta^- := \{-j(x_1) - \cdots - j(x_{g-1}) : x_1, \ldots, x_{g-1} \in C\},$$

since the divisor $\Theta + \Theta^-$ has a useful symmetry from the additive inverse. It turns out that $\Theta + \Theta^-$ is also ample, and the corresponding height function $h_{\Theta+\Theta^-}$ takes non-negative values. We obtain by restriction of $h_{\Theta+\Theta^-}$ to $j(C)$ a notion of arithmetic complexity for the rational and algebraic points on the curve C.

Remarkably, by a construction of Néron and Tate, one may modify $h_{\Theta+\Theta^-}$ by a bounded function to obtain a height function on $\mathrm{Jac}(C)$ which has encodes properties of the group structure on $\mathrm{Jac}(C)$. We define the *canonical* or *Néron-Tate height* associated to $\mathrm{Jac}(C)$ to be

$$\hat{h}_\Theta(P) := \lim_{n\to\infty} \frac{h_{\Theta+\Theta^-}(nP)}{n^2}.$$

This canonical height is also non-negative, and transforms well with the group law: we have $\hat{h}_\Theta(nP) = n^2\hat{h}_\Theta(P)$ for all algebraic points P of the Jacobian. Together with the second positivity property of the height machine, it follows that $\hat{h}_\Theta(P) = 0$ if and only if P is a torsion point of the Jacobian. It also satisfies the 'parallelogram law'

$$\hat{h}_\Theta(P+Q) + \hat{h}_\Theta(P-Q) = 2\hat{h}_\Theta(P) + 2\hat{h}_\Theta(Q).$$

From the parallelogram law it follows that

$$\langle P, Q\rangle := \frac{1}{2}\left(\hat{h}_\Theta(P+Q) - \hat{h}_\Theta(P) - \hat{h}_\Theta(Q)\right)$$

defines an inner product on the algebraic points of the Jacobian, and $|P|^2 = \hat{h}_\Theta(P)$ gives a norm

In this way we have used the geometry of the curve - encoded via the theta divisor - to construct a height function, which in turn can be used to construct a norm on the algebraic points of the Jacobian. This norm provides for us a notion $|P - Q|$ of the arithmetic distance between two algebraic points of the curve.

5.4. **The inequalities of Mumford and Vojta.** Let's think for a moment about how we might use this norm to control the number of rational points on a curve C of genus $g \geq 2$. In 1965 Mumford [35] realized that the genus of the curve controls the geometry of the points with respect to the inner product associated to the height defined by the theta divisor as described in 5.3. To see this, let Δ denote the diagonal in $C \times C$, and h_Δ an associated height function, which by the positivity property for effective divisors we may assume is non-negative off the diagonal. We associate C with its image in $\mathrm{Jac}(C)$ under the Abel-Jacobi embedding associated to any rational point (if C has no rational points, we don't have much to prove for Mordell's conjecture). Mumford computed an explicit formula to relate h_Δ to the inner product, a simplified version of which states that for any algebraic points $P \neq Q$ of C, we have

$$h_\Delta(P, Q) = \frac{1}{2g}|P|^2 + \frac{1}{2g}|Q|^2 - \langle P, Q\rangle + \mathcal{O}(|P| + |Q| + 1)).$$

As Mumford realized, the key feature here is that when $g > 1$, the leading term

$$\frac{1}{2g}|P|^2 + \frac{1}{2g}|Q|^2 - \langle P, Q\rangle$$

is an indefinite quadratic form, taking negative values when $\langle P, Q\rangle$ is close to $|P||Q|$. This can be played off the non-negativity of the height $h_\Delta(P, Q)$ to impose restrictions on the real geometry (with respect to the inner product) of the algebraic points on higher genus curves. To demonstrate this simply, we'll consider the case of points which are linearly dependent as elements of $\mathrm{Jac}(C) \otimes \mathbb{R}$ equipped with the inner product described above, an example inspired by the exposition of Bombieri and Gubler [2]. Suppose that $P \neq Q$ are rational points of C which satisfy $\langle P, Q\rangle = |P||Q|$; let's assume without loss of generality that $|P| \geq |Q|$. Since $P \neq Q$, $h_\Delta(P, Q) \geq 0$, and by Mumford's formula we have

$$\langle P, Q\rangle \leq \frac{1}{2g}|P|^2 + \frac{1}{2g}|Q|^2 + \mathcal{O}(|P| + |Q| + 1)),$$

so by assumption,

$$|P||Q| \leq \frac{1}{2g}|P|^2 + \frac{1}{2g}|Q|^2 + \mathcal{O}(|P| + 1)).$$

Dividing both sides by $|P||Q|/2g$ we see that

$$2g \leq \frac{|P|}{|Q|} + \frac{|Q|}{|P|} + \mathcal{O}\left(\frac{1}{|Q|}\right).$$

If $g > 1$, then either the first or third term on the right must be sufficiently large for this inequality to hold, so we deduce that either $|P|$ is substantially larger than $|Q|$, or $|Q|$ is not too large.

This is (obviously) not a finiteness proof, but we see in this computation how the geometric features of the inner product control the ability of a higher genus curve to contain rational points; this particular computation forces a spacing between linearly dependent rational points of sufficient height. In fact, the same argument holds under the weaker assumption $\langle P, Q\rangle > (1-\epsilon)|P||Q|$ for a fixed small choice of ϵ, which is to say that large points can only be *nearly* parallel with respect to the inner product if the ratio $|P|/|Q|$ is sufficiently large.

Vojta's approach to prove Faltings' theorem worked with the same basic tools, using height functions associated to well-chosen divisors to understand the geometric spacing of rational points on higher genus curves, according to the inner product associated to the theta divisor on their Jacobian. Instead of the diagonal divisor Δ, Vojta worked with a family of divisors to get a sufficiently strong result, choosing a divisor based on the properties of P and Q; however, one then has the substantial technical issue of understanding the lower bound for the divisor - more precisely, controlling an effective representative for the divisor, both in complexity and in vanishing multiplicity at (P, Q). For this the elementary theory of heights is insufficient, and Vojta draws from ideas in arithmetic intersection theory and Diophantine approximation to prove a geometric separation of rational points known as *Vojta's inequality.*

Theorem 5.1 (Vojta's inequality)**.** *Let C be a curve of genus g defined over a number field K, embedded into its Jacobian by an Abel-Jacobi embedding associated to a rational point. Write $\langle\cdot,\cdot\rangle$ for the inner product associated to the canonical height relative to the theta divisor on the Jacobian. There exist constants C_1, C_2 so that if $P, Q \in C(\overline{K})$ are two points satisfying $|Q| \geq C_1$ and $|P| > C_2|Q|$, then*

$$\langle P, Q\rangle \leq \frac{3}{4}|P||Q|.$$

To understand how Faltings' theorem follows, note that for any $\epsilon > 0$, $\mathrm{Jac}(C)(K) \otimes \mathbb{R}$ can be covered by finitely many sectors of the form $\langle P, Q\rangle \leq (1-\epsilon)|P||Q|$, since $\mathrm{Jac}(C)(K)$ is finitely generated. Thus we really only need to know that each sector of this type contains finitely many K-rational points. If a sector of this type (with ϵ sufficiently small) contains any rational point Q with $|Q| \geq C_1$, then Vojta's inequality asserts that the sector contains no algebraic points P satisfying $|P| > C_2|Q|$, so we need only know that the set of rational points with bounded size is finite. Since $\Theta + \Theta^-$ is ample, we know that the set of rational points for which $h_{\Theta+\Theta^-}$ is bounded is finite; since $\hat{h}_\Theta$ differs from $h_{\Theta+\Theta^-}$ by a bounded constant, we have the desired finiteness statement.

6. The work of Dimitrov-Gao-Habegger and Kühne

Both proofs of the Mordell conjecture - Faltings' p-adic approach, and Vojta's geometric approach - are susceptible to effective refinements that might allow (partial) uniform control of the size of the finite set of rational points on a higher genus curve. As noted in 4.3, the uniform result of Dimitrov-Gao-Habegger and Kühne follows the structure of Vojta's argument. So let's begin by examining the discussion of 5.4 with an eye towards effective and uniform improvements.

6.1. **Uniformity in Vojta.** Vojta's inequality itself yields two constants, C_1 and C_2, where C_1 denotes the lower bound on height for both points, and C_2 the spacing bound. As our simple computation when $|P||Q| = \langle P, Q\rangle$ suggests, C_2 is easily shown to be chosen to depend only on the genus of the curve C. However, as Vojta varies the divisor away from Δ, the $\mathcal{O}$ bound for the linear term in $h_\Delta(P, Q)$ varies too, and so C_1 depends on the curve C (more particularly, on a suitable notion of height for C). Given Vojta's inequality, we then used two finiteness statements: the finite cover by narrow sectors of $\mathrm{Jac}(C)(K)$, and the finite points of bounded height. The number of narrow sectors has a simple bound 7^ρ, where ρ is the rank of the $\mathrm{Jac}(C)(K)$ (or any finite rank subgroup we might wish to consider). The set of points of bounded height, on the other hand, cannot be bounded independently of C by elementary height techniques, since the bound in question is related to C_1 and so depends on C itself.

Recall that the uniform Theorem 4.3 asserted a bound on $|C(\overline{Q}) \cap \Gamma|$ of the form $c(g)^{1+\rho}$ when C is viewed as embedded in $\mathrm{Jac}(C)$, where ρ is the rank of Γ. The discussion of the previous paragraph suggests we break the points of $|C(\overline{Q}) \cap \Gamma|$ into two types:

(1) small points: $P \in C(\overline{Q}) \cap \Gamma$ satisfying $\hat{h}_\Theta(P) \leq B(C)$, and
(2) large points: $P \in C(\overline{Q}) \cap \Gamma$ satisfying $\hat{h}_\Theta(P) > B(C)$,

where $B(C)$ is a constant which depends on a suitable notion of height for C. The techniques of Vojta noted above provide a uniform upper bound of the form $c(g)^{1+\rho}$ on the set of large points, and so the work of Dimitrov-Gao-Habegger and Kühne deals with counting points of small height.

