ICM + STAIX 2026 Notes

October 22, 2026

[SJ: Feel free to edit.]

[SJ: Sunay's compact schedule is in the filetree.]

[SJ: We will add notes for talks we missed via notes from the proceedings / recordings.]

July 23

Yu Deng laudatio:

A *finite-dimensional Hamiltonian system* (q, v) with Hamiltonian $H(q, v)$ satisfies

$$\dot{q} = \nabla_v H, \qquad \dot{v} = -\nabla_q H.$$

A basic infinite-dimensional Hamiltonian system is the *periodic defocusing NLS*

$$i\partial_t u + \Delta u = |u|^{p-2}u, \qquad u : \mathbb{R} \times \mathbb{T}^d \to \mathbb{C}.$$

[SJ: Spatially 1-periodic functions.] Writing $u = q + iv$, this is Hamiltonian with

$$H(u) = \frac{1}{2}\int_{\mathbb{T}^d} |\nabla u|^2 \, dx + \frac{1}{p}\int_{\mathbb{T}^d} |u|^p \, dx = \text{KE} + \text{PE}.$$

If the RHS is set to zero, then formally

$$u(t) = e^{it\Delta}u_0.$$

Formally, statistical equilibrium should be described by the infinite-dimensional Gibbs measure

$$d\mu(u) = \frac{1}{Z}e^{-\beta H(u)} \, du.$$

The challenge is to *construct* μ. Recall that there is no infinite-dimensional Lebesgue measure du. (Indeed, in the case of $\ell^2(\mathbb{R})$, consider the disjoint unit balls centered at $(e_n)_{n\in\mathbb{Z}}$, all contained in

$B_2(0)$; subadditivity fails). One instead introduces a "Fourier cutoff" μ_N (a low-pass filter), and attempts to prove that $\mu_N \Rightarrow \mu$.

A typical Gaussian Free Field (GFF) sample (?) lives in the *negative Sobolev space* [SJ: Why relevant to the above?]

$$u \in H^s(\mathbb{T}^d) \quad \text{for every } s < 1 - \frac{d}{2}.$$

The *flow map* corresponding to the NLS is

$$\Psi_t : u_0 \longmapsto u(t).$$

The goals are to *construct* Ψ_t for μ-almost every rough u_0 and prove invariance of μ: $(\Psi_t)_{\#}\mu = \mu$.

A basic decomposition is

$$u(t) = z(t) + w(t), \qquad z(t) = e^{it\Delta}u_0,$$

where z remains rough and random, while the nonlinear remainder w is smoother. Deng–Nahmod–Yue constructed the dynamics for 2-dim'l defocusing NLS and proved invariance of its Gibbs measure in the "Φ^4_2 case".

This belongs to the broader *constructive field theory* program associated with Simon, Fröhlich, Aizenman, and Duminil–Copin.

Kinetic theory and Hilbert's sixth problem:

A second, related theme is deriving kinetic equations from microscopic dynamics. For N hard spheres of diameter ε, moving freely and undergoing elastic collisions, the *Boltzmann–Grad scaling* is $N\varepsilon^{d-1} \sim 1$. The limiting one-particle density $f(t, x, v)$ [SJ: ?] is expected to solve the *Boltzmann equation*

$$\partial_t f + v \cdot \nabla_x f = Q(f, f),$$

where Q is the Boltzmann collision operator.

Lanford proved this derivation only for sufficiently short times. Deng–Hani–Ma proved it for arbitrarily long times on which the corresponding Boltzmann equation has a regular solution. Their proof controls the full collision history using collision diagrams.

Sobolev spaces. The space $H^s(\mathbb{T}^d)$ measures L^2-smoothness. For integers $k \geq 0$, $H^k = \{u : \partial^\alpha u \in L^2,\ |\alpha| \leq k\}$. Thus $H^0 = L^2$ and $H^1 = \{u \in L^2 : \nabla u \in L^2\}$.

For any $s \in \mathbb{R}$, using the Fourier series $u(x) = \sum_{n\in\mathbb{Z}^d} \widehat{u}(n)e^{in\cdot x}$, define

$$\|u\|^2_{H^s} = \sum_{n\in\mathbb{Z}^d} (1 + |n|^2)^s |\widehat{u}(n)|^2.$$

For example, $\|e^{in\cdot x}\|_{H^s} = (1 + |n|^2)^{s/2}$, so $s > 0$ penalizes high-frequency oscillations. Larger s means greater smoothness.

Negative Sobolev spaces. For $r > 0$, H^{-r} has norm

$$\|u\|^2_{H^{-r}} = \sum_{n\in\mathbb{Z}^d} (1 + |n|^2)^{-r} |\widehat{u}(n)|^2.$$

The negative weight suppresses high frequencies, allowing rough functions (and distributions). In particular, $H^r \subset L^2 \subset H^{-r}$ and $\|e^{in\cdot x}\|_{H^{-r}} = (1+|n|^2)^{-r/2}$.

Jacob Tsimerman laudatio:

On which algebraic varieties V can rational points be Zariski-dense? (Latter means, no additional algebraic equation can capture the points?) [SJ: ?]

Mordell conjecture proven by Faltings in 1983 settles curves of genus 0 (projective lines) and genus 1 (elliptic curves), in which Zariski-density is equivalent to infinitude of (the group of?) rational points. Faltings also showed that this can happen only for translates of abelian subvarieties within abelian varieties (?). For general varieties, it is wide open.

Tsimerman proved the analogue for *Shimura varieties,* which are complex algebraic varieties that are also moduli spaces. Example: CM points in the moduli space $\mathcal{A}_g$ of abelian varieties, like $\mathcal{A}_g = \mathrm{Sp}_{2g}(\mathbb{Z})/\mathbb{H}_g$ (?).

