MATH 232: Topics in Probability

Lecturer: Professor Amol Aggarwal

Notes by: Andrew Lin

Spring 2026

1 March 31, 2026

The topic of this quarter will be **probabilistic analysis through determinantal point processes and symmetric functions**. The reason for giving this course is that in probability, one often tries to reduce problems to questions about random walks (which are easy to solve via Fourier or exact methods). So it would be great if there were a family of models that we could also directly solve with bare hands because we know everything about them, and we'll try to explain a repertoire of techniques here of that type. Things will be rather algebraic (hence the determinantal point processes and symmetric functions).

Example 1

As an illustrative example, consider **last passage percolation (LPP)**. In this model, we have an $N \times N$ square (think of it as the first quadrant) where each cell is filled with some random variable $w_{ij} > 0$. We're interested in the **passage time** from a given cell to some other cell. The idea is that these random variables tell us the reward (or penalty in some other cases) for passing through each cell, so for each path $P = (v_1, \cdots, v_k)$ of adjacent cells we assign to it a weight

$$w(P) = \sum_{v \in P} w_v.$$

For any point (m, n) we can thus consider the **last passage time**

$$L(M, N) = \max_{\substack{(0,0) \overset{P}{\to} (M,N) \\ \text{directed up-right paths}}} w(P),$$

and we then have to consider how $L(M, N)$ behaves as $M, N \to \infty$.

This sits in the same general world of models as **first passage percolation (FPP)**, in which we have the exact same model but do not impose the directedness assumption and instead consider the **first passage time**

$$T(M, N) = \min_{\substack{(0,0) \overset{P}{\to} (M,N) \\ \text{arbitrary paths}}} w(P).$$

(So an optimal path which minimizes the total weight might wind around a little bit.) These questions indeed look very similar in nature, and they both tell us something about reward or cost of traversing a landscape. But we know **very little** about first passage percolation and **quite a lot** about last passage percolation, and that's because the latter turns out to be much more exactly solvable. So then we can try to conjecture things about FPP using LPP, such as the following:

Theorem 2 (Johansson, 1998)

Suppose the weights w_{ij} in the cells are iid geometric random variables, meaning that $\mathbb{P}(w_{ij} = k) = q^k(1-q)$ for some $q \in (0,1)$. (This, along with some siblings like exponential random variables, is basically the only instance of weights where we can solve the problem.) Then there exist constants $c_1, c_2 > 0$ (depending on q) such that

$$\mathbb{P}\left(\frac{L(N,N) - c_1 N}{c_2 N^{1/3}}\right) \xrightarrow{N\to\infty} \text{Tracy-Widom distribution.}$$

(The same also holds if M, N are different but we'll keep the notation simple.) So in other words, $L(N,N)$ has a law of large numbers behavior of $c_1 N$, along with fluctuations of order $N^{1/3}$.

Before we say more about what this Tracy-Widom distribution, we'll first give a general idea of the algebraic structure underlying the model. The "solvability" of this particular relies on the theory of **Schur functions** (also called **Schur polynomials**), which are indexed by **partitions** $\lambda = (\lambda_1, \cdots, \lambda_\ell)$ with $\lambda_1 \geq \cdots \geq \lambda_\ell > 0$ (ordered sequences of positive integers; we call ℓ the **length** of the partition). They are given by ratios of two determinants:

$$s_\lambda(x_1, \cdots, x_\ell) = \frac{\det\left[x_i^{\lambda_j + \ell - j}\right]_{1\leq i,j\leq \ell}}{\det\left[x_i^{\ell-j}\right]_{1\leq i,j\leq \ell}}.$$

We won't go further into this, but these turn out to be characters of irreducible representations of the unitary group $U(\ell)$. Such functions satisfy the **Cauchy identity**

$$\sum_\lambda s_\lambda(x_1, \cdots, x_\ell) s_\lambda(y_1, \cdots, y_\ell) = \prod_{1\leq i,j\leq \ell} \frac{1}{1 - x_i y_j} \quad \text{if } |x_i y_j| < 1 \text{ for all } i, j.$$

(note that the left-hand side is an infinite sum over all partitions). We will call this quantity $Z(\vec{x}, \vec{y})$, and the basic idea is that we can embed the last passage time in terms of the **largest part of a random partition**. There's something called a **Schur measure**, introduced by Okounkov in 2000, which is a probability measure on partitions which exactly assigns the corresponding summand to each partition:

$$\mathbb{P}[\lambda] = \mathbb{P}_{\mathrm{SM}(\vec{x},\vec{y})}[\lambda] = \frac{s_\lambda(\vec{x}) s_\lambda(\vec{y})}{Z(\vec{x}, \vec{y})}$$

(if the xs and ys are positive, this does turn out to always be positive and thus actually give us a measure). So to relate this to LPP, it really comes in two parts:

Theorem 3

Set all coordinates of the length-N vectors $\vec{x}$, $\vec{y}$ to be $\sqrt{q}$ (all that matters is that $x_i y_j = q$ for all i, j).

(a) Sample λ under the Schur measure $\mathrm{SM}(\vec{x}, \vec{y})$. Then

$$L(N,N) \stackrel{d}{=} \lambda_1,$$

where λ_1 is the largest part of the partition. (This is really the Robinson-Schensted-Knuth correspondence.)

(b) We have

$$\frac{\lambda_1 - c_1 N}{c_2 N^{1/3}} \xrightarrow{N\to\infty} \text{Tracy-Widom distribution.}$$

Example 4

We're now at a good point to discuss what the Tracy-Widom distribution really is. Consider an $N \times N$ matrix with random complex entries specified to be Hermitian; specifically, we have $H = (H_{ij})_{1 \le i,j \le N}$, where $H_{ij} = \overline{H_{ji}}$, off-diagonal upper-triangular entries are iid $\mathcal{N} + i\mathcal{N}'$ for $\mathcal{N}, \mathcal{N}'$ standard normal random variables, and the diagonal entries are just standard normal (scaled by a factor of $\sqrt{2}$, but it doesn't actually have an impact on the answer). A simpler description here is to let $X = (X_{ij})$ have iid complex normal entries (that is, $\mathcal{N} + i\mathcal{N}'$) and let $H \sim X + X^\dagger$ be its Hermitianization.

This is the simplest model we can have of a random Hermitian matrix, and we're generally interested in the spectrum (eigenvalues). Heuristically, if we want to think about how big the eigenvalues can get, they should roughly scale as the row sums of H (if v is the all-ones vector or something else kind of "delocalized or flat"). But a row sum of H is a sum of N iid normal random variables, hence scaling like $N^{1/2}$. The full spectrum turns out to look as the following:

Theorem 5 (Wigner, 1958)

Recall that the eigenvalues of H will all be real. We have the **semicircle law**

$$\frac{1}{N} \# \left\{ \text{eigenvalues of } H \text{ in the interval } N^{1/2}[a,b] \right\} \xrightarrow{N \to \infty} \int_a^b \rho_{\text{sc}}(x)\,dx,$$

where $\rho_{\text{sc}} = \frac{1}{2\pi}\sqrt{4 - x^2}$ is the density of a semicircle.

So we might guess (though it's not a consequence of this result) that the top eigenvalue λ_1 is about $2\sqrt{N}$. (Wigner's result only really tells us about positive proportions of eigenvalues.) But it is true, and we can describe the fluctuations:

Theorem 6 (Tracy–Widom, 1993; Gaudin, 1960s)

We have $N^{-1/6}$-order fluctuations of the top eigenvalue:

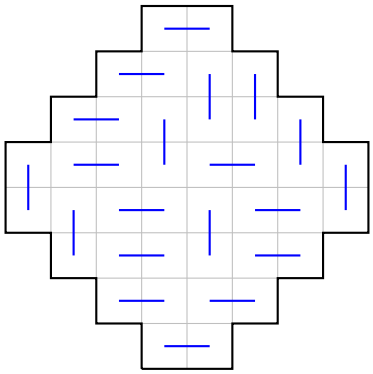
$$(\lambda_1 - 2\sqrt{N})N^{1/6} \xrightarrow{N \to \infty} \text{Tracy-Widom distribution.}$$

Remember that in the LPP case, our law-of-large-numbers gave us order N. So if we wanted to scale the random matrix case up, we'd have to multiply by $N^{1/2}$, and so the $N^{-1/6}$-order fluctuations scale up to $N^{1/3}$ (which is what we expect).

Remark 7. *This matrix model H above is called the* **Gaussian unitary ensemble (GUE)**, *and it has the property that $H \sim UHU^{-1}$ for any fixed unitary matrix U (just by properties of Gaussians). And the solvability of the GUE is closely related to the solvability of LPP (and of Schur processes), though we might notice that these results predate the introduction of Schur processes. Indeed, Schur processes were introduced as an abstraction which ended up being able to solve a wide class of other models, so this is the more modern approach to the technique of determinantal point processes.*

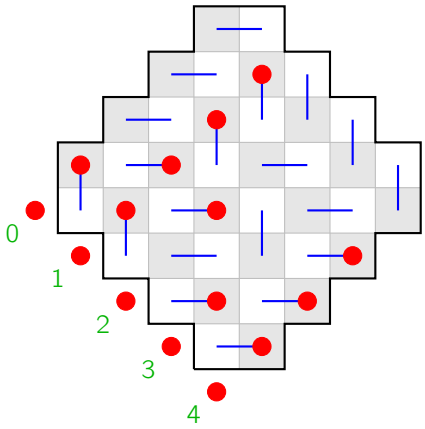
Example 8

For a further appearance of Schur polynomials, we can think about **random tilings**. Consider the Aztec diamond A_N shown below for $N = 4$ (consisting of four quadrants each of which is an $N \times N$ triangle). We are interested in the random tiling of A_N by dominos (either a pair of squares vertically or horizontally).



There are $2^{N(N+1)/2}$ possible tilings of A_N (due to Elkies, Kuperberg, Larsen, and Propp in 1991), and so that's a good sign that we can actually study it more explicitly. When we pick one uniformly at random and ask "how the tiling behaves," there's some interesting behavior as N tends to infinity. The idea is that outside of the inscribed circle for the limiting square, the dominos end up being fixed with high probability (on the left and right they are vertical in a shingle pattern, and on the top and bottom they are horizontal instead). But inside the circle things are much more rough and random, so there is a phase transition between "solid" and "liquid" regions. To relate things to the $N^{1/3}$ number (which isn't the focus of the course, but it's worth pointing out), there is really a random interface after which we see the brick pattern, and that also fluctuates at order $N^{1/3}$ with Tracy-Widom.

We won't talk about asymptotics here, but to see how the Schur functions pop up, we can shade the Aztec diamond in a checkerboard fashion. So there are really four kinds of dominos, vertical with shaded above or below, and horizontal with shaded left or right. Place a dotted circle on the shaded cell every time we see a **vertical shaded above or a horizontal shaded right**, and also shade in the cells in the bottom-left side.



The point is that once we know the circles, we can fill in the tilings, and now we can think about these circles as particles living on sites from 0 to N at $(N+1)$ different diagonals. On the ℓth diagonal from the top, we can denote the locations of the particles $(p_1, \cdots, p_\ell)$ in decreasing order (it does turn out we will always have i particles); these are strictly ordered instead of weakly ordered, we can turn these into partitions by defining $\lambda_j = p_j - \ell + j$. We claim that for each $0 \le \ell \le N$, the resulting partition $\lambda^{(\ell)}$ for that diagonal is distributed according to (a close relative of) the Schur measure. Without worrying about the power-of-2 factor in front, we really have

$$\mathbb{P}[\lambda^{(\ell)}] = 2^{(\cdots)} s_\lambda(1, 1, \cdots, 1) s_{\lambda'}(1, 1, \cdots, 1),$$

where λ' is the transpose of the partition λ (meaning that we draw partitions as Young diagrams with left-justified rows of $\lambda_1, \lambda_2, \cdots$ boxes, and then we "transpose" that diagram to just consider the columns instead – for example, we

can check that $(5,4,4,2,1)$ becomes $(5,4,3,3,1)$). So if we just look at the tiling problem, picking a tiling uniformly at random and defining the associated particles, then the slices are distributed exactly up to some normalization factor as a product of Schur polynomials.

Remark 9. *To see why there are exactly i particles on the ith diagonal, notice that along the very top-right one, we have to go from having rightward pointing dominos to upward pointing ones at some point, and then the "missing hole" will be exactly where the first particle is. Then we can split up the next diagonal into the parts before and after that particle and apply the same logic; there will need to be a particle in each of those parts. So this actually also means that the locations of the particles must "interlace."*

Fact 10

To connect things to the six-vertex model, for each tiling of A_N we can produce an **alternating sign matrix** of size $(N+1)\times(N+1)$ as follows. First, write down a matrix where the ith row corresponds to the ith diagonal from bottom to top, read from right to left, with a $+1$ if we have a particle and 0 otherwise. In this case our matrix would be $\begin{bmatrix}1&1&1&1&1\\1&1&0&1&1\\1&0&1&1&0\\1&0&0&1&0\\0&0&0&1&0\end{bmatrix}$. Now subtracting the $(i+1)$th row from the ith row yields $\begin{bmatrix}0&0&1&0&0\\0&1&-1&0&1\\0&0&1&0&0\\1&0&0&0&0\\0&0&0&1&0\end{bmatrix}$. (It's worth keeping in mind though that this last move is **not** a bijection – any ASM has a specific number of tilings, which is $2^{\#(-1)\text{s appearing}}$.) This is an alternating sign matrix, meaning that the sum of the entries in each row and column is 1, and the nonzero entries also alternate between 1 and -1 in each row and column.

So if we take a random alternating sign matrix, weighted in terms of the number of -1s it has, then that would correspond to a random uniform tiling of the Aztec diamond. Thus we might ask about a similar question if we don't do that weighting and just look at the alternating sign matrices. We know that $\sum_{\text{ASMs } A} 2^{\#(-1)\text{s appearing in } A} = 2^{N(N+1)/2}$, and there is a nice numeration for ASMs themselves too:

Theorem 11 (ASM conjecture)

The number of $(N+1)\times(N+1)$ alternating sign matrices is given by

$$\prod_{i=0}^{N}\frac{(3i+1)!}{(N+i+1)!}.$$

This conjecture was proposed by Mills-Robbins-Ramsey in 1982, proved by Zeilberger in 1992 and Kuperberg 1995, and actually the two proofs created two different fields of math (automated proofs for certain combinatorial identities, and quantum integrability for algebraic combinatorics).

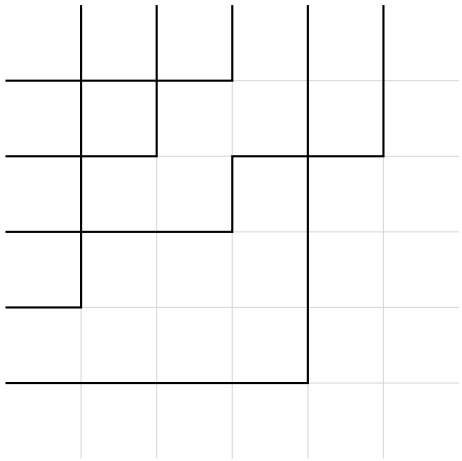
2 April 2, 2026

Last time, we gave a list of models with some kind of algebraic structure. We'll pick up there and connect to some other related models before getting into techniques.

We introduced LPP and domino tilings last time, seeing that both were related to Schur measures. It was also mentioned that domino tilings are in correspondence with alternating sign matrices, and so let's mention a few more

aspects of alternating sign matrices and their deeper connections to statistical mechanics. Recall that alternating sign matrices have entries in $\{-1, 0, 1\}$, so that each row and column has sum 1 and the nonzero entries alternate $1, -1, 1, -1, \cdots, 1$.

Consider replacing each entry of an alternating sign matrix M by a vertex, and then we put up-right directed paths so that a 1 corresponds to going right then up, -1 corresponds to going up then right, and 0 turns out to correspond to going straight. The result is as shown for the example $\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$ from last time (by first setting the turning points and then forcing everything else to go straight up or right:



We can immediately see a few properties:

- Paths enter through the left side and exit through the top side (since the row and column sums tell us we always have one more up-turning-point than right-turning-point).
- If we think of the paths as "bouncing off of each other" instead of passing through each other, then paths can touch but not share edges.

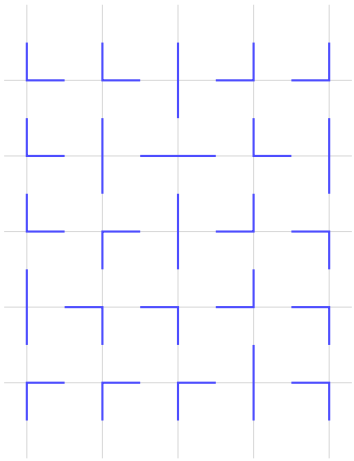
This is sometimes known as the **six-vertex model** (under domain-wall boundary conditions), since there are six possibilities for what each vertex can look like (one corner vertex, the other one, and then four ways of having straight paths through a vertex depending on whether the right and up ones exist – however remember that if both straight paths exist we actually think of them as bouncing off each other). We can check that there is actually a bijection between ASMs and six-vertex configurations (we can just read off the $-1, 0, 1$s from the local vertex configurations).

This is a very rich model but actually not a lot is known about it, and let's go through a bit of the history by thinking of this model as **ice**. If we think of the vertices in our six-vertex model as **oxygen atoms**, we can imagine having a hydrogen atom at the middle of each edge of our configuration. Water consists of oxygen atoms together with two hydrogen atoms, so there are $\binom{4}{2} = 6$ total ways each oxygen atom can connect with two adjacent edges.

We can now consider all ways that the oxygens can each bond with exactly two hydrogen atoms to form water molecules. The correspondence is as follows:

Hydrogens	Paths
left, down	empty vertex
up, right	full vertex
left, up	up-down straight path
down, right	left-right straight path
left, right	up-right corner
up, down	right-up corner

Basically, we copy the bond from the top and complement the bond on the bottom, and similarly we copy the bond on the right and complement it on the left. So notice that for the boundary conditions to match the domain-wall boundary conditions, the oxygens should be allowed to bond with the boundary hydrogens on the top and bottom, but not the left and right. So then we'll have exactly twice as many hydrogens as oxygens and everything will get perfectly matched. The blue lines indicate the bonds that we end up drawing (with hydrogens at the midpoints of the edges):



Notice that on the bottom row, we have to go from down-right bonds to vertical at some point, and then swap to down-left bonds, to match all of the hydrogens on the bottom, and then we can keep building things upward from there. So we get the same "interlacing" phenomenon as last time here.

This model was introduced by Pauling to study the **residual entropy** of square ice. The idea is that if we have a bunch of oxygens and hydrogens in a tank, then even if we cool to absolute zero we can still have entropy (lots of configurations in which the molecules combine). So the idea is that if we put things in this square grid (we can also map to a three-dimensional domain), then at absolute zero we must bond with the closest hydrogens to minimize energy. So the total configuration count is exactly the number of models from above, and it would be useful to be able to do enumeration. It turns out that things scale exponentially as c^{N^2}, and experimentally it was found that $c \approx 1.52$ for actual ice. Pauling did a heuristic back-of-the-envelope calculation to see how c would look for our vertex model:

- There are 6^{N^2} possibilities a priori, since each of our N^2 vertices can be bonded in six ways.
- But only two of the four ways of connecting each hydrogen can actually work, so we're "halving the number of possibilities" per edge.
- Well, there are $2N^2$ total edges, so we should divide by 2^{2N^2}. That gets us a total of 1.5^{N^2} configurations.

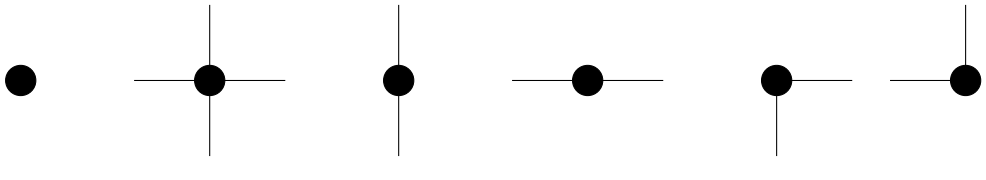
30 years later, Lieb saw this and noticed that it was shockingly good for a mathematical heuristic, so the **Lieb square ice constant** was computed (note that this is for the non-domain-wall case, so don't get it confused with the example we've been drawing so far). It was found that if f_N is the number of configurations for the system at size N,

then

$$\lim_{N\to\infty} f_N^{1/N^2} = \left(\frac{4}{3}\right)^{3/2} \approx 1.54.$$

Remark 12. *Recall that the number of alternating sign matrices would give us the count of configurations for domain-wall, which turns out to also be c^{N^2} but with a smaller constant (forcing the paths to take a certain arrangement significantly reduces the number of possible configurations). Domain-wall configuration does arise in models of graphene, since that's an arrangement of atoms which is "layered," and we end up seeing some similar physical phenomena with "arctic regions."*

Getting back to business with the six-vertex model now, recall that the six-vertex model corresponds right-up corners to $+1$ and up-right corners to -1, and all other arrangements are given 0. More generally, we can actually assign a weight to each possible vertex and weight configurations based on the product of these vertex weights:



We talked about this a bit last time as a "caution:" ASMs are like assigning $a_i = b_i = c_i = 1$, and domino tilings are like assigning specifically $c_1 = 2$ and all other constants 1. These two cases are actually of different complexities, and we can explain that through the parameter

$$\Delta = \frac{a_1a_2 + b_1b_2 - c_1c_2}{2\sqrt{a_1a_2b_1b_2}}$$

(this contains a lot of the algebraic structure behind the model). It turns out that $\Delta = 0$ is the "free fermion point" of the model, meaning that we get many determinants and we can solve everything by determinantal formulas (like Schur polynomials). So domino tilings are "very nice" because $\Delta = 0$, but alternating sign matrices are less nice and we have $\Delta = \frac{1}{2}$ (there are few, if any, detemrinants in this case). So one of the focuses of this course is "how can we deal with these determinants to get asymptotics results."

Example 13

We can now use this all to introduce the **stochastic six-vertex model**, which is the case where

$$a_1 = a_2 = 1, \quad b_1 + c_1 = 1, \quad b_2 + c_2 = 1.$$

In such a case, we can sample the paths in a Markovian fashion as follows: with some boundary condition (like starting all paths from the left), we can decide what to do with each path one step at a time. Indeed, if the path is moving right, we flip a coin to keep going straight with probability b_2 and turn up with probability c_2, and if the path is moving up, we flip a coin to keep going straight with probability b_1 and turn right with probability c_1. We can keep doing this until all paths exit (though we can't specify if they exit from the top or right).

In this case, we have

$$\Delta = \frac{1 + b_1b_2 - (1-b_1)(1-b_2)}{2\sqrt{b_1b_2}} = \frac{b_1+b_2}{2\sqrt{b_1b_2}} \geq 1.$$

So this will be a somewhat harder model to study, but we can still say some things about it:

Theorem 14 (Borodin–Corwin–Gorin, 2014)

Fix some vertex (x, y) (with x, y of the same order). The number of arrows passing to the right of (x, y) (meaning they cross the y-horizontal line after the x-vertical line) is some number between 0 and y, inclusive. Then

$$\frac{\#\text{ paths} - c_1 y}{c_2 y^{1/3}} \to \text{Tracy-Widom distribution.}$$

So we see the same fluctuations as in the other models we've introduced so far! This result is very specific to a few things though: we can only look at one point at a time, and it's only for specific initial data for the domain-wall case. There are results for arbitrary boundary conditions as well (For entering from the two axes); the fluctuations will still be of order $y^{1/3}$ but we get something more general than Tracy-Widom (this is where we get the KPZ fixed point).

The algebraic underpinning in this case is sometimes called the **Hall-Littlewood measure**, which is a generalization of the Schur measure. But this goes beyond the realm of determinants, so we won't talk about it much. (We defined the Schur measure explicitly in terms of Schur polynomials s_λ defined using determinants, but for Hall-Littlewood we replace s_λ with the Hall-Littlewood polynomials P_λ.)

Turning to techniques now, we'll begin with the "oldest modern" algebraic technique for studying determinants, which may be useful beyond the specific models we've described (e.g. appearing in random surfaces, growth models, random matrices). The plan for the next few weeks will be to start with a probabilistic model (a variant of LPP), play with it a bit, and get a determinant. We'll then "go algebraic" and set up some theory of symmetric functions and determinantal point processes, at which point we can go back and see how the model fits in so we can analyze it.

Example 15 (Polynuclear growth model (PNG))

The polynuclear growth model is a random height function (or random interface) $h(x, t)$ with x denoting space and t denoting time. We take the first quadrant and rotate it 45 degrees to get an upward-pointing wedge. At each vertex, we place an independent random variable; we'll be evolving the system upward, so we will index these vertices / variables as $w(i, j)$ if we're indexing by the 45-degree lattice and $\varpi(i, j)$ if we index the usual way. (So the closest point to the origin is $w(1, 1) = \varpi(0, 2)$, and we have $w(1, 2) = \varpi(-1, 3)$.)

So in our rotated grid (with ϖ coordinates), horizontal is space and vertical is time. Notice that $\varpi(x, y)$ is not set if $x + y$ is odd; thus we **define** $\varpi(i, j) = 0$ **for all of those vertices**. The law of the remaining variables $w(i, j)$ can be any of the following:

- w are iid geometric variables, so $\mathbb{P}(w(i, j) = k) = q^k(1 - q)$ for some $q \in (0, 1)$.
- More generally, we have

  $$\mathbb{P}(w(i, j) = k) = c_{ij}^k(1 - c_{ij})$$

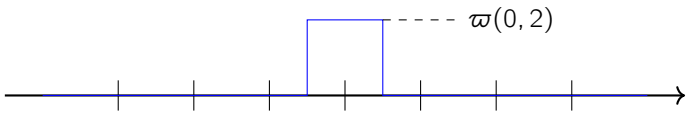
  where $c_{ij} = a_i b_i$ for some parameters so that $a_i b_i \in (0, 1)$. So in this case, the variables are not-independent geometrics with parameter coupled together in a certain way. (This will eventually be easier to use to see structure, but for simplicity we can start by thinking about the iid case where all as and bs are $\sqrt{q}$.)

So to define the growth model itself, we have the initial condition $h(x, 0) = 0$, and to evolve in time we have

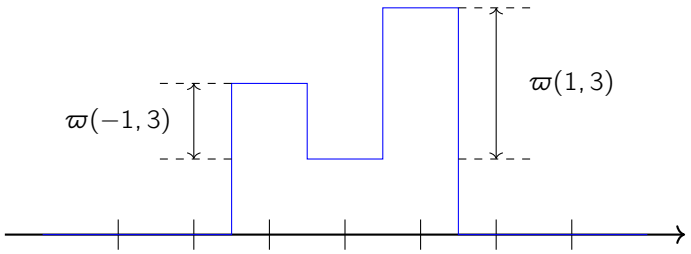
$$h(x, t + 1) = \max\{h(x - 1, t), h(x, t), h(x + 1, t)\} + \varpi(x, t + 1).$$

So the picture we can imagine is that we have a horizontal line moving upward, and the height starts off at 0 everywhere.

But then the first time something happens is that we have a nonzero weight at $w(1,1) = \varpi(0,2)$, so suddenly at $t = 2$ the height goes up at that one point. At subsequent times, this height will then "expand out horizontally." So at fixed time $t = 2$ the height $h(x, 2)$ will look something like the following (everything is only evaluated at integer values of x, but we'll draw the height with horizontal line segments by extending the value by $1/2$ in each direction):



But then at time $t = 3$, the values at $\varpi(-1,3)$ and $\varpi(1,3)$ kick in, so we get something like this:

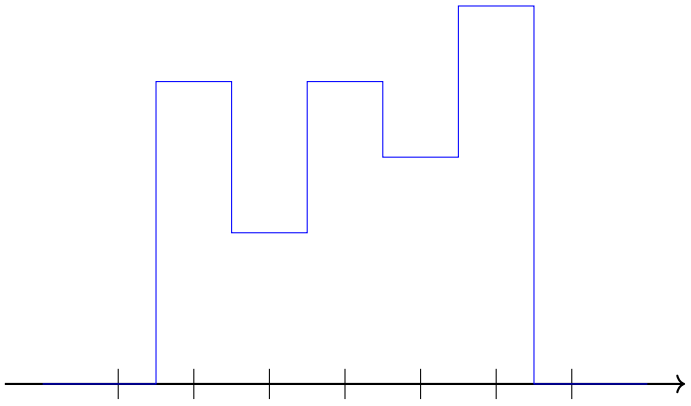


So plateaus "expand out by 1 to the left and right at each step, and whenever they expand out, they also increase by the associated ϖ at each point which is expanded to" (and the highest growth is taken). We'll see this in more detail next time, but this is exactly in correspondence to last passage percolation:

$$h(x,t) = \max_{\substack{P:(0,0)\to(x,t) \\ \text{directed up-left/up-right paths}}} \sum_{v\in P} \varpi(v).$$

3 April 7, 2026

We described the polynuclear growth model last time, which we can think of as growing a one-dimensional height function over time. Given some random variables at integers points in the first quadrant, we rotate those values 45 degrees counterclockwise and then use them to increment the height function via $h(x, t+1) = \max\{h(x-1,t), h(x,t), h(x+1,t)\} + \varpi(x, t+1)$. Pictorally, at each time step, island expand outward by 1 unit and also gain height boosts at the expanded locations (that is, at x with the same parity as t). For example, here is what the interface might look like at time $t = 4$ (after one time step from $t = 3$ in the previous lecture):



If we stare at these equations a bit, we'll see that this is exactly the same equation as the recursion for LPP (we might be concerned about having three different previous heights in the recursion, but because).

Fact 16

There was a question last time about whether this has anything to do with tropical geometry, and the answer is yes – there turns out to be a probabilistic interpretation of "tropicalization." So we'll spend a few minutes on an aside about this.

The idea is that in $\mathbb{R}_+$ we have the operations $+, \times$ taking (a, b) to $a + b$ and ab, respectively. We can then tropicalize this (it also goes under the name of **ultradiscretization**) by turning $+$ into max and $\times$ into $+$. Morally this turns a, b into $e^{A/\varepsilon}$ and $e^{B/\varepsilon}$ for ε small, so that $a + b = e^{A/\varepsilon} + e^{B/\varepsilon}$ is dominated by whichever of A and B are larger and hence is basically $e^{\max(A,B)/\varepsilon}$, and ab is exactly $e^{(A+B)/\varepsilon}$. So really we can think about $(a, b) \mapsto a + b$ as $(A, B) \overset{\approx}{\mapsto} \max(A, B)$ and $(a, b) \mapsto ab$ as $(A, B) \mapsto A + B$.

