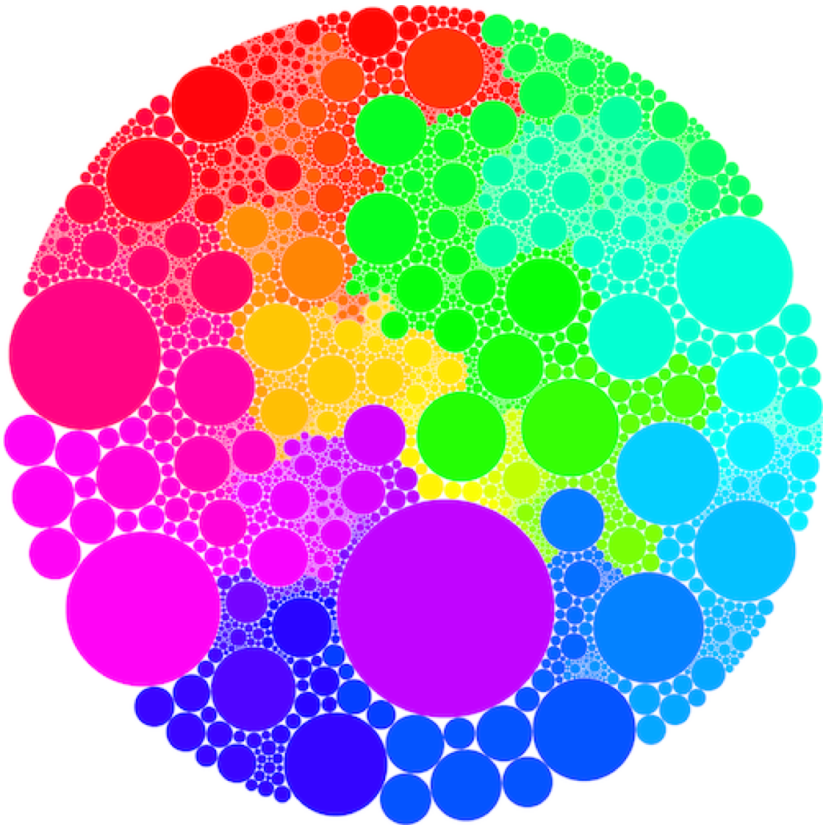
Gaussian free field, Liouville quantum gravity and Gaussian multiplicative chaos

Nathanaël Berestycki Ellen Powell

[Draft lecture notes: March 16, 2021]



Contents

Introduction

These lecture notes are intended to be an introduction to the two-dimensional continuum Gaussian free field, Liouville quantum gravity and the general theory of Gaussian multiplicative chaos. Topics covered include

- **Chapter 1:** the definition and main properties of the GFF;
- **Chapter 2:** the construction of the Liouville measure, its non degeneracy and conformal covariance, and applications to the KPZ formula;
- **Chapter 3:** a comprehensive exposition of the construction and properties of general Gaussian multiplicative chaos measures;
- **Chapter 4**: an introduction to random planar maps - the discrete counterparts of Liouville quantum gravity - and bijections with trees;
- **Chapter 5**: the Neumann GFF, the general notion of quantum surfaces, and in particular the so-called "thick quantum wedge";
- **Chapter 6:** Sheffield's quantum zipper theorem.

The final topic above is quite technical, and readers are advised that it will be of most use to people who are actively working in this area. (While the arguments in Chapter 6 follow the main ideas of [She16a], some are new or simplified, and the overall structure of the proof has been rethought. The result is, we hope, that some of the key ideas are more transparent and will be helpful to others.)

The theory is in full blossom and attempting to make a complete survey of the field would be hopeless, so quickly is it developing. Nevertheless, as the theory grows in complexity and applicability, it has appeared useful to summarise some of its basic and foundational aspects in one place, especially since complete proofs of some facts can be spread over a multitude of papers.

Clearly, the main drawback of this approach is that many of the important subsequent developments and alternative points of view are not included. For instance: the expansive body of work on random planar maps and their connection with Liouville quantum gravity, the Brownian map, imaginary geometry, and so on. Having said that, a future version of these notes will also touch upon the critical regime for Gaussian multiplicative chaos, the rigorous construction of Liouville conformal field theory, the peanosphere or "mating of trees" description of Liouville quantum gravity surfaces, and Liouville Brownian motion.

What's new?

Beside a streamlined and essentially self-contained presentation, several arguments in these notes can be considered novel.

- A new proof for the dimension of thick points of the GFF (especially the lower bound), see Exercise 4 in Chapter 2.
- The construction of Liouville measure which combines the general arguments of [Ber17] together with a few additional features of the GFF - giving an even simpler proof of convergence in this case.
- A reworking of Shamov's approach to GMC, see Section 3.3, in a unified language with the rest of the chapter.
- A somewhat simplified and explicit construction (inspired by [RV10a]) of scale-invariant log-correlated fields in all dimensions. This is used in particular to prove the existence of moments, positive and negative, in the most general set-up of Gaussian multiplicative chaos (with a general base measure).
- Computation of loop-erased random walk exponent using KPZ (thanks to Xin Sun for the suggestion to include this).
- A detailed definition of the Neumann GFF and thorough treatment of its analytic properties. Building on this, an explicit derivation of relationships between different variants of the GFF, and alternative Markov properties, see Section 5.4.
- A new proof based on Girsanov's theorem for one of the most delicate steps in the construction of the quantum zipper, see Lemma 6.20

Acknowledgements

An original version of these notes was written by the first author, in preparation for the LMS / Clay institute research school on *Modern Developments in Probability* taking place in Oxford, July 2015. The draft was revised in April 2016 on the occasion of the Spring School on *Geometric Models in Probability* in Darmstadt, then in July 2016 for the Probability Summer School at Northwestern (for which the chapter on statistical physics on random planar maps was added), and one more time in December 2017 for the Lectures on Probability and Statistics (LPS) at ISI Kolkata. In all cases we thank the organisers (respectively, Christina Goldschmidt and Dmitry Beliaev; Volker Betz and Matthias Meiners; Antonio Auffinger and Elton Hsu; Arijit Chakrabarty, Manjunath Krishnapur, Parthanil Roy and especially Rajat Subhra Hazra), for their invitations and superb organisation. Thanks also to Benoit Laslier for agreeing to run the exercise sessions accompanying the lectures at the initial school.

The Isaac Newton institute's semester on *Random Geometry*, was another important influence and motivation for these notes, and we would like to thank the INI for its hospitality. In fact, this semester served as the second author's initiation into the world of the Gaussian free field and Liouville quantum gravity. The resulting years of discussions together with the first author has led to the present expanded and revised version.

We would like to thank many of the INI programme participants for enlightening discussions related to aspects of the notes; especially, Omer Angel, Juhan Aru, Stéphane Benoît, Bertrand Duplantier, Ewain Gwynne, Nina Holden, Henry Jackson, Benoit Laslier, Jason Miller, James Norris, Gourab Ray, Scott Sheffield, Xin Sun, Wendelin Werner and Ofer Zeitouni. Special thanks in particular to Juhan Aru, Nina Holden and Xin Sun for many inspiring discussions over the years.

We would also like to thank the participants of two reading groups at the University of Bonn and ETH Zürich/University of Zürich respectively (particularly the organisers Nina Holden and Eveliina Peltola) which followed these notes, and led to many helpful comments.

Finally, comments on versions of this draft have been received at various stages from Jacopo Borga, Nina Holden, Henry Jackson, Benoit Laslier, Eveliina Peltola, Gourab Ray, Mark Sellke, Huy Tran, Fredrik Viklund, and Henrik Weber. We are grateful for their input which helped to correct minor problems, as well as to emphasise some of the subtle aspects of the arguments.

Thanks also to Jason Miller and to Jérémie Bettinelli and Benoit Laslier for letting us use some of their beautiful simulations which can be seen on the cover and in Chapter 4.

The work of the first author was supported during various stages of the writing by EPSRC (via grants EP/I03372X/1 and EP/L018896/1) and the FWF (via grant P33083), while the second author has been supported by the SNF grant 175505. This support is gratefully acknowledged.

1 Definition and properties of the GFF

1.1 Discrete case

⋆ *The discrete case is included here only for the purpose of guiding intuition when we come to work in the continuum.*

Consider a finite, weighted, undirected graph $\mathcal{G} = (V, E)$ (with weights $(w_e)_{e\in E}$ on the edges). For instance $\mathcal{G}$ could be a finite portion of the Euclidean lattice $\mathbb{Z}^d$ with weights $w_e \equiv 1$. Let ∂ be a distinguished set of vertices, called the boundary of the graph, and set $\hat{V} = V \setminus \partial$. Let $(X_t)_{t\geq 0}$ be the random walk on $\mathcal{G}$ in continuous time, meaning it jumps from x to y at rate $w_{x,y}$, and let τ be the first time that X hits ∂.

Write $Q = (q_{x,y})_{x,y\in V}$ for the Q-matrix of X. That is, its infinitesimal generator, so that for each $x \in V$, $q_{x,y} = w_{x,y}$ for $y \neq x$ and $q_{x,x} = -\sum_{y\sim x} w_{x,y} < \infty$. Note that the uniform measure $\pi(x) \equiv 1$ is reversible for X.

Definition 1.1 (Green function)**.** *The Green function $G(x, y)$ is defined for $x, y \in V$ by setting*

$$G(x, y) = \mathbb{E}_x \left(\int_0^\infty \mathbf{1}_{\{X_t = y; \tau > t\}} dt \right).$$

In other words $G(x, y)$ is the expected time that X spends at y, when started from x, before hitting the boundary. Note that with this definition we have $G(x, y) = G(y, x)$ for all $x, y \in \hat{V}$, since $\mathbb{P}_x(X_t = y;\, \tau > t) = \mathbb{P}_y(X_t = x;\, \tau > t)$ by reversibility of X with respect to π.

An equivalent expression for the Green function when working with the random walk in discrete time $Y = (Y_n)_{n\geq 0}$ (which jumps from x to y with probability proportional to $w_{x,y}$) is

$$G(x, y) = \frac{1}{q_y} \mathbb{E}_x \left(\sum_{n=0}^{\infty} \mathbf{1}_{\{Y_n = y; \tau > n\}} \right),$$

where $q_y = \sum_{y\sim x} w_{x,y} = -q_{y,y}$. Indeed, X and Y can be coupled so that for each $y \in \hat{V}$ and each visit of Y to y, X stays at y for an exponentially distributed time with mean $1/q_y$.

The Green function is a basic ingredient in the definition of the Gaussian free field, so the following elementary properties will be important to us.

Proposition 1.2. *Let $\hat{Q}$ denote the restriction of Q to $\hat{V} \times \hat{V}$. Then*

1. *$(-\hat{Q})^{-1}(x, y) = G(x, y)$ for all $x, y \in \hat{V}$.*
2. *G is a symmetric non-negative semi-definite function. That is, one has $G(x, y) = G(y, x)$ for all $x, y \in V$, and if $(\lambda_x)_{x\in V}$ is any vector of length $|V|$, then $\sum_{x,y\in V} \lambda_x \lambda_y G(x, y) \geq 0$. Equivalently, all of G's eigenvalues (when viewed as a matrix) are non-negative.*
3. *$G(x, \cdot)$ is discrete harmonic in $\hat{V} \setminus \{x\}$; more precisely $QG(x, \cdot) = -\delta_x(\cdot)$ for all $x \in \hat{V}$.*

Here, we use the natural notation $Qf(x) = \sum_{y\sim x} q_{xy}(f(y) - f(x))$ for the action of the generator Q on functions. Viewed as an operator in this way, Q is often referred to as the discrete Laplacian in continuous time.

Proof. Note that if $\hat{P}^t(x,y) = \mathbb{P}_x(X_t = y, \tau > t)$ then by the backward Kolmogorov equation, $(d/dt)\hat{P}^t(x,y) = \hat{Q}P^t(x,y)$, so that $\hat{P}^t(x,y) = e^{\hat{Q}t}(x,y)$. It then follows, by Fubini, that

$$\begin{aligned} G(x,y) &= \mathbb{E}_x(\int_0^\infty \mathbf{1}_{\{X_t=y;\tau>t\}}dt) \\ &= \int_0^\infty \hat{P}^t(x,y)dt \\ &= \int_0^\infty e^{\hat{Q}t}(x,y)dt \\ &= (-\hat{Q})^{-1}(x,y). \end{aligned} \tag{1.1}$$

The justification for the last equality comes from thinking about the action of the operator $\int_0^\infty e^{\hat{Q}t}$ on a single eigenfunction of $\hat{Q}$ and noting that all eigenvalues of $\hat{Q}$ are negative. This last fact is the analogue, and actually consequence, of the more well known statement that all eigenvalues of a discrete time Markov chain are less than 1. Also note that since $\hat{Q}$ is symmetric it is diagonalisable, so evaluating operators against eigenfunctions of $\hat{Q}$ is enough to identify them.

For the second point, we have already mentioned that $G(x,y) = G(y,x)$. Since $G(x,y) = 0$ whenever $y \in \partial$ it suffices to show that the restriction of G to $\hat{V}$ is positive definite. For this, we can use again that all the eigenvalues of $-\hat{Q}$, and hence of $(-\hat{Q})^{-1}$ are positive. This gives that G is positive definite when restricted to $\hat{V}$, by (1.1). (See (1.5) for a different proof in the continuous case which can also easily be adapted to this discrete setting).

Let us finally check the third point. This can be seen as a straightforward consequence of the first point, but we prefer to also include a probabilistic proof which is a simple application of the Markov property. Let $L(x) = \int_0^\infty \mathbf{1}_{\{X_t=x;\tau>t\}}dt$. If $y \neq x$ then we can write, by considering the first jump time J out of y (an exponential random variable with rate $q_y = -q_{y,y}$):

$$\begin{aligned} G(x,y) = G(y,x) &= \mathbb{E}_y L(x) \\ &= 0 \times \mathbb{P}(J > t) + \int_0^t q_y e^{-q_y s} ds \sum_z \frac{q_{y,z}}{q_y} \mathbb{E}_z(L(x)) \end{aligned}$$

so that, taking the derivative on both sides and again invoking symmetry:

$$0 = \sum_z q_{y,z} G(z,x) = \sum_z q_{y,z} G(x,z).$$

This means (for fixed x, viewing $G(x,y)$ as a function $g(y)$ of y only) that $Qg(y) = \sum_z q_{y,z} g(z) = 0$. Hence $G(x,\cdot)$ is harmonic in $\hat{V} \setminus \{x\}$.

When $y = x$ a similar argument can be made, but now the event $\{J > t\}$ contributes to $L(x)$, namely:

$$G(x,x) = \mathbb{P}(J > t)(t + G(x,x)) + \int_0^t q_x e^{-q_x s} ds \sum_{z \neq x} \frac{q_{x,z}}{q_x} \mathbb{E}_z(L(x))$$
$$= e^{-q_x t}(t + G(x,x)) + \int_0^t e^{-q_x s} \sum_{z \neq x} q_{x,z} G(x,z) ds.$$

Taking the derivative of both sides at $t = 0$ gives

$$0 = -q_x G(x,x) + 1 + \sum_{z \neq x} q_{x,z} G(x,z),$$

and hence

$$\sum_z q_{xz} G(x,z) = -1.$$

□

Remark 1.3. In the case where $\hat{V}$ has constant degree these facts are slightly easier to derive, since then (up to an unimportant scaling constant) we have $G(x,y) = \mathbb{E}_x(\sum_{n=0}^{\infty} \mathbf{1}_{\{Y_n = y; \tau > n\}})$.

For instance, (1.1) becomes simply the fact that

$$G(x,y) = \sum_{n=0}^{\infty} \hat{P}^n(x,y) = (I - \hat{P})^{-1}.$$

Definition 1.4 (Discrete GFF)**.** *The (zero boundary) discrete Gaussian free field on $\mathcal{G} = (V,E)$ is the centred Gaussian vector $(h(x))_{x \in V}$ with covariance given by the Green function G.*

Remark 1.5. This definition is justified. Indeed, suppose that $(C(x,y))_{x,y \in V}$ is a given function. Then there exists a centred Gaussian vector X having covariance matrix C if and only if C is symmetric and non-negative semi-definite (in the sense of property 2 above).

Note that if $x \in \partial$, then $G(x,y) = 0$ for all $y \in V$ and hence $h(x) = 0$ almost surely.

Usually for Gaussian fields, looking at the covariance structure is the most useful way gaining intuition. However in this case, the joint probability density function of the $|V|$ components of h is perhaps more illuminating.

Theorem 1.6 (Law of the GFF and Dirichlet energy)**.** *The law of the discrete GFF is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^V$. The joint density with respect to Lebesgue measure is proportional to*

$$\exp\left(-\frac{1}{4} \sum_{x,y \in V} q_{x,y}(h(x) - h(y))^2\right)$$

at any point $(h(x))_{x \in V}$ with $h(x) = 0$ for $x \in \partial$.

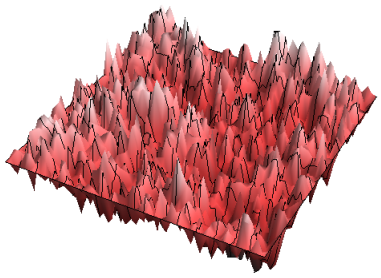


Figure 1: A discrete Gaussian free field

Remark 1.7. The way to read this formula is as follows: if $V = \{x_1, \dots, x_n\}$ and $Y_i := h(x_i)$, then $\mathbb{P}((Y_1, \dots, Y_n) \in A) = \int_A \frac{1}{Z} \exp(-\frac{1}{4}\sum(y_i - y_j)^2) \prod_i dy_i$ where the sum is over all i, j such that $x_i \sim x_j$ and $Z = \int_{\mathbb{R}^n} \exp(-\frac{1}{4}\sum(y_i - y_j)^2) \prod_i dy_i$. This holds for any Borel set A contained in the hyperplane $\{(y_1, \dots, y_n) : y_i = 0 \text{ for all } i \text{ s.t. } x_i \in \partial\}$.

For a given function $f : V \to \mathbb{R}$, the sum $\sum_{x\sim y} q_{x,y}(f(x) - f(y))^2$ is known as the **Dirichlet energy** of f, and is a discrete analogue of $\int_D |\nabla f|^2$.

Proof. This follows from the fact that for a centred Gaussian vector $(Y_1, \dots, Y_n)$ with invertible covariance matrix Σ, the joint probability density function on $\mathbb{R}^n$ is proportional to $f(y_1, \dots, y_n) = \exp(-\frac{1}{2}y^T\Sigma^{-1}y)$.

For us, the vertices $x \in V$ play the roles of the indices $1 \le i \le n$ above, and the values $h(x)$ for $x \in V$ play the roles of the y_i. To get a non-degenerate covariance matrix we restrict ourselves to vertices in $\hat{V}$, in which case G is invertible by Proposition 1.2. Note that since we are only considering h with $h(x) = 0$ for $x \in \partial$, it suffices to show that

$$-\frac{1}{2}h(\hat{\mathbf{x}})^T G^{-1} h(\hat{\mathbf{x}})) = -\frac{1}{4}\sum_{x,y\in V} q_{x,y}(h(x) - h(y))^2, \quad \text{for} \quad h(\hat{\mathbf{x}}) = (h(x))_{x\in\hat{V}}.$$

Recall that $(-\hat{Q})^{-1}(x, y) = G(x, y)$ for $x, y \in \hat{V}$, so that $G^{-1}(x, y) = -q_{xy}$. Hence

$$h(\hat{\mathbf{x}})^T G^{-1} h(\hat{\mathbf{x}}) = \sum_{x,y\in\hat{V}} G^{-1}(x, y)h(x)h(y) = \sum_{x,y\in\hat{V}} -q_{x,y}h(x)h(y).$$

Moreover, as we only consider h with $h(x) = 0$ for $x \in \partial$, this can be rewritten as

$$-\sum_{x,y\in V} q_{x,y}h(x)h(y) = \frac{1}{2}\sum_{x,y\in V} q_{x,y}(h(x) - h(y))^2 - \frac{1}{2}\sum_{x,y\in V} h(x)^2 q_{x,y} - \frac{1}{2}\sum_{x,y\in V} h(y)^2 q_{x,y},$$

where since $\sum_{y\in V} q_{x,y} = 0$ and $q_{x,y} = q_{y,x}$ for all x, y, the terms

$$\sum_{x,y\in V} h(x)^2 q_{x,y} \text{ and } \sum_{x,y\in V} h(y)^2 q_{x,y}$$

are both equal to 0. Note that in this final line of reasoning it is important to sum over all of V and not just $\hat{V}$. Thus

$$-\frac{1}{2}h(\mathbf{x})^T G^{-1} h(\hat{x}) = -\frac{1}{2} \times \frac{1}{2} \sum_{x,y \in V} q_{x,y}(h(x) - h(y))^2,$$

as required. □

Now, the Dirichlet energy of functions is minimised for harmonic functions. This means that the Gaussian free field is a "Gaussian perturbation of harmonic functions": as much as possible, it "tries" to be harmonic. In fact, this is a little ironic, given that in the continuum it is not even a function (see the next section).

This heuristic is at the heart of the Markov property, which is without a doubt the most useful property of the GFF. We state it here without proof, as we will soon prove its continuum counterpart (which is very similar).

Theorem 1.8 (Markov property of the discrete GFF)**.** *Fix $U \subset V$. The discrete GFF $h = (h(x))_{x \in V}$ can be decomposed as*

$$h = h_0 + \varphi,$$

where h_0 is Gaussian free field on U and φ is harmonic in U. Moreover, h_0 and φ are independent.

By a Gaussian free field in U we mean the GFF on the graph (V, E) but now with $\partial = V \setminus U$.

In words, this theorem says that conditionally on the values of h outside U, the field can be written as the sum of two independent terms. One of these is a zero boundary GFF in U, and the other is just the harmonic extension into U of the values of h outside U.

1.2 Continuous Green function

We will follow a route which is similar to the previous, discrete case. First we need to recall the definition of the Green function. We will only cover the basics here, and readers who want to know more are advised to consult, for instance, Lawler's book [Law05] which reviews important facts in a very accessible way. The presentation here will be somewhat different.

For $D \subset \mathbb{R}^d$ an open set, we define $p_t^D(x, y)$ to be the transition probability of Brownian motion killed when leaving D. In other words, $p_t^D(x, y) = p_t(x, y)\pi_t^D(x, y)$ where $p_t(x, y) = (2\pi t)^{-d/2} \exp(-|x - y|^2/2t)$ is the Gaussian transition probability, and $\pi_t^D(x, y)$ is the probability that a Brownian bridge from x to y of duration t remains in D.

Definition 1.9 (Continuous Green function)**.** *The Green function $G_0(x, y) = G_0^D(x, y)$ is defined by*

$$G_0(x, y) = \pi \int_0^\infty p_t^D(x, y) dt$$

for $x \neq y$ in D.

The factor π in front is purely so as to make some later computations in the two dimensional setting slightly easier. In particular, our normalisation is chosen so that when $D \subset \mathbb{C}$ we will have $G_0^D(x,y) \sim \log(|x-y|^{-1})$ as $y \to x$ (see Proposition 1.11), which is the standard set-up for the construction of Gaussian multiplicative chaos measures (see Chapters 2 and 3). Note that different authors use other conventions for this definition.

We will often drop the notational dependence of G_0^D on D when the domain is clear from the context. The subscript 0 refers to the fact that G has **zero boundary conditions**; equivalently, that G is defined from a Brownian motion killed when leaving D.

When $d \geq 2$, it is easy to see that $G_0(x,x)$ is typically ill-defined ($= \infty$) for all $x \in D$. This is since $\pi_t^D(x,x) \to 1$ as $t \to 0$ and so $(2\pi t)^{-d/2}\pi_t^D(x,x)$ cannot be integrable. However $G_0(x,y) < \infty$ as soon as $x \neq y$ and D is a **regular** domain (it is also necessary that $D \neq \mathbb{R}^d$ if $d \leq 2$). By regular, we mean that Brownian motion starting from a point $x \in \partial D$ will hit D^c instantaneously.In dimension $d = 1$, we will see that $G(x,y)$ is actually finite even when $x = y$.

Example. Suppose $D = \mathbb{H} \subset \mathbb{C}$ is the upper half plane. Then it is not hard to see that $p_t^{\mathbb{H}}(x,y) = p_t(x,y) - p_t(x,\bar{y})$ by a reflection argument. Hence one can deduce that

$$G_0^{\mathbb{H}}(x,y) = \log\left|\frac{x-\bar{y}}{x-y}\right| \tag{1.2}$$

for $x \neq y$ (see Exercise 1 for a hint on the proof).

In the special case $d = 2$, G inherits a conformal invariance property from the corresponding property of Brownian motion.

Proposition 1.10. *Let $D, D' \subset \mathbb{C}$ be simply connected domains. If $T : D \to D'$ is a conformal map (i.e., holomorphic and one-to-one), $G_0^{T(D)}(T(x), T(y)) = G_0^D(x,y)$.*

Proof. Essentially this is a change of variables. The Jacobian term $|T'(y)|^2$ arising from the change of variables just cancels the term $|T'(B_s)|^2$ arising from Itô's formula. More precisely, let ϕ be a test function and let $x' = T(x)$. Then

$$\int_{D'} G_0^{D'}(x',y')\phi(y')dy' = \pi\mathbb{E}_{x'}\left(\int_0^{\tau'} \phi(B'_{t'})dt'\right)$$

where B' is a Brownian motion and τ' is its exit time from D'. On the other hand, the change of variable formula applied to the left hand side gives us, letting $y' = T(y)$ and $dy' = |T'(y)|^2 dy$:

$$\int_{D'} G_0^{D'}(x',y')\phi(y')dy' = \int_D G_0^{D'}(T(x),T(y))\phi(T(y))|T'(y)|^2 dy.$$

Now we apply Itô's formula to the right hand side. This allows us to write $B'_{t'} = T(B_{F^{-1}(t)})$ where $F(t) = \int_0^t |T'(B_s)|^2 ds$ for $s \leq \tau$, τ is the first exit time of D by B, and F^{-1} is the

cadlag inverse of F. Moreover $\tau' = F^{-1}(\tau)$. Therefore,

$$\begin{aligned}\pi\mathbb{E}_{x'}\Big(\int_0^{\tau'} \phi(B'_{t'})dt'\Big) &= \pi\mathbb{E}_x\Big(\int_0^{F^{-1}(\tau)} \phi(T(B_{F^{-1}(t)}))dt\Big)\\ &= \pi\mathbb{E}_x\Big(\int_0^{\tau} \phi(T(B_s))F'(s)ds\Big)\\ &= \pi\mathbb{E}_x\Big(\int_0^{\tau} \phi(T(B_s))|T'(B_s)|^2 ds\Big)\\ &= \int_D G_0^D(x,y)\phi(T(y))|T'(y)|^2 dy.\end{aligned}$$

Hence identifying the left and right hand sides, since the test function ϕ is arbitrary, we conclude that $G_0^{D'}(T(x), T(\cdot)) = G_0^D(x,\cdot)$ as distributions. The result follows. □

Note that together with (1.2) and the Riemann mapping theorem, this allows us to determine G_0^D in any simply connected proper domain $D \subset \mathbb{C}$.

We state below some basic and fundamental properties of the Green function in two dimensions.

Proposition 1.11. *For any regular domain $D \subset \mathbb{R}^2$, and any $x \in D$:*

1. *$G_0^D(x,y) \to 0$ as $y \to y_0 \in \partial D$;*
2. *$G_0^D(x,\cdot)$ is harmonic in $D \setminus \{x\}$; and as a distribution $\Delta G_0^D(x,\cdot) = -2\pi\delta_x(\cdot)$;*
3. *$G_0^D(x,y) = -\log(|x-y|) + O(1)$ as $y \to x$.*

Proof. For the first point, observe that harmonic functions stay harmonic under composition with a conformal map, and the Green function in $\mathbb{H}$ is clearly harmonic off the diagonal (as the difference of two harmonic functions). For the second point, we can again use the explicit form of $G_0^{\mathbb{H}}$ on $\mathbb{H}$ and conformal invariance. □

Remark 1.12. In fact, the above result also holds in other dimensions with appropriate changes. One can check that in any dimension $d \geq 1$, for any domain D such that the Green function in D is finite, and for any fixed $x \in D$:

1. $G_0^D(x,y) \to 0$ as $y \to y_0 \in \partial D$;
2. $G_0^D(x,\cdot)$ is harmonic in $D \setminus \{x\}$ with $\Delta G_0^D(x,\cdot) = -2\pi\delta_x(\cdot)$ as distributions;
3. as $y \to x$,
$$G_0^D(x,y) = \begin{cases} G_0^D(x,x) + o(1) & d = 1 \\ -\log(|x-y|) + O(1) & d = 2 \\ \frac{2\pi}{A_d}|x-y|^{2-d} + O(1) & d \geq 3 \end{cases}$$
where A_d is the $(d-1)$-dimensional surface area of the unit ball in d dimensions.

In fact, the above properties characterise the Green function and thus may be used to identify it explicitly. See the exercises for this chapter, or [WP20, Lemma 3.7], for more on this.

Example. If $d = 1$ and $D = (0, 1)$ then

$$G_0^D(s, t) = 2\pi s(1 - t) \tag{1.3}$$

for $0 < s \leq t < 1$.

Actually, one can be a bit more precise about the behaviour of the Greens' function near the diagonal. That is, about the value of the error term $O(1)$ in Proposition 1.11.

For example, in two dimensions it holds that

$$G_0^D(x, y) = -\log(|x - y|) + \log R(x; D) + o(1) \tag{1.4}$$

as $y \to x$, where $R(x; D)$ is the **conformal radius** of x in D. That is, $R(x; D) = |f'(0)|$ for f any conformal map taking $\mathbb{D}$ to D and satisfying $f(0) = x$. (Note that this unambiguously defined). To see this, first note (using the explicit Möbius transform $\varphi(z) = (i - z)/(i + z)$ which maps the upper-half plane to the unit disc) that if $D = \mathbb{D}$ is the unit disc, we have

$$G_0^{\mathbb{D}}(0, z) = -\log|z|.$$

This makes (1.4) obvious for $D = \mathbb{D}$ and $x = 0$, and so (1.4) follows immediately in the general case by conformal invariance and definition of the conformal radius. Note that by the classical **Köbe quarter theorem**, we have

$$\operatorname{dist}(x, \partial D) \leq R(x; D) \leq 4 \operatorname{dist}(x, \partial D)$$

so the conformal radius is essentially a measure of the Euclidean distance to the boundary.

The conformal radius appears in Liouville quantum gravity in various formulae which will be discussed later on in the course. The reason it shows up in these formulae is usually because of (1.4).

The last property we need from the Green function, as in the discrete case, will be to say that $G_0(x, y)$ is a non-negative semi-definite function. We will see this in the next section.

1.3 GFF as a stochastic process

Essentially, as in the discrete case, we would like to define the GFF as a Gaussian "random function" with mean zero and covariance given by the Green function. However (when $d \geq 2$) the divergence of the Green function on the diagonal means that the GFF cannot be defined pointwise, as the variance at any point would have to be infinite. So instead, we define it as a random distribution, or generalised function in the sense of Schwartz[1]. More precisely, we

[1]This conflicts with the usage of distribution to mean the law of a random variable but is standard and should not cause confusion.

will take the point of view that it assigns values to certain measures with finite Green energy. In doing so we follow the point of view in the two sets of lecture notes [BN11] and [WP20]. The latter in particular contains a great deal more about the relationship between the GFF, SLE, Brownian loop soups and conformally invariant random processes in the plane, which will not be discussed in these notes. The foundational paper by Dubédat [Dub09] is also an excellent source of information regarding basic properties of the Gaussian free field. We should point out that the rest of this text is particularly focused on the case $d = 2$ but we will include results relevant to other dimensions if there is no cost in doing so.

Let $D \subset \mathbb{R}^d$ be a domain in which the Green function is finite – such a domain is called **Greenian**. As we have already mentioned, this is the case as soon as D is a proper domain of $\mathbb{R}^d$ (if $d \leq 2$) and if D is regular. Let $\mathcal{M}_0^+$ denote the set of (non-negative) measures with compact support in D such that $\int \rho(dx)\rho(dy)G_0^D(x,y) < \infty$. Note that when $d = 2$, due to the logarithmic divergence of the Green function on the diagonal, $\mathcal{M}_0^+$ includes the case where $\rho(x) = f(x)dx$ and f is continuous, but does not include Dirac point masses. Denote by $\mathcal{M}_0$ the set of signed measures that can be written in the form $\rho = \rho_+ - \rho_-$ for $\rho_\pm \in \mathcal{M}_0^+$. For test functions $\rho_1, \rho_2 \in \mathcal{M}_0$, we set

$$\Gamma_0(\rho_1, \rho_2) := \int_{D^2} G_0^D(x,y)\rho_1(dx)\rho_2(dy)$$

and also define $\Gamma_0(\rho) = \Gamma_0(\rho, \rho)$.

We now quickly recall some elementary facts from measure theory. Let I be an index set. A **stochastic process indexed by** I is just a collection of random variables $(X_i)_{i\in I}$, defined on some given probability space. The law of the process is a measure on $\mathbb{R}^I$, endowed with the product topology. It is uniquely characterised by its finite-dimensional marginals, via Kolmogorov's extension theorem. Note that in such a setting, it might not be possible to "simultaneously observe" more than a countable number of random variables. In order to do so we will have to rely on the existence of suitable versions with nice continuity properties (a version is another stochastic process indexed by the same set and with the same finite–dimensional marginals). The following is both an easy theorem and the definition of the GFF with zero boundary conditions on a domain.

Theorem 1.13 (Zero boundary or Dirichlet GFF)**.** *There exists a unique stochastic process $(h_\rho)_{\rho\in\mathcal{M}_0}$, indexed by $\mathcal{M}_0$, such that for every choice of $\rho_1, \dots, \rho_n$, $(h_{\rho_1}, \dots, h_{\rho_n})$ is a centred Gaussian vector with covariance structure $\mathrm{Cov}(h_{\rho_i}, h_{\rho_j}) = \Gamma_0(\rho_i, \rho_j)$.*

We will use the terminology "Dirichlet GFF", "zero boundary GFF" and "GFF with zero/Dirichlet boundary conditions" interchangeably throughout. This slightly abuses the notion of Dirichlet boundary conditions, which often means *any* specified boundary conditions, but it will be made clear in the sequel if we wish to talk about anything other than the zero case.

Definition 1.14. *The process $(h_\rho)_{\rho\in\mathcal{M}_0}$ is called the Gaussian free field in D (with Dirichlet or zero boundary conditions).*

Proof. We need to check several things:

- the finite-dimensional distributions exist and are uniquely specified;
- they are consistent.

The consistency is an immediate consequence of the Gaussianity of the finite dimensional marginals: indeed, the restriction of a Gaussian vector to a subset of coordinates is still a Gaussian vector with the same covariance structure.

For the first point, symmetry is a consequence of the reversibility of the heat kernel $p_t^D(x,y) = p_t^D(y,x)$. To check the non-negative semi-definite character of Γ, we need to check that for every $\rho_1, \dots, \rho_n \in \mathcal{M}_0$ and every $\lambda_1, \dots, \lambda_n \in \mathbb{R}$, we have that

$$\sum_{i,j} \lambda_i \lambda_j \Gamma_0(\rho_i, \rho_j) \geq 0.$$

However, by linearity of integration, this sum is nothing but $\Gamma_0(\rho)$ for $\rho = \sum_i \lambda_i \rho_i$. Hence it suffices to prove that

$$\Gamma_0(\rho, \rho) \geq 0 \text{ for } \rho \in \mathcal{M}_0. \tag{1.5}$$

A simple proof relies on the following observation. By the Markov property, we have

$$p_t^D(x,y) = \int_D p_{t/2}^D(x,z) p_{t/2}^D(z,y) dz,$$

and hence by symmetry again, we can deduce that

$$G_0(x,y) = 2 \int_D dz \int_0^\infty p_u^D(x,z) p_u^D(y,z) du.$$

Consequently,

$$\begin{aligned} \iint G_0(x,y) \rho(dx) \rho(dy) &= \int_D 2dz \int_0^\infty \iint \rho(dx) \rho(dy) p_u^D(x,z) p_u^D(y,z) du \\ &= \int_D 2dz \int_0^\infty \left(\int \rho(dx) p_u^D(x,z) \right)^2 du \geq 0. \end{aligned} \tag{1.6}$$

for any $\rho \in \mathcal{M}_0$. This proves (1.5) and so finishes the proof of the theorem. □

In the rest of this text, we will write (h, ρ) for h_ρ, and we will think of (h, ρ) as "h integrated against ρ".

Example. Suppose that $d = 1$ and $D = (0,1)$. Then by (1.3) we know that $G_D(x,y) = 2\pi x(1-y)$ for $0 < x \leq y < 1$, and this turns out to be the covariance of $\sqrt{2\pi}$ times a **Brownian bridge** $(b_s, 0 \leq s \leq 1)$ (see Chapter 1.3 of [RY99]). More precisely, since

$\mathbb{E}(b_s b_t) = s(1-t)$ for $0 \le s \le t \le 1$, we deduce that the Gaussian free field h in one dimension satisfies

$$h = \sqrt{2\pi} b,$$

in the sense of stochastic processes indexed by test functions, say.

Other boundary conditions than zero will also be relevant in practice. For this, we make the following definition (in the case $d = 2$ for simplicity). Suppose that f is a (possibly random) continuous function on the *conformal boundary* of the domain (equivalent to the Martin boundary of the domain for Brownian motion). Then the GFF with boundary data given by f is the random variable $h = h_0 + \varphi$, where h_0 is an independent Dirichlet GFF, and φ is the harmonic extension of f to D.

Remark 1.15. Note that (h, ρ) is linear in ρ: if $\rho, \rho' \in \mathcal{M}_0$, and if $\alpha, \beta \in \mathbb{R}$ $(h, \alpha\rho + \beta\rho') = \alpha(h, \rho) + \beta(h, \rho')$, almost surely (which can be seen by checking that the variance and mean of the difference are both zero). Hence an alternative definition of the GFF is simply as the unique stochastic process (h, ρ) which is linear in ρ and such that (h, ρ) is a centred Gaussian random variable with variance $\Gamma_0(\rho)$.

1.4 Integration by parts and Dirichlet energy

Let $\mathcal{D}_0(D)$ denote the set of compactly supported, C^∞ functions in D, also known as **test functions**. The set $\mathcal{D}_0(D)$ is equipped with a topology in which convergence is characterised as follows. A sequence $(f_n)_{n\ge 0}$ converges to 0 in $\mathcal{D}_0(D)$ if and only if there is a compact set $K \subset D$ such that $\mathrm{supp} f_n \subset K$ for all n and f_n and all its derivatives converge to 0 uniformly on K. A continuous linear map $u : \mathcal{D}_0(D) \to \mathbb{R}$ is called a **distribution** on D. Thus, the set of distributions on D is the dual space of $\mathcal{D}_0(D)$. It is denoted by $\mathcal{D}_0'(D)$ and is given the weak-$*$ topology. Thus $u_n \to u$ in $\mathcal{D}_0'(D)$ if and only if $u_n(\rho) \to u(\rho)$ for all $\rho \in \mathcal{D}_0(D)$.

The Gaussian free field can be understood as a random distribution. In order to do this, our first step is to relate the covariance of the GFF to the Dirichlet energy of a function (as in the discrete case). The following **Gauss–Green formula**, which is really just an integration by parts formula, will allow us to do so.

Lemma 1.16. *Suppose that D is C^1 smooth. If f, g are smooth functions on $\bar{D}$, then*

$$\int_D \nabla f \cdot \nabla g = -\int_D f \Delta g + \int_{\partial D} f \frac{\partial g}{\partial n} \tag{1.7}$$

(Note that this identity extends by definition to the case where say $f \in \mathcal{D}_0'(D)$ and $g \in \mathcal{D}_0(D)$; the boundary term on the right hand side is then equal to zero).

With this lemma in hand, we can now rewrite the variance $\Gamma_0(\rho, \rho)$ of (h, ρ) in terms of the Dirichlet energy of an appropriate function f.

Lemma 1.17. *Suppose $f \in \mathcal{D}_0(D)$ and that ρ is a smooth function such that $-\Delta f = 2\pi\rho$. Then $\rho \in \mathcal{M}_0$ and*

$$\Gamma_0(\rho, \rho) = \frac{1}{2\pi} \int_D |\nabla f|^2. \tag{1.8}$$

Proof. By the Gauss–Green formula (Lemma 1.16), noting that there are no boundary terms arising in each application,

$$\begin{aligned}
\Gamma_0(\rho) &= -\frac{1}{2\pi}\int_x \rho(x)\int_y G_0(x,y)\Delta_y f(y)dydx && \text{since } \rho(y) = -\frac{1}{2\pi}\Delta_y f(y),\\
&= -\frac{1}{2\pi}\int_x \rho(x)\int_y \Delta_y G_0(x,y) f(y)dydx && \text{by integration by parts (twice)},\\
&= \int_x \rho(x) f(x)dx && \text{since } \Delta G_0(x,\cdot) = -2\pi\delta_x(\cdot)\\
&= -\frac{1}{2\pi}\int_D (\Delta f(x)) f(x)dz\\
&= \frac{1}{2\pi}\int_D |\nabla f|^2 && \text{by the Gauss–Green formula again.}
\end{aligned}$$

□

Note that this gives another proof that $\Gamma_0(\rho,\rho) \geq 0$, and therefore that the GFF is well defined as a Gaussian stochastic process (at least when indexed by smooth functions ρ). Indeed, when ρ is smooth one can always find a smooth function f such that $-\Delta f = 2\pi\rho$: simply define

$$f(x) = \int G_0(x,y)\rho(y)dy. \tag{1.9}$$

In the general case where $\rho \in \mathcal{M}_0$ is not necessarily a smooth function, we can still deduce that $\Gamma_0(\rho) \geq 0$ by the following approximation argument: if we define $\rho_\varepsilon(x) = \int_D f_\varepsilon(x-x')\rho(dx')$ with f_ε a smooth approximation of identity (such as the heat kernel $f_\varepsilon(z) = p_\varepsilon(0,z)$), then $\Gamma_0(\rho_\varepsilon) \to \Gamma_0(\rho)$ as $\varepsilon \to 0$. This itself can be seen from uniform integrability assuming $\rho \in \mathcal{M}_0^+$, which is relatively easy to establish from the assumption.

1.5 Reminders about functions spaces

One drawback of defining the GFF as a stochastic process, is that we cannot realise (h,ρ) for all $\rho \in \mathcal{M}_0$ **simultaneously**. For example, it will not always be possible to define (h,ρ) when $\rho \in \mathcal{M}_0$ is random.

With this in mind, it is often useful to work with **versions** of the GFF that almost surely live in some "function" space. For example, it turns out to be possible to define a version of the GFF that is a random variable taking values in the space of distributions, or generalized functions. In fact, versions of the GFF taking values in much nicer Sobolev spaces (with negative index) can also be defined.

For completeness we include some brief reminders in this section.

Definition 1.18 (Dirichlet energy, Sobolev space H_0^1)**.** *For $f,g \in \mathcal{D}_0(D)$, introduce the Dirichlet inner product:*

$$(f,g)_\nabla := \frac{1}{2\pi}\int_D \nabla f \cdot \nabla g.$$

By definition, the Sobolev space $H^1_0(D)$ is the completion of $\mathcal{D}_0(D)$ with respect to this inner product; this is the Sobolev space *of index 1, and consists of all distributions which are $L^2(D)$ functions, have zero "trace" on the boundary and whose gradient is also an $L^2(D)$ function. See, for example, [Eva10, §6].*

Recall that if $D \subset \mathbb{R}^d$ is an open set, and if $\alpha = (\alpha_1, \dots, \alpha_d)$ is a multi-index, then $C^{|\alpha|}(D)$ is the set of functions for which

$$D_\alpha u = \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \dots \left(\frac{\partial}{\partial x_d}\right)^{\alpha_d} u \tag{1.10}$$

is well-defined as a function. An integrable function $v : D \to \mathbb{R}$ is said to be the αth weak derivative of u, or $D_\alpha u = v$, if

$$\int_D \varphi v = (-1)^{|\alpha|} \int_D u D_\alpha \varphi,$$

for all $\varphi \in C_0^{|\alpha|}(D)$. If u is in fact an element of $C^{|\alpha|}(D)$, then the weak derivative $D_\alpha u$ exists and is the classical derivative (1.10).

Eigenbasis of H^1_0. When D is bounded, it is easy to find a suitable orthonormal eigenbasis for $H^1_0(D)$. Let $(f_n)_{n\geq 1}$ be the eigenfunctions of $-\Delta$ on D, with zero or Dirichlet boundary conditions, and let λ_n denote the corresponding eigenvalues, so the f_n satisfy

$$\begin{cases} -\Delta f_n = \lambda_n f_n & \text{in } D \\ f_n = 0 & \text{on } \partial D \end{cases}$$

for each n. The $(\lambda_n)_{n\geq 1}$ are non-negative, ordered in non-decreasing order and $\lambda_n \to \infty$. See, e.g., [Cha84, Theorem III.9.11] for existence and other elementary facts about $(\lambda_n)_{n\geq 1}$. Moreover the Gauss–Green formula (1.7) implies that for $\lambda_n \neq \lambda_m$,

$$(f_n, f_m)_\nabla = \lambda_m \int_D f_n f_m = \lambda_n \int_D f_n f_m.$$

Hence $(f_n, f_m)_\nabla = 0$, and the eigenfunctions corresponding to different eigenvalues are orthogonal. Indeed, elementary spectral theory of operators (going slightly beyond the scope of these notes) tells us that $(f_n)_{n\geq 1}$, properly normalised, form an orthonormal basis of $H^1_0(D)$.

Often, the eigenfunctions of $-\Delta$ are normalised to have unit L^2 norm, since they also form an orthogonal basis of $L^2(D)$ for the standard L^2 inner product (again by the Gauss–Green formula). If $(e_j)_j$ are normalised in this way, then the above considerations imply that setting

$$f_j = (2\pi/\lambda_j)^{1/2} e_j \tag{1.11}$$

for each j, we get an orthonormal basis $(f_j)_j$ of $H^1_0(D)$.

Fractional powers of the Laplacian. The information in (1.11) can be usefully repackaged in terms of fractional powers of the operator $(-\Delta/2\pi)$. Namely, since applying $-\Delta$ to e_j gives $\lambda_j e_j$, we would like to think of f_j as $(-\Delta/2\pi)^{-1/2} e_j$ for each j.

More generally, let $s \in \mathbb{R}$ and for $f = \sum_{n\geq 1} \alpha_n e_n \in L^2$ let us define, formally for now,

$$(-\Delta)^s f = \sum_{n\geq 1} \alpha_n \lambda_n^s e_n. \tag{1.12}$$

We claim that if $f \in \mathcal{D}_0(D)$ then the series (1.12) converges for every $s \in \mathbb{R}$, for example, in the L^2 sense. Indeed this is clear for $s = 0$ and thus also for $s \leq 0$ since $\lambda_n \geq 1$ for n large enough. To deal with positive s, notice that Δf is also in $\mathcal{D}_0(D) \subset L^2(D)$, and so

$$\sum_{n\geq 1} (\Delta f, e_n)^2 < \infty,$$

where (f, g) without any index refers to the standard L^2 inner product. By the Gauss Green identity (1.7), this implies that $\sum_{n\geq 1} \lambda_n^2 \alpha_n^2 < \infty$ and thus the series (1.12) also converges in L^2 for $s = 2$. Since $\lambda_n \geq 1$ for n large enough, it must therefore converge for any $s \leq 2$. Iterating this argument, we find that (1.12) converges in L^2 for all $s \in \mathbb{R}$. Moreover, if $f \in \mathcal{D}_0(D)$, then $(-\Delta)^s(-\Delta)^{-s} f = (-\Delta)^{-s}(-\Delta)^s f = f$.

Sobolev spaces of general index $H_0^s(D), s \in \mathbb{R}$. We may thus define H_0^s to be the Hilbert space completion of $\mathcal{D}_0(D)$ with respect to the inner product

$$(f, g)_s = \frac{1}{(2\pi)^s}((-\Delta)^{s/2} f, (-\Delta)^{s/2} g)$$

where, again, (f, g) without any index refers to the standard L^2 product. (In the PDE literature, when $s \leq 0$, this space is often denoted as $H^s(D)$ instead of $H_0^s(D)$ – very roughly speaking, this is because the restriction to compactly supported test functions is not relevant in that case.) Thus for the distribution f defined by the series $f = \sum_{n\geq 1} \alpha_n e_n$, we have

$$f \in H_0^s \iff \frac{1}{(2\pi)^s} \sum_{n\geq 1} \alpha_n^2 \lambda_n^s < \infty \tag{1.13}$$

and if $g = \sum_{n\geq 1} \beta_n e_n$ is such that $\sum_{n\geq 1} \beta_n^2 \lambda_n^s < \infty$ as well, then

$$(f, g)_s = \frac{1}{(2\pi)^s} \sum_{n\geq 1} \alpha_n \beta_n \lambda_n^s < \infty. \tag{1.14}$$

Furthermore, when $s = 1$, and $f = \sum \alpha_n e_n, g = \sum \beta_n e_n \in \mathcal{D}_0(D) \subset L^2$, the identity

$$(f, g)_1 = \frac{1}{2\pi} \sum_n (\alpha_n \sqrt{\lambda_n})(\beta_n \sqrt{\lambda_n}) = \frac{1}{2\pi}(f, -\Delta g) = (f, g)_\nabla$$

holds by (1.7). Thus the space H_0^s with $s = 1$ as defined above is the same space as that defined above in Definition 1.18.

This calculation actually shows that the map

$$T : f \in L^2(D) \mapsto g = (-\Delta/2\pi)^{-1/2} f \in H_0^1(D)$$

is an isometry of Hilbert spaces. Indeed,

$$\|g\|_\nabla^2 = \|g\|_{s=1}^2 = ((-\Delta/2\pi)^{1/2} g, (-\Delta/2\pi)^{1/2} g) = (f, f) = \|f\|^2.$$

This explains and generalises (1.11).

Remark 1.19. Since $-\Delta$ is an operator which takes two derivatives of f, we have that, intuitively speaking, $(-\Delta)^s$ is an operator which takes $2s$ derivatives of f. In particular, we should not be surprised that for $s = 1$, the image of L^2 by $(-\Delta)^{-1/2}$ gives us $H_0^1(D)$ (up to a factor of 2π). For $s \leq 0$, $(-\Delta)^s$ takes "anti-derivative" of a distribution and so smoothes it. So to say that $f \in H_0^s$ with $s > 0$ is to say that f is slightly more regular than an L^2 function; while to say that $f \in H_0^s$ with $s < 0$ is to say that it is slightly less regular: indeed, "taking $|s|$ anti-derivatives" we would get an actual L^2 function.

Lemma 1.20. *Let $s \in \mathbb{R}$ and let $f \in H_0^s$. Then the map $g \in H_0^{-s} \mapsto (f, g)$ is a linear continuous functional.*

In other words, H_0^{-s} is the dual space of H_0^s.

Proof. For $f, g \in \mathcal{D}_0(D)$, writing $f = \sum_{n\geq 1} \alpha_n e_n$ and $g = \sum_{n\geq 1} \beta_n e_n$ we have by the Cauchy–Schwarz inequality:

$$(f, g) = \sum_{n\geq 1} \alpha_n \beta_n \leq \left(\sum_{n\geq 1} \alpha_n^2 \lambda_n^{-s}\right)^{1/2} \left(\sum_{n\geq 1} \beta_n^2 \lambda_n^s\right)^{1/2} = \|f\|_s \|g\|_{-s}.$$

The lemma follows by density of $\mathcal{D}_0(D)$ in H_0^{-s} and then in H_0^s. □

It will be useful in what follows to rephrase the expression for $\mathrm{Var}(h, f)$ (when h is a GFF) in terms of Sobolev norms. Recall that by (1.8), if $\Delta f = 2\pi\rho$ for $f, \rho \in \mathcal{D}_0(D)$, then

$$\Gamma_0(\rho, \rho) = \frac{1}{2\pi} \int_D |\nabla f|^2. \tag{1.15}$$

In other words:

Lemma 1.21.

$$\mathrm{Var}(h, \rho) = \|\rho\|_{s=-1}^2. \tag{1.16}$$

Consequently, if ρ is a measure, then $\rho \in \mathcal{M}_0$ if and only if $\rho \in H_0^{-1}(D)$. In particular, if $\rho \in \mathcal{M}_0$ we can find a sequence of smooth functions $\rho_n \in \mathcal{D}_0(D)$ such that $\rho_n \to \rho$ in $H_0^{-1}(D)$. Equivalently, $\mathrm{Var}(h, \rho_n - \rho) \to 0$, so that $(h, \rho_n) \to (h, \rho)$ in $L^2(\mathbb{P})$.

1.6 GFF as a random distribution

At this stage we do not yet know that the GFF may be viewed as a random distribution (i.e., as a random variable in $\mathcal{D}_0'(D)$). The goal of this section will be to prove that such a representation exists. Guided by (1.8) we will find an expression for the GFF as a random series, which we will show converges in the distribution space $\mathcal{D}_0'(D)$. In fact, we will show that it converges in a Sobolev space of appropriate index.

Suppose that $-\Delta f = 2\pi\rho$ with ρ and f smooth compactly supported functions in D. Then on the one hand we have $\mathrm{Var}((h,\rho)) = \Gamma_0(\rho)$ by definition, and on the other, we have

$$\|f\|_\nabla^2 = \Gamma_0(\rho)$$

by (1.8). Formally this says that $\mathrm{Var}((h,f)_\nabla) = \|f\|_\nabla^2$, so by the polarisation identity for inner products, the covariance between $(h,f)_\nabla$ and $(h,g)_\nabla$ should be given by $(f,g)_\nabla$. This property is sometimes taken as the definition of the GFF; for example, in [She07]. For a more precise formulation, see Corollary 1.23 below.[2]

With this in mind, we let $(f_n)_{n\geq 1}$ denote an orthonormal basis of $H_0^1(D)$. Since f_n, f_m are orthogonal by definition, the random variables $(h,f_n)_\nabla$ and $(h,f_m)_\nabla$ will be uncorrelated (and hence independent) Gaussian random variables, with unit variances. This suggests the following representation for h as a random series:

$$h = \sum_{n=1}^{\infty} X_n f_n = \lim_{N\to\infty} \sum_{n=1}^{N} X_n f_n. \tag{1.17}$$

It is not clear at this point in what sense (if any) this series converges. We will see below in Theorem 1.24 that it converges in an appropriate Sobolev space and hence in the space of distributions. Note however that the series does **not** converge a.s. in $H_0^1(D)$, since the H_0^1 norms of the partial sums tend to infinity almost surely as $N\to\infty$ (by the law of large numbers).

We start with the following observation. Set $h_N = \sum_{n=1}^N X_n f_n$, and let $f \in \mathcal{D}_0(D)$ or more generally let $f \in H_0^1(D)$. Then

$$(h_N, f)_\nabla = \sum_{n=1}^{N} X_n (f_n, f)_\nabla \tag{1.18}$$

does converge almost surely and in $L^2(\mathbb{P})$, by the martingale convergence theorem. Its limit is a Gaussian random variable with variance $\sum_{n\geq 1}(f_n,f)_\nabla^2 = \|f\|_\nabla^2$ by Parseval's identity. This defines a random variable which we call $(h,f)_\nabla$, which has the law of a mean zero Gaussian random variable with variance $\|f\|_\nabla^2$. Hence while the series (1.17) does *not* converge in H_0^1, when we take the inner product with a given $f \in H_0^1$ then this does converge almost surely.

[2] In the language of Gaussian Hilbert spaces (see e.g. [Jan97]) this is saying that the GFF is the canonical Gaussian random variable "in" the Hilbert space H_0^1 (the quotation marks are added since in fact h does not live in H_0^1).

Let $\rho \in \mathcal{D}_0(D)$, and let $f \in \mathcal{D}_0(D)$ be such that $-\Delta f = 2\pi\rho$. Then $(h_N, \rho) = (h_N, f)_\nabla$ converges almost surely. Let us call $(\tilde{h}, \rho)$ this limit. Then $(\tilde{h}, \rho)$ defines a stochastic process indexed by smooth test functions, which has the same law as a GFF restricted to $\mathcal{D}_0(D)$. Using a density argument we will now show the following:

Theorem 1.22 (GFF as a random Fourier series)**.** *Let D be a Greenian domain. Set $h_N = \sum_{n=1}^N X_n f_n$ to be the truncated series in (1.17). Then for any $\rho \in \mathcal{M}_0$,*

$$(h_N, \rho) \to (h, \rho)$$

in $L^2(\mathbb{P})$ (and hence in probability as well) as $N \to \infty$, where (h, ρ) is a Gaussian random variable with variance $\Gamma_0(\rho, \rho)$.

Proof. Let $\nu \in \mathcal{D}_0(D)$. We claim that for every $N \geq 1$,

$$\mathrm{Var}(h_N, \nu) \leq \mathrm{Var}(h, \nu). \tag{1.19}$$

To see this, observe that

$$\mathrm{Var}(h_N, \nu) = \sum_{n=1}^N (f_n, \nu)^2$$

When ν is smooth the sum is equal to $\sum_{n=1}^N (f_n, f)_\nabla^2$, where $2\pi\nu = -\Delta f$, and so converges to $\|f\|_\nabla^2$. As we have already seen, this is equal to $\iint G_0(x,y)\nu(x)\nu(y)dxdy$ and hence equal to $\mathrm{Var}(h, \nu)$. Hence (1.19) is proved for ν smooth. This can be extended by density to arbitrary $\nu \in \mathcal{M}_0$, as follows.

Let $\nu \in \mathcal{M}_0$, and recall by Lemma 1.21 that $\mathcal{M}_0 \subset H_0^{-1}$, so that there exists $\nu_\varepsilon \in \mathcal{D}_0(D)$ such that $\nu_\varepsilon \to \nu$ in H_0^{-1} as $\varepsilon \to 0$. By (1.19), we have

$$\mathrm{Var}(h_N, \nu_\varepsilon) \leq \mathrm{Var}(h, \nu_\varepsilon).$$

Now, by definition of H_0^{-1} norm, $\Gamma_0(\nu_\varepsilon) \to \Gamma_0(\nu)$ as $\varepsilon \to 0$. On the other hand, $\mathrm{Var}(h_N, \nu_\varepsilon) = \sum_{n=1}^N (f_n, \nu_\varepsilon)^2$ and since each f_n is smooth, we see that $(f_n, \nu_\varepsilon) \to (f_n, \nu)$ as $\varepsilon \to 0$, which shows (1.19).

Now let $\rho \in \mathcal{M}_0$. Let ρ_ε be an approximation of ρ in the H_0^{-1} sense (again by Lemma 1.21), and let $\nu_\varepsilon = \rho - \rho_\varepsilon$. Then $\mathrm{Var}(h, \nu_\varepsilon) \to 0$, so that applying (1.19) to $\nu = \nu_\varepsilon$ we deduce that (h_N, ν_ε) converges to 0 in L^2 and in probability as $\varepsilon \to 0$, uniformly in N. The result follows since for smooth ρ_ε, we already know (as a consequence of the martingale convergence argument) that $(h_N, \rho_\varepsilon) \to (\tilde{h}, \rho_\varepsilon)$ in L^2, and this has the same law as (h, ρ_ε). □

It is worth reiterating what we have just learnt:

Corollary 1.23. *The restriction of a Gaussian free field h with zero boundary conditions to $\mathcal{D}_0(D)$ is the unique stochastic process such that for all $f \in \mathcal{D}_0(D)$, $(h, f)_\nabla$ is a Gaussian centred variable with variance $(f, f)_\nabla$, where we define $(h, f)_\nabla$ by $-(2\pi)^{-1}(h, \Delta f)$ for each f.*

We finally address convergence of the series (1.17):

Theorem 1.24 (GFF as a random variable in a Sobolev space)**.** *Suppose D is bounded. If $(X_n)_{n\geq 1}$ are i.i.d. standard Gaussian random variables and $(f_n)_{n\geq 1}$ is any orthonormal basis of $H_0^1(D)$, then the series $\sum_{n\geq 1} X_n f_n$ converges almost surely in $H_0^s(D)$, where*

$$s = 1 - \tfrac{d}{2} - \varepsilon,$$

for any $\varepsilon > 0$. In particular, for $d = 2$, the series converges in $H_0^{-\varepsilon}(D)$ for any $\varepsilon > 0$.

Observe that by Theorem 1.22, the law of the limit is uniquely defined.

Proof. Let us take $(e_m)_{m\geq 1}$ to be an orthonormal basis of $L^2(D)$ which are eigenfunctions for $-\Delta$, as in Section 1.5. This is possible since D is bounded. As usual we write λ_m for the eigenvalue corresponding to e_m; so that $((\lambda_m/2\pi)^{-s/2} e_m)_{m\geq 1}$ is an orthonormal basis of $H_0^s(D)$ and $((\lambda_m/2\pi)^{-1/2} e_m)_{m\geq 1}$ is an orthonormal basis of $H_0^1(D)$. In some cases λ_m can be computed explicitly: e.g., when D is a rectangle or the unit disc. In general, we will make use of the following fundamental estimate due to Weyl (see e.g., [Cha84, Note III.15] for a proof):

Lemma 1.25. *We have*

$$\lambda_m \sim c m^{2/d}$$

as $m \to \infty$, in the sense that the ratio of the two sides tends to 1 as $m \to \infty$, where $c = (2\pi)^2/(A_d \operatorname{Leb}(D))^{2/d}$.

The upshot is that if $(f_n)_{n\geq 1}$ is *any* orthonormal basis of $H_0^1(D)$, we can control the expectation

$$\mathbb{E}(\| \sum_{n=M}^{N} X_n f_n \|_{H_0^s}^2) = \mathbb{E}(\sum_{n=M}^{N} \sum_{m=M}^{N} X_n X_m (f_n, f_m)_{H_0^s}) = \sum_{n=M}^{N} \|f_n\|_{H_0^s}^2 \tag{1.20}$$

as $N, M \to \infty$.

Indeed, by applying Parseval's identity, then (1.14), and then Parseval again, we have that

$$\begin{aligned} \sum_{n\geq 1} \|f_n\|_{H_0^s}^2 &= \sum_{n\geq 1} \sum_m (f_n, (\frac{\lambda_m}{2\pi})^{-s/2} e_m)_{H_0^s}^2 \\ &= \sum_m (\frac{\lambda_m}{2\pi})^{-1+s} \sum_{n\geq 1} (f_n, (\frac{\lambda_m}{2\pi})^{-1/2} e_m)_{H_0^1}^2 \\ &= \sum_m (\frac{\lambda_m}{2\pi})^{-1+s} < \infty \end{aligned}$$

where we have used positivity and Fubini to interchange the order of summation. The finiteness of the last sum follows by Lemma 1.25, since $(2/d)(-1+s) = -1-\varepsilon(2/d) < -1$. In particular, the sequence $\sum_{n=1}^{N} \|f_n\|_{H_0^s}^2$ is Cauchy in N and so (1.20) tends to 0 as $N, M \to \infty$. This shows that $\sum_{n=1}^{N} X_n f_n$ is Cauchy in $L^2(\mathbb{P})$, and hence converges as desired.

□

Remark 1.26. The above implies that the series $\sum_{n=1}^{N} X_n f_n$ converges almost surely in the space of distributions $\mathcal{D}_0'(D)$ whenever D is bounded. However in two dimensions, conformal invariance allows us to drop this boundedness assumption. Indeed, take D any simply connected domain of $\mathbb{C}$, and let ϕ be a conformal map from $\mathbb{D} \to D$. Then by conformal invariance of the Dirichlet inner product, if $(f_n)_{n\geq 1}$ is an orthonormal basis of $H_0^1(D)$, the images $(f_n \circ \phi)_{n\geq 1}$ form an orthonormal basis of $H_0^1(\mathbb{D})$. So if $(X_n)_{n\geq 1}$ are i.i.d. standard Gaussians, we have that $\sum_{n\geq 1} X_n (f_n \circ \phi)$ converges a.s. in $\mathcal{D}_0'(\mathbb{D})$, and consequently that $\sum_{n\geq 1} X_n f_n$ converges in $\mathcal{D}_0'(D)$.