We are looking for functions $u : H_S \to S$ that are Γ_S-invariant. Simplest example is j-invariant of an elliptic curve $\mathbb{C}/\Gamma$, which is $\mathrm{SL}_2(\mathbb{Z})$-invariant. Another simple example is a Siegel modular form (?).

The height of a special point: the Galois orbit of a CM j-invariant $\sigma \in \mathbb{C}$ for the right of integers of $\mathbb{Q}(\sqrt{-d})$ is given by a class number, estimated by Brauer–Siegel estimate, $[\mathbb{Q}(\sigma) : \mathbb{Q}] = h(-d) \gg d^{1/2-\varepsilon}$. [SJ: See notes from Katy?]

Hong Wang laudatio:

Contributions: Kakeya conjecture for $\mathbb{R}^3$, local smoothing for two-dimensional wave equations, Stein's restriction problem.

Notion of Kakeya set: contains every direction. In 1971, Charles Fefferman realized the connection between Kakeya and convergence of truncated Fourier integrals $S_N f$ (again, low-pass filter with cutoff $|\xi| \leqslant N$) to f in $L^p(\mathbb{R}^d)$. [SJ: The analogue of $S_N f$ in the Fourier series setting is convolution with the Dirichlet kernel, which Mihalis once said was at the root of all technical difficulties with the pointwise / L^p-convergence of Fourier series? And which motivated weaker notions of convergence, like Cesàro summation / Fejér kernels?]

Riesz showed convergence when $d = 1$. [SJ: Proof is in Stein–Shakarchi volume 4 chapter 2, via a clever holomorphic extension?]

Fefferman's ball multiplier theorem: for $d \geqslant 2$, no convergence except in L^2.

[SJ: Is non-convergence equivalence to non-boundedness of some (sequence of) operator(s)?]

[SJ: Precisely, what is connection between Kakeya and convergence?]

Kakeya conjecture: if $\mathcal{K} \subseteq \mathbb{R}^d$ is Kakeya, then Hausdorff dimension is d.

Idea behind Wang's proof: lower bound the volume of unions of δ-tubes. Multiscale analysis to exploit the structure of unions of tubes, "induction on scales", improving bounds incrementally.

Theorem 0.1 (Discrete Kakeya estimate)**.** *If $\mathbb{T} = \{T\}$ is a collection of δ-tubes with δ-separated directions, then $|\bigcup T| \geqslant c\delta^{\varepsilon}(|\mathbb{T}| \cdot |T|)$.*

Since union is large, tubes don't intersect too much, hence Minkowski dimension (which is the rate at which the Lebesgue measure of such δ-neighborhoods grows) is large.

Overall, Wang has worked on a family of conjectures related to properties of the Fourier transform (FT), a "tower resting on a needle" as Quanta put it:

- local smoothing for wave equation,
- Bochner–Riesz conjecture (generalizing the "ball multiplier operator" that maps f to $S_N f$, i.e., suppressing certain frequencies in the FT?),
- Stein's Fourier restriction conjecture (when does $\hat{f}$ restrict to S^{d-1} when $f \in L^p(\mathbb{R}^d)$?),
- Kakeya conjecture

Local smoothing $\implies$ B-R $\implies$ Restriction $\implies$ Kakeya

The last implication is due to Bourgain, using "decoupling theory"?

Shayan Oveis-Gharan laudatio: (by Spielman)

MCMC for sampling:

Fundamental question is how long to run the MC? Here, we care about sampling random combinatorial objects, which serve as "algorithmic primitives" for approximation algorithms [SJ: ?].

Matroids generalize the notion of the "set of spanning trees of a graph". Imagine jumping between spanning trees via random mutations.

Matroids also generalize independent vectors. How to find a uniformly random basis? Apply MCMC via random mutations (drop and add elements).

Mihail–Vazirani conjecture from 1989.

Proof techniques rely on log concave polynomials, high-dim'l expanders. Obtain sharp bound on eigenvalues that bound mixing and expansion.

Simplicial complex of independent sets; connect sets to their subsets. We care about walks on "horizontal slices" consisting of sets with a fixed size?

Trickle-down theorem from 2017.

Log-concave polynomials: generating polynomial of a matroid is

$$p(x_{1:m}) = \sum_{\text{basis } S} \prod_{i \in S} x_i = x_1x_2x_3 + x_1x_2x_4 + \dots$$

Fact: the generating polynomial of a matroid is *strongly log-concave* (SLC).

(Further, SLC implies all "link graphs" are "0-spectral expanders".)

Example: Hardcore model; can sample up to the "uniqueness threshold" (see discussion with Austen M about random colorings of d-regular graphs).

Example: Traveling Salesman Problem (TSP). It is NP hard to get an shortest tour. Here, we consider approximation algorithms.

Ideas: they use a random spanning tree, polygonal rep'n of min cuts, real stable polynomials / strong Rayleigh distributions.

Shayan Oveis-Gharan talk:

Currently, just have pictures in the `shayan_0724_talk` folder. Can transfer soon.

July 24

Nima Anari and Shayan Oveis-Gharan on sampling and high-dim'l expansion:

[SJ: Missing first half; see slides.]

HDX method: analyze MCs in "small pieces".

Spectral / entropic independeence is related to log-concavity of polynomials.

Spectral independence (SI) characterizes fast mixing in rw's on combinatorial structures (?). [SJ: I missed the intro that motivated SI.]

Related to *zero-freeness* of certain polynomials in certain subsets (sectors) of $\mathbb{C}^n$.

To prove spectral independence of a distribution, there are some techniques (complex-analytic and others).