6.2. **Variation of canonical height.** The basic constructions of Section 5 were universal in the sense that they were independent of the higher genus curve under consideration: the construction of the theta divisor, of the canonical height and associated norm, and of the Vojta divisors leading to Vojta's inequality require only the structure of the group law on the Jacobian and the divisor group of the product of the curve with itself, and so these constructions are amenable to being done in families. Silverman [38] explored this variation of canonical height in families of abelian varieties; to illustrate, let's consider a basic example that follows from his work.

Let E_t denote the *Legendre elliptic curve* for t, the set of solutions to

$$y^2 = x(x-1)(x-t),$$

where $t \in \mathbb{C} \setminus \{0, 1\}$. We can consider the two-dimensional space $\mathcal{E}$ consisting of points of the form (t, P), where $t \in \mathbb{C} \setminus \{0, 1\}$ and P is any point on E_t; note that we have a natural projection map $\pi : \mathcal{E} \to \mathbb{C} \setminus \{0,1\}$ which sends $(t, P) \mapsto t$, and we should think of $\pi^{-1}(t)$ as a copy of E_t.

When $t \in \overline{\mathbb{Q}}$, the elliptic curve E_t admits a Néron-Tate height on the algebraic points of the curve, which we will denote by $\hat{h}_{E_t}$. We can therefore define a function $h_{\mathcal{E}}(\overline{\mathbb{Q}}) : \mathcal{E}(\overline{\mathbb{Q}}) \to \mathbb{R}$ by $h_{\mathcal{E}}((t,P)) = \hat{h}_{E_t}(P)$. This is not a height function associated to a very ample divisor as described in Section 5; for example, since the point $(0,0)$ is a torsion point of order 2 for any choice of t, we have $h_{\mathcal{E}}((t,(0,0))) = 0$ for all t, and so the points of bounded height are not a finite set.

On the other hand, $(0,0)$ was a rather special choice of point, torsion for every parameter t. If instead we consider for each t the point

$$P_t := (2, \sqrt{2(2-t)})$$

for some choice of square root, then the canonical height of P_t will grow as t does. This relationship persists for other choices of varying point P_t, known as *sections* of the family. Silverman proves (a generalization of the following result):

Theorem 6.1 (Silverman). *Given a section σ, there is a geometric quantity $\hat{h}_{\mathcal{E}_\eta}(\sigma_\eta)$ so that*

$$\lim_{h(t)\to\infty} \frac{\hat{h}(\sigma(t))}{h(t)}$$

exists and equals $\hat{h}_{\mathcal{E}_\eta}(\sigma_\eta)$.

In this setting, this geometric quantity $\hat{h}_{\mathcal{E}_\eta}(\sigma_\eta)$ is a function field height. The key thing to notice is the difference between the two examples of sections that we considered above. If the section is taken to be the constant point $(0,0)$, which is *persistently* torsion for the family, then the limiting value in Silverman's theorem is 0, and we can make no finiteness deduction about points of small height associated to this section: this is bad news for controlling points of small heights uniformly in this family (and indeed, elliptic curves contain infinitely many torsion points). On the other hand, when the limiting value of Silverman's theorem is non-zero, a point of small height must live above a parameter t of bounded height, and so we can hope to control small points of this type via Silverman's theorem.

This is precisely the approach of Dimitrov-Gao-Habegger and the work of the authors that precedes it, which includes the substantial technical improvement of allowing points to live in a subvariety of any dimension rather than a section. The novelty in their resolution of this question is a complete and natural description of a geometric quantity associated to the subvariety which plays a role analogous to that of $\hat{h}_{\mathcal{E}_\eta}(\sigma_\eta)$ in Silverman's theorem. We will outline now, omitting a number of technicalities (see [17] for a more technical survey), their construction.

6.3. **Moving to the universal family.** Roughly speaking, curves of genus g are parametrized by an algebraic family $\mathbb{M}_g$ and we have a universal curve $\mathfrak{C}_g \to \mathbb{M}_g$, where the preimage in $\mathfrak{C}_g$ of a point in $\mathbb{M}_g$ is isomorphic to the curve parametrized by the point. We have also a universal abelian variety $\mathfrak{A}_g \to \mathbb{A}_g$, where $\mathbb{A}_g$ is a parameter space for abelian varieties of dimension g and above each $t \in \mathbb{A}_g$ lies a copy of the abelian variety corresponding to t (note that for readability, we are not discussing principal polarization and level structure).

We connect these two universal objects as follows. For $m \geq 1$ and any parameter t of $\mathbb{M}_g$ we may define a map

$$(P_0, \ldots, P_m) \to (P_1 - P_0, \ldots, P_m - P_0);$$

the image is a list of m points on the Jacobian of the curve. This induces the *Faltings-Zhang* map

$$\mathfrak{C}_g^{m+1} \to \mathfrak{A}_g^m$$

on (products of) the universal objects. Just as in Silverman's setting, we have a natural notion of height on $\mathfrak{A}_g^m$ arising from the Néron-Tate height on each abelian variety.

$\mathbb{A}_g$ itself is quasi-projective and defined over a number field so admits a height function $h : \mathbb{A}_g(\overline{\mathbb{Q}}) \to \mathbb{R}$, from which we obtain a notion of height for the Jacobian of a curve defined over a number field; we write $h([\mathrm{Jac}(C)])$ for this quantity. The constant $B(C)$ which discriminates small from large points in fact takes the form $c_1 \max\{1, h([\mathrm{Jac}(C)])\}$ for a constant c_1 depending only on g (there is also a dependence on the choice of height function we place on $\mathbb{A}_g$, but this is no obstruction to a uniformity result).

In this language, Theorem 4.3 comes down to a statement of the following form, which provides uniform control over the small points.

Theorem 6.2. *Given $g \geq 2$, there exist constants c_1 and c_2 depending only on g so that for any curve C of genus g and any algebraic point P of C, we have*

$$\left|\left\{Q \in C(\overline{\mathbb{Q}}) : \hat{h}(Q - P) \leq c_1 \max\{1, h([\mathrm{Jac}(C)])\}\right\}\right| < c_2.$$

There are two key inputs to the proof of this theorem, which we'll describe with the remaining two subsections.

6.4. **Degeneracy and the Betti form.** The first key input, due to Dimitrov-Gao-Habegger, is a proof of Theorem 6.2 under the assumption that $h([\mathrm{Jac}(C)])$ is not too small. This follows from a statement very similar to that of Silverman's variation of canonical height, controlling the height of a point in the total space in terms of the height of the parameter. In the case of Silverman's theorem, we can use such a statement to deduce that a point of small height sits above a parameter of small height precisely when the limit - that geometric quantity which is a function field height - is non-zero. In the full generality of Dimitrov-Gao-Habegger, they replace this geometric quantity with a notion called non-degeneracy, and crucially provide both analytic and algebraic descriptions of when a subvariety is non-degenerate.

This notion of non-degeneracy can be analytically described by the *Betti rank* of the subvariety at any smooth point. If $\mathcal{A} \to S$ is a family of abelian varieties of dimension g parametrized by S, we may at any point $s \in S$ choose a basis $\omega_1(s), \ldots, \omega_{2g}(s)$ of the period lattice for A_s, the abelian variety parametrized by s_0. From the discussion of subsection 3.1, this choice varies holomorphically with the parameter s on a simply connected neighborhood $\Delta \subset S$, and for any $s \in \Delta$ and any point $x \in A_s$ we can write

$$x = \sum_{i=1}^{2g} b_i(x)\omega_i(s),$$

where $(b_1(x), \ldots, b_{2g}(x)) \in \mathbb{R}^{2g} \mod \mathbb{Z}^{2g} =: \mathbb{T}^{2g}$ The map

$$b_\Delta : \Delta \to \mathbb{T}^{2g}$$

is the Betti map associated to Δ (though it is only unique up torus automorphisms).

As a very basic example to see that the Betti map might tell us something about torsion and height along subvarieties of $\mathcal{A}$, notice that if we consider an algebraic torsion section X in $\mathcal{A}$ - for example, the subvariety of identity points - then the Betti map on simply connected $\Delta \subset X$ will take a constant value. To that end, we define the *Betti rank* of an irreducible subvariety X of $\mathcal{A}$ at a smooth point $x \in X(\mathbb{C})$ to be the ($\mathbb{R}$-)rank of $\mathrm{d}b_\Delta \mid_{X^{\mathrm{sm,an}}}$ at x, for any appropriate choice of Δ and b_Δ.

We then say X is *non-degenerate* if there is any smooth point with maximal Betti rank; that is, the Betti rank is $2\dim(X)$ at x for some smooth point x of X. This notion of non-degeneracy is most easily understood in the situation when $\mathcal{A}$ is a family of *simple* abelian varieties of dimension g; that is, the generic element of the family has no proper abelian subvarieties, assuming the family is finite over the universal abelian variety. In that case, non-degeneracy of a subvariety X is simply the restriction that the group generated by X is not contained in any proper subvariety of $\mathcal{A}$ (note that the identity example above fails this requirement) and that $\dim X \leq g$. In practice, the constraint $\dim X \leq g$ can be circumvented by taking fibered products.

There is a useful characterization of non-degeneracy in differential geometric terms. By a construction of Mok [33], there is a closed semi-positive smooth $(1,1)-$form ω on $\mathcal{A}^{\mathrm{an}}$ so that for any subvariety X of $\mathcal{A}$ and any smooth point x of X, we have maximal Betti rank at x in X if and only if $(\omega|_X^{\wedge \dim X})_x \neq 0$.

