18.677: Topics in Stochastic Processes

Lecturer: Professor Alexei Borodin

Notes by: Andrew Lin

Spring 2021

Introduction

This semester's class will cover **topics in integrable probability**. Remote teaching makes it harder to transfer information in online instruction than in a normal semester, and this becomes even more true with high-level material (because it's difficult to listen to a technical proof online). This class will cover a lot of material that is in papers or folklore but not books, so it will be difficult to pick things up outside of lecture (since reading the original papers take a lot of time). Overall, this means that the outcome will be to **not teach too many difficult proofs** at the beginning, instead making this more of a seminar class or conversation.

We'll start with a few lectures surveying various objects that Professor Borodin likes to think about – these are often referred to as **solvable** or **integrable** probabilistic models. Some of us may be familiar with some of the models, but we'll focus here on what kinds of questions are usually asked or answered about the objects. From there, each of us will choose one of the objects and read more in-depth about it (beyond the pictures and impressions that are initially presented): we'll each then present what we can learn during class. (This way, each of us can learn something by reading and talking about an object, rather than trying to listen and potentially getting distracted.)

This class will be joint with a seminar that was run last semester on integrable probability: the talks will happen more rarely in that we'll have outside speakers in these lectures every once in a while. These talks will be instructional (the speakers will make sure not to go too fast), and the first will occur next week.

The grading requirements for this class are basically to choose a topic, read about it, and present it (if we think about this as a seminar). The hope is that this class is not too imposing but will still be entertaining!

Remark 1. *This class will not have problem sets, because the material does not have a well-defined textbook outline. Instead, we'll have a much more "chatty" lecture experience than usual, and this will be (again) more of a seminar class.*

1   February 16, 2021
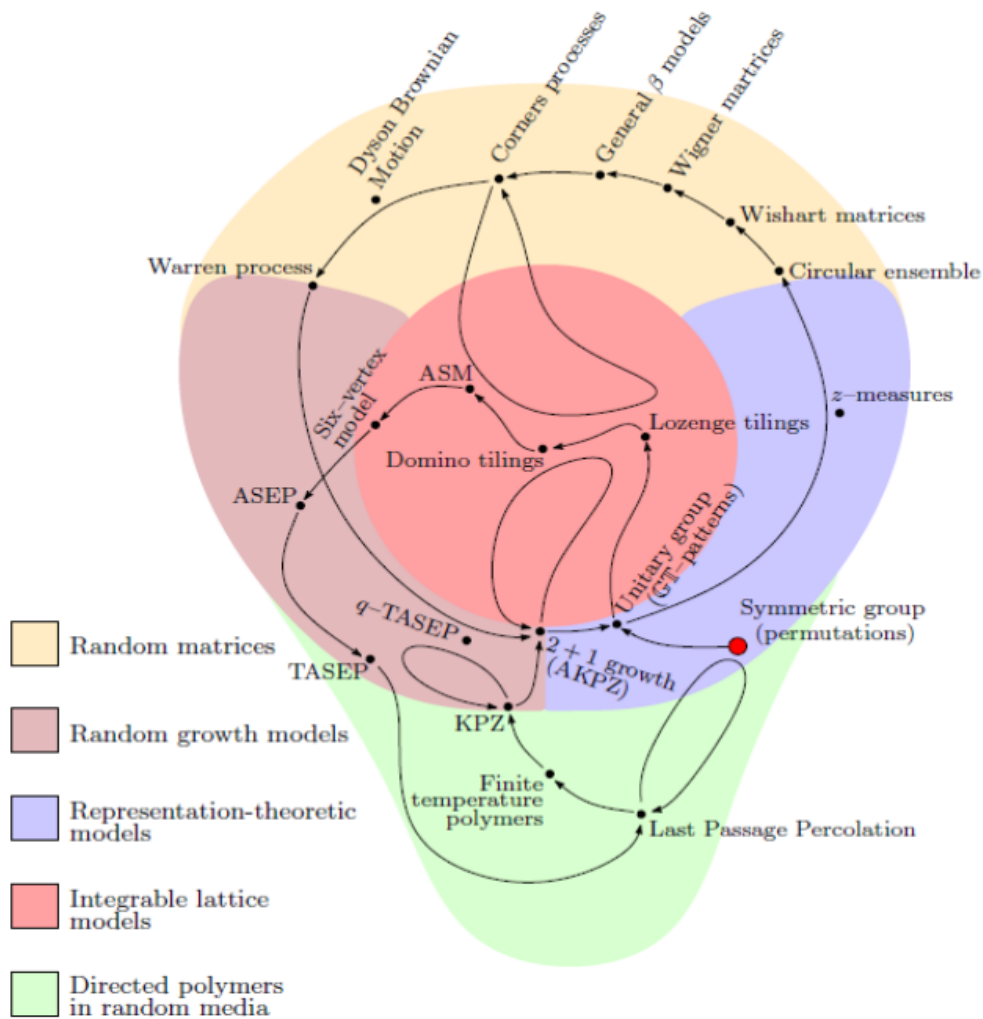
The picture above is a diagram of many objects we'll talk about in the next few classes. The symmetric group will be our starting point, which is why that point is thick and red. From there, we'll basically follow the winding path along the arrows (and we'll switch between different domains of mathematics and mathematical physics, as we can see in the colored legend).

Basically, the point is that there is no natural way to separate the systems into different fields, and that's why the subfield is called "integrable probability:" there are various types of analysis that will come up in our study.

We'll clarify some of the acronyms that appear here just so we have English words (and we can search them up if we'd like, but hopefully they'll make more sense in the next few weeks):

- **ASM**: Alternating Sign Matrices,

- **(T)ASEP**: (Totally) Asymmetric Simple Exclusion Process,

- **(A)KPZ**: (Anisotropic) Kardar-Parisi-Zhang (three names – Professor Kardar is in the physics department at MIT). The KPZ acronym is also used in another context in probability to refer to three Russian physicists.

This picture is basically a map or landscape of how we'll walk in this class, and we'll return to it every time we switch topics. So let's get started now:

We often write these permutations in different ways: one is to indicate where each of the integers $1, 2, \cdots, n$ goes, but a more informative encoding is to use **cyclic structure**.

Right now, we can focus on the sizes of these cycles: this permutation has cycles of sizes $(3, 2, 2, 1)$, and it's useful to think of this collection of cycle sizes as a "partition" of the total number of elements, 8:

One pictoral way to represent partitions is to use **Young diagrams** (also known as **Ferrers shapes**). For example, the partition $(3, 2, 2, 1)$ is represented as shown below (in the **English** and **French** notation, respectively):
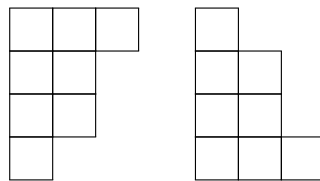
(We usually call the horizontal components **arms** and the vertical ones **legs**.) Such diagrams can also be represented by rotating the (French) diagram by 45 degrees to get the Russian notation:
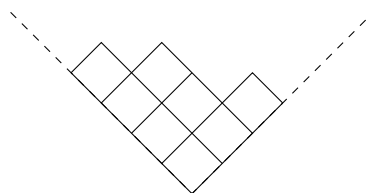
The question of "which objects probabilists find interesting" is broad, but there are basically two classes of probability measures that are worth mentioning:

- Take a bunch of simple objects (such as n copies of $\{0, 1\}$), and we introduce some simple probability measure for each set. Then we can construct the product measure and then do a certain operation on that product set (projection, mixing, and so on). For example, the sum of all such digits gives us a more interesting object than the individual parts, and we're **reducing the dimension** of the object in this way. This is a common thing to do in statistical physics, where the simple sets are the state spaces for atoms or molecules, and some average characteristic is the interesting measurement.

- Start with an object like $S(n)$, and select one of the choices at random. (So we could pick a permutation $P(\sigma)$ uniformly at random with probability $\frac{1}{n!}$.) Doing this kind of choice is often difficult in practice, but it's often nice to say things about them (with high probability). A question like "what happens to a permutation with high probability if chosen uniformly at random" can be difficult to answer, and the question of cycle structure is a good one to start with: **what is the typical cycle structure for a random permutation** $\sigma \in S(n)$, if we choose from the uniform measure?

Answering this question is basically an exercise in algebra, so we won't spend too much time on it. Letting $\mathbb{Y}_n$ be notation for the set of all partitions of n (here the "Y" comes from "Young diagrams"), we have

$$\boxed{\mathbb{P}(\text{cycle structure } \mu) = \prod_{i=1}^{\ell(\mu)} \frac{1}{\mu_i} \prod_{j \geq 1} \frac{1}{m_j!}}$$

for any $\mu \in \mathbb{Y}_n$, where

$$m_j = \#\{k : \mu_k = j\}$$

are the multiplicities appearing in the parts of the partition. (So another notation to represent partitions is

$$\mu = 1^{m_1} 2^{m_2} \cdots,$$

and for example $(3, 2, 2, 1)$ would be the partition $1^1 2^2 3^1$.) So now the boxed expression above is a probability measure on $\mathbb{Y}_n$, and we've discarded a lot of information from $S(n)$ already, getting a complicated-looking set of weights. And it's difficult to prove directly that this sum of weights is 1 without referring to the symmetric group, so the point is that we've created an interesting object already.

Here are the cycle structures of two permutations of 100 that were chosen uniformly at random: the partitions are $(65, 28, 3, 2, 2)$ and $(28, 25, 18, 17, 4, 4, 2, 1, 1)$.
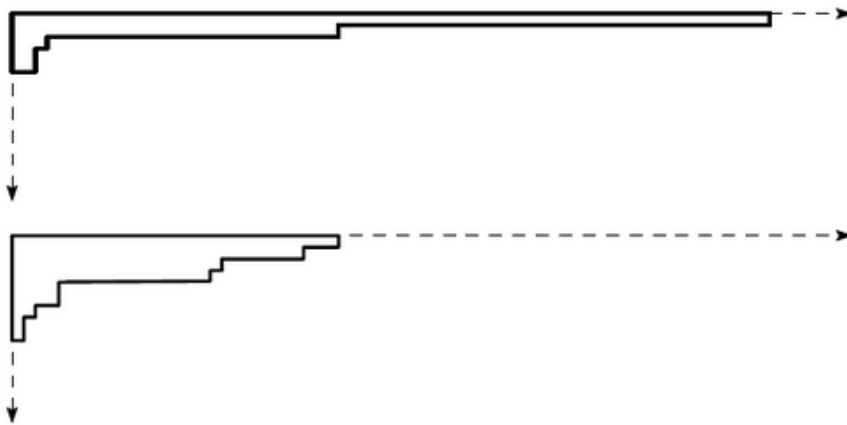
There are a few conjectures that we can make about "typical" properties of permutations. Here's a few suggestions:

- With high probability, $\lambda_1 \geq \ell(\lambda)$ (that is, the largest component of the partition is bigger than the number of parts).

- There is some constant c so that $\frac{\lambda_1}{\lambda_2} \approx c$ with high probability. (This one is perhaps harder to believe, because $\frac{65}{28}$ and $\frac{28}{25}$ are far apart from each other.)

- The quantity $\lim_{m \to \infty} \frac{\lambda_m}{\lambda_{m+1}}$ has some limit for "large enough n." (This one does turn out to be true if stated more formally, and the ratio is e. Understanding where this e comes from the symmetric group is a topic that we can talk about in our presentations!)

4

There are other ways to produce measures on partitions starting from uniformly random permutations, too. For example, consider the permutation of $S(8)$ sending $(1,2,3,4,5,6,7,8) \to (5,7,2,3,1,6,8,4)$ as an example: we'll project $S(8)$ onto $\mathbb{Y}_8$ by making λ_1 the length of the **longest increasing subsequence**. For example, $(2,3,6,8)$ is a subsequence of length 4 (so the terms don't need to be consecutive).

> **Fact 5**
>
> There's an interesting algorithmic way to find the longest increasing or decreasing subsequence, and this is sort of related to the game of "trying to sort a deck of cards as fast as possible."

Let's think of $(5,7,2,3,1,6,8,4)$ as the "cards in our deck," with 5 on the top and so on. We put 5 in its own pile, and then because 7 is greater than 5, we add 7 to its pile as well. After that, 2 is less than 7, so we start a new pile for it. From there, we **place cards as far to the left as possible**, as long as the cards are increasing from top to bottom. This gives us the following end result (where the columns can be thought of as piles):

5	2	1
7	3	4
8	6	

And now we can collect the cards in an ordered manner: if we keep grabbing the highest card that appears on top, we can get $8,7,6,5,4,3,2,1$ in that order, and thus we've properly sorted the cards. (It turns out this is a really good algorithm to do in practice, and Professor Borodin has beaten many people with it – it's called **patience sorting**, and it was a form of entertainment from times when people didn't have computers.)

The number of piles that we get here is actually equal to the longest **decreasing** subsequence, $(5,2,1)$, so if we wanted to get the longest **increasing** subsequence, we should make sure the cards decrease instead of increase. That gives us the following shape:

5	7	6	8
2	3	4	
1			

The longest increasing subsequence problem is actually related to the **airplane boarding** problem, which we can state as follows:

> **Example 6**
>
> Suppose we have an airplane with 8 seats all in a row, and there are a line of 8 people boarding. Each of them has a ticket for a particular seat, and the order in which the people line up is a uniformly random permutation.

People board the airplane with large suitcases, so it takes time to fit the carry-on suitcase above their seat: suppose that it takes 1 minute for each person to place the suitcase. If we assume that the passengers are trying to sit in seats $(5,7,2,3,1,6,8,4)$, in order of entrance, then the person at seat 7 needs to wait for a minute for the person at seat 5 to place their suitcase, but the person at seat 2 does not need to wait. So the relevant question is **how long the total boarding time takes** here, and it turns out that this answer is exactly the length of either the longest increasing or decreasing subsequence (exercise).

The study of longest increasing subsequences was actually advertised to airlines by a friend of Professor Borodin, but airlines weren't too impressed by the math... (In fact, boarding quickly is not super important in real life, because the split of the boarding queue into zones increases boarding time.) But there are other applied problems in this direction as well (like the scheduling of requests for hard drives) that may be interesting to study.

Anyway, returning to the probability: if we let λ_1 be the length of the longest increasing subsequence, we can then define $\lambda_1 + \lambda_2$ to be the largest sum of the lengths of two disjoint increasing subsequences. (For example, we could try picking $(2, 3, 6, 8)$ and also $(5, 7)$, and our goal is to maximize the sum of those two lengths.) Ultimately, this gives us a partition $\lambda = (\lambda_1, \lambda_2, \cdots) \in \mathbb{Y}_n$ in a different way than by looking at cycle structure. But we'll think a bit more about this next time!

2 February 18, 2021

Last time, we started studying the symmetric group, looking at partitions, cycles structure, and coming up with probability measures on partitions starting with permutations. At the end, we discussed how to use the longest increasing subsequence(s) of a permutation to construct the largest entries of a partition: for example, $\lambda_1 + \lambda_2$ would be the maximum sum of lengths for two disjoint increasing subsequences in our $\sigma \in S(n)$, and we define $\lambda_1 + \cdots + \lambda_k$ similarly. The resulting $(\lambda_1, \lambda_2, \cdots) \in \mathbb{Z}_{\geq 1}$ is then indeed a partition of n (though proving the statement that $\lambda_1 \geq \lambda_2 \geq \cdots$ is nontrivial).

Since we want to talk about probability, the next question is what we can say about partitions $\lambda(\sigma)$ when σ is chosen uniformly at random from $S(n)$. One such random sample is shown below, where the boxes of the partition have been scaled down by $\sqrt{n}$:
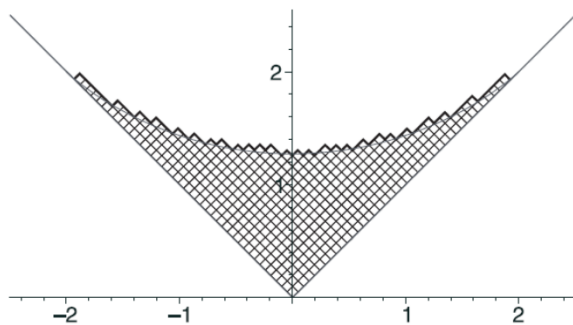
We'll notice that the partitions formed in this case are drastically different from the first method we discussed last week, but we'll understand why in a few weeks (the curve that looks to be traced out goes under the names of Logan-Shepp and Vershik-Kerov). For now, we should remember that we're pushing the uniform measure forward via our map $S(n) \to \mathbb{Y}_n$, and this distribution is typically called the **Plancherel** measure. Plancherel was a Swiss mathematician who studied Fourier analysis — his name is under the theorem for extending the Fourier transform to the L^2 spaces. So we'll spend a bit of time explaining why these two results are actually related!

Fact 7

The Prancherel probability of finding a given partition λ can be given by

$$\mathbb{P}_{\text{Plancherel}}(\lambda) = \frac{\dim^2 \lambda}{n!}.$$

(Here, $\dim \lambda$ is also often denoted as f^λ in combinatorics, and we'll define it below.)

The division by $n!$ should seem natural, because we started with the uniform measure $\frac{1}{n!}$ on permutations. But let's first define $\dim \lambda$ combinatorially:

For example, there is a single path from the empty Young diagram to $\square$, but there are two paths from the empty diagram to $\square\square$ (because there are two ways to add boxes), so $\dim \square = 1$ and $\dim \begin{smallmatrix}\square\square\\\square\end{smallmatrix} = 2$. And what we're claiming above is that at any given level $\mathbb{Y}_n$, we must have

$$\frac{1}{n!} \sum_{\lambda \in \mathbb{Y}_n} (\dim \lambda)^2 = 1.$$

This is much like the problems yesterday in that it is very difficult to prove from scratch. But we can also see another definition for $\dim \lambda$: such directed paths are in bijection with **standard Young tableaux**:

An example of a Young tableaux of shape $(3, 1, 1)$ is as shown below:

1	2	5
3		
4		

We can think of the number k being written down as its corresponding box being added at the kth step. So the dimension of a partition is the number of standard Young tableaux, but to understand why we use the word "dimension," we need to do a bit of representation theory.

A representation is then a way for groups to act on a space. For example, there is a natural way for $S(n)$ to act on $\mathbb{C}^n$: specifically, a permutation $\sigma \in S(n)$ permutes the coordinates of an n-dimensional vector via

$$T(\sigma)(x_1, \cdots, x_n) = (x_{\sigma^{-1}(1)}, x_{\sigma^{-1}(2)}, \cdots, x_{\sigma^{-1}(n)})$$

(the inverse here is to make sure we have a homomorphism rather than an antihomomorphism).

> **Definition 12**
>
> An **irreducible representation** is a representation that has no nontrivial invariant subspaces $U \subseteq V$ (meaning that $T(G)U \subseteq U$).

In other words, we're trying to find the smallest vector space on which G acts.

> **Example 13**
>
> The natural representation of $S(n)$ on $\mathbb{C}^n$ described above is **not** irreducible, because there is an invariant subspace
>
> $$U = \{\vec{x} : x_1 + \cdots + x_n = 0\}$$
>
> (and so is $U^\perp = \{(c, c, \cdots, c)\}$). But the representation is indeed irreducible beyond that.

It is a fact that every finite group has finitely many irreducible representations – the number of representations is the same as the number of conjugacy classes of G. And because conjugacy classes are exactly the cycle structure classes for $S(n)$, which is exactly the number of partitions of n, we can begin to see a connection here!

In particular, we can parameterize these irreducible representations using elements of $\mathbb{Y}_n$ **nicely** in the $S(n)$ case (even if we can't do this generally): for each $\lambda \in \mathbb{Y}_n$, we have the map

$$T^\lambda : G \to GL(V_\lambda), \quad \dim \lambda = \dim V_\lambda.$$

So the "dimension" of a partition is the dimension of its corresponding irreducible representation, and in fact basis elements of V_λ can be naturally parameterized by standard Young tableaux, with an explicit action of certain generators of $S(n)$. (If we want to learn more about this, we can search for **Young's orthogonal form**.)

But our goal is still to get back to the Plancherel name, and to understand that, we should note that there's a **version of the Fourier transform** for any finite locally compact group.

> **Definition 14**
>
> The **character** of a representation $T : G \to GL(V)$ is a function $\chi^T : G \to \mathbb{C}$ defined as
>
> $$\chi^T(g) = \text{tr}(T(g))$$
>
> for any $g \in G$.

Because trace is invariant under conjugation (using that $\text{tr}(AB) = \text{tr}(BA)$), characters are constant on the conjugacy classes of G. And in fact, the characters of irreducible representations form an **orthonormal basis** of the space of functions on G which are constant on conjugacy classes, if we use the scalar product

$$\langle f_1, f_2 \rangle = \frac{1}{|G|} \sum_{g \in G} f_1(g) f_2(g).$$

And whenever we have an orthonormal basis, we can break up an arbitrary function into components along those orthonormal basis elements. So we compute the scalar product with respect to a certain basis, and then we reassemble those coefficients together, but that's actually basically what the Fourier transform does (if e^{-ikx} are thought of as our "basis elements")! So now we can define the Fourier transform:

Here, the Plancherel measure $\frac{\dim^2 T}{G}$ can be thought as the equivalent of the rescaled Lebesgue measure $\frac{d\mu}{2\pi}$ in the Fourier transform. So we can think of the scalar product term above as the "Fourier transform," and the final composition is done by the Plancherel measure $\frac{\dim^2 T}{G}$. (This basically relates to **Burnside's theorem** in representation theory, which says that for any finite group G, $\sum_{T \in \mathrm{Irr}(G)} \dim^2 T = |G|$.)

Even though representations won't be central to what we want to talk about, they'll make appearances occasionally for different groups G. **And often behind algebraically analyzable probability measures, there is some backbone which will be frequently related to representation theory!**

Returning back to our combinatorial definition, we can say some more about the dimension of a partition $\dim \lambda$:

For example, the red box in the picture below has hook-length 4:
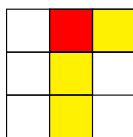
and we can compute that the number of Young tableaux of shape $(3, 3, 1)$ (and thus the dimension of $(3, 3, 1)$) is $\frac{7!}{2^3 \cdot 3 \cdot 4 \cdot 5}$. Furthermore, we can rewrite the above hook-length expression also as

$$= \frac{n!}{\prod_{i<j}(\lambda_i + n - i)!} \prod ((\lambda_i - i) - (\lambda_j - j)) = \frac{n!}{\prod_i \ell_i!} \prod_{i<j} (\ell_i - \ell_j),$$

where we define $\ell_i = n + \lambda_i - i$. We're being a bit lazy with the indices here, but the formula is flexible in that padding by zeros will not change the final answer. The point of rewriting the expression in this new way is that we have the **Vandermonde determinant**

$$\prod_{i<j}(\ell_i - \ell_j)$$

appearing in the expression, and we also have some other multiplicative factor. Both of these types of expressions will show up in our future calculations as well, so we can keep an eye out!

There are other interesting things we can say about the symmetric group, and we'll talk about some things that aren't typically considered when discussing the Plancherel measure. If we look again at $\mathbb{Y}_n$ and consider a **uniform** measure on it (instead of a uniform measure on $S(n)$, as we've been doing so far), we need $|\mathbb{Y}_n|$, the number of partitions of n (which is also often written as $p(n)$).

Remark 17. *There is no clean formula for this expression – people tried for many years, and there's a famous Ramanujan asymptotic formula that we do know. But there is a recent advance in which $p(n)$ was given in terms of certain algebraic numbers connected to modular forms.*

On the other hand, the generating function $\sum_{n\geq 0} p(n)t^n = 1 + t + 2t^2 + 3t^3 + 5t^4 + \cdots$ (we can check that these first few coefficients are correct) is quite easy to write down:

$$\sum_{n\geq 0} p(n)t^n = \prod_{n\geq 1} \frac{1}{1-t^n}.$$

There are indeed natural questions that lead to considering uniform measures on partitions, though: solving equations in finite groups and counting coverings.

Fact 18

Let $K_1, \cdots, K_r$ be (some of) the conjugacy classes (possibly with replacement) in a finite group G, and let $n(K_1, \cdots, K_r)$ denote the number of r-tuples $(g_1, \cdots, g_r) \in K_1 \times \cdots \times K_r$ such that $g_1 g_2 \cdots g_r = e$. (For example, we can imagine picking one permutation of each cycle type so that we end up with the identity.) Then

$$n(K_1, \cdots, K_r) = \frac{|K_1| \cdots |K_r|}{|G|} \sum_{\chi \text{ irred.} \in \hat{G}} \frac{\chi(g_1) \cdots \chi(g_r)}{(\dim \chi)^{r-2}},$$

where $g_i \in K_i$ for all i.

This is an example where characters can be useful even if we don't care about the representation theory itself – the initial problem doesn't necessarily talk about representations! And in particular, we can use this to compute **Hurwitz numbers** for Riemann surfaces:

Example 19

Let S^2 be the two-dimensional sphere, which is also the Riemann sphere, and suppose we have a map $S^2 \to S^2$ given by $f : z \mapsto z^d$. This function has branch points at 0 and ∞, and at those points the function has a unique preimage (but everywhere else there is a preimage of d points). This is an example of a **ramified covering** over two points, meaning that the covering space is basically d copies of $\mathbb{C}$ that are glued together at 0 and at ∞.

If we imagine a loop around 0 on the image of f, then we will go through a **long cycle** on the preimage, which switches between the different d sheets in some permutation: this is related to the **monodromy** around 0 and ∞ (which we can loosely define as the permutation that arises on the sheets of the covering).

But we can ask a more general problem: if we have d surfaces X and Y, we have a map $f : X \to Y$ of degree 2, and we puncture a few **ramification points** in Y, then the neighborhood of a regular point in Y will have a preimage of d neighborhoods, and the "long cycle" idea around ramification points on Y is related to the permutations that we travel through on X!

Fact 20

The number of coverings of Y by X, normalized by the number of automorphisms (for example, avoiding cyclic shifts), of a prescribed degree d and prescribed conjugacy classes of monodromies, is

$$\sum_{\lambda \in \mathbb{Y}_d} \left(\frac{\dim \lambda}{d!} \right)^{2-2g} \prod_i \frac{|K_i| \chi^\lambda(g_i)}{\dim \lambda},$$

where g is the genus of the target Y.

We should see the similarity between the two formulas here, but we're in a totally different domain of mathematics! (And for example, when $g = 1$ and we're covering the torus, the first term disappears and we basically have an **average over the uniform measure on partitions**.) We're now counting equivalence classes of covering maps with prescribed monodromies, which takes us into certain meaningful topological applications. So this is a bit far-out in that we're moving off the initial map, but next lecture we'll move into the next two stations on our roadmap.

3 February 23, 2021

As a reminder, the Thursday lecture will be a seminar by Jimmy He on **limit theorems for descents of Mallows permutations**.

In the first two lectures, we've thought about the symmetric group and related objects in a variety of ways, and we're now moving on to **representation theory of the unitary group** (mostly combinatorics and randomness, though the unitary group itself will come up later on). Recall that we often represent partitions combinatorially using a Young diagram, and we'll now talk about something related. We can think of a partition as dividing up a line segment, and our next step is to divide up a rectangle:

Definition 21

A **plane partition** is specified by a set of nonnegative integers written in a (possibly infinite) rectangular grid, weakly decreasing in both directions and with finitely many zeros.

Here is an example (which can be extended infinitely to the bottom right by adding 0s):

5	4	4	0
2	1	1	0
0	0	0	0

Similarly to how we described partitions using a two-dimensional Young diagram, we can think of plane partitions as a "height" map. A good starting point to work with is the picture for the **plane partition of all 0s**:
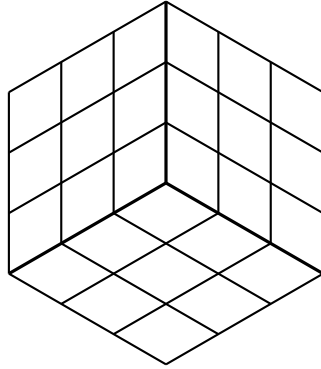
This picture can be either interpreted as the corner of a room, or as the "top vertex" of a cube, and both ways of seeing the picture are acceptable. And then if we put a box in the corner of the room (so our plane partition now has a 1 in the top left box and 0s everywhere else), we'll notice that the picture slightly changes:
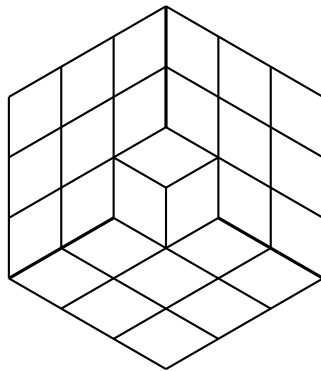
Below is a more complicated example which we'll use throughout the lecture – in particular, we can notice that each 3 corresponds to a column of boxes of height 3.
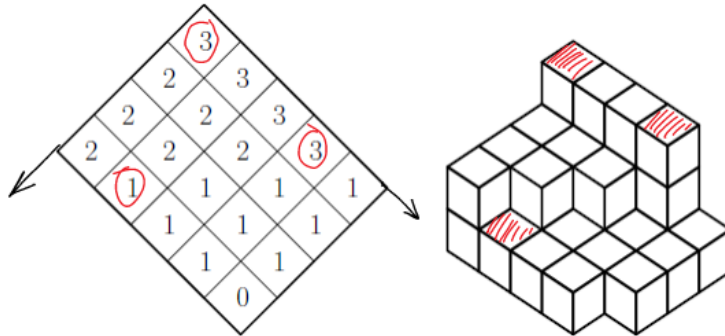
These staircase-like pictures can be interpreted in a few different ways, and we'll go through some of those here. First of all, even though the picture can be thought of as a three-dimensional structure, we can also think of it as a **tiling of part of the plane by rhombi** in three different orientations (horizontal faces and side faces in two directions). And furthermore, if we break up each rhombus into two equilateral triangles (drawing the short diagonals), another way to think about these tilings is to think about connecting these triangles. And then, instead of looking at the triangles as the grid, we can form the **dual grid** by looking at the hexagonal lattice formed by the centers of those triangles:
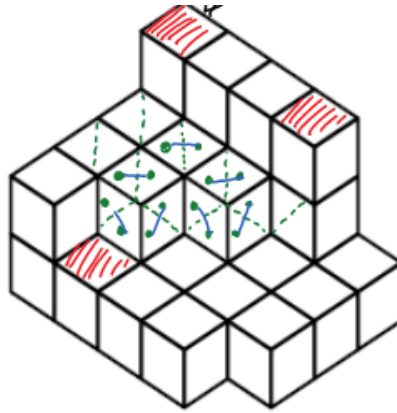
Forming rhombi then comes by gluing two of these centers together, and this forms what we call a **dimer**. When we have a plane partition, then, we can think about it as a **perfect matching** or **dimer covering** of the hexagonal lattice. To study these kinds of models, then, we often place the uniform measure on these dimer coverings and try to see what we can say about them. (And this has a statistical physics interpretation in terms of molecules, which is not super accurate, but the model still exhibits interesting features – in particular, we'll see **phase transitions** soon!)

In order to make the problem finite, we often place all of our rhombus tilings inside of an ambient (equiangular) hexagon, as shown below:
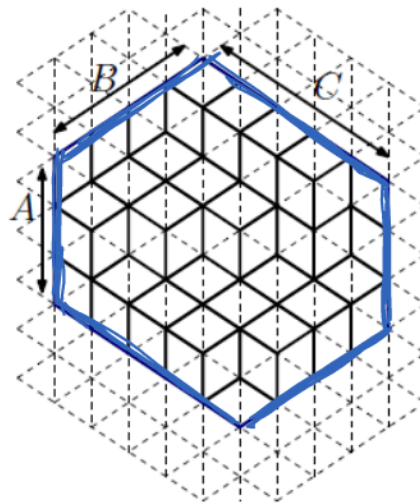
We can think of this as placing our three-dimensional picture in a box, where the side lengths A, B, C (labeled on the diagram above) limit the size of our plane partition. We can then erase some of the features of this tiling and keep other important ones – for example, in the picture below, we take all rhombi that do not have their long diagonal oriented horizontally, and we draw the horizontal midlines:
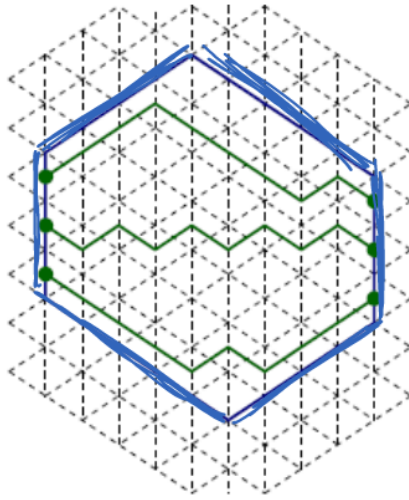
And we can see that these green paths determine the plane partition uniquely – one way to think about it is that the top green path traces out the boundary of the "height-3" boxes in the original plane partition, the middle one traces out the "height-2" boxes, and the bottom traces out the "height-1" boxes. But alternatively, each path is a random walk which either moves up or down (vertically) for each unit of horizontal "time." So that gives us an interesting correspondence:

> **Proposition 22**
>
> Boxed plane partitions are in 1-to-1 correspondence with nonintersecting Bernoulli paths which start and end next to each other.

Once we start talking about nonintersecting Bernoulli paths, things become interesting, because probabilists often expect a Brownian motion to come up nearby. And the corresponding probabilistic model if we replace these random walks with Brownian motions is the **Gaussian unitary ensemble** and **Dyson Brownian motion** on that ensemble, but we'll talk more about this later.

Below is a uniform sample from the boxed plane partitions in a $30 \times 30 \times 30$ grid (with coloring indicating either the type of rhombus or the "height"):
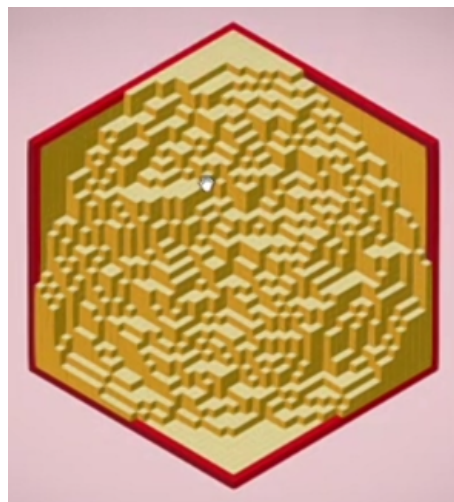
We'll notice that near all six of the corners of the hexagon, we have "flat regions", and there is a chaotic region in the middle. So that's the phase transition phenomenon: we have regions in the cube with very different behavior (the sides are often called **frozen**, and the middle is called **liquid**). And if we draw a $100 \times 100 \times 100$ hexagon (uniformly

sampling plane partitions), conditioned to be fixed inside a white rhombus, looks as shown:
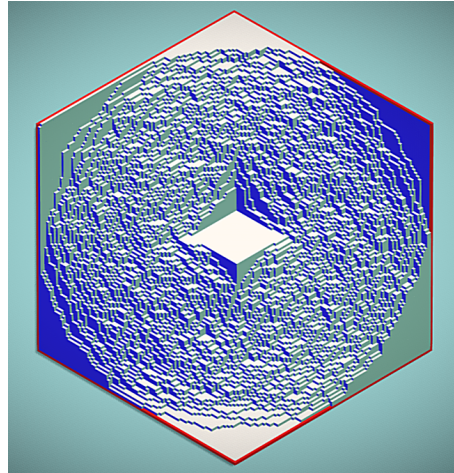
This is related to the "soap film" phenomenon that we may have seen in museums before: if we fix the boundary of the hexagon as a wire frame, and we have a table in the corner of the room, then placing a lot of dust in the room will give us a picture that looks as above, if dust arranges uniformly across all possibilities. And there are many questions that we can start asking about these types of systems, but it is often difficult to prove their answers! We'll learn how to predict the answers to those questions as this class goes on.

Recall that the Plancherel distribution on partitions can come from representation theory, and we'll try to understand the origin of the corresponding distribution on plane partitions now. We can cut our hexagon into two pieces as shown below:
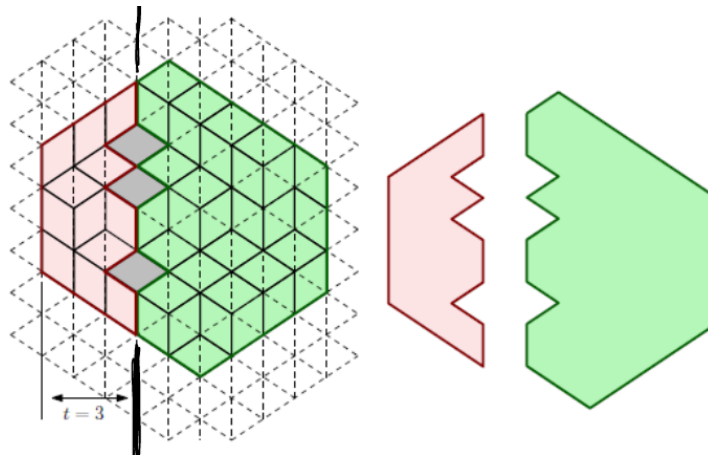
Basically, we pick some horizontal coordinate t, and we count where the gray rhombi (also called **lozenges**) appear. To describe this more explicitly, we need coordinates. Let ℓ_k^n be the position of the kth highest horizontal lozenge at the nth vertical section:
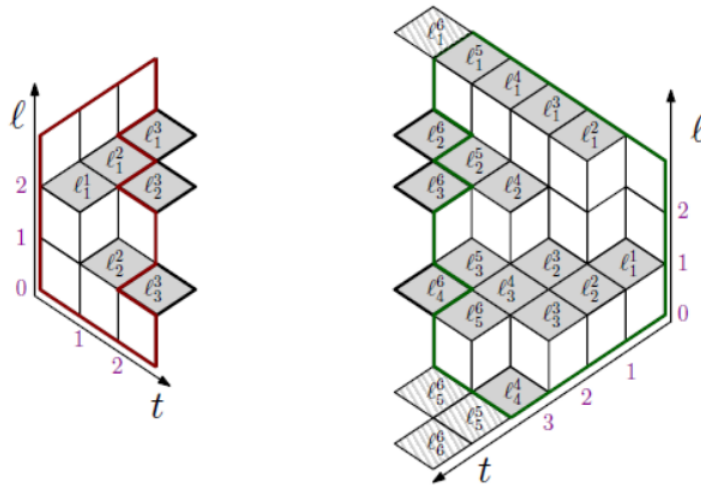
We'll notice that there are $1, 2, 3$ horizontal lozenges at the first, second, and third vertical sections, and we can prove (by induction on plane partitions and adding boxes, for example) that the number of horizontal lozenges in each vertical section is fixed. And just like plane partitions are described exactly by the green paths (which pass through the non-horizontal lozenges), we can specify some of the ℓ_k^n to know our plane partition exactly.

Proposition 23

The number of possibilities for the shape on the left, given a fixed $\ell_1^3, \ell_2^3, \ell_3^3$, is the number of integer solutions for the following inequalities:

$$\ell_3^3 \qquad\qquad \ell_2^3 \qquad\qquad \ell_1^3$$
$$\ell_2^2 \qquad\qquad \ell_1^2$$
$$\ell_1^1$$

It is often more convenient to solve these inequalities with a small shift of coordinates: if we define

$$\lambda_k^n = \ell_k^n + k - n,$$

then we get a similar triangular lattice but with **weak inequalities** everywhere.

Definition 24

A triangular array of interlacing integers as above is called a **Gelfand-Tsetlin pattern**.

Fact 25

The name of "Gelfand-Tsetlin pattern" comes from the fact that these can be placed in a bijection with basis vectors of an irreducible representation of $U(n)$ (where n is the number of rows), parameterized by the top row.

Gelfand and Tsetlin wanted to describe the action of the unitary group (or its corresponding Lie algebra) on irreducible representations, so they produced a basis and the explicit action on that basis using the generators of the Lie algebra. (And relatedly, standard Young tableaux were in bijection with basis vectors of irreducible representations

16

of $S(n)$ – this is a related story!) And the dimension of that irreducible representation of $U(n)$, which is the number of Gelfand-Tsetlin patterns of a fixed row, is

$$\text{Dim}_n(\ell_1^n, \cdots, \ell_n^n) = \prod_{1 \le i < j \le n} \frac{\ell_i - \ell_j}{j - i}.$$

(This is actually a special case of **Weyl's dimension formula**.) Again, we can recognize the Vandermonde determinant in the numerator – these will continue to pop up from time to time.