Given a measure μ on $2^{[n]}$, recall the *generating polynomial* $g_\mu(z_{1:n}) = \sum_S \mu(S) z^S$, where $z^S := \prod_{i \in S} z^i$.

Down-up walk is often the right MC to use; delete something and add something back randomly.

Example: matroids. $\mu(S) = \mathbf{1}\{S \text{ is a basis}\}$. Down-up rw by dropping and adding back uar.

(Something about find matching in complement...)

Example: similarly, a random walk on monomer set. Delete two vertices and replace them with any valid configuration.

Spectral independence vs entropic independence. Fractional log-concavity is equivalent to spectral independence.

Now, suppose our algorithms "work". Are they computationally-efficient?

Sampling is an algorithmic task: time $= t_{\text{mix}} \cdot$ (per-step cost). The latter factor is important.

We think of rw's by moving between "horizontal" levels in a "layered graph" whose vertices are structures (see Shayan's talk yesterday); down and up again.

Example: sampling spanning trees, by dropping an ege and adding it back.

Corresponding to the down-up walk is the *dual down-up walk.* (Add a new edge first, then delete one. This is *slightly* more work, as you need to find the cycle formed by the added edge, and then delete a uar edge from that unique cycle.)

Example: sampling directed Eulerian tours.

Besides down-up walk for sampling, can consider *flip-repair walk*... for sampling random colorings?...

Joel Tropp on applied random matrix theory:

Want to understand eigenvectors, eigenvalues, singular vectors, singular values, operator norms, etc.

Wishart introduced RMT to statistics. How to study the sample covariance matrix $\hat{\Sigma} = \frac{1}{n}\sum X_i X_i^\top$, where $X_i \sim N(0, \Sigma)$ iid.

von Neumann and Goldstine also used RMT to understand rounding errors in numerical analysis, e.g., computing LU decomposition.

[SJ: Side-note by Tropp about the "recent" resurgence of probability / randomness in numerical linear algebra.]

Wigner studied spacing between eigenvalues of random Hamiltonians, inspired by nuclear physics. In an Annals paper, he showed that the spectrum of a Wigner matrix (symmetric with iid gaussians above diagonal and zero on diagonal) converges weakly to the semicircle distribution. [SJ: His techniques are combinatorial in flavor, via moment matching and trace expansions.]

Classical RMT continues Wigner's program, studying asymptotics of highly "structured" RMs. But in practice, RMs are not so structured, and we want non-asymptotic bounds.

Example: approximate a matrix with a sparse one. (E.g., Chatterjee's universal singular value thresholding algorithm.)

Key non-asymptotic technique: *matrix concentration.* Want bounds on $\|X - \mathbb{E}X\|$, in spectral norm (largest absolute singular value).

We work in the *independent sum model* $X = \sum_k S_k$, where S_k are independent RMs (not necessarily iid).

Example: sample covariance matrices, Wigner matrices.

[SJ: Which RMs belong to independent sum model? Matrix analogue of "sum of three gaussians" theorem?]

Most useful scalar concentration inequality: Bernstein's inequality. If S_k are centered independent and bounded by L a.s., then for $t > 0$, we have

$$\mathbb{P}[|\sum_k S_k| \geqslant t] \lesssim e^{-c(\frac{t^2}{v} \wedge \frac{t}{L})}.$$

(For small t, we have subgaussian behavior with variance proxy v; for large t, we have subexponential behavior due to a few rare large values of the S_k.)

The *matrix Bernstein inequality* of Oliveria–Tropp makes the same assumptions on RMs living in $\mathbb{R}^{d_1+d_2}$. It requires a *matrix variance proxy,* given by

$$v := \max\{\|\sum \mathbb{E}[S_k S_k^\top]\|, \|\sum \mathbb{E}[S_k^\top S_k]\|\}.$$

The tail bound itself is exactly the same (with a prefactor given by the sum of the dimensions $d_1 + d_2$).

Integrating the tail bound, up to constants, we get

$$\frac{1}{\sqrt{v}} \lesssim \mathbb{E}\|\sum S_k\| \lesssim \sqrt{v \log(d_1 + d_2)} + L \log(d_1 + d_2)$$

Proof relies on *matrix Lieb concavity* (?).

[SJ: I asked what this is. Tropp said it is the fact that the "trace exponential" $f(H) = \mathrm{tr}(\exp(A + \log H))$, for fixed A? It first arose in Lieb's study of the quantum stability of matter.)]

[SJ: Jiaoyang: are the log factors necessary in matrix Bernstein? Tropp says that it has to do with the *non-commutativity* of the individual matrices. Intuitively, why? Well, if the matrices were all cI, then it reduces to scalar Bernstein, in which case the constants should be one. Does the proof rely there some expansion that is easier to analyze in the commutative case? Or is it related to free probability?]

[SJ: Random graphs and RMs are of course related. Oliveira has a paper on random lifts of graphs via RMs.]

Back to example of matrix sparsification. [SJ: Isn't the "canonical" solution to project onto the space of same-dimensional rank $\leqslant k$ matrices?]

Simplest sparse RM model: sample $S \propto a_{ij} e_i e_j^\top$ wp p_{ij} for each entry (i, j), where $A = (a_{ij})$.

It is clearly unbiased, and for variance reduction, average r iid copies of S. This gives an estimator $\hat{A}_r := \frac{1}{r}\sum_k S_k$. Optimizing the matrix Bernstein RHS wrt $p = (p_{ij})$, we arrive at the sampling distribution $p_{ij} \propto |a_{ij}| + |a_{ij}|^2$.

Example: fast Laplacian solvers. I.e., if L is a graph's Laplacian, we want to solve $Lu = f$. (Applications include numerical PDE.)

