ICM 2018

Notes by James H. Davenport

August 2018

Contents

Chapter 1

2 August 2018

1.1 Donaldson

1.1.1 Part I: Kähler

$n = 2m$ and X a complex manifold. Define a skew-symmetric $\omega(v_1, v_2) = \langle v_1, v_2 \rangle_g$. The Kähler metric is a fixed cohomology class $[\omega] \in H^2(X, \mathbf{R})$ are parametrised by a function.

ICM 1954 (Amsterdam) where Calabi initiated the study of existence question.

Kähler–Einstein metrics have Ricci $= \lambda g$. These can only exist when the first Chern class $c_1(M)$ is positive, negative or zero, depending on the sign of λ. In 1983 he introduced more general *extremal metrics*, which include constant scalar curvature Kähler (CSCK) metrics.

Example 1 (Classical) *$m = 1$: any compact Riemann surface admits a metric of constant Gauss (i.e. scalar) curvature, unique up to holomorphic automorphisms. Basically the Uniformisation Theorem.*

Yau (1978) solved the case $\lambda \leq 0$. Looks like the $m = 1$ case. $\lambda = 0$ gives Calabi–Yau metrics with Ricci=0 is especially important. The structure group of the tangent bundle is reduced to $SU(m) \subset U(m)$.

This existence problem can be set up as a second order nonlinear PDE (complex Monge–Ampère):

$$\det \left(\frac{\partial^2 \phi}{\partial z_a \partial z_b} \right) = e^{-\lambda \phi}.$$

So the main technique is PDE methods.

Positive λ gives Fano manifolds. But there are various obstructions. This is now understood as *K-stability*: Tian (1997) and Donaldson (2002). $(X, [\omega])$ is K-stable if, for all non-trivial degenerations χ of X have $\mathrm{Fut}(\chi) > 0$.

Conjecture 1 (YTD) *X admits a CSCK metric in the class $[\omega]$ iff $(X, [\omega])$ is K-stable.*

The Fano case what settled in 2013 by Chen–Donaldson–Sun in 2013, see 2014 ICM Székelyhidi. There are now a variety of proofs of this. Chen–Cheng have done recent work on extremal metrics.

Infinite-dimensional geometry. Let $\mathcal{H}$ be the space of Kähler metrics in a fixed class. Then there's a Mabuchi metric on this space:

$$||\delta\phi||^2 = \int_\chi (\delta\phi)^2 \operatorname*{vol}_\phi .$$

The Mabuchi functional $\mathcal{F} : \mathcal{H} \to \mathbf{R}$, whose extremal points are the CSCK. See lecture of Berndtsson in this ICM.

Theorem 1 (Berman–Berndtsson) *Convexity ...*

Roughly speaking, degenerations of X give points at infinity of $\mathcal{H}$. If there is no critical point of $\mathcal{F}$ in $\mathcal{H}$, then this can be detected by the behaviour of $\mathcal{F}$ at such points at infinity.

Riemannian convergence theory. Gromov, Cheeger–Colding. Let (M_i, g_i) be a sequence of Riemannian manifolds with volume 1, bounded diameter and bounded Ricci curvature. Then there is a subsequence that converges in the Gromov–Hausdorff sense to a metric space M_∞. This has an open dense *regular set* $M_\infty \setminus S$ with a $C^{1,\infty}$ Riemannian metric. [Fujita2015] showed "projective space maximises volume". But the proof is algebro-geometric.

Singularities and tangent cones. Let Z be a Gromov–Hausdorff limit as above. Then it is naturally a complex algebraic variety and the metric tangent cone at a point $p \in Z$ can be described in terms of a valuation ν_p on the local ring

$$\nu_p(f) = \lim_{r\to 0} \frac{\log \max_{B_{r,p}} |f|}{\dots}.$$

see Sun at this ICM. These ideas can be compared with Deligne–Mumford theory.

1.1.2 Part 2: exceptional holonomy

$\mathbf{R}, \mathbf{C}, \mathbf{H}, \mathbf{O}$=octonians are the four normed division algebras. Note that cross products only exist in $\mathbf{R}^3$ and $\mathbf{R}^7$. $\mathbf{O} = \mathbf{R}^1 \oplus \mathbf{R}^7$. The corresponding 7-manifolds are important in physics. [Bryant,ICM1986] had local examples, and [Joyce,ICM1998] had compact examples. Little is known systematically about such questions.

Let Y ve a reducible Calabi–Yau 3-fold Then there is a parallel 2-form ω on Y and a parallel holomorphic 3-form Θ ....

[Kovalev2003] Let $\overline{W}$ be a complex 3-fold with a "Lefschetz fibration". Then there are a finite number of critical values of $\overline{\pi}$, where the fibres have ordinary double point singularities.

Corti et al. have millions of deformation classes of such matching building blocks. There are 7-manifolds which are homeomorphic but not diffeomorphic.

1.2 Sylvia Serfaty: Systems of Points with Coulomb Interactions

The N-particle Coulomb kernel is $w(x) = \frac{1}{|x|^{d-2}}$ for $d \geq 3$ and $w(x) = \log|X|$ if $d = 2$.Consider the energy

$$H_n(x_1, \ldots, x_N) = \frac{1}{2} \sum_{1 \leq i \neq j \leq N} w(x_i - x_j) + N \sum_{i=1}^{N} V(x_i),$$

where $x_i \in \mathbf{R}^d$. We want minimisers (critical points) of H_n, Also evolutions $\dot{x}_i - \frac{1}{N} \nabla_{x_i} H_N$ (gradient flow), or $\ddot{x}_i - \frac{1}{N} \nabla_{x_i} H_N$ (Newton flow).

Motivation 1 (Fekete points (Approximation theory)) *Given μ, since $x_1, \ldots, x_N$ s.t.*

$$\left| \frac{1}{N} \sum_{i=1}^{N} f(x_i) - \int f(x) d\mu(x) \right|$$

is small for all regular functions f. On closed manifolds (spheres) they are maximisers of $\prod_{1 \leq i \neq j \leq N} |x_i - x_j|$.

If $s \to \infty$ this tends to close packing problems.

Motivation 2 (Vortices) *Vortices in the Ginzburg–Landau model of superconductivity, in superfluids and Bose–Einstein condensates. Vortices are the zeros of $\psi : \mathbf{R}^2 \to \mathbf{C}$ with degree in the asymptotics $\epsilon \to 0$. This can (with a lot of effort) be reduced to discrete problem.*

Motivation 3 (Statistical and Quantum mechanics) *Various:*

$d = 1, 2$ *logarithmic case and eigenvalues of random matrices problems.*

$d \geq 2$ *Classical Coulomb gas. Toy model for matter.*

$d = 2$ **logarithmic** *is "two-component plasma", with particles of $\pm$ charges we get theoretical physics models (sine-Gordon etc.)*

Problem 1 (sSmale's 7th) *compute a minimiser on the sphere up to an error $\log N$, in polynomial time.*

1.2.1 II Mean-field Limit

Limits for empirical measures $\frac{1}{N} \sum \delta_{x_i}$?

We get Frostman equilibrium measure as the unique minimiser among probabilities of

$$\mathcal{E}(\mu) = \frac{1}{2} \int_{\mathbf{R}^d \times \mathbf{R}^d} w(x - y) d\mu(x) d\mu(y) + \int_{\mathbf{R}^d} V(x) d\mu(x).$$

Makes sense only if w integrable near $0 \Leftrightarrow s < d$. Exists only if V grows fast enough at ∞.

$V(x) = |x|^2$, is the Coulomb case, then $\mu_V = \frac{1}{c_d}\mathbf{1}_{B_1}$ (**circle law**).

Potential generated by a distribution μ

$$h^\mu(x) := w * \mu = \int_{\mathbf{R}^d} w(x-y)d\mu(y).$$

Mean field force is $\nabla(h^\mu + V)$. Then $\partial_t \delta_{x(t)} = -\operatorname{div}\cdots$. But a formal proof has difficulties with passing to the limit $N \to \infty$ of the nonlinear products. Proves [Speaker2018].

For Newton's law, the formal limit is

$$\partial_t f + v \cdot \nabla_x f - \nabla(h^\mu + V) \cdot \nabla_v f = 0.$$

There are convergence proofs if the singularity is less than Coulomb, but the Coulomb case is still open.

But none of these mean field results are really specific to Coulomb .... They work for Riesz $s < d$, on integrable or ... So what is specific to Coulomb? The Coulomb kernel is the fundamental solution for the Laplace operator. so the main tool is rewriting the interaction energy. We reformulate the energy in terms of the Coulomb potential $h^\mu = w * \mu$, where $\Delta w = c_\mu \delta_0$.

1.2.2 Beyond the Mean-field limit

So what is the next order term in the expansion for H_N. Main approach is to expand the energy around μ_V and compute the energy via the potential h_N. Rewrite the next-order energy as ....

$$H_N(x_1, \ldots) = N^2 \mathcal{E}(\mu_V) - \frac{N}{2d}\log N + N^{1+\frac{s}{d}}\overline{W} + o(N^{1+\frac{s}{d}}).$$

If $d = 1$ the minimum of W over all possible configurations is achieved for the lattice $\mathbf{Z}$.

Conjecture 2 (Cohn–Kumar) *If $d = 2$ we have the triangular lattice. Also E_8 and the Leech lattice.*

These are really hard crystallisation problems. There is an announced solution ($n = 8, 24$), not yet on arXiv. Note $n = 2$ is not yet solved. But the conjecture is supported by experimental observations in Abrikosov lattice observations. We know it is the minimum in the class of lattices of fixed volume. [CasslesRankinEnnolaDiananda1950s].

In $d = 3$ we believe BCC has the min mum, but still not known.

1.2.3 With temperature

$\beta = 1/\text{temperature}$. We assume $Z_{N,\beta} = \int_{\mathbf{R}^d)^N} e^{-\beta H_N(x_1,\ldots,x_N} dx_1 \ldots, dX_N$ as the partition function. There is a "large deviations" principle.

The Gibbs measure minimises "energy $+\frac{1}{\beta}$entropy".

Theorem 2 (Leblé–Speaker2015) *Gibbs concentrates on configurations whose limiting processes P^x after zoom around x minimise*

$$\mathcal{F}_\beta(P) := \int)\Sigma(W(P^x) + \frac{1}{\beta}\,\mathrm{ent}[P^x[\Pi])dx$$

where Π *is Poisson.*

Then

$$\log Z_{N,\beta} - -\beta N^{2-\frac{n}{d}}\mathcal{E}(\mu v) + \underbrace{\frac{\beta N}{2d}}_{\text{interesting}} + \cdots$$

The distribution of the points is following the equilibrium behaviour very closely.

1.2.4 Conclusion

The analysis of Coulomb systems is at the intersection of several branches (analysis , PDE, probability, number theory, geometry) Large Coulomb systems exhibit an macroscopic behaviour which can be understood by mean field theory. The microscopic behaviour can be understood ...

1.3 Rahul Pandharipande: Geometry of the moduli of curves

Given four points, we can take the first three to $0, 1, \infty$, so $M_{0,4} \cong \mathbf{C}\P^1\backslash\{0, 1, \infty\}$ — essentially cross-ratio.

What happens in higher genus, For $g = 0$ the complex structure is unique, but this is no longer the case in higher g $\mathcal{C}$ can be viewed as an algebraic curve defined by the zero locus in $\mathbf{C}^2$ of a single polynomial $F(x, y) = 0$. Plot in $\mathbf{R}^2$ of the solutions in $\mathbf{C}^2$. $\mathcal{M}_g$ is the moduli space of Riemann surfaces of genus g, $[C] \in \mathcal{M}_g$.

Riemann essentially know that $\mathcal{M}_g$ is a complex manifold of dimension $3g-3$ ($g \geq 2$). "Riemann constructs the variations of complex structure, states this dimension, and defines the term moduli, all in a single sentence " [shown]. Last 30 years have all been about the cohomology of the moduli space.

1.3.1 Cohomology

Let $S \subset |^n \times \mathrm{Gr}(r, n)$ be the universal subbundle. $H^*(\mathrm{Gr}(n, r)\mathbf{Q})$ is generated by the Chern classes of S, which measure how much S twists. $\lim_{n\to\infty} H^*(\mathrm{Gr}(rmn)\mathbf{Q}) = \mathbf{Q}[c_1, \ldots, c_r]$.

1.3.2 What is the analogue of S

The answer is the "universal curve". $C \cong \mathcal{M}_{g,1}$. We will construct cohomology classes from an intrinsic complex line bundle on $\mathcal{C}$. Let $calL$ be the cotangent line over the universal curve.

Question 1 *Is* $R^*(\mathcal{M}_g) = H^*(\mathcal{M}_g, \mathbf{Q})$*?*

The answer is No, but Yes stably! (Mumford's conjecture).

Question 2 *What is the structure of* $R^*(\mathcal{M}_g)$*?*

We are interested in the full ideal of relations of $R^*(\mathcal{M}_g)$.

Conjecture 3 (Faber–Zagier) *Let* p *be the set of variables* p_i *where* $i \not\equiv 2 \pmod 3$*. Let*

$$\Psi(t,p) = (1 + tp_3 + t^2 p_g + \cdots) \ldots + () \ldots .$$

Define the constants

$$\log \Psi = \sum_{\sigma} \sum_{r=0}^{\infty} C_r^{tZ}(\sigma^r) p^{\sigma}$$

In particular, do the Faber–Zagier relations span the ideal of all relations. For $g < 24$ shown by Faber (lots of computer work), but otherwise open.

There are essentially three proof of Faber–Zagier, all via Gromov–Witten theory and the virtual fundamental class. We proved it via Witten's 3-spin class and Giventah–Teleman classification of semisimple CohFTs. Then [Janda2015] proved all suitable semisimple CofHTs yield exactly the Faber–Zagier relations.

A CohFT is .....

The genus 0 3-pointed map $\Omega_{0,3}$ determines a quantum product $(v_1 * v_2, v_3) = \Omega_{0,3}(v_1, v_2, v_3)$.

We seem to have a class of pure dimension, but Givental–Teleman gives a CohFT of pure dimension, and the two actually agree precisely because of Faber–Zagier.

Question 3 *What are the relations in* $\overline{\mathcal{M}}_{g,n}$*?*

Claims that the boundary strata of the moduli $\overline{\mathcal{M}}_{g,n}$ *of fixed topological type* correspond to stable graphs. He can go from graphs to the corresponding stratum (?1–1?).

The initial relation is the cross-ratio. First non-trivial found in 1996 (Chicago). A relation between seven graphs. Speaker+Belorusski1998 found a genus 2 one connection 20 graphs with small rational coefficients. So is there any structure to these formulae or coefficients? How does it connect to Faber–Zagier?

1.3.3 Pixton's relations on $\overline{\mathcal{M}}_{g,n}$

We define tautological classes $\mathcal{R}^d_{g,A}$. We have already seen series B_0, B_1. They controlled the original set of Faber–Zagier relations. They also control Pixton's relations.

$$\mathcal{R}^d_{g,A} = \sum \Gamma \in G_{g,n} \frac{1}{2^{h^1(\Gamma)}} [\Gamma, \Pi K_\nu \Pi \Psi_{\mathcal{E}} \Pi \Delta_e]_d$$

A key tool is $B_0(T)B_1(-T) + B_1(T)B_0(-T) = 2$.

Theorem 3 (Pixon–speaker–someone)

$$R^d_{g,A} = 0 \in H * 2d(\overline{\mathcal{M}}_{g,n}, \mathbf{Q}).$$

Question 4 *Are Pixton's relations complete?*

1.4 Mądry: Gradients and Flows

Core primitive: given x, find Δ such that $f(x+\Delta) < f(x)$. We're assuming our functions are locally Taylor.

Problem 2 (Maximum Flow) *Think of a directed graph G, with capacities on each edge, and special nodes S[ource] and T[arget]. Constraints are conservation at each interior node and the capacities, and we wish to maximise the total flow. If all capacities are 1, we are looking for arc-disjoint paths. Especially interested in sparse graphs.*

Classical Evan–Tarjan 1975, Karzanov1975 where $O(n^{3/2})$ for unit capacities. [GoldbergRao1998] $\tilde{\mathcal{O}} * n^{3/2} \log U)$ for general case.

Modern Let's look for ϵ-approximate algorithms. [Madry2010] etc. $\tilde{\mathcal{O}}(n\epsilon^{-1})$ time for undirected and $(1+\epsilon)$-approximate. Not have $\tilde{\mathcal{O}}((nU)^{10/7})$ for undirected case, which breaks the $\Omega(n^{3/2})$ belief.

Setting Think of flow as a vector f with f_e being flow along edge e: fix an orientation and use sign of f_e to indicate direction. Therefore we should solve

$$\min_f ||\Pi_\gamma(f) - f||_2 \text{ s.t. } ||f||_\infty \leq 1.$$

Then can do this by gradient descent. Naïvely$\tilde{\mathcal{O}}(n^2/\epsilon)$. Standard techniques get $O(n^{3/2}/\epsilon)$. Then problem is that we are trying to minimise l_∞, but our tools are l_2.

1.4.1 Directed Version

The good news is that you can transfer directed to undirected. But only works fro (near)-exact solutions. In this regime, gradient descent doesn't give enough accuracy. So consider path-following interior points methods [Kar84a, Kar84b]. But these are not natural. He has a basically classical method with three insights from interior-point and this gets the complexity.

These insights would for a wide range of graph problems. Is this a major change of paradigm for graph algorithms?

Q Implementations?

A Most graphs succumb well to classical, so it is not clear what the practical gain is.

Q Which problems?

A "If there is a linear relation in the combinatorics, come and call me".

1.5 Ambainis: Understanding Quantum Algorithms via Query Complexity

See [Amb18][1].

Notation 1 *$D(f)$ =# queries for a deterministic algorithm; $R_0(f)$ for a zero-error randomised algorithm and R_2 for a bounded-error one. Corresponding for quantum we use Q_E and Q_2.*

[Feynman1981]: simulating quantum processes on a classical computer requires exponential memory. We have search [Gro96], Logic evaluation [FKT16]. We now have 20-qubit devices by IBM, and larger (50–70) bits under construction. At this point they are too hard to simulate.

So my question is: what is the biggest advantage we can get? Note that $BQP \subset PSPACE$, so we can't expect much progress. [Simon1994] $P^A \neq BQP^A$ for an oracle A. [RazTal2018] as well.

1.5.1 Decision Trees/Quantum Algorithms

. Task, compute $f(x_1, \ldots, x_N)$ where the x_i are accessed via queries. Complexity=#queries.