   We mentioned that we can use these Gelfand-Tsetlin patterns to count the number of shapes of the pink domain below (type I) and it also makes sense to ask about the green domain. That's why we have some extra lozenges at the boundary of the right shape: those lozenges force freezing in the part of the domain that we don't care about, and then we can get the same expression for how to enumerate tilings in the green domain.
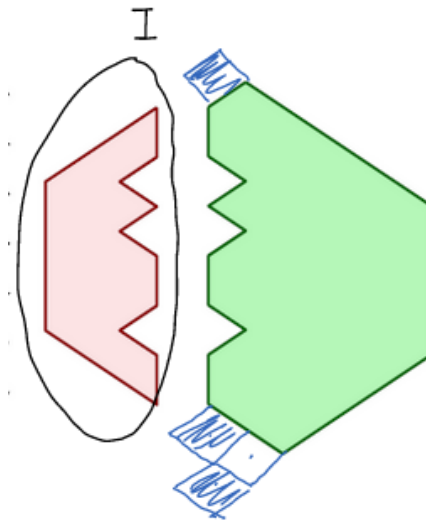
   So we'll again get a Vandermonde determinant, up to a constant, though the phantom lozenges do need to be accounted for. And what we find is that **under the uniform measure of rhombus tilings**, putting the pink and the green shapes together,

$$\mathbb{P}(\ell_1 > \ell_2 > \cdots > \ell_n) = c \prod_{1 \le i \le j \le n} (\ell_i - \ell_j)^2 \prod_{i=1}^{n} (\ell_i + 1)_{C-n} (A + n - \ell_i)_{B-n}$$

(where c contains all of the factorials, the factorials in the Weyl's dimension formula, and so on). Here, we're using the notation for the **Pochhammer symbol**

$$(a)_m = a(a+1)\cdots(a+m-1),$$

and these factors are coming from differences between the "phantom lozenges" that we added at the top and bottom of the green region. (Recall that A, B, C are the dimensions of our boxed plane partitions, and thus the $C - n$ comes from "bottom phantom lozenges," while $B - n$ comes from "top phantom lozenges.") And if we return to the symmetric group and the Plancherel measure there, we can recall that $\dim \lambda$ has a Vandermonde determinant times a product of multiplicative factors – the Plancherel measure being the square of this dimension then also has a squared Vandermonde determinant coming up.

Remark 26. *Depending on which vertical section we're looking at, we may need to place phantom lozenges on both the left and the right picture, but all of this is still accounted for if we track the different multiplicative factors coming*

17

from $C - n$ and $B - n$ – the probability distribution **does look a bit different** *in the three cases.*
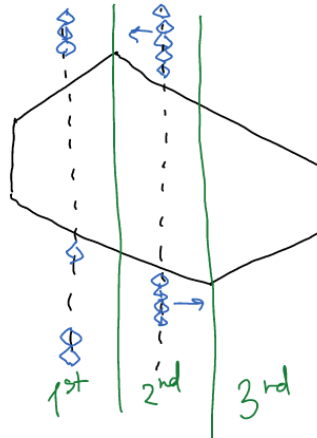
We can make a few modifications to the uniform measure as well:

- Use a q^{volume} weighting factor for each plane partition (where the volume is the number of boxes). Here's a simulation for $q = 0.96$ for a $60 \times 60 \times 60$ box:
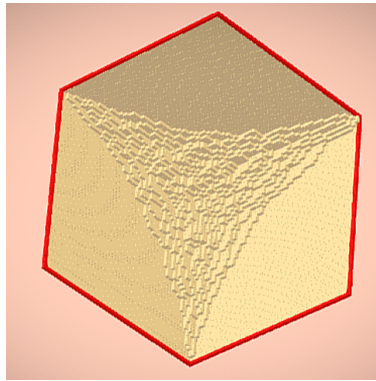
This is a good way to set up the problem when our plane partitions are not restricted to a box, so that we get a convergent sum. And the Vandermonde determinant in our formulas becomes something like $\prod_{i<j}(q^{\ell_i} - q^{\ell_j})$ instead.

- Modify the back wall of our partition, so that the base shape on which we're adding boxes looks (for example) like
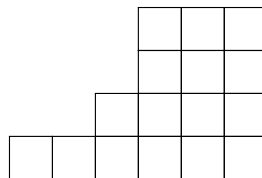
This creates what is called a **skew plane partition** (because the base shape is called a **skew partition**). And here is a numerical simulation of the q^{volume} measure with a nontrivial back wall:
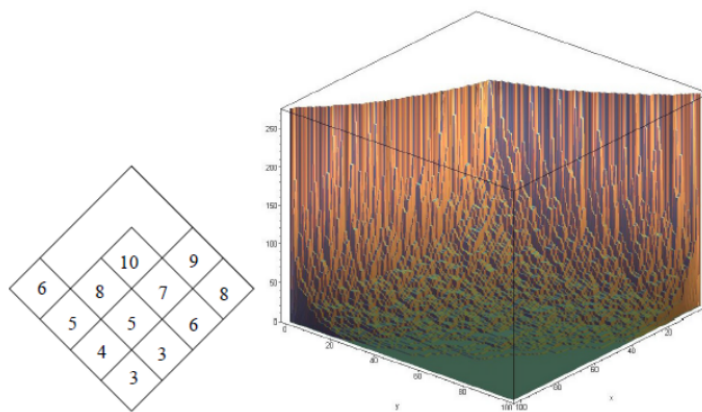
This is all we'll say about Gelfand-Tsetlin patterns and lozenge tilings for now – we'll come back more to the representation theory of the unitary group later on, but for now we'll move on to **domino tilings**. To give a natural bridge at the level of definitions, recall that we described our rhombus tilings as **dimer models** (perfect matchings on a hexagonal lattice). We can then also define a similar thing on the square lattice (the matchings are drawn in blue):
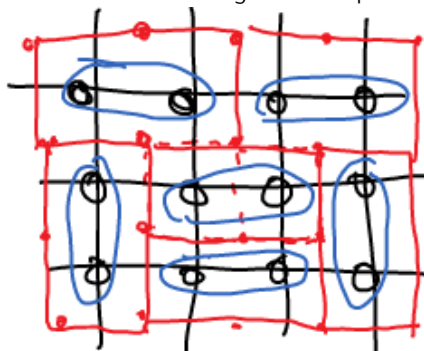
The elementary cells of the dual lattice (in red) are then squares, and we connect two red squares if we have a matching between them. And the 1×2 shapes we end up with look exactly like dominoes (in other words, dimers on the square lattice are in one-to-one correspondence with domino tilings of the dual lattice), and we'll talk about this object next time! It'll lead us to other combinatorial topics (like the alternating sign matrices and the six-vertex model).

Remark 27. *Dimers on the triangular lattice are much more difficult to study than the above two cases, because the graph is not bipartite! And because of this, we basically don't know any results for the triangular lattice that are close in power to those for the hexagonal and square and other bipartite lattices.*

4 February 25, 2021

Today's class is a seminar by **Jimmy He**, titled **A central limit theorem for descents of a Mallows permutation and its inverse**. We'll start by explaining how the Mallows model fits into integrable probability (there's been some work which connects this to particle systems, which we'll make sure to discuss).

We'll use this to define the Mallows model:

In particular, this converges to the uniform distribution as $q \to 1$, so the Mallows measure is a deformation of that uniform distribution.

This Mallows measure is connected to integrable probability through a group theory object:

Additionally, it turns out that Weyl groups are always Coxeter groups, and the affine symmetric group $\tilde{S}_n$ is an example of an infinite Coxeter group. And the reason we care about these Coxeter groups is that we can define a certain q-deformation of their group algebra – the result is called the **Hecke algebra**, and we can define dynamics called **Hecke algebra walks** whose stationary distribution is the Mallows measure. Furthermore, if we project these dynamics onto certain quotients, we can interpret the results as particle systems (for example, we can get ASEP with closed or half-open boundary conditions from various Coxeter groups).

> **Example 33**
>
> Elements of the parabolic quotient group $S_n/(S_k \times S_{n-k})$ correspond to "having two types of particles" (we forget about the differentiation between $(1, 2, \cdots, k)$, and also between $(k + 1, \cdots, n)$, which we can think of as a filled-in hole and an empty hole.

Then we can interpret this as a system with k particles on a line of length n, and the dynamics consist of swapping the particles at two sites (with some probability depending on q). And the Mallows measure does turn out to be the stationary measure of this system under an appropriate description.

We can note a few properties of the Mallows measure:

- The stationary distribution μ_q approaches the uniform distribution as $q \to 1$, and it concentrates to the identity element as $q \to 0$.

- When $q < 1$ is held constant, this distribution behaves very differently from the uniform distribution: for example, $P(w(1) = 1) \to 1 - q$ as $n \to \infty$, while for a uniform permutation this probability is $\frac{1}{n}$. but in general, it's hard to study $P(w(i) = j)$, and there aren't any tractable formulas.

- The inverse of a Mallows permutation has the same distribution with the same parameter q. (This can be more easily thought of in the Coxeter group formulation, because our s_is are all involutions. So thinking about the group structure sometimes leads to probabilistic properties more easily than studying the combinatorics!)

- The reverse permutation w^{rev}, obtained by reversing the one-line notation of the permutation, is distributed as $\mu_{1/q}$ if w^{rev} is distributed as μ_q. (This last fact will be useful for restricting ourselves to $q < 1$.)

> **Fact 34**
>
> Previous work on the Mallows model has been done: there is a lot known about the distribution of the cycle lengths, as well as the longest increasing subsequence, and there's a **phase transition** going on in both of these cases. Specifically, there's a Tracy-Widom distribution here for the uniform measure (so no central limit theorem), but it turns out that there is a central limit theorem for the Mallows measure! And there's been some work done to connect this to Markov chains and to define a **Mallows process**.

Remark 35. *The Mallows model is supposed to be a non-uniform measure on "ranked data," where we have reason to believe that there's concentration around one permutation. And people do indeed write papers comparing real data to this model, and one benefit here is that the algebraic properties makes the Mallows measure easy to sample from.*

With that, we'll move into new work. We'll talk about the distribution of descents in Mallows permutations, and that's because descents are possibly the most well-understood statistic in permutation. They form what's called a (determinantal) **one-dependent** process under the uniform measure:

> **Fact 36**
>
> Consider the random variables $d_i(w) = \begin{cases} 1 & w \text{ has a descent at } i \\ 0 & \text{otherwise} \end{cases}$. Then we know that $d_i(w), d_j(w)$ are independent if $|i - j| \geq 2$.

This fact is even true for the Mallows permutations, and thus proving a central limit theorem for descents is very easy (there is weak dependence between the d_is). But trying to undersatnd the joint distribution between the descents

of w and w^{-1} is more difficult. It turns out that **we do get independent Gaussians** (of the same mean and variance) for the uniform distribution, and we'll now present a new result for the Mallows distribution:

Theorem 37

Let μ and σ be the mean and variance of $\text{des}(w) + \text{des}(w^{-1})$, where w is Mallows distributed with parameter q. Then under a certain metric d_W, and as long as $nq \to \infty$ (even if q varies on n),

$$d_W \left(\frac{\text{des}(w) + \text{des}(w^{-1}) - \mu}{\sigma}, Z \right) \le C q^{-1/2} n^{-1/2},$$

and this is sharp in the dependence on q.

This is nice because it helps us get convergence results regardless of how q goes to 1 or 0, and there is a "phase transition" of some sort (but not of the same type as described above) if $q = \frac{\lambda}{n}$, because we go from a Gaussian behavior to a Poisson behavior. (This is because the descents of w and w^{-1} become equal in the limit: w is an involution with high probability.) The proof of this result uses **Stein's method** (with a particular coupling), and the main difficulty is estimating the covariance terms.

Furthermore, we can get a joint covariance result if we keep q constant: it turns out that normalizing the descents of w and w^{-1} gives us convergence to some joint Gaussian $N \left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right)$. (So the more descents we have for w, the more descents we'll have for w^{-1}.) It's a fact that $0 < \rho < 1$ for any $q \in (0, 1)$ (so there is some positive covariance), but calculating this is difficult because it's hard to compute the covariance quantitatively.

So we'll talk a bit about the proof details now: first, we'll discuss Stein's method. (Everything here can be done with continuous distributions, but we'll stick to discrete ones.)

Definition 38

Let X be a (nonnegative) discrete random variable with positive expectation. A random variable X^* has the **size-bias** distribution if

$$\mathbb{P}(X^* = x) = \frac{x \mathbb{P}(X = x)}{\mathbb{E}(X)}.$$

In other words, we bias for a distribution to be large. This actually comes up (for example) if we randomly arrive at a bus stop and want to know the expectation value of how long we need to wait (the **waiting time paradox**).

Theorem 39

Let (X, X^*) be a coupling where X^* has the size-bias distribution with respect to X. Then under the same d_W metric as above,

$$d_W \left(\frac{X - \mu}{\sigma}, Z \right) \le \frac{\mu}{\sigma^2} \sqrt{\frac{2}{\pi}} \sqrt{\text{Var}(\mathbb{E}(X - X^*|X))} + \frac{\mu}{\sigma^3} \mathbb{E}[(X - X^*)^2].$$

The idea is that we expect normal behavior if X and X^* are "close" to each other. So we can prove a CLT by computing variance elements of various random variables, which is usually easier to do!

So we'll bias a permutation as follows: if w is a permutation, we define w_i^* to be the permutation which has a descent at i (by swapping $(i, i+1)$ if necessary), and similarly define w_{-i}^* to be this permutation for w^{-1}, which we then invert again (basically swap the numbers whose image are i and $i+1$). Then we define w^* to be the random permutation where we pick $w_{\pm i}^*$ for $\pm$ and $i \in \{1, \cdots, n-1\}$ uniformly at random.

> **Example 40**
>
> The permutation 23145 becomes 23415 if we pick $3, +$ and 24135 if we pick $3, -$.

And we claim that doing this creates a good inverse:

> **Proposition 41**
>
> If w is Mallows distributed, then $\mathrm{des}(w^*) + \mathrm{des}((w^*)^{-1})$ is size-bias distributed with respect to $\mathrm{des}(w) + \mathrm{des}(w^{-1})$.

This kind of strategy works as long as our statistic depends only on a bounded neighborhood (and also for a Coxeter group if we use the Coxeter Dynkin diagram as a measurement of "distance"), and the proof of the result relies heavily on the algebraic structure of Mallows permutations. And there's a more general method for constructing size-biased couplings: the idea is to take $\sum X_i$, pick an X_i, size-bias that particular X_i, and then adjust the other X_js so that the conditional distribution $X_j' | X_i'$ is the same as $X_j | X_i$. To prove that our proposition is valid, we're using the indicator functions

$$x_i = 1\{w \text{ has descent at } i\}$$

(and similar for negative i), and what we need to check is that if we condition on i and $+$, the distribution of w^* is the same as if we conditioned w to have a descent at i (and note that the X_i''s are always equal to 1). This is true because swapping $(i, i+1)$ gives us a map $w \to w(i, i+1)$ which bijects permutations with and without a descent at i, and this map only changes the probability by a factor of q (because we add exactly 1 descent). And now, if w has a descent at i, we have $w^* = w$, and otherwise we have $w^* = w(i, i+1)$. And in both cases, the conditional distributions are the same!

So now to show Theorem 39, we can condition on w to make the coupling useful, and the only source of randomness left in that case is based on our choice of i and $\pm$. So the details from there are to calculate various covariance terms: for example, $\mathrm{des}(w) - \mathrm{des}(w_i^*)$ is independent of $\mathrm{des}(w) - \mathrm{des}(w_j^*)$ if $|i - j| \geq 4$ (because we're caring about the relative order of only a few numbers). But it's harder to deal with w and w^{-1} simultaneously, and it turns out that we deal with variables like the indicator of $w(i) - w(i+1) = 1$. Correlations of those kinds of events are weak for constant q — for example, the correlation decays exponentially in the distance $|i - j|$.

And now we're ready to return to the bivariate distribution: it turns out the difficulty here is that we need to show that $\mathrm{Cov}(\mathrm{des}(w), \mathrm{des}(w^{-1})) / \mathrm{Var}(\mathrm{des}(w))$ has a limit, and we do this with the Mallows process:

> **Definition 42**
>
> The **Mallows process** is a random bijection $w : \mathbb{N} \to \mathbb{N}$ as follows: $w(1)$ is a random variable distributed as $\mathrm{Geom}(1 - q)$, and given the values $w(1), \cdots, w(i)$, we define $w(i+1)$ to be the Nth largest element in $\mathbb{N}$ that hasn't already been chosen, where N is again distributed as $\mathrm{Geom}(1 - q)$.

If we then restrict w to $\{1, 2, \cdots, n\}$ and replace $w(i)$ with the relative rank, we get a permutation $w^{(n)}$, and it turns out this is Mallows distributed (so this is a "finite-dimensional distribution" of the Mallows process, in some sense). This process is **regenerative**, which means that the first time $w(i) \leq T$ for all $i \leq T$ (meaning we get a real permutation of the numbers 1 through T),

$$(w(i + T) - T)_{i=1}^{\infty} = (w(i))_{i=1}^{\infty}.$$

If T_i is the time between these kinds of regenerations, and w_i are then the induced permutations on the corresponding intervals, then the T_is and w_is are iid, and this fact was used to prove a Central Limit Theorem for the longest

increasing subsequence (under the Mallows distribution). The point is that as long as we can write quantities as sums over these individual intervals of length T_i (because of the underlying renewal times), we just need to do a few moment bounds for the T_is!

5   March 2, 2021

Last time, we spoke mostly about plane partitions (which can also be viewed as cubed surfaces in 3D, as rhombus tilings, or as dimers on a hexagonal lattice). We'll talk more in detail about domino tilings and dimers on the **square lattice** today – while the structure appears similar, the simplest domains that we're studying tend to be **finite** (while it was infinite in the hexagonal case). So we'll be looking at a single finite domain, known as the **Aztec diamond**, for much of our study today. Here's an example of an Aztec diamond of size 4:
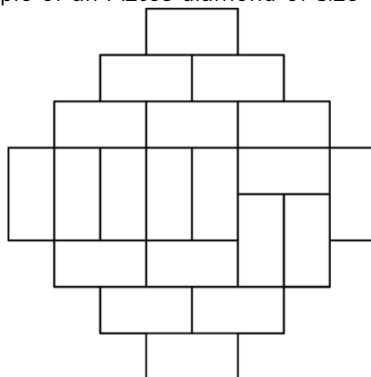
This is basically a square rotated by 45 degrees, but it is important how the boundary is specified – in fact, the behavior of the random tilings are very different from something like a regular square. To do probability on an object like this, we will consider a random **uniform** domino tiling (especially because there are no easy product measures to construct immediately).

It turns out that counting the number of domino tilings is not so difficult:

Fact 43

There are $2^{n(n+1)/2}$ different tilings of the Aztec diamond of size n.

We can verify this directly for $n = 1$, where the Aztec diamond is just a 2×2 square, or for $n = 2$, whose shape is found below:
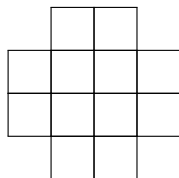
(We can verify ourselves that there are 8 domino tilings in the picture above.) Surprisingly nice numbers like this usually come with a reason, and there is in fact a lot of structure behind this object!

Working only with horizontal and vertical dominoes doesn't tell us very much, so we often color the square lattice in a chessboard-like pattern. Then we find that there are actually **four types of distinguishable dominoes** instead of two:
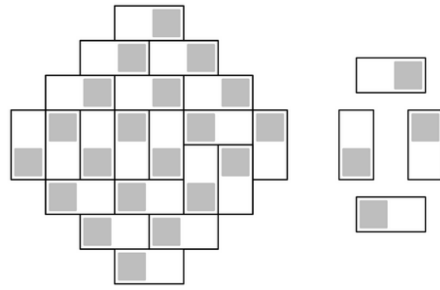
These four types of dominoes are often called "north", "east," "south," and "west", corresponding to the corner of the Aztec diamond that they fit into. And in fact, the picture is often colored based on the four types of dominoes, to make them more distinguishable from each other:
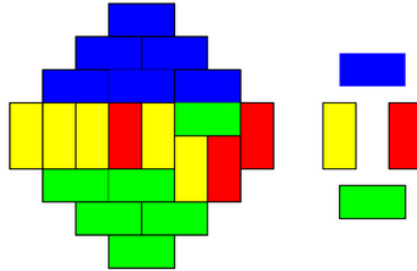
(One of the yellow dominoes should really be red.) And now it is much easier for us to distinguish between the four kinds of dominoes by eye, and we can start seeing how this is related combinatorially to some of the other objects we've already discussed.

Much like with the rhombus tilings, we can set up a useful coordinate system (unfortunately, the coloring scheme has been flipped from the above example):
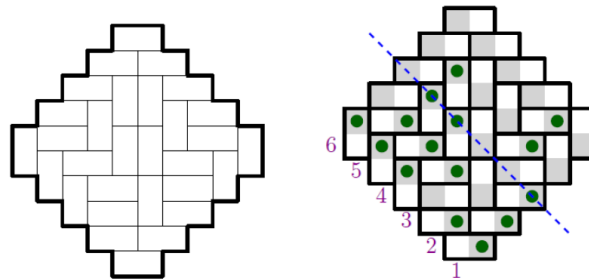
Here, we have an Aztec diamond of size 5. Along each dashed line (of dark checkerboard squares), we only look at dominoes crossing the line of two types, and we mark them with a green dot. We can see that there are 5 possible dashed lines, and there are 6 possible locations where green dots (which we can think of as "particles") can go. So that gives us an $(n+1) \times n$ matrix, and we'll in fact add an extra layer of dimers in the bottom left corner so that we have a 6×6 matrix, and those will always give us green dots.
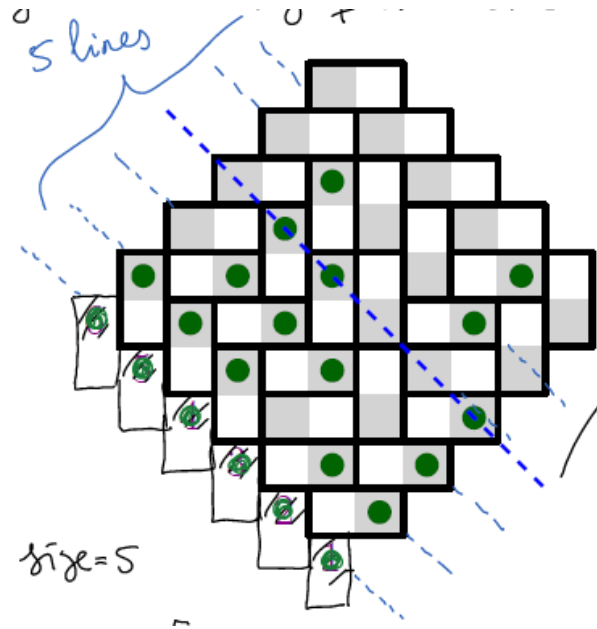
Rotating the picture 135 degrees in the clockwise direction makes the top row always all 1s, and then the rest we can read off from the picture:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

We notice that this matrix has $6, 5, 4, 3, 2, 1$ zeros in the different rows, from top to bottom, and this is in fact a general phenomenon. And if we record the positions of 1s, we get something similar to a Gelfand-Tsetlin pattern:

$$
\begin{array}{ccccccccccc}
1 & & 2 & & 3 & & 4 & & 5 & & 6 \\
& 1 & & 2 & & 4 & & 5 & & 6 & \\
& & 1 & & 3 & & 4 & & 5 & & \\
& & & 1 & & 4 & & 5 & & & \\
& & & & 2 & & 5 & & & & \\
& & & & & 2 & & & & &
\end{array}
$$

Notice that, just like in the Gelfand-Tsetlin patterns, we have weak interlacing, but there is a small difference (after all, if the pictures would be the same otherwise). And the difference is that **the rows are distinct** in our description here, while coinciding numbers were okay in the Gelfand-Tsetlin pattern. We call this type of triangular array a **monotone** or **strict** or **semi-strict Gelfand-Tsetlin pattern**. But the point is that we have a similar combinatorics as before, and we'll reinforce that feeling now.

(This basically tells us about the positions of the 1s in the kth row of our matrix.) Again, we see the squared Vandermonde determinant, as well as a multiplicative factor where the binomial distribution plays a role. (And that's in fact directly related to the powers of 2 for the total number of tilings!)

Remark 45. *If we bias horizontal dominoes more than vertical dominoes, we can get a similar expression as above, but with a biased binomial distribution.*

Just like with the lozenge tilings, domino tilings can be described with something called a **height function**. Recall that we started with a three-dimensional connotation when we defined plane partitions and lozenge tilings – the idea there was that each lozenge is some height above a chosen plane, and then the distance of the centers of the faces gives us a function on the flat picture. It may not be clear how domino tilings can also have a "height function" or a three-dimensional representation, and it's a pretty remarkable fact that these types of matchings can indeed be made into a three-dimensional object for an **arbitrary bipartite planar graph**. But for now, we'll just mention the result: we define the height function on every vertex of our Aztec diamond by defining them differently for each domino:
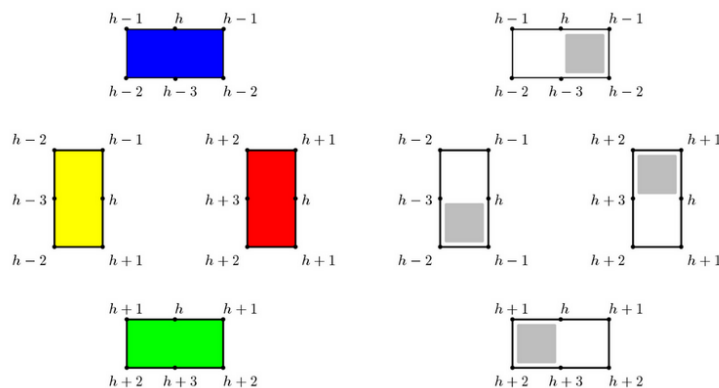
The idea is that if we go around a boundary of a domino and we see a **dark square on our left**, then we increase the height function by 1. But if we see a light square, we decrease the height function by 1. So then setting the height function at 0 at some base point means that we can compute the height function everywhere else, and the tricky thing to prove is that this is actually well-defined. (It turns out to not be true if the domain has holes!)

Above is a uniformly random sample of the domino tilings – we see again a **phase transition** between the (liquid) circle and the (frozen) corners of the Aztec diamond. (It is actually a circle – we'll get to that later on)! And here's the three-dimensional representation of that same picture:
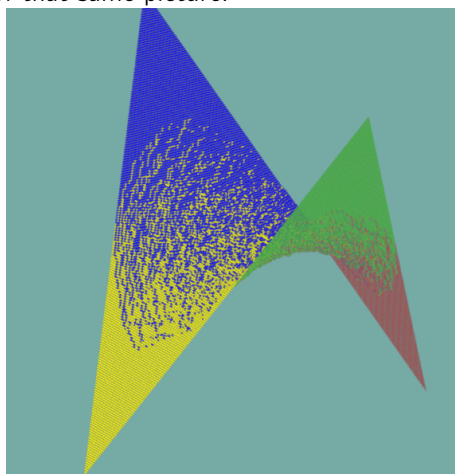
And in fact, we can get a third phase (the **gaseous** phase) if we change the weights of our dominos, biasing in a periodic manner: we end up with an almost perfectly flat piece in the middle of our three-dimensional picture, and the structure of the "landscape" is much more independent than in the liquid phase. (But we can play around with that at the link above!)

Remark 47. *And with this three-dimensional visualization, we can understand why the ordinary square and the Aztec diamond are different: there is a saddle in the Aztec diamond, while the landscape basically has "flat boundary conditions" in the ordinary square.*

We'll now continue on our landscape, talking about **alternating sign matrices** (ASMs) and how they connect to the domino tilings. Alternating sign matrices are a generalization of the permutation matrices (and this is a sign that we're increasing in the level of abstraction from where we started):

> **Definition 48**
>
> An **alternating sign matrix (ASM)** has matrix elements $0, 1, -1$, such that the sum of the entries in each row and column is 1, and such that 1s and -1s must interlace in each row and column (zeros between them are fine).

(Notice that this definition is the same as for a permutation matrix if we didn't allow ourselves -1s.) For example, if we take the 6×6 matrix that we produced from our domino tiling above, we can **turn it into an alternating sign**

matrix by subtracting row $(i+1)$ from row i from top to bottom:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(We can check that the interlacing condition for ASMs is the same as the interlacing condition for our monotone Gelfand-Tsetlin patterns.)

Remark 49. *ASMs actually appeared earlier on in a completely different context, called the* **Dodgson algorithm** *(Dodgson is the same person who wrote Alice in Wonderland under a different name): if we start with an $n \times n$ matrix whose determinant is not very easy to compute, we can look at the 2×2 minors and compute those determinants to get an $(n-1) \times (n-1)$ matrix. Then if we repeat this process again to get an $(n-2) \times (n-2)$ matrix, and then divide that matrix by the middle $(n-2) \times (n-2)$ part of our original matrix, we get the next step of our process. Then we can repeat this process by doing 2×2 minors of the current matrix and then dividing by the element-wise middle part of the previous matrix, and eventually we end up at the determinant of the $n \times n$ matrix.*

A natural generalization is to replace the definition of the 2×2 determinant $\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$ by the **λ-determinant**, given by

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}_\lambda = ad + \lambda bc.$$

(So $\lambda = -1$ gives us the usual answer.) If we repeat the Dodgson algorithm, instead of getting the determinant $\sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) \prod_{i=1}^n a_{i,\sigma(i)}$, we end up with a sum over ASMs! Mills, Robbins, and Rumsey used this in 1982 (at the dawn of symbolic computation by computers) to conjecture that the number of $(n+1) \times (n+1)$ alternating sign matrices is the expression

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

and this was proved by Zeilberg (1992) and then by Kuperberg (1995). It turns out the number of ASMs is much smaller than the number of domino tilings $2^{n(n+1)/2}$, so the correspondence we described cannot be a bijection: instead, the preimage of each ASM (to the domino tilings) is $2^{\text{number of } -1\text{s in the ASM}}$, and it turns out that each factor of 2 comes from a 2×2 **square in the original dimino tiling** (meaning that we have two possibilities for how to orient them, and this flipping doesn't change the point particle configuration).

> **Fact 50**
>
> The expression $\prod_{i=0}^n \frac{(3i+1)!}{(n+1+i)!}$ also happens to be the number of **TSCPPs (totally symmetric self-complementary plane partitions)**, which are lozenge tilings of the hexagon symmetric with respect to the full dihedral group D_6 (of order 12). This was a well-known open problem in enumerative combinatorics, and a paper just last year showed a proof of this fact.

Although we've touched a lot on combinatorics recently, the topics will now take a turn into the world of chemistry. The **six-vertex model** comes from the following story: in the 1930s, chemists were measuring **entropy** (in the chemical sense, not the mathematical one), which can be done using calorimeters (measuring the change in heat drawn away from a material) or using spectroscopy.

Near the absolute-zero temperature, entropy is supposed to vanish (this is the third law of thermodynamics), and this is often how we calibrate entropy in real measurements. But there ended up being **residual entropy** in substances such as ice, and for a long time it was unclear why this occurred. The first answer came from **Linus Pauling**, an American chemist:
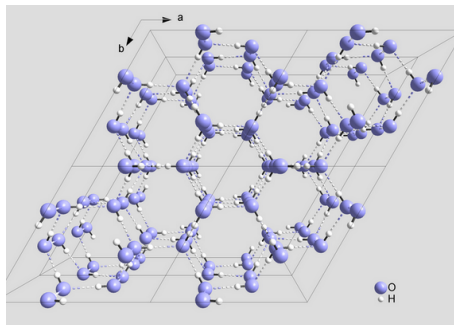
In the picture above, we see the most common crystal structure of ice, with oxygens in blue and hydrogens in white. There are four bonds for each oxygen, but the resolution (with what we know about valence electrons) is that everything is in hydrogen bonds, and two of the hydrogens are close and two are far apart from each oxygen. So there are $\binom{4}{2} = 6$ ways to arrange which two bonds are closer, and it turns out that this is indeed the source of residual entropy (because there is still this freedom of configurations)! The number of possibilities across the picture is of course not 6^N for N molecules, because we can't freely and consistently choose the combinations. But Pauling still guessed this value of 6, and the result turns out to be within 5 percent of the residual entropy! This was the starting point of the six-vertex model, and we'll understand how to relate it to the ASMs next lecture.

6 March 4, 2021

Last time, we started discussing the residual entropy of ice, using the number of legal configurations of hydrogens and oxygens to explain residual entropy. Linus Pauling estimated 6^n as an overestimate, but the actual asymptotics are still unknown today. And the point of this story is to give a transition into the **six-vertex model**.

Pauling's work was done in 1936, and in the 1950s, some calculations were done (numerically) to estimate the accuracy of the 6^n number. But dealing with a three-dimensional lattice is rather difficult, so for simplicity the model was replaced with a square lattice.

Definition 51

Consider a square lattice, with an oxygen at each site and a "hydrogen bond" at each edge. Each hydrogen bond is close to one of its vertices, and a valid configuration is one where each vertex has two hydrogen bonds close to it – this is known as **square ice**.
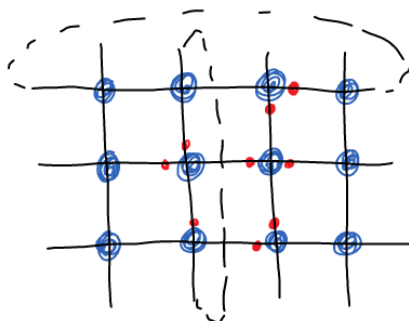
Configurations like tori or squares could indeed be numerically analyzed, and in fact Lieb computed the asymptotics for the torus in 1964. (We often impose periodic boundary conditions for simplicity.) And this was one of the primary reasons for the emergence of the field of integrable lattice models:

<div style="border:1px solid red; padding:10px;">

Definition 52

The **six-vertex model** is a probability measure on the valid configurations of O and H on a planar domain, where a configuration has weight proportional to $w_1^{\text{vertices of type 1}} w_2^{\text{vertices of type 2}} \cdots w_6^{\text{vertices of type 6}}$, where the 6 types of vertices are the $\binom{4}{2}$ different ways to pick two edges to be the hydrogen bonds close to the vertex. (Here, $w_1, \cdots, w_6$ are nonnegative real numbers.)

</div>

This model contains a range of different phenomena for different values of w: it seems like we only have 5 free parameters, because multiplying all parameters by c multiplies the total sum of weights by c^n (for some fixed number of vertices n), which doesn't change the probability measure. And in fact, there are three other conservation laws as well which are harder to see, and there are thus only two parameters left. It turns out one of them is important — it dictates the behavior of the model — and the other is also important in that it can be varied without changing the behavior much (which helps with the **solvability** of the model).

<div style="border:1px solid green; padding:10px;">

Example 53

Square ice is an example of the six-vertex model with the uniform measure $w_i = 1$ for all i.

</div>

Recall that we got to this point by considering alternating sign matrices, and it turns out that alternating sign matrices biject to configurations of the six-vertex model with certain boundary conditions:
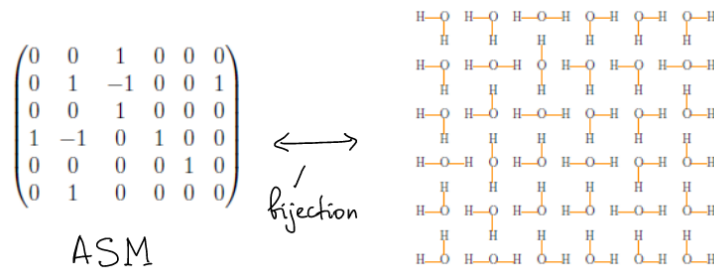
We can notice that there are all Hs on the left and right of our diagram, and there are no Hs on the top and bottom. And we can immediately see how the bijection works: -1s correspond to vertical molecules, and $+1$s correspond to horizontal molecules. (And it turns out these boundary conditions constrain the remaining possibilities.)

But remember that even our alternating sign matrices came from a certain map between domino tilings of the Aztec diamond and ASMs, and it turns out that the domino tilings of the Aztec diamond can also be mapped to the six-vertex model with the same boundary conditions (but with different ws), and this gets us to the **free-fermion six vertex model**.

<div style="border:1px solid orange; padding:10px;">

Fact 54

The reason that Lieb was able to solve the square ice asymptotic problem was because of a connection to a model in quantum mechanics called the **Heisenberg model for ferromagnetism**.

</div>

Heisenberg was one of the founders of quantum mechanics — we won't talk about it much, but the basic setup for quantum mechanics is that we have a Hilbert space $\mathcal{H}$ and a Hamiltonian H, which is a (typically unbounded) energy

operator $\mathcal{H} \to \mathcal{H}$. We're often interested in solving the Schrodinger equation

$$\frac{d\psi(t)}{dt} = H\psi(t)$$

for a time-dependent **wavefunction** ψ. One of the first models that came up in this study for multiple particles was the Heisenberg model, in which we have a bunch of particles at n sites with dipoles. Then $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$ (the Hilbert space is spanned by the "up" and "down" states), and the Hamiltonian for the magnetic interaction is

$$H = -\frac{1}{2} \sum_{j=1}^{n} \left(J_1 \sigma_j^x \sigma_{j+1}^x + J_2 \sigma_j^y \sigma_{j+1}^y + J_3 \sigma_j^z \sigma_{j+1}^z \right) + h \sum_{j=1}^{n} \sigma_j^z,$$

where J_1, J_2, J_3, h are numbers, the σs are the **Pauli matrices**

$$\sigma^x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma^y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma^z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

and σ_j^x is the tensor product $I \otimes I \otimes \cdots \otimes \sigma^x \otimes 1 \otimes \cdots \otimes I$, with the σ^x in the jth spot. (In the classical case, we could imagine having the analogous expression $H = \sum_j \sigma_j \sigma_{j+1}$ with $\sigma_j \in \{-1, 1\}$, which is the **Ising model**.) When $J_1 = J_2 = J_3$, we call this system the **isotropic Heisenberg model** or **XXX model**, and there is also an **XXZ** and **XYZ model** (for the cases where $J_1 = J_2 \neq J_3$ and when they are all different, respectively).

Remark 55. *We can also set $J_3 = 0$, which gets us the* **XY model** *and is actually the free-fermionic case for the model.*

It'll take a bit of time for us to get to a pretty exact answer for why the Heisenberg model is related to the six-vertex model, but the idea is to **read the vertices of the six-vertex model layer by layer**, viewing the vertical occupations as 0s and 1s. Then each row gives us a matrix in $(\mathbb{C}^2)^{\otimes n}$ – we don't quite get the Heisenberg model like this, but we can take a certain logarithmic derivative and it'll get us to the XXZ model. (And the XYZ model is related to something called the **8-vertex model**, which we won't get into here.)

But the point is that the XXZ and XXX model were already approached in the 1930s, and the Schrodinger equation was solved in those cases.

Fact 56

Physicists often consider correspondences between a two-dimensional statistical model (like the six-vertex model) and a (1+1)-dimensional quantum model (treating one of the directions as time). But we won't be going too much in this direction, either.

Instead, we can consider a different case of this six-vertex model:

Definition 57

The **stochastic six-vertex model** is defined as the six-vertex model with the weights shown below.

$$1 \quad 1 \quad b_1 \quad b_2 \quad 1-b_1 \quad 1-b_2$$
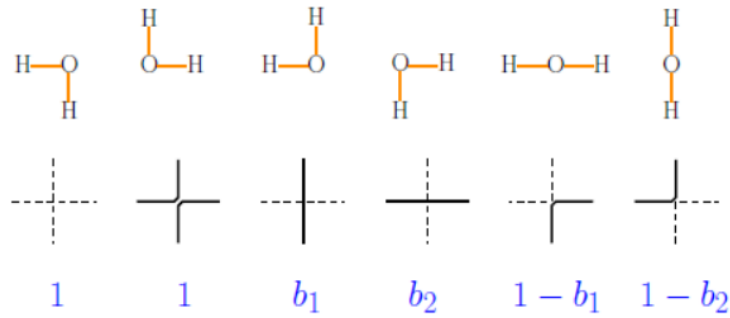
Between the weights and the molecules above, we can see that there is a new way of representing our configurations (as a set of dashed and thick lines). It turns out that this new way of looking at the picture is useful, because it helps us tell whether a particular configuration is allowed or not with the naked eye:
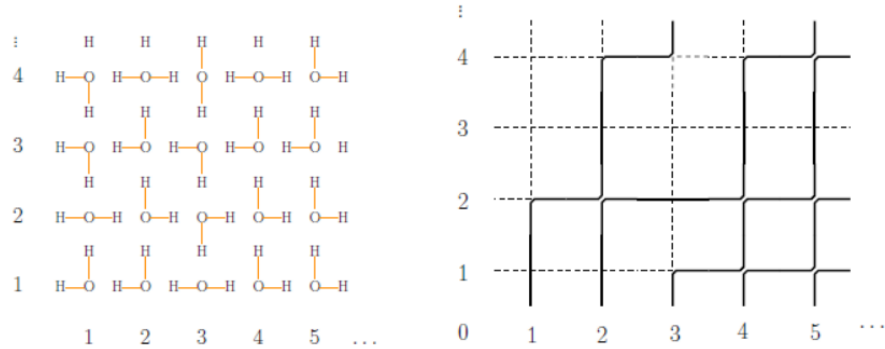
The idea is that this configuration turns into a **sequence of up-right paths** that are allowed to touch, and in fact the weights that we chose above allow us to sample this configuration step by step! The idea is that the weights b_1 and $1-b_1$ come up in the following way: if we enter a vertex from the bottom, it has probability b_1 of going up and a probability $1-b_1$ of going to the right. So we can start with a bunch of paths originating from the bottom and evolve them step by step!

