Contents

1 Disordered systems, rank-one matrix estimation and Hamilton-Jacobi equations (Jean-Christophe Mourrat) 1
1.1 Definitions . . . 3
1.2 Interlude on Hamilton-Jacobi equation . . . 5
1.3 Back to Curie-Weiss . . . 9
1.4 Matrix estimation problem . . . 10
1.5 Main steps . . . 12

2 Critical and near-critical percolation (Gady Kozma) 14

3 Branching random walks: some recent results and open questions (Nina Gantert) 21

4 Gibbsian line ensembles in integrable probability (Ivan Corwin) 30
4.1 Non-touching geometric random walks . . . 31
4.2 Airy line ensemble . . . 35
4.3 KPZ line ensemble . . . 38

5 Mixing and hitting times for Markov chains (Perla Sousi) 42

These are notes from the Online Open Probability School (OOPS) 2020.

- Webpage for OOPS: `https://www.math.ubc.ca/Links/OOPS/`.
- Source is available at `https://github.com/holdenlee/oops`. Contributions, corrections, and comments welcome; feel free to send a pull request. Any errors are probably due to me! You can ping me on zulip or email me at `holden.lee@duke.edu`.

  A direct link to these notes is `https://www.dropbox.com/s/u4lpayi98r06un5/oops.pdf?dl=0`.

1 Disordered systems, rank-one matrix estimation and Hamilton-Jacobi equations (Jean-Christophe Mourrat)

We consider the problem of estimating a large rank-one matrix, given noisy observations. This inference problem is known to have a phase transition, in the sense that the partial recovery of the original matrix is only possible if the signal-to-noise ratio exceeds a (non-zero) value. We will present a new proof of this fact based on the study of a Hamilton-Jacobi equation. This alternative argument allows to obtain better rates of convergence, and also seems more amenable to extensions to other models such as spin glasses.

References:

- Paper: [Mou18], `https://cims.nyu.edu/~jcm777/HJinfer.pdf`

  Approach based on weak solutions: [Mou19], `https://cims.nyu.edu/~jcm777/HJrank.pdf`
- Book for background material: [FV17], available at `https://www.unige.ch/math/folks/velenik/smbook/index.html`
- Other approaches
  - Lelarge-Miolane: [LM19], `https://arxiv.org/abs/1611.03888`
  - Barbier-Macris: [BM19], `https://arxiv.org/abs/1705.02780`

2020/5/18 Lecture 1

Suppose students are assigned one of two dormitories. They put on a sorting hat, which decides which dorm they go in.

The students are $i \in \{1, \ldots, N\}$. An assignment is $\sigma \in \{\pm 1\}^N$. The sorting had optimizes the quality of interaction between i and j, J_{ij}. Suppose (J_{ij}) are independent standard Gaussians. The larger J_{ij} is, the more that i and j like to be together. We want to maximize $\sigma \mapsto \sum J_{ij} \mathbb{1}_{\{\sigma_i = \sigma_j\}}$. By a linear transformation this is equivalent to maximizing $\sum J_{ij}\sigma_i\sigma_j$.

What is $\max_{\sigma\{\pm 1\}^N} \sum J_{ij}\sigma_i\sigma_j$ as $N \to \infty$? Because the J_{ij} can be positive or negative, we can't make all the students happy. Thus we can say there are **frustrations** in the problem. These models are called spin glasses in the literature. It's difficult to find the optimum: making local moves, you may have to decrease the objective before increasing it.

I want to consider a softer version of the maximum. We look at the **spin glass model**[1]

$$\mathbb{E}\frac{1}{N} \log \sum_{\sigma \in \{\pm 1\}^N} \exp\left(\frac{\beta}{\sqrt{N}} \sum_{i,j=1}^N J_{ij}\sigma_i\sigma_j\right)$$

If β is large this is dominated by the maximum. This is like a relaxation of the problem. We should expect what's inside is order N, so we divide by N.[2]

Parisi in the late 70's (1979) proposed an answer for what this becomes as $N \to \infty$. It's a fairly complicated formula.

Guerra 03 and Talagrand 06 proved it rigorously. I find it mysterious; I want to think about a slight variation of the problem. Instead of connections between each i, j, think of them organized in two layers; there are interactions between but not within the layers (the graph is bipartite). This seems an innocent modification, but I could not understand what to write instead of the complicated formula!

[1] Note that the normalization is $\frac{1}{N}$ instead of $\frac{1}{\sqrt{N}}$ because here the $J_{ij} = 1$.

[2] Can we fix the number of $+1$'s and -1's? You can change the reference measure; this can be encoded as changing the reference measure. $\beta = 0$ is summing over reference measure. $\beta \to \infty$ recovers the maximum. β small is high temperature, β large is small temperature.

Now I consider rank-one matrix esimation/inference. The question is statistical: we only observe a noisy version of a rank-one matrix. Can we recover information about it?

Here are some concrete settings where this is useful:

- You are Netflix, you want to make recommendations for your customers. A simple model is that whether or not a person likes a movie is captured by a few parameters of the movie (action, introspection, sad/happy, etc.) and customer, and is a linear function of the parameters. Then you have a large low-rank matrix. A simplification is that it's a matrix of rank 1. I'll describe rank 1, but it's not hard to generalize.

- Community detection: The US is polarized, and there is a binary variable that will predict whether two people will be friends.

The common thread is the relation with certain partial differential equations, called **Hamilton-Jacobi equations**.

The **Curie-Weiss model** is a simple model that can be solved in many ways. I want to emphasize the method that uses intuition with Hamilton-Jacobi equations. Next when we turn to rank-1 matrix estimation, the proof will be almost the same.

Our derivation is not standard; if you want to see a more standard derivation see Friedli and Velenik [FV17].

Can you meaningfully recover information about the rank-1 matrix? In the Ising model there is a phase transition between an ordered and disordered state. In this inference problem there is also a phase transition. When signal-to-noise ratio is too small (weak), you cannot recover meaningful information. After the threshold, you can recover partial information.

1.1 Definitions

We want to study the probability measure that to each $\sigma \in \{\pm 1\}^N$, associates a weight proportional to

$$\exp\left(\frac{t}{N}\sum_{i,j=1}^{N}\sigma_i\sigma_j + h\sum_{i=1}^{N}\sigma_i\right).$$

The second term doesn't have interaction; it "tilts" the σ_i, giving each a preference. Here $t > 0$ but $h \in \mathbb{R}$. Define the expected value

$$\langle f(\sigma)\rangle_{k,h} := \frac{\sum_\sigma f(\sigma)\exp\left(\frac{t}{N}\sum_{i,j=1}^{N}\sigma_i\sigma_j + h\sum_{i=1}^{N}\sigma_i\right)}{\sum_\sigma \exp\left(\frac{t}{N}\sum_{i,j=1}^{N}\sigma_i\sigma_j + h\sum_{i=1}^{N}\sigma_i\right)}.$$

The subscripts are omitted when clear.

Define the free energy

$$F_N(t,h) = \frac{1}{N}\log\sum_\sigma \exp\left(\frac{t}{N}\sum_{i,j=1}^{N}\sigma_i\sigma_j + h\sum_{i=1}^{N}\sigma_i\right)$$

You might say: this is the normalization constant, we care about the measure. This is misleading because the normalization constant is the generating function of quantities we care about. You are calculating the exponential (moment) generating function of these variables. If you understand the mgf, you understand these quantities.

Moment generating function. Differentiating gives

$$\partial_h F_N = \frac{1}{N} \frac{\sum_\sigma \left(\sum_{i=1}^N \sigma_i\right) \exp\left(\frac{t}{N}\sum_{i,j=1}^N \sigma_i\sigma_j + h\sum_{i=1}^N \sigma_i\right)}{\sum_\sigma \exp\left(\frac{t}{N}\sum_{i,j=1}^N \sigma_i\sigma_j + h\sum_{i=1}^N \sigma_i\right)} = \frac{1}{N}\left\langle \sum_i \sigma_i \right\rangle \tag{1}$$

$$\partial_t F_N = \frac{1}{N}\left\langle \frac{1}{N}\sum \sigma_i\sigma_j \right\rangle = \left\langle \left(\frac{1}{N}\sum \sigma_i\right)^2 \right\rangle. \tag{2}$$

This model is simple; I can rewrite $\partial_t F_N$ in a simple way. The derivatives are all order 1. It's a good starting point to notice that

$$\partial_t F_N - (\partial_h F_N)^2 = \left\langle \left(\frac{1}{N}\sum \sigma_i\right)^2 \right\rangle - \left\langle \frac{1}{N}\sum \sigma_i \right\rangle^2.$$

This is the mean magnetization, the variance of the magnetization. Idea: The variance is lower-order, so as $N \to \infty$, F_N solves the equation with 0 on the right.

F_N is the mgf of $\sum \sigma_i$. So in particular it should encode the variance of the variable in some way. I should find a way to express it in terms of F_N. Looking at the second derivative is a good idea.

$$\partial_h^2 F_N = \frac{1}{N}\left\langle \left(\sum \sigma_i\right)^2 \right\rangle - \frac{1}{N}\left(\left\langle \sum \sigma_i \right\rangle\right)^2.$$

So we have shown

$$\partial_t F_N - (\partial_h F_N)^2 = \frac{1}{N}\partial_h^2 F_N. \tag{3}$$

This is a very important observation: everything is expressed in terms of F_N. We can forget about the probability measure, definition in terms of probability measures, and just think about what F_N satisfies this equation, and what happens when N becomes large. It also suggests that as $N \to \infty$, the RHS will vanish.

I think of this as an evolution equation; think of t as time. It will be useful to understand what happens when $t = 0$, the initial conditions.

$$\begin{aligned} F_N(0,h) &= \frac{1}{N}\log \sum_\sigma \sum_\sigma \exp\left(h\sum_i \sigma_i\right) && (4)\\ &= \frac{1}{N}\log \sum_\sigma \prod_{i=1}^N \exp(h\sigma_i) && (5)\\ &= \frac{1}{N}\log\left(e^h + e^{-h}\right)^N = \log(e^h + e^{-h}) && (6)\\ F_N(0,h) &= F_1(0,h) =: \psi(h). && (7) \end{aligned}$$

This does not depend on N.

The most important connection is that the h-derivative is the mean magnetization (1). What do we do with (3) and (7)?

1.2 Interlude on Hamilton-Jacobi equation

Let's take a step back and think about what the equation is saying. We need to think about what it means to be a solution of

$$\partial_t f - (\partial_h f)^2 = 0. \tag{8}$$

The first thing to look for is a C^1 function that solves the equation pointwise. What's the problem with this?

The problem is that there is a phase transition in the Ising model. When t is small nothing impressive happens. For fixed small t, $F_N(t, h)$ will be smooth.

But for larger t, if h is positive, then the mean magnetization is positive and away from 0, and if h is tiny negative, then the mean magnetization is negative and away from 0. There will be a jump in the derivative of the function; it looks like $|h|$. The equation is not solved pointwise at $h = 0$.

QA:

- If you change the measure, you can create lots of discontinuities. Considering P with bounded support on $\mathbb{R}$, we can consider

  $$\int \exp(\cdots)\, dP^{\otimes N}(\sigma).$$

  You can play with P to create more corners in the limit. This changes what this ψ function.

- What's the notion of convergence for solutions as $N \to \infty$? All functions are uniformly Lipschitz. By Arzela-Ascoli there are convergent subsequences. We can take uniform convergence as the topology. If you prove convergence for some topology, you can bring it to $C^{0,1}$ topology.

- HJ equation can be solved by characteristics. Is there a probabilistic interpretation of the PDE method of characteristics? I'll try to bypass it. Barbier and Macris [BM19] use different techniques: construct characteristics for finite N. The method I present is more convenient, you don't need to follow characteristics closely, just look at whether characteristics are contracting or expanding.

2020/5/19 Lecture 2

Recall that we showed $\partial_t F_N - (\partial_h F_N)^2 = \frac{1}{N}\partial_h^2 F_N$ (3), and hoped the limit object will solve the equation (8),

$$\partial_t f - (\partial_h f)^2 = 0.$$

We have to think about what it means to solve the equation. Because of phase transitions, we don't expect solutions to be C^1. We need to lower our expectations about what being a solution means.

Every Lipschitz function is differentiable almost everywhere. What if we just ask the equation to be satisfied almost everywhere? The problem is that the solution is not unique (given the initial value at $t = 0$, there are multiple possible solutions). Here are some examples.

- 0 is a solution.
- $(t, h) \mapsto t + h$ or $t - h$.
- From these 3 solutions, we can construct another solution with the "tent function,"

$$f(t,h) = \begin{cases} t+h, & -t \le h \le 0 \\ t-h, & 0 \le h \le t \\ 0, & \text{else.} \end{cases}$$

  We have an infinite number of solutions by taking combinations of translations of these.

But these are not the solutions we care about! We know the solution should be convex in the h variable; the tent function is not convex. If we add in this condition, then we get a unique solution.

Definition 1.1: We say that a Lipschitz function $\mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ is a **weak solution** of the HJ equation (8) if

1. (HJ) is satisfied almost everywhere.
2. For every $t \ge 0$, the mapping $h \mapsto f(t,h)$ is convex.[3]

Proposition 1.2 (Uniqueness)**:** If f and g are two weak solutions of the HJ equation with $f(0,\cdot) = g(0,\cdot)$, then $f = g$.

Why do we want this? This proposition says that two solutions with same initial condition are equal. What we want is actually a refinement of the statement: we want to compare finite-N solution (to (3)) for N large to the solution of the HJ equation. Instead of two solutions being equal, we want to show the almost-solution and the solution are close.

Proof sketch. Denote $w = f - g$. We have that almost everywhere

$$\begin{aligned} \partial_t w &= (\partial_h f)^2 - (\partial_h g)^2 \\ &= \underbrace{(\partial_h f + \partial_h g)}_{=:b} \partial_h w. \end{aligned}$$

Then

$$\partial_t w - b \partial_h w = 0 \text{ (a.e.)}.$$

[3] It suffices to be semi-convex: there is a lower bound on the Hessian.

Because f, g are convex in h, the derivative of b is positive.[4] The rough idea is to look at how the integral $I(t)$ evolves:

$$I(t) = \int w(t,h)\,dh.$$

Suppose this is well-defined. Then integrating by parts (suppose there are no boundary terms)

$$\begin{aligned}\partial_t I(t) &= \int \partial_t w = \int b\partial_h w \\ &= -\int \underbrace{\partial_h b}_{\geq 0}\, w\end{aligned}$$

This says that $I(t)$ wants to come back to 0. Some problems with this proof:

1. We pretended w has a fixed sign.
2. We integrated over whole space. This is a problem because of boundary terms when integrating by parts.

Now let's be more rigorous. Let $\phi(x) = \frac{x^2}{1+x^2}$, $v = \phi(w) = \phi(f-g)$. Showing $w = 0$ is equivalent to showing $v = 0$. We have by the chain rule that

$$\partial_t v - b\partial_h v = 0.$$

Now v has a fixed (positive) sign. This solves item 1.

