Random Structures

Jan Olucha Fuentes

May 22, 2025

Contents

Notation and how to study these notes

Difficulty notations:

☕: denotes a proof or idea that is hard or hard to reproduce.

♪: denotes a proof or idea that requires partially unmotivated tools or machinery that is hard to pull out of thin air, but the rest of the argument is easy.

♩: denotes a proof or idea that is easy to see knowing some other results that help motivate it.

♫: denotes a proof or idea that is easy and no particularly clever ideas are needed. For revision purposes reading it a couple times will suffice.

Other notation:

$\lesssim$: same this as big O, i.e: $f \lesssim g$ iff $f = O(g)$ iff $f \leq Cg$

A quick word

This script consists of the lecture notes that I compiled during the academic year 2024-2025 at the University of Cambridge for Professor Wendelin Werner's course *Random Structures on finite dimensional space*. Naturally, none of the ideas, (nor the brilliant choice of topics and their presentation) in this script are mine, whereas all errors and typos certainly are. The object of this script is to present a beautiful story about randomness. When one starts to learn about probability, one first learns a few stories about general enumeration and combinatorics, and at some point, we realise that the interesting things to focus on really do occur when the *randomness* gets accumulated and goes to infinity. From these ideas we eventually get the Laws of Large Numbers, the Central Limit Theorem, Martingales, Markov Chains, etc. The key thing to realise, is that with these first *random structures* that we get, randomness piles up with time. Sometimes it can be that randomness comes in the form of a sequence of independent random variables, sometimes it comes with some structured dependence, such as a Markov Chain or a Martingale, but there always is this structure of causality and time. In this course, we explore randomness that accumulates not in a temporal way, but spatially, that is to say, we will explore random structures defined on spaces like $\mathbf{Z}^d$ or $\mathbf{R}^d$ for $d \geq 2$. Examples of the models we will explore include Percolation Theory, the Ising Model, Gaussian Free Field and others. These models are often quite simple to define, yet they present remarkably difficult problems, many of which, are still open questions. As one last comment from myself, as always in my notes, I wish the reader the best of luck, for the ideas in this script are certainly not easy. I hope these notes are of help to anyone who decides to read them, as they certainly were of help to me when studying this topic.

Your falsely,

JOF

Chapter 1

Percolation Theory

1.1 Percolation on $\mathbf{Z}^d$ and its phase transition

Definition 1.1 (Percolation Model) Let $G=(V,E)$ be a graph and $p\in(0,1)$. Consider a collection of independent and identically distributed Ber(p) random variables $\{\omega(x): x\in V\}$. If $\omega(x)=1$ for some $x\in V$, we say that the site is **open** or occupied. Otherwise its **closed** or unoccupied. We say that two open sites x and y are **connected**, denoted by $x\longleftrightarrow y$ if there exists a path $x=x_0\sim x_1\sim\ldots\sim x_n=y$ where each $\omega(x_i)=1$. By the **cluster** C_x of x we refer to

$$C_x=\{y\in V: y \text{ is open and } y\longleftrightarrow x\}$$

Remark 1.2 (A remark on measurability) We have to be a bit careful in regard to which statements we can even talk about. We can model our probability space as $(\{0,1\}^V,\mathscr{F},\mathbf{P}_p)$ where $\mathscr{F}$ is the sigma algebra generated by the π-system $\bigcup_{|S|<\infty}\underbrace{\sigma(\omega(x): x\in S)}_{\mathscr{F}_S}$, i.e: generated by all the events generated by finitely many of our random variables. That is to say $\mathscr{F}$ is the sigma algebra generated by events that depend on finitely many vertices, and finally $\mathbf{P}_p$ is the corresponding Ber$(p)^{\otimes\mathbf{Z}^d}$ product measure, where each coordinate is Bernoulli p.

The kinds of questions we are interested in asking are to do with the large scale connectivity properties of the Percolation model: do we see infinitely big clusters? Does every cluster have finite size? With what probability do these statements occur? We will from now on focus on the graph $\mathbf{Z}^d$. Let us define the following events: for $x\in\mathbf{Z}^d$, we define $A_x=\{$there is an infinite cluster at $x\}$. First of all, note that this statement is indeed measurable with respect to our σ-algebra:

$$A_x=\bigcap_{n\geq 0}\{y\longleftrightarrow x \text{ for some } y \text{ at distance } n \text{ from } x\}.$$

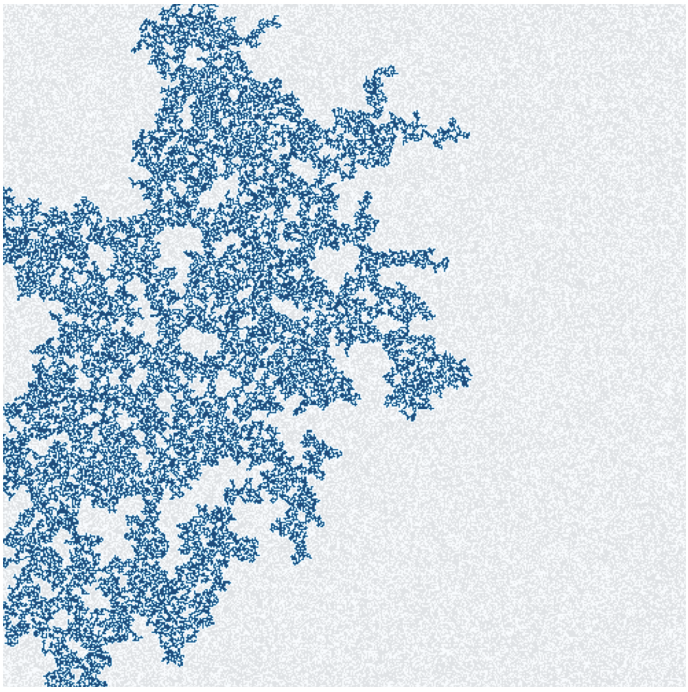


Figure 1.1: Simulation of the percolation model in Python, highlighted is the largest cluster

Naturally the events we are taking an intersection over only depend on finitely many vertices, as they can be determined by looking at the box $[-n:n]^d$, which contains only finitely many vertices. Then we can define the event $A = \{\text{there is an infinite cluster}\}$, and we can similarly see this is $\mathscr{F}$-measurable, for

$$A = \bigcup_{x \in \mathbf{Z}^d} A_x$$

Thus we see that we are safe to talk about the probability $\mathbf{P}_p[A]$. We now have some observations on $\mathbf{P}_p[A]$.

Proposition 1.3 (Preliminary facts of the probability of an infinite cluster) Let $\{A_x : x \in \mathbf{Z}^d\}$ and A be as above. Then we have the following.

- All the events A_x for $x \in \mathbf{Z}^d$ have the same probability.
- $\mathbf{P}_p[A] > 0$ if and only if $\mathbf{P}_p[A_0] > 0$.
- $\mathbf{P}_p[A] \in \{0, 1\}$.

Main idea: The first statement follows due to the fact that the law of the collection of all "coins" is translation invariant. The second claim follows simply by the fact that $A = \bigcup_x A_x$ and then using part 1. The third claim is due a to Kolmogorov's zero-one type-argument.

Proof of first claim. First note that since the random variables $\{\omega(x) : x \in \mathbf{Z}^d\}$ are i.i.d, the law of $\{\omega(x) : x \in \mathbf{Z}^d\}$ is the same law as that of $\{\omega(x+y) : x \in \mathbf{Z}^d\}$ for $y \in \mathbf{Z}^d$, and $\{\omega(x) : x \in \mathbf{Z}^d\} \in A_y$ if and only if $\{\omega(x+y) \in \mathbf{Z}^d\} \in A_0$. ♡

Proof of second claim. Note that if $\mathbf{P}_p[A_0] = 0$, then by the first claim, each A_x for $x \in \mathbf{Z}^d$ has zero probability under $\mathbf{P}_p$ and so by taking their countable union we also get an event of zero $\mathbf{P}_p$ probability. Conversely, if $\mathbf{P}_p[A] = 0$, given that $A_0 \subseteq A$ and hence $\mathbf{P}_p[A_0] \leq \mathbf{P}_p[A]$, we will also have that $\mathbf{P}_p[A_0] = 0$. ♡

Proof of third claim. Suppose $\mathbf{P}_p[A] \neq 0$ and recall that $\mathcal{G} = \bigcup_S \mathcal{F}_S$ (where the union is taken over finite subsets of our configuration space, and $\mathcal{F}_S$ is the sigma algebra generated by these vertices) is a π-system that generates $\mathcal{F}$. Let B be any set in $\mathcal{G}$. Since A is the event that there exists some infinite cluster, A is independent of whatever value ω takes on finitely many vertices, and so A is independent of B. From this we gather that the maps

$$B \mapsto \mathbf{P}_p[A \cap B] \qquad B \mapsto \mathbf{P}_p[A]\mathbf{P}_p[B]$$

agree for all $B \in \mathcal{G}$. However, it is easy to see that these two maps are measures, so by the uniqueness of measures lemma, it follows that they agree on the entire σ algebra $\mathcal{F}$. However, since $A \in \mathcal{F}$ it also follows that $\mathbf{P}_p[A \cap A] = \mathbf{P}_p[A]^2$. ♡

Remark 1.4 From this we have learnt that to understand $\mathbf{P}_p[A]$, all we need to do is understand what happens at the origin, and from this we immediately conclude the probability of A occurring. Of course, we can also note that when $p = 0$, $\mathbf{P}_p[A] = 0$, and when $p = 1$, we have that $\mathbf{P}_p[A] = 1$, so this could lead us to believe that there is some magical value of p for which suddenly $\mathbf{P}_p[A]$ jumps from 0 to 1, thus producing a **phase transition**.

Remark 1.5 Note that $\mathbf{P}_p[A_0]$ takes values in $[0,1]$, not just $\{0,1\}$ because it is not a $0-1$ type event, as it indeed does depend on the behavior of finitely meany vertices, for example, if there is a "cycle" around the origin of all closed vertices, we conclude A_0 does not occur.

Of course we haven't yet shown that $p \mapsto \mathbf{P}_p[A]$ just keeps bouncing up and down from zero to one without order, but if there is still a God in this world, this should not be the case: if we increase p, there is no reason to believe that A would become less likely. The key to formalising this argument is to construct the so-called monotone coupling.

Proposition 1.6 (Monotonicity of probability of infinite cluster) Let A be the event that there is an infinite cluster, then the function $\theta(p) = \mathbf{P}_p[A]$ is non-decreasing.

Main idea: Construct a coupling of the law $\mathbf{P}_p$ as a collection of indicator functions of the event that some Unif$[0,1]$ random variable exceeds the value p.

Proof. Define on a probability space $(\Omega, \Xi, \mathbf{Q})$ a family of independent identically distributed random variables $\{X(x) : x \in \mathbf{Z}^d\}$ that take uniform values in the interval $[0,1]$, and are indexed in the graph $\mathbf{Z}^d$. It is clear that the family of random variables $\{U_p(x) : x \in \mathbf{Z}^d\}$ defined by $U_p(x) = \mathbf{1}_{X(x) \leq p}$ has law $\mathbf{P}_p$. However, it is also clear that if $p \leq p'$, then $U_p(x) \leq U_{p'}(x)$ for all $x \in \mathbf{Z}^d$. Now finally we simply note that $U_p(x)$ takes values one or zero representing whether the vertex x is open or closed. Naturally, since $U_p(x) \leq U_{p'}(x)$, under $U_{p'}$ there will at least as many open vertices on $\mathbf{Z}^d$ that under U_p, therefore, since adding more open vertices cannot destroy an infinite cluster, it must be that if $U_p \in A$, then $U_{p'} \in A$, and so

$$\mathbf{P}_p[A] = \mathbf{Q}[U_p \in] \leq \mathbf{Q}[U_{p'} \in A] = \mathbf{P}_{p'}[A]$$

♡

Remark 1.7 (Critical probability) Having established monotonicity of the map $\theta(p)$, and noting again that $\theta(0) = 0$ and $\theta(1) = 1$, we conclude the existence of some value $p_c \in [0,1]$ for which $\theta(p)$ experiments a phase transition: it goes from almost surely not containing any infinite clusters, to almost surely containing some. Of course, we have yet not argued that p_c should not be either zero or one. It could be that as soon as your percolation parameter increases above zero, the infinitude of vertices will make it so that an infinite cluster will appear, or alternatively, it could just as well be the case that if the percolation parameter is even just a tiny bit below one, the amount of closed sites will already be sufficiently big to prevent any cluster. We however reach a much more interesting answer, there is indeed a special value somewhere in between that triggers the transition.

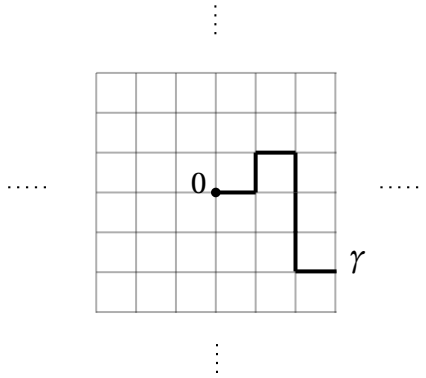


Figure 1.2: A self avoiding walk

Definition 1.8 With the function $\theta(p) = \mathbf{P}_p[A_0]$, we define the **critical probability** p_c as

$$p_c = \sup\{p : \theta(p) = 0\}$$

Of course this will be a function of the dimension. Naturally by definition we have that whenever $p < p_c$, $\mathbf{P}_p[A_0] = 0$ and whenever $p > p_c$, $\mathbf{P}_p[A_0] = 1$.

Theorem 1.9 (Non-triviality of critical probability) Consider the percolation model on $\mathbf{Z}^d$, $d \geq 2$. Then the value of $p_c = p_c(d)$ is in the interval $(0,1)$.

The heuristics are: if A_0 holds, then you have arbitrarily long open self avoiding paths, and if p is too small, this can't happen. Conversely, if A_0 does not hold, then you will find a closed loop of vertices, and this can't happen if p is large enough.

Main idea: (Lower bound) To prove p_c is bounded away from zero, we note that if A_0 holds, then for each n, one has a self-avoiding-path of length n consisting of open vertices. This gives that in fact for all $p < \frac{1}{2d}$, $\mathbf{P}_p[A_0] = 0$, which in turn implies that $p_c > 0$.

Proof that $p_c > 0$. Suppose A_0 holds, then for each $n \in \mathbf{N}$, there will be a self-avoiding walk γ of length n that only visits open vertices. (The reason why we want self-avoiding walks, is that if we didn't require this, the walk could just go in circles for as long as it wants without actually going far away). We define Ω_n to be the set of self-avoiding walks of length n. Then we have the

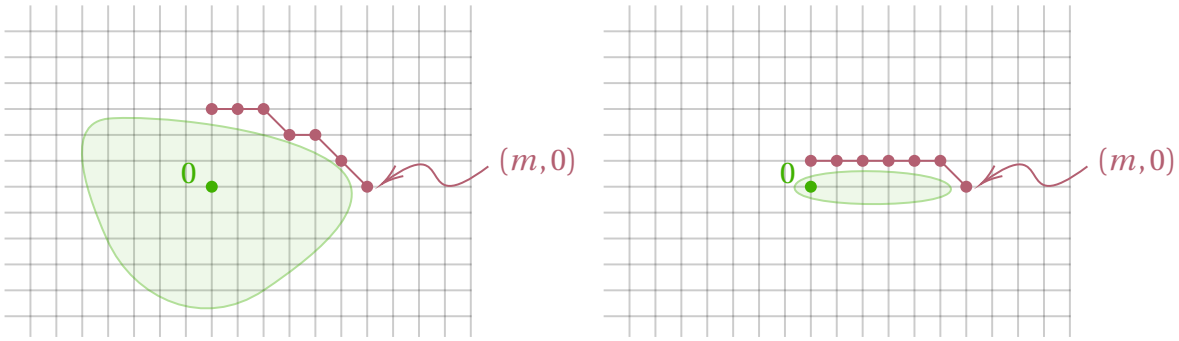


Figure 1.3: The diagram that says it all: proof that $p_c < 1$. Two possible cases of clusters about zero.

following simple calculation. For any $n \in \mathbf{N}$:

$$\mathbf{P}_p[A_0] \leq \mathbf{P}_p[\forall n, \text{there is some } \gamma \in \Omega_n \text{ with only open vertices}] \tag{1.1}$$

$$= \mathbf{P}_p\left[\bigcap_n \bigcup_{\gamma \in \Omega_n} \{\gamma \text{ only goes through open vertices}\}\right] \tag{1.2}$$

$$\leq \lim_{n\to\infty} \sum_{\gamma \in \Omega_n} \mathbf{P}_p[\gamma \text{ only goes through open vertices}] \tag{1.3}$$

$$\leq \lim_{n\to\infty} |\Omega_n| p^n \tag{1.4}$$

$$\leq \lim_{n\to\infty} (2dp)^n \tag{1.5}$$

Where step (3) comes from a union bound and the fact that the events

$$\bigcup_{\gamma \in \Omega_n} \{\gamma \text{ only goes through open vertices}\}$$

form a decreasing (in n) sequence of events. Step (4) comes from the fact that the configurations are iid Bernoulli p and so the probability that the entire γ consists only of open vertices equals p to the power of the length of γ. Finally step (5) comes from the brutal bound that the amount of self-avoiding walks of length n is at most $(2d)^n$. If $p < 1/2d$, then this quantity goes to zero as $n \to \infty$, therefore ♡

Main idea: (Upper bound) The key is that if A_0 does not hold, then you can find a "closed loop" around the origin. So in particular, we can find a path of closed sites of some length m. There are at most 8^m such paths and each path has probability $(1-p)^m$. By choosing p close enough to 1, one can make $\sum_{m\geq 0}(8(1-p))^m < 1$.

Proof that $p_c < 1$. We will proceed similarly, by showing that for all values of p larger than some value less than one, we will have that $\mathbf{P}_p[A_0] = 1$. Moreover, we will show this for the case that $d = 2$, because $\mathbf{Z}^2$ can be seen as a subgraph of $\mathbf{Z}^3$, and so if on this restricted plane there is a

path to infinity, then so will there be a path on the larger $\mathbf{Z}^d$. Now suppose A_0^c holds, meaning that the cluster at zero is finite, then there will be some $(m,0)$ for some m large enough for which $\omega(m,0)$ is closed. The idea now is to "circumvent the cluster". But since we don't know the shape of the cluster, we can only guarantee that we can get to a point with x-coordinate equal to zero, (since it could be that the open cluster about zero is only a thin strip of length $m-1$. That is to say, we can guarantee a self-avoiding-walk of length m (We can actually guarantee a larger walk, of at least $2n+4$ vertices, corresponding to the second picture in the figure that says it all, which would improve our bounds but we don't care). There are at most 8^m such walks, because at each point in the walk, you have 8 at most choices of vertices to move to (the ones at Euclidean distances 1 and $\sqrt{2}$). The probability that these are all closed is of $(1-p)^m$, and so

$$1-\mathbf{P}_p[A_0]=\mathbf{P}_p[A_0^c]\leq\sum_{m\geq 0}\big(8(1-p)\big)^m$$

If p is close enough to 1, this sum can be made strictly less than 1, and so it will follow that $\mathbf{P}_p[A_0]>0$, which by the Kolmogorov-type argument we did, means that $\mathbf{P}_p[A_0]=1$. ♡

Remark 1.10 We have therefore shown that percolation exhibits a phase transition. The range $p\in[0,p_c)$ is called the subcritical phase, $p=p_c$ is called critical percolation, and $p\in(p_c,1]$ is called the supercritical phase. It is actually a major open problem to show that $\mathbf{P}_{p_c}[A_0]=0$ in $\mathbf{Z}^d$ for $d=3$. Notice that even though $\theta(p)$ can be shown to be right-continuous, this does not imply $\theta(p_c)=1$, this is because p_c is defined as $\sup\{p>0:\theta(p)=0\}$.

Remark 1.11 (A remark on the self-avoiding-walk) On the Proof of Theorem 1.9, we invoked the seemingly obvious fact that if C_0 is infinite, then there is a self-avoiding-walk consisting of open sites that connects 0 with infinity. Why is this the case? Well, since C_0 is infinite, it must be that $C_0\setminus\{x\}$ contains an infinite connected component (indeed, if all the connected components of $C_0\setminus x$ were finite, then C_0 itself would be finite), such that at least one of the neighbours of x belongs to this new connected infinite component. If we call γ our self-avoiding-path, then we can set $\gamma(0)=0$, and $\gamma(1)$ to be this neighbour. Then we can repeat the procedure: since $\gamma(1)$ belongs to an infinite component $C(1)$, then $C(1)\setminus\gamma(1)$ contains a connected infinite component, etc.

1.2 Number of infinite clusters

In this section we will focus on the range of p for which there is $\mathbf{P}_p$ almost surely an infinite cluster. In this case, we are interested in asking the question of how many infinite clusters we can see. For the remaining of this chapter, define $N : \{0,1\}^{\mathbf{Z}^d} \to \mathbf{N} \cup \{0, \infty\}$ be the random variable that for a given configuration $\omega \in \{0,1\}^{\mathbf{Z}^d}$, returns the number of infinite clusters of that configuration. We will work towards the following goal:

Theorem 1.12 (Uniqueness of infinite cluster) Let p be in the range for which there is $\mathbf{P}_p$ almost surely an infinite cluster. Then $\mathbf{P}_p[\{\omega : N(\omega) = 1\}] = 1$, i.e: the infinite cluster is almost surely unique.

One of the ideas that we will use in the proof of this fact is that of translation invariance. This can be seen as a stronger statement than the one we saw to show that $\mathbf{P}_p[A] \in \{0,1\}$.

Definition 1.13 (Translation-Invariant Events) An event $A \in \mathscr{F}$ is said to be invariant under translation by $e = (1,0,\ldots,0) \in \mathbf{Z}^d$ if for any configuration $\omega \in \{0,1\}^{\mathbf{Z}^d}$, we have that

$$\big(\omega(x) : x \in \mathbf{Z}^d\big) \in A \iff \big(\omega(x+e) : x \in \mathbf{Z}^d\big) \in A$$

A clear example of such an event is the existence of an infinite cluster. A non-example of a translation-invariant event is the existence of an infinite cluster at the origin. As our intuition would tell us, translation-invariant events also satisfy a zero-one law.

Lemma 1.14 (Zero-one law for Translation-Invariant events) If $A \in \mathscr{F}$ is translation invariant by e, then $\mathbf{P}_p[A] \in \{0,1\}$.

The proof of this Lemma in turn relies on this approximation fact, which is intuitively true:

Lemma 1.15 (Approximation Lemma) Let $A \in \mathscr{F}$. Then for any $\epsilon > 0$ there exists some $B \in \mathscr{G} = \bigcup_S \mathscr{F}_S$, where S is taken over finite subsets of the vertices, such that

$$\mathbf{P}_p[A \Delta B] < \epsilon.$$

Proof. Let $\mathscr{G}_n$ be the σ-algebra generated by the events that can be determined by observing the $[-n,n]^d$ box, i.e: $\mathscr{G}_n = \sigma\big(\omega(x) : x \in [-n,n]^d\big)$. Now let A be the event to be approximated. Consider the following Martingale:

$$(\mathbf{P}_p[A \mid \mathscr{G}_n])_n$$

It is clear that this Martingale is $\mathscr{L}^1$ bounded. Indeed: $\mathbf{E}_p[|\mathbf{P}_p[A \mid \mathscr{G}_n]|] = \mathbf{P}_p[A] \leq 1$. Therefore by the UI Martingale Convergence Theorem, we have that almost surely,

$$\mathbf{P}_p[A \mid \mathscr{G}_n] \to \mathbf{P}_p[A \mid \mathscr{F}] = \mathbf{1}(A)$$

Now we can construct the following approximating sets:

$$A_n = \{\mathbf{P}_p[A \mid \mathscr{G}_n] > 1/2\}$$

Then we have that $\lim_{n\to\infty} \mathbf{1}(A_n) = \mathbf{1}(A)$ almost surely and in $\mathscr{L}^1$. Indeed: by almost sure convergence, we have that for n large enough, $|\mathbf{P}_p[A \mid \mathscr{G}_n] - \mathbf{1}_A|$ is very very small with probability 1. Since $\mathbf{1}(A)$ can only take the values zero and 1, it follows that for n large enough, $\mathbf{P}_p[A \mid \mathscr{G}_n] > 1/2$ if and only if $\mathbf{1}_A = 1$ with probability 1. From this we have that $\mathbf{1}_{A_n} \to \mathbf{1}_A$ almost surely and in $\mathscr{L}^1$. Now finally we just note that

$$\mathbf{P}_p[A_n \Delta A] = \mathbf{E}_p|\,\mathbf{1}(A_n) - \mathbf{1}(A)| \to 0.$$

♡

Remark 1.16 (Why symmetric difference?) It may not be clear at the start why a small symmetric difference indeed means that our set can be well-approximated. Simply note that the following relation holds in general: $\mathbf{P}[A] = \mathbf{P}[A \cap B] + \mathbf{P}[A \setminus B]$. And similarly for B, therefore, if our symmetric difference is small, then we can write

$$\begin{aligned} |\mathbf{P}_p[A] - \mathbf{P}_p[B]| &= |\mathbf{P}_p[A \cap B] + \mathbf{P}_p[A \setminus B] - \mathbf{P}_p[A \cap B] - \mathbf{P}_p[B \setminus A]| \\ &= |\mathbf{P}_p[A \setminus B] - \mathbf{P}_p[B \setminus A]| \\ &\leq |\mathbf{P}_p[A \setminus B] + \mathbf{P}_p[B \setminus A]| \\ &= |\mathbf{P}_p[A \Delta B]| < \epsilon. \end{aligned}$$

We can actually already see this statement from the last step of our proof above. Since we showed $\mathbf{E}|\,\mathbf{1}(A_n) - \mathbf{1}(A)| \to 0$ and $|\mathbf{P}_p[A_n] - \mathbf{P}_p[A]| = |\mathbf{E}[\mathbf{1}(A_n) - \mathbf{1}(A)]| \leq \mathbf{E}|\,\mathbf{1}(A_n) - \mathbf{1}(A)|$

We are now ready to prove the zero-one law for translation invariant events:
Main idea: By using the approximation Lemma, show that $\mathbf{P}_p[A] \leq \mathbf{P}_p[A]^2$.

Proof of Lemma 1.14. Let $A \in \mathscr{F}$ be a translation-invariant event. Denote by τ_x the translation map by x units in direction $(1, 0, \cdots, 0)$. By hypothesis of translation-invariance, $\mathbf{P}_p[A] = \mathbf{P}_p[\tau_x(A)]$. Then, let $\epsilon > 0$ be given, and by the Approximation Lemma 1.15, choose an event B that depends

on the sites of a finite subset $E \subset \mathbf{Z}^d$, such that $\mathbf{P}_p[A \Delta B] < \epsilon$. Since B depends only on a finite set E of vertices, we can find some x large enough, so that $E \cap \tau_x(E) = \varnothing$, so that in particular, the translated event $\tau_x(B)$ is independent of B. Then:

$$\begin{aligned}
\mathbf{P}_p[A] &= \mathbf{P}_p[A \cap A] && (1.1)\\
&= \mathbf{P}_p[A \cap \tau_x(A)] && (1.2)\\
&\leq \mathbf{P}_p[B \cap \tau_x(B)] + 2\epsilon && (1.3)\\
&= \mathbf{P}_p[B]\mathbf{P}_p[\tau_x(B)] + 2\epsilon && (1.4)\\
&= \mathbf{P}_p[B]^2 + 2\epsilon && (1.5)\\
&\leq \mathbf{P}_p[A]^2 + 4\epsilon + \epsilon^2. && (1.6)
\end{aligned}$$

Where step 1.2 is due to the fact that $A = \tau_x(A)$, and 1.5 is due to the fact that the measure $\mathbf{P}_p$ itself is translation invariant. Thus sending $\epsilon \to 0$ gives that $\mathbf{P}_p[A] \leq \mathbf{P}_p[A]^2$ which means $\mathbf{P}_p[A] \in \{0, 1\}$. ♡

Now we can actually rule out most possibilities of the number of clusters that there could almost surely be:

Corollary 1.17 Letting $N : \{0,1\}^{\mathbf{Z}^d} \to \mathbf{N} \cup \{0 \cup \infty\}$ be the number of clusters of a given percolation configuration, we have that either

$$\mathbf{P}_p[N = 0] = 1 \quad \text{or} \quad \mathbf{P}_p[N = 1] = 1 \quad \text{or} \quad \mathbf{P}_p[N = \infty] = 1$$

Main idea: The key here is that if we are working on the event that there are finitely many clusters, we can intersect all of them with a large enough box, and then by resampling within this new box, there is a positive probability of merging all these finitely many clusters into one big cluster.

Proof. For notational simplicity, let $\mathscr{E}_k$ be the event that $N = k$. Suppose that $\mathbf{P}_p[\mathscr{E}_k] = 1$ for some $k \geq 2$ finite. Let F_n be the event that all clusters intersect the box $[-n, n]^d$. Clearly, we have that

$$\mathbf{P}_p[\mathscr{E}_{\leq 1}] \geq \mathbf{P}_p\left[F_n \cap \{\omega(x) = 1 \text{ for all } x \in [-(n-1), n-1]^d\}\right]$$

Indeed, if all the clusters, however there may be, intersect the box $[-n, n]^d$, and all the sites in the $[-(n-1), n-1]^d$ box are open, then all clusters are connected so there is either one cluster or no cluster. However, whether the infinite cluster reaches the $[-n, n]^d$ box is independent to whatever happens in the smaller box (think of the infinite cluster coming from infinity into the graph, while the smaller box changes configurations, the probability that the cluster stops "growing inwards"

before reaching the n-box is independent of whatever happens inside the smaller box), therefore

$$\mathbf{P}_p[\mathscr{E}_{\leq 1}] \geq \mathbf{P}_p[F_n] p^{\text{something}}$$

Since we are assuming $\mathbf{P}_p[\mathscr{E}_k] = 1$, we can however choose an n large enough so that $\mathbf{P}_p[F_n] > 0$. Therefore $\mathbf{P}_p[\mathscr{E}_{\leq 1}] > 0$, which means $\mathbf{P}_p[\mathscr{E}_{\leq 1}] = 1$, but it is not possible that $\mathbf{P}_p[\mathscr{E}_{\leq 1}] = \mathbf{P}_p[\mathscr{E}_k] = 1$. Therefore either $\mathbf{P}_p[\mathscr{E}_\infty] = 1$ or $\mathbf{P}_p[\mathscr{E}_{\leq 1}] = 1$. ♡

Remark 1.18 A cleaner proof is by Harris Inequality.

We are now ready to finish off the proof of the unicity of the infinite cluster. The only thing we need is to rule out the possibility of there being infinite clusters. Of course, it would be outrageous to believe in the almost sure infinitude of clusters, the philosophy being that in $\mathbf{Z}^d$, there is simply not enough space to accommodate too many disjoint infinite clusters. We will need one last definition:

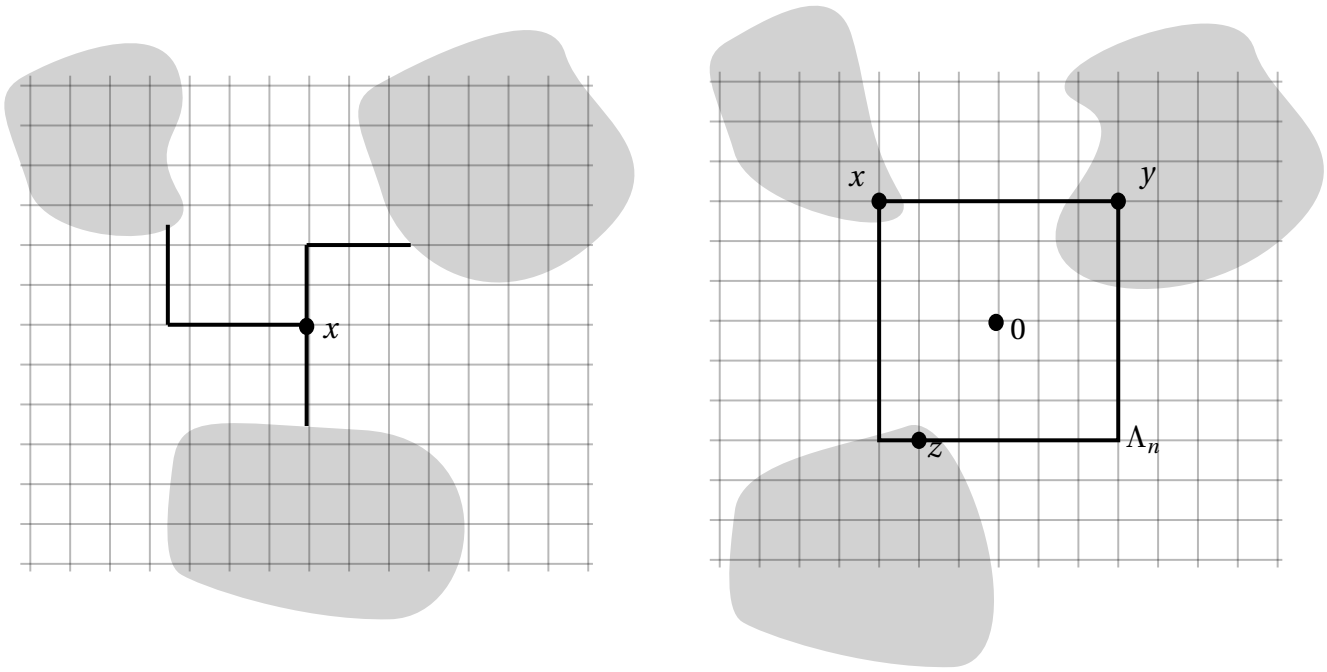


Figure 1.4: A trifurcation on the left, and the diagram that says it all on the right

Definition 1.19 (Trifurcation) Let $\omega \in \{0,1\}^{\mathbf{Z}^d}$ be a configuration. A vertex $x \in \mathbf{Z}^d$ is said to be a trifurcation of ω if x is contained in an infinite cluster C, and $C \setminus \{x\}$ is comprised of three disjoint infinite clusters.