Some physics models fall under the category of **equilibrium statistical physics**: after some local interactions, the system settles to some equilibrium. But this is an example of **nonequilibrium statistical physics** instead: we can think of the left edge as being "empty," the bottom edge as being "full," and we can think of time as evolving in the up-right direction (and the question being how the particles mix over time). We get a picture that looks something like this:
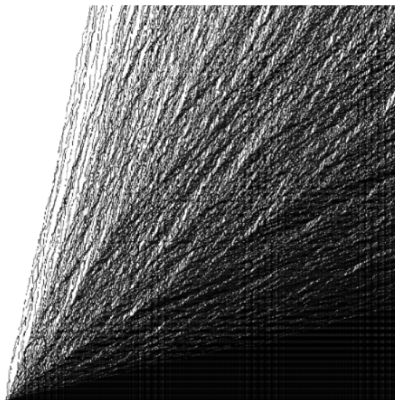
We can notice in the picture above that below a certain slope, all of the paths are packed (there are no empty edges), and above a certain slope, there are no lines at all. So this is somewhat similar to the tilings from previous

lecture, but it turns out that the statistical properties of this picture are very different from the others! (And again, there is a key difference between the equilibrium situation where we fix boundary conditions, like in the Aztec diamond, and the nonequilibrium situation where we start with something separated and allow things to mix.)

For now, we'll think about this picture in another way: we place particles on the vertical edges of our paths, and we send time upward (there are other choices of direction as well).
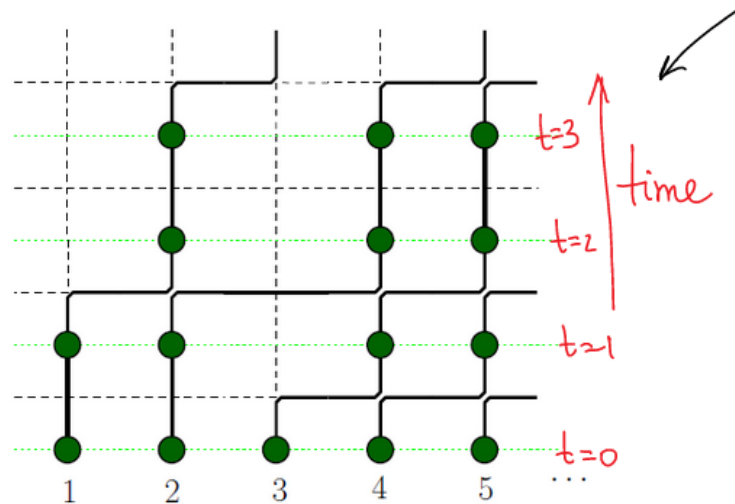
The idea is to degenerate this model so that particles want to travel diagonally (meaning that our particles step up, then right, then up, then right, and so on). And we can do this by sending the probability of going straight up or right to 0, which means we set $b_1 = b_2 = 0$. If we just do this, nothing interesting happens, because everything does indeed travel diagonally. But if we do this more carefully by taking $b_1 = q\varepsilon$ and $b_2 = p\varepsilon$, then looking at times on the order of ε^{-1} will give us a constant-order number of straight jumps (approximating a Poisson process).

Indeed, rescaling time by ε^{-1} and taking $\varepsilon \to 0$ means that our particles evolve according to a **continuous-time Markov process**. If we get rid of the constant drift term, meaning that we look at our coordinates in a moving-diagonal coordinate system, our green particles will then live on a one-dimensional lattice, jumping to the right at rate p (corresponding to two consecutive steps to the right) and jumping to the left at rate q (corresponding to two consecutive jumps upward). (In other words, we wait for an **exponentially distributed** amount of time proportional to $ae^{-at}dt$ for all $t \geq 0$, where $a = p + q$, and then we jump to the left with probability $\frac{q}{p+q}$ and to the right with probability $\frac{p}{p+q}$.) However, we sometimes have a situation where a particle's destination position is already occupied – then those jumps are forbidden and nothing happens.

Definition 58

The description above (with particles jumping to the left or right with some rate, as long as the spot is not occupied) is a Markov chain called the **Asymmetric Simple Exclusion Process** (ASEP).

The "asymmetric" part of ASEP comes from $p \neq q$, "simple" means that particles are all equivalent, and "exclusion process" means that there is at most one particle per site. This system first appeared in biology in the late 1960s, and the limit transition from the six-vertex model to ASEP is the same transition as the one to the (quantum mechanical) XXZ model – the **generator** of the ASEP is actually conjugated to the XXZ model's Hamiltonian. The operator e^{itH} is a unitary operator that describes the quantum evolution of a system, and (in a parallel description) e^{tH} evolves the particle system forward. (So the imaginary unit is the main difference, and ASEP is sometimes called the "XXZ model in imaginary time.") Unfortunately, the behavior of e^{itH} and e^{tH} are completely different from each other, but luckily the eigenfunctions are the same whether or not we have this i.

We'll continue moving along, giving a survey and talking about connections at a high level – again, once all of the pieces are on the board, we can pick and choose for our own presentations. There's yet another interpretation of the ASEP that we can think about: if we have particles on the $\mathbb{Z}$ lattice (notice that we've dropped the dimension from 2 to 1), we can add a second dimension by drawing a graph above the lattice, as shown below, with midpoints projecting to the nodes of the lattice and slopes ± 1 (based on occupation).
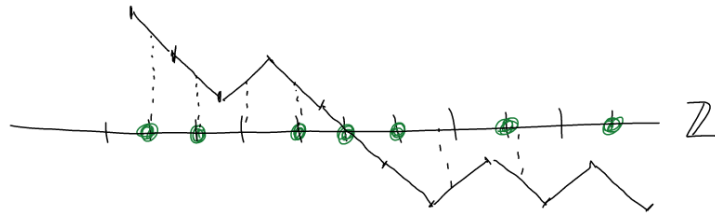
Drawing this line is equivalent to giving a point configuration, and we can check that particles jumping to the right and left adding or removing a box from the bottom:
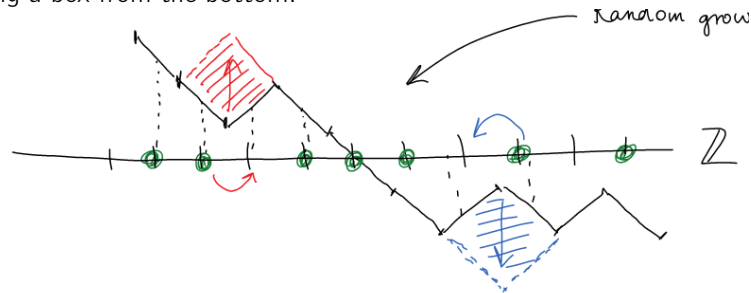
This is known as a **random growth model**: we have an **interface** which splits the plane into two parts, and the interface randomly changes. And it's an interesting question to think about how such an interface evolves over time (especially in soft matter in physics), because there are many physical phenomena which demonstrate easily-modeled behavior.

Example 59

The simplest case of ASEP is called **TASEP**, and it corresponds to the case where we set $q = 0$ (meaning that particles only jump to the right, and thus the interface always moves upward).

It turns out that TASEP is useful because it connects to a process called **Last Passage Percolation** (LPP), which we'll think about in the next lecture!

7 March 11, 2021

Last time, we described the growth-model interpretation of the ASEP (Asymmetric Simple Exclusion Process), in which particles being present or not on a one-dimensional lattice corresponds to adding or removing a box from a two-dimensional picture. We mentioned that a simple case of ASEP, TASEP, comes up when particles only jump to the right (and thus the interface only grows upward). Here is a moment in time of a TASEP simulation, with particle locations displayed at the bottom and correpsonding interface above, with initial conditions particles being placed at every other site. This is called the **flat initial condition**:

As time runs on, the particles continue moving to the right, and the red line is the law-of-large-numbers predicted behavior for the growth interface (which generally moves upward). We can also use other initial conditions − for example, here is a snapshot of TASEP evolution, where particles are initially placed at all negative lattice points. This is called the **step initial condition**:
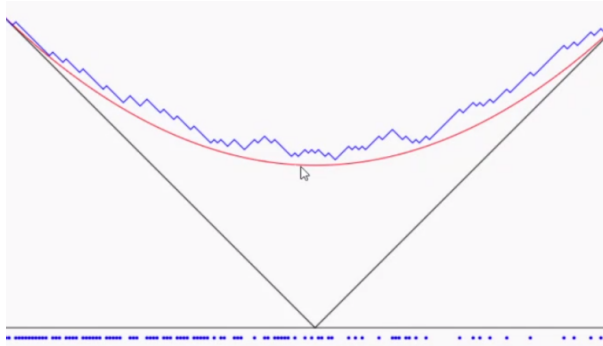
Finally, here is a snapshot from the **half flat initial condition**, in which particles are initially placed at every other site but only at the negative lattice points:

It turns out that the interface generally lies above the law-of-large-numbers predicted behavior, and if we subtract off the height from the red line, we get certain distributions of fluctuations (which are different for different initial conditions but of the same order of magnitude $\sqrt[3]{t}$). But we'll get to that later on.

Another interesting modification we can make to this model is to slow down the first particle, which causes a kind of "traffic jam:"
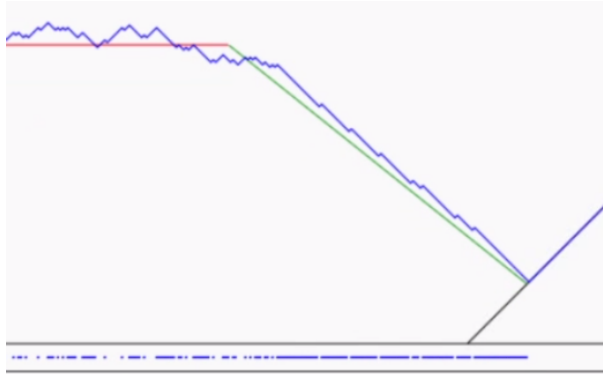
The fluctuations of the traffic-jam region turn out to be pretty large compared to the fluctuations in the other parts (we'll have $\sqrt{t}$ instead of $\sqrt[3]{t}$)! And the reason for this is that the fluctuations mostly come from the slow particle at the front, and the central limit theorem behavior dominates there.

But conclusions about fluctuations of the interface, traffic jams, and so on turn out to be more broadly applicable, and we'll see that soon! For now, we'll return to our journey through objects in integrable probability, and in particular we'll relate this growing interface model to **percolation**. We'll use the step initial condition from TASEP (meaning particles are packed on one side of the lattice), and we'll add boxes to our interface like in TASEP. The idea is that in each box that is about to be "eaten" by the interface, there is an absorption time before the box joins the interface:
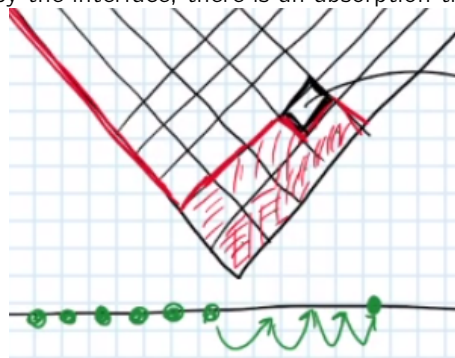
We can notice that the boxes in the right-most row correspond to the jumps of the rightmost particle, and thus the absorption times are basically the amount of time required before the particles jump! So this allows us to fill a quadrant with iid random variables w_{ij} (for positive integers i, j), and in the continuous-time TASEP case, they're exponentially distributed random variables.

Remark 60. *This kind of structure doesn't work for ASEP, because the interface can move up and down, meaning that there are boxes that can be potentially added or removed several times (so absorption and deletion times are not clear for each box).*

So from now on, we'll forget about the particles and just encode the absorption times. It's then reasonable to ask how long it will take to absorb a particular box, given those times:

37
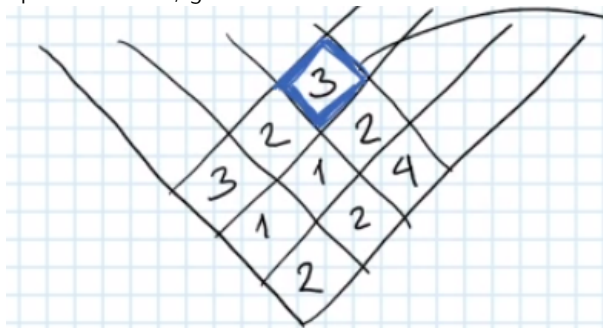
(Remember that these absorption times are random, so we're picking particular values of those random variables for illustration here.) The bottom box here is easy to compute, because it will take time 2 to absorb it, and then the 2 and 1 above that bottom box take $2 + 2 = 4$ and $2 + 1 = 3$ total time to be absorbed, respectively.

But then the picture becomes more complicated after that: we can only absorb the 1 directly above the corner 2 once all three of the boxes below it are absorbed, which happens only at time $\max(3, 4) + 1 = 5$. So this gives us a general answer: the absorption time of any box is the **maximum sum of absorption times over all directed paths between the bottom corner and the target point box**, and this is known as the **last passage percolation time**.

To explain the word "percolation" here, we can imagine a liquid that is trying to invade through the boxes, and we want to know the worst time it takes to percolate to a given box. And this is a useful formulation for a few reasons: first of all, it makes sense to define this model for arbitrary random variables w_{ij}, which means that the model can have much more structure (in either an integrable or asymptotic manner). So this class of percolation problems is diverse, and this is a large domain of modern probability.

We'll stick to the integrable realm, though, and we'll take a particular limit of our w_{ij} random variables now. Suppose that the w_{ij}s are **often zero**, which we'll do by taking a Bernoulli random variable

$$\mathbb{P}(w_{ij} = 0) = 1 - \varepsilon^2, \quad \mathbb{P}(w_{ij} = 1) = \varepsilon^2$$

and taking $\varepsilon \to 0$. Then most of the boxes are absorbed immedaitely (because $w_{ij} = 0$), and there is are particules with $w_{ij} = 1$ of density ε^2. We can then scale down our lattice by ε^{-1}, so that we always have the expected number of w_{ij}s equal to 1 constant in a finite domain $\mathcal{D}$, and the number of w_{ij}s will be Poisson distributed with parameter equal to the area of the domain:

$$\mathbb{P}(\#\{(i, j) \in \mathcal{D} : w_{ij} = 1\} = k) = e^{-\mathcal{D}} \frac{|\mathcal{D}|^k}{k!}.$$

We can still ask for the last-passage percolation time for any point in such a picture, and we calculate this by looking at the set of paths that connect the root to that particular point (always fitting in the appropriate cones), and finding the **maximum number of points** that we can pass through in a single path.
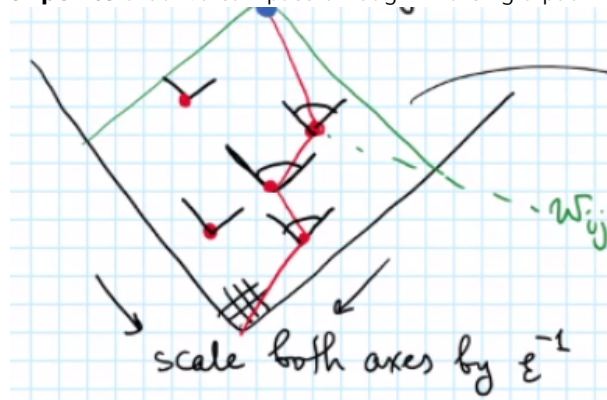
An alternative way to understand the locations of the points where $w_{ij} = 1$ is that we first sample the number of points within our domain, and then we place them in the domain uniformly at random with respect to the Lebesgue measure. But because we know that paths must respect the appropriate cones, we can think about sending rays from each particle and canceling them out when they meet other rays:
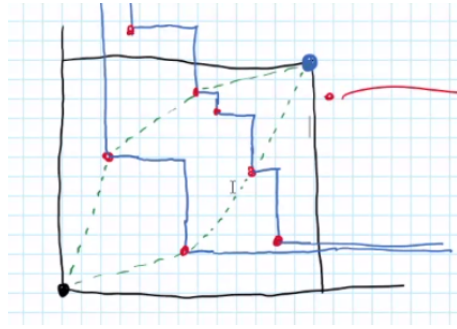
This forms a set of broken lines that separate the root and the target point, so we get the optimal answer by passing through one per line! We can draw these blue rays with a different interpretation of this situation: we can think of the two rays as **spacetime trajectories**, which are represented by a left and right wall that appears at the top of the picture below.
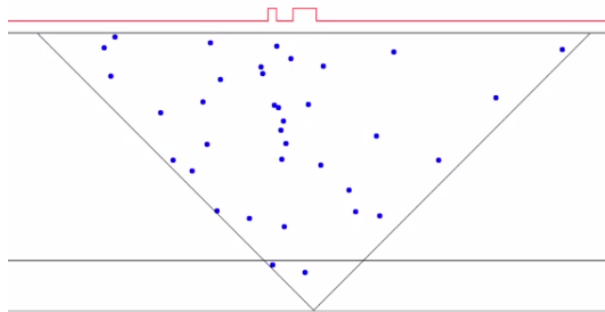
As rays meet, they cancel each other out by disappearing, and slowly we will see a surface grow as time goes on (and our horizontal scan line moves up vertically):
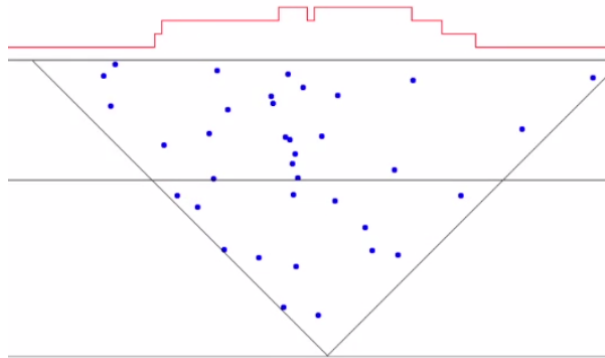
This is called the **polynuclear growth process**, and we can understand that because of the following reason: if we have an oversaturated gas above a table that can condensate, there will be a nucleation seed that appears on the surface of the table. That nucleation seed then grows (by condensation), because things will stick to it, and that is very similar to what is happening in our description above.

And then the last passage percolation times correspond to the heights of the growing interface. But there's another connection we can make as well: once we pick our N random points in the (typically square) domain, we can associate a permutation to them by looking at the relative ordering in the x- and y-coordinates:
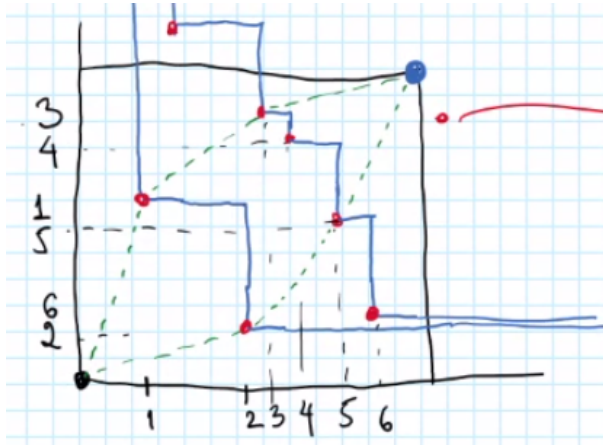
The permutation here is 265143, and if we look at the green optimal paths, we find that we go through points $(2, 5)$ in x-coordinate in one case and $(1, 3)$ in the other. And this is because **optimal paths choose longest increasing subsequences** for the corresponding permutations! In particular, the last-passage percolation time, as well as the height of the polynuclear growth model, both have the same distribution as the longest increasing subsequence for uniformly random permutations, except that the size of the permutation is now Poissonized (because we haven't fixed the total number of particles).

And just like in our earlier lectures, it makes sense to connect this back to partitions (under the Plancherel measure) and try to look for λ_2, λ_3, and so on, now that we already have λ_1 (the length of the longest increasing subsequence) from our picture. It turns out that the answer is yes, and what we need to do is create new second-level nucleation points instead of cancelling out the rays:
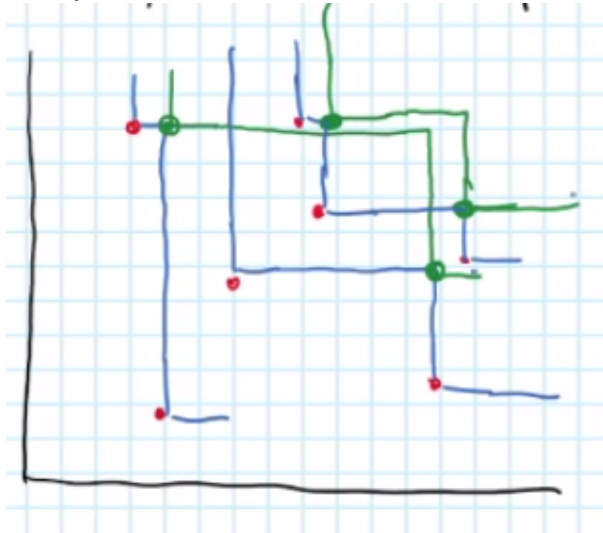
Then the second-level nucleations give us more lines to cross from the root to the target point, and we can keep repeating this process: the kth element λ_k of the partition is then the number of lines formed by the kth level nucleations! (This construction is known as the **matrix ball RSK algorithm**.) And we can implement this on the polynuclear growth picture by adding multiple levels of interfaces, and having the boundaries drop down a nucleation to the next level instead of disappearing:
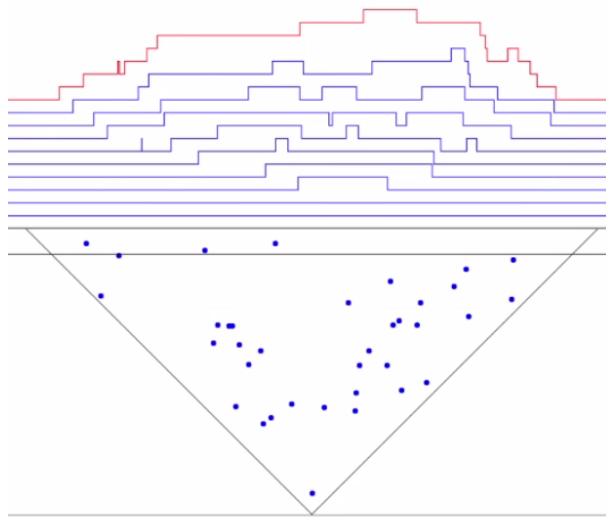
And conveniently, the parts of the partition can be represented by looking at vertical sections of the diagram above. Adding more lines and more points shows us a limit shape, and it turns out this is actually an ellipse:
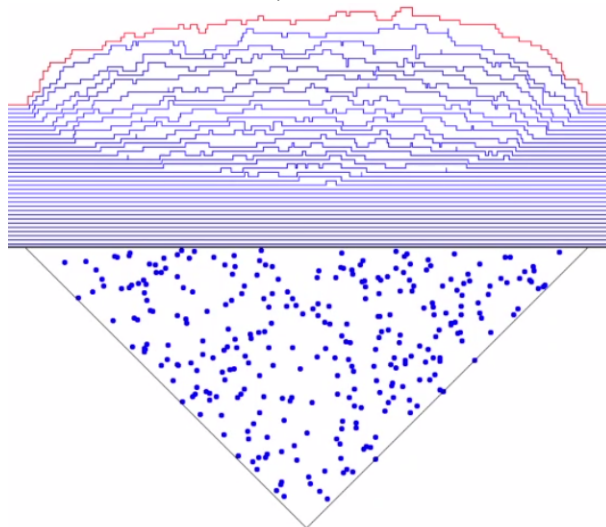
We've now seen two simulations of growth models (TASEP and polynuclear growth), and it makes sense to ask whether there is a larger class of growing interfaces which naturally includes both of these.

> **Fact 61**
>
> For a physics connection, if we spill tea (a solution) on a white surface, the color of the stain will be uniform on the whole area. But if we spill coffee (a suspension), the coffee stain will have a darker rim which is harder to clean, and that's because there are larger particles in coffee that move towards the edge!

The first step in developing such a theory, if we're trying to think like a physicist, is to try to come up with a model for this kind of system. The simplest model is that we might have a bunch of boxes that drop down onto a surface:
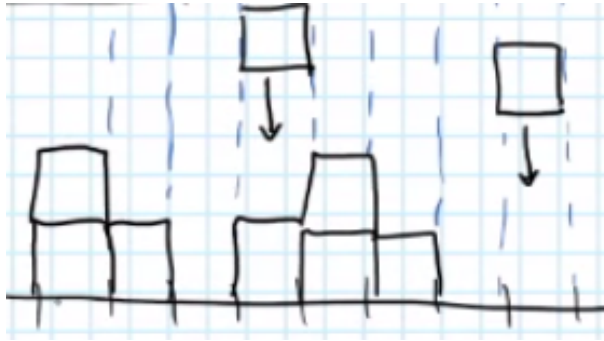
This system is boring, because the different columns do not talk to each other, and thus the description is governed by the one-dimensional Central Limit Theorem, and the fluctuations of the heights of boxes above the interface are of size $\sqrt{t}$. But we can make a modification to this model: we'll have the boxes still fall into discrete columns, but we'll introduce an interaction by saying that the particle will look right and left by one unit and find the lowest possible location it can drop to (by falling down).
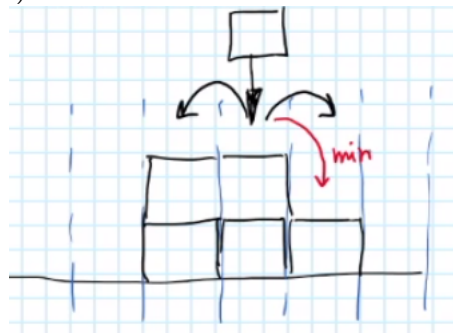
We'll explore next lecture whether this changes the behavior of the system significantly!

8   March 16, 2021

Last time, we discussed some simple interface growth models, and there are many different physical systems that can be modeled by some form of growth. We'll try to think about how physicists may think about these systems today.

Our last example in the previous lecture was a system which has boxes dropping down in various columns, where we add communication between columns by allowing boxes to fill spots with minimal height in the immediately adjacent spots. Here's another potential setup that we could have:

> **Example 62**
> Suppose that boxes fall in various columns, but they are sticky on all sides, so they can settle without landing on the ground, as seen below.
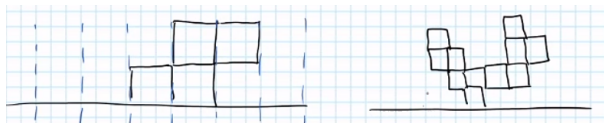
We can then consider the interface (formed by the highest box in each column), and think about whether it looks significantly from the previous examples (boxes falling straight down, or falling in adjacent spots). Here are some pictures taken from a book by Barabasi and Stanley, **Fractal Concepts in Surface Growth**:

Random deposition   Random deposition with   Ballistic deposition
                     relaxation
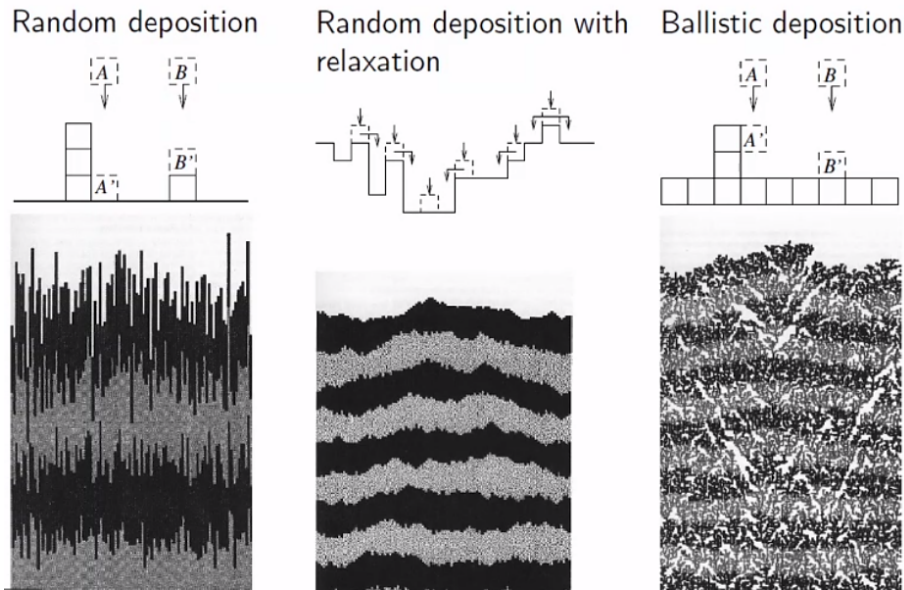
The picture on the left is simulated with no interaction between the columns (dark and light represents progress over time), and we'll see that the fluctuations are of order $t^{1/2}$ by the Central Limit Theorem. In contrast, the picture in the middle (with relaxation) makes the interface much smoother – if we measure the fluctuation size, it turns out that they are of order $t^{1/4}$. Finally, the sticky-boxes model that we've just described gives a roughness that's in between the two cases, and it turns out the fluctuations here are of order $t^{1/3}$.

> **Fact 63**
>
> It's natural to ask why we have $t^{1/n}$ for various natural numbers n here – there are other **critical exponents** that can come out of these pictures, such as the distance along the interface that we need to travel to get independent values. It turns out those exponents are $0, 1/2, 2/3$ respectively, and this in fact has to do with the universal scaling of random walks.

The natural next question is why these models act so different – in other words, if we have some other model that we create, we want our conclusions for the models above to be robust (not dependent on the details of the model). This is the concept of **universality**: we care about understanding what general features of a model make them look similar.

As we've said already, the left model is described by the 1-dimensional CLT, but the middle and right models have their own universality classes: they are the (1+1)-dimensional **Edwards-Wilkinson** and **KPZ (Kardar-Parisi-Zhang)** universality classes, respectively. It thus makes sense to see whether there are particular features that make growth models fit into one of these two classes, and here are some features that cause a model to be in the **KPZ** class:

- Growth must be **local** – distant parts of the interface grow independently of each other. (In other words, boxes falling in distant places are not correlated, so we can't have things like "even/odd boxes only fall at even/odd times" – there is no long-range structure in how the growth occurs.)

- **Relaxation** should occur. This is not as rigorous as the point above: the idea is that a vertical stick falling should have some mechanism that pulls it down. This way, there cannot be "true fractals" at the boundary (like the behavior that occurs when oil is pumped into water, which is often called **fractal fingers**).

  We know that this kind of relaxation does happen in the middle picture, since the particles falling to lower heights automatically smooths out the surface. And in the picture on the right, there is a different kind of relaxation: we can't have deep holes, because the sticky boxes are likely to quickly cover them.

If we go back to the TASEP growth model, we know that the slopes of the interface are always 1 or -1, so there is no way for deep holes to develop. So the relaxation there is built into the way the model is defined!

- Finally, **lateral growth** must occur, which means that the interface grows sideways on sloped surfaces. More concretely, the average **vertical** speed of growth depends on the local **slope**.

For example, TASEP's flat initial condition starts with an interface of slope $1, -1, 1, -1, 1, -1, \cdots$. If we look at a horizontal strip of length L, about half of the spots are available for boxes to fall in, so the average growth is $\frac{L}{2}$. On the other hand, if we start with an initial condition which is biased upward (so for example it starts $1, 1, 1, 1, -1, 1, 1, 1, 1, -1, \cdots$), there are many fewer spots for boxes to fall! So TASEP does indeed exhibit this lateral growth property.

But the middle model in our picture above, random deposition with relaxation, does **not** satisfy this property. After all, if we consider a strip of length L, all boxes that don't fall exactly on the endpoints will stay within that strip, so the rate of deposition does not change based on the local slope.

And the ballistic deposition model on the right does indeed grow sideways: the initial condition for this picture is just a tall stick in the middle, but over time, this has been flattened out. (Basically, some of the particles growing on the tree are contributing to sideways growth!)

Definition 64 (Vague)

The **KPZ universality class** is the set of models that satisfy local growth, relaxation, and lateral growth. The **EW universality class** is the set of models that satisfy only local growth and relaxation.

With this "definition" in mind, we can probably figure out whether these three properties above hold, and this will tell us whether we'll fall into one universality class or another. And it's possible of course that there are models that fall in one of these classes but with different critical exponents, but no one has found a counterexample yet (including Professor Borodin)! And there are only a few concrete models, like TASEP and ASEP, that have been presently studied in enough detail to get proofs of these kinds of results.

What Kardar, Parisi, and Zhang did was build an analytical model with these three properties of locality, relaxation, and local growth. For the simplest model in mathematical physics, we use differential equations to measure the height $h(x, t)$ of an interface. The simplest model then looks something like (since $\partial_t h$ is the vertical velocity)

$$\partial_t h = v \partial_x^2 h.$$

This is the **heat equation**, and it models how heat spreads in a one-dimensional homogeneous medium (heat should diffuse so that the temperature spreads out). But we also need a part responsible for the lateral growth, and we do that by adding a term

$$\partial_t h = \nu \partial_x^2 h - \lambda F(\partial_x h),$$

since $\partial_x h$ is the slope and F is some function of the slope. And we also need to add some randomness (to give us locality of growth), so what we end up with is the equation

$$\partial_t h = \nu \partial_x^2 h - \lambda F(\partial_x h) + \sqrt{D}\eta(x, t),$$

where η is a **two-dimensional white noise** (which assigns a normal random variable at every point in space and time, which are δ-correlated).

But we don't know what the function F is, and the idea is to use its Taylor approximation $F(s) = F(0) + F'(0)s + \frac{1}{2}F''(0)s^2 + \cdots$. Here, a deterministic shift by a constant doesn't change our behavior very much, so adding $F(0)$ doesn't change anything about our system. And the linear term can also be removed by a deterministic change, so the quadratic one is the first relevant term.

Definition 65

The **KPZ equation** is the differential equation

$$\partial_t h = \nu \partial_x^2 h - \lambda(\partial_x h)^2 + \sqrt{D}\eta.$$

This equation is just one model in the KPZ universality class, but what's important is that this will (provably) be consistent with the discrete models in the KPZ universality class! And this model has the advantage (over TASEP or other models) that the spatial coordinate is continuity.

Fact 66

The Edwards-Wilkinson universality class corresponds to the case here where $\lambda = 0$ (because we're removing the lateral growth), and there is a pretty satisfying theory there with Gaussian fluctuations.

However, if we turn on $\lambda > 0$, we get an issue where we are trying to square the derivative of a Brownian motion, and the square of a generalized distribution doesn't always make sense! So figuring out how to make the KPZ equation well-posed in mathematics was a big challenge for a while.

We're now ready to return to the LPP (Last Passage Percolation) and polynuclear growth models from last lecture: recall that we defined the LPP time for a spot in our grid (m, n) to be

$$\max_{\text{directed paths } (1,1)\to(m,n)} \sum_{\text{path}} w_{ij}.$$

Quantities like this, where we maximize some function $H(x)$ over $x \in X$, can often be rewritten as

$$\max_{x \in X} H(x) = \lim_{\beta \to \infty} \frac{1}{\beta} \ln \sum_{x \in X} e^{\beta H(x)}.$$

This is because the terms with the largest H will dominate at large β, and thus there will be a largest term that dominates the sum (if the maximum is only achieved at finitely many points). And the reason for writing an expression like this is that this gives us the **Gibbs ensemble** in statistical physics: $H(x)$ can be interpreted as the energy of a given configuration x of our system, and β is often thought of as the **inverse temperature** (so that taking $\beta \to \infty$

means we're freezing our system at $T = 0$, and thus we are settling the system into one of a few states where the energy is maximized).

So if we apply this type of thinking to the LPP model, we can (change notation a bit and) define the **partition function**

$$Z(T, x) = \sum_{(1,1)=b_1 \nearrow b_2 \nearrow \cdots \nearrow b_{x+T-1}=(T,x)} \prod_{k=1}^{x+T-1} d_k.$$

Basically, for each potential path $(1, 1)$ to (T, x), we calculate its contribution to the total Z by multiplying the values of the d_is (which are positive random variables) along the path. And we'll explain how this object is related to the KPZ equation next time!

9   March 18, 2021

We started discussing the transition from Last Passage Percolation to directed polymers last lecture: the setup is that we have a grid of squares, each of which has a random variable, and we'll use different coordinates to describe this system this time:
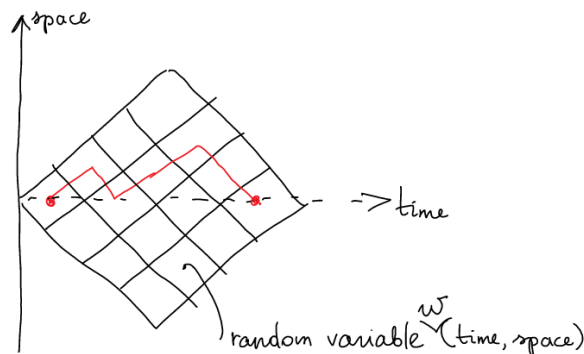
In other words, paths that get us from the starting to ending box are paths that vary in space over time, and we can define a random variable at each point in (time, space). We can then define the **last passage percolation time** as

$$\text{LPP} = \max_{\text{paths } \{b_i\}} \sum_{i=1}^{n} w(i, b(i)),$$

where the conditions on our paths is that $b(1) = b(n) = 0$ and $|b(t-1) - b(t)| = 1$ for all t. This is the "zero-temperature" model for our system − the more general "finite-temperature" model has us instead computing a sum over all paths, rather than just finding the maximum sum: then we have something like

$$Z_\beta(n, 0) = \sum_{\{b(i)\}} e^{\beta \sum_i w(i, b(i))},$$

where $(n, 0)$ is the final point that paths end at. (As mentioned, we obtain the LPP time by taking the limit $\lim_{\beta \to \infty} \frac{\log Z_\beta}{\beta}$, and here β is the inverse temperature in the Gibbs formulation.) The object Z_β is known as a **partition function** for the **directed polymer** in a random medium: "directed" corresponds to the red path, and the random medium is the set of random variables $W(i, b(i))$ we've defined.

It may seem artificial why we're introducing the inverse temperature like this, and here's a way to introduce a probabilistic model that computes it:

Example 67

Consider a system of massive particles on $\mathbb{Z}$, where different sites can have different nonnegative masses. This system evolves in discrete (integer) time as follows: we input random variables $w(t, x) > 0$, and at each time step t, we multiply the mass of each particle at x by $w(t, x)$, and then each particle splits into two twins of the same mass as the original particle and go left and right by one unit.

We can think of the masses as populations of plankton, and w as describing how favorable the conditions are for the plankton at the given time and position. And there is some kind of random walk done to look for food. (This model is in fact actually used, and it is easy to see how to generalize it to more dimensions.) We often start this system with a single particle at mass 1 at the origin, and it's interesting to think about the long-term expected behavior.

Fact 68

For this system, it matters a lot whether there are locations in the lattice where w are maximal (for example, if the weight is equal to 2 at one point and less than 1 everywhere else).

Then mass will slowly die off everywhere where the conditions are unfavorable, but it will also eventually move to the positions where there is favorable growth. So we might expect a spike when the weights w are high, but the question of "where does the mass actually settle at large times" is often difficult to answer (because we don't know whether far-out, large spikes will be largely populated), even if w's are independent of time.

And adding time-dependence in can further complicate the story – we can end up with **intermittent behavior** – there may be tiny islands with huge distributions of mass, and this is unfortunate for probability because calculating the distribution is difficult if **we can't determine it from the moments** (the distribution is dominated by peaks that we almost never see). Physicists like to speculate that the distributions around us are specified by this kind of intermittent behavior (such as the distribution of mass in the universe)!