Denote $L = \|\partial_h f\|_{L^\infty} + \|\partial_h g\|_{L^\infty} + 1$. Fix $T < \infty$, and study

$$\begin{aligned}J(t) :&= \int_{-L(T-t)}^{L(T-t)} v(t,h)\,dh \\ &= \int_{-R_t}^{R_t} v\end{aligned}$$

Note J is Lipschitz in t, so it has a derivative a.e. By the Fundamental Theorem of Calculus,

$$\partial_t J = \int_{-R_t}^{R_t} \partial_t v - L(v(t,R_t) + v(t,-R_t))$$

$$\int_{-R_t}^{R_t} \partial_t v = \int_{-R_t}^{R_t} b\partial_h v = -\int_{-R_t}^{R_t} (\partial_h b)v + [bv]_{-R_t}^{R_t}.$$

[4]For higher dimensions, the divergence of this vector field has a fixed sign, which says that the flows of the vector field being convergent or divergent.

The term $[bv]_{R_t}^{-R_t}$ may be positive. However, it will be compensated by the $-$ terms because $|b| \leq L$.

$$\begin{aligned}
\partial_t J &\leq -\int_{-R_t}^{R_t} \underbrace{(\partial_h b)}_{\geq 0} \underbrace{v}_{\geq 0} \leq 0 \\
\partial_t J &\leq 0.
\end{aligned}$$

We know $J(0) = 0$ and $J \geq 0$. We conclude that $J \equiv 0$, and $v = 0$ a.e. Because v is Lipschitz, $v = 0$ and $f = g$. $\square$

To be fully rigorous, we should justify interchanging the derivative and integral, and differentiating b (since f, g are only Lipschitz, and may not be twice differentiable).

Exercise 1.3 (for credit)**:** Make this proof rigorous.
Hint: convolve with a function to make it smooth. See the book [Eva10].

What makes the proof work is that the nonlinearity H is convex:

$$\partial_t f - H(\nabla f) = 0.$$

Difficulty comes in when H is not convex or concave.

Any limit point of a solution should be a solution: You can try to run the proof, show the sequence of functions is a Cauchy sequence, and the limit is a solution.

We can write down a formula for the solution.

Proposition 1.4 (Hopf-Lax formula)**:** Let ψ be convex and Lipschitz.[5] The function

$$f(t,h) = \sup_{h' \in \mathbb{R}} \left(\psi(h - h') - \frac{(h')^2}{4t} \right)$$

is the weak solution of

$$\begin{aligned}
\partial_t f - (\partial_h f)^2 &= 0 \\
f(0, \cdot) &= \psi.
\end{aligned}$$

[6] Note the last term is related to the convex dual of the nonlinearity: the convex dual of $p \mapsto p^2$ is $q \mapsto \frac{q^2}{4}$. See [Eva10].

Exercise 1.5: Suppose that f solves $\partial_t f - (\partial_h f)^2 = 0$ and $f(0,h) = \psi(h) = \log(e^h + e^{-h})$.
Show that for $t > 0$ small that $\partial_h f(t, 0) = 0$. You can use the Hopf-Lax formula.
For $t < \infty$ large, $\partial_h^+ f(t,0) > 0 > \partial_h^- f(t,0)$; there is a phase transition in the derivative.

[5] It suffices to be locally semi-convex: for every $\delta > 0$, there exists $C_\delta < \infty$ such that for all $\geq \delta$, $h \mapsto f(t,h) + C_\delta h^2$. (Note the lower bound is degenerate as $t \to 0$.) More generally, this is true under mild regularity assumptions, though we have to define weak solution differently.

[6] Is there a modification for the pre-limit PDE in F_N? There would be some Brownian motion. f will be the some expecation of some exponential of Brownian motion. For our application, we will not have a completely closed formula for F_N, so we cannot write a formula of this form before passing to the limit.

1.3 Back to Curie-Weiss

In general it's not true that convergence of functions implies convergence of their derivatives, so we need to prove the following.

Proposition 1.6: If (t, h) is a point of differentiability (in h) of f, and if $F_N \to f$ (pointwise), then

$$\partial_h F_N(t, h) \to \partial_h f(t, h).$$

Proof. F_N is convex in h, so

$$F_N(t, h') \geq F_N(t, h) + \underbrace{\partial_h F_N(t, h)}_{\text{bounded}}(h' - h).$$

Because the derivatives are bounded, we can take a subsequence along which $\partial_h F_N(t, h) \to p$. Then

$$f(t, h') \geq f(t, h) + p(h' - h),$$

and p must be $\partial_h f(t, h)$. □

Convergence of F_N. We have

$$\partial_t F_N - (\partial_h F_N)^2 = \frac{1}{N}\partial_h^2 F_N.$$

We want F_N to be close to the HJ solution. Let $w = F_N - f$; then

$$\partial_t w - b\partial_h w = \frac{1}{N}\partial_h^2 F_N$$
$$\text{where } b = \partial_h F_N + \partial_h f.$$

Let $v = \phi(w)$. Before, the difference solved the same equation, but now we have to be more careful.

$$\partial_t v - b\partial_h v = \phi'(w)\frac{1}{N}\partial_h^2 F_N.$$

The argument is similar, but the RHS is different here. Define

$$J(t) = \int_{-R_t}^{R_t} v(t, h)\, dh.$$

We have an extra term

$$\partial_t J \leq \underbrace{\int_{-R_t}^{R_t} b\partial_h v - L(v(t, R_t) + v(t, -R_t))}_{\leq 0} + \int_{-R_t}^{R_t} \underbrace{\phi'(w)}_{\leq 1} \frac{1}{N} \underbrace{\partial_h^2 F_N(t, h)}_{\geq 0}\, dh$$

$$\partial_t J \leq \frac{1}{N}\int_{-R_t}^{R_t} \partial_h^2 F_N(t, h)\, dh = \frac{1}{N}[\partial_h F_N]_{-R_t}^{R_t} \leq \frac{2}{N}.$$

where $|\partial_h F_N| \leq 1$ follows from (1). Recall $J(0) = 0$. So $J(t) \leq \frac{2t}{N}$.

Exercise 1.7: Clean up this proof! How do you get pointwise convergence?

In the rank-1 case, we find some quantities similar to F_N. We have a similar situation with a function which satisfies a similar equation, and want to show it converges to the true solution. We need some L^1 estimate.

Key point: If there are "error terms on the right hand side," we need to estimate it in L^1 in the h variable uniformly in h (locally). We want a "local L^∞_t, L^1_h estimate."

2020/5/21 Lecture 3

1.4 Matrix estimation problem

We saw how to justify the solution in the limit for the Curie-Weiss model, showing error term disappears. Now we consider the matrix estimation problem. This has already been solved by other people. The first that gave a complete solution of this problem is [LM19]. Shortly after, [BM19] gave another proof.

We consider the problem in [Mou18] using the approach in [Mou19].

Problem: Let $\overline{x} = (\overline{x}_1, \ldots, \overline{x}_N)$ be a vector of bounded independent random variables with law $P_N = P^{\otimes N}$. For some $t > 0$, we observe

$$Y = \sqrt{\frac{2t}{N}}\overline{x}\overline{x}^\top + W$$

where $W = (W_{ij})$ are independent standard Gaussians. (We don't assume symmetry.)

For example, we want to understand

$$\begin{aligned}\mathrm{mmse}_N(t) &= \frac{1}{N^2}\inf_{\widehat{\theta}}\mathbb{E}\left[|\overline{x}\overline{x}^\top - \widehat{\theta}(Y)|^2\right]\\ &= \frac{1}{N^2}\mathbb{E}\left[|\overline{x}\overline{x}^\top - \mathbb{E}[\overline{x}\overline{x}^\top|Y]|^2\right].\end{aligned}$$

where infimum over all estimators. We assume we can find the best possible estimator given the data, which is the conditional expectation given Y.

Informally,

$$\mathbb{P}\left[\overline{x} = x \text{ and } Y = y\right] = dP_N(x)\,dy\,\exp\left(-\frac{1}{2}\left|Y - \sqrt{\frac{2t}{N}}xx^\top\right|^2\right)$$

Now compute the conditional law,

$$\mathbb{P}[\overline{x} = x|Y] = \frac{\exp\left(-\frac{1}{2}\left|Y - \sqrt{\frac{2t}{N}}xx^\top\right|^2\right)\,dP_N(x)}{\int\exp\left(-\frac{1}{2}\left|Y - \sqrt{\frac{2t}{N}}x'x'^\top\right|^2\right)\,dP_N(x')}$$

Expand the exponent and remove $-\frac{1}{2}|Y|^2$ (which doesn't depend on t, x) to obtain the following.

Definition 1.8: Define

$$H_N^\circ(t,x) = \sqrt{\frac{2t}{N}} Y \cdot xx^\top - \frac{t}{N}|x|^4.$$

Here Y is the sum of 2 terms, $\sqrt{\frac{2t}{N}}\overline{x}\overline{x}^\top + W$. Substituting and expanding gives

$$H_N^\circ(t,x) = \sqrt{\frac{2t}{N}} x \cdot Wx + \frac{2t}{N}(x\cdot\overline{x})^2 - \frac{t}{N}|x|^4.$$

The first term is the important term; it is like the spin glass model. The other terms will help us. They aren't negligible, but they are here to make out life easier. In summary, the conditional law of $\overline{x}|Y$ has the form of a Gibbs measure, with the most important part looking like the spin glass case.

Denote

$$\langle f(x)\rangle = \frac{\int f(x)\exp(H_N^\circ(t,x))\,dP_N(x)}{\int \exp(H_N^\circ(t,x))\,dP_N(x)}.$$

The important difference with the Curie-Weiss model is that $H_N^\circ(t,x)$ is still random.

Exercise 1.9 (for credit)**:** Make this rigorous: Show that

$$\langle f(x)\rangle = \mathbb{E}[f(\overline{x})|Y].$$

Why is this model, which looks like the spin glass model, simpler than the spin glass model? The important property is the following.

Proposition 1.10 (Nishimori)**:**

$$\begin{aligned}
\mathbb{E}\langle f(x)\rangle &= \mathbb{E}[f(\overline{x})] \\
\mathbb{E}\langle f(x)g(x')\rangle &= \mathbb{E}[\langle f(x)\rangle \langle g(x)\rangle] \\
&= \mathbb{E}[\langle f(x)\rangle \mathbb{E}[g(\overline{x})|Y]] \\
&= \mathbb{E}\langle f(x)g(\overline{x})\rangle .
\end{aligned}$$

where x' is an independent copy of x under $\langle\cdot\rangle$. More generally,

$$\mathbb{E}\langle f(x,x')\rangle = \mathbb{E}\langle f(x,\overline{x})\rangle$$

What is x and why is it different from $\overline{x}$? Some intuition: Suppose that if we look at Y, we learn nothing about $\overline{x}$; then x is a resample. If Y reveals perfect information on $\overline{x}$, then $x=\overline{x}$. The general case is like an interpolation between these trivial cases.

The independent copy x' is called a **replica**.

This is not trivial, it's not saying $x=\overline{x}$. It is only true because I average outside.

This will help us identify the partial differential equation; we can link it back to the original signal.

We will identify the PDE and compute the mmse by looking at the derivative in the limit, you can recover the phase transition.

Here $\overline{x}$ is the original random vector, and x is our best guess for $\overline{x}$ given what we've seen.

We want to take the log and look for a PDE. But I only have t in this function. If I describe the Curie-Weiss model and had not thought of adding the h; then I would be stuck, I can differentiate in t many times and not find a equation. The relation was between derivatives in t and h. I have t, but what do I relate it to? We play the same game, add a h parameter. Sufficient to relate derivatives.

We have a quadratic term; I would like to add a linear term in x in the model. We want to do this with a constraint: I don't want to destroy the Nishimori property. For Curie-Weiss this was easy, but for spin glasses it's subtle, not easy to figure out in general.

We need to enrich the model somehow!

It has to be rich enough to relate derivatives, and simple enough to compute what happens. Suppose we also observe,

$$\sqrt{2h}\overline{x} + z$$

where z is a standard Gaussian vector. Now I recompute the conditional law of $\overline{x}$ given this rich observation,

$$H_N(t, h, x) = H_N^{\circ}(t, x) + \sqrt{2h}x \cdot z + 2hx \cdot \overline{x} - h|x|^2.$$

The important part is the term $\sqrt{2h}x \cdot z$.

Now I can define the free energy,

$$\begin{aligned} F_N(t, h) &= \frac{1}{N}\mathbb{E} \log \int \exp(H_N(t, h, x))\, dP_N(x) \\ \overline{F}_N(t, h) &= \mathbb{E} F_N(t, h) \end{aligned}$$

I take the expectation because the expression inside depends on the randomness in the noise.

Theorem 1.11. *$\overline{F}_N \to f$ (local uniform convergence) which is a solution of*

$$\begin{aligned} \partial_t f - (\partial_h f)^2 &= 0 \qquad\qquad \text{on } \mathbb{R}_+ \times \mathbb{R}_+ \\ f(0, \cdot) &= F_1(0, \cdot) \end{aligned}$$

Note $\overline{F}_N(0, \cdot)$ does not depend on N.

1.5 Main steps

The proof has two parts. First we relate the derivates to a variance.

Proposition 1.12:

$$\partial_t \overline{F}_N - (\partial_h \overline{F}_N)^2 = \frac{1}{N^2}\mathbb{E}\left\langle (x \cdot \overline{x} - \mathbb{E}\left\langle x \cdot \overline{x}\right\rangle)^2 \right\rangle.$$

Next, we find a way to express this variance in terms of other things. In the Curie-Weiss model we also described this in terms of derivatives. Here we don't have an equality.

Proposition 1.13: We have

$$\mathbb{E}\left\langle (x\cdot\overline{x} - \mathbb{E}\left\langle x\cdot\overline{x}\right\rangle)^2\right\rangle \le \frac{1}{N}\partial_h^2\overline{F}_N + \mathbb{E}[(\partial_h F_N - \partial_h\overline{F}_N)^2]$$

The last term is still random, so we have to control it.

I will justify why this is true, and how to conclude the main result. For the first part, we use the following.

Lemma 1.14.

$$\partial_h\overline{F}_N = \mathbb{E}\left\langle \frac{x\cdot\overline{x}}{N}\right\rangle$$
$$\partial_t\overline{F}_N = \mathbb{E}\left\langle \left(\frac{x\cdot\overline{x}}{N}\right)^2\right\rangle.$$

The key ingredient is Itô calculus without Itô. Letting X be a standard gaussian,

$$\mathbb{E}[\exp(\sqrt{2t}X - t)] = 1.$$

(Taking the expectation of Brownian motion and subtracting t, we get a constant. Here $\sqrt{2t}X$ is like B_{2t}.) Differentiating with respect to t we should get 0.