This kind of dictionary plays a role in probability because in last pasage percolation we have

$$L(M, N) = \max_{\substack{(0,0) \xrightarrow{P} (M,N) \\ \text{directed up-right paths}}} w(P).$$

IF we try to "de-tropicalize" this, then the max becomes a sum and the sum becomes a product, and thus we have

$$T(M, N) \sim \sum_{\substack{(0,0) \xrightarrow{P} (M,N) \\ \text{directed up-right paths}}} \prod_{u \in P} w(u)$$

and this is exactly the **polymer model** in probability. Specifically, we can define a polymer measure on paths (where the weight of a path is the product of the weights along that path) and then we sample a path proportional to the path weight. We're then often curious about the behavior of the normalizing constant, and in this case it is $Z = T(M, N)$. So we can get an analog of the Tracy-Widom results we talked about last week too:

Theorem 17

Assume the vertex weights w_{ij} are Gamma random variables, meaning that w_{ij} are iid and have density $x^{\theta-1}e^{-x}dx$ for some $\theta > 0$. Take $Z = T(N, N)$. Then $\frac{\log Z - c_1 N}{c_2 N^{1/3}}$ converges to the Tracy-Widom (GUE) distribution.

And if we take $\theta \to \infty$ in some suitable sense, this recovers LPP because the tropicalization becomes accurate – this is sometimes called the zero-temperature limit (so that the variables are so large and so different from each other that it's just the biggest one that matters). So in general taking a limit allows a $(+, \times)$ model to specialize to the $(\max, +)$ model.

Remark 18. *If we try doing this with first-passage percolation, it's a bit tricky because we have to write* $\min(a, b) = -\max(-a, -b)$ *and so we end up getting the harmonic mean* $\frac{1}{\frac{1}{A}+\frac{1}{B}}$. *But if we try to repeat this operation for a big sum it becomes quite involved, so this hasn't really been studied much. Usually polymer models are "inherently max objects" though.*

Note that the governing objects behind last-passage percolation are **determinants and Schur polynomials**, while the governing objects behind polymers are ***q*-Whittaker functions**. But there are few (if any) determinants in the latter case, so our goal will be to focus more on things like the former and use them when they're available.

Turning back to the polynuclear growth model, we'll need to add a bit more structure so that the determinants become apparent. First of all, notice that when we merge two islands together, we forget some of the heights because they get overwritten. So after some time, just looking at the height function will not tell us all of the variables it took

to get there, and therefore we'll make our model richer by considering the **multi-layer PNG**. The idea is that instead of a single interface, we'll have many of them:

Example 19

The **multi-layer polynuclear growth model** is a family of height functions (we'll also call them **interfaces**) $h_\ell(x,t)$ for $\ell \in \mathbb{N}$, $x \in \mathbb{Z}$ (extended to $\mathbb{R}$ by extending each value by $1/2$ in each direction) and $t \in \mathbb{Z}_+$.

- First of all, $h_1(x,t) = h(x,t)$ for all x, t.
- Now for $h_\ell(x,t)$, we set $h_\ell(x,0) = 1-\ell$, and then

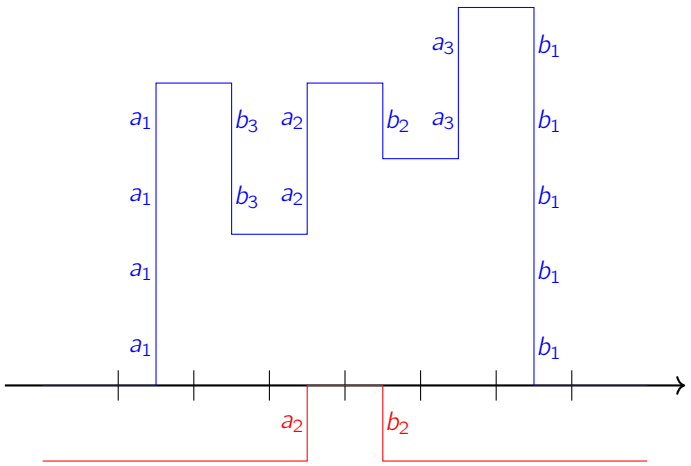
$$h_\ell(x,t+1) = \max\{h_\ell(x-1,t), h_\ell(x,t), h_\ell(x+1,t)\} + \varpi_\ell(x,t+1),$$

where $\varpi_\ell(x,t+1)$ is the "overlap at x at level $\ell-1$." This is easiest to understand with an explicit example: suppose we have height $3,2,5$ at three adjacent integer points, and the islands of height 3 and 5 expand outward. Then the overlap created in the middle would be 1, because that's the height above 2 which came from both sides.

The idea is that these height functions all lie below the original one, and whenever two islands merge, we want a disturbance in the function below so that we can still keep track of it. (The second height function at time $t = 4$ is shown in red below.) This is the minimal way to keep track of all of the weights, and secretly this is some "local toggle RSK" and also the Yang-Baxter equations (though we won't go there – we can see the paper "Spin q-Whittaker polynomials and deformed quantum Toda" by Mucciconi and Petrov for more).

By induction we see that the heights satisfy $h_1 > h_2 > h_3 > \cdots$ (because if $h_{\ell+1}$ increased by some amount, then h_ℓ needs to increase by at least that amount). So this creates a series of non-intersecting paths, also sometimes called a **line ensemble**. Our goal now is to understand the joint law of these paths, and the advantage of LPP is that we can actually compute them exactly.

Recall that our $\varpi(i,j)$ random variables are geometric with parameter $c_{ij} = a_j b_i$ (so we have parameters for each row and column of the original quadrant), meaning that $\mathbb{P}(w_{ij} = k) = c_{ij}^k(1-c_{ij})$. Each of our paths goes up and down, and for every time the path goes up or down, we will label it with an a or b parameter, with increasing index from the left and from the right, respectively.



So we assign a_is to increments and b_is to decrements, and we also assign weight 1 to the horizontal edges of the interfaces (where they stay flat). So every edge in a path gets some weight, and for each fixed x, t we know exactly what the variables will be (it's some straightforward linear combination).

Lemma 20 (Johansson '02)

Fix some time t (so we know what variables to consider). For any ensemble (h_ℓ), assign to it a weight where we take the product of all edges in the paths. (So in our example above, we would have $a_1^4 a_2^3 a_3^2 b_1^5 b_2^2 b_3^2$.) Then the law of (h_ℓ) is given by random walks which go either up, to the right, or down, weighted by $w(h_\ell)$, conditioned to stay ordered (non-intersecting) and with the boundary conditions $h_\ell(\pm\infty, t) = 1 - \ell$.

We won't prove this (it's left as an exercise to us), but it's not particularly deep and can be done by induction. But it shouldn't be too surprising because the probability of $w_{ij} = k$ is proportional to $(a_j b_i)^k$, so adding a box at a given coordinate gives us a factor of $a_j b_i$, which is exactly the new edges that we would accumulate.

So the point is that we have a family of non-intersecting walks weighted by some particular weight, and thus the determinant will pop out. Indeed, we know that the probability of any configuration is proportional to its weight, so all we need is the normalizing factor (partition function). Abstractly, we have a fairly broad situation:

Proposition 21 (Lindström–Gessel–Viennot lemma)

Let (V, E) be a directed planar graph without cycles, and let $P(u, v)$ be the set of directed paths from u to v. To each $e \in E$, associate a weight $w(e)$, and then for any path P let $w(P) = \prod_{e \in P} w(e)$. Let $W(u, v) = \sum_{P \in P(u,v)} w(P)$.

Now suppose we have n starting points $u_1, \cdots, u_n$ and n ending points $v_1, \cdots, v_n$. Define $P(\vec{u}, \vec{v}) = P(u_1, v_1) \times \cdots \times P(u_n, v_n)$ to be the collection of paths that start at the u_is and end at the v_is, and let $P_{\text{NI}}(\vec{u}, \vec{v})$ be the subset of those paths which are non-intersecting, meaning that they share no common vertices. Define

$$W_{\text{NI}}(\vec{u}, \vec{v}) = \sum_{\substack{(P_1, \cdots, P_n) \\ \text{in } P_{\text{NI}}(\vec{u}, \vec{v})}} \prod_{j=1}^{n} w(P_j),$$

and assume there exists exactly one permutation $\sigma \in S_n$ where non-intersecting paths from u_i to $v_{\sigma(i)}$ can exist – without loss of generality relabel so σ is the identity. Then we can compute the total non-intersecting weight via a determinant

$$W_{\text{NI}}(\vec{u}, \vec{v}) = \det\left[W(u_i, v_j)\right]_{i,j=1}^{n}.$$

In our setting, the set of directed edges are alternating up and down vertically depending on the parity of x, as well as all right edges. The weights we assign are the as and bs along the vertical edges and 1 along the horizontal edges. We then want to set $u_i = (-\infty, 1 - i)$ and $v_i = (+\infty, 1 - i)$.

In particular, the assumption of a unique σ automatically holds if our graph is planar and "directed properly" (for example, the directed $\mathbb{Z}^2$ lattice), and all of the examples we'll consider in this course will fall into that category.

Proof. Opening up the determinant, we have

$$\begin{aligned}
\det\left[W(u_i, v_j)\right]_{i,j=1}^{n} &= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} W(u_i, v_{\sigma(i)}) \\
&= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} \sum_{\substack{P_{i,\sigma(i)} \\ \text{connecting} \\ u_i \to v_{\sigma(i)}}} w(P_{i,\sigma(i)})
\end{aligned}$$

Swapping the sum and the product, we thus sum over all families of walks with the given connections, and thus we

have

$$\det\left[W(u_i, v_j)\right]_{i,j=1}^n = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \left(\sum_{\{P_{i,\sigma(i)}\}} \prod_{i=1}^n w(P_{i,\sigma(i)}) \right),$$

meaning we have a weighted sum of collections of path weights. We can split this parenthetical term into two parts: the weight of the collection of non-intersecting paths, plus the weight of the collection of paths intersecting in at least one vertex. It remains to show that the latter part is zero, since the non-intersecting parts will only come from $\sigma = \mathrm{id}$ by assumption.

For this, suppose we fix the traces of the paths, meaning that we fix the set of edges used but we aren't told which edge belongs to which path. For any intersecting family – this property is visible from the trace because some vertex will have two incoming edges – consider the first such problematic vertex. Then we have two choices for how we can assign the edges to paths, since for any path collection we can flip what happens before that first intersection point. More specifically, if a point has e_1, e_2 incoming and $\overline{e}_1, \overline{e}_2$ outgoing, then we can have $e_1 \mapsto \overline{e}_1$ and $e_2 \mapsto \overline{e}_2$, or $e_1 \mapsto \overline{e}_2$ and $e_2 \mapsto \overline{e}_1$. But this rewiring keeps the weight of the intersecting paths the same, and it switches the permutation dictating the endpoints by a transposition. Thus the $\operatorname{sgn}(\sigma)$ term makes terms cancel out, and because we always do this at the first intersection point this pairs up all nonintersecting paths and has everything cancel out. □

4 April 9, 2026

Our current goal is connecting our model of interest (multi-layer polynuclear growth) to determinants and then to use symmetric functions to study those determinants. Last time, we described the Lindström–Gessel–Viennot lemma, which says that if we want to count the total weight of nonintersecting paths from u_i to v_i, then we take the determinant of the matrix encoding the individual weights $W(u_i, v_j)$.

We can apply that to our multi-layer polynuclear growth model by assigning upward and downward oriented edges weights of a_is and b_is, respectively; we described last time that the weight of any configuration is proportional to the product of weights $w(h_\ell)$. But the point is that if we want to understand something like $h_1(0)$ (which corresponds to the value of LPP at a particular point), it suffices to know the joint law of $(h_1(0), h_2(0), \cdots)$, which turns out to be easier to study directly because they come from non-intersecting paths connecting $(-\infty, 1-i)$ to $(\infty, 1-i)$. So we can write (NI means non-intersecting)

$$\mathbb{P}(h_1(0) = k_1, h_2(0) = k_2, \cdots) \sim (\text{weight of NI paths } (-\infty, 1-i) \to (0, k_i)) \times (\text{weight of NI paths } (0, k_i) \to (\infty, k_i)),$$

and we can write this out using the LGV lemma to write it as a determinant:

$$\mathbb{P}(h_1(0) = k_1, h_2(0) = k_2, \cdots) \sim \det\left[W((-\infty, 1-i) \to (0, k_j))\right]_{i,j=1}^n \det\left[W((0, k_i) \to (\infty, 1-j))\right]_{i,j=1}^n.$$

So we have a product of two determinants, which gives us an indication that this might have to do with s_λ functions where λ is somehow determined by the values of k. That's what we'll get into more now – we'll take a detour into **symmetric function theory** for a few lectures.

Example 22

Suppose we have N variables $x_1, x_2, \cdots, x_N$. Let Sym_N^n be the linear vector space of symmetric polynomials in the variables $x_1, \cdots, x_N$, homogeneous of degree n.

For example, Sym_N^1 is just spanned by $x_1 + \cdots + x_N$, and Sym_N^2 contains things like $x_1^2 + \cdots + x_N^2$ (the power sum),

$(x_1 + \cdots + x_N)^2$ (the square of the previous power sum), and $\sum_{i<j} x_i x_j$ (this is the elementary symmetric function). There are plenty of other examples, but of course they are not all linearly independent.

The way we will usually index these functions is via partitions, and here's the first way we might do that:

Definition 23

For any multi-index $\alpha = (\alpha_1, \cdots, \alpha_N) \in \mathbb{Z}^N_{\geq 0}$ (not necessarily a partition), set $x^\alpha = x_1^{\alpha_1} \cdots x_N^{\alpha_N}$, and then define the **monomial symmetric polynomial** to be the symmetrization

$$m_\lambda(x) = \sum_{\substack{\text{distinct permutations } \alpha \\ \text{of } \lambda}} x^\alpha$$

(we use x as shorthand for $(x_1, \cdots, x_N)$). Note that if $\lambda = (\lambda_1, \cdots, \lambda_\ell)$ has length $\ell > N$, then we automatically set $m_\lambda = 0$ (because it doesn't make sense to have more variables than we have), and if it has length $\ell < N$, we pad it with $N - \ell$ zeros and then things will make sense.

For degree 1, we have $m_{(1)}(x) = x_1 + \cdots + x_N$ (because the partition gets augmented by a bunch of 0s), and likewise the powersum $x_1^2 + \cdots + x_N^2$ is just $m_{(2)}(x)$ and $\sum_{i<j} x_i x_j$ is $m_{(1,1)}(x)$.

If we keep ourselves to having just N variables, we end up having some irritating dependencies with linear or algebraic dependence, so what we really want to do is take $N \to \infty$. To set that up, define the space of **all symmetric polynomials in N variables**

$$\mathrm{Sym}_N = \bigoplus_{n=0}^{\infty} \mathrm{Sym}^n_N.$$

Now for any $N > N'$, we have a projection map $\mathrm{pr}_{N,N'} : \mathbb{C}[x_1, \cdots, x_N] \to \mathbb{C}[x_1, \cdots, x_{N'}]$ where we just set the last $N - N'$ variables equal to zero (in words, if we have a polynomial with 10 variables, then just setting the last 3 equal to zero leaves us with the terms only involving the first 7 variables). Furthermore, $\mathrm{pr}_{N,N'}(x^\alpha) = x^\alpha$, where the x^α on the left-hand side has N variables and one on the right-hand side has N' variables. Thus we can define the projective limit

$$\mathrm{Sym}^n = \varprojlim \mathrm{Sym}^n_N,$$

which formally consists of sequences $(f_1, f_2, \cdots)$ where $f_n \in \mathrm{Sym}^n_N$ and $\mathrm{pr}_{N,N'}(f_N)$ for all $N > N'$. The idea is that we start with a symmetric function on a huge number of variables and then "extend it infinitely;" it's pretty easy to imagine how we would do it for the examples we described. Finally, we can just take a direct sum to get the full space of symmetric polynomials

$$\mathrm{Sym} = \bigoplus_{n=0}^{\infty} \mathrm{Sym}^n$$

(this is the set of all finite linear combinations of homogeneous polynomials in infinitely many xs). That means that we can define the monomial symmetric polynomials $m_\lambda(x)$ in Sym in a similar way to what we did in the finite case: for example it is true that we just have

$$m_{(1,1)}(x) = \sum_{1 \leq i < j} x_i x_j.$$

On the other hand, something like $(1 + x_1)(1 + x_2) \cdots$ will not be in Sym, because expanding out the sum would require an infinite linear combination $1 + \sum_i x_i + \sum_{i<j} x_i x_j + \sum_{i<j<k} x_i x_j x_k + \cdots$. (We could try to make sense of this by doing something like $(1 + tx_1)(1 + tx_2) \cdots$ for "t small" and defining some valuation, but we just won't worry about that here.)

Definition 24

The **elementary symmetric polynomials** are defined as follows. For any n, we have

$$e_n(x) = \sum_{i_1<\cdots<i_n} x_{i_1}x_{i_2}\cdots x_{i_n},$$

and then for any partition λ set $e_\lambda(x) = e_{\lambda_1}(x)e_{\lambda_2(x)}\cdots e_{\lambda_\ell}(x)$.

These e_ns have an explicit generating function: from the discussion above, we have

$$E_t(x) = \sum_{n=0}^{\infty} t^n e_n(x) = \prod_{i=1}^{\infty}(1+tx_i)$$

even though both sides aren't actually in Sym.

Lemma 25

The polynomials e_n are algebraically independent.

We can see the value in taking $N \to \infty$ here; indeed if we only had v variables, as soon as $n > v$ we would require $e_n = 0$ because the sum would be empty and the lemma would be false.

Proof. To show algebraic independence, we must show that any nonzero polynomial in the e_ns is nonzero. But the monomials in the e_ns are exactly the e_λs, so it suffices to show that the e_λs are linearly independent.

For this, we know that the m_λs are linearly independent (by definition they were obtained by symmetrizing linearly independent things). We want an invertible change-of-basis from the m_λs to the e_λs, which we can describe as follows. For any partition λ, recall that its **transpose** (also **conjugate**) is obtained by reflecting its Young diagram about its 45-degree diagonal so that rows become columns. The basic claim now is that

$$e_{\lambda'} = m_\lambda + \sum_{\mu<\lambda} c_{\lambda,\mu} m_\mu,$$

where $\mu < \lambda$ is under the lexicographic ordering (meaning there exists some i such that $\mu_1 = \lambda_1, \cdots, \mu_{i-1} = \lambda_{i-1}$ and then $\mu_i < \lambda_i$) and $c_{\lambda,\mu}$ are some real numbers. For example, we have $\lambda' = (1,1)$ and $\lambda = 2$, and indeed

$$e_{\lambda'} = e_1 e_1 = \left(\sum_i x_i\right)^2 = \sum_i x_i^2 + 2\sum_{i<j} x_i x_j.$$

More generally, we can show this claim by writing out

$$e_{\lambda_1'}e_{\lambda_2'}\cdots = \left(\sum_{i_1<\cdots,i_{\lambda_1'}} x_{i_1}\cdots x_{i_{\lambda_1'}}\right)\left(\sum_{i_1<\cdots,i_{\lambda_2'}} x_{i_1}\cdots x_{i_{\lambda_2'}}\right)\cdots.$$

We can then expand this out as a product of monomial symmetric polynomials. We get the one with the biggest possible λ by ensuring the most repetitions (since that makes λ_1 as big as possible, then λ_2 as big as possible afterward, and so on). Thus the corresponding term will come from taking the symmetrization of taking $(x_1x_2\cdots x_{\lambda_1'})(x_1x_2\cdots x_{\lambda_2'})\cdots$ from the product. But then we can check that this is exactly $x_1^{\lambda_1}x_2^{\lambda_2}\cdots$, which is exactly m_λ, and all other terms are for smaller partitions.

So we get the basis by first shuffling the m_λs via transpose $\lambda \mapsto \lambda'$, and then we get this triangular change-of-basis to get to the e_λs, as desired. □

The other family of symmetric polynomials that is often quite useful is the following:

Definition 26

The **complete symmetric polynomials** are like the elementary symmetric polynomials, but with weak ordering instead of strict ordering: for any n we have

$$h_n(x) = \sum_{i_1 \le \cdots \le i_n} x_{i_1} \cdots x_{i_n},$$

and then for any partition λ we have $h_\lambda(x) = h_{\lambda_1(x)} \cdots h_{\lambda_\ell(x)}$.

For example, we have

$$h_2(x) = \sum_i x_i^2 + \sum_{i<j} x_i x_j.$$

We can again write down a generating series for this: we have

$$\begin{aligned} H_t(x) &= \sum_{n=0}^{\infty} t^n h_n(x) \\ &= \sum_{n=0}^{\infty} \sum_{i_1 \le \cdots \le i_n} (tx_{i_1}) \cdots (tx_{i_n}) \\ &= \prod_{i=1}^{\infty} \sum_{k=0}^{\infty} (tx_i)^k \\ &= \prod_{i=1}^{\infty} \frac{1}{1-tx_i}. \end{aligned}$$

(The tricky step here is the third line: there's no overcounting factor because if we want any given monomial, there's only one way to order them in how they show up in $x_1, x_2, \cdots$.) Notice in particular that our two generating series satisfy

$$E_{-t}(x) H_t(x) = \prod_{i=1}^{\infty} (1-tx_i) \prod_{i=1}^{\infty} \frac{1}{1-tx_i} = 1,$$

and therefore if we look at the coefficient of t^n on the left-hand side we have

$$\sum_{r=0}^{n} (-1)^r e_r h_{n-r} = \begin{cases} 1 & \text{if } n = 0, \\ 0 & \text{otherwise.} \end{cases}$$

But because the e_rs are algebraically independent (and in fact form an algebraic basis of Sym, since the e_λs span Sym), we can define a homomorphism $\omega : \text{Sym} \to \text{Sym}$ which maps e_r to h_r. And the relation above tells us that $\omega^2 = 1$, since applying ω to it tells us that

$$\sum_{r=0}^{n} (-1)^r h_r \omega(h_{n-r}) = \begin{cases} 1 & \text{if } n = 0, \\ 0 & \text{otherwise.} \end{cases}$$

The key point is that by induction, knowing all of the es and also knowing $h_0, \cdots, h_{N-1}$ automatically specifies h_N. Therefore there's only one possible solution here and we must actually have $\omega(h_{n-r}) = e_{n-r}$, which gives us the following:

Corollary 27

The complete symmetric polynomials (h_n) are algebraically independent.

(Our map ω will also be useful for other purposes later on.) We'll do one more family before we get to the Schur polynomials:

Definition 28

The **power sum polynomials** are defined by setting

$$p_r(x) = \sum_{i=1}^{\infty} x_i^r$$

for any n and then letting $p_\lambda(x) = p_{\lambda_1}(x) \cdots p_{\lambda_\ell}(x)$.

The (p_r)s are algebraically independent, for example by finding an upper-triangular change-of-basis from m_λs to p_λs like we did for the e_ns. Another way to do this is via the **Jacobian criterion**: if we want to check that a family $(f_1, \cdots, f_n)$ are algebraically independent, we can examine the determinant $\det \left[\frac{d}{dx_j} f_i\right]_{1 \le i,j \le n}$ and see if it is nonzero, and this is exactly a multiple of the Vandermonde determinant. And a third way we can check this more directly is to take a slight twist of the generating series

$$\begin{aligned} p_t(x) &= \sum_{n=1}^{\infty} p_n \frac{t^n}{n} \\ &= \sum_{n=1}^{\infty} \sum_{i=1}^{\infty} \frac{x_i^n t^n}{n} \\ &= \sum_{i=1}^{\infty} \frac{(tx_i)^n}{n} \\ &= \sum_{i=1}^{\infty} -\log(1 - tx_i) \\ &= \log \left(\prod_{i=1}^{\infty} \frac{1}{1 - tx_i} \right) \\ &= \log H_t(x). \end{aligned}$$

So now differentiating in x yields $P_t'(x) = \frac{H_t'(x)}{H_t(x)}$, and now multiplying both sides by $H_t(x)$ and opening everything up lets us express the p_rs in terms of the h_ns. We thus find that $\{p_r\}_{r=0}^n$ algebraically generate the same space as $\{h_r\}_{r=0}^n$, so the transcendence degrees must be the same; since h_rs are algebraically independent so are the p_rs.

5 April 14, 2026

We introduced the ring Sym of symmetric functions last time, and we described some particular sets of elements of them which have nice independence properties (the monomial, elementary, complete, and power sum symmetric polynomials) – all of these are indexed by partitions λ. We'll now talk about another set of them, the Schur polynomials. These polynomials are naturally defined directly in terms of the partitions:

Definition 29

For a partition λ of length at most n (padding with zeros to length n), define $a_\lambda(x) = \det\left[x_i^{\lambda_j}\right]_{1\le i,j\le N}$. This is an antisymmetric polynomial in the xs. Let $\delta = \delta_N = (N-1, N_2, \cdots, 0)$ be the "staircase partition," and define the **Schur polynomial**

$$s_\lambda(x) = \frac{a_{\lambda+\delta}(x)}{a_\delta(x)} = \frac{\det\left[x_i^{\lambda_j+N-j}\right]_{1\le i,j\le N}}{\det\left[x_i^{N-j}\right]_{1\le i,j\le N}}.$$

Notice that the denominator here is just the Vandermonde determinant $\prod_{i<j}(x_i - x_j)$ (because it is antisymmetric in the xs, meaning it divides each $x_i - x_j$, and has degree $\binom{n}{2}$). And the reason for this δ is basically that parts of λ might be equal, but adding δ means they will be spread out so that we don't have two columns of the determinant equal.

We've previously shown that e_n, h_n, p_ns form algebraic bases of Sym so that $e_\lambda, h_\lambda, p_\lambda$ form linear bases. This turns out to also be true for the Schur polynomials s_λ:

Proposition 30

The polynomials s_λ form a linear basis of Sym.

Proof. The polynomials $a_\delta s_\lambda$ are antisymmetric in x by definition, and expanding out the determinant yields

$$a_\delta s_\lambda = \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma) x^{\sigma(\lambda+\delta)}.$$

But observe that $[x^\mu]_\mu$ spanning all polynomials in x implies that the antisymmetrization $\sum_\sigma \operatorname{sgn}(\sigma) x^{\sigma(\mu)}$ span all **antisymmetric** polynomials in x. And now the only difference in what we have in $a_\delta s_\lambda$ is that we only take μ of the form $\lambda + \delta$; that is, we only consider strict partitions. But again if μ had any two parts equal, then $\sum_\sigma \operatorname{sgn}(\sigma) x^{\sigma(\mu)}$ is already zero. Thus we do indeed form a basis over anti-symmetric polynomials using $a_\delta s_\lambda$s.

But now any antisymmetric polynomial necessarily has to divide all $x_i - x_j$ factors; thus dividing by the Vandermonde determinant is exactly an isomorphism from antisymmetric to symmetric polynomials. Thus if $a_\delta s_\lambda$ span antisymmetric polynomials, then s_λ must span Sym. □

An alternative proof of this result is to write out

$$s_\lambda = m_\lambda + \sum_{\mu<\lambda} K_{\lambda\mu} s_\mu$$

where this time $\mu < \lambda$ is under the **dominance ordering**, meaning that $\sum_{i=1}^r \mu_i \le \sum_{i=1}^r \lambda_i$ for all r. (These $K_{\lambda\mu}$s are called the Kostka numbers, but there are lots of things called that so we should be careful.) Since we just did the other proof, we won't derive the details of this strategy.

Having defined these s_λs, we can now prove the Cauchy identity that we stated in the first lecture:

Proposition 31 (Cauchy identity)

We have

$$\sum_\lambda s_\lambda(x) s_\lambda(y) = \prod_{x_i\in x, y_j\in y} \frac{1}{1 - x_i y_j}.$$

In the finite case we can interpret this as taking a sum over all partitions λ of length at most N and then having $x = (x_1, \cdots, x_N)$ and $y = (y_1, \cdots, y_N)$. But the way we write this is really "setting $N = \infty$."

Remark 32. *Notice that neither side of this equation actually resides in $Sym(x) \otimes Sym(y)$, since the degrees are unbounded on both sides if we write the fractions as geometric series. These expressions instead reside in the completion, which includes elements of the form $\sum_{\lambda,\mu} c_{\lambda\mu} m_\lambda(x) m_\mu(y)$ with no decay constraints on the constant $c_{\lambda\mu}$. In our case it turns out that as long as all x_i, y_js are less than 1, both sides will converge, though.*

Proof. Since the left-hand side involves determinants, we also want to express the right-hand side in a similar form. For simplicity of notation, we'll express the proof in the finite N case. We'll begin by proving the **Cauchy determinant identity**

$$\begin{aligned}\det\left[\frac{1}{1-x_iy_j}\right]_{1\le i,j\le N} &= \frac{\prod_{1\le i<j\le N}(x_i-x_j)(y_i-y_j)}{\prod_{1\le i,j\le N}(1-x_iy_j)} \\ &= \frac{a_\delta(x)a_\delta(y)}{\prod_{1\le i,j\le N}(1-x_iy_j)}.\end{aligned}$$

Indeed, this is true by a similar argument to Vandermonde: after clearing denominators $\prod_{1\le i,j\le N}(1-x_iy_j)$, we're left with a polynomial on both sides and the degrees of the two sides are both equal to $2\binom{N}{2} = N(N-1)$. But then if $x_i = x_{i'}$ or $y_i = y_{i'}$ for some $i \neq i'$, then they must be equal; thus we do get the Vandermonde factors and the constants match up.