1.7 Itô's isometry for the GFF

⋆ *This section will not be used in the rest of the text and the reader may wish to skip it on a first reading.*

In this section we describe an observation which emerged from joint discussions with James Norris. It is closely linked to Lemma 1.21, which implies that for a zero boundary GFF h in a bounded domain D, and for any $f \in H_0^{-1}(D)$, the quantity (h, f) makes sense almost surely. That is, as the almost sure (and $L^2(\mathbb{P})$) limit of (h, f_n) for any sequence f_n converging to f in $H_0^{-1}(D)$.

In other words, even though h is only almost surely defined as a continuous linear functional on $H^{d/2-1+\varepsilon}(D)$ for $\varepsilon > 0$ (Theorem 1.24), we can actually test it against fixed functions that are much less regular. Namely, we can test it against any fixed function $H_0^{-1}(D)$. Note that this agrees with (in fact slightly extends) our previous definition of h as a stochastic process, since we have seen that $\mathcal{M}_0$ is precisely the set of signed measures that are elements of H_0^{-1}.

In this section we will essentially formulate the above discussion in terms of an isometry. To motivate this, it is useful to recall the following well known analogy within Itô's theory of stochastic integration. Let B be a standard Brownian motion. Even though dB does not have the regularity of a function in L^2 (in fact, it is essentially an element of $H^{-1/2-\varepsilon}$ for any $\varepsilon > 0$), it makes perfect sense to integrate it against a test function in L^2. This is thanks to the fact that the map

$$f \mapsto \int f_s dB_s$$

defines an isometry of suitable Hilbert spaces. Thus much flexibility has been gained: *a priori* we don't even have the right to integrate against functions $H^{1/2}$, and yet, taking advantage of some almost sure properties of Brownian motion – namely, quadratic variation – it is possible to integrate against functions in L^2 (and actually much more).

A similar gain can be seen in the context of the GFF: *a priori*, as an element of $H_0^{1-d/2-\varepsilon}$ ($\varepsilon > 0$), it would seem that integrating against an arbitrary test function $f \in L^2$ is not even allowed when $d \geq 2$. Yet, as discussed above, we can a.s. integrate against much rougher objects, namely distributions in H_0^{-1}:

Theorem 1.27 (Itô isometry)**.** *The map X sending $f \in \mathcal{D}_0(D)$ to the random variable $X_f = (h, f)$ can be viewed as a linear map between $\mathcal{D}_0(D)$ and the set of random variables $L^2(\Omega, \mathcal{F}, \mathbb{P})$ viewed as a function space. If we endow $\mathcal{D}_0(D)$ with the $H_0^{-1}(D)$ norm and $L^2(\Omega, \mathcal{F}, \mathbb{P})$ with its L^2 norm then X is an isometry:*

$$\|f\|_{H_0^{-1}(D)} = \|X_f\|_2 = \mathbb{E}((h, f)^2)^{1/2}.$$

In particular, since $\mathcal{D}_0(D)$ is dense in $H_0^{-1}(D)$, X_f extends uniquely as an isometry from $H_0^{-1}(D)$ into $L^2(\Omega, \mathcal{F}, \mathbb{P})$. Hence if $f \in H_0^{-1}(D)$, then we can set (h, f) to be the unique limit in $L^2(\mathbb{P})$ of (h, f_n) where f_n is any sequence of test functions that converge in $H_0^{-1}(D)$ to f.

Proof. This is a direct consequence of Lemma 1.21. □

Remark 1.28. Note that although (h, f) makes sense as an a.s. limit for any fixed $f \in H_0^{-1}(D)$, or indeed for any countable collection of such f, this does not mean that h is an element of H_0^1 or that we can test h against every element of H_0^{-1} simultaneously. For example, writing

$$h = \lim_{N\to\infty} h_N := \lim_{N\to\infty} \sum_{n=1}^{N} X_n \sqrt{\frac{2\pi}{\lambda_n}} e_n$$

with $(X_n)_n \sim \mathcal{N}(0, 1)$ i.i.d. and $(e_n)_n$ an orthonormal basis of Laplacian eigenfunctions for $L^2(D)$, we have $h_N \to h$ a.s. in $H_0^{-1}(D)$ but $\mathrm{Var}(h, h_N) = \sum_{n=1}^{N} \lambda_n^{-1} \to \infty$ (at least when $d \geq 2$). So there do exist random elements of $H_0^{-1}(D)$ that cannot be tested against h.

1.8 Cameron–Martin space of the Dirichlet GFF

⋆ *This section will not be used until Chapter 5 and the reader may wish to skip it on a first reading.*

In this section, we will address the following question:

- for h a Dirichlet (zero) boundary condition GFF in D and F a (deterministic) function on D, when are $h + F$ and h mutually absolutely continuous?[3]

The answer is that this holds whenever $F \in H_0^1(D)$. This question can be phrased for general Gaussian processes, and the space of "F" for which absolute continuity holds is known as the **Cameron–Martin space** of the process. Thus, the lemma below says that $H_0^1(D)$ is the Cameron–Martin space of the (Dirichlet boundary condition) GFF.

Proposition 1.29. *Let h be a GFF in a bounded domain D with Dirichlet (zero) boundary conditions. Then h and $h + F$ are mutually absolutely continuous, as stochastic processes*

[3] as stochastic processes indexed by $\mathcal{M}_0$.

indexed by $\mathcal{M}_0$, if and only if $F \in H_0^1(D)$. When this holds, the Radon–Nikodym derivative of $(h+F)$ with respect to h is given by

$$\frac{\exp((h,F)_\nabla)}{\exp((F,F)_\nabla/2)}.$$

Proof. Let $(e_i)_i$ be an orthonormal basis of $L^2(D)$ consisting of eigenfunctions of the Laplacian, with associated eigenvalues $(\lambda_i)_i$. We write $g_i := (\sqrt{\lambda_i}/\sqrt{2\pi})e_i$, so that the $(g_i)_i$ form an orthonormal basis of $H_0^{-1} \supset \mathcal{M}_0$. Recall that

$$F \in H_0^1(D) \Leftrightarrow \sum (F,g_i)^2 < \infty.$$

For $n \in \mathbb{N}$, we consider the finite vector $((h,g_i))_{1\le i\le n}$, which by definition of the GFF is just a vector of independent $\mathcal{N}(0,1)$ random variables.

This is convenient to work with because of the following elementary fact: if $(X_i)_{1\le i\le n}$ are i.i.d. standard normals and $(a_i)_{1\le i\le n}$ are real numbers, then $(X_1, X_2, ..., X_n)$ and $(X_1 + a_1, X_2 + a_2, ...X_n + a_n)$ are mutually absolutely continuous. Moreover, the RN derivative of the latter with respect to the former is given by $e^{\sum a_i X_i}/e^{\sum a_i^2/2}$.

In our context, this means that the law of $((h,g_i))_{1\le i\le n}$ is mutually absolutely continuous with that of $((h+F,g_i))_{1\le i\le n}$, if and only if $(F,g_i) < \infty$ for $1 \le i \le n$. Furthermore when this does hold, the Radon-Nikodym derivative of $((h+F,g_i))_{1\le i\le n}$ with respect to $((h,g_i))_{1\le i\le n}$ is equal to

$$\frac{\exp(\sum_{i=1}^n (h,g_i)(F,g_i))}{\mathbb{E}(\exp(\sum_{i=1}^n (h,g_i)(F,g_i)))} = \frac{\exp(\sum_{i=1}^n (h,g_i)(F,g_i))}{\exp(\sum_{i=1}^n (F,g_i)^2/2)}. \tag{1.21}$$

Now, for h and $h+F$ to be mutually absolutely continuous, the family of random variables on the right hand side of (1.21) must be uniformly integrable (in n). Indeed, they should be the conditional expectations, with respect to a family of sub σ-algebras, of the Radon–Nikodym derivative of $(h+F)$ with respect to h. This family is *not* uniformly integrable if $F \notin H_0^1(D)$, i.e., $\sum_{i\ge 1}(F,g_i)^2 = \infty$. Hence we obtain the necessity of the condition $F \in H_0^1(D)$ in the proposition.

For the sufficiency, we observe that when $F \in H_0^1(D)$, the random variables on the right hand side of (1.21) converge in $L^1(\mathbb{P})$ to

$$\frac{\exp(\sum_{i\ge 1}(h,g_i)(F,g_i))}{\exp(\sum_{i\ge 1}(F,g_i)^2/2)} = \frac{\exp((h,F)_\nabla)}{\exp((F,F)_\nabla/2)}$$

as $n \to \infty$. We also know by Theorem 1.22 that whenever $\rho \in \mathcal{M}_0$, $\sum_{i=1}^n 2\pi\lambda_i^{-1}(\rho,g_i)(h,g_i)$ converges to (h,ρ) a.s. This implies that for any $\rho_1, ..., \rho_m \in \mathcal{M}_0$ and any $\psi : \mathbb{R}^m \to \mathbb{R}$ continuous and bounded:

$$\mathbb{E}\left(\psi((h+F,\rho_1), ..., (h+F,\rho_m))\right) = \mathbb{E}\left(\frac{\exp((h,F)_\nabla)}{\exp((F,F)_\nabla/2)}\psi((h,\rho_1), ..., (h,\rho_m))\right).$$

But this is exactly the statement that $h+F$ is absolutely continuous with respect to h, as a stochastic process indexed by $\mathcal{M}_0$, with the desired Radon–Nikodym derivative. Since the inverse of the Radon–Nikodym derivative is also in L^1, we obtain the mutual absolute continuity. □

1.9 Markov property

We are now ready to state one of the main properties of the GFF, which is the (domain) Markov property. As in the discrete case, informally speaking, it states that conditionally on the values of h outside of a given subset U, the free field inside U is obtained by harmonically extending $h|_{D\setminus U}$ into U and then adding an independent GFF with Dirichlet boundary conditions. Note that in this case, however, it is not at all clear that such a harmonic extension is well defined.

Theorem 1.30 (Markov property)**.** *Fix $U \subset D$ open. Let h be a GFF (with zero boundary conditions on D). Then we may write*

$$h = h_0 + \varphi,$$

where:

1. *h_0 is a zero boundary condition GFF in U, and is zero outside of U;*
2. *φ is harmonic in U; and*
3. *h_0 and φ are independent.*

This makes sense whether we view h as a random distribution or a stochastic process indexed by $\mathcal{M}_0$.

Proof. The key point is the following Hilbertian decomposition:

Lemma 1.31. *Let U be as in Theorem 1.30. We have*

$$H_0^1(D) = \mathrm{Supp}(U) \oplus \mathrm{Harm}(U),$$

where $\mathrm{Harm}(U)$ *consists of harmonic functions in U, and* $\mathrm{Supp}(U) \equiv H_0^1(U)$ *consists of functions supported in U (the closure with respect to the Dirichlet inner product of compactly supported smooth functions in U).*

Proof. First observe that any function $f \in \mathrm{Supp}(U)$ is orthogonal to any function in $\mathrm{Harm}(U)$, by the Gauss–Green formula. Note that this holds for any U (however regular) by definition of $\mathrm{Supp}(U)$ (indeed, just take a sequence of converging compactly supported functions). Thus we are left to check that the sum spans the entire space.

Let us suppose to begin with that U is C^1 smooth. For $f \in H_0^1(D)$, let f_0 denote the orthogonal projection of f onto $\mathrm{Supp}(U)$. Set $\varphi = f - f_0$: our aim is to show that φ is harmonic in U. Note that φ is (by definition) orthogonal to $\mathrm{Supp}(U)$. Hence for any test function $\psi \in \mathcal{D}_0(U)$, we have that $(\varphi, \psi)_\nabla = 0$. By the Gauss–Green formula (and since U is C^1 smooth), we deduce that

$$\int_D (\Delta\varphi)\psi = \int_U (\Delta\varphi)\psi = 0$$

and hence $\Delta\varphi = 0$ as a distribution in U. Elliptic regularity arguments (going beyond the scope of these notes) show that a distribution which is harmonic in the sense of distributions must in fact be a smooth function, harmonic in the usual sense. Therefore $\varphi \in \mathrm{Harm}(U)$ and we are done.

If U does not have C^1 boundary, let $(U_n)_{n\in\mathbb{N}}$ be a sequence of increasing open subsets of U with C^1 boundaries, such that $\cup U_n = U$. For $f \in H_0^1(D)$, by the previous paragraph, we can write $f = f_0^n + \varphi^n$ for each $n \in \mathbb{N}$, where f_0^n is the projection of f onto $\mathrm{Supp}(U_n)$ and $\varphi^n \in \mathrm{Harm}(U_n)$. Then we just need to show that: (a) $f_0^n \to f_0$ as $n \to \infty$ for some $f_0 \in \mathrm{Supp}(U)$; and (b) that $f - f_0$ is harmonic in U. For (a), we observe that $\mathrm{Supp}(U) = \overline{\cup_n \mathrm{Supp}(U_n)}$ (by definition of $\mathrm{Supp}(U)$ as the closure of $\mathcal{D}_0(U)$ with respect to the Dirichlet inner product) and so the projections f_0^n of f onto $\mathrm{Supp}(U_n)$ converge, with respect to $\|\cdot\|_\nabla$, to $f_0 \in \mathrm{Supp}(U)$. For (b), notice that by definition of f_0, $f - f_0$ is the limit of φ_n as $n \to \infty$, with respect to $\|\cdot\|_\nabla$. In particular, it is clear that when restricted to any U_n, $f - f_0 = \lim_n \varphi_n$ is harmonic in the distributional sense, and thus harmonic by elliptic regularity. Since this holds for any n, it follows that $f - f_0$ is harmonic in U. □

Having this decomposition in hand, we may deduce the Markov property in a rather straightforward way. Indeed, let $(f_n^0)_n$ be an orthonormal basis of $\mathrm{Supp}(U)$, and let $(\phi_n)_n$ be an orthonormal basis of $\mathrm{Harm}(U)$. For $((X_n, Y_n))_n$ an i.i.d. sequence of independent standard Gaussian random variables, set $h_0 = \sum_n X_n f_n^0$ and $\varphi = \sum_n Y_n \phi_n$. Then the first series converges in $\mathcal{D}_0'(D)$ since it is a series of a GFF in U. The sum of the two series gives h by construction, and so the second series also converges in the space of distributions. In the space of distributions, the limit of harmonic distributions must be harmonic as a distribution, and hence (by the same elliptic regularity arguments as above) a true harmonic function. This proves the theorem. □

Remark 1.32. It is worth pointing out an important message from the proof above: any orthogonal decomposition of $H_0^1(D)$ gives rise to a decomposition of the GFF into independent summands.

Example. When $d = 1$. this is the statement that if $(b_s)_{s\in[0,1]}$ is a Brownian bridge from 0 to 0 and $[a, b] \subset [0, 1]$, then conditionally on $(b_s)_{s\in[0,a]\cup[b,1]}$, the law of $(b_s)_{s\in[a,b]}$ is given by a the linear interpolation of b_a and b_b, plus an independent Brownian bridge from 0 to 0 on $[a, b]$.

1.10 Conformal invariance

In the remainder of this chapter, we restrict ourselves to dimension $d = 2$.

In this case the GFF possesses the important additional property of **conformal invariance**, which follows almost immediately from the construction in the previous section. Indeed, a straightforward change of variable formula shows that the Dirichlet inner product is conformally invariant: if $\varphi : D \to D'$ is a conformal map, then

$$\int_{D'} \nabla(f \circ \varphi^{-1}) \cdot \nabla(g \circ \varphi^{-1}) = \int_D \nabla f \cdot \nabla g.$$

Consequently, if $(f_n)_n$ is an orthonormal basis of $H_0^1(D)$, then $(f_n \circ \varphi^{-1})_n$ defines an orthonormal basis of $H_0^1(D')$. So by Theorem 1.24:

Theorem 1.33 (Conformal invariance of the GFF)**.** *If h is a random distribution on $\mathcal{D}_0'(D)$ with the law of the Gaussian free field on D, then the distribution $h \circ \varphi^{-1}$, defined by setting $(h \circ \varphi^{-1}, f) = (h, |\varphi'|^2 (f \circ \varphi))$ for $f \in \mathcal{D}_0'(D')$, has the law of a GFF on D'.*

Recently, a kind of converse was shown in [BPR20a, BPR20b]: if a field h with zero boundary conditions satisfies conformal invariance and the domain Markov property, as well as a moment condition ($\mathbb{E}((h,\phi)^{1+\varepsilon}) < \infty$ for some $\varepsilon > 0$ and all $\phi \in \mathcal{D}_0(D)$), then h must be a multiple of the Gaussian free field. See [BPR20a, BPR20b] for details.

1.11 Circle averages

An important tool for studying the GFF is the process which describes its average values on small circles centred around a point $z \in D$. This is known as the **circle average process** around z.

More precisely, fix $z \in D$ and let $0 < \varepsilon < \operatorname{dist}(z, \partial D)$. Let $\rho_{z,\varepsilon}$ denote the uniform distribution on the circle of radius ε around z, and note that $\rho_{z,\varepsilon} \in \mathcal{M}_0$. We set $h_\varepsilon(z) = (h, \rho_{z,\varepsilon})$ The following theorem, which is a consequence of the Kolmogorov-Čentsov continuity theorem (a multidimensional generalisation of the more classical Kolmogorov continuity criterion), will not be proved here. The interested reader is directed to Proposition 3.1 of [DS11] for a proof.

Proposition 1.34 (Circle average is jointly Hölder)**.** *There exists a modification of h such that $(h_\varepsilon(z), z \in D, 0 < \varepsilon < \operatorname{dist}(z, \partial D))$ is a.s. jointly Hölder continuous of order $\eta < 1/2$ on all compact subsets of $\{z \in D \text{ s.t. } 0 < \varepsilon < \operatorname{dist}(z, \partial D)\}$.*

In fact it can be shown that this version of the GFF is the same as the version which turns h into a random distribution in Theorem 1.22. The reason circle averages are so useful is because of the following result.

Theorem 1.35 (Circle average is a Brownian motion)**.** *Let h be a GFF on D. Fix $z \in D$ and let $0 < \varepsilon_0 < \operatorname{dist}(z, \partial D)$. For $t \geq t_0 = \log(1/\varepsilon_0)$, set*

$$B_t = h_{e^{-t}}(z).$$

Then $(B_t, t \geq t_0)$ has the law of a Brownian motion started from B_{t_0}.

Proof. Various proofs can be given. For instance, the covariance function can be computed explicitly (this is a good exercise)! Alternatively, we can use the Markov property of the GFF to see that B_t must have stationary and independent increments. Indeed, suppose $\varepsilon_1 > \varepsilon_2$, and we condition on h outside of $B(z, \varepsilon_1)$. That is, we write $h = h^0 + \varphi$, where φ is harmonic in $U = B(z, \varepsilon_1)$ and h^0 is a GFF in U that is independent of $(h_\varepsilon(z))_{\varepsilon \geq \varepsilon_1}$. Then $h_{\varepsilon_2}(z)$ is the

sum of two terms: $h^0_{\varepsilon_2}(z)$; and the circle average of φ on $\partial B(z, \varepsilon_2)$. By harmonicity of φ the latter is nothing else than $h_{\varepsilon_1}(z)$. This gives that the increment can be expressed as

$$h_{\varepsilon_2}(z) - h_{\varepsilon_1}(z) = h^0_{\varepsilon_2}(z)$$

and hence, since h^0 is independent of $(h_\varepsilon(z))_{\varepsilon \geq \varepsilon_1}$, the increments are independent. Moreover, by applying the change of scale $w \mapsto (w - z)/\varepsilon_1$, so that the outer circle is mapped to the unit circle, we see that the distribution of $h_{\varepsilon_2}(z) - h_{\varepsilon_1}(z)$ depends only on $r = \varepsilon_2/\varepsilon_1$. This means that they are also stationary.

To show from here that $h_{e^{-t}}(z)$ is a Brownian motion, it suffices to compute its variance. That is, to check that if h is a GFF in the unit disc $\mathbb{D}$ and $r < 1$, then $h_r(0)$ has variance $-\log r$.

For this, let ρ denote the uniform distribution on the circle $\partial(r\mathbb{D})$ at distance r from the origin, so that

$$\mathrm{Var}(h_r(0)) = \int_{\mathbb{D}^2} G^{\mathbb{D}}_0(x, y)\rho(dx)\rho(dy). \tag{1.22}$$

The point is that by harmonicity of $G^{\mathbb{D}}_0(x, \cdot)$ in $\mathbb{D} \setminus \{x\}$ and the mean value property, the above integral is simply

$$\mathrm{Var}(h_r(0)) = \int_{\mathbb{D}} G^{\mathbb{D}}_0(x, 0)\rho(dx), \tag{1.23}$$

which completes the proof since $G^{\mathbb{D}}_0(x, 0) = -\log|x| = -\log r$ on $\partial(r\mathbb{D})$.

To check (1.23) rigourously, first consider for a fixed $\eta > 0$, the double integral

$$I_\eta = \int_{\mathbb{D}^2} G^{\mathbb{D}}_0\big((1 + \eta)x, y\big)\rho(dx)\rho(dy).$$

Then I_η converges clearly to the right hand side of (1.22) as $\eta \to 0$, and it is now rigourous to exploit the mean-value property for the harmonic function $G^{\mathbb{D}}_0((1 + \eta)x, \cdot)$ in the entire ball $B(0, r)$ to deduce that

$$I_\eta = \int_{\mathbb{D}} G^{\mathbb{D}}_0\big((1 + \eta)x, 0\big)\rho(dx).$$

Letting $\eta \to 0$ proves (1.23).

□

So, as we "zoom in" towards a point, the average values of the field oscillate like those of a Brownian motion. This gives us a very precise sense in which the field cannot be defined pointwise.

1.12 Thick points

An important notion in the study of Liouville quantum gravity is that of thick points of the Gaussian free field. Indeed, although these points are atypical from the point of view of Euclidean geometry, we will see that they are typical from the point of the associated quantum geometry.

Definition 1.36. *Let h be a GFF in $D \subset \mathbb{C}$ simply connected and let $\alpha > 0$. We say a point $z \in D$ is α-thick if*

$$\liminf_{\varepsilon \to 0} \frac{h_\varepsilon(z)}{\log(1/\varepsilon)} = \alpha.$$

In fact, the lim inf in the definition could be replaced with a lim sup or lim. Note that a given point $z \in D$ is almost surely not thick: the typical value of $h_\varepsilon(z)$ is of order $\sqrt{\log 1/\varepsilon}$ since $h_\varepsilon(z)$ is a Brownian motion at scale $\log 1/\varepsilon$. At this stage, the most relevant result is the following fact due to Hu, Miller and Peres [HMP10] (though it was independently and earlier proved by Kahane in the context of his work on Gaussian multiplicative chaos).

Theorem 1.37. *Let $\mathcal{T}_\alpha$ denote the set of α-thick points. Then almost surely,*

$$\dim(\mathcal{T}_\alpha) = (2 - \frac{\alpha^2}{2})_+$$

and $\mathcal{T}_\alpha$ is empty if $\alpha > 2$.

Heuristics. The value of the dimension of $\mathcal{T}_\alpha$ is easy to understand and to guess. Indeed, for a given $\varepsilon > 0$,

$$\begin{aligned}\mathbb{P}(h_\varepsilon(z) \geq \alpha \log(1/\varepsilon)) &= \mathbb{P}(\mathcal{N}(0, \log(1/\varepsilon) + O(1)) \geq \alpha \log(1/\varepsilon)) \\ &= \mathbb{P}(\mathcal{N}(0,1) \geq \alpha \sqrt{\log(1/\varepsilon) + O(1)}) \leq \varepsilon^{\alpha^2/2}\end{aligned}$$

using scaling and the standard bound $\mathbb{P}(X > t) \leq \text{const} \times t^{-1} e^{-t^2/2}$ for $X \sim \mathcal{N}(0,1)$. Suppose without loss of generality that $D = (0,1)^2$ is the unit square. Then the expected number of squares of size ε such that the centre z satisfies $h_\varepsilon(z) \geq \alpha \log 1/\varepsilon$ is bounded by $\varepsilon^{-2+\alpha^2/2}$. This suggests that the Minkowski dimension is less or equal to $2 - \alpha^2/2$ when $\alpha < 2$ and that $\mathcal{T}_\alpha$ is empty if $\alpha > 2$. □

Rigourous proof of upper bound. We now turn the above heuristics into a rigourous proof that $\dim(\mathcal{T}_\alpha) \leq 2 - \alpha^2/2 \vee 0$, which follows closely the argument given in [HMP10]. The lower bound given in [HMP10] is more complicated, but we will obtain an elementary proof in the next chapter, via the Liouville measure: see Exercise 4 of Chapter 2.

To start the proof of the upper bound, we begin by stating an improvement of Proposition 1.34, which is Proposition 2.1 in [HMP10]. This is the circle average analogue of Lévy's modulus of continuity for Brownian motion.

Lemma 1.38. *Suppose D is bounded with smooth boundary. Then for every $\eta < 1/2$, for every $\zeta > 0$ and $\varepsilon > 0$, there exists $M = M(\eta, \zeta, \varepsilon) < \infty$ a.s. such that*

$$|h_r(z) - h_s(w)| \leq M \left(\log \frac{1}{r}\right)^\zeta \frac{(|z-w| + |r-s|)^\eta}{r^{\eta+\varepsilon}}$$

for every $z, w \in D$ and for all $r, s \in (0,1)$ such that $r/s \in [1/2, 2]$ and $B(z,r), B(w,s) \subset D$.

See Proposition 2.1 in [HMP10] for a proof.

Without loss of generality, we will now work in the case where D is bounded with smooth boundary. This yields the proof in the general case by the domain Markov property, and since $\dim(\mathcal{T}_\alpha) = \lim_{n\to\infty} \dim(\mathcal{T}_\alpha \cap D_n)$ for a sequence of smooth, bounded domains D_n with $\cup D_n = D$.

In this setting, the above lemma allows us to "discretise" the set of ε and z on which it suffices to check thickness. More precisely, set $\varepsilon > 0$, $K > 0$ and consider the sequence of scales $r_n = n^{-K}$. Fix $\zeta < 1$, and $\eta < 1/2$ arbitrarily (say $\zeta = 1/2, \eta = 1/4$), and let $M = M(\eta, \zeta, \varepsilon\eta)$ be as in the lemma. Then for any $z \in D$, we have that if $r_{n+1} \le r \le r_n$,

$$\begin{aligned} |h_r(z) - h_{r_n}(z)| &\le MK^\zeta (\log n)^\zeta \frac{(r_{n+1} - r_n)^\eta}{r_n^{\eta(1+\varepsilon)}} \\ &\lesssim (\log n)^\zeta n^{K\eta(1+\varepsilon)-(K+1)\eta} \lesssim (\log n)^\zeta \end{aligned}$$

if we choose $\varepsilon = K^{-1}$. Thus any point $z \in D$ is in $\mathcal{T}_\alpha$ if and only if

$$\lim_{n\to\infty} \frac{h_{r_n}(z)}{\log 1/r_n} = \alpha.$$

Now for any $n \ge 1$ let $\{z_{n,j}\}_j = D \cap r_n^{1+\varepsilon}\mathbb{Z}^2$ be a set of discrete points spaced by $r_n^{1+\varepsilon}$ within D. Then if $z \in B(z_{n,j}, r_n^{1+\varepsilon})$ we have for the same reasons

$$|h_{r_n}(z) - h_{r_n}(z_{n,j})| \lesssim (\log n)^\zeta.$$

Thus for fixed $\delta > 0$ let

$$\mathcal{I}_n = \{j : h_{r_n}(z_{n,j}) \ge (\alpha - \delta) \log 1/r_n\}$$

Then for each $N \ge 1$, each $\delta > 0$,

$$\mathcal{T}'_\alpha = \bigcup_{n>N} \bigcup_{j\in\mathcal{I}_n} B(z_{n,j}, r_n^{1+\varepsilon})$$

is a cover of $\mathcal{T}_\alpha$. Consequently, if $\mathcal{H}_q$ denotes q-dimensional Hausdorff measure for $q > 0$,

$$\mathbb{E}(\mathcal{H}_q(\mathcal{T}_\alpha)) \le \mathbb{E}\left(\sum_{n>N}\sum_{j\in\mathcal{I}_n} \mathrm{diam} B(z_{n,j}, r_n^{1+\varepsilon})^q\right) \lesssim \sum_{n>N} r_n^{-2-2\varepsilon} r_n^{q(1+\varepsilon)} \max_j \mathbb{P}(j \in \mathcal{I}_n).$$

For a fixed n and a fixed j, as argued in the heuristics, $\mathbb{P}(j \in \mathcal{I}_n) \lesssim \exp(-(\alpha-\delta)^2/2\log(1/r_n)) = r_n^{(\alpha-\delta)^2/2}$ where the implied constants are uniform over D. We deduce

$$\mathbb{E}(\mathcal{H}_q(\mathcal{T}_\alpha)) \le \sum_{n>N} r_n^{-2-2\varepsilon+(\alpha-\delta)^2/2+q(1+\varepsilon)}.$$

As $r_n = n^{-K}$ and K can be chosen arbitrarily large, the right hand side tends to zero as $N \to \infty$ as soon as the exponent of r_n in the above sum is positive, or, in other words, if q is such that

$$-2 - 2\varepsilon + (\alpha - \delta)^2/2 + q(1+\varepsilon) > 0.$$

Thus we deduce that $\mathcal{H}_q(\mathcal{T}_\alpha) = 0$ a.s. (and hence $\dim(\mathcal{T}_\alpha) \leq q$ whenever $q(1+\varepsilon) > 2 + 2\varepsilon - (\alpha-\delta)^2/2$. So

$$\dim(\mathcal{T}_\alpha) \leq \frac{2 + 2\varepsilon - (\alpha-\delta)^2/2}{1+\varepsilon},$$

a.s. Since $\varepsilon > 0, \delta > 0$ are arbitrary, we deduce

$$\dim(\mathcal{T}_\alpha) \leq 2 - \alpha^2/2,$$

almost surely, as desired. □

The value $\alpha = 2$ corresponds informally to the maximum of the free field, and the study of the set $\mathcal{T}_2$ is, informally at least, related to the study of extremes in a branching Brownian motion (see [ABBS13, ABK13]).

1.13 Exercises

Discrete GFF

1. Describe the GFF on a binary tree of depth n, where ∂ is the root of the tree.
2. Using an orthonormal basis of eigenfunctions for $\hat{P}$, show that the partition function Z in Theorem 1.6 is given by
$$Z = \det(I - \hat{P})^{-1/2}$$
3. Prove that the minimiser of the discrete Dirichlet energy is discrete harmonic.
4. Prove the spatial Markov property of the discrete GFF.

Continuous GFF

1. Show that on the upper half pane,
$$G_0^{\mathbb{H}}(x,y) = \log\left|\frac{x-\bar{y}}{x-y}\right|.$$
Hint: use that $p_t^{\mathbb{H}}(x,y) = p_t(x,y) - p_t(x,\bar{y})$ by symmetry, and use the formula $e^{-a/t} - e^{-b/t} = t^{-1}\int_a^b e^{-x/t}dx$.
Deduce the value of G_0 on the unit disc.

2. Let $p_t(x,y)$ be the transition function of Brownian motion on the whole plane. Show that $\pi \int_0^1 p_t(x,y)dt = -\log|x-y| + O(1)$ as $x \to y$. Then use this to argue that $G_D(x,y) = -\log|x-y| + O(1)$ as $x \to y$, recovering the third property of Proposition 1.11.

3. Let D be a bounded domain and $z_0 \in D$. Suppose that $\phi(z)$ is harmonic in $D \setminus \{z_0\}$ and that

$$\phi(z) = -(1+o(1))\log|z-z_0| \text{ as } z \to z_0 \quad ; \quad \phi(z) \to 0 \text{ as } z \to z_0 \in \partial D.$$

Show that $\phi(z) = G_0^D(z_0, z)$ for all $z \in D \setminus \{z_0\}$. (*Hint: use the optional stopping theorem.*)

4. Consider $\tilde{h}_\varepsilon(z)$: the average value of the GFF on a square at distance ε from z. Is this a Brownian motion as a function of $t = \log 1/\varepsilon$? If not, how can you modify it so that it becomes a Brownian motion? More generally, what about the average of the field on a scaled contour $\varepsilon\lambda$, where λ is a piecewise smooth loop (the so-called potato average...)?

5. **Radial decomposition.** Suppose $D = \mathbb{D}$ is the unit disc and h is a GFF. Then show that h can be written as the sum

$$h = h_{\text{rad}} + h_{\text{circ}}$$

where h_{rad} is a radially symmetric function, h_{circ} is a distribution with zero average on each disc, and the two parts are independent. Specify the law of each of these two parts.

6. Let D be a proper simply connected domain and let $z \in D$.

(a) Show that

$$\log R(z;D) = \mathbb{E}_z(|B_T - z|)$$

where $T = \inf\{t > 0 : B_t \notin D\}$. (*Hint: let g be a map sending D to $\mathbb{D}$ and z_0 to 0. Let $\phi(z) = \frac{g(z)}{z-z_0}$ for $z \neq z_0$ and $\phi(z_0) = g'(z_0)$; and consider* $\log|\phi|$.)

(b) Deduce the following useful formula: let $D \subset \mathbb{C}$ be as above, let $U \subset D$ be a subdomain and for $z \in U$ let ρ_z be the harmonic measure on ∂U as seen from z. Then show that $\rho \in \mathcal{M}_0$ and that

$$\text{Var}(h, \rho) = \log \frac{R(z;D)}{R(z,U)}.$$

7. Show that the constraints in Remark 1.12 uniquely identify G^D when $d \geq 3$. For $x \in D$, defining $H_x(y) = (2\pi/A_d)|x-y|^{2-d}$, let h_x be the unique harmonic extension of $H_x|_{\partial D}$ into D. Show that the function $H(x,y) = H_x(y) - h_x(y)$, defined for $x \neq y \in D$, satisfies the constraints of Remark 1.12. Deduce that $G^D = H$. Show this directly by proving that the transition probability $p_t^D(x,y)$ solves the heat equation in D.

2 Liouville measure

In this chapter we introduce the Liouville measure, which (informally speaking) is the measure μ_γ defined by

$$\mu_\gamma(dz) = e^{\gamma h(z)} dz, \tag{2.1}$$

where h is a GFF in two dimensions and $\gamma \geq 0$ is a parameter. The construction will be generalised in Chapter 3 which is devoted to **Gaussian multiplicative chaos**, which are measures of the form (2.1) but for generic log-correlated fields Gaussian fields h. While the Gaussian free field in two dimensions is of course an example of such a field, so that Liouville measure really is just a particular case of the theory of Gaussian multiplicative chaos, some arguments specific to the GFF can be used to simplify the presentation and introduce relevant ideas in a clean way, without the need to introduce too much machinery. This is the reason why we have chosen to do the construction of Liouville measure (i.e., in the case of the GFF) in this separate chapter.

Heuristics. The informal definition (2.1) should be interpreted as follows. Some abstract Riemann surface has been parametrised, after Riemann uniformisation, by a domain of our choice – perhaps the disc, assuming that it has a boundary, or perhaps the unit sphere in three dimensions if it doesn't. In this parametrisation, the conformal structure is preserved: i.e., curves crossing at an angle θ at some point in the domain would also correspond to curves crossing at an angle θ in the original surface. However, in this parametrisation, the metric and the volume do not correspond to the ambient volume and metric of Euclidean space. Namely, a small element of volume dz in the domain really corresponds to a small element of volume $e^{\gamma h(z)} dz$ in the original surface. Hence points where h is *very big* (e.g., thick points) correspond in reality to relatively big portions of the surface; while points where h is very low are points which correspond to small portions of the surface. The first points will tend to be typical from the point of view of sampling from the volume measure, while the second points will be where geodesics tend to travel.

Rigourous approach. Let $D \subset \mathbb{R}^2$ be an open set and let h be a Dirichlet (or zero boundary) GFF on D. When we try to give a precise meaning (2.1), we immediately run into a serious problem: the exponential of a distribution (such as h) is not *a priori* defined. This corresponds to the fact that while h is regular enough to be a *distribution*, so small rough oscillations cancel each other when we average h over macroscopic regions of space, these oscillations become highly magnified when we take the exponential and they can no longer cancel each other out. In fact, giving a meaning to (2.1) will require non-trivial work, and will be done via an approximation procedure, using

$$\mu_\varepsilon(dz) := e^{\gamma h_\varepsilon(z)} \varepsilon^{\gamma^2/2} dz, \tag{2.2}$$

for $\varepsilon > 0$, where $h_\varepsilon(z)$ is a jointly continuous version of the circle average. (More general regularisations will be considered in Chapter 3). It is straightforward to see that μ_ε is a (random) Radon measure on D for every ε. Our goal will be to prove the following theorem.

Theorem 2.1. *Suppose $0 \leq \gamma < 2$. If D is bounded, then the random measure μ_ε converges weakly almost surely to a random measure μ, the* (bulk) Liouville measure*, along the subsequence $\varepsilon = 2^{-k}$. μ a.s. has no atoms, and for any $A \subset D$ open, we have $\mu(A) > 0$ a.s. In fact, $\mathbb{E}(\mu(A)) = \int_A R(z, D)^{\gamma^2/2} dz \in (0, \infty)$.*

We remind the reader that the notation $R(z, D)$ above stands for the conformal radius of D seen from z. That is, $R(z, D) = |f'(0)|$ where f is (any) conformal map taking $\mathbb{D}$ to D and 0 to z. If D is unbounded then weak convergence can be replaced by vague convergence with exactly the same proof.

In this form, the result is due to Duplantier and Sheffield [DS11]. It could also have been deduced from earlier work of Kahane [Kah85] who used a different approximation procedure, together with results of Robert and Vargas [RV10b] showing universality of the limit with respect to the approximating procedure. (In fact, these two results would have given convergence in distribution of μ_ε rather than in probability; and hence would not show that the limiting measure μ depends solely on the free field h. However, a strengthening of the arguments of Robert and Vargas due to Shamov [Sha16] has recently yielded convergence in probability.) Earlier, Høegh-Krohn [HK71] had introduced a similar model in the context of quantum field theory, and analysed it in the relatively easy L^2 phase when $0 \leq \gamma < \sqrt{2}$. Here we will follow the elementary approach developed in [Ber17], which works in the more general context of Gaussian multiplicative chaos (see Chapter 3), but with the simplifications that are allowed by taking the underlying field to be the GFF.

2.1 Preliminaries

Before we start the proof of Theorem 2.1 we first observe that this is the right normalisation.

Lemma 2.2. *We have that* $\mathrm{Var}(h_\varepsilon(x)) = \log(1/\varepsilon) + \log R(x, D)$*. As a consequence,* $\mathbb{E}(\mu_\varepsilon(A)) = \int_A R(z, D)^{\gamma^2/2} dz \in (0, \infty)$.

Proof. The proof is very similar to the argument in Theorem 1.35 and is a good exercise. Fix $x \in D$. By definition,

$$\mathrm{Var}(h_\varepsilon(x)) = \Gamma(\rho_{x,\varepsilon}) = \int \rho_{x,\varepsilon}(dz)\rho_{x,\varepsilon}(dw)G(z, w).$$

For a fixed z, $G(z, \cdot)$ is harmonic on $D \setminus \{z\}$ and so $\int \rho_{x,\varepsilon}(dw)G(w, z) = G(x, z)$ by the mean value property and an approximation argument similar to (1.23). Therefore,

$$\mathrm{Var}(h_\varepsilon(x)) = \Gamma(\rho_{x,\varepsilon}) = \int \rho_{x,\varepsilon}(dz)G(z, x).$$

Now, observe that $G(x, \cdot) = -\log|x - \cdot| + \xi(\cdot)$ where $\xi(\cdot)$ is harmonic and $\xi(x) = \log R(x; D)$. Indeed let $\xi(\cdot)$ be the harmonic extension of $-\log|x - \cdot|$ on ∂D. Then $G(x, \cdot) - \xi(\cdot)$ has zero boundary values on ∂D, and is bounded and harmonic in $D \setminus \{x\}$. Hence it must be zero in

all of D by uniqueness of solutions to the Dirichlet problem among bounded functions (e.g., by the optional stopping theorem). Note that $\xi(x) = \log R(x; D)$ by (1.4). Therefore, by harmonicity of ξ and the mean value property,

$$\mathrm{Var}(h_\varepsilon(x)) = \int G(x,z)\rho_{x,\varepsilon}(dz) = \log(1/\varepsilon) + \xi(x)$$

as desired. □

We now make a couple of remarks:

1. Not only is the expectation constant, but we have that for each fixed z, $e^{\gamma h_\varepsilon(z)}\varepsilon^{\gamma^2/2}$ forms a martingale as a function of ε. This is nothing but the exponential martingale of a Brownian motion.

2. However, the integral $\mu_\varepsilon(A)$ is **not** a martingale. This is because the underlying filtration in which $e^{\gamma h_\varepsilon(z)}\varepsilon^{\gamma^2/2}$ is a martingale depends on z. If we try to condition on all the information $h_\varepsilon(z), z \in D$, then this is too much information, and we lose the martingale property.

2.2 Convergence and uniform integrability in the L^2 phase

The bulk of the proof consists in showing that for any fixed bounded Borel subset S of D (including possibly D itself), we have that $\mu_\varepsilon(S)$ converges almost surely along the subsequence $\varepsilon = 2^{-k}$ to a non-degenerate limit. We will then explain in Section 2.3, using fairly general arguments, why this implies the a.s. weak convergence of the sequence of measures μ_ε along the same subsequence.

Let us now fix S and set $I_\varepsilon = \mu_\varepsilon(S)$. We first suppose that $\gamma \in [0, \sqrt{2})$. In this case, the so-called L^2 phase, it is relatively to check the convergence (which actually holds in L^2), but difficulties arise when $\gamma \in [\sqrt{2}, 2)$. (As luck would have it this coincides precisely with the phase which is interesting from the point of view of statistical physics on random planar maps).

Proposition 2.3. *If $\gamma \in [0, \sqrt{2})$ and $\varepsilon > 0$, $\delta = \varepsilon/2$, then we have the estimate $\mathbb{E}((I_\varepsilon - I_\delta)^2) \leq C\varepsilon^{2-\gamma^2}$. In particular, I_ε is a Cauchy sequence in $L^2(\mathbb{P})$ and so converges to a limit in probability as $\varepsilon \to 0$. Along the sequence $\varepsilon = 2^{-k}$, this convergence occurs almost surely, and the limit is almost surely strictly positive.*

Proof. For ease of notations, let $\bar{h}_\varepsilon(z) = \gamma h_\varepsilon(z) - (\gamma^2/2)\,\mathrm{Var}(h_\varepsilon(z))$, and let

$$\sigma(dz) = R(z, D)^{\gamma^2/2} dz.$$

The idea is to say that if we consider the Brownian motions $h_\varepsilon(x)$ and $h_\varepsilon(y)$ (viewed as a function of $\varepsilon = e^{-t}$, then they are (approximately) identical until $\varepsilon \leq |x - y|$, after which time they evolve (exactly) independently.

Observe that by Fubini's theorem,

$$\begin{aligned}\mathbb{E}((I_\varepsilon - I_\delta)^2) &= \int_{S^2} \mathbb{E}\left((e^{\bar{h}_\varepsilon(x)} - e^{\bar{h}_\delta(x)})(e^{\bar{h}_\varepsilon(y)} - e^{\bar{h}_\delta(y)})\right)\sigma(dx)\sigma(dy)\\ &= \int_{S^2} \mathbb{E}\left(e^{\bar{h}_\varepsilon(x)+\bar{h}_\varepsilon(y)}(1 - e^{\bar{h}_\delta(x)-\bar{h}_\varepsilon(x)})(1 - e^{\bar{h}_\delta(y)-\bar{h}_\varepsilon(y)})\right)\sigma(dx)\sigma(dy).\end{aligned}$$

By the Markov property, $h_\varepsilon(x) - h_\delta(x)$ and $h_\varepsilon(y) - h_\delta(y)$ are independent as soon as $|x-y| \geq 2\varepsilon$: indeed, we can apply the Markov property in $U = B(x,\varepsilon) \cup B(y,\varepsilon)$, which allows us to write $h = \tilde{h} + \varphi$ where φ is harmonic in U and $\tilde{h}$ is an independent GFF in U. Since U is a disjoint union of two balls in U, the restriction of $\tilde{h}$ to each of these balls is independent. Hence all three terms in the above product are independent: indeed, the first is measurable with respect to h outside of U (i.e., depends only on φ), the second term depends only on $\tilde{h}$ in the ball $B(x,\varepsilon)$, and the third term depends only on $\tilde{h}$ in $B(y,\varepsilon)$.

Hence if $|x-y| \geq 2\varepsilon$, then the expectation in the above integral is simply

$$= \mathbb{E}(e^{\bar{h}_\varepsilon(x)+\bar{h}_\varepsilon(y)})\mathbb{E}(1 - e^{\bar{h}_\delta(x)-\bar{h}_\varepsilon(x)})\mathbb{E}(1 - e^{\bar{h}_\delta(y)-\bar{h}_\varepsilon(y)})$$

But both second and third terms are equal to zero, because of the pointwise martingale property. Therefore the expectation is just 0 as soon as $|x-y| > 2\varepsilon$.

Also note that by the martingale property for a fixed x,

$$\begin{aligned}\mathbb{E}((e^{\bar{h}_\varepsilon(x)} - e^{\bar{h}_\delta(x)})^2) &= \mathbb{E}(e^{2\bar{h}_\delta(x)} - e^{2\bar{h}_\varepsilon(x)})\\ &\leq \mathbb{E}(e^{2\bar{h}_\delta(x)}) \leq C\mathbb{E}(e^{2\bar{h}_\varepsilon(x)})\end{aligned}$$

for some $C > 0$. Hence using Cauchy–Schwarz in the case where $|x-y| \leq 2\varepsilon$,

$$\begin{aligned}\mathbb{E}((I_\varepsilon - I_\delta)^2) &\leq \int_{|x-y|\leq 2\varepsilon} \sqrt{\mathbb{E}((e^{\bar{h}_\varepsilon(x)} - e^{\bar{h}_\delta(x)})^2)\mathbb{E}((e^{\bar{h}_\varepsilon(y)} - e^{\bar{h}_\delta(y)})^2)}\sigma(dx)\sigma(dy)\\ &\leq C\int_{|x-y|\leq 2\varepsilon} \sqrt{\mathbb{E}(e^{2\bar{h}_\varepsilon(x)})\mathbb{E}(e^{2\bar{h}_\varepsilon(y)})}\sigma(dx)\sigma(dy) \qquad (2.3)\\ &\leq C\int_{|x-y|\leq 2\varepsilon} \varepsilon^{\gamma^2} e^{\frac{1}{2}(2\gamma)^2\log(1/\varepsilon)}\sigma(dx)\sigma(dy)\\ &\leq C\varepsilon^{2+\gamma^2-2\gamma^2} = C\varepsilon^{2-\gamma^2}.\end{aligned}$$

To see that $\mathbb{P}(\lim_{\varepsilon\to 0} I_\varepsilon > 0) = 1$ we will appeal to Kolmogorov's 0-1 law. We already know that $\mathbb{P}(\lim_{\varepsilon\to 0} I_\varepsilon > 0) > 0$, since $\mathbb{E}(\lim_{\varepsilon\to 0} I_\varepsilon) = \lim_{\varepsilon\to 0}\mathbb{E}(I_\varepsilon) > 0$. Moreover, notice that if $(f_i)_{i\geq 1}$ is an orthonormal basis of $H_0^1(D)$, then $\{h_\varepsilon(x) : x \in D, \varepsilon > 0\}$ and therefore $\lim_{\varepsilon\to 0} I_\varepsilon$, is a function of the sequence of coefficients $(h, f_i)_\nabla$. Now, we have seen that these coefficients are independent standard Gaussians, and it is clear that the event $\{\lim_{\varepsilon\to 0} I_\varepsilon > 0\}$ is in the tail σ-algebra generated by the sequence (since this event is invariant under resampling any finite number of terms). Thus it has probability zero or one, and since the probability is positive, it must be one. This concludes the proof of the proposition. □

The moral of this proof is the following: while I_ε is not a martingale in ε (because there is no filtration common to all points x such that $e^{\bar{h}_\varepsilon(x)}$ forms a martingale), we can use the pointwise martingales to estimate the second moment of the increment $I_\varepsilon - I_\delta$. Only for points x, y which are very close (of order ε) do we get a non-trivial contribution.

We defer the proof of the general case $\gamma \in [0, 2)$ until a bit later (see Section 2.5); and for now show how convergence of masses $\mu_\varepsilon(S)$ towards some limit implies a.s. weak convergence of the sequence of measures μ_ε.

2.3 Weak convergence to Liouville measure

We now finish the proof of Theorem 2.1 (assuming convergence of masses of fixed bounded Borel subsets $S \subseteq D$ toward some limit that is strictly positive with probability one) by showing that the sequence of measures μ_ε converges in probability for the weak topology towards a measure μ. This measure will be defined by the limits of quantities of the form $\mu_\varepsilon(S)$, where S is a cube such that $\bar{S} \subset D$. These arguments are borrowed from [Ber17].

Note that since $\mu_\varepsilon(D)$ converges a.s. we have that the measures μ_ε are a.s. tight in the space of Borel measures on D with the topology of weak convergence (along the subsequence $\varepsilon = 2^{-k}$, which we will not repeat). Let $\tilde{\mu}$ be any weak limit.

Let $\mathcal{A}$ denote the π-system of subsets of $\mathbb{R}^2$ of the form $A = [x_1, y_1) \times [x_2, y_2)$ where $x_i, y_i \in \mathbb{Q}, i = 1, 2$ and such that $\bar{A} \subset D$, and note that the σ-algebra generated by $\mathcal{A}$ is the Borel σ-field on D. Observe that $\mu_\varepsilon(A)$ converges almost surely to a limit (which we call $\mu(A)$) for any $A \in \mathcal{A}$, by the part of the theorem which is already proved (or assumed in the case $\gamma \geq \sqrt{2}$). Observe that this convergence holds a.s. simultaneously for all $A \in \mathcal{A}$, since $\mathcal{A}$ is countable.

Let $A = [x_1, y_1) \times [x_2, y_2) \in \mathcal{A}$. We first claim that

$$\mu(A) = \sup_{x_i', y_i'} \{\mu([x_1', y_1'] \times [x_2', y_2'])\} \tag{2.4}$$

where the sup is over all $x_i', y_i' \in \mathbb{Q}$ with $x_i' > x_i$ and $y_i' < y_i, 1 \leq i \leq 2$. Indeed, clearly the left hand side is a.s. greater or equal to the right hand side, but both sides have the same expectation by monotone convergence (for $\mathbb{E}$). Likewise, it is easy to check that

$$\mu(A) = \inf_{x_i', y_i'} \{\mu((x_1', y_1') \times (x_2', y_2'))\} \tag{2.5}$$

where now the inf is over all $x_i', y_i' \in \mathbb{Q}$ with $x_i' < x_i$ and $y_i' > y_i, 1 \leq i \leq 2$.

We aim to check that $\tilde{\mu}(A) = \mu(A)$, which uniquely identifies the weak limit $\tilde{\mu}$ and hence proves the desired weak convergence.

Note that by the portmanteau lemma, for any $A = [x_1, y_1) \times [x_2, y_2)$, and for any $x_i', y_i' \in \mathbb{Q}$ with $x_i' < x_i$ and $y_i' > y_i, 1 \leq i \leq 2$, we have:

$$\begin{aligned} \tilde{\mu}(A) &\leq \tilde{\mu}((x_1', y_1') \times (x_2', y_2')) \\ &\leq \liminf_{\varepsilon \to 0} \mu_\varepsilon((x_1', y_1') \times (x_2', y_2')) \end{aligned}$$

$$= \mu((x_1', y_1') \times (x_2', y_2')).$$

(The portmanteau lemma is more classically stated for probability measures, but there is no problem in using the theorem here since we already know convergence of the total mass, so we can equivalently work with the normalised measures $\mu_\varepsilon/\mu_\varepsilon(D)$).

Since the x_i', y_i' are arbitrary, taking the inf over the admissible values and using (2.5) we get

$$\tilde{\mu}(A) \leq \mu(A).$$

The converse inequality follows in the same manner, using (2.4). We deduce that $\tilde{\mu}(A) = \mu(A)$, almost surely, as desired. As already explained, this uniquely identifies the limit $\tilde{\mu}$. Hence μ_ε converges a.s. weakly to μ on D. □

2.4 The GFF viewed from a Liouville typical point

Let h be a Gaussian free field on a domain D, with associated Liouville measure μ for some $\gamma < 2$. An interesting question is the following: if z is a random point sampled according to the Liouville measure, normalised to be a probability distribution (this is possible when D is bounded), then what does h look like near the point z? This gives rise to the concept of *rooted measure* in the terminology of [DS11] or to the Peyrière measure in the terminology of Gaussian multiplicative chaos.

We expect some atypical behaviour: after all, for any given fixed $z \in D$, $e^{\gamma h_\varepsilon(z)}\varepsilon^{\gamma^2/2}$ converges a.s. to 0, so the only reason μ could be non-trivial is if there are enough points on which h is atypically big. Of course this leads us to suspect that μ is in some sense carried by certain thick points of the GFF. It remains to identify the level of thickness. As mentioned before, a simple back-of-the-envelope calculation (made slightly more rigorous in the next result) suggests that these points should be γ-thick. As we will see, this in fact a simple consequence of Girsanov's lemma: essentially, when we bias h by $e^{\gamma h(z)}$, we shift the mean value of the field by $\gamma G(\cdot, z) = \gamma \log 1/|\cdot - z| + O(1)$, thereby resulting in a γ-thick point.

Theorem 2.4. *Suppose D is bounded. Let z be a point sampled according to the Liouville measure μ, normalised to be a probability measure. Then, a.s.,*

$$\lim_{\varepsilon \to 0} \frac{h_\varepsilon(z)}{\log(1/\varepsilon)} = \gamma.$$

In other words, z is almost surely a γ-thick point ($z \in \mathcal{T}_\gamma$).

When D is not bounded we can simply say that $\mu(\mathcal{T}_\gamma^c) = 0$, almost surely. In particular, μ is singular with respect to Lebesgue measure, a.s.

Proof. The proof is elegant and simple, but the first time one sees it, it is somewhat perturbing. We require the following important but elementary lemma, which can be seen as a (completely elementary) version of Girsanov's theorem.

Lemma 2.5 (Tilting lemma / Girsanov / Cameron–Martin)**.** *Let $X = (X_1, \dots, X_n)$ be a Gaussian vector under the law $\mathbb{P}$, with mean μ and covariance matrix V. Let $\alpha \in \mathbb{R}^n$ and define a new probability measure by*

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{e^{\langle \alpha, X\rangle}}{Z},$$

where $Z = \mathbb{E}(e^{\langle \alpha, X\rangle})$ is a normalising constant. Then under $\mathbb{Q}$, X is still a Gaussian vector, with covariance matrix V and mean $\mu + V\alpha$.

It is worth rephrasing this lemma in plain words. Suppose we weigh the law of a Gaussian vector by some linear functional. Then the process remains Gaussian, with unchanged covariances, however the mean is shifted, and the new mean of the variable X_i say, is

$$\mu_i' = \mu_i + \mathrm{Cov}(X_i, \langle \alpha, X\rangle).$$

In other words, the mean is shifted by an amount which is simply the covariance of the quantity we are considering and what we are weighting by.

Proof. Assume for simplicity (and in fact without loss of generality) that $\mu = 0$. It is simple to check it with Laplace transforms: indeed if $\lambda \in \mathbb{R}^n$, then

$$\begin{aligned}\mathbb{Q}(e^{\langle \lambda, X\rangle}) &= \frac{1}{Z}\mathbb{E}(e^{\langle \lambda+\alpha, X\rangle}) \\ &= \frac{1}{e^{\frac{1}{2}\langle \alpha, V\alpha\rangle}} e^{\frac{1}{2}\langle \alpha+\lambda, V(\alpha+\lambda)\rangle} \\ &= e^{\frac{1}{2}\langle \lambda, V\lambda\rangle + \langle \lambda, V\alpha\rangle}\end{aligned}$$

The first term in the exponent $\langle \lambda, V\lambda\rangle$ is the Gaussian term with variance V, while the second term $\langle \lambda, V\alpha\rangle$ shows that the mean is now $V\alpha$, as desired. □

Let $\mathbb{P} = \mathbb{P}(dh)$ be the law of the GFF, and let Q_ε denote the joint law on (z, h) defined by:

$$Q_\varepsilon(dz, dh) = \frac{1}{Z} e^{\gamma h_\varepsilon(z)} \varepsilon^{\gamma^2/2} dz \mathbb{P}(dh).$$

Here Z is a normalising (non-random) constant depending solely on ε. Note that the marginal law of h is weighted by $\mu_\varepsilon(D)$ under Q_ε, and given h, the point z is sampled proportionally to μ_ε.

Define also $Q(dz, dh) = \mu_h(dz)\mathbb{P}(dh)$ where by μ_h we mean the Liouville measure which is a.s. defined by h. Note that Q_ε converges to Q weakly with respect to the product topology induced by the Euclidean metric for z and the Sobolev H^{-1} norm for h, say, or, if we prefer the point of view that h is a stochastic process indexed by $\mathcal{M}$, then the meaning of this convergence is with respect to the infinite product $D \times \mathbb{R}^{\mathcal{M}}$: that is, for any fixed $m \geq 1$ and $\rho_1, \dots, \rho_m \in \mathcal{M}$, and any continuous bounded function f on D,

$$\mathbb{E}\left((h, \rho_1)\dots(h, \rho_m)\int f(z)\varepsilon^{\gamma^2/2} e^{\gamma h_\varepsilon(z)} dz\right) \to \mathbb{E}\left((h, \rho_1)\dots(h, \rho_m)\int f(z)\mu_h(dz)\right).$$

This can be verified exactly with the same argument which shows the weak convergence of the approximate Liouville measures. For simplicity we will keep the point of view of a stochastic process for the rest of the proof.

Note that under the law Q_ε, the marginal law of h is simply

$$Q_\varepsilon(dh) = \frac{1}{Z}\mu_\varepsilon(D)\mathbb{P}(dh) \tag{2.6}$$

so it has the law of a GFF biased by its total mass, and we deduce that $Z = \mathbb{E}(\mu_\varepsilon(D)) = \int_D R(z,D)^{\gamma^2/2}dz$ does not even depend on ε (in fact there are small effects from the boundary which we freely ignore).

Furthermore, the marginal law of z is

$$Q_\varepsilon(dz) = \frac{1}{Z}dz\mathbb{E}(e^{\gamma h_\varepsilon(z)}\varepsilon^{\gamma^2/2}) = \frac{dz}{Z}R(z,D)^{\gamma^2/2}.$$

Here again, the law does not depend on ε and is nice, i.e. absolutely continuous, with respect to Lebesgue measure. Finally, it is clear that under Q_ε, given h, the conditional law of z is just given by a sample from μ_ε.

We will simply reverse the procedure, and focus instead on the *conditional distribution of* h given z. We start by explaining the argument without worrying about its formal justification, and add the justifications where needed afterwards.

Note that by definition,

$$Q_\varepsilon(dh|z) = \frac{1}{Z(z)}e^{\gamma h_\varepsilon(z)}\varepsilon^{\gamma^2/2}\mathbb{P}(dh).$$

In other words, the law of the Gaussian field h has been reweighted by an exponential linear functional. By Girsanov's lemma, we deduce that under $Q_\varepsilon(dh|z)$, h is a field with the same covariances as under $\mathbb{P}$, and *non-zero* mean at point w given by

$$\mathrm{Cov}(h(w), \gamma h_\varepsilon(z)) = \gamma\log(1/|w-z|) + O(1).$$

(More rigorously, we apply Girsanov's lemma to the Gaussian stochastic process $(h,\rho)_{\rho\in\mathcal{M}}$ and find that under Q_ε, its covariance structure remains unchanged, while its mean has been shifted by $\mathrm{Cov}((h,\rho);\gamma h_\varepsilon(z))$.)

In the limit as $\varepsilon\to 0$, this amounts to adding the function $\gamma G_D(\cdot,z)$ to the field $h(\cdot)$. We now argue that this must coincide with the law of $Q(dh|z)$. To see this, we use the previous paragraph to write for any $\varepsilon>0$, and for any $m\geq 1$, $\rho_1,\dots,\rho_m\in\mathcal{M}$, $\psi\in C_b(D)$:

$$\mathbb{E}_{Q^\varepsilon}((h,\rho_1)\dots(h,\rho_m)\psi(z))$$
$$= \int_D dz\psi(z)R(z,D)^{\frac{\gamma^2}{2}}\mathbb{E}_h\left(((h,\rho_1)+\mathrm{Cov}((h,\rho_1),\gamma h_\varepsilon(z)))\dots((h,\rho_n)+\mathrm{Cov}((h,\rho_m),\gamma h_\varepsilon(z)))\right).$$

Invoking the weak convergence of Q_ε to Q, we see that the left hand side of the above equality converges to $\mathbb{E}_Q((h,\rho_1)\dots(h,\rho_m)\psi(z))$ as $\varepsilon\to 0$. At the same time, an application of the dominated convergence theorem shows that the right hand side converges as $\varepsilon\to 0$ to

$$\int_D dz\psi(z)R(z,D)^{\frac{\gamma^2}{2}}\mathbb{E}_h((h+\gamma G(z,\cdot),\rho_1)\dots(h+\gamma G(z,\cdot),\rho_m)).$$

Hence the law of $Q(dh|z)$ is as claimed.

To summarise, under Q and given z, a logarithmic singularity of strength γ has been introduced at the point z. Hence we find that under $Q(dh|z)$, a.s.,

$$\lim_{\delta\to 0}\frac{h_\delta(z)}{\log(1/\delta)}=\gamma,$$

so $z\in\mathcal{T}_\gamma$, a.s. as desired. In other words, $Q(\mu_h(\mathcal{T}_\gamma^c)=0)=1$.

We conclude the proof of the theorem by observing that the marginal laws $Q(dh)$ and $\mathbb{P}(dh)$ are mutually absolutely continuous with respect to one another, so any property which holds a.s. under Q holds also a.s. under $\mathbb{P}$. (This absolute continuity follows simply from the fact that $\mu(S)\in(0,\infty),\mathbb{P}-$a.s.) □

2.5 General case

To address the difficulties that arise when $\gamma\geq\sqrt{2}$, we proceed as follows. Roughly, we claim that the second moment of I_ε blows up because of rare points which are *too thick* and which do not contribute to the integral in an a.s. sense, but nevertheless inflate the value of the second moment. So we will remove these points by hand. To see which points to remove, we appeal the considerations of the previous section: this suggests that we should be safe to get rid of points that are strictly more than γ-thick.

Let $\alpha>0$ be fixed (it will be chosen $>\gamma$ and very close to γ soon). We define a good event $G_\varepsilon^\alpha(x)=\{h_\varepsilon(x)\leq\alpha\log(1/\varepsilon)\}$. This is the good event that the point x is not too thick at scale ε.

Lemma 2.6 (Liouville points are no more than γ-thick)**.** *For $\alpha>\gamma$ we have*

$$\mathbb{E}(e^{\bar{h}_\varepsilon(x)}1_{G_\varepsilon^\alpha(x)})\geq 1-p(\varepsilon)$$

where the function p may depend on α and for a fixed $\alpha>\gamma$, $p(\varepsilon)\to 0$ as $\varepsilon\to 0$, polynomially fast. The same estimate holds if $\bar{h}_\varepsilon(x)$ is replaced with $\bar{h}_{\varepsilon/2}(x)$.