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{2t}}X - 1\right)\exp(\sqrt{2t}X - t)\right] = 0.$$

How to see this without Itô calculus? It's something about gaussians. $x\exp(-\frac{x^2}{2})$ want to come together and be integrated out. Using integration by parts,

$$\int x\exp(\sqrt{2t}x - t)\exp\left(-\frac{x^2}{2}\right)\,dx = \int \sqrt{2t}\exp(\sqrt{2t}x - t)\exp\left(-\frac{x^2}{2}\right)\,dx.$$

This is Gaussian integration by parts.

Exercise 1.15 (for credit)**:**

1. Prove Lemma 1.14 using Gaussian integration by parts and Nishimori.
2. Relate mmse with $\partial_t\overline{F}_N$.

For Proposition 1.13, letting A be a random variable, notice that

$$\mathbb{E}\left\langle (A - \mathbb{E}\left\langle A\right\rangle)^2\right\rangle = \mathbb{E}\left\langle (A - \left\langle A\right\rangle)^2\right\rangle + \mathbb{E}[(\left\langle A\right\rangle - \mathbb{E}\left\langle A\right\rangle)^2].$$

We do expectation one step at a time. The first term looks like Curie-Weiss, and produces $\frac{1}{N}\partial_h^2\overline{F}_N$. The second term is new, and produces $\partial_h F_N - \partial_h\overline{F}_N$.

The new ingredient is a classical concentration estimate. The Efron-Stein inequality shows $\mathbb{E}[|F_N - \overline{F}_N|^2]$ is small.

Phase transition in t: For small t, we have $\mathbb{E}[\overline{x}_1] = 0$, $\partial_t f = 0$; then it detaches itself from 0, $\partial_t f > 0$. Before the transition, you have maximal error; then the error reduces.

We've talked about best estimation when there are unlimited resources? People believe there is another regime where it is possible but hard to learn something about $\overline{x}$.

Convergence rate: In the Curie-Weiss, we got the rate for free. In this case, if you follow through, you can get rates. In the paper [Mou19] I was not careful. It's not straightforward to reconstruct the argument in this talk for paper. In the finite-rank case it's not clear that $\overline{F}_N$ is convex. I only know lower bound on convexity, which blows up when $h \to 0$. In the rank-1 case it's convex. From the argument here ,you can get good, N^{-a} convergence (check integration by parts). Certainly you can get $N^{-1/4}$ without a problem.

For p spin, we consider $\partial_t f - (\partial_h f)^p$ instead. Perhaps you can get $|\partial_t f - (\partial_h f)^3| \le \frac{C}{N}(\partial_h^2 f)^{1/2}$.

2 Critical and near-critical percolation (Gady Kozma)

Critical and near-critical percolation is well-understood in dimension 2 and in high dimensions. The behaviour in intermediate dimensions (in particular 3) is still largely not understood, but in recent years there was some progress in this field, with contributions by van den Berg, Cerf, Duminil-Copin, Tassion and others. We will survey this recent progress (and a few older but not sufficiently known results).

Prerequisites: The Fortuin-Kasteleyn-Ginibre (FKG) and van den Berg-Kesten (BK) inequalities. See e.g. Chapter 2 of Duminil-Copin's lecture notes `https://www.ihes.fr/~duminil/publi/2017percolation.pdf` or almost any book on percolation theory.

Slides: `https://www.math.ubc.ca/Links/OOPS/slides/Kozma_1.pdf`

2020/5/25 Lecture 1

Let's start with basic definitions. Examine the graph $\mathbb{Z}^d$ for $d \ge 2$, with edges between nearest neighbors (in L^1 norm). I'm most interested in dimension 3.

For $p \in [0,1]$, keep every edge with probability p and delete it withprobability $1-p$, independently for each edge. For background, see Grimmett's book. (The book is a little outdated; I'll discuss results not in the book.)

There is some $p_c \in (0,1)$ (critical p) such that for $p < p_c$, all components ("clusters") of the resulting graph are finite, while for $p > p_c$, there is a unique infinite cluster. Existence of p_c is a consequence of the 0-1 law.

This course is about what happens at p_c. The behaviour at and near p_c is not well understood, except if $d = 2$ or $d > 6$. We're interested in the statistics of finite clusters, which has very peculiar properties in all cases we know.

We'll focus on recent advances. The Aizenman-Kesten-Newman argument from the 80's, originally used to prove uniqueness, plays an important role. The argument is in a paper of Cerf in 2015. This is an important piece to understand what is happening at criticality.

What's charming about percolation theory, is that you have to say what is happening at p_c without knowing what the value of p_c is. From a combinatorial point of view: you have some quantity which has an exponential and polynomial correction; you want to understand the polynomial correction without knowing the exponential part. This seems impossible.

Theorem 2.1. $\mathbb{E}_{p_c}(|\mathscr{C}(0)|) = \infty$.

Notation: 0 is the point with all coordinates 0, $\mathscr{C}(x)$ is the cluster containing x. The expected size of the cluster is infinity. Even if it is always finite, it has large tails—it has no finite first moment.

Proof. Fix p and denote $\chi = \mathbb{E}_p(|\mathscr{C}(0)|)$. Let

$$\varepsilon < \frac{1}{4d\chi}$$

We will show that at $p+\varepsilon$ there is no infinite cluster. Consider $p+\varepsilon$ percolation as if we (1) take p-percolation and (2) then "sprinkle" each edge with probability ε. For a vertex x and a sequence of directed edges $e_1, \ldots, e_n$, denote by $E_{x,e1,\ldots,en}$ the event that 0 is connected to x by a path γ_1 in p-percolation from 0 to e_1 then e_1 is sprinkled, then there is a path γ_2 from e_1 to e_2 then e_2 is sprinkled and soon. We end with a path γ_{n+1} from e_n to x. We require all the γ_i to be *disjoint*. Clearly $0 \leftrightarrow x$ in $p+\varepsilon$-percolation if and only if there exist some $e_1, \ldots, e_n$ (possibly empty) such that $E_{x,e_1,\ldots,e_n}$ hold. Then $(0 \leftrightarrow x) = \bigcup_{n=0}^{\infty} \bigcup_{e_1,\ldots,e_n} E_{x,e_1,\ldots,e_n}$ and by the union bound

$$\mathbb{P}_{p+\varepsilon}(0 \leftrightarrow x) \leq \sum_{n=0}^{\infty} \sum_{e_1,\ldots,e_n} \mathbb{P}(E_{x,e_1,\ldots,e_n}).$$

This is the probability in the setting where edge has 3 states: p, ε, or closed (missing).

By the BK inequality, this is

$$\leq \sum_{n=0}^{\infty} \sum_{e_1,\ldots,e_n} \mathbb{P}_p(0 \leftrightarrow e_1^-)\mathbb{P}_p(e_1^+ \leftrightarrow e_2^-) \cdots \mathbb{P}_p(e_n + \leftrightarrow x)\varepsilon^n.$$

(The paths are disjoint witnesses.)

Summing over all x gives

$$\chi(p+\varepsilon) \leq \sum_{n=0}^{\infty} \varepsilon^n \sum_{e_1,\ldots,e_n} \mathbb{P}_p(0 \leftrightarrow e_1^-)\mathbb{P}_p(e_1^+ \leftrightarrow e_2^-) \cdots \mathbb{P}(e_n^+ \leftrightarrow x).$$

(The number of x that e_n^+ is connected to is the size of $|(0)|$.) Summing over x gives one $\chi(p)$ which we can take out of the sum

$$\sum_{n=0}^{\infty} \varepsilon^n \chi(p) \sum_{e_1,\ldots,e_n} \mathbb{P}_p(0 \leftrightarrow e_1^-)\mathbb{P}_p(e_1^+ \leftrightarrow e_2^-) \cdots \mathbb{P}(e_{n-1} + \leftrightarrow x).$$

Even though e_n^+ does not appear in the summand we are summing over it. e_n^+ has $2d$ possibilities. Summing over e_n^- gives another χ term. Taking both out of the sum gives

$$= \sum_{n=0}^{\infty} \varepsilon^n \cdot 2d\chi(p)^2 \sum_{e_1,\ldots,e_{n-1}} \mathbb{P}_p(0 \leftrightarrow e_1^-) \cdots \mathbb{P}_p(e_{n-2}^+ \leftrightarrow e_{n-1}^-)$$

$$= \sum_{n=0}^{\infty} \varepsilon^n (2d)^n \chi(p)^{n+1} < \infty.$$

This shows $p + \varepsilon_c$, contradiction.

(Set of p's for which it is finite is an open set, so at the boundary of the set χ cannot be finite. We used finiteness at the end of the proof.)

The argument also gives

$$\chi(p) \geq \frac{1}{4d(p_c - p)}$$

for all $p < p_c$. □

van den Berg-Kesten inequality: Let E be an event on $\{\pm 1\}^n$. We say that an $A \subseteq \{1, \dots, n\}$ is a witness for E (this is an event itself, denote by ω the parameter) if any ω' such that $\omega'_i = \omega_i$ for all $i \in A$ satisfies that $\omega' \in E$. For example, for the event $0 \to x$, any path between 0 and x can serve as a witness.
For two events E and F we denote by $E \circ F$ the event that they have disjoint witnesses. Then

$$\mathbb{P}(E \circ F) \leq \mathbb{P}(E)\mathbb{P}(F).$$

(van den Berg-Kesten 1985 for increasing E and F, Reimer 1997 for arbitrary E and F)

We saw we can argue at p_c using a argument by contradiction. We'll see another example of this argument.

For a set $S \subset \mathbb{Z}^d$ denote by ∂S the set of $x \in S$ with a neighbour $y \notin S$.

Theorem 2.2. *Let $S \subset \mathbb{Z}^d$ be some finite set containing 0. Then*

$$\sum_{x \in \partial S} \mathbb{P}_{p_c}(0 \overset{S}{\leftrightarrow} x) \geq 1.$$

Proof sketch. Let $x \in \mathbb{Z}^d$. If $0 \leftrightarrow x$ then there exists $0 = y_1, \dots, y_n = x$ such and open paths γ_i such that

1. γ_i is from y_i to y_{i+1} and is contained in $y_i + S$.

2. The γ_i are disjoint.

We have $n \geq r|x|$ for some number $r > 0$ that depends on S. (Here $|x|$ is L^1 norm.) A calculation similar to the previous proof shows that

$$\mathbb{P}(0 \leftrightarrow x) \leq \sum_{n \geq r|x|} \left(\sum_{y \in \partial S} \mathbb{P}_{p_c}(0 \overset{S}{\leftrightarrow} y) \right)^n.$$

If the value in the parenthesis is smaller than 1, then $\mathbb{P}(0 \leftrightarrow x)$ decays exponentially in $|x|$, contradicting the previous theorem. □

A full proof can be found in H. Duminil-Copin and V. Tassion [DT16].

Theorem 2.3 (Menshikov, Aizenman-Barsky). *For any $p < p_c$, $\chi(p) < \infty$.*

This is a deeper theorem.

We need to know at the first step that there's one phase transition. Otherwise, there could be p_1, p_2, such that below p_1 all clusters are finite and expectations are finite, between p_1 and p_2 there is all clusters are finite but expectations are infinite, and $> p_2$, there is an infinite cluster with positive probability. That would be a completely different picture!

What happens for $d > 6$? Remote paths of clusters develop independently, as if the other doesn't exist. There's a lot of space. They never get to meet.

2020/5/27 Lecture 2: The Aizenman-Kesten-Newman argument

Today we'll be getting out of the '80's (though this argument is from the 80's).

We start with the idea of using exploration. (This is the one most important takeaway from today's lecture.) We demonstrate this idea with a simple lemma. Recall $\mathscr{C}(0)$ is the cluster containing 0 (everything connected to 0 via open edges).

Lemma 2.4. *Let E be the number of open edges in $\mathcal{C}(0)$ and let B be the number of closed edges in the boundary. There are C and c such that for any n,*

$$\mathbb{P}_p(B + E \leq n, |(1-p)E - pB| > \lambda\sqrt{n}) \leq Ce^{-c\lambda^2}.$$

The important part is the $(1-p)E - pB$ part. Consider $p = \frac{1}{2}$, so we have roughly the same amount of open and closed edges.

Proof. We define sets of edges $\phi = S_0 \subset S_1 \subset \cdots$ for $i \leq n$ as follows. Assume at step i there exists some edge $e \notin S_i$ such that there is an open path in S_i from 0 to one of the vertices of e. We choose one such e arbitrarily (e.g., by lexicographic order) and define $S_{i+1} := S_i \cup \{e\}$. (Expose the cluster one edge at a time.) If no such e exists (and this happens when $|S_i| = B + E$), let $S_{i+1} = S_i$.

Let

$$X_i = (1-p)\,(\#\text{open edges in } S_i) - p\,(\#\text{closed edges in } S_i)\,.$$

The key point is X_i *is a martingale* ($\mathbb{E}[X_{i+1}|X_i] = 0$).

The lemma follows from Azuma-Hoeffding. □

Theorem 2.5 (Azuma-Hoeffding). *If X_i is a martingale and if $|X_{i+1} - X_i| \leq m$ for some numbers m_i, then for any M,*

$$\mathbb{P}(|X_n - X_0|) \leq 2\exp\left(\frac{-M^2}{2\sum_{i=1}^N m_i^2}\right).$$

This is a flexible argument. You can start from a set of vertices (not just one), or you can add additional stopping conditions. I'll show one variation.

Lemma 2.6. *Let $S \subset \Lambda$ be the set of vertices connected to the boundary. Let E be the number of open edges between vertices of S and let B be the number of closed edges with at least one vertex in S and both vertices in Λ. Let $X = (1-p)E - pB$. Then*

$$\mathbb{P}(|X| > \lambda n^{d/2}) \leq Ce^{-c\lambda^2}.$$

Start from the boundary and examine edges inside the box in arbitrarily order until I discover all vertices connected to the boundary by open paths.

Definition 2.7: Let A, B be subsets of $E \subseteq \mathbb{Z}^d$. We denote by

$$A \overset{E}{\leftrightarrows\!\!\!\leftrightarrow} B$$

the event that there are 2 disjoint clusters in E which intersect both A and B. We write $A \leftrightarrows\!\!\!\leftrightarrow$ to mean $A \overset{E}{\leftrightarrows\!\!\!\leftrightarrow} \partial E$.

This is not the same as having 2 disjoint paths. It's stronger. Not only are they disjoint, they are in disjoint clusters.

Theorem 2.8. *Let V be the number of edges (x, y) in Λ_n such that $\{x, y\} \leftrightarrows\!\!\!\leftrightarrow \partial\Lambda_n$, i.e., both x and y are connected to $\partial\Lambda_n$ but $x \overset{\Lambda_n}{\not\leftrightarrow} y$. Then $\mathbb{E}(V) < Cn^{d-\frac{1}{2}}\sqrt{\log n}$.*

Clusters can meet outside Λ_n. This proof is a simplification due to Gandolfi-Grimmett-Russo [GGR88] (a nice 4-page paper).