But the point is that a lot depends on the noise, so we'll make a distinction here: there is a difference between **strong disorder** and **weak disorder**. To explain this, if we have no noise at all, then our mass will behave like a random walk (weak disorder – the picture doesn't change substantially from the random walk). But the Last Passage Percolation story has contributions coming from large ws, and thus the optimizing paths have a much different behavior – for example, in (1+1) dimensions, the paths are very diffusive and move far away from the diagonal much more than the random walk trajectory.

We'll consider the simpler limit where noise is small but not negligible. We might remember that we had a limit of small noise in LPP (the polynuclear growth limit), which was the picture where we had nucleation events in a quadrant and then looked for paths that maximized the number of paths. It turns out that if we want to consider these directed polymers, the correct scaling is to take $\beta \to \frac{\beta}{n^{1/4}}$. This means that we can approximate

$$e^{\beta n^{-1/4} w(t, x)} \approx 1 + \beta n^{-1/4} w(t, x),$$

and let's assume here that $w(t, x)$ have mean 0 and variance 1 for concreteness. Then the average over all random walks is

$$\mathbb{E}_{\text{walks } w} \prod_{i=1}^{n} \left(1 + \beta n^{-1/4} w(i, b(i))\right) \approx 1 + \beta n^{1/4} \mathbb{E}_{\text{walks } w} \sum_{i=1}^{n} w(i, b(i)).$$

Swapping the order of summation turns this into

$$= 1 + \beta n^{-1/4} \sum_{t=1}^{n} \sum_{x \in \mathbb{Z}} w(t, x) \mathbb{P}\left(\text{walk hits } x \text{ at time } t\right).$$

It turns out that $n^{-1/4}$ was chosen specifically because this object has a well-defined limit for $n \to \infty$: the expected value of the double sum is 0 because the ws each have mean 0, and we can understand the variance by noticing that paths typically cover a vertical width of roughly $\sqrt{n}$ (probabilities beyond that become tiny). Then the total number of points covered by random walks with non-negligible probability is of order $n^{3/2}$, and the probabilities are each of order $\frac{1}{\sqrt{n}}$, so the variance calculation gives us a rough estimate of $\frac{1}{(n^{1/4})^2} \cdot \frac{1}{n} \cdot n^{3/2} \sim 1$. And there is in fact a Central Limit Theorem argument here to show that the expression written above is actually a normal random variable!

But the issue with this argument is that we can't really replace the product with the first-order terms – if we keep all orders in the product, we end up with the object

$$1 + \sum_{k-1}^{n} \beta^k n^{-k/4} \sum_{1 \le t_1 < \cdots < t_k \le n} \sum_{\vec{x} \in \mathbb{Z}^k} w(t_1, x_1) \cdots w(t_k, x_k) \mathbb{P}(\text{random walk hits } (t_1, x_1), \cdots, (t_k, x_k)).$$

where k sums over the power of the monomial in w.

Theorem 69 (Alberts, Khamin, Quastel 2011)

The above expression converges as $n \to \infty$ to a similar sum, but with $w(t, x)$ replaced by the **space-time white noise** $\eta(t, x)$ and random walk probabilities replaced with Brownian motion probabilities.

The idea is that replacing a bunch of iid random variables with mean 0 and variance 1 with a white noise is a natural thing to do (this is the representation of independent noise variables in the continuous limit). Of course, we're skipping over the normalization constants and many other details, because it will then take a lot more time to state the result exactly. But we should try reading the paper ourselves if we're curious! The main point is that the noise w hasn't disappeared, and neither has the random walk probability, so neither the "entropy" (noise) term or the "energy" (diffusion of random walks) dominates in this particular limit. So that leads us to what is called the **intermediate disorder regime**: we can in fact write down a continuous version of the partition function,

$$Z = \mathbb{E}_{\text{Brownian bridge } B(t) \text{ from 0 to 1}} e^{\int_0^1 \eta(t, B(t)) dt}.$$

This expression doesn't make sense literally, because the integral of a white noise over a Brownian motion is not well-defined as stated. So we often do a renormalization, and thus we actually need to do a **Wick ordering** to get to some expression $Z = \mathbb{E}_{\text{Brownian bridge } B(t) \text{ from 0 to 1}} : e^{\int_0^1 \eta(t, B(t)) dt} :$, which basically means the sum

$$= 1 + \sum_{k=1}^{\infty} \sum_{0 \le t_1 < \cdots, < t_k} \int_{x \in \mathbb{R}^k} \eta(t, x_1) \cdots \eta(t_k, x_k) \mathbb{P}(\text{Brownian bridge hits the correct points}).$$

And this partition function Z can be computed in yet another way: we can use the partial differential equation

$$\frac{\partial}{\partial t} Z(t, x) = \frac{1}{2} \frac{\partial^2}{\partial x^2} Z(t, x) + \eta(t, x) Z(t, x),$$

which is a linear stochastic partial differential equation (specifically, the **stochastic heat equation** with multiplicative noise). If we think of η as being some deterministic function, then we can solve the resulting linear PDE with standard tools, and we then find an expression for the solution as this type of series described above.

And there is a connection with the KPZ equation here (which did not have this multiplicative factor). In general, the point is that now that we've moved off the lattice and into the continuous limit, we have to be more careful. But if we substitute $Z = e^U$ formally into this equation (since we want to consider $\log Z$ to get back to the LPP time),

ignoring the proof that Z is positive with probability 1, we find that

$$\frac{\partial}{\partial t}U(t,x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}U(t,x) + \frac{1}{2}\left(\frac{\partial}{\partial x}U(t,x)\right)^2 + \eta(t,x),$$

and thus U satisfies the KPZ equation. So this is a way to give meaning to KPZ solutions: we do this exponential substitution to make sense of a differential equation which may not be well-posed as stated, and the results are called **Hopf-Cole** solutions. So we've basically returned to KPZ now, understanding that this equation shows up from both random polymers and growing surface models.

> **Example 70**
>
> ASEP is an example of a particle system which can converge to the KPZ equation: recall that this is a system with particles on $\mathbb{Z}$ jumping left and right at certain jump rates.

More specifically, if ASEP's initial conditions are such that particles initially occupy all negative integers, and the jump rates are $q \in (0,1)$ to the left and 1 to the right.

> **Fact 71**
>
> Note that if we set $q = 1$, we get the **symmetric simple exclusion process** or **SSEP**, and this is a member of the Edwards-Wilkinson universality class, because there is no growth-dependence on the slope.

We'll scale ASEP as follows: take $\varepsilon > 0$ to be a small parameter, and let $q = 1 - \varepsilon$, $t = \varepsilon^{-4}\hat{t}$, $x = \varepsilon^{-3}\hat{x}$ (so we rescale our space and time accordingly), so that $\frac{\hat{x}}{\hat{t}} \in (0,1)$. Then it turns out that we can consider the limit

$$\lim_{\varepsilon \to 0}\left(\varepsilon^{-2}\hat{\sigma} - \log\varepsilon - \varepsilon\text{height}_{\text{ASEP}}(x,t)\right),$$

where we should remember that the height of ASEP is the number of particles to the right of location x at time t, and $\hat{\sigma} = \frac{(\hat{t}-\hat{x})^2}{4\hat{t}}$. This limit then turns out to be a function

$$= \frac{T}{24} - (\text{solution to KPZ at } (t,x) = (T,0)),$$

where $T = 2\hat{t}^3$. This is a pretty advanced statement – it is hard to see where the scalings come from, especially because the exponent scaling is more complicated than the Brownian scaling exponents of 1 and 2, and there are lots of other strange terms popping up! So this is not an easy statement to prove, and it would take a few lectures to prove (even with advances from the past 10 years). But the first work done to show this convergence was by Bertini and Giacomin (1997), and the convergence also works for a class of initial conditions (not just particles at all negative integers).

Remark 72. *Taking $q = 1 - \varepsilon$ and the limit $\varepsilon \to 0$ is known as the* **weak asymmetry limit** *(since the parameters are getting pretty close to the symmetric case). It turns out that for a particle system to have any asymptotic behavior that is described by the KPZ equation, we need some sort of a parameter in the model that scales in a special way. (So if q does not go to 1 as time and space are scaled to ∞, then ASEP will not converge to the KPZ equation: there will still be a universal limit described by random matrix theory, though.) And this is similar to the limit of* **weak noise** *$\beta \to \frac{\beta}{n^{1/4}}$ that we used for the directed polymer!*

The KPZ equation itself thus "knows something" about how to scale systems, and we'll discuss this (how to deal with the noise or asymmetry) in a future lecture. There are also quite a few papers that prove convergence to KPZ in a variety of setups, but they often require integrability structure, and what people can do is not far from the integrable world today (even if we should hope that there are generic arguments that work for convergence)!

10 March 25, 2021

Today's class is a seminar by **Harriet Walsh**, titled **Schur measures, unitary matrix models, and multicriticality**. This talk will basically give a way of going outside the KPZ universality class, and we'll do that by looking at models of random integer partitions.

Recall that integer partitions are weakly decreasing sequences of positive integers ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{\ell(\lambda)} > 0$), and these partitions index both the conjugacy classes of the symmetric group (by cycle type) and the bases for the symmetric functions Λ_n. There are various choices: for example, we have the powersum symmetric functions

$$p_{4,2,1}(x) = \sum_{i,j,k} x_i^4 x_j^2 x_k,$$

the elementary symmetric functions

$$e_{4,2,1}(x) = \left(\sum_{i<j<k<\ell} x_i x_j x_k x_\ell \right) \left(\sum_{i<j} x_i x_j \right) \left(\sum_i x_i \right),$$

and the Schur symmetric functions

$$s_{4,2,1}(x) = \sum_{i \neq j \neq k} x_i^4 x_j^2 x_k + x_i^3 x_j^3 x_k + 2x_i^3 x_j^2 x_k^2.$$

This last **Schur basis** is the most subtle: we define a sum over semistandard Young tableaux

$$s_\lambda = \sum_{\text{SSYT}} x^T,$$

where x^T means that we take $\prod_i x_i^{\text{number of } i\text{s}}$. These integer partitions can then map to **fermion configurations**, which are sequences of strictly decreasing half-integers: we map

$$\lambda \mapsto \left\{ \lambda_1 - \frac{1}{2}, \lambda_2 - \frac{3}{2}, \lambda_3 - \frac{5}{2}, \cdots \right\},$$

and these configurations then tell us about the locations of particles in a one-dimensional lattice. And it turns out that the **Schur measures** $\mathbb{P}(\lambda) \propto s_\lambda(x)s_\lambda(y)$ correspond to certain free models, where the probability of a given configuration $\mathbb{P}(\{u_1, u_2, \cdots, u_n\} \subset S(\lambda))$ is given by the determinant of a certain kernel.

We'll look at Schur measures from the fermion perspective from now on: we can define a **determinantal point process** by looking at an infinite wedge space, which has infinitely many sites and an antisymmetric exchange relation.

Definition 73

We can then define the **creation** ψ_k and **annihilation** ψ_k^* operators on this wedge space, which create or remove a particle at site k if possible and kill the state completely otherwise.

We then get the **indicator operator** $\psi_k \psi_k^*$ for whether there is a particle at spot k, and we can also define the jump operator

$$a_r = \sum_{k \in \mathbb{Z} + \frac{1}{2}} \psi_{k-r} \psi_k^*.$$

(We can then think of these jump operators as **bosonic operators**.) We also have the anticommutation relation to

finish the construction of the wedge space:

$$\psi_k\psi_\ell + \psi_\ell\psi_k = \psi_k^*\psi_\ell^* + \psi_\ell^*\psi_k^* = 0, \quad \psi_k\psi_\ell^* + \psi_\ell^*\psi_k = \delta_{k\ell}.$$

These bosonic operators a_r will now allow us to evolve the system: if we start with a system with all particles to the left of the origin, then we can build our Young diagram and add r boxes in a ribbon by using the operator a_{-r}. So in the picture below, we just applied two bosonic operators a_{-1} and a_{-6} to the ground state:
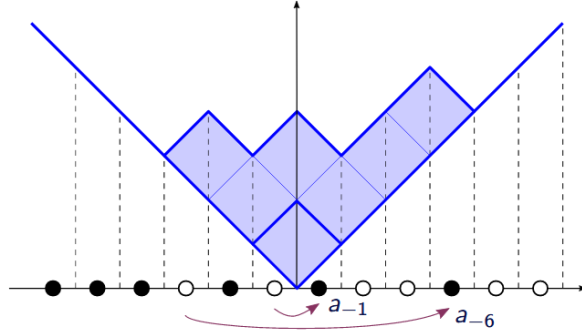
We can then use a set of parameters $\{t_r\}$ and $\{t_r'\}$ to find the correlation functions

$$\rho(X) = \mathbb{P}(X \subset S) = \frac{1}{Z} \left\langle \phi \left| e^{\sum_r t_r' a_r} \prod_{k \in X} \psi_k \psi_k^* e^{\sum_r t_r a_r} \right| \phi \right\rangle,$$

and we can use Wick's theorem to find the commutation relation (setting the two exponentials to be Γ_+ and Γ_-, respectively)

$$\Gamma_+(t')\Gamma_-(t) = e^{\sum_r \frac{t_r t_r'}{r}}\Gamma_-(t)\Gamma_+(t'),$$

which gives us the formula

$$\rho(X) = \det_{k,\ell \in X} K(k,\ell),$$

where the kernel is

$$K(k,\ell) = \left\langle \phi \left| \Gamma_+(t')\Gamma_-(t)^{-1}\psi_k\psi_\ell^*\Gamma_+(t')^{-1}\Gamma_-(t) \right| \phi \right\rangle.$$

(This is basically a correlator for finding a particle at ℓ, and killing it and moving it to k.) We can also then find the normalization factor to be $Z = e^{\sum_r \frac{1}{r}t_r t_r'}$.

And now if we have a given partition λ, we can find a determinantal expression for the probability

$$\mathbb{P}(\lambda) = \oint \frac{dz}{2\pi i} \det_{1 \leq i,j \leq \ell(\lambda)} e^{\sum_r \frac{t_r}{r}z^r},$$

and this is a generating function if we think of the t_r as the powersum polynomials $p_r(x)$. And then we get the Schur measures via

$$\mathbb{P}(\lambda) = e^{-\sum_r \frac{1}{r}t_r t_r'}s_\lambda[t']s_\lambda[t],$$

and this defines a determinantal point process (measure) if $\mathbb{P}(\lambda) \geq 0$ for all λ and we have a finite normalization $e^{\sum_r \frac{1}{r}t_r t_r'}$. So for a given **specialization** $\{t_r\}$ (plugging in values), we want $s_\lambda[t']s_\lambda[t] \geq 0$, and we can do this by either setting $s_\lambda[t'] = 0$ for all partitions λ or by just setting $t' = t$ (which is more physically motivated because it means we have a Hermitian system where the creation and annihilation behaviors are related to each other).

The canonical case for this, which existed before the Schur measures came up, is the **Plancherel measure** where $t_1 = t_1' = \theta$ and all other $t_{>1}$ and $t_{>1}'= 0$. Then we only have the bosonic operators a_1 and a_{-1}, and thus we can only add one box at a time in the process above. And then the number of ways to form a given diagram λ is the number of standard Young tableaux of shape λ, which gives us the familiar d_λ^2 term in the Plancherel measure.

We can then define the **Poissonized Plancherel measure**

$$\mathbb{P}_\theta(\lambda) = d_\lambda^2 \frac{\theta^{2|\lambda|}}{|\lambda|!} e^{-\theta^2},$$

where $|\lambda|$, the size of the partition, is a Poisson random variable with mean θ^2.

It turns out that we get a particular limit shape in the limit where $\theta \to \infty$, called the Vershik-Kerov-Logan-Shepp limit shape, and the fluctuations along the edge of the shape satisfy

$$\mathbb{P}\left[\frac{\lambda_1 - 2\theta}{\theta^{1/3}} > s\right] = F_2^{\text{TW}}(s)$$

for the Tracy-Widom distribution. These fluctuations are universal in the KPZ class, and they were first discovered for the largest eigenvalue for a GUE random Hermitian matrix. And we can see from the GUE ensemble where the fermionic connection comes in: since we get the measure $e^{-\frac{1}{2}\text{tr}(M^2)}$ for a matrix M, using the unitary symmetry gives us a Jacobian which is a Vandermonde determinant, and that allows us to write the partition function as

$$Z = \int \det p_j(\xi_i) \prod d\xi_i e^{\frac{1}{2}\xi_i^2},$$

where the ξ_is are the eigenvalues of the GUE matrix. And we can pick out the polynomials here so that they are orthogonal with respect to the measure, so that $\langle p_j | p_i \rangle = \delta_{ij}$, at which point we can diagonalize our determinant and find the **Hermite polynomials** (from quantum mechanics).

Talking more about the edge fluctuations described above, we have an asymptotic **Fredholm determinant**

$$F_{\text{TW}}(s) = \det(1 - \mathcal{A})_{[s,\infty)} = 1 + \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \int_s^\infty dx_1 \int_s^\infty dx_2 \cdots \int_s^\infty dx_n \det_{1 \leq i,j \leq n} \mathcal{A}(x_i, x_j)$$

for the **Airy kernel** $\mathcal{A}(x, y) = \int_0^\infty d\mu \text{Ai}(x + \mu)\text{Ai}(y + \mu)$.

But now we'll introduce some new results: we'll generalize away from Plancherel measures and introduce the **multicritical measures**

$$\mathbb{P}_{n,\theta}^\gamma(\lambda) = e^{-\theta^2 \sum_r \frac{\gamma_r^2}{r}} s_\lambda[\theta\gamma_1, \theta\gamma_2, \cdots]^2,$$

where the $\{\gamma_r\}$ must satisfy certain equations

$$2 \sum_r r^{2p}\gamma_r = \delta_{p,0} b + \delta_{p,n}(-1)^{n+1}(2n)! d.$$

(We can notice that this is a Hermitian measure to get something physically sensible, much like setting $t' = t$ above. And it turns out that we can't have Schur positivity so that the s_λs are nonnegative, unless we try to add something nonphysical!) Once the specialization satisfies those conditions, we get a generalization of the Plancherel measure in the following way:

If λ is a random partition distributed as $\mathbb{P}_{n,\theta}^\gamma$, then the distribution of the first part of the partition follows

$$\lim_{\theta\to\infty} \mathbb{P}_{n,\theta}^\gamma\left[\frac{\lambda_1 - b\theta}{(d\theta)^{1/(2n+1)}} < s\right] = F(2n + 1; s) = \det(1 - \mathcal{A}_{2n+1})_{[s,\infty)}.$$

Here, we have the **generalized** Airy kernel

$$\mathcal{A}_{2n+1}(x, y) = \int_0^\mu \mathrm{Ai}_{2n+1}(x + \mu) \mathrm{Ai}_{2n+1}(y + \mu),$$

where the generalized Airy function is

$$\mathrm{Ai}_{2n+1}(x) = \int_{i\mathbb{R}+\delta} \frac{d\zeta}{2\pi i} e^{-\frac{(-1)^{n+1}}{2n+1}\zeta^{2n+1} - x\zeta}.$$

For some physical intuition, these functions come from getting fermions in "flat traps" of potential $V(x) \sim x^{2n}$

$$\left[p + (-1)^n \frac{d^{2n}}{dp^{2n}} \right] \mathrm{Ai}_{2n+1}(p) = 0.$$

To understand more about how to work with the left-hand side, we can consider the classical involution where we reflect a Young diagram around the diagonal. This maps the polynomial bases via

$$s_\lambda = \det_{1 \le i,j \le \ell(\lambda)} h_{\lambda_i - i + j} \mapsto s_{\lambda'} = \det_{1 \le i,j \le \ell(\lambda)} e_{\lambda_i - i + j}, \quad p_i \mapsto (-1)^{i-1} p_i.$$

This gives us the following lemma:

Lemma 76

The conjugate partition λ' is distributed with the same measure as λ, if we make a change $\gamma_r \mapsto (-1)^{r-1}\gamma_r$ to the specialization.

There are two particular families of multicritical Hermitian Schur measures to consider: the **odd-even** multicritical measures $\mathbb{P}_{n,\theta}^{oe}(\lambda)$ have

$$\gamma_r = (-1)^r \binom{2n}{n-r} / \binom{2n}{n-1}$$

for all $r = 1, 2, \cdots, n$ and $\gamma_r = 0$ otherwise, while setting $b = \frac{n+1}{n}$ and $d = \binom{2n}{2n+1}$, and the **odd** multicritical measures $\mathbb{P}_{n,\theta}^{o}(\lambda)$ having

$$\gamma_{2r-1} = (-1)^r \binom{2n-1}{n-r} / \binom{2n-1}{n-1},$$

all other γ_rs being zero, while setting $b = 2^{4n-1} n^{-1} \binom{2n}{n}^{-2}$ and $d = \binom{(2n-1)!!}{(2n-2)!!}$. (Basically, these are the simplest possible cases that give the multicritical behavior.) Notice that when $n = 1$, both of these recover the Plancherel measures, and when $n > 1$, these are not Schur-positive measures. And we can see the limiting fermion **density profile** for $\theta \to \infty$ in both cases as well: setting $u = \frac{k}{\theta}$ and looking within a fixed interval $[-\tilde{b}, b]$, we can compute the one-point function

$$\rho^{oe}(u) = \frac{1}{\pi} \arccos \left(1 - \frac{1}{2} \binom{2n}{n-1}^{1/n} (b - u)^{1/n} \right),$$

and a less closed expression for the odd case:

$$\rho^{o}(u) = \frac{1}{\pi} \chi(u), \quad \int_0^{\chi(u)} d\phi (2\sin\phi)^{2n-1} = (-1)^{n+1} \binom{2n-1}{n} u.$$

We can connect this back to matrix models now:

53

Proposition 77

The length of a multicritical partition satisfies the identity

$$e^{\sum_i \theta_i^2/i}\mathbb{P}(\ell(\lambda) \le \ell) = \mathbb{E}_{U \in U_\ell}\left[e^{\theta \operatorname{tr}\sum_r \frac{\gamma_r}{r}(U^r + U^{*r})}\right] = \det_{1 \le i,j \le \ell}[f_{j-i}]$$

where $\mathbb{E}_{U \in U_\ell}$ is the expectation with respect to the Haar measure on the unitary group, and we have a **Toeplitz determinant** here with the **symbol** $\sum_{k \in \mathbb{Z}} f_k z^k = e^{\theta \sum_r \frac{\gamma_r}{r}(z^r + z^{-r})}$.

To explain the second equality here, we can evaluate the Toeplitz detrerminant using **Heine's identity**, which gives us an integral through Fourier decomposition

$$\det_{1 \le i,j \le \ell}[f_{j-i}] = \int d\theta_1 \int d\theta_2 \cdots \int d\theta_\ell \prod_{i<j} |e^{\theta_i} - e^{\theta_j}|^2 f(\theta_1) \cdots f(\theta_\ell).$$

which we can read as an integral over unitary matrices (taking the θ_is as eigenvalues). And a similar expression can be found for $\mathbb{P}(\lambda_1 \le \ell)$, using the classical involution lemma – we expect that as $\theta \to \infty$, we get the phase transition

$$\mathbb{P}(\lambda_1 \le \ell) = 1\{\ell < b\theta\},$$

and the point of the result above is that we can understand more closely the transition between the occupied and unoccupied states (around b).

We'll conclude with a sketch of how to look at the asymptotics for this problem, and where the continuous kernels come from. Heuristically, the reason for the steeper-than-harmonic-traps coming up is that the multicritical Schur measure $\mathbb{P}_{n,\theta}^\gamma(\lambda)$ corresponds to a free fermion with the Hamiltonian

$$H = -\sum_{r>0} \theta(\gamma_r a_r + \gamma_r a_{-r}) + a_{-1}a_1,$$

(where we should think of the kernel $K(k,\ell) = \langle \Omega | \psi_k \psi_\ell^* | \Omega \rangle$ as acting on the ground state of the Hamiltonian Ω), which gives us eigenfunctions

$$J_n = \frac{1}{2\pi i} \oint \frac{dz}{z^{n+1}} e^{\theta \sum_r \frac{\gamma_r}{r}(z^r - z^{-r})}$$

which can be thought of as generalized Bessel functions. And more rigorously, the way we can think about the kernel is to write

$$K(k,\ell) = \frac{1}{(2\pi i)^2} = \oint \oint \frac{dz}{z^{k+1/2}} \frac{dw}{w^{-\ell+1/2}} \frac{e^{\theta \sum_r \frac{\gamma_r}{r}(z^r - z^{-r})}}{e^{\theta \sum_r \frac{\gamma_r}{r}(w^r - w^{-r})}},$$

for an integral along $|w| = 1 - \delta$ and $|z| = 1 + \delta$, looking at the saddle points of the contour integral (because they dominate the contributions to the integral as $\theta \to \infty$) and how they coalesce for various values of n. The central point of multicriticality is that there is more coalescence!

We can also get the limiting density profile by considering $\rho(u) = \lim_{\theta \to \infty} K(u\theta, u\theta)$ – when $u > b$ and $u < -\tilde{b}$, those saddle points are not on the unit circle, so the integral converges to 0 and 1 respectively, and in between we have $\rho(u) = \frac{\chi(u)}{\pi}$ as above.

If we look around $u \sim b$ (near the edge of the limit shape), we then need to capture the edge fluctuations by considering

$$K(k_i(\theta), k_j(\theta)), \quad k_i(\theta) = b\theta + (d\theta)^{1/(2n+1)}.$$

If we then expand out the actions and compute the integral in the large θ limit, using the appropriately scaled contours

$$|z| = e^{\zeta(\theta/d)^{-1/(2n+1)}}, \quad |w| = e^{\omega(\theta/d)^{-1/(2n+1)}},$$

54

then in the limit of large θ, we require ζ and ω to go along contours parallel to $i\mathbb{R}$, and the evaluation of the integral converges to the generalized Airy kernel that we described above.

11 March 30, 2021

Last lecture, we discussed convergence to the KPZ equation and stochastic heat equation, which are stochastic differential equations that come up when we scale certain surface growth models. In particular, we discussed how to scale the random polymer partition function and ASEP height function, and we'll start today by discussing how to discover these scaling limits.

> **Fact 78**
>
> If we're looking at a set or semigroup of transformations, and we expect that these transformations lead to a limit, then the limit is invariant under the action of that set or semigroup.

For example, if we have a sequence of iid Bernoulli random variables ξ_i, which we can represent by a sequence of sites on the positive x-axis that are filled or unfilled, we can consider the "recentering" of the sequence by shifting the whole sequence by M units to the left and taking $M \to \infty$. The limit is then a sequence of iid Bernoulli random variables on the whole x-axis, rather than the positive x-axis, and the limiting shape must be **translation-invariant**. And as another example, if we want to evaluate the continued fraction $\frac{1}{1+\frac{1}{1+\frac{1}{1+\frac{1}{\ddots}}}}$, we can call that expression x and notice that we must have $x = \frac{1}{1+x}$. (Here, the transformation is the function $f(x) = \frac{1}{1+x}$.)

So if we apply this logic to our interfaces, and we consider a $(1 + 1)$-dimensional interface (like cubes falling), we have an x-coordinate (representing the space of potential spots to fall) and an h-coordinate (representing the height of the boxes), plus a time-coordinate (for time-evolution). So we'll need to **scale all three coordinates** if we want to observe a limit: specifically, t, x, y must all grow larger so that we get a specific limiting shape.

> **Example 79**
>
> Recall that if we have a (simple) random walk trajectory, and we scale space by M and height by $\sqrt{M}$, then the limit gives us Brownian motion trajectories (which are invariant under scaling of the two coordinates by this M-$\sqrt{M}$ factor). So if we want a specific interface to converge to Brownian motion or something absolutely continuous with respect to it, that can tell us a lot about how we should be scaling our interface.

If we now return to the KPZ equation

$$\frac{\partial U(x, t)}{\partial t} = \nu \frac{\partial^2}{\partial x^2} U(x, t) + \lambda \left(\frac{\partial}{\partial x} U(x, t) \right)^2 + D\eta(x, t),$$

and we try scaling via

$$U^\varepsilon(x, t) = \varepsilon^b U(\varepsilon^{-1} x, \varepsilon^{-z} t),$$

then we have a fully general power-like scaling (of course, there might be logarithms or other factors, but we'll work with this for now). ε is then the parameter with which we scale x, and then time t is scaled by some power of ε and so is the height U. If we plug this into the KPZ equation, we find that U^ε will satisfy

$$\frac{\partial U^\varepsilon(x, t)}{\partial t} = \varepsilon^{2-z} \nu \frac{\partial^2}{\partial x^2} U^\varepsilon(x, t) + \varepsilon^{2-z-b} \lambda \left(\frac{\partial}{\partial x} U^\varepsilon(x, t) \right)^2 + \varepsilon^{b - \frac{z}{2} + \frac{1}{2}} D\eta(x, t),$$

55

where we've divided through by ε^{b+z} to get the left-hand side to match the original equation, and where we're using the fact that the white noise scales as $\eta(x, t) = \varepsilon^{-\frac{z+1}{2}}\eta(\varepsilon^{-1}x, \varepsilon^{-z}t)$. So the point is that the limiting object must be invariant with respect to the scaling, but notice that there is no way to make the equation look like the original KPZ equation without any blue ε factors by just picking appropriate powers (because we must have $z = 2, b = 0$, and that doesn't make the last term work out). But the red coefficients ν, λ, D (which come from relaxation, lateral growth, and noise) **can also be scaled as part of our limiting process**, and that will make the convergence possible!

Example 80

When we dealt with the random polymer convergence last lecture, we did not need to normalize the partition function Z, so the scaling of height was $b = 0$. And we had to scale space and time in a **Brownian** manner: we had $z = 2$ for the random polymer.

So substituting these back in gives us an extra $\varepsilon^{-1/2}$ on the noise term, and thus we actually need to have $D_\varepsilon = \varepsilon^{1/2}D$. In other words, if we hope to observe the KPZ equation in the limit, our noise needs to be reduced, and indeed we did this by rescaling the noise term $w(t, x)$ with a factor of $n^{-1/4}$.

Example 81

Similarly, when we dealt with ASEP's convergence to the KPZ equation, we scaled with $b = \frac{1}{2}$, $z = 2$. (We only showed convergence of ASEP to KPZ at a single point, so we didn't talk about the details of the limiting shape.)

This time, the first and third term on the right-hand side are good, but we need to adjust the λ term by taking $\lambda_\varepsilon = \varepsilon^{1/2}$. So because the lateral growth comes from the differences in jump rates to the left and right, taking $\lambda_\varepsilon \to 0$ is reflecting the **weak asymmetry** assumption that we talked about last time.

There is also a third system where the details aren't worked out yet:

Definition 82

The q-**TASEP** is a generalization of the ASEP, in which particles only jump to the right by 1 with rate $1 - q^{\text{gap}}$, where the gap is the number of empty spaces in front of the particle.

If we have $q \to 1$ in q-TASEP, then the asymmetry goes away, and it turns out this system also converges to the KPZ equation if we scale the parameters correctly.

But in all three examples above, we need to tune some parameter to make convergence work out (either ν, λ, or D). So for both TASEP and for the "sticky boxes" example in our interface growth a few lectures ago, **because there are no tunable parameters**, there is no way to get convergence to the KPZ equation. Instead, the large-time limit for systems when we don't scale any parameters is conjectured to be an object called the **KPZ fixed point** (a specific Markov process), and this convergence happens when we scale as $z = \frac{3}{2}, b = \frac{1}{2}$. So the KPZ fixed point is invariant if we scale x, t, h with those appropriate exponents, and this is often called the **1:2:3** or **KPZ scaling** because $(x, t, h) \mapsto (\varepsilon^{-1}x, \varepsilon^{-3/2}t, \varepsilon^{-1/2}h) = (A^2x, A^3t, Ah)$ if $A = \varepsilon^{-1/2}$.

Remark 83. *With similar notation, we have* **1:2:4 scaling** *for the Edwards-Wilkinson class, and in both of these two situations the "1:2" part is the Brownian scaling. The idea is that if we look at the ASEP interface at large time, the result looks like Brownian motion in both cases, but the actual fluctuations of the height are different!*

There is very little actually proved about the KPZ fixed point, though – there are no generic classes of processes that are known to converge, and the only results that are currently proven have lots of algebraic structure. And this

is sort of like how we know that the simple random walk or Bernoulli sequences converge to the normal distribution through the Central Limit Theorem through combinatorial estimates – trying to generalize that statement to more general random walks requires a different level of analysis.

For now, we'll move to the next step on our roadmap and talk about **random growth models in 2+1 dimensions**. So far, all of our growth models have been in (1+1)-dimension (with an interface given by a height function $h(x, t)$), and a (2+1)-dimensional interface would then be a growing surface. So it's tempting for us to start by trying to take pictures that look like lattice-like structures in 3 dimensions – for example, we can imagine adding cubes to plane partitions, independently at each possible point, and this would keep all of the requirements of the KPZ class but increase the dimension by 1.

Fact 84

But there's nothing known about this (2+1)-dimensional object at large times – all we can do right now is simulate this system with some set of initial conditions with a computer! Physicists are usually the ones who guess and find formulas for this kind of model, but so far there has been no success.

We know that the fluctuations have a power-like time-dependence in the $(1 + 1)$-dimensional case ($t^{1/4}$ or $t^{1/3}$ in the EW and KPZ classes), but the exponent looks like it has to be something like 0.24, with error bars indicating that it is not $\frac{1}{4}$. And this 0.24 number appears in different models – it appears there is universality – but we do not know anything about where it comes from.

However, there **are** $(2+1)$-dimensional growth models which people are able to handle, and the reasons come from representation theory.

Example 85

In one such system, instead of taking all possible cubes and adding them to a plane partition, we can imagine placing $1 \times 1 \times n$ sticks (with the long end always in one fixed direction) in our current interface, so that we never have overhangs and so that we're constrained to a triangular region of the plane.

Below is an example of a shape that may appear:
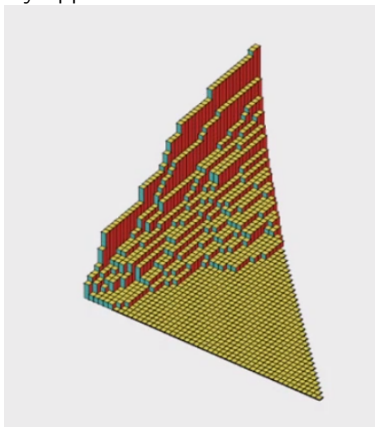
We'll draw this picture in another way now – instead of thinking about adding to the height at each point in our region, we can imagine pushing our particles to the right. (The way that we usually describe this is that every particle has an exponential clock, and when it rings, it moves if it can and resets its clock otherwise.)
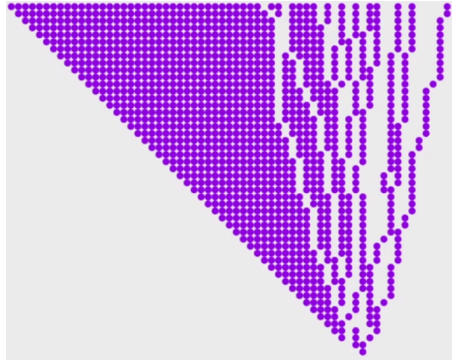
We then have the rule of **interlacing** – if a particle jumps to the right, then it pushes particles above it to the right if it would otherwise violate interlacing. And the particles at the bottom have "maximum" priority here: the particle $(1,1)$ at the very bottom does a random walk, the particle $(2,1)$ in the second row on the right does a random walk except that it may be pushed by $(1,1)$, and the particle $(2,2)$ on the second row on the left does a random walk except that it may be blocked by $(1,1)$.

Fact 86

Experimentally, we can indeed find the fluctuations around the average surface for this model, and the computer will find that the fluctuations are smaller than any power of t. And in fact, it's a theorem that the fluctuations are of order $\sqrt{\log t}$.

The reason that there is such a big difference is that if we try to write down a $(2+1)$-dimensional KPZ equation (which is **meaningless** in the sense that there is no way to regularize the solutions), it should look something like

$$\frac{\partial h}{\partial t}(x, y, t) = \nu \Delta_{x,y} h + \lambda Q \left(\frac{\partial h}{\partial x}, \frac{\partial h}{\partial y} \right) + D\eta(x, y, t).$$

Basically, the nonlinearity term $\left(\frac{\partial}{\partial x} U(t, x) \right)^2$ came from the Taylor decomposition of the slope-dependent growth $F(s)$, and now Q is a **quadratic form** in two variables, and this quadratic form has a **signature** (how many $+$s and $-$s we have after simplifying by scaling). The conjecture is then that there are two different universality classes: $(+, +), (-, -)$ lead to the **isotropic KPZ universality class** (which contains the $1 \times 1 \times 1$ boxes and is where the 0.24 exponent comes from), and $(+, -)$ leads to the **anisotropic KPZ universality class** (which contains the $1 \times 1 \times n$ sticks).

So next time, we'll look at the leftmost particles in our pushing-particle model above, which will turn out to evolve as **TASEP**. It turns out that this falls into the KPZ class – pictorally, the heights of the growing boxes corresponding to that row of leftmost particles (which is the row closest to us in the 3-dimensional picture above) will follow a parabola limit shape, with $t^{1/3}$ fluctuations.
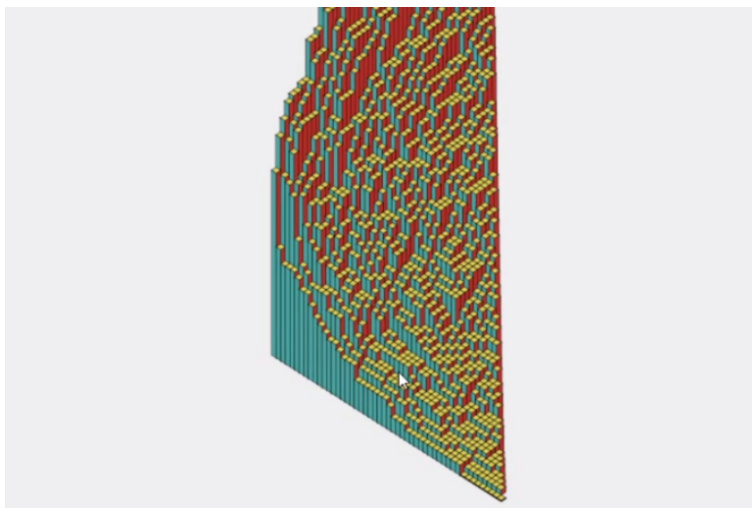
12   April 1, 2021

Last time, we started talking about dynamics which could be viewed as random interface growth models in (2+1)-dimensions: the model that we described at the end of last lecture is a way to unify a lot of the topics we've been describing in this class so far, and we'll try to outline why that is the case.

First, let's define the system more formally: consider a Markov chain on a 2D array of particles, indexed by $\{x_m^n : 1 \leq m \leq n, 1 \leq n\}$. (Particles cannot occupy the same location.) These particles sit within $\mathbb{Z}^2$, and we can draw them in a skewed way so that the y-axis is offset 120 degrees from the x-axis.

We'll focus on the single initial condition where $x_m^n = m - 1$, which corresponds to the following picture:
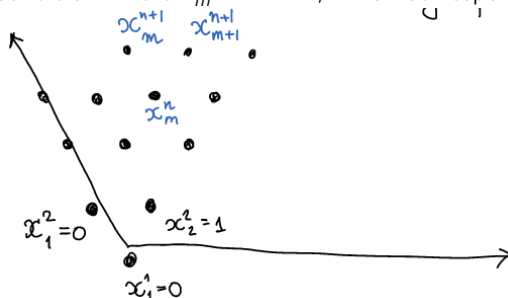
The rule for evolution is that the particles need to interlace: because n indexes the row and m indexes the horizontal position within each row, we need

$$x_m^{n+1} \leq x_m^n < x_{m+1}^{n+1},$$

and then the arrays of particles will be in one-to-one correspondence with Gelfand-Tsetlin patterns by setting

$$\lambda_m^n = x_m^n - (m - 1),$$

so that at the start all λ_m^n are zero and we have the weak inequalities $\lambda_m^{n+1} \leq \lambda_m^n \leq \lambda_{m+1}^{n+1}$.