Remark 33. *The usual way the Cauchy identity is written in terms of $\frac{1}{x_i-y_j}$ instead of $\frac{1}{1-x_iy_j}$. But we can get from one form to another by replacing y_j with $\frac{1}{y_j}$ and doing a bit of rearrangement.*

So now to prove the proposition, it suffices to prove the same thing if we multiply both sides by $a_\delta(x)a_\delta(y)$ (remember we're fixing N so δ is exactly the staircase $(N-1, N-2, \cdots)$): by the Cauchy determinant identity, that means we're left to prove

$$\begin{aligned}&\sum_\lambda a_\delta(x)s_\lambda(x)a_\delta(y)s_\lambda(y) \stackrel{?}{=} \det\left[\frac{1}{1-x_iy_j}\right] \\ \iff &\sum_\lambda a_{\lambda+\delta}(x)a_{\lambda+\delta}(y) \stackrel{?}{=} \det\left[\frac{1}{1-x_iy_j}\right].\end{aligned}$$

Opening up the determinant on the right-hand side yields

$$\begin{aligned}\det\left[\frac{1}{1-x_iy_j}\right] &= \det\left[\sum_{k=0}^\infty (x_iy_j)^k\right] \\ &= \sum_{\sigma\in S_N}\sum_{k_1,\cdots,k_N} \operatorname{sgn}(\sigma)\left(x_1y_{\sigma(1)}\right)^{k_1}\cdots\left(x_Ny_{\sigma(N)}\right)^{k_N}.\end{aligned}$$

Now if $k_i = k_j$ for any $i \neq j$, the contribution is zero by antisymmetry; otherwise we can order the k_is by a permutation ρ so that they are in decreasing order, meaning that $k_{\rho^{-1}(1)} > k_{\rho^{-1}(2)} > \cdots > k_{\rho^{-1}(N)}$. Call these new indices $\ell_1, \cdots, \ell_N$. Thus our sum becomes

$$\begin{aligned}&\sum_{\sigma\in S_N}\sum_{\rho\in S_N}\sum_{\ell_1>\cdots>\ell_n} \operatorname{sgn}(\sigma)\left(x_1y_{\sigma(1)}\right)^{\ell_{\rho^{-1}(1)}}\cdots\left(x_Ny_{\sigma(N)}\right)^{\ell_{\rho^{-1}(N)}} \\ &= \sum_{\sigma\in S_N}\sum_{\rho\in S_N}\sum_{\ell_1>\cdots>\ell_n} \operatorname{sgn}(\sigma)\left(x_{\rho(1)}y_{\sigma\rho(1)}\right)^{\ell_1}\left(x_{\rho(N)}y_{\sigma\rho(N)}\right)^{\ell_N}\end{aligned}$$

where in the last line we've "changed variable order" from x_i to $x_{\rho(i)}$. But now we can call $\tau = \sigma\rho$ and relabel the sum;

this leaves us with

$$\sum_{\rho\in S_N}\sum_{\tau\in S_N}\sum_{\ell_1>\cdots>\ell_n}\operatorname{sgn}(\rho)\operatorname{sgn}(\tau)\left(x_{\rho(1)}y_{\tau(1)}\right)^{\ell_1}\left(x_{\rho(N)}y_{\tau(N)}\right)^{\ell_N}.$$

The permutation parts are now decoupled, so if we swap the summation order we can also write this as

$$\sum_{\ell_1>\cdots>\ell_N}\left(\sum_{\rho\in S_N}\operatorname{sgn}(\rho)x_{\rho(1)}^{\ell_1}\cdots x_{\rho(N)}^{\ell_N}\right)\left(\sum_{\tau\in S_N}\operatorname{sgn}(\tau)y_{\tau(1)}^{\ell_1}\cdots y_{\tau(N)}^{\ell_N}\right)=\sum_{\ell}a_\ell(x)a_\ell(y).$$

In particular, since ℓ will always be a strictly decreasing sequence, we can always write it as $\ell=\lambda+\delta$ for a partition λ, so this is exactly the desired left-hand side from above. □

There's also a GSV proof of this result, but that will become more natural once we write the Schur polynomials as determinants rather than just ratios of determinants:

Proposition 34 (Jacobi–Trudi formula)

Recall that the h_ns are the complete symmetric functions. We have

$$s_\lambda=\det\left[h_{\lambda_i-i+j}\right]_{1\le i,j\le N},$$

where we set $h_n=0$ if $n<0$.

We first state a claim:

Lemma 35

Let $(f_0,f_1,\cdots)$ be coefficients (just think of them as being in some formal ring) and set $f(y)=\sum_{i=0}^{\infty}f_iy^i$ to be their generating series. Then

$$f(y_1)\cdots f(y_N)=\sum_{\lambda}\det\left[f_{\lambda_i-i+j}(y)\right]s_\lambda(y).$$

(The idea is that the left-hand side is a symmetric polynomial, so we can write it as some linear combination of the s_λs, and we claim the coefficients are determinants.) First, let's see how we can use this to prove Jacobi–Trudi:

Proof of Proposition 34. Set $f_i=h_i(x)$ in Lemma 35, so that for a single variable we have

$$f(y)=\sum_{i=0}^{\infty}h_i(x)y^i=\prod_{i=1}^{N}(1-x_iy)^{-1}$$

by the generating series for the complete symmetric polynomials. Thus Lemma 35 yields

$$\prod_{i,j}\frac{1}{1-x_iy_j}=\sum_{\lambda}s_\lambda(y)\det\left[h_{\lambda_i-i+j}(x)\right].$$

But the left-hand side, by the Cauchy identity, is $\sum_\lambda s_\lambda(x)s_\lambda(y)$. Since the s_λs form a basis, that means the coefficients of $s_\lambda(y)$ on both sides must be the same, which is exactly what we wanted to show. □

Proof outline of Lemma 35. This is really just linear algebra; we use the only trick we have, which is to open up the determinant. Multiplying both sides by $a_\delta(y)$ and expanding the left-hand side yields (via expanding the determinant

for a_δ and writing out the generating series)

$$\begin{aligned} &a_\delta(y) f(y_1) \cdots f(y_N) \\ &= \sum_{\sigma \in S_N} \operatorname{sgn}(\sigma) y_1^{N-\sigma(1)} y_2^{N-\sigma(2)} \cdots y_N^{N-\sigma(N)} \sum_{m_1, m_2, \cdots, m_N} \left(f_{m_1} y_1^{m_1}\right) \cdots \left(f_{m_N} y_N^{m_N}\right). \end{aligned}$$

We can now collect the factors of y to get

$$\sum_{\sigma \in S_N} \sum_{m_1, m_2, \cdots, m_N} \operatorname{sgn}(\sigma) f_{m_1} \cdots f_{m_N} y_1^{N+m_1-\sigma(1)} \cdots y_N^{N+m_N-\sigma(N)}.$$

So now for any $y_1^{k_1} \cdots y_N^{k_N}$, for $k_i = N + m_i - \sigma(i) \iff m_i = k_i + \sigma(i) - N$, its corresponding coefficient in this whole expression is

$$\sum_{\sigma \in S_N} \operatorname{sgn}(\sigma) f_{m_1} \cdots f_{m_N} = \sum_{\sigma \in S_N} \operatorname{sgn}(\sigma) f_{k_1+\sigma(1)-N} \cdots f_{k_N+\sigma(N)-N},$$

and this is exactly the determinant $\det\left[f_{k_i+j-N}\right]_{1 \le i,j \le N}$. So if we want the contribution to be nonzero, all k_is should be distinct. Ordering the ks then gives us a strict permutation which we can write as $\lambda + \delta$; we get a coefficient of $\sum_{\sigma \in S_n} \operatorname{sgn}(\rho) y_j^{\rho(\lambda+\delta)}$, which is exactly $\det\left[f_{(\lambda+\delta)_i - N + j}\right]$. Then by checking indices we can confirm that this is exactly the $\det\left[f_{\lambda_i - i + j}\right]$ that we wanted. □

Remark 36. *The "messiness" of this whole strategy comes from doing everything from scratch. There are ways of proving this which are more visual with Yang-Baxter equations, but with formulas this is probably the most direct proof.*

Fact 37

In a similar vein to our proof that $s_\lambda = \det\left[h_{\lambda_i - i + j}\right]$, we can also show that

$$s_\lambda = \det\left[e_{\lambda'_i - i + j}\right],$$

where λ' is the transpose of λ.

This is actually most easily checked by verifying equality of the two determinants directly, and we'll sketch that here. Recall from the generating functions that

$$\sum_{r=0}^{n} (-1)^r e_r h_{n-r} = \begin{cases} 1 & \text{if } n = 0, \\ 0 & \text{otherwise.} \end{cases}$$

and so we can define the (upper-triangular) matrices $H = [h_{i-j}]$ and $E = [(-1)^{i-j} e_{i-j}]$. By our identity above we have $H^{-1} = E$, meaning that the minors of E are exactly the cofactors of H. (Even though we have infinite matrices, the cofactors are only going to have "finite disruptions" because things are upper triangular.) Then these minors and cofactors turn out to exactly index those two different determinants above (by choosing the rows and columns to correspond to exactly the ones giving us the $\lambda_i - i + j$s and the $\lambda'_i - i + j$s). Normally this would give us a factor of $\det H$, but because the diagonal entries of H are $h_0 = 1$, we have $\det H = 1$ and everything lines up. For a further reference, we can see section 2.9 of Macdonald's book "Symmetric Functions and Hall Polynomials."

The reason for pointing this all out is that we have the involution ω mapping $h_n \mapsto e_n$, and this identity tells us that $\omega(s_\lambda) = s_{\lambda'}$. Thus the Cauchy identity

$$\sum_\lambda s_\lambda(x) s_\lambda(y) = \prod_{x_i \in x, y_j \in y} \frac{1}{1 - x_i y_j}$$

with ω applied to the x-variables will turn into

$$\sum_{\lambda} s_{\lambda'}(x)s_{\lambda}(y) = \prod_{y_j \in y} \omega\left(\prod_{x_i \in x} \frac{1}{1-x_i y_j}\right),$$

and now because ω turns the generating series for the h_ns ($\frac{1}{1-xy}$) into the generating series for the e_ns ($1+xy$), the right-hand side now simplifies:

Proposition 38 (Dual Cauchy identity)
We have

$$\sum_{\lambda} s_{\lambda'}(x)s_{\lambda}(y) = \prod_{x_i \in x, y_j \in y} (1+x_i y_j).$$

6 April 16, 2026

Now that we've developed some symmetric function theory, we can get closer to returning to PNG. We've previously seen that the LGV lemma gets us formulas for nonintersecting walks through determinants, and in particular multilayer PNG is defined by non-intersecting walks, so the locations of those paths at some fixed vertical slice $x = 0$ can be determined by a product of two determinants. And last time, we also introduced the Jacobi–Trudi formula, which wrote our Schur polynomials as a determinant of complete symmetric functions h_n.

So one way to match directly is that our multilayer PNG formula had determinants involving partition functions for a single walk (from $(-\infty, 1-i)$ to $(0, k_i)$) – those turn out to be exactly the hs in the Schur polynomials. But that's not the route we'll take because it misses a bit of structure that will be useful for us, and we'll develop a few more aspects of symmetric functions that will make the appearance of Schur in PNG very clear.

Recall that the Schur functions s_λ have some set of variables which we can denote by $X = (x_1, x_2, \cdots)$. We'll now augment this polynomial to a symmetric function $s_\lambda(X, Y)$ jointly symmetric in the variables (X, Y), and then we can expand that as a symmetric function in Y whose coefficients are symmetric in X. But then we know that a basis of symmetric functions is Schur polynomials, so that means we can write

$$s_\lambda(X, Y) = \sum_{\mu} s_{\lambda/\mu}(X)s_\mu(Y).$$

These $s_{\lambda/\mu}$ coefficients are known as **skew Schur polynomials**.

Remark 39. *These turn out to be related to vertex models in the following way: if we have a bunch of paths coming in through the bottom and depositing at the top of a lattice grid, in the case where all paths enter at the same spot then the exits are dictated by Schur functions, but if they are not entering at the same spot then we get skew Schur functions instead. In this context the formula above is sometimes called the* **branching rule**.

We'll now state a bunch of properties which we won't prove (though Section 1.5 of the Macdonald book is a good reference):

Proposition 40 (Jacobi–Trudi formula for skew Schur functions)
We have (similarly to Proposition 34)

$$s_{\lambda/\mu} = \det\left[h_{\lambda_i - i - \mu_j + j}\right]_{1 \le i,j \le N}$$

This gives rise to the interpretation of these functions as **minors of Toeplitz matrices**. A **Toeplitz matrix** is a matrix where along each main diagonal, the entries are constant (so we can write the entries on the diagonal as a_0, directly above the diagonal as a_1, directly below as a_{-1}, and so on). So what we can do is consider the Toeplitz matrix where $a_i = h_i$, where $h_i = 0$ for i negative, and so the matrix looks like

$$M = \begin{bmatrix} 1 & h_1 & h_2 & \ddots \\ 0 & 1 & h_1 & \cdots \\ \vdots & \vdots & \ddots & h_1 \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Then the minors of this matrix are exactly skew Schur functions with rows and columns indexed by the shifted λ and μ (because $\lambda_i - i$ and $\mu_j - j$ will be strictly ordered).

For a bit of intuition on how we might prove this version of Jacobi–Trudi, consider the **symbol** of a Toeplitz matrix

$$T_A(z) = \sum_{i \in \mathbb{Z}} a_i z^i.$$

The symbol multiplies well, since for any two Toeplitz matrices A, B we have $T_A(z)T_B(z) = T_{AB}(z)$ by matrix multiplication. So now if our Toeplitz matrix of entry is the $M(x)$ matrix above (where x is implicitly the variables in the symmetric functions), we have

$$T_{M(x)}(z)T_{M(y)}(z) = T_{M(x)M(y)}(z),$$

but we also know that $T_A(z) = \sum_{i=0}^{\infty} h_i z^i = \prod_i \frac{1}{(1-zx_i)}$ by the generating function for the hs. Thus the symbol of $T_{M(x)M(y)}(z)$ is

$$\prod_i \frac{1}{1 - zx_i} \prod_i \frac{1}{1 - zy_i},$$

but this can also be thought of as $T_{M(x,y)}(z)$ where we treat xs and ys collectively as a single set of symmetric variables. So because the symbols are equal, actually we have $M(x, y) = M(x)M(y)$. (This is kind of like a discrete Fourier transform. In this case, it's not actually that necessary to use generating functions, since we could also just do the matrix multiplication out, but we'll see that other properties that are a little harder to read off will become clearer.)

So if we want to check consistency of Jacobi–Trudi by plugging into $s_\lambda(X, Y) = \sum_\mu s_{\lambda/\mu}(X)s_\mu(Y)$, we can use the Cauchy-Binet formula to compute the determinant of a product of two matrices and we end up with what we would expect – the real general formula is actually that

$$s_{\lambda/\nu}(X, Y) = \sum_\mu s_{\lambda/\mu}(X)s_{\mu/\nu}(Y).$$

Proposition 41

We have the following properties of skew Schur functions:

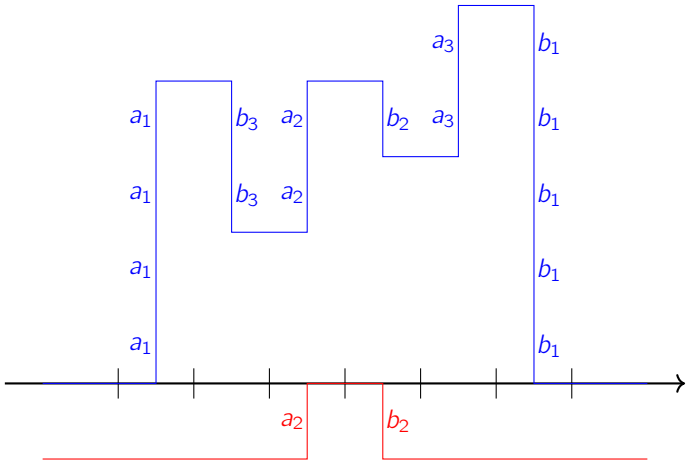
1. For $\mu = \varnothing$ the partition of length 0, we have $s_{\lambda/\varnothing} = s_\lambda$
2. We have $s_{\lambda/\mu} = 0$ unless $\mu \subseteq \lambda$, meaning that $\mu_i \le \lambda_i$ for all i. In other words, the Young diagram for μ must be entirely contained in the Young diagram for λ if we justify at the top left corner.
3. Now suppose that $\mu \subseteq \lambda$ and $\lambda/\mu = \bigsqcup_{i=1}^m \Theta^{(i)}$. (Here, we define the Young diagram λ/μ, sometimes called the **skew shape**, to be the Young diagram of λ without the boxes of the the Young diagram of μ. Then each $\Theta^{(i)}$ is a connected component of the skew shape, where touching at corners does not count as connected, and we can set the "top left corner" of each of those parts wherever we'd like.) Then, thinking of $\Theta^{(i)}$s as skew shapes, we have
$$s_{\lambda/\mu} = \prod_{i=1}^m s_{\Theta^{(i)}}.$$
4. Suppose $x = (x_1, x_2, \cdots) = (a, 0, 0, \cdots)$. Then $s_{\lambda/\mu}(x)$ is only nonzero if λ/μ is a horizontal strip, meaning that no two boxes of λ/μ are directly on top of each other. (For example, λ could be $(5,3,3)$ and μ could be $(3,3,1)$.) In such a case, $s_{\lambda/\mu}((a,0,0,\cdots)) = a^{|\lambda|-|\mu|}$.

These all come out of Jacobi–Trudi; in particular (3) has to do with the matrix of hs decomposing into diagonal blocks. Unfortunately, notice that this identity doesn't actually let us inductively compute the skew Schur functions (for example if λ/μ is just a single component) – we should really just do it via Jacobi–Trudi directly. And identity (4) is the main one we'll use.

Remark 42. *To generalize (4), we can actually use the branching rule $s_{\lambda/\nu}(x,y) = \sum_\mu s_{\lambda/\mu}(x) s_{\mu/\nu}(y)$. So for the left-hand side to be nonzero, both λ/μ and μ/ν must be horizontal strips, and then we can iterate to get the corresponding results (or to compute Schur functions in more variables by peeling one off at a time).*

Example 43

Turning back to PNG now – the example below is a copy of the example above Lemma 20 – recall that what we're interested in is the heights of the paths at a particular vertical slice.



Along any vertical slice, we have locations $(\ell_1, \ell_2, \cdots)$ of the top-most path, second-top path, and so on; we know that $\ell_1 > \ell_2 > \cdots$ and that $\ell_i = 1 - i$ for sufficiently large i (because if we go down far enough we won't have any

bumps). So to create a partition out of these, we can define

$$\lambda_i = \ell_i + i - 1,$$

which will indeed give us $\lambda_1 \geq \lambda_2 \geq \cdots$ with parts eventually zero. Basically λ_i is how much the ith path went up by at that slice; in our example above we have the Young diagrams (4), (2), (4, 1), (3), and (5) at x-locations $-2, -1, 0, 1, 2$. We will call these partitions, from left-to-right, $\lambda^{(1)}, \mu^{(1)}, \lambda^{(2)}, \mu^{(2)}, \lambda^{(3)}$.

So the point is that under this bijection, multi-layer PNG is exactly equal to a sequence of partitions

$$\varnothing \subseteq \lambda^{(1)} \supseteq \mu^{(1)} \subseteq \lambda^{(2)} \supseteq \mu^{(2)} \cdots \subseteq \lambda^{(m)} \supseteq \varnothing,$$

where m is determined by the time that we run the PNG for. We can easily compute the weight of such a configuration in terms of the partitions, since we gain an a every time we go up and a b every time we go down, and the number of steps is exactly given by the differences $|\lambda^{(i)}| - |\mu^{(i)}|$ and $|\lambda^{(i+1)}| - |\mu^{(i)}|$. Thus the contribution always looks like a power of an a_i or a b_i.

The key point now is that each $\lambda^{(i)}/\mu^{(i)}$ or $\lambda^{(i+1)}/\mu^{(i)}$ **is always a horizontal strip**, since otherwise our paths would share an edge. Thus the weight is something we can compute with Proposition 41, part (4): we have a product of one-variable Schur functions

$$W(\vec{\lambda}, \vec{\mu}) = s_{\lambda^{(1)}}(a_1) s_{\lambda^{(1)}/\mu^{(1)}}(b_m) s_{\lambda^{(2)}/\mu^{(1)}}(a_2) s_{\lambda^{(2)}/\mu^{(2)}}(b_{m-1}) \cdots s_{\lambda^{(m)}}(b_1).$$

This is an example of something called a **Schur process** – we go up and down, indexed by time in some way, and we get a product of Schur functions. But now we have a bunch of terms together, and so what helps is if we can get the total partition function by summing over all possible λs and μs. We can deal with local pieces (adjacent terms) as follows:

Proposition 44 (Skew Cauchy identity)
Fix partitions λ, ν. We have

$$\sum_{\mu} s_{\mu/\lambda}(X) s_{\mu/\nu}(Y) = \prod_{x \in X, y \in Y} \frac{1}{1 - xy} \sum_{\kappa} s_{\lambda/\kappa}(Y) s_{\nu/\kappa}(X).$$

In particular, the left-hand sum is infinite but the right-hand sum is finite.

To prove this, we can expand the skew Schur function and product of Schur functions in the Schur basis and get two sets of coefficients via

$$s_{\lambda/\mu}(x) = \sum_{\nu} c_{\mu\nu}^{\lambda} s_{\nu}(x),$$

$$s_{\mu}(x) s_{\nu}(x) = \sum_{\lambda} d_{\mu\nu}^{\lambda} s_{\lambda}(x).$$

(Notice that the indexing is a little bit different – we're expanding over ν in one case and λ in the other.) The ds are sometimes called **Littlewood-Richardson coefficients**, and in fact so are the others:

Lemma 45
We have $c_{\mu\nu}^{\lambda} = d_{\mu\nu}^{\lambda}$ for all μ, ν, λ.

Proof. Define, from the right-hand side of the Cauchy identity,

$$H(X;Y) = \prod_{x\in X, y\in Y} \frac{1}{1-xy}.$$

We want to get a bunch of products of the form $s_\mu s_\nu$, so we'll multiply various Hs together. Let Z be a new variable set, and observe that

$$\begin{aligned} H(X;Z)H(Y;Z) &= H(X\cup Y;Z) \\ &= \sum_\lambda s_\lambda(X\cup Y)s_\lambda(Z) \end{aligned}$$

by the Cauchy identity. Now we can expand out the right-hand side to get skew Schur functions, which yields

$$\sum_{\lambda,\mu} s_{\lambda/\mu}(X)s_\mu(Y)s_\lambda(Z) = \sum_{\lambda,\mu,\nu} c^\lambda_{\mu\nu} s_\nu(X)s_\mu(Y)s_\lambda(Z)$$

after using the definition of the cs. But on the other hand, we can do another computation by applying the Cauchy identity to $H(X;Z)$ and $H(Y;Z)$ separately, which gets us

$$\begin{aligned} H(X;Z)H(Y;Z) &= \sum_\nu s_\nu(X)s_\nu(Z)\sum_\mu s_\mu(Y)s_\mu(Z) \\ &= \sum_\nu s_\nu(X)\sum_\mu s_\mu(Y)\sum_\lambda d^\lambda_{\mu\nu}s_\lambda(Z) \\ &= \sum_{\lambda,\mu\nu} d^\lambda_{\mu\nu}s_\nu(X)s_\mu(Y)s_\lambda(Z) \end{aligned}$$

by using the definition of the ds, and then everything matches up. Since each set of Schur polynomials forms a basis in its own variable, the products form bases of functions that are jointly symmetric in each variable set separately; thus the coefficients must be equal. □

Notice that a priori it's not even clear that the cs are symmetric in μ and ν, but we get that as a corollary of this result.

Proof of Proposition 44. We'll add a bunch of variables and consider $H(X;Y)H(X;U)H(Z;Y)H(Z;U)$ in two ways. First of all, we can combine and then use the Cauchy identity to get

$$\begin{aligned} H(X\cup Z, Y\cup U) &= \sum_\rho s_\rho(X\cup Z)s_\rho(Y\cup U) \\ &= \boxed{\sum_{\rho,\lambda,\mu} s_{\rho/\lambda}(X)s_\lambda(Z)s_{\rho/\mu}(Y)s_\mu(U)}. \end{aligned}$$

But alternatively, we could take the product and apply Cauchy identity to the last three terms, which yields

$$\begin{aligned} &H(X,Y)\sum_{\sigma,\nu,\tau} s_\sigma(X)s_\sigma(U)s_\nu(Y)s_\nu(Z)s_\tau(Z)s_\tau(U) \\ &= H(X,Y)\sum_{\sigma,\nu,\tau,\lambda,\mu} s_\sigma(X)s_\nu(Y)d^\mu_{\sigma\tau}s_\mu(U)d^\lambda_{\nu\tau}s_\lambda(Z). \end{aligned}$$

But now the ds are equal to the corresponding cs, and we can pair $d^\mu_{\sigma\tau}$ with $s_\sigma(X)$ and $d^\lambda_{\nu\tau}$ with $s_\nu(Y)$ to simplify to

$$\boxed{H(X,Y)\sum_{\tau,\lambda,\mu} s_{\mu/\tau}(X)s_{\lambda/\tau}(Y)s_\mu(U)s_\lambda(Z)}.$$

Now taking the two expressions in boxes as variables in U and Z (fixing λ, μ) and looking at the coefficients in X and Y exactly yields the desired result. □

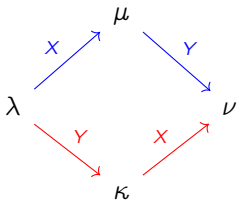
7 April 21, 2026

We introduced Schur functions and their skew analogs last time, and in particular we established the Cauchy and skew Cauchy identities and introduced the Schur process. We'll connect this to the PNG model today – we've already outlined this via the LGV lemma, but now we can use our extra machinery to make the connection more direct.

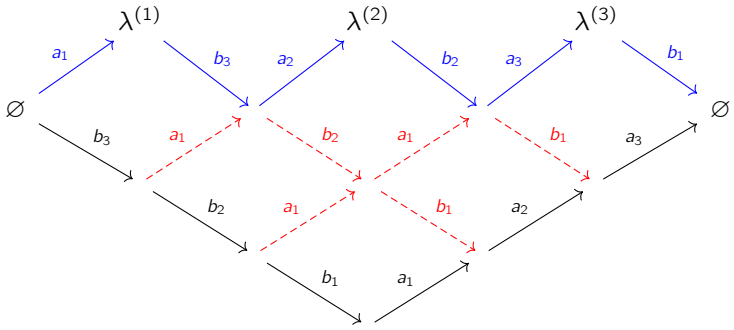
To do that, we associated to the Schur process a measure on sequences of partitions. The idea is to start from the empty partition $\varnothing$, and then we go up to $\lambda^{(1)}$ with some parameter a_1, then down to $\mu^{(1)}$ with some parameter b_n, then up to $\lambda^{(2)}$ with parameter a_2, down to $\mu^{(2)}$ with parameter b_{n-1}, and so on, until we end up with some partition $\lambda^{(n)}$ and then go down to $\varnothing$ again. The weight of such an arrangement was then proprtional to a product of skew Schur functions

$$s_{\lambda^{(1)}}(a_1) s_{\lambda^{(1)}/\mu^{(1)}}(b_n) s_{\lambda^{(2)}/\mu^{(1)}}(a_2) \cdots s_{\lambda^{(n)}}(b_1),$$

since each of these terms is just a power of the variable a_i or b_i. And we can now compute the normalizing constant as well using the skew Cauchy identity from Proposition 44 – the picture to have in mind is the following "corner-flip:"



We can thus iterate this skew Cauchy identity one step at a time to take our up-down zigzag sequence of partitions and flip all the boxes down until we just get a path going all the way down, then all the way up. Here's the diagram for $n = 3$:



Then the $\prod \frac{1}{1-xy}$ factor gives us an overall factor of

$$\frac{1}{1-a_1 b_n} \frac{1}{1-a_2 b_{n-1}} \cdots \frac{1}{1-a_n b_1} \times \frac{1}{a_1 b_{n-1}} \frac{1}{a_2 b_{n-2}} \cdots \times \cdots = \prod_{1 \le i \le j \le n} \frac{1}{1-a_i b_{n-j+1}}$$

(basically any case where the indices add to $(n+1)$ or less), and then we still have to sum over all possible partitions along the remaining path. But if we start with the empty partition and go down, we have to always be empty at every step, and similarly if we end up with the empty partition then everything along the path has to be $\varnothing$. That means the

sum of Schur functions we get is just supported on a single term, and that just ends up being 1. In short, that means our normalization factor is exactly the $\prod_{1\le i\le j\le n}\frac{1}{1-a_ib_{n-j+1}}$ from before, and so the probability in our Schur process is exactly

$$s_{\lambda^{(1)}}(a_1)s_{\lambda^{(1)}/\mu^{(1)}}(b_n)s_{\lambda^{(2)}/\mu^{(1)}}(a_2)\cdots s_{\lambda^{(n)}}(b_1)\prod_{1\le i\le j\le n}(1-a_ib_{n-j+1}).$$

(Notice that all of this would work for something like Hall-Littlewood as well, since it just relies on the skew Cauchy identity.)

Fact 46

The question now is how we can use this to study the marginal on a single slice (say at the very middle, since that's what we care about for solving LPP). That means we're curious about the law of a single partition, and we claim that will take the form

$$\mathbb{P}(\lambda)=\frac{s_\lambda(a_1,\cdots,a_{\lceil\frac{n}{2}\rceil})s_\lambda(b_1,\cdots,b_{\lceil\frac{n}{2}\rceil})}{H(\vec{a},\vec{b})},\quad H(\vec{a},\vec{b})=\prod_{i,j}(1-a_ib_j)^{-1}.$$

Indeed, this is a very similar idea to the previous proof. (We don't need to track the constants as long as we end up with something proportional to $s_\lambda(a_1,\cdots,a_n)s_\lambda(b_1,\cdots,b_n)$, since we've already established the normalizing factor by the Cauchy identity.) We want to sum over everything except $\lambda^{(m)}$, so we flip the diagram down as much as possible while keeping that in place. The result is then the furthest possible dip down then up from $\varnothing$ to $\lambda^{(m)}$, and then again the further possible dip down then up from $\lambda^{(m)}$ to $\varnothing$; we then have to sum over all sequences of partitions that go up from $\varnothing$ to $\lambda^{(m)}$ using the branching rule, yielding the $s_\lambda(a_1,\cdots,a_n)$ term, as well as the same going down from $\lambda^{(m)}$ to $\varnothing$, yielding the $s_\lambda(b_1,\cdots,b_n)$ term. (And also remember that the Schur polynomials are symmetric, so it doesn't matter how we order them.)

Thus the full PNG is described by a Schur process, and the marginal on a single partition is a Schur measure. That means we'll want to study this Schur measure more carefully, and for that we'll need to study **determinantal point processes**.

Fact 47

Before we do this, it's worthwhile to think about some structural properties of measures that arise in this Schur process manner, where we have products of skew Schur polynomials. In our case, it's fairly transparent that things are nonnegative, since the Schur function evaluated at a single variable is just some power. But those Schur functions could take in more general variables, and we might want to know what is allowed to still guarantee that we get a probability measure (nonnegativity). This is called **Schur positivity**.

Definition 48

Let $\rho:\mathrm{Sym}\to\mathbb{C}$ be a homomorphism. We say that ρ is **Schur positive** if $\rho(s_\lambda)$ (also written $s_\lambda(\rho)$) is nonnegative for all λ.

(ρ corresponds to the character of some infinite-dimensional symmetric group.) Note that this is different from the "positivity of coefficients" notion that is seen in other (non-probabilistic) settings.

- For example, ρ might be the **evaluation map** where we set each x_i equal to some a_i, and if a_is are all nonnegative then actually we get a Schur positive homomorphism by the branching rule (since $s_\lambda(a_1,\cdots,a_r)=$

$\sum_\mu s_{\lambda/\mu}(a_r)s_\mu(a_1,\cdots,a_{r-1})$, so we can induct on the number of boxes on the partition and use that $s_{\lambda/\mu}(a_r)$ is a power of a_r, hence nonnegative).