Main idea: The key idea is that if there are infinitely many clusters, then there is a positive probability that the origin is a trifurcation. However, by an isoperimetric argument, that uses the fact that there cannot be more trifurcation points in Λ_n than there are points in $\partial\Lambda_n$, the probability that the origin is a trifurcation is actually zero.

Proof of Theorem 1.12. All there is to show is that $\mathbf{P}_p[\mathscr{E}_\infty] = 0$. Suppose there are indeed infinitely many clusters, we are going to reach a contradiction. Given $\mathscr{E}_\infty$ holds, it is clear that for any fixed k,

$$\mathbf{P}_p\left[\bigcup_n \{k \text{ clusters intersect } [-n,n]^d\}\right] = 1$$

Which by taking out the union as a limit means that there is some $n(k)$ large enough so that $\mathbf{P}_p\left[\{k \text{ clusters intersect } [-n,n]^d\}\right] \geq 1/2$. (The one half is irrelevant, we just want some large enough number) Let us call this event as F for simplicity. An important observation to make is that the event F is independent of whatever happens in the interior of the box $\Lambda_n := [-n,n]^d$ (you may modify the local structure of the clusters if you alter the interior of the box, but the fact that k clusters touch the box will remain intact, for a cluster cannot touch the box if it doesn't

touch its boundary). With this powerful observation in mind, we proceed to choose k in a clever way. Choose k large enough so that there are three vertices in the boundary of the box $\partial\Lambda_n$, say x, y, z that are a distance at least 3 apart from each other, and that are all connected to infinity, of course, with this choice of k and n, we still have that $\mathbf{P}_p[F] \geq 1/2$. The idea is that we are now going to modify the interior of the box to create a trifurcation. Say by connecting x, y, z in any trifurcation like path of open sites to the origin, and closing everything else, call T' the event that this construction holds. Let $T(0)$ denote the event that we have a trifurcation at the origin, then we obviously have that

$$\mathbf{P}_p[T(0)] \geq \mathbf{P}_p[T(0) \cap F] \geq \mathbf{P}[T' \cap F] = p^{\text{something}}(1-p)^{|\Lambda_n| - \text{something}}\mathbf{P}[F] > 0$$

The punchline will come now when we show that actually $\mathbf{P}_p[T(0)] = 0$, which will give us our desired contradiction. The observation to make is that clearly, the event $T(x)$ - that $x \in \mathbf{Z}^d$ is a trifurcation - is independent of x. So the expected number of trifurcations in the box Λ_n is precisely $\mathbf{P}_p[T(0)] \times |\Lambda_n|$. On the other hand, we have the observation (This should really require a proof of its own, but we just claim that it is intuitive enough) that there cannot be more trifurcations in Λ than points on the boundary $\partial\Lambda_n$ (Intuitively, no matter how cleverly you try to make a trifurcation, you will always end up having to make one "arm" of the trifurcation leave to infinity in a new path to infinity that has not been used before, so a new point of the boundary has to be crossed). Therefore, combining all this:

$$\mathbf{P}_p[T(0)] = \frac{\mathbf{E}\#\{\text{trifurcations in } \Lambda_n\}}{|\Lambda_n|} \leq \frac{|\partial\Lambda_n|}{|\Lambda_n|} \to 0$$

as $n \to \infty$, since the number of vertices in the box grows like $O(n^d)$ whereas the size of the boundary grows like $O(n^{d-1})$. ♡

Remark 1.20 (On the ommitted step) In the proof we claimed that there couldn't be more trifurcation points inside a box than there were points on its boundary. Let us give a more clear but still not rigorous explanation of this fact: Suppose that we have an already existing trifurcation in our box, and we wish to add a new trifurcation. We have a few cases to consider:

A The new trifurcation point y is in one of the prior "arm", in which case, is has already two open neighbours, which when y is removed, will each be in an infinite cluster disjoint from the other. Therefore the third arm of y has to be a path going to infinity, disjoint from the already existing connected component, thus occupying one more point in the boundary.

B The new trifurcation point z is on "empty space", then it can either have all its arms going off to infinity, opening 3 new vertices in the boundary, or one of its arms can be connected

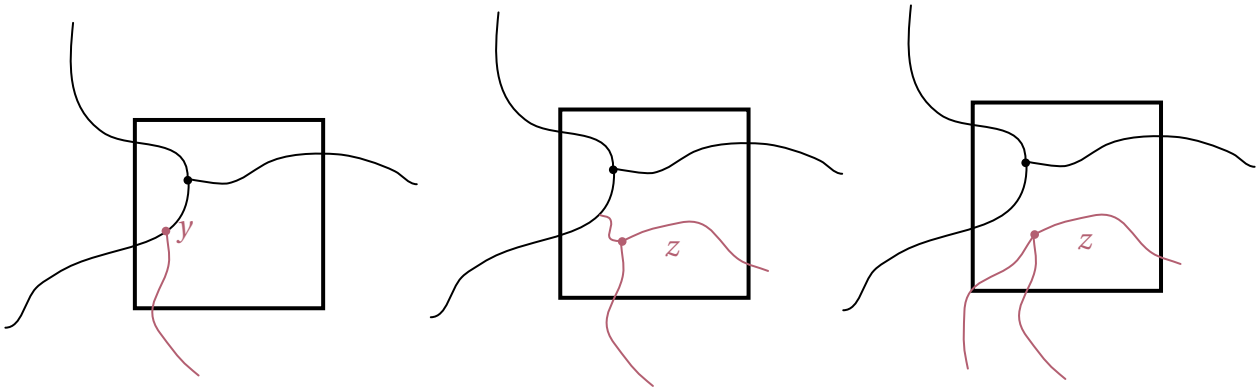


Figure 1.5: The diagram that says it all of Remark 1.20

to the already existing component, in which case we have 2 new paths to infinity.

An important comment to make is that of course this proof only holds because we are in $\mathbf{Z}^d$, in particular, we used the isoperimetric fact of the ratio between the boundary and size of the box Λ_n. As a fun final remark, note that on trees, this result is not true.

Summary of the proof:

1. We first show that if there are a finite number of infinite clusters, we may merge them together and thus have one or zero infinite clusters.

2. To show that there cannot be infinitely many clusters, we show that if there are infinitely many clusters, we can deduce that the probability of a trifurcation at the origin is non-zero by doing a "surgery" at a small box around the origin, and then contrast this with the observation that by an isoperimetric argument, the probability of a trifurcation at the origin is indeed of zero.

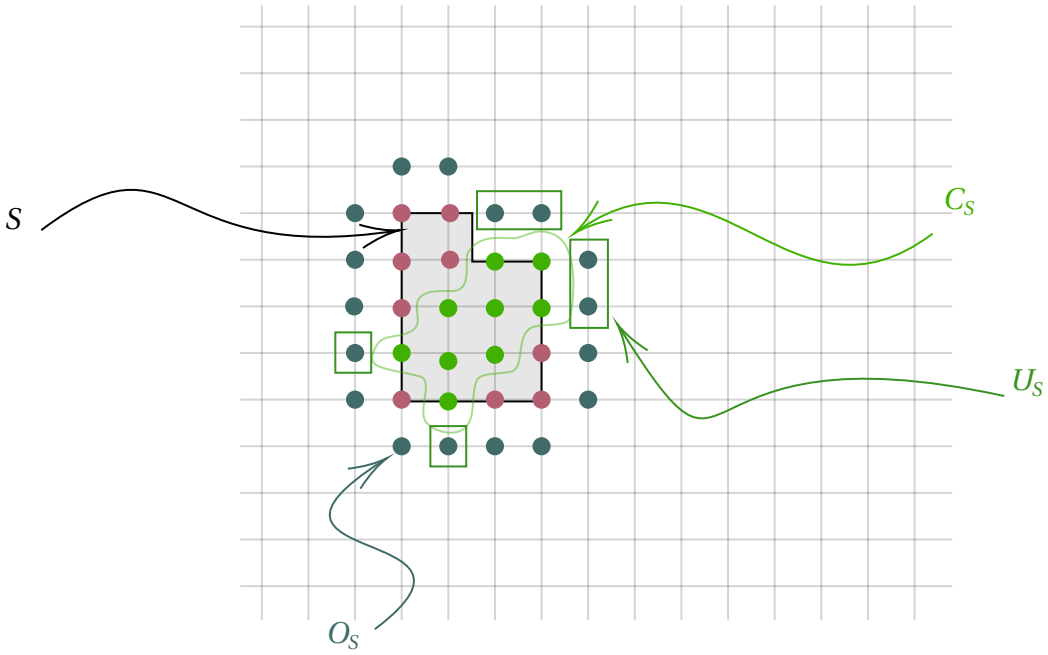
1.3 Exponential decay in the subcritical regime

In the subcritical regime, i.e: $p < p_c$ we know that almost surely there will be no infinite cluster, metaphorically, a big ocean of closed sites with some smaller islands of open resistance. How big will these islands be? Can we give bounds on the probability that their radius is at least some number n? How will this probability decay? The title of this section should give a spoiler as to what to expect. We will work towards proving the following goal:

Theorem 1.21 (Exponential decay of subcritical percolation) If p is in the subcritical regime, then there is some $\psi = \psi(p)$ for which

$$\mathbf{P}_p[0 \longleftrightarrow \partial\Lambda_n] \lesssim \exp(-\psi n)$$

Before we start proving this Theorem, we need to introduce a bunch of notation, please observe the following diagram:



Let $\mathscr{S}$ be the set of sets $S \subseteq \mathbf{Z}^d$ that are finite and connected, that contains the origin, and with $\mathbf{Z}^d \setminus S$ being itself connected this is a discrete analogue of a simply connected finite domain, so for example, S cannot be the boundary of a box). We then perform percolation on S, and denote C_S to be the **C**luster containing the origin, O_S denotes the **O**utside points of distance 1, finally U_S denotes the neighbours of the cluster C_S in O_S, i.e: U_S is the set of available exit points of the cluster. We then define $\varphi_p(S)$ to be the expected of such exit points, i.e: $\varphi_p(S) = \mathbf{E}_p[\#U_S]$. Finally, note that if D is a subset of S that contains the origin and is connected, the event that $C_S = D$ depends only on the value of the percolation inside D (namely all the sites in D must be open) as well as the values of the percolation on the neighbours of D (in the sense that the neighbours must all be closed). Therefore if we let

$\tilde{D} = \tilde{D}(D,S)$ denote the union of D with its neighbours in S, we see that $\{C_S = D\} \in \sigma\big(\omega(x) : x \in \tilde{D}\big)$. The main idea in the proof of this result will be a close connection between $\varphi_p(S)$ being bounded from above or below for large S, and the regime being supercritical or subcritical.

Lemma 1.22 If there is some $S \in \mathscr{S}$ such that $\varphi_p(S) < 1$, then $\mathbf{P}_p[0 \longleftrightarrow \partial\Lambda_n]$ decays exponentially in n.

Main idea: What the calculation we are going to do now is morally telling us is that since we can fit S inside some box Λ_{n_0}, for all $n \geq n_0$, we have that if $0 \longleftrightarrow \partial\Lambda_n$, then for each possible exit point of S, we must be able to move a distance of $n - n_0$. This is the content of Diagram 6.1. This will give that $\mathbf{P}_p[0 \longleftrightarrow \partial\Lambda_n] \leq \mathbf{E}[\#U_S]\mathbf{P}_p[0 \longleftrightarrow \partial\Lambda_{n-n_0}]$, so by iterating this we get the exponential decay, since we are assuming that $\mathbf{E}[\#U_S] < 1$.

Proof. Start by fixing some n_0 so that $S \subseteq \Lambda_{n_0-1}$ and choose $n \geq n_0$. We make the following observation (**OBS 1**): if $0 \longleftrightarrow \partial\Lambda_n$, then we have a path γ of open sites started at 0 that reaches $\partial\Lambda_n$. In particular, by assumption of $n \geq n_0$, there is a point after which γ ceases forever to be inside C_S. Say $x = \gamma(j) \in C_S$ and then onwards we have that γ stays outside C_S. In particular, we have that after this point γ stays out of $\widetilde{C}_S = C_S \cup \{\text{neighbours of } C_S \text{ inside } S\}$ (it may actually return to S, but not to C_S or any of its neighbours by definition of x being the last one). Moreover, we have the second rather crude observation (**OBS 2**) that for a set D, percolation on D is independent

to percolation on $\widetilde{D}^c$. With this we can now condition on how C_S looks:

$$\mathbf{P}_p[0 \longleftrightarrow \partial\Lambda_n] = \sum_{D\subseteq S} \mathbf{P}_p[\{C_S = D\} \cap \{0 \longleftrightarrow \partial\Lambda_D\}] \tag{1.1}$$

$$= \sum_{D\subseteq S} \mathbf{P}_p\left[\{C_S = D\} \cap \bigcup_{y\in O_S} \{0 \longleftrightarrow y\} \cap \left\{y \overset{\widetilde{D}^c}{\longleftrightarrow} \partial\Lambda_n\right\}\right] \tag{1.2}$$

$$\leq \sum_{D\subseteq S} \sum_{y\in O_S} \mathbf{P}_p\left[\{0 \longleftrightarrow y\} \cap \{C_S = D\} \cap \left\{y \overset{\widetilde{D}^c}{\longleftrightarrow} \partial\Lambda_n\right\}\right] \tag{1.3}$$

$$= \sum_{D\subseteq S} \sum_{y\in O_S} \mathbf{P}_p\left[\{0 \longleftrightarrow y\} \cap \{C_S = D\}\right] \mathbf{P}_p\left[y \overset{\widetilde{D}^c}{\longleftrightarrow} \partial\Lambda_n\right] \tag{1.4}$$

$$\leq \sum_{D\subseteq S} \sum_{y\in O_S} \mathbf{P}_p\left[\{0 \longleftrightarrow y\} \cap \{C_S = D\}\right] \mathbf{P}_p\left[y \longleftrightarrow \partial\Lambda_n\right] \tag{1.5}$$

$$\leq \sum_{D\subseteq S} \sum_{y\in O_S} \mathbf{P}_p\left[\{0 \longleftrightarrow y\} \cap \{C_S = D\}\right] \mathbf{P}_p\left[y \longleftrightarrow \partial\Lambda_{n-n_0} + y\right] \tag{1.6}$$

$$= \sum_{D\subseteq S} \sum_{y\in O_S} \mathbf{P}_p\left[\{0 \longleftrightarrow y\} \cap \{C_S = D\}\right] \mathbf{P}_p\left[0 \longleftrightarrow \partial\Lambda_{n-n_0}\right] \tag{1.7}$$

$$= \mathbf{P}_p[0 \longleftrightarrow \partial\Lambda_{n-n_0}] \sum_{y\in O_S} \mathbf{P}_p[0 \longleftrightarrow y] \tag{1.8}$$

$$= \mathbf{P}_p[0 \longleftrightarrow \partial\Lambda_{n-n_0}] \mathbf{E}[\#\{y \in O_S : 0 \longleftrightarrow y\}] \tag{1.9}$$

$$\leq \mathbf{P}_p[0 \longleftrightarrow \partial\Lambda_{n-n_0}] \mathbf{E}[\#U_S] \tag{1.10}$$

This clearly completes the proof by assumption that $\varphi_p(S) := \mathbf{E}[\#U_S] < 1$ by iterating it. Now let us justify the steps. Step (1.1) is trivial. Step (1.2) comes from **(OBS 1)**. Step (1.3) is a union-bound. Step (1.4) uses **(OBS 2)**. Step (1.5) uses the fact that if we restrict on y reaching $\partial\Lambda_n$ strictly via $\widetilde{D}^c$, then we obviously have that y reaches $\partial\Lambda_n$. Step (1.6) uses the fact that y is in O_S and that $S \subseteq \Lambda_{n_0}$ (you may wish to look at the diagram). Step (1.7) comes from the fact that $\mathbf{P}_p$ is a translation-invariant measure. Step (1.8) is trivial. Step (1.9) is trivial. Step (1.10) comes from the fact that the number of vertices in O_S that are connected to 0 are no more than the possible exit points (the subtlety is that vertices in U_S need not be open, whereas the set that we are taking an expectation of in step (1.9) need to be open). ♡

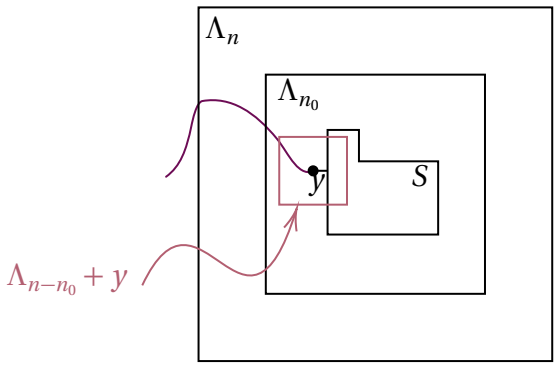


Figure 1.6: Diagram of Step (1.6) in the proof above

Before continuing with the proof of the Theorem of this section, we need one more tool: Russo's Formula.

Definition 1.23 (Increasing event) An event A that depends on the outcomes of percolation on finitely many vertices - i.e: say $A \in \sigma(\omega(x) : x \in \Lambda_n)$ for some n - is said to be increasing if whenever $\omega \in A$, and $\omega \leq \omega'$, then we also have that $\omega' \in A$.

There's this observation we can notice: consider the coupling ω_p introduced before, namely, a probability space $(\Omega, \Xi, \mathbf{Q})$ in which a collection of iid uniform $[0,1]$ random variables $(X(x) : x \in \mathbf{Z}^d)$ exist, and we define $(\omega_p(x) : x \in \mathbf{Z}^d)$ to be the collection of random variables with

$$\omega_p(x) = \mathbf{1}(X(x) \leq p)$$

We can think of increasing p just a little bit. What would be required for ω_p to not belong to A but suddenly, after having increased p by just a little bit, it belonged to A. Clearly, some vertices must have switched on, and some of those must have been vital, i.e: without those specific ones, it might not have been possible for ω_p to suddenly belong to A. We refer to these special vertices as pivotal. In some sense, they are the vertices that hold the key to the outcome.

Definition 1.24 (Pivotal vertex) Let $\omega \in \{0,1\}^{\mathbf{Z}^d}$ and let A be an event that depends on finitely many vertices. We say that a vertex v is A-pivotal for ω if whenever ω with v closed does not belong to A, but ω with v open does belong to A. With symbols we may write something like

$$\omega^{v,0} \notin A, \text{ but }, \omega^{v,1} \in A.$$

Naturally, the more pivotal vertices there are, the more sensitive we expect the probability of $\omega_p \in A$ being, because there will be more chance that one of the key vertices will flip. This intuitive relation is formulated in Russo's Formula, and the proof will reflect very clearly this intuition too.

Proposition 1.25 (Russo's Formula) Let A be an increasing event that depends on the outcomes on some box Λ_n. Then we have that

$$\frac{\mathrm{d}}{\mathrm{d}p}\mathbf{P}_p[A] = \mathbf{E}_p[\#\{A\text{-pivotal vertices}\}]$$

Proof. Once again consider the coupling $(\omega_p(x) : x \in \mathbf{Z}^d)$ defined on the common probability space $(\Omega, \Xi, \mathbf{Q})$. Then

$$\mathbf{P}_{p+\epsilon}[A] - \mathbf{P}_p[A] = \mathbf{Q}[\{\omega_{p+\epsilon} \in A\} \cap \{\omega_p \notin A\}]$$

As outlined above, it must be that some pivotal sites have "turned on", and so we can start by decomposing this probability in terms of the pivotal sites. Of course it could be that a large number of sites were pivotal meaning that all needed to be turned on, so we will decompose probability as follows: (the sums are taken over distinct values in Λ^n)

$$\begin{aligned}\mathbf{Q}[\{\omega_{p+\epsilon} \in A\} \cap \{\omega_p \notin A\}] = &\sum_{x \in \Lambda_n} \mathbf{Q}[\{\omega_{p+\epsilon} \in A\} \cap \{\omega_p \notin A\} \cap \{x \text{ is pivotal}\}] \\ &+ \sum_{x,y \in \Lambda_n} \mathbf{Q}[\{\omega_{p+\epsilon} \in A\} \cap \{\omega_p \notin A\} \cap \{(x,y) \text{ are pivotal}\}] \\ &+ \sum_{x,y,z \in \Lambda_n} \mathbf{Q}[\{\omega_{p+\epsilon} \in A\} \cap \{\omega_p \notin A\} \cap \{(x,y,z) \text{ are pivotal}\}] \\ &+ \cdots\end{aligned}$$

(There's a little subtlety here: when we say $\{x \text{ is pivotal}\}$ we mean that if x is off, then A doesn't hold. If x is on, then A holds. However, when we say $\{(x,y,z) \text{ is pivotal}\}$, we mean that if all x, y, z are on, then A holds, but if one of them fails, then A doesn't hold. Note therefore that the events $\{x \text{ is pivotal}\}$ and $\{y \text{ is pivotal}\}$ are disjoint). Observe that if say $(x_1, \cdots, x_k)$ are the pivotal vertices, and we are also working on the event that $\omega_p \notin A$ but $\omega_{p+\epsilon} \in A$, then this is equivalent to saying that have all switched on as the probability increased. Formally, what this means is that

$$\begin{aligned}&\mathbf{Q}[\{\omega_{p+\epsilon} \in A\} \cap \{\omega_p \notin A\} \cap \{(x_1, \cdots, x_k) \text{ are pivotal}\}] \\ &= \mathbf{Q}[\{X(x_1) \in [p, p+\epsilon]\} \cap \cdots \cap \{X(x_k) \in [p, p+\epsilon]\} \cap \{(x_1, \cdots, x_k) \text{ are pivotal}\}] \\ &= \epsilon^k \mathbf{Q}[(x_1, \cdots, x_k) \text{ are pivotal}] = O(\epsilon^k) = O(\epsilon^2)\end{aligned}$$

where this last equality is if $k \geq 2$. Therefore we have that

$$\mathbf{Q}[\{\omega_{p+\epsilon} \in A\} \cap \{\omega_p \notin A\}] = \sum_{x \in \Lambda_n} \epsilon \mathbf{Q}[x \text{ is pivotal}] + O(\epsilon^2) \tag{1.1}$$

Figure 1.7: An illustrative example of $\{x \text{ is pivotal}\}$ (left) and an example of $\{(x,y) \text{ are pivotal}\}$ (right), for the event that the two sausages are connected.

Where we used the fact that $\mathbf{Q}[X(x) \in [p, p+\epsilon], \{x \text{ is pivotal}\}] = \epsilon \mathbf{Q}[x \text{ is pivotal}]$ because the event of x being pivotal can be determined without knowing the value of x, hence $\{x \text{ is pivotal}\}$ is independent of $X(x)$. Putting this all together:

$$\frac{\mathbf{P}_{p+\epsilon}[A] - \mathbf{P}_p[A]}{\epsilon} = \sum_{x \in \Lambda_n} \mathbf{Q}[x \text{ is pivotal}] + O(\epsilon)$$

which when taking $\epsilon \downarrow 0$ gives the desired claim for the right derivative. The left derivative follows similarly. ♡

We are now ready to finish the proof of the Theorem.

Proof of Theorem 1.21. We will finish proving the Theorem by showing the following:

$$\text{if } \inf_{S \in \mathscr{S}} \varphi_{p_0}(S) > 0 \text{ then } p_0 \geq p_c.$$

The conclusion of the Theorem will indeed hold, because if $\inf_{S \in \mathscr{S}} \varphi_{p_0}(S) < 1$, then by the Lemma, we have exponential decay, and if $\inf_{S \in \mathscr{S}} \varphi_{p_0}(S) \geq 1$, then in particular it will be strictly greater than zero and so $p_0 \geq p_c$, i.e: we are not subcritical. For this we make the following observation **(OBS 1)**: the map $p \mapsto \varphi_p(S)$ is non-decreasing (larger p, more exit sites out of S), so if $\inf_{S \in \mathscr{S}} \varphi_{p_0}(S) > 0$, then there is some $\alpha > 0$ so that for any $p \geq p_0$ and for any $S \in \mathscr{S}$ we also have that $\varphi_p(S) \geq \alpha$. Now let $p_1 \in (p_0, 1)$, we will show that $\mathbf{P}_{p_1}$-almost surely, there is an infinite connected cluster. We start by using Russo's Formula:

$$\frac{\mathrm{d}}{\mathrm{d}p} u_n(p) = \mathbf{E}_p \#\text{pivotal sites for } u_n \tag{1.1}$$

$$= \frac{1}{1-p} \mathbf{E}_p \#\text{ closed pivotal sites for } u_n \tag{1.2}$$

where we have defined $u_n(p) = \mathbf{P}_p[0 \longleftrightarrow \partial\Lambda_n]$ and for convenience we also let u_n be the event $\{0 \longleftrightarrow \partial\Lambda_n\}$. Step (1.1) is Russo's formula and step (1.2) comes from the fact that a site x being pivotal is independent of the value at the site, so that

$$\mathbf{E}_p \#\text{closed pivotal sites for } u_n = \sum_{x \in \Lambda_n} \mathbf{E}_p \, \mathbf{1}_{\omega(x)=0} \mathbf{1}_{x \text{ is pivotal}}$$

and now you split the expectation using independence. The idea is that we will explore what's

happening at Λ_n from the boundary inwards: let U be the random set of all points in Λ_n that are connected to the boundary. Then the event $\{U = V\}$ is measurable with respect to V (whose sites must all be open) and the neighbours of V in Λ_n. We call the union of these two sets $\tilde{V}$. Of course, we still haven't discovered what happens on the complement of $\tilde{V}$. Define now $S(V)$ to be the connected component containing the origin in $\tilde{V}^c$ (don't get confused, this is not the open connected cluster, just the connected component, see diagram). We will now decompose $\mathbf{E}_p\#$ closed pivotal sites for u_n in terms of the possible values that U can take:

$$\mathbf{E}_p\#\text{ closed pivotal sites for } u_n = \sum_{0\notin V} \mathbf{P}_p[U=V]\mathbf{E}_p\left[\sum_{y\in\Lambda_n} \mathbf{1}(y \text{ is pivotal and closed for } u_n)\middle| U=V\right] \tag{1.1}$$

$$= \sum_{0\notin V} \mathbf{P}_p[U=V]\varphi_p(S(V)) \tag{1.2}$$

$$\geq \alpha \sum_{0\notin V} \mathbf{P}_p[U=V] \tag{1.3}$$

$$= \alpha \mathbf{P}_p[0 \not\longleftrightarrow \partial\Lambda_n] \tag{1.4}$$

Where (1.1) is a simple conditioning, step (1.2) is the saucy one: given that $U = V$, then y being pivotal for the event that $\{0 \longleftrightarrow \partial\Lambda_n\}$ is equivalent to saying that there is an open path connecting zero to a neighbour of y in $S(V)$ (see diagram), which gives the equality with $\varphi_p(S(V))$ (note that this also takes care of the "pivotal AND closed" part, because the immediate outside of $S(V)$ contains all closed vertices). From then (1.3) comes from the discussion above that used **(OBS 1)**. Step (1.4) comes from the fact that if you sum over all the events that the components in Λ_n that reach the boundary equals a set that does not contain the origin, this equals the probability that zero does not reach the boundary. Therefore we have that for all $p \in [p_0, p_1]$, where $p_1 \in (p_0, 1)$, we have that

$$\frac{\mathrm{d}}{\mathrm{d}p}\mathbf{P}_p[0 \longleftrightarrow \partial\Lambda_n] \geq \frac{\alpha}{1-p}\mathbf{P}_p[0 \not\longleftrightarrow \partial\Lambda_n] \geq \frac{\alpha}{1-p}(1-p_1)$$

where the last inequality comes from the fact that if the origin is closed, which occurs with probability $1-p \geq 1-p_1$, then the origin is not connected to $\partial\Lambda_n$. We can of course also get one

Figure 1.8: The diagram that says it all: second part of exponential decay

more inequality where we forget about that denominator. Therefore:

$$\mathbf{P}_{p_1}[0 \longleftrightarrow \infty] = \lim_{n\to\infty} \mathbf{P}_{p_1}[0 \longleftrightarrow \partial\Lambda_n] \tag{1.1}$$

$$= \lim_{n\to\infty} \mathbf{P}_{p_0}[0 \longleftrightarrow \partial\Lambda_n] + \int_{p_0}^{p_1} \frac{\mathrm{d}}{\mathrm{d}p}\mathbf{P}_p[0 \longleftrightarrow \partial\Lambda_n]\mathrm{d}p \tag{1.2}$$

$$\geq \lim_{n\to\infty} \int_{p_0}^{p_1} \alpha(1-p_1)\mathrm{d}p \tag{1.3}$$

$$\geq \alpha(1-p_1)(p_1-p_0) > 0 \tag{1.4}$$

And thus we have that $p_1 \geq p_c$. But since p_1 was chosen to be any $1 > p_1 > p_0$, it follows that $p_0 \geq p_c$. ♡

Summary and study tips: In this chapter we covered:

1. How the existence of a set S whose $\varphi_p(S) := \mathbf{E}[\#\text{outside escape points}]$ is < 1 implies exponential decay: the proof in one line is that for you to reach Λ_n, you must first reach one of these escape points and then travel the remaining distance. Thus, by fitting S inside Λ_{n_0}, and letting n be arbitrary and large, we have that $\mathbf{P}_p[0 \longleftrightarrow \partial\Lambda_n] \leq \mathbf{P}_p[0 \longleftrightarrow \partial\Lambda_{n-n_0}]\varphi_p(S)$. This is the content of diagram 6.1

2. To finish the proof we explored Russo's formula, which says that for an increasing event such as $\{0 \longleftrightarrow \partial\Lambda_n\}$, the derivative $\frac{d}{dp}\mathbf{P}_p[0 \longleftrightarrow \partial\Lambda_n] = \mathbf{E}[\#\text{pivotal sites for the event}]$. The proof of the second part of the Theorem was to show that if there was some $\alpha > 0$ such that for all S, $\varphi_p(S) \geq \alpha$, meaning that there is a large enough number of possible escape routes, then there will be a lot of pivotal points for the event $0 \longleftrightarrow \partial\Lambda_n$ (for this see diagram 1.8), and as such $\mathbf{P}_p[0 \longleftrightarrow \partial\Lambda_n]$ will be bounded from below by a constant independent of n, and so there will be an infinite cluster with non-zero probability.

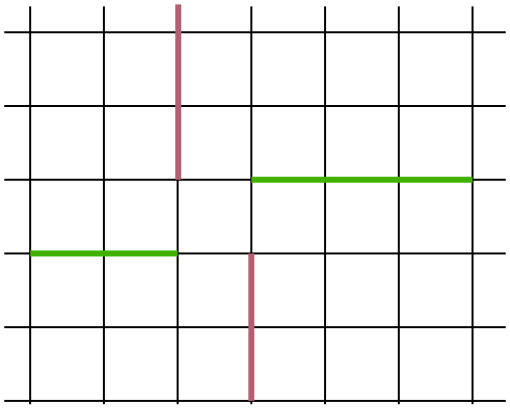


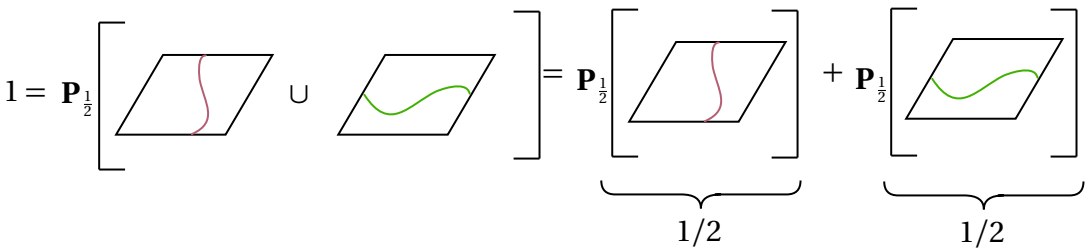
Figure 1.9: Why the triangular lattice: in $\mathbf{Z}^2$ it is not necessary that there is either a top-to-bottom or left-to-right connection

1.4 The value of p_c on the triangular lattice

We are now going to focus on the value of the critical percolation parameter on the triangular lattice $\mathcal{T}$. The main goal of this section is to prove the following Theorem:

Theorem 1.26 (p_c on triangular lattice) For site percolation on $\mathcal{T}$, we have that $p_c = p_c(\mathcal{T}) = 1/2$. Moreover, $\mathbf{P}_{1/2}[\text{there is an infinite cluster}] = 0$.