Example 2 (Period finding) *We are told the sequence is periodic with period $R < \sqrt{N}$. [Sho94] this is $O(\log \log N)$ quantumly, as oppose to $O(\sqrt[4]{N})$ classically. This is the key to his factoring.*

Example 3 (Grover's search) *$\exists ?i : x_i = 1$: classically $O(N)$ for determinism.*

Deterministic we look at decision trees. A quantum state is a unit vector in C^d, with basis states $|1\rangle$ etc. Generally $|\psi\rangle = \sum_{i=1}^{d} \alpha_i |i\rangle$ with $\sum \alpha_i^2 = 1$. A quantum query algorithm inputs $|\psi\rangle$ and does a sequence of queries and quantum transformations (not depending on x_i). For partial functions, there are huge speedups, but for total the best we can do is Grover.

Example 4 ([+Aaronson]) *Task that requires 1 query quantumly, and $\Omega(\sqrt{N}/\log N)$ classically. This is "Forrelation" = Fourier Correlation. Let $|\psi\rangle = \sum x_i |i> + \sum y_i |i + N\rangle$, which is one query, then use standard SWAP technique. The classical lower bound is fairly standard.*

Theorem 4 *Any 1 query quantum algorithm can be simulated probabilistically with $O(\sqrt{N})$ queries. [BBC+01].*

Theorem 5 ([BBC+01]) *Let $p(x_1, \ldots, x_n)$ be a polynomial of degree p, evaluating in $[0, 1]$ for values in $[0, 1]$*

[1] `https://eta.impa.br/dl/043.pdf`.

Theorem 6 *Any k-query quantum algorithms can be simulated probabilistically with $O(N^{1-1/2k})$ queries.*

Conjectures this is optimal, but not proved, even for $k = 2$.

So why are partial function so much better? For total functions we have an example where it is $O(N)/O(N^{1/4})$.

And/Or trees. [Snir1985,SaksWidgerson1986]. Randomly $O(n^{0.5537\cdots})$, but deterministically $O(n)$. [GPW15]. Communication versus partition number. Want $D(f)$ to be large, but $f = 1$ easy to certify. Function of mn variables, $f = 1$ iff there is a unique all-1 column. So a certificate is $O(m + n)$ — all values in the column, and *one* zero in every other column. There's a subtle arrangement [GPW15] of linking these non-zeros.

Theorem 7 ([ABB$^+$17]) *Gaps:*

- *There exists a total Boolean function f with $Q_2(f) = \tilde{\mathcal{O}}(D^{1/4}(f))$.*
- *There exists a total Boolean function f with $R_0(f) = \tilde{\mathcal{O}}(D^{1/2}(f))$.*
- *There exists a total Boolean function f with $R_2(f) = \tilde{\mathcal{O}}(R_0^{1/2}(f))$.*

1.5.2 Computing f on most inputs

Partial: $Q = 1, R = \Omega(\sqrt{N}/\log N)$ Total functions: $Q = \Omega(\sqrt[6]{D})$.

Conjecture 4 *$D_\epsilon(F) = O(Q_\delta^c(f))$ where ϵ means computing on $1-\epsilon$ of inputs.*

1.5.3 Symmetric functions

Counting and search tend to be symmetric. Period-finding is very non-symmetric.

Conjecture 5 *For a symmetric function, then R is polynomial in Q.*

Can prove it in the case where we can also permute values.

1.6 Lower bounds for Subgraph Isomorphism problems: Rossman

PnNP and NC^{1}vP. Average case. n is size of graph, k is fixed (or slowly growing).

k-Clique Does the graph contain a complete graph of size k. Trivial n^k, or $n^{0.79k}$ with fast matrix algorithms. $n^{\Omega(k)}$ would imply $P \neq NP$.

k-Cycle on layered graphs best space complexity is $O(\log(k)\log(n))$. If this were Ω we would prove $NC^1 \neq P$.

general: SUB(G) Given a graph X and a vertex-colouring $V(X) \to V(G)$, does X contain a properly-coloured G-subgraph. Includes previous two.

Circuit complexity is an approach to complexity theory.

Theorem 8 ([Sha49]) *almost all functions on n variables require $\Theta(2^n/n)$-size circuits.*

Note that formulae are weaker than circuits as they lack memory. Formalising this is $NC^1 \neq P$.

AC^0 circuit size PARITY_n is $2^{\Theta(n^{1/(d-1)})}$. Note that our problems are monotone in the graphs: if X has it, so does any supergraph. Monotone circuits have many strong properties, and mon-P≠mon-NP.

However, though both our key problems are monotone, the "graph" is pretty jagged. So let's look at slice functions. A function $f : \{0,1\}^n \to \{0,1\}^n$ is called a k-slice iff $f(x)$ equals the O-vector if x has less than k ones and $f(x)$ equals the l-vector if x has more than k ones. That means the interesting part of f happens when x has exactly k ones.

Theorem 9 ([Ber82]) *The monotone circuit complexity of slice functions cannot be much larger than the circuit (combinatorial) complexity of these functions for arbitrary complete bases.*

1.6.1 Average case k-Clique

Seem to have a phase transition issue for "$G(n,p)$ contains a k-clique".

Conjecture 6 (after Karp) *Iterated greedy is the asymptotically optimal k-clique finding algorithm in $G(n.p)$.*

Theorem 10 (Speaker 2008) *AC^0 circuits solving k-Clique on $G(n,p)$ with probability 0.51 requires size $\Omega(n^{k/4})$.*

Theorem 11 (Speaker 2008) *AC^0 circuits solving k-Clique on $G(n,p)$ and $G(n, p + p^{1-o(1)})$ with probability 0.51 requires size $\Omega(n^{k/4})$.*

Folklore: $n^{\text{treewidth}(G)+1}$.

By divide and conquer we can do k-Cycle with a circuit of size $n^{o(1)}$ depth $O(\log k)$. They become AC^0 formulas of size $n^{O(\log k)}$.

The formula size of SUB(G) is soluble by monotone AC^0 formulas of size $O(n^{\text{tree-depth}(G)})$. Also a lower bound of $n^{\Omega(\text{tree-depth}(G)^{const})}$.

1.7 Kalai: Delegating Computation via Non-Signalling Strategies

Scenario of weak devices and powerful cloud with the data largely stored there. To do verified computation, we are going to import ideas from cryptography and quantum mechanics, even though the problem specification requires neither.

Many functions computable in time T do not have proofs verifiable in times $\prec\prec T$. [GMR88] introduced the idea of interactive proofs. The verifier can

make random choices, and only requires high probability. Also [BenOretal1988] examples of multi-prover interactive proofs. [BabaiFortnowLund1990] any proof can be made exponentially shorter.

Open Problem: Interactive $polylog(T)$ proof for any T-time S-space computation, Can do this for bounded-depth computation. Also bounded-space.

Since these are open, we'll take ideas from the two-prover setting. Hence soundness will only hold against computationally bounded adversaries. Then from quantum we can reduce to two queries, the first of which is independent of the computation, so we're almost back to the non-interactive setting.

[BielMeyerWetzel1999]: send the one server the two queries, but encrypted under FHE [Gen09]. But there were counter-examples. We had soundness when A_1 was only a function of Q_1 etc. The guarantee that we actually get is that A_1 doesn't reveal Q_2 etc., which is not the same thing. This is "non-signalling", as defined in quantum mechanics.

[KalaiRothblumRaz2014] $\forall$ T-time f, where verifier's runtime is ??-bounded, ...

There's also work on non-deterministic computations, on delegating memory and on actual efficiency. Also note that this has been a real bridge between theory and practice

Q FHE?

A You don't need the full power of FHE in practice. All you need is enough non-signalling.

Chapter 2

3 August 2018

2.1 Okounkov: New worlds for Lie Theory

See `http://math.columbia/edu/~okounkov/icm.pdf`.In particular susy gauge theories in < 4 (especially 3) space-time dimensions. I want to share my excitement about a field that is still forming. Today I am talking about supersymmetric QFT. There is a powerful idea of *duality*, generalising Langlands. So which highlights of late XXth are being generalised.

Weyl groups

braid group $B = \pi_1(\mathbf{C}^n_{reg}/W$

Hecke algebra

2.1.1 MacDonald–Cherednik Theory

Replace linear differential equation by q-difference equations. Solutions are generalisation of q-hypergeometric. Remarkable index symmetry: $P_n(q^m) = P_m(q^n)$; $P_n(x) = x^n$ are kindergarten examples.

2.1.2 Kazhdan–Lustzig theory

Describes the character of irreducible highest-weight modules.

2.1.3 Yang–Baxter equations

Started in 2D statistical mechanics. The degrees of freedom live in vector space V_i attached to the edges of a grid, and their interactions are described by a matrix R of weights attached to each vertex. R-matrices with a spectral parameter define an action of non affine Weyl group of type A by q-difference operators.

Look at Index := Even Fermion number − Odd. At the lowest energies (very large B) the states of the QFT can be described by a modulated vacuum, that

is a map f from B to the moduli space X of vacua of the theory. The amount of supersymmetry that we want makes X, ideally, a hyperkähler manifold and f a holomorphic map. This now looks like enumerative geometry. The index is the Euler characteristic of a certain coherent sheaf (or virtual $\hat{A}$-genus) on the moduli space f such. The additional grading on this index by the degree of the map can be viewed as a character of the Kähler torus $\mathcal{Z} = Pic(X) \otimes \mathbf{C}^*$.

For example, susy gauge theories contains gauge fields for a compact form of a Lie group G, matter fields in a symplectic representation M of G. Then $X - \mu^{-1}(o)//G$. Ultimately the q becomes the automorphism of $\mathcal{B} \simeq \mathbf{C}$. There are two ways of thinking of this:

- These q-difference equations generalise what we have seen before
- The whole enumerative theory may be described using certain new geometric representation theory.

Unlike Langlands duality, there is no reason why the Kähler equivariant roots have to live in spaces of the same dimension. There is a full elliptic theory of [Agaagic–O]. It controls the roots of unity analogues of characteristic $p\rangle\rangle 0$ quantisation questions for finite p.

2.2 Lawler: Critical Phenomena in Statistical Physics

$\beta = 1/T$ (T = temperature). For high temperature ($\beta < \beta_c$) long range interactions decay quickly. I am interested in behaviour around β_c. Physicists use Coulomb gas techniques to give non-rigorous predictions of critical exponents for 2D systems. From the late 1990s mathematicians have been able to make some of this rigorous, and give greater insight. Define a discrete model, let it become large (or lattice spacing $\to 0$) to approximate continuum.

Imagine a random walk in $\mathbf{Z}^2 = \mathbf{Z} + i\mathbf{Z}$. Brownian (heat equation) scaling has $\Delta T \sim (\Delta x)^2$. For a self-avoiding walk, we believe (open!) that the length $\sim n^{4/3}$. This converges to a probability measure on continuous curves $\gamma(t)$.

What happens to Loop-Erased Random Walks, which are, of course, self-avoiding. These can also be considered as Laplacian random walks. Need a random walk loop measure. "Loop soup" is a Poissonian realisation from the loop measure. At intensity λ a loop of length l appears with probability $\lambda m(l)$.

Have a (wired) uniform spanning tree chosen at random. Wilson's algorithm chooses a tree with probability $F(A)\left(\frac{1}{4}\right)^{\#(A)}$. This is the same for all trees, so we recover Kirchoff's theorem on the number of trees. $4^{\#(A)} \det(I - Q)$. Brownian loop soup measure has a restriction property: if $D \subset D'$, soup(D)=soup(D')$|_D$.

Many of these have the Domain Markov property: In the measure $\mu_D(z, w)$, given an initial segment$\gamma(s) : 0 \le s \le t$, then the distribution of the remainder depends only on $\gamma(t)$, not the history of getting to t.

For $\kappa < 8$, these are supported on curves of Hausdorff dimension$\alpha = 1 + \frac{\kappa}{8}$. Our parametrisation of Hölder u continuous for $u \leq \frac{1}{\alpha}$. α-dimensional Hausdorff measure tends to be 0. α-dimensional Minkowski content is more helpful:

$$Cont_\alpha(V) = \lim_{\epsilon \downarrow 0} \epsilon^{\alpha-2} Area\{z : dist(z, V) < \epsilon\}.$$

SLE_κ is weakly Hölder-continuous up to $\frac{1}{\alpha}$. It's still an open problem to prove the convergence of self-avoiding walks with $\kappa = \frac{8}{3}$.

Take any finite simply-connected $\subset \mathbf{Z}^2$. containing origin. There exists c_*, u such that the probability that a LERW from x to w (both on boundary) goes through the origin is $\ldots(\theta)$, where θ is the angle between z and w at the origin.

There are other loop measures.

Take a very fine lattice of infinitesimal length $1/N$ (non-standard analysis) can choose mesoscopic and multi-scale analysis by choosing $N \rangle\rangle M \rangle\rangle 1$.

2.2.1 Higher dimensions

For LERW and SAW the upper critical dimension is 4. This is because random walk paths have fractal dimension2, and the critical dimension for the intersection of two-dimensional sets is $d = 4$. For $d = 4$ we expect convergence to Brownian motion with logarithmic corrections to the scaling. For LERW the mean square distance after n steps is ....

Three dimensions is the domain for ICMs of the future, Conformal invariance doesn't work. The mean square distance is predicted to be $2^{2\nu}$. SAW: originally believed $\nu = 3.5$, but now .588.... For LERW we get $\nu = 0.61\ldots$, and a rigorous proof that $\nu > 0.6$.

2.3 Moreira: Dynamical Systems, Fractal Geometry and Diophantine Approximations

Poincaré's restricted three-body work was the start of dynamical systems. Hyperbolic system introduced by Smale, especially Smale Horseshoe.

Conjecture 7 (Palis–Smale) *The structurally stable dynamical systems are the hyperbolic ones.*

Proved by[Mañé1988] for a special case. Homoclinic bifurcations are the most important mechanisms for creating complicated systems from simple ones.

Consider $\phi_\mu : M^2 \to M^2$, hyperbolic fro $\mu < 0$. Homoclinic tangency at $\mu = 0$.

Fractal sets appear naturally. Hence needs Hausdorff dimension $:= \inf\{s > 0; \inf_{X \subset \cup B(x_n, r_n)} \sum r_n^s\} = 0\}$. Regular Cantor sets, e.g. the usual ternary set. A Horseshoe Λ in a surface is locally diffeomorphic to the Cartesian product of two regular Cantor sets: the stable and unstable ones K^s/K^u. The $HD(\Lambda) := HD(K^s) + HD(K^u)$ is important.

Use equivalent of Erdős probabilistic method.

2.3.1 II: Diophantine

Dirichlet's Theorem. Hurwitz improvement. $\left|\alpha - \frac{p}{q}\right| < \frac{1}{\sqrt{5}q^2}$ and $\sqrt{5}$ is best possible. But fix α, and let best possible be $k(\alpha)$. Indeed $k\left(\frac{1+\sqrt{5}}{2}\right) = \sqrt{5}$. Note $k(\alpha) = \infty$ is possible, indeed these have Lebesgue measure 1.

Theorem 12 ([Markov,1975]) *$L \cap \{-\infty, 3) = \{$finite set$\}$. These are $\sqrt{9 - \frac{4}{z^2}}$ when $z : \exists x, y : x^2 + y^2 + z^2 = 3xyz$.*

Note there's a unicity conjecture for $x^2 + y^2 + z^2 = 3xyz$.

The biggest half-line contained in L is $[c, \infty]$ with $c \approx 4,527 = \frac{A+B\sqrt{452}}{C}$. So there's a gap.

$k(\alpha) = \limsup(\alpha_n + \beta_n)$ where α_n/β_n is the residuum of the CF approximation. Replace lim sup by sup and get the Markov set $M \supset L$. Now have some results on $HD(M \setminus L)$. Can interpret these sets as maximum heights (resp. asymptotic heights) in a modular space. Nice film demonstrating these.

2.4 Ventakesh (Fields): Cohomology of Arithmetuc Groups

Consider a form like $10x^2 - 14xy + 5y^2$. This is the same as $u^2 + v^2$ with $x := v - 2u; y := v - 3u$. All my QFs are +ve definite. All forms are equivalent to $ax^2 + bxy + cy^2$, with $|b| \leq a \leq c$. Minkowski generalised to n variables. Image of Cassels' *Rational Quadratic Forms*. The group $\Gamma = S :_n (\mathbf{Z})$ of invertible integral linear transformation acts on the space S of positive definite quadratic forms, and we are looking for a *fundamental domain*, i.e. a region that tessellates space. The number of faces is at least $\sum$ Betti(S/T). kth Betti number is $\dim H^k(S/T, \mathbf{R})$. Group cohomology assigns a family of vector spaces to a group Γ. $H^1(\Gamma, \mathbf{R}$ is the space of homomorphisms: $\Gamma \to \mathbf{R}$.

Example 5 *2×2 matrices with bottom left divisible by 11.*

If S/T is a s a complex algebraic variety, we know more in this case. A precise conjecture, based on a small number of examples which all check out.

For $x^3 - x - 1 \pmod p$ the algorithm computes the discrete log $\log_p x$ in terms of a derived Hecke operator.

2.5 Chenyang Xu: Interaction between singularity theory and the minimal model program

Given $x \in X$ a singularity, we would resolve it: find a morphism $f : Y \to X$ with Y smooth, and a lower-dimensional $Z \subset X$ such that f is isomorphic over

$X \setminus Y$, Then X ad T are birational. In characteristic 0 this is always possible by Hironaka's Resolution of Singularities Theorem. However, when $\dim \geq 3$, there is often no "best" (i.e. minimal) Y.

Given a normal space X, and $\mathbf{Q}$-Gorenstein (i.e.the class given by K_X is torsion in the class group). Let $f : Y \to X$ be a resolution, such that $Ex(f)$ is a simple normal crossing. Write $K_{Y/X} = \sum_i a_i(X : E_i)E_i$ then we say X is

terminal $a_i(X : E_i) > 0$

canonical $]a_i(X : E_i) \geq 0$

Kawanata log terminal $]a_i(X : E_i) > -1$

log canonical $]a_i(X : E_i) \geq -1$

for all i.

Each class is preserved under the minimal model. Giving an explicit classification is too complication for $\dim > 3$, We can do the same for a pair (X, Δ).

Define a regular cell complex $D(\Delta)$ to characterise how the E_i intersect each other. For each E_i we put a vertex v_j for each component $E_i \cap E_j$, an edge for each component $E_i \cap E_j \cap E_k$ etc. So $\dim D(\Delta) \leq \dim(X) - 1$. We have shown $D(\Delta)$ does not depend on $X^{\cdots}$: call it $DMR(x \in X)$.

Theorem 13 *$D(Ex(f))$ admits a strong deformation retract to $DMR(x \in X)$.*

There is a general local–global principle that there is a correspondence between klt singularities and Fano, strict lc singularities and Calabi–Yao.