These descriptions of non-degeneracy are fundamentally analytic; however, one would hope - and indeed Gao proved [16] - that they have an algebraic interpretation which (very!) roughly amounts to a statement like: 'when we throw away all the pieces of X which come from abelian subvarieties, there is something left over'. The proper phrasing of this statement is non-trivial; to quote Gao [17], we will say only that Gao proves that to each subvariety X of $\mathcal{A}$, there is an intrinsically defined algebraic subvariety X^{deg} with the property that $X \neq X^{\mathrm{deg}}$ if and only if X is non-degenerate in the Betti sense.

Under reasonable geometric conditions (generalizing the finiteness and group-generation properties in the simple abelian case) on X, we have two methods to construct non-degenerate subvarieties from X: one by taking mth fibered powers of X and $\mathcal{A}$ for $m \geq \dim S$, and the other as the image of the Faltings-Zhang map for $m \geq \dim X$.

We can now formulate the generalization of Silverman's theorem proved by Dimitrov-Gao-Habegger; rather, a slight improvement from the exposition of [17]. Recall our setup that $\pi : \mathcal{A} \to S$ is a family of abelian varieties over an irreducible projective variety S, on which we define a fiberwise height function $\hat{h}(P) := \hat{h}_{A,\Theta}(P)$, where A is the abelian variety parametrized by $\pi(P)$. As a projective variety we may also choose a height function h_S on the base S.

Theorem 6.3. *[Dimitrov-Gao-Habegger] Let X be an irreducible subvariety of $\mathcal{A}$ defined over $\overline{\mathbb{Q}}$, and let $X^* = X \setminus X^{deg}$. There exist constants $c > 0$ and c',* *depending only on X and the choice of height functions, so that*

$$\hat{h}(P) \geq ch_S(\pi(P)) - c' \qquad \forall P \in X^*(\overline{\mathbb{Q}}).$$

Note here the resemblance to Silverman's theorem - away from a suitable locus (in Silverman's case, the sections where $\hat{h}_{\mathcal{E}_\eta}(\sigma_\eta) = 0$), the fiberwise height $\hat{h}$ of a point must grow as the corresponding parameter does. From this, they deduce Theorem 6.2: when the height of the Jacobian of a curve is sufficiently large, so must be the Néron-Tate height of the difference of points on the curve.

Crucial to this deduction is the construction of a suitable non-degenerate subvariety arising from curves inside their Jacobian. While the theorem is trivial if X is degenerate, the statement of the theorem is carefully constructed this way to allow the proof to proceed by induction on the dimension of the base S. Their argument uses the geometric information of non-degeneracy of X to deduce positivity properties of well-chosen divisors (contrast this with the discussion of subsection 5.4 and the choice of the diagonal divisor) which in turn provide lower bounds on associated height functions; a key tool here is Siu's criterion,which allows the positivity to be checked via intersection theory.

From this work already a strong uniformity statement of Mordell-Lang type can be deduced: that Theorem 4.3 holds under the assumption that the height of the Jacobian of the curve is not too small. The work of Kühne [26] on the uniform Bogomolov conjecture provides sufficient control over points of bounded height on families of curves in their Jacobians to strengthen this to the full statement of Theorem 6.2 and so of Theorem 4.3.

6.5. **Small points and equidistribution.** An abelian variety is a complex torus $\mathbb{C}^g/\Lambda$ for a lattice Λ and therefore comes with a natural notion of volume: the *Haar measure* on the abelian variety, a translation-invariant measure which can be normalized to have total volume 1. For any $N \in \mathbb{N}$, the torsion points of order dividing N are the images under the quotient map of points $z \in \mathbb{C}^g$ with $Nz \in \Lambda$, and these are evenly distributed with respect to the Haar measure: the number of N-torsion points contained in a subset of a fundamental domain is asymptotic to the volume of the subset, as $N \to \infty$.

This is an illustration of a much broader phenomenon of equidistribution of points of small height, initiated by the work of Szpiro-Ullmo-Zhang [40] and Zhang [43] studying points of small height on abelian varieties. First, we make precise the notion that a sequence of points is 'evenly distributed' with respect to a measure. In the case of torsion points above, notice that we clearly must take the complete set of N-torsion points to obtain an even distribution; in the general setting, we consider Galois orbits of points.

Torsion points have Néron-Tate height 0; more generally, if A is an abelian variety defined over a number field K equipped with Néron-Tate height $\hat{h}_A$, we say that a sequence $x_n \in A(\overline{K})$ is *of small height* if $\hat{h}(x_n) \to 0$ as $n \to \infty$. Since A is defined over a number field K, for each $x \in A(\overline{K})$ we have a finite set $\mathrm{Gal}(\overline{K}/K)x$ called the *Galois orbit* of x; these are the points of $A(\overline{K})$ which satisfy the same polynomial relations with coefficients in K as x does for example, since the group law is defined over K, this means that a torsion point will have Galois orbit consisting of torsion points of the same order. In this setting, Zhang proved the following equidistribution theorem.

Theorem 6.4. *[Zhang] Let $\{x_n\}$ be a sequence of points of small height in $A(\overline{K})$, and assume that no subsequence is contained in a proper subvariety of A. Then the Galois orbits are equidistributed with respect to the Haar measure μ_A on A; that is, for any continuous function f on A we have*

$$\frac{1}{|\mathrm{Gal}(\overline{K}/K)x|} \sum_{y \in \mathrm{Gal}(\overline{K}/K)x} f(y) \to \int_A f \; d\mu_A$$

as $n \to \infty$.

This equidistribution statement provides an alternative proof of the Manin-Mumford conjecture (Raynaud's theorem) via contradiction: if a curve defined over a number field has infinitely many torsion points when embedded in its Jacobian, then that sequence satisfies the requirements of the equidistribution theorem, so has Galois orbits which equidistribute to the Haar measure on the Jacobian. This measure is supported everywhere on the Jacobian; in particular, the small points cannot be trapped on the curve itself, the needed contradiction. In fact Zhang proved a more general equidistribution theorem: that for any subvariety $X \subset A$, there is a unique probability measure to which Galois orbits of any small and generic sequence of points can equidistribute.

A clever double application of this equidistribution theorem allowed Zhang to conclude a stronger statement which had been known as the *Bogomolov conjecture*: that if X is a subvariety of A which is not a torsion translate of an abelian subvariety, then there is $\epsilon > 0$ so that the set

$$\{x \in X(\overline{\mathbb{Q}}) : \hat{h}_A(x) \leq \epsilon\}$$

is contained in a proper subvariety of X. If X is a curve, this is therefore a finite set.

It is this work that Kühne upgraded to the setting of families of abelian varieties, using the notion of non-degeneracy and the Betti form to replace the requirements on X and the Haar measure on A. More precisely, Kühne proved for a family $\mathcal{A} \to S$ of abelian varieties defined over $\overline{\mathbb{Q}}$ that for any non-degenerate subvariety X, any generic sequence of points of small height (here, the fiberwise Néron-Tate height) in X Galois-equidistributes to the measure $(\omega|_X)^{\wedge \dim X}$, where ω is the Betti form on $\mathcal{A}$.

This equidistribution theorem was first proved in a special case by DeMarco-Mavraki [8] and has since been generalized broadly by Yuan-Zhang [42]. Following the 'double equidistribution' strategy of Zhang, Kühne deduced a uniform Bogomolov statement for curves in their Jacobians. As in Theorem 6.2, the constants depend on choices of heights, which does not concern us for the Mordell-Lang application.

Theorem 6.5. *[Kühne] Given $g \geq 2$, there exist constants c_3, c_4 depending only on g so that for any curve of genus g and any algebraic point P of C, we have*

$$\left|\{Q \in C(\overline{\mathbb{Q}}) : \hat{h}(Q - P) \leq c_3)\}\right| < c_4.$$

In rough outline, the double application of equidistribution works as follows. We may construct a family $\mathcal{A} \to \mathfrak{C}_g$ over the universal curve and a subvariety $X \subset \mathcal{A}$ so that the fiber of $X \to \mathfrak{C}_g$ over a point P is precisely the curve containing P, which via $X \subset \mathcal{A}$ is embedded into its Jacobian via the Abel-Jacobi embedding based at P. By the construction of non-degenerate subvarieties given in the previous subsection, taking mth fibered powers of X and $\mathcal{A}$ for m suitably large yields a non-degenerate subvariety Y of a fiber power of $\mathcal{A}$.

Now, this non-degenerate subvariety comes with a natural measure from the Betti form which by Kühne's work is the limit of any small generic sequence of points, and the same is true taking further fibered products of Y. On the other hand, on the $(n+1)$st fiber power of Y we have the nth Faltings-Zhang map, and the image this map comes with its own unique measure via the equidistribution theorem (in fact, Kühne here needs to consider a product of the identity and Faltings-Zhang map to obtain non-degeneracy). Pulling this measure back under that map, we get a second measure on a large power of Y. Examining these measures at the (power of the) smooth point where full Betti rank is achieved by the non-degeneracy of Y, one can show that they are not equal.

Since the measures are distinct, they can be distinguished by some continuous function. Therefore, equidistribution tells us that a sequence of points of small height cannot have image under the Faltings-Zhang map which consists of points of small height. But differences of points and Néron-Tate height are controlled by the parallelogram law (see subsection 5.3); in particular, we have

$$\hat{h}(P-Q) \leq 2\hat{h}(P) + 2\hat{h}(Q),$$

and so small points *do* have Faltings-Zhang images which are small height. Carefully leveraged, this yields Theorem 6.5.

With some additional manipulation of constants (see section 9 of [17]), this combines with Theorem 6.3 to provide Theorem 6.2, and the uniform Mordell-Lang conjecture follows.