Proof. For $S \subseteq \Lambda_n$ define $X(S)$ to be

$$\begin{aligned} &(1-p)\,(\#\text{ open edges between 2 vertices of } S) \\ &-p\,(\#\text{ closed edges with at least 1 vertex in } S\text{, both in } \Lambda_n) \end{aligned}$$

Let $\mathcal{C}_1, \mathcal{C}_2, \ldots$ be all the clusters in Λ_n that touch the boundary. Then

$$X\left(\bigcup_i \mathcal{C}_i\right) - \sum_i X(\mathcal{C}_i) = pV.$$

This formula is the essence of the proof.

- For an open edge: It's necessarily in one cluster. It contributes $1-p$ in both terms, which cancel out.
- For a closed edge with both points in the same cluster: It contributes p to both terms.
- What's left is the event that the closed edge goes between 2 clusters: It is counted twice in the sum, so contributes $-p + 2p = p$.

The exploration argument shows that with high probability

$$\left|X\left(\bigcup \mathcal{C}_i\right)\right| < Cn^{d/2}\sqrt{\log n} \qquad\qquad |X(\mathcal{C}_i)| < C\sqrt{|\mathcal{C}_i|}\sqrt{\log n}$$

for all i.

By Cauchy-Schwarz,

$$\sum_i \sqrt{|\mathcal{C}_i|} \le \sqrt{\sum_i |\mathcal{C}_i|}\sqrt{\sum_i 1} \le \sqrt{n^d}\sqrt{n^{d-1}} = n^{d-\frac{1}{2}}.$$

(The number of clusters is bounded by the size of the boundary.) With high probability can be made to mean "with probability $> 1 - n^{-1/2}$" and we are done. $\square$

You can get an estimate for a single edge, by using the theorem for $2n$.

Corollary 2.9. *For x a neighbour of 0,*

$$\mathbb{P}(\{0,x\} \not\leftrightarrow \partial\Lambda_n) < C\sqrt{\frac{\log n}{n}}.$$

By changing from where you explore you can get all kinds of results. For example, if L is the union of all clusters reaching the left side of Λ_n and R is the union of all clusters reaching the right side of Λ_n then

$$X(L \cup R) - X(L) - X(R)$$

teaches something about edges connected to both the left and the right. Hutchcroft has a version where one explores from random points.

Theorem 2.10. *Let $S \subset \mathbb{Z}^d$ be some finite set containing 0. Then $\sum_{x\in\partial S} \mathbb{P}_{p_c}(0 \overset{S}{\longleftrightarrow} x) \geq 1$.*

Two applications:

Lemma 2.11 (K-Nachmias, 2011)**.** *For any $x \in \partial\Lambda_n$, $\Lambda_n := [-n,n]^d$,*

$$\mathbb{P}_{p_c}(0 \overset{\Lambda_n}{\longleftrightarrow} x) \geq c\exp(-C\log^2 n).$$

Lemma 2.12 (Cerf, 2015)**.** *For any $x, y \in \Lambda_n$,*

$$\mathbb{P}_{p_c}(x \overset{\Lambda_{2n}}{\longleftrightarrow} y) \geq cn^{-C}$$

We can take $C = 2d^2 - 2d$.

All constants c, C might depend on the dimension.

Proof. Assume first that x, y are on the same line (and the distance is even): $x - y = (2k, 0, \ldots, 0)$, $k \leq n$. By the theorem there exists $z \in \partial\Lambda_k$ such that

$$\mathbb{P}(0 \overset{\Lambda_k}{\longleftrightarrow} z) \geq \frac{1}{2d|\partial\Lambda_k|} \geq \frac{c}{k^{d-1}}.$$

By rotation and reflection symmetry we may assume z is in some face of Λ_k, for example $z_1 = k$. Let $\overline{z}$ be the reflection of z in the first coordinate i.e. $z = (-z_1, z_2, \ldots, z_d)$. By reflection symmetry we also have $\mathbb{P}(0 \overset{\Lambda_k}{\longleftrightarrow} \overline{z}) \geq ck^{1-d}$. Translating z to x and $\overline{z}$ to y gives

$$\mathbb{P}(x \overset{x+\Lambda_k}{\longleftrightarrow} x+z), \mathbb{P}(y \overset{y+\Lambda_k}{\longleftrightarrow} y+\overline{z}) \geq \frac{c}{k^{d-1}}.$$

But $x + z = y + \overline{z}$.

Since $x + \Lambda_k \subset \Lambda_{2n}$ and ditto for $y + \Lambda_k$, we can write

$$\mathbb{P}(x \overset{\Lambda_{2n}}{\longleftrightarrow} x+z), \mathbb{P}(y \overset{\Lambda_{2n}}{\longleftrightarrow} y+\overline{z}) \geq \frac{c}{k^{d-1}}.$$

By FKG,

$$\mathbb{P}(x \xleftrightarrow{\Lambda_{2n}} y) \geq \mathbb{P}(x \xleftrightarrow{\Lambda_{2n}} x+z), \mathbb{P}(y \xleftrightarrow{\Lambda_{2n}} y+\overline{z}) \geq \frac{c}{k^{2d-2}}.$$

With a slightly smaller c, we can remove the requirement that the distance between x and y is even. If they are not on a line, we define

$$x = x_0, \dots, x_d = y$$

such that each couple x_i, x_{i+1} differ by only one coordinate. Hence $\mathbb{P}(x_i \xleftrightarrow{\Lambda_{2n}} x_{i+1}) \geq cn^{2-2d}$. Using FKG again gives

$$\begin{aligned}\mathbb{P}(x \xleftrightarrow{\Lambda_{2n}} y) &\geq \mathbb{P}(x_0 \xleftrightarrow{\Lambda_{2n}} x_1, \cdots, x_{d-1} \xleftrightarrow{\Lambda_{2n}} x_d) \\ &\geq \prod_{i=1}^{d} \mathbb{P}(x_{i-1} \xleftrightarrow{\Lambda_{2n}} x_i) \geq \frac{c}{n^{2d^2-2d}}.\end{aligned}$$

$\square$

Theorem 2.13 (Fortuin-Kasteleyn-Ginibre, Harris)**.** *A function $f : \{\pm 1\}^n$ is called increasing if it (weakly) increases in every coordinate. For any 2 increasing functions f, g,*

$$\mathbb{E}[fg] \geq \mathbb{E}[f]\mathbb{E}[g].$$

This was recently improved to cn^{-d^2} by van den Berg and Don. Their proof has an interesting topological component (Brouwer's fixed point theorem).

Theorem 2.14 (Cerf, 2015)**.** *$\mathbb{P}_{p_c}(\Lambda_{n^c} \overset{\leftrightarrow}{\leftrightarrow} \partial\Lambda_n) \leq Cn^{-c}$ for $c > 0$ small enough. More precisely,*

$$\mathbb{P}_{p_c}(\Lambda_{n^{1/(8d^2+8d)-o(1)}} \overset{\leftrightarrow}{\leftrightarrow} \partial\Lambda_n) \leq Cn^{-1/4}$$

Prove using the lemma.

- The theorem actually holds for all p.
- Cerf had a scheme for improving the exponents: repeat the following.
  1. Get a better estimate for the number of clusters from $\partial\Lambda_{2n}$ to $\partial\Lambda_n$.
  2. Looking at $\sum \sqrt{|\mathcal{C}|}$, get a better estimate for $\mathbb{P}(\{0, x\} \overset{\leftrightarrow}{\leftrightarrow} \partial\Lambda)$.
  3. Get a better estimate for $\mathbb{P}(\Lambda_{n^c} \overset{\leftrightarrow}{\leftrightarrow} \partial\Lambda_n)$. (Use a patching argument.)

  Unfortunately, the end result was not a big improvement over $\frac{1}{2}$. (The exponent is $-\frac{2d^2+3d-3}{4d^2+5d-5}$.)

2020/5/28 Lecture 3

3 Branching random walks: some recent results and open questions (Nina Gantert)

We give an introduction to branching random walks and their continuous counterpart, branching Brownian motion. We explain some recent results on the maximum of a branching random walk and its relation to point processes, as well as a connection with fragmentations. The focus will be on open questions.

Preparatory reading: Lyons and Peres, Probability on Trees and Networks `http://pages.iu.edu/~rdlyons/prbtree/prbtree.html`, Chapters 5.1 (Galton-Watson branching processes) and 13.8 (Tree-indexed random walks)

Further reading:

- Zhan Shi, Branching Random Walks, `https://www.lpsm.paris/pageperso/zhan/brw.html`
- Julien Berestycki, Topics on Branching Brownian motion, `http://www.stats.ox.ac.uk/~berestyc/Articles/EBP18_v2.pdf`
- Ofer Zeitouni, Branching Random Walks and Gaussian Fields, `http://www.wisdom.weizmann.ac.il/~zeitouni/pdf/notesBRW.pdf`

2020/6/1 Lecture 1

There are two ingredients for a branching random walk (BRW):

1. Offspring law $p(\cdot)$ with $\sum_{k=1}^{\infty} p(k) = 1$. Assume $\sum_{k=1}^{\infty} kp(k) > 1$. Often we assume $p(0) = 0$ to simplify.
2. Displacement law: Random variable X with $\mathrm{Var}(X) > 0$.

Then

1. Start with one particle at 0, particles produce offspring according to $p(\cdot)$.
2. Offspring take displacements according to X (all particles behave independently).

How does this crowd of particles behave?

There is a continuous analogue, branching Brownian motion (BBM) (see talk on Friday). It follows Brownian motion and after an exponential lifetime, it splits into 2 particles which continue to follow Brownian motion (and splitting after exponential lifetime, and so on).

We first consider BRW. I will look at it as a tree-indexed random walk in the following sense. Note the Browian motion and branching are independent.

First build a Galton-Watson (GW) process according to $p(\cdot)$. Then label edges with iid random variables distributed as X.

Let D_n be the vertices in generation n. Then $|D_n|$ is the Galton-Watson process.

The position of a particle v is the sum of the random variables (displacements) on the edges leading to the vertex v:

$$S_v = \sum_{e \in [0,v]} X_e$$

Recall the following.

Theorem 3.1. *If $m := \sum_{k=0}^{\infty} kp(k) > 1$ then $\mathbb{P}(T \text{ infinite}) > 0$.*

Here m is the reproduction number.

If $p(0) > 0$, we can look at $\mathbb{P}^*[\cdot] = \mathbb{P}[\cdot || D_n| > 0, \forall n]$. This event has positive probability.

The collection $(S_v)_{v \in D_n}$ are random variables which are not independent.

There is a more general model: Particles produce "offspring and displacements" at once according to some point process.

For example, the point process could be:

- produce particles at position 1 and -1,
- produce 3 particles at positions 3,

each with probability $\frac{1}{2}$. There is dependence between siblings, and displacements are not independent of tree.

Many things I will say will generalize to the general model.

The first question I will address is the behavior of the rightmost particle:

$$M_n = \sup_{v \in D_n} S_v.$$

This is the maximum of variables with are not independent. Assume the exponential moment condition

$$\mathbb{E}[e^{\lambda X}] < \infty$$

for some $\lambda > 0$ and define a large deviation rate function

$$I(y) = \sup_{\lambda} [\lambda y - \log \mathbb{E}[e^{\lambda X}]]$$

It is a fact that for $y > \mathbb{E}[X]$,

$$\frac{1}{n} \log \mathbb{P}[S_n \geq ny] \to -I(y).$$

You can prove one direction quite easily:

$$\begin{aligned} \mathbb{P}(S_n > ny) &\leq \mathbb{E}[e^{\lambda S_n}] e^{-\lambda ny} \\ &= e^{-nI(y)} \end{aligned}$$

by Chebyshev's inequality and optimizing the RHS over λ. Define

$$x^* = \sup \{ s \geq \mathbb{E}[X] : I(s) \leq \log m \}.$$

We have the following old result.

Theorem 3.2 (Biggins, Hammersley, Kingman)**.**

$$\frac{M_n}{n} \to x^*, \quad \mathbb{P}^*\text{-}a.s.$$

The method of proof is useful.

Exercise 3.3: Assume the theorem.

1. Let $X \overset{d}{=} N(0,1)$. Compute x^*.
2. Suppose $p(3) = 1$, $\mathbb{P}[X = 0] = \frac{1}{2} = \mathbb{P}[X = 1]$. Compute x^*. Do we have $\mathbb{P}[M_n = n, \forall n] > 0$?
3. Same as (ii) if $p(2) = 1$.

I'll give my favorite proof of the theorem.

Intuition: At time n, we have $\approx m^n$ particles. For each $v \in D_n$, $\mathbb{P}(S_u \geq ny) \approx e^{-nI(y)}$. We have 2 competing effects: the probability decays exponentially but the number of particles increase exponentially. These effects shold balance: if

$$e^{-nI(y)} \underbrace{e^{n \log m}}_{m^n} = 1$$

then $y = x^*$.

Proof. 1. First moment method: This will give an upper bound.

$$\begin{aligned} \mathbb{P}[M_n \geq ny] &\leq \mathbb{E}\left[\sum_{v \in D_n} I_{S_v \geq ny}\right] \\ &= \underbrace{\mathbb{E}[|\Delta_n|]}_{m^n} \underbrace{\mathbb{P}[S_n \geq ny]}_{\leq e^{-nI(y)}}. \end{aligned}$$

using the independence assumption. Hence, if $I(y) > \log m$, then

$$\begin{aligned} &\sum_u \mathbb{P}[M_u \geq ny] < \infty \\ \implies &\limsup \frac{M_n}{n} \leq y. \end{aligned}$$

2. Embedded tree: Assume $y < x^*$. Choose $\varepsilon > 0$ such that $I(y) - 2\varepsilon < \log m$. Then using the lower bound

$$\mathbb{P}[S_k \geq ky] \geq e^{-k(I(y)-\varepsilon)}$$

for $k \geq k_0(\varepsilon)$. I don't have strict inequality $e^{-kI(y)}$, I have an ε.

Now I'm doing a kind of percolation:

- keep $v \in D_k$ if $\frac{S_v}{k} \geq y$
- delete v otherwise
- go at level $2k$, etc.