[SJ: Why only for Laplacians and not other psd matrices? Due to *closure properties* of Laplacians under Schur complement (?).]

For a graph on n vertices and m edges, classical algorithms (triangular factorization $L = CC^\top$) take time $O(n^3)$.

Building on work of Spielman and others, Kyng–Sachdeva introduced a practical fast algorithm called `SparseCholesky` in $O(m)$ time.

Their idea: approximate $L \approx CC^\top$. Then do preconditioned conjugate gradient. [SJ: ?]

Recall Cholesky reduces a matrix to triangular form by repeatedly killing the first row and column, via "repeated rank-one deflation".

Idea: randomly sparsify the rank-one update. [SJ: ?] Analysis depends on matrix Bernstein inequality.

Carola Schönlieb on AI for inverse problems:

Inverse problems are everywhere, especially in medical / biological imaging. (Think MRI, CT, or cryo-EM.)

[SJ: Is the field of inverse problems necessarily statistical, or is it more general? The latter.]

[SJ: Why are "classical" methods being replaced by AI? In Bayesian language, do classical methods place too strong priors on the true signal, whereas DNNs can learn "richer" priors? Equivalently, do classical methods use the wrong regularization?]

Example: CT (computed tomography). Estimate an object $u : \mathbb{R}^3 \to$ from its line integrals $y(\theta, s) = \int_{\ell(\theta,s)} u dx + \eta$, where η is noise. This can be interpreted as an integral operator $T(u) + \eta$. [SJ: X-ray / Radon transform in Stein–Shakarchi volume 1?]

Example: cryo-EM. (See my notes from Amit Singer.) [SJ: Are these two isomorphic problems?]

Generally, given an object u living in a Banach space, and T a linear unbounded operator [SJ: Why?], given an observation $y = T(u) + \eta$, where η is "noise", recover the underlying signal u.

The answer is always minimizing a regularized loss of the form

$$\hat{u} = \text{argmin}_u \{D(T(u), y) + \alpha R(u)\}.$$

(Why use any regularization? Because typically the system $T(u) = y$ is highly *underdetermined.* You can always try to solve for the pseudoinverse of T, e.g., via filtered back-projection (FBP) algorithm, but it will underperform a decent regularizer.)

Example: DNNs. Specifically, the continuum limit of residual NNs. (Alex's questions about this paradigm.) [SJ: Shouldn't we be using a *delay differential equation* in a particular scaling limit?]

NNs model prior information; a *learnable prior* R_θ. Or, the NN parametrize the updates of an iterative solver $u_{k+1} = \Psi_\theta(u_k, \nabla D(Tu_k, y))$, like "learned primal-dual" algorithms; but we will

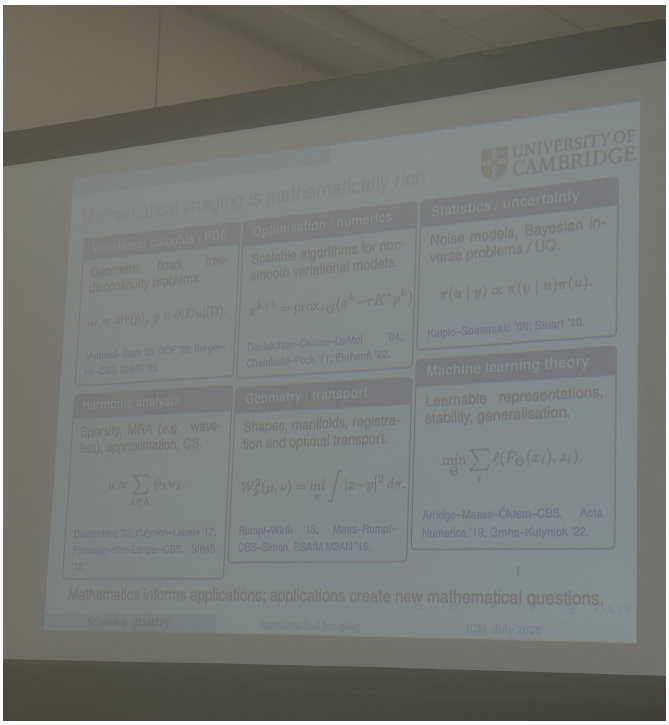


Figure 1: Perspectives.

not discuss this here.

[SJ: Ask her my + Shervin's question about non-physical artifacts (e.g., organs appear that don't exist) in a slightly different setting, of synthetic medical image data augmentation? She says this has *always* been an issue, even in the pre-generative days, when things like TV regularization induced "staircase" artefacts. However, the difference now is that with learned priors / deep generative models, we are not sure *why* these novel artefacts occur, we can't trace their causes. See Davis' comments on data attribution.]

When learning the regularizer R_θ, we want to impose constraints like Lipschitzness / convexity / etc.

There are heuristic approaches to UQ in the generative setting, which is also naturally Bayesian.

Alexandra Carpentier on statistical theory for ranking:

In this talk, we consider a few problems where there are *statistical-computational gaps* / tradeoffs. In certain problems, there is a nonempty region above the IT limit, but below the poly-time threshold.

The field started with Rigollet's paper on the link between planted clique detection and sparse PCA.

Carpentier found many false conjectures.

Ranking in tournament is given by match outcome matrix, sampled from the distribution defined by the underlying structure. (E.g., Bradley–Terry–Luce, Maxine's generalizations, etc.)