References

1. E. Bombieri, *The mordell conjecture revisited*, Annali della Scuola Normale Superiore di Pisa-Classe di Scienze **17** (1990), no. 4, 615–640.
2. E. Bombieri and W. Gubler, *Heights in diophantine geometry*, no. 4, Cambridge university press, 2006.
3. C. Chabauty, *Sur les points rationnels des courbes algébriques de genre supérieura l'unité*, CR Acad. Sci. Paris **212** (1941), no. 1, 882–885.
4. R. F Coleman, *Effective chabauty*, Duke Math. J. **52** (1985), no. 3, 765–770.
5. S. David, M. Nakamaye, and P. Philippon, *Bornes uniformes pour le nombre de points rationnels de certaines courbes*, Diophantine geometry **4** (2007), 143–164.
6. S. David and P. Philippon, *Minorations des hauteurs normalisées des sous-variétés des puissances des courbes elliptiques*, International Mathematics Research Papers **2007** (2007), rpm006.
7. L. DeMarco, H. Krieger, and H. Ye, *Uniform manin-mumford for a family of genus 2 curves*, Annals of Mathematics **191** (2020), no. 3, 949–1001.
8. L. DeMarco and N. M. Mavraki, *Variation of canonical height and equidistribution*, American Journal of Mathematics **142** (2020), no. 2, 443–473.
9. ———, *Dynamics on $\mathbb{P}^1$: preperiodic points and pairwise stability*, Preprint (2023).
10. V. Dimitrov, Z. Gao, and P. Habegger, *Uniform bound for the number of rational points on a pencil of curves*, International Mathematics Research Notices **2021** (2021), no. 2, 1138–1159.

11. ______, *Uniformity in mordell–lang for curves*, Annals of Mathematics **194** (2021), no. 1, 237–298.
12. G. Faltings, *Endlichkeitssätze für abelsche varietäten über zahlkörpern.*, Inventiones mathematicae **73** (1983), 349–366 (ger).
13. ______, *Diophantine approximation on abelian varieties*, Annals of Mathematics (1991), 549–576.
14. C. Favre and J. Rivera-Letelier, *Équidistribution quantitative des points de petite hauteur sur la droite projective*, Mathematische Annalen **335** (2006), 311–361.
15. Z. Gao, *Generic rank of betti map and unlikely intersections*, Compositio Mathematica **156** (2020), no. 12, 2469–2509.
16. ______, *Mixed ax–schanuel for the universal abelian varieties and some applications*, Compositio Mathematica **156** (2020), no. 11, 2263–2297.
17. ______, *Recent developments of the uniform mordell-lang conjecture*, arXiv preprint arXiv:2104.03431 (2021).
18. Z. Gao, T. Ge, and L. Kühne, *The uniform mordell-lang conjecture*, arXiv preprint arXiv:2105.15085 (2021).
19. Z. Gao and P. Habegger, *Heights in families of abelian varieties and the geometric bogomolov conjecture*, Annals of Mathematics **189** (2019), no. 2, 527–604.
20. T. Gauthier and G. Vigny, *The geometric dynamical northcott and bogomolov properties*, arXiv preprint arXiv:1912.07907 (2019).
21. P. Habegger, *Special points on fibered powers of elliptic surfaces*, Journal für die reine und angewandte Mathematik (Crelles Journal) **2013** (2013), no. 685, 143–179.
22. M. Hindry, *Autour d'une conjecture de serge lang*, Inventiones mathematicae **94** (1988), no. 3, 575–603.
23. M. Hindry and J. H. Silverman, *Diophantine geometry: an introduction*, vol. 201, Springer Science & Business Media, 2013.
24. A. Hurwitz, *Über algebraische gebilde mit eindeutigen transformationen in sich*, Mathematische Annalen **41** (1892), no. 3, 403–442.
25. E. Katz, J. Rabinoff, and D. Zureick-Brown, *Diophantine and tropical geometry, and uniformity of rational points on curves*, Algebraic geometry: Salt Lake City 2015 **97** (2018), 231–279.
26. L. Kühne, *Equidistribution in families of abelian varieties and uniformity*, arXiv preprint arXiv:2101.10272 (2021).
27. S. Lang, *Division points on curves*, Annali di Matematica Pura ed Applicata **70** (1965), 229–234.
28. B. Lawrence and A. Venkatesh, *Diophantine problems and p-adic period mappings*, Inventiones mathematicae **221** (2020), 893–999.
29. D. Masser and U. Zannier, *Torsion points on families of squares of elliptic curves*, Mathematische Annalen **352** (2012), no. 2, 453–484.
30. David W Masser, *Specializations of finitely generated subgroups of abelian varieties*, Transactions of the American Mathematical Society **311** (1989), no. 1, 413–424.
31. B. Mazur, *Arithmetic on curves*, Bulletin of the American Mathematical Society **14** (1986), no. 2, 207–259.
32. J.-F. Mestre, *Courbes elliptiques de rang ≥ 12 sur $\mathbb{Q}(t)$*, Comptes rendus de l'Académie des sciences. Série 1, Mathématique **313** (1991), no. 4, 171–174.
33. N. Mok, *Aspects of kähler geometry on arithmetic varieties*, Several complex variables and complex geometry, Part **2** (1991), 335–396.
34. L. J. Mordell, *On the rational resolutions of the indeterminate equations of the third and fourth degree*, Proc. Cambridge Phil. Soc., vol. 21, 1922, pp. 179–192.
35. D. Mumford, *A remark on mordell's conjecture*, American Journal of Mathematics **87** (1965), no. 4, 1007–1016.
36. M. Raynaud, *Sous-variétés d'une variété abélienne et points de torsion*, Arithmetic and Geometry: Papers Dedicated to IR Shafarevich on the Occasion of His Sixtieth Birthday Volume I Arithmetic (1983), 327–352.
37. G. Rémond, *Décompte dans une conjecture de lang*, Inventiones mathematicae **142** (2000), no. 3, 513–545.
38. J. H. Silverman, *Heights and the specialization map for families of abelian varieties.*, (1983).

39. M. Stoll, *Uniform bounds for the number of rational points on hyperelliptic curves of small mordell-weil rank.*, Journal of the European Mathematical Society (EMS Publishing) **21** (2019), no. 3.
40. L. Szpiro, E. Ullmo, and S.-W. Zhang, *Équirépartition des petits points*, Inventiones mathematicae **127** (1997), 337–347.
41. P. Vojta, *Siegel's theorem in the compact case*, Annals of Mathematics **133** (1991), no. 3, 509–548.
42. X. Yuan and S.-W. Zhang, *Adelic line bundles over quasi-projective varieties*, arXiv preprint arXiv:2105.13587 (2021).
43. S.-W. Zhang, *Equidistribution of small points on abelian varieties*, Annals of mathematics **147** (1998), no. 1, 159–165.

DPMMS, Wilberforce Road, Cambridge UK CB30WB

Email address: hkrieger@dpmms.cam.ac.uk

PASSING A CURVE THROUGH N POINTS – SOLUTION OF A 100-YEAR-OLD PROBLEM (BY E. LARSON AND I. VOGT)

RAVI VAKIL

ABSTRACT. Through two randomly chosen points in the plane, indeed in n-space, there is a line. Through five randomly chosen points in the plane, there is a conic. But in higher-dimensional space, through even four randomly chosen points, there can't be a conic, because all conics have to lie on a plane, and four randomly chosen points don't. Through four randomly points in the plane, you can find a cubic $y = ax^3 + bx^2 + cx + d$, and if you're not looking just for a graph of an equation, you can find a cubic $ax^3 + bx^2 + \cdots + hx + iy = 1$ through nine randomly chosen points.

For centuries, the "interpolation problem" has arisen in many contexts: is there a curve of some "type" through a bunch of generally chosen points? Even making this precise has led to important definitions and notions and theorems. In this talk, I'll discuss Eric Larson and Isabel Vogt's proof of the interpolation problem (in its modern incarnation) in full generality, bringing on beautiful ideas both old and new. We will start with some elementary observations and intuitions, and gently build up to some of the ideas behind their tour-de-force solution.

CONTENTS

Date: November 8, 2022.

1. Introduction

Last year, Eric Larson and Isabel Vogt posted the final paper [LV2] in their solution to the "interpolation problem" for curves in space. This current note (and accompanying lecture) is intended to give a broad mathematical audience an introduction to why this is such a fascinating problem, of longstanding interest, and to give some insight into the ideas behind both the problem and the solution. Not incidentally, I hope to give the reader some sense of the kinds of thinking that go into such algebro-geometric problems and solutions, and especially of the perspectives that carry over into other fields of mathematics.

For rigorous detail, I refer the reader to Larson and Vogt's papers, starting with [LV2] and working backwards. Better still is to watch a lecture by one of the authors (either in person, or one of the recorded versions online). For a broader overview aimed at a more general audience, I recommend Jordana Cepelewicz's excellent article in Quanta [Ce].

2. What is the interpolation problem, and what is the motivation for it?

The problem is about interpolating curves through points. Let me begin with some versions of this, and then see how the problem naturally extends.

Suppose I have n distinct points in the plane $\mathbb{R}^2$, and I want to pass a curve through it (see Figure 1). Obviously I can do so, given a pencil and enough hand-eye coordination. That's not interesting. But can I pass a *line* through them simultaneously? If $n > 2$, then I'd have to be very lucky to be able to — the points would have to be collinear. If $n = 1$, it is very easy, and there are infinitely many lines that will do the trick (and if $n = 0$, it is easier still). Clearly $n = 2$ is the important "edge-case" — I can manage it, but only one line does the job (Figure 1).