There is a key observation, unique to $\mathcal{T}$ as opposed to $\mathbf{Z}^2$, which will come in handy later. If we consider Λ_n to be a rhombus of side-length $2n+1$, the probability of an open horizontal crossing is identical to the probability of a top-to-bottom closed crossing. Indeed: consider the sites that are connected to the top boundary, if this set stretches out to the bottom boundary, then there is a top-to-bottom open connection, and therefore a left-to-right closed connection cannot exist, and if there is not a top-to-bottom open connection, it must be the case that there is a left-to-right closed connection blocking the open path. This is a property intrinsic to the triangular lattice, since it is allowed to use the diagonals. In $\mathbf{Z}^2$ for example it could be the case that neither there is a left-to-right closed or top-to-bottom open connections (see diagram). We have established that one of the two events must always hold. Of course, they cannot both happen at the same time, and since we are working at $p = 1/2$, both events have the same probability by symmetry. Therefore we have the following:



Now we can easily prove half of our goal:

Lemma 1.27 On percolation in $\mathscr{T}$, one has that $p_c \leq 1/2$.

Proof. Assume that $p_c > 1/2$. We will reach a contradiction using our observation as well as exponential decay (note that the exponential decay Theorem can also be shown to work in $\mathscr{T}$ with essentially the same proof). By our observation, we have that

$$\begin{aligned}
\frac{1}{2} &= \mathbf{P}_{\frac{1}{2}}[\text{there is a left-to-right open crossing in } \Lambda_n] \\
&\leq \mathbf{P}_{\frac{1}{2}}[\text{there is a point on the left boundary of } \Lambda_n \text{ in a cluster of diameter at least } 2n+1] \\
&\leq \sum_{x \in \text{ left boundary of } \Lambda_n} \mathbf{P}_{\frac{1}{2}}[\operatorname{diam}(C_x) \geq 2n+1] \\
&\leq (2n+1)\exp(-\psi(p)n) \to 0
\end{aligned}$$

Where in the last step we used exponential decay. Thus we reach a contradiction and $p_c \leq 1/2$. ♡

In order to continue our quest for p_c, we need to make a small detour and develop a tool that will prove crucial later on:

FKG Inequality

In this subsection we will prove an inequality referred to as the FKG or Harris' inequality. Recall that an event A is said to be increasing if $v \in A$ and $v \leq v'$ implies that $v' \in A$. Then Harris' inequality states that

Proposition 1.28 (Harris' Inequality) Let A and B be two increasing events that depend on the outcome of finitely many sites. Then $\mathbf{P}[A \cap B] \geq \mathbf{P}[A]\mathbf{P}[B]$.

One could prove this formula via combinatorial arguments, performing induction on the number of sites for which A and B depend on, but the proof that we present here, quite beautiful indeed, uses a coupling of suitably chosen Markov chains, whose invariant distributions are the measures of interest to us.

Proof. We start by writing S for a large enough finite set for which the outcomes of A and B depend on percolation only inside S. That is to say, $A \in \mathscr{F}_S$ and $B \in \mathscr{F}_S$. We are going to run two Markov Chains, X_n and Y_n on $\{0,1\}^S$. Call $\mathbf{Q}$ the probability measure of the probability space on which these two Markov chains live. Let us first define the chain $(X_t)_{t\geq 0}$ as follows:

- X_0 is the configuration where every site is open, i.e: the ones vector.
- At time n, choose one of the $\#S$ sites uniformly at random (here is why we need in this proof that A and B depend on a finite amount of vertices, otherwise we would not be able to pick one at random). And resample the site, opening it with probability p, and closing it with probability $1-p$.

We have the following preliminary observations about $X = (X_n)_{n\geq 0}$:

- X is an irreducible Markov Chain. Indeed, it is easy to see that one can go from any point in $\{0,1\}^S$ to any other point with positive probability by waiting enough time.
- X is aperiodic. In fact it is lazy, which as we know implies aperiodicity.
- X is reversible with respect to the percolation measure $\mathbf{P}_p$, and hence $\mathbf{P}_p$ is the invariant measure, not very surprising! To see reversibility one can manually check the detailed balance equations, which shouldn't be too terrible (I've done it I swear).

It follows from these observations, by the Fundamental Theorem of Markov Chains, that $\mathscr{L}(X_n)$ converges to $\mathbf{P}_p$ in Total Variation. This Markov chain is the one that will give us part of the proof. Now we need to couple this Markov Chain to another. Define $(Y_n)_{n\geq 0}$ as follows. Similarly, set Y_0 to be the ones vector. By the fact that B is an increasing an non-empty set, it is clear that $Y_0 \in B$. Now we sample Y_n just as X_n (i.e: we use the outcomes of the sampling of X_n to determine what Y_n will be, thus using the same source of randomness). However, if at time say $n+1$, one of these updates makes Y_{n+1} leave the set B, we cancel this update and set $Y_{n+1} = Y_n$. Thus by our construction, we see that $Y_n \in B$ for all n. Now we have the following observations about the chain $Y = (Y_n)_{n\geq 0}$:

- Y is an irreducible and aperiodic chain **in** B.
- Y is reversible with respect to the conditioned measure $\mathbf{P}_p[\cdot \mid B]$, and hence this is its invariant measure. Indeed:

$$\frac{\mathbf{Q}[Y_n = v^x, Y_{n+1} = v_x]}{\mathbf{Q}[Y_n = v_x, Y_{n+1} = v^x]} = \frac{1-p}{p} = \frac{\mathbf{P}_p[v_x]}{\mathbf{P}_p[v^x]} = \frac{\mathbf{P}_p[v_x \mid B]}{\mathbf{P}_p[v^x \mid B]}$$

  Where the only mysterious inequality could be the last one, and it simply follows because both v^x and v_x must be in B, and so $v^x \cap B = v^x$ and same for v_x.

- The last observation, which is quite trivial, is that by construction $X_n \leq Y_n$, and so if A is an increasing event, and $X_n \in A$, then $Y_n \in A$.

Putting this all together, we get

$$\begin{aligned}\mathbf{P}_p[A \mid B] &= \lim_{n\to\infty} \mathbf{Q}[Y_n \in A] \\ &\geq \lim_{n\to\infty} \mathbf{Q}[X_n \in A] \\ &= \mathbf{P}_p[A].\end{aligned}$$

And so by expressing $\mathbf{P}_p[A \mid B] - \mathbf{P}_p[A \cap B]/\mathbf{P}_p[B]$ we are done. Of course in this proof we have implicitly assumed that the probability of B is non-zero, but if it were, the claim would have followed trivially. ♡

Remark 1.29 We can now extend this to some other cases. Obviously by induction we have that if $A_1, \cdots, A_n$ are all increasing, then $\mathbf{P}_p\left[\bigcap_i A_i\right] \geq \prod_i \mathbf{P}_p[A_i]$. The second observation, is that since if A is increasing, then A^c is decreasing, (I had proof for this but now I forgot, it went by contradiction I believe) and so if A is increasing and B is decreasing, we have that $\mathbf{P}_p[A \cap B] \leq \mathbf{P}_p[A]\mathbf{P}_p[B]$. Similarly, if a collection $A_1, \cdots, A_n$ are all decreasing, we have that $\mathbf{P}_p\left[\bigcap_i A_i\right] \geq \prod_i \mathbf{P}_p[A_i]$.

The value of p_c on the triangular lattice

We now return to the main quest of this section. Armed with Lemma 1.27 and Harris' Inequality, we are ready to finish off the proof of Theorem 1.26.

Proof of Theorem 1.26. It is left to show that there is almost surely no infinite cluster at $p = 1/2$. If we show this, it will follow that $p_c \geq 1/2$, and by Lemma 1.27, the Theorem will follow. Recall that we define Λ_n as the rhombus of side-length $2n+1$. For each of its sides, $L_1, \cdots, L_4$, we define $E_i(n)$ to be the event that there is an open path from L_i to infinity that stays outside the rhombus. Then it is clear that

$$\bigcup_{i=1}^{4} E_i(n) = \{\Lambda_n \longleftrightarrow \infty\}$$

Note that the events $E_i(n)^c$ are all decreasing, and hence all positively correlated. Suppose now that there is an infinite cluster at $p = 1/2$. Working on this event, we naturally have that $\mathbf{P}_p[\Lambda_n \not\longleftrightarrow \infty] \to 0$ as $n \to \infty$, so putting this all together, we have that

$$\mathbf{P}_p[E_1(n)^c]^4 = \prod_{i=1}^{4} \mathbf{P}_p[E_i(n)^c] \leq \mathbf{P}_p[\Lambda_n \not\longleftrightarrow \infty] \to 0.$$

Where the first equality is due to symmetry of the rhombus, and the inequality is due to Harris' Inequality. From this it follows that for large enough n, $\mathbf{P}_p[E_1(n)^c] < 1/8$. Now note the following, since we are working at $p = 1/2$, the probability that there is an open path from L_i to infinity is the same as the probability that there is a closed path from L_i to infinity. Therefore, since

$$\mathbf{P}_p\left[\bigcap_{i=1}^{4} E_i(n)\right] = 1 - \mathbf{P}_p\left[\bigcup_{i=1}^{4} E_i(n)^c\right] \geq 1 - \sum_{i=1}^{4} \mathbf{P}_p[E_i(n)^c] \geq \frac{1}{2}$$

We have that with probability at least $1/2$, the four following events occur simultaneously: there is an open path from L_1 to infinity, there is a closed path from L_2 to infinity, there is an open path from L_3 to infinity, and there is a closed path from L_4 to infinity. Since the percolation on $\partial\Lambda_n$ and outside is independent of percolation in Λ_{n-1}, we can resample on the inside, and close all sites, so that we isolate the two open infinite clusters emanating from L_1 and L_3, thus creating, with probability strictly greater than zero, two infinite disjoint open clusters. But we know that this occurs with probability zero. Thus reaching a contradiction and we get that there is almost surely no infinite cluster at $p = 1/2$.

♡

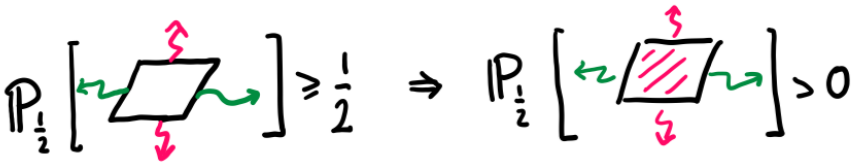


Figure 1.10: The diagram that says it all, proof of Theorem 1.26

Chapter 2

Conformal invariance of critical percolation on $\mathcal{T}$

The main goal of this chapter will be Smirnov's Theorem, sometimes called Smirnov's proof of Cardy's formula, that determines the limit when the mesh-size goes to zero of the probability of crossing quadrilaterals. Whenever we write $\mathbf{P}_p$ we mean $\mathbf{P}_{\frac{1}{2}}$. We will make a slight detour before talking about this, and prove a tool that will be key later: the RSW bounds, which will tell us about the probability of crossing "rectangles of k-aspect ratio", and will also be used to give a lower bound for the probability of closed circuits around a point.

2.1 Russo-Seymour-Welsh bounds

For convenience we look at the triangular lattice rotated by 90 degrees so that an axis of symmetry lies vertically. We define the rectangles of the triangular lattice as follows. For any $a>0$ and any $b>0$, we define $R(a,b)$ the set of sites in the rotated triangular lattice with $(x,y)\in[0,a]\times[0,b]$. We denote by $H(a,b)$ the event that there exists a horizontal open crossing of $R(a,b)$, i.e: a nearest-neighbour path of open sites in the rectangle that joins a point with x-coordinate zero to a point with x-coordinate a. Therefore its easy to see that the map $a\mapsto\mathbf{P}_p[H(a,b)]$ is non-increasing in a. We want a way of comparing the left-to-right crossing probabilities of glued domains.

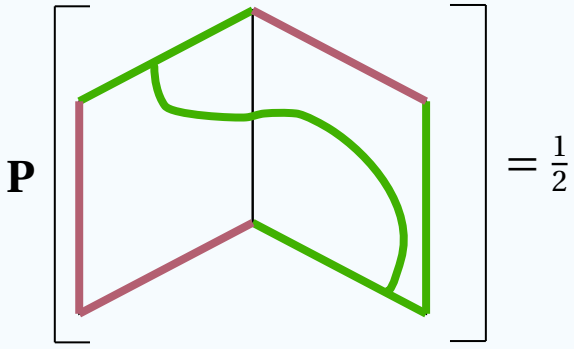
Lemma 2.1 (Russo-Seymour-Welsh) One has that

$$\mathbf{P}_p[H(2a,b)]\geq\frac{\mathbf{P}_p[H(a,b)]^2}{4}$$

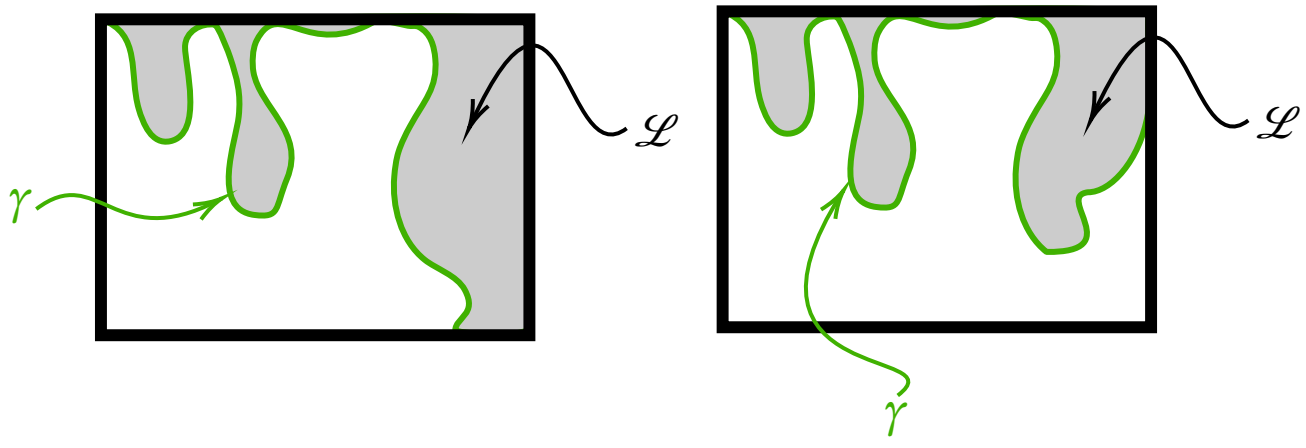
In general, the proof will work for any symmetric domain (we will work with domains symmetric with respect to a vertical axis since we are working with the rotated lattice). First we have the following observation, if we have a symmetric domain D, we can divide the boundary into four pieces, and we

ask the question, what is the probability that there is an open crossing from the top left boundary to the middle right or bottom right components? Well by symmetry, and using arguments similar to what we did before, namely using the fact that we are on a triangular lattice, it must be that there is exactly either a top-left-bottom right open crossing, or a bottom-left-top-right closed crossing, by the symmetry of the domain both of these things have the same probability so we have that

Lemma 2.2 Consider a symmetric domain (not necessarily two rhombus as shown in the picture), then



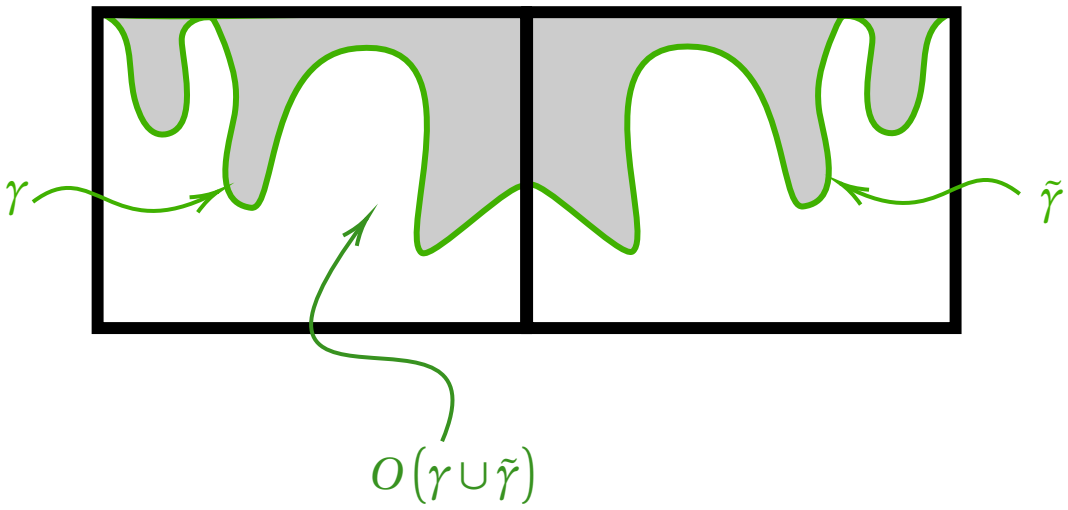
Proof of Russo-Seymour-Welsh bound. The key to this proof is to explore the state of the different sites in a suitable way. Clearly, in order to cross $R(2a, b)$, one has to first cross the left-half rectangle $R(a, b)$. One way to discover if there is a left-right open crossing of $R(a, b)$ is from top-to-bottom, by looking at the interface between open and closed sites. More precisely, let $\mathscr{L}$ be the (random) cluster of closed sites that are connected to the top boundary. For each possible fixed set L, the event $\mathscr{L} = L$ means that the sites of L are all closed, whereas all sites that are at a distance 1 from L are open. (Note from this that the event $\mathscr{L} = L$ is measurable with respect to the state of the sites in the union of L and its neighboring sites. The "lower" boundary of $\mathscr{L}$ contains a random curve γ consisting of open cells only This could look something like this:



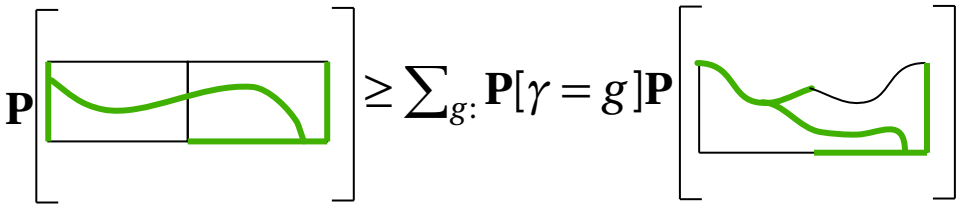
On the left we see a case where there is no left-to-right open crossing, meaning that γ does touches the bottom side, and on the right, we have a case where there is a left-to-right open crossing, as shown by γ touching the right-hand side. Thus we have that

$$\{\text{there is a left-to-right open crossing}\} = \{\gamma \text{ hits the right hand side}\}.$$

Let's now return to our case of the two glued rectangles. Recall that for $H(2a,b)$ to hold, it is necessary that $H(a,b)$ holds. So suppose we have explored our left rectangle and found that the random curve γ described above is equal to some curve g that touches the right hand-side. Then we can artificially draw the symmetric image of g on the second rectangle, say $\tilde{g}$, and now we look at the connected component of $R(2a,b)$ that contains the new top boundary consisting of $g \cup \tilde{g}$, call this domain $O(g \cup \tilde{g})$:

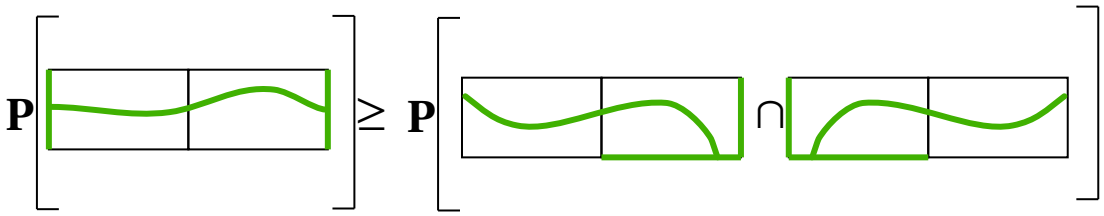


Now notice that $O(g)$ looks a lot like the preliminary warmup observation we did in Lemma 2.2, and given that the law of the sites underneath γ is that of percolation, since we haven't revealed anything, we can ask, conditional on $\gamma = g$, what is the probability that there is an open path from the top-left boundary of $O(g)$ to the bottom-right or right-most boundary of $O(g)$. By the Lemma we know this is $1/2$. Therefore, putting this all together we conclude that the probability of having a crossing in $R(2a,b)$ that goes from the left-most side to the bottom-right or right-most side can be bounded below:



Where the sum over g is taken over paths that have a right-crossing. However, by Lemma 2.2, we know that the right-most term in this sum is $1/2$, and so we have that our original probability is greater than or equal to $\frac{1}{2}\mathbf{P}_p[H(a,b)]$. This is close to our goal but not yet quite, because

we want a crossing from our left-most side to our right-most side, we don't want to include the bottom-right side, this however, can be easily fixed by using the FKG Inequality, indeed: a possible way to realise a left-to-right crossing is as follows:



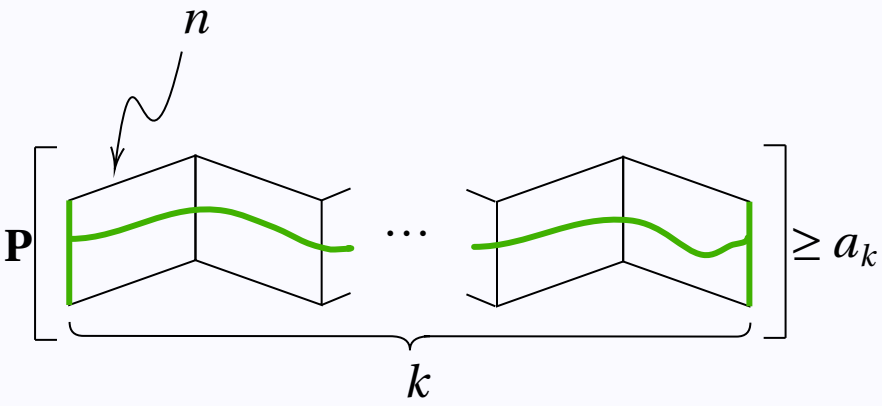
And now we can use the FKG inequality using the fact that both of the events that are being intersected are increasing, and we get the final result:

$$\mathbf{P}_p[H(2a,b)] \geq \frac{\mathbf{P}_p[H(a,b)^2]}{4}$$

♡

From this we have an immediate corollary. Recall that in the case of the Rhombus Λ_n, we had seen that the probability of a left-to-right open crossing was exactly equal to $1/2$. Moreover, in this proof we did not use any specific facts about the rectangle. The only thing we did use is that it is the union of a set, namely $R(a,b)$ and its symmetric image. Therefore, we have the following

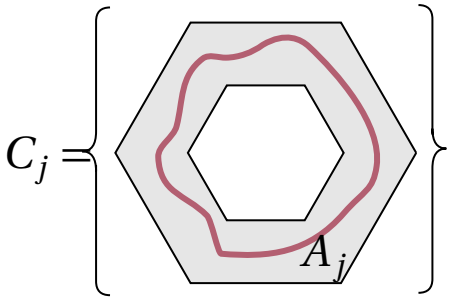
Corollary 2.3 Let $k, n \in \mathbf{N}$ be given where k is a power of two and n is odd. Then there exists a constant a_k independent of n for which the probability of having a left-to-right crossing on k glued rhombus of side-length n is bounded below by a_k



Proof. We have that the probability of a left-to-right crossing on two glued rhombus is at least $\left(\frac{1/2}{2}\right)^2 = \frac{1}{16}$ now keep iterating the RSW bound. ♡

We already saw in Chapter 1 how we could give a proof by using the FKG inequality there was almost surely no infinite cluster at criticality in the triangular lattice, using the RSW bound, we can give an alternative proof. We will present the key result as a result of its own as it is of independent

interest. Define now by Λ_n the hexagon centered at the origin with graph radius n (this is the natural analog of the ball of radius n in the triangular lattice). We then define the concentric disjoint annuli $A_j = \Lambda_{2j+1} \setminus \Lambda_{2j}$.

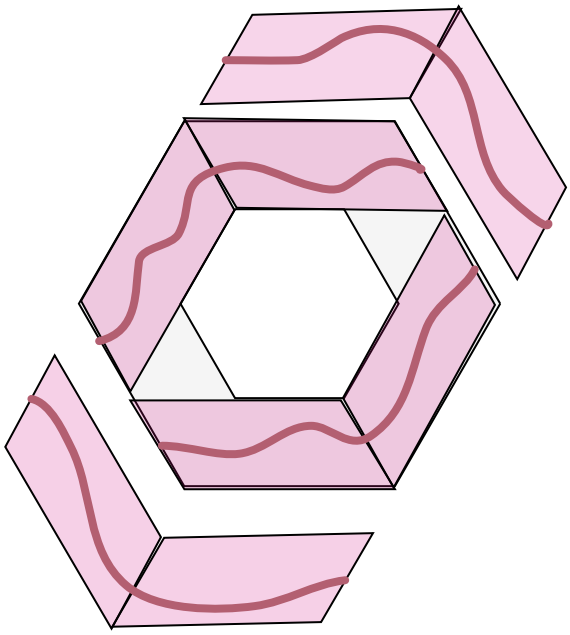


And then we define C_j to be the event that there is a loop in A_j formed of closed sites, thus disconnecting the origin from infinity. Then we have the following:

Corollary 2.4 (Loops in the hexagon) Consider the events C_j as defined above. Then there exists some $\alpha > 0$ such that for all $j \in \mathbf{N}$,

$$\mathbf{P}_p[C_j] \geq \alpha.$$

Proof. The idea is to "cover" the annulus A_j by four large enough glued rhombuses. Here's a messy picture which hopefully doesn't confuse the reader more than need be:



As attempted to depict in this picture, one can cover the hexagon with 4 "double rhombuses" which will overlap (I have drawn two of the double rhombuses outside the hexagon, but they should be in the annulus, but otherwise the overlap would not allow the reader to see all the shapes). We know regardless of the size of the rhombuses, the probability of a left-right-crossing of closed sites on each of thse double rhombuses is at least of $1/16$. Label R_i the event that there is a left-right

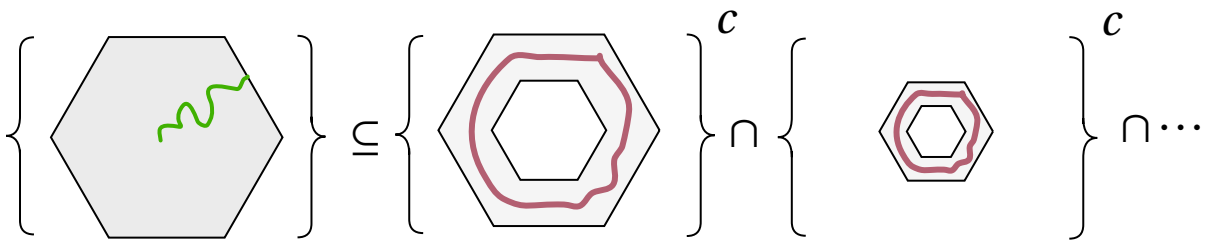


Figure 2.1: Alternative Proof to Theorem 1.26, the picture that says it all

crossing made of closed sites on the i^{th} double-rhombus. Then it is clear that

$$\mathbf{P}_p[C_j] \geq \mathbf{P}_p[R_1 \cap \cdots \cap R_4]$$

Indeed, if we have a closed path on each of the double-rhombuses, we can just jump from path to path at the intersection points of the paths and create the loop. Now the point is that each R_i is a decreasing event, because if you turn off more sites, it is more likely that you have a closed crossing. Therefore:

$$\mathbf{P}_p[C_j] \geq \mathbf{P}_p[R_1]^4 \geq \left(\frac{1}{16}\right)^4$$

Once again, we emphasize that there is no dependence on j here because we can make the rhombuses as big as we want. ♡

Now we can present an alternative proof for the non-existence of infinite cluster at criticality.

Alternative Proof to Theorem 1.26. Let us study the event that zero is connected to the boundary of the hexagon of side-length 2^{j+1}. If this event holds, it must mean that it is not the case that any of the j annuli has a closed loop around it. It is also clear however, that percolation in these concentric annular regions is independent from each other, and so we deduce that, using the α of Corollary 2.4, we have that

$$\mathbf{P}_p[0 \longleftrightarrow \partial\Lambda_n] \leq (1-\alpha)^j$$

Therefore as we take $n \to \infty$ the right hand side rapidly decays to zero. ♡

2.2 Smirnov's proof of Cardy's formula

Before stating the Theorem we will first provide some motivation and later review a well-known result in Complex Analysis.

Remark 2.5 (Consequence of RSW bounds) Recall that if you take a rectangle $R(kN,N)$ of aspect ratio $k \in \mathbf{N}$, then the probability that there is a left-to-right crossing, $\mathbf{P}_p[H(kN,N)]$ is bounded below by a constant $c_k > 0$ independent of N. Similarly, by now considering the probability of a closed top-to-bottom crossing, we also get an upper bound. That is to say, there is some $\epsilon > 0$ for which

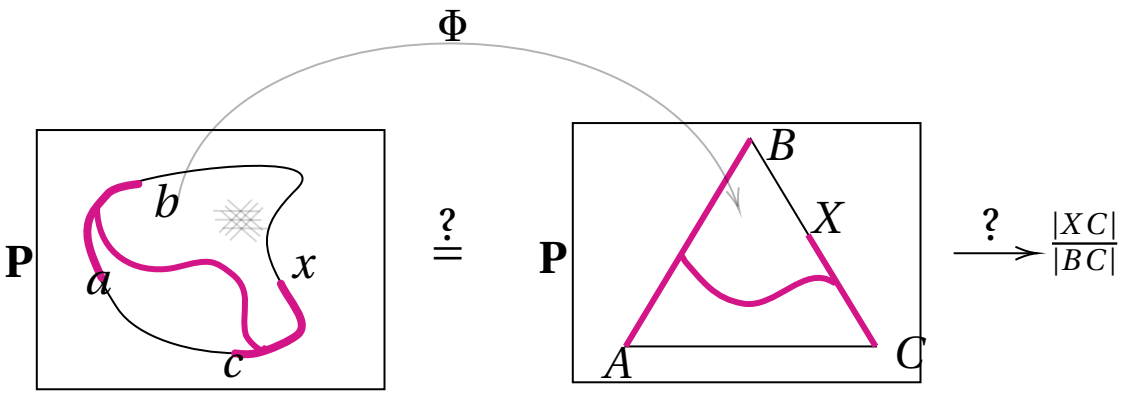
$$\epsilon < \mathbf{P}_p[H(kN,N)] < 1-\epsilon.$$

Therefore, we can find a subsequence (N_t) along which these probabilities converge somewhere in $(0,1)$. A way to interpret this statement of $N \to \infty$ is either that the rectangle gets very large, or if on the other hand, "we move away from the plane" with the rectangle, we can interpret this as the mesh size getting very small. In this section we will prove a formula that helps us understand these kind of crossing probabilities as mesh sizes tend to zero.

Now let us state the following:

Theorem 2.6 (Riemann's Mapping Theorem) Let D be a simply connected open domain in the complex plane with $D \neq \mathbf{C}$. Let T be the unit length equilateral triangle, i.e: the equilateral triangle with vertices $A = 0, B = 1$, and $C = \exp(i\pi/3)$. Fix three points $a, b, c \in \partial D$. Then there exists a unique conformal (i.e: angle preserving) map $\Phi : \bar{D} \to T$ that is a bijection and maps the points a, b, c to the points A, B, C respectively.

Therefore, if we pick a fourth point $x \in \partial D$, say on the arc between a and c, then the image $\Phi(x) = X$ is prescribe and belongs to the side AC. We call (D, a, b, c, x) a **Conformal rectangle** . We can ask then the following question: in the scaling limit where the mesh size of the triangular percolation goes to zero, what is the crossing probability from the arc ab to the arc xc? It turns out that there is a very elegant answer: very simply, $|XC|/|BC|$ where these are the conformal images onto the triangle.