Let E over X have $a(E) = -1$, then we can find a model $g : Z \to X$ such that $g^*(K_X + c \cdot V(f)) = K_Z + c \cdot g_v^{-1}(V(f)) + E$, and restricting to E we get a log Calabi–Yao pair.

Theorem 14 *If $I \subset [0, 1]$ is a set satisfying DCC, then*

- *$V(l, n) := (volK_x + \Delta))$ where $\dim(X) = n$, coefficients of Δ and in I, (X, Δ) is log canonical) form a DCC set. In particular $V(l, n) \cap (0, \infty)$ has a minimum.*
- *There exists $N = N(l, n)$ such that if $K_x + \Delta$ is big ...*

Conjecture 8 (Stable Degneration) *Given any klt singularity $x \in X$ up to a rescaling there is a unique (up to scaling) minimiser $v \in Val_{X,x}$ over $\check{vol}_{X,x}$ which is quasi-monomial, with a finitely-generated associated graded ring $R_0 = gr_v(R) := \bigoplus \ldots$.*

We know existence (Blum), and (us) semi-stability implies minimising. Answering a conjecture of Donaldson–Sun, uniqueness among K-semistable valuations. The missing part is that minimiser is quasi-monomial, and the associated graded ring is finite-dimensional.

Question 5 *How to check an example of a Fano variety if K-(semi,poly)stable.*

Question 6 *Using K-(semi,poly)stability to construct a projective moduli space of X-polystable Fano varieties.*

2.6 Kurdyka/ From continuous rational to regulous functions

Example 6 (Cartan's umbrella) $x^3 - z(x^2 + y^2) = 0$. *Looks like two components, but not distinguishable.*

So we are searching for real analogues of complex utopia: arc-symmetric sets and arc-analytic functions.

A subset $E \subset \mathbf{R}^n$ is *arc-symmetric* if for every analytic arc $\gamma : (-1, 1) \to \mathbf{R}^n$, with $\gamma((-1, 0)) \subset E$, we have $\gamma((-1, 1)) \subset E$.

Theorem 15 (BierstoneMilman) *Let X be a* smooth *real algebraic set. A semialgebraic function $f : X \to \mathbf{R}$ is arc-analytic iff it is blow-Nash.*

Let $X \subset \mathbf{R}^n$ be a real algebraic set and $f : W \to \mathbf{R}$ be a function defined on $W \subset X$. f is (regular) at $w \in W$ if there are two polynomial functions p, q on $\mathbf{R}^n$ ....

Let Y be the Zariski closure of W in X. A rational function R on Y is said to be a rational representation of f if there is a Zariski open dense subset $Y^0 \subset Y \setminus Pole(R)$ such that $f|_{W \cap Y^0} = R|_{W \cap Y^0}$. Then the following conditions are equivalent ....

Example 7 (Kollár) $S := (x^3 - (1 + z^2)y^3 = 0) \subset \mathbf{R}^3$ *and let* $f : S \to \mathbf{R}$. $f(x, y, z) = (1 + z^2)^{1/3}4$. $Sing(S) = z$*-axis and* $f(x, y, z) = x/y$ *on* $S \setminus z$*-axis. f is continuous and has a rational representation,* not regulous *since $f|z$-axis is not rational.*

We say f is curve-regulous if for every irreducible algebraic curve $C \subset X$ the restriction $F|_{W \cap C}$ is regulous. Arc-regulous if this happens in neighbourhoods. Say k-regulous of regulous and C^k. Let $R^k(U)$ be the ring of k-regulous functions on an open $U \subset |R^n$. Such functions are semi-algebraic and arc-analytic. $R^k(U)$ is not Noetherian of $n \geq 2$. ∞-regulous = regular, so we only consider k finite.$A \subset |R^n$ s *constructible* f it belongs to the Boolean algebra generated by the algebraic subsets of $\mathbf{R}^n$. The Euclidean closed constructible subsets of $\mathbf{R}^n$ are precisely the closed sets of a Noetherian topology.

Theorem 16 (Fichoi et al, 2015) *For a subset $E \subset \mathbf{R}^n$, take*

1. *$E = Z(I)$ for some ideal I of $R^k(\mathbf{R}^n)$*
2. *$E = Z(f)$ for some function $f \in R^k(\mathbf{R}^n)$*
3. *...*

Also let I be an ideal of $R^k(\mathbf{R}^n)$. If $f \in R^k(\mathbf{R}^n)$ vanishes on $Z(I)$ then some power f^m belongs to I.

The last line is almost a Nullstellensatz.

Then Cartan's Theorems A and B are available for k-regulous sheaves. Note that this isn't obvious: they fail in the obvious (?) translation.

Conjecture 9 *Let X be a compact real algebraic set. For a continuous map $f : X \to \S^p$, the following are equivalent:*

1. *f can be approximated by regular maps;*
2. *f is homotopic to a regular map.*

$2 \Rightarrow 1$ is the hard part.

Conjecture 10 *For any pair (np) of positive integers, each continuous map $\mathbf{S}^p \to \mathbf{S}^p$ can be approximated by regular maps.*

Theorem 17 (Bochnak+K) *These two are true when $p = 1, 2, 4$.*

2.7 Global symmetry from local information: the Graph Isomorphism Theorem: Babai

This is something that essentially doesn't happen, yet here it does. Isomorphism is a bijection of vertices that preserves adjacency. Trivial bound $n!$. $\exp(O(\sqrt{n \log n}))$ [Luks1983] – moderately exponential. Group theory is the asymptotic theory of permutation groups, combinatorics of highly regular objects (Kirkman, Bose, Schur, Higman etc.). This is a project I have worked on for more than three decades, but Eureka 14 September 2015. Graphs are universal over all explicit finite structures, so semigroups is efficiently reducible to graph isomorphism. $P \subset NP \cup coNP$. $GI \in NP$, but not known to be in coNP. factoring $\in NP \cap coNP$ [Pratt1975]. NP-complete — "the hardest NP-problems". 30-colourability, Hamiltonicity of graphs. But there are problems (like GI) which are not known to be in P, or to be NP-complete.

Groups: Sym and Alt. $G_x = \{\sigma \in G : x^\sigma = x\}$ (stabiliser). Let $\phi : G \to Alt(\Gamma)$ be an epimorphism. Suppose $G \leq Sym(\Omega)$. $x \in \Omega$ is *affected* by ϕ if $\phi(G(x) \in Alt(\Gamma)$.

Theorem 18 (Unaffected Stabiliser Theorem) *Let U be the set of unaffected elements of Ω and $G_{(U)}$ the pointwise stabiliser of U. Then ... if $m \geq 2 + \log n$ (strict!).*

Therefore at least one point is affected. Uses CSFG via Schreier's Hypothesis: the outer automorphism group of every finite simple group is solvable. Not using this ends up with $m > polylog(n)$.

PATRIOTMENU is an anagram of PERMUTATION. G-isomorphism if the permutation comes from G. $\exp(\tilde{\mathcal{O}}(\sqrt{n}))$ [Babai1983], now quasi-polynomial. Note that coset intersection, centraliser in coset etc. are equivalent to string isomorphism under Karp reduction. Note that a Graph with n vertices can be encoded in a string of length $\binom{n}{2}$.

D+C: a moderate number of significantly smaller ($\leq 90\%$) instances. Branching factor $q(n)$. If q is quasipolynomial, so is f.

We can get a canonical by degree, refined by colours of neighbours until stable. This is a functor from Graphs to coloured sets. Weisfeller–Leman refine to ordered pairs by counting triples with shared base and same colour composition. There's a k-version of this. [ImmermannLaner1980s]. Computable in $n^{\Theta(k)}$ But there are CFI pairs of non-isomorphic graphs indistinguishable by k-ary WL unless $k = \Omega(n)$.

Theorem 19 *GI for vertex-coloured with bounded colour classes is in Las Vegas polynomial time.*

This actually solved CFI graphs.

"Isomorphism of graphs of bounded degree can be solved in polynomial time" [Luks1980] really used deep group theory. This works by string isomorphism on the composition factors.

- reduce to orbits
- descend to a subgroup (multiplicative cost is the index of the subgroup)
- typically descent to the kernel of action on blocks of imprimitivity.

The Luks bottleneck is Sym/Alt (Giant), see [Cameron1981] — CFSG. Then want either

- confirm $\mathrm{Aut}_G(x) \to Giant(\Gamma)$ or
- break symmetry of Γ; find $M \leq \mathrm{Sym}(\Gamma)$ s.t. ...

So use unaffected stabilisers to find canonical$k$-ary structure on Γ where $k = 3 - \log_2 n$. Then use Split-or-Johnson. Johnson Graph $J(s,t)$ with $s \geq 2t + 1$. Vertex set $= \binom{t}{t}$.. We analyse CC(coherent configuration) which are colourings of pairs stable under WL.

2.7.1 Local Certificates

Take a test set $T \subset \Gamma$. $|T| = k > 2 + \log_2 n$. Restrict G to G_T. Sat=y "T is full" if $\mathrm{Aut}_{G_T}(x) \to Giant(T)$. So we either have a fullness certificate, or converse, which is $M(T) \leq \mathrm{Sym}(T)$ s.t. ....

Local Certificates Algorithm

- $W := \emptyset$
- while (condition
- $W := Aff(A(G_T, W))$ points affected by current $A(G_T, W)$
- update $A(G_T, W)$
- end while

Why do we stop?

1. $A(G_T, H)$ becomes too small. Then $M(T) := \phi(A(G_T, W))$ and I have non-fullness

2. window stops growing. Then we have fullness, and we have deduced this global property from local information.

2.8 Chang: Conformal Geometry on 4-manifolds

[1] (M^n, g) a compact Riemannian manifold. A metric $\hat{g}$ is conformal to g if $\hat{g} = \rho g$ for some $\rho > 0$. Set $\rho = e^{2w}$, and $g_w = e^{2w} g$. Conformal means "angle preserving". In Geometric Analysis we use PDE etc. methods to study problems such as the sign of the curvature.

2.8.1 Introduction Yamabe problem

On a compact surface (M^2, g), K_g the Gaussian curvature. Gauss–Bonnet formula. Uniformisation Theorem classifies orientable (M^2, g) according to sign of $\int_M K_g dv_g$. It's really

-1 torus

1 ?

0 sphere

One can solve $K_{g_w} = c$ by a variational approach and Moser's function J_g. Also Ray–Singer–Polyakov formula.

On $M^n, g)$ with $n \geq 3$, the conformal Laplace operator $L_g = \Delta_g + c_n R_g$ when $c_n = \frac{n-2}{4(n-1)}$ and R_g is the scalar curvature. Under a conformal change $\hat{g} = u^{4/(n-2)} g$ with $u > 0$, the $L_g u = cn \ldots$

2.8.2 Compact closed 4-manifolds

Gauss–Bonnet–Chern formula:

$$8\pi^2 \chi(M) \int_m \frac{1}{2} |W_g|^2 dv_g + \int_m \frac{1}{6} (R_g^2 - 3|\operatorname{Ric}_g|^2) dv_g,$$

where χ is the Euler characteristic of M, W_g the Weyl curvature, R_g the scalar curvature and Ric the Ricci. W measures the obstruction to being conformally flat. On (m^n, g), $W_g \equiv 0$ in a neighbourhood of a point iff $g = e^{2w} \ldots$. Let $\sigma_2 = \frac{1}{6}(R_g^2 - 3|ric_g|^2)$ and conclude that $g \to \int_M \sigma_2(g) dv_g$ is also an integral conformal invariant. Then Schouten tensor $A := Rig_g - \frac{R}{2(n-1)} g$. On (M^4, g)

$$\sigma_2(g) - \sigma_2(A_g) = \frac{1}{6}(\cdots).$$

[1] Emmy Noether lecture. She gave the first ICM Plenary by a woman at ICM 1932 — Zürich.

To solve the "generalised Yamabe" problem. $\sigma_2(A_{g_w}) = const.$ This is a fully non=linear equation.

When $n = 2$ we have J_g, and $n \neq 4$ we have F_2. For $n = 4$ we need new ideas. σ_2 is linked to the Panietz operator and Q-curvature. Panietz operator (1983) on (M^n, g) for $n \geq 5$,

$$P_4^n = (-\Delta)^2 + \delta(a_n Rg + b_n \operatorname{Ric})d + \frac{n-4}{2} D_4^n.$$

He used $\hat{g} = u^{\frac{4}{n-4}} g$, hence the restriction. $P_4^n(1) = \frac{n-4}{2} Q_4^n$ so Q_4^n is immediate.

Branson pointed out that we can define these when $n = 4$. $g_w = e^{2w} g$.

$2Q_g = -\frac{1}{6}\delta R_g + \sigma_2(A_g)$. Following Moser, we define .... Theorem: $g \in \mathcal{A}$ iff there is some $g_w \in [g]$ .... There is also a uniqueness result. If $g \in \mathcal{A}$ and not S^4, then $g_w \in [g]$ with $g_w \in ....$ par There are also diffeomorphism theorem. If ..., then M is diffeomorphic to either S^4 or $\mathbf{RP}^4$. Use the Signature formula

$$12\pi^2 \tau = \int_{M^4} (||W^+||^2 = ||W^-||^2) dv$$

where $\tau = b_2^* - b_2^-$. Then there's a perturbation theorem on $\mathbf{CP}^2$. It would be an ambitious program to find the entire class of 4-manifolds with metrics in $\mathcal{A}$, and classify their diffeomorphism types by the relative size of their conformal invariants.

2.8.3 Conformal invariants on compact 4-manifolds with boundary

In 2D for $(X^2, M * 1, g)$ where the metric g is defined on $X^2 \cup M * 1$, we have a Gauss-Bonnet formula

$$2\pi ....$$

We construct (P_3, T) where P_3 has bidegree $(0, 3)$. In general the formula for T is lengthy, but when (X^4, g) is with totally geodesic boundary, that is the second fundamental term vanishes, then $T = \frac{1}{12}\frac{\partial}{\partial n} R$.

2.8.4 Conformal Compact Einstein manifolds

Given a compact (M^4, h) when is it the boundary of a conformally compact Einstein manifold (X^{n+1}, g) with $r^w g^+|_m = h$.

On a CCE manifold special r can be chosen such that $r^2 g^+$ is with totally geodesic boundary.

Example 8 *On* $(\mathbf{R}_+^{n+1}, \mathbf{R}^n, g_{\mathbf{H}})$

There are many existence (and non-existence) results.

There is a 'renormalised volume" [Maldacena1998]. For n even,

$$Vol_{g^+}(\{r > \epsilon\}) = \text{series in } \epsilon + V \cdots + L \cdots .$$

2.8.5 Compactness results for Einstein manifolds of dimension 3+1

Question 7 *Does the entire class of metrics (S^3, h) with positive scalar curvature allow CCE filling in B^4?*

The class is path-connected by a result of [F.Marques2012].

The difficulty lies in the existence of a non-local term.

$$g := r^2 g^+ = h + g^{(2)} r^2 + g^{(3)} r^3 + g^{(4)} r^4 + \cdots$$

where ....

For convenience we choose $h = h^Y \in [h]$, the Yamabe metric. But what is a good choice of g. g^Y doesn't work as we can't control its behaviour.

Example 9 (B^4, S^3, g), $g^* = e^{(1-(x)2)} \ldots$

We get a bunch of equivalent conditions. Then there is some $\epsilon > 0$ such that for (B^4, S^3, h), if $||h - g_c||_{C^\infty} < \epsilon$, the CCE filling of (B^4, S^3, h) is unique.

Chapter 3

4 August 2018

3.1 Luigi Ambrosio: Calculus, heat flow and curvature-dimension bounds in metric measure spaces

Many recent developments rest on Eulerian (gradients, Laplacians, Hessians) versus Lagrangian (1D curves) duality. We tend to overlook this in Calculus II.

Example 10 *The ODE $\dot{\gamma}_t = b(t, \gamma_t), \gamma_0 = x$ corresponds in Eulerian terms, to $\partial_t u + \operatorname{div}(bu) = 0 \qquad (t, x) \in \mathbf{R} \times X$. This connection when b is not smooth, was studied by [DiPerma–Lions] then in general metric spaces by us.*

Example 11 *Compressible Euler equations in fluid mechanics*

$$\partial_t v_(V \cdot \nabla) v = -\nabla p; \operatorname*{div}_x v = 0$$

corresponds to

$$\ddot{\gamma}_t = -\nabla - \cdots$$

3.1.1 I: Weakly differentiable functions

Lagrangian minimisation:

$$\min \left\{ \int_a * bL(t, \gamma_y \dot{\gamma}_t) dt : \gamma_a = A, \gamma_b = B \right\}.$$

There are three approaches

1. distribution; $\int_{\mathbf{R}^n} f \nabla \phi dx = -\int_{\mathbf{R}^n} F \phi dx$
2. approximation by smooth functions (Lipschitz in our setting).
3. identify by prescribing behaviour on "p-almost all" curves. Introduced by Beppo Levi (1901).
$$f(\gamma_1) - f(\gamma_0) = \int_\gamma F$$

for p-almost every curve $\gamma : [0,1] \to \mathbf{R}^n$.

Theorem 20 *In any m.m.s. (X, d, m), the three definitions, properly adapted, are equivalent and define the same "gradient".*

The theory of optimal transport, pioneered by [Monge1781] then [Kantorovich1939] has been confined to probability and linear programming. Now we see more connections. The new ideas are

- New nonlinear interpolation between probability measures
- new geometric way of looking at the space of probability measures.

Monge's formulation

Let $\mu.\nu \in \mathcal{P}(X)$ and $c : X \times X \to [0, \infty)$ a Bole cost function. The minimise $\int_X c(x, T(x))d\mu(x)$ among all *admissible transport maps* pushing μ tp ν. $\mu(T^{-1}(E)) = \cdots$.

K's formulation

Minimise $\int_{X^2} c(x, y)d\Sigma(x, y)$ among all couplings $\Sigma \in \mathcal{P}(X \times X)$ of μ and ν, so $|\sigma(A \times X) = \mu(A)$ etc.

McCann's displacement interpolation

If T is an optimal transport maps in $\mathcal{P}_2(\mathbf{R}^n)$, then

$$\mu_t := (T_t)_{\#\mu} \in Geo(\mathcal{P}_2(\mathbf{R}^n))$$

and T_t is optimal. With this, McCann proved Renyi is optimal.

Dynamic formulation

More generally we focus on geodesic spaces.. Let $Geo(X) := \{$constant speed geodesics$\gamma : [0,1] \to X\}$.

Otto Calculus

$\partial_t \mu + \operatorname{div}(v\mu) = 0$ the continuity equation, Otto linked infinitesimal variations $s \in T_\mu \mathcal{P}_2(\mathbf{R})$to gradient velocities $v = \nabla\phi$ and defined a formal Riemannian metric. The induced Riemannian distance is precisely $W_2^2(\mu, \nu)$. Many effort to make this rigorous, but it's still insightful.