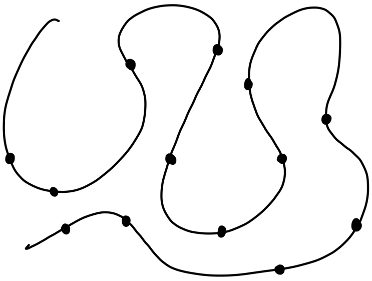


FIGURE 1. Passing a curve through 13 points in $\mathbb{R}^2$

But if my curve can be of higher degree, I can reach more points.

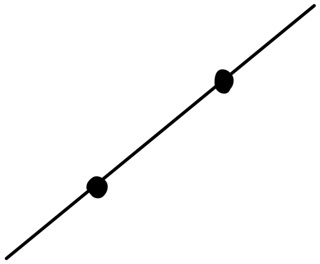


FIGURE 2. There is a unique line through two distinct points in the plane (Euclid's *Elements*, c. 300 BCE)

A generalization can be interpreted as "Lagrange interpolation": if I am looking for a polynomial $f(x)$ satisfying n constraints

$$f(x_1) = y_1, \ f(x_2) = y_2, \ \dots \ f(x_n) = y_n,$$

can I find such a polynomial of degree d? For a general such problem you can find a polynomial of degree $d = n - 1$ or greater, but not of any smaller degree.

Let's unpack this a bit, because there are a number of lessons hidden here.

I can make up an example where this is *never* possible, if you choose your constraints badly — there is no polynomial for which $f(0) = 1$ and $f(0) = 2$ for example. And if you choose your constraints very well, you may be able find an interpolating polynomial of degree *smaller* than $n - 1$ — it is not hard to find a polynomials satisfying $f(x) = x$ for the integers between 1 and 5 for example (Figure 3). But there *is* a correct answer that applies for a "generally chosen" set of constraints. We can make this precise in a number of ways — "almost all" choices of constraints, a "randomly chosen" choice of constraints, a Zariski-dense subset of the parameter space of constraints, etc. All that matters is that you see that the proper formulation of the question (at least to first approximation) is that the constraints be "general" in some sense we can make precise on another day. So from now on we will make that assumption.

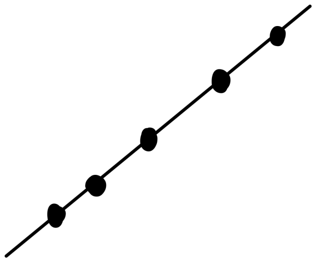


FIGURE 3. If you are lucky, you may be able to find a line through five points in the plane (Vakil's *Elements*, c. 2024 CE)

An interesting but tangential point (or should it be "tangential line"?) is that if $d = n-1$, then (i) there is precisely *one* polynomial of degree d through these points. And also, (ii) we can write down explicitly what that polynomial is. For example, the unique quadratic $f(x)$ in $\mathbb{C}[x]$ satisfying $f(x_1) = y_1$, $f(x_2) = y_2$, and $f(x_3) = y_3$ is

$$f(x) = \frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)}y_1 + \frac{(x-x_1)(x-x_3)}{(x_2-x_1)(x_2-x_3)}y_2 + \frac{(x-x_1)(x-x_2)}{(x_3-x_1)(x_3-x_2)}y_3.$$

These two facts (i) and (ii) are tightly intertwined, by a basic algebro-geometric principle. (Similarly, when there were are solutions to an algebro-geometric problem, you might expect to need a single square root in the solution, and the quadratic equation is the first example of this principle.)

Next, consider five generally chosen points $(x_1, y_1), \ldots, (x_5, y_5)$ in $\mathbb{R}^2$. Lagrange interpolation shows that we can string a "degree 4 polynomial" $y - f(x) = 0$ through them. But we can manage something of lower degree — of degree 2 (see Figure 4). Let me show this to you in two different ways, which will later give us two different insights.

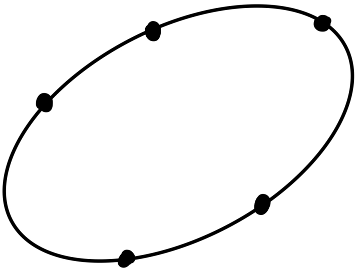


FIGURE 4. There is a (unique) conic through five generally chosen points (Pappus, c. 340 CE)

First, we can try to find a quadratic in two variables that vanishes at these five points. In other words, we're looking for 6 numbers a, b, c, d, e, f such that our five points (x_i, y_i) satisfy the equation

$$ax^2 + bxy + cy^2 + dx + ey + f = 0. \tag{1}$$

It looks like we have one too many unknowns. But multiplying the equation of a conic by a constant yields the same curve. (Translation: the plane conics are parametrized not by $\mathbb{R}^6$, but by $\mathbb{RP}^5$. Notice how we are forced to consider the notion of projective geometry, just as the ancients were!) So we may as well set one of the variables to 1, say $f = 1$. (Remember that our points are randomly chosen, so none of them is $(0, 0)$.) We then have five linear equations in the five unknowns $a, b, \ldots, e$. These turn out to have a unique solution. (It is worth thinking through why.) So we can find a degree 2 curve through the points.

Now let us instead think of the problem parametrically, by asking if there is a way of passing a pencil through these points, in a "degree 2 way". The right way of posing the question turns out to be the following: Are there three degree two polynomials $X(t) = at^2 + bt + c$, $Y(t) = dt^2 + et + f$, and $Z(t) = gt^2 + ht + i$ such that the parametrized curve

$$(X(t)/Z(t), Y(t)/Z(t)) \tag{2}$$

passes through the five given points? This turns out to be the same question, because such parametrized curves are precisely the plane conics. It looks like we have 9 unknowns $a, \dots,$ i, which is well over our five constraints. We realize quickly that we can knock this down to 8 unknowns, because we can assume $i = 1$ (in a similar way to what we did with plane conics), but we are still off by three unknowns. It turns out that there is a 3-dimensional way of parametrizing a conic, and that is where these 3 extra choices come from. (This 3 is precisely the dimension of the automorphism group of $\mathbb{P}^1$. This was also well-known to the classical Greeks, although not in this language.)

A new behavior arises in degree 3. There is a single cubic passing through 9 generally chosen points $(x_1, y_1), \dots, (x_9, y_9)$:

$$ax^3 + bx^3y + \cdots + i = 0. \tag{3}$$

As with the case of the conic, this becomes a question of linear algebra. However, there is not a degree 3 *parametrized curve* passing through 9 generally chosen points — you can only pass one through 8 generally chosen points, and in this case there turn out to be 12 of them (not 1, as in all of the other problems we have considered so far). (Incidentally, this 12 is the magic 12 that turns up whenever elliptic curves are lurking in the shadows.) What is going on?

The answer is helpfully explained by the complex picture. Every parametrized curve (2) of any degree is basically the Riemann sphere (except we remove finitely many points where the denominators $Z(t)$ are 0). But a cubic curve (3) is torus (a "genus 1 Riemann surface"), again minus finitely many points. So the difference in behavior in the *problem over* $\mathbb{R}$ is explained by the *same problem over* $\mathbb{C}$ — and is yet another reason why we decided to work over the complex numbers. The key additional information here is that our curves now have a notion of "genus" (the number of holes), and cubic curves can have genus 0 or genus 1. Genus 0 cubic curves can interpolate 8 general points in the plane, and genus 1 cubic curves can interpolate 9 general points in the plane. (See Figure 5 for a sketch of a genus 2 Riemann surface.)

Working with the complex numbers resolves a problem that you might have asked about even in degree 2. A conic over $\mathbb{R}$ can have two "pieces" if it is a hyperbola. If the conic through our five points is a hyperbola, and the points are not all on one branch, does that count as stringing a single curve through the five points? For now let's say "yes". And one important and good way around this is to work over the complex numbers, rather than the real numbers.

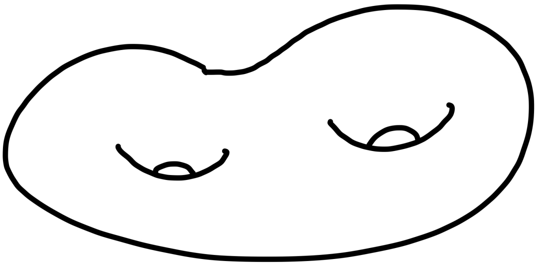


FIGURE 5. A genus 2 Riemann surface

(There are many clues in algebraic geometry telling us that it is easier to work over the complex numbers, even if we initially care about the real numbers. Frankly the real numbers are much harder to deal with sometimes, and may not be worth the trouble.) So from now on we work over $\mathbb{C}$ rather than $\mathbb{R}$ (and it is important to note that this doesn't change any of our earlier discussion).

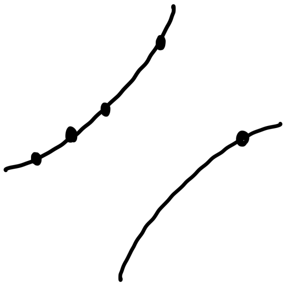


FIGURE 6. This counts as a single curve (conic) through 5 points

Now I can tell you the full story of degree d curves in the plane. Such curves can have genus g anywhere between 0 and $(d-1)(d-2)/2$ inclusive. (A generally chosen degree d polynomial in two variables has genus $(d-1)(d-2)/2$. At the other extreme, a "parametrizable" degree d curve, known also a rational curve, has genus 0.)

2.1. **Theorem (interpolation for degree d genus g plane curves).** — *Plane curves of degree d and genus g can interpolate up to $3d-1+g$ generally chosen points in the plane.*

Our tour so far has taken us through a vast territory and chronology of mathematics, from millennia past through the late nineteenth century. (For background, see [Cr, W, Ca, HB, B, L].)

3. The interpolation in higher dimension: Curves in space

When we start to consider curves in space (and higher-dimensional space), things get even more interesting. If the previous section connects many important themes in mathematics leading up to 1900, this section connects central ideas in algebraic geometry in the twentieth century. I will continue to focus on the mathematics rather than giving complete citations and history.