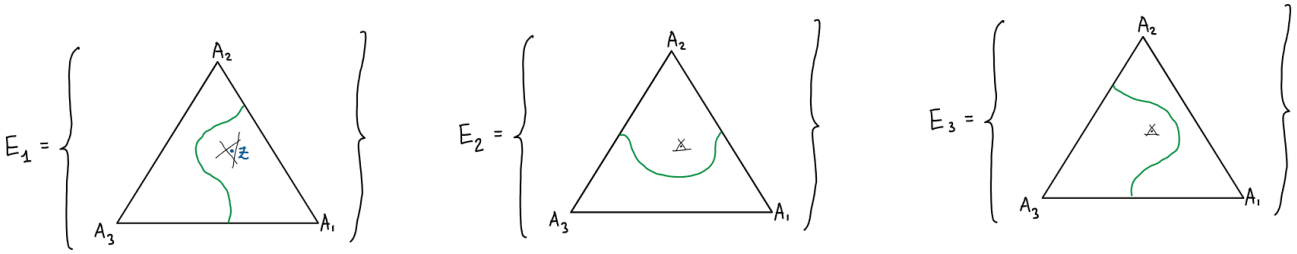
Theorem 2.7 (Smirnov-Cardy's Formula) If D is conformally equivalent to the equilateral triangle ABC, and if the four boundary points a, b, c, x are respectively mapped to A, B, C, X and $X \in [CB]$, then if one takes a lattice approximation of (D, a, b, c, x) and performs critical percolation on it, as the mesh size goes to zero the probability that there exists a crossing in D from ab to cx converges to $|XC|/|BC|$.

Let us give an intuitive overview of the proof. We will prove the special case of the statement for the triangle T. We will actually consider some sort of generalisation of crossing probabilities, by moving the "target point" to the inside of the triangle. This will yield three events, the events $E_1^\delta(z), E_2^\delta(z), E_3^\delta(z)$, with $E_i^\delta(z)$ corresponding to the event that there is an open crossing isolating z in the region that contains the i^{th} vertex from the rest of the triangle. Then the probabilities of these events will be considered. The idea is that in the case of E_2 below, if we move the point z towards the right-hand-side of the boundary, we will actually get the probability we were interested in from the start. The outline of the proof will be

1. Prove that the functions $H_i^\delta(z)$ have subsequential limits as $\delta \to 0$.
2. Prove a combinatorial identity called the "colour-switching Lemma".
3. With this Lemma, we will show that if the limit $\delta \to 0$ exists, the contour integrals of triangles of the functions $H_1 + H_2 + H_3$ and $H_1 + \tau H_2 + \tau^2 H_3$ vanish, thus showing they are analytic.
4. Using analyticity of these functions, we will derive some properties about the H_i's, namely that they are Harmonic and some boundary conditions that will uniquely determine their structure. In particular, we will show that
$$H_2(z) = \text{Height}(z)$$
5. In summary, we will have shown that there is only one possible accumulation point, and since the functions are on a compact set, they do converge to the specified limit.

Tightness via RSW

Let us formalise some of what we have said above. For a face z on the triangular lattice, consider the events $E_i^\delta(z)$ for $i = 1, 2, 3$, where δ is the size of the mesh, to be defined as follows:



$E_1^\delta(z)$ is the event that there is a line from A_1A_2 to A_1A_3 that separates z from A_2A_3, and the other events are defined similarly by "rotating" the starting and ending points as well as the "blocking goal" by $2\pi/3$. We now define three functions, $H_i^\delta(z)$ given by $H_i^\delta(z) = \mathbf{P}_p[E_i^\delta(z)]$. These functions can be extended to be continuous by interpolation. One of the main steps of this proof will be to show that these functions have a limit as $\delta \to 0$. We will now see how RSW helps us show that there will in fact be sub-sequential limits, the final idea will be to "push out" z towards the boundary. Let us recall a Theorem from Analysis:

Theorem 2.8 (Arzela-Ascoli (in $\mathbf{R}$)) A set $\mathbf{F}$ of functions in $C([a,b])$ that is uniformly bounded and satisfies a Holder condition of order β:

$$|f(x) - f(y)| \leq M|x-y|^\beta$$

for some uniform M, is relatively compact in $C([a,b])$. In particular, every sequence $\{f_n\} \subseteq \mathbf{F}$ has a converging subsequence.

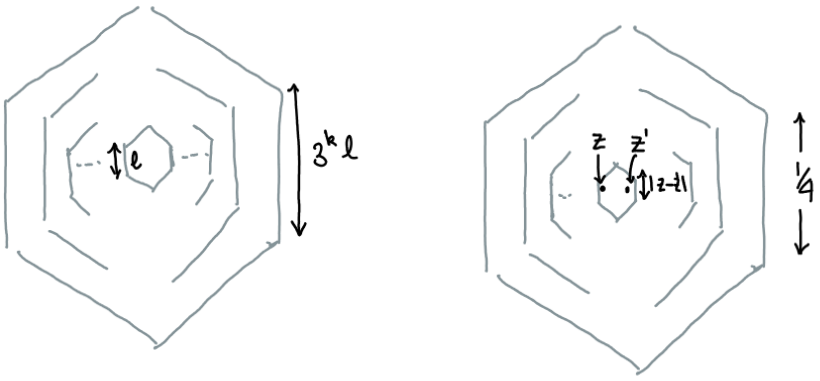
This Theorem is easily generalised to Euclidean domains such as the one we are working in. We now have the following.

Lemma 2.9 The collection of functions $\{H_i^\delta\}$ as defined above satisfies a Holder condition.

Proof. Let z and z' be two faces inside the triangle that are close (say $|z - z'| < 1/100$, this is only so that we can fit some hexagons as we will see now). Notice that if for percolation on the triangle with mesh-size δ we have that if there exists an open (or closed) circuit inside the triangle that surround both z and z', then $E_i^\delta(z)$ and $E_i^\delta(z')$ either both hold or they both don't hold. Then, we have that

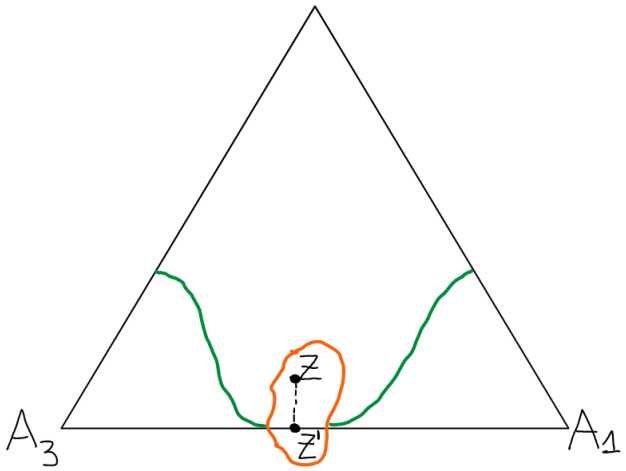
$$\begin{aligned}
|H_i^\delta(z) - H_i^\delta(z')| &= |\mathbf{P}_p[E_i^\delta(z)] - \mathbf{P}_p[E_i^\delta(z')]| && (2.1)\\
&\leq \mathbf{P}_p[E_i^\delta(z) \setminus E_i^\delta(z') \cup E_i^\delta(z') \setminus E_i^\delta(z)] && (2.2)\\
&\leq \mathbf{P}_p[\text{no open or closed circuit around } z, z'] && (2.3)\\
&\leq 2\mathbf{P}_p[\text{no open circuit in any annuli of hexagons around } z, z'] && (2.4)
\end{aligned}$$

Recall from the RSW bound, that if you have series of hexagonal annuli of this form:



Then there are k concentric annuli of the form discussed in Corollary 2.4, and so the probability of not having an open loop in any of the annuli is precisely of $(1-\alpha)^k$. Now we think of fitting one large hexagon of a fixed side-length, and creating concentric annuli until we circle z and z': Naturally, the close z and z' are, the more hexagons we'll be able to fit and so the smaller the probability will be. Precisely, since we'll have $-\frac{\log(|z-z'|)}{\log(N)}$ such annuli and so after rearranging, we'll have that step (2.4) can be bounded above by $M|z-z'|^\beta$ for some M and β. Thus showing the Holder condition. ♡

Now we have shown that if δ_n is a sequence tending to zero, we have that the sequence of functions $H_i^{\delta_n}$ converges uniformly to some sub-sequential limit. The goal moving on will be to find the only possible limit. Before doing that, note that we can obtain one more piece of information about this limit. Consider for an illustrative case that we are talking about the event E_2. The claim is that as z approaches A_3A_1, $H_2^\delta(z)$ goes to zero uniformly in δ. Indeed: we can consider z' to be the "projection" of z down to A_3A_1. If there is a closed loop that contains both z and z', then it is not possible for the open path to go below z and connect A_2A_3 with A_2A_1 by going below z:



However, since $H_2^\delta(z')=0$ (because z' is on A_3A_1 so you cannot go below z'), it follows that $H_2^\delta(z)\leq$

$M|z-A_3A_1|^\beta \to 0$. Therefore the sub-sequential limits H_i will also satisfy this property when z approaches the corresponding sides.

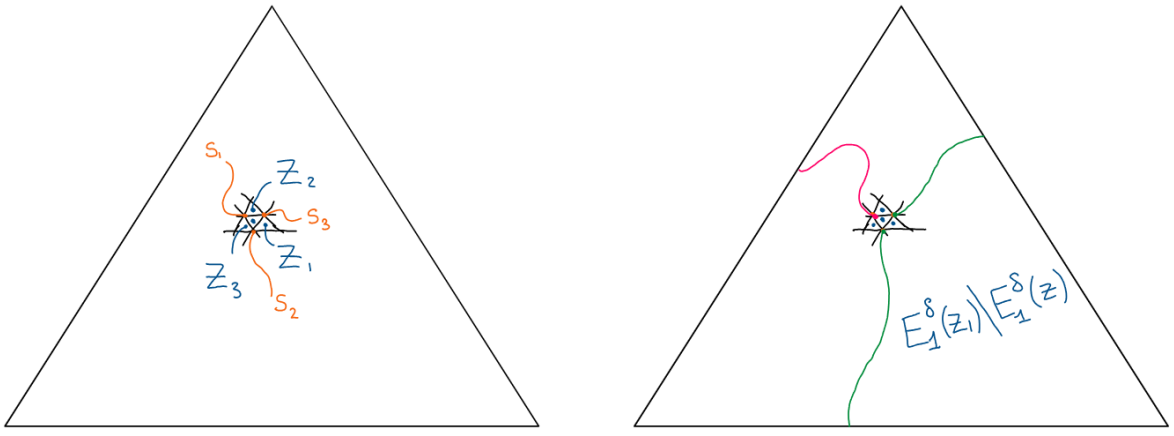
Discrete exploration process and the colour-switching Lemma

Spoiler, the main technique that will allow us to understand these limits of the H_i^δ's is that a certain combination of them will be analytic. Let us recall the following fact about complex analytic functions:

Lemma 2.10 (Characterisation of analytic functions) Let $F:\mathbf{C}\to\mathbf{C}$ be a C^1 function. Then the following are equivalent:

1. F is analytic in an open domain D.
2. The following derivatives coincide on D, $\partial(\Re(F))/\partial 1 = \partial(\Re(F/\tau))/\partial\tau = \partial(\Re(F/\tau^2))/\partial\tau^2$ where $(\partial F/\partial\eta)(z)$ denotes the limit
$$\lim_{h\to 0, h\in\mathbf{R}} \frac{F(z+\eta h)-F(z)}{h}$$
3. (Morera's Theorem)For all triangles T in D with one side parallel to the real axis,
$$\oint_T F(z)dz = 0.$$

With this in mind, we aim to explore the "derivative" of H_i^δ in these three directions. For this let us introduce some notation.

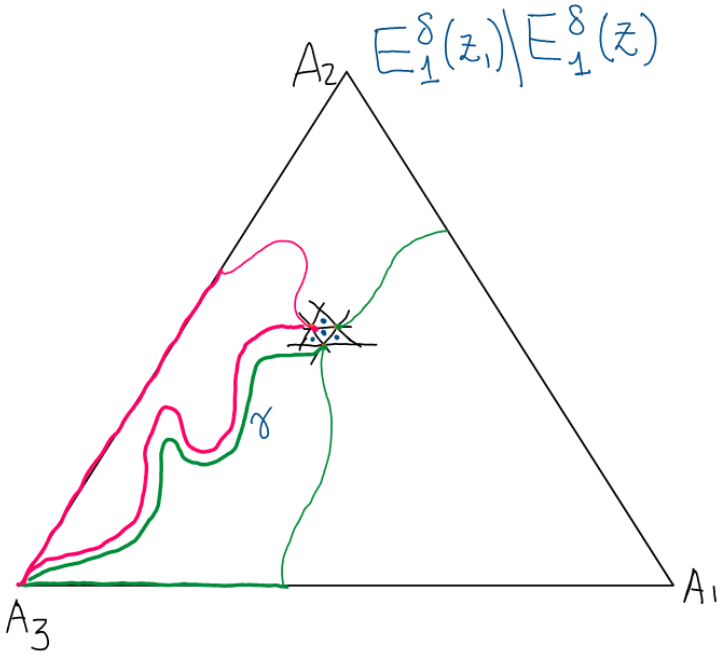


For a face z in the triangular lattice, we call z_1, z_2, z_3 the faces of the lattice that are in directions of the vertices A_1, A_2, A_3 respectively, and we call the sites s_1, s_2, s_3 to be the sites "opposite" to z_1, z_2, z_3 respectively. Exploring these directional derivatives is equivalent to understanding the probability of the event $E_1^\delta(z_1)\setminus E_1^\delta(z)$. It is easy to see that this event is precisely the event that:

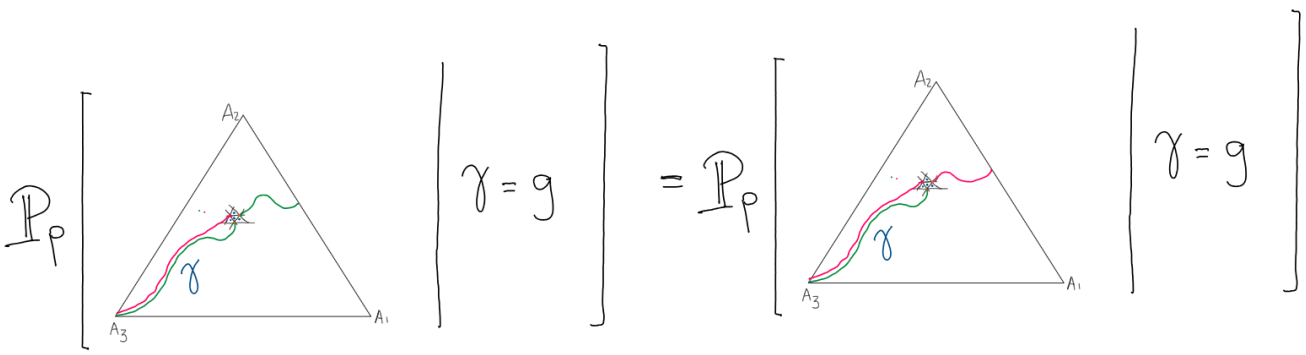
1. There are two open paths, l_2 and l_3 from s_2 and s_3 that connect to A_3A_1 and A_2A_1 respectively.

2. There is a closed path from s_1 to A_2A_3.

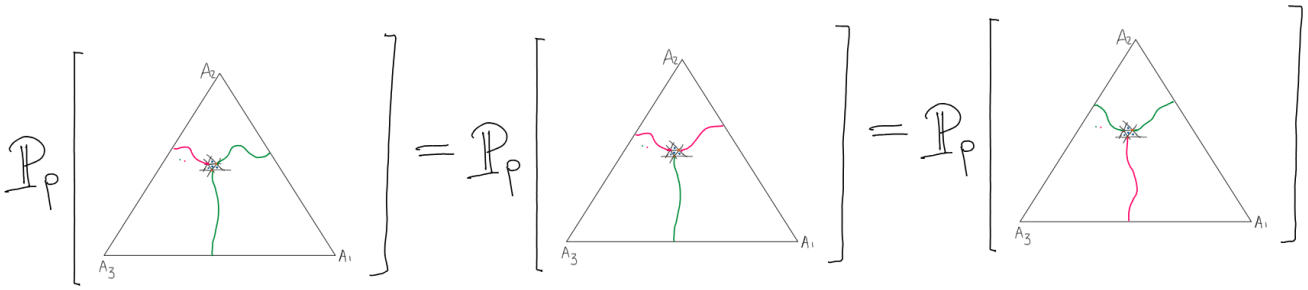
A way to determine if this event has occurred is by starting an exploration process at A_3 by considering the random curve γ that separates the open cluster connected to A_3A_1 and the closed cluster connected to A_2A_3, in other words, you consider γ to be the random curve obtained if one "artificially colours" A_2A_3 closed, A_1A_3 open, and simply follows a path that leaves closed to its left and open to its right. The event described above will hold if and only if this exploratory process γ reaches the face z by connecting the closed path to s_1 and the open path to s_2:



and there is an open path from s_3 to A_2A_1 in the unexplored domain. The key observation is that since the exploratory process has not revealed the state of percolation on the unexplored region, conditional on the event that our exploration process γ yields a given curve g, i.e; conditional on $\{\gamma = g\}$, we have that the probability that there is an open path from s_3 to A_2A_1 is the same probability that there is a closed path from s_3 to A_2A_1. That is to say:



But if we now average over the possible values of γ and then use the fact that $p = 1/2$ so we can flip all colours, we have that:



From this, we have arrived at the following conclusion:

Lemma 2.11 (Colour-Switching Lemma) One has the following two equalities:

$$\mathbf{P}_p[E_1^\delta(z_1) \setminus E_1^\delta(z)] = \mathbf{P}_p[E_2^\delta(z_2) \setminus E_2^\delta(z)] = \mathbf{P}_p[E_3^\delta(z_3) \setminus E_3^\delta(z)]$$

Discrete Contour Integrals

Now we will start working towards determining information about the limit $H_i(z)$. So assume that the $H_i^\delta(z)$ converge to some $H_i(z)$. The goal will be to show, using Morera's Theorem, that

Lemma 2.12 If the limits H_1, H_2, H_3, exist, then the functions $H_1(z)+H_2(z)+H_3(z)$ and $H_1(z)+\tau H_2(z)+\tau^2 H_3(z)$ are analytic in the triangle T.

Proof. The ultimate goal will be to show that for any equilateral contour Γ inside the original equilateral triangle T, the discrete contour integrals of H_j^δ are very closely related. Define $h_j^\delta(z,\eta)$ to be $\mathbf{P}[E_j^\delta(z+\eta) \setminus E_j^\delta(z)]$ where η is chosen so that $z+\eta$ is one of the three neighbours of z (loosely speaking we will say that $\eta = i, i\tau, i\tau^2$ when z is in a triangle "pointing downwards"), and assume that Γ is "perfectly fitted" to consist of unions of small triangular faces of our lattice with mesh-size δ (One can do away without this assumption and then explain that one can well-approximate Γ by a contour Γ^δ that is actally fitted to be part of the triangles in the lattice, but we don't really care). Suppose moreover that Γ is "looking-upwards". Let $\mathscr{D}$ denote the set of centres of triangular faces which are looking downwards. We have the following remarks:

1. Since $\mathbf{P}[A]-\mathbf{P}[B] = \mathbf{P}[A \setminus B]-\mathbf{P}[B \setminus A]$, we have that

$$\begin{aligned} H_j^\delta(z+\eta)-H_j^\delta(z) &= \mathbf{P}[E_j^\delta(z+\eta)]-\mathbf{P}[E_j^\delta(z)] \\ &= \mathbf{P}[E_j^\delta(z+\eta) \setminus E_j^\delta(z)]-\mathbf{P}[E_j^\delta(z) \setminus E_j^\delta(z+\eta)] \\ &= h_j^\delta(z,\eta)-h_j^\delta(z+\eta,-\eta) \end{aligned}$$

2. In this language, the colour-switching Lemma takes the form:

$$h_1^\delta(z,\eta) = h_2^\delta(z,\eta\tau) = h_3^\delta(z,\eta\tau^2)$$

3. Recall that $h_j^\delta(z,\eta)$ is the probability that $E_j^\delta(z+\eta)$ occurs but not $E_j^\delta(z)$. We saw that if there is an open and a closed circuit surrounding both z and some other point z', then either both $E_j^\delta(z)$ and $E_j^\delta(z')$ occur, or they both don't occur. Therefore $h_j^\delta(z,\eta)$ is at most the probability that there is not an open circuit surrounding $z+\eta$ and z or there is not a closed circuit doing the same thing, and we saw that the probability of this occurring is precisely $M|(z+\eta)-z|^\beta$ for some M,β, and so since η was chosen for $z+\eta$ to be a neighbour of z, we have that $h_j^\delta(z,\eta) \lesssim \delta^\beta$ for some $\beta > 0$.

With these three observations, we note the following:

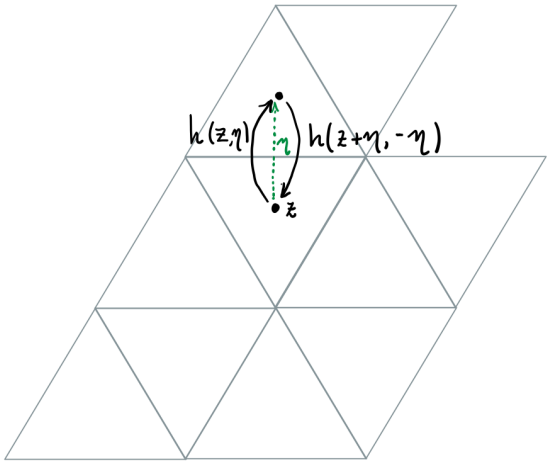
- For $\eta = i, i\tau, i\tau^2$, one has that

$$\sum_{z\in\mathcal{D}} H_1^\delta(z+\eta) - H_1^\delta(z) = \sum_{z\in\mathcal{D}} H_2^\delta(z+\eta\tau) - H_2^\delta(z) + O(\delta^{\beta-1}) \tag{2.5}$$

Indeed:

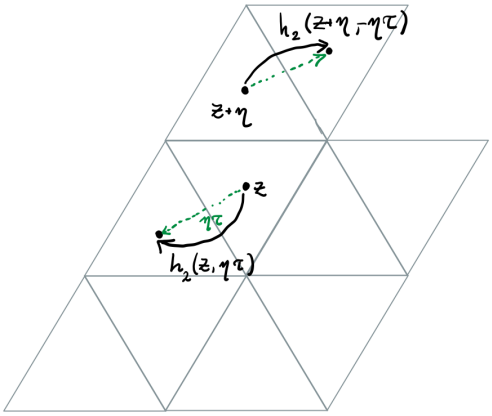
$$\begin{aligned}
\sum_{z\in\mathcal{D}} H_1^\delta(z+\eta) - H_1^\delta(z) &= \sum_{z\in\mathcal{D}} h_1^\delta(z,\eta) - h_1^\delta(z+\eta,-\eta) && (2.1)\\
&= \sum_{z\in\mathcal{D}} h_2^\delta(z,\eta\tau) - h_2^\delta(z+\eta,-\eta\tau) && (2.2)\\
&\overset{!}{=} \sum_{z\in\mathcal{D}} h_2^\delta(z,\eta\tau) - h_2^\delta(z+\eta\tau,-\eta\tau) + O(\delta^\beta\cdot\delta^{-1}) && (2.3)\\
&= \sum_{z\in\mathcal{D}} H_2^\delta(z+\eta\tau) - H_2^\delta(z) + O(\delta^\beta\cdot\delta^{-1}) && (2.4)
\end{aligned}$$

Step (2.1) comes from Remark 1 above, and step (2.2) is the colour-switching lemma as seen in Remark 2 above. Naturally step (2.4) also comes from the same reason as (2.1) Let us explain in detail step (2.3). The summands in step (2.1), before performing the colour-switching, look something like this, say for $\eta = i$:

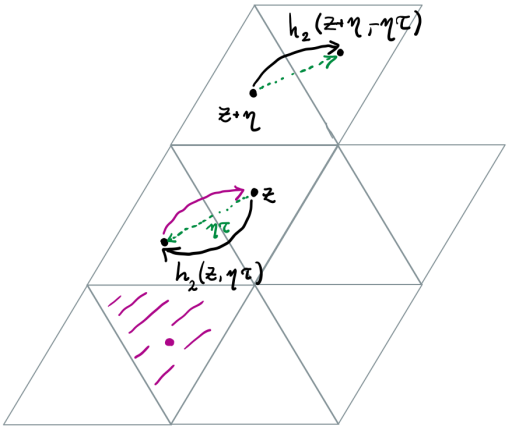


$(\eta = i)$

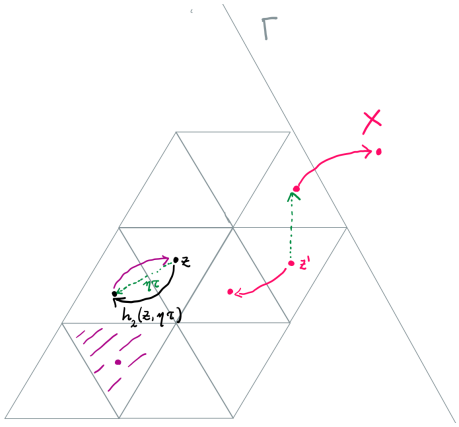
and so after performing the colour switching, in order to end up with $H_2^\delta(z+\eta\tau)-H_2^\delta(z)$, we would like the same picture but with the arrows rotated by τ. However, if we draw on the diagram what we see in step (2.2), we see something like this:



Which is clearly not what we wanted to end up with. But not all is lost, because if you think about it, even though the top arrow is emanating from the wrong starting point, there will be some other face in this diagram, that will "substitute" the wrong arrow:



so actually, in the grand scheme of things (i.e: when all the sum is taken into account), step (2.2) already gives mostly the right thing. The only problem comes from those faces that are adjacent to the edge of the big triangular contour Γ, since they will not have other faces outside Γ to cover up for the missing arrows. This might look something like this:



But that's not a huge problem either, we can simply do a bit of bookkeeping and account for this error. Each error we make costs $O(\delta^\beta)$ to correct, as seen by Remark 3 above, and the errors will come from the sides of the triangle, each having length δ^-1, and so our total error is $O(\delta^{\beta-1})$. We are now ready to prove the main goal. We now consider the following expression:

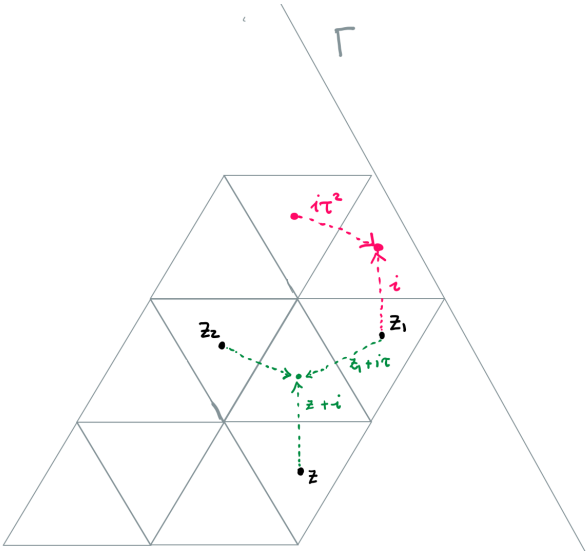
- **EXPRESSION 2**

$$\sum_{z\in\mathcal{D}}\left(H_1^\delta(z+i)-H_1^\delta(z)\right)+\tau\left(H_1^\delta(z+i\tau)-H_1^\delta(z)\right)+\tau^2\left(H_1^\delta(z+i\tau^2)-H_1^\delta(z)\right) \tag{2.1}$$

$$=\sum_{z\in\mathcal{D}}-H_1^\delta(z)\left(1+\tau+\tau^2\right)+H_1^\delta(z+i)+\tau H_1^\delta(z+i\tau)+\tau^2 H_1^\delta(z+i\tau^2) \tag{2.2}$$

$$=\sum_{z\in\mathcal{D}}H_1^\delta(z+i)+\tau H_1^\delta(z+i\tau)+\tau^2 H_1^\delta(z+i\tau^2) \tag{2.3}$$

The main observation to do with this expression is that actually most of this sum is equal to zero. In fact, only the sum over the contour of Γ is left. This is because of the following fact. Pick any face z that is "well inside" the triangle, then $H_1^\delta(z+i)$ will manifest as $H_1^\delta(z_1+i\tau)$ and $H_1^\delta(z_2+i\tau^2)$ for some z_1 and z_2. Thus when the terms get paired up, a factor of $(1+\tau+\tau^2)$ will appear in front. Thus the only surviving terms in this sum will be the sum over the faces in the "boundary of Γ". This is an illustration of this argument



So if we now let $\mathcal{U}_d, \mathcal{U}_l, \mathcal{U}_r$, denote the set of "upwards-pointing triangles" for the down, left, and right sides of Γ respectively, we have that our sum above really becomes

$$\sum_{z\in\mathcal{U}_d}(\tau+\tau^2)H_1^\delta(z)+\sum_{z\in\mathcal{U}_r}(1+\tau^2)H_1^\delta(z)+\sum_{z\in\mathcal{U}_l}(1+\tau)H_1^\delta(z)$$

which of course is nothing but

$$-\left(\sum_{z\in\mathcal{U}_d}H_1^\delta(z)+\sum_{z\in\mathcal{U}_r}\tau H_1^\delta(z)+\sum_{z\in\mathcal{U}_l}\tau^2 H_1^\delta(z)\right)$$

And this can now clearly be seen to be almost a discrete contour integral over Γ. Indeed, it is summing the values of the function H_1^δ around the contour, and multiplying it by the increments of the direction that our path around Γ is making. Therefore, as $\delta\to 0$, assuming these functions converge (uniformly), we have that

$$\delta\sum_{z\in\mathcal{D}}H_1^\delta(z+i)+\tau H_1^\delta(z+i\tau)+\tau^2 H_1^\delta(z+i\tau^2)\to-\oint_\Gamma H_1(z)dz$$

Now you might be wondering what was the point of doing all that **EXPRESSION 1** business, well, if you look back at **EXPRESSION 2**, we can now use 2.5 and obtain that

$$\begin{aligned}\delta\sum_{z\in\mathcal{D}}H_1^\delta(z+i)+\tau H_1^\delta(z+i\tau)+\tau^2 H_1^\delta(z+i\tau^2)&=\delta\sum_{z\in\mathcal{D}}H_2^\delta(z+i\tau)+\tau H_2^\delta(z+i\tau^2)+\tau^2 H_2^\delta(z+i)+O(\delta^\beta)\\&=\delta\frac{1}{\tau}\sum_{z\in\mathcal{D}}H_2^\delta(z+i)+\tau H_2^\delta(z+i\tau)+\tau^2 H_2^\delta(z+i\tau^2)+O(\delta^\beta)\end{aligned}$$

and a similar expression with H_3 can be obtained by using 2.5 once again and having a $1/\tau^2$ instead

of a $1/\tau$. By taking $\delta \to 0$ we have that

$$\oint_\Gamma H_1(z)dz = \frac{1}{\tau}\oint_\Gamma H_2(z)dz = \frac{1}{\tau^2}\oint_\Gamma H_3 dz$$

So from this we have that

$$\oint_\Gamma H_1(z)+H_2(z)+H_3(z)dz = \oint_\Gamma H_1 dz(1+\tau+\tau^2) = 0$$

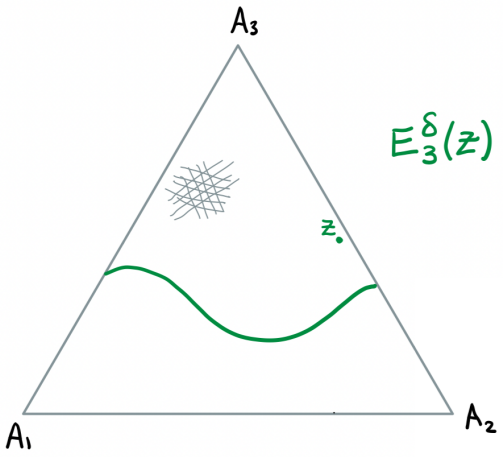
and similarly

$$\oint_\Gamma H_1(z)+\tau H_2(z)+\tau^2 H_3(z)dz = \oint_\Gamma H_1+\tau^2 H_1(z)+\tau^4 H_1(z)dz = 0$$

Thus by Morera's Theorem we are done. ♡

Identification Of the Limit

Let us summarise what we have discovered so far. We had our functions H_i^δ. We saw that they had subsequential limits as $\delta \to 0$, and moreover, we saw that for any subsequential limit (H_1, H_2, H_3), the combinations $H_1+H_2+H_3$ and $H_1+\tau H_2+\tau^2 H_3$ are analytic on our triangle T. We will now determine using this information that there is only one possible limit. Therefore we will have shown that there is only one accumulation point, and since the functions take values on a compact set, the limit will be uniquely determined. Recall that we are in a situation like this:



Then we can now state the final lemma which finishes the proof of Smirnov's Theorem for the triangular domain.