Proof. Note that

$$\mathbb{E}(e^{\bar{h}_\varepsilon(x)}\mathbf{1}_{\{G_\varepsilon^\alpha(x)\}})=\tilde{\mathbb{P}}(G_\varepsilon^\alpha(x)),\ \text{where}\ \frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}=e^{\bar{h}_\varepsilon(x)}.$$

By Girsanov's lemma, under $\tilde{\mathbb{P}}$, the process $X_s=h_{e^{-s}}(x)$ has the same covariance structure as under $\mathbb{P}$ and its mean is now $\gamma\operatorname{Cov}(X_s,X_t)=\gamma s+O(1)$ for $s\leq t$. Hence it is a Brownian motion with drift γ, and the lemma follows from the fact that such a process does not exceed αt at time t with high probability when t is large (and the error probability is exponential in t, or polynomial in ε, as desired).

Changing ε into $\varepsilon/2$ means that the drift of X_s is $\gamma s+O(1)$ over a slightly larger interval of time, namely until time $t+\log 2$. In particular the same argument as above shows that the same estimate holds for $\bar{h}_{\varepsilon/2}(x)$ as well. □

We therefore see that points which are more than γ-thick do not contribute significantly to I_ε in expectation and can be safely removed. To this end, we fix $\alpha > \gamma$ and introduce:

$$J_\varepsilon = \int_S e^{\bar{h}_\varepsilon(x)} \mathbf{1}_{G_\varepsilon(x)} \sigma(dx); \quad J'_{\varepsilon/2}(x) = \int_S e^{\bar{h}_{\varepsilon/2}(x)} \mathbf{1}_{G_\varepsilon(x)} \sigma(dx) \tag{2.7}$$

with $G_\varepsilon(x) = G_\varepsilon^\alpha(x)$, and where we recall that $\sigma(x) = R(x,D)^{\gamma^2/2}dx$. Note that a consequence of Lemma 3.5 is that

$$\mathbb{E}(|I_\varepsilon - J_\varepsilon|) \le p(\varepsilon)|S| \to 0 \text{ and } \mathbb{E}(|I_{\varepsilon/2} - J'_{\varepsilon/2}|) \le p(\varepsilon)|S| \to 0 \tag{2.8}$$

as $\varepsilon \to 0$.

Lemma 2.7. *We have the estimate $\mathbb{E}((J_\varepsilon - J'_{\varepsilon/2})^2) \le \varepsilon^r$ for some $r > 0$. In particular, I_ε is a Cauchy sequence in L^1 and so converges to a limit in probability. Along $\varepsilon = 2^{-k}$, this convergence occurs almost surely.*

Proof. The proof of this lemma is virtually identical to that in the L^2 phase (see Proposition 2.3). The key observation there was that if $|x-y| \ge 2\varepsilon$, then the increments $h_\varepsilon(x) - h_{\varepsilon/2}(x)$ and $h_\varepsilon(y) - h_{\varepsilon/2}(y)$ are independent of each other, and in fact also of $\mathcal{F}$: the σ-algebra generated by h restricted to the complement of $B(x,\varepsilon) \cup B(y,\varepsilon)$. Since the events $G_\varepsilon(x)$ and $G_\varepsilon(y)$ are both measurable with respect to $\mathcal{F}$, we may therefore deduce from *that* proof (see (2.3)) that

$$\mathbb{E}((J_\varepsilon - J'_{\varepsilon/2})^2) \le C \int_{|x-y|\le 2\varepsilon} \sqrt{\mathbb{E}(e^{2\bar{h}_\varepsilon(x)} \mathbf{1}_{G_\varepsilon(x)}) \mathbb{E}(e^{2\bar{h}_\varepsilon(y)} \mathbf{1}_{G_\varepsilon(y)})} \sigma(dx)\sigma(dy).$$

Now,

$$\begin{aligned} \mathbb{E}(e^{2\bar{h}_\varepsilon(x)} \mathbf{1}_{G_\varepsilon(x)}) &\le \mathbb{E}(e^{2\bar{h}_\varepsilon(x)} \mathbf{1}_{\{h_\varepsilon(x) \le \alpha \log(1/\varepsilon)\}}) \\ &\le O(1)\varepsilon^{-\gamma^2} \mathbb{Q}(h_\varepsilon(x) \le \alpha \log 1/\varepsilon) \end{aligned}$$

where by Girsanov's lemma, under $\mathbb{Q}$, $h_\varepsilon(x)$ is a normal random variable with mean $2\gamma \log(1/\varepsilon) + O(1)$ and variance $\log 1/\varepsilon + O(1)$. This means that

$$\mathbb{Q}(h_\varepsilon(x) \le \alpha \log 1/\varepsilon) \le O(1) \exp(-\frac{1}{2}(2\gamma - \alpha)^2 \log 1/\varepsilon)$$

and hence

$$\mathbb{E}((J_\varepsilon - J'_{\varepsilon/2})^2) \le O(1)\varepsilon^{2-\gamma^2} \varepsilon^{\frac{1}{2}(2\gamma-\alpha)^2}.$$

Again, choosing $\alpha > \gamma$ sufficiently close to γ ensures that the bound on the right hand side is at most $O(1)\varepsilon^r$ for some $r > 0$, as desired. This finishes the proof of the lemma. It also concludes the proof of Theorem 2.1 in the general case $\gamma < 2$, by (2.8), and recalling that $p(\varepsilon)$ decays polynomially in ε for fixed α, so we can apply Borel–Cantelli to get a.s. convergence along the sequence $\varepsilon = 2^{-k}$. □

We note that the a.s. convergence over the entire range of ε (not just the dyadic values $\varepsilon = 2^{-k}$) was proved by Sheffield and Wang [SW16].

2.6 The phase transition for the Liouville measure

The fact that the Liouville measure $\mu = \mu_\gamma$ is supported on the γ-thick points, $\mathcal{T}_\gamma$, is very helpful to get a clearer picture what changes when $\gamma = 2$. Indeed recall that $\dim(\mathcal{T}_\gamma) = (2 - \gamma^2/2)_+$, and $\mathcal{T}_\gamma$ is empty if $\gamma > 2$. The point is that $\mu = \mu_\gamma$ does not degenerate when $\gamma < 2$ *because* there are thick points to support it. Once $\gamma > 2$ there are no longer any thick points, and this makes it in some sense "clear" that any approximations to μ_γ must degenerate to the zero measure. When $\gamma = 2$ however, $\mathcal{T}_\gamma$ is not empty, and there is therefore a hope to construct a meaningful critical Liouville measure μ. Such a construction has indeed been carried out in two separate papers by Duplantier, Rhodes, Sheffield, and Vargas [DRSV14b, DRSV14a]. However the normalisation must be done more carefully – see these two papers for details, as well as the more recent preprints [JS17, HRV18, Pow18].

2.7 Conformal covariance

Of course, it is natural to wonder in what way the conformal invariance of the GFF manifests itself at the level of the Liouville measure. As it turns out these measures are not simply conformally invariant. This is easy to believe intuitively, since the total mass of the Liouville measure has to do with total surface area (measured in quantum terms) enclosed in a domain (for example via circle packing), and so this must grow as the domain grows.

However, the measures are **conformally covariant**: that is, to relate their laws under conformal mappings one must include a correction term accounting for the inflation of the domain under the conformal map. This term is naturally proportional to the derivative of the conformal map.

To formulate the result, it is convenient to use the following notation. Suppose that h is a given distribution – perhaps a realisation of a GFF, but also perhaps one of its close relatives (for example, the GFF plus some smooth deterministic function) – and suppose that its circle average process is well defined. Then we define μ_h to be the measure, if it exists, given by $\mu_h(dz) = \lim_{\varepsilon \to 0} e^{\gamma h_\varepsilon(z)} \varepsilon^{\gamma^2/2} dz$. Of course, if h is just a GFF, then μ_h is nothing else but the measure we have constructed in the previous part. If h can be written as $h = h_0 + \varphi$ where φ is deterministic, h_0 is a GFF and $\mathrm{e}^{\gamma\varphi} \in L^1(\mu_{h_0})$, then $\mu_h(dz) = e^{\gamma\varphi(z)} \cdot \mu_{h_0}(dz)$ is absolutely continuous with respect to the Liouville measure μ_{h_0}.

Theorem 2.8 (Conformal covariance of Liouville measure)**.** *Let $f : D \to D'$ be a conformal map, and let h be a GFF in D. Then $h' = h \circ f^{-1}$ (where we define this image in the sense of distributions) is a GFF in D', and*

$$\begin{aligned} \mu_h \circ f^{-1} &= \mu_{h \circ f^{-1} + Q \log |(f^{-1})'|} \\ &= e^{\gamma Q \log |(f^{-1})'|} \mu_{h'}, \end{aligned}$$

where

$$Q = \frac{\gamma}{2} + \frac{2}{\gamma}.$$

In other words, pushing forward the Liouville measure μ_h by the map f, we get a measure which is absolutely continuous (with density $|f^{-1}(z)|^{\gamma Q}$ at $z \in D'$) with respect to the Liouville measure on D'.

Informal proof. The reason for this formula may be understood quite easily. Indeed, note that $\gamma Q = \gamma^2/2 + 2$. When we use the map f, a small circle of radius ε is mapped *approximately* into a small circle of radius $\varepsilon' = |f'(z)|\varepsilon$ around $f(z)$. So $e^{\gamma h_\varepsilon(z)}\varepsilon^{\gamma^2/2}dz$ approximately corresponds to

$$e^{\gamma h'_{|f'(z)|\varepsilon}(z')}\varepsilon^{\gamma^2/2}\frac{dz'}{|f'(z)|^2}$$

by the usual change of variable formula. This can be rewritten as

$$e^{\gamma h'_{\varepsilon'}(z')}(\varepsilon')^{\gamma^2/2}\frac{dz'}{|f'(z)|^{2+\gamma^2/2}}$$

Letting $\varepsilon \to 0$ we get, at least heuristically speaking, the desired result. □

Proof of Theorem 2.8. Of course, the above heuristic is far from a proof, and the main reason is that $h_\varepsilon(z)$ is not a very well-behaved approximation of h under conformal maps. It is better to instead work with a different approximation of the GFF, using an orthonormal basis of $H_0^1(D)$ as in Section 1.6, which has the advantage of being conformally invariant.

In view of this, we make the following definition: suppose $h = \sum_n X_n f_n$, where X_n are i.i.d. standard normal random variables, and f_n is an orthonormal basis of $H_0^1(D)$. Set $h^N(z) = \sum_{i=1}^N X_i f_i$ to be the truncated series, and define

$$\mu^N(S) = \int_S \exp\left(\gamma h^N(z) - \frac{\gamma^2}{2}\operatorname{Var}(h^N(z))\right)\sigma(dz)$$

where we recall that $\sigma(dz) = R(z, D)^{\gamma^2/2}dz$. Note that $\mu^N(S)$ has the same expected value as $\mu(S)$. Furthermore, $\mu^N(S)$ is a non-negative martingale with respect to the filtration $(\mathcal{F}_N)_N$ generated by $(X_N)_N$, so has an almost sure limit which we will call $\mu^*(S)$.

Lemma 2.9. *Almost surely,* $\mu^*(S) = \mu(S)$.

Proof. When we take the circle averages of the series we obtain

$$h_\varepsilon = h_\varepsilon^N + h'_\varepsilon$$

where h'_ε is independent from h^N, and h_ε^N denotes the circle average of the function h^N. Hence

$$\varepsilon^{\gamma^2/2}e^{\gamma h_\varepsilon(z)} = e^{\gamma h_\varepsilon^N(z)}\varepsilon^{\gamma^2/2}e^{\gamma h'_\varepsilon(z)}.$$

Consequently, integrating over S and taking the conditional expectation given $\mathcal{F}_N$, we obtain that

$$\mathbb{E}(\mu_\varepsilon(S)|\mathcal{F}_N) = \mu_\varepsilon^N(S) := \int_S \exp\left(\gamma h_\varepsilon^N(z) - \frac{\gamma^2}{2}\operatorname{Var}(h^N(z))\right)\sigma(dz).$$

When $\varepsilon \to 0$, the right hand side converges to $\mu^N(S)$, since h^N is a nice smooth function. Consequently,

$$\mu^N(S) = \lim_{\varepsilon \to 0} \mathbb{E}(\mu_\varepsilon(S)|\mathcal{F}_N).$$

Since $\mu_\varepsilon(S) \to \mu(S)$ in L^1, we have $\mu^N = \lim_{\varepsilon \to 0} \mathbb{E}(\mu_\varepsilon(S)|\mathcal{F}_N) = \mathbb{E}(\mu(S)|\mathcal{F}_N)$ and so by martingale convergence, $\mu^N(S) \to \mu(S)$ as $N \to \infty$. Hence $\mu(S) = \mu^*(S)$, as desired. □

To finish the proof of conformal covariance (Theorem 2.8) we now simply recall that if f_n is an orthonormal basis of $H_0^1(D)$ then $f_n \circ f^{-1}$ gives an orthonormal basis of $H_0^1(D')$. Hence if $h' = h \circ f^{-1}$, then its truncated series h'_N can also simply be written as $h'_N = h^N \circ f^{-1}$. Thus, consider the measure μ^N and apply the map f. We obtain a measure $\tilde{\mu}'_N$ in D' such that

$$\begin{aligned}\tilde{\mu}'_N(D') &= \int_{D'} \exp\{\gamma h^N(f^{-1}(z')) - \frac{\gamma^2}{2}\operatorname{Var}(h^N(f^{-1}(z')))\} R(f^{-1}(z'), D)^{\gamma^2/2} \frac{dz'}{|f'(f^{-1}(z'))|^2} \\ &= \int_{D'} d\mu'_N(z') e^{(2+\gamma^2/2)|(f^{-1})'(z')|},\end{aligned}$$

where $d\mu'_N$ is the approximating measure to $\mu_{h'}$ in D'. (The second identity is justified by properties of the conformal radius). Letting $N \to \infty$, and recalling that $d\mu'_N$ converges to $d\mu_{h'}$ by the previous lemma, we obtain the desired statement of conformal covariance. This finishes the proof of Theorem 2.8. □

2.8 Random surfaces

The notion of **random surface** is a way of identifying Gaussian free field type distributions that give rise to different "parametrisations" of the same Liouville measure. Essentially, we want to consider the surfaces encoded by μ_h and by $\mu_h \circ f^{-1}$ to be "the same" for any given conformal map $f : D \to D'$. By the conformal covariance formula (Theorem 2.8) if h is a GFF, we have $\mu_h \circ f^{-1} = \mu_{h'}$ a.s., where $h' = h \circ f^{-1} + Q \log |(f^{-1})'|$. Thus we should think of h and h' as encoding the same quantum surface.

In fact, (when h is a GFF) this equality holds almost surely for *all* D' and **all** conformal maps $f : D \to D'$ simultaneously. This result was proved by Sheffield and Wang in [SW16].

This motivates the following definition, due to Duplantier and Sheffield [DS11]. Define an equivalence relation between (D_1, h_1) and (D_2, h_2), where (D_i, h_i) consists of a simply connected domain and an element of $\mathcal{D}'(D_i)$ for $i = 1, 2$, if there exists $f : D_1 \to D_2$ a conformal map such that

$$h_2 = h_1 \circ f^{-1} + Q \log |(f^{-1})'|.$$

It is easy to see that this is an equivalence relation.

Definition 2.10. *A (random) surface is a pair (D, h) consisting of a domain and a (random) distribution $h \in \mathcal{D}'(D)$, where the pair is considered modulo the above equivalence relation.*

Observe that this definition of (random) surface depends on the parameter $\gamma \geq 0$ of the Liouville measure (since Q depends on γ).

Interesting random surfaces arise, among other things, when we sample a point according to the Liouville measure (either in the bulk, or on the boundary for a free field with a non-trivial boundary behaviour, see later), and we 'zoom in' near this point. Roughly speaking, these are the *quantum cones* and *quantum wedges* introduced by Sheffield in [She16a]. A particular kind of wedge will be studied in a fair amount of detail later on in these notes (see Theorem 5.53).

2.9 Exercises

1. Explain why Lemma 3.7 and Lemma 3.5 imply uniform integrability of $\mu_\varepsilon(S)$.

2. Let μ be the Liouville measure with parameter $0 \leq \gamma < 2$. Use uniform integrability and the Markov property of the GFF to show that $\mu(S) > 0$ a.s.

3. How would you normalise $e^{\gamma h_\varepsilon(z)}$ if you are just aiming to define the Liouville measure on some line segment contained in D? Show that with this normalisation you get a non-degenerate limit. What is the conformal covariance in this case?

4. Recall the events $G_\varepsilon(z) = \{h_r(z) \leq \alpha \log 1/r, \text{ for all } r \in [\varepsilon_0, \varepsilon]\}$ from the proof of uniform integrability of the Liouville measure in the general case. Show that for any $0 \leq d < 2 - \gamma^2/2 < 2$, then

$$\mathbb{E}\left(\int_{S^2} \frac{1}{|x-y|^d} e^{\bar{h}_\varepsilon(x)} \mathbf{1}_{G_\varepsilon(x)} \sigma(dx)\ e^{\bar{h}_\varepsilon(y)} \mathbf{1}_{G_\varepsilon(y)} \sigma(dy)\right) \leq C < \infty$$

   where C does not depend on ε. Deduce that

$$\dim(\mathcal{T}_\gamma) \geq 2 - \gamma^2/2.$$

   Conclude with a proof of Theorem 1.37.

3 Properties of Gaussian multiplicative chaos

In this chapter we take a slightly more general point of view than in the previous one, and construct Gaussian multiplicative chaos (i.e., exponential measures) for a wider class of logarithmically correlated fields. The construction is slightly more involved than the one presented in previous chapter for the Gaussian free field, since we can no longer take advantage of the domain Markov property. The argument here relies mainly on ideas of [Ber17].

⋆ *The beginning of this chapter could be skipped by a reader interested only in the GMC measures associated to the Gaussian free field (i.e., the Liouville measures). In this case the reader may wish to skip to Section 3.6, although tools such as Kahane's inequality (Theorem 3.15) will be needed.*

Following the construction we will present a set of useful tools for the study of Gaussian multiplicative chaos measures; in particular, Kahane's powerful **convexity inequality**. Finally, we will apply these tools to the study of moments of GMC and in particular derive the **multifractal spectrum of GMC** and discuss applications to the famous **KPZ formula**.

3.1 Setup for Gaussian multiplicative chaos

We consider a more general setup than before and in particular for the rest of this chapter we do not assume we are working exclusively in two dimensions. Let $D \subset \mathbb{R}^d$ be a domain. Consider a nonnegative definite kernel $K(x, y)$ of the form

$$K(x, y) = \log(|x - y|^{-1}) + g(x, y) \tag{3.1}$$

where g is continuous over $\bar{D} \times \bar{D}$. Set

$$\mathcal{M}_+ = \{\rho \text{ a nonnegative measure in } D \text{ such that: } \iint |K(x, y)|\rho(dx)\rho(dy) < \infty\},$$

and set $\mathcal{M}$ to be the set of signed measures of the form $\rho = \rho_+ - \rho_-$, where $\rho_\pm \in \mathcal{M}_+$. Note that $\mathcal{M}$ contains all smooth compactly supported functions in D. Let h be the centered Gaussian generalised function with covariance K. That is, we view h as a stochastic process indexed by $\mathcal{M}$, characterised by the two properties that: (h, ρ) is linear in $\rho \in \mathcal{M}$ in the sense that $(h, \alpha\rho_1 + \beta\rho_2) = \alpha(h, \rho_1) + \beta(h, \rho_2)$ almost surely; and for any $\rho \in \mathcal{M}$,

$$(h, \rho) \text{ is a centered Gaussian random variable with variance } \iint K(x, y)\rho(dx)\rho(dy).$$

We will write $\int h(x)\rho(dx)$ for the random variable (h, ρ) with an abuse of notation. Note that this setup covers the case of a Gaussian free field in two dimensions with Dirichlet boundary conditions. In fact, it also covers the case of the Gaussian free field with free or Neumann boundary conditions, see Chapter 5, by changing γ into 2γ if necessary. We

extend the definition of h outside of D by setting $h|_{D^c} = 0$, so for any measure ρ such that $\rho|_D \in \mathcal{M}$, by definition $(h, \rho) = (h, \rho|_D)$.

Let σ be a Radon measure on $\bar{D}$ of dimension at least $\mathfrak{d}$ (where $0 \leq \mathfrak{d} \leq d$), in the sense that,

$$\iint_{\bar{D}\times\bar{D}} \frac{1}{|x-y|^{\mathfrak{d}-\varepsilon}}\sigma(dx)\sigma(dy) < \infty \tag{3.2}$$

for all $\varepsilon > 0$ (so for example, if σ is Lebesgue measure, then $d = \mathfrak{d}$). In particular σ is a finite measure since $\bar{D}$ is bounded. Note that $\mathfrak{d} \geq 0$ and may be equal to 0, but the statement of the theorem below will be empty in that case. In particular, we will only care about the case $\mathfrak{d} > 0$, which prevents σ from having any atoms. Throughout this chapter, when $\mathfrak{d} > 0$ we will fix a number $0 < \mathbf{d} < \mathfrak{d}$, such that

$$\iint_{\bar{D}\times\bar{D}} \frac{1}{|x-y|^{\mathbf{d}}}\sigma(dx)\sigma(dy) < \infty. \tag{3.3}$$

Remark 3.1. Many results of this chapter (e.g., Theorem 3.2, Theorem 3.26) are stated under an assumption of a strict inequality involving $\mathbf{d}$; since $\mathbf{d}$ can be chosen arbitrarily close to $\mathfrak{d}$ the same results could be stated by replacing $\mathbf{d}$ by $\mathfrak{d}$.

Let θ be a fixed nonnegative Radon measure on $\mathbb{R}^d$ supported in the unit ball $B(0,1)$, such that $\theta(\mathbb{R}^d) = 1$ and

$$\int |\log(1/|x-y|)|\theta(dy) \leq C < \infty \tag{3.4}$$

where C does not depend on $x \in B(0,5)$. It is easy to check that the condition (3.4) is satisfied whenever θ has an L^p Lebesgue density supported in $B(0,1)$ for some $p > 1$, but also in many other cases, for example, when θ is uniform distribution on the unit circle.

For $\varepsilon > 0$, set $\theta_\varepsilon(\cdot)$ to be the image of the measure θ under the mapping $x \mapsto \varepsilon x$, i.e. $\theta_\varepsilon(A) = \theta(A/\varepsilon)$ for all Borel sets A. We view this as an approximation of the identity based on θ (and will sometimes write $\theta_\varepsilon(x)dx$ for the measure $\theta_\varepsilon(dx)$ with an abuse of notation). We also write $\theta_{x,\varepsilon}(\cdot)$ for the measure θ_ε translated by x. For $x \in D$, note that by (3.4), the translated measure $\theta_{x,\varepsilon} \in \mathcal{M}$, so we can define an ε-regularisation of the field h by setting for ε small,

$$h_\varepsilon(x) = h * \theta_\varepsilon(x) = \int h(y)\theta_\varepsilon(x-y)dy = \int h(y)\theta_{x,\varepsilon}(dy), x \in D. \tag{3.5}$$

One can check that $\mathrm{Var}(h_\varepsilon(x) - h_\varepsilon(x')) \to 0$ as $|x - x'| \to 0$ for a fixed ε, so there exists a version of the stochastic process h such that $h_\varepsilon(x)$ is almost surely a Borel measurable function of $x \in S$ (see e.g. Proposition 2.1.12 in [GN16]). Hence for any Borel set $S \subset D$ and $\gamma \geq 0$ we may define

$$\mu_\varepsilon(S) = I_\varepsilon = \int_S e^{\gamma h_\varepsilon(z) - \frac{\gamma^2}{2}\mathbb{E}(h_\varepsilon(z)^2)}\sigma(dz).$$

In the previous chapter where h was the 2d Gaussian free field, our choice for σ was $\sigma(dz) = R(z,D)^{\gamma^2/2}dz$, and our choice for the measure θ was the uniform distribution on the unit circle (so $h_\varepsilon(z)$ was the usual circle average process of h).

3.2 Construction of Gaussian multiplicative chaos

With these definitions we can state the result that guarantee the existence of Gaussian multiplicative chaos. For simplicity we assume D bounded, and let $S \subset D$ be a Borel subset (which may be equal to D itself).

Theorem 3.2. *Let $0 \leq \gamma < \sqrt{2\mathbf{d}}$ (equivalently, $0 \leq \gamma < \sqrt{2\mathfrak{d}}$). Then $\mu_\varepsilon(S)$ converges in probability and in $L^1(\mathbb{P})$ to a limit $\mu(S)$. The random variable $\mu(S)$ does not depend on the choice of the regularising kernel θ subject to the above assumptions. Furthermore, the collection $(\mu(S))_{S\subset D}$ defines a Borel measure μ on D, and μ_ε converges in probability towards μ for the topology of weak convergence of measures on D.*

Let us assume without loss of generality that $\mathbf{d} > 0$, so that σ has no atoms.

As before, the main idea will be to pick $\alpha > \gamma$ and consider the normalised measure $e^{\gamma h_\varepsilon(x)}dx$, but *restricted to good points*; that is, points that are not too thick. We will check that the L^1 contribution of bad points is negligible (essentially by the above Cameron–Martin–Girsanov observation), while the remaining part is shown to remain bounded and in fact convergent in $L^2(\mathbb{P})$. The key will be to take a good and slightly more subtle definition of the notion of *good points*, that makes the relevant L^2 computation very simple.

In [Ber17], uniqueness of the limit was obtained by comparing to a different approximation of the field, arising from the Karhunen–Loeve expansion of h. This gives another approximation of the measure which turns out to be a martingale, and hence also has a limit. [Ber17] then showed that the two measures must agree, thereby deducing uniqueness. Here we present a slightly simpler argument based on a remark made Hubert Lacoin (private communication).

3.2.1 Uniform integrability

The goal of this section will be to prove:

Proposition 3.3. *I_ε is uniformly integrable.*

Proof. Let $\alpha > 0$ be fixed (it will be chosen $> \gamma$ and very close to γ soon). We will use the following notation in the rest of the article: for $r > 0$ we define

$$\bar{r} = e^{\lceil \log r \rceil} = \inf\{e^k : k \in \mathbb{Z}, e^k > r\} \tag{3.6}$$

to be the closest upper e-adic approximation of r. We define a **good event**

$$G_\varepsilon^\alpha(x) = \{h_{\bar{r}}(x) \leq \alpha \log(1/\bar{r}) \text{ for all } r \in [\varepsilon, \varepsilon_0]\} \tag{3.7}$$

with $\varepsilon_0 \leq 1$ for instance. This is the good event that the point x is never too thick up to scale ε. Further let $\bar{h}_\varepsilon(x) = \gamma h_\varepsilon(x) - (\gamma^2/2)\mathbb{E}(h_\varepsilon(x)^2)$ to ease notations.

Lemma 3.4 (Ordinary points are not thick)**.** *For any $\alpha > 0$, we have that uniformly over $x \in S$, $\mathbb{P}(G_\varepsilon^\alpha(x)) \geq 1 - p(\varepsilon_0)$ where the function p may depend on α and for a fixed $\alpha > \gamma$, $p(\varepsilon_0) \to 0$ as $\varepsilon_0 \to 0$.*

Proof. Set $X_t = h_\varepsilon(x)$ for $\varepsilon = e^{-t}$. Then a direct computation from (3.1) (see below in Lemma 3.6, and more precisely (3.10)), implies that

$$|\operatorname{Cov}(X_s, X_t) - s \wedge t| \leq O(1), \tag{3.8}$$

where the implicit constant is uniform. In particular $\operatorname{Var}(X_t) = t + O(1)$.

Note that for each $k \geq 1$, $\mathbb{P}(X_k \geq \alpha k/2) \leq e^{-\alpha^2 k^2/(8 \operatorname{Var}(X_k))}$ which decays exponentially in k by the above, and so is smaller than $Ce^{-\lambda k}$ for some $\lambda > 0$. Hence

$$\mathbb{P}(\exists k \geq k_0 : |X_k| \geq \alpha k) \leq \sum_{k \geq k_0} Ce^{-\lambda k}$$

We call $p(\varepsilon_0)$ to be the right hand side of the above for $k_0 = \lceil -\log(\varepsilon_0) \rceil$ which can be made arbitrarily small by picking ε_0 small enough. This proves the lemma. □

Lemma 3.5 (Liouville points are no more than γ-thick)**.** *For $\alpha > \gamma$ we have*

$$\mathbb{E}(e^{\bar{h}_\varepsilon(x)} 1_{G_\varepsilon^\alpha(x)}) \geq 1 - p(\varepsilon_0).$$

Proof. Note that

$$\mathbb{E}(e^{\bar{h}_\varepsilon(x)} \mathbf{1}_{\{G_\varepsilon^\alpha(x)\}}) = \tilde{\mathbb{P}}(G_\varepsilon^\alpha(x)), \text{ where } \frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} = e^{\bar{h}_\varepsilon(x)}.$$

By the Cameron–Martin–Girsanov lemma, under $\tilde{\mathbb{P}}$, the process $(X_s)_{-\log \varepsilon_0 \leq s \leq t}$ has the same covariance structure as under $\mathbb{P}$ and its mean is now $\gamma \operatorname{Cov}(X_s, X_t) = \gamma s + O(1)$ for $s \leq t$. Hence

$$\tilde{\mathbb{P}}(G_\varepsilon^\alpha(x)) \geq \mathbb{P}(G_\varepsilon^{\alpha-\gamma}(x)) \geq 1 - p(\varepsilon_0)$$

by Lemma 3.4 since $\alpha > \gamma$. □

We therefore see that points which are more than γ-thick do not contribute significantly to I_ε in expectation and can therefore be safely removed. We therefore fix $\alpha > \gamma$ and introduce:

$$J_\varepsilon = \int_S e^{\bar{h}_\varepsilon(z)} \mathbf{1}_{\{G_\varepsilon(z)\}} \sigma(dz) \tag{3.9}$$

with $G_\varepsilon(x) = G_\varepsilon^\alpha(x)$. We will show that J_ε is uniformly integrable from which the result follows.

Before we embark on the main argument of the proof, we record here for ease of reference an elementary estimate on the covariance structure of $h_\varepsilon(x)$. Roughly speaking, the role of the first estimate (3.10) is to bound from above (up to an unimportant constant of the form $e^{O(1)}$) the contribution to $\mathbb{E}(J_\varepsilon^2)$ coming from points x, y that are close to each other. That will suffice to prove uniform integrability. The role of the finer estimate (3.11) is to get a more precise estimate to the contribution to $\mathbb{E}(J_\varepsilon^2)$ coming from points x, y which are macroscopically far away, which we will be able to assume thanks to (3.10). This time the error in the covariance up to an additive term $o(1)$ will translate into an error up to a factor $e^{o(1)} = 1 + o(1)$ in the estimation of this contribution. In turn this will imply convergence.

Lemma 3.6. *We have the following estimate:*

$$\mathrm{Cov}(h_\varepsilon(x), h_r(y)) = \log 1/(|x-y| \vee r \vee \varepsilon) + O(1). \tag{3.10}$$

Moreover, if $\eta > 0$ and $|x-y| \geq \eta$, then

$$\mathrm{Cov}(h_\varepsilon(x), h_\delta(y)) = \log(1/|x-y|) + g(x,y) + o(1) \tag{3.11}$$

where $o(1)$ tends to 0 as $\delta, \varepsilon \to 0$, uniformly in $|x-y| \geq \eta$.

Proof. We start with the proof of (3.10). Assume without loss of generality that $\varepsilon \leq r$. Note that

$$\begin{aligned}\mathrm{Cov}(h_\varepsilon(x), h_r(y)) &= \iint K(z,w)\theta_{x,\varepsilon}(dw)\theta_{y,r}(dz) \\ &= \iint -\log(|w-z|)\theta_{x,\varepsilon}(dw)\theta_{y,r}(dz) + O(1)\end{aligned} \tag{3.12}$$

We consider the following cases: (a) $r \leq |x-y|/3$, and (b) $r \geq |x-y|/3$.

In case (a), $|x-y| \leq \varepsilon + |w-z| + r \leq 2r + |w-z| \leq (2/3)|x-y| + |w-z|$ by the triangle inequality, so $|w-z| \geq (1/3)|x-y|$ and we get

$$\mathrm{Cov}(h_\varepsilon(x), h_r(y)) \leq -\log|x-y| + O(1)$$

as desired in this case.

The second case (b) is when $r \geq |x-y|/3$. Then by translation and scaling so that $B(y,r)$ becomes $B(0,1)$, the right hand side of (3.12) is equal to

$$\log(1/r) + \iint -\log|w-z|\theta_{\frac{x-y}{r},\frac{\varepsilon}{r}}(dw)\theta(dz)$$

Conditioning on w (which is necessarily in $\bar{B}(0,4)$ under the assumptions of case (b)), we see that by the assumption (3.4) on θ, the second term is bounded by $O(1)$, uniformly, so that

$$\mathrm{Cov}(h_\varepsilon(x), h_r(y)) \leq -\log r + O(1)$$

as desired in this case. This proves (3.10).

The proof of (3.11) is similar but simpler. Indeed, we get (as in (3.12)),

$$\mathrm{Cov}(h_\varepsilon(x), h_\delta(y)) = \iint -\log|w-z|\theta_{x,\varepsilon}(dw)\theta_{y,\delta}(dz) + g(x,y) + o(1) \tag{3.13}$$

where the $o(1)$ term tends to 0 as $\varepsilon, \delta \to 0$, coming from the continuity of g, and hence is uniform in x, y (not even assuming $|x-y| \geq \eta$). Now note that

$$\big|\log|w-z| - \log|x-y|\big| \leq \frac{4\max(\varepsilon,\delta)}{|x-y|}$$

as soon as $\max(\varepsilon,\delta) \leq \eta/4 \leq |x-y|/4$. Therefore the right hand side of (3.13) is $-\log|x-y| + g(x,y) + O(\max(\varepsilon,\delta)) + o(1)$ when $|x-y| \geq \eta$, which proves the claim (3.11). □

Lemma 3.7. *For $\alpha > \gamma$ sufficiently close to γ, J_ε is bounded in $L^2(\mathbb{P})$ and hence uniformly integrable.*

Proof. By Fubini's theorem,

$$\begin{aligned}\mathbb{E}(J_\varepsilon^2) &= \int_{S\times S} \mathbb{E}(e^{\bar{h}_\varepsilon(x)+\bar{h}_\varepsilon(y)}\mathbf{1}_{\{G_\varepsilon(x)\cap G_\varepsilon(y)\}})\sigma(dx)\sigma(dy)\\ &= \int_{S\times S} e^{\gamma^2\operatorname{Cov}(h_\varepsilon(x),h_\varepsilon(y))}\tilde{\mathbb{P}}(G_\varepsilon(x)\cap G_\varepsilon(y))\sigma(dx)\sigma(dy)\end{aligned}$$

where $\tilde{\mathbb{P}}$ is a new probability measure obtained by the Radon-Nikodym derivative

$$\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} = \frac{e^{\bar{h}_\varepsilon(x)+\bar{h}_\varepsilon(y)}}{\mathbb{E}(e^{\bar{h}_\varepsilon(x)+\bar{h}_\varepsilon(y)})}.$$

Note that since σ has no atoms, we may assume that $x \neq y$. By Lemma 3.6 (more precisely by (3.10))

$$\operatorname{Cov}(h_\varepsilon(x), h_\varepsilon(y)) = -\log(|x-y|\vee\varepsilon) + g(x,y) + O(1). \tag{3.14}$$

Also, if and $\varepsilon \le e^{-1}\varepsilon_0$ and $|x-y| \le e^{-1}\varepsilon_0$ (else we bound the probability below by one), we have

$$\tilde{\mathbb{P}}(G_\varepsilon(x)\cap G_\varepsilon(y)) \le \tilde{\mathbb{P}}(h_r(x) \le \alpha\log 1/r)$$

where

$$r = \overline{\varepsilon\vee|x-y|} \tag{3.15}$$

(recall our notation for $\bar{r} = \inf\{e^k : k\in\mathbb{Z}, e^k > r\}$, see (3.6).) Furthermore, by Cameron–Martin–Girsanov, under $\tilde{\mathbb{P}}$ we have that $h_r(x)$ has the same variance as before (therefore $\log 1/r + O(1)$) and a mean given by

$$\operatorname{Cov}_{\mathbb{P}}(h_r(x), \gamma h_\varepsilon(x) + \gamma h_\varepsilon(y)) = 2\gamma\log 1/r + O(1), \tag{3.16}$$

again by Lemma 3.6 (more precisely, by (3.10)). Consequently,

$$\begin{aligned}\tilde{\mathbb{P}}(h_r(x) \le \alpha\log 1/r) &= \mathbb{P}(\mathcal{N}(2\gamma\log(1/r), \log 1/r) \le \alpha\log(1/r) + O(1))\\ &\le \exp(-\frac{1}{2}(2\gamma-\alpha)^2(\log(1/r)+O(1))) = O(1)r^{(2\gamma-\alpha)^2/2}.\end{aligned} \tag{3.17}$$

We deduce

$$\mathbb{E}(J_\varepsilon^2) \le O(1)\int_{S\times S} |(x-y)\vee\varepsilon|^{(2\gamma-\alpha)^2/2-\gamma^2}\sigma(dx)\sigma(dy). \tag{3.18}$$

(We will get a better approximation in the next section). Clearly by (3.2) this is bounded if

$$(2\gamma-\alpha)^2/2 - \gamma^2 > -\mathbf{d}$$

and since α can be chosen arbitrarily close to γ this is possible if

$$\mathbf{d} - \gamma^2/2 > 0 \text{ or } \gamma < \sqrt{2\mathbf{d}}. \tag{3.19}$$

This proves the lemma. □

To finish the proof of Proposition 3.3, observe that $I_\varepsilon = J_\varepsilon + J'_\varepsilon$. We have $\mathbb{E}(J'_\varepsilon) \leq p(\varepsilon_0)$ by Lemma 3.5, and for a fixed ε_0, J_ε is bounded in L^2 (uniformly in ε). Hence I_ε is uniformly integrable. □

3.2.2 Convergence

As before, since $\mathbb{E}(J'_\varepsilon)$ can be made arbitrarily small by choosing ε_0 sufficiently small, it suffices to show that J_ε converges in probability and in L^1. In fact we will show that it converges in L^2, from which convergence will follow. To do this we will show that $(J_\varepsilon)_\varepsilon$ forms a Cauchy sequence in L^2, and we start by writing

$$\mathbb{E}((J_\varepsilon - J_\delta)^2) = \mathbb{E}(J_\varepsilon^2) + \mathbb{E}(J_\delta^2) - 2\mathbb{E}(J_\varepsilon J_\delta). \tag{3.20}$$

Our basic approach is thus to estimate better than before $\mathbb{E}(J_\varepsilon^2)$ from above and $\mathbb{E}(J_\varepsilon J_\delta)$ from below. Essentially, the idea is that for x, y which are at a small but macroscopic distance, we can identify the limiting distribution of $(h_r(x), h_r(y))_{r\leq\varepsilon_0}$ under the distribution $\mathbb{P}$ biased by $e^{\bar{h}_\varepsilon(x)+\bar{h}_\delta(y)}$. On the other hand when x, y are closer than that we know from the previous section that the contribution is essentially negligible.

Lemma 3.8. *We have*

$$\limsup_{\varepsilon\to 0} \mathbb{E}(J_\varepsilon^2) \leq \int_{S\times S} e^{\gamma^2 g(x,y)} \frac{1}{|x-y|^{\gamma^2}} g_\alpha(x,y)\sigma(dx)\sigma(dy)$$

where $g_\alpha(x,y)$ is a non-negative function depending on α, ε_0 and γ such that the above integral is finite.

Proof. Recall that from (3.14) we already know

$$\mathbb{E}(J_\varepsilon^2) = \int_{S^2} e^{\gamma^2 \operatorname{Cov}(h_\varepsilon(x),h_\varepsilon(y))} \tilde{\mathbb{P}}(G_\varepsilon(x) \cap G_\varepsilon(y))\sigma(dx)\sigma(dy).$$

We simply have to estimate better $\tilde{\mathbb{P}}(G_\varepsilon(x) \cap G_\varepsilon(y))$. We fix $\eta > 0$ arbitrarily small (in particular, η may and will be smaller than $e^{-1}\varepsilon_0$). If $|x-y| \leq \eta$ we use the same bound as in (3.18). The contribution coming from the part $|x-y| \leq \eta$ can thus be bounded, uniformly in ε, by $f(\eta)$ (where $f(\eta) \to 0$ as $\eta \to 0$ and the precise order of magnitude of $f(\eta)$ is determined by (3.2), and is at most polynomial in η). We thus focus on the contribution coming from $|x-y| \geq \eta$.

Then observe that for any fixed $\varepsilon_1 \leq \varepsilon_0$, as $\varepsilon \to 0$, and uniformly over $x \in S$ and $r \geq \varepsilon_1$,

$$\operatorname{Cov}(h_r(x), h_\varepsilon(x)) \to \int_D K(x,z)\theta_r(x-z)dz \tag{3.21}$$

and likewise, uniformly over $x, y \in S$ such that $|x-y| \geq \eta$, and over $r \geq \varepsilon_1$, as $\varepsilon \to 0$:

$$\operatorname{Cov}(h_r(x), h_\varepsilon(y)) \to \int_D K(z,y)\theta_r(x-z)dz \tag{3.22}$$

(Note that both right hand sides of (3.21) and (3.22) are finite by (3.10).) Consequently, by Cameron–Martin–Girsanov, the joint law of the processes $(h_r(x), h_r(y))_{r\le\varepsilon_0}$ under $\tilde{\mathbb{P}}$ converges to a joint distribution $(\tilde{h}_r(x), \tilde{h}_r(y))_{r\le\varepsilon_0}$ whose covariance is unchanged and whose mean is given by the sum of (3.21) and (3.22) times γ. This convergence is for the weak convergence on compacts of $r \in (0, \varepsilon_0]$, and is uniformly in $|x-y| \ge \eta$. Let $\tilde{G}(x)$ be the event that $\tilde{h}_r(x) \le \alpha \log(1/r)$ for all $r \le \varepsilon_0$. Then it is not hard to deduce, uniformly in $|x-y| \ge \eta$,

$$\tilde{\mathbb{P}}(G_\varepsilon(x) \cap G_\varepsilon(y)) \to g_\alpha(x,y) := \mathbb{P}(\tilde{G}(x) \cap \tilde{G}(y)) \quad (\varepsilon \to 0). \tag{3.23}$$

Indeed, by (3.10), under $\tilde{\mathbb{P}}$, the drifts of $h_r(x)$ and of $h_r(y)$ (with $r \ge \varepsilon$) are each $\gamma \log(1/r) + O(1)$ where the $O(1)$ term is uniform in $|x-y| \ge \eta$. Because of this, up to an error in $\tilde{\mathbb{P}}$ probability that is arbitrarily small (uniformly in $|x-y| \ge \eta$), the events $G_\varepsilon(x), G_\varepsilon(y)$ as well as $\tilde{G}(x), \tilde{G}(y)$ depend only on the "macroscopic" behaviour of $h_r(x)$ and $h_r(y)$; that is, depend only on $(h_r(x), h_r(y))_{r\ge\varepsilon_1}$ for some ε_1. Consequently, as $\varepsilon \to 0$, after applying Lemma 3.6 (and more specifically (3.11)), we deduce (using (3.18) to justify the use of dominated convergence):

$$\int_{S^2;|x-y|\ge\eta} e^{\gamma^2 \operatorname{Cov}(h_\varepsilon(x), h_\varepsilon(y))} \tilde{\mathbb{P}}(G_\varepsilon(x), G_\varepsilon(y)) \sigma(dx)\sigma(dy) \to \int_{S^2;|x-y|\ge\eta} \frac{e^{\gamma^2 g(x,y)}}{|x-y|^{\gamma^2}} g_\alpha(x,y)\sigma(dx)\sigma(dy). \tag{3.24}$$

Since we already know that the piece of the integral coming from $|x-y| \le \eta$ contributes at most $f(\eta) \to 0$ when $\eta \to 0$, it remains to check that the integral on the right hand side of (3.24) remains finite as $\eta \to 0$. But we have already seen in (3.17) that for $|x-y| \le \varepsilon_0/3$, $\tilde{\mathbb{P}}(G_\varepsilon(x) \cap G_\varepsilon(y)) \le O(1)|x-y|^{(2\gamma-\alpha)^2/2-\gamma^2}$; hence this inequality must also hold for $g_\alpha(x,y)$. Hence the result follows as in (3.19). □

Lemma 3.9. *We have*

$$\liminf_{\varepsilon,\delta\to 0} \mathbb{E}(J_\varepsilon J_\delta) \ge \int_{S\times S} e^{\gamma^2 g(x,y)} \frac{1}{|x-y|^{\gamma^2}} g_\alpha(x,y)\sigma(dx)\sigma(dy).$$

Proof. In fact, the proof is almost exactly the same as in Lemma 3.8, except that $\tilde{\mathbb{P}}$ is now weighted by $e^{\bar{h}_\varepsilon(x)+\bar{h}_\delta(y)}$ instead of $e^{\bar{h}_\varepsilon(x)+\bar{h}_\varepsilon(y)}$. But this changes nothing to the argument leading up to (3.23) and hence (3.24) still holds. Since we get a lower bound by restricting ourselves to $|x-y| \ge \eta$, we deduce immediately that

$$\liminf_{\varepsilon,\delta\to 0} \mathbb{E}(J_\varepsilon J_\delta) \ge \int_{S^2;|x-y|\ge\eta} e^{\gamma^2 g(x,y)} \frac{1}{|x-y|^{\gamma^2}} g_\alpha(x,y)\sigma(dx)\sigma(dy).$$

Since η is arbitrary, the result follows. □

Proof of convergence in Theorem 3.2. Using (3.20) together with Lemmas 3.8 and 3.9, we see that J_ε is a Cauchy sequence in L^2 for any $\varepsilon_0 > 0$. Combining with Lemma 3.5, it therefore follows that I_ε is a Cauchy sequence in L^1 and hence converges in L^1 (and also in probability) to a limit $I = \mu(S)$. The proof of weak convergence follows by the argument in Section 2.3. □

Remark 3.10. Note that $\lim_{\varepsilon\to 0} \mathbb{E}(J_\varepsilon^2)$ depends on the regularisation θ, even though, as we will see next, $\lim_{\varepsilon\to 0} I_\varepsilon$ does not.

Proof of uniqueness in Theorem 3.2. To prove uniqueness, we take $\tilde{\theta}$ another Radon measure on $\mathbb{R}^d$ satisfying (3.4). Let $\tilde{h}_\delta(x) = h * \tilde{\theta}_\delta(x)$, and let $\tilde{J}_\delta$ be defined as J_δ but with $\tilde{\theta}$ instead of θ: that is,

$$\tilde{J}_\delta = \int_S e^{\gamma \tilde{h}_\delta(z) - (\gamma^2/2)\mathbb{E}(\tilde{h}_\delta(z)^2)} \mathbf{1}_{\{\tilde{G}_\delta(z)\}} \sigma(dz)$$

where the good event $\tilde{G}_\delta(z)$ is as in (3.7), with $\tilde{\theta}$ in place of θ. Then the argument of Lemma 3.9 can be used to show that the same conclusion holds for J_δ replace by $\tilde{J}_\delta$: that is,

$$\liminf_{\varepsilon,\delta\to 0} \mathbb{E}(J_\varepsilon \tilde{J}_\delta) \geq \int_{S\times S} e^{\gamma^2 g(x,y)} \frac{1}{|x-y|^{\gamma^2}} g_\alpha(x,y)\sigma(dx)\sigma(dy).$$

Hence we deduce $\lim_{\varepsilon\to 0,\delta\to 0} \mathbb{E}((J_\varepsilon - \tilde{J}_\delta)^2) = 0$ and this implies that the limits associated with θ and $\tilde{\theta}$ are a.s. the same.

3.3 Shamov's approach to Gaussian multiplicative chaos

An alternative and powerful viewpoint on Gaussian multiplicative chaos was also developed in Shamov [Sha16]. It is closely related to the generalisation of "rooted measures" for the GFF: see Section 2.4. In what follows h will be a centred Gaussian field with logarithmically diverging covariance kernel K as in (3.1) (although the original paper [Sha16] works in a more general setting).

Before stating the result, let us make an observation about changes of measure for the field h. If $\rho \in \mathcal{M}$ we write $T\rho(x) = \int_D K(x,y)\,\rho(dy)$. Then by Girsanov's theorem, Lemma 2.5, it follows that the field $h + T\rho$ is absolutely continuous with respect to h, with associated Radon–Nikodym derivative

$$\frac{\exp((h,\rho))}{\exp(\frac{1}{2}(\rho, T\rho))}. \tag{3.25}$$

Note the connection with Section 1.8 in the case of the zero boundary GFF: when $\rho \in \mathcal{M}_0$ ($\mathcal{M}_0$ corresponding to the zero boundary condition Green function) then $(h,\rho) = (h,F)_\nabla$, where F is defined by $-\Delta F = 2\pi\rho$ and is an element of $H_0^1(D)$. By (1.9) this is exactly the statement that $F = T\rho$, and the above expression is equal to $\exp((h,F)_\nabla)/\exp(\frac{1}{2}(F,F)_\nabla)$ as in Proposition 1.29. See [Aru17] for more on Shamov's approach when the field is the planar GFF.

Definition 3.11 (Shamov's definition of GMC)**.** *Let h be as above and σ as in (3.2). Let $\gamma \in (0,2)$. A measure μ_γ is a γ-multiplicative chaos measure for h, with background measure σ if:*

- *μ_γ is measurable with respect to h as a stochastic process indexed by $\mathcal{M}$ (note that this allows us to write $\mu_\gamma(dx) = \mu_\gamma(h, dx)$);*
- *$\mathbb{E}(\mu_\gamma(S)) = \sigma(S)$ for all Borel sets $S \subset D$;*

- *For every fixed, deterministic Borel measurable function ξ such that $\xi(x) = T\rho(x)$ σ-a.e. with $\rho \in \mathcal{M}$,*

$$\mu_\gamma(h + \xi, dx) = \exp(\gamma\xi(x))\mu_\gamma(h, dx) \text{ almost surely.} \tag{3.26}$$

Note that although ξ is only defined almost everywhere with respect to σ (for example when the field is a GFF with Dirichlet boundary conditions in D, then ξ will only be an element of $H_0^1(D)$), the measure $\exp(\gamma\xi(x))\mu_\gamma(h, dx)$ still makes sense unambiguously. Indeed, the assumption that $\mathbb{E}(\mu_\gamma) = \sigma$ implies that if $\tilde{\xi}$ is such that $\xi(x) = \tilde{\xi}(x)$ for σ-almost every x, then by Fubini's theorem

$$\mathbb{E}(\int_S 1_{\{\xi(x)\neq\tilde{\xi}(x)\}} \exp(\gamma\xi(x))\mu_\gamma(dx)) = 0.$$

It follows that $\int_A 1_{\{\xi(x)\neq\tilde{\xi}(x)\}} \exp(\gamma\xi(x))\mu_\gamma(dx) = 0$ a.s. simultaneously for all Borel sets $A \subset S$. This implies that on an event of probability one,

$$\int_A \exp(\gamma\tilde{\xi}(x))\mu_\gamma(dx) = \int_A \exp(\gamma\xi(x))\mu_\gamma(h, dx)$$

for all $A \subset S$. Hence the measures $\mu_\gamma(h + \xi, dx)$ and $\mu_\gamma(h + \tilde{\xi})$ agree with probability one, and so are unambiguously defined.

Theorem 3.12 (Shamov, [Sha16])**.** *Assume the setup of Definition 3.11. Then a γ-multiplicative chaos measure for h with background measure σ exists. Moreover, it is unique.*

We note that the uniqueness part of Theorem 3.12 may be particularly useful if one wants to identify some limit as being a GMC measure, since the conditions are in many contexts relatively easy to check. Actually, these conditions can be slightly weakened so as to restrict ξ to an appropriately dense subspace; for instance, in the case where h is the GFF with Dirichlet boundary conditions, it suffices to know (3.26) for smooth functions with compact support, see Remark 3.14.

Remark 3.13. As we will see in the proof below, the condition (3.26) ensures that the effect of weighting the law of the field by $\mu_\gamma(D)$ is to add the singularity $\gamma K(x, \cdot)$ to the field at a point x chosen from μ_γ. So essentially, Shamov's approach characterises the GMC measure as a certain Radon–Nikodym derivative for the field.

Given Theorem 3.2 the existence part of this theorem is clear. Indeed we can check that the GMC measure constructed in Theorem 3.2 does satisfy the stated conditions, since the limit holds in probability and in L^1. (In particular, given the uniqueness of Theorem 3.12, it follows that the measures of Theorem 3.12 and Theorem 3.2 are the same). It remains to prove the uniqueness.

Proof of uniqueness. Suppose that a measure μ_γ satisfying the constraints of Definition 3.11 exists. We will consider the probability measure

$$Q(dh, dx) = \frac{\mathbb{P}(dh)\mu_\gamma(h, dx)}{\mathbb{E}(\mu_\gamma(D))} \tag{3.27}$$

and show that under Q, the marginal law of x has density proportional to $\sigma(dx)$, and that given x, the conditional law of the field (viewed as a stochastic process indexed by $\mathcal{M}$) is that of h plus the deterministic function $\gamma K(x, \cdot)$. Observe that this completely characterises the the law Q and thus, by disintegration, the conditional law of x given h under Q. On the other hand, the definition of Q means that this conditional law is exactly $\mu_\gamma(h, dx)$ and so we have identified μ_γ uniquely (note that this doesn't identify only the law of μ_γ but really the joint law of h and μ_γ).

To show the claim concerning Q, it is enough to prove that the Q marginal law of x is equal to $\sigma(dx)/\sigma(D)$, and that for any $\rho_1, \cdots, \rho_m \in \mathcal{M}$ and $a_1, \ldots, a_m \in \mathbb{R}$ the Q conditional law of $(a_1(h, \rho_1) + \ldots + a_m(h, \rho_m))$ given x is a normal random variable with the correct mean and covariance. In other words (using linearity of h on the space $\mathcal{M}$) it suffices to show that for any $g \in L^1(\sigma)$ on D, and $\rho \in \mathcal{M}$

$$\mathbb{E}_Q(\mathrm{e}^{(h,\rho)} g(x)) = \mathbb{E}(\int_D \mathrm{e}^{(h+\gamma K(x,\cdot),\rho)} g(x) \frac{\sigma(dx)}{\sigma(D)}). \tag{3.28}$$

Note that by Fubini's theorem, the right-hand side of the above is equal to

$$\sigma(D)^{-1} \int_D \mathrm{e}^{\frac{1}{2}\operatorname{Var}((h,\rho)) + \gamma \int K(x,y)\rho(dy)} g(x)\sigma(dx) = \sigma(D)^{-1} \int_D \mathrm{e}^{\frac{1}{2}(T\rho,\rho) + \gamma T\rho(x)} g(x)\sigma(dx)$$

(recalling the notation $T\rho$ when $\rho \in \mathcal{M}$). Furthermore, the left-hand side of (3.28) (using the assumption that $\mathbb{E}(\mu_\gamma(D)) = \sigma(D)$ and the definition of Q) is equal to

$$\sigma(D)^{-1}\mathbb{E}(\int_D \mathrm{e}^{(h,\rho)} g(x)\mu_\gamma(h, dx)).$$

However, using the observation (3.25) and (3.26), we have

$$\begin{aligned}
\mathbb{E}(\int_D \mathrm{e}^{(h,\rho)} g(x)\mu_\gamma(h, dx)) &= \mathbb{E}(\int_D \mathrm{e}^{\frac{1}{2}(T\rho,\rho)} g(x)\mu_\gamma(h + T\rho, dx)) \\
&= \mathbb{E}(\int_D \mathrm{e}^{\frac{1}{2}(T\rho,\rho)} \mathrm{e}^{\gamma T\rho(x)} g(x)\mu_\gamma(h, dx)) \\
&= \int_D \mathrm{e}^{\frac{1}{2}(T\rho,\rho) + \gamma T\rho(x)} g(x)\sigma(dx),
\end{aligned} \tag{3.29}$$

where in the last line we again used the assumption that $\mathbb{E}(\mu_\gamma(h, dx)) = \sigma(dx)$. Dividing by $\sigma(D)$ this is the same as the right hand side of (3.28), so we get the desired result. □

Remark 3.14. Note that in the case where the field is a GFF in some domain D, the assumption (3.26) can be weakened and only assumed to hold for smooth functions $\xi \in \mathcal{D}_0(D)$ with compact support. Indeed, by Lemma 1.21, we know that any $\rho \in \mathcal{M}_0$ can be approximated by such functions, with respect to H_0^{-1} norm $f \mapsto (f, Tf)$. This implies that if $\rho \in \mathcal{M}$ we can find a sequence $\rho_n \in \mathcal{D}_0(D)$ such that (h, ρ_n) converges in probability for $\mathbb{P}$ to (h, ρ). Since Q is absolutely continuous with respect to $\mathbb{P}$, this also holds under Q. From the proof of Theorem 3.12, we see that we have characterised the law of $(h, \rho)_{\rho \in \mathcal{D}_0(D)}$ under Q. Using the above density argument we have therefore also characterised the law of $(h, \rho)_{\rho \in \mathcal{M}}$ under Q, which proves the claim. The same argument should work for more general fields but would require first proving an analogue of Lemma 1.21.

3.4 Kahane's convexity inequality

We now present a fundamental tool in the study of Gaussian multiplicative chaos, which is Kahane's convexity inequality. Essentially, this is an inequality that will allow us to "compare" the GMC measures associated with two slightly different fields. Such comparison arguments are very useful in order to do scaling arguments and so compute moments and multifractal spectra, which is our next goal. This inequality was actually crucial to Kahane's construction of Gaussian multiplicative chaos [Kah85], although modern approaches such as the one presented just above (coming from [Ber17]) do not rely on this.

More precisely, the content of Kahane's inequality is to say that given a **convex** function f, and two centred Gaussian fields $X = (X_s)_{s \in T}$ and $Y = (Y_s)_{s \in T}$ with covariances Σ_X and Σ_Y such that $\Sigma_X(s,t) \leq \Sigma_Y(s,t)$ pointwise, we have

$$\mathbb{E}(f(\mu_X(D)) \leq \mathbb{E}(f(\mu_Y(D)))$$

for μ_X, μ_Y the Liouville measures associated with X and Y. The precise statement of the inequality comes in different flavours depending on what one is willing to assume about f and the fields. A statement first appeared in [Kah86], which had an elegant proof but relied on the extra assumption that f is increasing. As we will see this assumption is crucially violated for us (for example, in the proof of Theorem 3.23 we will use $f(x) = -x^q$ with $q < 1$, so f is convex but decreasing). The assumption of increasing f is removed in [Kah85], whose proof we will follow roughly here.

Theorem 3.15 (Kahane's convexity inequality)**.** *Suppose that $D \subset \mathbb{R}^d$ is bounded and that $(X(x))_{x \in D}, (Y(x))_{x \in D}$ are a.s. continuous centred Gaussian fields with*

$$K_X(x,y) := \mathbb{E}(X(x)X(y)) \leq \mathbb{E}(Y(x)Y(y)) =: K_Y(x,y) \text{ for all } x, y \in D.$$

Assume that $f : (0, \infty) \to \mathbb{R}$ is convex with at most polynomial growth at 0 and ∞, and σ is a Radon measure as in (3.2). Then

$$\mathbb{E}\left(f\left(\int_D \mathrm{e}^{X(x) - \frac{1}{2}\mathbb{E}(X(x)^2)}\sigma(dx)\right)\right) \leq \mathbb{E}\left(f\left(\int_D \mathrm{e}^{Y(x) - \frac{1}{2}\mathbb{E}(Y(x)^2)}\sigma(dx)\right)\right).$$

Proof. The proof is closely related to a Gaussian Integration by Parts formula (see e.g. [Zei15]). Define, for $t \in [0,1]$:

$$Z_t = \sqrt{1-t}X + \sqrt{t}Y.$$

Thus $Z_0 = X$ and $Z_1 = Y$. Since the fields X and Y are assumed to be continuous, the maxima and minima of X and Y on D have sub-Gaussian tails by Borell's inequality (see for example [Zei15, Theorem 2]). This means that if f is as in the statement of theorem, we have

$$h(t) := \mathbb{E}\left(f(\int_D Q_t(x)\,\sigma(dx))\right) := \mathbb{E}\left(f\left(\int_D \mathrm{e}^{Z_t(x)-\frac{1}{2}\mathbb{E}((Z_t(x))^2)}\,\sigma(dx)\right)\right) < \infty \text{ for all } t \in [0,1].$$

In fact, since f is convex, it is differentiable at all but countably many points. This means we can actually differentiate the above expression and obtain that

$$\frac{dh}{dt} = \frac{1}{2}\mathbb{E}\left(f'\left(\int_D Q_t(x)\,\sigma(dx)\right)\int_D \sigma(dy)\left(\frac{-X(y)}{\sqrt{1-t}} + \frac{Y(y)}{\sqrt{t}} + K_X(y,y) - K_Y(y,y)\right)Q_t(y)\right).$$

Here we have "differentiated under the integral sign" twice (once for the derivative of $\int_D Q_t(x)\sigma(dx)$ and once for the expectation) which is permitted since $\int_D Q_t(x)\sigma(dx)$ has sub-Gaussian tails and f has at most polynomial growth at 0 and ∞ (since f' is increasing this means that f' also has at most polynomial growth at 0 and ∞).

Consequently, by Fubini's theorem, it suffices to show that for any fixed y :

$$\mathbb{E}\left(\left(\frac{-X(y)}{\sqrt{1-t}} + \frac{Y(y)}{\sqrt{t}} + K_X(y,y) - K_Y(y,y)\right)Q_t(y)\,f'\left(\int_D Q_t(x)\,\sigma(dx)\right)\right) \geq 0. \tag{3.30}$$

Indeed, this then implies that h is increasing and so $h(0) = \mathbb{E}(f(\int_D \mathrm{e}^{X(x)-(1/2)\mathbb{E}(X(x)^2)}\sigma(dx)))$ is less than or equal to $h(1) = \mathbb{E}(f(\int_D \mathrm{e}^{Y(x)-(1/2)\mathbb{E}(Y(x)^2)}\sigma(dx)))$, as desired.