The dynamics of the Markov processes are then defined as follows: every x_m^n has an independent exponential clock of rate 1. When a particle's clock rings, a particle tries to jump to the right by 1. If we violate the interlacing condition with the lower level, meaning that $x_m^n + 1 > x_{m-1}^{n-1}$, then the jump is suppressed. Otherwise, the jump is allowed, and then if the jump violates the other interlacing condition $x_m^n + 1 = x_{m_1}^{n+1}$, then we also have the higher-level particle x_{m+1}^{n+1} jump to the right at the same time, and we check the next interlacing condition as well (so that a whole row of

particles could jump at the same time).

We outlined last time that this system has a connection to TASEP if we look at the left boundary $\{x_1^n\}$. Since those particles all start at position 0, we look at particle locations $\{x_1^n - n\}$ (so that the TASEP initial condition has particles at $-1, -2, -3, \cdots$). And because there is no pushing here that affects the $m = 1$ particles, we just need to make sure that particles do not occupy the same location, and that's exactly the TASEP evolution condition.

We'll now make a second connection to a system known as **PushTASEP** – we look at the particles $\{x_n^n\}$, which are initially at locations $0, 1, 2, 3, \cdots$. This is basically a "rude version of TASEP:" each particle tries to jump to the right by 1, with exponential waiting time 1, but if the spots are blocked, then we "push" the tail of occupied spots each 1 spot to the right. We can alternatively describe this system as **long range TASEP**, in which a particle jumps to the first unoccupied location when it wants to move (this also gives us the same occupation states, because pushing the whole tail and jumping over it are equivalent when the particles are indistinguishable).

We might want to ask whether these systems are in the KPZ universality class, and we can figure this out now. Much like TASEP, we can imagine adding boxes to a $(1 + 1)$-dimensional interface, and this time we can notice that we add $1 \times n$ rectangles of boxes in one direction (each with intensity 1).

Remembering that the conditions for KPZ universality are locality, slope-dependent growth, and relaxation, it seems that we may have issues where there are long strings of particles that can move at the same time. But if we have an initial condition such as Bernoulli occupation, then the probability of long strings will decay exponentially at the beginning. And because those Bernoulli distributions are invariant for this evolution, we'll see that the long strings will indeed be broken up, so we do have locality after a while! (The location of the rightmost particles in the diagram below correspond to the number of steps that the particles have moved.)
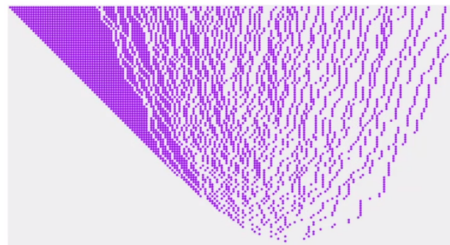
Relaxation automatically follows just like in ASEP because our slopes are always ± 1, and slope-dependent growth also holds because the more "negative" our slope is, the more possibilities we have for adding boxes, and the larger the rectangles can be when we add boxes each time. So our conclusion is that **this model should indeed be in the KPZ universality class**, and this turns out to be true – we have asymptotic calculations for certain initial conditions.

Remark 87. *For $(1 + 1)$-dimensional systems, it's often interesting to think about the dyanmics of the holes instead of the particles. The dynamics of TASEP holes are like the evolution of TASEP particles in the opposite direction, but PushTASEP holes evolve by taking the place of a particular hole – each hole can jump to any of the occupied locations before overcoming the next hole.*

We can now turn to another model which appear in this $(2 + 1)$-dimensional growth model, the domino tilings of the Aztec diamond. We will need to make a **discrete-time** version of the particle evolution, which we do as follows: in each second, each particle could potentially jump to the right by 1, and this happens in the following way. We look at the bottom particle, flip a coin, and if it flips heads, the particle jumps (otherwise it doesn't). Then we look at the

next level – if a particle was forced to jump by a particle in the lower level, then it jumps (and cannot do anything else). Otherwise, if a particle is forbidden from jumping because of the interlacing conditions, then it does nothing. And in the last case, if the particle can jump but hasn't been pushed yet, we flip a fair coin and jump if it comes up heads. (This process can then continue similarly for all other levels.)

So we now have a discrete-time version of the particle system, with time denoted by $T \in \mathbb{Z}_{\geq 0}$, and we'll look at the configuration by shifting time layer-by-layer: we look at x_1^1 at time T, then x_1^2 and x_2^2 at time $(T-1)$, then x_1^3, x_2^3, x_3^3 at time $(T-2)$, and so on, until layer T (which is evaluated at time 0, so that everything must be packed). In other words, we define a new array via

$$y_m^n = x_m^n(T - n + 1).$$

This array then satisfies the weak interlacing conditions $y_m^{n+1} \leq y_m^n \leq y_{m+1}^{n+1}$, but additionally in every layer we have strict inequalities between neighbors. So this is exactly what we saw with the **monotone Gelfand-Tsetlin patterns** in the domino tilings – at time T, we get a domino tiling of some size (which is larger for larger T), and we might ask whether we can find dynamics for evolving T forward. We can do this with the **shuffling algorithm** on the y_m^n's.

> **Fact 88**
>
> We can see a Youtube video at https://www.youtube.com/watch?v=Yy7Q8IWNfHM, in which it is mentioned that the Aztec tilings can be sampled uniformly at random but that no implementation was readily available. Within a few days of the video, there were dozens of implementations that were submitted, and https://charlymarchiaro.github.io/magic-square-dance/ is one of them.

Basically, we know that each of the four types of domino tilings have different preferred corners, so at each time step $T \to T + 1$, dominos will move towards their preferred corners. And if there are contradictions with two dominos trying to move to the same spot, they cancel each other out (these are exactly the 2×2 bad squares that explain why the ASMs and domino tilings do not biject!). And once all the dominos move, the remaining empty space is uniformly filled with more dominos.
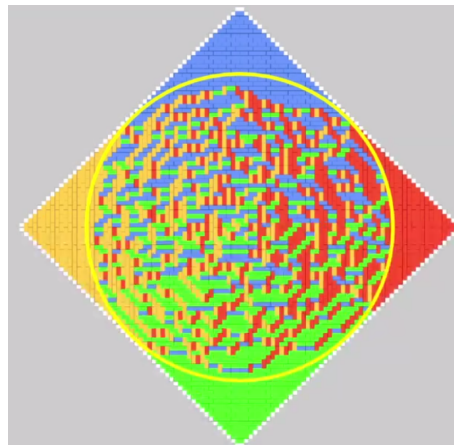
(Blue, red, green, and yellow dominos will move up, right, down, and left.) And in fact, if we bias the appearances of new dominos between vertical and horizontal dominos, the limit shape becomes an inscribed ellipse instead of an inscribed circle! But we should try going through the different implementations ourselves, and the thing we should understand is that random domino tilings can be grown by some dynamics in the anistropic KPZ universality class. (So the fluctuations that we'll see around the limit shape should be the same as in other dynamics in the class!)

And our next step will be to connect to lozenge tilings, which is more straightforward to do than for the domino tilings: we place a vertical rhombus at each location where we have a particle, and we fill in the rest with horizontal rhombi.
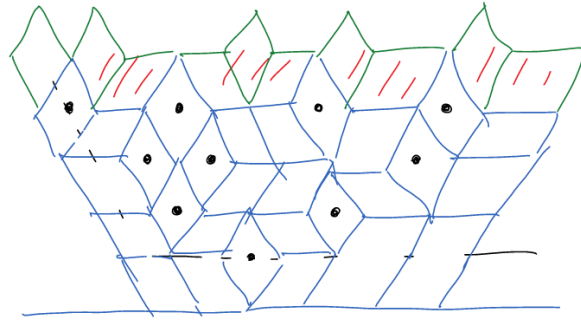
Notice that the domain on which we draw rhombi is now infinite, and in this case, if we want to restrict to a finite set of rows, what happens at the boundary is more complicated. But the point is that we have a probability measure on infinite tilings for every $t \in \mathbb{R}_{\geq 0}$, and it makes sense to ask whether there is an independent way to describe this measure (like how we had uniformly random tilings).

But because we have infinite tilings, we cannot have a uniform measure – instead, the answer turns out to come from the uniform measure on the **tilings of a hexagon** with lengths A, B, C. Basically, if we take the limit where $A, B, C \to \infty$, but $\frac{AB}{C}$ converges to some t, we'll observe the measure for the infinite tilings near the AB vertex of the hexagon. We then have the "frozen" region which is usually an inscribed ellipse inside the hexagon, and we will degenerate from the ellipse to a parabola in the limit: this parabola is the one that describes the asymptotics of TASEP!
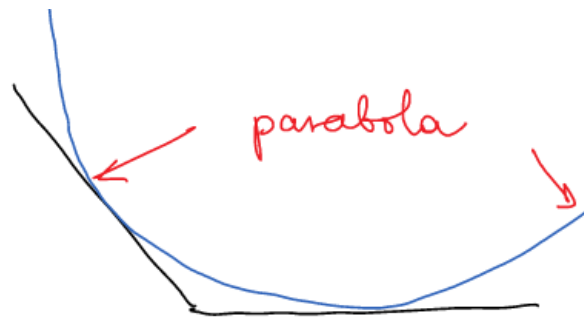
It's then natural to ask whether we can define dynamics on this finite hexagon, and indeed we can do this by expanding one of the side lengths (which gives us "paths" of differently-oriented rhombi that emerge)! And next lecture, we'll relate these objects from today to representation theory of the unitary group.

13 April 6, 2021

Our path will now turn to representation theory and random matrices, and we'll discuss our probability objects with more mathematics (and less pictures) from now on. Recall that at the beginning of 18.677, we started discussing the symmetric group $S(n)$ and its representations: in general, if we have a group G, recall that a **representation** of G is a group homomorphism $T : G \to GL(V)$. Here, the presentation will usually be taken to be finite-dimensional, meaning elements of $GL(V)$ can be represented as matrices, complex, meaning that V is a vector space over $\mathbb{C}$, and unitarizable, meaning that the image of T is a subset of the unitary operators on V.

We mentioned that a representation T is **irreducible** if it has no nontrivial invariant subspaces (that is, none other than V or $\{0\}$). Representation theory basically has a few different stages: in the first stage, we want to classify all irreducible representations, because other problems can often be broken up into these smaller breaking blocks. The

next stage has us decomposing "natural" representations on irreducibles into

$$V = \bigoplus_{\lambda \in \text{Irr } G} m_\lambda V_\lambda.$$

For example, we found that if we look at the space of all functions on $\mathbb{C}[G]$ with natural action

$$(T(g)f) = f(g^{-1}x)$$

for any $f : G \to \mathbb{C}$, we get the **regular representation**, and it turns out that we can break this up into irreducible representations as

$$\mathbb{C}[G] = \bigoplus_{\lambda \in \text{Irr } G} = \dim(\lambda) \cdot T_\lambda.$$

We then get Burnside's formula by looking at dimensions on both sides: we find that

$$|G| = \sum_{\lambda \in \text{Irr } G} \dim^2 \lambda.$$

Last time we visited this, we used it to justify the Plancherel measure, but we're going to think about other classes of groups and take things in a different direction now. We can build a satisfying theory of irreducible representations of **compact groups**, and the main difference is that the space of functions is infinite-dimensional (so we won't get any dimension-counting formulas in the same way). And this time, we can look at the **unitary groups**

$$\mathcal{U}(N) = \{U \in \text{Mat}(N, \mathbb{C}) : UU^* = I\},$$

and the key is that there are many different ways to represent $\mathcal{U}(N)$:

Proposition 89

The representations of $\mathcal{U}(N)$ are in bijection with representations of $GL(N, \mathbb{C})$.

One direction of this is easy: we go from a representation of $GL(N, \mathbb{C})$ to $\mathcal{U}(N)$ through restriction, but the other direction is called the **unitary trick of Weyl** and is basically an analytic continuation. Basically, we can show abstractly that a representation of $\mathcal{U}(N)$ lifts to one of $GL(N, \mathbb{C})$, in a way that preserves many of the important properties.

Remark 90. *The representation of $GL(N, \mathbb{R})$ is very different from $GL(N, \mathbb{C})$: the interesting theory is infinite-dimensional, but it has no direct relation to unitary groups.*

Theorem 91

All irreducible representations of compact (Lie) groups are finite-dimensional.

The idea is then that we'll describe irreducible representations of $\mathcal{U}(N)$ by restricting to the set of diagonal unitary matrices (also called the **maximal torus**)

$$\mathcal{H}(N) = \left\{ \text{diag}(e^{i\phi_1}, \cdots, e^{i\phi_N}), \phi_1, \cdots, \phi_N \in \mathbb{R} \right\}.$$

This is a maximal abelian subgroup, and these subgroups play a big role in representation theory because irreducible representations of abelian groups are always one-dimensional (due to simultaneous diagonalization, and it turns out that for compact groups Jordan blocks never arise in diagonalization). So if we take some $T : \mathcal{U}(N) \to GL(V)$, where $V \cong \mathbb{C}^m$, and we restrict T to the maximal torus $\mathcal{H}_N$, then $T : \mathcal{H}_N \to GL(V)$ is diagonalizable, and we can break up

$\mathbb{C}^m$ into m one-dimensional subspaces so that

$$T(\text{diag}(e^{i\phi_1}, \cdots, e^{i\phi_N}))v_j = t_j(e^{i\phi_1}, \cdots, e^{i\phi_n})v_j,$$

where the eigenvalues $t_j(e^{i\phi_1}, \cdots, e^{i\phi_n})$ are continuous homomorphisms of the maximal torus $(S^1)^N$. Because the most general **continuous** homomorphisms of the n-dimensional torus into $\mathbb{C}$ look like

$$t_j(e^{i\phi_1}, \cdots, e^{i\phi_N}) = e^{ik_1 phi_1} \cdots e^{ik_N \phi_N}$$

for integers $(k_1, \cdots, k_N)$ (such an N-tuple is called a **weight** of T), and each representation should only have finitely many weights, this helps us classify all irreducible representations of $\mathcal{U}(N)$:

Theorem 92 (Weyl)

Irreducible representations of $\mathcal{U}(N)$ over $\mathbb{C}$ are parameterized by N-tuples $\lambda = (\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N) \subset \mathbb{Z}^N$, where the correspondence asks λ to be the highest weight of the representation. Furthermore, given the highest weight, the generating function of the weights for a representation is given by **Weyl's character formula** for the special case $\mathcal{U}(N)$:

$$\sum_{(k_1, \cdots, k_N) \text{ weight of } T_\lambda} z_1^{k_1} \cdots z_N^{k_N} = \text{tr}\left(T_\lambda(z_1, \cdots, z_N)\right) = \frac{\det\left[z_i^{N+\lambda_j-j}\right]_{i,j=1}^N}{\det\left[z_i^{N-j}\right]_{i,j=1}^N}$$

(where the first equality here is because "trace is the sum of the eigenvalues").

Again, we notice that there is a Vandermonde determinant on the right-hand side. And notice that if we think about our earlier discussion of symmetric groups, we defined the **character** of a representation $T : S(n) \to GL(V)$ to be $\text{tr}(T) : S(n) \to \mathbb{C}$. But there is no explicit formula for this character – if we take an irreducible representation of $S(n)$ and we want to know the values of the character on a given permutation, there is no known formula for computing that value. So it's interesting that we get an explicit formula for the unitary group – there are differences between study of finite and compact groups.

Definition 93

The **Schur polynomial** $s_\lambda(z_1, \cdots, z_N)$ is the character $\text{tr}(T_\lambda)$.

These were introduced initially in the context of harmonic analysis, but they now play major roles in a variety of areas (the homology of the Grassmannian and so on). In representation theory, we also have the **Schur-Weyl duality**, in which we consider the n-fold tensor product

$$\mathbb{C}^N \otimes \cdots \otimes \mathbb{C}^N.$$

This space is a representation of the unitary group $\mathcal{U}(N)$, since

$$T(U)(v_1 \otimes \cdots \otimes v_n) = Uv_1 \otimes \cdot \otimes Uv_n.$$

But on the other hand, this is also a representation of $S(n)$, since for any permutation σ we can take

$$\Pi(\sigma)(v_1 \otimes \cdots \otimes v_n) = v_{\sigma^{-1}(1)} \otimes \cdots \otimes v_{\sigma^{-1}(n)}.$$

These actions clearly commute with each other, so we have a representation of $\mathcal{U}(N) \times S(n)$, and in fact this

representation decomposes as a direct sum over λs:

$$\mathbb{C}^N \otimes \cdots \otimes \mathbb{C}^N = \bigoplus_{\lambda = (\lambda_1 \geq \cdots \geq \lambda_N) \geq 0, \sum_i \lambda_i = n} T_\lambda \otimes \Pi_\lambda.$$

> **Fact 94**
>
> If we assume that $\lambda_N \geq 0$, then we often picture λ as a partition of length at most N (using a Young diagram).
> So this is a duality that is summing over partitions, and there is a unique representation of the symmetric group
> corresponding to a representation of the unitary group in this way above.

We'll now use all of this structure of representation for a more concrete problem: take an irreducible representation T_λ of $\mathcal{U}(N)$ and restrict it to $\mathcal{U}(N-1)$ (by only looking at the upper $(N-1) \times (N-1)$ submatrix), we must have a decomposition as a direct sum over highest weights

$$T_\lambda|_{\mathcal{U}(N-1)} = \bigoplus_{\mu = (\mu_1 \geq \cdots \geq \mu_{N-1})} m_\mu T_\mu^{(N-1)},$$

and it's a problem in harmonic analysis to try to find the m_μs. And it turns out that the answer is quite simple: taking traces on both sides, we have the Schur polynomial coming up, and

$$s_\lambda(z_1, \cdots, z_{N-1}, 1) = \sum_\mu m_\mu s_\mu(z_1, \cdots, z_{N-1}).$$

So if we take the Schur polynomial and set the last variable to 1, and we find how the Schur polynomial decomposes into Schur polynomials with one fewer variable, we can find the m_λs, and the computation tells us in fact that we have the **branching rule**

$$s_\lambda(z_1, \cdots, z_{N-1}, 1) = \sum_{\mu \prec \lambda} s_\mu(z_1, \cdots, z_{N-1}),$$

where $\mu \prec \lambda$ means that we have the **interlacing**

$$\lambda_N \leq \mu_{N-1} \leq \lambda_{N-1} \leq \cdots \leq \lambda_2 \leq \mu_1 \leq \lambda_1.$$

So all of the coefficients turn out to be either 0 or 1, and they are 1 exactly when the partitions interlace! And if we now pass to the shifted coordinates, so that

$$\lambda \mapsto (\lambda_1 + N - 1, \lambda_2 + N - 2, \cdots, \lambda_N)$$

are pairwise distinct, and similarly

$$\mu \mapsto (\mu_1 + N - 2, \mu_2 + N - 3, \cdots, \mu_{N-1}),$$

our interlacing condition becomes similar to what we had on our two-dimensional dynamics: we're saying that if we place dots at the shifted partition locations, and draw vertical rhombi at each dot, then interlacing is equivalent to the existence of such a two-layer rhombus tiling.
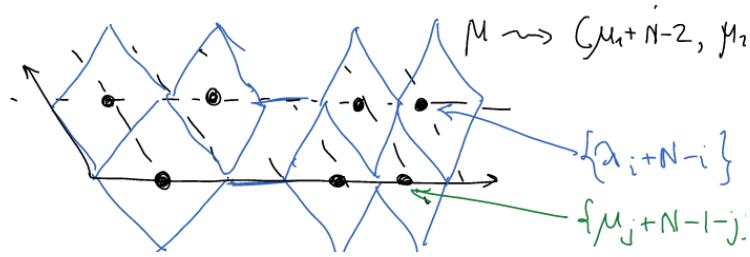
From here, we want to take a representation T_λ and restrict $T_\lambda^{(N)}$ further to $\mathcal{U}(N-1)$, then $\mathcal{U}(N-2)$, then $\mathcal{U}(N-3)$, and so on. Eventually, we'll end up the set of diagonal matrices, and on the diagonals we have the **weights of our representation** (since we went all the way down to $\mathcal{U}(1)$, which acts on S^1 (the unit circle), and we can only have one-dimensional irreducible representations. Since we have a unique way of restricting down at each step, we find a basis of $T_\lambda^{(N)}$ this way, parameterized by this restrictions

$$\lambda = \lambda^{(N)} \succ \lambda^{(N-1)} \succ \cdots \succ \lambda^{(1)},$$

and this is the return of our **Gelfand-Tsetlin patterns** (and this was the setting of how they first came up in the 1950s)! So these basis vectors of $T_\lambda^{(N)}$ are in one-to-one correspondence with triangular arrays, and we can describe these arrays as the rhombus tilings of a strip of width N. That leads us to the following:

Proposition 95

Tilings of a hexagon are in one-to-one correspondence with basis vectors of a particular representation (since we can require some rhombi to be tiled in the corners so that we recover the usual strip shape).
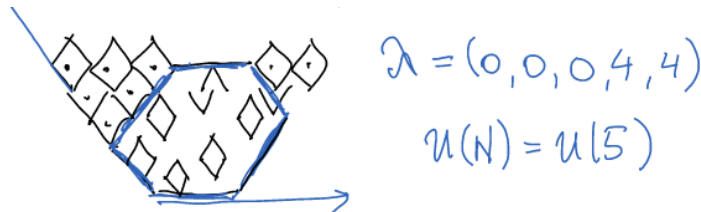
So we have a connection between representation theory (starting with a representation on a group and breaking it up into irreducible components) and probability (random lozenge tilings of a hexagon), and thus we might think that there could be a way to think of these probabilistic objects in representation theoretic language. And indeed, this exists: if we look at a horizontal slice like the third row of rhombi, the horizontal locations of the vertical rhombi corresponds to $\mathcal{U}(3) \subset \mathcal{U}(5)$, and we have

$$T_\lambda|_{\mathcal{U}(3)} = \bigoplus_{\mu_1 \geq \mu_2 \geq \mu_3} m(\mu) T_\mu,$$

and to get probability out of this we need the probability of a particular partition, which is given by

$$\boxed{\mathbb{P}(\mu) = \frac{m(\mu) \dim \mu}{\dim \lambda}}.$$

And there are methods that people have used for decomposing classical groups into irreducibles: for example, we can use representations of SL_3 to understand the orbitals of chemical atoms and other topics in physics! So there are tools for how to find, for example, these dimensions of restrictions. But we'll get to that later on.

Remark 96. *The above formula is a special case of Weyl's dimension formula, since we can compute the dimension*

of λ via $s_\lambda(1, 1, \cdots, 1)$ and then apply L'Hopital's rule multiple times – we find that

$$\dim \lambda = \prod_{1 \le i \le j \le n} \frac{\lambda_i - \lambda_j - i + j}{j - i}.$$

If we now remind ourselves about the measure on Gelfand-Tsetlin patterns that showed up in the (2+1)-dimensional dynamics, we should remember that those come up from a corner of the $A \times B \times C$ hexagon as $A, B, C \to \infty$ in a particular ratio. But then we might want to ask whether the object we get is an infinite-dimensional analog of the unitary group – the complication is that we can't define representations in the same way as for finite-dimensional theory. So we'll talk more next lecture about what that limiting object is in this case and how to properly define representations!

14   April 8, 2021

Today's class is a seminar by **Mackenzie Simper**, titled **Induced Probability Distributions on Double Cosets**. We'll start with some background on double cosets and probability distributions on them, looking at some particular examples and focusing on the Fisher-Yates distribution on contingency tables. We'll then build off of that and construct a Markov chain on the space of contingency tables using these double cosets.

Definition 97

Let G be a finite group and H and K be subgroups of G. Then the double-cosets $H \backslash G / K$ are the sets of equivalence classes under the relation

$$s \sim t \iff s = h^{-1}tk, \quad h \in H, k \in K.$$

The double cosets containing some $s \in G$ is denoted HsK.

We can think about double cosets as dictating symmetries under two subgroups.

Example 98

If $G = S_4$ and $H = K$ are both the set of permutations that fix element 4, then there are two double cosets in G. They are $H \mathrm{id} H = H$, the set of all permutations that fix 4, and $H \sigma H$ (where $\sigma = 4123$), which is the set of all permutations that don't fix 4.

Even though this is a group theory construction, we can ask questions about how many double cosets there are, how large they are, how likely a uniform element $g \in G$ is to be in a particular double coset, and so on. (For more detail and algebra background, we can check out "Statistical Enumeration of Groups by Double Cosets" on arxiv.). For example, in the first example above, the induced distribution on double cosets is

$$p(H(\mathrm{id})H) = \frac{1}{4}, \quad P(H\sigma H) = \frac{3}{4}.$$

> **Example 99** (Mallows measure on permutations)
>
> Let $G = GL_n(\mathbb{F}_q)$ (where $\mathbb{F}_q$ is the field on q elements), and let $H = K = B$ both be the set of lower triangular matrices in G. We have by the **Bruhat decomposition** that
>
> $$G = \bigcup_{\omega \in W} B\omega B,$$
>
> where W is the group of permutation matrices.

In other words, double cosets on $GL_n(\mathbb{F}_q)$ are indexed by permutations, and the induced measure turns out to be

$$p_q(\omega) = \frac{q^{\ell(\omega)}}{[n]_q!},$$

where $\ell(\omega)$ is the number of inversions in the permutation, and

$$[n]_q! = (1 + q)(1 + q + q^2) \cdots (1 + q + \cdots + q^{n-1}).$$

In other words, the size of the double coset depends on the number of inversions of the corresponding permutation. And we can understand that by combining the following combinatorial facts, which we can do as an exercise:

- The number of lower-triangular matrices is $|B| = (q - 1)^n q^{\binom{n}{2}}$.

- The size $|G|$ is $|B| \prod_{i=1}^{n-1}(1 + q + \cdots + q^i) = |B| \cdot [n]_q!$.

- The size of $B\omega B$ is $|B|q^{\ell(\omega)}$.

> **Example 100**
>
> Let $G = S_{2n}$, and let $H = K = B_n$ be the group of symmetries of an n-dimensional hypercube – in other words, we have the **centrally symmetric permutations** where $\sigma(i) + \sigma(2n + 1 - i) = 2n + 1$ for all i.

For example, $B_2 \subset S_4$ is the permutations $\{1234, 4231, 1324, 4321, 3142, 2143, 3412, 2413\}$, and in general $|B_n| = 2^n n!$. Then it turns out that the double cosets $B_n \backslash S_{2n} / B_n$ are indexed by **partitions** of n, and the induced measure is the **Ewen's sampling formula**

$$p_\theta(\lambda) = \frac{n!}{\theta(\theta + 1) \cdots (\theta + n - 1)} \cdot \frac{\theta^{\ell(\lambda)}}{\prod_{i=1}^n i^{a_i} a_i!},$$

with $\theta = \frac{1}{2}$. (If we instead set $\theta = 1$, we notice that this actually gives us the distribution of cycles induced by a uniform random permutations.)

We can make the mapping to partitions as follows: for each permutation $\sigma \in S_{2n}$, we draw a graph $T(\sigma)$ with vertices $\{1, 2, \cdots, 2n\}$ and edges $\{\varepsilon_i, \varepsilon_i^\sigma\}$, where ε_i joins the vertices $(2i - 1)$ and $2i$ (color these edges red) and ε_i^σ joins vertices $\sigma(2i - 1)$ and $\sigma(2i)$ (color these edges blue). For example, if $n = 3$ and $\sigma = 612543$, then we get edges between $(1, 2), (3, 4), (5, 6)$, as well as $(6, 1), (2, 5), (4, 3)$.

It turns out that this process always makes a graph which is a disjoint set of cycles of even length, and if we divide all of those lengths by 2 we get a partition of n. (In the above case, we get the partition $(2, 1)$ for $n = 3$, since we get the cycles $(1256)(34)$.) And then the Ewen's sampling formula basically comes from doing the combinatorics of how many ways there are to make a permutation map to a particular permutation – it's not too challenging for us to verify.

Example 101 (Young subgroups of S_n)

Let $(\lambda_1, \lambda_2, \cdots, \lambda_I)$ be a partition of n. Then the **Young subgroup** S_λ is the set of permutations which permute the first λ_1 elements among themeslves, the next λ_2 elements among themselves, and so on, so that

$$S_\lambda \cong S_{\lambda_1} \times S_{\lambda_2} \times \cdots \times S_{\lambda_I}.$$

We wish to study the double cosets $S_\lambda \backslash S_n / S_\mu$.

It turns out that we can index these double cosets through **contingency tables**, which are $I \times J$ arrays of nonnnegative integers with fixed row sums $\lambda_1, \cdots, \lambda_I$ and fixed column sums $\mu_1, \cdots, \mu_J$. (These are usually useful ways of encoding two distinct categories of data in a two-way table.) To make this mapping, suppose that we have our two partitions $\lambda = (\lambda_1, \cdots, \lambda_I)$ and $\mu = (\mu_1, \cdots, \mu_J)$. Then we can define the sets

$$L_1 = \{1, \cdots, \lambda_1\}, L_2 = \{\lambda_1 + 1, \cdots, \lambda_1 + \lambda_2\}, \cdots, L_I = \{n - \lambda_I + 1, \cdots, n\},$$

and analogously define sets $M_1, \cdots, M_J$. Then we can let T_{ij} be the **number of elements in M_j that occur at positions** L_i.

Example 102

If we let $n = 5, \lambda = (3, 2), \mu = (2, 2, 1)$, then $L_1 = \{1, 2, 3\}, L_2 = \{4, 5\}$, and $M_1 = \{1, 2\}, M_2 = \{3, 4\}, M_3 = \{5\}$. Suppose we want the contingency table for $\sigma = 12345$.

We want our contingency table to have two rows and three columns with the row and column sums above, and we calculate the entries as follows: in the first $|L_1| = 3$ entries of the permutation, there are two elements of M_1, so $T_{11} = 2$. Similarly, in the first three entries, there is 1 element of M_2 and 0 elements of M_3. Continuing in this way, we find the contingency table

$$\sigma = 12345 \implies \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

(Here, we can see the double coset symmetries in action: two elements map to the same table if the only change between them is the action of S_λ on the indices of the permutation, as well as the action of S_μ on the elements of the permutation.) It turns out that there are five such possible tables, each with their own double coset representative.

So now we want to return to probability and ask for the distribution on double cosets induced by the uniform distribution on permutations, finding the chance of a particular contingency table. Again, this comes from the size of the double cosets:

Definition 103

The **Fisher-Yates distribution** on contingency tables for $I \times J$ contingency tables with fixed row sums $\lambda_1, \cdots, \lambda_I$ and column sums $\mu_1, \cdots, \mu_J$ is

$$P_{\lambda, \mu}(T) = \prod_{j=1}^{J} \binom{\mu_j}{T_{1j}, \cdots, T_{Ij}} \Big/ \binom{n}{\lambda_1, \cdots, \lambda_I} = \frac{\prod_i (\lambda_i)! \prod_j (\mu_j)!}{n! \prod_{i,j} (T_{ij})!}$$

Combinatorially, we can think of this as a **sampling-without-replacement** distribution as follows: suppose we have an urn with J different colors, μ_j of type j. Then we make I draws without replacement – the first time, we draw λ_1 balls, the next time we draw λ_2, and so on, so that after I draws there are no balls left in the urn. And then we always

set T_{ij} to be the number of balls of color j on our ith draw. And this should convince us why the uniform distribution on S_n induces the Fisher-Yates distribution: we can color our balls based on μ and record the sequence of n draws of balls through our permutation $\sigma \in S_n$.

Remark 104. *We might notice that when we have a 2×2 table, this distribution is the hypergeometric distribution, and for the $2 \times J$ table, we have the multivariate hypergeometric distribution.*

Remark 105 (Independence model)**.** *Suppose we drop balls in an $I \times J$ table by dropping one in cell (i, j) with probability $p_i q_j$. Then the measure on tables is*

$$\mathbb{P}(T) = \frac{n!}{\prod_{i,j} T_{ij}!} \prod_{i,j} (p_i q_j)^{T_{ij}} = \frac{n!}{\prod_{i,j} T_{ij}!} \prod_{i=1}^{I} p_i^{\lambda_i} \prod_{j=1}^{J} q_j^{\mu_j}.$$

*In particular, if we fix the row sums and column sums λ_i, μ_j, and we condition the independence model on those sums, we get the Fisher-Yates distribution regardless of our initial p_i, q_j. And this type of setup is useful for **testing a hypothesis that two different traits are independent** (using the chi-squared distribution).*

It turns out that the contingency table corresponding to the largest double coset (and thus the highest probability) is the one that is closest to the **independence table** where $T_{ij}^* = \frac{\lambda_i \mu_j}{n}$ (which is the expected value). So we do get some centering around some central behavior here!

So the rest of our talk will now be focused on defining a **Markov chain on contingency tables**, looking at mixing time, eigenvalues, and eigenfunctions. We'll start with a simple Markov chain:

Example 106 (Random transpositions)

Think of a permutation as a deck of cards, and on each move, we pick one card with our left hand and one with our right hand (potentially the same card) and swap their positions. Then the transition matrix is

$$P(x, y) = \begin{cases} 2/n^2 & y = (ij)x, i \neq j \\ 1/n & y = x \\ 0 & \text{otherwise.} \end{cases}$$

In other words, we randomly swap some i and j in our permutation. We can use this to define the following Markov chain on contingency tables:

Definition 107

The **random transpositions Markov chain** on the space of tables $T_{\lambda,\mu}$ (where row and column sums are given by λ, μ) is defined as follows: pick an $x \in S_n$ so that $T = T^x$ (the contingency table induced by x), make one move in the random transpositions Markov chain to $y \in S_n$, and then move to the table $T' = T^y$.

It turns out that this transition probability **does not depend on the choice of double-coset representative** – this is Dynkin's condition – and thus we have a well-defined Markov chain, and in fact the Fisher-Yates distribution is the stationary distribution. We'll state Dynkin's condition below:

Lemma 108

If X_t is a Markov chain on a state space Ω with equivalence relation $\sim$ on Ω, and

$$P(x, [y]) = \sum_{y' \sim y} P(x, y) = P(x', [y])$$

for all $x \sim x'$, then we can define the Markov chain $[X_t]$ on the space of equivalence classes $\tilde{\Omega}$ with transition probability $\tilde{P}([x], [y]) = P(x, [y])$. This gives us a **lumped chain** with stationary distribution $\tilde{\pi}([x]) = \sum_{x' \sim x} \pi(x)$ (where π is the stationary distribution of X_t).

Example 109

If we stick with our example $\lambda = (3, 2), \mu = (2, 1, 1)$, and we start with the table $T = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$, we can pick some permutation, say $x = 12345$, and then make a random transposition to (for example) $y = 14325$. Then we end up with $T' = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 0 & 1 \end{bmatrix}$.

We may notice that the net change in the contingency table is $\begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \end{bmatrix}$, and this kind of change always preserves the column sums. And in fact, those kinds of moves are the only types of moves that will ever happen: to get to the table $F_{(i_1, j_1), (i_2, j_2)} T$ from T, we basically subtract 1 from (i_1, j_1) and (i_2, j_2) and add 1 to (i_1, j_2) and (i_2, j_1), and it turns out that the probability of making this transition is

$$\mathbb{P}(T, F_{(i_1, j_1), (i_2, j_2)} T) = \frac{T_{i_1 i_2} T_{j_1 j_2}}{n^2}$$

(since this transformation of the contingency table basically depend on being able to pick out one of the elements that originally was in $T_{i_1 i_2}$ and one that was originally in $T_{j_1 j_2}$). This Markov chain is in fact similar to another one which has the same kind of move, but instead of having the probabilities depend on the values in the table, we pick the rows and columns for $(i_1, j_1), (i_2, j_2)$ **uniformly at random**. So this time we just have a **symmetric chain**, and thus we have the uniform stationary distribution

$$\mathbb{P}(T, F_{(i_1, j_1), (i_2, j_2)} T) = \frac{1}{I \cdot J}.$$

We can run this Markov chain for a long time and that will help us sample from the uniform distribution on tables, and thus that's a motivation for asking about the **mixing time** of the Markov chain:

Definition 110

Let X_t be a discrete-time Markov chain. The **total variation distance** is defined as

$$d(t) = \sup_{x \in \Omega} ||P^t(x, \cdot) - \pi(c)||_{TV} = \sup_{x \in \Omega} \max_{A \subset \Omega} |P(X_t \in A : X_0 = x) - \pi(A)|$$

(this is decreasing with time). Then the **mixing time** is

$$t_{\text{mix}} = \left\{ \inf\{t > 0 : d(t) < \frac{1}{4}\} \right.$$

Basically, the mixing time is an asymptotic way to understand how long it takes for a Markov chain to converge to its stationary distribution. The results we'll be able to state here are related to the eigenvalues and eigenfunctions of

the Markov chain: probabilistically, the definition is that

$$\mathbb{E}[f_i(X_1) : X_0 = x] = \beta_i f_i(x).$$

We can choose our eigenfunctions to be orthonormal with respect to the stationary distribution, meaning that

$$\sum_{x \in \Omega} f_i(x) f_j(x) \pi(x) = \delta_{ij}.$$

So we have the inner product space $\ell^2(\pi)$, and now we can also define the **chi-square distance**

$$||P^t(x, \cdot) - \pi(\cdot)||_2^2 = \sum_{y \in \Omega} \frac{|P^t(x, y) - \pi(y)|^2}{\pi(y)} = \sum_{i=1}^{|\Omega|-1} \beta_i^{2t} f_i^2(x).$$

This then gives us a bound on the total variation, which turns out to be at most $\frac{1}{4}$ of this chi-square distance. But to do this, we still need to evaluate the functions f_i at particular points x, so we need to understand how they look over the whole state space.

We can do so as follows: we know that the chain on contingency tables is a lumped version of the chain on permutations, which is good because we can lift eigenfunctions on our lumped chain to eigenfunctions on our original chain (with the same eigenvalue), and conversely we can take an eigenfunction on our original chain (which is constant on equivalence classes) and get a projection to our lumped chain. And we know the eigenvalues for the random transpositions chain:

Proposition 111 (Diaconis and Shashahani, 1981)

The eigenvalues β_ρ for the random transpositions chain are indexed by partitions $\rho = (\rho_1, \cdots, \rho_k)$ of n, and they are given by

$$\beta_\rho = \frac{1}{n} + \frac{1}{n^2} \sum_{j=1}^{k} \left[(\rho_j - j)(\rho_j - j + 1) - j(j-1) \right].$$

In additional, the multiplicity of β_ρ is d_ρ^2, where d_ρ is the **hook-length** of the partition ρ.

So we know that the eigenvalues of the contingency table must be a subset of these, but we need to know which show up and with what multiplicity. And here is the result that we have:

Theorem 112

If β is an eigenvalue of the random transpositions Markov chain on the space of contingency tables, then

$$\beta = 1 - \frac{2m(n+1-m)}{n^2}$$

for some $0 \le m \le \lfloor n/2 \rfloor$. The eigenbasis for β is the set of **orthogonal polynomials** for the Fisher-Yates distribution of degree m (meaning that $\langle f, g \rangle_\pi = 0$).

This formula corresponds to the formula on the previous page with a two-part partition $(n - m, m)$ (though we do not know yet what the multiplicities are, since we need the number of orthogonal polynomials).

> **Corollary 113**
>
> The second-largest eigenvalue of the chain is $\beta_1 = 1 - \frac{2}{n}$, and it always has multiplicity $(I-1)(J-1)$. The basis for the space is given by
> $$f_{ij}(x) = x_{ij} - \frac{\lambda_i u_j}{n}.$$

(If we now specialize to $2 \times J$ tables, the stationary distribution is multivariate hypergeometric, so we know more in this case – we can write down all of the eigenfunctions and multiplicities more explicitly.) But to get back to the mixing time and bounding the total variation distance, we want to be able to analyze $f_i(x)$ for every x. And again, we can do this in the case where we have $2 \times J$ tables, so that $\lambda = (n - k, k)$ for some $k \leq \lfloor n/2 \rfloor$. Suppose that $\mu = (\mu_1, \cdots, \mu_J)$, and further suppose that $\mu_j > k$. If we let $k\mathbf{e}_j$ be the table whose second row is just k in the jth column and 0 everywhere else, the orthogonal polynomials will simplify, and we can get an explicit bound for the chi-square distance in terms of n:

> **Proposition 114**
>
> We have a bound on the distance $||\tilde{P}^t(k\mathbf{e}_j, \cdot) - \pi(\cdot)||_2^2 \leq e^{-c}$ for $t \geq \frac{n}{4}\left(c + \log\left(\frac{kn(n-\mu_j)}{(n-k)(\mu_j-k)}\right)\right)$.