3.1.2 Heat Flow

$\partial_t u = \Delta u$; $u(0,\cdot) = u_0$; $\int_{\mathbf{R}^n} u_0(x)dx = 1$.

$$S(\mu := \begin{cases} \int_{\mathbf{R}^n} \rho \log \rho de & \text{if } \mu = \rho \mathcal{L}^n \\ +\infty & otherwise \end{cases}$$

This gave rise to many more interpretations of conservative PDEs. A key property is that the roles of d and m are nicely decoupled, unlike was happens for **D**. In m.m.s. (X, d, m),

$$Ent_m(\mu) := \begin{cases} \int_X \rho \log \rho dm & \text{if } \mu = \rho m \\ +\infty & otherwise \end{cases} \text{ for } \ldots$$

3.1.3 Curvature/Dimension Bounds

Bounds on the Ricci tensor are at the heart of many functions/geometric inequalities. $\mathrm{Ric}_m := \mathrm{Ric}_g + \nabla^2 V \geq Kg$ for $m = e^{-v} vol$, and the upper bound N on dimension.

Synthetic theories (going beyond the weighted Riemannian setting) emerge from diffusion operators and .... Need Gromov–Hausdorff convergence for metric spaces. By Gromov's precompactness theorem, geometrically m.m.s. can arise as (measured) Gromov–Hausdorff limits of Riemannian manifolds M^n.

3.1.4 Curvature–dimension

Bakry–Émery theory. Bochner identity

$$\frac{1}{2}\Delta_g |\nabla f|^2 - \langle \nabla f, \nabla \Delta_g f \rangle = |Hess(f)|^2 + \mathrm{Ric}_g(\nabla f, \nabla f).$$

In the case $M = \infty$ we get gradient contractivity and Logarithmic Sobolev inequality as well as Transport inequalities.

In the Riemannian setting, the link between Ricci curvature and displacement convexity in $\mathcal{P}_2(X)$ goes back to work of many. Relies in (K, N)-concavity.

A key idea is to average these inequalities along the geodesics selected by the optimal transport problem..

The CD theory (unlike B–É) has stability w.r.t. measured Gromov–Hausdorff convergence and its variants. B–É is more "Riemannian".

Theorem 21 *Assume that (X^n, d^n, m^n) are $CD(K, \infty)$ and m-GH converge to (X, d, m). Then the Cheeger energies, heat flows and Laplacians all converge as well.*

3.2 Lai-Sang Young: Dynamical systems evolving

Time evolution of systems. Began with Poincaré. Geometric/qualitative theory of ODEs. Foundations: Ergodic theory (probabilistic approach), KAM theory for quasi-periodic systems, Hyperbolic theory for chaotic systems. From the 1970s on, the field has matured and diversified. So this lecture is five snapshots of my work

3.2.1 Entropy, Lyapunov exponents and fractal dimension

Cantor set Λ is invariant set of $f(x) := 3x \pmod 1$.

$$HD(\Lambda) = \frac{\log 2}{\log 3} : \frac{\text{2=degree of branching complexity}}{\text{3=derivative of map}}. \tag{3.1}$$

Entropy is a measure of predictability of dynamical events. Let $\alpha = \{A_1, \ldots\}$ a partition of X. The degree of uncertainty $H(\alpha) = -\sum_i \mu(A_i) \log \mu(A_i)$.

$$h_\mu(T) = \sup_\alpha \left[\lim_{n\to\infty} H(a| \bigvee_i^n T^{-1}\alpha \right].$$

Lyapunov exponents are the rates of separation of nearby orbits.

$$\lambda * x, v) = \lim_{n\to\infty} \frac{1}{n} \log |Df_x^n(v)|.$$

Theorem 22 *There are "partial dimensions" δ_i such that $h_\mu(f = \sum_{i=1}^r \lambda_i \delta_i$ and $\dim(\mu|W^u) = \sum_{i=1}^r \delta_i$. This generalises (3.1).*

3.2.2 Correlation decay and geometry

Geometry may be hyperbolic $\begin{pmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}$, elliptic $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ or parabolic $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$,

Known since the 1970s that for purely hyperbolic equations, the decay was exponential. Markov tower construction provides a unified view for a class of dynamic systems with identifiable source of nonhyperbolicity.

3.2.3 Observable chaos

Chaotic behaviour = instability = rapid separation f nearby orbits = positive Lyapunov exponents. By "observable" we mean positive Lebesgue measure. Saddle fixed points are not observable for example. Strange attractors [but note this phrase has no formal agreed definition] are observable.

SRB constructed a measure μ s.t. for every continuous $\phi : U \to \mathbf{R}$,

$$\frac{1}{n}\sum_{i=0}^{n-1}\phi(f^i x) \to \int \phi d\mu$$

for Lebesgue a.e. $x \in U$. Existence of such measures provide observable chaos.

Consider a generic supercritical Hopf bifurcation. A pair of complex eigenvalues cross the imaginary axis at $\mu = 0$. Define twist number $\tau = \frac{\Im k_1(0)}{-\Re k_1(0)}$. where $\dot{z} = k_0(\mu)z + k_1(\mu)z^2\overline{z} + k_2(\mu)z^2\overline{z}^2 + \cdots$.

Theorem 23 *Setting ... Then for $|\tau| \cdot ||\pi^c \kappa(0)|| \cdot \mu^{-1/2}$ large enough, there is a positive Lebesgue measure set of $T >> 1$ for which the flow map $\Phi_{m,T}$ has ....*

This can explain shear-induced chaos, where the chaos is induced by external forcing magnifying shear in nonchaotic systems.

3.2.4 Applications 1: dynamics of infectious diseases

This looks at networks divided into nodes which are healthy (and susceptible), infected etc. Question is effectiveness of response. Not directed at any specific disease. Suppose the strategy is isolation, with imperfect implementation. Identify affected individual with probability p, τ days after infection. Quarantine duration κ, and need to add (have done) latency etc. Let S/I/Q = fraction health/infected/quarantine. System of delay–differential equations. β=transmission rate and *gamma* is recovery rate. Get a C^1 semi-flow on $C = C([-\tau - \kappa, 0], \mathbf{R}^3)$. $r = \frac{\beta m}{\gamma}$ is the disease reproductive number. $r < 1$ implies disease dies out.

Theorem 24 *Suppose $(S, I, Q) = (1, 0, 0)$. Contained if $\epsilon > 1 - \frac{1}{r}$ where $\epsilon = pe^{-\gamma r}$. In particular need $p > p_c = 1 - \frac{1}{r}$. $\tau < \frac{1}{\gamma}\log\frac{p}{p_c}$ tells what speed of quarantine we need.*

Table for various diseases. Note for smallpox, $p_c = 0.79$. Can look at endemic states, but in practice p, τ improve over time.

3.2.5 Applications 2: Dynamics of the brain

The brain is a structured dynamic network of 10^{11} neurons. I also expect biological systems to provide the sort of impetus that celestial mechanics did 100 years ago. I look at primate visual cortex. Eyes connect to LGN to visual cortex. Retina and LGN are tiled with cells that are excited when receptive field goes from dark to light (or v.v.). Once in the cortex the interaction between neurons first produces edges, and then recognises shapes, motion etc. We only have partial knowledge of this dynamical system, so we have an inverse problem of deducing the system from experiments. There's a major area of networks of interacting dynamical systems. Chaos doesn't make much sense, should look for emergent phenomena. We should also look at competition of subpopulations with "opposing" actions. This is predator/prey and much else.

3.3 Peter Scholze: Period maps in p-adic geometry

Introduction: Arrived at Bonn U. aged 19, BSc in 18 months and MSc in further 18.

Does a (system of) polynomials have integer solutions. We can do $\mathbf{R}$, and (mod p) (and then p-adic). Classic tool is to study these. p-adic geometry started in the 1960s by John Tate.

A period map takes the moduli spaces of curves etc, into a Flag Variety, e.g. $\mathbf{P}^n$, $Gr(d,n)$ etc. A complicated map from a complicated space to a simple space.

3.3.1 Over C

We have a Hodge structure $(H, H \times \mathbf{C} = F^0 \supseteq F^1 \supseteq \cdots)$. $H^i(X) \otimes_{\mathbf{Z}} \mathbf{C} \simeq H^i_{\mathbf{R}}(x) = \mathbf{H}^i(X, O_X \mapsto \Omega^1_X \mapsto \cdots \Omega^d_X))$.

Example 12 (Elliptic Curves) *Compact Riemann surface of genus 1. Subgroups of* $\mathbf{C}$ *quotiented out by the two periods. Period domain* $\mathbf{P}^1(\mathbf{C}) \setminus \mathbf{P}^1(\mathbf{R})$ *and fundamental domain.*

Shimura varieties.

3.3.2 Period maps for p-adic

Locally, over $\mathbf{C}$ the period maps had power series, and the coefficients (if nice) were rational or algebraic. So we can consider these p-adically. They only converge on small discs. There is no global maps that unify these.

Example 13 (Elliptic Curves) *Fix* E_0 *over* $\overline{\mathbf{F}}_p$.

$$D \simeq S = \{E/\mathbf{C}_p \textit{with reduction} E_0 \{\subset \{\textit{all} E/\mathbf{C}_p\}.$$

If E_0 *is ordinary,* log *maps to* $\mathbf{A}^1_{\mathbf{C}_p} \subseteq \mathbf{P}^1_{\mathbf{C}_p}$. *For supersingular ....*

Again we remove $\mathbf{P}^1$(base), this gets Drinfeld upper half-plane.

There is a duality of local Shimura varieties. $(G, \mu, b) \in B*(G, \mu)$ the basic (not supersingular) case. The dual space is $(\check{G}, \check{\mu}, \check{b})$

3.3.3 p-adic Hodge Theory

Always degenerates, without Kähler. Problem: $H^n_{\mathbf{R}}(x) \ncong H^n_{et}(X, \mathbf{Z}_p) \otimes_{\mathbf{Z}_p} \mathbf{C}_p$ (same dimension, so isomorphic, but not canonically so).

The there is a Hodge–Tate spectral sequence

$$H^i(X, \Omega^j_X) \Rightarrow H^{i+j}_{et}(X, \mathbf{Z}_p) \otimes_{\mathbf{Z}_p} \mathbf{C}_p$$

3.3.4 Existence of global period maps

Hodge–Tate period maps. $\pi_{HT} : \tilde{S} \to Flag$.

Example 14 (Elliptic Curves) $E/\mathbf{C}_p + H*1(E, \mathbf{Z}_p) \cong \mathbf{Z}_p^2 \to^{\pi_{HT}} \mathbf{P}^1_{\mathbf{C}_p}$. *If you're on an Igusa curve, the image is constant for a long time, but not always.*

This leads to a p-adic of Riemann's theorem. This says that the Hodge–Tate period map is "elliptic curves $E \mapsto p$-divisible group".

"No torsion in cohomology of Shimura varieties" (not literally true!). The "generic" part of $H^i(Sh_\Gamma, \mathbf{F}_l)$ is concentrated is degree $\dim Sh_\Gamma$. "generic" means we localise at "generic" maximal ideal of the Hecke algebra.

3.3.5 Galois Representations

For any system m of Hecke eigenvalues in $H^i(X_\Gamma, \overline{\mathbf{F}}_p)$ there is a unique continuous semisimple Galois representation $p_m : Gal(\overline{\mathbf{F}}/\mathbf{F}) \to GL_n(\overline{\mathbf{F}}_p)$.

3.3.6 Converse Theorem

Theorem 25 *Let F be a CM field.*

1. *Let E be an elliptic curve over F. Then Sato–Tate conjecture, meromorphic continuation of $L(E, s)$*
2. *Let Π be weight 2 cuspidal automorphic representation for O_2/F implies Ramanujan conjecture*

3.4 Special Event

Note the theft of a Fields medal. Videos to police and local media. SM presented a replacement Fields Medal, noting ironically that the GA had just decided that no-one could receive more than one Fields medal.

Recipient: if I were destroyed by such small events I wouldn't be here. One side-effect has been that many more people have heard of the Fields medal.

3.5 Figalli: Property of Interfaces in Phase Transitions via Obstacle Problems

Example: ice melting.

Problem 3 (Stefan) *Cylinder containing ice and water, with a free boundary. Also boundary conditions. $\theta(t, x)$ is temperature, assumed 0 is the ice. $\partial_i \theta = \Delta\theta$ i the water. $\dot{x}^(t) = -\nabla\theta$ on the boundary — Stefan condition*

Duvaut's transform: $u(t, x) = \int_o^t \theta(s, x) ds$ Then $u \geq 0$ and $\partial_t u \geq 0$. u solves the *parabolic obstacle problem* $\partial_t u - \Delta u = -\chi_{(u>0)}$.

Q1 Regularity of u. Can we classify the limits of these functions. x_0 is a *regular point* if, up to a sqq (?) of radii .... Also singular points. Is that all?

Q2 Regularity of the interface $\partial(u > 0)$.

To simplify the analysis, we look at the stationary problem.

3.5.1 Elliptic obstacle

$\Delta u - \chi_{(u>0)}; u \geq 0$. This minimises $\min_{v \geq 0} \left\{ \int_\omega \frac{1}{2} |\nabla v|^2 + v : V|_{\partial\Omega} = f \right\}$.

Theorem 26 (Weiss1999) *Let $n = 2$ Then*

$$||u(\cdots) \leq Cr^2$$

Recently extended to $n \geq 2$ but $< Cr^2 |\log r|^{-s}$.

Theorem 27 (Us) *Use a different method.*

- *For $n = 2$ $\Sigma_1 \subset C^2$ curve*
- *for $n \geq 3$*

(a) $\Sigma_{n-1} = \Sigma^g_{n-1} \cup \Sigma^a_{n-1}$ *where*

- Σ^g_{n-1} *["g=generic"] is also a subset of $C^{1,1}$ surface*
- Σ^a_{n-1} *["a=anomalous"] is small.*

(b)

Does Σ^a_{n-1} actually occur? Also are the estimates best possible.
Answers Yes/Yes.

Example 15 *Consider an axially-symmetric, reflection-symmetric problem. At the origin, C^1 is the best you can get, not $C^{1,\alpha}$ for any α.*

3.5.2 Parabolic Obstacle

This looks like a trivial extension, but we are still struggling.

Theorem 28 (In progress) *Let u solve the parabolic obstacle problem in $\mathbf{R}^+ \times \mathbf{R}^n$. Set $\Sigma_t := \{$singular points of $\partial \dots \}$.*

3.6 Raghavendra/Steurer: high-dimension estimation via sums-of-squares proofs

see proceedings article with ??

3.6.1 Steurer

Problem 4 (Estimation) *Given output Y of a randomised process with input X^*, to recover (approximately) X^*.*

Example 16 (Clustering) *Given a lot of Y_i, each of which is a perturbed x_j, recover the $x_j : j = 1 : m$. WE are given m, but not the mapping $Y \mapsto X$. There's a nice matrix formulation, PCA says X^* is a rank-1 matrix etc. Note generalisation to Tensor PCA/*

Best known guarantees until recently:

statistical $\Theta(\log k)$ but exponential time.

computational $O(k^{1/2})$ polynomial time.

Meta-algorithm: sums of squares. There are strong limitations of SOS for tensor PCA, which might imply that the gap is inherent.

Let $\Omega \subset \mathbf{R}^n$ be the set of possible signals (inputs). Given $Y = X * + W$ fro $X^* \in \Omega$ and W Gaussian. Need this to be ϵ-identifying, i.e. the true X^* is at most ϵ-away from the true one.

Let $S = \{X \in \mathbf{R}^N | p_1(X) \geq 0, \ldots, p_m(x) \geq 0\}$. A degree-$\ell$ sum of squares proof of the statement $\forall X \in S, q(X) \geq 0$, by expressing $q = \sum r_j(x)^2 \prod p_i$ of degree at most ℓ.

In the context of clustering, our polynomial constraints are $W = Y - X \in \mathbf{R}^{d \times n}$ "looks Gaussian", i.e. $\forall t \leq \ell$ the moments match.

The meta-algorithm says that the existence of low-degree proofs directly implied efficient d^{ℓ^2} algorithms. So make $\ell \approx \log n$. A key ingredient is the stability of the mean of the Gaussian under restriction.

3.6.2 Raghavendra: a lens on average case complexity

Tensor PCA. Given a 4-tensor $T : [n]^4 \mapsto \mathbf{R}$, $T = \lambda \cdot x^{\otimes 4} + W$, and the goal is to recover $x \in \mathbf{R}^n$. Depends on signal–noise ratio, and conjecture a sharp transition. A sum-of-squares lower bound almost always implies no known algorithm. For tensor PCA, recently shows that degree of SoS depends on SNR.

A natural way to recover x is to maximise $T(x) = \sum_{i,j,k,l} x_i x_j x_k x_l = \langle T, x^{\otimes 4} \rangle$. A related problem is injective norm certification. Given a random tensor can we bound how much signal it might have.

Standard ϵ-net algorithms shown $||T||_{inj} = O(\sqrt{n})$, but this isn't efficient: best we can do efficiently is $O(n)$. This is via $x^{\otimes 2} A x^{\otimes 2}$, where A is a $n^2 \times n^2$ matrix.

Use symmetry of $x^{\otimes 2}$.

More simply

1. Compute a $M^{(k)}$ whose entries are polynomials of input
2. Compute the spectrum of $M^{(k)}$

Theorem 29 *If an SoS algorithms succeeds* robustly*, then the simpler algorithm succeeds.*

This can be used to produce lower bounds. This has been done by hand, e.g. for random CSPs and k-clique.

Pseudocalibration

Don't forget the planted distribution. Here the SOS is feasible, as we have planted the solution. So we have a function F : tensor $\to$ solution. Approximate it by low-degree polynomials, and apply to the null case.

This recovers many existing constructions. But it's all conjectural.

Conjecture 11 *If low-degree $< D$ SoS semi-definite algorithms distinguishes two distributions, iff low-degree polynomials degree $D \log n$) do that.*

3.7 Poonen: Heuristics for the Arithmetic of Elliptic Curves

Work with many others.

$y^2 = x^3 + Ax + B$ properly scaled over $\mathbf{Z}$. $\mathrm{ht}(E) = \max(|4A^3|, |27B^2|)$. $\mathcal{E}_{\leq H} := \{E : \mathrm{ht}(E) \leq H\}$. $|\mathcal{E}_{\leq H}| \approx H^{5/6}$. MW theorem, and torsion is fully understood [Maz77]. Conjecture finite rank initially, but Cassels queried this. [RS00] would imply quadratic twists of a fixed E have rank ≤ 8. But there are greater ranks. [Granville2006publ2014] suggested all but finitely many twists have $r \leq 7$. [Watkins2015] suggests all but finitely many are ≤ 21.