Continue only vertices which I kept at the first step, where the mean is large enough. Now I have an embedded GW-tree $\widetilde{T} \subset T$. If $\widetilde{T}$ is infinite, the maximum is at least y. Question: what is $\widetilde{m}$? We have

$$\begin{aligned}\widetilde{m} &= m^k \mathbb{P}[S_k \geq ky] \\ &\geq e^{k \log m} e^{-k(I(y)-\varepsilon)} \geq e^{\varepsilon k} > 1\end{aligned}$$

for k large enough. Then

$$\mathbb{P}[\liminf_{n\to\infty} \frac{M_n}{n} \geq y] > 0.$$

Now we need a 0-1-law for inherited properties. Call a property A of trees **inherited** if each finite trees has A, and if T has A, then all descendant trees of the children of the root have it. Then $\mathbb{P}^*[T \text{ has } A] \in \{0,1\}$. Thus we conclude the probability is 1. □

Proof of 0-1 law. We have

$$\begin{aligned}\mathbb{P}[T \text{ has } A] &= \mathbb{E}[\mathbb{P}[T \text{ has } A||D_1|]] \\ &\leq \mathbb{E}[\mathbb{P}[T^{(1)} \text{ has } A, \ldots, T^{(|D_1|)} \text{ has } A]||D_1|] && \text{inherited property} \\ &= \mathbb{E}[\mathbb{P}[T \text{ has } A]^{|D_1|}] && \text{independence}\end{aligned}$$

Hence

$$\gamma := \mathbb{P}[T \text{ has } A] \leq f(\mathbb{P}[T \text{ has } A])$$

where $f(s) = \sum_{k=0}^{\infty} s^k p(k) = \mathbb{E}[s^{|D_1|}]$. So $q \leq \gamma \leq f(\gamma)$. On the other hand

$$\begin{aligned}\mathbb{P}[T \text{ has } A] &\geq 0 \\ q :&= \mathbb{P}[\lim_n |D_n| = 0]\end{aligned}$$

This means that

$$\mathbb{P}[T \text{ has } A] \in \{q, 1\}.$$

(The probability is in $[q,1]$ because it is at least as big as the probability of finiteness. The function $f(s)$ is convex—its second derivative is positive—so its graph is below the line connecting the two fixed points.) But if we condition, we get

$$\mathbb{P}^*[T \text{ has } A] \in \{0,1\}$$

□

2020/6/2 Lecture 2

Remark 3.4: If $\mathbb{P}[X \geq t] = e^{-ct^r}$ where $0 < r < 1$ (stretched exponential tails) and $\mathbb{E}[X^2] < \infty$ (finite second moment) then

$$\frac{1}{n} \log \mathbb{P}[S_n \geq n^{1/r} y] \to -cy^n \text{ as } n \to \infty.$$

The same arguments give that

$$\frac{M_n}{n^{1/r}} \to x^* = \left(\frac{\log m}{c}\right)^{1/r}.$$

Note the position of the rightmost particle grows faster than linearly. x^* is such that the two effects balance: the number of particles growing exponentially, and the probability decreasing exponentially (with a power). See the talk "Stretched exponential tails," on Friday, by Piotr Dyszewski.

Remark 3.5: Let $X, X_1, X_2, \dots$ be iid, $S_n = \sum_{i=1}^n X_i$. Let $S_n, S_n^{(1)}, S_n^{(2)}, \dots$ be iid. Let $\widetilde{M_n} = \max_{1 \leq i \leq |D_n|} S_n^{(i)}$. We showed by a union bound that

$$\mathbb{P}[\widetilde{M_n} \geq ny] \leq m^n \mathbb{P}[S_n \geq ny]$$
$$\limsup \frac{\widetilde{M_n}}{n} \leq x^*.$$

We can compare M_n (from the BRW) and $\widetilde{M_n}$: The maximum of the branching random walk is smaller than the maximum of the independent random walks.

Lemma 3.6. $M_n \leq \widetilde{M_n}$. *(Here $\leq$ means stochastically dominated.)*

The result is not surprising because there are more independent random variables involved in $\widetilde{M_n}$.

Lemma 3.7. *Let $(X_i)_{i\geq 1}$, $(Y_i)_{i \geq 1}$ be independent, and $(Y_i)_{i\geq 1}$ identically distributed. Then*

$$\max_{1\leq i \leq n}(X_i + Y_1) \leq \max_{1 \leq i \leq n}(X_i + Y_i).$$

Exercise 3.8: Prove Lemma 3.7.

Proof of Lemma 3.6. Proof is by induction. Apply the induction hypothesis to the subtrees of the root. By lemma 2, we can replace the identical variables in the edges coming from the root with iid variables, and get something larger. □

Exercise 3.9:

1. Show $\frac{\widetilde{M_n}}{n} \to x^*$.
2. Suppose $\mathbb{P}[X = 1] = \alpha = 1 - \mathbb{P}[X = -1]$. Will the cloud visit 0 infinitely often? I.e.,

$$\gamma = \mathbb{P}[S_v = 0 \text{ for some } v \in D_u, \text{ for infinitely many } n].$$

Find conditions on α and m such that $\gamma = 0$ or such that $\gamma > 0$.

3. Do the same for a tree-indexed (branching) Markov chain. We have a Galton-Watson tree with a Markov chain:
   (a) particles reproduce according to $p(\cdot)$
   (b) offspring takes a step according to $q(x, \cdot)$, where $q(x, y)_{x,y \in S}$ are transition probabilities of an irreducible Markov chain, with S countable.

   Take $0 \in S$, define γ as before. Find conditions on m, $q(\cdot, \cdot)$ such that $\gamma = 0$ or $\gamma > 0$.

   This is a generalization of the previous example.

To solve the recurrence-transience questions, re-run the argument we saw yesterday. I don't claim they are necessary and sufficient in all cases, but in almost all cases. (The arguments do not determine what happens at the critical point.)

What is the second order term, i.e., $M_n - nx^* \sim$? E. Aïdékou proved the following.

Theorem 3.10. *Suppose*

- *X is non-lattice (this is important),*
- *X has finite exponential moments, $\mathbb{E}[e^{\lambda X}] < \infty$ for all $\lambda \in \mathbb{R}$.*
- *$\log m \in int\{y : I(y) < \infty\}$ (int means interior).*
- *$\mathbb{E}[|D_1|^{1+\delta}] < \infty$ for some $\delta > 0$.*

Then

$$\mathbb{P}[M_n - x^* n + \frac{3}{2\overline{\lambda}} \log n \leq t] \xrightarrow{n\to\infty} \mathbb{E}[\exp(-CZ_\infty e^{-t})]$$

where $\overline{\lambda} = I'(x^*)$ and Z_∞ is the a.s. limit of $Z_n = -\sum_{v \in D_n}(S_v - nx^*)e^{\overline{\lambda}(S_v - nx^*)}$.

This is the Gumbel distribution with a random shift.

Exercise 3.11: Check that (Z_n) is a martingale with respect to $(\mathcal{F}_n)$, $\mathcal{F}_n = \sigma$ ("up to level n"), the "derivative martingale."

The limit of the martingale is strictly positive. Showing it converges is more difficult; the result goes back to Biggins.

The theorem was first proved for Brownian motion, and took 20 years before it was shown for branching random walks.

An easier theorem is the following.

Theorem 3.12. *Assume the same assumptions, $p(m) = 1$. Then*

$$\mathbb{P}[\widetilde{M}_n - x^* n + \frac{1}{2\overline{\lambda}} \log n \leq t] \xrightarrow{n\to\infty} \exp(-Ce^{-\overline{\lambda} t}).$$

The reference model also gives a Gumbel distribution. Note the second term has a different constant. I'll give a proof sketch; it can easily be made precise in the gaussian case.

Proof sketch. Note that for $a_n = o(\sqrt{n})$, using the non-lattice assumption, by the Bahaduv-Rao Theorem,

$$\mathbb{P}[S_n > nx^* - a_n] \approx \frac{C}{\sqrt{n}} e^{-nI(x^* - \frac{a_n}{n})}.$$

(Here $\approx$ means if you divide you will have bounds by constants.) But

$$nI(x^* - \frac{a_n}{n}) = n \underbrace{I(x^*)}_{\log m} - \underbrace{I'(x^*)}_{\overline{\lambda}} a_n + o(1).$$

Plugging in and noting $\widetilde{M}_n$ is a maximum of independent S_n's gives

$$\mathbb{P}[\widetilde{M}_n \leq nx^* - a_n] \sim \left(1 - \frac{C}{m^n \sqrt{n}} e^{\overline{\lambda} a_n + o(1)}\right)^{m^n}.$$

Choose $a_n = \frac{\log n}{2\overline{\lambda}} - t$ to get this probability converges:

$$\mathbb{P}[\widetilde{M}_n \leq nx^* - a_n] \sim \exp(-Ce^{-\overline{\lambda} t} + o(1)).$$

□

Exercise 3.13: 1. Make this argument precise for $X \stackrel{d}{=} N(0,1)$.

2. Prove that $\mathbb{E}\left[\frac{\widetilde{M}_n}{n}\right] \to x^*$ and conclude

$$\limsup_{n\to\infty} \mathbb{E}\left[\frac{M_n}{n}\right] \leq x^*.$$

There are plenty of open questions for $d \geq 2$. The field starts with BBM: J Berestycki, B. Mallein, J. Schweinsberg...

2020/6/4 Lecture 3

Today there will be less proofs and more open questions.

Consider an example of a BRW with killing: $p(2) = 1$, $\mathbb{P}[X = 0] = 1-p$ and $\mathbb{P}[X = 1] = p$. By the exercise in Lecture 1, we know how the maximum will behave. I'll ask a different question now.

A ray is $v_0, v_1, v_2, v_3, \ldots$ where v_0 is the root and v_{i+1} is a child of v_i. Is there a nearly optimal ray? Define

$$\begin{aligned}\rho(\varepsilon, p) &= \mathbb{P}[\exists \text{ infinitely ray such that } S_{v_j} \geq (x^* - \varepsilon)j, \forall j \geq 1] \\ &= \mathbb{P}[\text{“nearly optimal ray”}]\end{aligned}$$

This comes up in computer science in search trees. I'm "killing" all particles below the line with slope $x^* - \varepsilon$.

Question: $\rho(\varepsilon, \rho) \sim ?$ for fixed p, as $\varepsilon \to 0$.

- For $p > \frac{1}{2}$, $x^* = 1$, with positive probability there is an optimal ray, so $\rho(\varepsilon, \rho) \not\to 0$ as $\varepsilon \to 0$.

- For $p = \frac{1}{2}$, this is much less clear. We don't have an infinite cluster if we keep edges with 1. We have $\rho(\varepsilon, p) \sim c\varepsilon$.

  This goes back to Robin Pemantle; he also formulated a conjecture for $p < \frac{1}{2}$.

  Take $\varepsilon = \frac{1}{n}$ ($n = \frac{1}{\varepsilon}$). To survive until time n, you need to have 1's up to time n. The probability is bounded by

  $$\mathbb{P}_{\frac{1}{2}}[\exists \text{ ray with only 1's up to level } n] \sim \frac{2}{n}.$$

  This is the probability a critical GW process suvives up to generation n, which is known to be $\sim \frac{2}{n}$.

  The argument gives an upper bound: If you survive infinitely you have to survive to level n. You can give a lower bound with a more involved branching process argument.

- For $p < \frac{1}{n}$,

  $$\log \rho(\varepsilon, p) \sim \frac{-c(p)}{\sqrt{\varepsilon}} \text{ for } \varepsilon \to 0.$$

  This is a result by NG, Y. Hu, and Z. Shi. The idea is similar but the implementation is more involved.

  Replace the large deviation estimate for *endpoint* of path, which has the same law as sum of iid random variable, with large deviations for the *whole path*.

  Theorem 3.14 (Mogulskii)**.** *Let $Y_1, Y_2, \ldots$ be iid with $\mathbb{E}[Y_i] = 0$, $\mathbb{E}[|Y_1|^{2+\delta}] < \infty$, $S_n = \sum_{i=1}^n Y_i$, $g_1(\cdot)$, $g_2(\cdot)$ continuous, $g_1(0) < 0 < g_2(0)$. Then*

  $$\frac{a_n^2}{n} \log \mathbb{P}[g_1\left(\frac{i}{n}\right) \leq \frac{S_i}{a_n} \leq g_2\left(\frac{i}{n}\right), 1 \leq i \leq n] \to \exists(g_1, g_2, \sigma^2)$$

  Here $a_n \to \infty$, but $\frac{a_n}{\sqrt{n}} \to 0$.

  This decays slower than exponential. This is at the heart of the survival probability.

I'll mention another model with selection, the N-BRW.

- Keep only the N individuals with the largest positions.

This was introduces by Buouet, Derrida, Mueller, and Meunier. Let $M_{n,N} = \max_{v \in D_{n,N}} S_v$. Then

$$\frac{M_{n,N}}{n} \xrightarrow{n \to \infty} x_N^* \text{ a.s.}$$

(Has the largest one has always been the largest, or do other particles pass it?) It is clear that $x_N^* \leq x^*$; the question is how does the difference behave?

Theorem 3.15 (J. Bérard, J.-B. Goueré)**.** *For $p(2) = 1$, $\mathbb{E}[e^{\lambda X}] < \infty$ for all $\lambda \in \mathbb{R}$. Then*

$$x^* - x_N^* \sim \frac{c}{(\log N)^2} \text{ for } N \to \infty.$$

J. Bérard and P. Maillard showed for heavy-tailed X that

$$\mathbb{P}[X > t] \sim \frac{1}{t^\alpha}, \quad 0 < \alpha < 2.$$

I'll give a heuristic argument of how this is related to the survival probability of BRW with killing.

The following two events are comparable:

1. BRW with killing at slope $x^* - \varepsilon$, starting with N particles, survives.
2. N-BRW moves at a speed $\geq x^* - \varepsilon$.

In (1) the probability is $1 - (1 - \rho(\varepsilon))^N$.

So in (2), $x^* - x_N^*$ should be of order $\varepsilon = \varepsilon(N)$ with ε such that $\rho(\varepsilon) \approx \frac{1}{N}$.

But since $\rho(\varepsilon) \approx e^{-\frac{c}{\sqrt{\varepsilon}}} \approx \frac{1}{N}$, then we have $\varepsilon \approx \frac{c}{(\log N)^2}$. The proof does not use this heuristic, but confirms it. The proof is relatively technical; it would be nice to have a simple proof.

I'll give another example with selection, the L-BRW.

- Keep only the individuals within (spatial) distance L to the maximum M_n, remove all the others.

Define

$$x_L^* = \lim_{n \to \infty} \frac{M_{n,L}}{n}.$$

It is clear that $x_L^* \leq x_L$.

Conjecture 3.16 (BDMM)**.** *$x_L^* \approx x^*$ with $L = \log N$.*

This is open! This theorem has been confirmed for L-BBM by M. Pain (recent).

BRW with interaction: As soon as you introduce interactions, the methods are less clear.

Q: Take a BRW with some $X \stackrel{d}{=} N(0, 1)$. Fix $R \in \mathbb{R}$.

- kill two particles if they come too close: closer to R.

Is the survival probability strictly positive? Even in 1-D the problem is completely open.

Fragmentation process. $(I_t)_{t\geq 0}$ is a collection of disjoint intervals $\subseteq (0,1)$. An interval (a,b) with $u = b-a$ splits at rate u^α into m subintervals $(a, a+\frac{u}{m}, a+\frac{2u}{m}, \ldots, b)$. Let $N_t(j)$ be the number of intervals (a,b) at time time t such that $b-a = m^{-j}$.

We can see this as a multitype branching process.

- Multitype branching process: Types in $\mathbb{N}_0$. Particles die at a rate q^j which depend on time, replaced by m offspring of type $j+1$.

  Let $N_t(j)$ be the number of particles of type j present at time t.

- The tree-indexed RW is not homogeneous anymore. On edges we have independence variables $(W_e) \stackrel{d}{=} \mathrm{Exp}(q^n)$ if e goes from level $n-1$ to level n. At time t, we have a certain number of particles which are present for some lifetime.

  Say v is alive at time t if $S_v \leq t$, $S_{v_i} > t$ for all $1 \leq i \leq m$ for all children v_i of v.