In the special case where $I = J = 2$, $k = \frac{n}{2} - 1$, and $\mu = \frac{n}{2}$, the argument inside the log is proportional to n. In comparison, it turns out that the chain of random transpositions on S_n mixes as $\frac{n}{2}\log n$, so in this case we have something that is a little bit faster! And we can also get a lower bound for the mixing time in general by applying Wilson's method to get a lower bound:

> **Proposition 115**
>
> If $(\lambda_1, \cdots, \lambda_I)$ and $(\mu_1, \cdots, \mu_J)$ are two partitions of n, then for any i, j with $n > 2(\lambda_i + \mu_j)$, we ave
> $$t_{\text{mix}} \geq \frac{n}{4}\left(\log\left(\min(\lambda_i, \mu_j) - \frac{\lambda_i\mu_j}{n}\right) + c\right)$$
> for some constant c.

15   April 13, 2021

Last lecture, we discussed how to view certain probability objects, like lozenge tilings, in a representation theoretic manner by bijecting them with certain basis vectors and looking at the corresponding irreducible representation. (If we then want to look at restrictions to a particular horizontal line of the tiling, we can look at a specific subspace and decompose that further into irreducible representations.) The connection is that the probability measure on a given horizontal line can be written in terms of the multiplicity of the representation, as well as the dimensions of the representations – while these calculations may not be simpler than counting, they do carry along tools from representation theory that can be applied.

The last thing we talked about last time was the stretched hexagon, in which we look at the AB corner of an $A \times B \times C$ equiangular hexagon and take $A, B, C \to \infty$ with $\frac{AB}{C} \to t$.
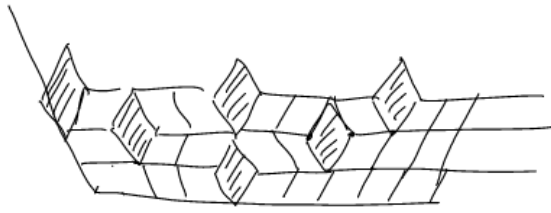
Then we only see the bottom levels of our lozenge tiling, and we are still interested in the position of the vertical lozenges – the question is whether this picture has any representation theoretic meaning. Remember that we have the formula for the homomorphisms T_λ

$$T_\lambda|_{\mathcal{U}(M)} = \bigoplus_{\mu_1 \geq \cdots \geq \mu_M} m(\mu) T_\mu,$$

we can take traces of both sides and find that

$$\chi_\lambda(z_1, \cdots, z_M, 1, \cdots, 1) = \sum_{\mu_1 \geq \cdots \geq \mu_M} m(\mu) \chi_\mu(z_1, \cdots, z_M),$$

where χ_λ on the left-hand side is the character of the group $\mathcal{U}(N)$ (specialized so everything that isn't the first M coordinates is 1) and the χ_μ on the right side is the character of the group $\mathcal{U}(M)$. Alternatively, we can look at normalized characters and get the equation

$$\frac{\chi_\lambda(z_1, \cdots, z_M, 1, \cdots, 1)}{\chi_\lambda(1, \cdots, 1)} = \sum_{\mu_1 \geq \cdots \geq \mu_M} \frac{\dim \mu \cdot m(\mu)}{\dim \lambda} \frac{\chi_\mu(z_1, \cdots, z_M)}{\chi_\mu(1, \cdots, 1)}.$$

And now if we take $N \to \infty$, meaning that our hexagon approximates a sector of the plane in a certain way, then it turns out that if we set the z_is to be points on the unit circle, we have

$$\lim_{N \to \infty} \frac{\chi_\lambda(z_1, \cdots, z_M, 1, \cdots, 1)}{\chi_\lambda(1, \cdots, 1)} = \exp\left(t \sum_{i=1}^{M} (z_i - 1) \right).$$

Our limiting probability measure is thus determined by the simple-looking functions

$$\boxed{\exp\left(t \sum_{i=1}^{M} (z_i - 1) \right) = \sum_{\mu_1 \geq \cdots \geq \mu_M} \text{Prob}(\mu) \cdot \frac{\chi_M(z_1, \cdots, z_M)}{\chi_M(1, \cdots, 1)}.}$$

Studying this requires infinite-dimensional representation theory, and we should remember that we have a nesting of groups $\mathcal{U}(1) \subset \mathcal{U}(2) \subset \cdots \mathcal{U}(m) \subset \mathcal{U}(m+1) \subset \cdots$, and we can construct the **infinite-dimensional unitary group** using the **inductive limit** or **direct limit** $\varinjlim \mathcal{U}(m)$, which is the set of infinite-dimensional matrices $\mathcal{U}(\infty)$ which has a finite-dimensional unitary matrix and 1s on the diagonal after a certain point.

But we need to be very careful when defining objects like representations on this group – instead of jumping into that, we'll just talk about the characters here. We know that we need normalized characters to talk about probability measures, and thus we want to think about the normalized traces of our irreducible representations. Here's how we can approach this:

> **Definition 116**
>
> A **normalized character** of a topological group G is a continuous function $\chi : G \to \mathbb{C}$ with the following conditions:
>
> - $\chi(e) = 1$ (**normalization**),
>
> - $\chi(ab) = \chi(ba)$ (**centrality** − we can alternatively think of this as a function constant on conjugacy classes, and this fact is true for traces),
>
> - For any $z_1, \cdots, z_k \in \mathbb{C}$ and $g_1, \cdots, g_k \in G$, $\sum_{i,j=1}^{k} z_i \overline{z_j} \chi(g_i g_j^{-1}) \geq 0$ (**positive definiteness**). We can think of this as saying that the matrix $\left(\chi(g_i g_j^{-1}) \right)_{i,j=1}^{k}$ is positive definite.

The motivation for wanting positive definiteness in this definition is that if G is a finite group, there should be some relevant to the representation theory of finite groups (in the language of invertible operators on finite-dimensional vector spaces). If we have a finite-dimensional representation $T : G \to GL(V)$, and we define

$$\chi(G) = \frac{\mathrm{Tr}(Tg)}{\dim T},$$

normalization and centrality, and we want to check positive definiteness. Indeed, we want to compute

$$\frac{1}{\dim T} \sum_{i,j} z_i \overline{z_j} \mathrm{Tr}(T(g_i g_j^{-1})),$$

and using linearity of trace and the fact that T is a homomorphism, we can simplify this to

$$= \frac{1}{\dim T} \mathrm{Tr}\left(\sum_{i,j} z_i \overline{z_j} T(g_i) T(g_j^{-1}) \right) = \frac{1}{\dim T} \mathrm{Tr}\left(\left(\sum_i z_i T(g_i) \right) \left(z_j T(g_j^{-1}) \right)^* \right)$$

where the $*$ means conjugate transpose, since $T(g_j^{-1}) = (T(g_j))^*$ for a unitary representation T (T is always unitarizable). And this last expression is of the form $\frac{1}{\dim T} \mathrm{Tr}(AA^*)$, which is indeed positive.

> **Fact 117**
>
> It turns out that this is a good characterization of our characters: any central, normalized, positive-definite function on a finite group G is always of the form
>
> $$\chi(\cdot) = \sum_{T \in \mathrm{Irr}(G)} C_T \frac{\mathrm{Tr}(T(\cdot))}{\dim T}$$
>
> for nonnegative constants C_T summing to 1.

In other words, characters are **convex combinations** of the normalized traces of the irreducible representations, and thus irreducible representations are in bijection with the **extreme points** of the convex set of characters (and it's indeed clear that if f and g satisfy the condition above, so does $af + (1-a)g$ for $a \in [0,1]$). This convex set turns out to always be a simplex, so that there is a unique decomposition of any point in the interior.

So far, we've only had to talk about functions on groups, rather than representation theory, and then we can make the analogous definition for infinite-dimensional groups as well. The theory of characters of the infinite-dimensional unitary group is deep and has many connections, but we'll just say a few things here: if χ is a character of $\mathcal{U}(\infty)$, then χ restricted to $\mathcal{U}(M)$ is a character of $\mathcal{U}(M)$ (since the definitions above do not change), and in fact we get a

decomposition onto extreme points given by

$$\chi|_{\mathcal{U}(M)} = \sum_{\mu_1 \geq \cdots \geq \mu_M} c_\mu^\chi \frac{\mathrm{Tr}(T_\mu)}{\dim \mu}.$$

But this equation is exactly the boxed equation from earlier on in lecture: on the left-hand side, we get contributions from eigenvalues in $\mathcal{U}(M)$, and on the right-hand side we have normalized characters. So $e^{t\sum_i(z_i-1)}$ turns out to be an extreme character, and it corresponds to a particular representation in $\mathcal{U}(\infty)$ which is important.

So far, we've only spoken about the unitary groups, but there are other Lie groups that form embeddings in the same way, like the (odd and even) orthogonal groups and the symplectic groups. The pictures we get in those cases look as below (90 degrees rotated, so that the top of our Gelfand-Tsetlin schemes are now on the right):
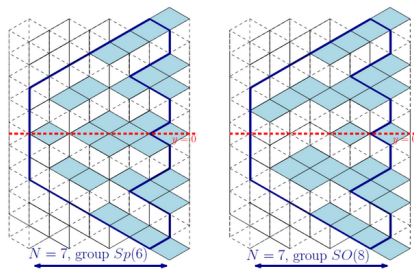
This time, on the left picture, we look at hexagon tilings that are symmetric under reflections around $y = 0$, so that we require $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4 \geq 0$, and on the right picture, we just require $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq |\lambda_4|$ (so λ_4 can be either positive or negative). So in both cases, we impose additional boundary conditions on our tilings!

We're now ready to turn to **random matrix theory**, which is a huge subject with many connections. This subject is like an octopus – there are not that many core facts that are basic in random matrix theory, but there are many different developments in different directions, and it can be difficult to understand the various topics. We'll try to spend our discussion understanding the connections with other things we've discussed in this class.

Recall that in our discussion of the symmetric group $S(n)$, we talked about the uniform measure on permutations and about representations of the group. We've talked about the latter point for $\mathcal{U}(N)$, and now we can say a few words about picking from the "uniform measure" on $\mathcal{U}(N)$. Since $\mathcal{U}(N)$ is a compact Lie group, we can define a unique measure on the unitary matrices that is invariant under both left and right multiplication, and this measure can be normalized – this is known as the **Haar measure** on the corresponding Lie group. So we'll do our best to study a uniformly random unitary matrix and understand its properties.

The first thing we did was to study the cycle length (the conjugacy classes of $S(n)$), and in this case unitary matrices can be diagonalized (and those eigenvalues can be permuted). So if we care about the distribution of conjugacy classes, the question is really the **distribution of the eigenvalues** of the matrix. (The set of all invertible matrices is not compact, so it's harder to put a measure on that – that's why we're considering the unitary ones for now.)

> **Theorem 118** (Weyl's integration formula)
> Let U be a Haar-random element of $\mathcal{U}(N)$. Then the distribution of the eigenvalues has a density with respect to the Lebesgue measure on $(S^1)^N$, equal to a constant times $\prod_{1 \leq i < j \leq N} |u_i - u_j|^2$.

This product can be rewritten so that the density is proportional to the quantity

$$\exp\left(2 \sum_{i<j} \ln(|u_i - u_j|)\right).$$

In other words, we can imagine the unit circle and N eigenvalues on it, and we have a **logarithmic** or (two-dimensional)

Coulomb potential between the eigenvalues, which are all electrically charged the same way. Then these charges want to repel each other, and the 2 is like putting a fixed value of the temperature on our probability ensemble.

The reason Weyl needed this formula was to integrate central functions over the unitary group – if we know that a function f only depends on the eigenvalues, we can just integrate

$$\int_{\mathcal{U}(N)} f(u)dU$$

over the N eigenvalues rather than the N^2 dimensions, and that gives us an extra Jacobian factor:

$$= \int_{(S^1)^N} f(u_1, \cdots, u_N) \prod_{i<j} |u_i - u_j|^2 d\text{Lebesgue}.$$

This probability measure is called the **Dyson circular ensemble**, named after the theoretical physicist who first applied this formula to probability. So random matrices come up naturally when we discuss conjugacy classes on the "uniform measure," but there are other contexts in which they come up as well:

Example 119

Suppose we have an observation (in multivariate statistics) which returns an N-dimensional vector, and suppose that we do M such observations and make an $N \times M$ matrix of those numbers (where each observation is a column). We can then compute the $N \times N$ **sample covariance matrix** given by $\mathcal{L} = XX^T$.

We often care about seeing whether we have independence in our statistical data, and studying the covariance matrix $\mathcal{L}$ is often a good way to understand that. A good starting point is to understand whether we see noise if our components of X are iid Gaussians:

Theorem 120 (Fisher-39, Hsu-39, Roy-39)

Let $M \geq N$ (more observations than coordinates), and suppose that the matrix elements of X are standard normal random variables. Then the distribution of the N ordered eigenvalues $x_1 \geq \cdots \geq x_N$ of the sample covariance matrix XX^T has density (with respect to the Lebesgue measure) a constant times $\prod_{1 \leq i < j \leq N} (x_i - x_j) \prod_{i=1}^N x_i^{(M-N-1)/2} e^{-x_i/2}$.

The main differences from the previous result are that our points x_i now live on the real line, but we still have the Vandermonde determinant (the logarithmic potential) present as a "repulsion" between particles, and we also have an additional **confining potential** term that prevents our eigenvalues from going to 0 or ∞.

This probability measure is known as the **Wishart ensemble**, though that name sometimes also refers to the case where the matrix elements X_{ij} are just independent, rather than Gaussian (so that we won't have a nice formula in the same way), and also as the **Laguerre orthogonal polynomial ensemble**.

Remark 121. *Basically, the Laguerre orthogonal polynomials $p_n(x)$ are real-valued polynomials (each x^n plus lower-order terms), satisfying*

$$\int_{\mathbb{R}} p_n(x)p_m(x)x^a e^{-x} dx = \gamma_n \delta_{mn}.$$

But we won't talk too much about these polynomials here.

The relevance of this type of analysis became more relevant as computers started being able to process data for large matrices – the idea is to use **principal component analysis**, looking at whether the largest eigenvalue of $\mathcal{L}$ is much larger than the one we expect in the Wishart ensemble. If so, that corresponds to a correlation, and we then try to remove that component from the matrix and see if there is noise left (and so on).

And there is a third historically important source of random matrix theory, coming from nuclear physics (again, the range of topics here has to do with the "octopus" of this area of mathematics). The story here starts with something similar to that like the crystal structure of ice, and we'll discuss the story of **Eugene Wigner** next time!

16 April 15, 2021

Today's class is a seminar by **Mustazee Rahman**, titled **A random growth model and its time evolution**. In the first half, we'll introduce some random growth models, like the polynuclear random growth model, and its KPZ scaling limit and some particular results related to it. We'll then discuss some new results about the multi-time distribution for this growth model, and we'll analyze the proof and some of the technical details of that argument in the second half.

> **Example 122**
>
> Planar growth models come up in real life all the time – for example, when we have snowfall on a landscape (like Boston), the different layers of snow will sketch out a growing height interface, and we want to understand how this height evolves. (Other basic examples include bacteria growth in a petridish, a coffee stain on a surface, the front of burning paper, "sticky Tetris," and so on.)

The goal is to come up with a mathematical model that helps us study these, and we'll do so now with the polynuclear growth model:

> **Definition 123** (Polynuclear Growth / Last Passage Percolation)
>
> Start with the positive quadrant, and attach a random weight ω_{ij} at every point (i, j) which are iid exponential random variables of mean 1. Then we define the **growth function**
>
> $$G(m, n) = \max\{G(m - 1, n), G(m, n - 1)\} + \omega_{mn},$$
>
> with boundary conditions $G(0, n) = G(m, 0) = 0$ for all $n, m \geq 0$.

Our growth function G indeed "grows" in the up-right direction, and we can view it as a height interface by turning the quadrant by 45 degrees and labeling the axes with x and t. We then get a **height function**

$$H(x, t) = G\left(\frac{t - x}{2}, \frac{t + x}{2}\right)$$

for all x, t of the same parity and $|x| \leq t, t \geq 0$. If we write down the recursion for H in terms of the recursion for G, we find that

$$H(x, t + 1) = \max\{H(x - 1, t), H(x + 1, t)\} + \eta_{x,t+1},$$

where $\eta_{x,t+1}$ is an iid exponential random variable with mean 1. We want to understand (1) the limit shape, (2) the fluctuations around that limit shape, and (3) the time-evolution (how the shapes at different times are related).

For point (1), it turns out that (Rost, 1980s) the macroscopic behavior looks like

$$\lim_{T \to \infty} \frac{H(Tx, Tt)}{T} = \frac{(\sqrt{t - x} + \sqrt{t + x})^2}{2}$$

almost surely, and this gives us a circular profile (as we might expect). This type of limit shape theorem exists for general iid weights – we can prove this with subadditive ergodic theorems – but computing it exactly is difficult, and

we only know the answer for a handful of them.

We next turn to point (2), looking at the microscopic behavior, we want to find the fluctuations of the height interface around the limit shape. It makes sense to first ask whether we have a central limit theorem of the form

$$\frac{H(0, T) - 2T}{\sqrt{T}} \to N(0, \sigma^2)$$

(we're looking at the **one-point fluctuations** for $x = 0$, since we know the expected value of the height is $2T$). But this is not a situation where the height function is the sum of independent random variables, so we might expect instead that we have

$$\frac{H(0, T) - 2T}{T^\alpha} \to \chi$$

for some exponent α and random variable χ. In addition, we may also care about how the height function is correlated at different spatial points; we can quantify this by asking at what critical length scale $|x_1 - x_2| = T^\beta$ where we find nontrivial correlations between $H(x_1, T)$ and $H(x_2, T)$. Alternatively, we may care about how correlated the height function is at different temporal points $H(0, T t_1)$ and $H(0, T t_2)$, or how all of the fluctuations depend on the weight distribution ω_{ij}.

To answer many of these questions, we go back to the work of the 1980s – Kardar, Parisi, and Zhang studied the **KPZ equation**, which is a stochastic partial differential equation

$$\partial_t h = \partial_x^2 h + (\partial_x h)^2 + \eta,$$

where h is a function of x, t, and η is a spacetime white noise. (This is a continuum analog of the polynuclear growth height interface from above!) From that study, they predicted that we in fact have $\alpha = \frac{1}{3}$ and $\beta = \frac{2}{3}$, which are very different from what we normally see in probability theory! In other words, the limiting fluctuations are not normal, and if we rescale our function as

$$H_T(x, t) = \frac{H(xT^{2/3}, Tt) - 2tT}{(tT)^{1/3}},$$

we should have a distributional limit $\lim_{T \to \infty} H_T(x, t)$ to some random interface. In fact, it's believed that these limit fluctuations don't depend on the ω_{ij} distributions, as long as they're iid and behave appropriately – this is a **universality** conjecture, and it's part of a much broader conjecture about the KPZ class.

It turns out that $H_T(0, 1)$ converges to the **GUE Tracy-Widom distribution** $H(0, 1)$, which appeared first in random matrix theory: it's the scaling limit of the largest eigenvalue of a GUE matrix. So the connection here was surprising, and the important thing about the GUE Tracy-Widom distribution is that its cdf is given by a **Fredholm determinant**:

$$\mathbb{P}(H(0, 1) \le a) \le \det(I - K_{\mathsf{Ai}})_{L^2(a, \infty)},$$

where

$$K_{\mathsf{Ai}}(u, v) = \int_0^\infty d\lambda \, \mathsf{Ai}(\lambda + u)\mathsf{Ai}(\lambda + v)$$

And even before we take the limit of H_T, it turns out that $H_T(0, 1)$ has the distribution of the largest eigenvalue of a $T \times T$ matrix from the **Laguerre ensemble**.

Soon after this, it was found that the spatial fluctuations converge as

$$H_T(x, 1) \to \mathsf{Airy}(x)$$

to the **Airy process**. This is because the finite-dimensional distributions of the Airy process are determined by Fredholm determinants:

$$\mathbb{P}(\mathsf{Airy}(x_k) \le a_k \quad \forall k = 1, \cdots, p) = \det(I - K_{\mathsf{Ai}}^{\mathsf{ext}}),$$

where $K_{\text{Ai}}^{\text{ext}}$ is an extended Airy kernel acting on the space $L^2(\{x_1, \cdots, x_p\} \times \mathbb{R})$. And this is because we can see the distribution of $H_T(x, 1)$ as a marginal of a determinantal process on Gelfand-Tsetlin patterns, and the Fredholm determinant appears naturally – we can then get a result by looking at the kernel of that determinant.

More recently, people have identified the spacetime fluctuations for H: for example, the finite-dimensional distributions of $H(x_1, t_1), \cdots, H(x_p, t_p)$ have been computed as contour integrals of Fredholm determinants. But we can also describe the limiting interface using the KPZ fixed point (a Markov process on the space of the interfaces), or as a random metric in a directed landscape. But we'll focus on the finite-dimensional distributions for now, and we'll do so by first looking at some combinatorics.

<div style="border: 2px solid red; padding: 10px;">

Definition 124

For any permutation σ, let $L(\sigma)$ be the length of the longest increasing subsequence of σ (for example, $L(531246) = 4$).

</div>

If we look at a random permutation σ, and define

$$L_N(t_k) = L(\sigma(1), \sigma(2), \cdots, \sigma(t_k N))$$

for times $0 < t_1 < t_2 < \cdots < t_p \leq 1$, then it turns out (after a lot of work) that the rescaled joint random variables scale as

$$\frac{L_N(t_k) - 2\sqrt{t_k N}}{(t_k N)^{1/6}} \to H(0, t_k).$$

So if we now want to think about the finite-dimensional distribution formula for H, it turns out that

$$\mathbb{P}(H(x_1, t_1) \leq a_1, \cdots, H(x_p, t_p) \leq a_p)$$

can be written as a $(p-1)$-fold contour integral

$$\frac{1}{(2\pi i)^{p-1}} \oint \cdots \oint d\theta_1 \cdots d\theta_{p-1} \frac{\det(I + F(\theta))_{\mathcal{H}}}{\prod_j (\theta_j - 1)},$$

where our contours over θ_j are circular with radius $R > 1$ and $\theta = (\theta_1, \cdots, \theta_{p-1})$. Here,

$$F(\theta) = [F_{i,j}(\theta)]_{1 \leq i \leq j \leq p},$$

so that this object acts on p copies of $L^2(0, \infty)$ (each as an **integral kernel**), and the kernel can be written as

$$F_{i,j}(\theta) = \sum_k c_k(\theta) F_{i,j}^{(k)}.$$

where there are roughly 2^p terms in this sum, defined in a particular combinatorial way, so that each $c_k(\theta)$ is a Laurent polynomial in the θ_is, and the basic kernels $F_{i,j}^{(k)}$ are basically like the Airy kernels: as an integral kernel, we can write

$$F_{i,j}^{(k)}(u, v) = \int_0^\infty \cdot \int_0^\infty d\lambda_1, \cdots, d\lambda_s A_1(u, \lambda_1) A_2(\lambda_1, \lambda_2) \cdots A_s(\lambda_s, v)$$

for basic kernels

$$A[t, x, a](\lambda, \nu) = t^{-1/3} \text{Ai}\left(x^2 + a + t^{-1/3}(\nu - \lambda)\right) e^{2x^3/3 + xa + xt^{-1/3}(\nu - \lambda)}.$$

Remark 125. *This means that in our contour integral above, we only have poles at 0 and 1 – the poles at 1 come from the denominator $\theta_j - 1$, and the poles at 0 come from the Laurent polynomials in the θ_is.*

We'll now go to the second part of the talk, in which we understand how the computations work. We start with

polynuclear growth, and our goal is to find the p-time probability distribution

$$\Pr = \mathbb{P}\left(G(m_k, n_k) \leq a_k \quad \forall k = 1, \cdots, p\right)$$

for $n_1 < \cdots < n_p$ and $m_1 < \cdots < m_p$. (We can then take a limit to get from this Pr to the distribution of H_T and then to H.) We'll do this computation in three steps:

- First, we express Pr as a contour integral of some $N \times N$ matrix (a finite matrix instead of a Fredholm determinant).

- From there, we use an **orthogonalization procedure** to represent this as a Fredholm determinant $\det(I + F_N(\theta))$ by using row and column operations, so that $F_N(\theta)$ has a $p \times p$ block structure (so we can embed things as a kernel into our Hilbert space $\mathcal{H}$), and it has nice properties under KPZ scaling.

- Finally, we can write the entries of $F_N(\theta)$ of contour integrals, and we can then use steepest descent analysis to get asymptotics.

We'll focus on the second and then the first of these bullet points. For orthogonalization, we note that the entries of $L_N(\theta)$ are given by the contour integrals

$$L_N(\theta; i, j) = \frac{1}{(2\pi i)^p} \oint \cdots \oint dz_1 \cdots dz_p \mathcal{G}(z_1, \cdots, z_p|(\text{parameters})),$$

(for some very complicated function $\mathcal{G}$, coming from generating functions of certain derivative operators) where things depend on the parameters $i, j, n_k, m_k, a_k, \theta$, our poles being only at $z_j = 0, 1$, and circular contours of integration with radius at least 1. From there, we separate out the pole where we have $z_j = 0$ for all j, so that we can write $L_N(\theta; i, j) = A(i, j) + B(i, j)$, where $A(i, j)$ is the same integral but over contours $|z_j| = r$ instead.

It turns out that A is always upper triangular, and it always has 1s on the diagonal, so the determinant of A is 1 and we can write $\det(L_N(\theta)) = \det(I + A^{-1}B)$.

Now, we can explicitly compute A^{-1} – it turns out to take on a form very similar to the same integral as above but with $\frac{1}{\mathcal{G}}$ instead of $\mathcal{G}$ (and some parameters modified), due to symmetries in the function. And to figure out the matrix B, we write it out as a sum of residues over all $2^p - 1$ poles (where each z_i is either 0 or 1 but not all 0). We can then describe the situation in a combinatorial way, and many of the residues turn out to be zero: it turns out that the θ-dependence factors in a way to get a sum over basic kernels $F_N^{(k)}$

$$A^{-1}B = \sum_k c_k(\theta) F_N^{(k)},$$

where the sum over k and the functions $c_k(\theta)$ are the same as in the limit result above, and the $F_N^{(k)}$s converge to $F^{(k)}$s.

So now we can turn to the determinantal expression for the probability and the contour integral: the idea is to take our polynuclear growth, look at a column m, and look at the vector-valued process

$$\vec{G}(m) = (G(m, 1), G(m, 2) \cdots, G(m, N)),$$

which takes values in the Weyl chamber $\mathbb{W}_N = \{(x_1, \cdots, x_N) \in \mathbb{R}^N : x_1 \leq \cdots \leq x_N\}$. Because the process here is Markovian in N (by definition and the recurrence relation for G), we want to write down the Markov transition probabilities, and the transition kernel was found to be

$$\mathbb{P}\left(\vec{G}(m) \in [y, y + dy] \mid \vec{G}(0) = x\right) = \det(D^{j-i} w_m(Y_j - x_i))_{i,j} \, dy,$$

where the determinant is of an $N \times N$ matrix, D is the derivative operator (and thus D^{-1} is the integral operator $\int_{-\infty}^{x} f(y)\,dy$), and we have the density

$$w_m(x) = e^{-x}\frac{x^m}{m!}1_{x\geq 0, m\geq 1}$$

(the density of the sum of the m exponential random variables, which is the Gamma density). (We can derive this formula for the transition kernel from the RSK algorithm.) But now that we have this Markovian relation, the Pr probability that we want to calculate is now a p-fold integral

$$\Pr = \int_{\mathbb{W}_n} dx^{(1)} \cdots \int_{\mathbb{W}_n} dx^{(p)} \prod_{k=1}^{p} \mathbb{P}\left[\vec{G}(m_k) \in [x^{(k)}, x^{(k)} + dx^{(k)} \mid \vec{G}(m_{k-1}) = x^{(k-1)}\right] \cdot 1_{x_{n_k}^{(k)} \leq a_k},$$

which simplifies to

$$\int_{\mathbb{W}_n} dx^{(1)} \cdots \int_{\mathbb{W}_n} dx^{(p)} \prod_{k=1}^{p} \det\left(f_k(x_j^{(k)} - x_j^{(k-1)}, i, j)\right) 1_{x_{n_k}^{(k)} \leq a_k}$$

for the same $f_k(x, i, j) = D^{j-i} w_{m_k - m_{k-1}}(x)$. (Here, remember that the xs are still N-dimensional vectors, so our determinants are still $N \times N$.) And now we can do the integral over the variable $x^{(p)}$, since

$$\int_{\mathbb{W}_N} dx \det(D^{j-i} w_i(x_j)) 1_{x_N \leq a}$$

can be done column-by-column to get

$$= \det(D^{j-i-1} w_i(a)).$$

So our p-fold integral becomes a $(p-1)$-fold integral, and the remaining product is complicated – we want to simplify it so that our entries $f_k(x, i, j)$ in our determinants only depend on x, with the exception that f_1 can also depend on i and f_p can also depend on j, and also suppose that there are no indicators. Then the Cauchy-Binet identity iteratively gives us an expression of the form

$$= \det\left(\int_{\mathbb{R}^{p-1}} dx f_1(i, x_1) f_2(x_1, x_2) \cdots, f_p(x_{p-1}, j)\right),$$

and that's what we'll aim for. To do that, we need to prove a determinantal identity

$$\int_{\mathbb{W}_N} dx \det(D^{j-a_i} f(i, x_j)) \det(D^{b_j-i} g(-x_i, j)) 1_{x_n \leq a} = \int_{\mathbb{W}_N} dx \det(D^{n-a_i} f(i, x_j)) \det(D^{b_j-n} g(-x_i, j)) 1_{x_n \leq a}$$

(so that we're removing the explicit i- and j-dependence). This is true if f and g are functions that vanish for sufficiently negative x, so that we can use integration by parts and move derivatives between columns of our matrix. Applying this gives us the correct form that we desire, but we still have some extra indicators $1_{x_{n_1} \leq a_1}, \cdots, 1_{x_{n_{p-1}} \leq a_{p-1}}$. And we deal with the indicators by using contour integrals: notice that for x in the Weyl chamber,

$$x_n \leq a \iff \sum_{j=1}^{N} 1_{x_j \leq a} \geq n$$

because our components are ordered, and then we can present this as a contour integral because we have the identity

$$1_{\ell \geq 0} = \frac{1}{2\pi i} \oint_{|\theta| = R > 1} d\theta \frac{\theta^{\ell}}{\theta - 1}.$$

So indeed, we get our p-fold contour integral, and the using Cauchy-Binet (as mentioned above) gives us a $p-1$-fold contour integral, as desired.

17   April 22, 2021

Today's class is a seminar by **Jonas Arista**, titled **Exit distributions associated with loop-erased walks and random matrices**. We'll start with some motivating background of the one-dimensional case:

> **Example 126**
>
> Consider a system of n independent Brownian motions conditioned not to intersect in the time interval $[0, t]$. This is equivalent to having an n-dimensional Brownian motion started at some point $(x_1, \cdots, x_n)$ and conditioned on staying in the Weyl chamber C.

The well-known **Karlin-McGregor formula** gives us the unnormalized transition density q_t given by

$$q_t(x, y) = \det_{n \times n}(p_t(x_i; y_j)) \quad \forall x, y \in C.$$

This is good because we like working with determinants, and it is also important that the transition density is given by a matrix with entries given by **unconstrained one-dimensional Brownian motion**. With this, we can now figure out the distribution of the n particles at time t if we start at $(0, 0, \cdots, 0)$: since the origin is not part of the chamber C, we can do this by taking an appropriate limit

$$\lim_{x_i \to 0, x_i \in C} \frac{1}{M_x} \det_{n \times n}(p_t(x_i; y_j)) = \frac{1}{M} e^{-\frac{1}{2t} \sum_{i=1}^n y_i^2} \prod_{1 \le i < j \le n} (y_i - y_j).$$

The important point here is that the right-hand side is the joint density of the eigenvalues in the **Gaussian Orthogonal Ensemble** (GOE), taking $\beta = 1$: basically, this density is the density for the eigenvalues of a random real symmetric matrix.

The way that we prove this determinant formula is to use the Feynman-Kac formula to show that the determinant solves the heat equation with appropriate boundary conditions, or to use a general method coming from the (classical) **reflection principle**.

> **Example 127**
>
> Suppose we have $n = 2$, so we want the determinant $\det \begin{bmatrix} p_t(x_1, y_1) & p_t(x_1, y_2) \\ p_t(x_2, y_1) & p_t(x_2, y_2) \end{bmatrix} = p_t(x_1, y_1) p_t(x_2, y_2) - p_t(x_1, y_2) p_t(x_2, y_1)$.

The idea is to write this as a sum of two terms

$$= \mathbb{P}(x_1 \to y_1, x_2 \to y_2) - \mathbb{P}(x_1 \to y_2, x_2 \to y_1)$$

for unconstrained Brownian motion, and now we can rewrite this as follows: if our paths in the first and second terms intersect, we can take their first intersection point and swap the paths after that point on. This then gives us an involution ϕ on paths that is mass-preserving, so that

$$\mathbb{P}(x_1 \to y_1, x_2 \to y_2, \text{paths intersect}) = \mathbb{P}(x_1 \to y_2, x_2 \to y_1, \text{paths intersect}),$$

and thus we can subtract this term off from the two terms above to get

$$q_t(x, y) = \mathbb{P}(x_1 \to y_1, x_2 \to y_2, \text{no intersection}) - \mathbb{P}(x_1 \to y_2, x_2 \to y_1, \text{no intersection}),$$

and now the second term is zero because paths must cross by continuity if $x_1 < x_2$ but $y_2 > y_1$. So the blue path is all that's left, and that gets us towards the formula that we mentioned above!

But the problem we're considering here is that we want to consider **two-dimensional planar processes** instead of one-dimensional Brownian motions. Then a few things become different from the one-dimensional case: most notably, paths can have loops, so they may intersect in space. (It even makes sense to have the traces of paths intersect, rather than just for the paths to hit each other at the same time.) But a key point is that Karlin-McGregor no longer works, since our traced out paths should not have loops. But there is indeed a generalization for **discrete planar processes**, which we'll describe now:

Theorem 128

Let $\Omega^\sharp$ be a discretization of a planar domain $\Omega \subset \mathbb{C}$, and let $\xi_1, \cdots, \xi_n$ be n independent simple random walks on $\Omega^\sharp$. Then the transition probability

$$\mathbb{P}(\xi : x_i \to y_i, \text{non-intersection}) = \det_{n \times n} h(x_i, y_j),$$

where x_i are points in the domain, y_i are exit points on the boundary, and $h(x, y) = \mathbb{P}_x(\xi_{T_{\Omega^\sharp}} = y)$ are given by hitting probabilities (this is the discrete Poisson kernel).

Here, the non-intersection condition looks like

$$\xi_j \cap LE(\xi_i) = \varnothing \quad \forall j > i,$$

where $LE(\xi_i)$ denotes the **loop-erased random walk** of ξ_i, which is the **chronological** loop-erasure from x_i to y_i. (Notably, chronological loop-erasure keep the initial and final point invariant.) The reason for this condition in the theorem above is that it allows us to **generalize the reflection principle**: the problem when we have loops is that we if we naively do the ϕ reflection from above, we won't obtain the original paths after applying the reflection twice. But if we instead make sure that ξ_i has loops erased, and then we consider the **first intersection along the loop-erased part of** ξ_i, we can indeed define an involution just like in the 1-D case, and then everything else works as we want.

Remark 129. *We can understand the symmetry that goes on here if we look at things in a different context, namely the* **(wired) uniform spanning tree***, meaning that we have a spanning tree where the whole boundary is considered as a single point. By* **Wilson's algorithm** *(which generates a uniformly spanning tree), we find that*

$$\mathbb{P}(\xi_i : x_i \to y_i, \xi_j \cap LE(\xi_i) = \varnothing \forall j > i) = \mathbb{P}(n \text{ branches of uniform spanning tree } x \to y).$$

and notably this doesn't depend on the order of the vertices of the tree 1 *through n.*

We can now see the connection with random matrices: if we let Ω be a simply connected domain in $\mathbb{C}$, and we let $x_1, \cdots, x_n, y_1, \cdots, y_n \in \partial\Omega$ be on the boundary $\partial\Omega$. If we now consider a discretization $\Omega \cap \delta\mathbb{Z}^2$ for some small δ, then we know from above that

$$\det_{n \times n}(h^\delta(x_i^\delta, y_j^\delta)) = \mathbb{P}(\xi_i^\delta : x_i^\delta \to y_i^\delta, \xi_j^\delta \cap LE(\xi_i^\delta) = \varnothing \quad \forall j > i),$$

and we can now take a **scaling limit** of both sides. For the left-hand side, we can show that determinants of the discretized Poisson kernel converge to the appropriate Poisson kernels (the **excursion Poisson kernels** of Brownian motion. But for the right-hand side, the calculations are more complicated: we know that simple random walks converge to Brownian motion, and we know that the loop-erasure of simple random walk converges to the **SLE(2)** process (SLE stands for **stochastic Loewner evolution**). But having n such paths with the non-intersecting conditions requires us

to think about having "multiple SLE," which can be more complicated. For our purposes, though, we can compute the distribution of the **exit points** more easily, because the start and end points don't change under loop-erasure, and we will be led to random matrix ensembles in this way.

First of all, we know that the excursion Poisson kernel in the positive xy-quadrant is

$$h(x, y) = \frac{4xy}{\pi(x^2 + y^2)^2}$$

(by using the reflection principle and the results known for the upper half-plane), and then a normalization gives us

$$\lim_{x_i \to 1} \frac{1}{M_x} \det_{n \times n}(h(x_i, y_j)) = \frac{1}{M} \prod_{1 \leq j \leq n} \frac{y_j}{(1 + y_j^2)^{n+1}} \prod_{1 \leq i < j \leq n} (y_i^2 - y_j)^2.$$

We can do something similar for the half-unit disk $|x| < 1, 0 < \theta < \pi$ as well, and then we find that

$$h(x, \theta) = \frac{1}{\pi} \frac{(1 - x^2) \sin \theta}{(1 - 2x \cos \theta + x^2)^2},$$

so that

$$\lim_{x_i \to 0} \frac{1}{M_x} \det_{n \times n}(h(x_i, \theta_j)) = \frac{1}{M} \prod_{1 \leq j \leq n} \sin(\theta_j) \prod_{1 \leq i < j \leq n} (\cos \theta_i - \cos \theta_j).$$

(It makes sense that these are related, because we can go from the positive quadrant to the half-unit disk by a conformal transformation.)

But the next question we can ask is how to deal with domains Ω that are **not simply connected**, such as the annulus.

Example 130

Suppose we start with n starting values on the inside boundary of the annulus, and we want the distribution of hitting points θ_i on the outer boundary when the inner radius goes to 0.

The idea here is to add a "zipper" and unfold the annulus, so that we get an infinite strip with periodic boundary conditions, and now we can find an affine version of the above formulas by looking at translations: the analogous calculation is to look at

$$\mathbb{P}\left(\xi_i^\delta : \nu_i^\delta \to T_j^m \theta_j^\delta, m \in \mathbb{Z}, \xi_j^\delta \cap LE(\xi_{j-1}^\delta) = \varnothing \quad \forall 1 < j \leq n, \quad \xi_1^\delta \cap LE(T_\delta \xi_n) = \varnothing\right)$$

$$= \sum_{\gamma \in S_n} \sum_{k_1 + \cdots + k_n = 0 \bmod n} \text{sgn}(\gamma) \prod_{i=1}^n h^\delta(\nu_i^\delta, T_\delta^\kappa \theta_{\gamma(i)})$$

(notice that we need to make sure that the nth path doesn't interact with the first). The idea with the sum being 0 mod n is that cyclic shifts of our ending points are also valid, and this expression turns out to also be a determinant:

$$= \det\left(\sum_{k \in \mathbb{Z}} e^{i2\pi x k} h^\delta(\nu_i^\delta, T_\delta^\kappa \theta_j^\delta)\right),$$

where x is 0 if n is odd and $\frac{1}{2}$ otherwise. And now if we take $\delta \to 0$, and we take a strip of radius $0 < r < 1$, we get the excursion Poisson kernel of Brownian motion in the strip:

$$h(\nu, \theta) = \frac{\pi}{4|\log r|^2} \text{sech}^2\left(\frac{\pi}{2|\log r|}(\theta - \nu)\right).$$

This leads us to the final result, agreeing with the eigenvalue distribution for the **COE (circular orthogonal ensemble)**:

> **Theorem 131**
>
> For any $\overline{\nu}, \overline{\theta} \in \mathbb{C}$, we have (taking the width of the strip to zero)
>
> $$\lim_{r \to 0} \frac{1}{M_{\overline{\theta}}} \det_{n \times n} \left(\sum_{k \in \mathbb{Z}} e^{i2\pi x k} h(\nu_i, \theta_j + 2\pi k) \right) = \frac{1}{M} \prod_{1 \le i < j \le n} |e^{i\theta_i} - e^{\theta_j}|.$$

And we can now replace the annulus with other domains that are not simply connected, and this will lead us to other Poisson kernels and other kinds of ensembles.