We produce the same result by a very different method. We are trying to get a model for the complete package (rank, Selmer, Shafarevich–Tate groups).

$$0 \to \frac{E(\mathbf{Q})}{nE(\mathbf{Q})} \to Sel_n(E) \to Ш[n] \to 0.$$

Setting $n = p^e$ and taking a limit gives results.

Conjecture 12 *The probability for the density of* $\dim Sel_p E = s$ *is* $\prod_{j \geq 0}(1 + p^{-j})^{-1} \prod j = 1^s \frac{p}{p^j - 1}$.

Compatible with all known results. But this is only an upper bound. Should look at Sel_{p^k} and its limit.

$$0 \to E(\mathbf{Q}) \otimes \frac{\mathbf{Q}_p}{\mathbf{Z}_p} \to Sel_{p^\infty} \to Ш[p^\infty] \to 0.$$

3.7.1 Models for $Ш[p^\infty]$

1. Define $A \in M_n(\mathbf{Z}_p)_{alt}$

2. view each A
3. Sample A
4. take the distribution of coker(A).
5.

Compatible with Goldfeld conjecture: 0 50%, 1 50%, > 1 0%.

3.7.2 The Model

1. choose n to be an integer of size about $\eta(H)$ of random parity.
2. Choose random $A_E \in M_n(\mathbf{Z})_{alt,\leq X(H)}$
3. Define random variables $Ш'_E := (\operatorname{coker} A)_{tors}$ and $\mathrm{rk}'_E = \mathrm{rk}_{\mathbf{Z}} \ker A$

* These should model $Ш(E)$ and $\mathrm{rk}\, E(\mathbf{Q})$ respectively.

With probability 1, this gives us $H^{20/24}$ for rank 0,1, $H^{19/24}$ for rank 2 etc. and $H^{o(1)}$ for rank 21, and only finitely many for rank > 21.

Elkies can prove infinitely many of rank ≥ 19, and *one* of rank 28.

3.7.3 Fix torsion subgroup?

Then we can use this model as well.

For function fields the rank can be arbitrarily large. For number fields, can be arbitrarily large (but maybe still finite for any fixed field). Example are anticyclotomic, or multiquadratic fields. These curves all come from subfields, rather than truly defined over the large field. If we exclude these, then it might still be true.

3.7.4 Higher dimension

Fix dimension and degree of number field. MW still applies.

Fix g. By restrictions of scalars and Zarhin's trick, one reduces to considering one algebraic family $\mathcal{F}_g$ of principally polarised abelian varieties over $\mathbf{Q}$. Then define some height. The number of such varieties is bounded by a polynomial in H. We assume there's a similar algebraic model, and *if* it applies, the same model works.

3.8 MoMath etc.

Various talks. MoMath is only one in North America. They actually brought two square-wheeled tricycles which are on tour in Brazil.

3.9 V.V. Williams: A Fine-Grained Approach to Algorithms and Complexity

[Wil18][1] The general question is "how fast can we solve problems in the worst case". $T_{f,A}(n)$ is the maximum number of basic operations that algorithm A needs to compute f on inputs of size n. $O(n)$ is asymptotically optimal assuming you're going to read the input. 1956 Letter Gödel to von Neumann asked how strongly one could do better than exhaustive search. Despite all the research elsewhere, in some areas not much progress for many years.

Example 17 (k-SAT) *m clauses, n variables. Basic search is 2^n. Best known are $O\left(s^{n-\frac{cn}{k}} m^d\right)$*

Example 18 (Longest Common Subsequence) *Given two strings on n letters, find a subsequence of both strings of maximal length. Applications in genetics. Best known is $O(n^2/\log^2 n)$.*

Definition 1 *A* hard *problem is one for which the obvious algorithm is (bad) $T(n)$ and there's no known $T(n)^{1-\epsilon}$.*

Theorem 30 *For any $c > 1$ there are problems solvable in $O(n^c)$ bot not in $O(n^{c-\epsilon})$ for any $\epsilon > 0$.*

But we don't know how to do this for most realistic practicable.

PvNP, and most people believe $P \neq NP$.

Theorem 31 ([Coo71]) *For $k \geq 3$ k-SAT is NP-hard.*

This doesn't quite meet the goals of Definition 1, but we use a similar methodology. My hardness hypothesis will be that H requires $h(n)^{1-o(1)}$ time on inputs of size n on a RAM.

Then I show that an $O(q(N)^{1-\epsilon})$ algorithm for problem Q would imply a $h(n)^{1-\delta})$ for H.

Conjecture 13 (ETH) Exponential Time Hypothesis. *There exists a $\delta > 0$ such that 3-SAT cannot be (worst case) solved in $< 2^\delta n$ time. Call this least δ s_3.*

Conjecture 14 (SETH) Strong Exponential Time Hypothesis. $\lim_{n\to\infty} s_n = 1$.

Three problems: Orthogonal Vectors, 3Sum and ASPS [All Pairs Shortest Paths].

[1] `https://eta.impa.br/dl/194.pdf`.

3.9.1 Polynomial Many-One Reduction

A is reducible to B is there is a polynomial time R that transforms any instance x of A into $R(x)$ of B such that $A(x) = 1 \Leftrightarrow B(R(x)) = 1$. But this is coarse-grained.

Turing reduction: solve A by polynomially many appeals to a B oracle. Again coarse-grained.

Definition 2 *A is (a, b)-reducible to B if $\forall \epsilon > 0 \exists \delta > 0$ and an $O(a(n)^{1-\delta})$ time algorithm that can solve A on instances of size n making call to an oracle for B with query lengths $n_1, \ldots, n_k$ such that $\sum_i b * n_i)^{1-\epsilon} < a(n)^{1-\delta}$.*

This is basically transitive.

Graph of many problems under this. Many n^3, n^3 relating APSP to many other graph problems, including second shortest path. Also "Negative Triangle" [in a weighted complete graph].

Since 2010 there has been great work in this area. Fine-grained space complexity, also approximability questions. Also what about average case?

Q Matrix Multiplication?

A People believe n^2. Such algorithm would do all sorts of things.

Q How close is this to real computers.

A Maybe we need a better model, and you would also care about the constants.

Q These all have integer exponents.

A Not a restriction. There are $n^{2.5}$ problems as well.

Q-JHD

A The following problem n a graph with coloured vertices: $\exists?(c_1, c_2, c_3)$:/ $\exists$triangle coloured(c_1, c_2, c_3). This is a super-problem that all three of mine reduce to.

3.10 Kayal: the quest for a polynomial that is hard to compute

Formally an arithmetic circuit, where addition gates compute $\lambda_1 x_1 + \lambda_2 x_2$. Care about depth (amount of parallelism) and number of nodes. We know (based on [Sha49]) that almost all are hard to compute, but we can't get even close to this. $(n \cdot \log d)^{\omega(1)}$ is what we're looking for. Look at MAJORITY (easy) and CLIQUE:

$$\sum_{S \in \binom{m}{m/2}} \prod_{i,j \in S} x_{i,j},$$

thought to be hard. Also PERMANENT, which is believed to be hard, even though it's very like DETERMINANT, which is easy to compute.

$IMM_{nd} := Tr(X_1 \dot{X}_2 \cdots X_d)$ where the X_i are symbolic matrices. There's an obvious D&C method on d. Split int chunks of size $t = d^{2/\Delta}$, where Δ is the desired depth. Size .... Part of a general depth-simulation result: $s^{O(d^{2/\Delta})}$.

3.10.1 Strategy

Suppose #Steps(f)=$n^{O(1)}$. Then apply depth-reduction,and $f = \sum^s T_i$ where each T_i is a product of $O(\sqrt{d})$-degree polynomial and s is $n^{O(\sqrt{d})}$.

Theorem 32 *Explicit* $\{f_n : N \geq 1\}$ *with* $n = d^2$ *such that* $s \geq N^{\Omega(\sqrt{d})}$.

Close but not quite.

Associate a matrix $M(g)$ to every polynomial g such that rk$(M(T_i))$ is small, *Linearity* $M(\alpha f + \beta g) = \alpha M(f) + \beta M(g)$ and rk$(M(f_n))$ is large. We would need to add a large number of $M(T_i)$ to attain the large rank.

To do this, we find a geometric property GP of $V(T_i)$, and express GP in terms of the rank of a big matrix. Note that a large upper triangular submatrix implies full rank. If the columns of M are almost orthogonal, the M has large rank.

If T is a product of low-degree polynomials, then the variety of a union of low-degree hypersurfaces, so lots of high-order singularities. Hence $V(\partial^k T)$ has lots of points.

Let $V = V(f_1, \ldots)$ be a variety, and G_ℓ a set of degree-ℓ polynomials. $G_\ell(V)$ those that vanish at each point of V. Let $I_\ell(V) = \{(a_1 f_1 + \cdots + a_m f_m)\text{degree} \leq \ell\} \subseteq G_\ell(V)$.

Theorem 33 (Hilbert) *If* V *has dimension* r, $I_\ell(V)$ *has asymptotic dimension* $\binom{n+\ell}{n} - \Theta(\ell^r)$

But we need to improve on this.

Theorem 34 ([FLMS14,KS15]) $IMM_{n,d}$ *with* $n = d^{10}$ *then* $s \geq n^{\Omega(\sqrt{d})}$.

which is a bit worrying as IMM is easy to compute.

Chapter 4

6 August 2018

4.1 Coifmam: Harmonic analytic geometry

$C(f)(z) := f_\Gamma(z) = \int_\Gamma \frac{f(\zeta)}{\zeta - z}$. Then $S = C(I + C - C^*)^{-1}$ where C^* is a perturbation of C. We want to use these techniques to discover intrinsic coordinate systems for data clouds.

Think of a million samples o fa molecules. Conceptually, we have a $10^6 \times 10^6$ matrix, and can look for the eigenvalues. The number we need is (at least) the number of states we are looking for. These eigenvectors are intrinsic, and provide a "universal library" of functions for building relevant models.

But a fundamental question is "what is the distance between two subsets of $\mathbf{R}^n$?". Classically $\min_{\text{maps}} ||x - f(x)||$. An alternative is via filtering. In terms of the clouds, machine learning will do nothing: we need to organise the data. We regard each point z as a question, with $C(f)(z)$ the answer.

4.2 Kronheimer/Mrowka: Knots, three-manifolds and instantons

Definition 3 *A knot K is an embedding of the circle in $\mathbf{R}^3$ with no self-intersections. Links are a set of knots.*

Example: 16n-63441, one of 1.3M 16-crossing knots.

Study the fundamental group of $\mathbf{R}^3 \setminus K$.

Theorem 35 (Dane's Lemma) *A knot is the unknot iff fundamental group is abelian.*

Hence try to map $\pi_1(\mathbf{R}^3 \setminus K$ into G with a non-abelian image. We have shown that $SO(3)$ is a sufficient G: dihedral groups don't work for 17 knots. In particular we map meridians to elements of order 2.

Example 19 (5,7 torus knot) *To every meridian curve we associate a point on 2-sphere* $\mathbf{RP}^2$*. Nice graphics.*

Theorem 36 *If K is a non-trivial know, the non-*$\mathbf{RP}^2$ *part of the representation is non-trivial.*

4.2.1 Three-manifolds and SO(3)

See instanton Floer Homology.$\{\rho : \pi_1(M) \to SO(3)\}$ gives $I(M)$. These ρ are flat connections in principal $SO(3)$ bundles over M. We actually work with $SU(2)$ bundles. But $SU(3) \to SO(3)$ has kernel $\{-1, 1\}$. Hence $[w]$ is an obstruction.

For gauge theory, we want all connections, not just flat ones. Flat connections are critical points of Chern–Simons connection: $CS_A = \int_M tr(A \wedge dA + \frac{2}{3} A \wedge A \wedge A)$. Gradient flows are instantons, solutions of Yang–Mills $(d/dt)A = - * (dA + A \wedge A)$.

4.2.2 Back to knots

4.2.3 Now looks at spatial graphs

We consider only trivalent graphs. A bridge is an edge such that there exists a plane which intersects this edge only. WE are interested in bridgeless graphs. Still look at embeddings into SO(3) mapping meridians to order 2 elements. Again get a representation variety $\mathcal{R}(G)$. Again use Floer's packaging via Morse theory to give an "Instanton Floer Homology" $J^{\#}(G)$. F.d. vector space over $\mathbf{Z}/2$. If G is bridgeless, this is non-empty.

These are connected by Tait colourings of the graph (edges) where each vertex has all three colours. Let $Tait(G)$ be the number of such. The snark is the smallest graph with no Tait colourings. $\rho_1 : \pi_1(\mathbf{R}^3 \setminus G) \to V_4$ correspond to Tait colourings, $\dim J^{\#}(G) \geq Tat(G)$ for planar graphs.

Conjecture 15 *This holds for all graphs.*

Conjecture 16 *Equality holds for planar graphs.*

Note that Conjecture 16 + non-vanishing implies the four-colour theorem. Tait-colouring the edges is equivalent to four-colouring the regions, and effectively the mapping between Tait colourings and 4-colourings is the multiplication table of V. Note that all proofs of the 4CT have required computer assistance.

4.3 Catherine Goldstein: Long-term history and ephemeral configurations

Poincaré: Mathematics is the art of giving the same name to different things (Science and Method, 1908); his example was "loop" (? and "uniform convergence").

Note that 70s/80s, when much of this started, were a time of societal change, when science was being challenged. Note that how mathematics is circulated, evaluated have changed. Note that Fermat didn't write on Diophantus, he wrote on a commentary on Diophantus.

Also need to reflect on "discipline", "school' etc. Looks at original Jahrbuch classification.

Example 20 *Did Charles Hermite invite Hermitian forms? Pictre of the famous, "reactionary" Hermite. But shows young Hermite, who had dropped out of Polytechnique as he didn't want to become a military engineer.*

Theorem 37 *A definite quadratic forms with $n+1$ variables, determinant D. Then $\exists$ $n+1$ integers with $0 < f < \left(\frac{4}{3}\right)^{n/2} D^{\cdots}$*

Applications to simultaneouse approximation of algebraic numbers. Also proved that forms of determinant 1 are sums of squares for $\dim \in 2 \ldots 6$. *This was based on his reading of Gauss's classification of binary forms. These two were actually closely related by Hermite. Sent to Jacobi who published these in Crelle. He wrote to Jacobi, explaining mathematics as a science of classification "just as descriptive natural history".*

Hermite started with $A = \sum_4 x_i^2$. Prove Adivides$\alpha^2 + \beta^2 + 1$. This gives a form of determinant A^4,and we apply his theorem.

Then (1855) he considers forms over two complex numbers, $Avv_0 + BVw_0 + B_0v_0w + Cww_0$.

So he didn't "invent" then: he detected them!

Segre was all his life in Torino, but closely connected to Klein. Many translations from German were organised by him. Notes of Segre published on web site, which have examples of Hermitian forms. But only reference is von Staudt *Geometrie der Lage 1847+supplements*. Main aim was to remove all measurement. But this wasn't complete. However, von Staudt's work went to Karl Culman (at ETH's predecessor) and his book Graphical Statics (draw lines on the blueprint itself rather than revert to a separate page of calculations). Drawing of Maths of Eiffel tower.

Risogimento and unification let to both a return to Romano-Greek scientific roots, but also a growth in engineering schools, and much geometry teaching. Segre wrote to Hurwitz drawing attention to him work. But note that at this time abstract geometry was not axiomatised. Hermitian forms were now geometric objects. Immense graph of papers and links (partly citations, but also choice of journals, citations, etc.

Note the common connections with classification programs.

As well as links etc., we need to understand discontinuities. General theories (Kuhn, Lakatos) don't really seem to apply.

4.4

4.4.1

set theory	functions
topology	continuous functions
differential geometry	smooth functions
Algebraic geometry	polynomial functions

Affine variety = zero set. We can ask does it have solutions: depends on the field: $t_1^2 + t_2^2 + 1 = 0$ for example. We assume now that K is algebarically closed, and that varieties are irreducible. Vanishing sets of polynomials define Zariski topology. Rational maps, birational maps. If $x \in X$ can assume x on an anaffice neigbouthood. Birational alalmaps can remove varieties over $\mathbf{C}$,

4.4.2 Algebraic Varieties

Fano: K_X anti-ample; Calabi–Yau if K_x if trivial.

Conjecture 17 *Each smmoth projective variety is brational to a projective Y with good singularities s.t.*

- *Y admits a Fano fibration*
- *Y admits a CY fibration*
- *Y is canonicallly polarised.*

Known in dimension 1, 2. Then Miminal Model Program (MMP). Problem is finding Y. A sequence of birational transfromations designed to make the canonical divisor as positive as possible. divisor contractions or flips. No flips in dimenson 2, and dimension 3 was Mori's work.

Conjecture 18 *Termination; Abundance (when it stops Y satisfies Conjecture 17)*

Known fro 1..3, and ≥ 4 in the general case.

For a smaooth projective W, the vector spaces $H^0(mK_W) = \{\text{rational functions}\alpha | \operatorname{Div}(\alpha) + mk_W \geq 0\} \cup \{0\}$. say so much about the geometry of W, for $m \in X$. This gives rise to canonical rings.

Also need to prove that flips exist in dimension ≥ 4.

4.4.3 Fano varietes

Theorem 38 (Borisov–Alexeev–Borisov (BAB) Conjecture) *For each $d \in \mathbf{N}$, $\epsilon \in \mathbf{R}^{>0}$ the set $\{X | X\epsilon - ldFano$ of dimension $d\}$ is bounded.*

Example 21 *For $n \geq 2$ consider $E \subset W_b \to^t X_0$. where*

X_b is the cobe over the rational curve of degree n

W_n is obtained by blowing up the vertex

E is the exceptonal curve.

Then X is $\frac{2}{n}$Fano.

MMP is open over finite characteristic. Note that we don't have complete resolution of singularities. Also arithmetic schemes. This connects with Manin's Conjecture, Lang's Conjecture etc.

4.4.4 Also

Let X be a variety with god singularities. $f : X \to Z$ be a surjective projectie morphism with connected fibres. We say X is aon over Z if K_X is anti-ample over Z. I the global case with $\dim Z = 0$ then X is a usual Fano variety. In the fibration case with $0 < \dim Z < \dim X$ then f is a Fano fibraton who general fibred are usual Fanos. The birational case with $\dim Z = \dim X$ is open

4.5 Wormald: Asymptotic Enumeration of Graphs with Given Degree Sequence

Why count

Proving existence by counting Properly 4-coloured 6-regular graph. It cannot be three-coloured becaus eof triangles. If we forbid triangles? No. Indeed even forbidding $\leq k$ cycles for any k doesn't work.