- However, there are other homomorphisms as well; we'll give one example now and another one later. In the **dual evaluation**, we fix parameters $a_1,\cdots,a_r \geq 0$ and set $\rho(s_\lambda) = s_{\lambda'}(a_1,\cdots,a_r)$. Such a ρ is Schur positive by the same branching rule, but what's not obvious is that this is an algebra homomorphism. But remember that we have the map $\omega : h_n \leftrightarrow e_n$ which exactly swaps s_λ to $s_{\lambda'}$, so in fact ρ is exactly ω followed by an evaluation.

Fact 49

If ρ is Schur positive, then it is also nonnegative on skew Schur polynomials, meaning that $\rho(s_{\lambda/\mu}) \geq 0$. One "hammer" way of proving this is that $s_{\lambda/\mu} = \sum c^\lambda_{\nu\mu} s_\nu$ for Littlewood-Richardson coefficients $c^\lambda_{\nu\mu}$ which are actually nonnegative; we won't prove this here though. But the point is that not only is the Schur measure nonnegative, so is the full Schur process.

With these two examples in mind, the next natural question to ask is whether there's a full classification:

Theorem 50 (Edrei 1953, Thoma 1964)

The Schur positive homomorphisms ρ are parametrized as follows. Fix $\vec{\alpha} = (\alpha_1,\alpha_2,\cdots)$, $\vec{\beta} = (\beta_1,\beta_2,\cdots)$ nonnegative sequences of real numbers such that $\sum\alpha_i, \sum\beta_i < \infty$, as well as $\gamma \geq 0$. Then any Schur positive homomorphism is described by its action on an algebraic basis, and we'll do so on the powersums $p_k = \sum_{i=1}^\infty x_i^k$. Then ρ always takes the form

$$\rho(p_1) = \gamma + \sum_i (\alpha_i + \beta_i),$$

and for $k \geq 2$

$$\rho(p_k) = \sum_{i=1}^\infty \alpha_i^k + (-1)^{k-1}\beta_i^k.$$

Intuitively, the αs should be thought of as coming from evaluation, the βs from dual evaluation, and γ as the limit of $\alpha = (\frac{\gamma}{N},\cdots\frac{\gamma}{N},0)$ as $N\to\infty$.

Recall that we've previously written out the generating function $h(z) = \sum_{k=0}^\infty h_k z^k = \frac{1}{\prod_{i=1}^\infty (1-x_i z)}$, and now we can ask how h (which governs the Cauchy identity) looks under an arbitrary ρ, which would be important for computing the normalization factor. It turns out to have a fairly clean form:

$$\rho(h(z)) = e^{\gamma z}\prod_{i=1}^\infty \frac{1+\beta_i z}{1-\alpha_i z}.$$

(Indeed, just having the αs gives us the Cauchy identity, and just having the βs gives us the dual Cauchy identity. And the $e^{\gamma z}$ matches up with the limit of $\alpha = (\frac{\gamma}{N},\cdots\frac{\gamma}{N},0)$ at the power series level too.)

These homomorphisms are also sometimes called **specializations**. The pure α specializations are what we needed for multi-layer PNG, and all of our a_is and b_js were αs (evaluations); the relevant graph for non-intersecting paths was the $\mathbb{Z}^2$ grid with as and bs the vertical edge weights. In general, a Schur measure would take the form $Z^{-1}s_\lambda(\rho_1)s_\lambda(\rho_2)$, and a Schur process would just replace all of the a_is and b_i evaluations with different specializations too.

So for context, it's worth also mentioning what we get if we do a pure β specialization instead for all ρ_i. We'll end up with nonintersecting paths on a different graph, which is the $\mathbb{Z}^2$ lattice with up edges replaced with up-right edges. Horizontal edges again have weight 1, and the diagonal edges have weights given by the βs. Since time is still indexed by vertical slices, the qualitative difference here is that we can only have one of each β, whereas we could

collect arbitrarily many as in the PNG model. (This is connected to us taking elementary symmetric functions, which only allow degree 1 per variable.). The growth process we get is a version of TASEP, and if we mix αs and βs we actually get domino tilings of the Aztec diamond.

Example 51

We'll conclude with a connection to the **longest increasing subsequence** problem. Recall that the top curve of multi-layer PNG corresponds to the point-to-point last passage times along some set of points, so for example $\mathrm{LPP}((1,1) \to (N,N))$ is the height of the top path in multi-layer PNG at the middle at some time. If we look now just at LPP, we can set all a_is and b_is equal to some κ, meaning that each random variable in the integer grid is geometric with parameter κ^2.

Taking $\kappa \to 0$ simultaneously as the dimension of the grid N grows, specifically $\kappa = \frac{\gamma}{N}$, means that the probability of a box being nonzero is roughly $\frac{\gamma^2}{N^2}$, and the probability of it being at least 2 is roughly $\frac{\gamma^3}{N^3}$. Since we have N^2 total boxes, the latter happens with shrinking probability, and the 1s form a (continuous) Poisson point process of rate γ^2 as $N \to \infty$. Call the resulting points $(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)$ (for N an exponential random variable).

The last passage time is then determined by the up-right path connecting bottom-left to top-right which passes through the maximum number of points in this process. But the LPP value only depends on the ordering of the coordinates, not their exact values! So if we order the points so that $x_1 < \cdots < x_N$, what matters is the permutation of the values $y_1, \cdots, y_N$, and in fact the path is exactly the longest increasing subsequence of that permutation. Thus a certain degeneration of our process gives us the distribution of the LIS for a uniformly random permutation, except of slightly random size roughly γ^2.

8 April 23, 2026

We'll be getting into some general probabilistic aspects today. So far, we've discussed Schur measures and processes (the former of which is the marginal of the later) and seen that the multi-layer PNG is a Schur process with top curve given by the Schur measure. The Jacobi–Trudi formula then tells us that the law of the top curve takes the form $Z^{-1} \det[h_{\lambda_i - i+j}(x)]_{i,j} \det[h_{\lambda_i - i+j}(y)]_{i,j}$ (where x encodes the a_is and y encodes the b_is).

This measure is an example of something known as a **biorthogonal ensemble** (a product of determinants of a certain form), and in this class and the next we'll give some examples of point processes in general and define what it means to be a "determinantal" point process (and a "biorthogonal" ensemble), as well as prove some things about them. We'll then be able to specialize to our particular instance and study PNG from there.

Definition 52

Let X be a topological space which is locally compact and separable (for our purposes, we can really just think of $X = \mathbb{Z}^d$ or $\mathbb{R}^d$ or open or closed or finite subsets of them). From X, we can form the **configuration space**

$$\mathrm{Conf}(X) = \text{discrete (locally finite) subsets of } X,$$

meaning that around any point, there is a small compact neighborhood around it whose intersection with the subset is finite. Such subsets are also sometimes called **point configurations**.

We'll be interested in probability measures on $\mathrm{Conf}(X)$; keep in mind that we can think of Schur measures as being of this type by letting X be $\mathbb{Z}$ and bijecting partitions λ to subsets $(\lambda_1 - 1, \lambda_2 - 2, \cdots) \subseteq \mathbb{Z}$ with the property that all

but finitely many entries are of the form $-n, -n-1, \cdots$, going down to $-\infty$. So the Schur measure gives a probability measure on λ, which can then be bijected to a measure on $\mathrm{Conf}(\mathbb{Z})$. The idea is somehow that in the limit, we'll take a limit of the λs and study the edge; we'll scale things down so that the scaling becomes of finite size and so this will approach some point process on $\mathbb{R}$.

Fact 53

Since X is a topological space, we can assume it has a Borel structure; call a relatively compact Borel subset $A \subseteq X$ a **window**. For any window A and point configuration $S \in \mathrm{Conf}(X)$, define $N_A(S) = |A \cap S|$; N_A is then a function $\mathrm{Conf}(X) \to \mathbb{Z}$; these then generate a Borel structure on $\mathrm{Conf}(X)$ (the minimal collection of open sets making all of these functions continuous) and so we can talk about probability on it.

The point of this definition is really just so that we can take expectations of random variables like $N_A(S)$; all we need to check is that the inverse of any singleton in $\mathbb{Z}$ under N_A is always open. We will also call elements of X **particles**; it's often common to represent these configurations via **Maya diagrams** by drawing balls in boxes at the appropriate locations.

Example 54

Recall that the **Poisson point process** with intensity μ (for μ some measure on X) has the following two properties:

- For any A, $\mathbb{P}(N_A = n) = e^{-\mu(A)} \frac{\mu(A)^n}{n!}$ (that is, N_A is a Poisson random variable of parameter $\mu(A)$).
- If $A_1, A_2, \cdots, A_k$ are disjoint, then N_{A_i} are mutually independent.

This is a special case of our general description – for example if $X = \mathbb{R}$ and μ is Lebesgue measure, then we can place M particles independently and uniformly at random on $[-\frac{M}{2}, \frac{M}{2}]$ and let $M \to \infty$ and it will verify our two properties. Indeed, for example (calling the point process P)

$$\mathbb{P}(|[0,1] \cap P| = k) = \binom{M}{k} \left(\frac{1}{M}\right)^k \left(1 - \frac{1}{M}\right)^{M-k} \xrightarrow{M \to \infty} \frac{1}{k!} \cdot \frac{1}{e}$$

because the count of how many lie in a given interval is binomial, and this is exactly the Poisson of parameter 1. So in this way we really do have a "point process" in that we are randomly placing points, but the way we're defining these processes is really by specifying the random variables N_A.

Example 55

For $X = \mathbb{Z}$, we can consider the **Bernoulli point process** P each $i \in \mathbb{Z}$ is independently in P with some probability $\rho \in [0,1]$ (or ρ_i depending on i). Then if we let $\rho \sim \varepsilon$ and consider an interval of size ε^{-1}, then send $\varepsilon \to 0$, we will again get a Poisson point process after rescaling space down.

The reason for mentioning this is that we have a natural limit transition turning a discrete process into a continuous one, and we'll see somewhat more advanced versions of that later on.

Definition 56

Let X be discrete. For any finite subset A of X, we can define the **correlation function** $\rho(A) = \mathbb{P}(A \in S)$, where S is sampled under the random process.

For example for $\lambda \mapsto (\lambda_1 - 1, \lambda_2 - 2, \cdots)$, the correlation functions essentially tell us the probability that some given locations are in the subset. We'll really be interested in whether some particular point is the top one in our partition, but that'll come later.

Fact 57

For X discrete, the correlation functions determine the random point process by inclusion-exclusion, since we'll know the joint law of $A \cap S$ over all A.

Definition 58

For general X, the ***n*th correlation measure** ρ_n is a measure on X^n which satisfies

$$\int_{X^n} f(\vec{x}) d\rho_n(\vec{X}) = \mathbb{E}\left[\sum_{\substack{x_1,\cdots,x_n \\ \text{distinct points in } P}} f(x_1, x_2, \cdots, x_n)\right] x$$

for any compactly supported f (if the right-hand side exists, then this determines ρ_n).

These two definitions actually don't match up exactly, but they are just rescaled versions of each other, and we'll use the latter renormalization going forward. It can be seen (though this is harder to see) that under certain growth conditions (specifically $\rho_n(\vec{X}) \leq n^{2n} C^n$) that these correlation measures determine P, which can be thought of as some version of the moment problem.

Fact 59

If ρ_n has a density, its density will be denoted $\rho_n(x_1, \cdots, x_n)$, and we call the density the **correlation function**. The dream for any point process is to compute these correlation functions, since then we can compute any probabilities we want to know. We can think of $\rho_n(x_1, \cdots, x_n)$ as the probability that there is a particle in $[x_i, x_i + dx_i]$ for all i.

Note that ρ_n is not a probability measure in general (and for example, the total integral of ρ_1 would tell us something about the sparsity of the point process overall).

Proposition 60

Suppose we have a point process P which always has N particles on X, and suppose P has a density $p(x_1, \cdots, x_N)$. We'll assume without loss of generality that p is symmetric (that is, "the particles are indistinguishable"). Then the correlation functions are given by essentially marginalizing out the density:

$$\rho_n(x_1, \cdots, x_n) = \frac{N!}{(N-n)!} \int_{X^{N-n}} p(x_1, \cdots, x_N) dx_{n+1} \cdots dx_N.$$

Proof. We have, by writing out the definition,

$$\int_{X^N} \sum_{\substack{i_1,\cdots,i_n \\ \text{distinct in } P}} f(x_{i_1}, \cdots, x_{i_n}) p(dx) = N(N-1)\cdots(N-n+1) \int_{X^N} f(x_1, \cdots, x_n) p(x_1, \cdots, x_N) d\vec{x}$$

by symmetry because there are N ways to choose which particle is i_1, $N-1$ ways to choose i_2, and so on. Then we can separate the integral over X^n and the integral over the remaining X^{N-n} to get

$$\int_{X^n} \left(\frac{N!}{(N-n)!} \int_{X^{N-n}} p(x_1, \cdots, x_N) dx_{n+1} \cdots dx_N \right) f(x_1, \cdots, x_n) dx_1 \cdots dx_n,$$

so by pattern matching with the definition of the nth correlation measure we see that the blue part must be ρ_n. □

An example of this setup is the case where we have a random $N \times N$ Hermitian matrix whose entries are complex Gaussians (in the sense of Example 4). Then the eigenvalues of the matrix always forms a point process with N particles, and in fact the density is given by

$$p(x_1, \cdots, x_N) = \prod_{i<j} (x_i - x_j)^2 \prod_{i=1}^{N} e^{-x_i^2/(2N)}.$$

So if we wanted to know the probability of a specific eigenvalue existing, we'd have to compute $N-1$ integrals, which is probably not very feasible.

For other examples to keep in mind, for the Bernoulli point process of probability ρ we can convince ourselves that $p(A) = \rho^{|A|}$ for any set A (which means the correlation measure $\rho_n(\vec{x})$ is ρ^n as long as $\vec{x}$ has no repeats and 0 otherwise), and for the Poisson point process we have $\rho_n = \mu \otimes \mu \otimes \cdots \otimes \mu$ for μ the intensity of the underlying measure.

Proposition 61

We have for any set B that

$$\mathbb{P}(\text{no particles in } P \text{ in some set } B) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \int_{B^n} \rho_n(x_1, \cdots, x_n) d\vec{x}.$$

For example if $B = [x, \infty)$, this tells us the probability that the largest particle is at most x, and so this is what we want to use to go from correlation functions to facts about partitions. And this formula is useful because we have a rapidly decaying sum depending only on finite pieces.

Proof. For simplicity of notation, take X to be discrete (but the same argument applies more generally but sums are replaced with integrals). We have

$$\begin{aligned}\mathbb{P}(|B \cap P| = 0) &= \mathbb{E}\left[\prod_{x \in P} (1 - \mathbf{1}\{x \in B\})\right] \\ &= 1 - \sum_{x \in P} \mathbb{E}[\mathbf{1}\{x \in B\}] + \mathbb{E}\left[\sum_{x_1 \neq x_2 \text{ ordered } \in P} \mathbf{1}\{x_1, x_2 \in B\}\right] - \cdots\end{aligned}$$

by expanding out and using linearity of expectation. And each of these terms can be written in terms of correlation functions; we can rewrite this as

$$\sum_{n \geq 0} (-1)^n \sum_{i_1 < \cdots < i_n} \mathbb{E}\left[\prod_{k=1}^{n} \mathbf{1}_{x_{i_k} \in B}\right] = \sum_{n \geq 0} \frac{(-1)^n}{n!} \sum_{i_1, \cdots, i_n \text{ distinct}} \mathbb{E}\left[\prod_{k=1}^{n} \mathbf{1}_{x_{i_k} \in B}\right],$$

and now the inner sum is exactly the integral of ρ_n over B^n, as desired. □

We'll now foreshadow what determinantal point processes are; a priori there's no reason to be able to compute ρ_n, but sometimes we do get something useful:

Definition 62

A point process is called a **determinantal point process** if there exists a **correlation kernel** $K : X^2 \to \mathbb{R}$ (or $\mathbb{C}$) such that

$$\rho_n(x_1, \cdots, x_n) = \det\left[K(x_i, x_j)\right]_{1 \le i,j \le n}$$

for all correlation functions ρ_n. (Note that K is not symmetric in general.)

The idea is that we've encoded everything into a single matrix K, and then we can compute everything we care about using minors of it. And what's nice is that if we want to take a limit now, all we need to do is take an appropriate limit of K instead of all of the ρ_ns. We'll see some examples next time!

9 April 28, 2026

We'll basically spend today on a computation. We defined the notion of a determinantal point process last time, which is a point process whose correlation functions ρ are given by determinants of some correlation kernel $K : X^2 \to \mathbb{R}$; specifically

$$\rho(x_1, \cdots, x_n) = \det\left[K(x_i, x_j)\right]_{1 \le i,j \le n}.$$

This is fundamental because everything we want to know is contained in these ρs, so if we can compress everything into a single matrix, that's good. Today, we'll show that the (general) Schur measure is a determinantal point process and compute its kernel.

Recall that PNG is described by a Schur measure; specifically, the top coordinate λ_1 of λ gives the distribution of the height of PNG. So we might ask for a clean description of the top point of a determinantal point process, and we can just plug into Proposition 61:

Lemma 63

Consider a determinantal point process P on X. For any $B \subseteq X$, we have

$$\begin{aligned}\mathbb{P}(P \cap B = \varnothing) &= \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \int_{B^n} \rho_n(x_1, \cdots, x_n) dx_1 \cdots dx_n \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \int_{B^n} \det[K(x_i, x_j)] dx_1 \cdots dx_n.\end{aligned}$$

We sometimes write this as the **Fredholm determinant** $\det[1 - K]_{L^2(B)}$.

If B is finite, then expanding the determinant out exactly gives us the previous expression, and in general we can think of this as the definition of the Fredholm determinant (what we really mean by the determinant of an infinite-dimensional object is the limit of finite-dimensional pieces). This will satisfy the "usual properties" of determinants as long as everything converges; for example it will be true that

$$\det[1 - AB] = \det[1 - BA].$$

Remark 64. *In the case where $X = \mathbb{R}$ and $B = [x, \infty)$, this tells us that*

$$\mathbb{P}(\max P \le x) = \det[1 - K]_{L^2[x,\infty)},$$

and so we can pin down the top point of our process by computing the right-hand side. And in fact the Tracy–Widom distribution is even defined by such a formula for a certain choice of K.

So what we'll do with the Schur measure is show that it's determinantal with a certain K, and that'll give us Tracy–Widom because of this last lemma. We can think about this for just Schur measures specifically, but to keep the computation relatively feasible it will be better to generalize to a more general class to understand what's going on.

Definition 65

A probability measure $\mathbb{P}$ on X^N is called a **biorthogonal ensemble** if there exist functions (ϕ_i, ψ_i) for $1 \leq i \leq N$ such that the density of $\mathbb{P}$ is, with respect to some reference measure μ,

$$\mathbb{P}(dx_1 \cdots dx_N) \sim \det\left[\phi_i(x_j)\right]_{1\leq i,j\leq N} \det\left[\psi_i(x_j)\right]_{i,j\in[1,N]} \prod_{i=1}^{n} \mu(dx_i).$$

For example, going back to the Schur measure, we have $X = \mathbb{Z}$, N the length of the partition (we can think of this as the number of variables a_i, b_i, or equivalently parametrized by the time we've run the PNG for), and μ the counting measure. The slightly confusing thing is that usually we write the Schur measure as proportional to $s_\lambda(x)s_\lambda(y)$, but if we write it out this would be $\det[x_i^{\lambda_j - j + N}] \det[y_i^{\lambda_j - j + N}]$ (divided by some normalization Vandermonde factor). So actually in this case we will have $\phi_i(a) = x_i^a$ and $\psi_i(a) = y_i^a$ exponential functions.

There are two other measures that also can be described by biorthogonal ensembles, which turn out to be special cases in some sense:

- The **circular unitary ensemble (CUE)** describes the eigenvalues of a random unitary matrix on $U(N)$ under Haar measure.
- The **Gaussian unitary ensemble (GUE)** describes the eigenvalues of a random Hermitian matrix with complex Gaussian entries.

We won't go through the derivations, but if we work out the calculations the CUE eigenvalues (which will all lie on the unit circle, hence are parametrized by angles) have density given by

$$\mathbb{P}(d\theta_1 \cdots d\theta_N) \sim \prod_{i<j} |z_i - z_j|^2 d\theta$$

for $z_j = e^{i\theta_j}$, and similarly the GUE eigenvalues (which will all be real) have density

$$\mathbb{P}(d\lambda_1 \cdots d\lambda_N) \sim \prod_{i<j} |\lambda_i - \lambda_j|^2 \prod_{i=1}^{N} e^{-\lambda_i^2} d\lambda_i.$$

In the latter case, notice that we have a Vandermonde appearing in the first factor twice, and so the GUE density can also be written as a product of two determinants $\det[\lambda_i^j]_{1\leq i,j\leq N} \det[e^{-\lambda_i^2}\lambda_i^j]_{1\leq i,j\leq N}$. (And the former case is similarly just putting in exponentials in place of the λ_is.)

Proposition 66

Any biorthogonal ensemble is a determinantal point process with kernel given by

$$K(x,y) = \sum_{1\le i,j\le N} (G^{-T})_{ij}\phi_i(x)\psi_j(y),$$

where $G \in \mathbb{R}^{N\times N}$ is the Gram matrix where $G_{ij} = \int \phi_i(x)\psi_j(x)d\mu(x)$, and $G^{-T} = (G^{-1})^T$.

This explains the name "biorthogonal:" if the ϕs are orthogonal with respect to the ψs, then G becomes the identity and the kernel becomes very explicit and we just take the sum $\phi_i\psi_i$. But it was realized quickly that the whole argument goes through in general too.

We'll prove this in generality and then specialize to the Schur measure:

Proof. We'll omit the reference measure μ everywhere for simplicity of notation. Whenever we have a concrete measure, the first thing we want to do is find the normalization constant; thus we need to compute

$$\begin{aligned}
&\int_{X^N} \det\left[\phi_i(x_j)\right]_{1\le i,j\le N} \det\left[\psi_i(x_j)\right]_{i,j\in[1,N]}\, dx \\
&= \int_{X^N} \sum_{\sigma,\tau} \mathrm{sgn}(\sigma)\mathrm{sgn}(\tau) \prod \phi_{\sigma(i)}(x_i)\psi_{\tau(i)}(x_i)dx \\
&= \sum_{\sigma,\tau} \left(\prod_{i=1}^N \int_X \phi_{\sigma(i)}(x_i)\psi_{\tau(i)}(x_i)dx_i\right) \mathrm{sgn}(\sigma\tau)
\end{aligned}$$

by opening up the determinants and then splitting up the independent integrals. But now by definition of the Gram matrix this is exactly

$$\sum_{\sigma,\tau} \left(\prod_{i=1}^N G_{\sigma(i)\tau(i)}\right) \mathrm{sgn}(\sigma\tau),$$

and now this is exactly the determinant of G except with an extra factor of $N!$ because σ,τ are redundant (for any σ there's one choice of τ that gives us the desired permutation). Thus $Z = N!\det G$.

Now the nth correlation function is given (using Proposition 60) by

$$\begin{aligned}
\rho_n(x_1,\cdots,x_n) &= \frac{N!}{(N-n)!}\int_{X^{N-n}} \mathbb{P}(dx_1\cdots dx_N)dx_{n+1}\cdots dx_N \\
&= \frac{N!}{(N-n)!}\cdot\frac{1}{N!\det G}\int_{X^N-n} \det[\phi_i(x_j)]\det[\psi_i(x_j)]dx_{n+1}\cdots dx_N
\end{aligned}$$

by plugging in the value of $\mathbb{P}$ including the normalization constant. Now we want to swap out ϕ and ψ with things that are "actually orthogonal:" let A and B be $n\times n$ matrices such that $AGB^T = \mathrm{id}$, and then use appropriate linear combinations

$$\Phi_k = (A\phi)_k = \sum_{\ell=1}^N A_{k\ell}\phi_\ell, \quad \Psi_k = (B\psi)_k = \sum_{\ell=1}^N B_{k\ell}\psi_\ell.$$

We can then observe that $(\Phi_k), (\Psi_k)$ are biorthogonal, in the sense that

$$\begin{aligned}\int_X \Phi_i(x)\Psi_j(x)dx &= \sum_{\ell,\ell'} A_{i\ell}B_{j\ell'}\int_X \phi_\ell(x)\psi_{\ell'}(x)dx \\ &= \sum_{\ell,\ell'} A_{i\ell}G_{\ell\ell'}B_{j\ell'} \\ &= (AGB^T)_{ij} \\ &= \delta_{ij}.\end{aligned}$$

So now we'll express ϕ, ψ in terms of Φ, Ψ as well; we have (notice that because G is a fixed matrix not depending on xs, A and B don't depend on the xs either)

$$\det[\phi_i(x_j)] = \det[\Phi_i(x_j)]\det[A^{-1}], \quad \det[\psi_i(x_j)] = \det[\Psi_i(x_j)]\det[B^{-1}].$$

Plugging back in, we thus have

$$\rho_n(x_1, \cdots, x_n) = \frac{1}{(N-n)!}\frac{1}{(\det G)(\det A)(\det B)}\int_{X^{N-n}} \det[\Phi_i(x_j)]\det[\Psi_i(x_j)]dx_{n+1}\cdots x_N,$$

and furthermore $\det G \det A \det B = \det(AGB^T) = 1$ so we can remove that factor.

Now to get a sense of what's happening, we'll first illustrate the proof in the case $n = 1$. We then want to compute

$$\rho_1(x_1) = \frac{1}{(N-1)!}\int_{X^{N-1}} \det[\Phi_i(x_j)]\det[\Psi_i(x_j)]dx_2\cdots dx_N$$

by again opening the determinants to get

$$\frac{1}{(N-1)!}\sum_{\sigma,\tau}\int_{X^{N-1}}\left(\prod_{i=2}^N \Phi_{\sigma(i)}(x_i)\Psi_{\tau_i}(x_i)dx_i\right)\Phi_{\sigma(1)}(x_1)\Psi_{\tau(1)}(x_1)\mathrm{sgn}(\sigma\tau).$$

But now (because the different x_is are decoupled) biorthogonality forces $\sigma(i) = \tau(i)$ for all $i > 1$, and therefore (this is the only special thing about $n = 1$) this also forces $\sigma(1) = \tau(1)$; call this number k. Removing the unnecessary terms, we thus have

$$\begin{aligned}\rho_1(x_1) &= \frac{1}{(N-1)!}\sum_{k=1}^N \sum_{\substack{\sigma,\tau: \\ \sigma(2,\cdots,N)=\tau(2,\cdots,N), \\ \sigma(1)=k}} \Phi_k(x_1)\Psi_k(x_1) \\ &= \frac{1}{(N-1)!}\sum_{k=1}^N (N-1)!\Phi_k(x_1)\Psi_k(x_1) \\ &= \sum_{k=1}^N \Phi_k(x_1)\Psi_k(x_1),\end{aligned}$$

and now we can re-express everything in terms of ϕ and ψ: writing the linear combinations back in, this yields

$$\begin{aligned}
&\sum_{k=1}^{N} \sum_{\ell,\ell'} A_{k\ell} B_{k\ell'} \phi_\ell(x_1) \psi_{\ell'}(x_1) \\
&= \sum_{\ell,\ell'} \phi_\ell(x_1) \psi_{\ell'}(x_1) \sum_{k=1}^{N} A_{k\ell} B_{k\ell'} \\
&= \sum_{\ell,\ell'} \phi_\ell(x_1) \psi_{\ell'}(x_1) (A^T B)_{\ell\ell'} \\
&= \sum_{\ell,\ell'} \phi_\ell(x_1) \psi_{\ell'}(x_1) (G^{-T})_{\ell\ell'}
\end{aligned}$$

since $AGB^T = \mathrm{id} \implies G = A^{-1}(B^T)^{-1} \implies G^{-1} = B^T A \implies G^{-T} = A^T B$. That gives us the correct "on-diagonal" kernel term $K(x,x)$, which is all we need when computing the first correlation function for a determinantal point process.

Now for the general n case, we use basically the same proof but with a bit heavier notation. We have, by opening up determinants again,

$$\rho_n(x_1, \cdots, x_n) = \frac{1}{(N-n)!} \sum_{\sigma,\tau} \mathrm{sgn}(\sigma\tau) \Phi_{\sigma(i)}(x_i) \Psi_{\tau(i)}(x_i) \int_{X^{N-n}} \prod_{i=n+1}^{N} \Phi_{\sigma(i)}(x_i) \Psi_{\tau(i)}(x_i) dx_i,$$

but this time we only fix $\sigma = \tau$ on the last $N-n$ coordinates, but now σ and τ can be different on the first n. Instead of a single k, we now have $j_1, \cdots, j_n$, and we can express this correlation function as

$$\begin{aligned}
\rho_n(x_1, \cdots, x_n) &= \frac{1}{(N-n)!} \sum_{j_1, \cdots j_n \text{ distinct}} \sum_{\substack{\sigma,\tau: \\ \sigma(n+1,\cdots,N) = \tau(n+1,\cdots,N), \\ \sigma(1,\cdots,n) = (j_1,\cdots j_n)}} \mathrm{sgn}(\sigma\tau) \prod_{i=1}^{n} \Phi_{j_i}(x_i) \Psi_{\tau(i)}(x_i) \\
&= \sum_{j_1 < \cdots < j_n} \det \Phi^{j_1,\cdots j_n} \det \Psi^{j_1,\cdots j_n}.
\end{aligned}$$

In this last step, the redundancy of σ and τ coming from the last $N-n$ coordinates cancels out the $(N-n)!$ factors, the sum over orderings of the j_is gives us the determinant for Φ, and then the remaining sum over τs on those coordinates gives us the determinant for Ψ. (We should think of these determinants as submatrices of $[\Phi_j(x_i)]$ and $[\Psi_j(x_i)]$ formed by columns of index $j_1, \cdots, j_n$.)

So now we use the **Cauchy-Binet formula**, which states that for $n \times N$ matrix A and an $N \times n$ matrix B, we have $\det(AB) = \sum_{S \subseteq [N], |S| = n} \det[A^S] \det[B^S]$, where A^S and B^S are the square matrices obtained by taking columns and rows indexed by S, respectively. That tells us that our expression is exactly $\det[\Phi\Psi^T]$, where Φ is the $n \times N$ matrix with entries $\Phi_j(x_i)$ and similar for Ψ. So this is what we wanted: we have a determinant of the form

$$\det\left[\sum_{k=1}^{N} \Phi_k(x_i) \Psi_K(x_j)\right]_{i,j},$$

and if we rewrite Φ and Ψ in terms of ϕ and ψ again we see that this (i,j) entry will be exactly $\sum_{\ell,\ell'} (G^{-}T)_{\ell\ell'} \phi_\ell(x) \psi_{\ell'}(x_j)$, which is the $K(x_i, x_j)$ we were looking for. $\square$

We'll specialize this next time to the case of Schur measures.