18 April 27, 2021

Last lecture, we started talking about random matrix theory – we mentioned two natural sources for random matrices, namely Weyl's integration formula (the projection of the Haar measure on the unitary group onto the eigenvalues on the N-dimensional torus) and multivariate statistics (thinking about the sample covariance matrix XX^t and looking at the distributions of the ordered eigenvalues).

But a third source comes from physics, and it gathered more attention than the first two in the probability world: back in the 1950s, nuclear physics was a hot topic, and in particular many experiments were being done on these nuclei.

> **Example 132**
>
> One measurement that people were curious about was **neutron resonance** – we take a heavy nucleus and bombard it with neutrons of low energy. The atom will then hold the neutron for a short amount of time, but due to instability the neutron will come out with a photon. We can then measure the frequency of that outgoing photon and look at the spectrum of frequencies.

What was find is that the emitted light has certain **spectral lines** corresponding to different energy levels. For small atoms (like hydrogen), it is easy to predict the location of those spectral lines using quantum theory, but it is much more difficult to do so for larger atoms like uranium – we need to investigate the spectrum of a complicated Hamiltonian operator.

But we can still do the experiment physically, and one measurement that was done was to calculate the nearest-neighbor spacing between the (≈ 100) energy lines. If these spectral lines were independently and uniformly distributed on some interval, then the nearest-neighbor distances will approximate an exponential (this is the **Poisson approximation**, since we have a Poisson point process). But the actual shape seen was very different, and thus an explanation is needed.

The physicist Eugene Wigner resolved this with the following logic: if the spectral levels are eigenvalues of the Hamiltonian, essentially a large matrix which is too complicated to calculate, we'll assume that the matrix is random and see what happens. Since the Hamiltonian needs to be self-adjoint, we need the random matrix to be symmetric or Hermitian, but in addition we need other conditions so that (for example) the Schrodinger equation is time-reversible. Furthermore, the Hamiltonian should be rotationally-symmetric in the degrees of freedom (to avoid biasing in a particular reference frame). So imposing all of these conditions forces us into the following setup:

> **Definition 133**
>
> The **Gaussian orthogonal ensemble** or **GOE** is a probability measure on real symmetric matrices X with density proportional to $e^{-c \cdot \text{tr}(X^2)}$ with respect to the Lebesgue measure.

Here, because we have a symmetric matrix, we have

$$\text{tr}(X^2) = \sum_{i,j} x_{ij} x_{ji} = \sum_{i=1}^{N} x_i^2 + \sum_{i>j} 2x_{ij}^2.$$

In other words, we pick each coordinate x_{ii} on the diagonal to be normal with some variance, and we pick x_{ij} to be normal with some other variance. This turns out to be basically the only measure that is under the orthogonal matrices (rotations).

While Wigner was not able to find the nearest-neighbor distance distribution, he did manage to find an approximation for it which looked something like xe^{-x^2}, and this gave an answer that matched the physical experiments pretty well! So replacing the Hamiltonian with a random matrix provided a reasonable answer.

Other early pioneers of the field asked questions like the joint distribution or the correlations between level lines, and Dyson and Mehta were after these particular questions.

Remark 134. *In an interview, Dyson mentioned that there are new applications to the theory because of simulated systems (larger than real-world atoms) that can be calculated on a computer. And he also mentioned that the distribution of zeros of the Riemann zeta function also have connections to random matrices, though the correspondence is still unproven.*

The first objects that we study in random matrix theory are the relevant determinants, and we'll use complex instead of real matrices for this result:

Theorem 135

Let X be an $N \times N$ complex matrix with iid entries distributed as $N(0,1) + iN(0,1)$. Then the distribution of eigenvalues $x_1 \geq x_2 \geq x_N$ for the self-adjoint matrix $Y = \frac{1}{2}(X + X^*)$ has density proportional to $\prod_{i<j}(x_i - x_j)^2 \prod_{i=1}^{N} e^{-x_i^2/2}$.

The matrix Y is then sampled from the **Gaussian unitary ensemble** or **GUE**, and it turns out to be mathematically simpler than the GOE. Again, we see the Vandermonde determinant squared in the GUE – a similar squared determinant shows up in the statistical setup described at the beginning of class, while the GOE has a single copy of the determinant.

If we look at the submatrices of Y that are formed by **nested corners**, meaning that we consider the top left 1×1 matrix Y_1, top left 2×2 matrix Y_2, and so on, we get a sequence $\{Y_1, Y_2, \cdots, Y_N = Y\}$, and we have eigenvalues for each Y_i. If we let $x_1^j \geq x_2^j \geq \cdots \geq x_j^j$ be the eigenvalues of Y_j for each j, then these sets of eigenvalues always interlace (this is a linear algebra exercise which is good for us to check): we have the deterministic statement

$$x_i^{j+1} \geq x_i^j \geq x_{i+1}^{j+1}$$

for any matrix Y. This should be reminiscent of the picture we had earlier with the lozenge tilings with its interlacing coordinates for vertical lozenges, and we can in fact draw a connection:

Theorem 136

Let $\ell_i^j(A, B, C)$ describe the positions of the jth vertical lozenges in the ith level of a lozenge tiling for an $A \times B \times C$ hexagon (where A is the horizontal length and B is the length along the direction 120 degrees counterclockwise to A). Then for a uniform random tiling of a regular hexagon of length L, we have

$$\frac{\ell_i^j(L, L, L) - \frac{L}{2}}{\sqrt{\frac{3}{8}L}} \xrightarrow{d} \text{eigenvalues } x_i^j \text{ for GUE corners.}$$

When we take our hexagon to be regular as described, we get six tangency points, and thus there are six processes that we can consider. The random matrix model can then describe a piece of the behavior for the lozenge tiling, and that's a common theme that we'll see.

The familiar expression $\prod_{i<j}(x_i - x_j)^\beta \prod_{i=1}^N w(x_i)$ has had exponent $\beta = 1, 2$ in all of our examples so far. But we can also define random matrices over quaternions to get $\beta = 4$, and thus $1, 2, 4$ correspond to objects over $\mathbb{R}, \mathbb{C}, \mathbb{H}$ respectively. It is often difficult to work with a more general $\beta > 0$, but we do have meaningful objects that come out of that study: the probability measures we get are called **log-gas**, because we can write them in the form

$$e^{\beta \sum \ln(x_i - x_j) + \ln w(x_i)} = e^{\beta H(x)},$$

where $H(x)$ is the **Hamiltonian** (or energy function) with a logarithmic interaction term, β is the familiar inverse temperature from statistical mechanics, and $\ln w(x_i)$ is a repulsion potential term. In a physical situation, then, it makes sense to not just define our systems for two particular values β. We can't look for random matrices over fields that are not $\mathbb{R}$, $\mathbb{C}$, or the quaternions, so instead we need a different strategy:

Theorem 137 (Dumitriu, Edelman 2002)

Let $a(n)$ and $b(n)$ be two sequences of random variables (for $n \geq 1$), and define a tridiagonal symmetric matrix via $M_{nn} = a(n), M_{n,n+1} = b(n)$ (so that our matrix is fairly constrained). If we then take each $a(n)$ to be independent $N(0, \frac{2}{\beta})$, and $b(n) = \frac{1}{\sqrt{\beta}}\chi\beta N$, then we have the limiting distribution

$$\chi_x \sim \frac{2^{1-x/2}}{\Gamma(x/2)}y^{x-1}e^{-y^2/2}dy \quad \forall y > 0,$$

and the eigenvalue density is proportional to $\prod_{i<j}(x_i - x_j)^\beta \sum_{i+1}^N e^{-\beta x_i^2/4}$.

But the result of arrays of eigenvalues for this tridiagonal matrix will not be very exciting: the dimensions of noise before the matrix is only-one-dimensional, so we won't get the same level of complexity with the eigenvalues as before. Our next step is to add time back into the picture: specifically, we replace the normal distribution $N(0, 1)$ for a given entry X_{ij} with a Brownian motion $N(0, \sqrt{t})$ (so that we do still have dependence in the entry between different times). The resulting evolution of the eigenvalues is called **Dyson Brownian motion**, and we can describe it using the stochastic partial differential equation

$$dW_i^N(t) = \sum_{\ell \neq i} \frac{1}{W_i^N(t) - W_\ell^N(t)} + dB_i^N, \quad 1 \leq i \leq N.$$

There is now a connection to the $(2+1)$-dimensional growth model that we described a few weeks ago: the bottom particle of that growth model, when **scaled diffusively**, gives us a Brownian motion (because instead of having a random walk, we scale it to large time and space). The two particles above it then evolve as Brownian motions conditioned to reflect off the particle on the bottom:

And similarly, level 3 particles do three independent Brownian motions, conditioned to reflect off the locations of the level 2 particles.

> **Fact 138**
>
> It turns out that if we restrict our attention to the level N particles, that gives us a Dyson Brownian motion of order N (in the diffusive limit). And if we take a fixed time snapshot at infinite time, we get a sample of the GUE corners process by looking at the first N levels.

But it turns out that if we don't restrict to a single time moment or a single limit, the correspondence with Brownian motions breaks down! So the situation can be a bit subtle. Next time, we'll see how random matrix theory connects more generally to representation theoretic models.

19   April 29, 2021

Today's class is a seminar by **Roger van Peski**, titled **Lozenge tilings and the Gaussian free field on a cylinder**. We'll give an expository description of what's known about random lozenge tilings and the Gaussian free field, and then we'll discuss some new work done on the cylinder (in which we see the Gaussian free field but also some discrete Gaussian corrections).

As we've seen in lecture, we can define **plane partitions** in various ways – two of them are as weakly decreasing arrays of nonnegative integers and as stacks of boxes forming a three-dimensional surface. And as seen previously in lectures, we can replace the plane partitions with other jagged "back walls" and get different three-dimensional pictures.

When we talk about lozenges, we often associate a **height function** which describes the stack of cubes at a given point in the xy-plane. What we're curious about is how the height function of a random tiling looks in the limit, and to answer that, we need to first pick how we pick a random tiling. If we have a finite domain, we can pick a uniformly random tiling (because there are finitely many plane partitions), or we can tile with measure proportional to q^{vol} for some $q \in (0, 1)$, which gives us a measure even for infinite domains.

To understand how limit shapes work, the following setup is useful:

> **Example 139**
>
> Suppose we have a simple random walk $Z_t : 0 \le t \le T$ conditioned to end at X (equivalently, we take the uniform measure on paths from $(0, 0)$ to (T, X)).

We know that as $X, T \to \infty$ with slope $\frac{X}{T}$ constant, our random walk's height will concentrate around a line of constant slope γ. One way to explain why this is is that the number of N-step random walks that end at a point γN

can be computed as a binomial coefficient with Stirling's approximation to be about $e^{N\left(-\frac{1+\gamma}{2}\log\frac{1+\gamma}{2}-\frac{1-\gamma}{2}\log\frac{1-\gamma}{2}\right)}$, where the term in parentheses gives us the Shannon entropy $h\left(\frac{1+\gamma}{2}\right)$. So we expect approximately

$$e^{\frac{T}{2}h\left(\frac{1+n}{2}\right)+\frac{T}{2}h\left(\frac{1+2\gamma-n}{2}\right)}$$

walks to go through a point $\left(\frac{T}{2}, n\frac{T}{2}\right)$, and now because the Shannon entropy is concave, this is maximized when $n = \gamma$.

More generally, our setup is to take a sequence of tileable domains $D_N \subset \mathbb{R}^2$, and suppose that $\frac{1}{N}D_N \to D$ as $N \to \infty$. Then our goal is to show that for each (x, y), the height function converges deterministically as

$$\frac{h(Nx, Ny)}{N} \to h_{\text{limit}}(x, y).$$

Results like this (finding the almost-sure convergence to limit shapes) have been proved in quite a bit of generality for domino tilings and more generally doubly periodic bipartite dimer models, as well as for q^{vol} ordinary plane partitions and the uniform measure for a fixed volume.

Returning to the simple random walk, though, we will want to ask about fluctuations around the limit shape, and in fact they should converge to a Brownian bridge. In particular, this means that the covariance between two points should converge to the covariance between two points of a Brownian bridge

$$\text{Cov}(B_s, B_{s'}) = \min(s, s')(1 - \max(s, s')) = G(s, s'),$$

which is the **Green's function** for a Laplacian $\Delta = \frac{\partial^2}{\partial s^2}$ on $[0, 1]$ with 0 periodic boundary conditions. More explicitly, we can "recover using a convolution"

$$\Delta f(s) = g(s) \iff f(s) = -\int_0^1 G(s, s')g(s')ds',$$

and in fact we can define such functions on a more general two-dimensional domain: if $D \subset \mathbb{C}$ is simply connected, and we still have our Laplacian Δ, we can define a Green's function with the same expression.

> **Example 140**
> We have the Green's function $G(z, w) = -\frac{1}{2\pi}\log\frac{z-w}{z-\bar{w}}$ for the upper half-plane $\mathbb{H}$.

Notably, as $w \to z$, the numerator blows up (and thus we're seeing something different happening in two dimensions). But we still want to talk about the limiting object in the same way as we did for the Brownian motion and the random walk, so we will informally define a **Gaussian free field** on a domain D as a random Gaussian "function" with

$$\text{Cov}(\Phi(z), \Phi(w)) = G(z, w).$$

More formally, we can define the random distribution Φ in terms of test functions

$$\text{Cov}\left(\int_D f_1\Phi, \int_D f_2\Phi\right) = \int_{D\times D} f_1(z)G(z, w)f_2(w).$$

Indeed, if f_1, f_2 are close to delta functions, these two definitions look very similar in the framework of "finding the value of a Gaussian free field at a point."

Returning to our setup of tilings on a region, the following conjecture was made in 2005 by Kenyon-Okounkov:

<div style="border: 1px solid blue;">

Conjecture 141

The centered height function in the liquid region $\mathcal{L}$ for a tiling of a simply connected planar region satisfies

$$\sqrt{\pi}(h(Nx, Ny) - \mathbb{E}[h(Nx, Ny)]) \to \Phi \circ \zeta(x, y)$$

for some (not necessarily conformal) map $\zeta : \mathcal{L} \to \mathbb{H}$ as $N \to \infty$, where Φ is the Gaussian free field on $\mathbb{H}$.

</div>

Unpacking this notation, what this means at the level of covariances is that (letting $\overline{h}$ denote the height function)

$$\mathrm{Cov}(\overline{h}(Nx_1, Ny_1), \overline{h}(Nx_2, Ny_2)) \to G(\zeta(x_1, y_1), \zeta(x_2, y_2)).$$

In other words, we can take our liquid region in our tiling and map it to the upper half-plane, in which we know the Green's function explicitly. So in general, we can then take our two coordinates in our original tiling, plug them into ζ, look at the two points in the upper half-plane, explicitly find the value of the Green's function in $\mathbb{H}$, and use that result to find the covariance in our original tiling.

Remark 142. *Remember that the liquid region is defined in the limit domain, so the (x, y) on the right-hand side are some fixed points in the rescaled limit. Meanwhile, the (Nx, Ny) on the left-hand side are coordinates for the domains D_N (which we need to rescale by $\frac{1}{N}$ eventually).*

We can describe this map explicitly for uniform lozenge tilings – it turns out to be parameterized by the limit shape. Basically, for each $(x, y) \in \mathcal{L}$ in the liquid region of the limiting domain, the surface will have some slope, which is equivalent to the local proportions of the three different kinds of lozenges. Given those lozenge proportions, we can then define the complex number

$$\zeta(x, y) = z(x, y),$$

where z is the third vertex of a triangle with vertices at $0, 1, z$ in the complex plane, with angles equal to π times the proportion of the three lozenges. And this is indeed the coordinate that appears in the Kenyon-Okounkov conjecture for uniform tilings. (What's going on here is that **we want the entropy to be holomorphic**.) And if we use the q^{vol} measure instead of the uniform measure, the image of ζ gains an additional e^{-cx} term in addition to $z(x, y)$.

While results for limit shapes are pretty general, the results for fluctuations are currently only know at the level of special cases: we know how to deal with some polygonal domains, q^{vol} plane partitions, domains with no frozen regions, and (notably not simply-connected) the hexagon with a hole.

We'll now move to the cylinder, which is no longer a simply connected region: we'll consider the q^{vol} measure on a room with boundary conditions as shown below (with leftmost and rightmost lozenges identified):
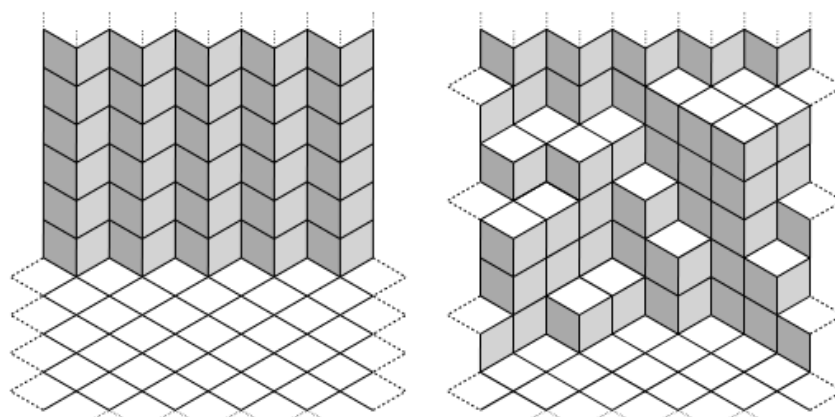
We'll then want to scale our system as follows: take the cylinder width to be $2N$ (it is naturally even based on the zigzag pattern of the wall), scale $q \to 1$ so that $q^N = t$ will remain constant, and then take $N \to \infty$. The result that we have is the following:

Theorem 143

In the setup above, we have

$$\frac{1}{N} h_N(N\tau, Ny) \to \begin{cases} 0 & y \leq \frac{\log 2}{\log t} \\ \int_{\log 2 / \log t}^{y} \frac{2 \arctan(\sqrt{4t^{-2u}-1})}{\pi} \, du & y \geq \frac{\log 2}{\log t}, \end{cases}$$

where y is "up" on the cylinder and τ is "around" the cylinder.

We know that the limit shape should be rotationally invariant, and when we go far down in the floor or high up on the wall, we expect very few lozenges. But we also want to know about the fluctuations around this limit shape:

Theorem 144

The centered height function $\sqrt{\pi}(h(N\tau, Ny) - \mathbb{E}[h(N\tau, Ny)])$ converges on the liquid region to the Gaussian free field $\Phi \circ \zeta$, using the Kenyon-Okounkov structure for the usual q^{vol} partitions.

(The map ζ sends the liquid region of the cylinder to a half-annulus in the upper half plane.) One important note is that is that there are tilings of the cylinder which only differ from the empty room in finitely many places (in terms of lozenges), but which we can't get by stacking boxes – in particular, we can shift every tile forward by one square, and thus there is the additional complication of **vertical shifts of cylindric partitions**:

$$\{\text{tilings}\} \text{ are in bijection with } \mathbb{Z} \times \{\text{cylindric partitions}\}.$$

So from the perspective of tiling models, the q^{vol} measure only works on cylindric partitions, and thus a notion of a **shift-mixed** q^{vol} measure is given by (S denotes the integer vertical shift and π denotes the cylindric partition)

$$\mathbb{P}(\pi, S) \propto (u^S q^{NS^2}) q^{\text{vol}(\pi)}.$$

This is a natural measure to impose because we can look at lozenge tilings as dimer models, and in particular we get a determinantal structure in this way.

It turns out that the shift-mixed measure still has the same limit shape, but we get a different situation for the fluctuations. Specifically, the centered height function $\overline{h}(N\tau, Ny)$ converges (on the liquid region) to $\frac{1}{\sqrt{\pi}}(\Phi \circ \zeta) - S\mathcal{H}'(y)$, where $\mathcal{H}$ is the limit shape and S is a **discrete Gaussian random variable** (an integer-valued random variable with mass function $e^{-C(x-m)^2}$). Here, the first term is the original non-shift-mixed limit shape, and the second term gives us some discrete shifts.

To understand more of where this comes from, we can switch to the hexagon-in-a-hole model (which is topologically equivalent to the cylinder). Note that the hole is fixed in location in the hexagon, but this does not mean the height of the hole is constant. So if we want to choose a uniformly random tiling, we can allow our hole height to vary, or we can condition our tiling as having a hole at a fixed height. The analogy we can then make is that fixing the height of the holey hexagon forces us to use **unshifted** cylindric partitions, while arbitary tilings of the holey hexagon are analogous to unrestricted tilings of the cylinder (allowing for vertical shifts in the cylinder).

It was shown that we get Gaussian free-field fluctuations in the Kenyon-Okounkov structure if we condition on a fixed hole height, but there are conjectures for general planar domains with holes. Basically, it is conjectured that the

limiting fluctuations for the **hole height** will be discrete Gaussian with parameters C, m, with C given by the **Dirichlet energy**

$$C = \frac{\pi}{2} \int_{\zeta(\mathcal{L})} ||\nabla g||^2 dx dy$$

for the unique harmonic function which is 0 on the outer boundary of our domain and 1 on the inner boundary (the hole). And indeed, our shift-mixed q^{vol} measure had

$$\mathbb{P}(S = x) \propto u^x q^{Nx^2},$$

and because $t = q^N$, this does indeed mean that S is distributed as a discrete Gaussian. We calculate that $C = \frac{|\log t|}{2}$, and this is in fact the Dirichlet energy in the conjecture that we just mentioned! So the pictures do line up on this point.

We'll end with some discussion about the proofs: q^{vol} plane partitions are distributed as a certain **Schur process** (measures on sequence of partitions), because we can view plane partitions as a sequence of integer partitions by looking at columns. It then turns out that the q^{vol} cylindric partitions are **periodic Schur processes**, which still have the nice properties of the usual Schur process, so we can use explicit properties and determinantal structure of the Schur process to get the joint moments (which we're trying to show converge to a Gaussian free field). And then what's left is an analysis of the asymptotics.

20   May 4, 2021

Today's class is a presentation by **Korina Digalaki**, titled **Evaluating Littlewood-Richardson coefficients via tilings**. This is an exposition based on a paper of the same title by Paul Zinn-Justin. We'll start by discussing Schur functions and Littlewood-Richardson coefficients, introducing the tiling model, Fock spaces, and transfer matrices, seeing how those come together, and relating this to other work.

Definition 145

Let λ be a partition. The **Schur functions** are defined via

$$s_\lambda(x_1, \cdots, x_n) = \frac{\det(x_i^{\lambda_j + n - j})_{i,j=1}^n}{\prod_{i<j}(x_i - x_j)}.$$

These Schur polynomials are characters of the polynomial irreducible representations of $GL(n)$, but we won't go into much detail in this direction. We can define these Schur functions alternatively as

$$s_\lambda(x_1, \cdots, x_n) = \sum_T \prod_{\text{box of } T} x_{T(b)},$$

where we vary T over all semi-standard Young tableaux of shape λ. Using this description, we can then also define a generalization, **semi-Schur functions**, via

$$s_{\lambda/\mu}(x_1, \cdots, x_n) = \sum_T \prod_{\text{box of } T} x_{T(b)},$$

where μ must be contained inside λ in terms of Young diagrams.

We'll now define the Littlewood-Richardson coefficients in three different ways: the first uses the fact that Schur functions form a basis for the symmetric functions.

It turns out that $c_{\mu,\nu}^{\lambda}$ are also the number of **Littlewood-Richardson tableaux** of skew shape λ/μ and weight ν, which means that the numbers filled in the tableaux must appear with multiplicity given by ν, and also that if we read the numbers from top to bottom, the number of 1s should be at most the number of 2s, which should be at most the number of 3s, and so on.

And we also have a third characterization, which we'll describe momentarily: recall that given a Young diagram, rotating by 45 degrees and assigning colored dots based on whether the diagram goes up or down gives us a sequence. And we can in fact compute our LR coefficients with the diagram below:
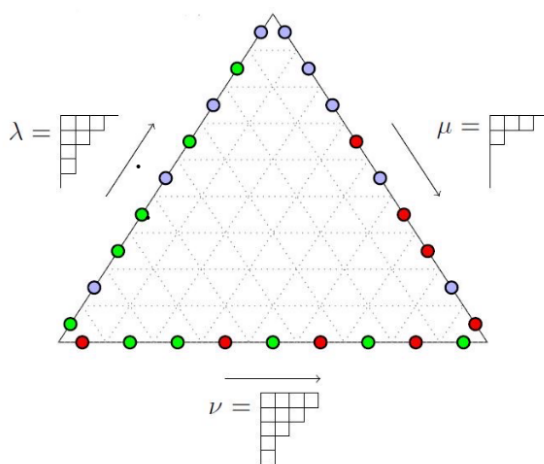

This shape is called a **puzzle**, and the LR coefficient $c_{\mu,\nu}^{\lambda}$ is then the number of ways to fill the triangle with tiles, in a way that forms paths that join red and green dots together. The set of allowed tiles, as well as an example of an allowed tiling, is below:
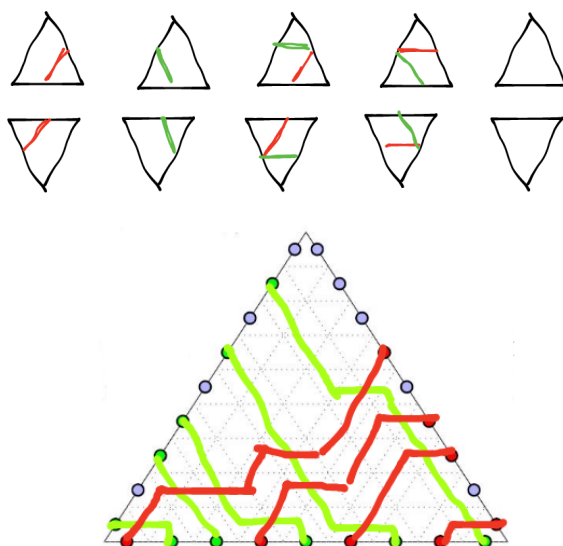
(Notice that our triangle can be expanded, and this will not change the number of allowed paths.) We will understand where this construction comes from in more detail, and we will do so by introducing a more general set of

tilings. But we'll do so by going back to our definition of skew Schur functions and considering the following diagram:
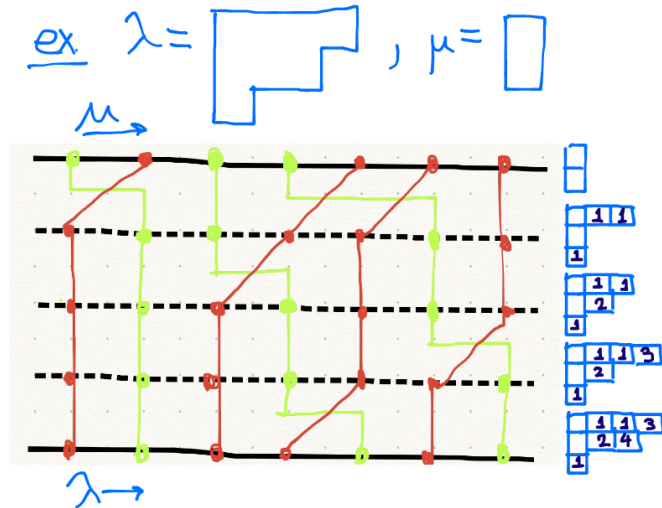
The dots at the bottom and top of the lattice represent the shape of λ and μ, and they are connected using the tiles introduced above (where each row is two series of tiles, using 45-45-90 triangles). Notice that tracking the series of red and green dots along each row gives us a Young diagram, and if we write in numbers coresponding to the row number when a box is added, that gives us a semi-standard Young tableau of shape λ/μ. (The set of allowed tiles only allow our green paths to move to the right, so we will always add boxes as we go down our diagram.) So this explains that it makes sense to expect some kind of correspondence!

We will formalize our discussion in the following way:

Definition 147

The **Fock space**, denoted $\mathcal{F}$, is the infinite-dimensional Hilbert space of sequences of green and red dots which are eventually green and red on the left and right, respectively.

In particular, all sequences arising from our Young diagrams are in the Fock space.

Definition 148

For each partition $\lambda \in \mathcal{F}$ the **transfer matrix** $T_{\text{free}}(x)$ corresponds to moving one row up in our tiling. In particular, we define the matrix elements

$$\langle f | T_{\text{free}}(x) | g \rangle$$

for any two basis elements $f, g \in \mathcal{F}$ to be the number of tilings with top row f and bottom row g, weighted by $x^{\text{number of right moves}}$.

For example, if f and g are the lowest and second lowest rows of our diagram above, respectively, then the matrix element between f and g is x (since only one red line moves to the right). Combining this definition with our previous discussion, we can notice that

$$\left\langle \mu \left| \prod_{i=1}^{n} T_{\text{free}}(x_i) \right| \lambda \right\rangle = S_{\lambda/\mu}(x_1, \cdots, x_n),$$

because each x_i corresponds to one of the rows that gets us from μ to λ and thus adding the boxes that are labeled with some particular number (remembering that skew Schur functions are also defined in terms of counting semistandard Young tableaux, so we can also write them as a sum of monomials of the form $\prod_{b \in T} x_{T(b)}$). Because Schur polynomials

are symmetric, this transfer matrix must satisfy $[T_{\text{free}}(x), T_{\text{free}}(y)] = 0$, and we can also write down the equation

$$S_{\lambda/\mu}(x_1, \cdots, x_n, y_1, \cdots, y_m) = \sum_{\rho} S_{\lambda/\rho}(x_1, \cdots, x_n) S_{\rho/\mu}(y_1, \cdots, y_m)$$

by making another combinatorial interpretation of having μ and λ be the top and bottom rows and sandwiching ρ between them n rows below the top. Unfortunately, all of this doesn't get us closer to L-R coefficients as stated: instead, we'll need to introduce new types of tiles to account for the other structure. Their labels are listed below:

Source (1)

new tiles
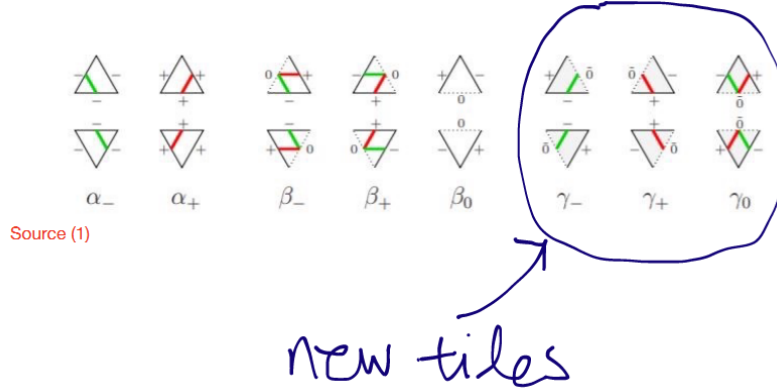
These tiles basically allow us to change directions, so that "green paths always move to the right" is no longer required. We'll now also allow ourselves to have **empty spots** in each row, rather than having just green and red dots, so that there are three possible states for each point in our horizontal line.

This new type of tiling now allows us to introduce a new **Fock space** $\mathcal{G} \supseteq \mathcal{F}$, which now allows for red, green, or empty dots but still requires that all dots are eventually green and red to the left and right, respectively. It turns out that we can take two elements of the Fock space $\mathcal{F}$ and get out an element of the Fock space $\mathcal{G}$: basically, instead of viewing $\mathcal{F}$ as a subset of $\mathcal{G}$ by inclusion, we can use a **concatenation map** by taking in $f_1, f_2 \in \mathcal{F}$ and outputting $f_1 \sqcup f_2$, taking the green entries of f_1 to the left of the origin 0, taking the red entries of f_2 to the right of the origin, and leaving all other spots blank. It becomes more subtle to define transfer matrices in this new settings, though, because here each row is shifted by a half time-step: alternate rows of our triangular tiling are at positions $\mathbb{Z}$, then $\mathbb{Z} + \frac{1}{2}$, then $\mathbb{Z}$, and so on. So we'll need to define many more transfer matrices:

- We define $T_{\pm}$ and $\tilde{T}_{\pm}$.

- In a $+$ matrix, the edges are shifted to the right by $1/2$ per time unit, and in a $-$ matrix, the edges are shifted to the left by $1/2$ per time unit.

- Tilde matrices allow all moves, including the new tiles, while non-tilde matrices allow only the original α, β type moves.

The key integrability of the model is as follows:

> **Lemma 149**
> The transfer matrices $T_+, T_-, \tilde{T}_+, \tilde{T}_-$ all commute with each other.

The proof of this result repeatedly applies the Yang-Baxter equation, which is an equation satisfied by many integrable models of this type. And another key fact that gets us closer to the L-R coefficients is the following:

In words, what's going on here is that no matter what initial state $|f\rangle$ we are in, after enough repeated applications of T, thinking about our diagrams from bottom to top, all of our crossings will happen, and the green lines will be to the left of the red lines. (This is importantly because the transformations T^2 are products of T_+ and T_-, which only allow for green lines to move upward left and/or red lines to move upward right.)

In particular, if our starting state $|f\rangle$ is a partition, then the right-hand side will only sum over partitions, and this is because T^{2k} applied on a partition will still have no empty dots, and also because partitions always have zero **charge** (associated with the $+$s and $-$s that label our tiles in our diagram above). Specifically, for any contribution to the sum, we must have $0 = C(f) = \frac{1}{2}(C(g) + C(h))$, and the emptiness number $0 = e(f) = \frac{1}{2}(C(g) - C(h))$, and thus this forces us to have $C(g) = C(h) = 0$.

This can be shown by combining the previous lemma, the commutation relations of $T_\pm$, $\tilde{T}_\pm$, and our original relation between transfer matrices and skew Schur functions. And to connect this back to our original drawing of of paths in a triangle, we can consider the following diagram:
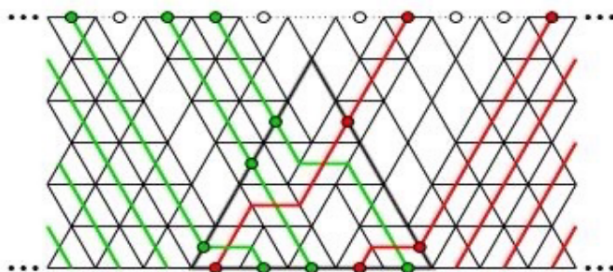
Basically, crossings must occur within a finite triangular region, and because we start with a partition on the bottom of our picture, there are no holes on the bottom edge of our triangle. And then we can fill in the empty spaces on the bounds of the triangle with purple dots, and this gets us back to the original diagram.

In summary, these integrable tilings allow us to find a direct proof of a combinatorial result (not relying on induction, which is what the original proof used).

21 May 6, 2021

We'll finish the landscape that we described in the first lecture today. Last lecture, we discussed a connection between rhombus tilings of a hexagon and the point of tangency of the frozen region: near that point of tangency, centering and rescaling the positions of the vertical lozenges gives us the eigenvalues of the GUE (top-left) corners. This may

seem like a coincidence – we can do the computations and see that it works – but there are in fact further connections that we can draw here.

Theorem 152 (Harish-Chandra 1957, Itzykson-Zuber 1980)

Fix N real eigenvalues of a matrix $a_1 > \cdots > a_N$, and let A be a uniformly random $N \times N$ Hermitian matrix with those eigenvalues $\{a_1, \cdots, a_n\}$. If B is a fixed Hermitian matrix with eigenvalues $b_1 > \cdots > b_N$, then

$$\mathbb{E}[e^{i\operatorname{Tr}(AB)}] = \det[e^{ia_k b_\ell}]_{k,\ell=1}^N \prod_{1 \leq i < j \leq N} \frac{j-i}{(a_i - a_j)(b_i - b_j)}.$$

Explaining how to generate such a random matrix A can be done in two ways: the first is to consider the map

$$\mathcal{U}(N) \to \operatorname{Herm}(N) \text{ given by } u \to u \begin{bmatrix} a_1 & & \\ & \ddots & \\ & & a_N \end{bmatrix} u^{-1}$$ and push the Haar measure forward onto $\operatorname{Herm}(N)$. This

gives us the **orbital measure**, with name coming from the fact that the set of conjugations of a diagonal matrix form an **orbit**. But the second is to think about this situation more geometrically: the orbital matrices form a manifold embedded in Euclidean space, so we can take the Lebesgue measure on the surface.

We can then write the expectation on the left-hand side alternatively as the **orbital integral**

$$\int_{\mathcal{U}(N)} \exp\left[i\operatorname{Tr}\left(U^{-1}\begin{bmatrix} a_1 & & \\ & \ddots & \\ & & a_N \end{bmatrix} U \begin{bmatrix} b_1 & & \\ & \ddots & \\ & & b_N \end{bmatrix}\right)\right] dU.$$

The expectation $\mathbb{E}e^{i\operatorname{Tr}(AB)}$ is then the Fourier transform of the orbital measure, because the ordinary Fourier transform is $\hat{f}(p) = \int_{\mathbb{R}^N} f(x)e^{ix \cdot p} dx$, and we can indeed think of Hermitian matrices as living in a finite-dimensional vector space.

We can compare this formula with the one for normalized Schur functions: if we compare what we have with the **normalized character**

$$\frac{s_\lambda(u_1, \cdots, u_N)}{s_\lambda(1, \cdots, 1)} = \frac{\det[u_i^{\lambda_j + n - j}]_{i,j=1}^N}{\prod_{i<j}(u_i - u_j)} \cdot \left(\prod_{i<j} \frac{\lambda_i - \lambda_j - i + j}{j - i}\right)^{-1},$$

where we've used Weyl's character formula and Weyl's dimension formula. We can notice that there are three Vandermonde determinants (the ones with $j - i$, $u_i - u_j$, and $(\lambda_i - i) - (\lambda_j - j)$), and that lines up with the formula above. What's left is the different expressions in the denominators, which are indeed different: we have u_js on the unit circle, but we have a_k, b_ℓ on the real line. It turns out that this is created by a **limit transition**, where the $u_j = e^{i\varepsilon a_j}$s are close to 1 and we take $\lambda_k = \varepsilon^{-1} b_k$ to avoid degeneracies. Our determinant then looks like

$$\det[u_j^{\lambda_k + N - k}]_{j,k=1}^N = \det\left[e^{i\varepsilon a_j(\varepsilon^{-1} b_k + N - k)}\right]_{j,k=1}^N,$$

and as $\varepsilon \to 0$ we can throw out the $(N - k)$ term, so this indeed becomes

$$\approx \det\left[e^{i\varepsilon a_j \varepsilon^{-1} b_k}\right]_{j,k=1}^N = \det\left[e^{ia_j b_k}\right]_{j,k=1}^N.$$

In this limit transition, we can check that the other Vandermonde determinants also line up: we have

$$u_k - u_\ell = e^{i\varepsilon a_k} - e^{i\varepsilon a_\ell} = i\varepsilon\left(a_k - a_\ell\right), \quad (\lambda_i - i) - (\lambda_j - j) = \varepsilon^{-1}(b_i - b_j) + o(\varepsilon^{-1}).$$

So indeed the formulas will converge to each other in the appropriate limit:

$$\lim_{\varepsilon \to 0} \frac{S_\lambda(\vec{u})}{S_\lambda(\vec{1})} = \mathbb{E}e^{i\text{Tr}(AB)}.$$

The next question we'll ask about is how to compare the following two problems:

1. Let M_α, M_β be two orbital measures (meaning that they live on matrices with spectrum $(\alpha_1, \cdots, \alpha_N)$ and $(\beta_1, \cdots, \beta_N)$, and these are unitarily invariant measures. If we add the two measures M_α, M_β, the when we decompose the result on orbital measures (because the result must still be unitarily invariant), we can get an integral over spectra $\int_{\gamma_1 > \cdots > \gamma_n} M_{\alpha\beta}^\gamma \mathbb{P}(d\gamma)$.

2. Let T_λ, T_μ be irreducible representations of the unitary group $\mathcal{U}(N)$ or of $GL(N, \mathbb{C})$, and decompose the tensor product $T_\lambda \otimes T_\mu$ into irreducibles $\bigoplus_\nu c_{\lambda\mu}^\nu T_\nu$, where those $c_{\lambda\mu}^\nu$ are the Littlewood-Richardson coefficients.