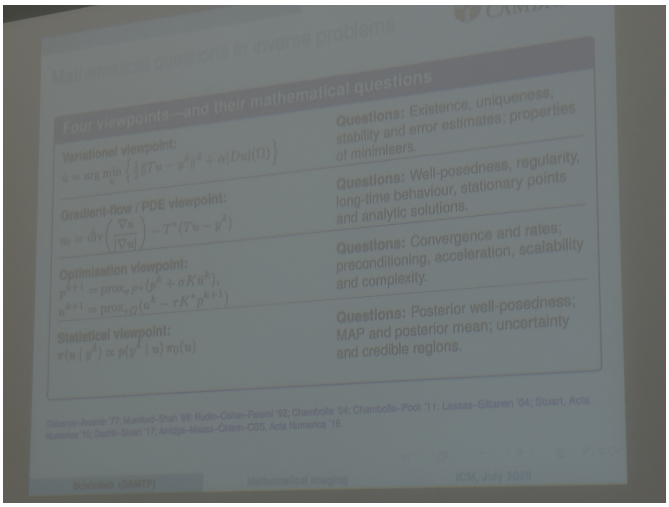


Figure 2: Problems.

Instead of studying "traditional" complexity theory, here, we use a relaxation called the *low-degree polynomial framework.* (Others include SoS relaxations, etc.)

Low-deg polynomails paradigm:

Given $Y = (Y_e)_{e \in \mathcal{E}}$ are observations, and z^* is the parameter of interest, define the loss $L(f) = \mathbb{E}[(f(Y) - z^*)^2]$. The key quantity $L^{*,D}$ is defined as the min over multivariate polynomials f in $\mathcal{P}^{\leq D}$.

"Low degree" means $D = O(\log |\mathcal{E}|)$. (Intuition based on decomposing a given f into its monomials.)

[SJ: Weird notation.]

Low-deg estimation framework encompasses many popular families of estimators, like spectral algorithms (see Fan's book), AMP, GD (approximately). It does not capture everything though.

Example: planted clique detection conjecture. [SJ: This is detection (aka testing), so it's easier than estimation; not quite an instantiation of the above?]

Schramm–Wein in 2022 characterized the IT and poly-time thresholds for estimating the support of the planted clique.

Proofs are based on choosing an ONB wrt some measure. Measure is simple to pick for For estimation it is much harder to pick the basis. See her recent papers on SBMs, clustering, etc. Clarifies the difference between detection and estimation.

Application to ranking.

Example: The observations are $Y_{i,k} \sim \text{Bern}(M_{i,k})$ for $i \in [n]$ and $k \in [d]$. Here, $M_{i,k}$reflects competency of expert i on question k, It was long known that there are no stat-comp gaps in the parametric setting. There is stat-comp gap in the *non-parametric* setting. (E.g., only impose shape constraints on $M_{i,k}$ (?).)

Key idea: can map from ranking to bipartite planted clique by embedding the clique as a rectangular submatrix of the binary observation matrix.

[SJ: See pics in Alex folder in the filetree.]

In the testing case, gap between $n^{3/4}$ and n (?).

In the estimation case, *no* stat-comp gap; immediately go from IT at

Complicated algorithm that stacks bipartite planted cliques.

So in planted clique we have SCTO for detection and estimation' whereas in ranking, SCTO for detection but not for estimation. The structure of the model matters.

Terry Tao on math in the age of AI:

See pics, as well as Alex Kontorovich's live tweets.

AI capability conjecture is still up in the air.

Terry's hypothesis, whose consequences he will discuss: "a reasonably strong version of the capability conjecture is true: AI tools will reasonably soon, become capable of performing a reasonable fraction of research-level math tasks, with reasonable levels of success."

Then what are our *goals and values?* Many are left implicit, and we need to make them explicit, lest we lose control of our field.

Technical goals are less important now, as AI begins to exceed us in this respect. We can't leave these questions to philosophers.

Attempt 1: solve as many problems as possible. (Could all be wrong.)

Attempt 2: generate as many verifiable solutions as possible. (Starting to happen with Erdős problems; but epistemologically, humans need to vet it.)

Attempt 3: generate as many understandable, well-written proofs as possible. (Currently AI models are bad at this. The AI places the emphasis on the wrong parts; spends same amount on time on the trivial and the crucial.)

In fact, Terry believes that proofs should contain some "natural friction". A proof that's "too slick" is not educational. Paradoxically, mistakes in human exposition can be helpful to the reader. Terry shows us his copy of a paper of the late Jean Bourgain on the Kakeya conjecture and Fourier restriction, which he read during grad school at Princeton. His experience is immortalized in the photo below.

Similarly with Thurston's "On proof and progress in mathematics" from 1994; human understanding is paramount.

But even if a proof obeys all of the above, it may not necessarily be *accepted / valued* by the community. You can't force it upon your colleagues. (This wasn't an issue until recently, since proofs were scarce. But now, we're flooded with proofs!) Having a *narrative* or showing the

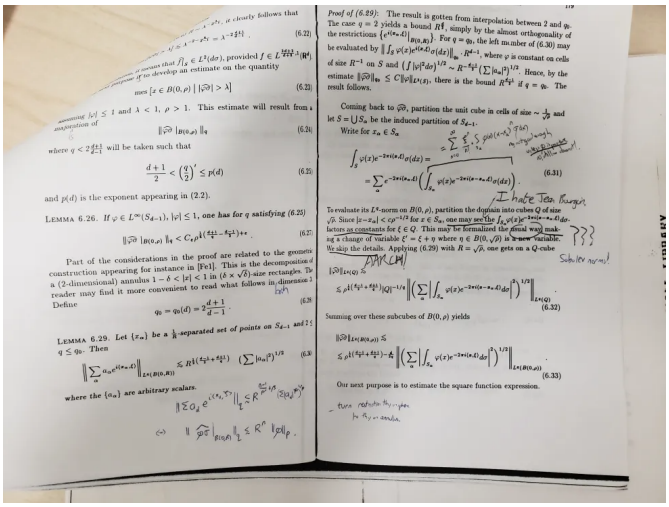


Figure 3: Bourgain's paper.

process helps.