Panbiogeography

Counting by degrees used fro random graphs Wigner's semicircle law for eigenvalues of a random d-regular graph as $D \to \infty$ (extending [McKay1981] for fixed d). Sandwiching $\mathcal{G}_{n,p} \leq R \leq \mathcal{G}_{n,p'}$

Structure of the Giant Component of $\mathcal{G}_{n,p}$. [Bollobas1984] has asymptotically $2\epsilon n$ vertices, and we proved the distribution is normal.

So what results do we have?

[Read1958] number of 3-regular graphs on n verstices.$g_3(n) \sim \frac{(3n)!e^{-2}}{(3n/2)!288^{n/2}}$. Further progress depends on Bollobas' "configurations". Formulae for graphs with small degreee $d_1, \ldots, d_n$. We have work of degrees $\sim cn$ with deviation $O(\sqrt{n})$.

$d := M_1/n$ average degree

$\lambda := d/(n-1)$ edge density

$\gamma_2 := \sum(d_j - d)^2/n^2$ (scaled variance).

Conjecture 19 (Binomial Conjecture) *Degree sequence of* $\mathcal{G}(n,m)$ *is asymptotically independent binomials* $\mathrm{Bin}(n-1,p)$ *subject to the sum being* $2m$. *[*$p = m/\binom{n}{2}$*]. Also degree sequence of* $\mathcal{G}(n,p)$ *... subject to the sum being even, even when* $\hat{p}$ *has a certain, tight, almost normal, distribution close to* p.

This would easily produce a lot of known results.

Last year showed the conjecture holds for all densities. Invovles work on "double edges". [Mckay1985] estimates $\frac{|C_i|}{|C_{i-1}|}$ using switching, so the distribution is asymptotically Poisson.

A new method compares numbers of graphs with different degreee sequences $R_{ab}(d) = \frac{g(d-e_a)}{g(d-e_b)}$. Seems to require very accurate estimation of products. Estimate ratios by "degree switching". But switching might create double edges or loops, and need to disallow these. There is work on d-regular graphs, and they extend to bipartite graphs. Working on hypergraphs.

Latin rectangles are edge-coloured bipartite graphs.

Q Latin Squares?

A Doesn't seem to go that far.

4.6 Atiyah: The Future of Mathematical Physics: New Ideas inn Old Bottles

Abel Prize Lecture.

Begin with π. Importance of groups and Lie groups. Noted that Sophus Lie was first to push for an Abel Prize. Most important formul: $e^{2\pi i} = 1$ (Euler). But there's a quaternionic analogue. Mentioned von Neumann, for whom Type I was Integer, Type II was Real.

Topology "Gang of four": Hirzebruck Atiyah, Bott, Singer.

Also Cartan (Henri), Serre, Grothendieck.

Unification: Gel'fand, Langlands, Penrose.

Witten, Octonians and M-Theory.

Chapter 5

Math Aspects of CS: 6 August

5.1 Jasmine Mathew: Hamilton Decomposition of Knodel and Fibonacci Graphs

Definition 4 *A decompisitionof graph G is a collection $\{H_i\}$ of non-empty subgraphs such that $H_i = \langle E_i \rangle$ for some non-empty subset of $G(G)$, E_i pairwise disjoint. If each H_i is a cycle, we call the decomposition a cycle one. Also Hamilton.*

Consider $K_{n,n}$. Connect $(0, j)$ to $(1, j')$ when $j - j' \equiv k \pmod{n}$. Call such edges a k-jump. The collection of all r-jumps is a perfect matching. Cann it G_r. Write $[r, l] := G_r \cup G_l$. For example $G_0 \cup G_1$ is a cycle, but $g_0 \cup G_2$ is a union of two cycles. $[r, l \simeq [0, l - r]$ for $r < l$ $\gcd(m, n) = d$ iff $[0, m]$ is the usin of d cycles of lenngth $\frac{2n}{d}$. Shows C_{12} decomposition of $K_{6,6}$.

The Knodel graph $W_{d,n}$ has vertices (i, j) with $0 \leq j \leq \frac{n-1}{2}$. Let $r = 2^k - 1, l = 2^{k+1} - 1$: "consecutive dimensions". Then $[r, l] \simeq [0, 2^k]$ so is Hamilton iff n is odd.

$F_{d,2n}$ is decomposable deending on congruence conditions.

5.2 Kumar: Two-stage hyper-chaotic system based image encryption in Wavelet packet dimain for wireless communication systems

Claim that we can generate random sequences through hyper-chaotic system. Need a fast secure algorithms. Claims that this works at the speed fo 4G. $\dot{x} = a(y - x) + u; \dot{y} = -xz + cy; \ldots$ is our hyperchaotic system. Essentially all vapourware. Some examples - encrypted image looks totally random.

5.3 Firer: Generalized free-column Distances for Convolution Codes

Context is ECC. $|⇳| = m$, $|§| = x$, $f : M \to X$ is an injection. Rate $r = \log m / \log x$. Normal setting $\mathbf{F}_q^n$.

Chapter 6

7 August 2018

6.1

6.1.1 Expanding graphs

Cheeger constant of $X := (V, E)$ a graph. $\min\left\{\frac{|E(W,\overline{W})|}{\min(|W|,|\overline{W}|)} | \emptyset \neq W \subsetneq V\right\}$. X is ϵ-expander if $h(x) \geq \epsilon$.

Proposition 1 (Pinkser) *Random k-regular graphs are expanders*

First exhibited construction [Margulis1970] If $\Gamma = \langle S \rangle$, $|S| < \infty$ has Kazhdan property (T) then $\{Cay(\Gamma/N; s) | N \triangleleft \Gamma\}$ is an expander.

Expansion $\approx$ Spectral Gap, i.e. $A = A_X$ is adjacency graph, eigenvalues $k = \lambda_0 > \lambda_1 \geq \cdots \geq -k$ (symmetric so all eigenvalues real), then X is ϵ-expander iff $\lambda_1 < k - \epsilon'$ [JHD he wrote ϵ' with no explanataion].

[Alon–Boppa] $\lambda_1 \geq 2\sqrt{k-1} + o_n(1)$. A graph is *Ramanujan* if all other eigenvalues are bounded by $2\sqrt{k-1}$.

In the 80s look at $\Gamma(I) \setminus PGL_2(F)/K$ where F is a local field. and K is a maximal group. and $\Gamma(I)$ is the congruence subgroup of an arithmetic group Γ. Depends on Deligne, Drinfeld, proofs of Ramanujan conjecture.

In general we fix k and ϵ, and want $|V| \to \infty$. The Zig-Zag product helps, but recent "interlacing polynomials" techniques are non-constructive.

6.1.2 High Dimension Theory

CS: hypergraphs, mathematicians simplicial complexes. A $(d+1)$-uniform hypergraph if a collection of subsets of V all of size $d+1$ [$d = 1$ these are edges in a graph!]. There are lots of options for "higher-dimensional expander", not all equivalent. Ramanujan complexes become quotients of the Bruhat–Tits buidlings $\mathcal{B}_d(F)$ associated with $PGL_{d+1}(F)$ as above. They satisy suitable optimal spectral gap bould (as in Alon–Boppana), Base don [Laurent] Lafforgue generalizations of Ramanujan conjecture.

Theorem 39 (Boros–Füredi) *Given $P \subset \mathbf{R}^2$ with $|P| = n$ $\exists z \in \mathbf{R}^2$ which is convered by $\left(\frac{2}{9} - o(1)\right)\binom{n}{3}$ of the $\binom{n}{3}$ triangles determined by P. Later, $\frac{2}{9}$ is optimal.*

Theorem 40 (Barany) *Fix d. $\exists c_d$ s.t. $\forall P \subset \mathbf{R}^d$ with $|P| = n$ $\exists z \in \mathbf{R}^d$ which is covered by at least $c_d\binom{n}{d+1}$ simplices from P.*

Let X be a d-dimensional simplicial complex. We say X has the geometric (resp. topological) ϵ-overlapping property if $\forall f : X(0) \to \mathbf{R}^d$ nad $\forall \tilde{f}$ affine (resp continuous) extension $\tilde{f} : X \to \mathbf{R}^d$, there exists $z \in \mathbf{R}^d$ with zc covered by ϵ-fraction of the images of $X(d)$.

Theorem 41 (Gromov 2010) *$\Delta_n^{(d)}$ has the topological ϵ-overlapping property.*

Look at $d = 1$ and assume X is ϵ-expander. Take $z \in \mathbf{R}$ such that $1/2$ of the images are more (say A) and $1/2$ less than z. As X is an expander, $E(A, \overline{A})$ is large and we overlapping property. In fact "expander" is stronger than topological overlapping.

Theorem 42 *$\forall d$ exists bounded degree (every vertex in contained in a bounded number of simplices) simplicial complexes of $\dim d$ with geometric overlapping. But these are not topological.*

Proof: either random or Ramanujan. We can now do it with bounded (but 10^{100}) degree, but starting in $\dim d$+1 which give us $\dim d$ topological expanders. $\dim d + 1$ is still open.

There's also "coboundary expanders". We need notation. Erdős–Renyi random graphs: edges with probability p, threshold for connectivity = vanishing of cohomology. Do the same for triangles etc. Replace connected by vanishing of cohomology. Relates to PCP Theorem.

Question 8 *Is there a random model than will give coboundary/topological expanders? Maybe not? Analogy with Mostow/Margulis rigidity.*

6.2 Nalini Anantharaman: Delocalization of Schrödinger eigenfunctions

6.2.1 History/Physics

1913 Bohr's model — momentum quantised; $J = nh : n \in \mathbf{N}$. Explains hydrogen (only)

1917 Einstein "Zun Quantensatz von Sommerfeld und Epstein". These only work for "'type A" — today we say "completely integrable"

1923 de Broglie wave/particle duality.

1925 Still can't do anything elses so Heisenberg, where physical observables are operators (matrices)

1925 Schrödinder equation $i\hbar\frac{d\psi}{dt} = \left(-\frac{\hbar^2}{2m}\Delta + V\right)\phi$

But Though the theories are mathematical equivalent, they are not physicaly equivalent.

1950s Wigner on heavy nuclie couldn't diagnalise S, so took random matrices.

1980s Simulation. Models of billiards: calssical $\phi^t : (x, \xi) \mapsto (x + t\xi, \xi)$. But Q: $i\hbar\frac{d\psi}{dt}\cdots$.

All Asymptotic.

Conjecture 20 (Bohigas...) *The spectrum of the quantum system resembles a large random matrix.*

Conjecture 21 (Quantun Unique Ergodicity) *Studhy* $|\psi(x)|^2$

Conjecture 22 *Show that* $\psi(x)$ *resembles a Gaussian process.*

Progress only on Conjecture 21.

Definition 5 (Quantum Variance)

$$Var_\lambda(K) = \frac{1}{N(\lambda)} \sum_{\lambda_k \leq \lambda} |\langle \psi_k, e \cdots \rangle| .$$

Get control by the L_2 norm. There is a result of hers that iplies, in constant negative curvature, the support of μ has dimension $d = \dim M$.

6.2.2 Toy models

These tend to be discrete. Instead of studying $\hbar \to 0$, we let $\dim \to \infty$. Look at k-regular graphs. Let G be a $(q+1)$-regular graph. If $F : V \to \mathbf{C}$, then $\Delta f(x) = \sum_{y \sim x}(f(y) - f(x))$ is the equivalent of the Laplacian. The adjacency matrix is $N \times N$, so can compare with Wigner. Also there is the random regular graph model.

Assume G_N has "few" short loops (= converges to a tree in the sense of ...). Shows examples which seems to match McKay's Law.

Also examples with Ramanuan graphs, and Cayley graphs.

Theorem 43 (Bauerschmidt+others) *Let* $d = q + 1 \geq 10^{20}$*, Then for the* $\mathcal{G}_{n,d}$*-model, with large probability as* $N \to +\infty$*, the small-scale McKay law holds.*

It's hard to get the quantifiers right here.

Theorem 44 *With probability* $1 - o(1)$ *as* $N \to \infty$ *on ehas, for all eigenfunctions* $\phi_i^{(N)}$*, for all diameters* $R > 0$ *the distribution of* $\phi_i^{(N)}|_{B(x,R)}$ *when* x *is chosen uniformly at random in* $V(\mathcal{G}_{N,d})$ *is ...*

Note that these don't apply to triangulations of surfaces, because of the "few small loops" assumption!

6.3 Arora: Mathematics of machine learning and deep learning

Note 2001 and 2010 Gödel prizes. Also author of a wonderful book.

Various gains, e.g. games and medical images, and getting close with translation and driving. This is *not* explicitly imitating human intelligence, instead it is creating machines that improve from experience and interaction. So we are asking Newton-like questions "what does it mean to ...".

6.3.1 Mathematical Formulation of ML

Learning patterns from data, e.g. curve fitting. For example, inflation versus unemployment in Japan: data points and Philips curve. Boyle's Law was determined originally by fitting a surface. Given a dataset of reviews [text] and ratings. Try to predict rating from a new text T. Note that we'll only get an approximate model. Linear model might be $\sum \text{words } w \in T \theta_w$ where θ is the sentiment score. Trivial least-squares fit. Got a student to do this,and $\theta_{love} = 1.1$ etc.

Definition 6 *Machine Learning is the task of finding a function mapping input to outputs, given data. We have a function with tunable $\theta \in \mathbf{R}^d$.*

Still aim at $\min \ell(\theta) = \sum_i \left(f_\theta(X_1^{(i)}) - Y^{(i)} \right)^2$. There are other definitions for the loss function. Then test the model on new data X_2. Standard rule is "train on 80% and test on 20%".

Fourier analysis allows you to learn a function. The problem is that it's practically infeasible. But exponential in #samples, and n. Hence training in terms of gradient descent. $\theta * (y+1) = \theta^{(t)} - \eta \nabla(\ell)$. Where η is the "learning rate", say 0.01. Note that real-life ℓ might well be non-convex: ouch! In practice many tricks.

Deep learning is alternations of linear (multiply by a matrix) and non-linear transformations, which is $v \mapsto \max(v, 0)$. Back propagation is basically a clever application of the chain rule. Why does it work? Linear enough to implement, non-linear enough to be powerful.

6.3.2 ML in action

Reviews "That's not how we do it". Word order matters. We already know what words mean. Unsupervised learning. Take a large corpus (Wikipedia), and try to predict from adjacent (say last five words). "Baby word2vec" [Mikolovetal2013a]. $\theta \in \mathbf{R}^{300}$: semantic vector. $Pr[w|w_1 \dots w_5] \propto \exp(\frac{1}{5} \sum_i \langle v_w, v_{w_i} \rangle)$. For semantic vectors, $\cos(v_1, v_2)$ relates to "similarity".

Sequential decision-making framework e.g. games, but most decision theory. Assume tree of moves, and opponent stochastic. Use a "move evaluation function", which is an ML result.

6.3.3 Towards mathematical understanding of deep learning

Last few years: me and others. Key questions.

1. When/how does it work?
2. Why deep, not shallow (can be thousands of layers)
3. Why doesn't training overfit (#parameters >> # training). Current deep models can achieve 0% loss on random data [Zhangetal2017a].

We know non-convex optimisation is NP-hard. This is a black box, as we lack any mathematical explanation for "why is this bunch of pixels a dog". NB: Deep Learning is a great way to motivate kids about calculus. Note that $\Delta \neq 0$ implies there's a descent direction, but large ∇^2 means this varies a lot. `offconvex.org`. We would like the direction of movement to be positively correlated with the direction to the global minimum (Lyapunov).

Theorem 45 ([EldanSmair2016a]) *There exist a function computable by a depth $d+1$ circuit of size S which is not approximable by a depth d [2] function.*

[+ChenHazan2018a] can replace l_4 regression by a depth-2 linear net. This accelerates optimisation.

A popular conjecture is that, on realistic data, the net's parameters are constrained — by problem or training – be be on a manifold of much lower dimension. Examined VGG19. Properly-trained nets have "noise stability": add Gaussian η to output x with $|\eta| = |x|$. Shows that the effect drops rapidly with number of layers. $|Mx|/|x| >> |M\eta|/|\eta|$. So the distribution of singular values in a filter of VGG19 is a few large eigenvalues. "Nearly low-dimensional".

6.3.4 Conclusion

Note that imitation doesn't always work, e.g. airplanes. He also claims that we have very little idea at an operational level of how humans think.

"I am optimistic that deep learning methods can be mathematically understood and/or simplified"

6.4 Donoho: From Blackboard to Bedside

Gauss Prize Lecture. Many acknowledgements.

6.4.1 Congressional Briefing

I am going to say much of what I said in my Congressional briefing. Also how did we go from Maths articles in 2006 to products in 2017.

His wife had brain surgery as a teenager (visible dimple), which has always affected his son, who became a neurosurgeon and has conducted the same operation. At a poor hospital with major waits for MRI scans. The MI takes 90 minutes, which can waste a person-day of surgical team time.

Also talked about prostate cancer. We don't have good techniques to identify the aggressive subtype accurately. MRI-guided 3D biopsy is $\times 5$ more accurate that alternatives. Example of man who had 10 standard biopsies all missing. Late treatment cost \$600K rather than early \$50K. Hence MRI can really make a difference.

GE Hypersense (recently approved) is $\times 8$ faster. Another 3+1D example is 25 seconds rather than 6 minutes.

Note that this is both personally and financially relevant to Congress.

One heartbeat rather than 9 for a child's scan.

"Federally funded mathematics played a key role". In particular some of the pioneers of compressed sensing (Terry Tao) were federally funded.

In 10^6 dimensional space. Consider a cone with a 10^4-D apex. The chance of a 9×10^5 plane slices the cone is very small. Equivalently, to recover 1 1000×1000 image with 10^4 nonzero wavelet coefficients only needs 10^5 elements. 40M MRI/year in the USA, at say \$500 each $=> 100\times$ Federal maths funding.

6.4.2 Tech transfer

Why so quick? The Maths was very welcome. There's a very active MRI research community. The community believed that speedups were possible (experiments). But people needed the certainty of the maths. Lustig's slide in MRI-speak.

- Study compressible signals with sparsifying transforms
- Undersample incoherently w.g. with quasi-randomness (uniform in the Grassmanian)
- Non-linearly reconstruct by promoting sparsity

6.4.3 Conclusions

CS could have been arrived at in many way. $\epsilon = k/N$, $\delta = n/N$

Chapter 7

8 August 2018

7.1 Naor: Metric Dimension Reduction: A Snapshot of the Ribe Program

Metrics that arise as a result of optimisation problems, optimal transport metrics etc. In fact a more tame world is normed spaces. These are in 1:1 correspondence with ??.

Phenomenon discovered by Ribe, analogy between "nice spaces" and real-world spaces.

Finite dimensional linear properties of normed spaced are actually metric properties in disguise.