   The idea with problem (1) is that we really have a **convolution** going on, so multiplying the Fourier transforms of M_α and M_β (which are functions on the dual variable) and then writing the result as an average of Fourier transforms of $M_{\alpha\beta}^\gamma$ will give us the result we want. And the idea for problem (2), we multiply characters $s_\lambda \cdot s_\mu = \sum c_{\lambda\mu}^\nu s_\nu$, and we can write this in terms of normalized characters as

$$\frac{s_\lambda}{s_\lambda(1, \cdots, 1)} \cdot \frac{s_\mu}{s_\mu(1, \cdots, 1)} = \sum_\nu \frac{s_\nu(1, \cdots, 1)}{s_\mu(1, \cdots, 1)s_\mu(1, \cdots, 1)} c_{\lambda\mu}^\nu \frac{s_\nu}{s_\nu(1, \cdots, 1)}.$$

The blue part of this expression must be a **probability measure**, because it sums to 1 if we plug in $\vec{u} = \vec{1}$ into both sides. And it turns out that if we take

$$\lambda \sim \varepsilon^{-1}(\alpha_1, \cdots, \alpha_N), \mu \sim \varepsilon^{-1}(\beta_1, \cdots, \beta_N),$$

we get a probability measure that converges to $P_{\alpha\beta}^\gamma$. So we can solve random matrix theory problems by solving the analogous representation theory problems first, and taking the labels λ, μ to ∞ accordingly – more generally, we have the phenomenon that "large representations of Lie groups behave as orbital measures on the (dual to their) Lie algebras." In representation theory, this is known as the **semiclassical limit** (and it is related to the transition from quantum mechanical to the classical limit in physics).

Fact 153

In fact, we sometimes have the Lie algebra (the commutative object) containing all of the information of the representation theory (the noncommutative object), and it works best for nilpotent Lie algebras and Lie groups.

And in our remaining time, we'll briefly go over some points that we would have discussed if there were more time in the class (since many of the remaining sessions will likely come from students in the class): the topic would have been **types of results that people look for in probabilistic systems**. There are a few classes of results that are usually being looked for – generically, they are often hard to get, but the integrable structures help.

- **Evaluating the partition function**. For example, if we weight plane partitions by a factor of $q^{\text{number of boxes}}$, then we may want to calculate

$$Z = \sum_{\text{plane partitions}} q^{\text{number of boxes}},$$

  which is the normalization constant in the measure $\mathbb{P}(\text{plane partition}) = \frac{q^{\text{number of boxes}}}{Z}$. It turns out that even though this may seem like a normalization constant, partition functions often encode a lot of information about the system – in fact, being able to compute them often means the system is integrable. For plane partitions,

the answer is **MacMahon's formula**

$$Z = \prod_{n \geq 1} \frac{1}{(1 - q^n)^n}.$$

A corollary of this formula is that if we take large plane partitions (sending $q \to 1$, specifically setting $q = e^{-L^{-1}}$), we get convergence in probability

$$\lim_{L \to \infty} \frac{\text{number of boxes in random plane partition}}{L^3} = 2\zeta(3).$$

This formula basically comes from the power of the partition function's ability to tell us how large the plane partitions are!

- **Law of Large Numbers results**. In an introductory probability class, this is the statement that the average of iid samples from a distribution converges to a fixed number, and more generally it is a result about **disappearance of randomness in the limit**.

- **Fluctuations**. Understanding these basically require us to subtract off the LLN behavior and understand what remains, but there are many different possible behaviors. For example, we can study **one-point fluctuations**, meaning that we look at a particular point on our limit curve and study the distribution of the deviation along a vertical section. But we can also study **global fluctuations** over the whole picture, and we saw some of this in the Gaussian free field behavior in cylindrical plane partitions, or **local fluctuations**, meaning that we restrict the range to a particular neighborhood of the limit shape (for example, looking at "edge fluctuations" or "bulk fluctuations"). **Mesoscopic (intermediate) fluctuations** are also studied, which are fluctuations occuring between the local and the global scale.

- **Classification of ergodic Gibbs measures**. To explain the words here, if we take a small subdomain from the hexagon-in-a-hole picture, and the configuration outside of that subdomain is frozen, we will still observe the uniform distribution on all possibilities. If we then zoom in on this subdomain inside the bulk liquid region, we get a translation-invariant object with the **Gibbs property**. Being able to classify these kinds of Gibbs measures abstractly means that we can get some information for describing our system.

> **Example 154**
>
> We can use rhombus tilings of the hexagon with a hole as an illustrative example of all of these different phenomenon that we can study.

It turns out we cannot compute the partition function in a nice form for this model. But we get law-of-large-numbers results coming up in the limit shapes (frozen versus liquid region), and one-point fluctuations come up by looking at deviations from the predicted smooth surface. Furthermore, the behavior of these fluctuations looks interestingly different near the boundary – we get a different type of fluctuation behavior at the edges because of the **Pokrovsky-Talapov law** (which basically says that the fluctuations are differentiable but not C^2). And looking at local fluctuations (such as the probability of seeing three rhombi of the same type next to each other) is also doable in this model, and there are nice formulas for those kinds of probabilities. Finally, we can get a correspondence with the GUE corners process by looking at positions of rhombi around the hole in the middle! So overall, there are many different limits occurring in a single picture of rhombus tilings.

If we use the list above, look at all of the objects we've studied in this class, and ask how each of the features show up in the models, there are typically nontrivial theorems for most of the items in our list. But we can always ask if we're curious about a particular model and what's known about it!

22   May 11, 2021

Today's class is a presentation by me (Andrew Lin), titled **Domino Tilings and the Arctic Circle Theorem**. I did not take notes on my own presentation (because I was giving it), but the handwritten notes I used to present can be found on my website.

23   May 13, 2021

Today's class is a presentation by **Carina Hong**, titled **A survey on dense O(n) loop models**.

Definition 155

Let H be a (potentially infinite) graph. A **loop configuration** is a spanning subgraph w where all vertices have degree 0 or 2 – we will denote $L(w)$ to be the number of (simple) cycles of w and $o(w)$ to be the number of edges.

We can define a probability measure on the loop configurations

$$\mathbb{P}_{H,n,x}(w) = \frac{x^{o(w)} n^{L(w)}}{Z_{H,n,x}^{\text{loop}}},$$

where n, x are parameters that can be changed and Z is the usual normalizing constant. While it is difficult to get accurate forms for various observables in this model, there have been some predictions made in the past: for example, Nienhuis (1982) predicts that a phase transition occurs at

$$x_c(n) = \frac{1}{\sqrt{2 + \sqrt{2 - n}}}$$

for $0 \leq n \leq 2$. Here, criticality means that for all $x < x_c$, the model is **subcritical**, meaning that we have exponential decay

$$\mathbb{P}(\text{loop through a given point has length } \geq t) < e^{-ct}$$

for some $c > 0$, while for all $x \geq x_c$, the model is **critical**, meaning that we have a power law

$$\mathbb{P}(\text{loop through a given point has length } \geq t) \geq t^{-c'}$$

for some $c' > 0$. There are also predictions about **conformally-invariant scaling limits** for the critical case, both for $x = x_c$ and $x > x_c$. Kager and Nienhuis (2004) conjectured that this scaling limit is the random SLE curve with $n = -2 \cos\left(\frac{4\pi}{\kappa}\right)$ – specifically, when $x = x_c$, $\kappa \in [\frac{8}{3}, 4]$, and when $x > x_c$, $\kappa \in [4, 8]$. And later, Sheffield (2009) conjectured that the limit is instead the CLE (intuitively, a loop version of the SLE) with the same parameters. (On the other hand, when $n > 2$, there are predictions that the model is always subcritical, and for n negative, we will have a signed measure, but we get the same predictions for the critical values x_c.)

We'll now survey some rigorous results that have already been proved:

Definition 156

Let T^0 be one of the three color classes of the canonical coloring of the dual graph of the hexagonal lattice. A domain H of the hexagonal lattice is of **type 0** if no edges of the hexagonal lattice border T^0 and have one vertex in H and one vertex in the complement of H.

(A type 0 domain then has a boundary which cannot just touch a single vertex of a T^0 hexagon.) The result is then that for a type 0 domain, there are no large loops:

Theorem 157

There exist constants $n_0, c > 0$ so that for all $n \geq n_0$, any type-0 domain H, and any $x \in (0, \infty]$, suppose that w is a loop configuration of H (chosen with parameters (n, x)). Then for all $u \in V(H)$ and an integer $k > 6$,

$$\mathbb{P}(\text{loop of length } k \text{ in } w \text{ surrounding } u) \leq n^{-ck}.$$

In other words, we have an exponential decay argument for these type-0 domains. We also have a similar type of theorem for loop connectivity:

Definition 158

Let u, v be two points. Then u and v are **loop-connected** if there is a path between u and v whose paths belong to loops in w.

Theorem 159

There exist $C, c > 0$ so that for all $n > 0$, any type-0 domain, and any $x \in (0, \infty]$, if we pick a loop configuration w (with parameters n, x), then

$$\mathbb{P}(u \text{ is loop-connected to some } v, \quad d(u, v) = k) \leq (C(n+1)x^6)^{ck}.$$

In other words, we cannot find loops that surround a given vertex, and a vertices are unlikely to be loop-connected to a nearby vertex. Those results are powerful when n, nx^6 are small, but there are also results for large values:

Definition 160

Let the 0-**phase ground state** be the collection of edges that bound T^0 hexagons. Two points u, v are **ground-connected** if there is a path between them whose vertices belong to loops both in w and in the 0-phase ground state.

It turns out that we get an inequality in the other direction as well:

Theorem 161

Taking the same C, c as above, and making the same assumptions,

$$\mathbb{P}(u \text{ ground-connected to some } v \text{ on the boundary of } H) \geq 1 - C(n \cdot \min(x^6, 1))^{-c}.$$

In general, remembering that x is the weighting factor for edges and n is the weighting factor for paths (which will be primarily hexagons), we get "packed" behavior for nx^6 large and "dilute" behavior for nx^6 small.

We'll now talk about results for large loops: if we let $\mathrm{LoopConf}(H, \zeta)$ be the set of configurations that coincide with some ζ configuration outside of H, then we define a **Gibbs measure**

$$\mathbb{P}^\zeta_{H,n,x}(w) = \frac{x^{o(w)} n^{L(w)}}{Z^\zeta_{H,n,x}}$$

(In other words, the probability measure conditioned on LoopConf(H, ζ) is $\mathbb{P}_{H,n,x}^{\zeta}$ for almost all loop configurations ζ.) There aren't results known in general about uniqueness of Gibbs measures, or whether there is a way to get a weak convergence of measures to a Gibbs measure, so things are still rather mysterious. But we'll see some results related to them soon:

Definition 162

Let Λ_k be the ball (in graph distance) in the triangle lattice of radius k around the origin, and let A_k be the annulus in the hexagonal lattice consisting of edges between vertices of a hexagon in $\Lambda_{2k} \setminus \Lambda_k$.

Theorem 163 (Dichotomy Theorem)

Let $n \geq 1$ and $x \leq n^{-1/2}$. Then there is a unique translationally-invariant Gibbs measure $\mathbb{P}_{n,x}$ with (almost surely) no infinite paths and with exactly one of the following two conditions:

1. There exists $c > 0$ so that for all $k \geq 1$, $\mathbb{P}_{n,x}(\text{loop surrounding origin of length} \geq k) \leq e^{-ck}$,

2. There exists $c > 0$ so that for all $k > 1$, we have for all loop configurations ζ that

$$\mathbb{P}_{A_k,n,x}^{\zeta}(\text{loop in } A_k \text{ surrounding origin}) \in [c, 1 - c].$$

We can also show that in fact the second condition holds in particular cases:

Theorem 164

For $n = 1, 2$ and $x = x_c(n)$, we have condition (2) above.

Proof sketch. Suppose that condition (1) were to hold, so that we are likely to have "no large loops" by the exponential decay argument. Then for these particular parameters, it is likely that two of the boundary paths of our domain will be connected by a path. Therefore, we can glue domains together, moving them so that the endpoints line up, and this forms a long loop. (Here, we need the assumption of "small loops" to deal with the boundary potentially being messy.) This leads us to a contradiction, so the Dichotomy Theorem tells us that we must have condition (2). □

Theorem 165

There exists $\delta > 0$ such that for all $n \in [1, 1 + \delta]$ and for all $x \in [1 - \delta, n^{-1/2}]$, condition (2) holds.

Proof sketch. The main idea is to use the **XOR trick**: in this regime, we claim that there is almost surely an infinite path, or every vertex is surrounded by infinitely many loops. Then by the dichotomy theorem, we can't have an infinite path, and we can't have condition (1) because there can't be infinitely many loops by the "exponential decay."

Definition 166

The loop configurations form a closed subgroup of $\{0, 1\}^{E(\mathbb{H})}$ (where $E(\mathbb{H})$ is the edge class of the hexagonal lattice), so we can define the **XOR operation** between a loop configuration w and a simple cycle Γ via

$$(w \oplus \Gamma)(e) = w(e) + 1_{e \in \Gamma} \mod 2.$$

We can then prove the above claim using the following lemma:

> **Lemma 167**
>
> Let Γ be a cycle that surrounds Λ_k, and let w be a loop configuration. Then either w or $w \oplus \Gamma$ contains an infinite path or a loop of diameter at least k surrounding the origin.

Proof of lemma. This is a combinatorial argument: let W_Γ is the set of loops in w that intersect T. If w does not have an infinite path or a loop of diameter at least k surrounding zero, then because XOR is an involution, all loops in $w_\Gamma \oplus \Gamma$ must intersect Γ (otherwise it wouldn't have existed in w_Γ in the first place, and also wouldn't be affected by toggling of Γ).

We've made the assumption that there is no loop in w_Γ that surrounds the origin (because Γ surrounds Λ_k already, so it is already a large cycle). So looking at the relation

$$w \oplus \Gamma = \Gamma \oplus w_\Gamma \oplus (w \setminus w_\Gamma) = (\Gamma \oplus w_\Gamma) \cup (w \setminus w_\Gamma),$$

we find that there must be a loop in $T \oplus w_\Gamma$ surrounding the origin (since there is a loop on the left-hand side, but not in $w \setminus w_\Gamma$). This loop is therefore in $w \oplus \Gamma$ by assumption, and it has diameter at least k as desired. $\qquad\square$

$\square$

> **Corollary 168**
>
> Suppose a configuration ζ has no infinite path. Then
>
> $$\mathbb{P}^\zeta_{\Lambda_k,1,1}(\text{exists a loop of diameter } \geq k \text{ surrounding } 0) \geq \frac{1}{2}.$$

Proof. If we let Γ be the boundary $\partial \Lambda_k$. Under $(n, x) = (1, 1)$, the XOR operation is measure-preserving, so $\mathbb{P}(w \text{ has a loop}) = \mathbb{P}(w \oplus \Gamma \text{ has a loop})$, and a union bound gives us the result. $\qquad\square$

Finally, we'll conclude by mentioning the significance of a few particular parameters (n, x) and their connections to some of the other objects we've studied:

- When $n = 0$, we are sampling a self-avoiding walk (with no cycles).

- When $n = 1$, we have the Ising model on the triangular lattice, with $x = e^{-2\beta}$. In other words, setting $x = 1$ gives us the Ising model at infinite temperature (giving rise to critical site percolation) – this is the regime of our previous corollary – and setting $x \to \infty$ gives us the anti-ferromagnetic Ising model at zero temperature, which is equivalent to the dimer model.

- When $n \to \infty$ and nx^6 is fixed, we get the hard-hexagon model, meaning that we have hexagons that do not intersect. (This can be connected to the grand canonical ensemble in statistical mechanics.)

24 May 18, 2021

Today's class is a presentation by **Matthew Everett Lerner-Brecher**, titled **The Driven q-Whittaker and Push Block Hall-Littlewood Processes on a Torus**. We'll start by reviewing some previous topics from earlier in this class, and then we'll use that motivation to understand some features of the new particle processes.

We can see a visualization of the 2+1 dimensional growth model at https://wt.iam.uni-bonn.de/ferrari/research/jsanimationakpz, in which the dynamics preserve the interlacing conditions of the particles. Basically, particles jump

to the right at exponential clock times, such that they never move in front of the particle below them and push all particles that live directly above them when they jump. (So this is related to where the "push-block" part of the title will come in.)

This particle model can then be thought of as a growth model. Essentially, everywhere that we have a particle in our particle process, we can place a slanted rhombus, and this naturally fills in the shape as shown:
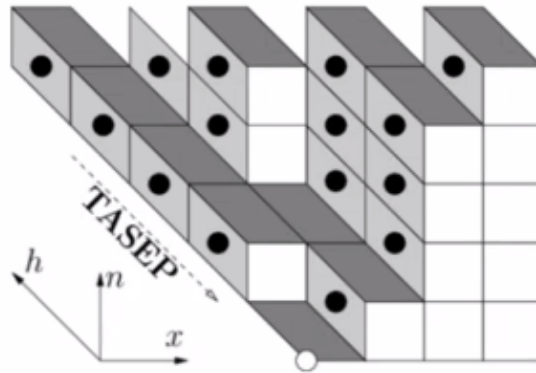
In this lecture, we'll be studying particle processes on a torus – the main points are that we'll still be able to define a height function, and the integrability will give us stationary measures (which aren't easy to come by in 2+1-dimensional growth models). We'll still focus on the particle description for the rest of this talk, though, because that's where the analysis is easier to do.

Our particle models will live on a torus $\mathbb{T}$ with length L and with N rows, such that each row has m_1 particles (this is required for interlacing). For simplicity, we'll refer to a particle by its location $x_{k,r}$ (for the kth particle on the rth row), and we want to be able to refer to things like $x_{k+1,r}$ and $x_{k,r+1}$ as the "right neighbor" and "top right neighbor," respectively. But we'll notice that we run into wrapping issues because the coordinates are defined on the torus, and thus we'll need to need an additional wrapping parameter which we'll denote by m_2 so that

$$x_{k,N+1} = x_{k+m_2,1}$$

(basically, so that the top and bottom rows interact nicely enough with each other).

Definition 169

Let Ω_{L,N,m_1,m_2} define the state space of all particle arrangements on the torus $\mathbb{T}_{L,N}$ with m_1 particles on each row and with shift parameter m_2.

It turns out that all particle models will be ergodic on this state space – as long as we have the same number of particles and the same wrapping parameter, we can get from any position to any other position.

Definition 170

In the **driven q-Whittaker system**, the particle $x_{k,r}$ will move to the right by one unit at an exponential clock of rate
$$\frac{(1 - q^{x_{k+1,r-1}-x_{k,r}-1})(1 - q^{x_{k,r}-x_{k-1,r}})}{1 - q^{x_{k,r}-x_{k,r-1}+1}},$$
and with the usual "pushing" mechanism described above.

Basically, we have a product of $(1 - q^n)$ terms, where n is a nearest-neighbor distance between various pairs of particles. We can check that the first term in the numerator is 0 if $x_{k+1,r-1} = x_{k,r}$ (meaning a particle is one unit

away from its lower-right neighbor). So indeed, we enforce the blocking behavior of the dynamics that we previously mentioned.

Theorem 171 (Corwin-Toninelli)

The stationary measure of the q-Whittaker dynamics on Ω_{L,N,m_1,m_2} is given by

$$\frac{1}{Z_{L,N,m_1,m_2}} \prod_{\substack{1 \leq k \leq m_1 \\ 1 \leq r \leq N}} \frac{(q;q)_{x_{k+1,r}-x_{k,r}}}{(q;q)_{x_{k+1,r-1}-x_{k,r}-1}(q;q)_{x_{k,r}-x_{k,r-1}-1}},$$

where $(q;q)_a$ is the q-Pochhammer symbol.

It turns out that this is a Gibbs measure, meaning that if we fix a certain set of particles around a given one, then the configuration of particles within the fixed region is independent of the distribution outside of that region. (The left, right, top-left, top-right, bottom-left, and bottom-right neighbors of a given particle are the only ones that affect its rate or have rate affected by it.)

Even though we have this stationary measure, the process is still difficult to study, but certain properties have been discovered in the **diffusive limit** where we scale time (which we'll denote by s because there will be another parameter later) and length L by ε^{-1}, take $q = e^{-\varepsilon}$, and keep N, m_1, m_2 fixed. We then assume an initial condition where the process has a "crystalline configuration" as follows: let C and D be fixed constants, and let

$$X_{k,m} = kD\varepsilon^{-1} + mC\varepsilon^{-1}.$$

If x is sufficiently close to $X_{k,m}$ initially, then we can define

$$\eta_{k,m}(s) = \sqrt{\varepsilon}(X_{k,r}(s\varepsilon^{-1}) - X_{k,r} - vs\varepsilon^{-1}).$$

Then it turns out (Borodin-Corwin-Toninelli) that $\{\eta_{k,r}\}$ will converge to the solution of the linear stochastic equation

$$\xi_{k,r}(s) = \sqrt{v}dW_s^{k,r} + \sum_{k',r'} A_{k,r}^{k',r'}\xi_{k',r'}dt.$$

With this, we can show that the particle gradients have Gaussian stationary measure, and we also have control over the spacetime fluctuations.

Definition 172

In the **push block Hall-Littlewood system**, particles move according to an exponential clock of rate

$$\frac{(t;t)_{a_{k,r}+1}(t;t)_{c_{k,r}}}{(1-t)(t;t)_{b_{k,r}}(t;t)_{d_{k,r}-1}} \cdot 1\{x_{k+1,r-1} = x_{k,r} + 1\}$$

where the coefficients $c_{(k,r)}, d_{(k,r)}$ represent the number of consecutive particles stacked in the row below and directly behind $x_{k,r}$, respectively, and $a_{(k,r)}, b_{(k,r)}$ represent the number of particles directly to the right and directly to the right a row below (where we ignore the gap in the column to the right of our current particle).

This may look like a complicated setup, but in order to have the interlacing condition satisfied, $c_{k,r} - (d_{k,r} - 1)$ will either be 0 or 1, so what we're left with is a relatively simple polynomial in t.

The stationary measure of this process can be described in the following way: we have a product of terms corresponding to pairs of adjacent rows of the torus. For each pair, we look at the different configurations of particles and assign a corresponding weight as shown:
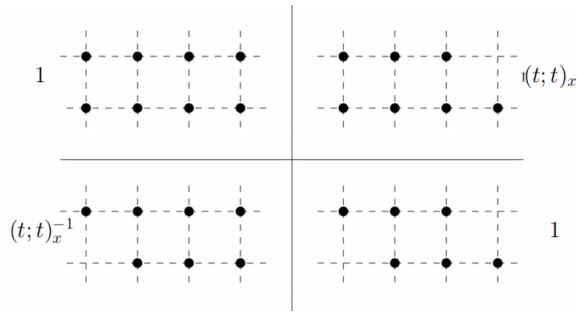
In particular, if we take t close to 1 and look at the diffusive limit, then an isolated particle contributes $(1 - t)$, but a pair of particles that are associated as part of a single configuration — frames 1 and 4 below — will give the larger weight $(1 - t)$ instead of $(1 - t)^2$ for two independent particles.
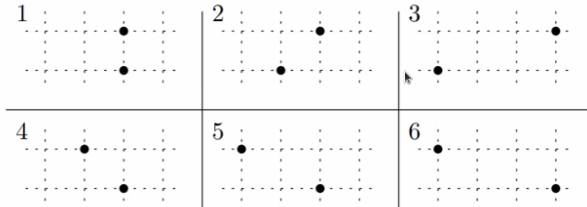
So this means that particles prefer to be grouped closer together — if we take $L \to \infty$, because the relative number of configurations with clumped groups is small, we expect an approximately uniform measure.

We may notice that the configurations are not very "local" in our figure above, so if we want to get a Gibbs measure, we will look at a dual process on the empty spaces of our torus, and we flip the diagram around so that particles still move to the right. We then get the new dual interlacing condition

$$0 \leq x_{k,r+1} - x_{k,r} \leq 1$$

(we can understand this in terms of partitions and dual partitions), and we get new rates for the particles. We then get the stationary measure

$$\frac{1}{Z} \prod_{\substack{1 \leq k \leq L - m_1 \\ 1 \leq r \leq N}} \left[(t; t)_{e_{k,r}} \right]^{e_{k,r} - e_{k,r+1}} ,$$

which gives us the same Gibbs property as before.

But if we return to the original push block Hall-Littlewood system and study the diffusive limit, we'll again scale time s and length L with ε^{-1} and take $t = e^{-\varepsilon}$, keeping N, m_1, m_2 fixed. We then want to look at

$$\eta_{k,m} = \sqrt{\varepsilon} \left(x_{k,m}(s\varepsilon^{-1}) - s\varepsilon^{-1} \right) .$$

We don't have a nice stochastic differential equation for this system, but these particles are expected to behave in the following way: particles will coalesce into vertical chains (attaching to lower-left neighbor), moving according to a Brownian motion. The reason these chains form is that particles in chain 1 can only break apart a finite number of times, but Brownian motion, which is what we get if we rescale the fluctuations of the jump process, hits 0 infinitely many times. And furthermore, when we're in frame 5 in the diagram above, the top particle has rate $O(\varepsilon^{-1})$ while the bottom has rate 1, so we will almost immediately go from frame 5 back to frame 4. So indeed, the top and bottom particle will follow the same Brownian behavior.

25   May 20, 2021

Today's class is a seminar by **Cesar Cuenca**, titled **Global asymptotics of particle systems at high temperature**. We'll discuss two particle examples, the Hermite ensemble and spectra for sums of random matrices, and then we'll sketch some ideas of the results and proofs.

Definition 173

The **Hermite N-particle ensemble** is defined via

$$\text{Herm}(x_1, \cdots, x_N) \propto \prod_{1 \leq i < j \leq N} (x_i - x_j)^2 \prod_{k=1}^{N} e^{-x_k^2/2},$$

determining a random N-tuple of reals that we write in increasing order.

We consider this probability measure because the tuple $x_1 \leq \cdots \leq x_N$ gives us the distribution of the eigenvalues for the Gaussian Unitary ensemble (GUE): basically, we let m_{ij} be distributed as $N(0,1) + iN(0,1)$, and take the matrix $M = [m_{ij}]_{i,j=1}^{N}$ and define $X = \frac{M+M^*}{2}$. Then X is sampled from the GUE and we get (random real) eigenvalues of X given by the Hermite ensemble above.

We're interested in **global asymptotics** of this distribution: if we consider the empirical measure

$$\mu_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{\frac{x_i}{\sqrt{N}}},$$

meaning that we place masses of $\frac{1}{N}$ at the rescaled eigenvalues, there is a result (Wigner '55) that these measure μ_N converge weakly in probability to the **semicircle distribution** $\frac{\sqrt{4-t^2}}{2\pi}$ on $t \in [-2, 2]$. (Graphically, we can imagine drawing histograms for the number of eigenvalues landing in each particular interval, and if our histogram width scales with N and we rescale so that the area is always 1, the shape will approach a semicircle.)

We can now generalize this ensemble:

Definition 174

The **Hermite N-particle β-ensemble** is defined via

$$\text{Herm}_\beta(x_1, \cdots, x_N) \propto \prod_{1 \leq i < j \leq N} (x_i - x_j)^\beta \prod_{k=1}^{N} e^{-x_k^2/2}.$$

where $\beta \geq 0$ is a real number.

We again get a random N-tuple of real numbers, and the reason for this generalization is that there are certain values where we have random matrix interpretations: $\beta = 1, 4$ give us the eigenvalue densities for the Gaussian Orthogonal Ensemble and Gaussian Symplectic Ensemble, respectively (meaning that we have symmetric matrices in the former case and self-adjoint matrices with quaternionic entries in the latter case). In general, β serves as an "inverse temperature," and these kinds of distributions are known as **log-gas systems** in physics.

We wish to study the global asymptotics of this system: it turns out that for any fixed $\beta > 0$, nothing changes, and the empirical measures still converge weakly in probability to a **semicircle distribution** (though the range of the semicircle is now an interval whose size depends on β). And there's also a special case at $\beta = 0$: in such a situation, the density is just $\prod_{k=1}^{N} e^{-x_k^2/2}$, meaning that the x_is are iid standard Gaussians. Therefore, the empirical measure will

instead converge weakly in probability to the **Gaussian distribution**. So there is a disconnect between the limits, and it makes sense to explore the middle ground between them. The following is a new result:
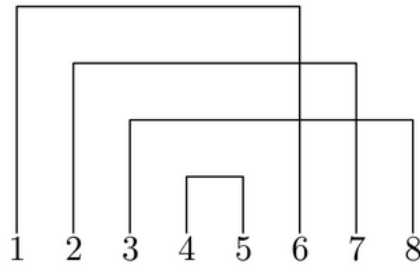

Theorem 175

If $x_1 \leq \cdots \leq x_N$ is $G\beta E$-distributed, denote the empirical measure by $\mu_{N,\beta}$. In the limit where $N \to \infty, \beta \to 0^+, \frac{N\beta}{2} \to \gamma \in (0, \infty)$, we have convergence of the measures weakly in probability to a limiting measure μ_γ.

In particular, μ_γ approaches the semicircle distribution as $\gamma \to \infty$ and the Gaussian distribution as $\gamma \to 0$. Instead of describing μ_γ in terms of the density, we'll describe them using moments in a combinatorial way:

Definition 176

Let $\pi = \{B_1, \cdots, B_n\}$ be a perfect matching of $\{1, \cdots, 2n\}$ (where each B_i contains a pair of elements). We draw the **arc diagram** for π by ordering B_1 through B_n in increasing order by smaller element, and then drawing an arc between the two elements of B_i so that B_1 is tallest and B_n is shortest (as shown below). Then the roof of π is defined as the number of roofs with no intersections.

For example, below is $B_1 = \{1, 6\}, B_2 = \{2, 7\}, B_3 = \{3, 8\}, B_4 = \{4, 5\}$, and the roof is 2:

Theorem 177

The moments of μ_γ are determined by

$$\int_{-\infty}^{\infty} x^k \mu_\gamma(dx) = \sum_{\text{perfect matchings } \pi \text{ of } \{1, \cdots, k\}} (\gamma + 1)^{\text{roof}(\pi)}.$$

In particular, we can notice that the kth moment is zero if k is odd, and if $\gamma \to 0^+$, the right-hand side approaches the number of perfect matchings of $\{1, 2, \cdots, 2n\}$, which is $(2n - 1)!!$. This again matches with the moment of the normal distribution. And on the other hand, if $\gamma \to \infty$ and we divide by γ^n, the right-hand side becomes the number of **noncrossing perfect matchings** of $\{1, 2, \cdots, 2n\}$, which is the Catalan number $C_n = \frac{1}{n+1}\binom{2n}{n}$. (This is because the noncrossing perfect matchings are the ones where all n arcs do not intersect.) This lines up with the moments of the semicircle distribution.

We'll now move to our second example of the talk: here is the setup. Let $a = (a_1 \leq \cdots \leq a_n)$ be an N-tuple of real numbers. Let A_N be a uniformly random complex Hermitian $N \times N$ matrix with a as its spectrum. (More precisely, we define $A_N = UDU^{-1}$, where D is the diagonal matrix with entries a_i, and U is chosen from the Haar distribution on $U(N)$.) We do this process with an N-tuple a and an N-tuple b, obtaining a matrix A_N and a matrix B_N independently. We're then curious about the (global asymptotics of the) eigenvalue distribution of $C_N = A_N + B_N$.

It turns out that the support of the eigenvalue distribution is given by **Horn's conjecture**: we have $\sum_i c_i = \sum_i (a_i + b_i)$ (by trace constraints), plus a variety of other inequalities such as $c_N \leq a_N + b_N$. But what we're curious

about is still the limits of the form

$$\frac{1}{N}\sum_{i=1}^{N}\delta_{c_i}.$$

Theorem 178 (Voiculescu '91)

Assume that $\frac{1}{N}\sum_{i=1}^{N}\delta_{a_i}\to\mu$ and $\frac{1}{N}\sum_{i=1}^{N}\delta_{b_i}\to\nu$ weakly. Then we have weak convergence in probability

$$\frac{1}{N}\sum_{i=1}^{N}\delta_{c_i}\to\tau=\mu\boxplus\nu,$$

where $\mu\boxplus\nu$ denotes the **free convolution** of μ and ν.

We can compute τ in the following way: letting $R_\mu(z)$ denote the **R-transform** of μ, we define

$$R_\tau(z)=R_\mu(z)+R_\nu(z).$$

But in the case where μ,ν are compactly supported, we can use moments (which will uniquely determine the measures): we have, for example, that

$$m_4^\tau=m_4^\mu+m_4^\nu+4m_2^\mu m_2^\nu-2(m_1^\mu)^2(m_1^\nu)^2+4m_1^\mu m_3^\nu+4m_3^\mu m_1^\nu+2(m_1^\mu)^2 m_2^\nu+2m_2^\mu(m_1^\nu)^2.$$

We will now define a β-analog of these, looking at the various limits and the "middle ground." We do get an issue, which is that we do not have β-random matrices in general. So we need to instead generalize the binary operator $(a,b)\mapsto c$, and we'll do that in the following way:

Definition 179

Let A,X be two complex Hermitian matrices. The **spherical Harish-Chandra-Itzykson-Zuber integral** is

$$I_N(A,X)=\int_{U(N)}e^{\text{Trace}(UAU^*X)}[Haar](dU)=\mathbb{E}\left[e^{\text{Trace}(UAU^*X)}\right],$$

where we integrate over the Haar probability measure over $U(N)$.

Because the Haar measure is $U(N)$-invariant, $I_N(A,X)$ only depends on the eigenvalues $a=(a_1,\cdots,a_n)$ and $x=(x_1,\cdots,x_n)$. Therefore, we can write $I_N(A,X)=I_N(a,x)$. It turns out that for all x, we have

$$\mathbb{E}[I_N(c,x)]=I_N(a,x)I_N(b,x)$$

(where c is random as in our construction above). We can prove this by noting that

$$I_N(a,x)I_N(b,x)=\mathbb{E}\left[e^{\text{Trace}(U\text{diag}(a)U^*X)}\right]\mathbb{E}\left[e^{\text{Trace}(V\text{diag}(b)V^*X)}\right]$$

where U,V are independent Haar-distributed and we are allowed to replace A with a diagonal matrix. We can then combine the two terms together by independence, giving us $\mathbb{E}[e^{\text{Tr}(C_N X)}]$, which simplifies to $\mathbb{E}[I_N(c,x)]$. So our generalization will be in how we define the spherical integral:

With this, we can define a beta-analog of sums:

It turns out that this definition has some caveats, because the existence of such a c is a conjecture (it's not known whether this exists for general β), and instead we just know that c is a generalized function (a distribution). But for simplicity, we'll call it a "random N-tuple" for now.

Again, what happens is that if β is positive and fixed, and $c = a +_\beta b$, and a and b's empirical measures converge weakly to μ, ν, we get $\frac{1}{N}\sum_{i=1}^n \delta_{c_i} \to \tau = \mu \boxplus \nu$ just like before. But we again have an outlier when $\beta = 0$: we have

$$B_N^{\beta=0}(a, x) = \frac{1}{N!} \sum_{\sigma \in S(N)} e^{x_1 a_{\sigma(1)} + \cdots + x_N a_{\sigma(N)}}.$$

So we can obtain $c = a +_0 b$ by picking

$$c = (a_1 + b_{\sigma(1)}, \cdots, a_N + b_{\sigma(N)})$$

for uniformly random σ. So in this case, we actually get convergence to $\tau = \mu * \nu$, the **usual convolution** of μ and ν. So we want to know again whether there's a middle ground, and this is a new result as well:

We can similarly compute the moments m_k^τ in terms of the moments of m_i^μ, m_j^ν, but they are complicated (even more in our formula above). So instead, we will consider another sequence with the same information as the moments, which we will call the **γ-cumulants**:
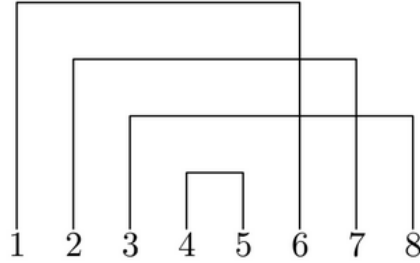
Below is the arc diagram for $B_1 = \{1, 3, 5, 7\}$, $B_2 = \{2, 4, 8\}$, $B_3 = \{6\}$:

For example, there are three roofs in the arc for B_1, but none of them have any intersections. There are two roofs for the arc for B_2, both of which intersect another roof. Finally, there are no roofs for the arc for B_3. Therefore, $p(1) = 0$, $p(2) = 2$, $p(3) = 0$, and in this case we get the weight

$$W_\gamma(\pi) = 0!(\gamma + 1)_3 \cdot 2!(\gamma + 3)_0 \cdot 0!(\gamma + 1)_0 = 2(\gamma + 1)(\gamma + 2)(\gamma + 3).$$

For example, there is only one possible set-partition for 1, so $m_1 = \kappa_1$. There are two possible set-partitions for 2, and it turns out that $m_2 = (\gamma + 1)\kappa_2 + \kappa_1^2$. We can then find that (by direct verification)

$$m_3 = (\gamma + 1)(\gamma + 2)\kappa_3 + 3(\gamma + 1)\kappa_2\kappa_1 + \kappa_1^3.$$

So the γ-cumulants and the moments encode the same information, and it turns out we have a clean characterization:

In the remaining time, we'll discuss some of the ingredients that went into proving these results. First of all, we should recall the classical Levy's continuity theorem, which tells us that we have weak convergence of probability measures $\mu_N \to \mu$ if and only if the characteristic functions, defined as the Fourier transforms

$$\phi(x_1, \cdots, x_d) = \int_{\mathbb{R}^d} e^{i(t_1 x_1 + \cdots + t_d x_d)} \mu_N(t_1, \cdots, t_d),$$

are such that $\phi_N \to \phi$ converges pointwise (which is often easier to prove than verification of weak convergence directly). This can be applied to the Central Limit Theorem, and the main philosophy is that convergence of measures is equivalent to convergence of functions. In the way we've defined our measures, though, our particles are correlated, so it makes sense to use the **(symmetric) Dunkl transform** instead of the Fourier transform:

$$G_N^\beta(x_1, \cdots, x_N) = \int_{\mathbb{R}^N} B_N^\beta((x_1, \cdots, x_N), (t_1, \cdots, t_N)) \mu_N(t_1, \cdots, t_N).$$

The idea is that we replace the function that we're taking the expectation of, $e^{i(t_1 x_1 + \cdots + t_n x_n)}$, with the new function $B_N^\beta((x_1, \cdots, x_N), (t_1, \cdots, t_N))$, and this means that instead of implying convergence relative to moments and polynomials, we need convergence relative to something else. So instead of taking derivatives $\frac{\partial^k}{\partial x_1^k} + \cdots + \frac{\partial^k}{\partial x_n^k}$, we use the **Dunkl operator** P_k^β, and this is useful because in both cases we obtain $t_1^k + t_2^k + \cdots + t_N^k$.

So we can apply P_k^β to both sides of our definition of G_N^β: since B_N^β is an eigenfunction of P_k^β, we get an eigenvalue out, and then setting all $x_i = 0$, we get

$$\boxed{P_k^\beta(G_N^\beta)\Big|_{x_1 = \cdots = x_N = 0} = \mathbb{E}_{\mu_N}[t_1^k + \cdots + t_N^k]},$$

since the multivariate Bessel function B_N^β evaluates to 1 at $x = 0$. This means that we tie probabilistic information about the information μ_N to analytic information about the Dunkl transforms G_N. From here, we obtain an analog of Levy's continuity theorem, and we get results like

$$\lim_{N \to \infty} \mathbb{E}_{\mu_N} \left[\frac{t_1^k + \cdots + t_N^k}{N} \right] = m_k,$$

$$\lim_{N \to \infty, \beta \to 0, N\beta \to 2\gamma} \frac{1}{(s-1)!} \frac{\partial^s}{\partial x_1^s} (\ln(G_N^\beta)) \Big|_{x_1 = \cdots = x_N = 0} = \kappa_s.$$

This allows us to connect moments m_k to γ-cumulants κ_s, and we do so using the combinatorial relations that we mentioned earlier. And an important point is that Dunkl transforms are more natural than characterstic functions in our two examples above: for the Hermite N-particle β-ensemble, we have

$$G_N^\beta(x_1, \cdots, x_N) = \exp\left(\frac{x_1^2 + \cdots + x_N^2}{2} \right),$$

and when μ_N is the distribution of $c = a +_\beta b$, we have

$$G_N^\beta(x_1, \cdots, x_N) = \mathbb{E}_{\mu_N}[B_N^\beta(c, x)] = B_N^\beta(a, x) \cdot B_N^\beta(b, x).$$