Lemma 2.13 Suppose the limits H_1, H_2, H_3 exist, then

$$H_3(z) = \frac{d(z, A_2A_1)}{d(A_3, A_2A_1)}$$

Proof. As we saw in the previous Lemma, the functions $F = H_1 + H_2 + H_3$ and $G = H_1 + \tau H_2 + \tau^2 H_3$ are analytic. The first one, being real, means that it must be constant. Noticing that due to the RSW estimates which were argued in the Tightness section, we have that $H_1(A_1) = 1$ and $H_2(A_1) = H_3(A_1) = 0$, which means that $H_1 + H_2 + H_3$ is constant and equal to one **(OBS 1)**. On the other hand, note that

$$\begin{aligned}\Re(G) &= H_1 + \Re(\tau H_2) + \Re(\tau^2 H_3) \\ &= H_1 - \frac{1}{2}H_2 - \frac{1}{2}H_3 \\ &= \frac{3}{2}H_1 - \frac{1}{2}\end{aligned}$$

Where in this last equality we used **OBS 1**. From this it can be seen that H_1 is the real part of an analytic function, and from complex analysis we know that this means that H_1 is Harmonic. Similarly we have that

$$\Re(G/\tau) = \frac{3}{2}H_2 - \frac{1}{2} \qquad \Re(G/\tau^2) = \frac{3}{2}H_3 - \frac{1}{2}$$

which by the same reasoning means that H_2 and H_3 are also harmonic. Moreover, we note that G can be extended analytically to a neighbourhood of the segments of the triangle T. This is because on the segment $[A_2A_3]$, H_1 is zero, so $\Re(G)$ maps the segment to $-1/2$. This means that G maps the segment to $-1/2 + i\mathbf{R}$, and so by Schwarz reflection principle, we have that G can be extended to a neighbourhood of the segment, so one can make sense of the derivatives of G on this segment. Since G is an analytic function, we have that

$$\frac{\partial}{\partial 1}\Re(G) = \frac{\partial}{\partial \tau}\Re(G/\tau) = \frac{\partial}{\partial \tau^2}\Re(G/\tau^2)$$

With this facts, we can actually show that the horizontal derivative of H_3 is zero on the sides $[A_1A_3]$ and $[A_2A_3]$. Indeed, since H_1 is identically zero on the line $[A_2A_3]$, its derivative must also be zero, but since $H_1 = 2/3\Re(G) + \frac{1}{2}$, we have that

$$\begin{aligned}\frac{\partial}{\partial \tau}H_1 &= \frac{2}{3}\frac{\partial}{\partial \tau}\Re(G) \\ &= \frac{2}{3}\frac{\partial}{\partial \tau}\Re(G\tau/\tau) \\ &= \frac{2}{3}\frac{\partial}{\partial 1}\Re(G\tau) = \frac{2}{3}\times\frac{3}{2}\frac{\partial}{\partial 1}H_3\end{aligned}$$

Therefore we have the following facts about H_3:

1. H_3 is equal to zero at $[A_1A_2]$.
2. H_3 is equal to 1 at A_3.
3. The horizontal derivative of H_3 is identically zero on the sides.
4. H_3 is harmonic, being the real part of some analytic function G.

A clear candidate that ticks all these boxes is $H_3(z)$ to be the normalised height of z with respect to the base of the triangle. It turns out that using the maximum principle, we can show that this is indeed the unique solution: Let $H(z)$ be the normalised height of z, and let $\tilde{H}(z)$ be some other function satisfying those same facts. In particular, we can consider the difference $H-\tilde{H}$, which can be seen as the real part of some analytic function $G-\tilde{G}$. Since $H-\tilde{H}$ is harmonic, by the maximum principle, it must attain its maximum on the boundary of T, say on the side $[A_2A_3]$, then since its a maximum, the derivative of $H-\tilde{H}$ on the direction A_2A_3 must be zero, but by one of the facts above, the horizontal derivative of $H-\tilde{H}$ must also be zero. Since we have that the derivative of $H-\tilde{H}$ is zero in two different directions, it must mean that the complex derivative must also be zero. By Cauchy-Riemann, if the complex derivative of the real part of an analytic function is zero, then the derivative of the analytic function itself is also zero at that point. Therefore, if the maximum is attained at some z_0, we have that $\partial/\partial z(G-\tilde{G})|_{z=z_0}=0$. This means that the power expansion of $G-\tilde{G}$ at z_0 is something like

$$a+b\,(z-z_0)^k+O\left((z-z_0)^k\right)$$

for some $k\geq 2$. But whenever one is in a situation like this, it must mean that there is some direction pointing inside the triangle for which the directional derivative of the real part of $G-\tilde{G}$ is positive, contradicting our prior conclusion. As an illustration, $\Re(z^2)=x^2-y^2$, so one can easily see that if you cut the complex plane by any half plane through the origin, there will always be a direction pointing to either side of the plane in which the real part is increasing. ♡

Extension to general domains

Suppose now that D is a simply connected domain, with $a=a_1, b=a_2, c=a_3$ three points on its boundary. The key is that most of the proof works the same way: the tightness estimates did not rely on the shape of a triangle, nor did the colour-switching Lemma, which was combinatorial argument based on microscopic behavior of percolation, similarly proof that $H_1+H_2+H_3$ and $H_1+\tau H_2+\tau^2 H_3$ were analytic only relied on computing the contour integrals of triangles inside D, so this also works. Hence, for any converging subsequential limit, one obtains a triplet (H_1,H_2,H_3) for which:

1. The function $H_1+H_2+H_3$ is constant and equal to 1.

2. The function $G = H_1 + \tau H_2 + \tau^2 H_3$ is analytic.

3. The function $H_j(x)$ tends to zero when x approaches the part of the boundary between $a_{j\tau}$ and $a_{j\tau^2}$

4. $H_j(x)$ tends to one when $x \to a_j$.

Let Φ be conformal map from D onto the equilateral triangle. Then $(H_1 \circ \Phi^{-1}, H_2 \circ \Phi^{-1}, H_3 \circ \Phi^{-1})$ solve the same problem as in the equilateral triangle and since conformal maps are analytic, these functions are the "heights" of the triangle, and so the claim for the general domain follows.

Chapter 3

The Ising Model

3.1 Introduction

We now discuss a model whose roots lie in the natural phenomenon of magnetism and thermodynamics. It will be the first example of the models we will see that exhibit interaction between the sites.

Definition 3.1 (The Ising Model) Let G be a finite graph, and let $\{-1,1\}^{V(G)}$ be the set of possible states for the sites of G. Fix a $\beta > 0$, the Ising Model is a probability measure $\mathbf{P}_\beta$ on $\{-1,1\}^{V(G)}$ defined as follows:

$$\mathbf{P}_\beta[(s_x)_{x\in V}] = \frac{1}{Z_\beta} \exp\left(-\beta \sum_{\{x,y\}:x\sim y} \mathbf{1}\{s_x \neq s_y\}\right)$$

The normalising constant Z_β is called the **partition function** and in the physical literature, the parameter β is known as the **inverse temperature**.

Remark 3.2 (Motivation) The motivation behind this model, is that we want some sort of "percolation-type" model, where each site on the graph decides to have one of two states, but now, as opposed to percolation, we don't want the decisions to be independent, we want to favour alignment between neighbours, therefore, whenever two neighbours have differing opinions, we make the model pay a price in the probability, and this price comes by dividing the probability by a factor. The natural choice is to encode this in the exponential above. Of course, when β is very high, we give a high importance to interactions, and a difference between neighbours carries a high penalty to the probability. If β is zero, then there is no interaction and we are back to a percolation-type model.

Remark 3.3 (Alternative phrasing) With our choice of states being ± 1, it is clear that we can express the model as the following sum over all states:

$$\mathbf{P}_\beta[s] = \frac{1}{Z_\beta} \exp\left(\frac{\beta}{2} \sum_{x \sim y} s_x s_y\right)$$

Glauber Dynamics

The reason why models like this are of physical relevance is that the Ising measure $\mathbf{P}_\beta$ is nothing but the stationary measure of a simple "local" Markov chain on the state space, meaning that the transition of the chain is decided only by local interactions at a point. Let us now define the Markov chain of interest, and we will later state some of its properties.

1. The chain starts at some $X_0 \in \{-1, 1\}^V$.
2. To decide X_{n+1}, we choose a site $x \in V$ at random and "forget its state". X_{n+1} will be equal to X_n at all sites except at x. We will resample the state of x in the most natural way. The resulting state could be $r+$, where x obtains the value $+1$, or $r-$, where x obtains the value -1. Then $X_{n+1}(x)$ is chosen to be

$$X_{n+1}(x) = \begin{cases} r_+ & \text{with probability } \mathbf{P}_\beta[r_+]/(\mathbf{P}_\beta[r_+] + \mathbf{P}_\beta[r_-]) \\ r_- & \text{with probability } \mathbf{P}_\beta[r_-]/(\mathbf{P}_\beta[r_+] + \mathbf{P}_\beta[r_-]) \end{cases}$$

If we wanted to be explicit we could just say that r_+ is chosen with probability

$$\frac{\exp\left(\beta \sum_{y \sim x} s_y\right)}{2\cosh\left(\beta \sum_{y \sim x} s_y\right)}$$

and r_- is chosen with an analogous expression but with a $-\beta$ on the numerator. It is clear that this chain is irreducible, as one can get to any state in $\{-1, 1\}^V$ given enough steps with non-zero probability, it is aperiodic, for there is a non-zero probability of staying in the same state, and moreover, it is trivially reversible with respect to $\mathbf{P}_\beta$ (I say it is a triviality because of the way the transition probabilities have been chosen, indeed if s and s' are two states that differ only at one coordinate, say $s_x = -s'_x$, and we label by Q the transition matrix of the chain, then by definition

$$\mathbf{P}_\beta[s]/\mathbf{P}_\beta[s'] = Q(s, s)/Q(s, s')$$

It follows that the law of X_n will converge in total variation to $\mathbf{P}_\beta$.

We now derive an FKG-type result for the Ising model.

Proposition 3.4 (FKG Inequality) Let A and B be two increasing events in $\{-1,1\}^V$, then

$$\mathbf{P}_\beta[A\cap B]\geq \mathbf{P}_\beta[A]\mathbf{P}_\beta[B]$$

Proof. We will use a similar proof as the one we used for the case of percolation. We will construct two coupled Markov chains $(X_n)_n$ and $(Y_n)_n$ living in a common probability space with probability measure $\mathbf{Q}$. The goal will be to have $\mathscr{L}(Y_n)$ to converge to $\mathbf{P}_\beta[\cdot \mid B]$ and to have $X_n \leq Y_n$ for all n. Then the increasing property of A will finish off the proof. Let us describe how the coupling is constructed: construct two sequences $(X_n)_n$ and $(U_n)_n$, the first one will be vertices of the graph chosen uniformly at random, and the second sequence will be independent uniform $[0,1]$ random variables.

1. The two chains start at $X_0 = Y_0 = (+1,\cdots,+1)$. This way both X_0 and Y_0 belong to B.
2. At time n, observe the randomly chosen vertex x_n, we will update X_n and Y_n as follows: all vertices distinct from x_n remain the same, and as for x_n,
   (a) We update $X_n(x_n)$ to be $+1$ if $U_n \leq \mathbf{P}_\beta[X_n^+]/(\mathbf{P}_\beta[X_n^+]+\mathbf{P}_\beta[X_n^-])$, otherwise we update it to be -1, this way we are following the transition rules of the "natural" Markov chain.
   (b) Similarly, we update $Y_n(x_n)$ to be $+1$ if $U_n \leq \mathbf{P}_\beta[Y_n^+]/(\mathbf{P}_\beta[Y_n^+]+\mathbf{P}_\beta[Y_n^-])$ and -1 otherwise, but we only do this step **IF** doing update results in Y_n still being in B, otherwise Y_n stays put.

It is then clear to see by induction, that $X_n \leq Y_n$ for all n. Indeed, the case holds for $n=0$. Now the only case where "X could start to overtake Y" is if X flips a spin to $+1$ but Y doesn't. Let us show that this cannot happen. Our inductive hypothesis is that $X_n \leq Y_n$. Then if X flips the spin at x_{n+1} to $+1$, this means that $U_n \leq \exp\left(\beta\sum_{y\sim x_n} X_n(y)\right)/2\cosh\left(\sum_{y\sim x_n} X_n(y)\right)$ but the key observation is that the map $x \mapsto \exp(x)/2\cosh(x)$ is increasing, and so the inductive hypothesis will give that if X flips x_n to be $+1$, then so will Y. All other cases won't affect the "hierarchy of X and Y". Now we simply note that Y is reversible with respect to the conditioned measure $\mathbf{P}_\beta[\cdot \mid B]$, (to check this one does the exact same procedure as we did in the proof of FKG for percolation). We also note that (Y_n) is irreducible on the state space $\{-1,1\}$ because one has non-zero probability to get to any state from any other state.

♡

Phases of the Ising Model in $\mathbf{Z}^d$

In this limited amount of time, we will not be able to explore many interesting questions about the Ising model. We will we interested in its phase transition for very large graphs, but first of all, let's try to explain what phase transition we are after. Let $\Lambda_n = [-n, n]^d$ be the box, and consider the three following probability measures:

1. $\mathbf{P}_{\beta,n}$, the Ising measure restricted to Λ_n.
2. $\mathbf{P}^+_{\beta,n}$, the Ising measure restricted to Λ_n, conditioned on the event that all spins are $+1$ on $\partial\Lambda_n$.
3. $\mathbf{P}^-_{\beta,n}$, the Ising measure restricted to Λ_n, conditioned on the event that all spins are -1 on $\partial\Lambda_n$.

Remark 3.5 From this it is clear, that for any increasing event $A \subseteq \{-1,1\}^{\Lambda_n}$, one has that $\mathbf{P}^-_{\beta,n}[A] \leq \mathbf{P}_{\beta,n}[A] \leq \mathbf{P}^+_{\beta,n}[A]$.

Let us now introduce the observable that we will be interested in studying and analysing whether it presents a phase transition:

Definition 3.6 (Magnetisation) We define $m^+_n(\beta)$ (respectively $m^-_n(\beta)$) to be

$$m^+_n(\beta) := \mathbf{E}^+_{\beta,n}[\sigma(0)] = 2\mathbf{P}^+_{\beta,n}[\{\sigma(0)=1\}] - 1$$

We can note some a priori things:

1. $m^+_n(\beta) \geq 0$. This is because as mentioned, $\mathbf{P}^+_{\beta,n}[A] \geq \mathbf{P}_{\beta,n}[A]$ for increasing events, and $\mathbf{P}_{\beta,n}[\{\sigma(0)=1\}] = 1/2$. Naturally $\{\sigma(0)=1\}$ is an increasing event.
2. If an event A depends only on the sites in Λ_n, then $\mathbf{P}^+_n[A] = \mathbf{P}_{n+1}[A \mid \sigma(x)=1$ for $x \in \partial\Lambda_{n-1}]$, so by the FKG Inequality we proved earlier, we have that

   $$\mathbf{P}^+_{n+1}[\sigma(0)=1] \leq \mathbf{P}^+_n[\sigma(0)=1]$$

   This is intuitive, since the further away the square of $+1$'s gets, the less influence the spin at zero will feel from it, so its magnetisation will get closer to zero.

From these two observations we get that m^+_n is a bounded from below, decreasing sequence, so it has a limit, $m^+(\beta)$. The goal of studying the phase transition will be to see for which values of β one has that $m^+(\beta) = 0$ and when one has that this quantity is positive. These questions are essentially related to knowing how much does the outer boundary of the box in the Ising model influence the decision of the spins on the inside, i.e: we want to understand the correlation properties.

3.2 The Random Cluster representation of the Ising Model

The goal of this section will be to show a way of coupling the Ising Model $\mathbf{P}_\beta$ with another model called the Random Cluster Model, $\mathbf{P}_p^{\mathrm{RC}}$. This connection will allow us to explore the correlation properties of the Ising Model through the connectivity properties of a "percolation-type" model on the edges of the graph. Let us first give a definition of the Random Cluster Model

Definition 3.7 (Random Cluster Model) Let $G=(V,E)$ be a finite graph. The Random Cluster Model is a measure on $\{0,1\}^E$. For an $\omega\in\{0,1\}^E$ Letting $o(\omega), c(\omega)$, and $k(\omega)$ be the number of open edges of ω, number of closed edges, and number of open connected components respectively (including single sites). Then for each $p\in[0,1]$, we define the Random Cluster Model measure as

$$\mathbf{P}_p^{\mathrm{RC}}(\omega)=\frac{1}{Z_p^{\mathrm{RC}}}p^{o(\omega)}(1-p)^{c(\omega)}2^{k(\omega)}$$

Remark 3.8 The Random Cluster model is simply a percolation model with parameter p with the modification that configurations with more connected components are favoured.

The goal is to construct a coupling μ of the two models, i.e. a measure on $\{-1,1\}^V\times\{0,1\}^E$ such that

$$\sum_{\sigma\in\{-1,1\}^V}\mu(\sigma,\omega)=\mathbf{P}_p^{\mathrm{RC}}(\omega)\qquad\sum_{\omega\in\{0,1\}^E}\mu(\sigma,\omega)=\mathbf{P}_\beta(\sigma)$$

Moreover, we want do this in a way that the two models "talk to each other". We will now describe how this coupling is achieved. We will actually construct two couplings, μ_1 and μ_2.

1. For μ_1, proceed as follows. Construct some probability space $(\Omega,\mathscr{F},\mathbf{P})$, in which we have an Ising model σ and a Bernoulli random variable (for example, just construct a probability space where you have an Ising model and a Lebesgue $[0,1]$ random variable and then sample the Bernoulli as usual), i.e: the law of the random variable σ is precisely $\mathbf{P}_\beta$. Now construct a random bond configuration as follows: for an edge $e=(x,y)\in E$, if $\sigma(x)\neq\sigma(y)$, then $\omega(e)=0$. If the two endpoints agree however, we toss a Bernoulli(p) random variable, where $p=1-\exp(-\beta)$, and then decide whether to open or close the edge. Then we let μ_1 be the joint law of (σ,ω), i.e: μ_1 is a measure on $\{0,1\}^E\times\{-1,1\}^V$. Then from the way we have constructed this:

$$\begin{aligned}\mu_1(s,v)&=\mathbf{P}[\omega=v\mid\sigma=s]\mathbf{P}[\sigma=s]\\&=\frac{\mathbf{1}_C}{Z_\beta}\left(p^{o(v)}(1-p)^{c(v)-\sum_{x\sim y}\mathbf{1}(s_x\neq s_y)}\times\exp\left(-\beta\sum_{x\sim y}\mathbf{1}(s_x\neq s_y)\right)\right)\\&=\frac{\mathbf{1}_C\,p^{o(v)}(1-p)^{c(v)}}{Z_\beta}\end{aligned}$$

Where C is the subset of pairs of (s,v) that are compatible, i.e: that v could have arised from s, or more specifically, that if $s(x)\neq s(y)$ for an edge (x,y), then $v(x,y)=0$. Naturally, we have that the Marginal (marginalising over ω), is the Ising Model.

2. For μ_2 we proceed as follows. As before, construct some probability space $(\Omega',\mathscr{F}',\mathbf{P}')$ where one has a Random Cluster Model ω with parameter $p=1-\exp(-\beta)$and a Bernoulli $1/2$ random variable defined. First sample ω, and then construct the following random spin configuration σ: for each open connected cluster of the configuration ω, assign the spin $+1$ or -1 with equal probability (i.e: paint all the cluster with the same colour). Then letting $\mu_2(s,v)$ be the joint law of (σ,ω), we now have that

$$\mu_2(s,v)=\mathbf{P}'[\sigma=s\mid\omega=v]\mathbf{P}'[\omega=v] \tag{3.1}$$
$$=\frac{\mathbf{1}_C}{Z^{\mathrm{RC}}}\left(p^{o(v)}(1-p)^{c(v)}2^{k(v)}\right)\times 2^{-k(v)} \tag{3.2}$$
$$=\frac{\mathbf{1}_C}{Z^{\mathrm{RC}}}p^{o(v)}(1-p)^{c(v)} \tag{3.3}$$

We have therefore seen that actually $\mu_1=\mu_2$ on events of the form (s,v), but therefore they agree on the whole space as this is a generating π-system. We therefore have that $\mu=\mu_1=\mu_2$ is a coupling of the Ising Model and the Random Cluster Model (indeed, its easy to see that marginalising μ over spins and using 3.1 gives random cluster measure, and vice-versa), and moreover, the way that these two models are coupled is done in a way that one can easily translate between "Ising events" to "connectivity events" and vice-versa. Let us see an example of this mechanism. If we look at the correlation of the spins at two sites x and y, it turns out that we can relate beautifully the expectation of this quantity to the probability that x and y are connected in the Random Cluster Model.

Proposition 3.9 (Relationship between correlation and connectivity) Let $x,y\in V$. Then with $p=1-\exp(-\beta)$

$$\mathbf{E}_\beta[\sigma(x)\sigma(y)]=\mathbf{P}_p^{\mathrm{RC}}[x\longleftrightarrow y]$$

where the expectation on the left corresponds to the expectation under the Ising measure.

And the proof really is quite simple thanks to this coupling:

Proof. Let $(\Omega, \mathscr{F}, \mathbf{P})$ be a probability space with the Ising-Random Cluster coupling (σ, ω), then:

$$\begin{aligned}\mathbf{E}_\beta[\sigma(x)\sigma(y)] &= \sum_{s\in\{-1,1\}^V} s(x)s(y)\mathbf{P}_\beta[s] \\ &= \sum_{v\in\{0,1\}^E}\sum_{s\in\{-1,1\}^V} s(x)s(y)\mu(s,v) \\ &= \sum_{v\in\{0,1\}^E}\sum_{s\in\{-1,1\}^V} s(x)s(y)\mathbf{P}[\sigma = s \mid \omega = v]\mathbf{P}_p^{\mathsf{RC}}[v] \\ &= \sum_v \mathbf{E}[\sigma(x)\sigma(y) \mid \omega = v]\mathbf{P}_p^{\mathsf{RC}}[v]\end{aligned}$$

Now the key is to observe that in this last step, this expectation is precisely $\mathbf{1}(x \overset{v}{\longleftrightarrow} y)$. This is because of the way that the coupling was constructed. If we condition on the bond configuration looking like v, then if the sites are connected, the spins take the same value, either -1 or $+1$, and so the product equals one. If the sites are not connected in v, then the spins at x and y were coloured at random with probability $1/2$ and so the expectation will be zero. Formally,

$$\mathbf{E}[\sigma(x)\sigma(y) \mid \omega = v] = \mathbf{E}[\sigma(x)\sigma(y)\mathbf{1}(x \overset{v}{\longleftrightarrow} y) \mid \omega = v] + \mathbf{E}[\sigma(x)\sigma(y)\mathbf{1}(\{x \overset{v}{\longleftrightarrow} y\}^c) \mid \omega = v].$$

In summary, we have that

$$\mathbf{E}_\beta[\sigma(x)\sigma(y)] = \sum_v \mathbf{1}(x \overset{v}{\longleftrightarrow} y)\mathbf{P}_p^{\mathsf{RC}}[v] = \mathbf{P}_p^{\mathsf{RC}}[x \longleftrightarrow y]$$

♡

Now we can use this same proof technique to understand our original question of the magnetisation of the origin. Suppose we have our box Λ_n, we can consider the "glued graph" obtained by identifying the boundary as one point, "morally, the graph $\Lambda_n/\partial\Lambda_n$". Call this new boundary point ∂_n. Then $m_n^+(\beta) = \mathbf{E}_\beta[\sigma(0)|\{\sigma(x) = 1 \text{ on } \partial_n\}]$. Let A be this conditioning event for simplicity. This expectation can now be in turn computed as follows:

$$\begin{aligned}\mathbf{E}_\beta[\sigma(0)|A] &= \widehat{\mathbf{E}}[\widehat{\mathbf{E}}[\sigma(0)|A,\omega]] \\ &= \widehat{\mathbf{E}}\Big[\mathbf{1}\{0 \overset{\omega}{\longleftrightarrow} \partial_n\}\widehat{\mathbf{E}}[\sigma(0)|A,\omega]\Big] + \widehat{\mathbf{E}}\Big[\mathbf{1}\{0 \overset{\omega}{\nleftrightarrow} \partial_n\}\widehat{\mathbf{E}}[\sigma(0)|A,\omega]\Big]\end{aligned}$$

Now we know that conditional on the bond configuration ω, σ assigns the same values to the clusters. On the event that $0 \longleftrightarrow \partial_n$, conditioned on $\sigma(\partial_n) = 1$, it will follow therefore that $\sigma(0) = 1$. Similarly, if we are on the event that 0 is not connected to ∂_n, then conditional on ω, $\sigma(0)$ will receive colour ± 1 with equal probability so the second expectation is zero. Therefore we are left with $\mathbf{E}_\beta^+[\sigma(0)] = \mathbf{P}_p^{\mathsf{RC}}[0 \longleftrightarrow \partial_n]$. This is the real punchline of the coupling, we have effectively translated our question about magnetisation, to a question about large-scale connectivity properties of the Random Cluster.

3.3 The Phase Transition of the Ising Model

The heart of the proof of the phase transition will be two results: firstly, that we can couple the Random Cluster Model with two Bernoulli Percolation models, and secondly, that Random Cluster models are monotone in their parameter. The latter will prove there is a transition, and the former will prove the critical temperature is non-trivial. Let us prove these statements and then combine them for the final punchline of the chapter.

Proposition 3.10 (Random-Cluster and Percolation coupling) Let ω_p^{RC} denote a realisation of the Random Cluster Model, i.e: a random variable on some probability space $(\Omega, \mathscr{F}, \mathbf{Q})$ such that $\mathbf{Q}[\omega_p^{\mathsf{RC}} \in \cdot] = \mathbf{P}_p^{\mathsf{RC}}[\cdot]$. Then we can find two realisations of Bernoulli bond percolations ω_p and $\omega_{p'}$ for some $p' < p$, such that

$$\omega_{p'} \leq \omega_p^{\mathsf{RC}} \leq \omega_p$$

almost surely.

Proof. The proof goes by a Markov Chain coupling argument. The key will be to describe a Markov Chain X_n whose invariant measure is $\mathbf{P}_p^{\mathsf{RC}}$, then the coupling with the Bernoulli Percolation will be evident. As usual, the way it goes is that at time n, one chooses an edge $e = (x, y)$ uniformly at random, and forgets its state, then we carefully define the resampling rules so that the detailed-balance equation is satisfied. Due to the fact that the Random Cluster Model penalises having a low number of connected components, we will define two kinds of resampling mechanisms.

- Case 1: suppose that the sites x and y are already connected in the graph regardless of the outcome of the resampling (this could be thought of as if there are two arms that loop around and connect x and y without making use of e. Then regardless of the outcome, no cluster will be created or destroyed, so we just need to respect $p/1-p$ rule, in particular, set

$$X_n(e) = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

  Let us verify the detailed balance equation, let Q be the transition matrix of this chain, and let η^1 denote the resulting configuration after setting $X_n(e) = 1$ and define η^0 analogously. We verify that

$$\frac{\mathbf{P}_p^{\mathsf{RC}}[\eta^1]}{\mathbf{P}_p^{\mathsf{RC}}[\eta^0]} = \frac{p}{1-p} = \frac{Q[\eta^0, \eta^1]}{Q[\eta^1, \eta^0]}$$

  so the detail-balance equation is satisfied in this case of the dynamics.

- Case 2: suppose that the sites x and y are connected only through edge e. This means that if e is open, the connected components C_x and C_y will merge into one larger component,

reducing k by one. Therefore one can now account for this by setting in this case

$$X_n(e) = \begin{cases} 1 & \text{with probability } \frac{p}{p+2(1-p)} \\ 0 & \text{with probability } \frac{2(1-p)}{p+2(1-p)} \end{cases}$$

Once again, let us verify that the detailed-balance equation holds true:

$$\frac{\mathbf{P}_p^{\mathrm{RC}}[\eta^1]}{\mathbf{P}_p^{\mathrm{RC}}[\eta^0]} = \frac{p}{2(1-p)} = \frac{Q[\eta^0, \eta^1]}{Q[\eta^1, \eta^0]}$$

as required.

Now we make the following observation,

$$\frac{p}{p+2(1-p)} = \frac{p}{2-p} \leq p.$$

So by inspecting the update dynamics of this Markov chain we have just described, we see that in both cases, the edge opens with probability at most p. Therefore, we can couple this chain (using the same coins, etc.) with a chain Y_n whose law converges to a Bernoulli p percolation, ω_p, and we will almost surely have that $X_n \leq Y_n$. Similarly, we can couple this to another chain Z_n whose law tends to that of Bernoulli percolation, now with parameter $p' = \frac{p}{2-p}$ and we will have that $Z_n \leq X_n$ almost surely. Therefore in the limit, by Strassen's Monotone Coupling Theorem we will have the desired realisations. ♡

Proposition 3.11 (Monotonicity of Random Cluster Measure) Let $0 \leq p_1 < p_2 \leq 1$. It is possible to couple two realisations $\omega_{p_1}^{\mathrm{RC}}$ and $\omega_{p_2}^{\mathrm{RC}}$ of the Random Cluster Model in a way that $\omega_{p_1}^{\mathrm{RC}} \leq \omega_{p_2}^{\mathrm{RC}}$ almost surely.

Proof. Consider the Markov Chain dynamics described the proof above, to obtain W^1 and W^2, chains that use the parameters p_1 and p_2 respectively. If we start both of them at the same state, then we obviously have that $W_0^1 \leq W_0^2$. Now by induction, one can easily check the different update cases, and since we are using the same uniform random variables to "sample the coins", as well as the fact that $p_1 \leq p_2$, $p_1' \leq p_2'$, and $p_1' \leq p_2$, as well as the fact that the inductive hypothesis forbids W_1 from updating with the coin p_1 and W_2 updating with the coin p_2', we are done. (This last comment I have done means that it cannot be the case that after choosing an edge $e = (x, y)$ to update, x and y are connected without using e in W_n^1 but not in W_n^2). In the limit we get the desired inequality for the laws. ♡

We are now ready to prove the "punchline" of this chapter.

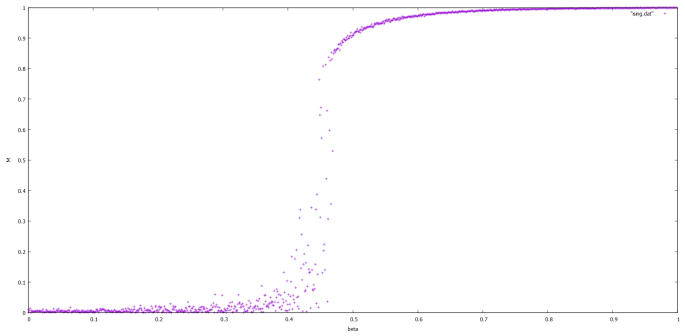


Figure 3.1: A simulation of the average magnetisation at the origin, thanks to some guy on Physics StackExchange for the image

Theorem 3.12 (Phase transition of the Ising model) There exists some $\beta_c \in (0,\infty)$, such that whenever $\beta < \beta_c$ one has that $m^+(\beta) = 0$, and whenever $\beta > \beta_c$, we have that $m^+(\beta) > 0$.

Proof. Let us recall the fundamental ingredient:

$$m_n^+(\beta) = \mathbf{P}_p^{\mathrm{RC}}[0 \longleftrightarrow \partial_n]$$

First of all, we have that at $\beta = 0$, $p = 0$, and obviously $\mathbf{P}_0^{\mathrm{RC}}[0 \longleftrightarrow \partial_n] = 0$, so taking the limit $n \to \infty$, we have that $m^+(0) = 0$. Now take formally $\beta = +\infty$, i.e. $p = 1$, it is also clear now that in this case the magnetization will be $+1$. By monotonicity, we therefore conclude that there is some critical $\beta_c \in [0, +\infty]$ for which there is a phase transition. We now show that β_c is neither of these two values. The idea is as follows: by the Markov chain coupling we have seen before, $m_n^+(\beta) = \lim_{n\to\infty} \mathbf{Q}[X_n \in \{0 \longleftrightarrow \partial_n\}]$. But due to the Bernoulli Percolation couplings we have seen,

$$\mathbf{Q}[Z_n \in \{0 \longleftrightarrow \partial_n\}] \leq \mathbf{Q}[X_n \in \{0 \longleftrightarrow \partial_n\}] \leq \mathbf{Q}[Y_n \in \{0 \longleftrightarrow \partial_n\}]$$

Taking limits we have that (recall that $\mathbf{P}_p$ denotes the Percolation law)

$$\mathbf{P}_{p'}[0 \longleftrightarrow \infty] \leq \mathbf{P}_p^{\mathrm{RC}}[0 \longleftrightarrow \infty] \leq \mathbf{P}_p[0 \longleftrightarrow \infty]$$

and now we are done, because we know that if $p < p_c(\mathbf{Z}^d)$, $\mathbf{P}_p^{\mathrm{RC}}[0 \longleftrightarrow \infty] = 0$ and so $m^+(\beta) = 0$, and if $p' > p_c(\mathbf{Z}^d)$, we have that $\mathbf{P}_{p'}[0 \longleftrightarrow \infty] > 0$, and so $m^+(\beta) > 0$. This finishes off the proof. ♡

3.4 Appendix: Infinite Ising Measures

3.4.1 Extending the Ising measures to infinite volumes

When we defined the Ising measure, we always worked with finite graphs, usually on the box Λ_n. One can however extend this to the infinite case.