To show (3.30), we fix y and write

$$U_t(y) := \frac{-X(y)}{\sqrt{1-t}} + \frac{Y(y)}{\sqrt{t}},$$

so that $U_t(y)$ is the time derivative of the interpolation $Z_t(y)$. Note that $\mathbb{E}(U_t(y)Z_t(x)) = K_Y(x,y) - K_X(x,y) \geq 0$ for all x. This means that we can decompose

$$Z_t(x) = A_t(x)U_t(y) + V_t(x)$$

for each $x \in D$, where $A_t(x) = (K_Y(x,y) - K_X(x,y))/\mathbb{E}(U_t(y)^2) \geq 0$ and $V_t(x)$ is centred, Gaussian and independent of $U_t(y)$. This corresponds to writing the conditional law of $Z_t(x)$ given $U_t(y)$. Let us rewrite the expectation in (3.30) in terms of $U_t(y)$ and $V_t(y)$. To start with, we decompose

$$Q_t(x) = \mathrm{e}^{A_t(x)U_t(y)-\frac{1}{2}A_t(x)^2\mathbb{E}(U_t(y)^2)}\mathrm{e}^{V_t(x)-\frac{1}{2}\mathbb{E}(V_t(x)^2)} \tag{3.31}$$

for each $x \in D$. Thus applying (3.31) with $x = y$, the expectation in (3.30) can be rewritten as

$$\mathbb{E}\left(\left(U_t(y) - A_t(y)\mathbb{E}(U_t(y)^2)\right) \mathrm{e}^{A_t(y)U_t(y) - \frac{1}{2}A_t(y)^2\mathbb{E}(U_t(y)^2)} \mathrm{e}^{V_t(y) - \frac{1}{2}\mathbb{E}(V_t(y)^2)} f'\left(\int_D Q_t(x)\sigma(dx)\right)\right).$$

Now, in order to write this an expectation involving the single Gaussian random variable $U_t(y)$, we consider the conditional expectation (now expanding $Q_t(x)$ as in (3.31) for clarity):

$$\mathbb{E}\left(\mathrm{e}^{V_t(y) - \frac{1}{2}\mathbb{E}(V_t(y)^2)} f'\left(\int_D \mathrm{e}^{A_t(x)U_t(y) - \frac{1}{2}A_t(x)^2\mathbb{E}(U_t(y)^2)} \mathrm{e}^{V_t(x) - \frac{1}{2}\mathbb{E}(V_t(x)^2)}\sigma(dx)\right) \,\middle|\, U_t(y)\right).$$

Since $U_t(y)$ is independent of $V_t(x)$ for each $x \in D$ (and thus, by Gaussianity, of $(V_t(x), x \in D)$), and since $A_t(x) \geq 0$ and f' is increasing, we see that the above conditional expectation is an a.s. increasing function of $U_t(y)$. Hence (3.30) can be written as

$$\mathbb{E}\left(g(U_t(y))\left(U_t(y) - A_t(y)\mathbb{E}(U_t(y)^2)\right) \mathrm{e}^{A_t(y)U_t(y) - \frac{1}{2}A_t(y)^2\mathbb{E}(U_t(y)^2)}\right),$$

where g is an increasing function. Approximating g by a positive linear combination of step functions and writing $a = A_t(y)$, $\sigma^2 = \mathbb{E}(U_t(y)^2)$ it therefore suffices to prove that

$$\int_x^\infty e^{-z^2/2\sigma^2}(z - a\sigma^2)\mathrm{e}^{az - \frac{a^2\sigma^2}{2}}\, dz \geq 0$$

for any $x \in \mathbb{R}$.

If $x \geq a\sigma^2$ then the above clearly holds by positivity of the integrand. On the other hand, if $x \leq a\sigma^2$ then the integral is greater than

$$\int_{-\infty}^\infty e^{-z^2/2\sigma^2}(z - a\sigma^2)\mathrm{e}^{az - \frac{a^2\sigma^2}{2}}\, dz = \frac{d}{da}\int_{-\infty}^\infty e^{-z^2/2\sigma^2}\mathrm{e}^{az - \frac{a^2\sigma^2}{2}} = \frac{d}{da}(1) = 0.$$

This concludes the proof. □

3.5 Scale-invariant fields

When we apply Kahane's convexity inequality we will want to compare our Gaussian field with an auxiliary Gaussian field enjoying an exact scaling relation. In this section we explain a modification, due to Rhodes and Vargas ([RV10a]) of a construction due to Bacry and Muzy ([BM03]), that will give us the desired scale-invariant field. (In the case of the two-dimensional GFF the Markov property gives a close analogue but would lead to extra technicalities.)

3.5.1 One-dimensional cone construction

We first explain the construction we will use in one dimension where things are easier. Fix $0 < \varepsilon < R$ and for $x \in \mathbb{R}$, consider the **truncated cone** $C_{\varepsilon,R}(x)$ in $\mathbb{R}^2$ given by

$$C(x) = C_R(x) = \{z = (y,t) \in \mathbb{R}^d \times [0,\infty) : |y - x| \leq (t \wedge R)/2\}. \tag{3.32}$$

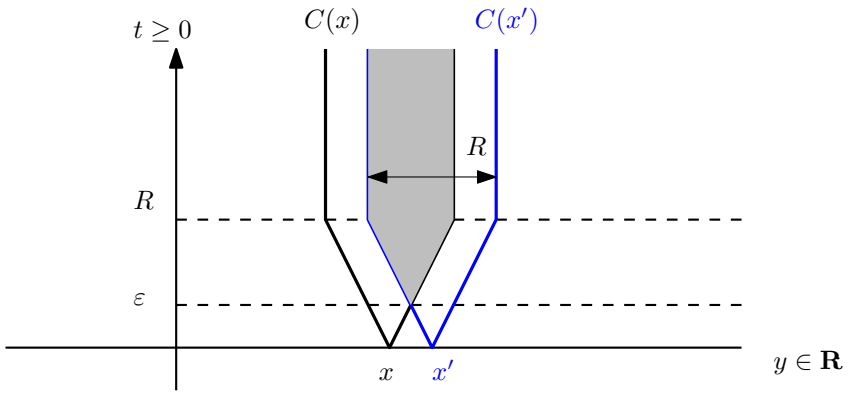


Figure 2: The truncated cones in the construction of the scale-invariant auxiliary field. The covariance of the field at (x, x') is obtained by integrating $dydt/t^2$ in the shaded area.

where $|y - x|$ denotes Euclidean norm in $\mathbb{R}$. Define a kernel

$$c_{\varepsilon,R}(x, x') = \int_{y\in\mathbb{R}} \int_{t\in[\varepsilon,\infty)} \mathbf{1}_{\{(y,t)\in C(x)\cap C(x')\}} \frac{dydt}{t^2}.$$

Note that since has been truncated at ε, the integral is finite. We claim that $c_{\varepsilon,R}$ is non-negative definite and so can be used to defined a Gaussian field $X_{\varepsilon,R}$ on $\mathbb{R}$ whose covariance is given by $c_{\varepsilon,R}$. Indeed, for any $n \geq 1$, for any $x_1, \ldots, x_n \in \mathbb{R}^d$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$,

$$\sum_{i,j=1}^{n} \lambda_i \lambda_j c_{\varepsilon,R}(x_i, x_j) = \sum_{i,j=1}^{n} \lambda_i \lambda_j \int_{\mathbb{R}} \int_{\varepsilon}^{\infty} \mathbf{1}_{(y,t)\in C(x_i)} \mathbf{1}_{(y,t)\in C(x_j)} \frac{dydt}{t^2} \tag{3.33}$$

$$= \int_{\mathbb{R}^d} \int_{\varepsilon}^{\infty} \left(\sum_{i=1}^{n} \lambda_i \mathbf{1}_{(y,t)\in C(x_i)} \right)^2 \frac{dydt}{t^2} \geq 0. \tag{3.34}$$

As the covariance kernel $c_{\varepsilon,R}$ is a nice continuous function of x, x' we can check (again using e.g. Proposition 2.1.12 in [GN16]) that there exists a centred Gaussian field $X_{\varepsilon,R}$ whose covariance is given by $c_{\varepsilon,R}$ and which is a.s. Borel measurable as a function on $\mathbb{R}$.

Remark 3.16. This computation showing that $c_{\varepsilon,R}$ is non-negative definite works because the covariance is defined to be of the form $c(x, x') = \int_S f_x(z) f_{x'}(z) \nu(dz)$, for some fixed function f_x of z associated to each $x \in \mathbb{R}$, where the integral can be on some arbitrary space S with measure ν. Here the space S is $\mathbb{R} \times (0, \infty)$, $\nu(dz) = \mathbf{1}_{\{t\geq\varepsilon\}} dydt/t^2$, and $f_x(z) = \mathbf{1}_{z\in C_{\varepsilon,R}(x)}$. We will use other choices when considering the higher dimensional case.

The key property of $X_{\varepsilon,R}$ (and the reason for introducing it) is that its covariance can be computed exactly. This not only shows that the field is logarithmically correlated, but enjoys an exact scaling relation, as follows.

Lemma 3.17. *Define the function*

$$g_{\varepsilon,R}(x) = \begin{cases} \log_+(R/|x|) & : |x| \geq \varepsilon \\ \log(R/\varepsilon) + 1 - (|x|/\varepsilon) & : |x| \leq \varepsilon, \end{cases} \tag{3.35}$$

where $\log_+(x) = \log(x) \vee 0$. *Then for all* $x, y \in \mathbb{R}$,

$$c_{\varepsilon,R}(x, x') = g_{\varepsilon,R}(x - x'). \tag{3.36}$$

In particular, $c_{\varepsilon,R}(x, x') = \log(R/|x - x'|) + O(1)$, *where the* $O(1)$ *term does not depend on* x, x', ε *or* R *(and is in fact bounded between* 0 *and 1).*

Proof. By translation invariance and symmetry we can assume that $x' = 0$ and $x > 0$. If $x \geq R$ there is nothing to prove, so assume first that $\varepsilon \leq x \leq R$. Then the two cones first intersect at height $x \geq \varepsilon$. Moreover the width of the intersection of these cones at height $t \geq x$ is $(t - x) \wedge (R - x)$, so

$$\begin{aligned} c_{\varepsilon,R}(0, x) &= \int_x^R (t - x)\frac{dt}{t^2} + \int_R^\infty (R - x)\frac{dt}{t^2} \\ &= \log(R/x) - x(\frac{1}{x} + \frac{1}{R}) + \frac{(R - x)}{R} \\ &= \log(R/x) \end{aligned}$$

as desired. When $x \leq \varepsilon$, the computation is almost the same, but the lower bound of integration for the first integral is ε instead of x, which gives the desired result. □

We now explain why this implies a scaling property. We fix the value R of truncation and write X_ε for $X_{\varepsilon,R}$ (often we will choose $R = 1$ and write Y_ε for $X_{\varepsilon,1}$).

Corollary 3.18. *For* $\lambda < 1$,

$$(X_{\lambda\varepsilon}(\lambda x))_{x \in B(0,R/2)} =_d (\Omega_\lambda + X_\varepsilon(x))_{x \in B(0,R/2)},$$

where Ω_λ *is an independent centred Gaussian random variable with variance* $\log(1/\lambda)$.

Proof. One directly checks that for all $x, x' \in \mathbb{R}$ such that $|x - x'| \leq R$ (and so also $|x - x'| \leq R/\lambda$ automatically),

$$c_{\lambda\varepsilon,R}(\lambda x, \lambda y) = c_{\varepsilon,R}(x, y) + \log(1/\lambda)$$

and hence the result follows. □

3.5.2 Higher dimensional construction

The one dimensional Bacry–Muzy construction presented above is beautiful and simple but does not trivially generalise to more than one dimension. This is because if one considers truncated cones in $\mathbb{R}^{d+1}$ (instead of $\mathbb{R}^2$) and integrates with respect to the scale-invariant

measure $dydt/t^{d+1}$, the volume of the intersection of two truncated cones based at x and x' does not lead to nice formulae which yield scale invariance in the sense of Corollary 3.18 (see [Cha06] for an article where this model is nevertheless studied).

To overcome this problem we follow (in a slightly simplified setting) a very nice construction proposed by Rhodes and Vargas [RV10a] in which the exact one-dimensional computation of Bacry and Muzy can be exploited to give a field in any number of dimensions satisfying both logarithmic correlations and exact scaling relations. The basic idea is to define the cones on $\mathbb{R}^d$ based at x and $x' \in \mathbb{R}^d$ by first applying a random rotation in order to preserve isotropy, and then applying the one-dimensional construction to the first coordinates of x and x'.

To be more precise, let $d \geq 1$ and consider $\mathcal{R}$ the unitary group of $\mathbb{R}^d$: that is, d-dimensional matrices M such that $MM^t = I$. (In two dimensions, $\mathcal{R}$ is nothing but the set of rotations). Let σ denote Haar measure on $\mathcal{R}$ normalised to be a probability distribution. (Thus in two dimensions σ can be identified with the uniform distribution on the unit circle). If $\rho \in G$ and $x \in \mathbb{R}^d$ let ρx denote the vector of $\mathbb{R}^d$ obtained by applying the isometry ρ to x, and let $(\rho x)_1$ denote its first coordinate. Define the **cone-like** set $\mathbf{C}_R(x)$ as follows:

$$\mathbf{C}_R(x) := \{(\rho, t, y) \in \mathcal{R} \times \mathbb{R} \times (0, \infty) : (t, y) \in C_R((\rho x)_1)\}.$$

where if $z \in \mathbb{R}$, $C_R(z)$ is the truncated cone of (3.32). Thus for any given ρ, we first apply ρ to x and consider the truncated cone (in two dimensions) based on the first coordinate of ρx. As in the previous section we define a field through its covariance kernel

$$\mathbf{c}_{\varepsilon,R}(x, x') = \int_{\mathcal{R}\times\mathbb{R}\times(0,\infty)} \mathbf{1}_{\{(\rho,t,y)\in \mathbf{C}_R(x)\cap\mathbf{C}_R(x')\}} \mathbf{1}_{\{t\geq\varepsilon\}} d\sigma(\rho) \otimes \frac{dydt}{t^2}.$$

We note that this is non-negative definite for the same reasons as (3.33) (see especially Remark 3.16). Hence, as before we can consider an a.s. Borel measurable function $x \in \mathbb{R}^d \mapsto X_{\varepsilon,R}(x) \in \mathbb{R}$ which is a centred Gaussian field with $\mathbf{c}_{\varepsilon,R}$ as its covariance kernel. Moreover, we have the following exact expression for the covariance. Recall the function $g_{\varepsilon,R}$ from Lemma 3.17:

$$g_{\varepsilon,R}(x) = \begin{cases} \log_+(R/|x|) & : |x| \geq \varepsilon \\ \log(R/\varepsilon) + 1 - (|x|/\varepsilon) & : |x| \leq \varepsilon. \end{cases}$$

We have

Lemma 3.19.

$$\mathbb{E}(X_{\varepsilon,R}(x)X_{\varepsilon,R}(x')) = \int_{\rho\in\mathcal{R}} g_{\varepsilon,R}((\rho x)_1 - (\rho x')_1) d\sigma.$$

Proof. This follows immediately from Fubini's theorem and the fact that $g_{\varepsilon,R}$ gives the covariance in the one-dimensional construction by Lemma 3.17. □

As a consequence, we directly obtain the desired scale-invariance. As before, we fix the value R (corresponding to the large scale truncation), and write X_ε for $X_{\varepsilon,R}$. Again, typically we will choose $R = 1$ in which case we write Y_ε for $X_{\varepsilon,1}$.

Corollary 3.20. *For $\lambda < 1$,*

$$(X_{\lambda\varepsilon}(\lambda x))_{x\in B(0,R/2)} =_d (\Omega_\lambda + X_\varepsilon(x))_{x\in B(0,R/2)},$$

where Ω_λ is an independent centred Gaussian random variable with variance $\log(1/\lambda)$.

Proof. This follows directly from the already observed fact that if $x, x' \in \mathbb{R}^d$ such that $|x - x'| \le R$ (and so also $|x - x'| \le R/\lambda$ automatically),

$$g_{\lambda\varepsilon,R}(\lambda x - \lambda y) = g_{\varepsilon,R}(x - y) + \log(1/\lambda)$$

which, as explained before, immediately implies the lemma. □

Remark 3.21. The covariance kernel takes a particularly nice form in a fixed neighbourhood of a given point when $\varepsilon \to 0$. Indeed, note that if $x \in B(0, R)$ and $|(\rho x)_1| \ge \varepsilon$, then writing $x = \|x\|e_x$ where e_x is the unit vector in the direction of x, we have (letting e_1 denote the unit vector in the first coordinate),

$$g_{\varepsilon,R}((\rho x)_1) = \log(R/\langle \rho x, e_1\rangle) = \log(R/\|x\|) + \log(R/\langle \rho e_x, e_1\rangle).$$

When we integrate against $d\sigma$, we can take advantage of rotational symmetry to note that

$$C = \int_{\rho\in\mathcal{R}} \log(R/\langle \rho e_x, e_1\rangle) d\sigma$$

does not in fact depend on x.

Therefore for *any* $x \in B(0, R)$,

$$\lim_{\varepsilon\to 0} \mathbf{c}_{\varepsilon,R}(x, 0) = \log(R/\|x\|) + C.$$

It follows from this observation that in $B(0, R)$ that the function $x \mapsto \log(R/\|x\|) + C$ is positive definite in $B(0, R)$. We can get rid of the constant C by changing the value of R and so we deduce that

$$x \mapsto K(x) := \log(R/\|x\|) \text{ is positive definite in a small neighbourhood of } 0,$$

a fact which appears to have been first proved for all dimensions in [RV10a].

Remark 3.22. By contrast, note that $x \mapsto \tilde{K}(x) = \log_+(R/\|x\|)$ is positive definite in the *whole space* if and only if $d \le 3$: see Section 5.2 of [RV10b] for a nice proof based on Fourier transform.

When $d = 1$ or $d = 2$ one can also show that $\tilde{K}(x)$ is not only positive definite but of σ-positive type in the sense of Kahane: that is, it is a sum $\tilde{K}(x) = \sum_{n=1}^{\infty} K_n(x)$ where the summands K_n are not only positive definite functions, but also pointwise non-negative ($K_n(x) \ge 0$). When $d = 3$ it is an open question to determine whether $\tilde{K}(x)$ is σ-positive.

3.6 Multifractal spectrum

We now explain how Kahane's convexity inequality can be used to obtain various estimates on the moments of the mass of small balls, and in turn to the multifractal spectrum of Gaussian multiplicative chaos. We take h, θ as in Section 3.1, and we assume that $d = \mathbf{d}$ and the reference measure σ is Lebesgue measure for simplicity.

Theorem 3.23 (Scaling relation for Gaussian multiplicative chaos)**.** *Let $\gamma \in (0, \sqrt{2d})$. Let $B(r)$ be a ball of radius r that is at distance at least ε_0 from the boundary of the domain D, for some ε_0. Then uniformly over $r \in (0, \varepsilon_0)$, and over all such balls, and for any $q \in \mathbb{R}$ (including $q < 0$) such that $\mu_\varepsilon(B(0,1))^q$ is uniformly integrable in ε,*

$$\mathbb{E}(\mu(B(r))^q) \asymp r^{(d+\gamma^2/2)q - \gamma^2 q^2/2}, \tag{3.37}$$

where the implied constants depend only on q, ε_0 and γ. The function

$$\xi(q) = q(d + \gamma^2/2) - \gamma^2 q^2/2$$

*is called the **multifractal spectrum** of Gaussian multiplicative chaos.*

Remark 3.24. In the next section, we will see that the assumption on q is equivalent to

$$q < \frac{2d}{\gamma^2}.$$

At this stage we already know it at least for $0 \leq q < 1$.

Remark 3.25. What is a multifractal spectrum? The above theorem characterises the multifractal spectrum of Gaussian multiplicative chaos. To explain the terminology, it is useful to consider the opposite case of a *monofractal* object. For instance, Brownian motion is a monofractal because its behaviour is (to first order at least) described by a single fractal exponent, $\alpha = 1/2$. One way to say this is to observe that for all q

$$\mathbb{E}(|B_t|^q) \asymp t^{q/2}.$$

(A variety of exponents however can be obtained by considering logarithmic corrections, see e.g. [MP10]). The monofractality of Brownian motion is thus expressed through the fact that its moments have a power law behaviour where the exponent is *linear* in the order of the moment q. By contrast, note that the function ξ in Theorem 3.23 is **non-linear**, which is indicative of multifractal behaviour. That is, several fractal exponents (in fact, a whole spectrum of exponents) are needed to characterise the first order behaviour of Gaussian multiplicative chaos. Roughly speaking, the multifractal formalism developed among others in [Fal14] is what allows the data of a non-linear function such as the right hand side of (3.37) to be translated into a knowledge about the various fractal exponents and their relative importance.

Proof of Theorem 3.23. As hinted previously, the idea will be to compare h_ε to the scale-invariant field Y_ε constructed in the previous section. Note that by Lemma 3.19, there exist constants $a, b > 0$ such that

$$c_{\varepsilon,R}(x,y) - a \leq \mathbb{E}(h_\varepsilon(x)h_\varepsilon(y)) \leq c_\varepsilon(x,y) + b \tag{3.38}$$

As a result it will be possible to estimate the moments of $\mu(B(r))$ up to constants by computing those of $\tilde{\mu}(B(0,r))$, where $\tilde{\mu}$ is the chaos measure associated to Y. More precisely, from (3.38) and Kahane's convexity inequality (applied to the fields h_ε and $Y_\varepsilon + \mathcal{N}(0,a)$ in one direction and to the fields Y_ε and $h_\varepsilon + \mathcal{N}(0,b)$ in the other direction, with the function f taken to be the concave or convex function $x \mapsto x^q$), we get:

$$\mathbb{E}((\mu_\varepsilon(S))^q) \asymp \mathbb{E}((\tilde{\mu}_\varepsilon(S))^q) \tag{3.39}$$

for $S \subset D$, where the implicit constants depend only on a, b and $q \in \mathbb{R}$, and not on S or ε, and where

$$\tilde{\mu}_\varepsilon(z) = \exp(\gamma Y_\varepsilon(z) - (\gamma^2/2)\mathbb{E}(Y_\varepsilon(z)^2))dz.$$

It therefore suffices (also making use of the translation invariance of Y) to study the moments of $\tilde{\mu}_\varepsilon(B(0,r))$.

We now turn to the proof of (3.37). Note that $\mathbb{E}(Y_\varepsilon(x)^2) = \log(1/\varepsilon) + O(1)$. Fix $\varepsilon > 0$, and $\lambda = r < 1$. Then

$$\begin{aligned}\tilde{\mu}_{r\varepsilon}(B(0,r)) &\asymp \int_{B(0,r)} e^{\gamma Y_{r\varepsilon}(z)}(r\varepsilon)^{\gamma^2/2}dz \\ &\asymp Cr^{d+\gamma^2/2}\int_{B(0,1)} e^{\gamma Y_{r\varepsilon}(rw)}\varepsilon^{\gamma^2/2}dw\end{aligned}$$

by the change of variables $z = rw$. Hence by Corollary 3.20,

$$\tilde{\mu}_{r\varepsilon}(B(0,r)) \asymp r^{d+\gamma^2/2}e^{\gamma\Omega_r}\tilde{\mu}'_\varepsilon(B(0,1)) \tag{3.40}$$

where $\tilde{\mu}'$ is a copy of $\tilde{\mu}$ and Ω_r is an independent $\mathcal{N}(0,\log(1/r))$ random variable. Raising to the power q, taking expectations and using eq. (3.39), we get:

$$\begin{aligned}\mathbb{E}(\mu_{r\varepsilon}(B(r)^q)) &\asymp \mathbb{E}(\tilde{\mu}_{r\varepsilon}(B(0,r)^q)) \\ &= r^{q(d+\gamma^2/2)}\mathbb{E}(e^{\gamma q\Omega_r})\mathbb{E}(\tilde{\mu}_\varepsilon(B(0,1))^q) \\ &\asymp r^{\xi(q)}\mathbb{E}(\mu_\varepsilon(B(0,1)^q))\end{aligned} \tag{3.41}$$

where

$$\xi(q) = q(d + \gamma^2/2) - \gamma^2 q^2/2$$

is the multifractal spectrum from the theorem statement. Suppose now that q is such that $\mu_\varepsilon(B(0,1))^q$ is uniformly integrable. Then

$$\mathbb{E}(\mu(B(r))^q) \asymp r^{\xi(q)},$$

as desired. □

3.7 Positive moments of Gaussian multiplicative chaos

We continue our study of GMC initiated above in $\mathbb{R}^d$ with the reference measure σ taken to be te Lebesgue measure, for a logarithmically correlated field h satisfying the general assumptions of Section 3.1. Let μ be the associated GMC measure. The goal of this section will be to prove the following theorem on its moments. (See Section 3.8 for similar results where σ is allowed to be more general than Lebesgue measure).

Theorem 3.26. *Let $S \subset D$ be bounded and open, and suppose that $\sigma(dx) = dx$ is the Lebesgue measure on $\mathbb{R}^d$. Let $\gamma \in (0,2)$ and $q > 0$. Then $\mathbb{E}(\mu(S)^q) < \infty$ if*

$$q < \frac{2d}{\gamma^2}. \tag{3.42}$$

In fact, the theorem shows that $(\mu_\varepsilon(S))^q$ is uniformly integrable in ε, so that Theorem 3.23 applies to this range of values of q.

Before starting the proof of this theorem, we note that from Theorem 3.26,

$$\mathbb{P}(\mu(S) > t) \leq t^{-2d/\gamma^2 + o(1)}; \quad t \to \infty. \tag{3.43}$$

In fact, much more precise information is known about the tail at ∞: a lower bound matching this upper bound can be obtained so that it becomes an equality. In fact, the $o(1)$ term in the exponent can also be removed and a constant identified: in the case of the two-dimensional GFF this was done by Rhodes and Vargas [RV19], and the universality of this behaviour (including the calculation of the constant itself) was shown subsequently in a paper by Mo-Dick Wong [Won19].

Proof. Note that we already know uniform integrability of $\mu_\varepsilon(S)$ (Theorem 3.2) so we can assume $q > 1$. For simplicity (and without loss of generality) we assume that S is the unit cube in $\mathbb{R}^d$. Let $\mathcal{S}_m$ denote the mth level dyadic covering of the domain $\mathbb{R}^d$ by cubes $S_i, i \in \mathcal{S}_m$ of side-length 2^{-m}. Given $q < 2d/\gamma^2$, we define $n = n(q) \geq 2$ such that $n - 1 < q \leq n$. We will show by **induction** on n that

$$M_\varepsilon := \mathbb{E}(\mu_\varepsilon(S)^q)$$

is uniformly bounded.

Let us consider the case $n = 2$ first. We first subdivide the cubes of $\mathcal{S}_m$ into $2d$ disjoint groups so that no two cubes within any given group touch (including at the boundary); thus any two cubes within a given group are at distance at least 2^{-m} from one another. The reader should convince themselves that this is actually possible (it is a generalisation of the usual checkboard pattern for $\mathbb{Z}^2$). Let $\mathcal{S}'_m$ denote one of these $2d$ groups of cubes of side-length 2^{-m}.

We will now take advantage of some convexity properties, using the fact that $q/2 \leq 1$ (recall that $n = 2$ and $n - 1 < q \leq n$ by definition). We write, for given m,

$$\Big(\sum_{i \in \mathcal{S}'_m} \mu_\varepsilon(S \cap S_i) \Big)^q = \left(\sum_{i,j \in \mathcal{S}'_m} \mu_\varepsilon(S_i)\mu_\varepsilon(S_j) \right)^{q/2}$$

$$\leq \sum_{i,j\in\mathcal{S}'_m} \mu_\varepsilon(S_i)^{q/2}\mu_\varepsilon(S_j)^{q/2}, \tag{3.44}$$

where we have used the elementary fact that $(x+y)^\alpha \leq x^\alpha + y^\alpha$ if $x, y > 0$ and $\alpha \in (0,1)$. (This is easily proven by writing $(x+y)^\alpha - x^\alpha = \int_x^{x+y} \alpha t^{\alpha-1} dt \leq \int_0^y \alpha t^{\alpha-1}\, dt = y^\alpha$, since the integrand $\alpha t^{\alpha-1}$ is decreasing in t).

We consider the on-diagonal and off-diagonal terms in (3.44) separately. We start with the on-diagonal terms (the estimate in this case works for general $q > 0$ so is not restricted to the case $n = 2$):

Lemma 3.27. *Assume the set-up of Theorem 3.26. Then there exist a constant c_q such that for all sufficiently large m, and for all $\varepsilon > 0$,*

$$\mathbb{E}\Big(\Big(\sum_{i\in\mathcal{S}'_m} \mu_\varepsilon(S_i)^q\Big)\Big) \leq c_q 2^{dm-\xi(q)m}\mathbb{E}(\mu_{\varepsilon 2^m}(S)^q). \tag{3.45}$$

Proof. By (3.41), applied with $r = 2^{-m}$, we have for each $i \in \mathcal{S}'_m$,

$$\mathbb{E}(\mu_\varepsilon(S_i)^q) \leq c_q 2^{-\xi(q)m}\mathbb{E}((\mu_{\varepsilon 2^m}(S))^q).$$

Since there are at most 2^{dm} terms in this sum, we deduce the lemma. □

For the off-diagonal terms, we simply observe that in the case where the two indices are distinct:

Lemma 3.28. *Assume the set-up of Theorem 3.26. Then for any fixed m and $q < 2$, there exists a constant $C_{m,q}$ independent of ε such that*

$$\mathbb{E}\Big(\sum_{i\neq j\in\mathcal{S}'_m} \mu_\varepsilon(S_i)^{q/2}\mu_\varepsilon(S_j)^{q/2}\Big) \leq C_{m,q}.$$

Proof. Note that by Jensen's inequality (since $q/2 \leq 1$),

$$\mathbb{E}\big(\mu_\varepsilon(S_i)^{q/2}\mu_\varepsilon(S_j)^{q/2}\big) \leq \mathbb{E}\big(\mu_\varepsilon(S_i)\mu_\varepsilon(S_j)\big)^{q/2}$$

for all $i \neq j \in \mathcal{S}'_m$. The expectation can easily be computed, and we have for some constant c_m,

$$\lim_{\varepsilon\to 0}\mathbb{E}\big(\mu_\varepsilon(S_i)\mu_\varepsilon(S_j)\big) \leq \int_{x\in S_i, y\in S_j} e^{\gamma^2 K(x,y)}dxdy \leq c_m < \infty$$

since the squares S_i and S_j are at distance at least 2^{-m} from one another. Taking the qth power and summing over all terms $i \neq j \in \mathcal{S}'_m$ gives the lemma. □

We put these two lemmas together as follows. First, note that for $q < 2d/\gamma^2$, $2d - \xi(q) < 0$. We can therefore choose m large enough that $c_q 2^{dm-\xi(q)m} < 1/(2d)^q$, where c_q is as in Lemma 3.27. From (3.44) we obtain

$$\mathbb{E}\Big(\Big(\sum_{i\in\mathcal{S}'_m} \mu(S_i)\Big)^q\Big) \leq \frac{1}{(2d)^q}\mathbb{E}(\mu_{2^m\varepsilon}(S)^q) + C_{m,q},$$

where $C_{m,q}$ comes from Lemma 3.28. Adding the contributions from all $2d$ groups (and using the fact that $(x_1 + \ldots + x_{2d})^q \leq (2d)^{q-1}(x_1^q + \ldots + x_{2d}^q)$ by convexity),

$$\mathbb{E}\big(\big(\sum_{i\in\mathcal{S}_m} \mu(S_i)\big)^q\big) \leq \frac{1}{2d}\mathbb{E}(\mu_{2^m\varepsilon}(S)^q) + (2d)^{q-1}C_{m,q}.$$

Therefore, recalling that $M_\varepsilon = \mathbb{E}(\mu_\varepsilon(S)^q)$, we have

$$M_\varepsilon \leq \frac{1}{2d}M_{2^m\varepsilon} + (2d)^{q-1}C_{m,q}.$$

Taking the sup over $\varepsilon > \varepsilon_0$, and since $2^m\varepsilon \geq \varepsilon$, we get

$$\sup_{\varepsilon>\varepsilon_0} M_\varepsilon \leq \frac{1}{2d}\sup_{\varepsilon>\varepsilon_0} M_\varepsilon + (2d)^{q-1}C_{m,q}$$

and hence

$$\sup_{\varepsilon>\varepsilon_0} M_\varepsilon \leq \frac{(2d)^q}{2d-1}C_{m,q}.$$

We conclude proof for $q \in (1,2]$ i.e., $n = 2$, by letting $\varepsilon_0 \to 0$ and Fatou's lemma.

We now consider the general case, which is in fact very similar to when $n = 2$. We use the fact that $q/n \leq 1$ and thus, arguing as in (3.44),

$$\Big(\sum_{i\in\mathcal{S}'_m} \mu_\varepsilon(S\cap S_i)\Big)^q \leq \sum_{i_1,\ldots,i_n\in\mathcal{S}'_m} \mu_\varepsilon(S_{i_1})^{q/n}\ldots\mu_\varepsilon(S_{i_n})^{q/n} \tag{3.46}$$

As before, we consider separately the on-diagonal (when all indices are equal) and off-diagonal terms. The on-diagonal terms were already estimated in Lemma 3.27, and we have the same upper bound (3.45) for all sufficiently large m and all $\varepsilon > 0$. For the off-diagonal terms, we obtain the following estimate.

Lemma 3.29. *Assume the set-up of Theorem 3.26. Then for any fixed m and $q < 2d/\gamma^2$, there exists a constant $C_{m,q}$ independent of ε such that*

$$\mathbb{E}\Big(\sum_{i_1,\ldots,i_n\in\mathcal{S}'_m : i_1\neq i_2} \mu_\varepsilon(S_{i_1})^{q/n}\ldots\mu_\varepsilon(S_{i_n})^{q/n}\Big) \leq C_{m,q}.$$

Proof. Note that by Jensen's inequality (since $q/n \leq 1$), if $i_1 \neq i_2 \in \mathcal{S}'_m$,

$$\mathbb{E}\big(\mu_\varepsilon(S_{i_1})^{q/n}\ldots\mu_\varepsilon(S_{i_n})^{q/n}\big) \leq \mathbb{E}\big(\mu_\varepsilon(S_{i_1})\ldots\mu_\varepsilon(S_{i_n})\big)^{q/n}$$

As before, this expectation can be computed exactly. To begin with, we rewrite the index set $\{i_1,\ldots,i_n\}$ in a way that takes into account which indices are equal and which are distinct. Thus let $\{i_1,\ldots,i_n\} = \{j_1,\ldots,j_p\}$ where the j_k are pairwise distinct and $2 \leq p \leq n$ (since at least two indices are distinct). Call m_k the multiplicity of j_k in $\{i_1,\ldots,i_n\}$, i.e., the number

of times j_k is present in that set, so that $m_1 + \ldots + m_p = n$ (with $m_k \geq 1$ by assumption). Then

$$\mathbb{E}\big(\mu_\varepsilon(S_{i_1}) \ldots \mu_\varepsilon(S_{i_n})\big) = \int_{x_1 \in S_{i_1}} \ldots \int_{x_n \in S_{i_n}} e^{(\gamma^2/2)\sum_{1 \leq k \neq \ell \leq n} K_\varepsilon(x_k, x_\ell)} dx_1 \ldots dx_n.$$

When $x_k \in S_{i_k}, x_\ell \in S_{i_\ell}$ and $S_{i_k} \neq S_{i_\ell}$, the term $K(x_k, x_\ell) = -\log|x_k - x_\ell| + O(1)$ is bounded above by a constant c_m since the cubes are separated by a minimum distance of 2^{-m}. Hence

$$\begin{aligned}\mathbb{E}\big(\mu_\varepsilon(S_{i_1}) \ldots \mu_\varepsilon(S_{i_n})\big) &\leq c'_m \prod_{k=1}^{p} \int_{S_{j_k}} \ldots \int_{S_{j_k}} e^{(\gamma^2/2)\sum_{1 \leq k \neq \ell \leq m_k} K_\varepsilon(x_k, x_\ell)} dx_1 \ldots dx_{m_k} \\ &= c'_m \prod_{k=1}^{p} \mathbb{E}(\big(\mu_\varepsilon(S_{j_k})\big)^{m_k})\end{aligned}$$

Now, since $m_k \leq n-1$ (as there are at least two distinct indices in the set $\{i_1, \ldots, i_n\}$), and since $S_{j_k} \subset S$, we have that

$$\mathbb{E}(\big(\mu_\varepsilon(S_{j_k})\big)^{m_k}) \leq \mathbb{E}(\mu_\varepsilon(S)^{n-1})$$

which, by the induction hypothesis, is uniformly bounded in ε, by a constant depending only on m and q. This concludes the proof of the lemma. □

Putting together (3.45) and Lemma 3.29, we conclude the proof that M_ε is uniformly bounded for arbitrary $q < 2d/\gamma^2$, as in the case $q < 2 \wedge (2d/\gamma^2)$. This finishes the proof of Theorem 3.26. □

We complement Theorem 3.26 with two results. The first one shows that the condition $q < 2d/\gamma^2$ is sharp for the finiteness of the moment of order $q > 0$. The second will show a partial result in the general framework of Gaussian multiplicative chaos with respect to a d-dimensional reference measure σ (i.e., satisfying (3.2)). We start with the first result.

Proposition 3.30. *Assume the set-up of Theorem 3.26 (in particular, that $\sigma(dx) = dx$ is the Lebesgue measure on $\mathbb{R}^d$). Let $q > 2d/\gamma^2$. Then*

$$\mathbb{E}(\mu(S)^q) = \infty.$$

Proof. The proof argues by contradiction, and has the same flavour as Theorem 3.26 but is much simpler (essentially, we can ignore the off-diagonal term). Suppose that for some $q > 2d/\gamma^2$, $\mathbb{E}(\mu(S)^q) < \infty$. By Kahane's inequality, there is no loss of generality in assuming that the Gaussian field h is in fact an exactly scale-invariant field X satisfying Corollary 3.20. Then for any cube S_i of side-length 2^{-m}, by (3.41) (or more precisely (3.40)),

$$\mathbb{E}((\mu(S_i))^q) \asymp 2^{-m\xi(q)} \mathbb{E}((\mu(S))^q).$$

On the other hand, keeping the same notations as in the proof of Theorem 3.26, and since $(x+y)^q \geq x^q + y^q$ for $q > 1$ and $x, y > 0$,

$$\mu(S)^q \geq \sum_{i \in \mathcal{S}_m} \mu(S_i)^q.$$

Hence, taking expectations,

$$\mathbb{E}(\mu(S)^q) \gtrsim 2^{dm - \xi(q)m} \mathbb{E}((\mu(S))^q)$$

However, when $q > 2d/\gamma^2$, we have that $d - \xi(q) > 0$. Since m is arbitrary and the implicit constant does not depend on m, we get the desired contradiction. □

3.8 Positive moments for general reference measures

We now introduce the second result complementing Theorem 3.26, which is an extension of Theorem 3.26 to the general setup of Gaussian multiplicative chaos relative to a $\mathbf{d}$-dimensional reference measure σ (satisfying (3.2)). In order to not make the exposition too cumbersome, we limit the proof to the case where $q < (2\mathbf{d}/\gamma^2) \wedge 2$ (hence, at least in the L^1 regime where $\gamma \in [\sqrt{\mathbf{d}}, \sqrt{2\mathbf{d}})$, there is no loss of generality at all).

Before doing so it may be useful to explain where the previous proof breaks down if σ is not Lebesgue measure. The main issue lies in the scaling argument of Lemma 3.27; when we consider a cube of S_i of $\mathcal{S}'_m$ (side-length 2^{-m}), blowing this up by a factor 2^m will of course still produce a cube of unit side-length, but the Gaussian multiplicative chaos is now with respect to a reference measure which is no longer σ, but instead reflects the local behaviour of σ near the cube S_i. For very inhomogeneous fractals this behaviour could be wildly different, and so the inequality in that Lemma has no reason to hold true.

Instead we will need a different approach that accounts for the possible inhomogeneities of the fractal supporting the reference measure σ. The proof below comes from work (written roughly in parallel with these notes) in [BSS14] and is reproduced here with permission of this paper's co-authors. It is based on Girsanov's theorem.

Proposition 3.31. *Let $S \subset D$ be bounded and open, and suppose that the reference measure σ satisfies the dimensionality condition (3.2). Then if $0 < q < 2 \wedge (2\mathbf{d}/\gamma^2)$,*

$$\mathbb{E}(\mu(S)^q) < \infty \tag{3.47}$$

and moreover, $\mu_\varepsilon(S)$ converges to $\mu(S)$ in L^q.

Proof. Again, we can assume without loss of generality that $q > 1$. Set $\delta = q - 1 \in (0, 1)$. Write

$$\mathbb{E}(\mu(S)^q) = \mathbb{E}(\mu(S)\mu(S)^\delta) = \mathbb{E}(\mu(S))\mathbb{E}^*(\mu(S)^\delta) = \sigma(S)\mathbb{E}^*(\mu(S)^\delta)$$

where $\mathbb{P}^*$ denotes the law of the field biased by $\mu(S)$. Using Girsanov's theorem (exactly as in (2.6)), we can rewrite this as

$$\mathbb{E}^*(\mu(S)^\delta) = \int_S \sigma(dx) \mathbb{E}\left(\left(\int_S e^{\gamma^2 K(x,y)} \mu(dy) \right)^\delta \right).$$

For each $n \geq 0$, let $A_n(x)$ denote the annulus at distance between 2^{-n} and 2^{-n-1} from x; that is, $A_n(x) = \{y : |y-x| \in [2^{-n-1}, 2^{-n})\}$. Then using the fact that $K(x,y) = -\log|x-y| + O(1)$ and the inequality $(a_1 + \ldots + a_n)^\delta \leq a_1^\delta + \ldots + a_n^\delta$ for $\delta < 1$ and $a_i > 0$, we see that

$$\begin{aligned}\mathbb{E}^*(\mu(S)^\delta) &\leq C \sum_{n=0}^{\infty} \int_S \sigma(dx) \mathbb{E}\left(\left(\int_{A_n(x)} e^{-\gamma^2 \log|x-y|} \mu(dy) \right)^\delta \right) \\ &\leq C \sum_{n=0}^{\infty} 2^{n\gamma^2\delta} \int_S \sigma(dx) \mathbb{E}(\mu(A_n(x))^\delta).\end{aligned}$$

For each fixed x, consider a field X which is exactly scale invariant around x as in Section 3.5.2: hence, for any $\lambda < 1$,

$$(X(x + \lambda z))_{z \in S} = (\tilde{X}(z) + \Omega_{-\log\lambda})_{z \in S}$$

where $\tilde{X}$ has the same law as X and Ω_r is a Gaussian with variance r independent of $\tilde{X}$. Write X_ε for the field truncated at level ε, as in Section 3.5.2.

Set $\lambda = 2^{-n} \leq 1$. Denote by $\sigma_{\lambda,x}(dz)$ the image measure of σ under the map $y = x + \lambda z \mapsto z$ (so that the total mass $\sigma_{\lambda,x}(A_1(0)) = \sigma(A_n(x))$). By applying Corollary 3.20 and changing variables $y \mapsto z$, we obtain:

$$\begin{aligned}\mathbb{E}\left(\left(\int_{A_n(x)} e^{\gamma X_{\lambda\varepsilon}(y)} (\lambda\varepsilon)^{\gamma^2/2} \sigma(dy) \right)^\delta \right) &\leq \lambda^{\gamma^2\delta/2} \mathbb{E}\left(\left(\int_{A_1(0)} e^{\gamma X_\varepsilon(x+\lambda z)} \varepsilon^{\gamma^2/2} \sigma_{\lambda,x}(dz) \right)^\delta \right) \\ &= \lambda^{\gamma^2\delta/2} \mathbb{E}\left(e^{\delta\gamma\Omega_{-\log\lambda}} \left(\int_{A_1(0)} e^{\gamma \tilde{X}_\varepsilon(z)} \varepsilon^{\gamma^2/2} \sigma_{\lambda,x}(dz) \right)^\delta \right) \\ &= \lambda^{\gamma^2\delta/2} e^{-\delta^2\gamma^2 \log(\lambda)/2} \mathbb{E}\left(\left(\int_{A_1(0)} e^{\gamma \tilde{X}_\varepsilon(z)} \varepsilon^{\gamma^2/2} \sigma_{\lambda,x}(dz) \right)^\delta \right) \\ &\leq \lambda^{\gamma^2\delta/2 - \delta^2\gamma^2/2} \mathbb{E}\left(\left(\int_{A_1(0)} e^{\gamma \tilde{X}_\varepsilon(z)} \varepsilon^{\gamma^2/2} \sigma_{\lambda,x}(dz) \right) \right)^\delta,\end{aligned}$$

where the last inequality is by Jensen's inequality since $\delta < 1$.

By Kahane's inequality (Theorem 3.15), there exists an absolute constant $C > 0$ such that

$$\mathbb{E}(\mu_{\varepsilon 2^{-n}}(A_n(x))^\delta) \leq C \lambda^{\gamma^2\delta/2 - \delta^2\gamma^2/2} \sigma(A_n(x))^\delta.$$

Letting $\varepsilon \to 0$, we get that for any $n \geq 0$,

$$\mathbb{E}(\mu(A_n(x))^\delta) \leq C 2^{-n(\gamma^2\delta/2 - \delta^2\gamma^2/2)} \sigma(A_n(x))^\delta.$$

We deduce that

$$\mathbb{E}^*(\mu(S)^\delta) \leq C \sum_n 2^{n(\gamma^2\delta/2 + \delta^2\gamma^2/2)} \int \sigma(dx) \sigma(A_n(x))^\delta.$$

To estimate the last integral, let $\bar{y}$ and $\bar{x}$ be two independent points distributed according to $\sigma(\cdot \cap S)/\sigma(S)$. Then note that

$$\sigma(A_n(x)) \leq \sigma(S)\mathbb{P}(|\bar{y} - x| \leq 2^{-n})$$

so that by Jensen's inequality again (as $\delta < 1$),

$$\begin{aligned}\int_S \sigma(dx)\sigma(A_n(x))^\delta &\leq \int_S \sigma(dx)\sigma(S)^\delta \mathbb{P}\Big(|\bar{y} - \bar{x}| \leq 2^{-n}\Big|\bar{x} = x\Big)^\delta \\ &\leq \sigma(S)^{\delta+1}\mathbb{E}(\mathbb{P}\Big(|\bar{x} - \bar{y}| < 2^{-n}\Big|\bar{x}\Big)^\delta) \\ &\leq \sigma(S)^{\delta+1}\mathbb{P}(|\bar{x} - \bar{y}| \leq 2^{-n})^\delta \\ &\leq \sigma(S)^{\delta+1}\mathbb{E}(|\bar{x} - \bar{y}|^{-\mathbf{d}})^\delta 2^{-n\mathbf{d}\delta}.\end{aligned}$$

by Markov's inequality. Now $\delta < 2\mathbf{d}/\gamma^2 - 1$ implies that $\gamma^2\delta/2 + \delta^2\gamma^2/2 - \mathbf{d}\delta = \delta(\gamma^2/2 - \mathbf{d} + \gamma^2\delta/2) < 0$. Putting everything together, we can find $c = c(\mathbf{d}, \gamma, \delta)$ such that

$$\mathbb{E}(\mu(S)^q) = \sigma(S)\mathbb{E}^*(\mu(S)^\delta) \leq c(\delta)\mathbb{E}(|\bar{x} - \bar{y}|^{-\mathbf{d}})^\delta < \infty$$

by (3.3). This concludes the proof. □

3.9 Negative moments of Gaussian multiplicative chaos

We now turn our attention to negative moments of the chaos measures. We will first show that in the general set up, $\mu(S)$ admits moments of order $q \in [q_0, 0]$ for some $q_0 < 0$. This proof is based on a similar argument appearing in [GHSS18]. We will then explain how to bootstrap this to get existence of all negative moments (see Theorem 3.36). Note that, in particular, this implies strict positivity of the measures with probability one.

We work in the general setting: σ is a Radon measure with dimension at least $\mathbf{d}$ ($0 < \mathbf{d} \leq d$) and $0 \leq \gamma < \sqrt{2\mathbf{d}}$. The first ingredient we will need concerns the β-dimensional energy of the measure μ.

Lemma 3.32. *Assume that $\sigma(D) > 0$. Suppose that $0 \leq \beta < \mathbf{d} \vee \sqrt{2\mathbf{d}}\gamma$, and x is a point chosen from the measure $\sigma(dx)$ in D (normalised to be a probability measure), independently of the field. Then*

$$\int_D |x - y|^{-\beta}\mu(dy) < \infty$$

a.s. and in fact, has finite rth moment for all $r > 0$ small enough.

Proof. If $\beta < \mathbf{d}$ then this energy will have finite expectation by assumption (3.2). So let us assume that $\gamma > \sqrt{\mathbf{d}/2}$, and $\beta < \sqrt{2\mathbf{d}}\gamma$. This means that for $r > 0$ small enough we will have $1 > 1 - r > 1/2 \vee \beta^2/(2\mathbf{d}\gamma^2)$. For such an r, we write

$$\mathbb{E}\left((\int_D |x - y|^{-\beta}\mu(dy))^r\right) \lesssim \mathbb{E}\left(\int_D \left(\int_D \mathrm{e}^{\beta K(x,y)}\mu(dy)\right)^r \frac{\sigma(dx)}{\sigma(D)}\right)$$

$$\lesssim \mathbb{E}(\mu(D)^r \mu_{\gamma^{-1}\beta}(D))$$

where $\mu_{\gamma^{-1}\beta}$ is the chaos measure of the field with parameter $\gamma^{-1}\beta$ rather than γ. This last inequality follows from Girsanov's theorem, as in the proof of Proposition 3.31.

Now by Hölder's inequality with $p = r^{-1}$ and $q = (1-r)^{-1}$, the above is less than or equal to

$$\mathbb{E}(\mu(D))^r \mathbb{E}(\mu_{\gamma^{-1}\beta}(D)^{\frac{1}{1-r}})^{1-r}.$$

By Proposition 3.31, this is finite as long as $(1-r)^{-1} \leq 2 \wedge 2\mathbf{d}/(\gamma^{-1}\beta)^2$, which is exactly our assumption on r. □

Corollary 3.33. *Take the same set up as in Lemma 3.32. Then there exists M large enough, depending only on γ and $\mathbf{d}$, such that*

$$\mathbb{P}(E_s) := \mathbb{P}\left(\int_{B(x,s^{-M})} \mathrm{e}^{\gamma^2 K(x,y)} \mu(dy) \leq \frac{1}{s}\right) \geq \frac{1}{2}$$

for all s sufficiently large.

Proof. By the assumption on K, we know that $\mathrm{e}^{\gamma^2 K(x,y)} \leq c|x-y|^{-\gamma^2}$ for some $c < \infty$. Writing $|x-y|^{-\gamma^2} = |x-y|^{-(\gamma^2+2/M)}|x-y|^{2/M}$, we therefore have that

$$\mathrm{e}^{\gamma^2 K(x,y)} \leq cs^{-2}|x-y|^{-(\gamma^2+2/M)} \text{ for all } y \in B(x, s^{-M}).$$

Hence

$$\mathbb{P}\left(\int_{B(x,s^{-M})} \mathrm{e}^{\gamma^2 K(x,y)} \mu(dy) \leq \frac{1}{s}\right) \geq \mathbb{P}\left(\int_D |x-y|^{\gamma^2+2/M} \mu(dy) \leq \frac{s}{c}\right).$$

If M is such that $\gamma^2 + 2/M < \mathbf{d} \vee \sqrt{2\mathbf{d}}\gamma$, then by Lemma 3.32, the right hand side converges monotonically to 1 as $s \to \infty$. □

From here the key observation is the following. If we write $\mathbb{P}^*$ for the field biased by $\mu(D)$ as before, then for $s > 0$,

$$\mathbb{E}^*(\exp(-s\mu(D))) = \sigma(D)^{-1}\mathbb{E}(\mu(D)\exp(-s\mu(D))) \leq \frac{e^{-1}}{\sigma(D)s}$$

simply because $x\mathrm{e}^{-sx} \leq e^{-1}/s$ for all positive x, s. This says that under $\mathbb{P}^*$, $\mu(D)$ is unlikely to be too small. Of course we would actually like such a statement under $\mathbb{P}$. The trouble is that the field has an extra log singularity under $\mathbb{P}^*$, and so it could be this that saves $\mu(D)$ from being very small. The work now is essentially to rule this out using Corollary 3.33.

Namely, we observe that if E_s is the event in Corollary 3.33, then

$$\mathbb{E}(\exp(-s^{M\gamma^2}\mu(D))\mathbf{1}_{E_s}) \leq \frac{C}{s\sigma(D)} \tag{3.48}$$

for some $C, s_0 < \infty$ (depending only on $\gamma, \mathbf{d}$) and all $s \geq s_0$. Indeed, by Girsanov again,

$$\mathbb{E}^*(\exp(-s\mu(D))) = \mathbb{E}(\exp(-s\int_D e^{\gamma^2 K(x,y)}\mu(dy)))$$

where under $\mathbb{P}$, x is a point chosen according to σ, independently of the field. Moreover, on the event E_s,

$$s\int_D \mathrm{e}^{\gamma^2 K(x,y)}\mu(dy) \leq 1 + cs^{M\gamma^2}\mu(D)$$

for some $c < \infty$. This implies that $\exp(-s^{M\gamma^2}\mu(D))\mathbf{1}_{E_s} \leq e^c \exp(-(s/c)\int_D e^{\gamma^2 K(x,y)}\mu(dy))$, and so (3.48) holds.

Note that if it weren't for the indicator function in (3.48), this would imply that $\mu(D)$ has *some* finite negative moments. On the other hand, we have shown in Corollary 3.33 that the event in the indicator function is rather likely. Putting these ideas together more precisely, we obtain the following.

Proposition 3.34. *Assume that $\sigma(D) > 0$ and $0 \leq \gamma < \sqrt{2\mathbf{d}}$. For some $q_0 < 0$, depending only on γ and $\mathbf{d}$, it holds that $\mathbb{E}(\mu(D)^{q_0}) < \infty$.*

Proof. Let us first observe that, without loss of generality, we may assume that $K(x,y) \geq 0$ for all $x, y \in D$. Indeed, we can always find some $D' \subset D$ with $\sigma(D') > 0$ and $K(x,y) \geq 0$ for all $x, y \in D'$ (since K diverges logarithmically near the diagonal), and then it clearly suffices to show that $\mathbb{E}(\mu(D')^{q_0}) < \infty$. Note that σ also has dimension at least $\mathbf{d}$ when restricted to D'.

The advantage of assuming this setup, is that we can make use of the following tool (see e.g., [Pit82]):

Theorem 3.35 (FKG inequality)**.** *Let $(Z(x))_{x\in D}$ be an a.s. continuous centred Gaussian field on $U \subset \mathbb{R}^d$ with $\mathbb{E}(Z(x)Z(y)) \geq 0$ for all $x, y \in U$. Then, if f, g are two bounded, increasing measurable functions,*

$$\mathbb{E}\big(f((Z(x))_{x\in U})g((Z(x))_{x\in U})\big) \geq \mathbb{E}\big(f((Z(x))_{x\in U})\big)\mathbb{E}\big(g((Z(x))_{x\in U})\big).$$

To apply this, we need to work with continuous fields, so let us consider the measure μ_ε, and denote by E_s^ε the event E_s with μ replaced by μ_ε (and we still define E_s^ε in terms of a point that is sampled independently of the field and with probability proportional to $\sigma(dx)$). Since X_ε is a.s. continuous and the functions $\mathbf{1}_{E_s^\varepsilon}$ and $\exp(-s^{M\gamma^2}\mu_\varepsilon(D))$ are both bounded, decreasing functions of the field, we can apply FKG to see that

$$\mathbb{E}(\exp(-s^{M\gamma^2}\mu_\varepsilon(D))\mathbf{1}_{E_s^\varepsilon}) \geq \mathbb{E}(\exp(-s^{M\gamma^2}\mu_\varepsilon(D)))\mathbb{P}(E_s^\varepsilon)$$

for all ε. (Recall that $\mathbb{E}$ is over the field as well as the independent random point x, so we actually apply the FKG inequality conditionally given x, then note that the first term in the right hand does not depend on x). By dominated convergence, we therefore obtain that

$$\mathbb{E}(\exp(-s^{M\gamma^2}\mu(D))\mathbf{1}_{E_s}) \geq \mathbb{E}(\exp(-s^{M\gamma^2}\mu(D)))\mathbb{P}(E_s).$$

Hence by Corollary 3.33 and (3.48), for M, s_0 large enough (depending only on $\gamma, \mathbf{d}$):

$$\mathbb{E}(\exp(-s^{M\gamma^2}\mu(D))) \leq \frac{2C}{s\sigma(D)} \qquad \forall s \geq s_0. \tag{3.49}$$

Finally, since $y^{-p} = \Gamma(p)^{-1}\int_0^\infty t^{p-1}\exp(-ty)\,dt$ for $p > 0$, this implies that as long as $(1/M\gamma^2) > p$, $\mathbb{E}(\mu(D)^{-p}) < \infty$. Note that this only depends on M, γ, so since M is a function of γ and $\mathbf{d}$, the obtained q_0 also depends only on γ and $\mathbf{d}$. □

Now we explain how to extend this to all negative moments, using an iteration procedure. This idea first appeared in the setting of multiplicative cascade measures (a simple case of multiplicative chaos) in [Mol96].

Theorem 3.36. *Suppose that $\sigma(D) > 0$. Then*

$$\mathbb{E}(\mu(D)^q) < \infty$$

for all $q < 0$.

We emphasise that we need only our standing assumptions on the measure σ here (as long as $\sigma(D) > 0$), and that there are no restrictions on d or $\mathbf{d} > 0$.

Proof. To begin with, note that since σ does not have any atoms, we can find two distinct points x_1, x_2 in the support of σ. Therefore we can find open sets D_1 and D_2 such that $x_1 \in D_1, x_2 \in D_2$, $\bar{D}_1 \cap \bar{D}_2 = \emptyset$ and $\sigma(D_1)\sigma(D_2) > 0$. Furthermore, by the assumption on K (more precisely, the continuity of g), we may assume that $K(x, y) \leq C$ whenever $x \in D_1$, $y \in D_2$.

The key point is that by Proposition 3.34, there exists $q_0 < 0$ such that $\mathbb{E}(\mu(D_i)^q) < \infty$ for all $q \in [q_0, 0]$ and $i = 1, 2$. Indeed we've seen that q_0 depends only on $\mathbf{d}, \gamma$, as long as the base measure has strictly positive mass.

The idea to make use of this is to note the trivial bound $\mu(D) \geq \mu(D_1) + \mu(D_2)$, and then apply the AM-GM inequality to see that $\mu(D) \geq \sqrt{\mu(D_1)\mu(D_2)}$. This gives that

$$\mathbb{E}(\mu(D)^q) \leq \mathbb{E}(\mu(D_1)^{q/2}\mu(D_2)^{q/2})$$

for $q < 0$. If $\mu(D_1)$ and $\mu(D_2)$ were independent, we could factorise the right hand side and choose $q = 2q_0$, therefore showing that negative moments exist with orders in the larger interval $[2q_0, 0]$. We could then iterate to get all negative moments.

The problem of course is that they are not actually independent. To get around this we will use the assumption that $K(x, y) \leq C$ for $x \in D_1, y \in D_2$, together with Kahane's inequality.

More precisely, let us denote our field restricted to $D_1 \cup D_2$ by X. Let us also define a Gaussian field Y on $D_1 \cup D_2$ by setting it equal to $Y_1 + Y_2 + Z$ where: Y_1, Y_2 are independent; Y_1 has the law of $X|_{D_1}$ on D_1 and is 0 on D_2; Y_2 has the law of $X|_{D_2}$ on D_2 and is 0 on D_1; and Z is an independent normal random variable with variance C. Then the covariance kernel of Y dominates (pointwise) the covariance kernel of X. Since polynomials of negative

order are convex, we can apply Kahane's inequality (Theorem 3.15) (and a limiting argument so that we can compare the respective GMC measures) to obtain that

$$\begin{aligned}\mathbb{E}((\mu(D_1)+\mu(D_2))^q) &\leq \mathbb{E}((\mu_Y(D_1)+\mu_Y(D_2))^q)\\ &\leq \mathbb{E}(\mu_Y(D_1)^{q/2}\mu_Y(D_2)^{q/2})\\ &= \mathbb{E}(\mathrm{e}^{\frac{q}{2}(\gamma Z-\frac{\gamma^2}{2}C)})\mathbb{E}(\mu_Y(D_1)^{q/2})\mathbb{E}(\mu_Y(D_2)^{q/2}),\end{aligned}$$

where we have applied AM-GM in the second line. By construction, if $q \in [2q_0, 0]$, the right hand side is finite. So we obtain that $\mathbb{E}(\mu(D)^q) < \infty$ for all $q \in [2q_0, 0]$. Repeating the argument one obtains the existence of any negative moment. □

Since $\mu(D)$ has finite negative moments of all orders (as shown by the previous theorem), we deduce that the tail at zero of $\mu(D)$ decays faster than any polynomial. It is natural to wonder whether the decay can be characterised precisely. A lognormal upper bound for this decay (meaning, $\mathbb{P}(\mu(D) \leq \delta) \leq \exp(-c(\log 1/\delta)^2)$ for some $c > 0$) was first established in [DS11], see also [Aru17]. In some one-dimensional cases of Gaussian Multiplicative Chaos, the exact law of the total mass is in fact known (this was obtained by Remy [Rem20], proving a well known conjecture of Fyodorov and Bouchaud [FB08]). In exercise (7), we propose a lognormal lower bound valid in great generality.

3.10 KPZ theorem

In this section, we consider the Gaussian free field with zero boundary conditions in a domain $D \subset \mathbb{R}^2$. The KPZ formula relates the "quantum" and "Euclidean" sizes of a given set A, which is either deterministic, or random but independent of the field. This often has a particularly natural interpretation in the context of discrete random planar maps and critical exponents; see Section 3.13. Concrete examples are given in Chapter 4.

We will first formulate the KPZ theorem using the framework of Rhodes and Vargas [RV11]. This article appeared simultaneously with (and independently from) the paper by Duplantier and Sheffield [DS11]. The results of these two papers are similar in spirit, but the version we present here is a bit easier to state, and in fact stronger. The formulation (and sketch of proof) corresponding to [DS11] will be given in Section 3.12. We will also include a version, due to Aru [Aru15]. Although this last statement is weaker, its proof is completely straightforward given our earlier work.

We first introduce the notion of *scaling exponent* of a set A (in the sense of [RV11]), starting with the Euclidean version. Let $A \subset D$ be a fixed Borel set and write $d_H(A)$ for the (Euclidean) Hausdorff dimension of A. Since $0 \leq d_H(A) \leq 2$, we may write

$$d_H(A) = 2(1-x), \tag{3.50}$$

for $x \in [0, 1]$. The number x is called the **(Euclidean) scaling exponent** of A.

We now define the quantum analogue of the scaling exponent. Let

$$C_\delta(A) := \inf\{\sum_i \mu(B(x_i, r_i))^\delta : \{B(x_i, r_i)\}_i \text{ is a cover of } A\},$$

so that $C_\delta(A)$ can be viewed as a (multiple) of the quantum Hausdorff content of A. We now define

$$d_{H,\gamma}(A) = \inf\{\delta > 0 : C_\delta(A) = 0\} \in [0,1]$$

and call $d_{H,\gamma}(A)$ its "quantum Hausdorff dimension". Finally, we define the **quantum scaling exponent** Δ by

$$\Delta = 1 - d_{H,\gamma}(A).$$

The terms "quantum Hausdorff dimension" and content should perhaps be qualified, for the following reasons.

1. Although it does not feature in these notes, a random metric associated with $e^{\gamma h}$ (h a GFF in D) has recently been constructed in a series of works culminating with [DDDF19, GM19b, GM19a]. The Hausdorff dimension d_γ of D equipped with this random metric is currently unknown, except for the special case $\{\gamma = \sqrt{8/3},\ d_\gamma = 4\}$. The general bound $d_\gamma > 2$ is also known, as well as more precise estimates: see [DG20, GP20].

2. Under this random metric, the actual value of the Hausdorff dimension of $A \subset D$ is then given by

   $$\dim_\gamma(A) = d_\gamma(1 - \Delta).$$

   Again it always holds that $\Delta \in [0,1]$, and note the analogy with (3.50).

3. Recently, a metric version of the KPZ formula has been obtained by Gwynne and Pfeffer [GP19]; more details concerning the relation between scaling exponent and Hausdorff dimension can be found there.

Remark 3.37. There is no consensus (even in the physics literature) about the value of d_γ. Until recently it seemed that the prediction

$$d_\gamma = 1 + \frac{\gamma^2}{4} + \sqrt{(1 + \frac{\gamma^2}{4})^2 + \gamma^2}$$

by Watabiki [Wat93] had a reasonable chance of being correct, but it has now been proved false - at least for small γ [DG19]. Simulations are notoriously difficult because of large fluctuations. As mentioned earlier, the only value that is known rigourously is when $\gamma = \sqrt{8/3}$. In this case the metric space is described by the Brownian map [Mie13, LG13, MS16d] and the Hausdorff dimension is equal to 4.