10 April 30, 2026

We'll compute the correlation kernel for the Schur measure today, using the computation from last time. We know that the kernel of a biorthogonal ensemble (where the density is proportional to the product of two determinants) is given by the explicit formula in Proposition 66, and in particular this formula simplifies in the case where the ϕs are orthogonal with respect to the ψs.

In the case of the Schur measure, we know that up to some overall normalizing constant, we have

$$s_\lambda(x)s_\lambda(y) \sim \det\left[x_i^{\lambda_j-j+n}\right]\det\left[y_i^{\lambda_j-j+n}\right],$$

where our maps ϕ_i and ψ_i will be the exponential maps $\phi_i(a) = x_i^a$ and $\psi_i(a) = y_i^a$. So we can plug into our formula for $K(a,b)$ (we'll use a,b for the kernel variables now to avoid confusion with x,y from the Schur functions). First we compute the Gram matrix: since we're working on $\mathbb{Z}_{\geq 0}$ with the counting measure μ, we have

$$\begin{aligned} G_{ij} &= \sum_a \phi_i(a)\psi_j(a) \\ &= \sum_a (x_iy_j)^a \\ &= \frac{1}{1-x_iy_j}. \end{aligned}$$

(This is called the "Cauchy matrix.") We now want to invert (and transpose) this matrix, and we'll use the Crámer rule to do the inversion: we have

$$(G^{-T})_{k\ell} = \frac{C(k,\ell)}{\det G},$$

where C is the cofactor matrix given by

$$C(k,\ell) = (-1)^{k+\ell}\det G^{(k,\ell)}, \quad G^{(k,\ell)} = G \text{ with row } k \text{ and column } \ell \text{ removed.}$$

So to compute the entries of G^{-T}, we essentially need to compute determinants of two Cauchy matrices. But we already did this (see the calculation above Remark 33); we know that

$$\det G = \frac{\prod_{i<j}(x_i-x_j)(y_i-y_j)}{\prod_{i,j}(1-x_iy_j)}$$

by matching up total degrees and including necessary factors by symmetry, and so because $G^{(k,\ell)}$ is the same as G but with x_k and y_ℓ "cut out," we have

$$\begin{aligned} (G^{-T})_{k\ell} &= \frac{(-1)^{k+\ell}\det G^{(k,\ell)}}{\det G} \\ &= (-1)^{k+\ell}\frac{\prod_{i<j,i,j\neq k}(x_i-x_j)\prod_{i<j,i,j\neq \ell}(y_i-y_j)}{\prod_{i\neq k,j\neq\ell}(1-x_iy_j)}\frac{\prod_{i,j}(1-x_iy_j)}{\prod_{i<j}(x_i-x_j)(y_i-y_j)} \\ &= \frac{\prod_{j=1}^N(1-x_jy_\ell)(1-x_ky_j)}{(1-x_ky_\ell)\prod_{j\neq k}(x_j-x_k)\prod_{j\neq\ell}(y_j-y_\ell)} \end{aligned}$$

(here the $(1-x_ky_\ell)$ is to avoid double counting in the numerator, and the factor of $(-1)^{k+\ell}$ cancels out with the

ordering of x_k and y_ℓ with the other). So now we have our Gram matrix and can plug into Proposition 66 to get

$$K(a,b) = \sum_{k,\ell} (G^{-T})_{k\ell} \phi_k(a) \psi_\ell(b)$$
$$= \sum_{k,\ell} (x_k)^a (y_\ell)^b \frac{\prod_{j=1}^N (1-x_j y_\ell)(1-x_k y_j)}{(1-x_k y_\ell)\prod_{j\neq k}(x_k - x_j)\prod_{j\neq \ell}(y_\ell - y_j)}.$$

We'll now massage this a bit by writing it as a contour integral:

$$K(a,b) = \frac{1}{(2\pi i)^2} \oint \oint \prod_{j=1}^N \frac{(1-x_j v)(1-y_j z)}{(z-x_j)(v-y_j)} \frac{z^a v^b}{1-zv} dz dv,$$

where the z-contour contains the x_js, the v-contour contains the y_js, and no other poles arise. Here the xs and ys are positive real numbers and we know that the products $x_i y_j$ are all less than 1 (by assumption when we set up the Schur measure), so we can always pick some contours so the products always stay inside the unit circle (just imagine a big oval around all of the points on the real axis). To see why this is the right formula, observe that the right-hand side above has poles exactly at points of the form $(z,v) = (x_k, y_\ell)$ (think of integrating over z first, then v, rather than simultaneously) and the residue we would get is

$$\frac{\prod_{j=1}^N (1-x_j y_\ell)(1-x_k y_j)}{\prod_{j\neq k}(x_k - x_j)\prod_{j\neq \ell}(y_\ell - y_j)} \frac{x_k^a y_\ell^b}{1-x_k y_\ell}$$

(with the $2\pi i$ factors canceling out), and then we sum over all k, ℓ and so it matches up directly. (This is kind of like taking a Fourier coefficient if we let $z = e^{i\theta}$ and $v = e^{i\psi}$.)

This integral still looks a bit messy for lending itself to computations like using the dual specialization, but now we'll rewrite it in a way that allows us to work with "more Schur-friendly" objects. Recall that the generating function for the complete symmetric functions is $H(x,z) = \prod_i \frac{1}{1-x_i z}$ and this is the right side of the Cauchy identity, so now we can write a lot of these things in terms of the Hs. First changing variables $z = \frac{1}{u}$ yields

$$K(a,b) = \frac{1}{(2\pi i)^2} \oint \oint \prod_{j=1}^N \frac{(1-x_j v)(1-u^{-1} y_j)}{(1-x_j u)(1-y_j v)} \frac{v^{b-N}}{u^{a-N+1}} \frac{du dv}{u-v},$$

where our contours for the variables u, v are such that $|v| < |u|$, u contains the x_j^{-1}s, and v contains the y_js. And now this simplifies to the following:

Theorem 67 (Okounkov '00)
The Schur measure is a determinantal point process with kernel

$$K(a,b) = \frac{1}{(2\pi i)^2} \oint \oint \frac{H(x,u)}{H(y,u^{-1})} \frac{H(y,v^{-1})}{H(x,v)} \frac{v^{b-N}}{u^{a-N+1}} \frac{du dv}{u-v},$$

For now this is the best we'll do – we'll take some specializations of this later and see how it evolves.

Remark 68. *The Schur process is also a determinantal point process, but the formula is quite a bit more involved – most things remain the same, but we now have a big "product of stuff" for a bunch of skew Schur functions which are all Hs. So if we care about more slices of the PNG, we just have to add in more of these H factors.*

There are other ways of proving this result as well, but this is the most elementary (just opening up determinants and using the Cauchy determinant identity); if we wanted to repeat this argument for the Schur process it would be more natural to use other proof methods.

Example 69

We'll now proceed to talking about how this relates to the longest increasing subsequence (LIS) and PNG problems. We've already described previously how to go from LPP to PNG to LIS: we put a geometric random variable of parameter $a_i b_j$ at integer grid points, then tilt and create upward paths, and then obtain LIS by a specific choice $a_i, b_j = \varepsilon$ and setting the size of the grid to $N = t\varepsilon^{-1}$ as $\varepsilon \to 0$ (so that most of the entries will be equal to 0 or 1, and a constant-order number of them will be 1 so that we get a permutation of roughly constant size).

But recall that the above limit is obtained via the "Plancherel specialization," which is the case where $H(x; z) = e^{tz}$ (by taking the appropriate limit of $\prod_{i=1}^{N} \frac{1}{1-x_i z} = (1 - \varepsilon z)^{t/\varepsilon}$). So in that special case, our formula from Theorem 67 simplifies:

Corollary 70

Under the above limit, the PNG at location 0 (where the top curve will be the length of the longest increasing subsequence of a permutation of length N) is a determinantal point process on the values $(\lambda_j - j)$ with kernel

$$K(a, b) = \frac{1}{(2\pi i)^2} \oint \oint e^{t(u-u^{-1})-t(v-v^{-1})} \frac{v^b}{u^{a+1}} \frac{dudv}{u - v},$$

where notice that we swap out $b - N$ and $a - N + 1$ for b and a because of our shift on the partitions $(\lambda_j - j$ instead of $\lambda_j - j + N)$, and where our contours are such that $|u| > |v|$, v contains 0, and u contains ∞.

So in other words, we can think of the v-contour as being a counterclockwise circle around 0 and the u-contour as a clockwise circle around 0 of larger radius. And this is not so bad – we're integrating an exponential against a polynomial, and we'll talk about how to massage this later on.

We'll now write this as a sum in terms of Bessel functions, and then we'll think about a slightly different probabilistic interpretation of all of this.

Lemma 71

The kernel above can be expressed also as

$$K(a, b) = \sum_{k=0}^{\infty} J_{a+k+1}(2t) J_{b+k+1}(2t),$$

where J is the **Bessel function** defined in one of the following three equivalent ways:

1. $J_n(2t) = \sum_{i=0}^{\infty} (-1)^i \frac{t^{2i+n}}{i!(i+n)!}$,
2. $J_n(2t) = \frac{1}{(2\pi i)} \oint \frac{e^{\frac{t}{2}(v-v^{-1})}}{v^{n+1}} dv$ for a v-contour containing 0,
3. J is the bounded solution to the differential equation $t^2 J'' + tJ' + (t^2 - n^2)J = 0$.

The three equivalent forms can be easily converted between each other by evaluating the residue in (2) and also comparing coefficients in the power series expansion for (3) to the expression for (1). And so if we believe that we can get a handle on the Bessel functions, this won't be too scary – we usually can't hope for better than a single sum when describing these models.

Proof sketch. We'll show that $K(a,b)$ is equal to the second expression above. First change variables $v \mapsto z^{-1}$ to get

$$\begin{aligned} K(a,b) &= \frac{1}{(2\pi i)^2} \oint \oint e^{t(u-u^{-1})+t(z-z^{-1})} \frac{z^{-b}}{u^{a+1}} \frac{dudz}{z^2(u-\frac{1}{z})} \\ &= \frac{1}{(2\pi i)^2} \oint \oint e^{t(u-u^{-1})+t(z-z^{-1})} \frac{dudz}{u^{a+2}z^{b+2}(1-\frac{1}{zu})} \\ &= \frac{1}{(2\pi i)^2} \oint \oint e^{t(u-u^{-1})+t(z-z^{-1})} \frac{dudz}{u^{a+2}z^{b+2}} \sum_{k=0}^{\infty} (zu)^{-k} \end{aligned}$$

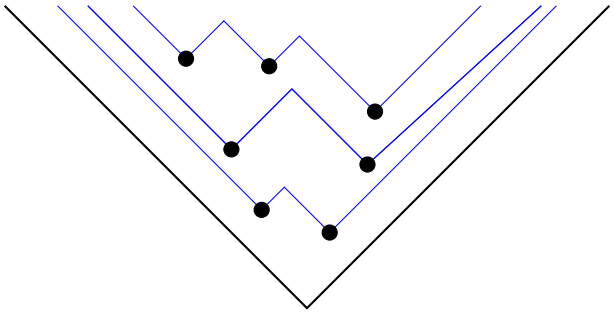
(we had to do the power series expansion this way because if $|u| > |v| = |z|^{-1}$, then $|uz| > 1$). And the advantage now is that the integral decouples as

$$\sum_{k=0}^{\infty} \frac{1}{2\pi i} \oint e^{t(u-u^{-1})} \frac{du}{u^{a+k+2}} \cdot \frac{1}{2\pi i} \oint e^{t(z-z^{-1})} \frac{dz}{z^{b+k+2}},$$

and now (after negating both integrals to swap the contour direction) we get exactly a sum of $J_{a+k+1}(2t)J_{b+k+1}(2t)$, as desired. □

Example 72

Taking a slightly different perspective now, this limit can be thought of as taking a continuous limit of PNG where we just have points in the wedge (not necessarily at integer points), and as we evolve time we form outwardly-growing plateaus which grow at rate 1, now at continuous locations and times. A slightly different way of writing this (if we work through the combinatorial description) is the following. Take each point and consider 45-degree rays to the northwest and northeast out of it; the fact that plateaus coalesce on the PNG side is like saying that "rays annihilate when they hit each other."



Observe then that this creates broken lines, and the **number of broken lines is exactly the length of the LIS** (since we want the maximal up-directed path from the bottom to the top collecting as many points as possible, and by definition we can collect at most one point per line). Furthermore, we get the number of such lines from the top point of a Schur measure. So we might then ask whether we can recover the other parts of the partition from this broken line construction – the answer is yes, but we won't go into that for now.

A different question we might ask is whether Schur can be replaced with something else and still have an interpretation of this kind. What we can do is say that the rays annihilate each other with probability $1-q$ but instead go through each other with probability q. (Such a thing can significantly change the dynamics beyond just a single ray.) We can now again think of having broken lines, where we interpret "passing through" as two broken lines touching but not crossing; and this time the number of broken lines will be the top point of something called the **Hall-Littlewood measure**. We won't give a complete definition, but the idea is that Hall-Littlewoods come from symmetric functions $Z^{-1}P_\lambda(x)Q_\lambda(x)$ as well. This is no longer a determinantal point process, and it's a fact that the number of broken lines

has fluctuations $t^{1/3}$ and scale to Tracy-Widom as long as $q < \frac{1}{2}$. But the moment we have $q \geq \frac{1}{2}$, a different story occurs; in particular we get Gaussian behavior at $q = \frac{1}{2}$ and can take $q = \frac{1}{2} - \frac{1}{\sqrt{t}}$ to interpolate between Tracy-Widom and Gaussian via the KPZ equation.

11 May 5, 2026

We computed the correlation function of the Schur measure last time, and we took the continuous limit to obtain the continuous PNG and longest increasing subsequence. We'll explore some of the probabilistic consequences and interpretations now, moving towards asymptotics.

Example 73

To begin, we'll think about some combinatorial interpretations of the continuous PNG. Recall that in the appropriate limit, we end up with some points in the plane, and what we really care about are their relative ordering among each other. In generic position, N such points will give rise to a permutation σ, where the ith point from the left is the $\sigma(i)$th lowest point.

We saw previously how to map models like this to line ensembles, and now we claim that those features still appear in this limit. One thing we can do is take the previous constructions we've done and take the limit of them, but here we'll do a raw description via the **RSK correspondence** (RSK stands here for "Robinson–Schensted–Knuth"). This correspondence maps any permutation to a pair of standard Young tableaux:

Definition 74

A **standard Young tableau (SYT)** is a filling of a Young diagram with n boxes of the numbers $\{1, 2, \cdots, n\}$, so that the numbers are increasing in rows (to the right) and columns (downward).

An example is shown below:

1	3	6
2		
4		
5		

The way this pair of standard Young tableaux is obtained from a permutation is the following:

1. Start with two empty standard Young tableaux.
2. At each step, add the next $\sigma(i)$ to the first SYT. We do so by trying to add to the first row, and if that doesn't work because it breaks the row-increasing rule, then we push whatever was in its way down a row instead, and repeat this until we can place the box.
3. Now add i to the second SYT so that it has the same shape as the first one.

For example, suppose our permutation in one-line notation is $(6, 2, 4, 3, 1, 5)$ (that is, 1 is mapped to 6, 2 mapped to 2, and so on). Then after the first iteration we have

6		1	

and then after adding 2 (which bumps 6 down) we have

2	1
6	2

.

Next adding 4 yields

2 4	1 3
6	2

,

adding 3 (so bumping 4 down, which then bumps 6 down) yields

2 3	1 3
4	2
6	4

,

adding 1 bumps the 2 (and 4 adn 6) down to get

1 3	1 3
2	2
4	4
6	5

,

and finally adding 5 gives us the pair of Young tableaux

1 3 5	1 3 6
2	2
4	4
6	5

,

Fact 75

The map described above is a bijection between the set of permutations S_n of size n and pairs (P, Q) of standard Young tableaux of n boxes of the same shape. In particular, letting $\dim \lambda$ be the number of standard Young tableaux of shape λ, this fact implies the identity

$$n! = \sum_{\lambda \text{ of size } n} (\dim \lambda)^2.$$

Here $\dim \lambda$ is also the dimension of the irreducible representation V_λ of S_n, so this is a combinatorial realization of a special case of Burnside's theorem $\sum_{V \text{ irrep of } G} (\dim V)^2 = |G|$.

We now claim that the length of the first row is also deterministically the longest increasing subsequence in σ (we can check this by checking that an LIS of length k implies at least k boxes in the first row, and also that k boxes in the first row implies a corresponding LIS). Thus we have the following corollary:

Corollary 76

Sample a partition λ under the measure $\frac{(\dim \lambda)^2}{n!}$. Then the law of λ_1 is the law of the LIS of a uniformly random permutation.

It turns out that in the continuum PNG limit (with the exponential specialization), the Schur measure is exactly $\frac{(\dim \lambda)^2}{N!}$ but with $N \sim \text{Poisson}(t)$, and in fact there is a "de-Poissonization" we can do to get this to a fixed N. But we won't go into that here.

We described $\dim \lambda$ in a rather non-explicit way, but there is actually a formula for it:

Proposition 77 (Hook-length formula)

For any partition λ of size n, we have

$$\dim \lambda = \frac{n!}{\prod_{(i,j)\in\lambda} h_{(i,j)}}.$$

Here the product in the denominator is over all boxes in the corresponding Young diagram, and $h_{(i,j)}$ is the **hook length** of the cell, which is the number of boxes weakly (directly) below or (directly) to the right of (i,j).

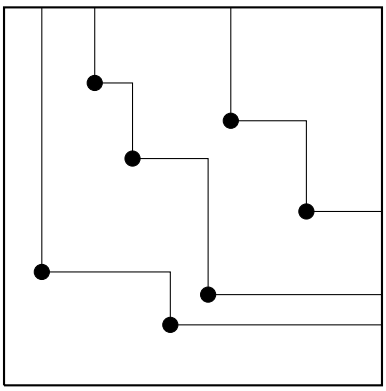
For example, the hook lengths for the partition $(3,3,2,1)$ are shown below:

6	4	2
5	3	1
3	1	
1		

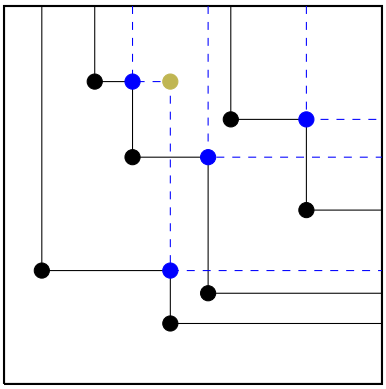
Example 78

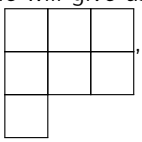
A more geometric witness of RSK in terms of the diagrams we drew last time is the following. Given a set of points in the plane, we draw horizontal and vertical rays up and to the right of each point, with intersections causing annihilations. This gives us a union of broken lines, and the number of such broken lines is λ_1. This also creates new points given by the intersection points of annihilation.

An example is shown below (with three total broken lines)

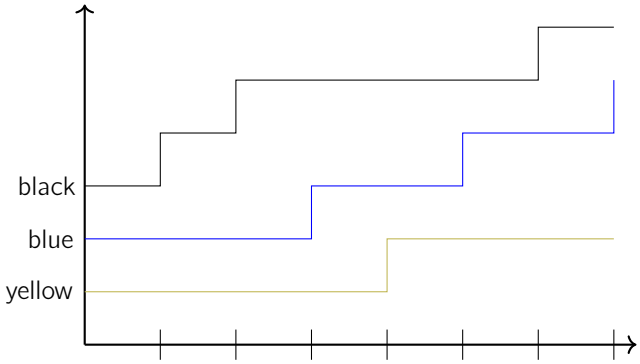


Once we do this, we can again repeat the construction but with the new points, which creates new broken lines in blue. It turns out the number of new broken lines will be λ_2:



Doing this one more time will give us one more broken line for the yellow point, hence $\lambda_3 = 1$. Thus in this case we get a partition of shape  , and the number of boxes is indeed exactly the number of points we started from. (This is sometimes called the **shadow lines picture** by Viennot, and it gives rise to the same λ from RSK.)

This actually gives rise to a line ensemble in the following way. We have three layers (the black, blue, and yellow), and so we can track, from left to right, the number of broken lines that we have as we scan through. We'll have a "tracker" for each of the colors, and we'll offset them by 1 so the black tracker starts at 0, the blue tracker starts at -1, and the yellow tracker starts at -2, and each time we see a line we increment by 1:



Notice that we only change how many levels / broken lines we've created each time we cross one of the n original points. And this line ensemble will be the limit of the PNG line ensemble if we set things up properly with the correct offset, though we won't get into the details.

Example 79

Next, there's an interpretation of longest increasing subsequence in terms of **airplane boarding**. Suppose we have a certain number of people who are trying to board a plane, and we have a certain number of seats from front to back. However, the order of boarding may differ from the order of seats. If the first person to board was at the back, the next person was the second from the back, and so on, then everyone can board and put their luggage up at the same time. But in general there will be some delays because earlier boarders are blocking later boarders from later seats, and in fact the amount of time it takes (in units of luggage time) is the longest increasing subsequence of the corresponding permutation.

Example 80

The LIS is also related to the idea of **patience sorting**, which is a method for sorting a deck of cards quickly by hand. This was introduced by Ross in the 1960s and named by Mallows; theoretical results were proved by Hammersley in 1971 and a paper by Aldous and Diaconis describes things in more detail.

Our algorithm is the following: we start with our unsorted cards, and we sort into piles as follows.

- Start with no piles.
- Sequentially, put each card into the leftmost pile whose top card is larger than it. (So this is kind of like RSK but with no bumping.)
- Once we finish this, the number of piles is the longest increasing subsequence. We can then get the cards in increasing order by repeatedly picking off cards from the top.

For example, if our cards are $4, 3, 5, 10, J, 2$, then after making the piles the first one will have $4, 3, 2$, the next will have 5, then the next 10, and the last one J. This kind of algorithm is also how we might compute the LIS in practice; it's essentially some variation of the greedy algorithm.

Turning to limits now, let's recall the kernel

$$K(a,b) = \frac{1}{(2\pi i)^2} \oint \oint \frac{w^a}{v^{b+1}} \frac{1}{v-w} \frac{e^{t(v-v^{-1})}}{e^{t(w-w^{-1})}}$$

(this is slightly offset from our definition last time, but it's really just a matter of convention of how our partitions are offset). We'll now think about what happens when we take t large, since that corresponds to the LIS of a very long permutation. This might seem a bit intimidating, but a "steepest descent" type calculation will suffice, and to understand that we'll begin with a simpler example:

Example 81

Let's think about the case where we have a one-dimensional real integral. Vladimir Arnold had a list of questions that he felt mathematicians should be able to answer, and one of them was to approximate $\int_0^\pi (\sin x)^{100} dx$ up to 10% without a computer.

We know that the integrand is very peaked at $\frac{\pi}{2}$, so we should approximate around that point: changing variables and then doing a Taylor expansion yields (set $n = 100$)

$$\begin{aligned}\int_{-\pi/2}^{\pi/2} (\cos x)^n dx &= \int_{-\pi/2}^{\pi/2} \left(1 - \frac{x^2}{2} + O(x^4)\right)^n \\ &\approx \int_{-1}^{1} \left(1 - \frac{x^2}{2}\right)^n,\end{aligned}$$

where it's accurate to restrict to a smaller interval because $\cos x$ is very small when we're far away from 0, hence the contributions there are negligible. And now we'll approximate by an exponential: we want to use that $\left(1 - \frac{a}{n}\right)^n \approx e^{-a}$, so rescaling by setting $x = \frac{y}{\sqrt{n}}$ yields

$$\frac{1}{\sqrt{n}} \int_{-\sqrt{n}}^{\sqrt{n}} \left(1 - \frac{y^2}{2n}\right)^n dy \approx \frac{1}{\sqrt{n}} \int_{-\sqrt{n}}^{\sqrt{n}} e^{-y^2/2} dy \approx \frac{1}{\sqrt{n}} \int_{-\infty}^{\infty} e^{-y^2/2} dy = \sqrt{\frac{2\pi}{n}}.$$

And indeed plugging in $n = 100$ lets us approximate this by roughly $\frac{1}{4}$, and the real answer is indeed something like 0.25004.

So the point was to find the maximum and rescale around the maximum so as to make the quantity being integrated of order 1. We know that typically a function is roughly quadratic around its maximum (because the first derivative vanishes), so that allows us to approximate around any "strict Hessian" point by a Gaussian integral and thus things become exactly doable. But sometimes we don't have a Gaussian and that's actually what gets us the $n^{1/3}$ instead in Tracy–Widom.

Instead of going straight there, we'll now try to repeat this reasoning on $K(a,b)$ directly. For simplicity, consider the case where $a=b=ut$ (where we think of u as the parameter); we thus have to evaluate

$$\frac{1}{(2\pi i)^2}\oint\oint \frac{1}{v-w}\exp\left(t(v-v^{-1}-w+w^{-1}+u\log w-u\log v)\right)dwdv,$$

where we're integrating over contours where $|v|>|w|$. We have an exponential and t is going to be huge, so it makes sense to try to maximize the quantity inside: defining $S(z)=z-z^{-1}-u\log z$, we want to compute

$$\frac{1}{(2\pi i)^2}\oint\oint \frac{1}{v-w}\exp\left(tS(v)-S(w)\right)dwdv.$$

Now if $\mathrm{Re}(S(v))<\mathrm{Re}(S(w))$, then the integral is super small and there's no issue. Thus we want to localize around a critical point of S, and $S'(z)=1+\frac{1}{z^2}-\frac{u}{z}$, so $S'(z)=0$ when

$$z^2-uz+1=0 \implies z=\frac{u\pm\sqrt{u^2-4}}{2}.$$

So immediately we see that there is a curiosity when $u=2 \implies z=1$, since that means we'll also have the quadratic term vanish. But let's not think about that case yet and just do the case where $u<2$; then the two critical points $z_\pm$ will lie on the unit circle. Here our contours for v and w are both counterclockwise around the origin and containing 0, but we haven't specified them further beyond that. Thus we can choose the contours so that we localize around z_+ and z_-.

Remark 82. *Notice that $Re(S(z))=0$ when $|z|=1$, since $z-z^{-1}$ is purely imaginary and $\log z$ is purely imaginary as well. Thus what we want to do is push both contours away from the unit circle in such a way that $Re(S(v))<Re(S(w))$ everywhere on the contours except at $z_\pm$. So ideally we'll make v and w lie mostly on "different level lines."*

12 May 7, 2026

We began computing asymptotics last time, and the basic formula we want to study is the correlation kernel for the partition parts (λ_i-i) (maybe shifted by 1), where λ is sampled under a certain Schur measure:

$$K(a,b)=\frac{1}{(2\pi i)^2}\oint\oint \frac{w^a}{v^{b+1}}\frac{dwdv}{v-w}\frac{e^{t(v-v^{-1})}}{e^{t(w-w^{-1})}}$$

for contours containing 0 with $|v|>|w|$.

Remark 83. *One way to think about λ_i-i is to take our partition Young diagram and rotate it by 135 degrees counterclockwise so that it "fits into the wedge" where we defined the PNG. Then the highest part of that figure will trace out a function, and if we draw a circle on every down-step its x-coordinates will be exactly λ_i-i up to a linear shift. This is particularly useful if we want to think about the exclusion process by adding boxes to the Young diagram.*

It was known long before the KPZ analysis that the longest increasing subsequence of a random permutation is $2\sqrt{n}+o(\sqrt{n})$; we'll see it's $n^{1/6}$ times Tracy–Widom. So the right-most down step will occur around $2\sqrt{n}$ and so that's really where we want to probe to understand the asymptotics.

We showed with a toy example last time that in problems like this, we want to Taylor expand around the critical points. And as we started doing previously, we'll start with examining $c\sqrt{n}$ for some $c < 2$ (taking $a = b = ct$; remember that our box will have t^2 particles so $n = t^2 \implies t = \sqrt{n}$). We should expect that if $c > 2$ we would have no particles (so $K(ct, ct) \to 0$), if $c < 2$ we have a positive density (so $K(ct, ct)$ tends to a constant), and if $c = 2$ something KPZ-like should occur (we'll see).

As we did last time, for fixed c we have

$$K(ct, ct) = \frac{1}{(2\pi i)^2} \oint \oint \frac{1}{v} \frac{dv\,dw}{v - w} \exp\left((S_c(v) - S_c(w))t\right), \quad S_c(z) = z - z^{-1} - c \log z.$$

We have $\mathrm{Re}(S(z)) = 0$ if $|z| = 1$, and we want to move our contours so that $\mathrm{Re}(S(v)) \leq \mathrm{Re}(S(w))$, with equality only at a few places so that the exponential decay can be neglected.

Example 84

First, we'll do the case $c > 2$. In this case we can actually push the contours so that we have strict inequality everywhere by pushing v away from the unit circle and pulling w inside it.

Indeed, if $z = a + bi$, we have

$$\mathrm{Re}(S(a + bi)) = a - \frac{a}{a^2 + b^2} - \frac{c}{2} \log(a^2 + b^2)$$

and now the gradient of this object (in (a, b)) is

$$\nabla \mathrm{Re}(S(a + bi)) = \left(1 - \frac{b^2 - a^2}{(a^2 + b^2)^2} - \frac{ac}{a^2 + b^2}, \frac{2ab}{(a^2 + b^2)^2} - \frac{bc}{a^2 + b^2}\right)$$

Taking z itself to be on the unit circle now, the denominators are all just 1 so we will have $\nabla \mathrm{Re}(S(a + bi)) = (1 + a^2 - b^2 - ac, 2ab - bc) = (2a^2 - ac, 2ab - bc) = (2a - c) \cdot (a, b)$. Thus if we perturb (a, b) radially outward in the normal direction by δ, the change in S will be $(2a - c)\delta$; in particular $(2a - c)$ is strictly negative for any point $a + bi$ on the unit circle, so S decreases as we move out and increases as we move in.

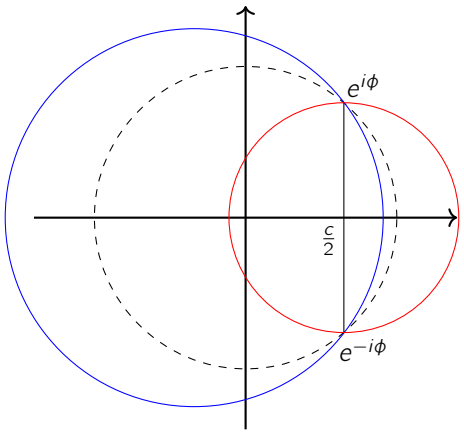
Thus we will take the v-contour to be the circle around the origin of radius $(1 + \delta)$ and the w-contour to be the circle of radius $(1 - \delta)$, chosen so that $\mathrm{Re}(S(v)) < -\varepsilon$ and $\mathrm{Re}(S(w)) > \varepsilon$ for some $\varepsilon > 0$. That means the exponential in the integrand decays as $e^{-2\varepsilon t}$; the size of the contour is order-1 so the whole thing will decay as $t \to \infty$. In words, the probability that the longest increasing subsequence is longer than $(2 + \delta)t$ decays exponentially in t.