- Recall that $\mathbf{P}_n^+$ is the Ising measure (we drop the dependency on β for convenience) with plus boundary conditions on the box Λ_n. We denote by σ_n^+ a random variable whose law is $\mathbf{P}_n^+$, and we define $\sigma_n^+(x) = +1$ for all $x \in \Lambda_n^c$ by convention and all of this in the same way for minus signs. The first step to extend these measures to the whole of $\mathbf{Z}^d$ is to show that one can couple all the processes $(\sigma_n^+)_{n\geq 1}$ in such a way that $\sigma_{n+1}^+ \leq \sigma_n^+$. This argument is very similar to the one we used for the FKG inequality but is slightly more subtle:

  1. By Kolmogorov's Extension Theorem, if we can couple any finite number of these processes, that is to say, for any N, find a probability space $(\Omega, \mathscr{F}, \mathbf{P})$ on which there are σ_n^+ for $n = 1, \cdots, N$ such that $\mathbf{P}$-almost surely $\sigma_{n+1}^+(x) \leq \sigma_n^+(x)$ for all $x \in \mathbf{Z}^d$, then we will be done.
  2. Define Markov chains $(X_m^{(n)})$ where $m \geq 0$ and $n = 1, \cdots, N$ that all start at $(+1, \cdots, +1) \in \{-1, 1\}^{\Lambda_N}$. Then we will run these processes simultaneously with Glauber dynamics to couple them.
  3. Suppose without loss of generality that on this space we have a collection $(U_m)_m$ of i.i.d Uniform $[0,1]$ random variables, and a collection $(x_m)_m$ of i.i.d Uniform random variables on Λ_N. Then for each m, we update $X_m^{(n)}$ for all n at the uniformly chosen vertex x. If x is outside Λ_n, then $X_m^{(n)}$ remains the same, otherwise, we use the toss of the U_m random variable to update $X_m^{(n)}$ as described earlier in these notes.
  4. This procedure gives immediately that $X_m^{(N)} \leq X_m^{(N-1)} \leq \cdots \leq X_m^{(1)}$ for all m, and by the Markov chain convergence Theorem, the laws of these chains will converge to their respective $\mathbf{P}_n^+$ measures. Therefore by Strassen's Monotone Coupling Theorem, there exists a probability space on which there are random variables $(\sigma_n^+)_n$ with $\sigma_{n+1}^+(x) \leq \sigma_n^+(x)$ for all $x \in \mathbf{Z}^d$ such that $\sigma_n^+ \sim \mathbf{P}_n^+$.

- With this coupling in mind, we now note that for each x, $\sigma_n^+(x)$ is a sequence that is bounded and almost surely non-increasing, from which we get that $\sigma^+(x) = \lim_{n\to\infty} \sigma_n^+(x)$ is almost surely well defined. Then we denote $\mathbf{P}^+$ the measure on $\{-1, 1\}^{\mathbf{Z}^d}$ given by the law of σ^+. We define $\mathbf{P}^-$ in a similar way.

With this coupling and measures, we can now make some initial observations, for example:

Proposition 3.13 Let $(\sigma^+(x))_{x\in\mathbf{Z}^d}$ be the random variables defined above. Then $\mathbf{P}[\sigma^+(x)=+1]$ does not depend on x.

Proof. For this we simply note that if e denotes a unit vector on $\mathbf{Z}^d$, then we have the following inclusion of boxes:

$$\Lambda_n \subseteq \Lambda_{n+1}+e \subseteq \Lambda_{n+2}.$$

Now let S be a finite set, and define the event

$$A_S = \{s(x)=+1 \text{ on all } x\in S\} \subseteq \{-1,+1\}^{\mathbf{Z}^d}.$$

Events of this form are a generating π-system, and we have that by our monotone coupling, we can obtain processes $\sigma_n^+, \sigma_{n+1}^{+,e}, \sigma_{n+2}^+$ to be Ising models with plus boundary conditions on the boxes above with the monotonicity property. We therefore have that

$$\mathbf{P}[\sigma_n^+\in A] \geq \mathbf{P}[\sigma_{n+1}^{+,e}\in A] \geq \mathbf{P}[\sigma_{n+2}^+\in A],$$

and so taking limits and noting that $(\sigma_{n+1}^{+,e}(x): x\in\Lambda_{n+1}+e)$ has the same distribution as $(\sigma_{n+1}^+(x+e): x\in\Lambda_{n+1})$, the claim follows. ♡

Another consequence of the coupling is the following proposition:

Proposition 3.14 Letting $\mathbf{P}^+$ and $\mathbf{P}^-$ be defined as above, we have that $\mathbf{P}^+=\mathbf{P}^-$ if and only if $\mathbf{P}^+[s(0)=+1]=1/2$.

Proof. Suppose that the two measures are indeed equal, then we have that

$$1=\mathbf{P}^+[s(0)=+1]+\underbrace{\mathbf{P}^+[s(0)=-1]}_{=\mathbf{P}^-[s(0)=+1]}$$

but by hypothesis we now have that $1=2\mathbf{P}^+[s(0)=+1]$. The converse is slightly more tricky: let $A\subseteq\mathbf{Z}^d$ be any finite subset. Then by π-system arguments, if $\mathbf{P}^+[\{s(x)=+1: x\in A\}]=\mathbf{P}^-[\{s(x)=-1: x\in A\}]$, we will deduce that the two measures are equal. The trick to this proposition is to consider the function

$$F_A(s)=\sum_{x\in A}\mathbf{1}\{s(x)=+1\}-\prod_{x\in A}\mathbf{1}\{s(x)=+1\}.$$

We make two preliminary observations:

1. $F_A(s)$ is an increasing function in s. Therefore, if we let $(\sigma^+(x): x\in\mathbf{Z}^d)$ and $(\sigma^-(x): x\in\mathbf{Z}^d)$ be the monotone coupling described in this section, we have that almost surely, $F_A(\sigma^+)\geq F_A(\sigma^-)$,

from which we deduce that

$$\mathbf{E}\left[\sum_{x\in A}\mathbf{1}\{\sigma^+(x)=+1\}\right]-\mathbf{E}\left[\sum_{x\in A}\mathbf{1}\{\sigma^-(x)=+1\}\right]\geq\mathbf{E}\left[\prod_{x\in A}\mathbf{1}\{\sigma^+(x)=+1\}\right]-\mathbf{E}\left[\prod_{x\in A}\mathbf{1}\{\sigma^-(x)=+1\}\right]. \tag{3.4}$$

2. the function $s\mapsto\prod_{x\in A}\mathbf{1}\{s(x)=+1\}$ is also increasing in s. Therefore the right-most side of 3.4 is bounded below by 0.

3. By assumption of $\mathbf{P}^+[s(0)=+1]=1/2$, we incidentally also have that $\mathbf{P}^-[s(0)=+1]=1/2$, and by the translation invariance we proved earlier, we indeed have that the left-most side of 3.4 is equal to zero. Therefore we deduce that

$$\mathbf{E}\left[\prod_{x\in A}\mathbf{1}\{\sigma^+(x)=+1\}\right]=\mathbf{E}\left[\prod_{x\in A}\mathbf{1}\{\sigma^-(x)=+1\}\right]$$

Which means that

$$\mathbf{P}^+[\{s(x)=+1:x\in A\}]=\mathbf{P}^-[\{s(x)=+1:x\in A\}],$$

as required.

♡

3.4.2 Infinite Ising measures

Definition 3.15 (Ising model with boundary conditions) Let $\xi:\partial\Lambda_n\to\{-1,1\}$. We define the Ising model on Λ_n with boundary conditions ξ and parameter β to be the Ising measure on Λ_n, conditioned on being equal to ξ on $\partial\Lambda_n$. Formally, we denote

$$\mathbf{P}_n^\xi[s]=\mathbf{P}_n[s\mid\{(x)=\xi(x)\text{ for all }x\in\partial\Lambda_n\}]$$

Definition 3.16 (Infinite Ising measure) We say that the law of a collection $(\sigma(x):x\in\mathbf{Z}^d)$ of random variables is an Infinite Ising Measure, if for any n, the conditional distribution of $(\sigma(x):x\in\Lambda_n)$ given $(\sigma(x):x\in\Lambda_{n-1}^c)$ is $\mathbf{P}_n^\xi$, where $\xi=\sigma$ on $\partial\Lambda_n$. That is to say, for any measurable $A\subseteq\{-1,1\}^{\Lambda_n}$:

$$\mathbf{P}[(\sigma(x):x\in\Lambda_n)\in A\mid\sigma(x):x\in\Lambda_{n-1}^c]=\mathbf{P}_n^{\sigma|\partial\Lambda_n}[A].$$

Remark 3.17 The definition of an Infinite Ising measure is simply introducing this sort of Spatial Markov Property: the conditional distribution of the spin configurations inside a box Λ_n given what happens on the entire outside, is the same as if we had an Ising measure on Λ_n with the given boundary conditions. In other words, given the value of the boundary, the outside of the box has no impact on the interactions inside of the box. Intuitively it makes sense that this property should hold in our usual infinite volume Ising measures $\mathbf{P}^+$ and $\mathbf{P}^-$, since the interactions they see are local, i.e: only are affected one neighbour away, thus we see that if we condition on the boundary, then the outside will not interact with the inside. In fact this is exactly the method of proof of the following proposition:

Proposition 3.18 The measure $\mathbf{P}^+$ (and $\mathbf{P}^-$) are infinite Ising measures.

Proof. For a finite set $S \subseteq \mathbf{Z}^d$, let us denote by $\mathscr{F}_S$ the sigma algebra generated by all the spins in S. Let σ be a random variable distributed according to $\mathbf{P}^+$, let n be given, and let f be a $\mathscr{F}_n$ measurable function. We wish to show that

$$\mathbf{E}\big[f((\sigma|\Lambda_n)\,|\,\mathscr{F}_{\Lambda_{n-1}^c}\big] = \mathbf{E}_{\Lambda_{n-1}}^{\sigma|\partial\Lambda_n}[f],$$

where $\sigma|\Lambda_n$ denotes σ restricted to Λ_n. Hence we need to show that for any set $E \in \mathscr{F}_{\Lambda_{n-1}^c}$, one has

$$\mathbf{E}\big[f(\sigma|\Lambda_n)\mathbf{1}_E\big] = \mathbf{E}\Big[\mathbf{E}_{n-1}^{\sigma|\partial\Lambda_n}[f]\mathbf{1}_E\Big].$$

By a π-system argument, it suffices to show the claim for any set $E \in \mathscr{F}_{\Lambda_N\setminus\Lambda_{n-1}}$, where N is some given integer. They key here is that the Hamiltonian of the Ising model separates nicely, indeed:

$$\mathbf{P}_N[s] := \frac{1}{Z_N}\exp\Big(-\beta\left\{H_{\Lambda_N\setminus\Lambda_{n-1}}(s) + H_{\Lambda_{n-1}}^{s|\partial\Lambda_n}(s)\right\}\Big).$$

Therefore we can now start computing (for notational convenience, sometimes we'll write $s \in \Lambda_k$ to indicate that s is a configuration of spins for Λ_k, so that in reality we should be writing $s \in \{-1,1\}^{\Lambda_k}$:

$$\begin{aligned}
\mathbf{E}\big[f(\sigma)\mathbf{1}_E\big] &= \frac{1}{Z_N}\sum_{s\in\Lambda_N} f(s|\Lambda_n)\exp\Big(-\beta\left\{H_{\Lambda_N\setminus\Lambda_{n-1}}(s) + H_{\Lambda_{n-1}}^{s|\partial\Lambda_n}(s)\right\}\Big)\mathbf{1}_E(s)\\
&= \frac{1}{Z_N}\sum_{s\in\Lambda_N\setminus\Lambda_{n-1}} \exp\big(-\beta H_{\Lambda_N\setminus\Lambda_{n-1}}(s)\big)\mathbf{1}_E(s)\left(\sum_{w\in\Lambda_{n-1}} f(w)\exp\Big(-\beta H_{\Lambda_{n-1}}^{s|\partial\Lambda_n}(w)\Big)\right)\\
&= \frac{1}{Z_N}\sum_{s\in\Lambda_N\setminus\Lambda_{n-1}} \exp\big(-\beta H_{\Lambda_N\setminus\Lambda_{n-1}}(s)\big)\mathbf{1}_E(s)\mathbf{E}_{\Lambda_{n-1}}^{s|\partial\Lambda_n}[f]\left(\sum_{w\in\Lambda_{n-1}} \exp\Big(-\beta H_{\Lambda_{n-1}}^{s|\partial\Lambda_n}(w)\Big)\right)\\
&= \mathbf{E}\Big[\mathbf{E}_{n-1}^{\sigma|\partial\Lambda_n}[f]\mathbf{1}_E\Big].
\end{aligned}$$

On the penultimate line, on the right-most side we have the sum over $w \in \Lambda_{n-1}$ corresponding to the partition function from $\mathbf{E}_{n-1}^{\sigma|\partial\Lambda_n}[f]$. ♡

We have one last proposition:

Proposition 3.19 $\mathbf{P}^+ = \mathbf{P}^-$ if and only if there exists a unique Ising measure.

Proof. The key is that if we let $\mathbf{P}$ be an infinite Ising measure, then for any cylindrical event $A \in \mathscr{F}_{\Lambda_n}$, we have that

$$\mathbf{P}^-[A] \leq \mathbf{P}[A] \leq \mathbf{P}^+[A],$$

this is because $\mathbf{P}[A] = \mathbf{E}[\mathbf{P}_n^{\sigma|\partial\Lambda_n}[A]]$, and by the monotone coupling argument, $\mathbf{P}_n^-[A] \leq \mathbf{P}_n^{\sigma|\partial\Lambda_n}[A] \leq \mathbf{P}_n^+[A]$. ♡

Chapter 4

The Discrete Gaussian Free Field

4.1 Introduction

The Discrete Gaussian Free Field (DGFF) is a model that can be thought of as a "generalisation" of the Ising model, to the case where each vertex of the graph, instead of adopting either a zero or a one, adopts a real number, which can be interpreted as a "height". Just as the Ising Model favoured configurations in which the opinions of neighbouring vertices didn't differ by much, the DGFF will favour configurations in which the heights of adjacent vertices don't differ by much. A nice interpretation of this is that each vertex is connected to its neighbours by a spring with quadratic energy, and the boundary of the graph is held in place at height zero. Let us introduce some notation:

- Let $d \geq 1$, for a subset $D \subseteq \mathbf{Z}^d$, we define ∂D, to be $\partial D = \{x \in \mathbf{Z}^d : \text{dist}(x, D) = 1\}$.
- We will denote $\mathbf{F}_{(D)}$ to be the set of functions from $\mathbf{Z}^d$ to $\mathbf{R}$ that are equal to zero outside of D.
- We define $E_{(D)}$ to be the set of edges with at least one endpoint in D.
- For an edge $e \in E_{(D)}$, and a function $F \in \mathbf{F}_{(D)}$, we define $|\nabla F(e)| = |F(x) - F(y)|$, where x, y are the endpoints of e.
- Finally, we define for a function $F \in E_{(D)}$, its Dirichlet Energy

$$\mathcal{E}_D(F) = \sum_{e \in E_{(D)}} |\nabla F(e)|^2$$

Naturally the Dirichlet Energy measures the total "heigh disagreement" of our vertices. We may now define the DGFF via its density function.

Definition 4.1 (The Discrete Gaussian Free Field) The DGFF in D with zero boundary conditions on ∂D is a random vector $(\Gamma_x)_{x\in D}$ whose density function on $\mathbf{R}^D$ at $(\gamma_x)_{x\in D}$ with respect to the Lebesgue measure is given by a constant multiple of

$$\exp\left(-\frac{1}{2}\times\frac{\mathcal{E}_D(\gamma)}{2d}\right)$$

with the convention that $\gamma_x = 0$ for all $x\in\partial D$.

Remark 4.2 (Gaussian vector) Note that we can think of the energy $\mathcal{E}_D(\gamma)$ as actually $\mathcal{E}_D(\gamma,\gamma)$ where

$$\mathcal{E}_D(F,G)=\sum_{e\in E_{(D)}}(F(x)-F(y))(G(x)-G(y))$$

so that $\mathcal{E}_D$ is actually a symmetric bilinear form on $\mathbf{R}^D\times\mathbf{R}^D$. Moreover, it is positive definite because if $\mathcal{E}_D(\gamma)=0$, then $|\nabla\gamma(e)|=0$ for all edges e, and since we have that on the boundary the value of γ is zero, it must be that γ is zero everywhere. Therefore we have that the density of our random vector is proportional to $\exp\left(-\frac{1}{2}\times(\text{symmetric bilinear form})\right)$ it follows that the random vector has the density function of a centered Gaussian random vector. (This is because we can express a symmetric bilinear, and positive definite form H on $\mathbf{R}^D\times\mathbf{R}^D$ as $H(x)=x^TAx$ for some symmetric positive definite matrix, now we can directly compare with the Gaussian vector density, in fact this A will be the inverse of our covariance matrix σ). Recall moreover, that the law of a centered Gaussian random vector is uniquely determined by its covariance matrix σ. As we will see soon, the covariance matrix has a nice interpretation.

Remark 4.3 (No phase transition) We could introduce a parameter β much like we did for the Ising model, so that our density now looks like some

$$\exp\left(-\frac{1}{2}\times\frac{\beta\mathcal{E}_D(\gamma)}{2d}\right)$$

heuristically, this parameter would control the "stiffness" of the springs, but if we inspect this expression, we could just introduce this β as a $\sqrt{\beta}$ into the energy, and so the law of this new Gaussian Free Field $(\Gamma^\beta_x)_{x\in D}$ is nothing but the law of $\left(\frac{1}{\sqrt{\beta}}\Gamma_x\right)_{x\in D}$, where Γ_x is the original field, so all that the introduction of this parameter does, is rescale the field. From this we see that there will be no qualitative phase transition.

4.2 Green's Function and Δ_D

Let us begin this section by giving some notation

- For a function $f : \mathbf{Z}^d \to \mathbf{R}$, we define $\bar{f}(x)$ to be the average of the neighbours of x, i.e:

$$\bar{f}(x) = \frac{1}{2d} \sum_{y : y \sim x} f(y)$$

- We also define the discrete Laplacian $\Delta f(x)$ to be the difference between the average of the neighbours and f (this definition makes sense when compared to the usual definition of Laplacian because $\Delta f = 0$ means that the value of the function is equal to the average of its neighbours), i.e.

$$\Delta f(x) = \bar{f}(x) - f(x)$$

- For our domain D, we may define $\Delta_D f$ to be the function equal to $\Delta f(x)$ for all $x \in D$ and zero otherwise.

Remark 4.4 (Inverse of Δ_D) Notice the following Δ_D can be seen as a linear operator $\mathbf{F}_{(D)} \to \mathbf{F}_{(D)}$, and it is easy to check that it is injective: let $x_0 \in D$ be such that $|F(x_0)| = \max_{x \in D} |F(x)|$. If $\Delta_D F = 0$, then $\bar{F}(x_0) = F(x_0)$, which means that for all neighbours y of x_0, the value of $F(y)$ is equal to the value of x_0. Repeating this argument we can find a sequence of neighbours $x_0, x_1, x_2, \cdots, y$ that lead to some point y on the boundary of D, and $F(x_0) = F(x_1) = \cdots = F(y) = 0$. Therefore the function F is identically zero. From this it follows that actually Δ_D is invertible.

It turns out that this inverse has a name:

Definition 4.5 (Green's Function) Let X be a simple random walk on $\mathbf{Z}^d$, with law denoted by $\mathbf{P}_x$ when it is started at x. Let $\tau_D := \inf\{t > 0 : X_t \notin D\}$, i.e: the first exit time. Then Green's function $G_D(x, y)$ is a function $D \times D \to \mathbf{R}$ defined by the expected number of visits to y by the random walk before leaving D when started from x, i.e:

$$G_D(x, y) = \mathbf{E}_x \left[\sum_{k=0}^{\tau_D - 1} \mathbf{1}\{X_k = y\} \right].$$

Proposition 4.6 (Green's function and Δ_D) Green's function G_D is the inverse of $-\Delta_D$.

Proof. If we label $D=\{x_1,\cdots,x_n\}$, then we can treat G_D,Δ_D and σ as $n\times n$ matrices. In particular, note that Δ_D in matrix form has the following description: $(\Delta_D)_{ii}=-1$ and for $i\neq j$, $(\Delta_D)_{ij}=\frac{1}{2d}$ if x_i is a neighbouring vertex of x_j and zero otherwise. From this, it is immediate that the matrix $I+\Delta_D$ is the transition matrix of a simple random walk on $\mathbf{Z}^d$ but restricted to D. Let us refer with P_D to this matrix. Then, since this random walk occurs restricted to D, we have that for all k,

$$\mathbf{P}_x[X_k=y,k<\tau_D]=(P_D)^k(x,y)$$

Now we can note that Green's function can be rewritten as

$$G_D(x,y)=\sum_{k\geq 0}\mathbf{P}_x[X_k=y,k<\tau_D]$$

and so we have that

$$G_D(x,y)=\sum_{k\geq 0}(P_D)^k(x,y)$$

From this, it is clear that $G_D(I-P_D)=I$ and hence the claim is proven. ♡

4.3 The resampling procedure and its consequences

In previous models, we discussed a natural dynamics that gave rise to an invariant distribution which was the model itself. In the case of the Gaussian Free Field, the same techniques will not work since the relevant dynamics would be on uncountable state space, but nonetheless we can discuss some natural resampling dynamics (which can be proven to converge to the Gaussian Free Field) and explore some of its consequences. Let us begin by stating the following fact:

Proposition 4.7 For any $x\in D$, one has that $\Gamma(x)-\bar{\Gamma}(x)\sim\mathcal{N}(0,1)$ and is independent of $\{\Gamma(y)\}_{y\neq x}$.

Proof. We begin by exploring the conditional distribution of $\Gamma(x)$ given the state of all other vertices. I.e: what's the conditional density of $\Gamma(x)$ given $\{\Gamma(y)\}_{y\neq x}=\{h(y)\}_{y\neq x}$? Well we know by elementary conditional probability that if f is the density of a random vector $(X_1,\cdots,X_n)$, then the conditional density of say X_1 given $(X_2,\cdots,X_n)=(y_2,\cdots,y_n)$ is precisely

$$\frac{f(x_1,y_2,\cdots,y_n)}{\int f(x_1,y_2,\cdots,y_n)\mathrm{d}x_1}$$

in our case, thanks to the fact that the density function f comes as a product of exponentials,

this expression heavily simplifies to saying that the conditional density of $\Gamma(x)$ conditioned on $\{\Gamma(y)\}_{y\neq x} = \{h(y)\}_{y\neq x}$ is

$$\frac{1}{c}\exp\left(-\frac{1}{2}\times\frac{1}{2d}\sum_{y:y\sim x}|\gamma(x)-h(y)|^2|\right)\mathrm{d}\gamma$$

where c is a normalising constant, and we wrote the $\mathrm{d}\gamma$ to emphasise that γ is the value to be taken by $\Gamma(x)$ and the density is with respect to Lebesgue measure. Now if one expands the bracket one sees that it is actually equal to

$$\frac{1}{c'}\exp\left(-\frac{1}{2}(\gamma(x)-\bar{h}(x))^2\right)\mathrm{d}\gamma$$

Indeed:

$$\begin{aligned}\frac{1}{c}\exp\left(-\frac{1}{2}\frac{1}{2d}\sum_{y\sim x}(\gamma(x)-h(y))^2\right) &= \frac{1}{c}\exp\left(-\frac{1}{2}\left(\gamma(x)^2-2\gamma(x)\bar{h}(x)+\text{something}^2\right)\right)\\ &= \frac{1}{c}\exp\left(-\frac{1}{2}\left(\gamma(x)^2-2\gamma(x)\bar{h}(x)+\bar{h}(x)^2+(-\bar{h}(x)^2+\text{something}^2)\right)\right)\\ &= \frac{1}{c'}\exp\left(-\frac{1}{2}(\gamma(x)-h(x))^2\right)\end{aligned}$$

which means that the conditional distribution of $\Gamma(x)$ is that of a $\mathcal{N}(\bar{h}(x),1)$ random variable. From this we immediately have that $\Gamma(x)-\bar{h}(x)$ is independent of the remaining values and has a $\mathcal{N}(0,1)$ distribution and so $\Gamma(x)-\bar{\Gamma}(x)$ is independent of $\{\Gamma(y)\}_{y\neq x}$ and moreover, its distribution is that of a $\mathcal{N}(0,1)$. ♡

Remark 4.8 (Resampling dynamics) We can make the following heuristic note on a resampling mechanism whose law tends to that of the Gaussian Free Field. Given a function $h\in\mathbf{R}^d$, choose one of the coordinates x at random, and resample the value at x by setting it equal to $\bar{h}(x)+N$ where N is an iid realisation of a standard Gaussian. It can be proven, although we will not show it here, that these dynamics converge to the DGFF law.

A more important consequence of the proposition above, is that it can be used to show that the covariance matrix of the DGFF is actually Green's function. This is quite a strong result, because since the law of a centered Gaussian random vector is fully determined by its covariance, this is saying that there is a direct relationship between the Gaussian Free Field on a graph G and a simple random walk on G.

Proposition 4.9 (Covariance matrix of DGFF) The covariance matrix $\sigma = \sigma(x,y)$ of the DGFF is equal to Green's Function.

Proof. By definition, the covariance matrix has the form $\sigma(x,y) = \mathbf{E}[\Gamma(x)\Gamma(y)]$. We can momentarily treat this as a function, for a fixed x, $y \mapsto \sigma_x(y) = \mathbf{E}[\Gamma(x)\Gamma(y)]$, which is a function in $\mathbf{F}_{(D)}$. Now we can make the following computation:

- If $y \neq x$, then

$$\begin{aligned}\sigma_x(y) &= \mathbf{E}[\Gamma(x)\Gamma(y)] \\ &= \mathbf{E}[\Gamma(x)\bar{\Gamma}(y)] + \mathbf{E}[\Gamma(x)(\Gamma(y)-\bar{\Gamma}(y))] \\ &\overset{(!)}{=} \mathbf{E}[\Gamma(x)\bar{\Gamma}(y)] \\ &= \frac{1}{2d}\sum_{z:z\sim y} \mathbf{E}[\Gamma(x)\Gamma(z)] = \bar{\sigma}_x(y)\end{aligned}$$

  Where in step (!) we used Proposition 4.7 to split the second expectation as a product and moreover used the fact that $\Gamma(y)-\bar{\Gamma}(y)$ is a centered Gaussian independent of $\Gamma(x)$. This shows that whenever $x \neq y$, $(\Delta\sigma)(x,y) = 0$.

- Otherwise:

$$\begin{aligned}\sigma_x(x) &= \mathbf{E}[\Gamma(x)\Gamma(x)] \\ &= \mathbf{E}[\Gamma(x)\bar{\Gamma}(x)] + \mathbf{E}[\Gamma(x)(\Gamma(x)-\bar{\Gamma}(x))] \\ &= \mathbf{E}[\Gamma(x)\bar{\Gamma}(x)] + \mathbf{E}[(\Gamma(x)-\bar{\Gamma}(x))^2] + \mathbf{E}[(\Gamma(x)-\bar{\Gamma}(x))\bar{\Gamma}(x)] \\ &\overset{(!)}{=} \mathbf{E}[\Gamma(x)\bar{\Gamma}(x)] + 1 + 0 \\ &= \frac{1}{2d}\sum_{y:y\sim x} \mathbf{E}[\Gamma(x)\Gamma(y)] = \bar{\sigma}_x(x) + 1\end{aligned}$$

  Where step (!) comes from the fact that since $\bar{\Gamma}(x)$ is really just a sum of $\Gamma(y)$ for vertices y different to x, we can still use the fact that $\Gamma(x)-\bar{\Gamma}(x)$ is independent of all other $\Gamma(y)$'s, and so we can split the expectation as before. This shows us that $(\Delta\sigma)(x,x) = -1$, this means that $(-\Delta_D)^{-1} = \sigma$, but by Proposition 4.6, it now follows that $\sigma = G_D$.

♡

Thus we can phrase this discovery as the following equivalent characterisation of the Gaussian Free Field

Corollary 4.10 The DGFF in D with zero boundary conditions on ∂D is the centered Gaussian random vector $(\Gamma(x))_{x\in D}$ with covariance matrix G_D on $D\times D$.

Remark 4.11 Not only does this provide a very interesting point of view on what the Discrete Gaussian Free Field is doing, but it allows one to extend the definition to infinite (proper) subsets D of $\mathbf{Z}^d$ as long as $d\geq 3$, because by transience, the walk will eventually exit D.

4.4 The Spatial Markov Property

When we talked about percolation, a tool we often used was the fact that if we revealed the state of some sites, the distribution of what remained was independent Bernoulli percolation. This property is some sort of Markov property but in the spatial sense, much like how in a random walk, if we condition on the value of the walk at a given time, the distribution of what's to come is simply a shifted random walk. Can we find some analogue of this property for the Discrete Gaussian Free Field? Before doing anything else, we note that we can extend the definition of the Gaussian Free Field to any setting of boundary conditions

Definition 4.12 (Discrete Gaussian Free Field, non-zero boundary conditions) The Discrete Gaussian Free Field on a domain D with boundary conditions f, where $f:\partial D\to\mathbf{R}$ is a function, is the random vector $\{\Gamma(x)\}_{x\in D}$ with density function proportional to

$$\exp\left(-\frac{1}{2}\times\frac{1}{2d}\mathscr{E}(\gamma)\right)\mathrm{d}\gamma$$

with the convention that $\gamma=f$ on ∂D.

Remark 4.13 We have some remarks about this definition:

- First of all, we note that the values of f are indeed used in calculating the above density, as they are used when computing the term $|\nabla\gamma(e)|$ when e is an edge in $\bar{D}$ that contains an endpoint in ∂D.
- Secondly, we note that this will still be a Gaussian random vector, just that it may not be a centered one.
- If $f=c$ identically for some constant c, then since translating does not affect the covariance

structure, $\{\Gamma(x)-c\}_x$ is still a Discrete Gaussian Free Field, but now with zero boundary conditions.

We now have our first version of the Markov Property

Proposition 4.14 (Spatial Markov Property V1) Let $O \subseteq D$ be a subset of our domain. Let Γ be a DGFF on D. Then conditional on the event that $\Gamma(x) = f(x)$ for all $x \notin O$, the distribution of $\{\Gamma(x)\}_{x\in O}$ is that of a Gaussian Free Field with boundary conditions f.

Proof. We once again recall that the density of the DGFF is given by

$$\rho(\gamma) = \frac{1}{c}\prod_{e\in E_{\vec{D}}} \exp\left\{-\frac{1}{4d}|\nabla\gamma(e)|\right\}$$

it is precisely this multiplicative form that will give us the behaviour we want. Suppose that we enumerate $D = \{x_1, \cdots, x_k, x_{k+1}, \cdots, x_n\}$ and without loss of generality, we can enumerate our domain O the points x_{k+1} onwards. If we condition on $\Gamma(x) = f(x)$ for $x = x_1, \cdots, x_k$, we can easily see that the conditional density will be

$$\rho_{|O^c}(\gamma) \propto \frac{\prod_{e\in E_{\vec{D}}} \exp\left\{-\frac{1}{4d}|\nabla\gamma(e)|\right\}}{\int \prod_{e\in E_{\vec{D}}} \exp\left\{-\frac{1}{4d}|\nabla\gamma(e)|\right\} \mathrm{d}x_{k+1}\cdots\mathrm{d}x_n}$$

Where it is understood that $\gamma(x) = f(x)$ for all $x \in O$. The multiplicative structure of the integral downstairs makes it so that whenever an edge e has both endpoints that are outside O, i.e: none of $x_{k+1}, \cdots, x_n$, we can take the corresponding exponential out and it will cancel with the corresponding exponential in the numerator, which means that the only edges in the numerator will be those whose both endpoints are in O or one of the endpoints is in ∂O. From this, it follows directly, that this density is simply that of a DGFF in O, with boundary conditions f. ♡

This is already a step in the right direction to a Spatial Markov Property, but ideally, we would like that the resulting conditional distribution was not that of a DGFF with altered boundary conditions, but rather simply that of a "standard" DGFF. Of course morally this cannot be true, but it turns out that we can achieve a close result: as we have just shown, after we reveal (or condition on) the state of the (standard = zero boundary conditions) Gaussian Free Field on some subset O^c, we are left with a DGFF on O that has boundary conditions f, on ∂O given by the values that we conditioned our original field to take on ∂O. As we will now see, we may decompose this field we have just obtained as a sum of the Harmonic extension to O of f, plus a standard Gaussian Free Field.

Theorem 4.15 (The Spatial Markov Property) If $\{\Gamma(x)\}_{x\in D}$ is a DGFF in a domain D with boundary conditions f, then letting F be the Harmonic extension of f to D, we have that

$\{\Gamma(x)-F(x)\}_{x\in D}$ is a standard DGFF.