We are now ready to state the KPZ theorem in this setup.

Theorem 3.38 (Almost sure Hausdorff KPZ formula)**.** *Suppose that A is deterministic and that $\gamma \in (0,2)$. Then, almost surely it holds that*

$$x = \frac{\gamma^2}{4}\Delta^2 + (1 - \frac{\gamma^2}{4})\Delta.$$

We will not prove this result and refer to [RV11] for details. (We will, however, soon see the proof of a closely related result due to Aru [Aru15]). We make a few observations.

1. $x = 0, 1$ if and only if $\Delta = 0, 1$.

2. This is a quadratic relation with positive discriminant so can be inverted.

3. In the particular but important case of uniform random planar map scaling limits (see Chapter 4), $\gamma = \sqrt{8/3}$ and so the relation is

$$x = \frac{2}{3}\Delta^2 + \frac{1}{3}\Delta. \tag{3.51}$$

As we have already mentioned, various forms of the KPZ relation have now been proved; the above statement comes from the work of Rhodes and Vargas [RV11]. Other versions can be found in the works of Aru [Aru15], Duplantier and Sheffield [DS11] which will both be discussed later in this chapter. See also works of Gwynne and Pfeffer [GP19] for a KPZ relation in the sense of metric (Hausdorff) dimensions; Gwynne, Holden and Miller [GHM20] for an effective KPZ formula which can be used rigourously for determining a number of SLE exponents, and Berestycki, Garban, Rhodes and Vargas [BGRV16] for a KPZ relation formulated using the Liouville heat kernel.

3.11 Proof in the case of expected Minkowski dimension

We now state Aru's version of the KPZ formula [Aru15] which, as already mentioned, has a straightforward proof given our earlier work. This statement uses an alternative notion of fractal dimension: Minkowski dimension rather than Hausdorff.

We will only state the case $d = 2$ of this result, even though the arguments generalise easily to arbitrary dimensions. We again use the notation $\mathcal{S}_n$ for the nth level dyadic covering of D by squares $S_i, i \in \mathcal{S}_n$ of side-length 2^{-n}. For $\delta > 0$, the (Euclidean) $(\delta, 2^{-n})$-Minkowski content of A is defined by

$$M_\delta(A; 2^{-n}) = \sum_{i \in \mathcal{S}_n} \mathbf{1}_{\{S_i \cap A \neq \emptyset\}} \operatorname{Leb}(S_i)^\delta,$$

and the (Euclidean) Minkowski dimension (fraction) of A is then

$$d_M(A) = \inf\{\delta : \limsup_{n \to \infty} M_\delta(A, 2^{-n}) < \infty\}.$$

Note that since we used $\operatorname{Leb}(S_i)$ in the definition of the Minkowski content rather than the more standard side-length 2^{-n} of S_i, the above quantity d_M is in $[0, 1]$ and is related to the more standard notion of Minkowski dimension D_M through the identity $d_M = D_M/2$. Finally, we define the **Minkowski scaling exponent**

$$x_M = 1 - d_M.$$

On the quantum side, we define

$$M_\delta^\gamma(A, 2^{-n}) = \sum_{i \in \mathcal{S}_n} \mathbf{1}_{\{S_i \cap A \neq \emptyset\}} \, \mu(S_i)^\delta,$$

and the quantum expected Minkowski dimension by

$$q_M = \inf\{\delta : \limsup_{n \to \infty} \mathbb{E}(M_\delta^\gamma(A, 2^{-n})) < \infty\}.$$

The quantum Minkowski scaling exponent is then set to be

$$\Delta_M = 1 - q_M.$$

The KPZ relation for the Minkowski scaling exponents is then $x_M = (\gamma^2/4)\Delta_M^2 + (1 - \gamma^2/4)\Delta_M$ (formally this is the same as the relation in Theorem 3.40). Equivalently, this can be rephrased as follows.

Proposition 3.39 (Expected Minkowski KPZ, [Aru15])**.** *Suppose $\bar{A}$ lies at a positive distance from ∂D and that A is bounded. Then*

$$d_M = (1 + \gamma^2/4)q_M - \gamma^2 q_M^2/4. \tag{3.52}$$

Proof. First recall Theorem 3.23 from earlier in this chapter, which implies that if $0 \leq q \leq 1$, then

$$\mathbb{E}(\mu(B(r))^q) \asymp r^{(d+\gamma^2/2)q - \gamma^2 q^2/2}$$

for balls $B(r)$ of Euclidean radius r lying strictly within the domain D.

Fix $d \in (0, 1)$ and let q be such that $d = (1 + \gamma^2/4)q - q^2\gamma^2/4$ and note that $q \in (0, 1)$. Therefore,

$$\mathbb{E}(\sum_{i \in \mathcal{S}_n} \mathbf{1}_{\{S_i \cap A \neq \emptyset\}} \mu(S_i)^q) \asymp \sum_{i \in \mathcal{S}_n} \mathbf{1}_{\{S_i \cap A \neq \emptyset\}} \operatorname{Leb}(S_i)^d$$

and consequently the limsup of the left hand side is infinite if and only if the limsup of the right hand side is infinite. In other words, d_M and q_M satisfy (3.52). □

3.12 Duplantier–Sheffield's KPZ theorem

We end this chapter with a short description of Duplantier and Sheffield's definitions of scaling exponents, as well as a sketch of proof of the resulting KPZ formula [DS11]. (The statement is a bit weaker than Theorem 3.38, since the notions of scaling exponents are slightly harder to use, and the formula holds only in expectation as opposed to almost surely).

In this section, the (Euclidean) scaling exponent of $A \subset D$ is the limit, if it exists, defined by

$$x' = \lim_{\varepsilon \to 0} \frac{\log \mathbb{P}(B(z, \varepsilon) \cap A \neq \emptyset)}{\log(\varepsilon^2)},$$

where $\mathbb{P}$ is the joint law of A (if it is random) and a point z chosen proportionally to Lebesgue measure in D. We will assume that D is bounded.

We need to make a few comments about this definition.

1. First, this is equivalent to saying that the volume of A_ε, the Euclidean ε-neighbourhood of A, decays like $\varepsilon^{2x'}$. In other words, A can be covered with $\varepsilon^{-(2-2x')}$ balls of radius ε, and hence typically the Hausdorff dimension of A is simply
$$\dim(A) = 2 - 2x' = 2(1 - x'),$$
consistent with our earlier definition of Euclidean scaling exponent. In particular, note that $x' \in [0, 1]$ always; $x' = 0$ means that A is practically the full space, $x' = 1$ means it is practically empty.

2. In the definition we divide by $\log(\varepsilon^2)$, because ε^2 is the volume (with respect to the Euclidean geometry on $\mathbb{R}^2$) of a ball of radius ε. In the quantum world, we would need to replace this by the Liouville area of a ball of radius ε - see below.

The quantum analogue of this is the following. For $z \in D$, we denote by $B^\delta(z)$ the quantum ball of mass δ: that is, the Euclidean ball centred at z whose radius is chosen so that its Liouville area is precisely δ. (In [DS11], this is called the *isothermal* ball of mass δ at z). The quantum scaling exponent of $A \subset D$ is then the limit, if it exists, defined by
$$\Delta' = \lim_{\delta \to 0} \frac{\log \mathbb{P}(B^\delta(z) \cap A \neq \emptyset)}{\log(\delta)},$$
where z is sampled from the Liouville measure μ normalised to be a probability distribution.

Theorem 3.40 (Expected Hausdorff KPZ formula)**.** *Suppose A is independent of the GFF, $\gamma \in (0, 2)$, and D is bounded. Then if A has Euclidean scaling exponent x', it has quantum scaling exponent Δ, where x and Δ' are related by the formula*
$$x' = \frac{\gamma^2}{4}(\Delta')^2 + (1 - \frac{\gamma^2}{4})\Delta'. \tag{3.53}$$

We will now sketch the argument used by Duplantier and Sheffield to prove Theorem 3.40, since it is interesting in its own right and gives a somewhat different perspective (in particular, it shows that the KPZ formula can be seen as a large deviation probability for Brownian motion).

Informal description of the idea of the proof. We wish to evaluate the probability $\mathbb{P}(B^\delta(z) \cap A \neq \emptyset)$, where z is a point sampled from the Liouville measure, and B^δ is the Euclidean ball of Liouville mass δ around z. Of course the event that this ball intersects A is rather unlikely, since the ball is small. But it can happen for two reasons. The first one is simply that z lands very close (in the Euclidean sense) to A – this has a cost governed by the Euclidean scaling exponent of A, by definition, since we may think of z as being sampled from the Lebesgue measure and then sampling the Gaussian free field given z, as in the description of the rooted measure Section 2.4. However, it is more economical for z to land relatively further away from z, and instead require that the ball of quantum mass δ have a bigger than usual radius. As the quantum mass of the ball of radius r around z is essentially governed by the size of the circle average $h_r(z)$, which behaves like a Brownian motion plus some drift, we find ourselves computing a large deviation probability for a Brownian motion. The KPZ formula is hence nothing else but the large deviation function for Brownian motion.

Sketch of proof of Theorem 3.40. Now we turn the informal idea above into more concrete mathematics, except for two approximations that we will not justify. Suppose z is sampled according to the Liouville measure μ. Then we know from Theorem 2.4 (see also exercise **??** in that chapter) that the joint law of the point z and the free field is absolutely continuous with respect to a point z sampled from Lebesgue measure, together with the field $h^0(\cdot) + \gamma \log|\cdot - z| + O(1)$, where h^0 is a GFF that is independent of z. (See Section 2.4). Hence the mass of the ball of radius ε about z is approximately given by

$$\begin{aligned}\mu(B(z,\varepsilon)) &\approx \varepsilon^{\gamma^2/2} e^{\gamma h_\varepsilon(z)} \times \varepsilon^2 \\ &\asymp e^{\gamma(h^0_\varepsilon(z) + \gamma \log 1/\varepsilon)} \varepsilon^{2+\gamma^2/2} \\ &= \varepsilon^{2-\gamma^2/2} e^{\gamma h^0_\varepsilon(z)}. \end{aligned} \tag{3.54}$$

It takes some time to justify rigorously the approximation in (3.54), but the idea is that the field h_ε fluctuates on a spatial scale of size roughly ε. Hence we are not making a big error by pretending that h_ε is constant on $B(z,\varepsilon)$, equal to $h_\varepsilon(z)$. In a way, making this precise is the most technical part of the paper [DS11]. We will not go through the arguments which do so, and instead we will see how, assuming it, one is naturally led to the KPZ relation.

Working on an exponential scale (which is more natural for circle averages) and writing $B_t = h^0_{e^{-t}}(z)$, we find that

$$\log \mu(B(z, \mathrm{e}^{-t})) \approx \gamma B_t - (2 - \gamma^2/2)t.$$

We are interested in the maximum radius ε such that $\mu(B(z,\varepsilon))$ will be approximately δ: this will give us the Euclidean radius of the quantum ball of mass δ around z. So let

$$\begin{aligned} T_\delta &= \inf\{t \geq 0 : \gamma B_t - (2-\gamma^2/2)t \leq \log \delta\} \\ &= \inf\{t \geq 0 : B_t + (\frac{2}{\gamma} - \frac{\gamma}{2})t \geq \frac{\log(1/\delta)}{\gamma}\}. \end{aligned}$$

where the second equality is in distribution. Note that since $\gamma < 2$ the drift is positive, and hence $T_\delta < \infty$ a.s.

Now, recall that if $\varepsilon > 0$ is fixed, the probability that z will fall within (Euclidean) distance ε of A is approximately $\varepsilon^{2x'}$. Hence, applying this with $\varepsilon = e^{-T_\delta}$ the probability that $B^\delta(z)$ intersects A is, approximately, given by

$$\mathbb{P}(B^\delta(z) \cap A \neq \emptyset) \approx \mathbb{E}(\exp(-2x' T_\delta)).$$

This is the second approximation that we will not seek to justify fully. Consequently, we deduce that

$$\Delta' = \lim_{\delta \to 0} \frac{\log \mathbb{E}(\exp(-2x' T_\delta))}{\log \delta}.$$

For $\beta > 0$, consider the martingale

$$M_t = \exp(\beta B_t - \beta^2 t/2),$$

and apply the optional stopping theorem at the time T_δ (note that this is justified). Then we get, letting $a = 2/\gamma - \gamma/2$, that

$$1 = \exp(\beta \frac{\log(1/\delta)}{\gamma}) \mathbb{E}(\exp(-(a\beta + \beta^2/2)T_\delta)).$$

Finally set $2x' = a\beta + \beta^2/2$, so that $\mathbb{E}(\exp(-2x'T_A)) = \delta^{\beta/\gamma}$. In other words, $\Delta' = \beta/\gamma$, where β is such that $2x' = a\beta + \beta^2/2$. Equivalently, $\beta = \gamma\Delta'$, and

$$2x' = (\frac{2}{\gamma} - \frac{\gamma}{2})\gamma\Delta' + \frac{\gamma^2}{2}(\Delta')^2.$$

This is exactly the KPZ relation. □

3.13 Applications of KPZ to critical exponents

⋆ *This section explains in a non-rigourous manner how the KPZ relation can be used to compute critical exponents in some models of statistical mechanics in two dimensions. This section can be skipped on a first reading, and is only relevant later in connection with the end of Chapter 4. This section also assumes basic familiarity with the notion of random planar maps and the conjectures related to their conformal embeddings, see Section 4.2.*

At the discrete level, the KPZ formula can be interpreted as follows. Consider a random planar map M of size N (where 'size' refers indistinctly to the number of faces, vertices or edges). Suppose that a certain subset A within M has a size $|A| \approx N^{1-\Delta}$, so that A is "fractal-like". We have in mind a set A which is defined conditionally independently given the map, and of course depends on N (but we do not indicate this in the notation). For instance, A could be the set of double points of a random walk on the map run until its cover time, or the set of pivotal edges for percolation on the map with respect to some macroscopic event. We may also consider the Euclidean analogue A' of A within a Euclidean box of area N (and thus of side length $n = \sqrt{N}$). Namely, A' is the set that one obtains when the map M is exactly this subset of the square lattice. In this case we again expect A' to be fractal-like, and so $|A'| \approx N^{1-x} = n^{2-2x}$ for some $x \in [0, 1]$. If A' has a scaling limit then this x is nothing but its Euclidean scaling exponent (indeed, the discrete size of A' is essentially the number of balls of a fixed radius required to cover a scaled version of it). Likewise, if A has a scaling limit then Δ is nothing but its quantum scaling exponent.

Hence the KPZ relation suggests that x, Δ should be related by

$$x = \frac{\gamma^2}{4}\Delta^2 + (1 - \frac{\gamma^2}{4})\Delta$$

in the limit as $N \to \infty$. Here γ refers to the universality class of the map; this assumes that A is (when embedded suitably in the plane) independent of the field h which represents the embedding of the map in the limit.

In particular, observe that the approximate (Euclidean) Hausdorff dimension of A' is then $2 - 2x$, consistent with our definitions. See Chapter 4 for concrete examples, where this is used, for instance, to guess the loop-erased random walk exponent.

3.14 Exercises

1. By considering the set of thick points or otherwise, show that the KPZ relation does not need to hold if the set A is allowed to depend on the free field. This type of example was first considered by [Aru15] who also considered the case of flow lines associated to the GFF.

2. Let $A \subset D$, and let $q \in [0,1]$. Show that $\mathbb{E}(\mu(A)^q)$ is a non-decreasing function of $\gamma \in [0, \sqrt{2\mathbf{d}})$.

3. Let $A \subset D$. For $\gamma < \sqrt{\mathbf{d}}$, show that $\mu(A)$ is admits a continuous modification in γ. (Hint: use the Kolmogorov continuity criterion.)

4. Use the scaling invariance properties developed in the proof of the multifractal spectrum to show that μ has a.s. no atoms. Observe that this also follows from the energy estimate in Exercise 4 of Chapter 2.

5. This exercise gives a flavour of Kahane's original pioneering argument for the construction of GMC in [Kah85]. Suppose that K is a covariance kernel of the form (3.1), that can be written in the form
$$K(x,y) = \sum_{n=1}^{\infty} K_n(x,y)$$
for all $x \neq y$ in $D \subset \mathbb{R}^d$, where for each n, $K_n : D \times D \to \mathbb{R}$ is positive-definite and satisfies $K_n(x,y) \geq 0$ for all $x, y \in D$. Such a covariance kernel was called by Kahane **σ-positive**. Show that there exists a sequence of centred Gaussian fields $(h^n)_{n\geq 1}$ such that the fields $(h^n - h^{n-1})_{n\geq 1}$ are independent centred Gaussian fields with covariances K_n for each n. Let σ be a reference Radon measure satisfying (3.2) for some $\mathbf{d} > 0$. For $0 \leq \gamma < \sqrt{2\mathbf{d}}$, we use this decomposition to construct a natural sequence of 'chaos approximations' μ_n by setting
$$\mu_n(A) = \int_A \exp\{\gamma h_n(x) - \tfrac{\gamma^2}{2}\mathbb{E}(h_n(x)^2)\}\sigma(dx),$$
for any Borel set A. Prove that $\mu_n(A)$ has an almost sure limit $\mu(A)$ as $n \to \infty$ which defines a random measure.

   Suppose we are given two σ-positive decompositions for K, say
$$K(x,y) = \sum_{n=1}^{\infty} K_n(x,y) = \sum_{n=1}^{\infty} K'_n(x,y),$$
and let μ and μ' be the associated chaos measures constructed above. Using Kahane's inequality (and without using Theorem 3.2), show that for any Borel set A, $\mathbb{E}(\mu(A)^q) \leq \mathbb{E}(\mu'(A)^q)$ for $q \in (0,1)$ (note that this argument does not require knowing that either μ or μ' is nonzero). Deduce that the laws of μ and μ' (as random measures) are identical. This is Kahane's theorem on uniqueness of GMC; Kahane's inequality [Kah86] was discovered for the purpose of this proof.

6. We now take the same setup as above, and assume the result of Theorem 3.2. Show that in the case $\gamma < \sqrt{2\mathbf{d}}$, the limit μ constructed above agrees with the GMC measure of Theorem 3.2.

7. If K is as in (3.1), define the linear operator T on $L^2(D)$ by setting

$$Tf(x) = \int_D K(x,y)f(y)\,dy$$

for each $f \in L^2(D)$. When D is bounded, one can show using standard operator theory that there exists an orthonormal basis $\{f_k\}_{k\geq 0}$ of $L^2(D)$, made up of eigenfunctions for T. The ordering can be chosen so that the associated eigenvalues $\{\lambda_k^{-1}\}_{k\geq 0}$ satisfy $0 < \lambda_1 < \lambda_2 \leq \lambda_3, \ldots$ Moreover, with this ordering, the first eigenfunction f_1 will be strictly positive in D.

   (a) Show that for each x in D,

$$\sum_{k=0}^{n} \lambda_k^{-1} f_k(x) f_k(\cdot) \to K(x,\cdot) \text{ in } L^2(D)$$

   as $n \to \infty$. Let h be the centred Gaussian field with covariance K. By considering the joint law of $\{\lambda_k^{1/2}(h, f_k)\}_{k\geq 1}$, show that for any test function f on D, if $h^n := \sum_{k=0}^{n}(h, f_k)f_k$, then (h^n, f) converges almost surely and in $L^2(\mathbb{P})$ to (h, f) as $n \to \infty$.

   Remark. *This decomposition of h is known as the* ***Karhuhen–Loeve expansion****.*

   (b) Show further that for $\gamma \geq 0$, the sequence of measures defined by

$$\mu^n(S) := \int_S \exp(\gamma h^n(z) - \gamma^2/2 \operatorname{Var}(h^n(z))\,dz \qquad S \subset D, n \geq 0$$

   has an almost sure limit with respect to the weak topology. When $\gamma < \sqrt{2\mathbf{d}}$, show that $\mu^n(S)$ is a uniformly integrable family for any fixed S. Use this to show that $\lim_n \mu^n(S)$ agrees almost surely with $\mu(S)$, where μ is the GMC measure for h constructed in Theorem 3.2.

   (c) Use the fact that f_1 is positive together with (3.49) to show that for $\delta > 0$, $\mathbb{P}(\mu^\gamma(D) \leq \delta) \geq c\mathbb{P}(Z \leq \delta)$ where Z is an appropriately chosen lognormal random variable and $c > 0$ does not depend on δ.

4 Statistical physics on random planar maps

4.1 Fortuin–Kasteleyn weighted random planar maps

In this chapter we change our focus from the continuous to the discrete, and describe the model of random planar maps weighted by critical Fortuin–Kasteleyn percolation. These maps can be thought of as canonical discretisations of Liouville quantum gravity.

We proceed as follows. We first recall the notion of planar map and **decorated planar map** before defining a probability measure on such maps (maps decorated by self-dual FK loops). In Section 4.2, we discuss aspects of the conjectured connection between this model of planar maps and **Liouville quantum gravity**. In Section 4.3 we focus on the case where the decoration is a spanning tree. Here we describe in detail a powerful **bijection** due (independently) to Mullin, Bernardi and Sheffield, between tree-decorated maps and pairs of independent, positive random walk excursions (equivalently, two-dimensional random walk excursions in the positive quadrant). This bijection is a convenient way to approach the question of scaling limits. We use it in Section 4.4 to compute the (quantum) scaling exponent of the **loop-erased random walk** (LERW). Using the KPZ relation of Section 3.10, we find that it agrees with various known properties of LERW on the square lattice, including the Hausdorff dimension of its scaling limit SLE_2. In Section 4.5, we discuss Sheffield's bijection, which is a generalisation of the aforementioned bijection to decorations which are no longer spanning trees but **densely packed loop configurations**. Again, this bijection is from decorated maps to pairs of excursions. In this case, however, the excursions are far from independent; this has an interpretation in terms of a **discrete mating of treesting of trees** which will be described in the continuum later. (To be added) This description is used in Section 4.6 to show the existence of an infinite volume local limit. A scaling limit result is discussed which, roughly speaking, shows that the limiting trees are correlated infinite CRTs.

Planar map, dual map. Recall that a **planar map** is a proper embedding of a (multi) graph with a finite number of edges in the plane $\mathbb{C} \cup \{\infty\}$ (viewed as the Riemann sphere), which is viewed up to orientation preserving homeomorphisms from the sphere to itself. Let $\boldsymbol{m}_n$ be a map with n edges and $\boldsymbol{t}_n$ be a subgraph spanning *all* of its vertices. We call the pair $(\boldsymbol{m}_n, \boldsymbol{t}_n)$ a **decorated map**. Let $\boldsymbol{m}_n^\dagger$ denote the *dual map* of $\boldsymbol{m}_n$. Recall that the vertices of the dual map correspond to the faces of $\boldsymbol{m}_n$ and two vertices in the dual map are adjacent if and only if their corresponding faces are adjacent to a common edge in the primal map. Every edge e in the primal map corresponds to an edge $e^\dagger$ in the dual map which joins the vertices corresponding to the two faces adjacent to e. The dual map $\boldsymbol{t}_n^\dagger$ is the graph formed by the subset of edges $\{e^\dagger : e \notin \boldsymbol{t}_n\}$ and all dual vertices. We fix an edge in the map $\boldsymbol{m}_n$, to which we also assign an orientation, and define it to be the root edge. With an abuse of notation, we will still write $\boldsymbol{m}_n$ for the rooted map; and we let $\mathcal{M}_n$ be the set of maps with n edges together with one distinguished edge called the root.

Canonical triangulation. Given a subgraph-decorated map $(\boldsymbol{m}_n, \boldsymbol{t}_n)$ with $\boldsymbol{m}_n \in \mathcal{M}_n$ and $\boldsymbol{t}_n$ spanning every vertex (in the sense that the vertex set of $\boldsymbol{t}_n$ is all of the vertex set of $\boldsymbol{m}_n$), one can associate to it a set of loops where in some sense each loop forms the interface between two clusters (connected components) of t_n and its planar dual. We will now define this carefully. To do so, we will need to discuss not only the dual planar map but also a couple of related maps that can be constructed from superposing the primal and dual maps.

We first consider an auxiliary map which we call the **Tutte map**, and which is formed by joining the dual vertices in every face of $\boldsymbol{m}_n$ with the primal vertices incident to that face. We call these edges **Tutte edges** (drawn in green in Figure 3). Thus the vertex set of the Tutte map consists of all primal and dual vertices, but note that its edge set does not contain any of the original edges of $\boldsymbol{m}_n$ or its dual. It is easy to check that this Tutte map is a quadrangulation, meaning each face has exactly four (refinement) edges surrounding it. Each of the original edges of $\boldsymbol{m}_n$ or $\boldsymbol{m}_n^\dagger$ is a diagonal of one of these quadrangles. In other words, every edge in $\boldsymbol{m}_n$ corresponds to a quadrangle in the Tutte map; this quadrangle can be viewed as the union of two triangles, one on either side of the edge.

In fact this construction defines a bijection between maps with n edges and quadrangulations with n faces, sometimes called the **Tutte bijection**.

Given a subgraph-decorated map $(\boldsymbol{m}_n, \boldsymbol{t}_n)$ define the **refinement** map $\bar{\boldsymbol{m}}_n$ to be formed by the union of $\boldsymbol{t}_n, \boldsymbol{t}_n^\dagger$ and the refinement edges; note that its vertex set is the same as the Tutte map, i.e. every primal and dual vertex of $\boldsymbol{m}_n$. The addition of $\boldsymbol{t}_n$ and $\boldsymbol{t}_n^\dagger$ makes the refinement map a triangulation: indeed, every quadrangle from the Tutte map has been split into two (either with a diagonal from $\boldsymbol{t}_n$ or from $\boldsymbol{t}_n^\dagger$). The root edge of $\boldsymbol{m}_n$ induces a **root triangle** on the refinement map, which is taken to be the triangle immediately to the right of the root edge of $\boldsymbol{m}_n$.

Note that every triangle consists of two Tutte edges and one edge from either $\boldsymbol{t}_n$ (primal edge) or $\boldsymbol{t}_n^\dagger$ (dual edge). For future reference, we call such a triangle in $\bar{\boldsymbol{m}}_n$ a **primal triangle** or **dual triangle** respectively (see Figure 6).

Loops. Finally, given $(\boldsymbol{m}_n, \boldsymbol{t}_n)$ we can define the loops induced by $\boldsymbol{t}_n$ as follows. For each connected component C of either $\boldsymbol{t}_n$ or $\boldsymbol{t}_n^\dagger$, we draw a loop surrounding it (meaning a closed curve in the complement of C in the sphere; the complement contains two components, and by convention we draw it in the "exterior" one that contains the point on the sphere designated to be ∞). If this loop is drawn sufficiently close to C it identifies a unique collection of triangles that are adjacent to C (in the sense that they share at least a vertex with it). We view the component C itself as an open cluster for a percolation configuration either on $\boldsymbol{m}_n$ or its dual, and will use the word "cluster" interchangeably from now on.

In what follows, one should visualise the loop of C as being a closed curve drawn sufficiently close to C in its complement, as above. However for precision, we will actually identify the loop with the collection of triangles through which it passes. See Figure 3 for an illustration. In this way, each loop is simply a collection of triangles "separating" a primal connected component of $\boldsymbol{t}_n$ from a dual connected component in $\boldsymbol{t}_n^\dagger$, or vice-versa. Note that the set of loops is "space filling" in the sense that every triangle of the refined map is

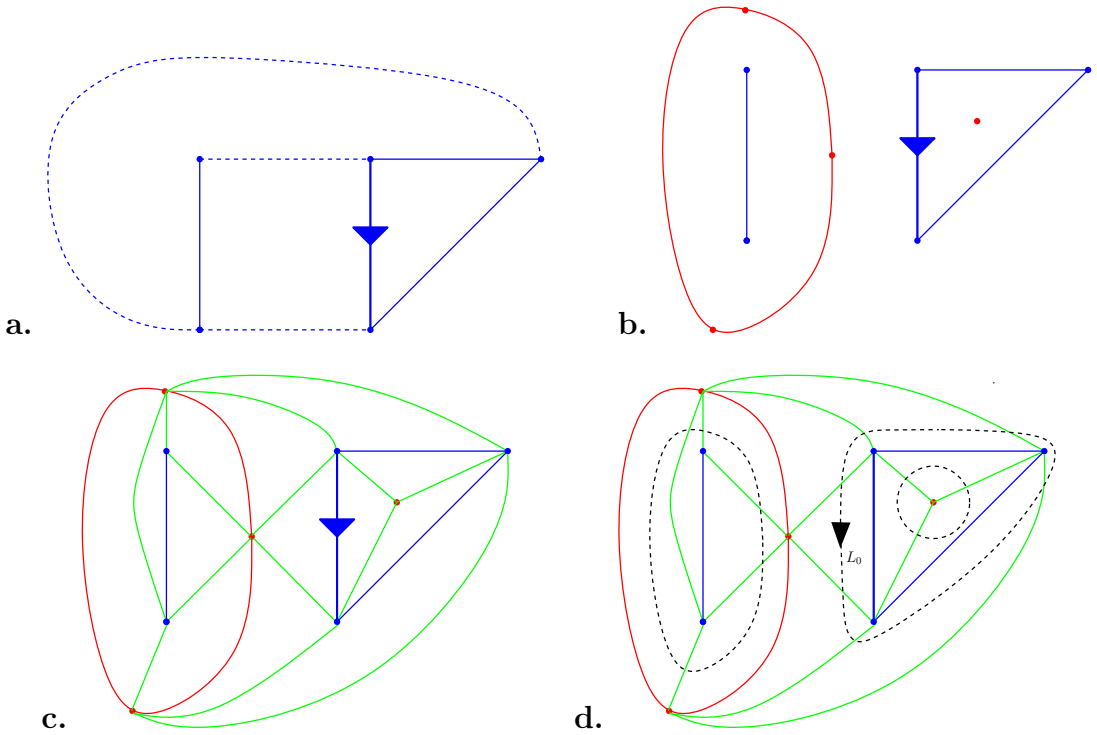


Figure 3: A map m decorated with loops associated to a set of open edges t. **a.** The map is in blue, with solid open edges and dashed closed edges. **b.** Open clusters and corresponding open dual clusters are shown in blue and red. **c.** Every dual vertex is joined to its adjacent primal vertices by a green edge. This results in an refined map $\bar{m}$ which is a triangulation. **d.** The primal and dual open clusters are separated by loops, which are drawn in black and are dashed. Each loop is identified with the set of triangles through which it passes: note that it crosses each triangle in the set exactly once. The oriented root edge of the map is indicated with a blue arrow in subfigures **a**, **b** and **c**. The loop L_0 is marked with an arrow in subfigure **d**, and the arrow indicates the orientation of the loop, parallel to the orientation of the root edge.

contained in a loop. We denote by L_0 the loop that is associated with the root triangle. It comes with a natural orientation, which is parallel to the orientation of the root edge of M_n.

Also, given the Tutte map and the collection of closed curves described above, one can recover the spanning subgraph $\boldsymbol{t}_n$ (hence also $\boldsymbol{t}_n^\dagger$) that generates it. Let $\ell(\boldsymbol{m}_n, \boldsymbol{t}_n)$ denote the number of loops corresponding to a configuration $(\boldsymbol{m}_n, \boldsymbol{t}_n)$. Note that this is equal to the number of clusters in $\boldsymbol{t}_n$ plus the number of clusters in $\boldsymbol{t}_n^\dagger$ minus one; indeed, each new cluster generates a new loop.

Fortuin–Kasteleyn model. The particular distribution on planar maps that we will now consider was introduced in [She16b]. Let $q \geq 0$ and let $n \geq 1$: we will define a random map $M_n \in \mathcal{M}_n$ decorated with a (random) subset T_n of edges that spans all of its vertices. As in the deterministic setting, this induces a dual collection of edges $T_n^\dagger$ on the dual map of M (see Figure 3). The law of (M_n, T_n) is defined by declaring that for any fixed planar map $\boldsymbol{m}$ with n edges, and $\boldsymbol{t}$ a given subset of edges of $\boldsymbol{m}$,

$$\mathbb{P}(M_n = \boldsymbol{m}, T_n = \boldsymbol{t}) \propto \sqrt{q}^{\ell}, \quad \ell = \ell(\boldsymbol{m}, \boldsymbol{t}). \tag{4.1}$$

Recall from above that ℓ is the (total) number of loops separating primal and dual clusters in $(\boldsymbol{m}, \boldsymbol{t})$.

Equivalently, the map M_n is chosen with probability proportional to the "partition function" of the self-dual Fortuin–Kasteleyn model on it, and given the map M_n, the collection of edges T_n is then sampled from this Fortuin–Kasteleyn model. This is in turn closely related to the critical q-state Potts model, see [Bax00]. Note that M_n is actually a rooted map (as all of our maps are) and with this definition, the root edge of the map and its orientation are chosen uniformly at random (given the unrooted version).

Uniform random planar maps. Observe that when $q = 1$, the FK model (4.1) has the property that the map M_n is chosen **uniformly** at random among the set $\mathcal{M}_n$ all of (rooted) maps with n edges, because the total number of possible configurations for $\boldsymbol{t}_n$ is 2^n *independently* of $\boldsymbol{m}_n$. Furthermore, given $M_n = \boldsymbol{m}_n$, T_n is chosen uniformly at random from the 2^n possibilities: this corresponds to each edge being present (open) with probability $1/2$, independently of one another. In other words, T_n corresponds to bond percolation with parameter $1/2$ given the map M_n. This is in fact the critical parameter for this percolation model, as shown in the work of Angel [Ang03].

The case of a uniformly chosen planar map in $\mathcal{M}_n$ is one in which remarkably detailed information is known about its structure. In particular, a landmark result due to Miermont [Mie13] and Le Gall [LG13] shows that, viewed as a metric space and rescaling edge lengths to be $n^{-1/4}$, the random map converges to a multiple of a certain universal random metric space known as the **Brownian map**. (In fact, the results of Miermont and Le Gall apply respectively to uniform quadrangulations with n faces and to p-angulations for $p = 3$ or p even, whereas the convergence result concerning uniform planar maps in $\mathcal{M}_n$ was established a bit later by Bettinelli, Jacob and Miermont [BJM14]). Critical percolation on a related

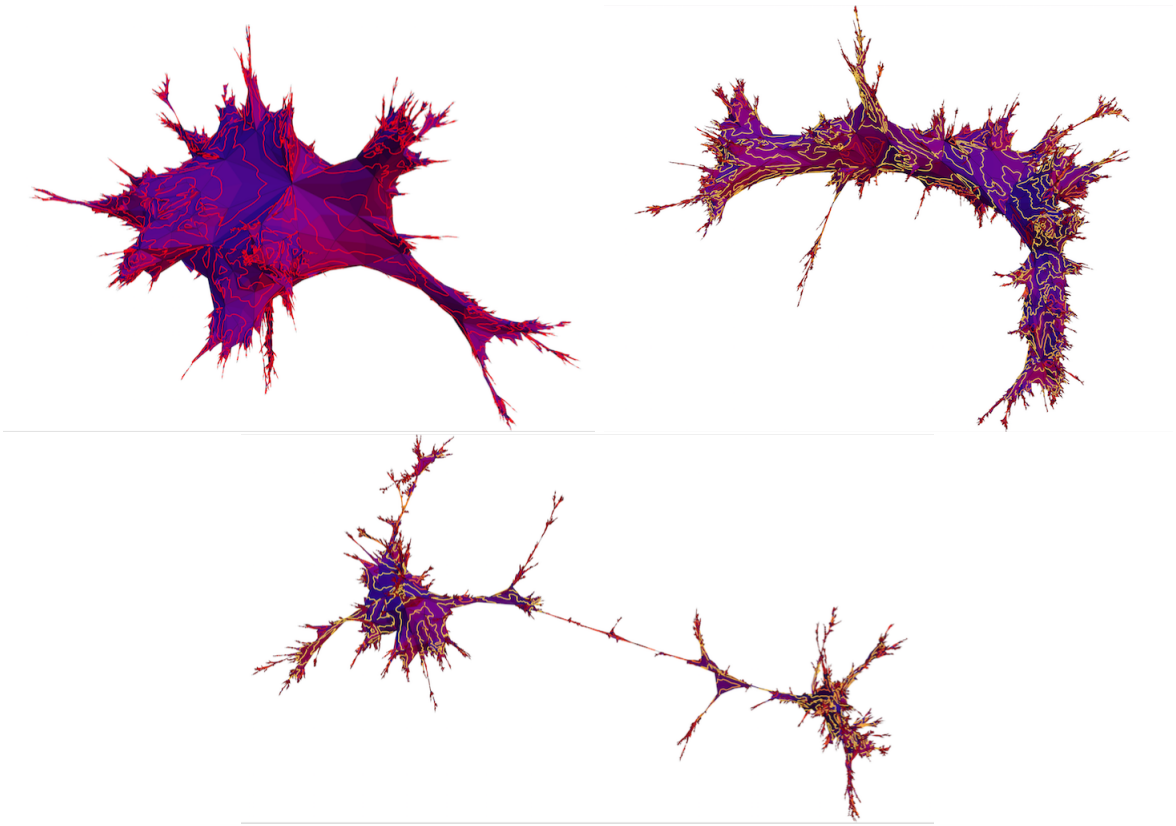


Figure 4: A map weighted by the FK model with $q = 0.5$, $q = 2$ (corresponding to the Ising model) and $q = 9$ respectively, together with some of their loops. Simulation by J. Bettinelli and B. Laslier. When $q > 4$ it is believed that the maps become tree-like, and the limiting metric space should be Aldous' continuum random tree.

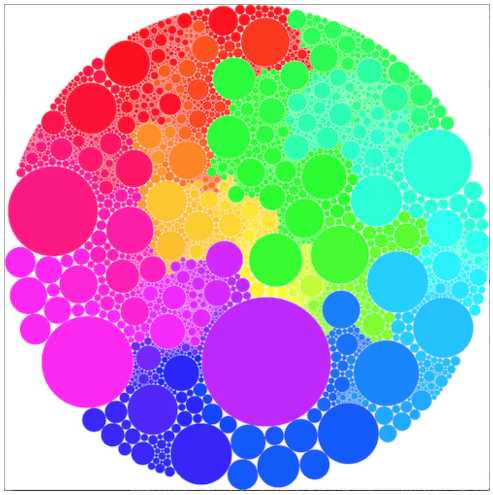


Figure 5: Circle packing of a uniform random planar map. Simulation by Jason Miller.

half-plane version of the maps has been analysed in a work of Angel and Curien [AC15], while information on the full plane percolation model was later obtained by Curien and Kortchemski [CK15]. Related works on loop models (sometimes rigorous, sometimes not) appear in [GJSZJ12, BBG12b, EK95, BBM11, BBG12a].

One reason for the particular choice of the FK model in (4.1) is the belief that for $q < 4$, after Riemann uniformisation, a large sample of such a map closely approximates a *Liouville quantum gravity* surface. We will try to summarise this conjecture in the next subsection.

4.2 Conjectured connection with Liouville quantum gravity

The distribution (4.1) gives us a natural family of distributions on planar maps (indexed by the parameter $q \geq 0$). As already mentioned, in this model, the weight of a particular map $\boldsymbol{m} \in \mathcal{M}_n$ is proportional to the *partition function* $Z(\boldsymbol{m}, q)$ of the critical FK model on the map.

Conformal Embedding. Now, it is strongly believed that in the limit $n \to \infty$, the geometry of such maps are related to Liouville quantum gravity with parameter γ, where

$$q = 2 + 2\cos\left(\frac{\gamma^2\pi}{2}\right). \tag{4.2}$$

To be more precise about this, one must relate the world of planar maps to the world of Liouville quantum gravity by specifying a "natural" embedding of the maps into the plane. There are various ways to do this, and a couple of the simplest are as follows.

- **Via the circle packing theorem.** By a theorem of Koebe–Andreev–Thurston (see the book by K. Stephenson [Ste05] for a comprehensive introduction), any planar map

can be represented as a circle packing. A circle packing is a collection of circles in the plane such that any two of the corresponding discs either are tangent to one another, or do not overlap. In the circle packing representation, the vertices of the map are given by the centres of the circles, and the edges correspond to tangent circles. See Figure 5 above. Each circle packing representation of a map gives an embedding in the plane, and when the map is a simple triangulation, this embedding is unique up to Möbius transformations.

- **Via the uniformisation theorem.** In this approach, a given map is viewed as a Riemann surface by declaring that each face of degree p is a regular p-gon of unit area, endowed with the standard metric, and specifying the charts near a vertex in the natural manner. This Riemann surface can then be embedded into the disc (say) by the uniformisation theorem (which is a generalisation of the Riemann mapping theorem from subsets of $\mathbb{C}$ to arbitary Riemann surfaces).

These embeddings are essentially unique up to Möbius transforms (in the first case, we can circle pack the refinement map $\bar{\boldsymbol{m}}_n$ instead of $\boldsymbol{m}_n$). The choice of Möbius transform can be fixed by requiring, for instance, that the root edge is mapped to $(0, 1)$.

Once an embedding has been chosen, a natural object to study is the measure μ_n in the plane which puts mass $1/N$ (N being the number of vertices in M_n) at the position of each embedded vertex. The conjecture alluded to above says that in the limit as $n \to \infty$, if M_n is sampled from (4.1), then this measure μ_n should converge to γ-Liouville quantum gravity. More precisely, if γ and q are related by (4.2), it should converge in distribution for the topology of weak convergence, to a *variant* of the Liouville measure μ_γ.

Remark 4.1. Note that when $q = 1$, which we have already discussed is the case of uniformly chosen random planar maps, we have $\cos(\gamma^2\pi/2) = -1/2$, i.e. $\gamma = \sqrt{8/3}$. Consequently, the limit of a (conformally embedded) uniformly chosen map should be related to Liouville quantum gravity with this parameter. This has been verified for a slightly different type of conformal embedding called the **Cardy embedding** in a recent breakthrough of Holden and Sun [HS19].

Loops and CLE The loops induced by the FK model (4.1) may be viewed as a decoration on the map. Indeed as we have already mentioned, given the map, they are the cluster boundaries of a self-dual FK percolation model on it with parameter q. It is therefore natural to wonder about their geometry in the scaling limit, after embeddings of the type discussed above. The widely shared belief is that they converge to so-called **conformal loop ensembles** $\mathrm{CLE}_{\kappa'}$ where the parameter κ' is given by

$$\kappa' = \frac{16}{\gamma^2}; \text{ and thus } q = 2 + 2\cos\left(\frac{8\pi}{\kappa'}\right). \tag{4.3}$$

In fact, one can also study the self-dual FK percolation model and its associated loops on a Euclidean lattice, and the same belief is held. That is, these loops are also conjectured

to converge to $\mathrm{CLE}_{\kappa'}$ in the scaling limit, where the relationship between q and κ' is the same as in (4.3). The fact that these two conjectures are the same should heuristically be considered as a consequence of conformal invariance. That is, if the scaling limit of FK loops is conformally invariant, it should be independent of the underlying metric: only their conformal type should matter.

For instance, we have already noticed that when $q = 1$, the associated FK model is just bond percolation. In this case we already know (at least in the case of the triangular lattice) that the scaling limit of the associated loops is given by CLE with parameter $\kappa' = 6$ ([Smi01], [CN08]). This is consistent with the value $\gamma = \sqrt{8/3}$ being the Liouville quantum gravity parameter for the scaling limit of uniform planar maps, as described in Remark 4.1.

Likewise, for $q = 2$ the associated FK model is the FK representation of the critical Ising model. It was proven in [KS16] (see also [CDCH+14] for interfaces and [BH19] for Ising loops) that the scaling limit of these loops is given by $\mathrm{CLE}_{16/3}$. The associated parameter γ is thus $\gamma = \sqrt{3}$.

A small summary of these values is provided in the table below.

FK Model (4.1)	q	γ	κ'
General $q \in [0, 4)$	$2 + 2\cos(\gamma^2\pi/2)$	$\gamma \in [\sqrt{2}, 2)$	$16/\gamma^2 \in (4, 8]$
Uniform map + critical bond percolation	1	$\sqrt{8/3}$	6
Spanning tree decorated map	0	$\sqrt{2}$	8
Critical Ising decorated map	2	$\sqrt{3}$	16/3

4.3 Mullin–Bernardi–Sheffield's bijection in the case of spanning trees

We will now discuss the case where the map $M_n \in \mathcal{M}_n$ is chosen with probability proportional to the number of spanning trees it admits. In other words, for any (rooted) map $\boldsymbol{m}_n \in \mathcal{M}_n$ with n edges and $\boldsymbol{t}_n$ a set of edges on it

$$\mathbb{P}(M_n = \boldsymbol{m}_n, T_n = \boldsymbol{t}_n) \propto \mathbf{1}_{\{\boldsymbol{t}_n \text{ is a spanning tree on } \boldsymbol{m}_n\}}. \tag{4.4}$$

This can be understood as the limit when $q \to 0^+$ of the Fortuin–Kasteleyn model discussed above in (4.1), since in this limit the model concentrates on configurations where $\ell = 0$, equivalently, $\boldsymbol{t}_n$ is a tree. In fact it is immediate in this case that given $M_n = \boldsymbol{m}_n$, $\boldsymbol{t}_n$ is a uniform spanning tree (UST) on $\boldsymbol{m}_n$. We will discuss a powerful bijection due to Mullin [Mul67] and Bernardi [Ber07, Ber08] which is key to the study of such planar maps. This bijection is actually a particular case of a bijection due to Sheffield, which is sometimes called the "hamburger–cheeseburger" bijection. Sheffield's bijection can be used for arbitrary $q \geq 0$, however the case $q = 0$ of trees is considerably simpler and so we discuss it first. (We will use the language of Sheffield, in order to prepare for the more general case later.) Although the hamburger-cheeseburger bijection is the only example we will treat in detail here, we mention that there are other powerful bijections of a similar flavour that can be used to

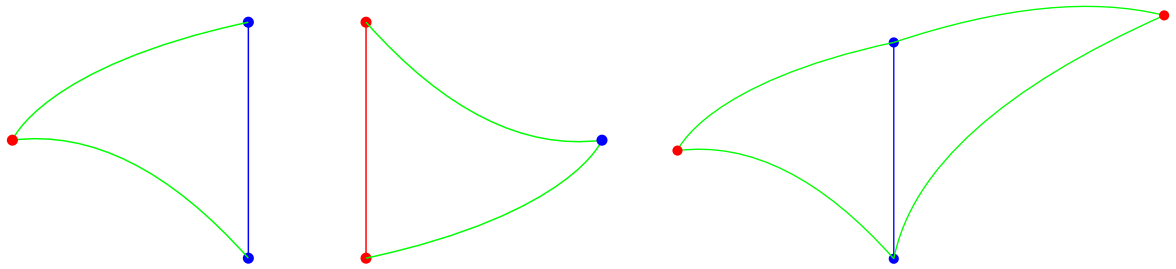


Figure 6: Refined or green edges split the map and its dual into primal and dual triangles. Each primal triangle sits opposite another primal triangle, resulting in a primal quadrangle as above.

connect random planar map models to Liouville quantum gravity and SLE: see for example [BHS17, LSW17, GKMW18, KMSW19].

To describe the $q = 0$ hamburger-cheeseburger bijection, we first fix a deterministic pair $(\boldsymbol{m}_n, \boldsymbol{t}_n)$ as above (with an oriented root edge chosen for $\boldsymbol{m}_n$ and $\boldsymbol{t}_n$ a spanning tree on $\boldsymbol{m}_n$) and describe how to associate with it a certain sequence of letters corresponding to "hamburgers" and "cheeseburgers". Recall that adding refinement edges to a map splits it into triangles of exactly two types: primal triangles (meaning two refined edges and one primal edge) or dual triangles (meaning two refined edges and one dual edge). For ease of reference, primal triangles will be associated with hamburgers, and dual triangles with cheeseburgers. Note that for a primal edge in a primal triangle, the triangle opposite that edge is obviously a primal triangle too. Hence it is better to think of the map as being split into quadrangles with either a primal or dual diagonal (see Figure 6).

We will reveal the map, triangle by triangle, by exploring it along a space-filling (in the sense that it visits every triangle once) path. When we do this, we will keep track of the first time that the path enters a given quadrangle by saying that either a hamburger or a cheeseburger is produced, depending on whether the quadrangle is primal or dual. Later on, when the path comes back to the quadrangle for the second and final time, we will say that the burger has been eaten. We will use the letters h, c to indicate that a hamburger or cheeseburger has been produced and we will use the letters H, C to indicate that a burger has been eaten (or *ordered* and eaten immediately). So in this description we will have one letter for every triangle.

It remains to specify in what order are the triangles visited; equivalently, to describe the space-filling path. In the case that we consider now, where the decoration $\boldsymbol{t}_n$ consists of a single spanning tree, the path is simply the contour path going around the tree (starting from the root), i.e., the unique loop L_0 separating the primal and dual spanning trees, with its orientation inherited from that of the root edge of $\boldsymbol{m}_n$. Hence in this case, we can associate to $(\boldsymbol{m}_n, \boldsymbol{t}_n)$ a sequence w (or **word**) made up of N letters in the alphabet $\Theta = \{\mathsf{h}, \mathsf{c}, \mathsf{H}, \mathsf{C}\}$. We will see below that subject to certain natural conditions on the word w, this map is actually a bijection.

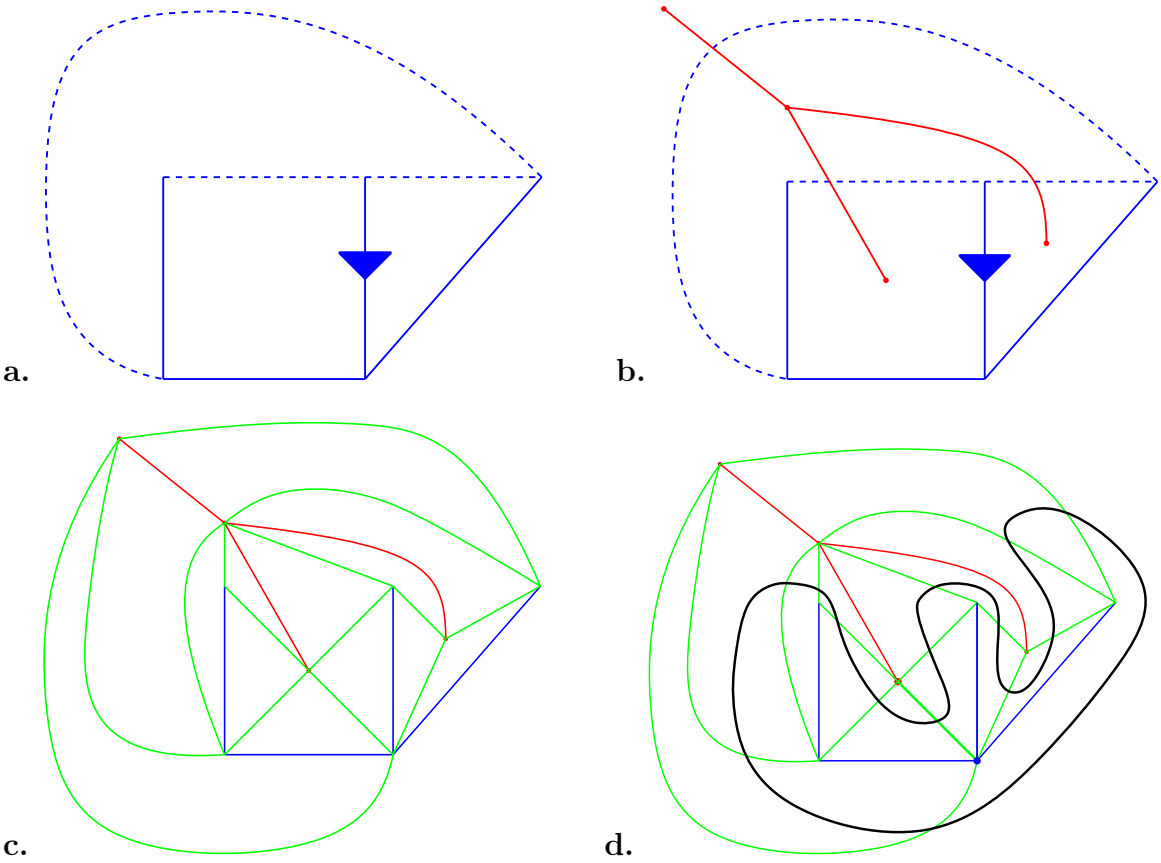


Figure 7: **a:** a map with a spanning tree. **b:** Spanning tree and dual tree. **c:** Refinement edges. **d:** Loop separating the primal and dual spanning trees, to which a root (refined) edge has been added in bold.

Observe that we always have $N = 2n$. To see why, recall that there is one letter for every triangle, so N is the total number of triangles. Moreover, each triangle can be identified with an edge (or in fact half an edge, because each edge is visited once when the burger is produced and once when it is eaten), and so

$$N = 2(E(\boldsymbol{t}_n) + E(\boldsymbol{t}_n^\dagger)) = 2(V(\boldsymbol{t}_n) - 1 + V(\boldsymbol{t}_n^\dagger) - 1).$$

Now $V(\boldsymbol{t}_n) = V(\boldsymbol{m}_n)$, since $\boldsymbol{t}_n$ is spanning, and $V(\boldsymbol{t}_n^\dagger) = V(\boldsymbol{m}_n^\dagger) = F(\boldsymbol{m}_n)$. This gives that

$$N = 2(V(\boldsymbol{m}_n) + F(\boldsymbol{m}_n) - 2), \tag{4.5}$$

and applying Euler's formula together with the fact $F(\boldsymbol{m}_n) = n$, we find that $N = 2n$.

To summarise, given $(\boldsymbol{m}_n, \boldsymbol{t}_n)$ a rooted, spanning tree decorated map with n edges, we can uniquely define a word w of length $2n$ in the letters $\{\mathsf{h}, \mathsf{c}, \mathsf{H}, \mathsf{C}\}$. Observe further that under the reduction rules

$$\overline{\mathsf{cC}} = \overline{\mathsf{hH}} = \emptyset, \ \overline{\mathsf{cH}} = \overline{\mathsf{Hc}} \text{ and } \overline{\mathsf{hC}} = \overline{\mathsf{Ch}},$$

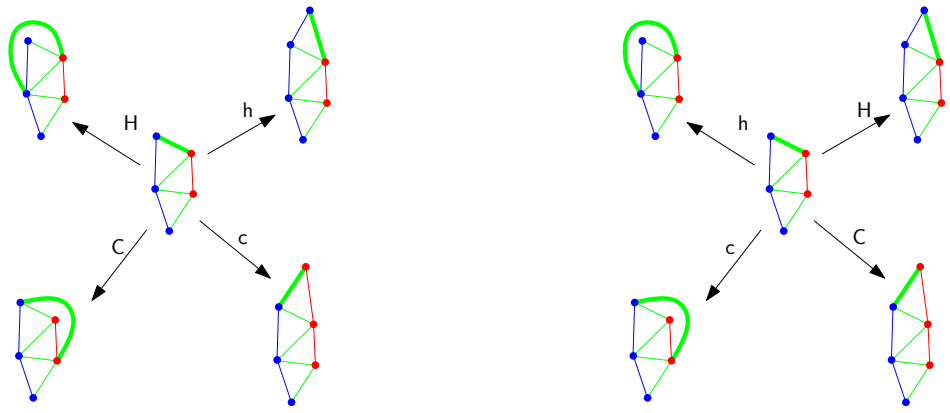


Figure 8: From symbols to map. The current position of the interface (or last discovered refined edge) is indicated with a bold line. Left: reading the word sequence from left to right or *into the future*. The map in the center is formed from the symbol sequence chc. Right: The corresponding operation when we go from right to left (or into the *past*). The map in the center now corresponds to the symbol sequence CHC.

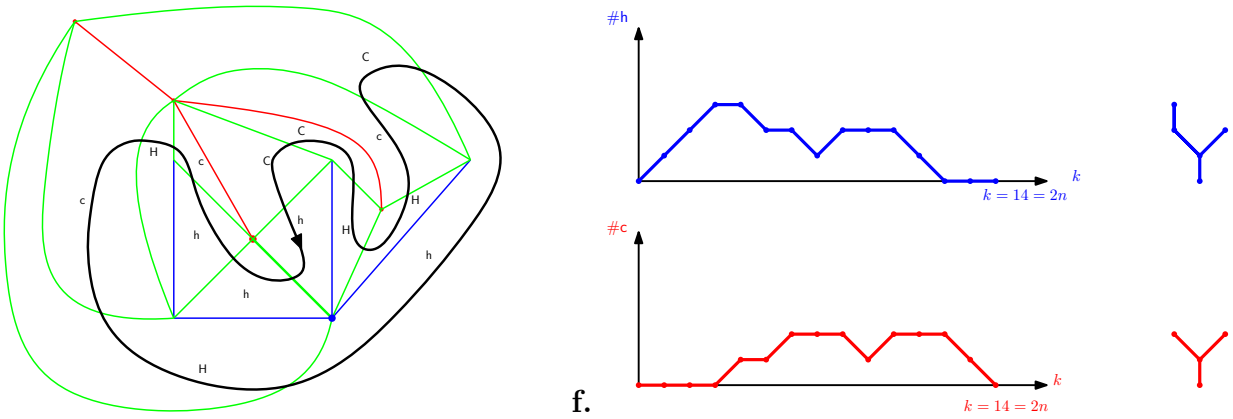
we have $\overline{w} = \emptyset$ (here $\overline{w}$ denotes the reduction of the word w). This corresponds to the fact that every burger produced is eaten, and every food order corresponds to a burger that was produced before. Subject to the condition $\bar{w} = \emptyset$, it is easy to see that the map $(\boldsymbol{m}_n, \boldsymbol{t}_n) \mapsto w$ is a bijection. See, e.g., Figure 8 for a proof by picture.

Now we go a step further, and associate to this word w a pair $(X_k, Y_k)_{1 \le k \le 2n}$, which count the number of hamburgers and cheeseburgers respectively in the stack at any given time $1 \le k \le 2n$ (i.e., the number of hamburgers or cheeseburgers which have been produced prior to time k but get eaten after time k). Note that (X, Y) is a process which starts from the origin at time $k = 0$, and ends at the origin at time $k = 2n$. Moreover, by construction X and Y both stay non-negative throughout. We call a process $(X_k, Y_k)_{0 \le k \le 2n}$ satisfying these properties a **discrete excursion** (in the quarter plane). So at this point, we have associated with any $(\boldsymbol{m}_n, \boldsymbol{t}_n)$ as above, a unique discrete excursion (X, Y) of length $2n$.

Conversely, given such a process (X, Y) we can associate to it a word w in the letters of Θ such that (X, Y) is the net burger count of w. Obviously w reduces to $\emptyset$ and so, as we have seen above, this word w specifies a unique pair $(\boldsymbol{m}_n, \boldsymbol{t}_n)$.

Another property which is easy to check (and easily seen on the picture) is that the excursions X and Y encode the spanning tree $\boldsymbol{t}_n$ and dual spanning tree $\boldsymbol{t}_n^\dagger$ in the sense that they are the contour functions of these trees: at a given time k, X_k denotes the height in the tree (distance to the root) of the last vertex discovered prior to time k.

Remark 4.2. It may be useful to recast the above connections in the language of **queues**, where customers are being served one at the time. More precisely, a queue (in discrete time) is a process where at each unit of time either a new customer arrives, or a customer at the front of the queue is being served and leaves the queue forever. Any queue can be equivalently described by a tree $\boldsymbol{t}$ or an excursion X. Indeed, a tree structure $\boldsymbol{t}$ can be defined from the queue, by declaring that any customer c arriving during the service of a



e. **f.**

Figure 9: **e:** The word associated to $(\boldsymbol{m}_n, \boldsymbol{t}_n)$ is: $w = \mathsf{hhhcHcHhCcHHCC}$ **f:** The hamburger and cheeseburger counts, as well as the trees encoded by these excursions (which are identical to the primal and dual spanning trees, respectively).

customer c' is a child of c'. An excursion X can be defined by simply counting the queue length at each time. Note that X is nothing but the contour function of the tree $\boldsymbol{t}$ (meaning the discrete process which measures the height of the tree $\boldsymbol{t}$ as it goes around it in depth-first order; see [LG05] for much more about this). In our case, the tree $\boldsymbol{t}$ is simply the spanning tree on the map.

When (M_n, T_n) are *random* and sampled according to (4.4), the corresponding random excursion (X, Y) is clearly chosen uniformly from the set of all possibilities. It therefore follows from classical results of Durrett, Ingleheart and Miller [DIM77] that as $n \to \infty$,

$$\frac{1}{\sqrt{n}}(X_{\lfloor 2nt \rfloor}, Y_{\lfloor 2nt \rfloor})_{0 \le t \le 1} \to (e_t, e'_t)_{0 \le t \le 1}$$

where e, e' are independent Brownian (one-dimensional) excursions (i.e., the pair (e, e') is Brownian excursion in the quarter plane). This property implies (see e.g. Lemma 2.4 in Le Gall's comprehensive survey [LG05]) that, in the Gromov–Hausdorff sense, the primal and dual spanning trees converge after rescaling the distances by a factor $n^{-1/2}$, to a pair of independent **Continuous Random Trees** (CRTs) [Ald93].

We summarise our findings, in the case of UST weighted random planar maps, in the following theorem.

Theorem 4.3. *The set of (rooted) spanning-tree decorated maps $(\boldsymbol{m}_n, \boldsymbol{t}_n)$ with n edges are in bijection with excursions $(X_k, Y_k)_{0 \le k \le 2n}$ in the quarter plane. When (M_n, T_n) is random and distributed according to (4.4), the pair of trees $(T_n, T_n^\dagger)$ converges after scaling distances (in each tree) by a factor $n^{-1/2}$, to a pair of independent Continuous Random Trees (CRTs).*

Note that the map M_n itself can then be thought of as a gluing of two discrete trees (i.e. the primal and dual spanning trees, which are glued along the space-filling path). In the scaling limit, this pair of trees becomes a pair of independent CRTs. As it turns out,

the procedure of gluing these two trees has a continuum analogue, which is described in the work of Duplantier, Miller and Sheffield [DMS14]. This is the **mating of trees** approach to LQG, and is an extremely powerful and fruitful point of view that we will describe in more detail later on. (To be added.)

4.4 The loop-erased random walk exponent

A loop-erased random walk (or LERW for short) is the process that one obtains when erasing the loops chronologically as they appear on a simple random walk trajectory. There is a well-known and very deep connection between uniform spanning trees and loop-erased random walks, which was discovered by Wilson [Wil96], and may be used to efficiently simulate such trees. This relation is known as **Wilson's algorithm**; see Chapter 4 of [LP16] for a thorough discussion. Here we will only need the following result, which may be seen as a straightforward consequence of Wilson's algorithm, but which was first discovered by Pemantle [Pem91] (prior to [Wil96]). We state and prove it here, since the proof is short and rather beautiful.

Theorem 4.4. *If x and y are two points of an arbitrary finite (connected, undirected) graph G, and if T is a UST on G, then the branch of T from x to y has the same distribution as a loop-erased random walk from x run until it hits y.*

Proof. For a possibly infinite path $\gamma = (\gamma_0, \gamma_1, \ldots)$ on the vertices V of the graph, let $T(\gamma)$ be the rooted tree obtained by retaining only the edges (γ_j, γ_{j+1}) which do not close a loop: i.e., keep this edge if and only if there is no $i < j$ such that $\gamma_i = \gamma_j$ (the root is the starting point γ_0 of the path). It is obvious that this generates an acyclic graph; if the path visits every vertex then $T(\gamma)$ is a spanning tree. Now suppose that $(X_n)_{n\geq 0}$ is a stationary random walk on G, so that X_0 is distributed according to its equilibrium measure π, and let $\gamma_n = (X_n, X_{n+1}, \ldots)$ be the path started from X_n. Then the claim is that $(T(\gamma_n), X_n)$ defines a Markov chain on rooted spanning trees. This is best seen by viewing the tree $T(\gamma_n)$ as rooted at the directed edge $e_n = (X_n, X_{n+1})$. Then one step of the Markov chain is as follows. First choose an e' uniformly at random from the neighbours of the head e^+ of the edge e. Add it to the tree T, this now creates a cycle (possibly a double edge). Then remove form the tree T the unique edge touching the tail e^- in this cycle. See [LP16, Chapter 4.4] for an illuminating alternative description of this argument.