Canonicalization of a proof is the final and slowest stage; deciding what become the "textbook" proofs.

In short, under our working hypothesis, we are transitioning from an era of *proof scarcity* to one of *proof abundance.* (A first stab at the question was the Leiden Declaration.)

We also need new workflows and infrastructures. And discuss these things openly, not merely passive recipients of the new technology.

[SJ: What's the situation for applied math? Try to envision the future of our field with Behrad.]

July 27

Hong Wang plenary:

Orthogonal projection problem.

Recall orthogonal projection operator $P_\theta : \mathbb{R}^2 \to \mathbb{R}$ along direction $\theta \in S^1$.

Given a subset $\Lambda \subseteq S^1$, how large is the Hausdorff dim of $P_\theta(E)$ for $\theta \in \Lambda$?

Recall the definition of *Hausdorff dimension;* the exponent of δ^{-1} in the number of δ-balls needed to cover E. (Ex: Cantor set has Hausdorff dimension $\log_3 2$.)

Marstrand's projection theorem: for a.e. θ, the Hausdorff dim is the smaller of $\dim E$ or 1.

Theorem 0.2 (Ren–Wang)**.** *if* $\dim \Lambda > 0$ *(in particular, can be "small"), then there exists* $\theta \in \Lambda$ *such that*

$$\dim P_\theta(E) \geqslant \min\{1, \dim E, \frac{\dim E + \dim \Lambda}{2}\}.$$

This is a special case of the *Furstenberg sets conjecture.*

Not true for $E \subseteq \mathbb{C}^2$ and $\Lambda = S^1 \subseteq \mathbb{C}P^1$. (Recall that $\mathbb{C}P^1$ is isomorphic to the space of lines through the origin in $\mathbb{C}^2$.) (See the bipartite construction.) [SJ: ?]

Kakeya sets. Recall the history. Connections to Fourier analysis, dynamics, number theory, etc.

Katz–Laba–Tao discovered the Heisenberg example in $\mathbb{C}^3$, showing that Wolff's estimate of $5/2$ is sharp. [SJ: Look into Heisenberg group example; Jef's class?]

The case of $n \geqslant 4$ is still open. [SJ: ?]

What's the connection between the projection problem and the Kakeya problem? Note that any Kakeya set can be written as $K = \bigcup_{(a,b)\in E} \ell_{a,b}$, where $\ell_{a,b} = \{(a,0) + (b,1)t : t \in (0,1)\}$.

Given $t \in (10,)$, consider the projection $P_t : (a,b) \mapsto (a+bt)$ from $\mathbb{R}^{2(n-1)}$ to $\mathbb{R}^{n-1}$. Thus by Fubini, $\dim K = \dim P_t(E) + 1$ for a "typical" t.

Wang's key result in this case is the Kakeya conjecture, that $\dim K = 3$ for any Kakeya set K in $\mathbb{R}^3$.

The proof strategy for these apparently distinct results relies on an inductive argument that bootstraps the argument for "sticky" sets E / K to *arbitrary* sets. Informally, sticky sets have the same structure at all scales.

Notion of AD regular set, a special type of sticky set. [SJ: ?]

Recall the δ covering number of E, denoted $|E|_\delta$. Bourgain showed that via pigeonholing, we may assume a given $E \subseteq [0,1]^n$ is *uniform.* [SJ: Add the definition of uniform via localization; roughly, "the δ covering number is the ρ covering number times the δ covering number of $E \cap B(x,\rho)$ for any $x \in E$, approximately"; i.e., $|E|_\delta \approx |E|_\rho \cdot |E \cap B_\rho(x)|_\delta$ for all $x \in E$.] The *branching function* $f_E(a) = \log|E|_\delta a / |\log \delta|$ is an n-Lipschitz function in $[0,1]$. This function is a better "summary statistic" of the set E than a *single* dimension. Recall that a Lipschitz function is differentiable a.e.; this will be key later. [SJ: See Stein–Shakarchi volume 3.]

Using the branching function, we can easily define different kinds of fractal sets; E is an AD regular set, stick set, Frostman, or Katz–Tao set if the shape of f_E is of a particular form.

We want to show the C-O-S-Wang-Zakharov theorem, which states that if Λ is AD regular, and $\dim \Lambda > 0$, then the same claim as above holds.

AD regular cannot be replaced by sticky. [SJ: ?]

Sticky case: we show that for $(1-\delta^\eta)$-fraction of $\theta \in \Lambda$, we have $|P_\theta(E)|_\delta > \delta^\varepsilon |E|_\delta$ ("same size").

Proof idea: "sum-product structure in extremal examples". Apply Bourgain's sum-product theorem to "distinguish $\mathbb{R}$ and $\mathbb{C}$", i.e., if $A \subseteq [1,2]$, then $\max\{|A+A|_\delta, |A \cdot A|_\delta\} > |A|_\delta^{1+\delta}$ whenever $|A|_\delta \ll \delta^{-1}$. [SJ: At least one of $A+A$ and $A \cdot A$ is "large".]

Finally, how do we go from sticky to general sets?

Bourgain showed that any E can be thought of as uniform, so we can define the branching function, which is Lipschitz, hence a.e. differentiable. Thus locally, f_E is a.e. close to a segment. Using this, Shmerkin showed that every uniform set is sticky at suitable scales. [SJ: ?]