  We can study $N_t = \sum_{n\geq 0} N_t(u)$.

- Work in progress with P. Dyszewski, S. Johnston, J. Puochno, and D. Schmid.

Theorem 3.17 (Brennan, Durrett). *$\mathbb{E}[N_t] \sim t^\beta$ where $\beta = \frac{\log m}{\log \frac{1}{q}}$, $q = m^{-\alpha}$.*

4 Gibbsian line ensembles in integrable probability (Ivan Corwin)

Many important models in integrable probability (e.g. the KPZ equation, solvable directed polymers, ASEP, stochastic six vertex model) can be embedded into Gibbsian line ensembles. This hidden probabilistic structure provides new tools to control the behavior and asymptotics of these systems. In my first talk, I will discuss the Airy line ensemble and its origins and properties. In my second talk, I will discuss the KPZ line ensemble and explain how this structure is used to probe the temporal correlation structure of the KPZ equation. In my final talk, I will zoom out and discuss the origins of this hidden structure.

- Webpage: `https://www.math.ubc.ca/Links/OOPS/abs_Corwin.php`
- Slides: `http://www.math.ubc.ca/Links/OOPS/slides/Corwin_1.png`
- Problems: `http://www.math.ubc.ca/Links/OOPS/slides/Corwin_problems.pdf`

2020/6/8 Lecture 1

Message: Gibbsian line ensembles are

- interesting: come up in many probabilistic/combinatorial models and having (nontrivial) universal scaling limits.

- useful: tool in establishing regularity and characterizing limit behavior.

Talks:

1. In this talk I'll focus on non-touching geometric random walk (RW) bridges and Schur process. I'll build from first principles. I'll talk about the structure as determinantal point process.
2. Airy line ensemble, the universal object.
3. KPZ line ensemble, and use it to understand temporal correlation of KPZ SPDE, a well-known stochastic PDE

4.1 Non-touching geometric random walks

Fix $M, N \geq 1$ and $a_1, \ldots, a_M; b_1, \ldots, b_N > 0$, $a_i b_j < 1$ (this comes from a normalization).

Hold the starting and ending point constant. Do a random walk, at times $1, \ldots, M+N-1$, have jumps with parameter $a_1, \ldots, a_M, b_N, \ldots, b_1$. Choose each of the jumps with a geometric distribution with the corresponding parameter.

More precisely, recall $X \sim \mathsf{geo}(q)$ if $\mathbb{P}(X = k) = (1-q)q^k$, $k \geq 0$. Then the $\mathsf{geo}(\vec{a}; \vec{b})$ RW bridge is

$$Y(s) - Y(s+1) \sim \begin{cases} \mathsf{geo}(a_s), & s = 1, \ldots, M, \\ \mathsf{geo}(b_{N+M-s}), & s = M, \ldots, M+N-1 \end{cases}$$

We consider $\{Y_i\}_{i=1}^{\infty}$ of $\mathsf{geo}(\vec{a}; \vec{b})$ bridges conditioned on:

- $Y_i(0) = -i = Y_i(M+N)$
- Non-touching: they need to keep distance 1 apart (no touching at a corner).

There are finitely many excited curves (curves that are not flat), because the number of excited curves is $\leq \min\{M, N\}$, and once a curve is flat, all subsequent curves are flat.

Picture when M is fixed and $N \to \infty$: look at a sloped window of width $M^{2/3}$ and height $M^{1/3}$, and rescale. What happens when $M \to \infty$? As $M \to \infty$, this converges to the Airy line ensemble (which is a stationary object) minus a parabola. I'll explain this in the first 2 lectures.

1. Today we'll connect the bridge to a Schur process, and use this to relate to a determinantal point process, and get convergence for finite dimensional distributions. This proves convergence at any fixed number of times. This doesn't give functional convergence, tightness, invariance under resampling.
2. (tomorrow) Tomorrow we'll show the Gibbs property, which results in tightness/functional convergence, which gives the Airy line ensemble.

Schur process It's the connection to Schur process that allows us to do calculations. Define $\lambda_i(0) \equiv 0$, and call $\lambda_i(s) = Y_i(s) + i$. For each s, $\lambda(s) = (\lambda_i(s))_i$ forms a partition:

$$\lambda_1(s) \geq \lambda_2(s) \geq \cdots \geq 0 \text{ integers}$$

For example, $(4, 4, 2, 1, 1, 0, 0, 0, \ldots)$ is a partition.

The size of a partition λ is $|\lambda| = \sum \lambda_i$.

Two partitions λ and μ are interlacing, $\lambda \geq \mu$ if $\lambda_1 \geq \mu_1 \geq \lambda_2 \geq \mu_2 \geq \cdots$

A **skew Schur polynomial** in one variable a is

$$S_{\lambda/\mu}(a) = \mathbb{1}_{\lambda \geq \mu} \cdot a^{|\lambda| - |\mu|}.$$

Exercise 4.1: If $\{Y_i\}$ is a $\mathsf{geo}(\vec{a}; \vec{b})$ bridge, then it pushes forward to a measure $\{\lambda_i\}$ with the following form

$$\mathbb{P}_{\vec{a};\vec{b}}(\vec{\lambda}) = Z(\vec{a}; \vec{b})^{-1} \prod_{i=1}^{M} S_{\lambda(s)/\lambda(s-1)}(a_s) \prod_{s=1}^{N} S_{\lambda(M+N-s)/\lambda(M+N-s+1)}(b_s).$$

Here, $Z(\vec{a}, \vec{b})$ is normalization.

This is cumbersome but useful. We had our ensemble of curves. At each time, we form a partition by looking at the point process and doing this affine shift. The measure on this collection is given by the interlacing times a, b's raised to the differences.

Exercise 4.2: The marginal law of $\lambda(M)$, called the **Schur measure** is given by

$$\mathbb{P}_{\vec{a},\vec{b}}(\lambda(M)) = Z(\vec{a}; \vec{b})^{-1} s_{\lambda(M)}(\vec{a}) s_{\lambda(M)}(\vec{b})$$

where the **Schur polynomial** is

$$S_\lambda(\vec{a}) = \sum_{(0,0,\ldots)=\phi=\lambda(0) \leq \cdots \leq \lambda(M)} \prod_{s=1}^{M} S_{\lambda(s)/\lambda(s-1)}(a_s).$$

Show that

- it is symmetric in the a variables (This is not obvious; it is a beautiful fact about invariance of the system.)
- (bialternant formula) $s_\lambda(\vec{a}) = \frac{\det(a_i^{\lambda_j + M - j})_{i,j=1}^M}{\det(a_i^{M-j})_{i,j=1}^M}$. (These are Vandermonde determinants.)
- (partition function) $Z(\vec{a}; \vec{b}) = \prod_{i,j}(1 - a_i b_j)^{-1}$. Cauchy-Littlewood identity: $\sum_\lambda S_\lambda(\vec{a}) S_\lambda(\vec{b}) = Z(\vec{a}; \vec{b})$.

  Idea: prove something locally, 1-variable skew Cauchy-Littlewood identity. If I look at the product of 2 and sum, $\sum S_{\lambda/\kappa}(a) S_{\mu/\kappa}(b) = (1 - ab) \sum_\nu S_{\nu/\lambda}(b) S_{\nu/\mu}(a)$. I can confirm this by hand. Interate this mn times, get out a factor of $(1 - ab)$ each time. Eventually you sum everything. This is a nice trick that generalizes.

If I look at a middle slice, the probability can be written as a product of Schur polynomials, which themselves admit formulas with determinants. Next I'll describe the connection with determinantal point processes.

The Schur process is a determinantal point process. I'll use this for asymptotics.

Determinantal point processes I'll show the Schur process is a DPP, to give asymptotics.

Theorem 4.3. *For $\lambda \sim \mathsf{Schur}(\vec{a};\vec{b})$, $\widetilde{Y} = \{\lambda_i - i + \frac{1}{2}\}_{i=1}^{\infty}$, a determinantal point process on $\mathbb{Z}+\frac{1}{2}$ with correlation kernel $K(i,j)$ where*

$$\sum_{i,j\in\mathbb{Z}+\frac{1}{2}} K(i,j)v^i w^{-j} = \frac{Z(\vec{a};v)Z(\vec{b};w^{-1})}{Z(\vec{b};v^{-1})Z(\vec{a};w)} \sum_{k=\frac{1}{2},\frac{3}{2},\dots} \left(\frac{w}{v}\right)^k.$$

This means that $\forall n$, $\{x_1,\dots,x_n\}\in\mathbb{Z}+\frac{1}{2}$ distinct

$$\mathbb{P}(\{x_1,\dots,x_n\}\in\widetilde{Y}) = \det(K(x_i,x_j))_{i,j=1}^n.$$

This will allow us to access asymptotic information about the system. The kernel is not getting more complicated as the system size grows.

Exercise 4.4:

$$\mathbb{P}(\lambda_1 \geq s) = \sum_{\ell=0}^{\infty} \frac{(-1)^\ell}{\ell!} \sum_{x_1,\dots,x_\ell > s} \det(K(x_i,x_j))_{i,j=1}^{\ell}.$$

There are two things left to do:

- Show where this formula comes from.
- Show where this formula goes (asymptotics).

We show where this formula comes from. The following theorem shows a certain class of measures is determinantal.

Definition 4.5: An N-point biorthogonal ensemble is a probability measure on $\{x_1,\dots,x_N\}$ of the form

$$\mathbb{P}_N(\{x_1,\dots,x_N\}) = c_N \det(\phi_i(x_j))_{i,j=1}^N \det(\psi_i(x_j))_{i,j=1}^N$$

under the condition that the Gram matrix $G_{ij} = \sum_x \phi_i(x)\psi_j(x)$ is finite, so that there is a normalizing constant c_N.

Theorem 4.6. *$\mathbb{P}_N$ is DPP with*

$$K(x,y) = \sum_{i,j=1}^{N} \phi_i(x)[G_{ij}]^{-\top}\psi_j(y)$$

Exercise 4.7: Prove this (use the Cauchy Binet theorem).

Exercise 4.8: Apply to Schur measure (bialternate formula). Prove the $K(i,j)$ generating function formula. (Hint: use Cauchy determinant.)

Asymptotics What are we after? Let's go back to nonintersecting geometric random walks, and simplify by setting $M = N$, $a_i = b_j = q$, $N \nearrow \infty$. (Things actually simplify when we keep the parameters. The more parameters, the fewer options. The theory of Schur polynomials is natural. Then we degenerate.)

There is a limit shape evolving, with two parts. One moving into open air, one invading the space beneath.

What are some interesting questions about the picture?

- Limit shape. Consider the midpoint line. How does the density interpolate from 1 to 0? It turns out that the bottom transition happens at $\frac{-2q}{1+q}N$; the top happens at $\frac{2q}{1-q}N$; and we can derive a formula for how it transitions.

  $$\text{Density}(u) = \frac{\arg(v_+)}{\pi}$$

  where $v_{+/-}$ are complex conjugate roots of

  $$f'(v) = 0$$
  $$f(v) = \log(1 - q/v) - \log(1 - qv) - u \log v.$$

  (I'll try to prove this.)

- Local limits

  - In the bulk, it has a limit, the discrete sine process.
  - At the edge, it has a scaling limit to the Airy point process. It is a progenitor for proving the finite dimensional distribution results we're after.

- Linear statistics, large deviation principles...

I'll try to explain where this comes from.

$K(x, x) = \mathbb{P}(x \in \widetilde{Y})$ is the density.

Given a generating function, by using Cauchy's residue theorem, you can rewrite in terms of a double contour integral.

Exercise 4.9: For $M = N$, $a_i = b_j = q$ for all i, j, letting $f(v) = \log(1 - q/v) - \log(1 - qv) - u \log v$.

$$K(uN, uN) = \left(\frac{1}{2\pi i}\right)^2 \oint \oint \frac{\sqrt{vw}}{(v - w)vw} \exp\{N[f(v) - f(w)]\}\, dv\, dw$$

The contour over w is a circle surrounding q; the contour over v is outside the w-circle, and not containing q^{-1}.

Density asymptotic: solve $f'(v) = 0$. ($f(v)$ as small as possible, $f(w)$ as large as possible, then can localize contribution of kernel. If I can't, I can find the main contribution. Given at critical points. So solving this is natural.) Solve discriminant$= 0$. $u = \frac{2q}{1-q}, -\frac{2q}{1+q}$.

- $u > \frac{2q}{1-q}$: Draw $\Re[f(v)]$: There are 3 curves upon which is 0. See picture
  As long as I'm in this region, $\text{density}_N(u) \searrow 0$ exponentially fast.

- $\frac{2q}{1+q} < u < \frac{2q}{1-q}$, $v_{+/-}$ complex conjugate pairs. 2 contours where $= 0$, with 2 points of intersection, v_+, v_-. Deform v contour so that it sits in the $-$ region. w is in the positive area. As I deformed, I encounter residues. We had $\int \frac{\sqrt{vw}}{(v-w)vw} \exp(\cdots)$. Using the residue theorem, get

$$\text{density} \approx \int_{v_-}^{v_+} \frac{1}{w}\, dw - \int\int \cdots .$$

  By the same argument as before, $\int\int \to 0$ at speed $\frac{1}{\sqrt{N}}$. The first part gives $\frac{\arg(v_+)}{\pi}$.

- $u < \frac{-2q}{1+q}$. Density is $1-$(something going to 0 exponentially fast).

2020/6/9 Lecture 2

4.2 Airy line ensemble

Top live at $\frac{2q}{1-q}N$. Bottom live at $\frac{-2q}{1+q}N$. Study the edge of the ensemble, get Airy line ensemble. Take window of length $N^{2/3}$, (mesoscopic compared to the order N) and height $N^{1/3}$. Take a finite number of times, subtract off a parabola to center, showed convergence of the finite dimensional distribution to limiting DPP.

Call the limiting ensemble $\mathcal{L}_1, \mathcal{L}_2, \dots$ (starting from the top curve). We don't know there exists a continuous distribution. Prove convergence of random walk, start with finite dimensional distributions, then strengthen to functional CLT. How to strengthen from FDD to functional CLT?

Let $\mathcal{A}_i(\cdot) = \mathcal{L}_i(\cdot) + x^2$. The $\{\mathcal{A}_i\}$ are the Airy line ensemble. There are constants. The reason you have parabolic behavior because the limiting shape has nonvanishing second derivative.

Theorem 4.10.

- *The above convergence can be strengthened to hold for curves, in locally uniform on compacts topology.*

- *The Airy line ensemble $\mathcal{A}_i = \mathcal{L}_i(x) + x^2$ is stationary process in x, with DPP FDD's.*

- *$\mathcal{L}$ enjoys non-intersecting Brownian Gibbs property:*[7]

  Fix $k_1 \le k_2$, $a < b$. Then

  $Law(\mathcal{L}_{[k_1,k_2]}[a,b] | \sigma_{ext})$
  $= Law(\mathcal{L}_{[k_1,k_2]}[a,b] | \sigma_{boundary})$
  $= Law(k_2 - k_1 + 1$ *Brownian bridges conditioned on non-int. with themselves and* $\mathcal{L}_{k_1-1}[a,b], \mathcal{L}_{k_2-1}$

[7] This comes from conditioning random walks from non-intersection. The property translates into the limit.