Remark 85. *In particular, this would let us answer the large-deviations question of how quickly these exponentials decay; we'd have to pick the contour more carefully to make the quantity $\mathrm{Re}(S(a + bi))$ "as negative as possible" and it's just a question about two real variables. But we won't do that too carefully here.*

Example 86

Now we'll do the $c < 2$ case. What will happen instead is that $(2a - c)\delta$ is no longer negative everywhere; instead, if $a \geq \frac{c}{2}$ there will be a sign change.

Let the points on the unit circle where $\mathrm{Re}(z) = \frac{c}{2}$ be $e^{i\phi}$ and $e^{-i\phi}$. (These points are actually the critical points of $S_c(z)$, which will become relevant for the Tracy–Widom case later.) To the left of $\frac{c}{2}$ we want to push v out and pull w in, but to the right of $\frac{c}{2}$ we instead want to pull v in and push w out. And our contours don't allow that if $|v| > |w|$. The key trick of this steepest descent technique now, though, is to do it anyway – "the exponential decay is much more important than respecting the contour," and we can see the price that we pay for doing it.



In the diagram above, the v contour is in blue and the w contour is in red (counterclockwise). We might be concerned about the intersection points, but we just have a pole of order 1 and that's okay to integrate in two dimensions. So the integral makes sense and we have $\operatorname{Re}(S(v)) \le \operatorname{Re}(S(w))$ everywhere on the contours, and the only way in which we've cheated is that "we pay for the poles where $v = w$" when we move the w-contour (while keeping the v-contour fixed), and the residue is actually not so bad to compute.

For concreteness, suppose the v-contour is actually exactly S^1. Then if we move w from inside S^1 to outside S^1, we'll only see a residue for $v = e^{i\theta}$ with $\theta \in [-\phi, \phi]$, and the residue will be 1. Therefore the price we pay is

$$\begin{aligned}\frac{1}{2\pi i} \oint_{S^1\text{-arc from } e^{-i\phi} \text{ to } e^{i\phi}} \frac{dv}{v} &= \frac{1}{2\pi} \int_{-\phi}^{\phi} d\theta \\ &= \frac{\phi}{\pi} \\ &= \frac{1}{\pi} \arccos \frac{c}{2},\end{aligned}$$

and this is exactly the density of particles at ct. Indeed, if $c = 2$ we get $\arccos 1 = 0$, so the density is indeed continuous. And so at $c = 0$ the density of particles is $\frac{1}{2}$, and near $c = -2$ we get a density of 1.

Fact 87

Once we've moved the contours past each other and paid the residue penalty, the remaining integral is actually very small and doesn't contribute. Indeed, we can say that only a small δ-ball around the two points $e^{i\phi}$ and $e^{-i\phi}$ will matter since everything else is exponentially decaying for the same reason as the $c > 2$ case, and within those δ-balls we're integrating an order-1 pole within volume δ^2; hence the contribution is of order δ. Thus if we take $t \to \infty$ then $\delta \to 0$, this part is indeed negligible.

So integrating out, we see that the limit shape of the (135-degree-rotated) partition is

$$\Omega(c) = \frac{2}{\pi} \left(c \arcsin \frac{c}{2} - \sqrt{4 - c^2} \right),$$

which is sometimes called the **Vershik–Kerov limit shape**. And we can even extract out the correction here, though people usually don't do so very carefully (really that's what we're going to do in the $c = 2$ case).

Example 88

In this analysis, we've looked at $a = b = ct$, which is exactly the particle density at a certain point. What we can do instead is consider the correlation $K(ct + x, ct + y)$ for x, y fixed, and everything goes through as before except with an extra $\frac{w^x}{v^y}$ in the integrand.

The exact same deformation of the contours works, but now the "price we pay" is the **sine kernel**

$$\frac{1}{2\pi i} \oint_{S^1\text{-arc from } e^{-i\phi} \text{ to } e^{i\phi}} v^{x-y-1} dv = \frac{\sin(\phi(x-y))}{\pi(x-y)}$$

(when $x = y$ we recover the usual $\frac{\phi}{\pi}$). So in particular, the probability that $ct, ct+1$ are both in the point process is the determinant

$$\det \begin{bmatrix} K(ct, ct) & K(ct, ct+1) \\ K(ct+1, ct) & K(ct, ct) \end{bmatrix} = \det \begin{bmatrix} \frac{\phi}{\pi} & \frac{\sin\phi}{\pi} \\ \frac{\sin\phi}{\pi} & \frac{\phi}{\pi} \end{bmatrix} = \frac{\phi^2 - \sin^2\phi}{\pi^2}.$$

Notice that this is smaller than what we'd get if the particles are independent, so **particles repel**. This is a very general phenomenon of point processes for any symmetric kernel (since we're subtracting off something squared when we compute this 2×2 determinant). And this also tells us that the rate at which the repelling happens is $\frac{1}{a^2}$ for distance a.

Example 89

We'll now start to move towards the $c = 2$ case; let's just state the result now and prove it next time.

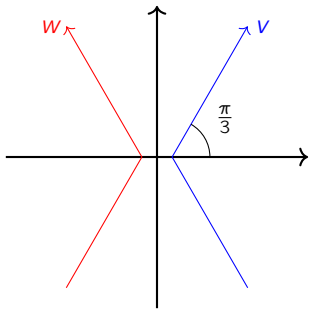
Instead of the sine kernel, we'll get the following object which is a process over the reals (since the fluctuations will not be discrete at the edge):

Definition 90

The **Airy kernel** K_{Airy} is given by, for $x, y \in \mathbb{R}$,

$$K_{\text{Airy}}(x, y) = \frac{1}{(2\pi i)^2} \iint \exp\left(\frac{v^3}{3} - \frac{w^3}{w} - vx + wy\right) \frac{dvdw}{v - w}.$$

where v and w are any fixed contours of the shape shown below (with all angles $\frac{\pi}{3}$).



The reason this converges is that on the lines we draw, v^3 mostly has negative real part and w^3 mostly has positive real part and so we have something exponentially decaying. The name "Airy" comes from the fact that we also have the representation

$$K_{\text{Airy}}(x, y) = \int_0^\infty \text{Ai}(x + \lambda)\text{Ai}(y + \lambda) d\lambda,$$

where Ai is the **Airy function**

$$\begin{aligned}\mathrm{Ai}(x) &= \frac{1}{\pi}\int \cos\left(\frac{z^3}{3}+zx\right)dx \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp\left(i\frac{z^3}{3}+izx\right)dz.\end{aligned}$$

Indeed, we can imagine plugging in this expression into the integral $\int_0^\infty \mathrm{Ai}(x+\lambda)\mathrm{Ai}(y+\lambda)d\lambda$, and then the integral over λ gives us $\frac{1}{z+w}$ and then we swap variables. And we can also further write things in a similar flavor as the sine kernel as

$$K_{\mathrm{Airy}}(x,y) = \frac{\mathrm{Ai}(x)\mathrm{Ai}'(y)-\mathrm{Ai}'(x)\mathrm{Ai}(y)}{x-y}.$$

This is kind of reminiscent of a Gaussian but with cubes instead of squares, so that gives us a sign we might see a $t^{1/3}$ term somewhere.

Definition 91

The **Tracy–Widom GUE distribution** is given by the expression

$$\begin{aligned}F_{\mathrm{GUE}}(s) &= \det\left[I - K_{\mathrm{Airy}}1\{(s,\infty)\}\right] \\ &= \sum_{n=0}^{\infty}\frac{(-1)^n}{n!}\int_s^\infty\cdots\int_s^\infty \det\left[K_{\mathrm{Airy}}(x_i,x_j)\right]_{1\le i,j\le n}dx.\end{aligned}$$

So perhaps we can start to see all of the pieces of the result coming together, and we'll do that next time!

13 May 12, 2026

Last time, we introduced the Airy kernel (which will arise as a scaling limit of our kernels at the edge). We were computing the density of particles at $c<2$ and $c>2$ last time, and now we're going to understand what happens at $c=2$ (where the density is very low and something different happens to the top particle). We wrote down an expression for the Tracy–Widom GUE distribution in terms of the Airy kernel; specifically, if X has that distribution, then we have $\mathbb{P}(X\le s) = F_{\mathrm{GUE}}(s)$.

Looking more carefully at the expression

$$F_{\mathrm{GUE}}(s) = \sum_{n=0}^{\infty}\frac{(-1)^n}{n!}\int_s^\infty\cdots\int_s^\infty \det\left[K_{\mathrm{Airy}}(x_i,x_j)\right]_{1\le i,j\le n}dx,$$

it turns out $K_{\mathrm{Airy}}(x,y)\to 0$ as $x,y\to\infty$ (it's not so transparent from the contour integral expression, but it's clearer from the representation in terms of the Airy function which decays exponentially in its argument). Thus as $s\to\infty$, we have $F_{\mathrm{GUE}}(s)\to 1$ (only the $n=0$ term still stays), which should indeed make sense. On the other hand, the fact that $F_{\mathrm{GUE}}(s)\to 1$ as $s\to-\infty$ takes quite a bit more work and we won't do it here.

Remark 92. *We might wonder why our infinite alternating sum even converges, since we have an n-fold integral which could be big. But what's useful in general is that we can estimate determinants via the* **Hadamard bound***: if $A=(A_{ij})$ is an $N\times n$ matrix with columns (A_j), then*

$$|\det A| \le \prod_{j=1}^{n}||A_j||_2$$

(because the determinant is a volume and the largest volume of the parallelopiped is if the columns are all orthogonal). So for any fixed j,

$$\det\left[K_{Airy}(x_i,x_j)\right]_{1\le i,j\le n}\le\prod_{j=1}^{n}||K_{Airy}(x_i,x_j))_i||_{L^2},$$

and we can use what we know about the Airy function to bound this in a way so that it grows only exponentially (and thus the $n!$ term dominates).

We can now state the result that we'll prove today.

Theorem 93

Let $K_t^{\mathrm{PNG}}(a,b)$ denote the PNG kernel coming from our particular (Poisson point process-type) specialization. Then

$$\lim_{t\to\infty}t^{1/3}K_t^{\mathrm{PNG}}(2t+t^{1/3}x,2t+t^{1/3}y)=K_{\mathrm{Airy}}(x,y).$$

This is in contrast to $K_t(ct,ct)$ converging either to 0 (exponentially in t) or to $\frac{\arccos\frac{c}{2}}{\pi}$ for $c>2$ and $c<2$, respectively. So we have a positive density of particles up until $c=2$, and now we see what happens right at the critical point even though we know that the density will tend to zero.

Interpreting the actual statement of the theorem, it will turn out that the proper scaling for variation in the limit is $t^{1/3}$ rather than constant order (as we had in the $c<2$ case), and the density of particles is roughly of order $t^{-1/3}$.

Example 94

Before we prove this, we'll make the connection to Tracy–Widom. We have to take all of the scaling into account, but what that means is if we center around $2t$ in a window of length $t^{1/3}$, then we'll see a point process $\frac{(\lambda_i-i-2t)_i}{t^{1/3}}$ on $\mathbb{R}$ instead of $\mathbb{Z}$, and the correlations are given by the Airy kernel. (The density correction from the rescaling accounts for the leading factor of $t^{1/3}$.)

Since $\det[1-K_{\mathrm{Airy}}1\{(s,\infty)\}]$ then tells us the position of the highest particle, we're indeed saying that the rightmost point in PNG has Tracy–Widom fluctuations:

$$\mathbb{P}(\lambda_i\le 2t+st^{1/3})\approx\mathbb{P}\left(\frac{\lambda_i-2t}{t^{1/3}}\le s\right)\to\det\left[I-K_{\mathrm{Airy}}(x,y)1\{x,y\le s\}\right]=F_{\mathrm{GUE}}(s).$$

Fact 95

We do have to worry about uniform convergence of K_{Airy} so that these infinite sums converge, but in practice that's not an issue and just makes the papers longer with more computations. It's not so involved to go from convergence of the kernel to convergence of the process; we need tightness and uniqueness of limits. Tightness usually holds just by reading off the kernel, and uniqueness requires correlation functions to pin down the process, which is actually what we're getting already. So tightness is really the only thing being swept under the rug here.

Proof of Theorem 93. We start with our explicit formula for the kernel

$$K_t(2t+t^{1/3}x,2t+t^{1/3y})=\frac{1}{(2\pi i)^2}\oint\oint e^{t(S(v)-S(w))}\frac{w^{yt^{1/3}}}{v^{xt^{1/3}+1}}\frac{dvdw}{v-w},\quad S(z)=z-z^{-1}-2\log z,$$

with contours such that $|v|>|w|$. Our goal is to take the limit $t\to\infty$, and now is when we're actually going to have to pull the Gaussian integral-type trick from Arnold's integration problem. With $S'(z)=\frac{1}{z^2}(z-1)^2$, we see that $z=1$

is the critical point of order 2, and the first nontrivial contribution is the cubic term (since $S(1) = S'(1) = S''(1) = 0$). Thus

$$S(1+\delta z) = \frac{\delta^3 z^3}{3} + O(\delta^4),$$

and so we will change our variables to have the δ^3 cancel out the t (much like we chose the rescaling parameter in the $\cos^N \theta$ problem to be of the correct "critical width" to make everything the right order). There we needed to do $\frac{1}{\sqrt{N}}$ for the quadratic behavior, and here we will need to do $\delta = \frac{1}{\sqrt[3]{t}}$ for the cubic behavior. Replacing $(v, w) \mapsto (1 + t^{-1/3}v, 1 + t^{-1/3}w)$ yields

$$K_t = t^{-1/3} \frac{1}{(2\pi i)^2} \oint \oint \exp\left(t\left(\frac{v^3}{3t} - \frac{w^3}{3t} + O(t^{-4/3}) \right) \right) \frac{(1+t^{-1/3}w)^{yt^{1/3}}}{(1+t^{-1/3}v)^{xt^{1/3}}} \frac{dvdw}{v-w}$$

where the $t^{-1/3}$ in the front comes from the Jacobian factor in $\frac{dvdw}{v-w}$, and where the contours for v and w now have to loop around $-t^{1/3}$ instead of $-\infty$ (which is really an irrelevant condition), and where v should be to the right of w near the main contribution of this integral. But in particular, the contours that we drew last class for the Airy kernel (at angle $\frac{\pi}{3}$) will work, and that's exactly what will make sense as $t \to \infty$.

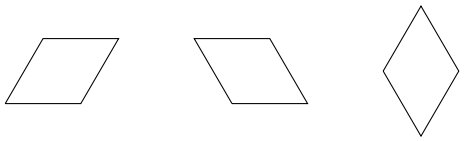
Now for the powers of w and v we use that $(1 + \frac{c}{n})^n \to e^c$ as $n \to \infty$. We have

$$\begin{aligned}
\lim_{t\to\infty} t^{1/3} K_t &= \lim_{t\to\infty} \frac{1}{(2\pi i)^2} \oint \oint \exp\left(\frac{v^3}{3} - \frac{w^3}{3} + O(t^{-1/3}) \right) \frac{(1+t^{-1/3}w)^{yt^{1/3}}}{(1+t^{-1/3}v)^{xt^{1/3}}} \frac{dvdw}{v-w} \\
&= \frac{1}{(2\pi i)^2} \oint \oint \exp\left(\frac{v^3}{3} - \frac{w^3}{3} \right) \frac{e^{wy}}{e^{vx}} \frac{dvdw}{v-w} \\
&= K_{\text{Airy}}(x, y),
\end{aligned}$$

as desired. We've neglected the contribution of the contour outside of this main region (that is, we've proven pointwise convergence instead of convergence of the integral), but we just use the fact that $\text{Re}(v) < \text{Re}(w)$ and that our specific choice of Airy contours allows for the dominated convergence theorem to apply. □

Fact 96

In our analysis above, we've picked c in such a way to get an order-2 zero instead of the generic order-1 zero, and we might ask what happens if we rig it in another problem to get something different. These limits also exist in the world of random tilings, and we'll make some remarks about that now. Under the name of **lozenge tilings**, we can tile a regular hexagon using three kinds of rhombi shown below, and we can consider a uniformly random such tiling.



Outside of the inscribed circle of the hexagon, we'll typically only see one kind of tile in each connected region (this is called the "Arctic circle phenomenon), and inside the inscribed circle we have some nontrivial density of each type of rhombus. The outer part is usually called the "frozen region," and the inner part is called the "liquid region."

Thus we might ask what happens right at the "Arctic boundary" of the circle (of course, one might complain that actually in the Arctic, the inside should be the frozen region). In order to explain that, we can also think about this model by drawing a dotted line between the midpoints of the two horizontal edges of each rhombus which has them (so two of the rhombi types will have a dotted line, and the last one does not); these dotted lines will form non-intersecting paths.

So if we zoom in around the rightmost point of the Arctic boundary, the vertical rhombi will turn into horizontal rhombi at some point near the boundary of the Arctic circle. Any horizontal slice will then intersect these dotted lines at some locations, and these locations form a point process, in fact even a determinantal one. This process is not a Schur measure, but in a somewhat similar flavor with a double contour integral we get an explicit kernel. With an appropriate scaling limit (again at the scale of $n^{1/3}$ fluctuations, where n is the sidelength of the hexagon) and the same "steepest descent" strategy, this kernel then converges to the Airy kernel near the circle. Thus in particular, the location at which the extreme path meets the slice is distributed as Tracy–Widom.

This prompts some natural questions:

1. We only looked at where the path meets a given slice, but what about the distribution of the whole path? (This is similar to asking for Brownian motion instead of just a Gaussian at each point.)
2. What about the full scaling limit of extreme paths together?
3. What if we use a different domain instead of the regular hexagon, but still use the same three rhombi to tile?

The names for the first two objects turn out to be the **Airy$_2$ process** and the **Airy line ensemble**, which we might talk about a bit more next time. But at a high level, we can describe the answer to the third question now.

Theorem 97 (Aggarwal–Huang)
For any generic polygon tiled by rhombi, we get a frozen region governed by an inscribed curve, along with some associated extreme paths. Then at any generic point on the Arctic boundary (we just want to avoid tangency locations, as well as cusps), we get the Airy line ensemble.

On the other hand, something different happens at the non-generic points. Zooming in around a cusp, the process describing the scaling limit of the extreme paths will be different. Indeed, at least on some domains, we get a determinantal point process with an explicit kernel. Applying steepest descent at this cusp point, the analog of the **action** $S(z)$ will instead have a third-order critical point instead of a second-order one. The Taylor expansion then gives us a quartic leading term, so that we have something like $e^{\frac{v^4}{4}-\frac{w^4}{4}+\cdots}$, leading to the **Pearcey kernel** and accordingly the **Pearcey process**:

Theorem 98 (Huang–Zhang)
Around a generic cusp in a generic domain, and under certain boundary conditions, we see the Pearcey process for the extreme paths.

In terms of natural probabilistic surface models, the Pearcey process is where it stops (and that's a provable statement). But we can get higher-order limits out of certain specific kernels, and that kind of phenomenon does matter in fields like string theory where these singularities are studied.

Remark 99. *The reason that we can have this kind of "cusp behavior" with Brownian motions is the following: if we imagine starting a bunch of nonintersecting Brownian motions in two separated groups, then they will spread out over time from each other until the two groups collide, and that collision is exactly what we should imagine happening at the cusp.*

14 May 14, 2026

We obtained the Tracy–Widom fluctuations for the longest increasing subsequence last time, and we also talked a bit about the relations to tilings, in particular leading to the Airy line ensemble for describing extreme paths. We'll discuss that line ensemble a bit more today and then see the connection to random matrices.

Example 100

Continuing the setting from last time, we tile a regular hexagon with one of three types of rhombi and draw dotted segments through the rhombi with horizontal edges. The resulting dotted lines near the inscribed circle are then the objects of interest.

We can see the Airy line ensemble through this model or multi-layer PNG, or we can do it through formulas. But a third interpretation is the Gibbs property, which is arguably the most powerful way of thinking about it. We'll go through all of those interpretations.

Recall that the multi-layer PNG is a set of nonintersecting paths in the integer lattice (again, see the example above Lemma 20 for a picture). We can imagine that if we run the PNG for a very long time, we'll have many nontrivial paths lifted above their initial flat state, and so we might predict that the top paths will look something like the dotted lines.

We'll be a bit heuristic about this, but the idea is to look at the top curve at location $x = 0$ and zoom in at the correct order of fluctuations, which is $N^{1/3}$. We might then imagine that because the multilayer PNG consists of (weighted) paths with certain allowed up, right, and down edges, conditioned to not intersect, in the absence of other paths we have a weighted random walk and so we should converge to Brownian motion under the usual diffusive scaling (dividing by $T^{1/2}$ if we have a horizontal scale T). Therefore to match the predicted $N^{1/3}$ fluctuations, we should be looking at a **horizontal window of width** $N^{2/3}$ around the origin, and then we'll get some nontrivial stochastic processes. That turns out to also be the correct width to consider in the Arctic circle example as well.

Theorem 101 (Prähofer–Spohn '01)

The heights of the top curves in the multilayer PNG model satisfy

$$\left(\frac{h_i(xN^{2/3}) - cN}{dN^{1/3}}\right)_{i\in\mathbb{Z}, x\in\mathbb{R}} \to \left(A_i(x) - x^2\right)_{i\in\mathbb{Z}, x\in\mathbb{R}}$$

for some constants c, d. (We sometimes call $(A_i(x))_{i\in\mathbb{N}}$ the stationary Airy line ensemble and $(P_i(x))_{i\in\mathbb{N}} = (A_i(x) - x^2)_{i\in\mathbb{N}}$ the parabolic Airy line ensemble.)

The reason for using the convention of $-x^2$ is that in the hexagon tiling model, the Arctic circle boundary curves downward and thus it looks parabolic locally. Since we're considering a window of smaller order than the radius, this approximation is what we want to do to correct for the curvature, and in fact this is what will make $A_i(x)$ translation-invariant in x.

Remark 102. *We can compute the values of c and d in a few ways. For the mean c, we can sometimes do it with steepest descent But sometimes with other strategies (like in Dyson Brownian motion or in the LIS model where things are already known before the fluctuation result). And the variance will just come from understanding the random walk.*

We can now try to express this as a determinantal point process. The Schur measure gave us a specific vertical slice (at a particular value t), and the Schur process gives us a multi-point distribution. The Schur process turns out to

also be a determinantal point process, meaning that we know all of the correlation functions between slices at different ts in terms of the kernel. And so we can use the joint correlations to form a determinantal point process for the Airy line ensemble as well, and that's what we'll try to do.

Proposition 103

The stationary Airy line ensemble is a sequence of random curves $(A_1(x), A_2(x), \cdots)$, ordered so that $A_1 > A_2 > \cdots$, defined by the probability density $\mathbb{P}\left(\bigcap_{i=1}^m \left\{(t_i + dt, y_i + dy) \cap \bigcup_{j\geq 1} A_i\right\} \neq \varnothing\right)$ which we'll write in shorthand as $d\mathbb{P}\left(\bigcap_{i=1}^m (t_i, y_i) \in \bigcup_j A_j\right)$. (In words, this is the probability density of m specific points lying on the curves.) That quantity can be written as a determinant via

$$d\mathbb{P}\left(\bigcap_{i=1}^m (t_i, y_i) \in \bigcup_j A_j\right) = \det\left[K(y_i, t_i; y_j, t_j)\right]_{1\leq i,j\leq m} dy\, dt,$$

where the kernel is given by the formula

$$K(x, s; y, t) = \begin{cases} \displaystyle\int_0^\infty e^{\lambda(t-s)}\mathrm{Ai}(x+\lambda)\mathrm{Ai}(y+\lambda)d\lambda & \text{if } s \geq t, \\ \displaystyle -\int_{-\infty}^0 e^{\lambda(t-s)}\mathrm{Ai}(x+\lambda)\mathrm{Ai}(y+\lambda)d\lambda & \text{if } s < t. \end{cases}$$

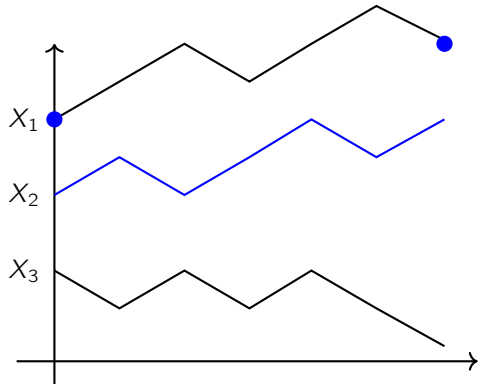
Notice that if we look along a single slice ($s = t$), this should all look familiar and we recover everything we've seen before: indeed, we have

$$K(x, t; y, t) = \int_0^\infty \mathrm{Ai}(x+\lambda)\mathrm{Ai}(y+\lambda)d\lambda = K_{\mathrm{Ai}}(x, y)$$

independent of t. And because the kernel only depends on $(t - s)$, this indeed shows translation-invariance of the process.

Remark 104. *As a sidenote, this line ensemble is no longer a determinantal point process in the locally finite sense, but we can always take a mesh to, and perhaps we can also think of it as a point process on functions via some reasonable topology. And in fact going from the kernel to the line ensemble is not actually trivial – it doesn't guarantee continuity in some natural sense, and we actually don't do it through determinantal point processes.*

Moving to a more probabilistic description now, recall that the parabolic Airy line ensemble $(A_1(x) - x^2, A_2(x) - x^2, \cdots)$ is constructed as a limit of nonintersecting paths coming from the tilings. We'll rotate the hexagon tiling 90 degrees so that the dotted-line paths are actually evolving horizontally. Knowing the random walks allows us to reconstruct the hexagon tilings (because the two types of increments give us the two types of horizontal – now vertical – rhombi, and then we fill in the gaps). Thus the bijection lets us think about nonintersecting random walks equivalently to lozenge tilings.



Well, let the top walk be X_1, the next one X_2, and so on. Suppose now that we condition on X_2 and also on the values of X_1 at two points s and t. We are taking the uniform measure on tilings, so the conditional law of X_1 in between s and t will then be a uniform (simple) random walk connecting $X_1(s)$ and $X_1(t)$. Thus in the limit we'll have a **Brownian bridge conditioned to stay above the other curves**, and this fact is sometimes called the **Gibbs property** (specifically for the tiling).

Remark 105. *Suppose we have a bunch of points $x_1, \cdots, x_n$ and we want to form Brownian motions from $(x_i, 0)$ to $(x_i, 1)$ for each i, so that the paths don't intersect. Because of the nonintersection condition, these paths will get pushed outward much like they do in the other models we've discussed, and in general it'll require a PDE to compute the limit shape of how these curves spread out. But in the case of the "watermelon" where all the starting points are essentially at the same point, it turns out the horizontal slices are given by a GUE eigenvalue distribution. We'll see the connection to that soon.*

We could have just as well made the same statement for the top two curves: fixing endpoints and the value of X_3, the top two curves would then be simple random walks with the specified start and end, conditioned to stay above X_3 and to not intersect. Putting these together lets us make the following definition in the limit:

Definition 106

A family of random functions $(L_i : \mathbb{R} \to \mathbb{R})_{i \in \mathbb{N}}$ satisfies the **Brownian Gibbs property** if the following hold:

1. The paths are ordered, meaning that we always have $L_i(x) > L_{i+1}(x)$.
2. Fix any $i \leq j$. Conditioned on $L_{i-1}(x)$ and $L_{j+1}(x)$ for $x \in [s, t]$, and also conditioned on the values of $L_i, L_{i+1}, \cdots, L_j$ at $x = s$ and $x = t$, the law of $(L_k(x))_{k \in [i,j], x \in [s,t]}$ is Brownian bridges connecting $L_k(s)$ to $L_k(t)$, conditioned to not intersect each other and to also not intersect L_{i-1} and L_{j+1}.

Fact 107

The parabolic Airy line ensemble $(A_i(x) - x^2)_{i \in \mathbb{N}} = (P_i(x))_{i \in \mathbb{N}}$ has the Brownian Gibbs property. This should hopefully not be a surprise because it holds in the discrete setting for the lozenge tiling random walks, but it does turn out to be a bit technical to actually prove it.

For more or less the same reason, most of the models we've studied also have the Brownian Gibbs property. And we can go the other way around too:

Theorem 108 (Characterization theorem)

Let $(L_1, L_2, \cdots)$ be a family of random functions satisfying the Brownian Gibbs property. Assume that L_1 "is approximately parabolic" – informally, $L_1(x) = -(1 + o(1))x^2$ with high probability, and more specifically we're asking that for all ε, δ, there is some constant $C = C(\varepsilon, \delta)$ such that

$$\mathbb{P}(|L_1(x) + x^2| > \varepsilon x^2 + C) < \delta.$$

(The constant is just to have some buffer at $x = 0$.) Then up to an affine shift, $(L_i)_{i\in\mathbb{N}}$ is the Airy line ensemble, meaning that there are some real numbers ℓ, r with $(L_i(x)) \stackrel{d}{=} (P_i(x) + \ell x + r)$.

(The affine shift is clearly necessary, since the parabolic condition respects it and Brownian bridges are still Brownian bridges after affine shifts. And in most systems in practice, we can pin down the values of ℓ and r.) So the point is that we always get the Tracy–Widom whenever we have this Brownian Gibbs property, even without needing to compute the exact distributions.

Fact 109

In a general combinatorial model, what we have to do to get Tracy–Widom is show the Gibbs property in the discrete and then find the right scaling so that the limit is still Brownian, but where we also "see the second curve."

So for example in the Arctic circle, we need R^2 horizontally and R vertically; if we wanted to figure out why $R = N^{1/3}$ without knowing it ahead of time, imagine shrinking the picture down so that the circle is of order 1, so we have a cell of width $\frac{R^2}{N}$ and height $\frac{R}{N}$ near the top of the circle. We then want the box to contain some significant part of the second curve too. But since we have a parabola (of constant leading coefficient) in this rescaling, we must want $\left(\frac{R^2}{N}\right)^2 \sim \frac{R}{N} \implies R = N^{1/3}$. And the way we would actually really justify this formally in many other models is through some variational problems which end up being invariant when we pass to the limit.