Proof. We will first prove some sort of "opposite fact": if Γ has Dirichlet (zero) boundary conditions, and F is the Harmonic extension of f to D, then $\Gamma+F$ is a DGFF with boundary conditions equal to f. For reasons that will become clear in a moment, let us define the following inner product on functions $\mathbf{Z}^d \to \mathbf{R}$ with finite support:

$$\langle F_1, F_2\rangle = \frac{1}{2}\times\frac{1}{2d}\sum_{x\in\mathbf{Z}^d}\sum_{y:y\sim x}\big(F_1(y)-F_1(x)\big)\big(F_2(y)-F_2(x)\big)$$

Then the first thing to notice is that we can think of Γ having a density on $\mathbf{R}^{\bar{D}}$ (including the boundary) of

$$\rho_\Gamma(\gamma) = \frac{1}{c}\exp\big(-\langle\gamma,\gamma\rangle\big)\mathbf{1}\{\gamma = 0 \text{ on } \partial D\}$$

(The reason why I'm including this last indicator term is to really emphasize that I am now thinking of the density as a density on configurations including potentially different values than f on ∂D, so to fix this, instead of asking a priori that this density is valid for γ that have value f on the boundary, I can allow this density on all γ and then just include an indicator function, this will make our life easier later). Now suppose that I introduce $\widetilde{\Gamma} = \Gamma + F$. Now we set off to investigate the density function of $\widetilde{\Gamma}$:Recall that if a random variable X has density $f(x)$, then the random variable $X+c$ has density $f(x-c)$

$$\rho_{\widetilde{\Gamma}}(\gamma) = \frac{1}{c'}\exp\big(-\langle\gamma-F,\gamma-F\rangle\big)\mathbf{1}\{\gamma-F=0 \text{ on } \partial D\}. \tag{4.1}$$

Let us now investigate what we have obtained in this inner product. A key step is to realise that we may rewrite the inner product, using symmetry, as follows:

$$\begin{aligned}
\langle F_1, F_2\rangle &:= \frac{1}{4d}\sum_{x\in\mathbf{Z}^d}\sum_{y:y\sim x}\big(F_1(y)-F_1(x)\big)\big(F_2(y)-F_2(x)\big)\\
&= \frac{1}{4d}\sum_{x\in\mathbf{Z}^d}\sum_{y:y\sim x} -F_1(y)\big(F_2(x)-F_2(y)\big)-F_1(x)\big(F_2(y)-F_2(x)\big)\\
&= \frac{1}{2d}\sum_{x\in\mathbf{Z}^d}\sum_{y:y\sim x} -F_1(x)\big(F_2(y)-F_2(x)\big)\\
&= \sum_{x\in\mathbf{Z}^d} F_1(x)\Delta F_2(x).
\end{aligned}$$

(Here these sums I'm writing as over $\mathbf{Z}^d$ but in reality we are only summing over $\bar{D}$) In particular, since we are working with Γ having Dirichlet Boundary conditions, we have that $\gamma = 0$ on ∂D, and since F is Harmonic on D, we have that for any x, regardless if x belongs to D or ∂D, $\gamma(x)\Delta F(x) = 0$, which means that $\langle\gamma, F\rangle = 0$. Now we can use bilinearity of the inner product on

Equation 4.1 and see that in fact

$$\rho_{\widetilde{\Gamma}}(\gamma) = \frac{1}{c''} \exp\left(-\langle \gamma, \gamma \rangle\right) \mathbf{1}\{\gamma = F \text{ on } \partial D\}.$$

From which we immediately read that $\widetilde{\Gamma}$ is a DGFF with boundary conditions F (and in particular f, as F is its extension). The Theorem follows now by a change of variable. ♡

In summary, we can reach the final observation now: Suppose Γ is a DGFF with Dirichlet boundary conditions on a domain D. Suppose that we condition on $\{\Gamma(x) = f(x) : x \in O^c\}$ for some subset $O \subseteq D$, then the conditional distribution of the remaining part is simply that of $\{\Gamma'(x) + F(x)\}_{x \in O}$ where Γ' is an independent DGFF on O with Dirichlet boundary conditions, and F is the Harmonic extension of f to O.

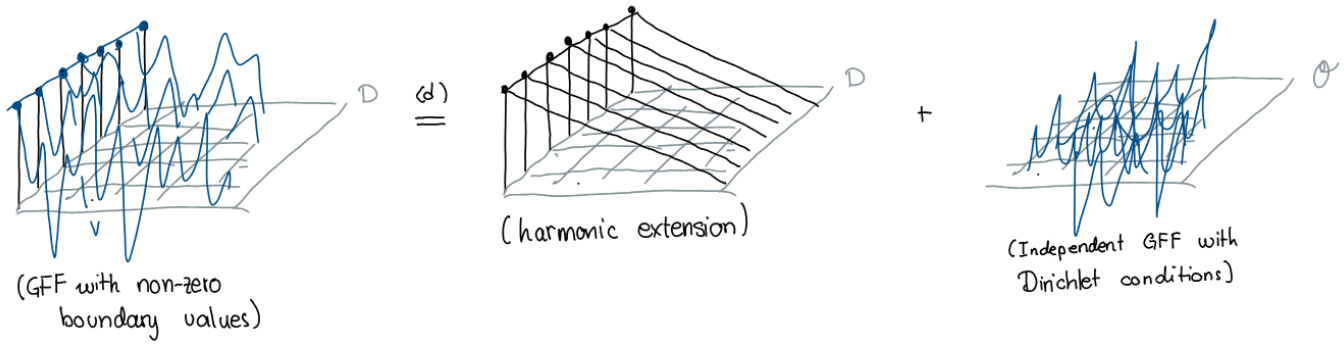


Figure 4.1: My best attempt at depicting the Spatial Markov Property

4.5 Brief comments: partition function and discovery procedure

Throughout this discussion of the Discrete Gaussian Free Field, we have been somewhat ignoring the normalising constant that makes the density $\exp\left(-\frac{1}{4d}\mathscr{E}(\gamma)\right)$ a density, if we are working in the Dirichlet boundary conditions case, it is easy to see that the normalising constant (the partition function in physics) is nothing but the density of the point $\gamma = (0, \cdots, 0)$. We will now describe an exploratory procedure of the Gaussian Free Field, that relies on the Spatial Markov Property discussed earlier, and will eventually give us a combinatorial result that will become of relevance in the later chapter of Uniform Spanning Trees.

Proposition 4.16 (Partition Function) The partition function of the Gaussian Free Field, i.e: the

density evaluated at zero, is equal to

$$\frac{1}{(2\pi)^{n/2}}\prod_{k=1}^{n} G_{\{x_k,\cdots,x_n\}}(x_k,x_k)^{-1/2}$$

in particular, the product is independent of the labelling of D.

Proof. In order for us to evaluate the density of the GFF at zero, we are going to need to study the quantity

$$\mathbf{P}\left[\bigcap_{i=1}^{n}\{\Gamma(x_i)\in[0,\epsilon]\}\right]$$

as $\epsilon\to 0$. To do this we simply note that the above quantity can be expressed as

$$\prod_{i=1}^{n}\mathbf{P}[\Gamma(x_i)\in[0,\epsilon]\,|\,\{\Gamma(x_1)\in[0,\epsilon]\}\cap\cdots\cap\{\Gamma(x_{i-1})\in[0,\epsilon]\}]$$

and to compute this quantity, we will proceed by induction, exploring each vertex one by one. Firstly, since we know that $\{\Gamma(x)\}_{x\in D}$ is a Gaussian vector with covariance matrix equal to Green's function, we have that $\Gamma(x_1)\sim\mathcal{N}(0,G_D(x_1,x_1))$. So indeed $\mathbf{P}[\Gamma(x_1)\in[0,\epsilon]]=\mathbf{P}\left[\mathcal{N}(0,1)\in\left[0,\frac{\epsilon}{\sqrt{G_D(x_1,x_1)}}\right]\right]$. Next one considers the quantity

$$\mathbf{P}\big[\Gamma(x_2)\in[0,\epsilon]\big|\Gamma(x_1)=\gamma_1\in[0,\epsilon]\big]$$

We know from our discussion of the Spatial Markov Property, conditioned on the event above, $\Gamma(x_2)=H_{\gamma_1}(x_2)+N_2$ where $N_2\sim\mathcal{N}(0,G_{\{x_2,\cdots,x_n\}}(x_2,x_2))$. This is once again, because conditional on the value of the field on some subset of points (in this case, the subset of points would be x_1, and we condition the field to take the value γ_1 there), the field on the remaining part has the distribution of a Dirichlet Gaussian free field plus the Harmonic Extension of the function that takes the value γ_1 on x_1 and zero on the boundary. Thus in effect,

$$\mathbf{P}\big[\Gamma(x_2)\in[0,\epsilon]\big|\Gamma(x_1)=\gamma_1\in[0,\epsilon]\big]=\mathbf{P}\big[N_2\in\big[-H_{\gamma_1},-H_{\gamma_1}+\epsilon\big]\big].$$

The key to estimate this, is to note that H_{γ_1} being a harmonic extension of a function that takes the values γ_1 and zero on some boundary points, must not be any greater than ϵ (since $\gamma_1\le\epsilon$), so that $[-H_{\gamma_1},-H_{\gamma_1}+\epsilon]\subseteq[-\epsilon,\epsilon]$ (imagine a sliding window of width ϵ within the larger frame of $[-\epsilon,\epsilon]$). From this it is clear (say by looking at the shape of the bell curve) that

$$\mathbf{P}[N_2\in[0,\epsilon]]\le\mathbf{P}\big[\Gamma(x_2)\in[0,\epsilon]\big|\Gamma(x_1)=\gamma_1\in[0,\epsilon]\big]\le\mathbf{P}[N_2\in[-\epsilon/2,\epsilon/2]]$$

The upper bound can be in turn upper bounded by $\epsilon\times\frac{1}{\sqrt{2\pi G_{\{x_2,\cdots,x_n\}}(x_2,x_2)}}$, where the ϵ comes from

the width of the window, and the second terms comes from the highest value of the curve, namely the value at zero. The lower bounded can be lower bounded by $\epsilon \times \frac{\exp\left(-\frac{1}{2}\frac{\epsilon^2}{G_{\{x_2,\cdots,x_n\}}(x_2,x_2)}\right)}{\sqrt{2\pi G_{\{x_2,\cdots,x_n\}}(x_2,x_2)}}$, as this is once again, the width of the window multiplied by the smallest value of the density on that window. The key is to realise that since the numerator of this lower bound in fact goes to one as $\epsilon \to 0$, we have that

$$\mathbf{P}\Big[\Gamma(x_2)\in[0,\epsilon]\Big|\Gamma(x_1)=\gamma_1\in[0,\epsilon]\Big] \overset{\epsilon\to 0}{\sim} \frac{\epsilon}{\sqrt{2\pi G_{\{x_2,\cdots,x_n\}}(x_2,x_2)}}$$

A similar argument on the computation of $\mathbf{P}[\Gamma(x_1)\in[0,\epsilon]]$ and induction gives that

$$\mathbf{P}\left[\bigcap_{i=1}^{n}\{\Gamma(x_i)\in[0,\epsilon]\}\right] \overset{\epsilon\to 0}{\sim} \epsilon^n \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi G_{\{x_i,\cdots,x_n\}}(x_i,x_i)}}$$

This means that the density at zero is in fact the product on the right hand side. Since the density of the Gaussian Free Field is of course independent of the way we label the domain D, it follows that the product is independent of the ordering of the x_i's. ♡

Chapter 5

Uniform Spanning Trees

In this short chapter, we will touch on the topic of spanning trees, and show how one can sample uniformly among the set of spanning trees. This model is of a very different nature to those we have explored until now, but has many connections with other models.

Definition 5.1 (Spanning Tree) Let $G=(V,E)$ be a connected graph. A spanning tree T is a subgraph of G, that is spanning (meaning its vertex set is the entirety of V), and is a tree (meaning it has no cycles)

We will focus on the case where G is a subset of $\mathbf{Z}^d$. For convenience, we will once again work with the graph obtained by identifying the boundary ∂G with one single boundary point ∂. We will often refer to this point as the root. We will now describe an algorithm that produces a randomly chosen spanning tree T. The main result of this section will be that the law of T is actually uniform on the set of all spanning trees and in doing so, we will showcase a connection with the DGFF.

5.1 Wilson's Algorithm

We now describe Wilson's Algorithm. For this, we need to introduce the concept of a loop erasure of a path:

Definition 5.2 (Loop erasure) Let $Z=(x_0,\cdots,x_m)$ be a path in $\mathbf{Z}^d$, i.e: $Z\in(\mathbf{Z}^d)^{m+1}$, such that $\text{dist}(x_i,x_{i+1})=1$. We define the loop-erasure L of Z to be the path

$$L(Z)=(L_0,L_1,\cdots,L_\sigma)$$

where the points L_i are define inductively as follows: $L_0=x_0$, and then $r_i=\max\{r\le m : x_r=L_{i-1}\}$, and $L_i=x_{r_i+1}$ and $\sigma=\min\{i : L_i=x_m\}$.

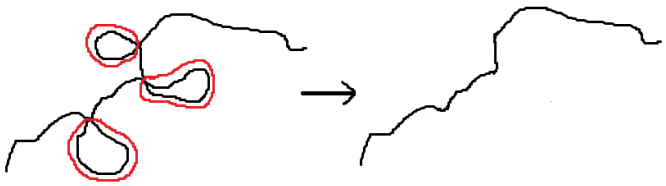


Figure 5.1: A loop erased path

Remark 5.3 (In plain English please) In plain English L_i is nothing but the point of the path Z that comes immediately after the time that the point L_{i-1} was visited for the very last time. There are many ways to perform loop erasure, but the one we have described above is the chronological loop erasure: you follow the path and whenever you encounter that you have finished making a loop, you erase the loop and keep moving.

We can now present the Algorithm.

1. Choose an ordering $\{a_1, \cdots, a_m\}$ of your graph G.
2. Define $x_1 = a_1$.
3. Start a simple random walk on G at x_1, and run it until it hits ∂. This gives a path $Z = \{x_1, x_2, \cdots, x_k, \partial\}$, perform loop erasure on Z, and call the resulting self avoiding walk S_1.
4. Now look at a_2. If $a_2 \in S_1$, then just set $S_2 = S_1$, then move on to a_3, otherwise, run a simple random walk on G started at a_2, and stop it until it hits S_1. Performing loop erasure on this new path gives a self-avoiding path S_2.
5. Generally, look at a_n, if $a_n \in \bigcup_{i=1}^{n-1} S_i$, then move on to a_{n+1}, otherwise run a simple random walk started at a_n until it hits $\bigcup_{i=1}^{n-1} S_i$. Perform loop erasure on the obtained path to get the self avoiding path S_n.

(I've been a bit careless here defining the indices that label each S_i, because the way I've written it I have made it so that there are $S_1, \cdots, S_m$. In reality, many of these will be the same set, so we will effectively have less, and we should use some more careful labelling to avoid this silly over counting, in the next proof, I will assume that I did the labelling properly, so that if out of $S_1, \cdots, S_m$ there are only j unique sets, then have the labels $S_1, \cdots, S_j$)

Remark 5.4 It is clear that once the algorithm has run through all $a_1, \cdots, a_m$, we obtain a subgraph $\bigcup_{i=1}^{m} S_m$ that is spanning, (indeed, $a_i \in S_i$), and is a tree. This is because of the loop erasure, which ensures there are no cycles, as well as the fact that we stop upon hitting the previous known structure.

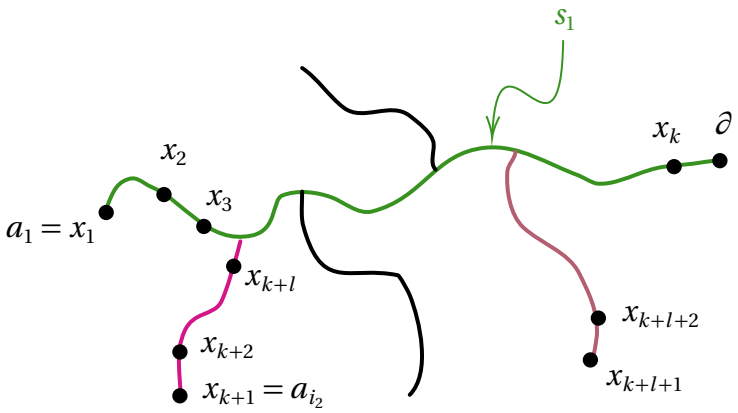


Figure 5.2: An illustration of the proof

Theorem 5.5 (Law of Wilson's Algorithm) Let T be the random tree obtained by performing Wilson's Algorithm on a graph G. Then $\mathscr{L}(T)$ is uniform among the set of all spanning trees on G.

Proof. For clarity, let $(\Omega, \mathscr{F}, \mathbf{P})$ be the probability space on which the random walks used in Wilson's Algorithm are defined. Let moreover t be a spanning tree for G. We are going to compute $\mathbf{P}[T = t]$. For this, assume that t is labelled as $t = \{a_1, a_2, \cdots, a_m\}$. Then we will start by doing as follows: let labelling $x_1 = a_1$, we define $s_1 = \{x_1, x_2, \cdots, x_k\}$ to be the branch of the tree t that connects a_1 to ∂. We then look amongst the list $t = \{a_1, a_2 \cdots, a_m\}$ and look for the next element in this list that is not in s_1. We may call for illustration purposes in order to match the figure a_{i_2}. Then once again, we look at the set of points $\{x_{k+1}, x_{k+2}, \cdots, x_{k+l}\}$ that define the branch s_2, that connects a_{i_2} to s_1. We now repeat this until we have found the branches $s_1, s_2, \cdots, s_j$. The key of doing this is that we have found exactly the self avoiding paths that Wilson's Algorithm should produce at each step in order for T to equal t. It is thus, that we see

$$\mathbf{P}[T = t] = \mathbf{P}\left[\bigcap_{i=1}^{j} \{S_i = s_i\}\right] = \prod_{i=1}^{j} \mathbf{P}\left[S_i = s_i \middle| \bigcap_{l=1}^{i-1} S_l = s_l\right].$$

Let us compute say $\mathbf{P}[S_1 = s_1]$. This corresponds to the probability that the random walk started at a_1 produces a walk whose loop erasure gives $s_1 = \{x_1, \cdots, x_k\}$. This can be calculated as follows. First the random walk, let's call it $X^{(1)}$ starts at a_1, it might then decide to do a few loops without exiting the graph (i.e: without hitting ∂) before jumping, with probability $1/\deg(a_1)$ (since we assume that this is $\mathbf{Z}^d$ we can just replace this by $1/2d$) on to x_2. Once at x_2, the walk might decide to make a few loops at x_2, but without hitting either ∂ **or** x_1, as we have assumed that $X^{(1)}$ did already all the loops at x_1. And so on. Since the walk is a simple random walk we can

multiply all the probabilities and see that

$$\begin{aligned}\mathbf{P}[S_1 = s_1] &= \left(\sum_{n\geq 0} P^n(x_1, x_1)\right) \times \frac{1}{2d} \times \cdots \\ &= G_D(x_1, x_1) \times \frac{1}{2d} \times G_{D\setminus\{x_1\}}(x_2, x_2) \times \frac{1}{2d} \cdots \\ &= \frac{1}{(2d)^k} \prod_{i=1}^{k} G_{D\setminus\{x_1,\cdots,x_{i-1}\}}(x_i, x_i)\end{aligned}$$

Then we can ask, what about the probability that $S_2 = s_2$ given that $S_1 = s_1$? Well its going to be essentially the same calculation, except that when we allow the walk to make loops, we will have to enforce that not only do they not exit D or hit the previously visited points, but we also have to constrain the walk to not hit s_1. From this we get that (using the Figure)

$$\mathbf{P}[S_2 = s_2 | S_1 = s_1] = G_{D\setminus\{x_1,\cdots,x_k\}}(x_{k+1}, x_{k+1}) \times \frac{1}{2d} \times \cdots$$

combining all this inductively, we get that (recall that $t = \{a_1, \cdots, a_m\}$):

$$\mathbf{P}[T = t] = \frac{1}{(2d)^m} \prod_{j=1}^{m} G_{D\setminus\{x_1,\cdots,x_{j-1}\}}(x_j, x_j)$$

Now, as we saw in the discussion of the density of the Gaussian Free Field at zero, we have that this product is independent of the ordering of the x_i's, from which we get that this is a number π, and so

$$\mathbf{P}[T = t] = \frac{\pi}{(2d)^m}$$

and so the Law is uniform. ♡

Chapter 6

A glimpse into the continuum

6.1 A *conformal Haar* measure on self-avoiding planar loops

In this section, we will briefly discuss in an informal manner how "conformal invariance" can be used to characterise the law of random curves, using the example of self-avoiding loops. This is the starting ingredient to a longer story, which we will not explain here, that allows one to connect the shape of the boundary of the clusters in critical percolation in the scaling limit to properties of planar Brownian motion. Let us denote $\mathfrak{L}$ for the set of self-avoiding loops that surround the origin in the (complex) plane. Since loops are nothing but compact sets, we can endow $\mathfrak{L}$ with the Hausdorff topology, where for two loops K and K', we had

$$d(K,K') = \max\left\{\max_{x\in K} d(x,K'), \max_{y\in K'} d(y,K)\right\}$$

and then turn this into a measurable space by considering the Borel sigma algebra generated by the open sets with respect to this topology. Now we can start talking about measures on $\mathfrak{L}$. The goal of this section will be to discuss how imposing that the measure satisfies a simple "conformal invariance" property leads to the fact that there is exactly one such measure. Let us indicate what this property is:

Definition 6.1 (Conformal restriction) A measure ν on $\mathfrak{L}$ satisfies conformal restriction if for any two conformally equivalent domains D and D' that contain the origin of the plain, and any conformal map $\Phi : D \to D'$ with $\Phi(0) = 0$, we have that

$$\Phi_*\left(\nu\big|_{\mathfrak{L}(D)}\right) = \nu\big|_{\mathfrak{L}(D')}$$

(where $f_*\mu = \mu \circ f^{-1}$)

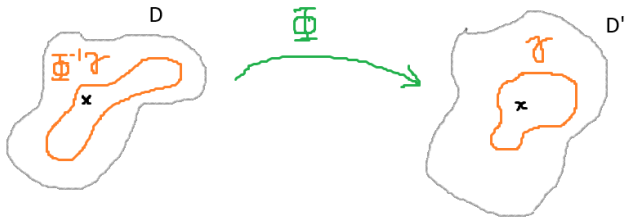


Figure 6.1: An illustration of the pushforward business

Remark 6.2 (Scale-invariance) Since we can in particular pick our domain D' to be a "blowup" of D, the conformal restriction property implies that ν is in fact scale-invariant, and so if our measure is non-trivial, meaning that there exists some $0<\delta<\Delta$ such that $\nu\{\gamma : \mathrm{diam}(\gamma)\in(\delta,\Delta)\}=c>0$, then we can simply apply scale invariance, and see that for any $k\in\mathbf{N}$, $\nu\{\gamma : \mathrm{diam}(\gamma)\in(k\delta,k\Delta)\}=c$, and so if we pick a sequence $\{r_k\}$ of stretching factors such that the annuli of inner radius $r_k\delta$ and outer radius $r_k\Delta$ are all disjoint, then we have found an infinite sequence of disjoint Borel sets, all with positive mass. Therefore ν must be of infinite mass.

As mentioned before, the goal of this section will be to outline the proof of the following Theorem:

Theorem 6.3 (Lawler, Werner) Up to multiplication by a positive constant, there exists a unique non-trivial measure ν on $\mathfrak{L}$ that satisfies conformal restriction.

Remark 6.4 (Heuristics of proof) The steps of the proof are the following:

1. Show that if a non-trivial measure ν on $\mathfrak{L}$ satisfies conformal restriction, then there exists $c>0$ such that for any bounded simply connected $D'\subseteq D$ on the plane that contain the origin, that are conformally equivalent via Φ, where $\Phi(0)=0$ and $\Phi'(0)\in\mathbf{R}_+$, then

$$\nu\{\gamma : \gamma\subseteq D, \gamma\not\subseteq D'\} = c\log\Phi'(0) \tag{6.1}$$

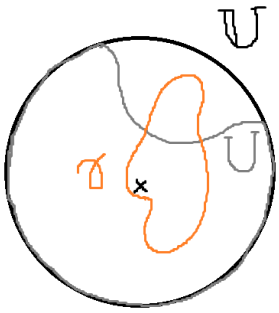
2. One then shows that for each $c>0$, there is at most one non-trivial measure on self-avoiding loops surrounding the origin that satisfies 6.1. This step is of a similar flavour as the "π-system uniqueness Lemma".

3. Finally, one shows that there is a way to construct a measure that satisfies conformal restriction by considering the outer boundaries of Brownian loops on the plain restricted to return to zero.

Step 1

Sketch of proof of step 1. Let ν be a non-trivial measure on $\mathfrak{L}$ that satisfies conformal restriction, to show that 6.1 holds, we first need to show an additivity result. First of all, by the conformal restriction property, we can just consider the case where $D = \mathbf{U}$ ($\mathbf{U}$ is defined to be the unit disk $\mathbf{U} = \{z \in \mathbf{C} : |z| \leq 1\}$), and D' is just some subset of $\mathbf{U}$. In addition, for a subset $U \subseteq \mathbf{U}$, we define $\Phi_U : U \to \mathbf{U}$ to be the conformal map with $\Phi_U(0) = 0$, $\Phi_U'(0) > 0$, and as $U \subseteq \mathbf{U}$, we have that $\Phi_U'(0) \geq 1$. Finally we define

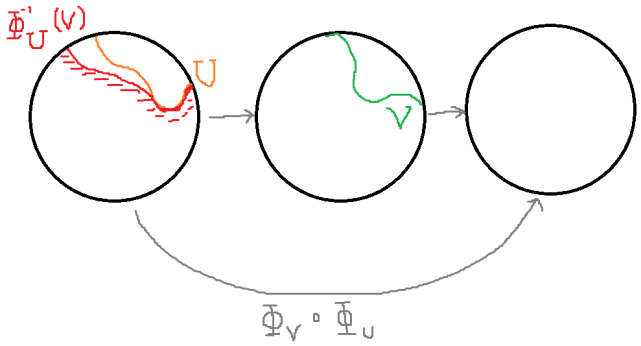
$$A(\Phi_U) = \nu\{\gamma : \gamma \subseteq \mathbf{U}, \gamma \not\subseteq U\}$$



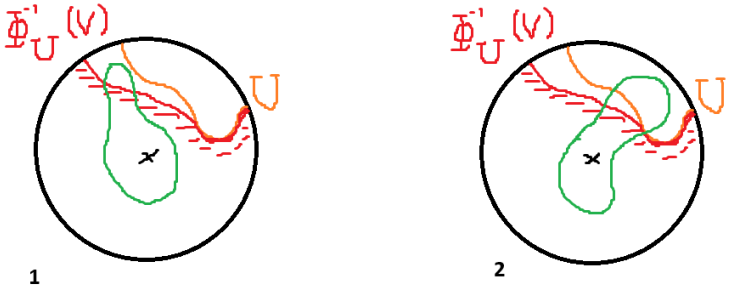
The first claim is that

$$A(\Phi_V \circ \Phi_U) = A(\Phi_V) + A(\Phi_U) \tag{6.2}$$

To make sense of this expression, one quickly notes that $\Phi_V \circ \Phi_U = \Phi_{(\Phi_V \circ \Phi_U)^{-1}(\mathbf{U})}$, indeed, one can see this from the following diagram



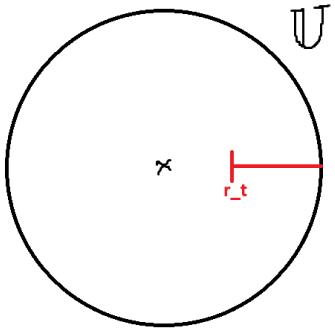
If we start on the right hand side with $\mathbf{U}$, we can go backwards one step to V by applying Φ_V^{-1}, and then we can go one step further and apply Φ_U^{-1} to ν and we obtain the red region on the left hand side. By reading this in the forward direction now, we see that indeed, the conformal map that sends $\Phi_U^{-1}\left(\Phi_V^{-1}(\mathbf{U})\right)$ is the composition of Φ_V with Φ_U. Therefore we can make sense of the left-hand side of equation 6.2. In fact 6.2 follows quite easily now from the diagram and conformal restriction: if a loop η containing the origin stays in $\mathbf{U}$ but leaves $\Phi_U^{-1}(V)$, then it is exactly of one of the following classes (1): $\eta \in \{\gamma : \gamma \subseteq U, \gamma \not\subseteq \Phi_U^{-1}(V)\}$ or (2) $\eta \in \{\gamma : \gamma \subseteq \mathbf{U}, \gamma \not\subseteq U\}$. Visually:



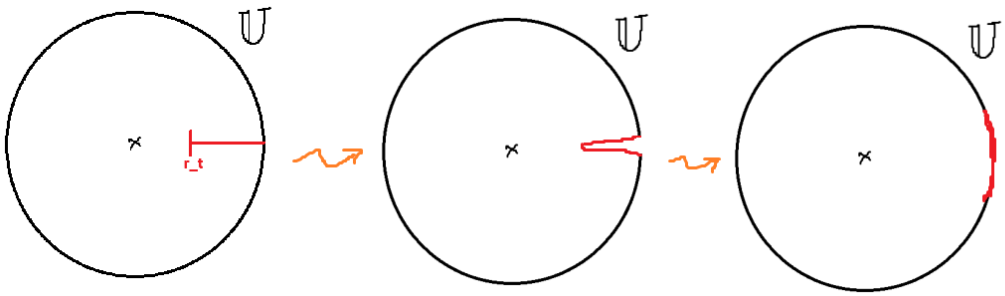
The key is that these two sets of loops are disjoint by definition, so we have that

$$A(\Phi_V \circ \Phi_U) = A(\Phi_U) + \nu\{\gamma : \gamma \subseteq U, \gamma \not\subseteq \Phi_U^{-1}(V)\}$$

But of course, we can just apply our conformal restriction property to this second term with the conformal map Φ_U, and conclude that in fact this second term is equal to $A(\Phi_V)$ and so 6.2 holds. The goal is to now deduce 6.1 from this additivity property. The first step is to consider a special family $(U_t)_{t\geq 0}$ of domains contained in $\mathbf{U}$: for each positive t, we define $U_t = \mathbf{U} \setminus [r_t, 1]$, which corresponds to nothing but $\mathbf{U}$ with a tiny slit cut as in the following diagram



it can be shown (hence why this is a proof sketch) that r_t can be chosen in a way that $\Phi'_{U_t}(0) = \exp(t)$. Heuristically, it makes sense that this can be done, because the derivative at zero in some sense indicates how much the domain U_t must be stretched to be deformed into $\mathbf{U}$, in fact one may visualise this conformal map as the following stretching procedure:

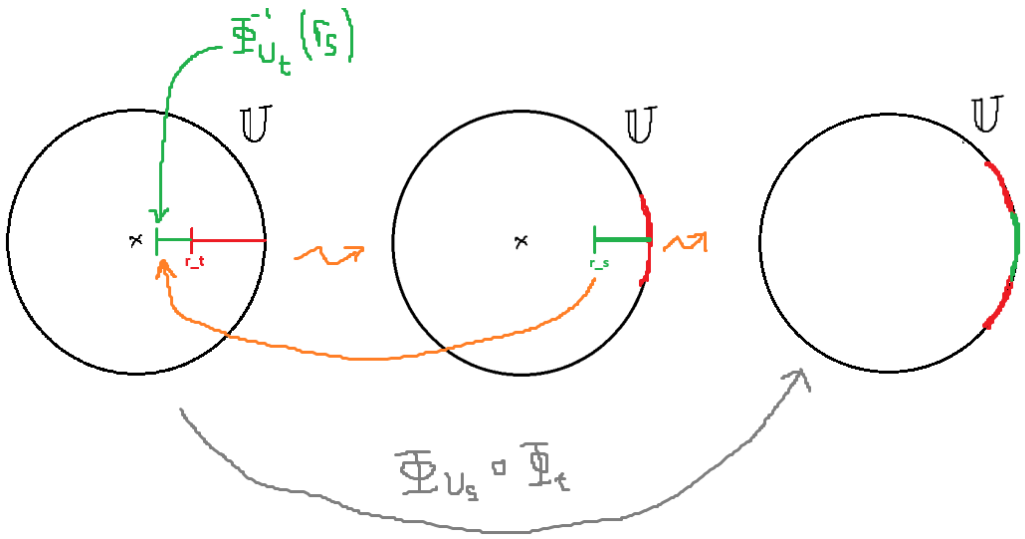


Now we note the following two important observations:

1. The family of conformal maps $\left(\Phi_{U_t}\right)_{t\geq 0}$ is in fact a semi-group, meaning that

$$\Phi_{U_t}\circ\Phi_{U_s}=\Phi_{U_r}$$

for some r. This can be seen by the following diagrams:



if we start with U_t and apply Φ_{U_t} we are sent to $\mathbf{U}$ middle one, and if on this new $\mathbf{U}$ we further slice a slit of width r_s, and then apply Φ_{U_s} we get sent further to $\mathbf{U}$ (the one on the right), in particular, since the r_s in the middle picture is on the real axis, we can apply $\Phi^{-1}_{U_t}$ to it, and we will go back to the left-most picture and obtain some point $\Phi^{-1}_{U_t}(r_s)$ on the real axis, from this we see that the composition $\Phi_{U_s}\circ\Phi_{U_t}$ does indeed correspond to Φ_{U_r} for some $r=\Phi^{-1}_{U_t}(r_s)$.