When X is stationary then so is $(T(\gamma_n), X_n)$. But the unique invariant (reversible) measure for such a chain is clearly $\tilde{\pi}(\boldsymbol{t}, v) = \deg v$, the degree of the root of T. This is known as the **Markov chain tree theorem**. This particular algorithm for generating a uniform spanning tree is known as the **Aldous–Broder algorithm**. . Hence, when we condition on $X_0 = v$, we get the conditional distribution of the unrooted tree T obtained from ignoring the root in the rooted tree $T(\gamma_0)$ as

$$\mathbb{P}(T = \boldsymbol{t} | X_0 = v) = \frac{\mathbb{P}((T(\gamma_0), X_0) = (\boldsymbol{t}, v))}{\pi(v)}$$

$$\propto \frac{\tilde{\pi}(\boldsymbol{t}, v)}{\deg(v)} \propto 1.$$

So T is a uniform (unrooted) spanning tree, and note that (remarkably) the law of T hence does not depend on the starting point v.

Now suppose the random walk X is started at $X_0 = x$, and consider the above construction of $T(X)$ which yields (as shown above) a uniform spanning tree. It is easy to check that the branch β of $T(X)$ connecting x to y is a simple random walk from x to y from which we have removed all the loops in a *reverse-chronological* order: that is, let γ be the path X run until it hits y, and let γ^* denote its time reversal. Then β is the loop-erasure $\Lambda(\gamma^*)$. The result is therefore proved if we show that $\Lambda(\gamma) = \Lambda(\gamma^*)$ in distribution, when viewed as a set of edges. The reason why this is true is because the ways the cycle are traversed can always be rearranged so that the backward erasure becomes the forward loop-erasure, and the probability of the whole path is unchanged by this reordering.

Another (perhaps more direct) way of seeing this is by considering a bi-infinite stationary version of the random walk $(\gamma_n, n \in \mathbb{Z})$ and looking at the tree associated with the time reversal $\hat{\gamma}_n = \gamma_{-n}$. Again, we refer to [LP16] for more details on this argument. □

Therefore, we may deduce from Theorem 4.3 the following result about the loop-erased random walk.

Theorem 4.5. *Let (M_n, T_n) be chosen as in (4.4) and let x, y be two vertices chosen independently and uniformly on M_n. Let $(\Lambda_k)_{0 \le k \le \xi_n}$ be a LERW starting from x, run until the random time ξ_n when it first hits y. Then*

$$\frac{\xi_n}{\sqrt{n}} \to \xi_\infty$$

in distribution, where ξ_∞ is a random variable that has a nondegenerate distribution (in the sense that $\xi_\infty \in (0, \infty)$ a.s.)

Proof. Let $(X_k, Y_k)_{1 \le k \le 2n}$ be the pair of excursions which describes the map (M_n, T_n). Then note that ξ_n may be identified with the value of the excursion X at a time k that is uniformly (and independently) chosen between 1 and $2n$. As a consequence, Theorem 4.5 holds with $\xi_\infty = e(U)$, where e is a Brownian excursion e and U is chosen uniformly from $(0, 1)$. □

Scaling exponent of LERW. We now explain how the above result can be used to compute an exponent for the loop-erased random walk. Let $\Lambda = \{\Lambda_0, \dots, \Lambda_{\xi_n}\}$ denote the loop-erasure of a random walk on M_n, run from a uniformly chosen vertex x until the hitting time of another uniformly chosen vertex y, as above. Then Λ may be viewed as an independent random "fractal" set on M_n, whose size is $|\Lambda| = n^{1/2+o(1)}$. Since M_n has $n^{1+o(1)}$ vertices, this means that Λ has a **quantum scaling exponent** given by

$$\Delta = 1/2$$

(recall our discussion from Section 3.13). We can therefore (at least informally) use the **KPZ relation** to compute the equivalent exponent for a loop-erased random walk on the square lattice. To do so, we must first find the correct value of γ: the constant in front of the GFF which describes the scaling limit of the conformally embedded planar map M_n. This is given by the relation (4.2) when $q = 0$ (which, as explained at the beginning of this section, indeed corresponds to the uniform spanning tree weighted map model of (4.4)). Plugging $q = 0$ yields

$$\gamma = \sqrt{2}.$$

Note that this is consistent with the conjecture (known to be true on the square lattice by results of [LSW04]) that the interface separating a uniform spanning tree from its dual, converges in the scaling limit to an SLE curve with parameter $\kappa' = 8$.

Therefore, the **Euclidean scaling exponent** x of the loop-erased random walk should satisfy

$$x = \frac{\gamma^2}{4}\Delta^2 + (1 - \frac{\gamma^2}{4})\Delta = 3/8.$$

In particular, we conclude that in the scaling limit, a loop-erased random walk on the square lattice has dimension

$$d_{\text{Hausdorff}} = 2 - 2x = 5/4.$$

This is in accordance with Beffara's formula [Bef08] for the dimension of SLE: indeed, in the scaling limit, LERW is known to converge to an SLE_κ curve with $\kappa = 2$. This is closely related to the above-mentioned scaling limit result for the UST, due to Lawler, Schramm and Werner [LSW04], and is proved in the same paper. Beffara's result [Bef08] states that the Hausdorff dimension of SLE_κ is $(1 + \kappa/8) \wedge 2$. In the case $\kappa = 2$ this is exactly $5/4$, as above.

In fact, this exponent for LERW had earlier been derived by Kenyon in a remarkable paper [Ken00], building on his earlier work on the dimer model and the Gaussian free field [Ken01].

4.5 Sheffield's bijection in the general case

We now describe the situation when $\bar{\boldsymbol{m}}_n \in \mathcal{M}_n$ but the collection of edges $\boldsymbol{t}_n$ is arbitrary (i.e., not necessarily a tree), which is more delicate. Note that in the case of spanning trees there was only one loop present, but now there will generally be more than one. These loops are **densely packed** in the sense that every triangle is part of some loop, as illustrated in Figure 3. Indeed, each triangle consists of an edge of some type and a vertex of the opposite type, so must contain a loop separating the two associated clusters. In this case we will see that we can still define a canonical space-filling interface (i.e., a curve which visits every single triangle exactly once). We will now describe this curve.

Recall that L_0 is the loop containing the root triangle of the map $\bar{\boldsymbol{m}}_n$, oriented parallel to the orientation of the root edge of $\boldsymbol{m}_n$. We view L_0 as an oriented collection of adjacent triangles (the triangles traversed by the loop). In general, L_0 does not cover every triangle of $\bar{\boldsymbol{m}}_n$, and we may consider the connected components $C_1, \ldots, C_k$ which are obtained by

removing all the triangles of L_0. Note that L_0 is adjacent to each of these components, in the sense for each $1 \leq i \leq k$, it contains a triangle that is opposite a triangle in C_i. For each i, let T_i be the *last* (with respect to the orientation of the loop and its starting point) triangle that is adjacent to C_i. The triangle opposite T_i is in C_i and together they form a quadrangle. In order to explore all of the map and not just L_0, we will first modify the map by *flipping* the diagonal of this quadrangle, for every $1 \leq i \leq k$. It can be seen that having done so, we have reduced the number of loops on the map by exactly k (each such flipping has the effect of merging two loops). We may then iterate this procedure until there is only a single loop left, the loop L_0 (which now fills the whole map). This loop separates primal and dual clusters of the modified map, in the sense that it has only primal clusters on one side, and dual clusters on the other (we will see below that these clusters are in fact spanning trees).

So we now have a canonical space-filling path which allows us to explore the map as in Section 4.3. As before, we can describe the type of triangles we see in this exploration using the symbols $\mathsf{h}, \mathsf{c}, \mathsf{H}, \mathsf{C}$. When we explore a triangle corresponding a flipped quadrangle for the first time, we record its type (either h, c) according to its type after having flipping the edge. However, when we visit its opposite triangle we record the fact that this is a special edge (which must be flipped to recover the original map) by the symbol F. The letter F stands for "flexible" or "freshest" order. (We will see below a more precise interpretation in terms of queues, or hamburgers and cheeseburgers.) In this way, we may associate to the decorated map $(\boldsymbol{m}_n, \boldsymbol{t}_n)$ a list w of $2n$ symbols $w = (X_i)_{1 \leq i \leq 2n}$ taking values in the alphabet $\Theta = \{\mathsf{h}, \mathsf{c}, \mathsf{H}, \mathsf{C}, \mathsf{F}\}$.

We will see below the properties of this word (essentially, it reduces to $\emptyset$ with the appropriate definition of reduction when there is an F) and that the map from $(\boldsymbol{m}_n, \boldsymbol{t}_n)$ to w subject to this constraints is a bijection. For now, we make the important observation that each loop corresponds to a unique symbol F, except for the loop through the root.

Inventory accumulation. Recall that we can interpret an element in $\{\mathsf{h}, \mathsf{c}, \mathsf{H}, \mathsf{C}\}^{2n}$ as a last-in, first-out inventory accumulation process in a burger factory with two types of product: hamburgers and cheeseburgers. Think of a sequence of events, occurring once per unit time, in which either a burger is produced (either ham or cheese) or there is an order of a burger (either ham or cheese). The burgers are put in a single **stack** and every time there is an order of a certain type of burger, the freshest burger in the stack of the corresponding type is removed. The symbol h (resp. c) corresponds to a ham (resp. cheese) burger production and the symbol H (resp. C) corresponds to a ham (resp. cheese) burger order.

The inventory interpretation of the symbol F is the following: this corresponds to a customer demanding the freshest or the topmost burger in the stack, irrespective of the type. In particular, whether an F symbol corresponds to a hamburger or a cheeseburger order depends on the topmost burger type at the time of the order. Thus overall, we can think of the inventory process as a sequence of symbols in Θ with the following reduction rules

- $\overline{\mathsf{cC}} = \overline{\mathsf{cF}} = \overline{\mathsf{hH}} = \overline{\mathsf{hF}} = \emptyset$,

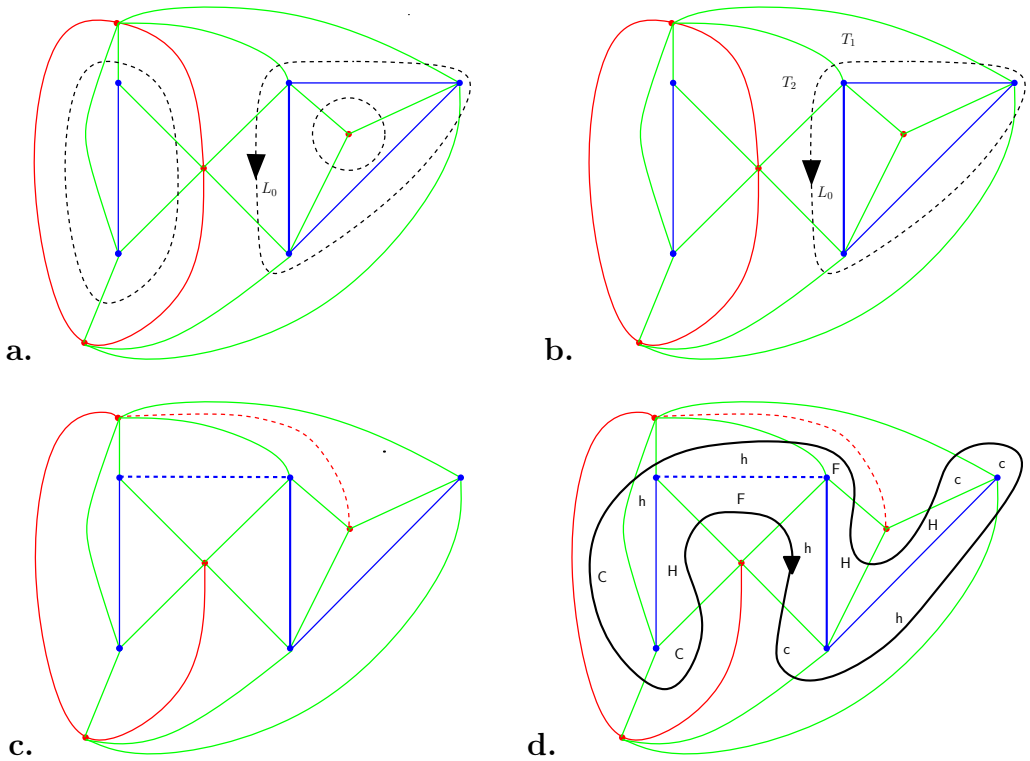


Figure 10: Generating a word from a decorated map in the general case. **a.** The decorated map is as in Figure 3, with the (oriented) root loop L_0. **b.** The complement of L_0 consists of two components, C_1 and C_2. T_1 and T_2 are the *last* triangles visited by the loop L_0 that share an edge with a triangle in C_1 and C_2 respectively. **c.** We flip the diagonals of the quadrangles associated with T_1 and T_2. **d.** We obtain a single space-filling loop (drawn in black). To this path we can again associate a word in {h, c, H, C}. However, we also record the second visit to a flipped quadrangle by replacing the symbol C or H by the symbol F. The word here is thus hchccHHFhhCCHF. Note the non obvious fact that after flipping, the primal and dual clusters have become trees.

- $\overline{\mathsf{cH}} = \overline{\mathsf{Hc}}$ and $\overline{\mathsf{hC}} = \overline{\mathsf{Ch}}$.

Given a sequence of symbols w, we denote by $\bar{w}$ the reduced word formed via the above reduction rule.

Reversing the construction. Given a sequence w of symbols from Θ, such that $\bar{w} = \emptyset$, we can construct a decorated map $(\boldsymbol{m}_n, \boldsymbol{t}_n)$ as follows. First, we convert all the F symbols to either an H or a C symbol depending on its order type. Then, we construct a spanning tree decorated map as described in Section 4.3 (see in particular Figure 8). The condition $\bar{w} = \emptyset$ ensures that we can do this. To obtain the original loop-decorated map, we simply flip the type of every quadrangle which has one of the triangles corresponding to an F symbol. That is, if a quadrangle formed by primal triangles has one of its triangles coming from an F symbol, then we replace the primal map edge in that quadrangle by the corresponding dual edge and vice versa. The interface is now divided into several loops (and the number of loops is exactly one more than the number of F symbols). In particular:

Theorem 4.6 (Sheffield, [She16b]). *The map $(\boldsymbol{m}_n, \boldsymbol{t}_n) \mapsto w$ (subject to $\overline{w} = \emptyset$) is a bijection.*

Two canonical spanning trees. It is not obvious but true that after flipping, the corresponding primal and dual decorations of the map have become two mutually dual spanning trees. One way to see this is as follows: observe that after flipping, we have (as already argued) a single space-filling loop which separates primal and dual clusters of the resulting modified map. These clusters are of course spanning, and they cannot contain non-trivial cycles, else the loop would either not be space-filling or consist of multiple loops. Therefore, we can again think of M_n as a gluing of two spanning trees, which are glued along the space-filling path (i.e., along their contour functions). Again, this perspective is a crucial intuition which guides the **mating of trees approach** to Liouville quantum gravity [DMS14]. We will survey this later on.

Generating FK-weighted maps. A remarkable consequence of Theorem 4.6 is the following simple way of generating a random planar map from the FK model (4.1). Fix $p \in [0, 1/2)$, which will be suitably chosen (as a function of q) below in (4.8). Let $(X_1, \dots, X_{2n})$ be i.i.d. with the following law

$$\mathbb{P}(\mathsf{c}) = \mathbb{P}(\mathsf{h}) = \frac{1}{4}, \mathbb{P}(\mathsf{C}) = \mathbb{P}(\mathsf{H}) = \frac{1-p}{4}, \mathbb{P}(\mathsf{F}) = \frac{p}{2}, \tag{4.6}$$

conditioned on $\overline{X_1, \dots, X_n} = \emptyset$.

Let (M_n, T_n) be the random associated decorated map (via the bijection described above). Then observe that since n hamburgers and cheeseburgers must be produced, and since $\#\mathsf{H} + \#\mathsf{C} = n - \#\mathsf{F}$,

$$\mathbb{P}((M_n, T_n) = (\boldsymbol{m}_n, \boldsymbol{t}_n)) = \left(\frac{1}{4}\right)^n \left(\frac{1-p}{4}\right)^{\#\mathsf{H}+\#\mathsf{C}} \left(\frac{p}{2}\right)^{\#\mathsf{F}}$$

$$\propto \left(\frac{2p}{1-p}\right)^{\#\mathsf{F}} = \left(\frac{2p}{1-p}\right)^{\#\ell(\boldsymbol{m}_n,\boldsymbol{t}_n)-1} \tag{4.7}$$

Thus we see that (M_n, T_n) is a realisation of the critical FK-weighted cluster random map model with

$$\sqrt{q} = \frac{2p}{(1-p)}. \tag{4.8}$$

Notice that $p \in [0, 1/2)$ corresponds to $q = [0, 4)$. From now on we fix the value of p and q in this regime. Recall that $q = 4$ is believed to be a critical value for many properties of the map; indeed later on we will later show that a phase transition occurs at $p = 1/2$ ($q = 4$) for the geometry of the map. Intuitively, it is perhaps not surprising that the value $p = 1/2$ marks a distinction from the point of view of inventory accumulation.

4.6 Infinite volume limit

The following theorem due to Sheffield [She16b], and made more precise later by Chen [Che17], shows that the decorated map (M_n, T_n) has a local limit as $n \to \infty$ in the local topology. Roughly two (decorated) maps are close in the local topology if the finite maps (and their decorations) near a large neighbourhood of the root are isomorphic as decorated maps.

Theorem 4.7 ([She16b, Che17]). *Fix $p \in [0, \frac{1}{2})$. We have*

$$(M_n, T_n) \xrightarrow[n\to\infty]{(d)} (M, T)$$

with respect to the local topology, where (M, T) is the unique infinite decorated map associated with a bi-infinite i.i.d. sequence of symbols $(X_n)_{n\in\mathbb{Z}}$ having law (4.6).

Sketch of proof. We now give the idea behind the proof of Theorem 4.7. Let $X_1, \dots, X_{2n}$ be i.i.d. with law given by (4.6), and denote by E_{2n} the event that $\overline{X_1 \dots X_{2n}} = \emptyset$.

A key step is to show the following.

Lemma 4.8 ([She16b, Che17]). *Let $X_1, \dots, X_{2n}$ be i.i.d. with law (4.6). Then $\mathbb{P}(E_{2n})$ decays subexponentially in n, i.e., $\log \mathbb{P}(E_{2n})/n \to 0$ as $n \to \infty$.*

We will not prove this statement (although we will later come back to it and explain it informally). Instead we explain how Theorem 4.7 follows.

Notice that uniformly selecting a symbol $1 \le I \le 2n$ corresponds to selecting a uniform triangle in $(\bar{M}_n, \bar{T}_n)$, which in turn corresponds to a unique oriented edge in M_n. Because of invariance of the decorated map (M_n, T_n) under re-rooting, we claim that it suffices to check the convergence in distribution of a large neighbourhood of the triangle corresponding to X_I in $\bar{M}_n$.

Let $r > 0$. We will first show that for any fixed word w of length $2r + 1$ in the alphabet Θ,

$$\mathbb{P}(X_{I-r} \dots X_{I+r} = w \big| E_{2n}) \to \mathbb{P}(w) := \mathbb{P}(X_{-r} \dots X_r = w). \tag{4.9}$$

To see (4.9), observe that that the conditional probability on the left hand side is equal to $f + o(1)$ as $n \to \infty$, where f is the fraction of occurrences of w in X, and the $o(1)$ term is uniform. Hence it suffices to check that $\mathbb{E}(f|E_{2n}) \to \mathbb{P}(w)$. To do this, for arbitrary $\varepsilon > 0$ we define $A_n = \{|f - \mathbb{P}(w)| \leq \varepsilon\}$, and write

$$\mathbb{E}(f|E_{2n}) = \mathbb{E}(f1_{A_n}|E_{2n}) + \mathbb{E}(f1_{A_n^c}|E_{2n}).$$

Now the first term $\mathbb{E}(f1_{A_n}|E_{2n})$ is equal to $(\mathbb{P}(w) + O(\varepsilon))\mathbb{P}(A_n|E_{2n}))$, while the second term satisfies

$$\mathbb{E}(f1_{A_n^c}|E_{2n}) \leq \mathbb{P}(A_n^c|E_{2n}) \leq \frac{\mathbb{P}(A_n^c)}{\mathbb{P}(E_{2n})}.$$

However, $\mathbb{P}(A_n^c) \to 0$ *exponentially fast* as $n \to \infty$, by basic large deviation estimates (Cramer's theorem). This means that $\mathbb{E}(f1_{A_n^c}|E_{2n})$ converges to zero by Lemma 4.8, and also that $\mathbb{P}(A_n|E_{2n}) \to 1$ as $n \to \infty$. We can conclude that $\mathbb{E}(f1_{A_n})$ and therefore $\mathbb{E}(f|E_{2n})$ converges to $\mathbb{P}(w)$ as $n \to \infty$, which proves (4.9).

To conclude the theorem, it remains to show that convergence of the symbols locally around a letter implies local convergence of the maps. This is a consequence of Exercise 1. □

One important feature arising from the proof of Lemma 4.8 in [She16b], is that every symbol in the i.i.d. sequence $\{X_i\}_{i\in\mathbb{Z}}$ has an almost sure unique **match**, meaning that every burger order is fulfilled (it corresponds to a burger that was produced at a finite time before), and every burger that is produced is consumed at some finite later time, both with probability 1. In the language of maps, this is equivalent to saying that the map M has no edge "to infinity". For future reference, let $\varphi(i)$ denote the match of the ith symbol. Notice that $\varphi : \mathbb{Z} \mapsto \mathbb{Z}$ defines an involution on the integers.

4.7 Scaling limit of the two canonical trees

We now state (without proof) one of the main results of Sheffield [She16b], which gives a scaling limit result for the geometry of the infinite volume map (M, T) defined in Theorem 4.7. Recall that (M, T) is completely described by a doubly infinite sequence $(X_n)_{n\in\mathbb{Z}}$ of i.i.d symbols in the alphabet Θ, having law (4.6). Associated to such a sequence we can define two processes $(H_n)_{n\in\mathbb{Z}}$ and $(C_n)_{n\in\mathbb{Z}}$ which count the respective number of hamburgers and cheeseburgers present in the queue at time $n \in \mathbb{Z}$. These numbers are defined relative to time 0, so $(H_0, C_0) = (0, 0)$.

This scaling limit is most conveniently phrased as a scaling limit for $H = (H_n)_{n\in\mathbb{Z}}$ and $C = (C_n)_{n\in\mathbb{Z}}$ (although the statement of Sheffield [She16b] concerns instead $H + C$ and the discrepancy $H - C$). We first state the result and then make some comments on its significance below.

Theorem 4.9. *Let $p \in [0, 1]$, and let C, H be as above. Then*

$$\left(\frac{H_{\lfloor nt\rfloor}}{\sqrt{n}}, \frac{C_{\lfloor nt\rfloor}}{\sqrt{n}}\right)_{-1\leq t\leq 1} \to (L_t, R_t)_{-1\leq t\leq 1}$$

as $n \to \infty$*, where* $(L_t, R_t)_{t\in\mathbb{R}}$ *is a two-sided Brownian motion in* $\mathbb{R}^2$*, starting from 0 and having covariance matrix given by*

$$\mathrm{Var}(L_t) = \mathrm{Var}(R_t) = \frac{1+\alpha^2}{4}|t| \quad ; \quad \mathrm{Cov}(L_t, R_t) = \frac{1-\alpha^2}{4}|t|$$

and

$$\alpha = \max(1-2p, 0).$$

See [She16b, Theorem 2.5] for a proof. We now make a few important remarks about this statement.

- This scaling limit result should be thought of as saying something about the large scale geometry of the map (M, T). Equivalently, what it looks like after scaling down by a large factor. However, what this actually means is not a priori obvious: really, the theorem only says that the pair of trees converge to correlated (infinite) CRTs. Later, we will see that this can be viewed (by definition) as convergence in the **peanosphere topology**, which is a fairly weak notion of convergence. In particular, it does not say anything about convergence of the metric on M.

- Notice that when $p \geq 1/2$ (corresponding to $q \geq 4$ in terms of the FK model (4.1), see (4.8)) we have $\alpha = 0$, so $L_t = R_t$ for all $t \in \mathbb{R}$. This is because the proportion of F orders is large enough that there can be no discrepancy in the scaling limit between hamburgers and cheeseburgers.

- However, when $p \leq 1/2$ (corresponding to $q \leq 4$), the correlation between L and R is non-trivial. When $p = 0$ (corresponding to $q = 0$) they are actually independent. This last case should be compared with the case of spanning-tree weighted maps (Theorem 4.3). In general, this suggests that the scaling limit of the map (M, T), if it exists, can be viewed as a gluing of two (possibly correlated) infinite CRTs; meaning that their contour (or alternatively their height) functions are described by a two-sided infinite Brownian motion (rather than a Brownian excursion of duration one). This fact is made rigourous (and will be discussed later on) in the **mating of tree approach to LQG** of [DMS14]. Note in particular that in the case $q \geq 4$, the two corresponding trees are identical, meaning that the map should degenerate to a CRT in the scaling limit. This is in contrast with the case $q < 4$, where the limit maps are expected to be homeomorphic to the sphere almost surely.

- H_n, C_n also have a geometric interpretation, as the boundary lengths at time n on the left- and right-hand sides of the space-filling interface (relative to time 0).

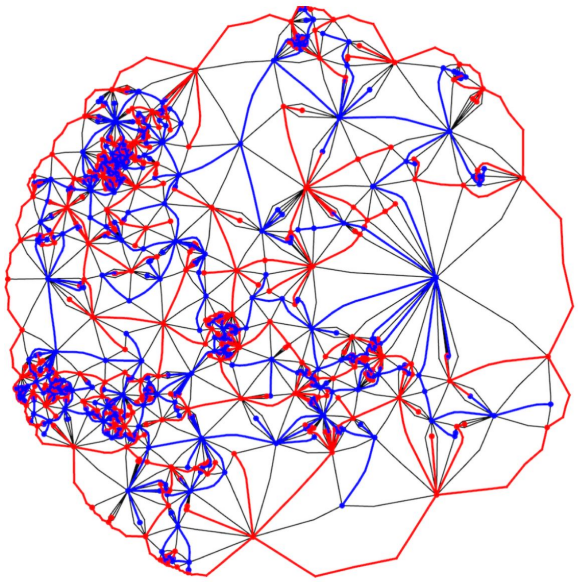


Figure 11: An random planar map with law (4.1) for $q = 1$ (uniform case), generated using Sheffield's bijection of Theorem 4.6. The map has been embedded using circle packing. Shown in blue and red are the primal and dual spanning trees. In the infinite volume limit and then in the scaling limit, these trees become correlated infinite CRTs. This is guaranteed by Theorem 4.9.

4.8 Exponents associated with FK weighted random planar maps

⋆ *In this short section, some critical exponents of random planar maps are computed heuristically. This section can be skipped on a first reading, as none of those results are needed later on.*

It is possible to use Theorem 4.9 to obtain very precise information on the geometry of loops on the map (M, T). In particular, it is possible to check that large loops have statistics that coincide with those of $\mathrm{CLE}_{\kappa'}$, where the value of κ' is related to $q \in (0, 4)$ via (4.3), thereby giving credence to the general conjectures formulated in Section 4.2. This line of reasoning has been pursued very successfully in a string of papers by Gwynne, Mao and Sun [GMS19, GS17, GS15]. We will present here a slightly less precise (but easier to state) result proved in [BLR17]. Let L_0 denote the loop containing the root triangle in the infinite map (M, T); here as usual we identify a loop with the set of triangles through which it passes. Let L denote its length (the number of triangles it passes through) and let A denote its area, meaning the number of triangles surrounded by it (where the "inside" of the loop is the connected component of the loop which does not contain infinity).

Let

$$p_0 = \frac{\pi}{4 \arccos\left(\frac{\sqrt{2-\sqrt{q}}}{2}\right)} = \frac{\kappa'}{8} \in (1/2, 1), \tag{4.10}$$

where q and κ' are related as in (4.2). The following is the main result in [BLR17].

Theorem 4.10. *Let $0 < q < 4$. The random variables L and A satisfy*

$$\mathbb{P}(\mathsf{L} > k) = k^{-1/p_0 + 1 + o(1)}, \tag{4.11}$$

and

$$\mathbb{P}(\mathsf{A} > k) = k^{-(1-p_0)+o(1)} \tag{4.12}$$

as $k \to \infty$.

Results in [GMS19, GS17, GS15] are analogous and more precise, in particular showing regular variation of the tail at infinity. (As a consequence, the sum of loop lengths and areas, in the order that they are discovered by the space-filling path, can be shown to converge after rescaling to a stable Lévy process with appropriate exponent).

A particular consequence of Theorem 4.10 is that we expect the longest loop in the map M_n to have size roughly $n^{p_0+o(1)}$; that is,

$$\max_{\ell \in (M_n, T_n)} |\ell| = n^{p_0+o(1)} \tag{4.13}$$

as $n \to \infty$. Indeed, first note that the loop L_0 containing the origin is biased by its length (since it contains 0) and so for a uniformly chosen loop L, we expect the length exponent to satisfy

$$\mathbb{P}(|L| \geq k) = k^{-1/p_0+o(1)}.$$

Heuristically, to derive (4.13), one then observes that M_n contains order n loops whose lengths are roughly i.i.d. with tail exponent $\alpha = 1/p_0$. The maximum value of this sequence of lengths is then easily shown to be of order $n^{1/\alpha+o(1)} = n^{p_0+o(1)}$.

We will not prove Theorem 4.10, but we will discuss in Exercise 4 an interesting application using the KPZ formula. These exponents are obtained (both in [BLR17] and [GMS19, GS17, GS15]) through a connection with a random walk in a cone. A simple setting, where it is easier to see this connection, is in the following result.

Proposition 4.11 ([GS17]). *Let $0 < q < 4$, and let E_{2n} be the event that the word $w = X_1 \dots X_{2n}$ reduces to $\bar{w} = \emptyset$. Then*

$$\mathbb{P}(E_{2n}) = n^{-2p_0-1+o(1)} = n^{-1-\kappa'/4+o(1)},$$

as $n \to \infty$. In particular, $\mathbb{P}(E_{2n})$ decays subexponentially.

Sketch of proof. We give a rough idea of where this exponent comes from, as it allows us to illustrate the connection to random walk in a cone, as mentioned above. A rigourous proof of this result may be found in [GS17].

The first step is to describe E_{2n} in terms of the burger count processes H and C of Theorem 4.9. In particular, we note that the event E_{2n} is equivalent to the conditions

- $C_i, H_i \geq 0$ for $0 \leq i \leq 2n$; and
- $C_{2n} = 0$, $H_{2n} = 0$

on H and C. Indeed, the first condition holds since if at some point $1 \leq k \leq 2n$ the burger count C or H becomes negative, this must be because of an order whose match in the bi-infinite sequence $(X_k)_{k\in\mathbb{Z}}$ was in the past, i.e. $\varphi(k) < 0$. Therefore, the event E_{2n} is equivalent to the process $Z_k = (C_k, H_k)_{1\leq k\leq 2n}$ being an excursion in the top right quadrant of the (C, H) plane, starting and ending at the origin.

This probability may be computed approximately (or rather, heuristically here) using Theorem 4.9. To do this it is useful to apply first a linear map of the (C, H) plane so as to deal with independent Brownian coordinates in the limit. More precisely, we apply the linear map Λ defined by

$$\Lambda = (1/\sigma) \begin{pmatrix} 1 & \cos(\theta_0) \\ 0 & \sin(\theta_0) \end{pmatrix},$$

where $\theta_0 = \pi/(2p_0) = 4\pi/\kappa' = 2\arctan(\sqrt{1/(1-2p)}$ and $\sigma^2 = (1-p)/2$. A direct but tedious computation shows that $\Lambda(L_t, R_t)$ is indeed a standard planar Brownian motion. (The computation is easier to do by reverting to the original formulation of Theorem 4.9 in [She16b], where it is shown that $C + H$ and $(C - H)/\sqrt{1-2p}$ converge to a standard planar Brownian motion). Note that the top right quadrant transforms under Λ, see Figure 12, into the cone $\mathcal{C}(\theta_0)$ of angle θ_0 with apex at zero.

We therefore consider the analogous question for two-dimensional Brownian motion. Namely, let B be a standard planar Brownian motion, starting from some point $z \in \mathcal{C}(\theta_0)$

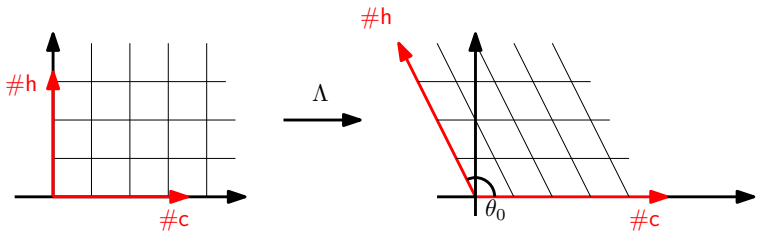


Figure 12: The coordinate transformation. In these new axis, the burger counts H and C become independent Brownian motions; the event E_{2n} then corresponds to $\Lambda(Z)$ making a long excursion in the cone $\mathcal{C}(\theta_0)$ of angle $\theta_0 = \pi/(2p_0) = 4\pi/\kappa'$, starting and and ending at its apex.

with $|z| = 1$. Let T be the first time that B leaves $\mathcal{C}(\theta_0)$. Then from Theorem 4.9 it is reasonable to guess that

$$\mathbb{P}(E_{2n}) \approx \mathbb{P}_z(T > t; |B_t| \leq 1), \text{ with } t = n^{1+o(1)}.$$

For this we first claim that

$$\mathbb{P}(T > t) = t^{-p_0+o(1)} \tag{4.14}$$

as $t \to \infty$. To see why this is the case, consider the conformal map $z \mapsto z^{\pi/\theta_0}$, which sends the cone $\mathcal{C}(\theta_0)$ to the upper-half plane. In the upper-half plane, the function $z \mapsto \Im(z)$ is harmonic with zero boundary condition, and so in the cone, the function

$$z \mapsto g(z) := r^{\pi/\theta_0} \sin\left(\frac{\pi\theta}{\theta_0}\right); \quad z \in \mathcal{C}(\theta_0),$$

is also harmonic. Applying the optional stopping theorem at time $T \wedge t$, the only nonzero contribution to $M_{T\wedge t}$ comes from the event $T > t$. On the other, conditionally on $T > t$, then $|B_t$ is likely to be at distance $\sqrt{t}$, in which case $M_{T\wedge t} \approx t^{\pi/(2\theta_0)} = t^{p_0}$. It is not hard to deduce (4.14).

We now claim that the desired probability

$$\mathbb{P}_z(T > t; |B_t| \leq 1) = t^{-2p_0-1+o(1)} \text{ as } t \to \infty. \tag{4.15}$$

To see this, we split the interval $[0, t]$ into three intervals of equal length $t/3$. In order for the event on the left-hand side to be satisfied, three things must happen during these three intervals.

- Over the interval $[0, t/3]$, B must not leave the cone. This has probability $t^{-p_0+o(1)}$ by (4.14).
- At the other extreme, if we reverse the direction of time, we also have a Brownian motion started close to the tip of the cone that must not leave the cone for time $t/3$. Again, this has probability $t^{-p_0+o(1)}$.

- Finally, given the behaviour of the process over $[0, t/3]$ and $[2t/3, t]$, the process must go from $B_{t/3}$ to $B_{2t/3}$ in the time interval $[t/3, 2t/3]$, and stay inside the cone. The latter requirement actually has probability bounded away from zero (because $B_{t/3}$ and $B_{2t/3}$ are typically far away from the boundary of the cone), so it remains to compute the probability to transition between these two endpoints. However this is roughly of order $t^{-1+o(1)}$, since we are dealing with a Brownian motion in dimension two.

Altogether, we obtain that $\mathbb{P}_z(T > t; |B_t| \leq 1) = t^{-2p_0-1+o(1)}$, as desired.

□

4.9 Exercises

1. Let $x_1, \ldots, x_{2n}$ be a sequence of $2n$ letters in the alphabet $\Theta = \{\mathsf{c}, \mathsf{h}, \mathsf{C}, \mathsf{H}, \mathsf{F}\}$ and suppose that the corresponding word $w = x_1 \ldots x_{2n}$ reduces to $\bar{w} = \emptyset$. For each $1 \leq i \leq 2n$, denote by $\varphi(i)$ the unique match of i: meaning that if i corresponds to production of a specific burger, then $\varphi(i)$ is the unique time at which this burger is consumed, and vice versa.

   Let us draw a map $\boldsymbol{m}$ as follows. Start with the line segment having vertices $1, \ldots, 2n$ and horizontal nearest neighbour edges. Draw an arc between i and $\varphi(i)$ for each $1 \leq i \leq 2n$; this arc is drawn in the upper half plane for a hamburger, and in the lower half plane for a cheeseburger.

   (a) Show that the arcs can be drawn in a planar way (so they don't cross one another).

   (b) Then show that this results in a planar map which is identical the planar map encoded by the word w via Sheffield's bijection. See [Che17] for additional details on this interpretation. Deduce that local convergence of maps is equivalent to local convergence of the symbols encoding them via Sheffield's bijection, as claimed in Theorem 4.7.

2. **The reduced walk.** Consider the infinite decorated planar map (M, T) of Theorem 4.7, and let $(X_n)_{n\in\mathbb{Z}}$ denote the bi-infinite sequence of symbols encoding it via Sheffield's bijection. Let us assume that $q > 0$ or equivalently $p > 0$, where p and q are related via (4.8) and p is the proportion of F symbols. Define a *backward* exploration process $(c_n, h_n)_{n\geq 1}$ of the map, which keeps track of the number of C and H in the reduced word, as follows. Let $(c_0, h_0) = (0, 0)$. Suppose we have performed n steps of the exploration and defined c_n, h_n and in this process, we have revealed triangles corresponding to symbols $(X_{-m}, \ldots, X_0)$. We inductively define the following.

   - If X_{-m-1} is a C (resp. H), define $(c_{n+1}, h_{n+1}) = (c_n, h_n) + (1, 0)$ (resp. $(c_n, h_n) + (0, 1)$).
   - If X_{-m-1} a c (resp. h), $(c_{n+1}, h_{n+1}) = (c_n, h_n) + (-1, 0)$ (resp. $(c_n, h_n) + (0, -1)$).
   - If X_{-m-1} is F, then we explore $X_{-m-2}, X_{-m-3} \ldots$ until we find the match of X_{-m-1}. Let $|\mathcal{R}_{n+1}|$ denote the number of symbols in the reduced word $\mathcal{R}_{n+1} =$

$\overline{X_{\varphi(-m-1)} \dots X_{-m-1}}$. Show that $\mathcal{R}_{n+1}$ contains only symbols of one type. If $\mathcal{R}_{n+1}$ consists of H symbols, define $(c_{n+1}, h_{n+1}) = (c_n, h_n) + (0, |\mathcal{R}_{n+1}|)$. Otherwise, if $\mathcal{R}_{n+1}$ consists of C symbols define $(c_{n+1}, h_{n+1}) = (c_n, h_n) + (|\mathcal{R}_{n+1}|, 0)$.

Show that the walk $(c_n, h_n)_{n \geq 0}$ is a sum of *independent* and identically distributed random variables. Note that this is in contrast to Theorem 4.9. It can be shown that these random variables are in fact centered when $q \leq 4$ (see [She16b]).

3. **Bubbles.** Consider the infinite decorated planar map (M, T) of Theorem 4.7, and let $(X_n)_{n \in \mathbb{Z}}$ denote the bi-infinite sequence of symbols encoding it via Sheffield's bijection. Let us assume that $q > 0$ or equivalently $p > 0$, where p and q are related via (4.8) and p is the proportion of F symbols. Let us condition on the event $X_0 = \mathsf{F}$. Let $\varphi(0) \leq 0$ denote the match of this symbol. The word $w = X_{\varphi(0)} \dots X_0$ encodes a finite planar map, called the **bubble** or envelope of the map at 0. This bubble corresponds to a finite number of loops of (M, T) (note that this is in general more than a single loop of (M, T) containing the root triangle, as there can be other F symbols in w). This notion was pivotal in [BLR17] where it was used to derive critical exponents of Theorem 4.10. This exercise gives one of the main steps in the derivation of this theorem.

   (a) Assume without loss of generality that $X_{\varphi(0)} = \mathsf{h}$. Give a description of the reduced word $\bar{w}$. By considering the random length $N = |\varphi(0)|$ of w and the random length K of the reduced word $\bar{w}$, describe the event $\{N = n, K = k\}$ in terms of a certain cone excursion for the reverse two-dimensional walk $(C_{-i}, H_{-i})_{0 \leq i \leq n}$. Explain why N is the area of the bubble and K the length of its outer boundary.

   (b) Arguing at the same level of rigour as in Proposition 4.11, show that there are exponents p_{area} and p_{boundary} such that

$$\mathbb{P}(N \geq n) = n^{-p_{\text{area}} + o(1)}; \quad \mathbb{P}(K \geq k) = k^{-p_{\text{boundary}} + o(1)}$$

   where $p_{\text{boundary}}/2 = p_{\text{area}} = p_0$, and

$$p_0 = \pi \left(4 \arccos \left(\frac{\sqrt{2 - \sqrt{q}}}{2} \right) \right)^{-1} = \frac{\kappa'}{8} \in (1/2, 1),$$

   was defined in (4.10).

   The next three exercises use exponents derived in this chapter together with the KPZ formulas of the previous chapter to give predictions (in some cases proved through other methods) about the value of certain critical exponents associated with random fractals which can be defined without any reference to random planar maps.

4. Use (4.13), the KPZ relation, and the relation

$$q = 2 + 2\cos(8\pi/\kappa')$$

   between q and κ', to recover (non-rigourously) that the dimension of $\text{SLE}_{\kappa'}$ is $1 + \kappa'/8$ for $\kappa' \in (4, 8)$.

5. Consider simple random walk on the (infinite) uniform random planar map. If $n \geq 1$, a pioneer point for the walk $(X_1, \dots X_n)$ is a point x such that x is visited at some time $m \leq n$ and is on the boundary of the unbounded component of $G \setminus \{X_1, \dots, X_m\}$. A beautiful theorem of Benjamini and Curien [BC13] shows that when such a simple random walk first exits a ball of radius R, it has had $\approx R^3$ pioneer points.

   Analogously, for $(B_s)_{s \geq 0}$ a planar Brownian motion, we define the set $\mathcal{P}_t$ for given $t > 0$ to be all points of the form B_s for some $0 \leq s \leq t$, such that B_s is on the "frontier" at time s (where by frontier we mean the boundary of the unbounded component of the complement of $B[0, s]$).

   Using a (non-rigourous) KPZ-type argument, derive the dimension of the Brownian pioneer points $\mathcal{P}_t$ for any fixed $t \geq 0$. (The answer is $7/4$, as rigourously proved in a famous paper of Lawler, Schramm and Werner [LSW01] using SLE techniques).

6. Consider a simple random walk (X_n) on the infinite local limit of FK weighted planar maps (as in Theorem 4.7), starting from the root. Try to argue using the KPZ relation (again without being fully rigourous), that the graph distance between X_n and X_0) must be approximately equal to $n^{1/D}$ where D is the dimension of the space. (Hint: the range of Brownian motion must satisfy $\Delta = 0$). In particular, on the UIPT, one conjectures that this distance is $\approx n^{1/4}$. This has now been proven rigourously in [GH20] and [GM17a].

5 Scale-invariant random surfaces

Here we introduce some variants of the planar Gaussian free field which define random surfaces that are scale-invariant in a precise sense: zooming in near a given point of the surface produces the same surface in distribution.

We first need to set up the scene correctly. This will require working with a slightly different version of the GFF, with non-zero boundary conditions: the so-called **Neumann** or **free boundary** GFF. In general if we wish to add boundary data to a GFF it is natural to simply add a function which is harmonic in the domain (though it can have relatively wild behaviour on the boundary). We will seek to impose **Neumann boundary conditions**. Recall that for a smooth function this means that the normal derivative of the function vanishes along the boundary (if the domain is smooth). Of course for an object as rough as the GFF it is a priori unclear what this condition should mean. Indeed, we will see that the resulting object is actually the same as when we don't impose any conditions at all (which is why the field can also be called a free boundary GFF, as is the case for example in the papers [She16a] and [DMS14]). Indeed, note that in the discrete, a random walk on a graph with Neumann/"reflecting" boundary conditions or no/"free" boundary conditions are really one and the same thing (and both converge to reflecting Brownian motion, whose generator is (1/2) the Laplace operator with Neumann boundary conditions).

5.1 The Neumann boundary GFF as a random distribution

Remark 5.1. We will first show how to define the Neumann GFF as a random distribution on $D \subset \mathbb{C}$, just as in Section 1.6 for the Dirichlet GFF. This allows for a straightforward deduction of several nice properties, which is why we present this point of view first. However, we will *not* take this as our final definition. In Section 5.3 we will go on to show that the Neumann GFF can be defined as a stochastic process (as in the Dirichlet case), and that this object coincides with the random distribution defined here when its index set is restricted appropriately.

One technical complication when working with the Neumann GFF, compared to the Dirichlet case, is that it is really only defined up to a global additive constant. This corresponds to the fact that if one tries to extend the Dirichlet inner product $(\cdot,\cdot)_\nabla$ to test functions which are not necessarily compactly supported in D, it is no longer an inner product. Indeed, functions that are constant on the domain will have zero Dirichlet norm. Alternatively (as we will see later) one can think of the additive constant as arising from the fact that the Green function with Neumann boundary conditions is *not* canonically defined (or equivalently, that Brownian motion reflected on the boundary of D is recurrent).

The consequence is that the Neumann GFF can either be viewed as a **distribution modulo constants** (two distributions are equivalent if their difference is a constant function) or, equivalently, as a linear form over test functions whose integral is required to be zero. Sometimes, it will also be useful to specify a particular representative of the GFF's

equivalence class modulo constants (e.g. by requiring that the average of the field over a specific region is zero). We will then speak of fixing a normalisation. Note that while this point of view may appear to be more concrete, fixing the additive constant for the free field in this way actually causes it to lose some useful properties, such as conformal invariance. When using the Neumann GFF, we will therefore always need to be careful to say whether we consider the modulo constants version, or a version that has been normalised in a particular way.

In the rest of this section, D can be any proper, simply connected domain of the complex plane.

Let $\bar{\mathcal{D}}(D)$ be the space of C^∞ functions in $\bar{D}$ with $(f,f)_\nabla < \infty$ ("finite Dirichlet energy"), defined **modulo constants**. It is clear that on this space, $(\cdot,\cdot)_\nabla$ really is an inner product. Hence we can define $\bar{H}^1(D)$ to be the Hilbert space closure of $\bar{\mathcal{D}}(D)$ with respect to $(\cdot,\cdot)_\nabla$.

We define a distribution modulo constants to be a continuous linear functional on the space of test functions $f \in \mathcal{D}_0(D)$ such that $\int_D f(x)\,dx = 0$, and denote the set of such test functions by $\tilde{\mathcal{D}}_0(D)$. We equip the space $\bar{\mathcal{D}}_0'(D)$ of distributions modulo constants with the topology of weak-$\star$ convergence. That is, a sequence T_n of distributions mod constants converges to a distribution T iff $(T_n, f) \to (T, f)$ for any test function $f \in \tilde{\mathcal{D}}_0(D)$.

Summary of notation:

- $\mathcal{D}_0(D)$ is the space of smooth test functions that are compactly supported in D.
- $H_0^1(D)$ is the closure of $\mathcal{D}_0(D)$ with respect to the Dirichlet inner product.
- $\mathcal{D}_0'(D)$ is the space of distributions on D: i.e., linear continuous functionals on $\mathcal{D}_0(D)$.
- $\tilde{\mathcal{D}}_0(D) \subset \mathcal{D}_0(D)$ is the set of smooth compactly supported test functions with average value 0.
- $\bar{\mathcal{D}}_0'(D)$ is the space of distributions modulo constants on D: i.e., linear continuous functionals on $\tilde{\mathcal{D}}_0(D)$.
- $\bar{\mathcal{D}}(D)$ is the space of smooth functions in D with finite Dirichlet energy, considered modulo constants (note that they do not have to be compactly supported).
- $\bar{H}^1(D)$ is the closure of $\bar{\mathcal{D}}(D)$ with respect to the Dirichlet inner product.

Definition 5.2 (The Neumann GFF as a distribution modulo constants)**.** *Suppose that $\{\bar{f}_j\}_{j\geq 1}$ is an orthonormal basis of $\bar{H}^1(D)$ and $\{X_j\}_{j\geq 1}$ are independent $\mathcal{N}(0,1)$ random variables. Then the random series*

$$\bar{h}_n := \sum_1^n X_j \bar{f}_j \tag{5.1}$$

converges almost surely in the space of distributions modulo constants.

Moreover, the law of the limit does not depend on the choice of orthonormal basis $\{\bar{f}_j\}_j$, and can be written as the sum of a Dirichlet boundary condition GFF on D and an independent harmonic function modulo constants. This defines the Neumann GFF as a random distribution modulo constants.

Remark 5.3. (*Neumann boundary conditions*) Suppose that $D = \mathbb{D}$. In defining $\bar{H}^1(\mathbb{D})$ we started from the space $\bar{\mathcal{D}}(\mathbb{D})$ of smooth functions (modulo constants) on $\mathbb{D}$ with no restriction on their boundary conditions. However, we could equally have started with the space of smooth functions (modulo constants) with Neumann boundary conditions, and ended up with the same space $\bar{H}^1(\mathbb{D})$ after taking the closure with respect to $(\cdot,\cdot)_\nabla$. Indeed, there exists an orthonormal basis of $L^2(\mathbb{D})$ made up of eigenfunctions of the Laplacian with Neumann boundary conditions (see for example [Jos02, Theorem 8.5.2]). Then omitting the first eigenfunction (which has eigenvalue 0) and dividing the rest by the square roots of their respective eigenvalues and considering them modulo constants, provides an orthonormal basis of $\bar{H}^1(\mathbb{D})$.

Thus, one can think of the Neumann GFF as either having no imposed ("free") boundary conditions, or as having Neumann boundary conditions.

The connection with Neumann boundary conditions will also become more apparent when we define the Neumann GFF as a stochastic process. Indeed, we will see that its covariance function is given by a Green function in the domain, with Neumann instead of Dirichlet boundary conditions. As already mentioned, note that in the discrete, a random walk on a graph with Neumann/"reflecting" boundary conditions or no/"free" boundary conditions are really one and the same thing. So the discrete Green's function will be the same if either free or Neumann boundary conditions are imposed.

Proof. We will carry out the proof of the claim in Definition 5.2 in two steps: first assuming that D is the unit disc $\mathbb{D}$; and then extending to general D by conformal invariance.

Step 1 ($D = \mathbb{D}$)**.** Write $\overline{\mathrm{Harm}}(\mathbb{D})$ for the space of harmonic functions on $\mathbb{D}$ with finite Dirichlet energy, viewed modulo constants. By the same reasoning as in Lemma 1.31, we can decompose

$$\bar{H}^1(\mathbb{D}) = H_0^1(\mathbb{D}) \oplus \overline{\mathrm{Harm}}(\mathbb{D})$$

as a direct orthogonal sum with respect to the Dirichlet inner product.

Given Theorem 1.22, Step 1 is therefore reduced to showing that if $\{\bar{e}_j\}_j$ is an orthonormal basis of $\overline{\mathrm{Harm}}(\mathbb{D})$ with respect to the Dirichlet inner product, and $\{\alpha_j\}_j$ are i.i.d. $\mathcal{N}(0,1)$, then:

(*) $\sum \alpha_j \bar{e}_j$ converges almost surely in the space of distributions modulo constants; and

(**) the limit has unique law, which is that of a harmonic function modulo constants.

For $f \in \overline{\mathrm{Harm}}(\mathbb{D})$, let us denote by Af the (modulo constants) equivalence class representative of f that satisfies $Af(0) = 0$. We claim that for any orthonormal basis $\{\bar{e}_j\}_j$ of $\overline{\mathrm{Harm}}(\mathbb{D})$,

$$\sum_{j=1}^{\infty} \|A\bar{e}_j\|^2_{L^2(\mathbb{D})} < \infty. \tag{5.2}$$

This implies that for any sequence $\{\alpha_j\}_j$ of i.i.d. $\mathcal{N}(0,1)$ variables,

$$\mathbb{E}\left(\| \sum_{j=m}^{n} \alpha_j A\bar{e}_j \|^2_{L^2(\mathbb{D})} \right) \to 0$$

as $m, n \to \infty$. Hence the series $\sum \alpha_j A\bar{e}_j$ converges almost surely in $L^2(\mathbb{D})$, and therefore also in the space of distributions. This in turn implies that $\sum_j \alpha_j \bar{e}_j$ converges almost surely in the space of distributions modulo constants.

To prove (5.2), let us first consider a specific choice of orthonormal basis. To define this basis, we observe that because any harmonic function on $\mathbb{D}$ is the real part of an analytic function, it admits a Taylor series expansion of the form $\sum_{j=1}^{\infty} b_j \Re(z^j) + \sum_{j=1}^{\infty} c_j \Re(iz^j)$ with $\{b_j, c_j\}_j \in \mathbb{R}$. One can also check that the functions $\{\Re(z^j)\}_{j\geq 0}$ and $\{\Re(iz^j)\}_{j\geq 1}$ are orthogonal with respect to the Dirichlet inner product, and

$$\|\Re(z^j)\|^2_{\nabla} = \|\Re(iz^j)\|^2_{\nabla} = \frac{j^2}{2\pi} \int_{\mathbb{D}} |z|^{2j-2}\, dz = j/2.$$

Hence, the functions $\sqrt{2/j}\,\Re(z^j)$ for $j \geq 0$ and $\sqrt{2/j}\,\Re(iz^j)$ for $j \geq 1$ (viewed modulo constants) define an orthonormal basis of $\overline{\mathrm{Harm}}(\mathbb{D})$.

With this choice of basis (by definition of the operator A), the corresponding series in (5.2) is equal to

$$\sum_{j=1}^{\infty} \frac{2}{j} \int_{\mathbb{D}} |\Re(z^j)|^2 + |\Re(iz^j)|^2\, dz = \sum_{j=1}^{\infty} \frac{4\pi}{j} \int_0^1 u^{2j+1} du = \sum_{j=1}^{\infty} \frac{2\pi}{j(j+1)} < \infty.$$

This concludes the proof of (*) for this basis.

To prove (*) for an arbitrary basis, we have to apply a slightly abstract argument, as we do not know how to calculate or estimate the size of the terms in (5.2) in general. Note that we used a similar argument in the proof of Theorem 1.24, but it was slightly more explicit since we did not have to worry about spaces of functions modulo constants.

The argument proceeds as follows. We first notice that by the Poincaré inequality (see for example [Eva10, §5.8.1])

$$A : (\overline{\mathrm{Harm}}(\mathbb{D}), (\cdot,\cdot)_{\nabla}) \to (L^2(\mathbb{D}), (\cdot,\cdot)_{L^2(\mathbb{D})})$$

defines a bounded linear operator between Hilbert spaces. This implies that there exists a unique adjoint operator A^* from $L^2(\mathbb{D})$ to $\overline{\mathrm{Harm}}(\mathbb{D})$, such that $(Af, g)_{L^2(\mathbb{D})} = (f, A^*g)_{\nabla}$ for any $f \in \overline{\mathrm{Harm}}(\mathbb{D})$ and $g \in L^2(\mathbb{D})$. Choosing an arbitrary orthonormal basis $\{g_k\}_k$ for $L^2(\mathbb{D})$, we then have

$$\sum_{j=1}^{\infty} \|A\bar{e}_j\|^2_{L^2(\mathbb{D})} = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} (A\bar{e}_j, g_k)_{L^2(\mathbb{D})} = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} (\bar{e}_j, A^*g_k)_{\nabla} = \sum_{k=1}^{\infty} \|A^*g_k\|^2_{\nabla}. \tag{5.3}$$

Since the right hand side does not depend on the choice of the orthonormal basis $\{\bar{e}_j\}_j$ for $\overline{\mathrm{Harm}}(\mathbb{D})$, we deduce that the left hand side (and so the sum in (5.2)) *does not depend on this*

choice either. Thus, the fact that we have proved (5.2) for a specific choice of orthonormal basis means that it actually holds for *any* orthonormal basis.[4]

For (**), we can use a very similar argument to (5.3) to see that the limit in $L^2(\mathbb{D})$ of $\sum_j \alpha_j A\bar{e}_j$ has to be independent of the choice of $\{\bar{e}_j\}_j$. Indeed, for any $g \in L^2(\mathbb{D})$ the same expansion gives that the L^2 product of g with this limit must be normally distributed, with mean 0 and variance $\|A^*g\|_\nabla^2$. Hence, we need only show that the almost sure limit of

$$\sum_1^n \sqrt{\frac{2}{j}} \Re((\alpha_j + i\beta_j)z^j) \; ; \; \alpha_j, \beta_j \overset{\text{i.i.d}}{\sim} \mathcal{N}(0,1) \tag{5.4}$$

is harmonic. This follows from the fact that $\sum_j \sqrt{2/j}(\alpha_j + i\beta_j)z^j$ almost surely defines a Taylor series with radius of convergence one (since, for example, $\limsup_n (\log n)^{-1} \max_{1\leq j\leq n} |\alpha_j + i\beta_j| \leq 1$ almost surely). Thus for any $r < 1$, the limit of (5.4) is the real part of an analytic function in $B(0,r)$, and so the limit is harmonic in $\cup_{r\in(0,1)} B(0,r) = \mathbb{D}$.

Step 2 (general D). Suppose that $D \subsetneq \mathbb{C}$ is simply connected, and let $\{\bar{f}_j\}_j$ be an orthonormal basis for $\bar{H}^1(D)$. We would like to show that $\bar{h}_n = \sum_1^n X_j f_j$ converges almost surely in $\bar{\mathcal{D}}_0'(D)$ (when the X_j are i.i.d. $\mathcal{N}(0,1)$).

For this, we are going to use Step 1 and conformal invariance. Let $T : \mathbb{D} \to D$ be a conformal map (which exists by the Riemann mapping theorem). Then by conformal invariance of the Dirichlet inner product, $\{\bar{f}_j \circ T\}_j$ forms an orthonormal basis of $\bar{H}^1(\mathbb{D})$. We therefore know, by Step 1, that $\bar{h}_n \circ T := \sum_1^n X_j(\bar{f}_j \circ T)$ converges almost surely in $\bar{\mathcal{D}}_0'(\mathbb{D})$. That is, with probability one, there exists $\bar{h}^{\mathbb{D}} \in \bar{\mathcal{D}}_0'(\mathbb{D})$ such that $(\bar{h}_n \circ T, g) \to (\bar{h}^{\mathbb{D}}, g)$ as $n \to \infty$ for all $g \in \tilde{\mathcal{D}}_0(\mathbb{D})$.

Since for $f \in \tilde{\mathcal{D}}_0(D)$ the function $g(z) = |T'(z)|^2(f \circ T)(z)$ is in $\tilde{\mathcal{D}}_0(\mathbb{D})$, this tells us - in particular - that with probability one:

$$(\bar{h}_n \circ T, |T'|^2(f \circ T)) \to (\bar{h}^{\mathbb{D}}, |T'|^2(f \circ T)) \text{ as } n \to \infty, \; \forall f \in \tilde{\mathcal{D}}_0(D).$$

Defining $\bar{h} \in \bar{\mathcal{D}}_0'(D)$ by $(\bar{h}, f) = (\bar{h}^{\mathbb{D}}, |T'|^2(f \circ T))$ for all $f \in \bar{\mathcal{D}}_0(D)$, this is exactly saying that with probability one, $(\bar{h}_n, f) \to (\bar{h}, f)$ as $n \to \infty$ for all $f \in \tilde{\mathcal{D}}_0(D)$. That is, $\bar{h}_n \to \bar{h}$ in $\bar{\mathcal{D}}_0'(D)$, almost surely as $n \to \infty$.

Finally, by the same argument, if $T : \mathbb{D} \to D$ is conformal then the law of $\bar{h}$ must be given by the law of $\bar{h}^{\mathbb{D}} \circ T^{-1}$, where $\bar{h}^{\mathbb{D}}$ is the (unique in law) limit of (5.2) when $D = \mathbb{D}$. Note that this does not depend on the choice of T, since the law of $\bar{h}^{\mathbb{D}}$ is conformally invariant (we can see this by applying the reasoning of the previous sentence with $D = \mathbb{D}$, together with the uniqueness in Step 1). Thus, the law of $\bar{h}$ is unique for general D.

Using the description of this law when $D = \mathbb{D}$ from Step 1, plus conformal invariance of the Dirichlet GFF (Theorem 1.33) and the fact that conformal maps preserve harmonicity, we see that in general the law of $\bar{h}$ satisfies the description in Definition 5.2. □

By conformal invariance of the Dirichlet inner product, we obtain the following (the details are spelled out in the proof above):

[4]This argument shows that A is a *Hilbert-Schmidt* operator and the expression in (5.2) is its Hilbert-Schmidt norm. This norm is always independent of the choice of orthonormal basis.

Corollary 5.4. *Let $\bar{h}^D$ be the Neumann GFF (viewed modulo constants) in D, as in Definition 5.2. Then the law of $\bar{h}^D$ is conformally invariant. That is, if $T : D \to D'$ is a conformal map between simply connected domains, then*

$$\bar{h}^{D'} \overset{(d)}{=} \bar{h}^D \circ T^{-1}$$

where $(\bar{h}^D \circ T^{-1}, f) := (\bar{h}^D, |T'|^2(f \circ T))$ for all $f \in \tilde{\mathcal{D}}_0(D')$.

Straight from the definition, we also know that if $\bar{h}$ is the Neumann GFF (viewed as a distribution modulo constants) in D, then $\bar{h}$ can be written as the sum $h + u$, where h has the law of a zero boundary GFF in D, and u is an independent harmonic function modulo constants in D. By applying the Markov property of the Dirichlet GFF (Theorem 1.30) to $\bar{h}$ we get an analogous decomposition for the Neumann GFF:

Theorem 5.5 (Markov property)**.** *Fix $U \subset D$, open. Let $\bar{h}$ be a Neumann GFF viewed as a distribution modulo constants in D, as in Definition 5.2. Then we may write*

$$\bar{h} = h_0 + \varphi$$

where:

1. *h_0 is a zero boundary condition GFF in U, and is zero outside of U;*
2. *φ is a harmonic function viewed modulo constants in U;*
3. *h_0 and φ are independent.*

Recall that we defined a distribution modulo constants to be a continuous linear functional on the space $\tilde{\mathcal{D}}_0(D)$ of test functions with average 0. Equivalently, we could define it to be an equivalence class of distributions (elements of $\mathcal{D}'_0(D)$), under the equivalence relation identifying distributions ϕ_1 and ϕ_2 whenever $\phi_1 - \phi_2 \equiv C$ for $C \in \mathbb{R}$.

Remark 5.6 (Fixing the additive constant)**.** With the latter perspective, it is quite natural (and will sometimes be useful) to fix the additive constant for the GFF in some way (i.e., to pick an equivalence class representative). For example, we could require it to have average zero when tested against some fixed test function.