Remark 110. *We might ask whether the assumption that L_1 is approximately parabolic is necessary. It turns out that there do exist collections of curves (L_i) satisfying the Brownian Gibbs property, such that $L_1(x) \sim -a|x|$ (and so in particular this is not the Airy line ensemble); in general, we can also construct things where the first K curves fall off linearly but then things are parabolic afterward. These are sometimes called* **Airy line ensembles with wanderers** *(explicit kernels are known for them), and it turns out the slopes on the two sides don't have to be the same as long as we have something of the form* $\begin{cases} -ax & \text{as } x \to \infty \\ bx & \text{as } x \to -\infty \end{cases}$ *with $a + b > 0$ (so some kind of upside-down V shape).*

Thus a natural conjecture is that if we have the "approximately parabolic" condition for L_{k+1} instead of L_1, then $(L_i)_{i\in\mathbb{N}}$ is an Airy line ensemble with wanderers up to an affine shift. A second natural conjecture is that if we're bounded between two parabolas, meaning there exist q, Q with

$$\mathbb{P}(-Qx^2 - C < L_1(x) < C - qx^2) > 1 - \delta,$$

then (L_i) are the parabolic Airy line ensemble up to some rescaling $\sigma^{-1}P_i(\sigma^2 x)$ which allows us to modify the curvature while preserving the Brownian structure, where σ can be a random variable valued within $[q, Q]$. And a third natural conjecture is that the same holds even if we only impose the upper bound.

In other words, we're conjecturing that nothing decays faster than quadratic, and we have to be a mixture of things that decay quadratically with the same rate on both sides. Indeed, this is because if we imagine taking a huge domain

of length N and length L, there will be some limit shape coming from the distribution of the number of curves above any point (conditioned on the boundary data). And this top curve has some curvature that varies smoothly along the box, so we should expect that with some regularity we get constant curvature in the quadratics too.

15 May 19, 2026

We discussed last time how to get the Airy line ensemble out of things like the PNG, and today (and next time) we'll move to a new topic of **random matrices**. The goal will be to look at the Gaussian unitary ensemble (GUE), show that it's a determinantal point process, derive its kernel, and then take limits (both in the bulk and at the edge, which will give rise to the Sine process and the Airy kernel, respectively.)

Example 111 (Gaussian unitary ensemble)
Let H be a random $N \times N$ Hermitian matrix sampled as $H \sim e^{-\mathrm{tr}(H^2)}$ relative to Lebesgue measure. In particular, if $H = (h_{ij})$, then the density of H is

$$\exp\left(-\sum_{i,j} |h_{ij}|^2\right) dh,$$

and so essentially all entries of h are complex Gaussian. So we can fill up the upper triangular entries with complex Gaussians and then fill out the rest so that it is Hermitian (recall Example 4).

Our goal will be to understand the spectrum of H; that is, we want to study the eigenvalues $(\lambda_1, \cdots, \lambda_N) \in \mathbb{R}^N$. We'll take the following without proof:

Fact 112
The probability density of the eigenvalues is given by

$$d\mathbb{P}(\vec{\lambda}) = Z_N^{-1} \prod_{1 \le i < j \le N} (\lambda_i - \lambda_j)^2 \prod_{i=1}^N e^{-\lambda_i^2} d\vec{\lambda}.$$

We wish to understand this object in the large-N limit, and to do that the first step is to realize this as a biorthogonal ensemble. Once we compute the kernel using a general formula for such ensembles, we'll be able to express the kernel as a double contour integral (like for the Schur measure) and take a limit with steepest descent methods.

Recall that we're always free to choose a reference measure μ with respect to which we have our determinants, and to have a biorthogonal ensemble we must have $\mathbb{P}(\vec{x}) \sim \det[\phi_i(x_j)]_{i,j=1}^N \det[\psi_i(x_j)]_{i,j=1}^N d\mu(x_1) \cdots d\mu(x_N)$ (leading to the correlation kernel $K(x,y) = \sum_{i,j=1}^N ((G^{-1})^T)_{ij} \phi_i(x) \psi_j(y)$ – see Proposition 66 for the details). For the Schur measure we didn't try to orthogonalize anything – we could compute the Gram matrix G explicitly using some identities – but here we'll make the analysis simpler.

In our case, we have a "decoupled" part with $\prod_{i=1}^N e^{-\lambda_i^2}$ as one of our terms, so we will take $\mu(x) = e^{-x^2}$ and we now have

$$\mathbb{P}(\vec{\lambda}) \sim \prod_{1 \le i < j \le N} (\lambda_i - \lambda_j)^2 \prod d\mu(\lambda_i).$$

Now the Vandermonde term (without a square) is already itself a determinant because $\prod_{1 \le i < j \le N} (\lambda_i - \lambda_j) = \det\left[\lambda_j^{i-1}\right]_{i,j=1}^N$. So we could take $\phi_i(x) = \psi_i(x) = x^{i-1}$, but this isn't quite as nice as it could be because powers of x are not orthogonal

under the Gaussian.

So to avoid carrying the factors coming from the Gram matrix G, we'll instead change basis and use the orthogonal polynomials With respect to the Gaussian measure. If f_i are polynomials of degree i and leading coefficient a_i, then $\det[f_i(x_j)]_{i,j=1}^N = \prod_{i=1}^N a_i \prod_{1\le i<j\le N}(x_i - x_j)$ by rescaling the leading coefficients and then using row reduction to get rid of the lower-order terms (by subtracting earlier rows from later ones to get rid of constants, then linear terms, and so on). That means we are indeed free to pick any family of polynomials that we want instead of just monomials, and the one we'll use is the Hermite polynomials:

Definition 113

The **Hermite polynomials** H_n are given by

$$H_n(x) = (-1)^n e^{x^2} \frac{d}{dx^n}\left(e^{-x^2}\right).$$

For example, $H_0(x) = 1$, $H_1(x) = 2x$, $H_2(x) = 4x^2 - 2$, and in general H_n will be a polynomial with leading term $2^n x^n$.

Lemma 114

For any m, n,

$$\int_{-\infty}^{\infty} H_m(x)H_n(x)e^{-x^2}dx = \begin{cases} 0 & \text{if } m \neq n, \\ 2^n n!\sqrt{\pi} & \text{if } m = n. \end{cases}$$

Proof. Plugging in the expression for H_m, we have that

$$\int_{-\infty}^{\infty} H_m(x)H_n(x)e^{-x^2}dx = (-1)^m \int_{-\infty}^{\infty} \frac{d}{dx^m}\left(e^{-x^2}\right) H_n(x)dx,$$

and now we can integrate by parts to push the m derivatives from one term to the other to get

$$\int_{-\infty}^{\infty} e^{-x^2} \frac{d}{dx^m}\left(H_n(x)\right) dx.$$

So in particular, if $m > n$ this is zero because H_n is of degree n and we take more than n derivatives; similarly we can just swap the roles of m and n if $m < n$. So the only remaining case is $m = n$, which would leave us with

$$\int_{-\infty}^{\infty} e^{-x^2} \frac{d}{dx^n}\left(2^n x^n + \cdots\right) dx = \int_{-\infty}^{\infty} e^{-x^2}(2^n n!)dx,$$

and the remaining Gaussian integral exactly gives us the factor of $\sqrt{\pi}$. □

Thus we have the biorthogonal ensemble representation (absorbing the leading-term factors into the proportionality)

$$\mathbb{P}(\vec{\lambda}) \sim \det[H_{i-1}(x_j)]_{i,j=1}^N \det[H_{i-1}(x_j)]_{i,j=1}^N d\mu(\vec{x}).$$

The Gram matrix is then diagonal, and our kernel will then be given by

$$K(x,y) = \sum_{n=0}^{N-1} \frac{H_n(x)H_n(y)}{2^n n!\sqrt{\pi}}$$

relative to the Gaussian measure. We now want to write this in a way that lets us apply steepest descent. Right now the formula for H_n requires us to take a lot of derivatives, but now we can use some of the other properties of the Hermite polynomials:

Fact 115

We have the following useful facts:

1. $H_n(x) = n! \sum_{m=0}^{\lfloor n/2 \rfloor} \frac{(-1)^m}{m!} \frac{(2x)^{n-2m}}{(n-2m)!}$. (This can be shown by induction.) In particular, the degrees of the nonzero monomials all have the same parity as n.
2. The following three-term recursion holds (and can be proven by integration by parts, but something general holds for orthogonal polynomial families):
$$H_{n+1}(x) - 2xH_n(x) + 2nH_{n-1}(x) = 0.$$
3. H_n satisfies the differential equation $H_n''(x) - 2nH_n'(x) + 2nH_n(x) = 0$ (replacing varying indices with varying derivatives).
4. $H_n(-x) = (-1)^n H_n(x)$ for all n.
5. $H_n'(x) = 2nH_{n-1}(x)$.
6. The generating function for the H_ns is given by
$$\sum_{n=0}^{\infty} H_n(x) \frac{z^n}{n!} = e^{2xz - z^2}.$$

The last of these properties will be the most useful for us – indeed, we used something similar with the case of the Schur measure. Indeed, the left-hand side of (6) is $e^{x^2} \sum_{n=0}^{\infty} \frac{(-z)^n}{n!} \frac{d}{dx^n}(e^{-x^2})$, but this is exactly the Taylor expansion of $f(x) = e^{-x^2}$ for $f(x-z)$. Therefore we can integrate both sides of the generating function expression and fish out the degree-n term appropriately:

Corollary 116

We can write the Hermite polynomials as contour integrals
$$H_n(x) = \frac{n!}{2\pi i} \oint \frac{e^{2xz - z^2}}{z^{n+1}} dz,$$
where the contours are any counterclockwise curves around 0.

Another integral representation will also be of use so that the double contour we end up with is a little easier to compute with:

Lemma 117

We also have the expression
$$H_n(x) = \frac{2^n e^{x^2}}{i\sqrt{\pi}} \int_{-i\infty}^{i\infty} s^n e^{s^2 - 2sx} dx,$$
where the contour is over the imaginary axis.

Proof. The idea will be to swap out the complicated expression e^{-x^2} (which is hard to take a derivative of) with something else with a simpler dependence on x:
$$e^{-x^2} = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{2itx - t^2} dt.$$

(This just comes out of Gaussian integration and completing the square, or equivalently this is the characteristic function of the Gaussian.) Therefore by changing variables $s = it$,

$$\begin{aligned}\frac{d^n}{dx^n}(e^{-x^2}) &= \frac{d}{dx^n}\left(\frac{1}{i\sqrt{\pi}}\int_{+i\infty}^{-i\infty} e^{2sx+s^2}ds\right)\\ &= \frac{2^n}{i\sqrt{\pi}}\int_{i\infty}^{-i\infty} s^n e^{2sx+s^2}ds,\end{aligned}$$

and then swapping $s \mapsto -s$ yields the desired claim. □

Therefore we can plug in this integral representation into our formula for the kernel (now relative to the Lebesgue measure instead of the Gaussian measure)

$$K(x,y) = \sum_{n=0}^{N-1} \frac{H_n(x)H_n(y)}{2^n n!\sqrt{\pi}} e^{-(x^2+y^2)/2}.$$

(The factor of 2 makes sense since $K(x,x)$ should be relative to $\mu(x) = e^{-x^2}$, and in general we need to split the Gaussian density across x and y.) We find, using the two different expressions for $H_n(x)$ and $H_n(y)$, that

$$K(x,y) = \frac{e^{-(x^2+y^2)/2}}{\sqrt{\pi}} \sum_{n=0}^{N-1} \left(\frac{2^n e^{y^2}}{i\sqrt{\pi}}\int_{-i\infty}^{i\infty} s^n e^{s^2-2sy}ds\right)\left(\frac{1}{2^n n!}\frac{n!}{2\pi i}\oint \frac{e^{2xz-z^2}}{z^{n+1}}dz\right).$$

(The reason for using the two expressions is that notice we now have a $+2xz$ and a $-2sy$ in the exponentials, and having the difference there is similar to how we did the $e^{-t(S(u)-S(v)}$ formulation in LIS before.) Simplifying a bit and swapping the order of integration yields

$$K(x,y) = \frac{e^{(y^2-x^2)/2}}{2(\pi i)^2}\int_{-i\infty}^{i\infty}\oint e^{s^2-z^2-2sy+2xz}\sum_{n=0}^{N-1}\frac{s^n}{z^{n+1}}dzds$$

Here the integral over the imaginary axis is uniformly converging, and the contour integral is over a compact interval with no poles, so swapping the integrals and sum causes no problems. The resulting kernel after computing the geometric sum is

$$\frac{e^{(y^2-x^2)/2}}{2(\pi i)^2}\int_{-i\infty}^{i\infty}\oint e^{s^2-z^2-2sy+2xz}\frac{s^n}{z^{n+1}}\frac{dzds}{z-s}\left(1-\left(\frac{s}{z}\right)^N\right).$$

This expression is okay; however, to avoid having to worry about this N dependence in this particular way (and so that we don't have issues with the tail, since actually s has to go off to infinity), we can instead massage things and "run the sum back to $-\infty$." Extending our finite sum to start from $n = -\infty$, we must take our contours so that $|z| < |s|$, so say that z is some small contourclockwise loop around 0 and s is from $\Delta - i\infty$ to $\Delta + i\infty$. Under those contours, adding in the negative n integer terms because the integrand $e^{s^2-z^2-2sy+2xz}\frac{s^n}{z^{n+1}}$ has no poles for negative values of n. We thus get

$$\begin{aligned}K(x,y) &= \frac{e^{(y^2-x^2)/2}}{2(\pi i)^2}\oint\int_{\Delta-i\infty}^{\Delta+i\infty} e^{s^2-z^2-2sy+2xz}\sum_{n=-\infty}^{N-1}\frac{s^n}{z^{n+1}}dsdz\\ &= \boxed{\frac{e^{(y^2-x^2)/2}}{2(\pi i)^2}\oint\int_{\Delta-i\infty}^{\Delta+i\infty} e^{s^2-z^2-2sy+2xz}\left(\frac{s}{z}\right)^N\frac{1}{s-z}dsdz}\end{aligned}$$

after computing this new geometric sum. From here, we are free to remove the $e^{(y^2-x^2)/2}$ term, since this amounts to conjugating the kernel K by the diagonal matrix $\mathrm{diag}(e^{-x^2/2})$, which doesn't change any determinants. So what we

will really work with is

$$\boxed{K(x,y) = \frac{1}{2(\pi i)^2} \oint \int_{\Delta - i\infty}^{\Delta + i\infty} \frac{e^{s^2 - z^2 - 2sy + 2xz}}{s - z} \left(\frac{s}{z}\right)^N ds dz}$$

This is what we'll deal with next time; we'll take limits similarly to the Schur measure to get different kernels in the bulk and in the edge.

16 May 21, 2026

We discussed the kernel for the GUE last time, getting it into a double contour expression. We'll write it this time as

$$K(x,y) = \frac{1}{2(\pi i)^2} \oint \int_{\Delta - i\infty}^{\Delta + i\infty} \left(\frac{w}{z}\right)^N \frac{e^{w^2 - 2wy}}{e^{z^2 - 2xz}} \frac{dwdz}{w - z},$$

where the z-contour is a counterclockwise circle around the origin and the w-contour is a vertical contour with $\Delta > |z|$. Today's goal will be to analyze this kernel in the large-scale limit $N \to \infty$, and we'll now get a continuous version of the sine kernel instead of a discrete one in the bulk, as well as the Airy kernel at the edge (which automatically gets us Tracy–Widom).

We'll be a lot more informal this time and be more heuristic about how the limits are actually taken, since we did it more carefully last time. The first step is to put everything under an exponential and see the critical points; calling the integrand $e^{I_N(w,z)} \frac{dwdz}{w-z}$, we have

$$I_N(w,z) = N(\log w - \log z) + w^2 - z^2 + 2xz - 2wy.$$

We want all terms in this to be of the same order, so we'll change variables $w = \sqrt{N}\tilde{w}$, $z = \sqrt{N}\tilde{z}$ to match the w^2, z^2 terms. We want to probe some point on our spectrum, and we gave a heuristic earlier that order $\sqrt{N}$ is the general width over which eigenvalues appear. So we'll want to probe at a particular location $c\sqrt{N}$, and because there are N eigenvalues we expect the spacings to be roughly of order $\frac{1}{\sqrt{N}}$. That motivates us to also scale $x = c\sqrt{N} + \frac{\tilde{x}}{\sqrt{N}}, y = c\sqrt{N} + \frac{\tilde{y}}{\sqrt{N}}$, yielding

$$I_N(\tilde{w}, \tilde{z}) = N(\log \tilde{w} - \log \tilde{z} + \tilde{w}^2 - \tilde{z}^2 + 2c\tilde{z} - 2c\tilde{w}) + 2\tilde{x}\tilde{z} - 2\tilde{y}\tilde{w}.$$

The idea will be that as we vary $\tilde{x}$ and $\tilde{y}$ by a constant, the integrand will actually see that effect but not in a way that dominates everything else. So to get it in the same form as with LIS, we define $S(x) = \log x + x^2 - cx$ so that

$$I_N(\tilde{w}, \tilde{z}) = N(S(\tilde{w}) - S(\tilde{z})) + 2\tilde{x}\tilde{z} - 2\tilde{y}\tilde{w}.$$

Computing the locations of the critical points, we have

$$S'(x) = \frac{1}{x}\left(1 + 2x^2 - 2cx\right),$$

and now observe that if $c = \sqrt{2}$ this simplifies as $\frac{1}{x}(x\sqrt{2} - 1)^2$. So this indicates that we get a double root at $x = \frac{1}{\sqrt{2}}$, which is kind of the analog of being at the edge before.

Example 118

We'll now take $|c| < \sqrt{2}$ and see that we get the sine kernel (which will then essentially indicate that for $|c| > \sqrt{2}$ the density is zero, though we won't go through the details of those calculations here).

Being much less careful than before, the roots of S' are at

$$\phi_\pm = \frac{c \pm \sqrt{c^2-2}}{2},$$

and for $c < \sqrt{2}$ these are always on the radius-$\frac{1}{\sqrt{2}}$ circle in the complex plane (we can imagine a similar picture to the one in Example 86 with those two roots). So we'll take $\tilde{w}$ and $\tilde{z}$ "close to a level curve of S" and pinch the curves closer or further away based on the sign of S'; it turns out that we pull $\tilde{w}$ in to the left of $\phi_\pm$ and $\tilde{z}$ in to the right. (But we do still need to make sure $\tilde{z}$ goes to $-i\infty$ and $+i\infty$.)

Once we do this, we have $\mathrm{Re}(S(\tilde{w})) < \mathrm{Re}(S(\tilde{z})) - \varepsilon$ except in a δ-ball around $\phi_\pm$. And when we integrate an order-1 pole $\frac{1}{\tilde{w}-\tilde{z}}$ over two-dimensional space, the price we must pay is the pole at $\tilde{w} = \tilde{z}$ we get at ϕ_- and ϕ_+. We are thus working with something of the form

$$\frac{1}{2(\pi i)^2} \oint \int e^{N(S(\tilde{w})-S(\tilde{z}))} \frac{e^{2\tilde{x}\tilde{z}-2\tilde{y}\tilde{w}}}{\tilde{w}-\tilde{z}} d\tilde{w} d\tilde{z},$$

and the resulting residue at $\tilde{w} = \tilde{z}$ is

$$\frac{1}{(\pi i)} \oint_{\phi_-}^{\phi_+} e^{2(\tilde{x}-\tilde{y})\tilde{w}} d\tilde{w}.$$

We can now evaluate this exactly: letting $\hat{w} = 2\tilde{w}(\tilde{x}-\tilde{y})$ yields

$$\frac{1}{2\pi i(\tilde{x}-\tilde{y})} \int_{2\phi^-\cdot(\tilde{x}-\tilde{y})}^{2\phi^+\cdot(\tilde{x}-\tilde{y})} e^{\hat{w}} d\hat{w} = \frac{\sin(2\theta(x-y))}{\tilde{x}-\tilde{y}},$$

where $\theta = \mathrm{Im}(\phi_+)$. The factor of 2 here is really just up to how we choose our scaling; it's qualitatively irrelevant because it basically comes down to the window at which we scale $\tilde{x}$ and $\tilde{y}$. (This was not true in the discrete setting where the point process was on the integers.)

Fact 119

To see more of a connection between the continuum and discrete, we can consider the measure on $(\lambda_i) \subseteq \mathbb{R}$ proportional to $\prod(\lambda_i - \lambda_j)^{2\beta} \prod V(\lambda_i)$ (where V is the Gaussian density) and take $\beta \to \infty$. Such measures are sometimes called β-ensembles and what we're doing is the $\beta = 1$ case.

This is no longer a determinantal point process, but it turns out the eigenvalues will locally crystallize. That is, if we look at $x = c\sqrt{N}$ and then look at the points in the process around that point (doing an appropriate scaling), everything will actually be equally spaced out (since we're essentially maximizing $\prod(\lambda_i - \lambda_j)$ under some kind of confinement from the Gaussian).

Remark 120. *We might also ask what happens if we restrict λ_i to the lattice and consider the same measure. Then this local crystallization can't happen, and actually it's not known what the local point process looks like – we might expect that it's a rounding of the even-spacing situation, but even that's not known.*

So far we've been looking at point processes on the line, but these models would also make sense in higher-dimensional domains. The following is sometimes called the Abrikosov conjecture:

Conjecture 121

Take $z_i \in \mathbb{C}$ distributed according to the density

$$Z^{-1} \prod_{1 \le i < j \le N} |z_i - z_j|^{2\beta} \prod_{i=1}^{N} V(z_i)$$

for any confining V of our choice (for example $e^{-|z|^2}$), and send $N \to \infty$, then $\beta \to \infty$. Then zooming in at a point z_0 to some scale where we expect to see $O(1)$ many particles, we will see a triangular lattice.

The triangular lattice is the optimal choice of centers for 2-dimensional sphere packing, and we also know a lot about sphere packings in dimensions $d = 8, 24$. Interestingly, the analog of this conjecture is actually known in those specific higher dimensions but not for $d = 2$ (we will converge to the Leech or E_8 lattices in those cases), and that uses "universal optimality."

Remark 122. *This conjecture is also meaningful on other manifolds; in fact it has some connections to complex geometry. The phenomenon is purely local, so the same statement makes sense and is expected to still be true. And we might also want to understand the behavior of the normalizing constant Z, which has some interesting connections as well. We'll make some comments about that next time.*

Remark 123. *We said above that $\theta = Im(\phi_+) \sim \sqrt{2 - c^2}$, and indeed θ is supposed to indicate the density of eigenvalues around $c\sqrt{N}$. And this is consistent with Wigner's semicircle law.*

Example 124

We'll now turn to studying the GUE at the edge. Scaling around $\sqrt{2N}$, the correct thing to do will be to set $x = \sqrt{2N} + N^{-1/6}\frac{\tilde{x}}{\sqrt{2}}$, $y = \sqrt{2N} + N^{-1/6}\frac{\tilde{y}}{\sqrt{2}}$. (Here $N^{-1/6}$ should be thought of as $\frac{\sqrt{N}}{N^{2/3}}$.)

Using the same scaling for w, z by $\sqrt{N}$, we thus want to calculate the double integral

$$\frac{N^{\cdots}}{2(\pi i)^2} \oint \int e^{N(S(\tilde{w}) - S(\tilde{z}))} \frac{e^{2(\tilde{x}\tilde{z} - 2\tilde{y}\tilde{w})N^{\cdots}}}{\tilde{w} - \tilde{z}} d\tilde{w} d\tilde{z}$$

where the $N^{\cdots}$ are some powers of N which we won't track out carefully. This time we have a double root $\phi_+ = \phi_- = \frac{1}{\sqrt{2}}$, and our contours will generally follow level lines but only be very close at this one point. If we then further rescale $\tilde{w}$ and $\tilde{z}$ around here, we take

$$\tilde{z} = \frac{1}{\sqrt{2}} + N^{-1/3}\zeta, \quad \tilde{w} = \frac{1}{\sqrt{2}} + N^{-1/3}\xi.$$

Indeed, the first two derivatives of S vanish, so it's the third derivative that will appear, and we want that to cancel out with the N in the exponential so that

$$N(S(\tilde{w}) - S(\tilde{z})) = N \cdot \frac{1}{6} S'''\left(\frac{1}{\sqrt{2}}\right)\left((N^{-1/3}\xi)^3 - (N^{-1/3}\zeta)^3\right) + \cdots.$$

Skipping the prefactors of N and so on, we end up wanting to use the 60-degree contours around our center point with ξ on the right and ζ on the left, and what we end up with looks like

$$\frac{N^{\cdots}}{2(\pi i)^2} \oint \oint \exp\left(\frac{S'''(\frac{1}{\sqrt{2}})}{6}(\xi^3 - \zeta^3)\right) \frac{\exp\left(\sqrt{2}(\tilde{x} - \tilde{y})N^{\cdots} + 2(\zeta\tilde{x} - \xi\tilde{y})N^{\cdots}\right)}{\xi - \zeta} d\zeta d\xi.$$

Now remember that the $\exp\left(\sqrt{2}(\tilde{x} - \tilde{y})N^{\cdots}\right)$ term is like a "diagonal conjugation" of our matrix, so it doesn't affect the determinant and we can ignore it. And it turns out the power of N in $2(\zeta\tilde{x} - \xi\tilde{y})N^{\cdots}$ was rigged to exactly be zero,

so that cancels out, and the leading order term ends up being $N^{-1/3}$. So indeed we have

$$\frac{N^{-1/3}}{2(\pi i)^2}\oint\oint \frac{e^{\text{const}(\xi^3-\zeta^3)+2(\zeta\tilde{x}-\xi\tilde{y})}}{\xi-\zeta}d\xi d\zeta,$$

which is exactly the Airy kernel up to rescaling by a constant! So in particular, the largest eigenvalue λ indeed satisfies

$$\mathbb{P}(\lambda_1 > \sqrt{2N} + N^{-1/6}s) \to \text{TW}.$$

Remark 125. *The Ginibre ensemble is also determinantal; it's similar to the β-ensemble we mentioned and it will again be biorthogonal. But we won't see Tracy–Widom; it's something with different scaling exponents that we get by localizing at the boundary of the circle. In particular notice that the eigenvalues will be spaced out differently in two-dimensional space than on a one-dimensional interval.*

17 May 26, 2026

We'll start with a student presentation:

Stanislav Krymskii

The title of this talk is "On the stationary measure for the open KPZ equation."

To start, we'll describe the KPZ equation: it's the stochastic PDE on $[0,1]$

$$\dot{h} = \frac{1}{2}(h'' + (h')^2) + \xi$$

where ξ is white noise, and with boundary conditions $h'(0,t) = u$ and $h'(1,t) = -v$. This equation can be interpreted either as a way to stochastically transport h' from left to right, or as a way to grow an interface while smoothed by diffusion.

It turns out that $h(x) - h(0)$ has a limit (even though h itself doesn't have a limit as $t \to \infty$, since the interface will move to infinity). This SDE is hard to solve, but if we represent $h(t,x) = \log z$ then we can write down a **Hopf–Cole solution** to the semilinear SDE. This allows us to use classical methods and relate to a continuous random polymer model to study instead.

Corwin and Knizel managed to obtain the precise solution and connect things to the open ASEP model, which we'll describe later. Those authors found that for each boundary conditions (u,v), there is a stationary measure for the open KPZ equation, and there is an exact formula in the case where $u+v \geq 0$. We can find that exact limit using a discrete process. (Furthermore we can bound $h_{u,v}$ between $h_{u,-u}$ and $h_{-v,v}$ via a coupling.)

Remark 126. *To describe the process $h_{u,v}(x)$, one way is to write it in terms of the Brownian motion as $B(x) + Y(0) - Y(x)$, where Y has some tilt* $\exp\left(2uY(0) + 2vY(1) - \int_0^1 e^{2Y(s)}ds\right)$ *relative to Brownian motion. And there are other formulas as well.*

We'll now apply the open ASEP to the KPZ equation:

Definition 127

In the **open ASEP** interacting particle system, we have a one-dimensional grid of sites where particles try to jump into unoccupied neighboring sites. Specifically, if a site has a particle but its right neighbor does not, then jumps happen at a rate 1; similarly if the left neighbor is empty then jumps happen at a rate q. Furthermore, some sites (the leftmost and rightmost) ones produce or destroy particles; specifically, there are some probabilities α, β of appearing and disappearing at site 0 and probability γ, δ of appearing and disappearing at site N.

This ASEP model generates a height function $h_N(t,x) = -2N(t) + \sum(2\tau_i(t) - 1)$, where $N(t)$ is the number of particles and τ_i is the indicator of whether the particle is present at site i or not. Such a height function can be thought of as a broken line with steps $(1,1)$ or $(1,-1)$. So then jumps correspond to adding or removing a rotated "box" from this function, and creating or disappearing particles correspond to flipping the leftmost or rightmost line segment up or down.

Fact 128

The limit behavior of ASEP can be described in terms of a phase diagram of the stationary density of the particles at the left and right boundaries, which we call ρ_L, ρ_R. (Here we're taking $N \to \infty$ but keeping $\alpha, \beta, \gamma, \delta$ fixed.) There are exact formulas for these densities:

- If $\rho_L > \frac{1}{2}, \rho_L < \rho_R < \frac{1}{2}$, then we have the "maximal current" phase where the current $J = \frac{1}{4}$ is maximized.
- In the "shock region" $\rho_L < \rho_R$, particles concentrate before being absorbed.
- In the region $\rho_L > \rho_R$, the density gradually decreases over time.

We can perform the **microscopic Hopf–Cole transformation** to turn this into an approximate solution to the KPZ equation

$$Z_t(x) = e^{\nu t - \lambda h_N(x)}, \quad \nu = (1 - \sqrt{q})^2, \quad \lambda = \ln \sqrt{q}.$$

The best known way to connect to the SHE currently requires the **Liggett condition** to hold, which is some condition between $\alpha, \beta, \gamma, \delta$. Then the "reservoirs" become standard ASEP sites, and we satisfy a certain discrete stochastic heat equation. But we also need to scale space, time, and our parameters accordingly; specifically take $q = e^{-2/\sqrt{N}}$ (which is $1 + O(N^{-1/2})$) and consider

$$h^{(N)}(t,x) = N^{-1/2} h_N(e^{N^{-1/2}} N^2 t/2, Nx) + (N^{-1}/2 + 1/24)t.$$

This then also requires boundary densities to be very close to $\frac{1}{2}$, and specifically we have $\rho_L = \frac{1+uN^{-1/2}}{2} + o(N^{-1/2})$ and $\rho_R = \frac{1-vN^{-1/2}}{2} + o(N^{-1/2})$. Then with Holder bounds on the initial data, we get convergence as $N \to \infty$. (Specifically, $h^{(N)}$ as a process converges to the KPZ solution, and we also have convergence to the stationary measure.)

The authors expect that uniqueness can be provable by studying the mixing time of ASEP, because if it is $O(N^{3/2}/(1-q))$ then after time N^2 the distribution tends to the stationary measure. So the expectation is that with $t = N^2\tau$ and taking $\tau \to \infty$, we will converge to the stationary measure, which would allow us to show that the continuous case also has convergence.