But there is a hiccup. Consider f_E that is continuous and piecewise linear with two pieces (specifically, $f(t) = t$ and then $f(t) =$. Then E is Katz–Tao and Frostman, but not sticky.

Goal: $|P_\theta(E)|_\delta \gtrsim \delta^{-1}$.

It is easy to see that $|P_\theta(E)|_\delta$ is the number of δ-*tubes* in the direction $\theta^\perp$ needed to cover E.

Key calculation: $|P_\theta(E)|_\delta \approx |P_\theta(E)|_\rho \cdot |P_\theta(E \cap T_\rho)|$, where T_ρ is a "tilted" ρ-tube that intersects many of the fibers (the δ-tubes in the direction $\theta^\perp$). We can understand these intersections. [SJ: ?]

[SJ: Got lost with the explanation of $\tilde{E}$ via T_ρ and "induction on scales".]

[SJ: Ask Evan for his input on the GMT topics based on his course with De Lellis.]

Eliran Subag on the geometry of spin glasses:

We consider sequences of gaussian fields. Consider a *stationary gaussian field* H_N on S^{N-1} with *covariance function* $\mathbb{E}H(x)H(y) = N\xi(x \cdot y)$. H is also called the *Hamiltonian.* (We drop subscripts when they are obvious from context.)

(The admissible ξ can be characterized.)

We want to study the *geometry of level sets.*

Example: pure p-spin model $H(x) = \sqrt{N}\langle J, x^{\otimes p}\rangle$, where J are iid $N(0,1)$. These generalize the disordered mean-field Sherrington–Kirkpatrick (SK) Hamiltonian. Also related to complexity of solving random polynomial systems. [SJ: See Montanari–Subag paper?]

Given inverse temperature β, recall the *free energy* $F = \frac{1}{N}\log Z$. If one can calculate the free energy, then one unlocks the key properties of the system (e.g., by taking various derivatives). By gaussian concentration, F concentrates. We are interested in its limiting expectation $F(\beta) = \lim \mathbb{E}F_{N,\beta}$. One can show that most of the contribution to $F_{N,\beta}$ coomes from approximately one level set $\{H = NE\}$. [SJ: Boltzmann's argument?]

The most celebrated result in the field is the *Parisi formula* for the (limiting) free energy, given by

$$F(\beta) = \inf_{\mu \in M_1([0,1])} \mathcal{P}_{\xi,\beta}(\mu),$$

where $\mathcal{P}$ is the Parisi functional [SJ: ?]. (Related breakthroughs / tools used in the rigorous proof by Guerra–Talagrand: Ghirlanda–Guerra identities, Aizenman-Sims-Starr scheme, ultrametricity conjecture by Panchenko. Also related to Hamilton–Jacobi equations.)

Recall the *Gibbs measure* $G(A) \propto \int_A e^{\beta H(x)} dx$ for $A \subseteq S^{N-1}$.

Asymptotic Gibbs measure: record the *infinite overlap array* $R = (\langle x^i, x^j \rangle)_{i,j}$ where x^i are iid "replicas". This array converges as $N \to \infty$. The limit is characterized by *Ruelle probability cascades*, with the Parisi measure as the *order parameter.*

We want a more intuitive and geometric understanding of the Gibbs measure, beyond the Parisi formula.

Given $B \subseteq \mathbb{R}$, consider the set of critical points $\mathrm{Crit}_N(B)$ with energy $H \in NB$. Seminal works by Auffinger, Ben Arous, Cerny, and Fyodorov. Analysis via Kac–Rice formula and the second moment method.

Theorem 0.3 (Informal, Subag)**.** *For the pure p-spin model with $\xi(t) = t^p$, for any $p \geqslant 3$, for large β: (1) the measure concentrates in thin bands around the few highest critical points; the center points are approximately orthogonal; and the free energy can be computed precisely, as $F = A + B \log N/N + O(1/N)$.*

Definition of $\mathrm{Band}(m, \delta)$ given $m \in \mathbb{B}^N$. For thin bands $\delta \ll 1$, for *pure models,* the Gibbs measure concentrates on such bands (around critical points). Talagrand in 2010 proved a "pure states decomposition" for which for *any* mixture ξ, the measure concentrates on a small number of such bands whp. See also the work of Jagannath 2017.

Q: can we characterize bands with non-exponentially-small Gibbs mass?

[SJ: See informal theorem by Subag, 2023 on which m have δ-bands of non-small Gibbs mass.]

Thouless–Anderson–Palmer (TAP) approach from 1977. This says that q is good iff F satisfies a particular fixed-point equation involving an *Onsager reaction term.*

Optimization algorithms.

As $\beta \to \infty$, the measure concentrates on maximizers of the Hamiltonian. [SJ: This is due to his backwards convention $e^{\beta H}$, as opposed to $e^{-\beta H}$.]

By *ultrametricity,* the ground states are arranged in a tree on S^{N-1} with orthogonal edges.

Algorithmic optimization of spin-glass Hamiltonians has been considered since the 1980s; see Kirkpatrick et al. 1983. (Physicists believe these spin-glass models are archetypical for non-convex optimization in high dimensions. [SJ: Has this been made precise?!])

If the model is *full-RSB* [SJ: which has a geometric interpretation in terms of the ultrametric tree?], then [SJ: Add.]

[SJ: Missed details on Sellke's algorithm.]

[SJ: Referenced Montanari collaboration on converting "solve $F_i(x) = c$ for all i" to optimizing a Hamiltonian of the form $E = \frac{1}{N}\sum_i (F_i - c)^2$. Related to Smale's 17th problem.]

[SJ: Related to Chewi's proposed meta-isomorphism between optimization algorithms and sampling algorithms, in the canonical non-convex setting of spin glass Hamiltonian optimization?]