2. A priori, we don't know what the value of r we have just introduced actually is, but we have the important fact (from Complex Analysis) that the derivative of the composition of two conformal maps corresponds to the multiplication of their derivatives, so that in fact $\left(\Phi_{U_r}\right)'(0)=\Phi'_{U_t}(0)\Phi'_{U_s}(0)=\exp(t+s)$. From this, we deduce that in fact $r=t+s$, so in particular, we have that $\Phi_{U_s}\circ\Phi_{U_t}=\Phi_{U_{s+t}}$. Combining this with the additive equation 6.2, one reaches the conclusion that

$$A\left(\Phi_{U_{t+s}}\right)=A\left(\Phi_{U_t}\right)+A\left(\Phi_{U_s}\right).$$

3. Another observation is that $A\left(\Phi_{U_t}\right)$ is in fact non-decreasing. This is because if t increases, the slit creeps further into the disk, and so there are more possible loops η that cross the slit (recall that crossing the slit indeed means that η stays in $\mathbf{U}$ but leaves U_t).

4. The final observation is that $A(\Phi_{U_0})$, which is equal to $A(\mathrm{Id})=0$, because there are no loops that stay within $\mathbf{U}$ but exit $\mathbf{U}$.

Let us summarise our observations. We have found a function of t, $A_t = A(\Phi_{U_t})$ that has the following properties:

1. $A(\Phi_{U_t+s}) = A(\Phi_{U_t}) + A(\Phi_{U_s})$, i.e: $A_{t+s} = A_t + A_s$
2. $t \mapsto A_t$ is non-decreasing,
3. $A_0 = 0$.

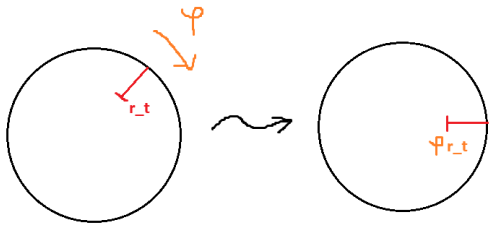
The only functions that satisfy this are indeed linear functions, so that $A_t = c\,t$ for some c. Now since $\Phi'_{U_t}(0) = \exp(t)$, rearranging gives that in fact

$$A(\Phi_{U_t}) = c \log \Phi'_{U_t}(0)$$

Which after plugging back into our definition of A, gives us that

$$\nu\{\gamma : \gamma \subseteq \mathbf{U}, \gamma \not\subseteq U_t\} = c \log \Phi'_{U_t}(0) \tag{6.3}$$

as promised. Of course we have just shown that formula 6.1 holds for these kinds of domains $(U_t)_{t\geq 0}$, we must now attempt to explain why this generalises to all kinds of domains $U \subseteq \mathbf{U}$. A first observation to make is that since ν satisfies conformal restriction, equation 6.3 immediately implies, (by using the conformal map $\varphi(z) = \exp(-i\theta z)$), that formula 6.1 also holds for domains U_t that are not just a straight cut along the real axis, but also a straight cut from any point on the circumference of the disk inwards:



Formally, if $\widetilde{U}_t$ is such a "tilted cut", then by conformal restriction

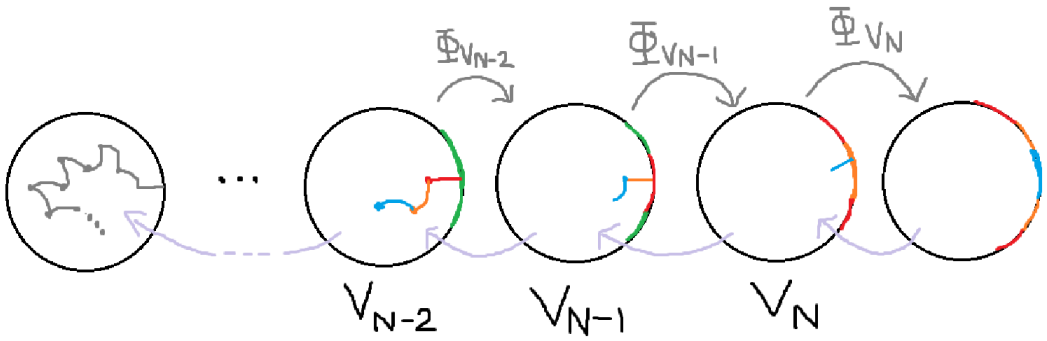
$$\begin{aligned}\nu\{\gamma \subseteq \mathbf{U}, \gamma \not\subseteq \widetilde{U}_t\} &= \nu\big(\varphi\{\gamma \subseteq \mathbf{U}, \gamma \not\subseteq \widetilde{U}_t\}\big) \\ &= \nu\{\gamma \subseteq \mathbf{U}, \gamma \not\subseteq U_t\} = c \log \Phi'_{U_t}(0) = c \log \Phi'_{\widetilde{U}_t}(0)\end{aligned}$$

where the very last step is because $\Phi_{\widetilde{U}_t} = \exp(i\theta) \circ \Phi_{I_t} \circ \exp(-i\theta)$, and since the derivative of compositions is the multiplication of the derivatives, and the derivative of $\exp(i\theta) = i\theta$, we have indeed that $\Phi'_{\widetilde{U}_t} = \Phi'_{U_t}$. For more clarity, let us call the domains $\widetilde{U}_t$ we have just described by U_t^θ. Now,

if we let $\mathscr{S}$ be the semi-group of conformal maps generated by the maps $\{\Phi_{U_t^\theta} : t > 0, \theta \in [0, 2\pi)\}$ and we let $\mathscr{U}$ be the corresponding family of domains, the idea is to argue that $\mathscr{S}$ is "sufficiently dense" in the class of all conformal maps Φ_U from simply connected domains U containing the origin to the unit disk $\mathbf{U}$, so that by a limiting argument, we see that 6.1 indeed holds for all such domains $U \subseteq \mathbf{U}$. The formal proof will be ommited, but follows from the classical theory developed by Loewner in the early 20th century. The main idea is that we are going to be able to well-approximate any such domain by a series of compositions of maps in $\mathscr{S}$. Notice that the formula 6.1 also holds for composition of such maps because of the following business:

$$\begin{aligned} A\left(\Phi_{U_{t_1}^{\theta_1}} \circ \Phi_{U_{t_1}^{\theta_1}}\right) &= A\left(\Phi_{U_{t_1}^{\theta_1}}\right) + A\left(\Phi_{U_{t_2}^{\theta_2}}\right) \\ &= c \log\left(\Phi'_{U_{t_1}^{\theta_1}}(0) \times \Phi'_{U_{t_2}^{\theta_2}}(0)\right) \\ &= c \log\left(\Phi_{U_{t_1}^{\theta_1}} \circ \Phi_{U_{t_2}^{\theta_2}}\right)'(0) \end{aligned}$$

where the last equality is once again because the derivative of a composition is the product of the derivatives. Now we sketch how a domain could be approximated by such compositions:



informally speaking, suppose one starts with several disks $V_1, \cdots, V_N$ (three of them have been depicted in the middle) and makes one "tilted cut" on each disk. We have the conformal maps $\Phi_{V_1}, \cdots, \Phi_{V_N}$ that send each V_i onto $\mathbf{U}$ in our "conformal manner". What we can consider is composing all of these inverse maps, so for example, if we start with (look at the diagram) $\Phi^{-1}_{V_N}(\mathbf{U})$ we get the picture with one blue cut in the diagram. We can then apply the next map and consider $\Phi^{-1}_{V_{N-1}} \circ \Phi^{-1}_{V_N}(\mathbf{U})$ and we would get the picture with the straight orange cut, and a perhaps "twisted" blue line, we can keep going and we will eventually reach a final shape, which consists of many "twisted cuts". The idea of the so called *Loewner Chains*, is that in fact, by choosing the cuts on each V_i appropriately, i.e: for an appropriate lenght, and an appropriate tilted angle, we can in fact well-approximate any domain $U \subseteq \mathbf{U}$ that we are interested in. To be a bit more precise, we can state (excuse the terminology)

Morally Speaking: for any simply connected $U \subseteq \mathbf{U}$ containing the origin, there exists a sequence

of conformal maps $\{\Phi_n\}_{n\geq 0}$ in $\mathscr{S}$, such that the corresponding sequence of "composed preimages" $(D_n)_{n\geq 0}$, has that D_n is an increasing family of sets, $D_n \subseteq U$ and $\bigcup_n D_n = U$.

(the way I have written this here is not precisely a true statement, but we are only attempting to sketch the main ideas). If one has this, then in the limit, since the sequence is increasing, we have that

$$A(\Phi_U) = \lim_{n\to\infty} A(\Phi_{D_n}) = c \lim_{n\to\infty} \log \Phi'_{D_n}(0) = c \log \Phi'_U(0)$$

(We have also not justified this last equality) and so, hopefully the reader is convinced that Step 1 is "proven". ♡

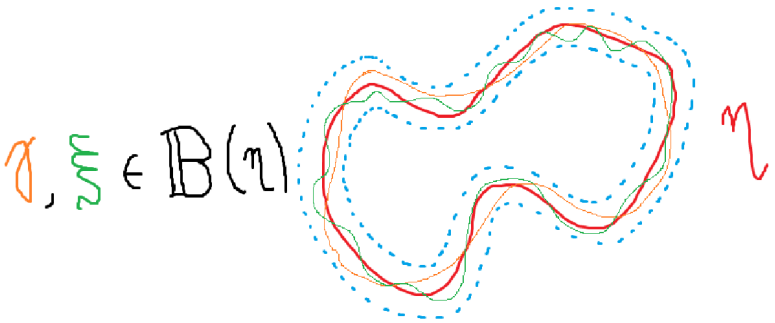
Step 2

Let us recall that now the goal is to show that any two measures on $\mathfrak{L}$ that agree on sets of the shape $\{\gamma : \gamma \subseteq \mathbf{U}, \gamma \not\subseteq U\}$ will in fact be equal on the whole of $\mathfrak{L}$. If we combine this with step one, it will follow that there is at most one measure that satisfies conformal restriction, since we have shown that any measure that satisfies conformal restriction must have values uniquely determined on these kinds of sets. (Up to choice of a scaling constant) This part will involve some simple measure-theoretic considerations, but most technicalities will be swept under the rug.

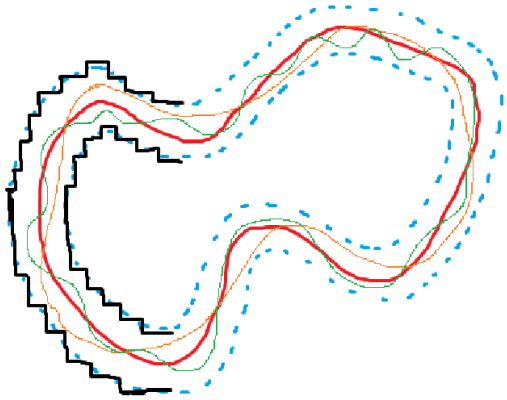
Sketch of proof of step 2. The main idea of this step is to show that if for two measures μ and ν, we have that for all domains $U' \subseteq U$, (for clarity, recall that bounded and simply connected is baked into the definition of domain) the masses

$$\mu\{\gamma \subseteq U, \gamma \not\subseteq U'\} = \nu\{\gamma \subseteq U, \gamma \not\subseteq U'\}$$

then the measures agree everywhere on $\mathfrak{L}$. We will do so by showing that this property actually forces the measures μ and ν to agree on the generating sets of the Borel sigma-algebra which we endowed $\mathfrak{L}$ with. The first thing to note, i.e: we need to show they agree on the open ball about a path η. Recall that we are working with Hausdorff distance, so that the open ball of η looks something like this:



The idea now is that one may discretize this "blue boundary" and obtain something like"



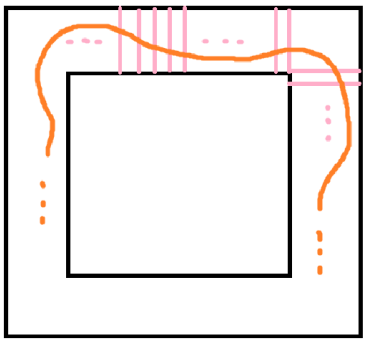
where the "discrete tube" is some set $A \subseteq 2^{-n}\mathbf{Z}^2$ for some n. We are then naturally interested in studying the measure of the sets $\mathcal{U}_A$ defined by

$$\mathcal{U}_A = \{\gamma : \gamma \text{ stays in the "annular region" } A\}.$$

The first thing to notice, is that the sets of the form $\mathcal{U}_A$ are indeed stable under intersections, the loops will just have to stay inside the intersection of the outer boundaries and the union of the inner boundaries. It could very well be that the intersection is the empty set. It is clear that any Hausdorff ball can be expressed as the countable intersection of these kinds of objects, and so it follows that the family of sets of the shape $\mathcal{U}_A$ (we now take n to be varying too) is a π-system generating the Borel sigma algebra. And so now we turn our attention to showing that the mass of the events $\mathcal{U}_A$ can indeed be determined using our events of the form

$$\{\gamma \subseteq U, \gamma \not\subseteq U'\}.$$

Then we will be done, modulo proving all of these claims rigorously. The idea is that for any set A like the ones above, i.e: $A \subseteq 2^{-n}\mathbf{Z}^2$ a "discrete tubular region", we can express $\gamma \in \mathcal{U}_A$ exactly as γ crossing all "slits" or crosscuts:

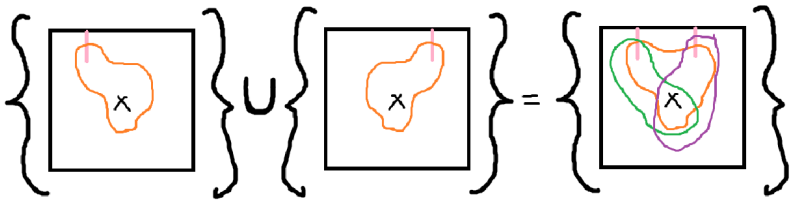


Let us now focus momentarily on the event that γ crosses one of these crosscuts, this is precisely of the shape that $\{\gamma \subseteq U, \gamma \not\subseteq U'\}$ for some U and U'!, in our particular diagram, U would be the

outer square, and U' would be the outer square with one such slit removed. The last problem however, is that the events $\{\gamma \subseteq U, \gamma \not\subseteq U'\}$ are sadly not closed under intersections, so even if we know the value of each event, we cannot know the value of the intersection, for it will not be in the shape of $\{\gamma \subseteq V, \gamma \not\subseteq V'\}$ for some V and V'. However, this is not the end of the world, because these events indeed are closed under countable unions, since

$$\{\gamma \subseteq U, \gamma \not\subseteq A\} \cup \{\gamma \subseteq U, \gamma \not\subseteq B\} = \{\gamma \subseteq U, \gamma \not\subseteq A \cup B\}$$

and we know the value of these events. Therefore by inclusion exclusion, we can know the value of intersections, and we are done!



♡

Remark 6.5 (Summary of Step 2) The summary of the steps were:

1. We can look at the Hausdorff Ball of a loop.
2. We can discretize the Hausdorff Ball.
3. Events of the type $\mathscr{U}_A$, i.e: loops that stay inside a tubular region A, are indeed a π-system generating the Borel sigma-algebra.
4. These events $\mathscr{U}_A$ can be expressed as the countable intersection of the event that the path hits a slit.
5. Using inclusion-exclusion, this intersection can be calculating by performing addition and subtraction, if we know the measure of the unions of events that the paths hit one of the slits.
6. Since events of the form $\{\gamma \subseteq U, \gamma \not\subseteq U'\}$ are indeed closed under unions, we can figure out the value of the measure of these unions and so work upwards again.

Step 3

So far, we have proven that there exists at most one measure ν on $\mathfrak{L}$ that satisfies the conformal restriction property. However, we haven't shown at all, that even such a measure should exist! For all we know, we could come up with a contradiction if we start exploring consequences of conformal restriction, but as it turns out, it is possible to construct one, and the main idea is to consider Brownian loops. The idea is to launch a Brownian Motion

6.2 The continuum Gaussian Free Field

In this last section, we will attempt to generalise the "random height function" we saw in our discussion of the discrete Gaussian Free Field to $\mathbf{R}^d$. Recall that in the discrete case, we ended up with the conclusion that the DGFF on a domain $D \subseteq \mathbf{Z}^d$ was the centered Gaussian vector whose covariance structure was given by the Harmonic function G_D. This gives a "seemingly straightforward" roadmap to define the GFF in the continuum, but a problem quickly arises! As we will discuss in a minute, if we want to discuss the Gaussian Free Field on $\mathbf{R}^d$, the only choice we'll have for the Green's function will not actually be well-defined on the diagonal, i.e: "$G_D(x,x) = +\infty$"! This unfortunately means that the naive approach to define $\{\Gamma(x)\}_{x\in\mathbf{R}^d}$ will not work. Intuitively, what happens is that the GFF will be a mess of $+\infty$ and $-\infty$ everywhere, and defining pointwise will prove impossible. The fix we will see, is that instead of "probing" the field at a point x and obtaining a random variable, we will probe the field by averaging its value with respect to certain measures. Thus in some sense, defining the GFF as what you would observe if you averaged out the field around some region. What we will have, morally speaking, is a random function

$$\text{“}\Gamma(\mu) = \int_{\mathbf{R}^d} \Gamma(x)\mathrm{d}\mu(x)\text{”}.$$

That, will be the continuum Gaussian Free Field, not defined pointwise, defined in terms of its integrals.

Recap on stochastic processes

Definition 6.6 (Stochastic process) A stochastic process indexed by a set $\mathscr{A}$ is a collection of random variables $\{X_a : a \in \mathscr{A}\}$ defined on the same probability space.

Remark 6.7 (Law of the process) The law of the process is the corresponding measure on $\mathbf{R}^{\mathscr{A}}$ endowed with the product sigma-algebra, and it is uniquely determined by its finite-dimensional distributions, meaning that laws of the random vectors $(X_{a_1}, \cdots, A_{a_k})$ for each $a_1, \cdots, a_k \in \mathscr{A}$.

We also have an important "converse" to this fact.

Theorem 6.8 (Kolmogorov's Extension Theorem) Suppose T is some interval and $n \in \mathbf{N}$. Consider a family of compatible finite dimensional distributions, that is to say: for each finite sequence of index times $t_1, \cdots, t_k$, we have a measure $\nu_{t_1,\cdots,t_k}$ on $(\mathbf{R}^n)^k$ that satisfy:

1. For any permutation $\sigma \in \mathfrak{S}_k$ and measurable sets $F_i \subseteq \mathbf{R}^n$, we have that

$$\nu_{t_1,\cdots,t_k}(F_1 \times \cdots \times F_k) = \nu_{t_{\sigma(1)},\cdots,t_{\sigma(k)}}(F_{\sigma(1)} \times \cdots \times F_{\sigma(k)}).$$

2. For all measurable sets $F_1 \subseteq \mathbf{R}^n$ and any $m \in \mathbf{N}$

$$\nu_{t_1,\cdots,t_k,t_{k+1},\cdots t_m}(F_1 \times \cdots \times F_k \times \mathbf{R}^n \times \cdots \times \mathbf{R}^n) = \nu_{t_1,\cdots,t_k}(F_1 \times \cdots \times F_k)$$

Then there exists a probability space $(\Omega, \mathscr{F}, \mathbf{P})$ and a stochastic process $X = \{X_t : t \in T\}$ such that the finite dimensional distributions of X are the $\nu_{t_1,\cdots,t_k}$'s

We have stated the Theorem above for the index set being an interval, but it still holds for general index sets. Let us recall another concept, which we already implicitly used in defining the DGFF

Definition 6.9 (Gaussian process) A stochastic process $(X_a)_{a\in\mathscr{A}}$ is a (centred) Gaussian process if its finite dimensional distributions are those of (centred) Gaussian vectors, meaning, that for any $a_1, \cdots, a_k \in \mathscr{A}$, and any $\lambda_1, \cdots, \lambda_k$ real constants, the random variable

$$\sum_{i=1}^{k} \lambda_i X_{a_i}$$

is a (centred) Normal random variable.

Remark 6.10 (Covariance structure) The law of a centred Gaussian process $(X_a)_{a\in\mathscr{A}}$ is uniquely determined by its covariance structure $\sigma : \mathscr{A} \times \mathscr{A} \to \mathbf{R}$:

$$\sigma(x, y) = \mathbf{E}[X_x X_y]$$

Remark 6.11 (Constructing a Gaussian process) Combining the previous fact, Kolmogorov's Theorem, and the definition, it can be shown that if $\sigma : \mathscr{A} \times \mathscr{A} \to \mathbf{R}$ is a real-valued symmetric that is positive-definite, meaning that for all $a_1, \cdots, a_n \in \mathscr{A}$ and all $\lambda_1, \cdots, \lambda_n \in \mathbf{R}$ $\sum_{i,j\leq n} \lambda_i \lambda_j \sigma(a_i, a_j) \geq 0$, then it is possible to construct a probability space with a process $X = \{X_a : a \in \mathscr{A}\}$ that is a Gaussian process whose covariance function is σ.

Remark 6.12 (Remark on independence) As a reminder, if X is a Gaussian process, it can be easily checked, by using the characteristic function, that if A_1 and A_2 are two subsets of the index set $\mathscr{A}$ of the process, for which whenever a_1 and a_2 belong to A_1 and A_2 respectively, one has that $\mathbf{E}[X_{a_1} X_{a_2}] = 0$, it follows that the processes $(X_{a_1})_{a_1\in A_1}$ and $(X_{a_2})_{a_2\in A_2}$ are in fact independent.

Remark 6.13 (A remark on measurability) Kolmogorov's Extension Theorem can be alternatively viewed as the fact that on the space $(\mathbf{R}^n)^T$, we can find a measure μ on the product sigma algebra, for which the marginals of μ coincide with the finite-dimensional distributions we prescribed. However, a limitation of this, is that we "are stuck" with the product sigma algebra, and so we can only make sense of events of the form $\{X_{a_1} \in A_1, \cdots, X_{a_k} \in A_k\}$ for measurable sets $A_1, \cdots, A_k$, that is to say, we can only "observe" at most a countable collection of the random variables X_a.

Basics on the Green's Function in the continuum

As mentioned at the start of this chapter, we will attempt to define the GFF on the continuum by borrowing ideas from what we saw in the discrete case. In the discrete case, we saw that the GFF was nothing but the centred Gaussian vector whose covariance structure was given by Green's function. We also saw that the function $y \mapsto G_D(x,y)$ was harmonic at all $y \neq x$, and that $G_D(x,y) = 0$ on all $y \in \partial D$. As it turns out, there is one unique function (up to multiplicative factors) in $\mathbf{R}^d$ that satisfies this properties, so this will be our candidate for Green's function.

Definition 6.14 (Green's function on the continuum) For $\mathbf{R}^d$ with $d \geq 3$, Green's function is defined to be

$$(x,y) \mapsto \frac{c_d}{|x-y|^{d-2}}$$

Remark 6.15 It can be proved that this is indeed the only function that satisfies these properties on $\mathbf{R}^d$. If we want to consider only bounded simply connected domains D of $\mathbf{R}^d$, then the only function that satisfies harmonicity at $y \neq x$ and is zero in the boundary ∂D, would be

$$G_D(x,y) = \underbrace{\frac{c_d}{|x-y|^{d-2}}}_{H_x(y)} - \mathbf{E}_y[H_x(B_T)]$$

where B is Brownian Motion and T is the hitting time of the boundary of D.

In particular, we have that if $D = \mathbf{R}^3$, Green's function takes the familiar form $\frac{c_3}{|x-y|}$, (Newtonian potential!). Now we find ourselves with a problem, we cannot simply define our GFF Γ to have covariance structure $G_D(x,y)$ because $\mathbf{E}[\Gamma(x)^2] = +\infty$. However, the following formal manipulation guides us onto where to go next: suppose that Γ could indeed be well-defined using this procedure, then we could define

$$I = \int_{\mathbf{B}(0;1)} \Gamma(x)\mathrm{d}x$$

and look at $\mathbf{E}[I^2]$:

$$\begin{aligned}
\mathbf{E}[I^2] &= \mathbf{E}\left[\int \Gamma(x)\mathrm{d}x \int \Gamma(y)\mathrm{d}y\right] \\
&= \int\int \mathbf{E}[\Gamma(x)\Gamma(y)]\mathrm{d}x\mathrm{d}y \\
&= c_3 \int\int \frac{\mathrm{d}x\mathrm{d}y}{|x-y|}
\end{aligned}$$

where the second equality came from Fubini's Theorem. Now note that since we are in $\mathbf{R}^d$ $d \geq 3$, this last integral is indeed finite, so perhaps this formal manipulation gives us a hint as to what the

GFF should be. Perhaps we cannot define Γ pointwise, but we can define its integral with respect to the Lebesgue measure or with respect to some other measures (these kind of objects are known as distribution functions)

Construction of the GFF

In light of the above comments, we have the following

Definition 6.16 Let

$$\mathscr{M}_D^+ = \left\{ \mu \geq \text{ a measure on } \mathbf{R}^d \text{ with } \operatorname{supp}(\mu) \subseteq D, \text{ such that } \int\int G_D(x,y)\mathrm{d}\mu(x)\mathrm{d}\mu(y) < \infty \right\}$$

and then take

$$M = \left\{ \mu : \mu = \mu^+ - \mu^-, \text{ where } \mu^+, \mu^- \in \mathscr{M} \right\}$$

since we will illustrate most of our discussion on $\mathbf{R}^3$, we will sometimes ommit the D in the subscript. We are now ready to define the Gaussian Free Field

Definition 6.17 (Continuum Gaussian Free Field) We say that $\Gamma = (\Gamma(\mu))_{\mu \in \mathscr{M}_D}$ is the Gaussian Free Field on D if it is a centered Gaussian process with covariance function

$$\sigma(\mu, \nu) = \int\int G_D(x,y)\mathrm{d}\mu(x)\mathrm{d}\nu(y)$$

Remark 6.18 This definition makes sense, as it can be checked that this does indeed provide a valid covariance structure. It is clearly symmetric, and one can show that $\sigma(\mu,\mu) \geq 0$, from which one can further show that for any $\mu_1, \cdots, \mu_n$ and $\lambda_1, \cdots, \lambda_n$, we have that

$$\sum_{i,j \leq n} \lambda_i \lambda_j \sigma(\mu_i, \mu_j) \geq 0.$$

Indeed, this just follows by taking $\mu = \sum_i \lambda_i \mu_i$ and using the fact that $\sigma(\mu,\mu) \geq 0$.

Hence, we have defined the Gaussian free field, not directly on $\mathbf{R}^d$ as we would have hoped, but in light of the formal computation we did for illustrative purposes, we "defined the GFF for each value of μ as the integral of the GFF with respect to μ" (I appreciate that I am talking in circles here, but hopefully the reader sees where this is coming from). As an example of what this is saying, suppose ν is the uniform measure on some box, then what we have is a random variable $\Gamma(\nu)$, and each realisation of this random variable, is what you would get as the average of the Gaussian Free Field on that box for a given realisation of the Gaussian Free Field.

Spherical averages

We now focus on the case where $D = \mathbf{R}^3$ so that Green's function is $G(x,y) = c_3/|x-y|$. Let us denote by $\lambda_{z,r}$ the uniform Lebesgue measure on on the boundary of the ball $\mathbf{B}(z,r)$ then we call $\gamma(z,r) := \Gamma(\lambda_{z,r})$ the spherical average about z of radius r, which is a well-defined random variable since $\lambda_{z,r} \in \mathscr{M}$.

Remark 6.19 (Preliminary remarks) Recall the fact that the average value of the map $y \mapsto 1/|x-y|$ over the ball $\mathbf{B}(z_0,r)$ for any $r < |z_0 - x|$ is equal to $1/|x-z_0|$. This is just what $y \mapsto 1/|x-y|$ being harmonic in $\mathbf{R}^3 \setminus \{x\}$ means. In fact, one can also show, since the "blowing up" of the function is too slow in $d=3$, that one can extend the radius to be actually $r \leq |x-z_0|$.

Recall that $\gamma(z,r)$ should be thought as the average of a random realisation of the Gaussian free field over the boundary of the ball of radius r about z, this quantity being in itself a random variable. We can try to explore some facts about these spherical averages about a point z_0 now. In particular, we will be interested in the behavior of the value of these spherical averages as the radius decreases. The first thing we will showcase, is that if we fix a radius r_0, the process of averages $(\gamma(z_0,r) - \gamma(z_0,r_0))_{r\in(0,r_0]}$ is indeed independent of the average of the field on some region different to the ball of radius r_0. More precisely:

Proposition 6.20 Let $\mu \in \mathscr{M}$ be a measure with $\mathrm{supp}(\mu) \cap \bar{\mathbf{B}}(z_0,r_0) = \varnothing$. Then the process $(\gamma(z_0,r) - \gamma(z_0,r_0))_{r\in(0,r_0]}$ is independent to $\Gamma(\mu)$.

Proof. By the harmonicity of Green's function:

$$\begin{aligned}\mathbf{E}\big[\gamma(z_0,r)\Gamma(\mu)\big] &= \int\int \mathrm{d}\mu(x) G(x,y)\mathrm{d}\lambda_{z_0,r}(y) \\ &= \int \mathrm{d}\mu(x)\left(\int G(x,y)\mathrm{d}\lambda_{z_0,r}(y)\right) \\ &= \int \mathrm{d}\mu(x) G(x,z_0)\end{aligned}$$

we used in this last equality that the support of μ and $\lambda_{z_0,r}$ are disjoint, therefore by Remark 6.19, the average of $G(x,y)$ when taking y over the boundary of the sphere centered at z_0 does indeed give us $G(x,z_0)$ (if the supports overlapped, then we could lose this averaging property due to the singularity at x) and this last quantity is independent of r, therefore: $\mathbf{E}\big[\big(\gamma(z_0,r) - \gamma(z,r_0)\big)\Gamma(\mu)\big] = 0$. Since the processes are Gaussian then zero covariance implies independence. ♡

Remark 6.21 In particular, $(\gamma(z_0,r)-\gamma(z_0,r_0))_{r\in(0,r_0]}$ will be independent of the sigma-algebra generated by $\Gamma(\mu)$ for μ supported away from $\mathbf{B}(z_0,r_0)$.

This shows some kind of spatial Markov property, although we will return later to what we really mean by the Spatial Markov property. Next on the roadmap is to look at the process $(\gamma(z_0,r))$ itself, and see what we can say about it. It turns out that it is a familiar friend:

Proposition 6.22 Let $B_{z_0}(t)=\frac{1}{\sqrt{c_3}}\gamma(z_0,1/t)$ with $B_{z_0}(0):=0$. Then B_{z_0} is a one dimensional Brownian Motion.

Proof. Take $r\le r'$, then

$$\begin{aligned}\mathbf{E}[\gamma(z_0,r)\gamma(z_0,r')]&=\int \mathrm{d}\lambda_{z_0,r'}(x)\left(\int \mathrm{d}\lambda_{z_0,r}(y)G(x,y)\right)\\&=\int \mathrm{d}\lambda_{z_0,r'}(x)G(x,z_0)=\frac{c_3}{r'}\end{aligned}$$

the second equality is once again using Remark 6.19 and the fact that $r\le r'$ so that the averaging property does indeed hold. Then the last equality is due to the fact that for any $x\in\partial\mathbf{B}(z_0,r')$, $G(x,z_0)=c_3/r'$. What we learn from this is that $\mathbf{E}[B(z_0,t)B(z_0,s)]=t\wedge s$. From this it quickly follows that in fact $\mathbf{E}[(B_t-B_s)^2]=t-s$. Using the covariance property we can now check that the increments are uncorrelated, and so we have a Gaussian Process with stationarity of increments, and whose increments are independent. That is to say, $B_{z_0}(t)$ does indeed have the law of a Brownian Motion. Just as like in the proof of Wiener's Theorem, we will now use Kolmogorov's Continuity Criterion to justify that this process admits a modification that makes it continuous, hence showing that $B_{z_0}(t)$ (admits a version) that makes it a true Brownian Motion. Recall Kolmogorov's Continuity Criterion for the Gaussian case: if $(X_a)_{a\in A}$ is a Gaussian Process for which there exists positive $\epsilon>0$ and $C>0$ such that

$$\mathbf{E}[(X_a-X_{a'})^2]\le C|a-a'|^{\epsilon},$$

then X admits a continuous modification. Since we have shown that $\mathbf{E}[(B_t-B_s)^2]=t-s$, we are done. ♡

Remark 6.23 Similar considerations can be used to show that one may vary z as well, and still have a continuous modification, showing that there exists a continuous modification of the process $(z,r)\mapsto\gamma(z,r)$ on $\mathbf{R}^d\times(0,\infty)$

The Spatial Markov Property

Maybe I finish this one day when I have energy.

Index