With the additive constant fixed, the Neumann GFF defines an element of $\mathcal{D}'_0(D)$, i.e., a distribution on D.

Note that the choice of constant/equivalence class representative changes the resulting element of $\mathcal{D}'_0(D)$, but not how it acts when tested against functions (with average 0) in $\tilde{\mathcal{D}}_0(D)$.

Remark 5.7. Although it is sometimes helpful to specify a normalisation of the Neumann GFF, one should take care with the conformal invariance and Markovian properties discussed above. In particular:

- if $\bar{h}$ is a Neumann GFF in D with additive constant fixed in some way, then it is *no longer* conformally invariant;
- one can still write $\bar{h} = h + u$ with h a Dirichlet GFF in D and u a harmonic function, but h and u *need not be independent*;
- on the other hand, if one starts with a Neumann GFF modulo constants, decomposes it as a Dirichlet GFF plus a harmonic function modulo constants, and then fixes the constant for the GFF in a way that only depends on the harmonic function (e.g., by specifying the value of the harmonic function at a point), then the two summands *will be* independent.

5.2 Covariance formula: the Neumann Green function

Recalling the definition of the Dirichlet GFF in a domain D, it is quite natural to guess that the Neumann GFF will have "covariance" given by a version of the Green function with Neumann boundary conditions in D. This is indeed the case, but with the caveat that the Green function with Neumann boundary conditions is not uniquely defined.

First, what exactly do we mean by covariance? We simply mean that we would like to specify a function G such that

$$\mathbb{E}((\bar{h}, \rho_1)(\bar{h}, \rho_2)) = \int_{D\times D} \rho_1(x) G(x,y) \rho_2(y)\, dxdy \tag{5.5}$$

when $\bar{h}$ is a Neumann GFF (viewed as a distribution modulo constants in D), and $\rho_1, \rho_2 \in \tilde{\mathcal{D}}_0(D)$.

Definition 5.8. *We call any G satisfying (5.5) a* valid choice of covariance *for the Neumann GFF in D.*

Observe that this need not uniquely define G, because the equality is only required to hold for test functions with average 0. This ill-definition is also an inherent property of the Neumann Green function (see below). Thus, if one instead tries to define the Neumann GFF as a stochastic process with covariance given by the Neumann Green function, it again only makes sense "modulo constants".

To define the Neumann Green function, we need to introduce the **Neumann problem** in D. This is the problem, given $\{\psi, v\}$, of

$$\text{finding } f \text{ such that: } \begin{cases} \Delta f = \psi & \text{in } D \\ \dfrac{\partial f}{\partial n} = v & \text{on } \partial D. \end{cases} \tag{5.6}$$

A requirement for the existence of a (weak) solution is that ψ, v satisfy the Stokes condition:

$$\int_D \psi = \int_{\partial D} v. \tag{5.7}$$

This condition comes from the divergence theorem; the integral of v along the boundary measures the total flux across the boundary, while the integral of the Laplacian of f inside the domain measures the sum of the flow at every point inside. This solution is then unique, up to a global additive constant. That is, this solution is unique in $\bar{H}^1(D)$. Existence of a solution is also known, for example, when D is smooth and bounded and $v = 0, \psi \in L^2(D)$, [Eva10, §6] (but we will not use any of these facts).

Definition 5.9 (Neumann Green function)**.** *We say that G is a (choice of) Neumann Green function in D, if for every $\rho \in \tilde{\mathcal{D}}_0(D)$:*

$$f(x) = \int_D G(x,y)\rho(y)dy \tag{5.8}$$

is a solution of the Neumann problem in D, with $\psi = -2\pi\rho$ and $v = 0$.

Proposition 5.10. *Suppose that $D \subset \mathbb{C}$ is simply connected and has C^1 smooth boundary. Then if G is a choice of Neumann Green function, it is a valid choice of covariance for the Neumann GFF $\bar{h}$ in D. That is for every $\rho \in \tilde{\mathcal{D}}_0(D)$*

$$\mathbb{E}((\bar{h},\rho)^2) = \int_{D\times D} \rho(x)G(x,y)\rho(y)\,dxdy. \tag{5.9}$$

Remark 5.11. Note that adding a constant to G will not affect whether f defined in (5.8) is a solution to the Neumann problem. This makes the lack of uniqueness for G apparent. However, two valid Neumann Green functions may also differ by a non-constant function (see examples 5.12 and 5.13).

Proof of Proposition 5.10. We need to check that if $\rho \in \tilde{\mathcal{D}}_0(D)$ and $\bar{h}$ is a Neumann GFF in D, then

$$\mathbb{E}((\bar{h},\rho)^2) = \int_{D\times D} \rho(x)G(x,y)\rho(y)\,dxdy. \tag{5.10}$$

Defining $f(x) = \int_D G(x,y)\rho(y)dy$, we will show that both sides are equal to $\|f\|_\nabla^2$.

Note that by assumption the right-hand side of (5.10) is equal to

$$\int_D -\frac{\Delta f}{2\pi}(x)f(x)\,dx,$$

which by applying the Gauss-Green formula and the Neumann boundary condition for f is equal to

$$\frac{1}{2\pi}\int_D \nabla f(x)\cdot\nabla f(x)\,dx = \|f\|_\nabla^2.$$

For the left-hand side we use the construction of $\bar{h}$ as the limit as $n\to\infty$ of $\sum_1^n X_j\bar{f}_j$ where the X_js are i.i.d. $\mathcal{N}(0,1)$ and the $\bar{f}_j$s are an orthonormal basis of $\bar{H}^1(D)$. Since this is an almost sure limit in the space of distributions modulo constants, we have that

$$(\bar{h},\rho) = \lim_{n\to\infty}\sum_{j=1}^n X_j(\bar{f}_j,\rho) \text{ a.s.}$$

Furthermore, by the Gauss-Green formula again, we have that $(\bar{f}_j, \rho) = (\bar{f}_j, f)_\nabla$ for each j, and so

$$\mathbb{E}((\sum_{j=1}^{n} X_j(\bar{f}_j, \rho))^2) = \sum_{j=1}^{n} (\bar{f}_j, f)_\nabla^2.$$

Note that this is bounded above by $\|f\|_\nabla^2$ for every n. Hence, $\sum_{j=1}^{n} X_j(\bar{f}_j, \rho)$ defines a martingale that is bounded in L^2, and so

$$\mathbb{E}((\bar{h}, \rho)^2) = \mathbb{E}(\lim_{n\to\infty} (\sum_{j=1}^{n} X_j(\bar{f}_j, \rho))^2) = \lim_{n\to\infty} \mathbb{E}((\sum_{j=1}^{n} X_j(\bar{f}_j, \rho))^2) = \|f\|_\nabla^2,$$

as desired. □

Example 5.12. *We can define a choice of Neumann Green function in the unit disc $\mathbb{D}$ by*

$$G^{\mathbb{D}}(x, y) = -\log|(x - y)(1 - x\bar{y})|; \quad x \neq y \in \mathbb{D}.$$

Indeed, a tedious but straightforward calculation can be used to verify that if $g_y(x) := G^{\mathbb{D}}(x, y)$ for fixed $y \in \mathbb{D}$, then

$$\begin{cases} \Delta g_y & = -2\pi\delta_y \\ \frac{\partial g_y}{\partial n} & = -1 \text{ on } \partial\mathbb{D}. \end{cases}$$

This implies that if $\rho \in \tilde{D}_0(\mathbb{D})$ then (5.8) is a solution of the Neumann problem with $\psi = -2\pi\rho$ and $v = 0$. Indeed,

- $\Delta f(x) = \int_{\mathbb{D}} \Delta g_y(x)\rho(y)\, dy = -2\pi \int_{\mathbb{D}} \delta_y(x)\rho(y)\, dy = -2\pi \int_{\mathbb{D}} \delta_x(y)\rho(y) dy = -2\pi\rho(x)$;
- *and for $x \in \partial\mathbb{D}$,* $(\partial f/\partial n)(x) = \int_{\mathbb{D}} (\partial g_y/\partial n)(x)\rho(y)\, dy = -\int_{\mathbb{D}} \rho(y)\, dy = 0$.

Hence $G^{\mathbb{D}}$ is a choice of Neumann Green function in $\mathbb{D}$, and so also a valid choice of covariance for the Neumann GFF in $\mathbb{D}$.

Example 5.13. *Define*

$$G^{\mathbb{H}}(x, y) = -\log|(x - y)| - \log|(x - \bar{y})|; \quad x \neq y \in \mathbb{H}. \tag{5.11}$$

In this case, defining the conformal map $T : \mathbb{H} \to \mathbb{D}$ by $T(z) = (i - z)(i + z)^{-1}$, we have that if $g_y(x) := G^{\mathbb{H}}(T^{-1}(x), T^{-1}(y))$, then $\Delta g_y = -2\pi\delta_y$ and $\partial g_y/\partial n = -2\pi\delta_{-1}$ on $\partial\mathbb{D}$. Similarly to in the previous example, this implies that $G^{\mathbb{H}}(T^{-1}(\cdot), T^{-1}(\cdot))$ is a valid choice of Neumann Green function on $\mathbb{D}$. Hence by Proposition 5.10, it defines a valid choice of covariance function for the Neumann GFF on $\mathbb{D}$.

Finally, by conformal invariance of the Neumann GFF (Corollary 5.4), we see that $G^{\mathbb{H}}$ is a valid choice of covariance function for the Neumann GFF in $\mathbb{H}$.

It may seem that we have taken a rather long-winded approach in example 5.13. Indeed, one can easily verify that if $g_y^{\mathbb{H}}(x) = G^{\mathbb{H}}(x,y)$ then $\Delta g_y^{\mathbb{H}} = -2\pi\delta_y$ on $\mathbb{H}$ and $\partial g_y^{\mathbb{H}}/\partial n = 0$ on $\mathbb{R}$. It is tempting to say that $G^{\mathbb{H}}$ therefore defines a choice of Green function on $\mathbb{H}$ and so by Proposition 5.10, a valid covariance for the Neumann GFF on $\mathbb{H}$. However, one needs to take care that there is an extra "point at ∞" on the boundary of $\mathbb{H}$ (where, as you can see from the calculations in example 5.13, we actually have a Dirac mass for $\partial g_y^{\mathbb{H}}/\partial n$). To make this example rigorous it is therefore necessary to map to the unit disc and appeal to conformal invariance - as carried out above.

Remark 5.14. Recall that the Green's function G_0^D for a GFF with zero boundary conditions on a domain D could be defined in terms of the expected occupation time of Brownian motion killed when leaving D:

$$G_0^D(x,y) = \pi \int_0^\infty p_t^D(x,y)\, dt \quad (x \neq y).$$

There is a similar relationship between the Neumann Green's function and Brownian motion reflected on the boundary of D. The fact that the Neumann Green's function is not uniquely defined is related to the fact that reflected Brownian motion is recurrent. This means if $\tilde{p}_t^D(x,y)$ is the transition density for this reflected Brownian motion, then $\int_0^\infty \tilde{p}_t^D(x,y)\, dt$ does not actually converge, so one needs to normalize in some way to obtain a finite quantity. There are many possible ways to do this - hence the non-uniqueness.

Let us describe this more precisely in the case where $D = \mathbb{H}$. Denoting by $p_t(x,y)$ the transition density of Brownian motion in $\mathbb{C}$, it is easy to see that $\tilde{p}_t^{\mathbb{H}}(x,y) = p_t(x,y) + p_t(x,\bar{y}) = (2\pi t)^{-1}\left(\exp(-|x-y|^2/2t) + \exp(-|x-y|^2/2t)\right)$, which does not have finite integral over $t \in [0,\infty)$. However, if we look at $\tilde{p}_t^{\mathbb{H}}(x,y) - \tilde{p}_t^{\mathbb{H}}(x_0,y)$ for some fixed $x_0 \in \mathbb{H}$ (for instance) then the corresponding integral does converge: to $G^{\mathbb{H}}(x,y)$ as defined in (5.11) plus the function $2\log|x_0 - y|$. It is straightforward to check that this integral, for any choice of x_0, does define a valid choice of Neumann Green function on $\mathbb{H}$.

Definition 5.15 (A choice of covariance for general D)**.** *By Corollary 5.4, we see that if G^D is a valid choice of covariance function for the Neumann GFF on some domain D, and $T : D' \to D$ is conformal, then $G^D(T(\cdot), T(\cdot))$ is a valid choice of covariance function for the Neumann GFF on D'.*

From this observation and the above examples, we obtain a recipe to define a valid covariance function for the Neumann GFF in any simply connected domain D. We emphasise that any valid choice gives the same value for $\mathbb{E}((\bar{h},\rho_1)(\bar{h},\rho_2))$ when $\bar{h}$ is a Neumann GFF (viewed as a distribution modulo constants) in D and $\rho_1, \rho_2 \in \tilde{\mathcal{D}}_0(D)$.

5.3 Neumann GFF as a stochastic process

In this section, we will define the Neumann GFF as a stochastic process, similarly to the definition of the Dirichlet GFF in Section 1.3. This will allow us to "test" the Neumann GFF against a wider range of functions (i.e. they need not be smooth). In particular, it allows

us to define circle and semi-circle averages of the field. We will see that the definition here agrees with the one in the previous sections, but is more general. Therefore, in everything that follows, we will assume that the Neumann GFF is as defined below.

Let D be a proper simply connected subdomain of $\mathbb{C}$. We define $\tilde{\mathcal{M}}^D = \tilde{\mathcal{M}}$ to be the space of signed Radon measures ρ on D, with finite and *equal* positive and negative mass (so $\int_D \rho(dx) = 0$), such that

$$\bar{f} \mapsto \int_D \bar{f}(x)\rho(dx)$$

defines a continuous linear functional on $\bar{H}^1(D)$. This is equivalent to the property that for any valid choice of covariance function G for the Neumann GFF on D (as in Proposition 5.10), ρ is an element of the Hilbert space completion of $\tilde{\mathcal{D}}_0(D)$ with respect to the inner product

$$(\rho, \rho)_G := \int_{D\times D} \rho(dx)G(x,y)\rho(dy)$$

(notice that any valid covariance G defines the *same* inner product on $\tilde{\mathcal{D}}_0(D)$). This follows from the fact that for any $\rho \in \tilde{D}_0(D)$, the norm of ρ as an element of the dual space of $\bar{H}^1(D)$ is equal to its norm with respect to $(\cdot,\cdot)_G$ (see Exercise 2 of this Chapter).

In particular, $(\cdot,\cdot)_G$ defines a positive definite bilinear form on $\tilde{\mathcal{M}}$: if $(\rho,\rho)_G = 0$ then for $\bar{h}$ a Neumann GFF as in (5.1) we must have $(\bar{h},\rho) = 0$ a.s. and so $(\bar{f}_j,\rho) = 0$ for all j. This implies that $\rho \equiv 0$.[5]

Remark 5.16. The reader should compare this with the definitions in Chapter 1. Recall that index set $\mathcal{M}_0$ for the Gaussian free field with zero boundary conditions (viewed as a stochastic process) is precisely the set of measures with finite Green energy. Equivalently, by Lemma 1.21, it is the set of measures lying in H_0^{-1}; i.e., the dual space or space of continuous linear functionals on H_0^1.

For later use, we further define

$$\mathcal{M} := \{\rho : \rho = \tilde{\rho} + f \text{ with } \tilde{\rho} \in \tilde{\mathcal{M}} \text{ and } f \in \mathcal{D}_0(D)\}.$$

This will play the role of the index set when we fix the additive constant for the Neumann GFF in some way (and hence may test it against things that do not have average 0).

Remark 5.17. If ρ is a signed Radon measure with finite negative and positive mass, that also:

- has support contained in some $D' \subseteq D$ which is bounded with smooth boundary; and
- integrates a log in the sense that $\int \log(|x-y| \wedge 1)|\rho|(dx)|\rho|(dy) < \infty$,

[5] In fact for $\rho \in \tilde{\mathcal{M}}$, $(\rho,\rho)_G$ is equal to the dual Hilbert space norm of ρ in the space of continuous linear functionals on $\bar{H}^1(D)$.

then $\rho \in \mathcal{M}$. To see this, by subtracting a function in $\mathcal{D}_0(D)$ and rescaling if necessary, we may assume that ρ is a probability measure (i.e., positive with mass one) and show that if $(\rho_\varepsilon)_{\varepsilon \geq 0} \in \tilde{\mathcal{D}}_0(D)$ are defined by convolving ρ with a smooth mollifier at scale ε and multiplying by a smooth function that is supported at distance ε away from the boundary of D', then $(\rho_\varepsilon - \rho, \rho_\varepsilon - \rho)_G \to 0$ as $\varepsilon \to 0$.

Notice that we may rewrite

$$\begin{aligned}(\rho_\varepsilon - \rho, \rho_\varepsilon - \rho)_G &= \mathbb{E}(G(X + U_X^\varepsilon, Y + U_Y^\varepsilon) - G(X + U_X^\varepsilon, Y) - G(X, Y + U_Y^\varepsilon) - G(X, Y)) \\ &=: \mathbb{E}(G_\varepsilon(X, Y))\end{aligned}$$

where: (X, U_X^ε) is independent of (Y, U_Y^ε); X, Y are distributed according to ρ; and $U_X^\varepsilon, U_Y^\varepsilon$ are random variables taking values in $B(0, \varepsilon)$ a.s. Then $G_\varepsilon(X, Y)$ converges to 0 as $\varepsilon \to 0$ almost surely, by continuity of the Greens function away from the diagonal and the fact that ρ cannot have any atoms. Moreover, $G_\varepsilon(X, Y)$ is almost surely bounded above by $C \log(|X - Y|^{-1})$ for all ε and some C deterministic depending only on D': this follows from the fact that $G(x, y)$ is bounded by a multiple of $\log|x - y|$ in D' (by Definition 5.15) and then by the same argument as in Lemma 3.6. The claim finally follows by dominated convergence and the assumption on ρ.

Definition 5.18 (Neumann GFF modulo constants as a stochastic process)**.** *There exists a unique stochastic process $(\bar{h}_\rho)_{\rho \in \tilde{\mathcal{M}}}$, indexed by $\tilde{\mathcal{M}}$, such that for every choice of $\rho_1, \cdots, \rho_n \in \tilde{\mathcal{M}}$, $(\bar{h}_{\rho_1}, \cdots, \bar{h}_{\rho_n})$ is a centered Gaussian vector with covariance $\mathrm{Cov}(\bar{h}_{\rho_i}, \bar{h}_{\rho_j}) = (\rho_i, \rho_j)_G$. We define this process to be the Neumann GFF in D, and denote*

$$\bar{h}_\rho := (\bar{h}, \rho)$$

for $\rho \in \tilde{\mathcal{M}}$.

Definition 5.18 agrees with the definition of the Neumann GFF as a random distribution modulo constants, in the following sense.

Lemma 5.19. *Let $\bar{h}_n = \sum_{j=1}^n X_j \bar{f}_j$ be as in (5.1). Then for any $\rho \in \tilde{\mathcal{M}}$, $(\bar{h}_n, \rho)$ has a limit in $L^2(\mathbb{P})$ and hence in probability as $n \to \infty$. This limit has the law of $(\bar{h}, \rho)$, where $\bar{h}$ is as in Definition 5.18.*

In particular:

Corollary 5.20. *Let $\bar{h}$ be as in Definition 5.18, and consider its restriction to a stochastic process*

$$(\bar{h}_\rho)_{\rho \in \tilde{\mathcal{D}}_0(D)} = (\bar{h}, \rho)_{\rho \in \tilde{\mathcal{D}}_0(D)}.$$

Then there exists a version of this stochastic process with the law of "the Neumann GFF as a random distribution modulo constants" from Definition 5.2.

Proof of Lemma 5.19. The proof follows very closely that of its analogue, Theorem 1.22, for the zero boundary condition GFF, but we include it for completeness. First, we may assume by conformal invariance that $\mathbb{D}$ is the unit disk.

As in the proof of Theorem 1.22 we begin by supposing that $\nu \in \tilde{\mathcal{D}}_0(\mathbb{D})$ and observe that

$$(\bar{h}_n, \nu) \to Z \overset{(d)}{=} (\bar{h}, \nu) \tag{5.12}$$

in $L^2(\mathbb{P})$ as $n \to \infty$. Indeed, this is exactly what is shown in the proof of Proposition 5.10.

In particular, the bound $\mathrm{Var}(\bar{h}_n, \nu) \leq \mathrm{Var}(\bar{h}, \nu)$ holds for every n. Notice that this bound can be extended to arbitrary $\nu \in \tilde{\mathcal{M}}$, by the same density argument as in the proof of Theorem 1.22. That is, using the fact that $\nu \in \tilde{\mathcal{M}}$ can be approximated by a sequence $\nu_\varepsilon \in \tilde{\mathcal{D}}_0(\mathbb{D})$ as $\varepsilon \to 0$, with respect to the norm $(\cdot, \cdot)_G$.

Using these ingredients, we can complete the proof. Let $\rho \in \tilde{\mathcal{M}}$ and let ρ_ε be an approximation of ρ with respect to the norm $(\cdot, \cdot)_G$. Then we have

$$(\bar{h}_n, \rho) = (\bar{h}_n, \rho_\varepsilon) + (\bar{h}_n, \rho - \rho_\varepsilon),$$

where the variance bound from the paragraph above implies that the second term converges to 0 in $L^2(\mathbb{P})$ as $\varepsilon \to 0$, *uniformly* in n. This allows us to conclude, since the first term converges to a random variable with the law of $(\bar{h}, \rho)$ in $L^2(\mathbb{P})$ as first $n \to \infty$ and then $\varepsilon \to 0$ (by (5.12) and the fact that $(\bar{h}, \rho_\varepsilon) \to (\bar{h}, \rho)$ in $L^2(\mathbb{P})$). □

We can now unambiguously take Definition 5.18 as our definition of the Neumann GFF modulo constants.

Remark 5.21 (Conformal invariance)**.** By Definition 5.15, the inner product $(\cdot, \cdot)_G$ is conformally invariant on $\tilde{\mathcal{M}}$. That is, if $T : D \to D'$ is conformal and $\rho \in \tilde{\mathcal{M}}^D$, then the push-forward $\rho \circ T^{-1}$ of ρ by T is in $\tilde{\mathcal{M}}^{D'}$, and $(\rho, \rho)_G = (\rho \circ T^{-1}, \rho \circ T^{-1})_G$.

This immediately implies that the Neumann GFF modulo constants is conformally invariant, in the sense that if $T : D \to D'$ is conformal and $\bar{h}^D$ is a Neumann GFF on D, then the stochastic process $(h^D, \rho \circ T)_{\rho \in \tilde{\mathcal{M}}^{D'}}$ has the law of a Neumann GFF on D'.

Remark 5.22 (Markov property)**.** Suppose that $U \subset D$ is an open connected set, and $\bar{h}^D$ is a Neumann GFF on D. Then

$$(\bar{h}, \rho)_{\rho \in \tilde{\mathcal{D}}_0(U)} = (h_0 + \varphi, \rho)_{\rho \in \tilde{\mathcal{D}}_0(U)}$$

where h_0 is a zero boundary GFF on U, φ and h_0 are independent, and φ defines a stochastic process indexed by $\tilde{\mathcal{D}}_0(U)$, that corresponds to integration against an almost surely harmonic function modulo constants.

With our new more general definition of the Neumann GFF, it still makes sense to then speak of *fixing a normalisation* for the field. This results in a stochastic process indexed by $\mathcal{M}$ rather than $\tilde{\mathcal{M}}$. We will always make it explicit when we are taking this point of view on the Neumann GFF.

Definition 5.23 (Neumann GFF with fixed normalisation)**.** *Suppose that $\rho_0 \in \mathcal{M} \setminus \tilde{\mathcal{M}}$. The Neumann GFF h with normalisation fixed so that $(h, \rho_0) = 0$ is the stochastic process defined from $\bar{h}$ in Definition 5.18 by setting*

$$(h, \rho) = (\bar{h}, \rho - \frac{\int_D \rho(dx)}{\int_D \rho_0(dx)} \rho_0)$$

for each $\rho \in \mathcal{M}$.

It is clear from Lemma 5.19 that this process, when restricted to $\mathcal{D}_0(D)$, has a version that is a.s. a distribution in D.

In the following, whenever we talk of a Neumann GFF with **arbitrary** fixed normalisation, we mean a Neumann GFF with fixed normalisation as defined above, for some $\rho_0 \in \mathcal{M} \setminus \tilde{\mathcal{M}}$ **deterministic.**

Example 5.24 (Semi-circle averages)**.** *Suppose that $D = \mathbb{H}$ and for $x \in \mathbb{R}, \varepsilon > 0$, let $\rho_{x,\varepsilon}$ be the uniform probability distribution on $\partial B(x, \varepsilon) \cap \mathbb{H}$ of radius ε about x. Then by Remark 5.17, we see that $\rho_{x,\varepsilon} \in \mathcal{M}$. Therefore if h is a Neumann GFF with a fixed normalisation, we can define the ε-semicircle average $(h, \rho_{x,\varepsilon})$ of h about x.*

Remark 5.25. Notice that if $\rho_1, \rho_2 \in \mathcal{M}$ with $\int_D \rho_1 = \int_D \rho_2$, then $\rho_1 - \rho_2 \in \tilde{\mathcal{M}}$. Hence we can define $(\bar{h}, \rho_1 - \rho_2)$ when $\bar{h}$ is a Neumann GFF modulo constants. We can also define $(h, \rho_1 - \rho_2)$ whenever h is a Neumann GFF with fixed normalisation, and its law will not depend on the normalisation: it will be exactly that of $(\bar{h}, \rho_1 - \rho_2)$.

5.4 Relationship with other boundary conditions

> $\star$ *This section could be skipped on a first reading. Its main purpose is to introduce the whole plane and Dirichlet–Neumann GFFs, and derive relationships with the Neumann and zero boundary condition versions. In particular, this enables the proof of some technical lemmas that will be useful later in the book.*

For two random variables X and Y taking values in the same measurable space $(E, \mathcal{E})$, we define the total variation distance between them by

$$d_{TV}(X, Y) = \sup_{A \in \mathcal{E}} |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)|.$$

If $(E, \mathcal{E})$ is a separable metric space, then there necessarily exists a coupling of X, Y such that $\mathbb{P}(X \neq Y) = d_{TV}(X, Y)$ [Che04, §5.1].

Lemma 5.26. *For $R > 1$, let h^R be a Dirichlet (zero) boundary condition GFF on $R\mathbb{D}$. Then as $R \to \infty$,*

$$\sup_{R_1, R_2 \geq R} d_{TV}(h^{R_1}|_{\mathbb{D}}, h^{R_2}|_{\mathbb{D}}) \to 0,$$

when $h^{R_1}|_{\mathbb{D}}$ and $h^{R_2}|_{\mathbb{D}}$ are considered as distributions modulo constants *in $\mathbb{D}$.*

Remark 5.27. It will be useful later on to be able to say that two GFF-like objects that are close in total variation distance can be coupled so that they are close with high probability. This could appear potentially problematic, since the space of (Schwarz) distributions is not a separable metric space. On the other hand, taking the specific setting of the lemma above to illustrate this point, we know that $h^R|_{\mathbb{D}}$ a.s. defines an element of the metric space $H^{-\varepsilon}(\mathbb{D})$ for every R and any $\varepsilon > 0$. Since $H^{-\varepsilon}(\mathbb{D})$ is a separable metric space, this implies that with arbitrarily high probability as $R_1, R_2 \to \infty$, we can couple $h^{R_1}|_{\mathbb{D}}$ and $h^{R_2}|_{\mathbb{D}}$ so that, as distributions, they differ by exactly a constant when restricted to $\mathbb{D}$.

Proof. Without loss of generality, assume that $R_2 \geq R_1$. Then by the Markov property of the Dirichlet GFF (Theorem 1.30), we can write $h^{R_2} = \tilde{h}^{R_1} + \varphi$, where $\tilde{h}^{R_1}$ has the law of h^{R_1}, and φ is independent of $\tilde{h}^{R_1}$ and a.s. harmonic in $R_1\mathbb{D}$. The proof of this lemma will essentially follow from the fact that, when viewed modulo constants and restricted to $\mathbb{D}$, φ is very small.

The key observation to this effect, is that if $\varphi_0 = \varphi - \varphi(0)$ then

$$\sup_{R_1,R_2\geq R} \sup_{w\in\partial(8\mathbb{D})} \mathrm{Var}(\varphi_0(w)) \to 0 \text{ as } R \to \infty \tag{5.13}$$

(note that it should not be obvious why this particular range of w has been chosen - this will become clear shortly). To justify (5.13), we use the decomposition $h^{R_2} = \tilde{h}^{R_1} + \varphi$ as an independent sum (plus harmonicity of φ), to see that $\mathrm{Var}(h_1^{R_2}(w) - h_1^{R_2}(0)) = \mathrm{Var}(h_1^{R_1}(w) - h_1^{R_1}(0)) + \mathrm{Var}(\varphi_0(w))$ for any $w \in \partial(8\mathbb{D})$. Since we know explicitly the Dirichlet Greens functions $G_0^{R_1\mathbb{D}}(x,y) = \log R_1 + \log|1 - (\bar{x}y/R_1^2)| - \log(|x-y|)$ and $G_0^{R_2\mathbb{D}}(x,y) = \log R_2 + \log|1 - (\bar{x}y/R_2^2)| - \log(|x-y|)$ for $x \neq y \in R_1\mathbb{D}$, (5.13) follows easily.

Now, note that h^{R_2} and $\tilde{h}^{R_1} + \varphi_0$ differ by exactly a constant in $R_1\mathbb{D}$. So we would be done if we could show that the laws of $\tilde{h}^{R_1} + \varphi_0$ and h^{R_1} (equivalently $\tilde{h}^{R_1}$) are close in total variation distance when restricted to $\mathbb{D}$ (uniformly in $R_2 \geq R_1 \geq R$ as $R \to \infty$). The idea for this is to use the explicit expression for the Radon–Nikodym derivative between a zero boundary GFF and a zero boundary GFF plus an H_0^1 function; see Proposition 1.29.

The first obstacle here is that φ_0 is obviously not in $H_0^1(R_1\mathbb{D})$. To get around this, we introduce $\tilde{\varphi}(z) = \psi(|z|)\varphi_0(z)$ for $z \in R_1\mathbb{D}$, where $\psi : [0, R_1] \to [0,1]$ is smooth, equal to 1 on $[0,1]$, and equal to 0 on $[2, R_1]$. Note that $\tilde{\varphi} \in H_0^1(R_1\mathbb{D})$ and that $\tilde{\varphi} = \varphi_0$ in $\mathbb{D}$. Moreover conditionally on $\tilde{\varphi}$, the Radon–Nikodym derivative between the laws of $\tilde{h}^{R_1}$ and $\tilde{h}^{R_1} + \tilde{\varphi}$ is given by

$$Z := \frac{\exp((\tilde{h}^{R_1}, \tilde{\varphi})_\nabla)}{\exp((\tilde{\varphi}, \tilde{\varphi})_\nabla)}, \tag{5.14}$$

see Proposition 1.29. By definition of the total variation distance, it suffices to show that (5.14) tends to 1 in $L^1(\mathbb{P})$, uniformly over $R_2 \geq R_1 \geq R$ as $R \to \infty$.

To show this, we will first prove that

$$\sup_{R_1,R_2\geq R} \mathbb{E}(e^{(\tilde{\varphi},\tilde{\varphi})_\nabla} - 1) \to 0 \tag{5.15}$$

as $R \to \infty$. Indeed, by harmonicity of φ_0, $(\tilde{\varphi}, \tilde{\varphi})_\nabla$ is less than some deterministic constant (depending on ψ) times $\sup_{z\in 4\mathbb{D}} |\varphi_0(z)|^2$, which in turn is less than another deterministic

constant times $\int_{\partial(8\mathbb{D})} |\varphi_0(w)|^2 \, \rho(dw)$, where ρ is the uniform measure on the circle $\partial(8\mathbb{D})$. Let us write C for the product of these two deterministic constants. Then we have that

$$\mathbb{E}(e^{(\tilde{\varphi},\tilde{\varphi})_\nabla} - 1) \leq \mathbb{E}(e^{C\int_{\partial(8\mathbb{D})} |\varphi_0(w)|^2 \, \rho(dw)} - 1)$$

which by Jensen's inequality is less than

$$\mathbb{E}(\int_{\partial(8\mathbb{D})} (e^{C|\varphi_0(w)|^2} - 1) \, \rho(dw)) \leq \int_{\partial(8\mathbb{D})} \mathbb{E}(e^{C|\varphi_0(w)|^2} - 1) \, \rho(dw).$$

Note that since C is a fixed constant and $\varphi_0(w)$ is centred and Gaussian with arbitrarily small variance (uniformly over $\partial(8\mathbb{D})$) as $R \to \infty$, these expectations will all be finite for $R_2 \geq R_1 \geq R$ large enough. Moreover, the right-hand side of the above expression will go to 0 uniformly in $R_2 \geq R_1 \geq R$ as $R \to \infty$. To conclude, we simply observe that conditionally on $\tilde{\varphi}$, the random variable Z from (5.14) is log-normal with parameters $(-(\tilde{\varphi},\tilde{\varphi})_\nabla^2/2, (\tilde{\varphi},\tilde{\varphi})_\nabla)$, meaning that

$$\mathbb{E}(|Z-1|^2) = \mathbb{E}(\mathbb{E}(|Z-1|^2 \,|\, \tilde{\varphi})) = \mathbb{E}(e^{(\tilde{\varphi},\tilde{\varphi})_\nabla} - 1).$$

By (5.15), this completes the proof.

□

Corollary 5.28 (Whole plane GFF in $\mathbb{D}$)**.** *There exists a law on distributions modulo constants in $\mathbb{D}$, such that if h^∞ is the distribution modulo constants with this law then*

$$d_{TV}(h^R|_{\mathbb{D}}, h^\infty) \to 0 \text{ as } R \to \infty.$$

Moreover, we can write $h^\infty = h^{\mathbb{D}} + \varphi$, where $h^{\mathbb{D}}$ has the law of a Dirichlet boundary condition GFF in $\mathbb{D}$, and φ is a harmonic function modulo constants that is independent of $h^{\mathbb{D}}$.

Remark 5.29. By scaling, it is possible to deduce the same result as above when $\mathbb{D}$ is replaced by any bounded subset D of $\mathbb{C}$. This gives rise to a well-defined distribution modulo constants h^∞ in $\mathbb{C}$ known as the whole plane GFF. For an equivalent construction, see the exercises for this section.

Lemma 5.30 (Neumann and whole plane GFFs)**.** *For h^∞ a whole plane GFF, define the* even *part of h^∞ on $\mathbb{D}_+ := \mathbb{D} \cap \mathbb{H}$ by*

$$(h^\infty_{\text{even}}, \rho) := \frac{(h^\infty, \rho) + (h^\infty, \rho^*)}{2} \text{ for } \rho \in \tilde{\mathcal{D}}_0(\mathbb{D}_+), \ \rho^*(z) = \rho(\bar{z}).$$

Write $h^{\mathbb{H}}$ for a Neumann GFF on $\mathbb{H}$. Then, as distributions modulo constants,

$$h^{\mathbb{H}}|_{\mathbb{D}_+} \overset{(law)}{=} \sqrt{2} h^\infty_{\text{even}}|_{\mathbb{D}_+}.$$

Proof. The fact that they have the same law as Gaussian processes indexed by $\tilde{\mathcal{D}}_0(\mathbb{D}_+)$ follows straight away by comparing covariances. Thus, there exist versions of both that have the same law as distributions modulo constants. □

Remark 5.31. One can also consider the odd part of a whole plane GFF, which turns out to be given by a Dirichlet GFF – see the exercises for this section.

Definition 5.32 (Dirichlet–Neumann GFF)**.** *Suppose that $h^{\mathbb{D}}$ is a Dirichlet GFF in $\mathbb{D}$. Then the Dirichlet–Neumann GFF, h^{DN}, is defined to be $\sqrt{2}$ times its even part*

$$(h^{\mathrm{DN}}, \rho) := \sqrt{2}(h^{\mathbb{D}}_{\mathrm{even}}, \rho) := \frac{(h^{\infty}, \rho) + (h^{\infty}, \rho^*)}{\sqrt{2}} \text{ for } \rho \in \mathcal{D}_0(\mathbb{D}_+)$$

which is a random distribution on $\mathbb{D}_+$.

Putting this together with Lemma 5.30 we see that $h^{\infty} = h^{\mathbb{D}} + \varphi$ has even part equal to $(1/\sqrt{2})$ times a Dirichlet–Neumann GFF, plus a harmonic function φ_{even} modulo constants in $\mathbb{D}_+$ with Neumann boundary conditions on $(-1, 1)$. (Note that the even part of a harmonic function in $\mathbb{D}$ defines a harmonic function in $\mathbb{D}_+$ with Neumann boundary conditions on $(-1, 1)$).

Proposition 5.33 (Boundary Markov property)**.** *Let $h^{\mathbb{H}}$ be a Neumann GFF on $\mathbb{H}$ (considered modulo constants). Then we can write*

$$h^{\mathbb{H}}|_{\mathbb{D}_+} = h^{\mathrm{DN}} + \varphi_{\mathrm{even}}$$

where the two summands are independent, h^{DN} has the law of a Dirichlet–Neumann GFF in $\mathbb{D}_+$, and φ_{even} is a harmonic function modulo constants in $\mathbb{D}_+$ with Neumann boundary conditions on $(-1, 1)$.

We conclude this section with one further comment, that will be useful at the end of this chapter and in Chapter 6. It can be used to say, roughly speaking, that any (nice enough) way of fixing the additive constant for a Neumann GFF in $\mathbb{H}$ will produce a field with the same behaviour when looking very close to the origin. Moreover, this will still be true if we condition on the realisation of the field far away from the origin.

Lemma 5.34. *Suppose that h is a Neumann GFF in $\mathbb{H}$, with normalisation fixed so that it has average 0 on the upper unit semicircle (this makes sense by Example 5.24). Let h^{DN} be an independent Dirichlet–Neumann GFF in $\mathbb{D}_+$. Then for any $K > 0$ the total variation distance between*

- *the joint law of $(h|_{K\mathbb{D}_+\setminus\mathbb{D}_+}, h|_{\delta\mathbb{D}_+})$ and*
- *the (independent product) law $(h|_{K\mathbb{D}_+\setminus\mathbb{D}_+}, h^{\mathrm{DN}}|_{\delta\mathbb{D}_+})$,*

tends to 0 as $\delta \to 0$. Note that the fields can be viewed as distributions here, rather than just distributions modulo constants.

Proof. The proof basically follows from taking even parts in Lemma 5.26.

More precisely, let $R \gg K$ be large and fixed, and write

$$\tilde{h}^{R\mathbb{D}} = h^{R\mathbb{D}} - h_1^{R\mathbb{D}(0)},$$

for $h^{R\mathbb{D}}$ a Dirichlet GFF in $R\mathbb{D}$ and $h_1^{R\mathbb{D}(0)}$ its unit circle average around 0. By Lemma 5.30 and Corollary 5.28, it suffices to prove that for any such R, and for $h^{\mathbb{D}}$ a Dirichlet GFF in $\mathbb{D}$ that is independent of $\tilde{h}^{R\mathbb{D}}$

$$d_{TV}\left((\tilde{h}^{R\mathbb{D}}|_{K\mathbb{D}\setminus\mathbb{D}}, \tilde{h}^{R\mathbb{D}}|_{\delta\mathbb{D}}), (\tilde{h}^{R\mathbb{D}}|_{K\mathbb{D}\setminus\mathbb{D}}, h^{\mathbb{D}}|_{\delta\mathbb{D}})\right) \to 0 \text{ as } \delta \to 0.$$

But it follows from the proof of Lemma 5.26 (after rescaling) that the Radon–Nikodym derivative, between the law of $h^{\mathbb{D}}|_{\delta\mathbb{D}}$ and the conditional law of $\tilde{h}^{R\mathbb{D}}|_{\delta\mathbb{D}}$ given $\tilde{h}^{R\mathbb{D}}|_{K\mathbb{D}\setminus\mathbb{D}}$, tends to 0 in $L^1(\mathbb{P})$ as $\delta \to 0$. This gives the desired claim. □

Remark 5.35. Note that the proof (and therefore the Lemma) will still hold if we replace h by $h + \mathfrak{h}$ where $\mathfrak{h}$ is a deterministic harmonic function in $\mathbb{D}^+$ with Neumann boundary conditions on $[-1, 1]$. Moreover the convergence will be uniform over $\{\mathfrak{h} : \sup_{z\in\mathbb{D}^+} |\mathfrak{h}(z)| \leq C\}$.

5.5 Semicircle averages and boundary Liouville measure

Let $\bar{h}$ be a Neumann GFF on $\mathbb{H}$ modulo constants (from now on we will use a bar in order to distinguish statements concerning the Neumann GFF modulo constants and Neumann GFFs with fixed normalisations). An immediate consequence of our previous considerations is the following fact.

Theorem 5.36. *For any $x \in \mathbb{R}$, the finite dimensional distributions of the process*

$$(X_t)_{t\in\mathbb{R}} := ((\bar{h}, \rho_{x,e^{-t}} - \rho_{x,1}))_{t\in\mathbb{R}}$$

are those of a two-sided Brownian motion with variance 2 (so $\mathrm{Var}(X_t) = 2|t|$*).*

Note that the statement of the theorem makes sense, since for any $\varepsilon > 0$, $\rho_{x,\varepsilon} - \rho_{x,1} \in \tilde{\mathcal{M}} = \tilde{\mathcal{M}}^{\mathbb{H}}$. By Remark 5.25, this also means that if h is a Neumann GFF in $\mathbb{H}$ with *any* fixed normalisation, and $h_\varepsilon(x) := (h, \rho_{x,\varepsilon})$, then

$$(h_{e^{-t}}(x) - h_1(x))_{t\in\mathbb{R}}$$

is a two-sided Brownian motion with variance 2.

Proof of Theorem 5.36. Without loss of generality we may take $x = 0$. Then by conformal invariance (actually just scale invariance) of $\bar{h}$, it follows that X has stationary increments. Moreover, by applying the Markov property (a scaled version of Proposition 5.33) in the semi disc of radius e^{-t} about 0 for any t, we see that $(X_r)_{r\leq t}$ and $(X_s - X_t)_{s\geq t}$ are independent. Hence, X has stationary and independent increments.

Since the increments are also Gaussian with mean zero and finite variance, it must be that $X_t = B_{\kappa t}$ for some $\kappa > 0$, where B is a standard Brownian motion. It remains to check that $\kappa = 2$, but this follows from the fact that $G^{\mathbb{H}}(0, y) = 2\log(1/\varepsilon)$ if $|y| = \varepsilon$: see (5.11). □

⋆ *The next corollary will only be used for the proofs of some technical results later on, and can safely be skipped by those initially wanting to learn more about the Neumann GFF.*

Corollary 5.37. *Suppose that h is a Neumann GFF in $\mathbb{H}$, with normalisation fixed so that it has average 0 on the upper unit semicircle. Also fix $\rho_0 \in \mathcal{M}^{\mathbb{H}} \setminus \tilde{\mathcal{M}}^{\mathbb{H}}$ with compact support contained outside of $\mathbb{D}_+ = \mathbb{D} \cap \mathbb{H}$ and let $\hat{h}$ be Neumann GFF in $\mathbb{H}$ with normalisation fixed so that $(\hat{h}, \rho_0) = 0$. Let $\mathfrak{h}$ be a deterministic harmonic function in $\mathbb{D}^+$ with Neumann boundary conditions on $(-1, 1)$. Then the total variation distance between the laws of $h|_{\delta\mathbb{D}_+}$ and $\hat{h} + \mathfrak{h}|_{\delta\mathbb{D}_+}$ tends to 0 as $\delta \to 0$. This convergence is uniform on the set $\{(\rho_0, \mathfrak{h}) : \|\mathfrak{h}\|_\infty \leq C_1, \mathrm{Var}((h, \rho_0)) \leq C_2\}$ for any $C_1, C_2 > 0$.*

Proof. This follows from Lemma 5.34, Remark 5.35 and Definition 5.23, together with the fact that a Brownian motion started from 0 and a Brownian motion started from any finite deterministic value can be coupled so that they agree after time T with arbitrarily high probability as $T \to \infty$. Uniformity on sets of desired form follows easily from the proof of Lemma 5.34. □

Having identified the "boundary behaviour" of the Neumann GFF, we can now construct a random measure supported on the boundary of $\mathbb{H}$. As it turns out, the measure of interest to us is again given by an "exponential of the Neumann GFF", but the multiplicative factor in the exponential is $\gamma/2$ rather than γ. The reason for this choice is rather deep, and has to do with the fact that we plan to use it to measure the "quantum length of an SLE". (Another justification comes from the fact that it satisfies the same KPZ equation as in the bulk case).

Theorem 5.38 (Boundary Liouville measure for the Neumann GFF on $\mathbb{H}$)**.** *Let h be a Neumann GFF in $\mathbb{H}$ with some arbitrary normalisation. Define a measure ν_ε on $\mathbb{R}$ by setting $\nu_\varepsilon(dx) = \varepsilon^{\gamma^2/4} e^{(\gamma/2)h_\varepsilon(x)} dx$. Then for $\gamma < 2$, the measure ν_ε converges a.s. along the dyadic subsequence $\varepsilon = 2^{-k}$ to a nontrivial, nonatomic measure ν called the boundary Liouville measure.*

Proof. This follows from Theorem 3.2 with h the (normalised) Neumann GFF, σ equal to Lebesgue measure on $\mathbb{R}$ and θ the semi-circle uniform measure. Equivalently, one can prove the convergence as in Chapter 2, using the Markov property. □

Note the scaling in ν_ε, which is by $\varepsilon^{\gamma^2/4}$. This is because, as proved in Theorem 5.36 (also see the discussion below), when $x \in \mathbb{R}$ and h is a Neumann GFF we have $\mathrm{Var}\, h_\varepsilon(x) = 2\log(1/\varepsilon) + O(1)$.

Remark 5.39. The law of ν above *does* depend on the choice of normalisation for h. If one starts with a Neumann GFF modulo constants, then the boundary Liouville measure can be defined as a measure *up to a multiplicative constant*.

For general D, h a Neumann GFF in D with arbitrary normalisation, and z, ε such that $B(z,\varepsilon) \subset D$, we can also define the circle average $(h, \rho_{z,\varepsilon}) =: h_\varepsilon(z)$. Although we use the same notation $h_\varepsilon(\cdot)$ for circle averages and semi-circle averages it should always be clear which one we refer to, depending whether the argument lies, respectively, in the bulk or on the boundary.

Definition 5.40 (Bulk Liouville measure for the Neumann GFF)**.** *When h is a Neumann GFF with some arbitrary normalisation and $\gamma < 2$, we can also define the bulk Liouville measure*

$$\mu(dz) := \lim_{\varepsilon \to 0} \varepsilon^{\gamma^2/2} \mathrm{e}^{\gamma h_\varepsilon(z)}\, dz,$$

exactly as for the Dirichlet GFF.

The existence of this limit follows from the construction of GMC measures for general log-correlated Gaussian processes in Chapter 3. The analogue of Remark 5.39 also applies in this case.

Remark 5.41. The results of Chapter 3 imply that for any fixed compact set of $\mathbb{R}$ (respectively $\mathbb{H}$) the boundary (respectively bulk) Liouville measure will assign finite and strictly positive mass to that set with probability one.

The conformal covariance properties of the boundary and bulk Liouville measures are not quite as straightforward as for the Dirichlet GFF. The first problem is that conformal invariance of the Neumann GFF only holds when we view it as a distribution modulo constants. The second is that we have only defined the boundary measure on the domain $\mathbb{H}$, where semicircles centered on the boundary can be defined. We could extend this definition to linear boundary segments of other domains, but it is unclear what to do when the boundary of the domain is very wild.

Let us start with the bulk measure, where we only need to deal with the first problem. In this case, the statement

$$\mu_h \circ T^{-1} = \mu_{h \circ T^{-1} + Q \log |(T^{-1})'|}$$

of Theorem 2.8 still holds (by absolute continuity with respect to the Dirichlet GFF) when $T : D \to D'$ is a deterministic, conformal map and we replace the Dirichlet GFF with a Neumann GFF h in D with some arbitrary normalisation. However, now $h \circ T^{-1}$ is a Neumann GFF in D' with a *different* normalisation. The exact analogue of Theorem 2.8 only holds if we consider Neumann GFFs modulo additive constants, and their associated bulk Liouville measures modulo multiplicative constants (see exercises).

Now for the boundary measure, suppose that h is a Neumann GFF on $\mathbb{H}$ with some arbitrary normalisation, and $T : \mathbb{H} \to D$ is conformal. Then $h' := h \circ T^{-1}$ is a Neumann GFF on D with another normalisation. Moreover, if ∂D contains a linear boundary segment $L \subset \partial D \cap \mathbb{R}$, the measure $\nu_{h'}(dx) = \lim_{\varepsilon \to 0} \mathrm{e}^{(\gamma/2) h'_\varepsilon(x)} \varepsilon^{\gamma^2/4}\, dx$ is well-defined and

$$\nu_h \circ T^{-1} = \nu_{h \circ T^{-1} + Q \log |(T^{-1})'|} = e^{\gamma Q \log |(T^{-1})'|} \nu_{h'}. \tag{5.16}$$

on L with probability one. In fact, by [SW16, Theorem 4.3], the measure is well-defined and the above formula holds with probability one for all conformal $T : \mathbb{H} \to D$ with $\partial D \cap \mathbb{R} \neq \emptyset$ simultaneously.

We will use this formula to *define* the boundary Liouville measure for GFF-like fields on the conformal boundary[6] of an arbitrary simply connected domain.

Definition 5.42 (Boundary Liouville measure for the GFF on D)**.** *Suppose that h is a random generalised function on D, and that for some conformal map $T : \mathbb{H} \to D$ the field $h \circ T + Q \log |T'|$ has the law of a Neumann GFF (with an arbitrary normalisation) plus an almost surely continuous function on some neighbourhood (in $\mathbb{H}$) of $L \subset \mathbb{R}$. Then the measure $\nu_{h \circ T + Q \log |T'|}$ is a.s. well-defined on L, and we may define*

$$\nu_h := \nu_{h \circ T + Q \log |T'|} \circ T^{-1} \tag{5.17}$$

to be the Liouville measure for h, on the part of the conformal boundary of D corresponding to the image of L under T. This defines the same measure simultaneously for all choices of T, with probability one.

Note that the behaviour of conformal maps near the boundary of a domain can be very wild. For instance if D is a domain whose boundary is only Hölder with a certain exponent, then the boundary Liouville measure defined as above may not be easy to construct directly by approximation.

5.6 Convergence of random surfaces

Recall that we defined a *random surface* to be an equivalence class of pairs (D, h) where D is a simply connected domain and h is a distribution on D, under the relation identifying (D_1, h_1) and (D_2, h_2) if for some $f : D_1 \to D_2$ conformal,

$$h_2 = h_1 \circ f^{-1} + Q \log |(f^{-1})'|.$$

The reason for this was that if h_1 is a Dirichlet Gaussian free field in D_1, then all members of the equivalence class of (D_1, h_1) describe the same Liouville measure up to taking conformal images.

Now, we have seen that the same thing is true when h_1 is a Neumann GFF with an arbitrary normalisation. And indeed if we want to view the Neumann GFF as a quantum surface then we have to fix a normalisation, since the definition of quantum surface involves distributions and not distributions modulo constants. But the Neumann GFF is only really *uniquely* defined as a distribution modulo constants. This manifests itself in the following problem: different ways of fixing the normalisation do not yield the same quantum surface in law (see Example 5.46 below). So if we want to view the Neumann GFF as a quantum

[6] The conformal boundary of a simply connected domain D, equivalent to the Martin boundary (see [BN11, §1.3]), is the set of limit points of D with respect to the metric $d(x, y) = d(\phi(x), \phi(y))$ for $\phi : D \to \mathbb{D}$ conformal.

surface, which normalisation should we pick? The lack of a canonical answer to this suggests that, at least when working with quantum surfaces, it is perhaps more natural to look at a slightly different object.

Another point of view is the following: if we consider a Neumann GFF h with some arbitrary normalisation, and also the field $h + C$ for some C, then the Liouville measure for $h + C$ is just $e^{\gamma C}$ times the Liouville measure for h. So we can think that the quantum surface described by $h + C$ is a "zoomed in" version of the quantum surface defined by h. (Note that this is distinct from rescaling space by a fixed factor and applying the change of coordinate formula, since by definition this does not change the quantum surface). And for some purposes, it will be very nice to have a quantum surface that is invariant (in law) under this zooming. Such a property can be thought of as a type of **scale invariance for quantum surfaces**.

In order to construct a surface $(\mathbb{H}, h)$ which does have this invariance property (once again, by Example 5.46 below this is not true when h is a Neumann GFF with some arbitrary normalisation), Sheffield [She16a] introduced the notion of **quantum wedge**. This will play an important role in our study of the *quantum gravity zipper* in next section. Roughly speaking, a quantum wedge is the limiting surface that one obtains by "zooming in" to a Neumann GFF close to a point on the boundary. Since this surface is obtained as a scaling limit, it automatically satisfies the desired scale invariance.

In order to make proper sense of the above discussion, we first need to provide a notion of convergence for random surfaces – and more precisely, for surfaces with marked points on their boundaries.

Definition 5.43 (Quantum surface with k-marked points)**.** *A quantum surface with k-marked boundary points is an equivalence class of tuples $(D, h, x_1, \cdots, x_k)$ where $D \subset \mathbb{C}$ is simply connected, $h \in \mathcal{D}'_0(D)$, and $x_1, \cdots, x_k$ are points on the (conformal) boundary of D, under the equivalence relation $(D, h, x_1, \cdots, x_k) \sim (D', h', x'_1, \cdots, x'_k)$ if and only if for some $T : D \to D'$ conformal with $T(x_i) = x'_i$ for $1 \leq i \leq k$ (note that T extends to a to map between conformal boundaries by definition):*

$$h' = h \circ T^{-1} + Q \log |(T^{-1})'|. \tag{5.18}$$

We recall that $Q = Q_\gamma = 2/\gamma + \gamma/2$ depends on the LQG parameter γ, and therefore so does the notion of *quantum surface*, but we drop this from the notation for simplicity. Note that since h is assumed to be in the space of distributions $\mathcal{D}'_0(D)$, this definition may be applied to a Neumann GFF with an arbitrary fixed normalisation.

In order to define a quantum surface $\mathcal{S}$ with k marked points, we need only specify a single equivalence class representative $(D, h, x_1, \cdots, x_k)$. We will call such a representative an **embedding** or **parametrisation** of the quantum surface.

This means that our usual topology on the space of distributions induces a topology on the space of quantum surfaces (with k marked points):

Definition 5.44 (Quantum surface convergence)**.** *A sequence of quantum surfaces $\mathcal{S}^n$ converges to a quantum surface $\mathcal{S}$ as $n \to \infty$ if there exist representatives $(D, h^n, x_1, \cdots, x_k)$ of $\mathcal{S}^n$ and $(D, h, x_1, \cdots, x_k)$ of $\mathcal{S}$, such that $h_n \to h$ in the space of distributions as $n \to \infty$.*

(We note that this notion of convergence is somewhat different from the notions used in [She16a] or [DMS14], but this definition has the advantage that it makes sense for all deterministic distributions viewed as quantum surfaces rather than a special class of random ones. It is also, in any case, the one that actually used to verify convergence statements for quantum surfaces, as will be discussed below.)

Now, when we are actually working with quantum surfaces, it will often be very useful to specify a surface by describing a particular *canonically chosen* embedding. Of particular interest are *random* surfaces (like the Neumann GFF or the quantum wedges to be defined below), and this allows for certain special choices of embedding (we will see several in the rest of this chapter and the next).

Example 5.45. *Suppose that h is equal to a continuous function plus Neumann GFF (with some normalisation) in D, and $z_0, z_1 \in \partial D$ are such that the* bulk *Liouville measure μ_h for h assigns finite mass to any finite neighbourhood of z_0, and infinite mass to any neighbourhood of z_1.*[7] *Then the doubly marked quantum surface (D, h, z_0, z_1) has a unique representative $(\mathbb{H}, \tilde{h}, 0, \infty)$ such that $\mu_{\tilde{h}}(\mathbb{D}\cap\mathbb{H}) = 1$. The distribution $\tilde{h}$ is called the* ***canonical description*** *of the quantum surface in [She16a] (but bear in mind that it is only well defined when h is in a particular class of distributions for which the Liouville measure makes sense). In fact, in practice this is a difficult embedding to work with and we usually prefer others - see the next section.*

Example 5.46 (Zooming in – important!)**.** *Let h be a Neumann GFF in $\mathbb{H}$, for concreteness, normalised to have average zero in $\mathbb{D}\cap\mathbb{H}$. Then the canonical descriptions of h and of $h+100$ (say), viewed as quantum surfaces in $\mathbb{H}$ with marked points at 0 and ∞, are* very different*. This can be confusing at first, since h is in some sense defined "up to a constant", but the point is that "equivalence as quantum surfaces" and "equivalence as distributions modulo constants"* are not the same*.*

Indeed to find the canonical description of h we just need to find the (random) r such that $\mu_h(B(0,r)\cap\mathbb{H}) = 1$, and apply the conformal map $z \mapsto z/r$; the resulting field

$$\tilde{h}(z) = h(rz) + Q\log(r)$$

defines the canonical description $\tilde{h}$ of the surface $(\mathbb{H}, h, 0, \infty)$. On the other hand, in order to find the canonical description of $h+100$, we need to find $s > 0$ such that $\mu_{h+100}(B(0,s)\cap\mathbb{H}) = 1$. That is, we need to find $s > 0$ such that $\mu_h(B(0,s)\cap\mathbb{H}) = e^{-100\gamma}$. The resulting field

$$h^*(z) = h(sz) + Q\log(s) + 100$$

defines the canonical description of $(\mathbb{H}, h+100, 0, \infty)$.

Note that in this example, the ball of radius s is much smaller than the ball of radius r. Yet in $\tilde{h}$, the ball of radius r has been scaled to become the unit disc, while in h^* it is the ball

[7] If h is just a Neumann GFF with arbitrary normalisation in an unbounded domain D, then this will be the case whenever $z_1 = \infty$ and z_0 is another ($\neq \infty$) boundary point where the boundary is smooth (say).

of radius s which has been scaled to become the unit disc. In other words, and since s is much smaller than r, the surface $(\mathbb{H}, h + 100, 0, \infty)$ is obtained by taking the surface $(\mathbb{H}, h, 0, \infty)$ and ***zooming in*** *at* 0.

5.7 (Thick) quantum wedges

As we will see very soon, a (thick) quantum wedge is the abstract random surface that arises as the $C \to \infty$ limit of the doubly marked surface $(h+C, \mathbb{H}, 0, \infty)$, when h is a Neumann GFF in $\mathbb{H}$ with some normalisation plus certain logarithmic singularity at the origin. Thus, as explained in the example above, it corresponds to zooming in near the origin of $(h, \mathbb{H}, 0, \infty)$.

In practice however, we prefer to work with a concrete definition of the quantum wedge and then prove that it can indeed be seen as a scaling limit. It turns out to be most convenient to define it in the infinite strip $S = \mathbb{R} \times (0, \pi)$ rather than the upper half plane, with the two marked boundary points being $+\infty$ and $-\infty$ respectively. A conformal map transforming $(S, \infty, -\infty)$ into $(\mathbb{H}, 0, \infty)$ is given by $z \mapsto -e^{-z}$. Then vertical line segments are mapped to semicircles. To be precise, the segment $\{z : \Re(z) = s\}$ is mapped to $\partial B(0, e^{-s}) \cap \bar{\mathbb{H}}$ for every $s \in \mathbb{R}$.

The following lemma will be used repeatedly in the rest of this chapter and the next.

Lemma 5.47 (Radial decomposition)**.** *Let S be the infinite strip $S = \{z = x + iy \in \mathbb{C} : y \in (0, \pi)\}$. Let $\bar{\mathcal{H}}_{\text{rad}}$ be the subspace of $\bar{H}^1(S)$ obtained as the closure of smooth functions which are constant on each vertical segment, viewed modulo constants. Let $\mathcal{H}_{\text{circ}}$ be the subspace obtained as the closure of smooth functions which have mean zero on all vertical segments. Then*

$$\bar{H}^1(S) = \bar{\mathcal{H}}_{\text{rad}} \oplus \mathcal{H}_{\text{circ}}.$$

Proof. Suppose that g_1 is a smooth function modulo constants in S, that is constant on vertical lines, and that g_2 is a smooth function in S that has mean zero on every vertical line. Then it is straightforward to check that $\iint_S \nabla g_1 \cdot \nabla g_2 = 0$. Indeed $\nabla g_1 = (\partial_x g_1, 0)$ and $\nabla g_2 = (\partial_x g_2, \partial_y g_2)$ where the partial derivative $\partial_x g_1$ is constant on vertical lines and $\partial_x g_2$ has average 0 on vertical lines. This means that $\nabla g_1 \cdot \nabla g_2$ has average 0 on every vertical line, and consequently has average 0 over S. By definition of $\bar{\mathcal{H}}_{\text{rad}}$ and $\mathcal{H}_{\text{circ}}$ (as closures with respect to $(\cdot, \cdot)_\nabla$) the two spaces are therefore orthogonal with respect to $(\cdot, \cdot)_\nabla$.

To check that they span $\bar{H}^1(S)$, note that if $f \in \bar{\mathcal{D}}(S)$ and we set $f_{\text{rad}}(z)$ to be the average of f on the line $\Re z + i[0, 2\pi]$, then $f_{\text{rad}} \in \bar{\mathcal{H}}_{\text{rad}}$. Moreover, defining $f_{\text{circ}} = f - f_{\text{rad}}$, it is clear that $f_{\text{circ}} \in \mathcal{H}_{\text{circ}}$. From this it follows that if $f \in \bar{H}^1(S)$ then we can write $f = \lim_n f_n = \lim_n ((f_n)_{\text{rad}} + (f_n)_{\text{circ}})$ where by orthogonality, the sequences $(f_n)_{\text{rad}}$ and $(f_n)_{\text{circ}}$ are each Cauchy and have individual limits $f_{\text{rad}} \in \bar{\mathcal{H}}_{\text{rad}}$ and $f_{\text{circ}} \in \mathcal{H}_{\text{circ}}$. Then $f = f_{\text{rad}} + f_{\text{circ}}$, and the two spaces do indeed span $\bar{H}^1(S)$. □

Similarly to the domain Markov property for the Neumann GFF (that we saw arises from the orthogonal decomposition $\bar{H}^1(D) = H_0^1(D) \oplus \overline{\text{Harm}}(D)$), this results in another representation of the Neumann GFF on S modulo constants. Namely, it can be written as $h = h_{\text{rad}}^{\text{GFF}} + h_{\text{circ}}^{\text{GFF}}$ where:

- $h_{\text{rad}}^{\text{GFF}}, h_{\text{circ}}^{\text{GFF}}$ are independent;
- $h_{\text{rad}}^{\text{GFF}}(z) = B_{2s}$ if $\Re(z) = s$, where B is an independent standard Brownian motion modulo constants (by Theorem 5.36 and conformal invariance);
- $h_{\text{circ}}^{\text{GFF}}(z)$ has mean zero on each vertical segment.