Example 22 *There is a constant $c > 0$ such that, for every n, for any unit vectors $x_1, \ldots, c_n$ one can find signs ϵ_i such that*

$$||\sum \epsilon_i x_i||_X \geq c\sqrt[3]{n}.$$

This is an f.d. property of infinite-dimensional normed spaces.

Theorem 46 (Ribe) *Let X and Y be normed such that they are homeomorphic as metric spaces with f, then ....*

7.1.1 Local Theory

Started by Grothendieck, James, ...

The study of isomorphic f.d. linear properties of normed spaces. All norms on $\mathbf{R}^n$ are equivalent, so this is essentially a quantitative theory. The Ribe theorem says that there is a hidden dictionary that translates linear properties that a priori only makes sense for normed spaces, into metric spaces. Formulated by Jean Bourgain.

Theorem 47 *$\exists$ a metric space $(M, d_.)$ such that for no (X, d_X) of nonpositive curvature there is an $f : M \to X$ satisfying*

$$\forall x, y \in M \omega(d_m(x, y)) \leq d_X(f(x), f(y)) \leq \Omega(\ldots).$$

"Dimension reduction" in one of the most important issues being tackled in statistics, machine learning etc, We look at this from a geometric point of view.

Example 23 *Given any 10^9 vectors in $\mathbf{R}^{1000000000}$, there are vectors $y_i \in \mathbf{R}^{329}$ such that $|x_i - x_j|$ and $|y_i - y_j|$ are within a factor of two, $\forall i, j$.*

When $d_M(x, y) \leq ||f(x)f(y)||_X \leq alphad_M(x, y)$: "$X$ embedded in M with distortion α". We ask whether we can do this, and aim for $\log n$ embeddings of spaces of dimension apparently n.

Recall definition of *expanders*. $|\{\{u, v\} \in E : \{u, , v\} \cap S \neq \emptyset \ldots\}|$ What is the average distance between people in this hall? All required $\binom{N}{2}$ calculations. So I impose the structure of 3-regular graph on the audience. Then I average the distances along the edges only. I want 100% certainty. Claims that, imposing the graph *in advance*, the audience can't defeat this. Claims this is guaranteed to work within $O(1)$ iff the graph is an $O(1)$-expander.

What is special about Euclidean geometry? The classical notion of an expander is "expander w.r.t. a Hilbert space". Clearly we need an expander (else can be defeated), the iff is a property of Euclidean metric. There are apparently nice metric spaces that do not admit any expander!

Theorem 48 *Any $O(1)$ expander w.r.t. a Hilbert space is an $O(\log n)$ expander* $\ldots.$

Hence any $O($1-expander does not admit a Lipschitz embedding into dimension $n^{o(1)}$ that preserves the average distance.

7.1.2 Geometric Graphs

Connect any two points of $||x, y|| < 1$ to get G. Then if k isn't large, the existence of graph means either that the average distance in V is small, or G is not an expander and we get a "partition" of the space. Hence the

Question 9 (Nearest Neighbour) *Given D consisting of n points in $\mathbf{R}^k$. Given a new point, what is approximately the nearest point of D? Can we do this in sublinear time?*

Q from speaker to self What happens outside l_1, l_2, l_∞?

7.2 Williamson: Representation Theory and Geometry

A representation is *faithful* if the map is injective. $\{S_n - \text{sets}\}$/isomorphism is subgroups of S_n under conjugation. S_3 permutes coordinates of $|R^3$, so invariant

spaces are "all equal" — L and "sub to zero" L. Hence $\mathbf{R}^3 = L \oplus H$. But in characteristic 3, $L \subset H$.

If representations are matter, simple representations are atoms, and semi-simple is "elements don't interact".

Fourier series are related to S^1, and Harmonic series to $SO(3)$.

This subject looks like algebra,but the deepest theorems have geometric proofs, often via invariant forms.

Theorem 49 (Maschke) *Any representation V of a finite group G over $\mathbf{R}$ or $\mathbf{C}$ is semi-simple.*

Example 24 $SU_2 = \left\{ A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} | AA* = I, \det(A) = 1 \right\}$, *unit quaternions. Natural representation on $\mathbf{C}^2$.*

Verma modules is an attempt to systematise these calculations. Has a basis space indexed by $\mathbf{N}$: $\{v_0, v_1, \ldots\}$. Move from v_i to v_{i-1} by $\lambda - i$, so if $\lambda \in \mathbf{N}$ we get a split.

g s a complex semi-simple Lie algebra, $h \subset g$ a Cartan subalgebra, and W the Weyl group, acting on h as a reflection group.. So $g = sl_n(\mathbf{C})$, $h =$ diagonals is an example. 'weight" $\lambda \in h$ corresponds to a Verma module Δ_λ. Ha a unique simple quotient L_λ, the simple highest weight module. In the *sl* example, $\lambda \notin \mathbf{N}$ gives $L_\lambda = \Delta_\lambda$, and $\lambda \in \mathbf{N}$ gives a finite dimensional L_λ

Conjecture 23 *Kazhdan–Lusztig, 1979]*

$$|\Delta_\lambda| = \sum_\mu P_{\lambda,\mu}(1)[L_\mu].$$

$P_{\lambda,\mu}$ are Kazhdan–Lusztig polynomials, and only depend on W.

First proved in the 1980s geometrically. Janzten conjecture says the graded multiplicity in $P_{\lambda,\mu}(v)$ and this again has a geometric proof Beilinson–Bernstein in the 1980. Algebraic proofs of point ?? by author+ in 2014/2016.

Weyl groups $\subset$ (mostly $=$) reflection groups $\subset$ (vast difference) Coxeter groups. $H := R/(R_+^W)$. d is the number of reflecting hypersurfaces in $h_{\mathbf{R}}$. There is an open cone $K \subset h_{\mathbf{R}}^*$.

Conjecture the KL polynomials have positive coefficients, again we have an algebraic proof (2013).

7.2.1 Modular representations

The analogy of KL is the Lusztig conjecture. Problem is that we don't have signature any more.

Conjecture 24 (Lusztig)

$$|\hat{\Delta}_\lambda| = \sum_B q_{A,B}(1)[\hat{L}_B].$$

True for large p (depending on the root system), e.g. $p > 1^{100}$ for SL_8 (we would like $p > 8$!). But I have proved (2013) it's false for primes growing exponentially in the rank

We can now work with ${}^p q_{A,B}$ as a o-KL polynomial.

Let V be a simple representation of S_n over $\mathbf{Q}$. Reduce modulo p to get modular representation $\mathbf{F}_p \otimes_{\mathbf{Z}} V$. So what are its multiplicities of simple modules? We only know or partitions with 1 or two rows

Conjecture 25 (Billiards) *Illustrated with a video, which implies that the behaviour of the coefficients of these p-KL polynomials are given by a dynamic system.*

7.3 Lubich: Dynamics, numerical analysis and some geometry

"Numerical Analysis works with algorithms on the real numbers, or their computer surrogates". Two major questions.

1. How can numerical methods be constructed that respect the geometry
2. What are the benefits of a structure-preserving algorithm

Consider Hamiltonian systems. Then (1) is Hamilton–Jacobi theory. $H(p(t), q(t)) = H(p(0), q(0))$ – i.e. energy conservation. Now what happens if we do Euler methods with step size h.

$$\frac{p_{n+1} - p_n}{h} = -\nabla_q H(p_{n+\alpha}, q_{n+\beta})$$

with $\alpha, \beta \in \{0, 1\}$. $\alpha = \beta = 0$ is explicit Euler, $\alpha = \beta = 1$ is implicit Euler, $\alpha \neq \beta$ in symplectic Euler. Applying to solar system, both implicit and explicit add spiral behaviour, but, even with $h \mapsto 10h$, the symplectic do much better.

> Thanks to the numerical experiments of the last two decades, we now know that the motion of the planets in the Solar System is chaotic — Laskar 2013.

Explicit Euler energy grows, implicit decays, and symplectic is oscillated around the true value, again even with $h \mapsto 100h$.

Let $y = (p, q)$, so $\dot{y} = J^{-1}\nabla H(y)$ with $J = \begin{pmatrix}) & I \\ -1 & 0 \end{pmatrix}$. Consider the time flow map $\phi_t : y(0) \mapsto y(t)$. $D\phi_t(y)^T J D\phi_y(y) = J$: symplectic property of ϕ and we want this preserved, hence these Euler methods. They were implicit in Hamilton–Jacobi, Numerical utility spotted in an unpublished preprint [de Vogelaere 1956], but published research only 1983, 1985 etc.

Note that a numerical solution is an exact solution of a modified problem, an asymptotic series in h. For a symplectic integrator applied to a Hamiltonian system, each of the perturbation terms on a Hamiltonian vector field, $f_j(y) =$

2D models in statistical mechanics
Quantum groups
temperature $q = 0$
Crystal Bases (199o) "$q = 0$" to all q
Global bases
LLTA theory — multiplicative properties
Quiver Hecke Algebras — Cluster Algebras
Monoidal categorifications of cluster algebras (2018)

$J^{-1}\nabla H_j(y)$ [Moser1968] locally and [E. Hairer1992] globally. 2nd half 1990s showed that errors grow only linearly, and near-preservation of KAM-tori.

But the backward error analysis needs $h\Omega << 1$. What happens when not?

Example of a nonlinear oscillator chain of springs. There's global conservation of energy, but transferred from spring to spring.

NF Standard solution

MFE Modulated Fourier Expansions in time. Embed the original system in a high-dimensional system. $q_t(t) \approx \sum_k z_j^k(t) e^{i(k\cdot\omega)t}$ with slowly-varying modulation functions z_j^k. Now use [E. Noether1918] to show that there are invariants. [FermiPastaUlam1955] was trying to address these points. Despite the fact that backward error analysis doesn't work, we can explain stability using the MFE process.

7.3.1 Dynamic low-rank approximation

Huge time-dependent matrices. $A(t) \in \mathbf{R}^{m\times n}$ Explicitly for solution of matrix ODE $\dot{A} = F(A)$. Approximate A by low-rank matrices $A(t) \approx Y(t) = USV$ where S is a small invertible matrix. Y is then in a low-rank manifold M. Modern idea in NA, but [Dirac1930] had it.

However, S might be ill-conditioned and S^{-1} features. This obstruction is geometric, $\frac{1}{\sigma_r}$ is the curvature of M at Y. So we split the tangent space protection.

$$P_Y Z = ZVV^T - UU^T ZVV^T + UU^T Z.$$

This reproduces rank-r matrices exactly, admits convergent error bounds that are independent of singular values. It is so robust because in each substep, the approximation moves along a flat sub-manifold, by analogy with a ruled surface.

7.4 Kashiwara: Crystal Bases and Categorifications

Chern Medal lecture.

$q = 0$ is absolute zero and something marvellous should happen.

$U_q(g)$ is the $\mathbf{C}(q)$-algebra generated by .... I thought it would have marvellous structure at $q = 0$ but couldn't find any. Not in $U_q(g)$, but in $U_q(g)$-modules.

Let $K = \mathbf{C}(q)$, V a K-vector space, (LB) is a local basis of (LB)$isalocalbasiso$ at $q = 0$ if

- L is a free A_o-modulo of V such that $V = K \times L$
- B is a basis of the $C(q)$-space L/Ql.

(LB) is a crystal basis of B if

- (L, B) is a local basis cxd
- ...

At $q = 0$ this complicated module structure becomes Combinatorics $\to$ Crystal Bases $\to$ Representation Theory.

7.4.1 Global bases

7.4.2 Quiver Hecke Algebras

$\{(R(n)\}$ is called the quiver Hecke algebra (KLR-algebra) associated with $(Q_{i,j})_{,j\in I}$. Let $R(n)$-gproj be the category of finite generated projective graded $R(n)$-modules. $R(m)$-gmod f.d. over k.

$K(R - gmod)$ is the Grothendieck ground of R-gmod. $[M] = [l] + [N]$ iff $P \to L \to M \to N \to$ is exact.

7.4.3 Cluster algebras

At $q = 1$, $A_1(n) \simeq D[c]$ is commutative. Hence we'd expect $b(A_q())$ to have nice multiplicative properties. Hence a conjecture (Berentstein–Zelevinsky. But Leclerc gave a counterexample (to the last clause). He conjectures rest are still true. Fomin–Zarevsky defined cluster algebras. Start with an initial cluster, ad apply "mutation" controlled by exchange matrix $\tilde{B} = (b_{ij})_{ij}$ Seed $C = \{X_1, \ldots, X_r\}$

7.4.4 Monoidal categorification

Assume Cartan matrix symmetric. Let $(C, \otimes)$ be an Abelian monoidal category. A simple $S \in C$ is real iff $S \otimes S$ is simple. Monoidal cluster $\{M_i\}_{1\le i\le r}$ is a finite let with $M_i \otimes M_i = m_{\otimes} M_i$. $\mu_k(\{M_i\}_{1\le i\le r}) = (M_1, \ldots, M_{k-1}, M'_k, M_{k+1}, \ldots, M_r)$ where M'_k is ..., is a mutation at k.

7.5 Pham Tiep: Representations of Finite Groups and Applications

Representation theory started in correspondence Frobenius–Dedekind.

$G = S_n$ the irreducible characters of G correspond to partitions of n. $\chi^\lambda \leftarrow \lambda \vdash n$. Irreducible representations correspond to strict partitions: $\lambda = (\lambda_1 > \lambda_2 > \cdots) \vdash$. What is the asymptotic?

Problem 5 *Given simple G and Field $\mathbf{F}$,*

1. *determine $\delta_p(G)$ and*
2. *classify irreducible $\mathbf{F}G$-representations of degree up to $\delta_p(G)$.*

Conjecture 26 (Alperin) $|\{$ *isomorphism classes of irreps of G over $\mathbf{F}G$* $\}| = \ldots$.

Conjecture 27 (McKay1971) *F a finite group and p a prime $P \in \mathrm{Syl}_p(G)$ There exists a bijection $\{\chi \in Irr(G) | P \not| \chi(1)\} \leftrightarrow^\pi \{\phi \in Irr(N_g(p), \ldots\}$*

Proved for various families of groups.

[MalleSpa"ath216] proved McKay conjecture for $p = 2$. Various results of the from "if all simple, then all groups". In particular need AWC-good. True for $Lie(p)$, Alt, Spor, but still need $Lie'(p)$, e.g. $PSL_p(q) : p \neq q$.

Problem 6 *Given a simple group G and $\mathbf{F}$*

1. *Determine smallest non-trivial degree $\delta_p(G)$ and*
2. *Classify ...*

Problem 7 *Let S be a finite simple group, $1 \neq g \in S$.*

1. *find an explicit, and as small as possible $0 < \gamma = \gamma(g) < 1$ such that $\frac{|\chi(g)|}{\chi(1)} \geq \gamma$ ...*
2. *...*

Fomin–Lulov case of S_n.

[Thorne2012a] defines the concept of "adequacy". If $P \not| |G|$, then G is adequate. If $p \geq 2n + 2$ then p is adequate.

Theorem 50 (Thompson1981) *If $G < GL(\ldots)$ ...*

I proved Thompson's conjecture holds with $C = 1184036$. conjecture it's in the 100s.

Applications to non-commutative Waring. $w(x, y) = xyx^{-1}y^{-1}$ is the Ore conjecture, now proved. $w = x^N y^N$ where $N = p^a q^b$ and produce of two prime powers. proved. $w = x^N y^N Z^N$ where N odd. Proved. Also results on random walks on groups.

7.6 Kohlenbach: Proof Theory

Origins in Hilbert's Programme. "Establish that uses of higher noneffective/transfinite ("ideal") principles I on proofs of combinatorial/finitistic ("real") propositions P can be eliminated, at least in principle." In principle, Gödel means that this is impossible in full generality.

Shift of emphasis (Kriesel $\geq$ 1951) use proof theory to extract new information from proofs of existential statements. Example: unwinding Littlewood's proof of $\pi(x) - li(x)$ by unwinding $RH \vee \neg RH$. Also bounds on Roth's theorem (first polynomial bounds).

Look at 'unwindings' ('proof minings') in analysis. Appropriate choice of *representations* of analytical objects suggested by the interpretations is crucial. There are interesting proofs that use WKL but allow for WKL-elimination doe to their logical form: uniqueness statement ($\in \forall \to \forall$).

About 2000, started to apply to abstract spaces, such as "Let X by a Banach space, $C \subset X$, $T : C \to C$ is nonexpansive ...". Numerous applications in fixed-point theory, convex optimisation, nonlinear semigroup theory etc. The finitary proof-theoretic analysis often generalises to geodesic settings (Hadamard spaces etc¿).

If $A^\omega[X, [[\cdot]]]$ proves a statement, then .... Works in metric, hyperbolic, CAT() etc. In the meta theorem, one can add a nonstandard boundedness axiom $\exists - UB^X$. Even though it's false, it leads to correct effective bounds.

Example 25 *A polynomial rate of asymptotic regularity in Bauschke's solution of the zero displacement conjecture. Previously known only for $N = 2$ or $Fix(T) \neq \emptyset$. Proof uses lots of abstract analysis, This is a $\forall\exists$ statement. Logical metatheorems therefore guarantee the extractability of a uniform rate of asymptotic regularity which only depends on $\epsilon > 0, N \in \mathbf{N}$ and majorants of $XinH$ and the projections.*

Example 26 (Proximal Point Analysis (PPA)) *H a real Hilbert space, $A : H \to 2^H$ a maximally monotone function. $F_{\gamma A} : -(Id + \gamma A)^{-1}$ be the resolvent of γA for $\gamma > 0$. Then $zerA = Fix(J_{\gamma\partial f})$.*

Theorem 51 (Mean Ergodic Theorem) *X Hilbert space, $f : X \to X$ linear and $||f(z)|| \leq |z||\forall Z$. $A_n(z) := \frac{1}{n+1}\sum_{i=0}^{n} f^{(I)}(x)$ for $n \geq 0$. Then for every $z \in X$, the sequence $(A_n(z))_n$ converges.*

We can say that $(A_n(z))$ admits at most $\cdots$ fluctuations. There are also nonlinear ergodic theorems.

Q Interactive Theorem Proofs?

A "Interactive" is the key. There is Coq work in this area. But currently pretty toy when fully automated, e.g. "$\sqrt{2}$ exists".

Chapter 8

9 August 2018

8.1 Kalai: Noise Stability, Noise Sensitivity and the Quantum Computer Puzzle

JHD missed the first part (Hotel checkout). See [Kal18].

8.1.1 Second Part

Model 4 is computation (Boolean circuits). Generated by NOT and AND, which are complete for classical computing. Copy slides from Widgerson@ICM2006. Multiplication is easy (polynomial time), but factoring is not known to be (the slide quoted $\exp(\sqrt{n})$, but speaker said $\exp(\sqrt[3]{n})$), and we all believe that not.