[SJ: My "applied" question: has anyone tried to approximate general non-convex landscape

(say, of the form in DL) with certain spin-glass Hamiltonians? His response: unrealistic.]

[SJ: There is a phenomenon of chaos with respect to the temperature.]

[SJ: See follow-up notes with Shervin about algorithmic stochastic localization for sampling from the SK model.]

Philippe Rigollet's talk on mean-field dynamics of transformers:

...

Richard Samworth's talk on nonparametric statistics under shape constraints:

...

Rina Foygel-Barber's talk on algorithmic stability and distribution-free statistics:

...

[SJ: Olivia's explanation of Ramon's new *strong convergence* notion for sequences of random matrices (it implies that the spectra converge in Hausdorff distance to the limiting spectrum a.s.), as well as the associated *polynomial method*. The idea is to use this notion to study things like the spectral gap of Ramanujan graphs. She also explained her work on *lower bounds* for the growth of radially-symmetric solutions to the Schrodinger equation via integral representations and oscillatory integrals. Austen drew a connection to exponential sums and van der Corput's method from analytic number theory, where similar lower bounds are needed. She also described Jacob Shapiro's iteration of Prof Aizenman's math-physics seminar.]

[SJ: Ramon's comment: when doing math-stat, make sure that when you critique an existing method, your shiny new method actually has a theoretical guarantee that *adds* something. Don't prove the same guarantee as before, and pretend that it explains the better empirical performance. In that case, it's better to simply drop the theorem from your paper. Fundamentally, statistics is about gaining a better understanding of data analysis, and that need not necessarily express itself in the form of a buttoned-up theorem.]

July 28

Gerard Ben-Arous' plenary:

...

Peter Bartlett's plenary:

...

Jiaoyang's talk:

...

Erdal Arikan's talk:

...

Laszlo Lovasz' Abel talk:

...

Aug 1

Misha Belkin talk:

...

Rajarshi Mukherjee talk:

...

Martin Wainwright talk:

...

Mengdi Wang talk:

...

Andrew Wilson talk:

...

Eric Xing talk:

...

Miscellaneous

Automorphic forms with Hari:

[SJ: Look at Lax's book on spectral theory.]

Hari was telling me about his group's research yesterday In number theory they study these functions called *automorphic forms.*

They are functions defined on manifolds that have important arithmetic properties. Like elliptic curves are surfaces ($\mathbb{C}/\Lambda$ for some lattice $\Lambda = \mathbb{Z}[\omega]$) or equivalently complex curves which have connections to some Diophantine equations.

Thought of as functions on the complex plane (often this can be done), automorphic forms definitely are periodic, and also have other cool symmetries and satisfy some differential equations. Examples roughly include the j invariant, the Weierstrass p-function (which along with its derivative parametrize elliptic curves, much like sine and cosine for circles), modular forms (which are of the form $f(z)/(az+b)^k$ for some holomorphic f and $k \in \mathbb{N}$, and Eisenstein series.

The way Hari says automorphic forms arise is via certain differential operators on manifolds. The most important of which is the *Laplacian operator,* which everyone in science is familiar with, though perhaps not on manifolds. We know from physics that eigenfunctions of the Laplacian have a close connection to Fourier analysis, and often provide a nice basis to work in. Think complex exponentials in Euclidean space or spherical harmonics on the sphere. As it turns out Laplacians also give rise to automorphic forms, though it's at this point that I begin to get lost.

He reminded me that there are three *model spaces* in geometry for manifolds of constant curvature, namely the *upper half plane,* and two of its quotients, the *Riemann sphere* and the *Poincaré ball.* The first is a zero curvature model space, the second positive, and the third negative (this stated formally is called the *uniformization theorem*).

The group of interest to him is $\mathrm{SL}_2(\mathbb{R})$, the group of conformal automorphisms of the upper half plane, as well as its quotients. In other words he's studying the symmetry groups of these model spaces. Why might he do that? I think he said it's because the action of $\mathrm{SL}_2(\mathbb{R})$ on each of these spaces induces an action on the functions out of those spaces, hence induces an action on the automorphic forms he cares about. [SJ: Confirm.]

He then reminded me that representation theory gives us the language to understand spaces of functions on groups. In his case the group is $G = \mathrm{SL}_2(\mathbb{R})$, and he wants to grok the representation theory of this finite [SJ: three?] dimensional infinite group, and also find a good basis for $L^2(G)$. He reminded me of the *Peter—Weyl theorem*, which states that the set of group characters form a basis (in the L^2 sense) for $L^2(G)$, generalizing the familiar fact for finite groups.

And jumping back to functions on the model spaces, he casually dropped the phrase *"the Laplacian is in the center of the universal enveloping algebra"*. Which on future prodding simply is the fact that the Laplacian commutes with many other interesting operators, hence its eigenfunctions diagonalize many operators simultaneously.

The point being, it's worth it to study model spaces, functions on those model spaces especially in dimensions 2 and 3 where there are clear arithmetic connections (Anshul is studying Laplacians on hyperbolic 3-manifolds and certain predictions coming from QFT), the eigenfunctions of the Laplacian on these model spaces, study $G = \mathrm{SL}_2(\mathbb{R})$ which (alongside its quotients) arises as a symmetry group of those model spaces, and study G's representation theory / symmetries via techniques from Lie groups.

Optimization for modern ML:

Peter Bartlett ICM plenary, Stephen Wright plenary, Mark Schmidt ICML talk, Jason Altschuler's reading group?, Behrad + Thomas's perspectives, deeper understanding of Adam and how to improve it, analogies to real analysis, dynamical systems, stochastic processes, and control, formalize Chewi's isomorphism between sampling and optimization...

References