Returning now to the asymptotics of open ASEP now, in the case $\rho_L = \rho_R$ we have the Bernoulli stationary measure where all sites have independent occupation probabilities. We know that if we increase coefficients of entrances or decrease the probabilities of exits, then the measure will become "more filled." But if they differ from $O(N^{-1/2})$ from the equilibrium, then Holder sandwiching of the ASEP measure allows the KPZ measure to be constructed as a

subsequential limit. So perturbing using monotonicity is the key step here, and uniqueness comes from the Laplace transform.

Finally, concluding Professor Aggarwal's part of the course, we'll instead conclude with a bit of a summary of the course and describe some of the ways where we could have alternatively approached these topics. This will witness a lot of the talks we'll give instead.

We started with LPP and TASEP and the desire to show convergence to Tracy–Widom; under a specific limit we have the LIS model. The way we did this was to study the multi-layer PNG, for which we found a bijection to non-intersecting paths and thus a determinantal formula via the LGV lemma. Specifically, we introduced Schur polynomials and Schur processes (whose arguments were the number of "up and down steps" in the walk paths). Via the framework of biorthogonal ensembles, we were able to understand what these Schur processes did (via symmetric function theory), and the key fact is that biorthogonal ensembles are DPPs with explicit kernels.

We thus had tools to study most of these objects, and once we got those exact formulas we could use steepest descent to get asymptotics. And then we saw that this whole story also holds for random matrices in the GUE model; in both cases we found Tracy–Widom behavior. And we also found an equivalence of nonintersecting paths with tilings, showing that the Airy line ensembles come up in the limit essentially because of the Gibbs property. So this Gibbs property is another tool for accessing limits (though we didn't actually do any mathematics with it).

Remark 129. *We only did PNG / LPP / TASEP for a very specific case, but for any choice of geometric a_i, b_j parameters we get a Schur measure and a biorthogonal ensemble. But if we don't have geometrics the step that's broken is that we don't get Schur measures in the first place.*

This is not the only way in which we could have approached a lot of these models, though. One interesting thing about this perspective is that there are many different-looking models which are all somehow equivalent, and if we use a different set of tools and expand our strategies.

For example, we can think about "fixing TASEP and asking what tools can be used to attack it," and then move on to other models from there. TASEP is the special case where particles can only move to the right, and we just saw ASEP in Stanislav's talk (which is no longer a DPP). What we instead use to study ASEP is the framework of **vertex models**, which in particular lets us pivot to the KPZ equation (a continuum PDE) by taking q a very specific nonzero quantity instead of zero. Vertex models are often studied via the **quantum inverse scattering method**.

Another pivot we can make is to the **box-ball system**, which is an interacting particle system with totally deterministic behavior but still with nontrivial limits. This is a special case of a family of deterministic integrable systems (perhaps with random initial conditions); another of these is the KdV equation which models water waves. The tool used for such systems is the **inverse scattering method** (the limit of the quantum one), which was somehow the "origin of integrability."

Additionally, we can more generally consider non-intersecting paths (families of lattice paths with some weights). We've seen so far that non-intersecting paths arise from lozenge tilings, but we can also consider random tilings on various lattices and get similar bijections. In particular, we get the Kasteleyn theory, which will be useful for dimer models in general. On the other hand, taking the limit of these paths in the continuum gets us to Dyson Brownian motion, which can be studied via a limit of DPPs or via "homogenization." That's what gets us to the GUE by thinking of how the eigenvalues evolve in time, and indeed we took for granted the distribution of the eigenvalues when describing the GUE (this is a certain change-of-variables calculation). A similar change-of-variables formula is the Hammersley (Kac–Rice) formula for computing the number of zeros of an analytic function; thus we can study other functions besides the characteristic function of a random matrix, in particular random analytic functions (power series with random coefficients).

Finally, yet another way to access results for TASEP if we care about limit shape (law-of-large-numbers) results is the family of tools for hydrodynamic methods such as the entropy method; this is purely analytic.

Many of these topics will be (or have been) talked about in student presentations.

18 May 28, 2026

Andrew Lin

I gave this talk, so I don't have notes, sorry! I discussed some parts of Sections 3, 5, and 6 of Kenyon's notes "Lectures on dimers," describing the Kasteleyn weighting and how it yields a determinantal point process, and then describing how we get formulas for the partition function and the inverse Kasteleyn matrix in terms of contour integrals.

James Stephens

This talk will be about **hydrodynamics and relative entropy**. We'll begin with a model of interacting particles, introduce the basic notions of getting laws of large numbers for it, and then describe a heuristic argument for getting a limiting PDE (along with some words about the proof).

Example 130

We'll consider identical particles on the periodic lattice $\mathbb{T}_N^d = \{0, \cdots, N-1\}^d$, where we can have multiple particles on a site so that our configuration space is $\mathbb{N}^{\mathbb{T}_N^d}$. Let $g : \mathbb{N} \to \mathbb{R}_+$ be a function and $p(\cdot, \cdot)$ some translation-invariant transition probabilities on $\mathbb{Z}^d$ (so $p(a, b) = p(b - a)$ for some function p). The dynamics are as follows: if we have k particles at a site x, then one of them jumps to y at rate $g(k)p(x, y)$.

The case $g(k) = k$ would correspond to many superimposed random walks with transition probability p, but if $g(k) > k$ then things will spread out. To get a limiting PDE which is second-order instead of the transport equation, we will assume that p has zero mean, meaning that

$$\sum_{x \in \mathbb{Z}^d} x_j p(x) = 0,$$

and also that there is a finite range.

We can write down the generator for this process, which is (here η denotes a configuration)

$$L_N f(\eta) = \sum_{x \in \mathbb{T}_d^N} \sum_{z \in \mathbb{T}_d^N} p(z) g(\eta(x)) \left[f(\eta^{x, x+z}) - f(\eta) \right]$$

where $\eta^{x,x+z}$ is η but with a particle at x replaced with one at $x + z$.

The first task for understanding the limiting behavior of this is to understand the invariant measure; this can be written down explicitly. We have

$$Z(\phi) = \sum_{k \geq 0} \frac{\phi^k}{g(1) \cdots g(k)},$$

which has some finite radius of convergence under some assumptions on g, and then we define $\overline{\nu}_{\phi,g}^N$ to be the product measure on the configuration space $\mathbb{N}^{\mathbb{T}_N^d}$ with marginals such that $\eta(x) = k$ with probability $\frac{1}{Z(\phi)} \frac{\phi^k}{g(1) \cdots g(k)}$. (So if

$g(k) = k$ the marginals are Poisson distributed at independent, which is what should happen.) This can explicitly be computed to be the invariant measure for those dynamics mentioned before.

We then define the **occupation variable**

$$R(\phi) = \mathbb{E}_{\overline{\nu}^N_{\psi,g}}[\eta(0)].$$

Letting $\Phi = R^{-1}$ yields a one-parameter family $\nu^N_\alpha = \overline{\nu}^N_{\Phi(\alpha)}$ of "global equilibrium" (invariant and translationally-invariant product measures), where α is somehow the density of particles. This is nice because Markovian dynamics preserve this number, and this is kind of the "usual situation" that we get in systems like this.

The philosophy now is that systems of interacting particles need not live on this global equilibrium, but locally they should look like this. Thinking of $\mathbb{T}^d_N$ as the "microscopic space," with $\frac{1}{N}$ the distance between particles, we will take $N \to \infty$ and understand profiles $\rho_0 : \mathbb{T}^d \to \mathbb{R}_+$. If $\nu^N_{\rho_0(\cdot)}$ is a measure on Π^d_N whose marginals are such that $\eta(x) = k$ has the same marginals as $\nu_{\overline{\rho_0}(x/N)}$, this gives us an initial condition with "slowly varying density" from which we will run our particle system.

The strongest sense of local equilibrium is that for some family of measures $\{\mu^N\}$, we have

$$\lim_{N\to\infty} \tau_{[uN]}\mu^N = \nu_{\rho_0}(u)$$

where $\tau_{[uN]}$ is translation for some $u \in \mathbb{T}^N$. That is, if we zoom in at the point uN, we get the continuous measure at point u. (This is kind of like how we zoomed in around a point ct for the Schur measure; instead of the sine kernel we get some product measure which is somehow universal because it is the stationary measure for the dynamics.) Proving such a statement is very hard, so one usually proves a weaker notion instead, which takes the form

$$\lim_{N\to\infty} \mu^N \left[\left| \frac{1}{N^d} \sum_{x\in\mathbb{T}^d_N} G(x/N)\tau_x\Psi(\eta) - \int_{\mathbb{T}^d} G(u)\mathbb{E}_{\nu_{\rho_0(u)}}(\Psi) \right| > \delta \right] = 0$$

for all δ, all continuous functions G, and bounded cylinder functions Ψ depending only on a finite number of coordinates. That is, as $N \to \infty$, if we zoom in at the point u and average nearby (think of G as a bump function near x), we get the corresponding thing in the continuum. For example, if we take $\Psi(\eta) = \eta(0)$ (it's not bounded, but oh well), then this condition really reads

$$\lim_{N\to\infty} \mu^N \left[\left| \frac{1}{N} \sum_{x} G(x/N)\eta(x) - \int_{\mathbb{T}^d} G(u)\overline{\Psi} \right| > \delta \right] = 0,$$

which occurs if empirical distributions $\frac{1}{N^d}\sum_x \eta(x)\delta_{x/N}(du)$ converge weakly in probability to $\nu_{\rho_0}(\cdot)$ (the histograms should look like our limiting measure).

We can now write down the limiting equation. Defining the matrix of correlations

$$\sigma_{i,j} = \sum_{x\in\mathbb{Z}^d} x_i x_j p(x)$$

allows us to define a second-order differential operator $\sum_{1\le i,j\le d} \sigma_{i,j}\partial_{u_i}\partial_{u_j}$, and then we can define the expected rate at which things leave the origin when we're at density α

$$\Phi(\alpha) = \mathbb{E}_{\nu_\alpha}[g(\eta(0))].$$

We then get the Cauchy problem

$$\begin{cases} \partial_t\rho = \frac{1}{2}\Delta_\sigma\Phi(\rho), \\ \rho(0,\cdot) = \rho_0(\cdot). \end{cases}$$

Then if $\rho_0(\cdot) \in C^{2+\varepsilon}(\mathbb{T}^d)$ and bounded above and below (meaning $0 < K_1 \leq \rho_0 \leq K_2$), then we get a solution which is C^2 in space, C^1 in time, and bounded above and below by the same constants. This will end up being the hydrodynamic limit of the interacting particle system.

So far, we've only described ρ and not the patricle system. Note that we need to speed up the dynamics to get a nontrivial limit; if we do things as we state we just get a constant. Instead, we have to speed up by N^2, and we get something weaker than local equilibrium:

Theorem 131

Let S_t^N be the evolution of the system for time $N^2 t$. Then (with G, Ψ the same as before)

$$\lim_{N\to\infty} \mathbb{E}_{S_t^N \mu^N}\left[\left||N^{-d}\sum_x G(x/N)\tau_x\Psi(\eta) - \int G(u)\mathbb{E}_{\nu_\rho(t,u)}[\Psi]du\right|\right] = 0.$$

That is, at macroscopic time t, we are at the local equilibrium of the profile $\rho(t,\cdot)$. This is only proved for some initial conditions μ^N, where the relative entropy satisfies $H(\mu^N|\nu_{\rho_0(\cdot)}) = o(N^d)$.

This weak statement of local equilibrium follows from the claim that the evolved state satisfies

$$H(S_t^N\mu^N|\nu_{\rho(t,\cdot)}) = o(N^d),$$

which comes from the variational description of relative entropy. And this is proved via regularity of the PDE solution (since we want to do a Taylor expansion), as well as some estimates which are probability facts about probability measures (large deviations results).The heuristic explanation for the N^2 time evolution is again a Taylor expansion where the linear term cancels out because p is of zero drift. (And actually once we have this, we can go back to proving local equilibrium.)

19 June 2, 2026

Rupert Li

We'll discuss the box-ball system, following the paper "Integrable structure of box-ball systems: crystal, Bethe ansatz, ultradiscretization and tropical geometry" by Inoue, Kuniba, and Takagi. We'll introduce the model, then introduce crystals and the KKR bijection (tools for study in this setting), and discuss a little bit of tropical geometry.

Example 132

In a box-ball system, we have a bunch of boxes arranged in a line (say on $\mathbb{Z}$) and balls in some of those boxes. A person with a backpack runs from left to right, picking up balls when they see one and depositing them in empty sites whenever they have one available.

For example, if we start off with balls in boxes 1, 2, 3, 8, then after one run through the balls will be in positions 4, 5, 6, 9. The next time, they'll be in spots 7, 8, 10, 11, and then after that they'll be in spots 9, 12, 13, 14. In particular, notice that the chunk of 3 started off moving at "speed 3" and the chunk of 1 moved at "speed 1," and other than the interaction when they collide (up to some phase shift) these groups essentially pass through each other.

Fact 133

This backpack operation is invertible. And as discussed above, a size-N cluster, called a **soliton**, moves at speed N, and when solitons collide they swap positions with some phase shift.

We can also define this system on $\mathbb{Z}/n\mathbb{Z}$ instead of $\mathbb{Z}$; the only ambiguity would be where the runner starts and it turns out that doesn't matter (the operation is still invertible and doesn't depend on the initial position).

Remark 134. *This box-ball system is obtained by crystallization (a $q \to 0$ limit of a different model) of* **quantum** *integrable systems like the six-vertex model. Those models were defined via spin Hall–Littlewood polynomials; sending $q \to 0$ recovers Schur functions so we might hope there is some Schur or DPP structure here. Unfortunately, the way we take the limit here is different from how Schur processes are specialized. On the other hand, we do still indeed have the Yang–Baxter equations so we get some nice properties anyway.*

The box-ball system can also be obtained by ultradiscretization of **classical** *integrable systems, and this perspective gives some other tools (outside the scope of this course). We'll describe this more at the end.*

Definition 135

Let I be an index set. A **crystal** is a set B equipped with **Kashiwara operators** $\tilde{e}_i, \tilde{f}_i : B \sqcup \{0\} \to B \sqcup \{0\}$, such that for all $b \in B$ and $i \in I$, there is some n with $\tilde{e}_i^n b = \tilde{f}_i^n b = 0$, and also satisfying

$$\tilde{e}_i 0 = \tilde{f}_i 0 = 0, \quad \tilde{f}_i b_1 = b_2 \iff \tilde{e}_i b_2 = b_1$$

(this is basically saying $\tilde{e}_i, \tilde{F}_i$ are inverses up to some many-to-one issues with zero, since $0 \notin B$). We define

$$\varepsilon_i(b) = \max\{m : \tilde{e}_i^m(b) \neq 0\}, \quad \phi_i(b) = \max\{m : \tilde{f}_i^m(b) \neq 0\}.$$

Example 136

Take $I = \{0, \cdots, n\}$ and let $B_\ell = \{x \in \mathbb{Z}_{\geq 0}^{n+1} : \sum_i x_i = \ell\}$. Think of elements of this crystal as contents of semistandard Young tableaux of shape $\lambda = (\ell)$; for example the element $(1, 3, 4)$ corresponds to the Young tableau $(1, 2, 2, 2, 3, 3, 3, 3)$ (and here $\ell = 8$ and $n = 2$). We then define

$$\tilde{e}_i(x) = (\cdots, x_i + 1, x_{i+1} - 1, \cdots),$$

$$\tilde{f}_i(x) = (\cdots, x_i - 1, x_{i+1} + 1, \cdots),$$

or zero otherwise if these operators do not keep us within B_ℓ.

We won't get into the technical details of this, but this allows us to generalize the box-ball system into a multi-species system where 1 would be like "no ball", 2 is the "lowest ball," and so on. Then ℓ will be the size of a single box (so the capacity is more than just a single ball).

Definition 137

Let B, B' be two crystals. Then the tensor product $B \otimes B'$ is another crystal with set $B \times B'$, where we define $x \otimes 0 = 0 \otimes y = 0$, and where the Kashiwara operators are

$$\tilde{e}_i(x \otimes y) = \begin{cases} \tilde{e}_i x \otimes y & \text{if } \phi_i(x) \geq \varepsilon_i(y) \\ x \otimes \tilde{e}_i y & \text{otherwise,} \end{cases}$$

$$\tilde{f}_i(x \otimes y) = \begin{cases} \tilde{f}_i x \otimes y & \text{if } \phi_i(x) > \varepsilon_i(y) \\ x \otimes \tilde{f}_i y & \text{otherwise.} \end{cases}$$

It can be checked that this is indeed a crystal and that the operation is associative. (There's a q-deformation of all of this, which is where crystals come from in the first place via quantum groups; these are actually the q-deformations of $e, f \in \mathfrak{sl}_2$ with $q \to 0$.)

Fact 138

In general, $B \otimes B'$ is in bijection with $B' \otimes B$ in the category theory sense, meaning there is some map $R_{BB'}$ between the two which commutes with $\tilde{e}_i, \tilde{f}_i$ which is uniquely defined by this via a universal property. We call this the **combinatorial R**.

Fact 139 (Nakayoshiki–Yamada '97)

The operator R can be defined via the following algorithm which we demonstrate by example. Take $R_{B_5 B_3}$ for $n = 6$ and suppose we want to compute R applied to $\begin{array}{|c|c|c|c|c|}\hline 1&3&3&4&7\\\hline\end{array} \otimes \begin{array}{|c|c|c|}\hline 1&3&5\\\hline\end{array}$. We have seven cells, so we can place particles in the corresponding locations $\begin{array}{|c|c|c|c|c|c|c|}\hline x& &xx&x& & &x\\\hline\end{array}$ and $\begin{array}{|c|c|c|c|c|c|c|}\hline x& &x& &x& & \\\hline\end{array}$.

In this case, we have more particles on the left than the right. So on the right argument, we look cyclically forward and pair particles with unpaired particles on the left. (So the particle in box 1 on the right pairs with the particle in box 7 on the left because of wraparound; box 3 on the right then pairs with box 1 on the left, and box 5 on the right pairs with box 4 on the left.) Then swap the locations of all of the unpaired particles across from left to right; we end up with $\begin{array}{|c|c|c|}\hline 1&4&7\\\hline\end{array} \otimes \begin{array}{|c|c|c|c|c|}\hline 1&3&3&3&5\\\hline\end{array}$. And for $\ell \leq \ell'$ we can do the analogous thing in the other direction.

Definition 140

The **energy function** $H(x \otimes y)$ is the number of winding pairs in this algorithm (in the case above it was 1).

Proposition 141 (Yang–Baxter equations)

As operators $B \otimes B' \otimes B'' \to B'' \otimes B' \otimes B$, we have

$$(R \otimes 1)(1 \otimes R)(R \otimes 1) = (1 \otimes R)(R \otimes 1)(1 \otimes R),$$

where R acts on two of the coordinates at a time.

Apparently these can be viewed as "fused vertex model states," again by taking the $q \to 0$ limit.

To see connections to the box-ball model When we take the collisions of two solitons, the passing operation can be given by R; for example for $\ell = n = 1$ it just says solitons pass through each other, but more generally it describes what happens to the makeup of the blocks.

Definition 142

The elements of $B_1^{\otimes L}$ are called **paths**. The **highest paths** satisfy the majorization inequality

$$\#\{\text{1s in first } k \text{ boxes}\} \geq \#\{\text{2s in first } k \text{ boxes}\} \geq \cdots,$$

and the **weight** of x is some partition of L of length $n+1$. The set of highest paths of weight λ will be denoted $P_+(L, \lambda)$.

Definition 143

The **rigged configurations** are the following tuples $(\mu^{(0)}, \cdots, \mu^{(n)})$ of partitions. We have $\mu^{(0)} = 1^L$. Define $m_j^{(a)}$ to be the number of length-j rows of $\mu^{(a)}$, and let $q_j^{(a)} = \sum_{i=1}^{j} (\mu^{(a)})_i'$ (where $'$ denotes the transpose) and $q_j^{(n+1)} = 0$. Define the **vacancy** $p_j^{(a)} = \Delta_a q_j^{(a)}$ for Δ_a the "second derivative;" we say that we have a **configuration** if $p_j^{(a)} \geq 0$ for all $a \in [n]$ and $j \geq 1$ with $m_j^{(a)} > 0$. In other words, we consider a partition to be a contiguous block of rows of the same length.

To such a configuration, we can define a **rigging**, which is a set of numbers $J_{j,\alpha}^{(a)} \in \mathbb{N}$ for all $\alpha \in [m_j^{(a)}]$. We index a block from bottom to top, and a rigged configuration is $(\mu, J)_L$, where for all (a, j) we have $0 \leq J_{j,1}^{(a)} \leq \cdots \leq J_{j,m_j^{(a)}}^{(a)} \leq p_j^{(a)}$. The **weight** is then the partition of x of L, where $\lambda_i = |\mu^{(i-1)}| - |\mu^{(i)}|$. The set of **rigged configurations** for a given L and a given λ is then written $RC(L, \lambda)$.

Theorem 144 (KKR bijection)

We have a bijection $RC(L, \lambda) \to P_+(L, \lambda)$, which in our specific box-ball setting reduces to just the RSK correspondence.

The reason we care about this is that it's kind of hard to think about how the box-ball system works directly, but we can linearize the dynamics using KKR:

Theorem 145

Take L sufficiently large so that the boundaries are irrelevant. Then for some state $p \in P_+(L, \lambda) \subseteq B_1^{\otimes l}$ (recall these paths are "box-ball like"), let $p' = T_\ell(p)$ be the configuration after one step of our model. Then in the bijection into the rigged configurations, $p \mapsto p'$ becomes

$$(\mu, J) \mapsto (\mu, J'), \quad J' = (J'^{(1)}, J^{(2)}, \cdots, J^{(n)}), \quad {J'_{j,\alpha}}^{(1)} = J_{j,\alpha}^{(1)} + \min(\ell, j)$$

(only the first coordinate of J ever changes).

So after n steps we just add some amount to ${J'_{j,\alpha}}^{(1)}$ and then reverse the bijection. The other constant things are called the "action variables" and the first coordinate of J is the "angle variable;" we can then read off information from the partition μ as well.

If we stare at the box-ball system, we can also define things via min and $+$, and in fact this relates to the exponentiation tropical geometry remark from earlier in the course. In the detropicalized setting, all of these things

work with geometric crystals instead, where R becomes a birational object instead of a combinatorial one (but still satisfies Yang–Baxter). The only thing that doesn't really carry over directly is the notion of rigged configurations; instead we get eigenvalues of a certain matrix which dictates the scattering theory.

Michael Ren

The topic for this talk will be **zeros of the iid Gaussian power series**, following a paper of Peres and Virág from 2005.

Example 146

Let $a_0, a_1, a_2, \cdots$ be iid standard complex Gaussians (so $X + iY$ with $X, Y \sim N(0, \frac{1}{2})$), and consider the power series

$$f(z) = \sum_{k \geq 0} a_k z^k.$$

We are interested in the zeros of this random power series f.

Proposition 147

f is almost surely holomorphic on the unit disk $\mathbb{D} = \{z : |z| < 1\}$, since by Borel–Cantelli all of the coefficients are stochastically all of order 1. The zero set Z_f is then a simple point process in $\mathbb{D}$ (since if we have a double root, the derivative must also vanish, but f' doesn't depend on a_0).

Theorem 148

The set Z_f is a determinantal point process with kernel the **Bergman kernel**

$$K_{\mathbb{D}}(z, w) = \frac{1}{\pi(1 - z\overline{w})^2}.$$

Note that f is a complex Gaussian process on the unit disk with covariance

$$\begin{aligned} C(z, w) &= \mathbb{E}[f(z)\overline{f(w)}] \\ &= \sum_{k \geq 0} (z\overline{w})^k \\ &= \frac{1}{1 - z\overline{w}} \\ &= 2\pi S_{\mathbb{D}}(z, w) \end{aligned}$$

(the process is determined by the covariance just like with usual Gaussian processes, but with a different inner product). $S_{\mathbb{D}}$ is called the Szegö kernel; we'll skip what this means in general but we can replace $\mathbb{D}$ with a general complex manifold and parts of this whole story port over.

The nice thing about this series is that not only do we have this DPP structure, but the whole function is conformally covariant in a certain sense and so is the zero set.

Proposition 149

If $g : \mathbb{D} \to \mathbb{D}$ is a conformal automorphism of the disk (bijective holomorphism with a holomorphic inverse), then we have

$$(g')^{-1/2}(f \circ g) \overset{d}{=} f$$

up to some factor which we'll write down soon. As a consequence, the zero set Z_f is conformally invariant since $g^{-1}(Z_f) \overset{d}{=} Z_f$ (the additional factors don't contribute any zeros).

Proof sketch. This can be directly checked because conformal maps are all of the form $e^{i\theta}\frac{z-\alpha}{1-\overline{\alpha}z}$ for some $\alpha \in \mathbb{D}$ and $\theta \in \mathbb{R}/2\pi\mathbb{Z}$. For notational convenience we'll ignore the θ factor and define $g(z) = T_\alpha(z) = \frac{z-\alpha}{1-\overline{\alpha}z}$. Then

$$g'(z) = \frac{1-|\alpha|^2}{(1-\overline{\alpha}z)^2},$$

so it makes sense to define the square root $\tau_\alpha(z) = \frac{1-\overline{\alpha}z}{\sqrt{1-|\alpha|^2}}$. Thus all we need to do to check that our complex Gaussian processes are equal in distribution is to compute the covariance

$$\begin{aligned} C_g(z,w) = C(g(z), g(w)) &= \frac{1}{1-g(z)\overline{g(w)}} \\ &= \frac{(1-\overline{\alpha}z)(1-\alpha\overline{w})}{1-|\alpha|^2}\frac{1}{1-z\overline{w}} \\ &= \tau_\alpha(z)\overline{\tau_\alpha(w)}C(z,w). \end{aligned}$$

by direct computation, so it's a multiple of the original kernel via some deterministic function of z and w, which we can absorb into the prefactor. □

So our zero set is conformally invariant, meaning that the DPP should also be conformally invariant. In particular, the density should generally follow the hyperbolic metric where we see more zeros near the boundary of the unit disk than in the center. Before starting the proof, we'll mention a few things we can deduce about Z_f if we assume the form of the kernel $K_\mathbb{D}$:

- $K_\mathbb{D}$ is Hermitian, so zeros repel each other (similar to what we've seen with the Schur processes with 2×2 determinants).
- The intensity of the measure with respect to Lebesgue measure on the disk is

  $$K_\mathbb{D}(z,z)dz = \frac{1}{\pi(1-|z|^2)^2}dz,$$

  which is $\frac{1}{4\pi}$ times the hyperbolic measure on $\mathbb{D}$ as expected.
- As we did before, we can analyze via a Fredholm determinant calculation the hole probability for Z_f. For any ball $B \subset \mathbb{D}$, we then get

  $$\mathbb{P}(Z_f \cap B = \varnothing) = \exp\left(\frac{\pi}{24}|B| + o(|B|)\right)$$

  for $|B|$ the hyperbolic area, via large deviations estimates.
- The sets of magnitudes of the zeros is equal in law to $\{U_k^{1/(2k)} : k \geq 1\}$ where U_i are iid uniform on $[0,1]$.

Proof sketch of main theorem. The idea will be to apply Hammersley's formula, or more generally the Kac–Rice formula, which is a way of writing down the zero intensity for random smooth functions and get a k-point correlation.

This will tell us (all matrices are $k \times k$)

$$p_k(z_1, \cdots, z_k) = \frac{\mathbb{E}\left[|f'(z_1)^2 \cdots f'(z_k)^2| \, | f(z_1) = \cdots = f(z_k) = 0\right]}{\det[\pi C(Z_i, Z_j)]}.$$

Usually this conditioning is difficult to compute, but for Gaussians we just have orthogonal projection in L^2, so via the conditioning lemma we actually have something nice. Specifically, the conditional law $f | f(z_1) = \cdots = f(z_k) = 0$ is equal in distribution to $T_{z_1}(z) \cdots T_{z_k}(z) f(z)$ for T_α the maps from before which "put a zero at α." (So that's nice because it lets us just compute the numerator nicely.) After that, we use lots of determinant identities, which we'll skip the details of here, such as the Borhchardt formula

$$\operatorname{perm}\left[\frac{1}{x_i + y_j}\right] \det\left[\frac{1}{x_i + y_j}\right] = \det\left[\frac{1}{(x_i + y_j)^2}\right]$$

which we get by expanding out the determinants.

Talking about those earlier steps in the proof now, suppose we have a random smooth function $F : \mathbb{R}^n \to \mathbb{R}^n$ so that generically the zero set is a point process. The intensity of the roots is then specified by asking, for each $A \subseteq \mathbb{R}^n$, the expected number of roots in A. The $n = 1$ case is easiest to explain: if we want to understand how many roots lie in an interval, we need to understand "$\sum_{x \in A} 1\{F(x) = 0\}$." Under some regularity conditions on F, we can do a smoothed out version of this and instead consider

$$\int_A 1\{|F(x)| < \delta\} \cdot \frac{F'(x)}{2\delta} dx$$

(since at slope $|F'(x)|$, $\frac{2\delta}{|F'(x)|}$ is the length of the interval around a zero at which f is within δ). If we take expectations of both sides, we then get

$$\begin{aligned}\mathbb{E}\left[\#\text{ roots of } A\right] &\approx \int_A \mathbb{E}\left[\frac{1\{|F(x) < \delta\}}{2\delta} |F'(x)|\right] dx \\ &\overset{\delta \to 0}{\to} \int_A p_{F(x)}(0) \mathbb{E}[|F'(x)| \, | F(x) = 0] dx.\end{aligned}$$

So we can see the form of the "derivatives conditioned on the zero set" appearing in Kac–Rice, and because we have a Gaussian process all densities are Gaussian and the "probability density for F term" $p_{F(x)}(0)$ comes in the denominator. So for $n > 1$, the expected number of roots in A will instead be

$$\int_A \mathbb{E}\left[|\det DF(\vec{x})| \, | F(\vec{x}) = \vec{0}\right] p_{F(\vec{x})}(\vec{0}) d\vec{x}.$$

And now the square comes from having complex Gaussians: if we apply this to the map $F : \mathbb{R}^{2k} \to \mathbb{R}^{2k}$ (also $\mathbb{C}^k \to \mathbb{C}^k$) with $F(z_1, \cdots, z_k) = (f(z_1), \cdots, f(z_k))$ we get exactly what is claimed up to some factorial factors. And the derivative of this map is essentially coordinate-by-coordinate; we have $DF(z_1, \cdots, z_k)$ block diagonal with k 2×2 blocks $f'(z_i)$ which are scale/rotation matrices, with determinant exactly being the product of those 2×2 determinants which are $|f'(z_i)|^2$.

The only thing left to explain in the formula is the probability density term, but $F(\vec{z})$ is a complex Gaussian process with covariance $C(z_i, z_j)$. So the density at zero will have a π coming from the density of the standard complex Gaussian, together with this covariance term. $\square$