P

NP and conjecture NP $\neq P$.

Shor shows quantum factoring is $O(n^2)$.

Note that the quantum class is different: $Q \setminus$ NP, NP $\setminus Q$, $Q \setminus$ co-NP, co-NP $\setminus Q$ are all (conjecturally) different.

Model 5 is quantum computing. A qubit is a unit vector in $\mathbf{C}^2$. Then the state of an n-qubit is in $(\mathbf{C}^2)^{\otimes n}$. Gates are now unitary transformations. IBM's 70-qubit computer uses 7 different types of gates. "measuring" is exactly a probability distribution.

Model 6 is noisy quantum computing. Every qubit is corrupted with probability t at each cycle, and each gate might produce a "nearby" transformation with probability t.

Theorem 52 (Threshold: various 1995) *If t is small enough, noisy quantum circuits allow universal quantum computing.*

8.1.2 Permanents, Determinants and noise sensitivity of boson sampling

Model 7 is boson sampling. Given a complex $n \times n$ matrix with orthonormal rows. Sample sub-multisets of columns according to the absolute value-squared of permanents. Quantum computers can perform boson sampling. Fermion sampling is sets rather than multisets.

Model 8 is noisy boson sampling. Given a matrix A, sample $(\sqrt{1-t}A+\sqrt{t}G)$ where G is a normalised complex Gaussian $n \times m$ noise matrix.

Theorem 53 *When the noise level is constant, distributions given by noisy Boson sampling are well-approximated by low-degree Fourier–Hermite. Hence can be approximated by bounded-depth polynomial-size circuits.*

Theorem 54 *When the level is larger than* $\frac{1}{n}$ *... see Figure 8.1.*

Fermion is well within P. When we introduce noise, the difference between boson and fermion sampling vanishes, and both are inside bounded-depth computation, and indeed a subset known as ...

8.1.3 The Quantum Computer Puzzle

NISQ = Noisy Intermediate Scale Quantum systems. Major experimental efforts are aimed at demonstrating "quantum supremacy" using pseudo-random circuits and building good quality quantum error-correcting codes. Surface codes are a major component.

Conjecture 28 *Theorems 53 and 54 extend to all NISQ systems, and to all realistic forms of noise. For a wide range of noise levels, NISQ systems are very sensitive to noise, with a vanishing*

Hence we predict

1. For a larger amount of noise, you will get robust experimental outcomes but they will represent low-degree polynomial distributions which are far away from the desired noiseless ones.

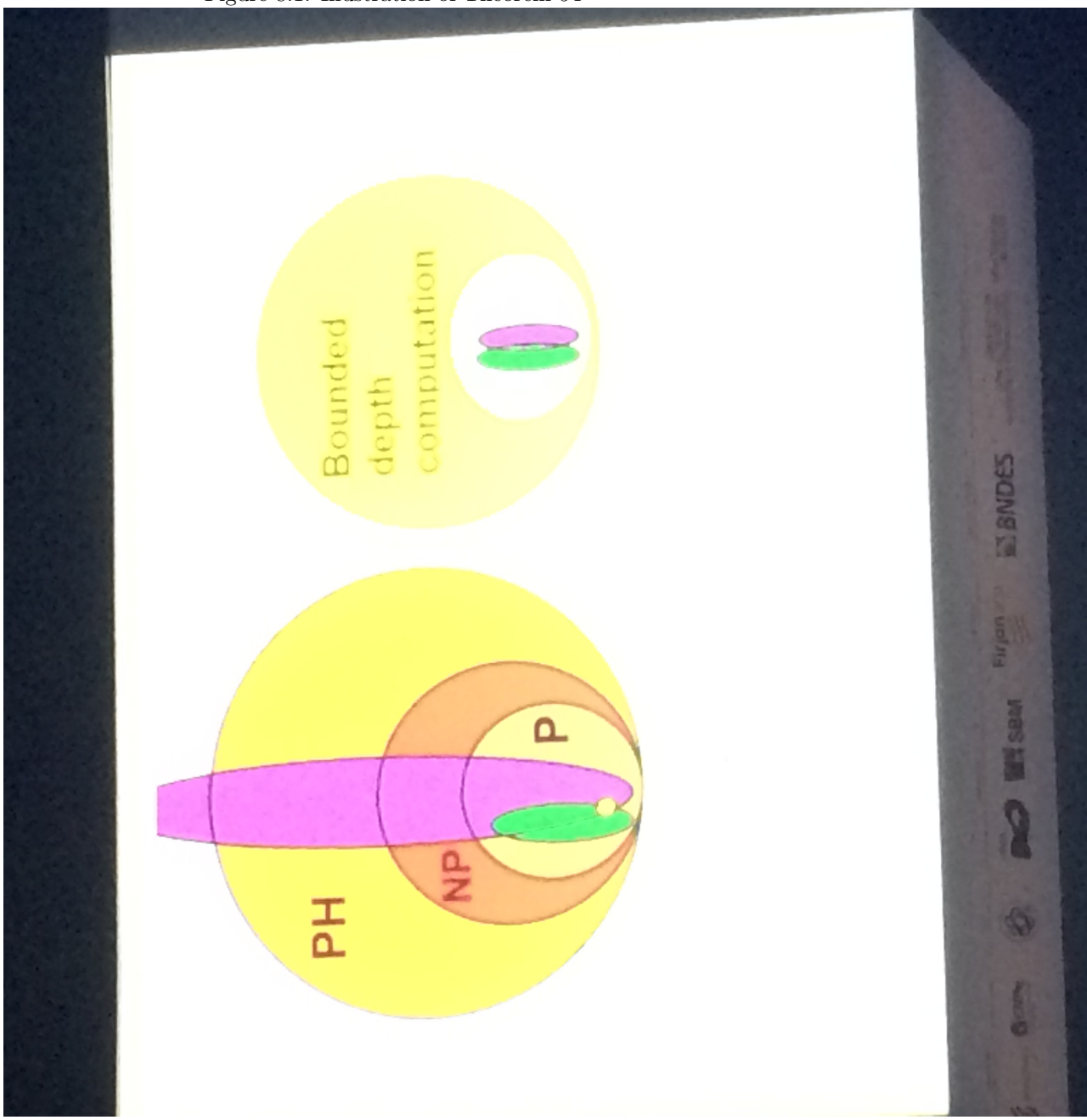
2. For a wide range of smaller amounts of noise, the outcomes will be chaotic. This means the resulting distribution will strongly depend on the fine properties of the noise,

3. The effort required to control i qubits to allow good approximations for the desired distribution will increase exponentially, and will therefore fail [Guess ≤ 20]

4. In the NISQ-regime, gated qubits will be subject to errors with large positive correlation. And so will any pair of entangled qubits. This will lead to a strong correlation of noise.

We will know by ICM 2022/2026.

Figure 8.1: Illustration of Theorem 54



8.1.4 Predictions

We have the following (at least as claims).

1. Quantum supremacy requires quantum error correction
2. Quantum supremacy is easier than quantum error correction

Hence Quantum supremacy is impossible. Related to high-dimensional expanders.

Note that classical computation requires error correction as well, but this is supported by very low-level computation, e.g. majority.

(A) Probability distributions described robustly by NISQ devices can be described by low-degree polynomials LDP is well inside AC0. See second photograph.

(B) Asymptotically low-level computation devices cannot lead to superior computation.

(C) Achieving QS is easier than achieving quantum error-correction.

So wait for the experiments!

Various possible laws.

0 Every quantum evolution is noisy (violates Quantum Mechanics, so discard it).

1. Time dependent noisy quantum evolutions are noisy
2. Noise (above the level allowing QC) is a necessary ingredient n modelling local quantum systems.
3. Quantum observables are noise-stable in the sense of Benjamini–Kalai–Schramm [BKS99].

Problem 8 *Study noise-sensitivity.*

1. *Prove that the crossing event in 3D percolation is noise-sensitive.*
2. *Study noise-stable versions of various models.*
3. *Study the math and physics above the fault-tolerance threshold.*
4.

Problem 9 ...

Final Note — SESAME: a third-generation synchrotron located in Jordan, part of a regional collaboration.

8.2 Jordan: Optimization, Computation and Statistics

Modern statistics has provided new challenges for optimisation. The computing platforms have changed (hence changing practical algorithms). I think ML/DL/AI/ etc are all poor buzzphrases for "algorithmic decision-making under uncertainty, in large-scale networks and markets". Note that von Neumann/Kolmogorov etc. could not have answered "are you a mathematician or a computer scientist.

8.2.1 Introduction

- gradients (Hessians are too hard)
- stochastics
- acceleration (out of Russian school: Nesterov *et al.*) [Nes83]

Problem 10 *Escape saddle-points (in large dimension, and efficiently). We escape down a strictly negative eigenvalue.*

"Bad local minima" were the problem of the 1980s. This has gone away as such, but many such implies many saddle points.

Theorem 55 *Asymptotically, gradient descent will avoid saddle points, but may take exponential time.*

Stochastic gradient descent can escape saddle points in polynomial time.

Basic gradient descent: $\min_{x\in\mathbf{R}^d} f(x)$ via $x_{t+1} = x_t - \eta\nabla f(x_t)$. The number of iterations is independent of the dimension.

Many real world problems are nonconvex, but have probably a single global optimum. E.g. PCA.

First Order Saddle Points. Assume f is ℓ-smooth (or gradient Lipschitz). $\forall x_1, x_2, ||\nabla f(x_1) - \nabla f(x_2)|| \leq \ell||x_1 - x_2||$.

Another assumption: f is p-Hessian gradient Lipschitz iff

$$\forall x_1, x_2, ||\nabla^2 f(x_1) - \nabla^2 f(x_2)|| \leq p||x_1 - x_2||.$$

Theorem 56 *PGD with $\eta = O(1/\ell)$ and proper r, find ϵ second order stationary points in $\tilde{\mathcal{O}}\left(\frac{\ell(f(x_0)-f^*}{\cdots}\right)$ in time $O(\log^4 d)$. [I believe the* log*, but the 4 is an artefact of the proof].*

Question: does Nesterov-style acceleration help you move faster past saddle-points? This requires a better understanding of acceleration.

8.2.2 Part II: Continuous Time

Newton differentiated, Laplace integrated and optimised. Physics uses both. Numerically, we have FE and Monte Carlo. Optimisation has probably not done as well.

Classical gradient descent $x_{k+1} = x_k - \beta\nabla(f(x_k))$ achieves a $O(1/k)$ convergence rate.

$$y_{k+1} = x_k - \beta\nabla(f(x_k)) \tag{8.1}$$
$$x_{k+1} = (1+\lambda_k)y_{k+1} - \lambda_k y_k. \tag{8.2}$$

Other two-phase descents now. $O(1/k^2)$ convergence rate for convex f. Here λ_k is an explicit function of the other problem parameters. In particular, accelerated gradient descent is not a descent algorithmthe function values can oscillate.

This is discretisation of $\dot{X}_t = -\nabla f(X_t)$. Nesterov becomes $\ddot{X}_t - \frac{3}{t}\dot{X}_t + \nabla f(X_t) = 0$.

$L(x, \dot{x}, t) = e^{\gamma_t+\alpha_t}(D_h(x) + e^{-\alpha_t}\ldots)$. Then Euler–Lagrange equation

$$\ddot{X}_t + (e^{alpha_t} - \dot{\alpha}_t)\dot{X}_t + e^{2\alpha_t+\beta_t}\left[\nabla^2 h(X_t + e^{-\alpha_t}\cdots\right]$$

Under ideal scaling, E-L equation has convergence rate $O(e^{-\beta_t})$. $\beta_t = \log t$ gives classical gradient descent, $\beta_t = \log^2 t$ gives Nesterov etc. Note that these all follow the same path. What about $\beta_t = t$?

8.2.3 Symplectic Integration

Discretisation that reflects the physical laws.

Shows his graph following Nesterov, but he can increase the step length whereas Nesterov goes unstable. In this case, $\times 5$. But we cheated and added a gradient flow term to the symplectic integrator.

8.2.4 Acceleration and Saddle Points

AGD is not a descent algorithm. If we lift to phase space, AGD is nearly a descent algorithm.

AGD is $\ddot{x} + \tilde{\theta}\dot{x} + \nabla f(x) = 0$. Seven line algorithm with 4/5=Nesterov, 2/3 is a perturbation step.

Theorem 57 *Perturbed AD converges as* $\tilde{\mathcal{O}}\left(\frac{\ell^{1/2}\rho^{1/4}(f(x_o)-f^*}{\epsilon^{7/4}}\right)$.

8.2.5 Acceleration and Stochastics

Some negative theorems, but I think irrelevant, as focused on overdamped diffusions.

Overdamped Langevin MCMC for an SDE $dx_t = -\nabla\cdots$. Very recently some guaranteed results in Total Variation, 2-Wasserstein or KL-divergence.

We look at underdamped. Quadratic improvement. So far been assuming $U(x)$ is m-strongly convex. New line of work on "convex outside a ball", e.g. mixture models. Ito calculus allows us to reduce to 1-D case.

Haven't said much about statistics. We don't have the population risk, only the sample risk. We can find a zeroth-order that finds ϵ-SOSP of F if $\nu \leq e^{1.5}/d$ which is provably optimal.

See my blog on AI. I am as cautious about AI as the previous talk was on Quantum.

8.3 Lafforgue: Global Langlands parameterization and shtukas for reductive groups.

L/number fields $\leftrightarrow$ analogy $\leftrightarrow$ L/function fields $\leftrightarrow$ L/l-adic sheaves $\leftrightarrow$ analogy L/D-modules $\leftrightarrow$ Conformal field theory.

Ring is commutative unitary ring. A number field is a finite extension of $\mathbf{Q}$.

$GL_n(\mathbf{Q})$. The vector space of automorphic forms is $L^2(\mathcal{L}, \mathbf{C})$ where $\mathcal{L} = GL_n(\mathbf{Z})/GL_n(\mathbf{C})$.

A *place* of a global field is a norm, and the completion of F for such a norm is called a *local field*, Either $\mathbf{R}$ for the Archimedean norm being $|\cdot|$, or p-adic $\mathbf{Q}_p$.

8.3.1 Idea of proof of theorem

Assume N is empty. We construct a commutative algebra $\mathcal{B}$ of "excursion operators" containing all the $\mathcal{H}_\nu$. Need the ℓ-adic cohomology of the stack of shtukas. Similar to Betti cohomology but has coefficients in $\mathbf{Q}_\ell$ or $\overline{\mathbf{Q}}_\ell$.

8.4 Closing Ceremony

121 years after the first, in ETH, ICM comes to the Southern Hemisphere. 630 travel grants to mathematicians from 73 countries, also 155 Brazilian applicants from 22/27 states. Just over 3000 registered participants including 238 accompanying. 40% Latin America 20% Europe, 14% North America. 5000 schoolchildren as well.

8.4.1 ICM 2022: Saint Petersburg

Visa-free arrival with ICM Registration tag (also free public transport with it) as with World Cup. Local support for young Mathematicians. ICM Satellites there or in Baltics. 2022 will be Year of Mathematics in Russia.

Bibliography

[ABB+17] A. Ambainis, K. Balodis, A. Belovs, T. Lee, M. Santha, and J. Smotrovs. Separations in query complexity based on pointer functions. *J. ACM*, 64(5):32:1–32:24, September 2017. URL: `http://doi.acm.org/10.1145/3106234`, `doi:10.1145/3106234`.

[Amb18] A. Ambainis. Understanding Quantum Algorithms via Query Complexity. *Proc. Int. Cong. of Math. 2018 Rio de Janeiro*, 3:3249–3270, 2018.

[BBC+01] R. Beals, H. Buhrman, R. Cleve, M. Mosca, , and R. de Wolf. Quantum lower bounds by polynomials. *J. ACM*, 48:778–797, 2001.

[Ber82] S.J. Berkowitz. On some relationships between monotone and nonmonotone circuit complexity. Technical Report Department of Computer Science University of Toronto, 1982.

[BKS99] I. Benjamini, G. Kalai, and O. Schramm. Noise sensitivity of Boolean functions and applications to percolation. *Inst. Hautes Études Sci. Publ. Math.*, 90:5–43, 1999.

[Coo71] S.A. Cook. The Complexity of Theorem-Proving Procedures. In *Proceedings of the 3rd Annual ACM Symposium on Theory of Computing*, pages 151–158, 1971.

[FKT16] E. Farhi, S. Kimmel, and K. Temme. A Quantum Version of Schöning's Algorithm Applied to Quantum 2-SAT. `https://arxiv.org/abs/1603.06985`, 2016.

[Gen09] C. Gentry. *A fully homomorphic encryption scheme*. PhD thesis, Stanford University, 2009.

[GMR88] S. Goldwasser, S. Micali, and R. Rivest. A Digital Signature Scheme Secure Against Adaptive Chosen-Message Attacks. *SIAM J. Comp.*, 17:281–308, 1988.

[GPW15] M. Göös, T. Pitassi, and T. Watson. Deterministic communication vs. partition number. In *Proceedings 2015 IEEE 56th Annual Symposium on Foundations of Computer Science - FOCS 2015. IEEE Computer Soc.*, pages 1077–1088, 2015.

[Gro96] L.K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings 28th Annual ACM Symposium on the Theory of Computing*, pages 212–219, 1996.

[Kal18] G. Kalai. Three Puzzles on Mathematics, Computation, and Games. `https://eta.impa.br/dl/PL008.pdf`, 2018.

[Kar84a] N.K. Karmarkar. A New Polynomial-Time Algorithm for Linear Programming. *Combinatorica*, 4:373–395, 1984.

[Kar84b] N.K. Karmarkar. A New Polynomial-Time Algorithm for Linear Programming. In *Proceedings 16th STOC*, pages 302–311, 1984.

[Maz77] B. Mazur. Rational Points on Modular Curves. *in Modular Functions of One Variable V*, pages 107–148, 1977.

[Nes83] Y.E. Nesterov. A method for solving the convex programming problem with convergence rate $O(1/k^2)$. *Soviet Math. Doklad.*, 27:372–376, 1983.

[RS00] K. Rubin and A. Silverberg. Ranks of elliptic curves. *Bull. Amer. Math. Soc. (N.S.)*, 39:455–474, 2000.

[Sha49] C.E. Shannon. The synthesis of two-terminal switching circuits. *Bell System Technical Journal*, 28:59–98, 1949.

[Sho94] P.W. Shor. Polynomial time algorithms for discrete logarithms and factoring on a quantum computer. In *Proceedings 1st Algorithmic Number Theory Symposium*, 1994.

[Wil18] V.V. Williams. On Some Fine-Grained Questions in Algorithms and Complexity. *Proc. Int. Cong. of Math. 2018 Rio de Janeiro*, 3:3431–3472, 2018.