We now consider curves, in space of any dimension. We continue to work over the complex numbers (although we will discuss other fields in §9). Rather than working in $\mathbb{C}^r$, it is more convenient (and logically equivalent) to work in $\mathbb{CP}^r$. (This is precisely the same way that the ancient Greeks knew that to understand conics, one should think of them in projective plane.)

Curves in projective space still have a "degree". One interpretation of the degree is as follows: intersect your curve with a random hyperplane in $\mathbb{CP}^r$ (or $\mathbb{C}^r$ if you prefer) and count the number of points.

But there are still very different sorts of curves of the same degree d. In particular, curves in $\mathbb{CP}^r$ still have a genus g (an "intrinsic" property of the curve) on top of their degree (something "extrinsic", relating to how it is sitting in r-space). So we can ask: for how many points can we string a degree d genus g curve in $\mathbb{CP}^r$ through? If $r = 2$, this is precisely the question discussed in §2.

This leads us to consider the "space" parametrizing such curves (a "moduli space" of degree d genus g curves in $\mathbb{CP}^r$). We call this space $\mathcal{M}_g(\mathbb{CP}^r, d)$. (Technically speaking, this notation should be reserved for the space of degree d maps of genus g smooth Riemann surfaces to $\mathbb{CP}^r$, which is almost the same thing.)

Another new feature of curves of different genus is that there is basically one curve of genus 0; and a one-parameter family of curves of genus 1 (parametrized by the "j-invariant"); and a $(3g-3)$-parameter family of curves of genus g. The space of genus g Riemann surfaces is denoted $\mathcal{M}_g$, so for example $\dim_{\mathbb{C}} \mathcal{M}_3 = 6$, and $\mathcal{M}_g \cong \mathcal{M}_g(\mathbb{CP}^r, 0)$.

And oddly, once the genus gets bigger, different curves of genus g can move differently in projective space (and can even move in different-dimensional families). This is central to the study of algebraic curves, and this area of study is called *Brill-Noether theory*. (Another fun side fact: the spaces $\mathcal{M}_g(\mathbb{CP}^r, d)$ are not only non-smooth, but are "as non-smooth as can be" in precise sense, which Mumford described as "Murphy's Law".)

The celebrated foundational result of Brill-Noether theory tells us whether there is a map from a generally chosen genus g curve to $\mathbb{CP}^r$ of degree d. We need to say this a bit more precisely, because we could map the curve to $\mathbb{CP}^2$ (or even to $\mathbb{CP}^1$), and then embed that $\mathbb{P}^2$

(or $\mathbb{P}^1$) linearly in any $\mathbb{CP}^r$ for $r \geq 2$. We wish to ask about maps that are "non-degenerate", which means that they don't map to any hyperplane inside $\mathbb{CP}^r$. The essential fact for our story is the following.

3.1. **Brill-Noether Theorem (Griffiths-Harris, 1980,** [GH]**). —** *There exists a nondegenerate degree* d *map from a generally chosen genus* g *curve* C *to* $\mathbb{CP}^r$ *if and only if the* **Brill-Noether number**

$$\rho(d, g, r) := (r+1)d - rg - r(r+1)$$

is nonnegative.

We now fit this into the language we are developing. The moduli space $\mathcal{M}_g(\mathbb{P}^r, d)$ has many different "pieces" (or "irreducible components"), and we are interested in those whose image is dense in $\mathcal{M}_g$ (i.e., which include a "general genus g curve"). That set (or more precisely, the closure of that subset of the moduli space) will be denoted $\mathcal{M}_g(\mathbb{P}^r, d)^{BN}$. Such curves in space are called **Brill-Noether curves**.

In this language, the Brill-Noether Theorem 3.1 the following

3.2. **Brill-Noether Theorem (reworded). —** *The space* $\mathcal{M}_g(\mathbb{P}^r, d)^{BN}$ *is nonempty if and only if* $\rho(d, g, r) \geq 0$.

We finally can have a well-posed interpolation question. Let $\mathcal{M}_{g,n}(\mathbb{P}^r, d)^{BN}$ be the parametrizing degree d Brill-Noether curves $C \to \mathbb{P}^r$ along with n points $p_1, \ldots, p_n$ of C.

3.3. *Question.* Suppose d, g, r, n are nonnegative integers with $\rho(d, g, r) \geq 0$. When does there exist a degree d genus g Brill-Noether curve through n generally chosen points in $\mathbb{CP}^r$? In other words, when is the image of the natural map

$$(p_1, \ldots, p_n) : \mathcal{M}_{g,n}(\mathbb{P}^r, d)^{BN} \longrightarrow (\mathbb{P}^r)^n$$

dense?

We can show using the deformation theory that the space $\mathcal{M}_g(\mathbb{P}^r, d)^{BN}$ of Brill-Noether curves has dimension

$$\dim_{\mathbb{C}} \mathcal{M}_g(\mathbb{P}^r, d)^{BN} = (r+1)d - (r-3)(g-1).$$

(All of our dimensions are "complex dimensions", since everything in sight will be a complex manifold, or essentially so. Hereafter we omit the subscript $\mathbb{C}$ on "dim".) We can use this to get a cheap bound on the answer to Question 3.3.

The space $\mathcal{M}_{g,n}(\mathbb{P}^r, d)^{BN}$ parametrizing Brill-Noether curves with n points $p_1, \dots, p_n$ has dimension

$$\dim \mathcal{M}_{g,n}(\mathbb{P}^r, d)^{BN} = (r+1)d - (r-3)(g-1) + n. \tag{4}$$

You should think of it as follows. The curve C can move in a family of dimension

$$\dim \mathcal{M}_g(\mathbb{P}^r, d)^{BN} = (r+1)d - (r-3)(g-1),$$

and at the same time, the n points are independently wandering around on the curve C.

3.4. *Turning out the lights.* If you put light bulbs on the n points $p_1, \dots, p_n$, and turn off the lights, you will see the n points moving around in space, with the curve barely visible in the dimness. You might hope that the n points are moving around completely independently. If they aren't, there should be some reason for it that is somehow visible, or that we can figure out. For example, if the curve C was a line ($d = 1$), we would see even with the lights out that those points were all obviously collinear as they wandered about.

3.5. *A necessary condition for interpolation.* So if those n points can sweep out all n-tuples of points in $\mathbb{CP}^r$, we necessarily need

$$(r+1)d - (r-3)(g-1) + n \geq nr \tag{5}$$

from which

$$(r-1)n \leq (r+1)d - (r-3)(g-1). \tag{6}$$

In other words, if inequality (6) is violated, there is no chance for a generally chosen Brill-Noether curve (of degree d, genus g, in $\mathbb{CP}^r$) to pass through n generally chosen points.

So the big question is: what if the inequality (6) is satisfied? Is it always possible to interpolate?

(To make sure you understand what we are doing, you should substitute $r = 2$ and compare it to our discussion in §2.)

After some thought you probably won't see any counterexamples, leading you to conjecture the following.

3.6. **Conjecture (naive interpolation). —** *If inequality* (6) *holds, then given* n *generally chosen points in* $\mathbb{CP}^r$*, there is a degree* d*, genus* g *Brill-Noether curve passing through them.*

Sadly, the world is not our friend.

4. The four exceptional cases, and the statement of Larson-Vogt's Interpolation Theorem

Here are four cases where inequality (6) is satisfied, but we *can't* string such a Brill-Noether curve through n points, and the reason is not as obvious as in the case of §3.4. The four cases are

$$(d, g, r) \in \{(5, 2, 3), (6, 4, 3), (7, 2, 5), (10, 6, 5)\}.$$

But in each of the four cases, there is a good explanation. I'll describe one case so you can see how more structure can emerge from the darkness.

For the case $(5, 2, 3)$, we are considering genus $g = 2$ curves of degree $d = 5$ in $\mathbb{CP}^3$ (so $r = 3$). From inequality (6), we see we can only hope to pass such a curve through n general points if $n \leq 10$. But it turns out that we *can't* pass a curve through 10 generally chosen points, because of a hidden structure present, which we now describe.

A genus 2 genus 5 curve C necessarily sits on a degree 2 surface in $\mathbb{CP}^3$ (a "quadric surface"). To show this requires some theory of Riemann surfaces.

Now if the coordinates of $\mathbb{C}^3$ are x, y, and z, a quadric surface has defining equation

$$?x^2 + ?xy + ?y^2 + ?yz + ?z^2 + ?zx + ?x + ?y + ?z + ? = 0,$$

which has ten unknowns. We can take the constant term to be 1 (as in our discussion of conics at equation (1), so we have only 9 degrees of freedom. Hence this quadric surface can only pass through 9 generally chosen points (and will then necessarily miss the tenth, which was after all randomly chosen), so there is no chance that genus 2 degree 5 curve C (which after all must lie on a quadric surface) can pass through 10 generally chosen points.

There are similar reasons for the other three counterexamples, and they all have to do with hidden geometric structure in the problem. The nature of that geometric structure depends on your point of view. From one point of view there is always some surface causing the problem (playing the role of the quadric surface). In each case, the surface is itself parametrizable, which is somehow not a coincidence (although in another sense all four exceptions are by definition exceptional and hence coincidences).

Faced with four counterexamples, we cleverly update our conjecture.

4.1. **Interpolation Conjecture.** — *If inequality* (6) *holds, then given* n *generally chosen points in* $\mathbb{CP}^r$, *there is a degree* d, *genus* g *Brill-Noether curve passing through them. Well, except for those four counterexamples.*

This is what Larson and Vogt prove.