Here, σ_{ext} has all information external to $\mathcal{L}_{[k_1,k_2]}[a,b]$, and the first inequality says the only dependence is on boundary data.

This is why it's called a Gibbs property. In lattice models, if you restrict to a region, the measure inside only depends on the boundary. There are only local interactions.

We'll sketch a "proof" of this theorem from geometric RW and give applications.

Exercise 4.11: A Brownian bridge (BB) enjoys a simple Brownian Gibbs property. The law between a, b is the Brownian bridge connecting the endpoints.

Use this to prove a.s. uniqueness of maximizer of BB.

One of the reasons the Giibs inequality is so useful is that it gives rise to a comparison inequality. There are others, like a FKG inequality that holds for this model, but I'll talk about this particular inequality

4.2.1 Stochastic monotonicity

If boundary data is coupled, then the same holds true for the law of nonintersecting Brownian bridge measure: If one set of boundary data is $\geq$ another set, then we can couple the conditional measures on Brownian bridges to also be ordered.

A similar monotonicity holds for non-touching geo RW bridges.

Poisson clock(x,k): When it rings, I flip a coin and change the height of each curve by ± 1 if it leads to non-touching. (Doing that for one curve might violate non-touching, then for that curve I don't make the move; but if another doesn't violate, then move that one.)

Exercise 4.12:

- Show this preserves order.
- Invariant measure = geometric RW bridge on touching. [8]

Exercise 4.13: Find a discrete RW (with 3 different choices) which violates stochastic monotonicity. In particular, for a 2-step bridge going from $0 \to 0$ and $0 \to 1$, there doesn't exist a coupling such that the bridge $0 \to 0$ is $\leq$ the bridge $0 \to 1$.

4.2.2 Construct the Airy line ensemble (4 steps)

Step 1: FDD convergence (Schur process). This doesn't rule out exceptional behavior at random times.

Step 2: No big max. Show on $[a, b]$ the top curve cannot get too high.

Consider a top curve between a and b. Let's say that "high" means getting to R. Define a stopping time χ, the first time between a and $\frac{a+b}{2}$ that I exceed R,

$$\chi = \inf_{x \in [0, \frac{a+b}{2}]} \left\{ x : \mathcal{L}_1^{(N)} \geq R \right\}.$$

[8]For general random walks, the condition on existence of this coupling is log-convexity.

(By continuity, it equals R at that point.) Say the surve is above $-R$. The Gibbs property holds true for random times, the Strong Gibbs property. $[\chi, b]$ is a stopping domain: knowledge of where it is is measurable with respect to the external field. The Strong Gibbs property says that on $[\chi, b]$, the law is that of the BB with $BB(\chi) = R$, $BB(b) = \mathcal{L}_1^{(N)}(b)$ and on avoiding $\mathcal{L}_2^{(N)}[\chi, b]$.

If χ exists there is a big max. We show that it's unlikely that χ exists.

If we drop $\mathcal{L}_2^{(N)}$, then the BB drops. On the event χ existing, look at what happens in between, and compare it to the linear interpolation between χ and b. The slope is bounded below.

A BB has probability $\frac{1}{2}$ of exceeding its linear interpolation at any given time. There is 50% chance it will exceed linear interpolation $R/2$ at the midpoint $\frac{a+b}{2}$. On the event χ exists, it exceed $R/2$ with probability $\geq \frac{1}{2}$. Exactly distributed as original line ensemble. By tightness, probability is going to 0. So

$$\mathbb{P}(\chi \text{ exist}) \geq 2\mathbb{P}\left(\mathcal{L}_2\left(\frac{a+b}{2}\right) \geq \frac{R}{2}\right).$$

The probability $\chi_1\left(\frac{a+b}{2}\right) \to 0$ as $R \to \infty$.

Exercise 4.14: Use this argument to show that for BB with $B(0) = B(1) = 0$,

$$\mathbb{P}\left(\max_{[0,\frac{1}{2}]} B(s) \geq R\right) \leq \mathbb{P}(B(1/2) \geq R/2).$$

We still need to establish tightness, not just that the max is not getting large. We need to argue curves are not getting too steep, because we want the distribution to be on continuous paths. Paths converge to something with absolutely continuous Radon-Nikodym derivative wrt Brownian motion.

Step 3: Good separation. If curves are not separated, Gibbs curve will sample small space, and there could be exceptional behavior.

At two points, we have good separation. This can be proved in two ways, using the Gibbs property, or DPP.

I want to argue I can take a little block with some width, and stick it between the bottom curve and point. Otherwise, the bottom curve becomes steeper, pushes itself up steeply, so we have issues (discontinuity) in limit. I need to rule this out.

Insert block at midpoint, spacer.

Shift picture over, and show it's true at a, b.

To show it's true at midpoint, use monotonicity argument. Replace by a needle with sits at the midpoint, or the curve at the interval of the block. The most the curves will drop is at most that conditioned on avoiding the interval.

This gives the following type of problem. Take a Brownian bridge, add a spike in the middle, condition to go above spike. You want to show that the distance between the spike and the BB isn't going to 0. It's not 0 as long as the height of the spike is bounded by R. There's an argument there.

Then you have the blockers.

Step 4: Extract limit Show that the Radon-Nikodym derivative

$$\frac{d\text{conditional measure}}{d\text{free RWs}} = \frac{\mathbb{1}_{\text{acceptance}}}{\mathbb{P}\left(\text{acceptance}\right)}.$$

Once I have good separation between paths, that implies tightness of the acceptance probability. The argument is that once I have blockers, I can force my paths into places that don't intersect. Force up into good region above R, and will avoid the bottom curve. So I have strictly positive probability of acceptance that will stay positive in the limit.

This implies tightness of the conditional measure.

Tightness of conditional measure and FDD implies functional convergence.

The Airy$-x^2$ has non-intersection BGP.

There's a lot of details.

4.2.3 Application

We went from discrete time regularity to regularity of the ensemble. As an application:

Theorem 4.15. *The top curve $\mathcal{L}_1(x) = \mathcal{A}_1(x) - x^2$ has a unique maximizer.*

The exercise was to show BB has the same property. Once we know this, all we need to do is the following. Take a big interval $[-R, R]$. The values there are small. Top curve a.s. has unique maximizer. With probability $\to 0$, there is no maximizer outside the interval. Show you're likely to be high $> -R^2/2$ in the middle, and likely to never go $> -R^2/2$ beyond $\pm R$. Take the union bound over intervals $[R, R+1]$, $[R+1, R+2]$,... Adapt the no big max approach. We need the probabilities to be absolutely summable, use tail bounds. $\mathcal{L}_1(0)$ is distributed according to the Tracy-Widom GUE,

$$\begin{aligned} \mathbb{P}(\mathcal{L}_1(0) \geq s) &\leq e^{-cs^{3/2}} \\ \mathbb{P}(\mathcal{L}_1(0) \leq s) &\leq e^{-cs^3}. \end{aligned}$$

We can also prove a localization result, $\mathbb{P}(|\text{maximizer}| > R) \leq e^{-cR^3}$.

Exercise 4.16: Fill in details.

2020/6/9 Lecture 3

4.3 KPZ line ensemble

The law itself is a random variable, measurable with respect to everything external.

There are other line ensembles with other Gibbs properties.

Another continuum, line ensemble that doesn't have such a strict condition of ordering of paths.

I'll talk about something that has a soft Brownian Gibbs property (BGP), the KPZ line ensemble.

Constructing infinite-volume Gibbs measures is not easy, it often takes a lot of work. You shouldn't think that if I specify any Gibbs property, there exists an infinite-volume version of it. I won't give the construction, it goes though more sophisticated methods than the Schur process.

Then I'll use the KPZ line ensemble to give temporal correlation of the KPZ equation.

4.3.1 KPZ line ensemble

Unlike the Airy line ensemble, as a stochastic process, I can't define it as a function of iid noise. This admits a description in terms of spacetime white noise.

Let $\xi(t, x)$ be Gaussian spacetime white noise.

An impressionistic definition of spacetime white noise: If I have 2 regions, the covariance of the integrals is the area of the overlap. $\sigma(t, x)$ is a generalized function with negative regularity.

I'll define something related to directed polymers. For $k = 0$, let $Z_0(t, x) \equiv \mathbb{1}$; for $k \geq 1$ consider k nonintersecting Brownian bridges $b_1, \ldots, b_k$ from $(0, 0)$ to (t, x). (A little work is needed to define that.)

$$Z_k(t, x) := p(t, x)^k \mathbb{E}_{b_1,\ldots,b_k} \left[: \exp : \left\{ \sum_{j=1}^{k} \int_0^t \xi(s, b_j)\, ds \right\} \right].$$

This is a continuum partition function. The noise is not a continuous function, so it's hard to integrate along a random path. Consider smoothing the noise: replace ξ by the mollified $\xi_\varepsilon = \xi * \delta_\varepsilon$. There is some regularization that makes this a martingale:

$$: \exp : \text{ means } \exp\left\{ \int_0^t \xi_\varepsilon(\cdot) - \delta_\varepsilon * \delta_\varepsilon(0)t/2 \right\} \text{ as } \varepsilon \searrow 0$$

I won't use this description, I'll use a fact about it which can be proved independently. $Z = Z_1$ soles the SHE

$$\partial_t Z = \frac{1}{2}\partial_{xx} Z + \xi Z$$

(HJ equation with additive white noise) This is related to directed polymers, related to Branching/dying RWs in random environments, density evolution in random environment in algae or disease. $h = \log Z_1$ solves the KPZ equation

$$\partial_t h = \frac{1}{2}\partial_{xx} h + \frac{1}{2}(\partial_x h)^2 + \xi,$$

about stochastic interface grown. For $k \geq 1$, t fixed, define KPZ^t line ensemble

$$h_k^{(t)}(x) = \log\left(\frac{Z_k(t, x)}{Z_{k-1}(t, x)} \right).$$

In particular, $h_1^{(t)} = h(t, x)$ is exactly the KPZ ("narrow wedge").

Look at the collection $\{h_k^{(t)}(\cdot)\}_{k=1}^{\infty}$. They are not necessarily ordered. They are parabolic because $h(x) + x^2$ is stationary in x. This is a cute way to see it: in the definition of Z_K, there is the heat kernel $p(t, x) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/(2t)}$. If I do an affine shift, I get back the law of

k BB's with different starting and ending points. White noise is also invariant to the shear. Stationarity is easy. The important point is that there is a hidden symmetry/invariant.

Going to the enhanced structure—looking at the bigger structure, embedding in the bigger ensemble—I have a Gibbs property.

Theorem 4.17. *For t fixed, KPZ^t line ensemble has a soft BGP: For any $k_1 \leq k_2$ and $a < b$ the law of $h^t_{[k_1,k_2]}[a,b]$ given σ_{ext} only depends on the boundary data $h^t_{k_1-1}[a,b]$, $h^t_{k_2+1}[a,b]$, and is invariant under the following resampling procedure.*

Choose $\mathcal{U} \sim \mathsf{Unif}[0,1]$ and sample $k_2 - k_1 + 1$ non-intersecting Brownian bridges (non-intersecting with themselves) $b_{[k_1,k_2]}$ from $h^t_{[k_1,k_2]}(a)$ to $h^t_{[k_1,k_2]}(b)$.

Accept the first set of bridges for which

$$\mathcal{U} \leq \exp\left\{\sum_{k=k_1}^{k_2} \int_a^b H(h^t_{k-1}(s) - b_k(s)) + H(b_k(s) - h^t_{k+1}(s))\, ds\right\}$$

where $H(x) = H_1(x)$, $H_r(x) = e^{-rx}$.

Consider the case when $r = \infty$. Then the energy is 0 if x positive, ∞ if x negative. Depending on the order, we get no energy, or infinite energy If in order, no energy, if out of order, infinite energy. If every go out of order, RHS is 0. This recovers the nonintersecting property. In the case of finite r we get penalization in exponential form, integrated over time out of order. Another way to interpret the formula is that it's a Radon-Nikodym derivative.

Conjecture 4.18. *If we scale transversally $t^{2/3}$ and vertically $t^{1/3}$ we get a line ensemble which has the $H_{t^{1/3}}$ BGP. As t $\nearrow \infty$ this should converge to the Airy line ensemble minus parabola which has the H_∞ BGP.*

We know one point convergence of top curve.

Theorem 4.19. ...

4.3.2 KPZ equation temporal correlation

It models growth interfaces. There are bumps that propagate in time and die out. How does KPZ growth decorrelate in (a long) time? The KPZ equation is a universal model, but its universality is in its long-time behavior.

Belief: correlation decays in $\varepsilon^{-3}t$ time scale, vary transversally $\varepsilon^{-2}x$, scale of fluctuations is ε^{-1}. The 2/3, 1/3 we saw earlier is here as well.

So we expect a nice scaling limit under this scaling:

Definition 4.20: Define 3:2:1 scaled KPZ equation

$$h_\varepsilon(t,x) := \varepsilon' \left[h(\varepsilon^{-3}t, \varepsilon^{-2}t) + \frac{\varepsilon^{-3}t}{24}\right]$$

We know one point GUE-TW convergence and (via H_1-BGP) spatial tightness. What about temporal behavior?

$$\mathrm{Corr}(X,Y) = \frac{\mathbb{E}[(X-\overline{X})(Y-\overline{Y})]}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}$$

Theorem 4.21. *For $\varepsilon \searrow 0$,*

$$\mathrm{Corr}(h_\varepsilon(1,0), h_\varepsilon(1+\beta)) \approx \begin{cases} \Theta(\beta^{-1/3}) & \text{as } \beta \nearrow \infty \text{ (remote)} \\ \Theta(\beta^{2/3}) & \text{as } \beta \nearrow \infty \text{ (remote)} \end{cases}$$

and any subsequential limit of $h_\varepsilon(1,x)$ is Hölder $\frac{1}{2}-$ in space and $\frac{1}{3}-$ in time.

(conjecture...)

4.3.3 Idea of proof

Three ingredients:

1. Two-time variational formula.

   Show relation between 2 times.

2. Use KPZ^t line ensemble to study.

3. One point tail bounds.

(1) Variational formula: For $s < t$ and $x, y \in \mathbb{R}$, let

$$Z(s,x;t,y) - \mathbb{E}_{b(s)=x, b(t)=y}\left[: \exp : \left\{ \int_s^t \xi(t, b(r))\, dr \right\}\right]$$

Then for every $s < t$, (Chapman-Kolmogorov property)

$$Z(t,y) = Z(0,0;t,y) = \int_{\mathbb{R}} Z(0,0;s,x) Z(s,x;t,y)\, dx$$

The two Z's are independent.