5. A first idea, and a better statement

Let's return to our picture of watching the points on the Brill-Noether curve move in the dark. We wish to show that the points move as freely as possible. In the case of our first counterexample, described in §4, how might we have seen that there was going to be a problem? How would we have discovered the quadric surface by just looking at the curve (and the n points, which really don't matter for this question)? The answer is that if you deform (wiggle) the curve C itself, it has a special set of deformations that keep it inside the quadric surface that we had yet to discover. So the clue is that when we look at the curve in 3-space, there is a 2-dimensional bit of surface near the curve that we can somehow see. This can be stated as follows — the curve has a rank 2 "normal" vector bundle, and inside it there is a special rank 1 "subbundle" (the normal bundle of the curve in the surface) that should have been identifiable in some way. And indeed this subbundle is identifiable by "cohomological means".

5.1. **The more correct statement.**

As always in mathematics, to better find the truth, we need to find the correct statement of the result. The following is a mild variation of the discussion at §3.5. If inequality (6) is satisfied, we might hope that the n points on a Brill-Noether curve can move freely in projective space (with the expected number of degrees of freedom), as the curve and points vary. This is the statement we should wish to prove, and we have to avoid counterexamples. As described earlier, there are four counterexamples we know of. But there is a fifth one. If $(d, g, r) = (6, 2, 4)$ (in English: genus 2 curves of degree 6 in $\mathbb{CP}^4$), then inequality (6) (or equivalently, inequality (5)) suggests that there is such a Brill-Noether curve passing through nine general points, and one generally chosen line. But this is not the case. (See [LV2, §2.1] for an explanation.)

6. The classical approach: "deformations"

Fix now g, d, r, and n. There is an approach to such questions that has been used for a very long time, presumably even in the 1800's or earlier. Suppose you had a single Brill-Noether curve (of genus g, degree d, in $\mathbb{CP}^r$), with n points on it, and if when you deform it ("jiggle it in all possible infinitesimal ways"), the n points can move in all possible directions. More precisely, suppose that the space of Brill-Noether curves has dimension D, and the choice of motion of points (in the direct sum of their "tangent spaces" in $\mathbb{CP}^r$, say $\oplus_{i=1}^n T_{\mathbb{CP}^r} p_i$) is a subvector space of dimension

$$\min\left(D + n, \dim\left(\oplus_{i=1}^n T_{\mathbb{CP}^r} p_i\right)\right) = \min(D + n, nr).$$

Then even though this is just is an infinitesimal statement, it is enough to show that the points can move in an honest family of that dimension, and if $D + n \le nr$, it means that the points can move basically everywhere independently. (Because everything is ″algebraic″, the locus they can move in is ″algebraically describable″, and the only algebraically describable sets containing an analytic open neighborhood are “almost everything”.) This means that they can move to pass through ″generally chosen″ n points, so ″we can interpolate″.

So now the problem has turned into finding a *single* Brill-Noether curve, with a *single* collection of n points, which can ″deform well″. This still seems very difficult, because it is hard to name *any* Brill-Noether curve of given genus and degree if g, d, and r are large numbers.

Let's look more closely at this problem. A neighborhood of the curve C in projective space looks a lot like the normal bundle of C in $\mathbb{CP}^r$ (the “tubular neighborhood theorem”). So a small deformation of the curve C should essentially be seen as a “section of the normal bundle”.

6.1. (For experts: It is worth taking a brief break to note that this is how the dimension of the space of Brill-Noether curves given in (4) can be found. One can show that $H^1(C, N_{C/\mathbb{CP}^r}) = 0$ for such curves; and then use the Riemann-Roch theorem to compute $h^0(C, N_{C/\mathbb{CP}^r}) - h^1(C, N_{C/\mathbb{CP}^r})$ to be the right side of (4).)

So if we want the n points to move freely in r-space as the curve moves freely, and the points move freely on the curve, we are hoping to see that given any elements of the normal vector spaces at those n points, there is a section of the normal bundle that takes on those values. (This is something you have to see to believe, which is why you need to attend the lecture. The argument will be in the form of an interpretative dance.)

Translation: The sections of the normal bundle form a vector space $H^0(C, N_{C/\mathbb{CP}^r})$. We want the map

$$H^0(C, N_{C/\mathbb{CP}^r}) \longrightarrow \oplus_{i=1}^n N_{C/\mathbb{CP}^r}|_{p_i}$$

to be surjective. Translation: we want the space of sections of the normal bundle which happen to vanish at the n points to be a codimension nr subspace in $H^0(C, N_{C/\mathbb{CP}^r})$ (the conditions imposed are “independent”). The space of sections of the normal bundle which happen to vanish at the n points can be interpreted as the sections of a slightly different bundle (the normal bundle “twisted by these n points”). So the question is now understanding the sections of this “modified normal bundle”.

Now comes the second trick, very much in Larson-Vogt's territory: our *curve* can be very special, and we see this in an example. Suppose we wished to show that conics in the plane can pass through five general points. I will now show you a very special conic passing through a very special set of five points.

For the conic C, I take the union of the x- and y-axes, i.e., the equation $xy = 0$. This is not a smooth curve, but it has the mildest possible singularity, what is called a "node" at the origin. (One definition of "node" is something analytically that looks like the union of the two axes at the origin.) For the points, I take $(1, 0)$, $(2, 0)$, $(3, 0)$, $(0, 1)$, and $(0, 2)$. You will notice that I have taken smooth points of the curve C (I have avoided the node), and that I have split up the five points as evenly as I can among the two "pieces" ("irreducible components") of C.

So we could apply our normal bundle strategy to this nodal curve, and hope for the best. The downside is that the tubular neighbourhood theorem no longer works where the curve isn't a manifold; but that doesn't matter much, because we have passed the stage of needing the tubular neighborhod theorem, and now only need to work with some modified normal bundle constructed in some abstract way. The notion of normal bundle turns out to make sense near the node (so that it is still a bundle, even though the tubular neighborhood theorem, again, fails).

But now we want to do some sort of induction, which means trying to change the problem into one about each of the two "pieces" ("irreducible components") of the degenerated conic, so we turn it into a question about lines, which are easier to understand.

The normal bundle to the x-axis in the plane is very similar to the restriction of the normal bundle to this degenerated conic in the plane — it is precisely the same except at $(0, 0)$, where it is some sort of modification. So at the hoped-for inductive step, instead of the normal bundle to the "simpler piece", we get a modified version of the normal bundle to the simpler piece, so the curve is simpler but the modifications needed to the normal bundle sound more severe.

But we have already resigned ourselves to trying to understand (the sections of) modified normal bundles, so at this point we realize that our induction is going to have to deal with curves in $\mathbb{CP}^r$ and given modifications of the normal bundle.

7. Completing the proof

All we have to do now is to make this precise, make the inductive statement and prove the theorem, and we could leave this as an exercise for the reader.

Unfortunately, this strategy roughly describes where Larson and Vogt were near the start of their approach. Except for their clever focus on the the fact that the interpolation problem was about modified normal bundles, and that this was something they could control well enough to induct on, it describes an attack that was undoubtedly tried by many authors in the 1980's or earlier.

8. Slalom skiing

I will conclude by describing the nature of the obstacles to making this work. The details of the analysis, and the tactical decisions that made them work, are less appropriate for an expository lecture.

For a Brill-Noether curve, the normal bundle is well-behaved — for example, it has no higher cohomology (in particular, H^1), as alluded to in §6.1. But as soon as you modify it, you are in danger of "unbalancing" the bundle to lose this niceness — or at the very least, to lose the ability to show that the bundle retains this niceness.

But on the other hand, you want to break the curve into simpler pieces, and eventually into lines, so you can end up with a base case you can work with.

And furthermore, when you break the curve into pieces, you want the resulting pieces to also be "Brill-Noether curves", and it is harder to make sense of what this means once the curves are so special as to be broken, particularly when the very definition of Brill-Noether curves is that they were "generally chosen". How can you tell when a highly degenerate curve is the limit of a Brill-Noether curve?

In the end, Larson and Vogt's induction resembles a difficult slalom, winding between these two dangers as your induction downwards careens toward your desired base case. As one might expect, as the induction gets closer and closer to the base case, their options for degeneration become more and more limited, and at many times it looked like there was no way forward, and they were guaranteed to crash into a tree. This is why there are particularly clever or delicate arguments in small-dimensional projective spaces, e.g., in $\mathbb{CP}^4$, see [LV1]. The interest of this paper isn't because $\mathbb{CP}^4$ is so gosh-darn interesting, but because the obstacles come hard and fast in that dimension.

9. Working over other fields

At the start, I gave reasons to work over $\mathbb{C}$ instead of over $\mathbb{R}$. A lot of arguments and intuitions involved "wiggling" or "deforming", so a priori we can't seem to make these arguments in other fields, where we can't use "convergence" or other analytic tools. Surprisingly, all of the techniques can extend to work over arbitrary algebraically closed fields (and even fields that are not algebraically closed if you say things with great care). Better yet, Larson and Vogt's arguments apply in this vast generality. Working in this generality actually helps (at least in some vague moral sense) because it makes clearer what geometrically is relevant to the problem, just as making the correct more general statement in §5.1 led us (or more correctly, Larson and Vogt) to the correct attack on the problem. And although the correct statement led to a fifth counterexample, knowing the fifth counterexample helped lead to a correct inductive strategy (since Larson and Vogt had to somehow dodge the fifth counterexample).

Similarly, if you extend the problem to arbitrary fields, you also get an additional family counterexamples to the statement they wished to prove, but this additional family helps think about the correct argument (at least in retrospect). The family of counterexamples is so basic that I am surprised (even amazed) that it wasn't well-known before — these aren't exotic curves.

In general, if you have a genus 0 (smooth) curve (over any algebraically closed field!), every vector bundle splits into a direct sum of line bundles, and you expect the "general one" to be "balanced" — the line bundles should either be all the same, or at least "differ by the least possible nonzero amount". (Translation for experts: the degrees of any two of them differ by at most one.)

9.1. *Example* [LV2, Thm. 1.4]. If k is an algebraically closed field of characteristic 2, and $C \subset \mathbb{CP}^r_k$ is a curve of genus 0, and degree d such that $d-1$ isn't divisible by $r-1$, then the normal bundle of C in $\mathbb{CP}^r_k$ is *not* balanced.

I leave this as a riddle for those experienced in this sort of arithmetic geometry: why would you expect this to be true?

10. Conclusion

The work of Larson and Vogt not only completes this long open programme, it really tied it up with a bow, giving a motivation and explanation for the counterexamples, as well as a structure for why the result holds.

I find Larson and Vogt's result very attractive, because the question is so motivated and multifaceted, and because the solution is a mixture of the concrete roll-up-your-sleeves get-your-hands-dirty argumentation, mixed with the use of judiciously deployed big and abstract machinery. Their general strategy is a frontal blunt unsubtle attack on the problem, but their tactics require precision and care. In the end, the approach is theoretical and pencil-and-paper, but at the same time they were unafraid of writing code to convince readers that they didn't miss any small extra edge case.

10.1. **Acknowledgements.** In preparing for the talk, I benefited from both Eric Larson and Isabel Vogt explaining their thoughts in detail, which ended up being far more for my own benefit than for yours, dear reader, because I tried to give you a birds-eye overview of the story for non-experts. I am also forever grateful to J. Harris, the teacher of the three of us, for showing us by example how to think about such beautiful things.

References

[B] G.D. Birkhoff, *General mean value value and remainder theorems with applications to mechanical differentiation and quadrature*, Trans. Amer. Math. Soc. **7** (1906), no. 1, 107–136.

[Ca] A.-L. Cauchy, *Cour d'analyse de l'École Royale Polytecnique; I.re Partie. Analyse algébrique*, Chez Debure Frères, Libraires du Roi et de la Bibliothèque du Roi, Paris, 1821.

[Ce] J. Cepelewicz, *Old problem about mathematical curves falls to young couple*, Quanta Magazine, Aug. 25, 2022.

[Cr] G. Cramer, *Introduction à l'analyse des lignes courbes algébriques*, Frères Cramer & Cl. Philbert, Geneva, 1750.

[GH] P. Griffiths and J. Harris, *On the variety of special linear systems on a general algebraic curves*, Duke Math. J. **47** (1980), no. 1, 233–270.

[HB] M. Ch. Hermite and M. Borchardt, *Sur la formule d'interpolation de Lagrange*, J. Reine Angew. Math. **84** (1878), 70-79.

[L] J. Lagrange, *Leçons élémentaires sur les mathématiques données à l'École Normale en 1795*, in *Oeuvres de Lagrange*, J. A. Serret, ed., tome 7, 183–288, Paris, 1867–1892.

[LV1] E. Larson and I. Vogt, *Interpolation for Brill–Noether curves in $\mathbb{P}^4$*, Eur. J. Math. **7** (2021), no. 1, 235–271.

[LV2] E. Larson and I. Vogt, *Interpolation for Brill–Noether curves*, Forum of Mathematics Pi, **11**, 2023, e25.

[W] E. Waring, *Problems concerning interpolations*, Philosophical Transactions of the Royal Society **69** (1779), 59–67.

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.

Friday, January 6, 2023 (Boston, MA)

Andrew Granville, Université de Montréal
Missing digits, and good approximations.

Christopher Eur, Harvard University
An essence of independence: recent works of June Huh on combinatorics and Hodge theory.

Henry Cohn, Massachusetts Institute of Technology
From sphere packing to Fourier interpolation.

Martin Hairer, Imperial College London
A stroll around the critical Potts model.

Friday, January 7, 2022 (Virtual)

Tom Scanlon, University of California, Berkeley
Tame geometry for Hodge theory

Elena Giorgi, Columbia University
The stability of black holes with matter

Anup Rao, University of Washington
Sunflowers: from soil to oil

Elamin Elbasha, Merck & Co., Inc.
Mathematics and the quest for vaccination-induced herd immunity threshold

January 18, 2021 (Virtual)

Abba Gumel, Arizona State University
Mathematics of the Dynamics and Control of the COVID-19 Pandemic

Ana Caraiani, Imperial College London
An excursion through the land of shtukas

Jennifer Hom, Georgia Institute of Technology
Getting a handle on the Conway knot

Richard Evan Schwartz, Brown University
Rectangles, Curves, and Klein Bottles

January 17, 2020 (Denver, CO)

Jordan S. Ellenberg, University of Wisconsin-Madison
Geometry, Inference, and Democracy

Bjorn Poonen, Massachusetts Institute of Technology
A p-adic approach to rational points on curves

Suncica Canic, University of California, Berkeley
Recent Progress on Moving Boundary Problems

Vlad C. Vicol, Courant Institute of Mathematical Sciences, New York University
Convex integration and fluid turbulence

January 18, 2019 (Baltimore, MD)

Bhargav Bhatt, University of Michigan
Perfectoid geometry and its applications

Thomas Vidick, California Institute of Technology
Verifying quantum computations at scale: a cryptographic leash on quantum devices

Stephanie van Willigenburg, University of British Columbia
The shuffle conjecture

Robert Lazarsfeld, Stony Brook University
Tangent Developable Surfaces and the Equations Defining Algebraic Curves

January 12, 2018 (San Diego, CA)

Richard D. James, University of Minnesota
Materials from mathematics

Craig L. Huneke, University of Virginia
How complicated are polynomials in many variables?

Isabelle Gallagher, Université Paris Diderot
From Newton to Navier-Stokes, or how to connect fluid mechanics equations from microscopic to macroscopic scales

Joshua A. Grochow, University of Colorado, Boulder
The Cap Set Conjecture, the polynomial method, and applications (after Croot-Lev-Pach, Ellenberg-Gijswijt, and others)

January 6, 2017 (Atlanta, GA)

Lydia Bieri, University of Michigan
Black hole formation and stability: a mathematical investigation.

Matt Baker, Georgia Tech
Hodge Theory in Combinatorics.

Kannan Soundararajan, Stanford University
Tao's work on the Erdos Discrepancy Problem.

Susan Holmes, Stanford University
Statistical proof and the problem of irreproducibility

January 8, 2016 (Seattle, WA)

Carina Curto, Pennsylvania State University
What can topology tell us about the neural code?

Lionel Levine, Cornell University and *Yuval Peres, Microsoft Research and University of California, Berkeley
Laplacian growth, sandpiles and scaling limits.

Timothy Gowers, Cambridge University
Probabilistic combinatorics and the recent work of Peter Keevash.

Amie Wilkinson, University of Chicago
What are Lyapunov exponents, and why are they interesting?

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

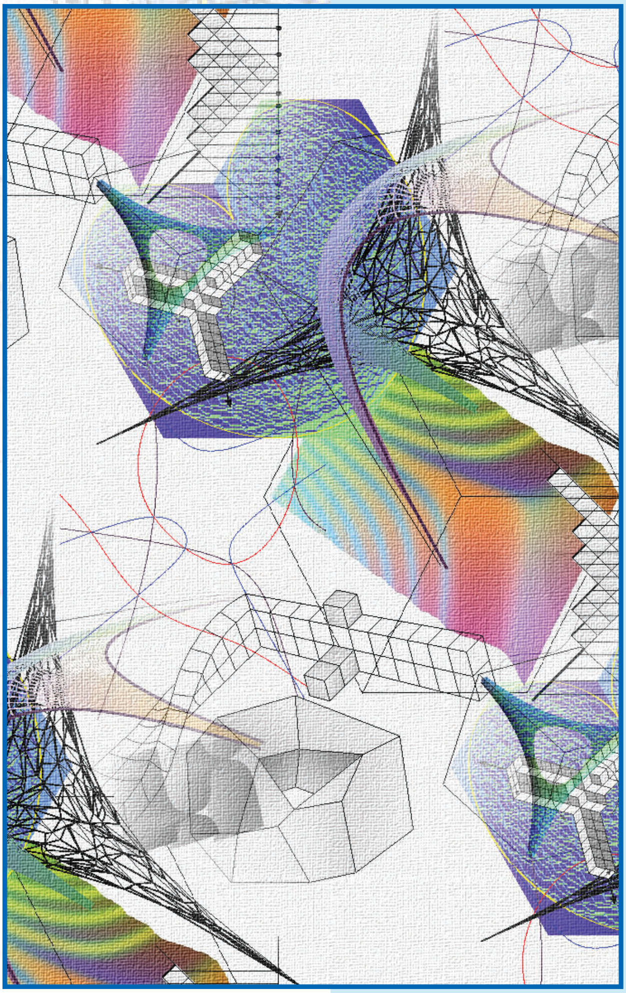
Edward Frenkel, University of California, Berkeley
Recent advances in the Langlands Program

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

2024 CURRENT EVENTS BULLETIN

Committee

Alejandro Adem, *University of British Columbia*
Hélène Barcelo, *Mathematical Sciences Research Institute*
David Eisenbud, *Chair*
Susan Friedlander, *University of Southern California*
Silvia Ghinassi, *University of Washington*
Abba Gumel, *University of Maryland*
Ursula Hamenstädt, *University of Bonn*
Helmut Hofer, *Institute for Advanced Study*
János Kollár, *Princeton University*
Andrei Okounkov, *Columbia University*
Peter Ozsvath, *Princeton University*
Avi Wigderson, *Institute for Advanced Study*



The back cover graphic is reprinted courtesy of Andrei Okounkov.

Cover graphic associated with Will Perkins's talk courtesy of Poike / iStock / Getty Images Plus via Getty Images.

Cover graphic associated with Hussein Mourtada's talk courtesy of the author.

Cover graphic associated with Holly Krieger's talk courtesy of Holly Krieger and Laura DeMarco.

Cover graphic associated with Ravi Vakil's talk courtesy of The American Mathematical Society.