Expressing everything in terms of h we get

$$\mathrm{Law}\{h_\varepsilon(1,0), h_\varepsilon(1+\beta,0)\} = \mathrm{Law}\left\{h_\varepsilon(1,0), \sup_x{}_\varepsilon(h_\varepsilon(1,x), \widetilde{h}_\varepsilon(\beta,-x))\right\}$$

where $\widetilde{h}_\varepsilon(\beta,\cdot) \stackrel{d}{=} \beta^{1/3} h_{\beta^{1/3}\varepsilon}(1, \beta^{-2/3})$ (independent copy) and

$$\sup_x{}_\varepsilon f(x) := \varepsilon \log\left(\int_{\mathbb{R}} e^{\varepsilon^{-1} f(\varepsilon^2 x)}\, dx\right) \xrightarrow{\varepsilon \searrow 0} \sup_x \beta(x).$$

See pic. Let's imagine $\sup_\varepsilon \to \sup$ and consider Airy $\mathcal{L}$ instead on h_ε. Focus on $\beta \nearrow \infty$ case:

$$\mathrm{Corr}(\mathcal{L}(0), \mathcal{L}(x^*) + \beta^{1/3}\widetilde{(}\beta^{-2/3}x^*))$$

Everything boils down to controlling small-scale oscillations.

4.3.4 Small scale oscillations

Claim:

$$\mathbb{P}\left(\sup_{x\in[0,\eta]} |\mathcal{L}(0) - \mathcal{L}(\eta)| \geq s\eta^{1/2} ||\mathcal{L}\{-2+\eta, 0, \eta, 2\}| \leq s\right) \leq ce^{-cs^2}.$$

(Condition on 4 deterministic times. Pay a small amount to control. That I can do using tail bounds.) Extreme jumps on small intervals has Gaussian-type decay. Two parts:

Fall:

Rise:

4.3.5 Questions and directions

- Origins and other integrable Gibbsian line ensembles:

  All stochastic vertex models and their degenerations can be embedded as top curve of discrete Gibbsian line ensembles. Consequence of dynamics which preserve spin HL/q-Whitter process

- Other initial data:

  - One point distribution expressible via variational formula
  - Hammond's Patchwork Quilt
  - Dauvergne-Orthman-Virag construct Airy sheet from Airy line ensemble

  Airy line ensemble contains the entire KPZ fixed point.

- Uniqueness:

  Conjectured that Airy line ensemble is characterized by its Gibbs property, stationarity, and ergocidicity.

  Dimitrov-Matetshi prove it is characterized by its top curve and the Gibbs property.

5 Mixing and hitting times for Markov chains (Perla Sousi)

Mixing times for Markov chains is an active area of research in modern probability and it lies at the interface of mathematics, statistical physics and theoretical computer science. The mixing time of a Markov chain is defined to be the time it takes to come close to equilibrium. There is a variety of techniques used to estimate mixing times, coming from probability, representation theory and spectral theory. In this mini course I will focus on probabilistic techniques and in particular, I will present some recent results (see references below) on connections between mixing times and hitting times of large sets.

Website: `https://www.math.ubc.ca/Links/OOPS/abs_Sousi.php`

2020/6/15 Lecture 1

I'll cover results from 3 papers.

1. Equivalence (up to constants) between mixing times and hitting times of large sets
2. Hitting times: comparison for different sizes of sets
3. Refined mixing and hitting equivalence

Let X be an irreducible Markov chain in a finite state space S. (You can go from any state to any other state in a finite number of steps with positive probability.) Let P be the transition matrix of X. Let $P^t(i,j) = \mathbb{P}_i(X_t = j)$ for all $i, j \in S$ (starting at i, get to j in t steps.

There exists an invariant distribution π, $\pi = \pi P$. If X is also aperiodic, then $P^t(x,y) \to \pi(y)$ as $t \to \infty$, forall x, y.

We use the total variation distance. Let μ and ν be 2 probability distributions on S. Let

$$\|\mu - \nu\|_{TV} = \max_{A \subset S} |\mu(A) - \nu(A)|.$$

Let

$$d(t) = \max_x \left\| P^t(x, \cdot) - \pi \right\|_{TV}.$$

(Take over the worst starting state.). For all $\varepsilon \in (0,1)$,

$$t_{\text{mix}} = \min\{t \geq 0 : d(t) \leq \varepsilon\}.$$

Define $t_{\text{mix}}\left(\frac{1}{4}\right)$.

X is called **reversible** if from the stationary distribution, running the Markov chain forwards or backwards in time is indistinguishable: for all x, y,

$$\pi(x)P(x,y) = \pi(y)P(y,x).$$

I'll mostly talk about the reversible case.

We always consider the lazy verion of the chain. The lazy version of X: either stay with probability $\frac{1}{2}$, or choose a state with respect to the transition matrix. So $P_L = \frac{P+I}{2}$.

Theorem 5.1 (Oliveira, Peres-S, 2012)**.** *For all $\alpha < \frac{1}{2}$, there exist positive constants c_α and c'_α such that for all reversible lazy Markov chains,*

$$c_\alpha t_H(\alpha) \leq t_{mix} \leq c'_\alpha t_H(\alpha).$$

Here $t_H(\alpha)$ is the maximum hitting time of sets of size at least α. I will also write this as $t_{\max} \asymp_\alpha t_H(\alpha)$. The lower inequality is the easy version: if we don't hit a big set, we cannot be mixed. The upper inequality is nontrivial, and where all of the work goes.

Proof of Theorem 5.1, lower bound. We show $t_{\text{mix}} \geq c_\alpha t_H(\alpha)$. We take $\alpha = \frac{1}{8}$, to make calculations easier.

We show $t := t_{\text{mix}}\left(\frac{1}{1}\right)6 \leq 3t_{\text{mix}}$. Use submultiplicative (see notes on webpage), lose a factor of 2.

For all x, A, $P^t(x, A) \geq \pi(A) - \frac{1}{16}$. Take A with $\pi(A) \geq \frac{1}{8}$, then $P^t(x, A) \geq \frac{1}{16}$ for all x.

So $\tau_A \leq t \cdot \text{Geo}\left(\frac{1}{16}\right)$: if we haven't hit A after time t, we have $\geq \frac{1}{16}$ chance of hitting it in the next t steps, and so on. Hence $\max_x \mathbb{E}_x[\tau_A] \leq 16t$. $\square$

(The lower bound doesn't need reversibility.) Related result: Aldous in 1982 showed that for all reversible lazy MC's,

$$t_{\text{mix}} \asymp \max_{x,A} \pi(A)\mathbb{E}_x[\tau_A]$$

(He showed it for continuous MC, but this is similar to a lazy MC.) Where is this maximum actually attained? The theorem says it is attained at large sets.

Remark 5.2: Reversibility is essential!

Exercise 5.3: Consider a biased RW on $\mathbb{Z}_n$ (with laziness). With probability $\frac{1}{3}$ move to $i-1$, with probability $\frac{2}{3}$ move to $i+1$, and take $P_L = \frac{P+I}{2}$.

Show $t_{\text{mix}} \asymp n^2$ and for all α, $t_H(\alpha) \asymp n$.

This is a counterexample when the chain is not reversible.

Remark 5.4: If $\alpha > \frac{1}{2}$, then the theorem is false. The constants explose as the constant $\nearrow \frac{1}{2}$.

Exercise 5.5: Consider K_n and K_n joined by an edge. Show that $t_{\text{mix}} \asymp n^2$ and $t_H(\alpha) \asymp n$ if $\alpha > \frac{1}{2}$.

Definition 5.6 (Mixing at a geometric time)**:** Let Z_t be a geometric r.v. of parameter $\frac{1}{t}$ taking values in $\{1, \ldots\}$ and independent of X. Define $d_G(t) = \max_x \|P_x(X_{Z_t} = \cdot) - \pi\|_{TV}$ and $t_G = \min\left\{t \geq 0 : d_G(t) \leq \frac{1}{4}\right\}$ as the **geometric mixing time**.

Remark 5.7: If instead of geometric, we take U_t to be uniform on $\{1, \ldots, t\}$, then this gives rise to **Cesaro mixing time**.

Exercise 5.8: Show that $d_G(t)$ is decreasing in t.

This is not true of the Cesaro mixing time.

Theorem 5.9. *For all (lazy) reversible chains, $t_G \asymp t_{mix}$.*

The ideas come from Aldous, and Lovász and Winkler.

Theorem 5.10. *For all chains (not necessarily lazy), $t_G \asymp_\alpha t_H(\alpha)$ for all $\alpha < \frac{1}{2}$.*

Proof of Theorem 5.1, upper bound. Immediate from Theorems 5.9 amd 5.10. $\square$

Proof of Theorem 5.10. $t_G \gtrsim_\alpha t_H(\alpha)$ is easy.

We prove $t_G \lesssim_\alpha t_H(\alpha)$. Take $\alpha = \frac{1}{8}$.

Let $t < t_G$. We want to find a set B with $\pi(B) \geq \frac{1}{8}$ such that $\max_x \mathbb{E}_x[\tau_B] \geq \theta t$ for some positive constant θ.

$t < t_G$ means that there exist z, A such that $\mathbb{P}_z(X_{Z_t} \in A) < \pi(A) - \frac{1}{4}$. We automatically get that $\pi(A) > \frac{1}{4}$. This is not the set B, but will be used to define B.

Define $B = \left\{y : \mathbb{P}_y(X_{Z_t} \in A) \geq \pi(A) - \frac{1}{8}\right\}$.

Claim. $\pi(B) > \frac{1}{8}$

$\pi = \pi P$ implies

$$\begin{aligned}
\pi(A) &= \underbrace{\sum_{y \in B} \pi(y) \underbrace{\mathbb{P}_y(X_{Z_t} \in A)}_{\leq 1}}_{\leq \pi(B)} + \underbrace{\sum_{y \in B^c} \pi(y)}_{\leq 1} \underbrace{\mathbb{P}_y(X_{Z_t} \in A)}_{\leq \pi(A) - \frac{1}{8}} \\
\pi(A) &\leq \pi(B) + \pi(A) - \frac{1}{8} \\
\implies \pi(B) &\geq \frac{1}{8},
\end{aligned}$$

proving the claim.

We will prove that assuming $\mathbb{E}_z[\tau_B] \leq \theta t$ for a suitable constant θ leads to a contradiction. B is the set of points, such that starting from there, we have substantial probability of hitting A. z is bad starting state, the fact that we hit quickly from z will contradict.

By Markov's inequality,

$$\mathbb{P}_z(\tau_B \geq 2\theta M t) \leq \frac{1}{2M}, \tag{9}$$

for $M \in \mathbb{N}$.

$$\mathbb{P}_z(X_{Z_t} \in A) \geq \underbrace{\mathbb{P}_z(X_{Z_t} \in A | Z_t \geq \tau_B, \tau_B < 2\theta M t)}_{\geq \min_{y \in B}(X_{Z_t} \in A)} \mathbb{P}_z(Z_t \geq \tau_B, \tau_B < 2\theta M t)$$

by the memoryless property of Z_t and strong Markov at τ_B.

$$\begin{aligned}
&\geq \left(\pi(A) - \frac{1}{8}\right) \underbrace{\mathbb{P}_z(Z_t > 2\theta M t, \tau_B < 2\theta M t)}_{\mathbb{P}_z(Z_t > 2\theta M t)\mathbb{P}_z(\tau_B < 2\theta M t)} \\
&\geq \left(\pi(A) - \frac{1}{8}\right) \left(1 - \frac{1}{t}\right)^{2\theta M t} \left(1 - \frac{1}{2M}\right) && \text{by (9)} \\
&\geq \left(\pi(A) - \frac{1}{8}\right) (1 - 2\theta M) \left(1 - \frac{1}{2M}\right) && 2\theta M t > 1.
\end{aligned}$$

Choosing $\theta = \frac{1}{4M^2}$ gives $\mathbb{P}_z(X_{Z_t} \in A) \geq \left(\pi(A) - \frac{1}{8}\right)\left(1 - \frac{1}{2M}\right)^2$. Taking M large enough shows $\mathbb{P}_z(X_{Z_t} \in A) > \pi(A) - \frac{1}{4}$ which is a contradiction (to the fact that z is a bad starting point for A. $\square$

The idea of geometric mixing is due to Oded Schramm.

Idea of Theorem 5.9: Let

$$t_{\text{stop}} = \max_x \min \{\mathbb{E}_x[\Lambda_x] : \Lambda_x \text{ is a randomised stopping time s.t. } \mathbb{P}_x(X_{\Lambda_x} = \cdot) = \pi(\cdot)\}.$$

Take the worst starting state x, find a randomised stopping time which achieves stationarity. Find the stopping time with minimum expectation. It is not at all obvious that it exists. It's clear that there is one achieving stationarity, it's not clear that the minimum is attained.

The filling rule: The construction goes back to Baxter and Chacon, '76, used by Aldous, Lovász-Winkler.

The easy direction is $t_{\text{stop}} \leq 8t_{\text{mix}}$. The hard direction is to show that $t_{\text{stop}} \gtrsim t_{\text{mix}}$.

Exercise 5.11: Prove that for reversible chains, $t_{\text{stop}} \leq 8t_{\text{mix}}$.

Hint: use separation distance to define an appropriate stopping time.

References

[BM19] Jean Barbier and Nicolas Macris. "The adaptive interpolation method: a simple scheme to prove replica formulas in Bayesian inference". In: *Probability Theory and Related Fields* 174.3-4 (2019), pp. 1133–1185. URL: https://arxiv.org/abs/1705.02780.

[DT16] Hugo Duminil-Copin and Vincent Tassion. "A new proof of the sharpness of the phase transition for Bernoulli percolation and the Ising model". In: *Communications in Mathematical Physics* 343.2 (2016), pp. 725–745.

[Eva10] Lawrence C Evans. *Partial differential equations*. Vol. 19. American Mathematical Soc., 2010.

[FV17] Sacha Friedli and Yvan Velenik. *Statistical mechanics of lattice systems: a concrete mathematical introduction*. Cambridge University Press, 2017. URL: https://www.unige.ch/math/folks/velenik/smbook/index.html.

[GGR88] Alberto Gandolfi, Geoffrey Grimmett, and Lucio Russo. "On the uniqueness of the infinite cluster in the percolation model". In: *Communications in mathematical physics* 114.4 (1988), pp. 549–552. URL: https://projecteuclid.org/euclid.cmp/1104160778.

[LM19] Marc Lelarge and Léo Miolane. "Fundamental limits of symmetric low-rank matrix estimation". In: *Probability Theory and Related Fields* 173.3-4 (2019), pp. 859–929. URL: https://arxiv.org/abs/1611.03888.

[Mou18] Jean-Christophe Mourrat. "Hamilton-Jacobi equations for mean-field disordered systems". In: *arXiv preprint arXiv:1811.01432* (2018). URL: https://cims.nyu.edu/~jcm777/HJinfer.pdf.

[Mou19] Jean-Christophe Mourrat. "Hamilton-Jacobi equations for finite-rank matrix inference". In: *arXiv preprint arXiv:1904.05294* (2019). URL: https://cims.nyu.edu/~jcm777/HJrank.pdf.