We will now explain this briefly. If we define $h_{\text{rad}}^{\text{GFF}}$ to be constant on each vertical segment with value equal to the average of h on that segment, then we know by Theorem 5.36 and conformal invariance that $h_{\text{rad}}^{\text{GFF}}$ has the distribution above. Thus it remains to justify is that $(h - h_{\text{rad}}^{\text{GFF}}, \rho)$ and $(h_{\text{rad}}^{\text{GFF}}, \rho)$ are independent for any $\rho \in \tilde{\mathcal{M}}^S$. For this, observe that if $(\bar{h}_n)_n$ is as in Lemma 5.19, then $(\bar{h}_n, \rho)$ converges in $L^2(\mathbb{P})$ and in probability, to a random variable with the law of (h, ρ). Moreover $((\bar{h}_n)_{\text{rad}}, \rho)$ and $(\bar{h}_n - (\bar{h}_n)_{\text{rad}}, \rho)$ are independent for every n, with $\text{Var}((\bar{h}_n)_{\text{rad}}, \rho) \leq \text{Var}((h_{\text{rad}}^{\text{GFF}}, \rho))$ and $\text{Var}(\bar{h}_n - (\bar{h}_n)_{\text{rad}}, \rho) \leq \text{Var}(h - h_{\text{rad}}^{\text{GFF}}, \rho)$. This implies that $(h - h_{\text{rad}}^{\text{GFF}}, \rho)$ and $(h_{\text{rad}}^{\text{GFF}}, \rho)$ are uncorrelated and hence, by Gaussianity, independent.

Note that the $h_{\text{rad}}^{\text{GFF}}$ part is defined modulo constants, while the $h_{\text{circ}}^{\text{GFF}}$ part has additive constant fixed. Also observe that all the roughness of h is contained in the $h_{\text{circ}}^{\text{GFF}}$ part, as $h_{\text{rad}}^{\text{GFF}}$ is a nice continuous function modulo constants. On the upper half plane, this would correspond to a decomposition of h into a part which is a radially symmetric continuous function (modulo constants), and one which has zero average on every semicircle (hence the notation).

Remark 5.48 (Translation invariance of $h_{\text{circ}}^{\text{GFF}}$)**.** Note that the Neumann GFF h on S is invariant under horizontal translations (modulo constants), as it is conformally invariant (modulo constants). Since the radial part is simply a two-sided Brownian motion, the translation invariance of this part modulo constants is also clear. Thus, we may deduce that the circular part $h_{\text{circ}}^{\text{GFF}}$ is translation invariant as well. (Note that the additive constant here is specified).

Let $0 \leq \alpha < Q = 2/\gamma + \gamma/2$. We will define an **$\alpha$-(thick) quantum wedge** by specifying separately its averages on vertical line segments, and what is left when we subtract these. The second of these components will be an element of $\mathcal{H}_{\text{circ}}$, having exactly the same law as the corresponding projection $h_{\text{circ}}^{\text{GFF}}$ of the standard Neumann GFF. It is only the "radially symmetric part" which is different.

The bound $\alpha < Q$ corresponds to the fact that we are defining so-called "thick" quantum wedges. When $\alpha > Q$ it is possible to define something called a "thin" quantum wedge, as introduced in [DMS14], but we will not touch on these in the present chapter.

Definition 5.49. *Let*

$$h_{\text{rad}}(z) = \begin{cases} B_{2s} + (\alpha - Q)s & \text{if } \Re(z) = s \text{ and } s \geq 0 \\ \widehat{B}_{-2s} + (\alpha - Q)s & \text{if } \Re(z) = s \text{ and } s < 0 \end{cases} \tag{5.19}$$

where $B = (B_t)_{t \geq 0}$ is a standard Brownian motion, and $\widehat{B} = (\widehat{B}_t)_{t \geq 0}$ is an independent Brownian motion conditioned so that $\widehat{B}_{2t} + (Q - \alpha)t > 0$ for all $t > 0$.

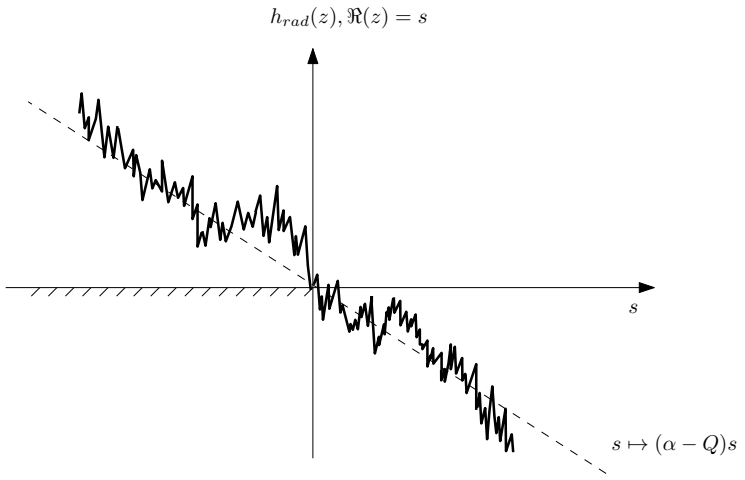


Figure 13: Schematic representation of the radially symmetric part of a quantum wedge in a strip. When $s < 0$, the function is conditioned to be positive.

*Let h_{circ} be a stochastic process indexed by $\mathcal{M} = \mathcal{M}^S$ that is independent of h_{rad} and has the same law as h_{circ}^{GFF}. Then $h = h_{\text{rad}} + h_{\text{circ}}$ (which since h_{rad} is just a continuous function can again be defined as a stochastic process indexed by $\mathcal{M}$) is called an α-**quantum wedge** in S.*

The conditioned process $\hat{B}$ can be defined as a limit, as $\varepsilon \to 0$, of a (speed 2)-Brownian motion with drift $(Q - \alpha)$, started from $\varepsilon > 0$ and conditioned to stay positive for all time.

To emphasise once more, our definition of quantum wedges is such that they come with a specific normalisation; we will *not* want to consider them as distributions modulo constants.

Remark 5.50. Observe that by the corresponding property of the Neumann GFF, if we restrict the index set of h defined above to $\mathcal{D}_0(S)$, it gives rise to a stochastic process having a version that a.s. defines a distribution in S i.e. an element of $\mathcal{D}_0'(S)$. We can then define the α-quantum wedge as a doubly marked random surface, by letting it be the equivalence class of $(S, h, +\infty, -\infty)$, in the sense of Definition 5.43. Using the change of coordinate formula we could thus also view it as being parametrised by the upper half plane, and we would obtain a distribution $\hat{h}$ defined on $\mathbb{H}$. However the expression for $\hat{h}$ is not particularly nice, and makes the following proofs more difficult to follow, which is why we usually take the strip S as our domain of reference.

Nonetheless, there is an embedding of the wedge in the upper half plane for which the associated field has a nice description in $\mathbb{D}_+$:

Remark 5.51. Note that when $s > 0$, $h_{\text{rad}}(s)$ is a Brownian motion with a drift of coefficient $\alpha - Q < 0$. This means that, embedding in the upper half plane using $z \mapsto -e^{-z}$ and taking into account the conformal change of variables formula, the obtained representative

$(\mathbb{H}, \hat{h}, 0, \infty)$ of the quantum wedge has a logarithmic singularity (for the field $\hat{h}$) of coefficient α near zero. In fact,

$$\hat{h}(z)\big|_{\mathbb{D}_+} \overset{(law)}{=} (h + \alpha \log 1/|z|)\big|_{\mathbb{D}_+},$$

where h has the law of a Neumann GFF in $\mathbb{H}$, normalised so that it has zero average on the semicircle of radius 1.

Remark 5.52 (Unit circle embedding)**.** Suppose that h is a distribution on S of the form $h_{\text{circ}}^{\text{GFF}} + h_r$ where h_r is constant on each vertical segment $\{\Re(z) = s\}$, and these constant values define a continuous function $h_r(s)$ that is positive for all $s \leq s_0$ small enough. Consider the unique translation of the strip so that the image of h_r under this translation hits 0 for the first time at $s = 0$, and let $\tilde{h}$ be the image of h after applying this translation, mapping to $\mathbb{H}$ using the map $z \mapsto -\mathrm{e}^{-z}$ and applying the change of coordinates formula.

If a quantum surface has a representative of the form $(S, h, \infty, -\infty)$ with h as above, then we call $(\mathbb{H}, \tilde{h}, 0, \infty)$ the **unit circle embedding** of this quantum surface.

The unit (semi)circle clearly plays a special role in this embedding since it is the image of the vertical segment with $\Re(z) = 0$ on the strip. Note that if $\hat{h}$ is defined as in Remark 5.51, then $(\mathbb{H}, \hat{h}, 0, \infty)$ is the unit circle embedding of the α-quantum wedge.

We can now state the result about the quantum wedge being the scaling limit of a Neumann GFF with a logarithmic singularity near the origin.

Theorem 5.53. *Fix $0 \leq \alpha < Q$. Then the following hold:*
(i) *Let $\tilde{h}$ be a Neumann GFF in $\mathbb{H}$ with normalisation fixed so that $(\tilde{h}, \rho_0)$ is equal to 0 for some $\rho_0 \in \mathcal{M}^{\mathbb{H}} \setminus \tilde{\mathcal{M}}^{\mathbb{H}}$ that is compactly supported away from the origin, and set $h(z) = \tilde{h}(z) + \alpha \log 1/|z|$. Let h^C be such that $(\mathbb{H}, h^C, 0, \infty)$ is the unit circle embedding of $(\mathbb{H}, h + C, 0, \infty)$, and let $(\mathbb{H}, h^{\text{wedge}}, 0, \infty)$ be the unit circle embedding of a quantum wedge. Then for any $R > 0$, $h^C|_{R\mathbb{D}_+}$ converges in total variation distance to $h^{\text{wedge}}|_{R\mathbb{D}_+}$ as $C \to \infty$.*
(ii) *If $(\mathbb{H}, h^{\text{wedge}}, 0, \infty)$ is an α-quantum wedge, then $(\mathbb{H}, h^{\text{wedge}}, 0, \infty)$ and $(\mathbb{H}, h^{\text{wedge}} + C, 0, \infty)$ have the same law as quantum surfaces.*

Recall that (ii) says a quantum wedge is invariant under rescaling, while (i) says a quantum wedge is the limit, zooming in near zero, of the surface described by $\tilde{h}(z) + \alpha \log 1/|z|$. The fact that the convergence holds in the strong sense of total variation is very useful (as we shall see in the next chapter).

Proof. We start with (i). We embed the field h into the strip S using the conformal map $z \in S \mapsto \phi(z) = -e^{-z} \in \mathbb{H}$, and apply the change of coordinates formula (5.18) to the field. The radial part of the resulting field can be written as the sum of a Gaussian random variable (according to the choice of normalisation) plus a standard two-sided Brownian motion equal to 0 at 0, with drift $\alpha - Q$ ($+\alpha$ coming from the logarithmic singularity of h, and $-Q$ from the change of coordinates). In general the Gaussian random variable and the Brownian motion are not independent. Nevertheless, the radial part will hit $-C/2$ at some finite time almost surely. We can therefore further apply a horizontal translation to the strip so that the new resulting field $h^{C/2}$ has radial part hitting $-C/2$ for the first time at time 0.

By Corollary 5.37 (and scaling) it follows that $h^{C/2}+C/2$ restricted to $S^+ = S \cap \{\Re(z) > 0\}$ converges in total variation distance, as $C \to \infty$, to the field

$$h^{\mathrm{GFF}}_{\mathrm{circ}}|_{S^+} + \mathbf{h}, \tag{5.20}$$

where $\mathbf{h}$ is independent of $h^{\mathrm{GFF}}_{\mathrm{circ}}$, and is constant and equal to $B_{2s}+(\alpha-Q)s$ on each vertical segment $\{\Re(z) = s\}$ (for B a standard Brownian motion with $B_0 = 0$).

Note that the first hitting time s_C of $-C/2$ by the process $(B_{2s} + (\alpha - Q)s)_{s\geq 0}$ tends to infinity almost surely as $C \to \infty$. In particular, for C large enough, it will be $> \log R$ with arbitrarily high probability (recall that for part (i) of the theorem we want to compare $h^C|_{R\mathbb{D}_+}$ and $h^{\mathrm{wedge}}|_{R\mathbb{D}_+}$ for R arbitrary but fixed as $C \to \infty$).

Write $h^{\mathrm{wedge}}_{\mathrm{rad}}$ for the radial part of h^{wedge} and $-s'_C$ for the horizontal coordinate of the leftmost vertical line segment where $h^{\mathrm{wedge}}_{\mathrm{rad}}$ is equal to $C/2$. Since $h^{\mathrm{GFF}}_{\mathrm{circ}}$ has the same law as $h^{\mathrm{wedge}} - h^{\mathrm{wedge}}_{\mathrm{rad}}$ by definition (and $h^{\mathrm{wedge}}_{\mathrm{rad}}$, $h^{\mathrm{wedge}} - h^{\mathrm{wedge}}_{\mathrm{rad}}$ are independent) it suffices for us to show that the processes

- $(B_{2(s+s_C)} + (\alpha - Q)(s + s_C) + C/2)$ for times $s \geq -s_C$; and
- $(h^{\mathrm{wedge}}_{\mathrm{rad}}(\{\Re(z) = s\}))$ for times $s \geq -s'_C$,

can be coupled so that they agree with arbitrarily high probability as $C \to \infty$. This follows because:

- the two processes have the same law for $s \geq 0$;
- the total variation distance between s_C and s'_C tends to 0 as $C \to \infty$;
- conditionally on s_C (resp. s'_C) the time reversal of the top (resp. bottom) process on the interval $[-s_C, 0]$ (resp. $[-s'_C, 0]$) has the law of a Browian bridge from 0 to $C/2$, conditioned to stay positive on this interval (in other words, a 3d-Bessel bridge).

The top and bottom bullet points above follow from the strong Markov property of Brownian motion. It is the middle point that requires some more justification. However, this is a result of the fact that a Brownian motion with positive drift, and a Brownian motion with positive drift conditioned to stay positive, can be coupled so that they agree after time t with arbitrarily high probability as $t \to \infty$.

Point (ii) of the theorem follows immediately, since scaling limits must be invariant under scaling. □

Remark 5.54. Note that the proof above is considerably simplified if we start with a Neumann GFF normalised to have zero average on the upper unit semicircle. Indeed, in its current, general form it is essentially made up of two steps: the first addition of $C/2$ takes us to close to the case of this "nicely normalised" GFF, and the second addition of $C/2$ takes us close to the quantum wedge.

As an example of application of this result, we mention that a quantum wedge field $\hat{h}$ with parameter $\alpha < Q$ has a well-defined Liouville bulk measure $\mu_{\hat{h}}$ and boundary measure $\nu_{\hat{h}}$, since it can be coupled with arbitrarily high probability to a Neumann GFF (with a given logarithmic singularity) plus a constant. Note that these measures are locally finite and atomless almost surely, by the results of Chapter 3 (with base measure σ incorporating the log-singularity).

Hence, we obtain the following strengthening of Theorem 5.53. We emphasise that we are making use of the strong convergence in total variation distance here, which allows us to couple things so that they are actually equal (when restricted to compacts) with high probability. We are also using that for a quantum surface parametrised by $\mathbb{H}$ with marked points at 0 and ∞, as in Example 5.45, the scaling map that determines the canonical parametrisation only depends on the field in a neighourhood of the origin with unit LQG area.

Corollary 5.55. *(i) Suppose that $\tilde{h}, h$ are as in Theorem 5.53. If $(\mathbb{H}, h_C, 0, \infty)$ is the canonical description of $(\mathbb{H}, h + C, 0, \infty)$ and $(\mathbb{H}, \hat{h}, 0, \infty)$ is the canonical description of an α-quantum wedge, then for any $R > 0$, $h_C|_{R\mathbb{D}_+} \to \hat{h}|_{R\mathbb{D}_+}$ in total variation distance as $C \to \infty$.*
(ii) Let $\mu_{h_C}, \mu_{\hat{h}}$ be the respective Liouville measures of $h_C, \hat{h}$ as in (i). Then for any $R > 0$ $\mu_{h_C}|_{R\mathbb{D}_+} \to \mu_{\hat{h}}|_{R\mathbb{D}_+}$ in total variation distance as $C \to \infty$.

We remark that Definition (ii) above was actually used in some of the earlier work of Sheffield, for example in [She16a], as the definition of convergence for quantum surfaces.

5.8 Exercises

1. *This exercise provides an alternative approach to the whole plane GFF than that presented in Section 5.4. In particular, it establishes a relationship with the Dirichlet GFF through its odd part.*

   Consider the Hilbert space completion $(\bar{H}_{\mathbb{C}}, (\cdot,\cdot)_\nabla)$ of the set of smooth functions modulo constants in $\mathbb{C}$ with finite Dirichlet norm.

   Let
   $$\bar{H}_{\text{even}} = \{h \in \bar{H}_{\mathbb{C}} : h(z) - h(0) = h(\bar{z}) - h(0), z \in \mathbb{C}\}$$
   and likewise let
   $$\bar{H}_{\text{odd}} = \{h \in \bar{H}_{\mathbb{C}} : h(z) - h(0) = -(h(\bar{z}) - h(0)), z \in \mathbb{C}\}$$
   (note that $h(z) - h(0)$ is well defined for a function modulo constants). Show that $\bar{H}_{\mathbb{C}} = \bar{H}_{\text{even}} \oplus \bar{H}_{\text{odd}}$. (Hint: orthogonality follows from the change of variables $z \mapsto \bar{z}$).

   The whole plane GFF is then the "standard Gaussian" associated to this Hilbert space:
   $$h^\infty = \sum_n X_n \bar{f}_n,$$

where X_n are i.i.d. standard normal random variables, and $\bar{f}_n$ is an orthonormal basis of $\bar{H}_{\mathbb{C}}$. Show that this series converges almost surely in the space $\mathcal{D}'_0(\mathbb{C})$, and if

$$h^\infty = h^\infty_{\text{even}} + h^\infty_{\text{odd}}$$

is decomposed into even and odd parts, then $h^\infty_{\text{even}}|_{\mathbb{H}}$ is $1/\sqrt{2}$ times a Neumann GFF in $\mathbb{H}$, and that $h^\infty_{\text{odd}}|_{\mathbb{H}}$ is (modulo additive constant) $1/\sqrt{2}$ times a Dirichlet GFF in $\mathbb{H}$.

2. Suppose $\rho \in \tilde{\mathcal{D}}_0(D)$ is viewed as an element of the Hilbertian dual of $\bar{H}^1(D)$ via the mapping $\rho : f \mapsto \int f(x)\rho(x)$ for $f \in \bar{\mathcal{D}}(D)$. Let g_ρ be the unique element in $\bar{H}^1(D)$ (which exists by the Riesz-representation theorem) such that $\rho(f) = (g_\rho, f)_\nabla$ for all $f \in \bar{H}^1(D)$. Show that if G is any valid choice of covariance function for the Neumann GFF modulo constants in D, then

$$(\rho, \rho)_G := \int \rho(x)G(x,y)\rho(y)\,dxdy = (g_\rho, g_\rho)_\nabla = \|\rho\|_{(\bar{H}^1(D))'}.$$

Deduce that $\tilde{\mathcal{M}}$ is a subset of the Hilbert space completion of $\tilde{\mathcal{D}}_0(D)$ with respect to $(\cdot,\cdot)_G$.

3. Prove (5.16) (see the proof of Theorem 2.8). Check that the boundary Liouville measure ν satisfies the same KPZ relation as the bulk Liouville measure.

4. Suppose we sample a point x from the boundary Liouville measure ν (restricted to (0,1) and renormalised so that it is a probability distribution). Is the point x thick for the field? If so, how thick?

5. Show that Theorem 5.53 (i) remains true if we replace h by $h = \tilde{h} + \alpha \log(1/|\cdot|) + \varphi$, where $\tilde{h}$ is a Neumann GFF on $\mathbb{H}$ with some normalisation and φ is a function which is independent of $\tilde{h}$ and continuous at 0: that is, show that if h is as above then as $C \to \infty$, the surfaces $(\mathbb{H}, h + C, 0, \infty)$ converge to an α-thick wedge.

6. Let $D = \{z : \arg(z) \in [0, \theta]\}$ be the (Euclidean) wedge of angle θ, and suppose that $\theta \in (0, 2\pi)$. Let h be a Neumann GFF in D. Show that by zooming in (D, h) near the tip of the wedge, we obtain a thick quantum wedge with $\alpha = Q(\theta/\pi - 1)$ (which satisfies $\alpha < Q$ if $\theta < 2\pi$).

7. (Proposition 4.7 in [DMS14]): show the following characterisation of quantum wedges. Fix $\alpha < Q$ and suppose that h is a fixed representative of a quantum surface parametrised by $\mathbb{H}$. Suppose that the following hold:

   (i) The law of $(\mathbb{H}, h, 0, \infty)$ (as a quantum surface with two marked points 0 and ∞) is invariant under the operation of multiplying its area by a constant. That is, if we fix $C \in \mathbb{R}$, then $(\mathbb{H}, h + C/\gamma, 0, \infty)$ has the same law as $(\mathbb{H}, h, 0, \infty)$.

   (ii) The total variation distance between the law of h restricted to $B(0, r)$ and the law of an α-quantum wedge field h_{wedge} (in its unit circle embedding in $\mathbb{H}$) restricted to $B(0, r)$ tends to 0 as $r \to 0$.

Then $(\mathbb{H}, h, 0, \infty)$ has the law of an α-quantum wedge; more precisely h has the law of h_{wedge}.

8. (**Quantum cones.**) By replacing the strip $S = \mathbb{R} \times (0, \pi)$ with the infinite cylinder $\bar{S} = \mathbb{R} \times (0, \pi) \mod \pi$ (so the top and bottom boundaries of the strip are identified, define for all $\alpha < Q$ a quantum surface $(\bar{S}, h)$ such that Theorem 5.53 holds, with the Neumann GFF h in that theorem replaced by a whole plane GFF h on $\mathbb{C}$. Such a surface is called a (thick) quantum cone [DMS14].

   State and prove the analogue of Exercise 6.

6 SLE and the quantum zipper

In this section we discuss some fundamental results due to Sheffield [She16a], which have the following flavour.

1. Theorem 6.1: An SLE_κ curve has a 'nice' coupling with $e^{\gamma h}$, when h is a certain variant of the Neumann GFF. This coupling can be formulated as a Markov property analogous to the domain Markov properties inherent to random maps. It makes the conjectures about convergence of random maps toward Liouville quantum gravity plausible, and in particular justifies that the "right" relationship between κ and γ is $\kappa = \gamma^2$.

2. Theorem 6.9: An SLE_κ curve can be endowed with a random measure which can roughly be interpreted as $e^{\gamma h} d\lambda$ for $d\lambda$ a natural length measure on the curve. (In fact, the measure $d\lambda$ is in itself hard to define, and the exponent γ needs to be changed slightly from $\sqrt{\kappa}$ to take into account the quantum scaling exponent of the SLE curve - see [BSS14] for a discussion - so we will not actually take this route to define the measure. We will instead use the notion of quantum boundary length. This has the advantage that measures on *either side* of the SLE_κ curve can be defined without difficulty, but we will have to do a fair bit of work to show that they are the same).

3. Theorem 6.31: An SLE_κ curve slits the upper half plane into two independent random surfaces, glued according to boundary length. Thus, SLE curves are solutions of natural random *conformal welding problems*. In fact, the existence of such solutions from a complex analytic view point is a highly non-trivial problem.

⋆ *We collect some relevant background material on SLE in Appendix A.1. Readers unfamiliar with the theory may wish to refer to this now.*

6.1 SLE and GFF coupling; domain Markov property

Here we describe one of the two couplings between the GFF and SLE. This was first stated in the context of Liouville quantum gravity (although presented slightly differently from here) in [She16a]. However, ideas for a related coupling go back to two seminal papers by Schramm and Sheffield [SS13] on the one hand, and Dubédat [Dub09] on the other.

Notational remark: Unless stated otherwise, in what follows the use of bars (e.g., $\bar{h}$) indicates a distribution that is considered modulo constants.

Let $\bar{h}$ be a Neumann GFF on $\mathbb{H}$ (viewed modulo constants), let $\kappa = \gamma^2 \in (0, 4)$, and set

$$\bar{h}_0 = \bar{h} + \varphi \qquad \text{where } \varphi(z) = \frac{2}{\gamma} \log |z|; \quad z \in \mathbb{H}. \tag{6.1}$$

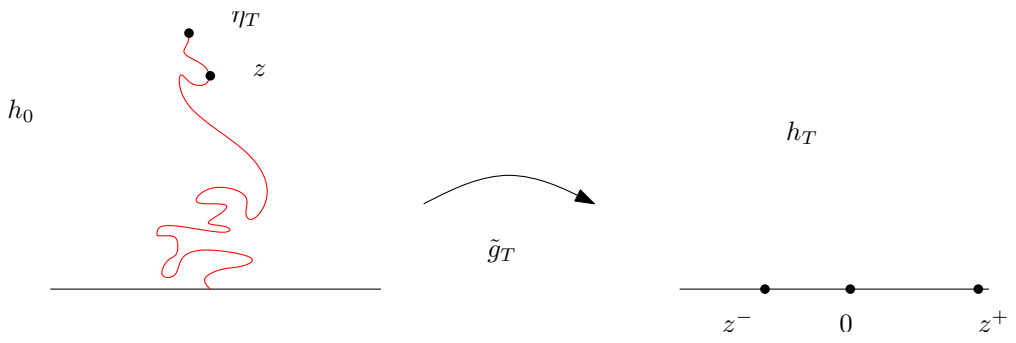


Figure 14: Start with the field $\bar{h}_0$ and an independent SLE_κ curve run up to some time T. After mapping $\bar{h}_0$, restricted to the complement of the curve H_T, by the Loewner map $\tilde{g}_T$ and applying the change of coordinate formula, we obtain a distribution modulo constants $\bar{h}_T$ in $\mathbb{H}$ which by the theorem has the same law as $\bar{h}_0$. This is a form of Markov property for random surfaces.

Hence $\bar{h}_0$ is a Neumann GFF from which we have *subtracted* (rather than added) a logarithmic singularity at zero. (The reason for the choice of multiple $2/\gamma$ will become clear only gradually.)

Let $\eta = (\eta_t)_{t\geq 0}$ be an independent chordal SLE_κ curve in $\mathbb{H}$, going from 0 to ∞ and parameterised by half-plane capacity, where $\kappa = \gamma^2$. Let g_t be the unique conformal map $g_t : \mathbb{H} \setminus \{\eta_s\}_{s\leq t} \to \mathbb{H}$ such that $g_t(z) = z + 2t/z + o(1/z)$ as $z \to \infty$ (we will call g_t the Loewner map). Then

$$\frac{dg_t(z)}{dt} = \frac{2}{g_t(z) - \xi_t}; \qquad z \notin \{\eta_s\}_{s\leq t}$$

where $(\xi_t)_{t\geq 0}$ is the Loewner driving function of η, and has the law of $\sqrt{\kappa}$ times a standard one-dimensional Brownian motion. Let $\tilde{g}_t(z) = g_t(z) - \xi_t$ be the *centred* Loewner map.

Theorem 6.1. *Let $T > 0$ be deterministic, and set*

$$\bar{h}_T = \bar{h}_0 \circ \tilde{g}_T^{-1} + Q \log |(\tilde{g}_T^{-1})'|, \text{ where } Q = \frac{2}{\gamma} + \frac{\gamma}{2}.$$

Then $\bar{h}_T$ defines a distribution in $\mathbb{H}$ modulo constants which has the same law as $\bar{h}_0$.

Remark 6.2. Here we have started with a field $\bar{h}_0$ with a certain law (described in (6.1)) and a curve η which is *independent* of $\bar{h}_0$. However, η is *not* independent of $\bar{h}_T$. In fact, we will see later on that $\bar{h}_T$ entirely *determines* the curve $(\eta_s)_{0\leq s\leq T}$. More precisely, we will see in Theorem 6.9 that when we apply the map $\tilde{g}_T$ to the curve $(\eta_s)_{0\leq s\leq T}$ the boundary lengths (measured with $\bar{h}_T$) of the two intervals to which η is mapped by $\tilde{g}_T$ must agree: that is, on Figure 14, the γ-quantum lengths (with respect to $\bar{h}_T$) of $[z^-, 0]$ and $[0, z^+]$ are the same. (Note that these quantum lengths are only defined up to a multiplicative constant but their ratio is well defined, so this statement makes sense.) Then, in Theorem 6.29, we will show

that given $\bar{h}_T$, the curve $(\eta_s)_{0\le s\le T}$ is determined by the requirement that $\tilde{g}_T^{-1}$ maps intervals of equal quantum length to identical pieces of the curve η. This is the idea of **conformal welding** (we are welding $\mathbb{H}$ to itself by welding together pieces of the positive and negative real line that have the same quantum length).

Remark 6.3. Suppose that instead of starting with $\bar{h}_0$, viewed modulo constants, we took h_0 to be an equivalence class representative of $\bar{h}_0$ with additive constant fixed in some arbitrary way (for example, so that $(h_0, \rho_0) = 0$ for some deterministic $\rho_0 \in \mathcal{M}$ with $\int \rho_0 = 1$). Then $h_T := h_0 \circ \tilde{g}_T^{-1} + Q \log |(\tilde{g}_T^{-1})'|$ would be such that

$$h_T - (h_T, \rho_0) \overset{\text{(law)}}{=} h_0 \quad \text{as distributions.}$$

In other words, the laws of h_T and h_0 would differ by a random constant.

Remark 6.4. The proof of the theorem (and the statement which can be found in Sheffield's paper [She16a, Theorem 1.2]), involves the (centred) reverse Loewner flow f_t rather than, for a fixed t, the map $\tilde{g}_t^{-1}$. In this context, the theorem is equivalent to saying that

$$\bar{h}_T = \bar{h}_0 \circ f_T + Q \log |f_T'|, \text{ where } Q = \frac{2}{\gamma} + \frac{\gamma}{2}.$$

Moreover, in this case the theorem is also true if T is a bounded stopping time (for the underlying reverse Loewner flow). The current formulation of Theorem 6.1 has been chosen because the usual forward Loewner flow is a simpler object and more natural in the context of the Markovian interpretation discussed below. On the other hand, the formulation in terms of the reverse flow will be the most useful when we actually come to prove things in this section.

Discussion and interpretation. Let $H_T = \mathbb{H} \setminus \{\eta_s\}_{0\le s\le T}$ and let h_0 be an equivalence class representative of $\bar{h}_0$ (defined by (6.1)), with additive constant fixed in some arbitrary way. In the language of random surfaces, Theorem 6.1 (more precisely, Remark 6.3) states that the random surface $(H_T, h_0|_{H_T}, \eta(T), \infty)$ has the same distribution, *up to multiplying areas by a random constant*, as $(\mathbb{H}, h_0, 0, \infty)$. This is because h_T is precisely obtained from h_0 by mapping its restriction to H_T through the centered Loewner map $\tilde{g}_T$ and applying the change of coordinate formula. The meaning of "up to multiplying areas by a random constant" corresponds to the fact that the laws of h_T and h_0 differ by a random constant: see Remark 6.3.

To rephrase the above, suppose we start with a surface described by $(\mathbb{H}, h_0, 0, \infty)$. Then we explore a small portion of it using an independent SLE_κ, started where the logarithmic singularity of the field is located (here it is important to assume that γ and κ are related by $\kappa = \gamma^2$). In this exploration, what is the law of the surface that remains to be discovered after some time T? The theorem states that, after zooming in or out by a random amount[8], this

[8]Recall from Section 5.6 that we can view the addition of a constant to the field describing a random surface, equivalently multiplying the area measure for the random surface by a constant, as "zooming" in or out of the surface.

law is the same as the original one. Hence the theorem can be seen as a **domain Markov property for Liouville quantum gravity**.

The fact that this invariance only holds up to additive constants for the field/multiplicative constants for the area measure, is because the Neumann GFF is only really uniquely defined modulo constants. *A more natural result comes if one replaces the Neumann GFF by a quantum wedge*, which is scale invariant by definition (meaning that if one adds a constant to the field, its law as a quantum surface does not change). In this context, we have a similar Markov property, but only if the exploration is stopped when the quantum boundary length of the curve reaches a given value: see Theorem 6.9 and Theorem 6.15. Of course at the moment, however, we do not even know that the quantum boundary length of SLE is well defined - this will be addressed in Section 6.2.

Connection with the discrete picture. This Markov property is to be expected from the discrete side of the story. To see this, consider for instance the uniform infinite half plane triangulation (UIHPT) constructed by Angel and Curien [AS03, Ang03, AC15]. This is obtained as the local limit of a uniform planar map with a large number of faces and a large boundary, rooted at a uniform edge along the boundary. One can further add a critical site percolation process on this map by colouring vertices black or white independently with probability $1/2$ (as shown by Angel, this is indeed the critical value). We make an exception for vertices along the boundary, where those to the left of the root edge are coloured in black, and those to the right in white. This generates an interface and it is possible to use that interface to discover the map. Such a procedure is called *peeling* and was used with great efficacy by Angel and Curien [AC15] to study critical percolation on the UIHPT. The important point for us is that conditionally on the map being discovered up to a certain point using this peeling procedure, it is straightforward to see that the rest of the surface that remains to be discovered also has the law of the UIHPT. An analogue also exists for FK models with $q \in (0, 4)$ in place of critical percolation.

This suggests that a nice coupling between the GFF and SLE should exist, recalling the discussion of Section 4.2. However, identifying the exact analogue in the continuum requires a little thought. First, observe that if one embeds the UIHPT into the upper half plane with the distinguished root edge sent to 0, there is a freedom in how the upper half plane is scaled. Roughly, it can be specified how many triangles should be mapped into the upper unit semi-disc. The natural scaling limit to consider is then the one that arises by letting this number of triangles go to infinity, and rescaling the counting measure on faces appropriately. Note that such a scaling limit will be a "scale invariant" random surface by definition. Indeed, it is expected to be the ($\gamma = \sqrt{8/3}$) LQG measure associated with a certain quantum wedge.

In fact, it is known that in the abstract "Gromov–Hausdorff–Prokhorov topology", the UIHPQ[9] equipped with its natural area measure converges under the rescaling described above to a metric measure space known as the Brownian half-plane [BMR19, GM17b]. Furthermore, the aforementioned quantum wedge can be equipped with metric in such a way that it agrees in law with the Brownian half plane as a metric measure space. Conjectures also hold for other models of maps, and correspondingly, for wedges associated with different

[9] quandrangulation rather than triangulation here

values of γ. This explains (arguably) why the most natural Markov property is actually the one that holds for quantum wedges.

Proof of Theorem 6.1. First, the idea is to use the *reverse Loewner flow* rather than the ordinary Loewner flow $g_t(z)$ and its centered version $\tilde{g}_t(z) = g_t(z) - \xi_t$. Recall that while $\tilde{g}_t(z) : H_t \to \mathbb{H}$ satisfies the SDE:

$$d\tilde{g}_t(z) = \frac{2}{\tilde{g}_t(z)}dt - d\xi_t,$$

in contrast, the reverse Loewner flow is the map $f_t : \mathbb{H} \to H_t := f_t(\mathbb{H})$ defined by the SDE:

$$df_t(z) = -\frac{2}{f_t(z)}dt - d\xi_t.$$

Note the change of signs in the dt term, which corresponds to a change in the direction of time. This Loewner flow is building the curve from the ground up rather than from the tip. More precisely, in the ordinary (forward) Loewner flow, an unusual increment for $d\xi_t$ will be reflected in an unusual behaviour of the curve near its tip at time t. But in the reverse Loewner flow, this increment is reflected in an unusual behaviour near the origin. Furthermore, by using the fact that for any fixed time $T > 0$, the process $(\xi_T - \xi_{T-t})_{0 \le t \le T}$ is a Brownian motion with variance κ run for time $T > 0$, the reader can check that $f_T = \tilde{g}_T^{-1}$ in distribution. Note that this is not necessarily true if T is a stopping time: we will see an example of this later on.

Lemma 6.5. *Suppose that $\gamma \in (0, 2)$ and $\kappa > 0$ is arbitrary. For $z \in \mathbb{H}$, let*

$$M_t = M_t(z) := \frac{2}{\sqrt{\kappa}} \log |f_t(z)| + Q \log |f_t'(z)|; \qquad Q = \frac{2}{\gamma} + \frac{\gamma}{2}.$$

Then for any fixed z, $(M_t(z); t \ge 0)$ is a continuous local martingale (with respect to the filtration generated by ξ) if and only if $\kappa = \gamma^2$. Furthermore, if $z, w \in \mathbb{H}$, then the quadratic cross variation between $M(z)$ and $M(w)$ satisfies

$$d[M(z), M(w)]_t = 4\Re(\frac{1}{f_t(z)})\Re(\frac{1}{f_t(w)})dt$$

Proof. Set $Z_t = f_t(z)$. Then $dZ_t = -2/Z_t dt - d\xi_t$. Set $M_t^* = \frac{2}{\sqrt{\kappa}} \log f_t(z) + Q \log f_t'(z)$, so that $M_t = \Re(M_t^*)$. Applying Itô's formula we see that

$$\begin{aligned} d \log Z_t &= \frac{dZ_t}{Z_t} - \frac{1}{2}\frac{d[\xi]_t}{Z_t^2} \\ &= -\frac{d\xi_t}{Z_t} + \frac{1}{Z_t^2}(-2 - \kappa/2)dt. \end{aligned}$$

To obtain $df_t'(z)$ we differentiate $df_t(z)$ with respect to z. We find that

$$df_t'(z) = 2\frac{f_t'(z)}{Z_t^2}dt,$$

and therefore

$$d\log f_t'(z) = \frac{df_t'(z)}{f_t'(z)} = \frac{2}{Z_t^2}dt.$$

Putting the two pieces together we find that

$$dM_t^* = -\frac{2d\xi_t}{\sqrt{\kappa}Z_t} + \frac{1}{Z_t^2}\left(\frac{2}{\sqrt{\kappa}}(-2-\kappa/2)+2Q\right)dt. \tag{6.2}$$

The dt term vanishes if and only if $2/\sqrt{\kappa}+\sqrt{\kappa}/2 = Q$. Clearly this happens if and only if $\gamma = \sqrt{\kappa}$ given the range of these two parameters.

Furthermore, taking the real part in (6.2), if z, w are two points in the upper half plane $\mathbb{H}$, then the quadratic cross variation between $M(z)$ and $M(w)$ is a process which can be identified as

$$d[M(z), M(w)]_t = 4\Re(\frac{1}{f_t(z)})\Re(\frac{1}{f_t(w)})dt,$$

and so Lemma 6.5 follows. □

One elementary but tedious calculation shows that if

$$G_t(z,w) = G^{\mathbb{H}}(f_t(z), f_t(w)) = -\log(|f_t(z) - \overline{f_t(w)}|) - \log(|f_t(z) - f_t(w)|)$$

then $G_t(z,w)$ is a finite variation process (in fact it is non-increasing) and furthermore:

Lemma 6.6. *We have that*

$$dG_t(z,w) = -4\Re(\frac{1}{f_t(z)})\Re(\frac{1}{f_t(w)})dt.$$

In particular, $d[M(z), M(w)]_t = -dG_t(z,w)$.

Proof. This is proved in [She16a, Section 4]. We encourage the reader to skip the proof here, which is included only for completeness. (However, the result itself will be quite important in what follows.)

Set $X_t = f_t(z)$ and $Y_t = f_t(w)$. From the definition of the Neumann Green function,

$$\begin{aligned} dG_t(x,y) &= -d\log(|X_t - \bar{Y}_t|) - d\log(|X_t - Y_t|) \\ &= -\Re(d\log(X_t - \bar{Y}_t)) - \Re(d\log(X_t - Y_t)). \end{aligned}$$

Now, $dX_t = (2/X_t)dt - d\xi_t$ and $dY_t = (2/Y_t) - d\xi_t$ so taking the difference

$$d(X_t - Y_t) = \frac{2}{X_t}dt - \frac{2}{Y_t}dt = 2\frac{Y_t - X_t}{X_tY_t}dt$$

and so

$$d\log(X_t - Y_t) = -\frac{2}{X_tY_t}dt; \quad d\log(X_t - \bar{Y}_t) = -\frac{2}{X_t\bar{Y}_t}dt.$$

Thus we get

$$dG_t(x,y) = -2\Re(\frac{1}{X_t Y_t} + \frac{1}{X_t \bar{Y}_t})dt. \tag{6.3}$$

Now, observe that for all $x, y \in \mathbb{C}$,

$$\frac{1}{xy} + \frac{1}{x\bar{y}} = \frac{\bar{x}\bar{y} + \bar{x}y}{|xy|^2} = \frac{\bar{x}(\bar{y}+y)}{|xy|^2} = \frac{2\Re(y)}{|xy|^2}\bar{x}.$$

Therefore, plugging into (6.3):

$$dG_t(x,y) = -4\frac{\Re(X_t)\Re(Y_t)}{|X_t Y_t|^2} = -4\Re(\frac{1}{X_t})\Re(\frac{1}{Y_t})$$

as desired. □

Equipped with the above two lemmas, we prove Theorem 6.1. Set $\bar{h}_0 = \bar{h} + \varphi = \bar{h} + \frac{2}{\gamma}\log|z|$, and let $(f_t; t \geq 0)$ be an independent reverse Loewner flow as above. Define

$$\bar{h}_t = \bar{h}_0 \circ f_t + Q\log|f_t'|.$$

Then, viewed as a distribution modulo constants, we claim that:

$$\bar{h}_t \text{ has the same distribution as } \bar{h}_0. \tag{6.4}$$

Let ρ be a test function with zero average, so $\rho \in \bar{\mathcal{D}}(\mathbb{H})$. To prove (6.4), it suffices to check that $(\bar{h}_t, \rho)$ is a Gaussian with mean (φ, ρ) and variance as in (5.5), i.e., $\sigma^2 = \int \rho(dz)\rho(dw)G(z,w)$ where $G(z,w)$ is a valid choice of covariance for the Neumann GFF in $\mathbb{H}$.

To do this, we take conditional expectations given $\mathcal{F}_t = \sigma(\xi_s, s \leq t)$, to see that

$$\mathbb{E}(e^{i(\bar{h}_t,\rho)}|\mathcal{F}_t) = e^{i(\frac{2}{\gamma}\log|f_t|+Q\log|f_t'|,\rho)} \times \mathbb{E}(e^{i(\bar{h}\circ f_t,\rho)}|\mathcal{F}_t)$$

Now we evaluate the term in the conditional expectation above. By conformal invariance, $\bar{h} \circ f_t$ is a Neumann GFF (modulo constants) in H_t. The integral $(\bar{h} \circ f_t, \rho)$ is *a priori* an integral over all of $\mathbb{H}$, but since the curve η has a.s. zero Lebesgue measure, we can view it as an integral only over H_t (and note that the restriction of ρ to H_t also has average value zero). Therefore, given $\mathcal{F}_t$, $(\bar{h} \circ f_t, \rho)$ is a Gaussian random variable with mean zero and variance

$$\int_{H_t} G^{H_t}(z,w)\rho(z)\rho(w)dzdw = \int_{\mathbb{H}} G^{\mathbb{H}}(f_t(z), f_t(w))\rho(z)\rho(w)dzdw,$$

where the equality follows by conformal invariance of the Green function. Hence, if we let $M_t(\rho) = \int M_t(z)\rho(z)dz$ we deduce that

$$\mathbb{E}(e^{i(\bar{h}_t,\rho)}|\mathcal{F}_t) = e^{iM_t(\rho)} \times e^{-\frac{1}{2}\int \rho(z)\rho(w)G_t(z,w)dzdw}. \tag{6.5}$$

Moreover, an application of Fubini's theorem (using for instance that $[M(z), M(w)]_t = G(z,w) - G_t(z,w) \leq G(z,w)$ for each t) gives that

$$[M(\rho)]_t = \int \rho(z)\rho(w) d[M(z), M(w)]_t dz dw.$$

and hence by Lemma 6.6,

$$\int \rho(x)\rho(y)G_t(x,y)dxdy = \int \rho(z)\rho(w)G(z,w)dzdw - [M(\rho)]_t.$$

Combining with (6.5) finally implies that

$$\mathbb{E}(e^{i(\bar{h}_t,\rho)}) = e^{-\frac{1}{2}\int \rho(z)\rho(w)G(z,w)dzdw}\mathbb{E}(e^{iM_t(\rho)+\frac{1}{2}[M(\rho)]_t}).$$

To conclude we observe that by Itô's formula, $e^{iM_t(\rho)+\frac{1}{2}[M(\rho)]_t}$ is an exponential local martingale, and it is not hard to see that it is a true martingale (since $[M(\rho)]_t \lesssim \int |\rho(z)||\rho(w)|G(z,w)\, dzdw$, which is finite, for all t). We deduce that the expectation in the right hand side above is equal to $\mathbb{E}(e^{i(M_0,\rho)}) = e^{i(\varphi,\rho)}$, and therefore

$$\mathbb{E}(e^{i(h_t,\rho)}) = e^{-\frac{1}{2}\int \rho(z)\rho(w)G(z,w)dzdw}e^{i(\varphi,\rho)}.$$

This proves (6.4). Arguing that f_t and $\tilde{g}_t^{-1}$ have the same distribution finishes the proof of the theorem. □

Remark 6.7. As mentioned earlier, since the proof relies on martingale computation and the optional stopping theorem, the theorem remains true if T is a (bounded) stopping time for the *reverse* Loewner flow.

Remark 6.8. This martingale is obtained by taking the real part of a certain complex martingale. Taking its imaginary part (in the case of the forward flow) gives rise to the imaginary geometry developed by Miller and Sheffield in a striking series of papers [MS16a, MS16b, MS16c, MS17].

6.2 Quantum length of SLE

We start with one of the main theorems of this section, which allows us, given a chordal SLE_κ curve and an independent Neumann GFF, to define a notion of quantum length of the curve unambiguously. The way this is done is by mapping the curve down to the real line with the centred Loewner map $\tilde{g}_t$, and using the quantum boundary measure ν (associated with the image of the GFF via the change of coordinates formula) to define the length. However, when we map away the curve using the map $\tilde{g}_t$, each point of the curve corresponds to two points on the real line - except for the tip of the curve which is sent to the origin since we consider the centred map. Hence, to measure the length of the curve, we need to know that measuring the length on one side of 0 almost surely gives the same answer as measuring the length on the other side of 0.

This is basically the content of the next theorem. For ease of proof, the theorem is stated in the case where h is not a Neumann GFF but rather the field of a certain wedge. However, we will see (Corollary 6.11) that this is no loss of generality.

Theorem 6.9. *Let $0 < \gamma < 2$ and let $(\mathbb{H}, h, 0, \infty)$ be an α-quantum wedge in the unit circle embedding, with $\alpha = \gamma - 2/\gamma$. Let ζ be an independent SLE_κ with $\kappa = \gamma^2$. Let $\tilde{g}_t$ be the (half-plane capacity parametrised) centred Loewner flow for ζ, and consider the distribution $h_t = h \circ \tilde{g}_t^{-1} + Q \log |(\tilde{g}_t^{-1})'|$ as before. Let ν_{h_t} be the boundary Liouville measure on $\mathbb{R}$ associated with the distribution h_t. Finally, given a point $z \in \zeta([0,t])$, let $z^- < z^+$ be the two images of z under $\tilde{g}_t$. Then*

$$\nu_{h_t}([z^-, 0]) = \nu_{h_t}([0, z^+]),$$

almost surely for all $z \in \zeta([0,t])$.

Remark 6.10. By Remark 6.3 and the fact that the slit domain formed by an SLE_κ with $\kappa < 4$ is a.s. Hölder continuous, we see that a Neumann GFF (with arbitrary normalisation) plus a $(\gamma - 2/\gamma)$ log-singularity in such a slit domain does satisfy the conditions of Definition 5.42. That is, the quantum boundary length on either side of the curve is well-defined by mapping down to the real line. Since a $(\gamma - 2/\gamma)$ quantum wedge in the unit circle embedding has the same law when restricted to $B(0,1) \cap \mathbb{H}$ as such a Neumann GFF (with normalisation fixed so that it has mean value 0 on the upper unit semicircle) this implies that the field h of the above theorem also satisfies the conditions of Definition 5.42, at least when restricted to $B(0,1)$. Scale invariance implies that this holds when the field is restricted to any large disc. In other words, the boundary Liouville measure ν_{h_t} for h_t is well defined.

Corollary 6.11. *Theorem 6.9 is still true when h is replaced by a Neumann GFF on $\mathbb{H}$, with arbitary normalisation. Indeed, by the discussion in the previous remark, it is true until the curve exits the upper unit semidisc, when the normalisation for the GFF is such that it has average 0 on the upper unit semicircle. This extends to arbitrary normalisations, since two Neumann GFFs with different normalisations (can be coupled so that they) differ by a random additive constant. Finally, scaling removes the need to restrict to the unit semidisc.*

Definition 6.12. *The quantity $\nu_{h_t}([\zeta(s)^-, 0]) = \nu_{h_t}([0, \zeta(s)^+])$ is called the **quantum length** of $\zeta([s,t])$ in the wedge $(\mathbb{H}, h, 0, \infty)$.*

False proof of Theorem 6.9. The following argument does not work but helps explain the idea and why wedges are a useful notion. Let ζ be the infinite SLE_κ curve parametrised by half-plane capacity. Let $L(t) = \nu_{h_t}([\zeta(t)^-, 0])$ be the quantum length of left hand side of the curve ζ up to time t (measured by computing the boundary quantum length on the left of zero after applying the map $\tilde{g}_t$) and likewise, let $R(t)$ be the quantum length of the right-hand side of ζ. Then it is *tempting* (but wrong) to think that, because SLE is stationary via the domain Markov property, and the Neumann GFF is invariant by Theorem 6.1, $L(t)$ and $R(t)$ form processes with stationary increments. If that were the case, we would conclude from Birkhoff's ergodic theorem for stationary increments processes that $L(t)/t$ converges almost surely to a possibly random constant, and $R(t)/t$ converges also to a random constant. We would deduce that $L(t)/R(t)$ converges to a possibly random constant. Finally, we would argue that this constant cannot be random because of tail triviality of SLE (i.e., of driving

Brownian motion) and in fact must be one by left-right symmetry. On the other hand by scale invariance, the distribution of $L(t)/R(t)$ is constant. Hence we would deduce that $L(t) = R(t)$.

This proof is wrong on at least two counts: first of all, it is not true that $L(t)$ and $R(t)$ have stationary increments. This does not hold, for instance, because h loses its stationarity (i.e., the relation $h_T = h_0$ in distribution does not hold) as soon as a normalisation is fixed for the Neumann GFF. Likewise the scale invariance does not hold in this case. This explains the importance of the concept of wedges, for which scale invariance holds by definition, as well as a certain form of stationarity (see Theorem 6.15). These properties allow us to make the above proof rigorous.

6.3 Proof of Theorem 6.9

Essential to the proof of Theorem 6.9 is the definition of two stationary processes: the *capacity zipper* and the *quantum zipper*. As in the original paper of Sheffield [She16a], once the existence and stationarity of these processes is proven, Theorem 6.9 follows relatively easily (in fact, using a similar argument to the "false proof" above).

In order to simplify notation in what follows, whenever f is a conformal map and h is a distribution or distribution modulo constants, we write

$$f(h) := h \circ f^{-1} + Q \log |(f^{-1})'|. \tag{6.6}$$

From now on we assume that $\gamma \in (0,2)$ is fixed, $Q = Q_\gamma$, and $\kappa = \gamma^2$. Recall that $\bar{\mathcal{D}}'_0(\mathbb{H})$ denotes the space of distributions modulo constants on $\mathbb{H}$, and we write $C([0,\infty), \mathbb{H})$ for the space of continuous functions from $[0,\infty)$ to $\mathbb{H}$.

Theorem 6.13 (Capacity zipper)**.** *There exists a two-sided* ***stationary*** *process $(\bar{h}^t, \eta^t)_{t\in\mathbb{R}}$, taking values in $\bar{\mathcal{D}}'_0(\mathbb{H}) \times C([0,\infty); \mathbb{H})$, such that:*

- **(Marginal law)** *$(\bar{h}^0, \eta^0)$ has the law of a Neumann GFF (modulo constants) plus the function $\varphi(z) = \frac{2}{\gamma}\log|z|$, together with an independent SLE$_\kappa$;*
- **(Positive time)** *there exists a family of conformal maps $(f_t)_{t\geq 0} : \mathbb{H} \to \mathbb{H} \setminus \eta^t([0,t])$, whose (marginal) law is that of a reverse SLE$_\kappa$ Loewner flow parametrised by capacity, and such that $\bar{h}^t|_{\mathbb{H}\setminus\eta^t([0,t])} = f_t(\bar{h}^0)$ and $\eta^t([t,\infty)) = f_t(\eta^0)$ for all $t \geq 0$;*
- **(Negative time)** *for $t < 0$, if $\tilde{g}_{-t}$ is the centred Loewner map corresponding to $\eta^0([0,-t])$ then $\eta^t = \tilde{g}_{-t}(\eta^0)$ and $\bar{h}^t = \tilde{g}_{-t}(\bar{h}^0)$.*

Thus given a field $\bar{h}^0$ and an independent SLE$_\kappa$ infinite curve η^0, we can either "zip it up" (weld it to itself) to obtain the configuration $(\bar{h}^t, \eta^t)$ for some $t > 0$, or "zip it down" (cut it open along η^0) to obtain the configuration $(\bar{h}^t, \eta^t)$ for some $t < 0$. See Figure 15. Beware that the relation between time t and time 0 is opposite to that of Theorem 6.1 - hence the change in notation from subscripts to superscripts for the time index.

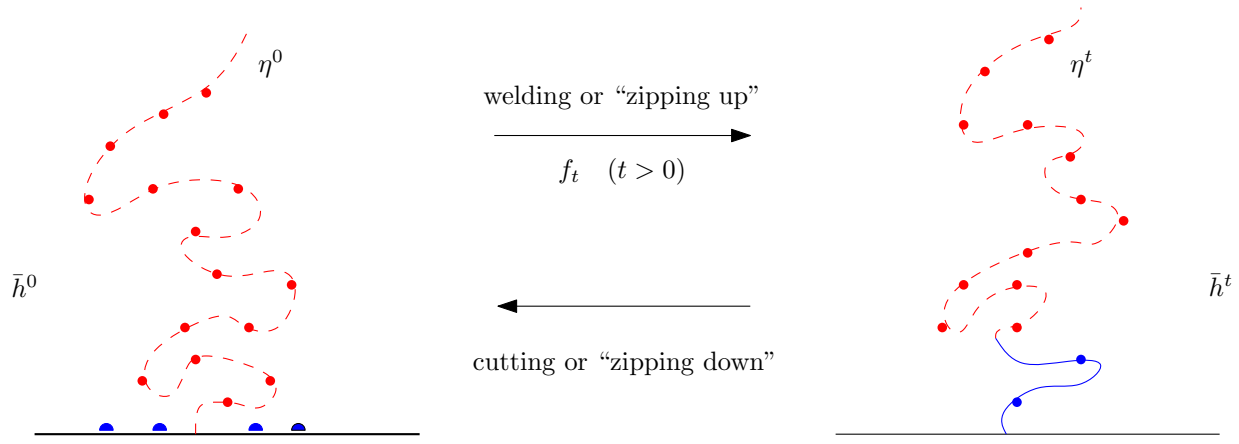


Figure 15: The capacity zipper

Also note that for $t > 0$, $\bar{h}^t|_{\mathbb{H}\setminus\eta^t([0,t])}$ uniquely defines $\bar{h}^t$ as a distribution modulo constants on $\mathbb{H}$ (since $\eta^t([0,t])$ is independent of $\bar{h}^t$ and has Lebesgue measure zero). The term "capacity" in the definition refers to the fact that in any positive time t, we are zipping up a curve with $2t$ units of half-plane capacity.

Remark 6.14. Note that the capacity zipper of Theorem 6.13 is defined to be a process taking values in $\bar{\mathcal{D}}'_0(\mathbb{H}) \times C([0,\infty), \mathbb{H})$. However, we can also define from $(\bar{h}^0, \eta^0, (f_t)_{t\geq 0})$ a version $(\tilde{h}^t, \eta^t)_{t\geq 0}$ of the capacity zipper indexed by positive times and taking values in $\mathcal{D}'_0(\mathbb{H}) \times C([0,\infty), \mathbb{H})$. That is, so that the field at any time is a distribution, not just a distribution modulo constants.

To do this, we can just fix a normalisation of $\bar{h}^0$ to obtain $\tilde{h}^0 \in \mathcal{D}'_0(\mathbb{H})$, and then for $t > 0$ set $(\tilde{h}^t, \eta^t) := (f_t(\tilde{h}^0), \mathbb{H} \setminus f_t(\mathbb{H} \setminus \eta^0))$. Note that this process will be no longer stationary: for given t, $\tilde{h}^t$ will have the law of $\tilde{h}^0$ plus a random constant.

Now we move on to the definition of the **quantum zipper**. For this, we need the notion of doubly marked surface-curve pair. This is just an extension of the definition of doubly marked surface, when the surface comes together with a chordal curve. More precisely, suppose that for $i = 1, 2$, D_i is a simply connected domain with marked boundary points (a_i, b_i), h_i is a distribution in D_i, and η_i is a simple curve (considered up to time reparametrisation) from a_i to b_i in D_i. We say that $(D_1, h_1, a_1, b_1, \eta_1)$ and $(D_2, h_2, a_2, b_2, \eta_2)$ are equivalent if there exists a conformal map $f : D_1 \to D_2$ such that $h_2 = f(h_1)$, $a_2 = f(a_1)$, $b_2 = f(b_1)$ and $\eta_2 = f(\eta_1)$. A doubly marked surface-curve pair (from here on in just *surface-curve pair*) is an equivalence class of (D, h, a, b, η) under this equivalence relation.

Theorem 6.15 (Quantum zipper)**.** *There exists a two-sided process*

$$(h^t, \zeta^t)_{t\in\mathbb{R}} = ((\mathbb{H}, h^t, 0, \infty), \zeta^t)_{t\in\mathbb{R}}$$

that is ***stationary*** *as a process of surface-curve pairs, and such that:*

- $(\mathbb{H}, h^0, 0, \infty)$ *is a quantum wedge in the unit circle embedding;*
- (h^0, ζ^0) *has the law, as a surface-curve pair, of a* $(\gamma - 2/\gamma)$ *quantum wedge together with an independent* SLE_κ*;*
- *for any* $t > 0$*, if* ζ^0 *is parametrised by half plane capacity, if*

$$\sigma(t) := \inf\{s \geq 0 : \nu_{h^0}\left(RHS \text{ of } \zeta^0([0,s])\right) \geq t\},$$

and if $\tilde{g}_{\sigma(t)}$ *is the centred Loewner map sending* $\mathbb{H} \setminus \zeta^0[0, \sigma(t)]$ *to* $\mathbb{H}$*, then we have that* $h^{-t} = \tilde{g}_{\sigma(t)}(h^0)$ *and* $\zeta^{-t} = \tilde{g}_{\sigma(t)}(\zeta^0)$*.*

Note that by stationarity, this defines the law of the process for all time (positive and negative).

So this is a similar picture to that of the capacity zipper (moving backwards in time corresponds to "cutting down" and hence moving forward in time corresponds to "zipping up") but now a segment of ζ^0 with *right* h^0 LQG-boundary length t is cut out between times 0 and $-t$. Hence the name "quantum zipper": the dynamic is parametrised by (right) quantum boundary length. Note that it makes sense to talk about the right boundary length of a segment of η, by conformally mapping to the upper half plane and applying the change of coordinate formula (see Remark 6.10). Also note the difference with the capacity zipper: here h^0 is a distribution (not a distribution modulo constants) while the stationarity is in the sense of quantum surface-curve pairs.

Assuming for now that Theorem 6.15 holds, we make the following claim.

Claim 6.16. *For any fixed* t*, the* ν_{h^0}*-boundary length of the* left-hand side *of* $\zeta^0([0, \sigma(t)])$ *is also equal to* t*.*

This means that the parametrisation is really, unambiguously, by quantum boundary length. It also immediately implies Theorem 6.9.

Proof of Claim 6.16, and hence Theorem 6.9, given Theorem 6.15. Denote by $L(t)$ the ν_{h^0}-boundary length of the left-hand side of $\zeta^0[0, \sigma(t)]$, so our aim is to show that $L(t) \equiv t$. We begin by making the following observations.

- By stationarity of the quantum zipper, we have that $(L(s+t) - L(s))_{t \geq 0}$ is equal in distribution to $(L(t))_{t \geq 0}$ for any fixed $s \geq 0$.
- By scale invariance of SLE_κ and the invariance property of quantum wedges (Theorem 5.53), for any $A > 0$ and $s < t$,

$$\frac{L(At)}{At} - \frac{L(As)}{As} \overset{(d)}{=} \frac{L(t)}{t} - \frac{L(s)}{s}.$$

The first point together with the ergodic theorem tells us that

$$\frac{L(n)}{n} \to X \text{ a.s. as } n \to \infty \text{ in } \mathbb{N}, \tag{6.7}$$

where X is some random variable. (Note that the theorem is usually stated under the assumption that $\mathbb{E}(|L(1)|) < \infty$, but it is straightforward to see, by a truncation argument and the monotone convergence theorem, that the conclusion is also true if we only know $L(1) \geq 0$ a.s., which is the case here.)

Together with this, the second point then implies that

$$\frac{L(t)}{t} - \frac{L(s)}{s} = 0 \text{ almost surely}$$

for any fixed $s, t \in \mathbb{Q}$ with $s \leq t$. Indeed, the law m of this difference is equal to that of $L(At)/At - L(As)/As$ for any A, and by taking a sequence $A_k \uparrow \infty$ such that $A_k t \in \mathbb{N}, A_k s \in \mathbb{N}$ for all k, we obtain a sequence of random variables all having law m, which by (6.7) tend to 0 as $k \to \infty$. Hence, with probability one we have that

$$\frac{L(t)}{t} = X \quad \forall t \in \mathbb{Q} \tag{6.8}$$

(where X is as in (6.7)). In particular, we have that

$$X = \lim_{t \downarrow 0, t \in \mathbb{Q}} \frac{L(t)}{t}.$$

Now by definition, the above limit (and therefore the random variable X) is measurable with respect to the σ-algebra

$$\mathcal{T} = \bigcap_{\varepsilon > 0} \sigma((h^0 - h^0_\varepsilon)|_{B(0,\varepsilon) \cap \mathbb{H}}, \zeta^0|_{B(0,\varepsilon) \cap \mathbb{H}})$$

(here h^0_ε is the ε-semi circle average of h^0 about the origin and can be subtracted since $L(t)/t$ is not affected by adding a constant to the field). On the other hand, since the h^0-right/left quantum boundary lengths along ζ^0 a.s. do not have atoms at $0^\pm$, X is also measurable with respect to

$$\sigma(\mathcal{A}) \; ; \; \mathcal{A} = \bigcup_{\varepsilon > 0} \sigma((h^0 - h^0_\varepsilon)|_{B(0,1) \setminus B(0,\varepsilon)}, \zeta^0|_{B(0,1) \setminus B(0,\varepsilon)}).$$

Hence the proof will be complete if we can show $\mathcal{T} \cap \sigma(\mathcal{A})$ is trivial, because then X must be almost surely constant, and by symmetry, this constant must be equal to 1.

For this final step, since $\mathcal{A}$ is a π-system, it suffices to show that for any $\varepsilon_0 > 0$, $A_0 \in \mathcal{T}$ and

$$A \in \sigma(h^0 - h^0_{\varepsilon_0}|_{B(0,1) \setminus B(0,\varepsilon_0)}, \zeta^0|_{B(0,1) \setminus B(0,\varepsilon_0)}),$$

we have $\mathbb{P}(A \cap A_0) = \mathbb{P}(A)\mathbb{P}(A_0)$. However, this follows by independence of h^0 and ζ^0, since the driving function of ζ^0 is a Brownian motion, and by Lemma 5.34.

□

The rest of this section will be dedicated to proving Theorem 6.13 and Theorem 6.15. In fact, Theorem 6.13 is straightforward to obtain from Theorem 6.1. The idea to then deduce Theorem 6.15 is to reparametrise time according to right quantum boundary length and appropriately "zoom in" at the whole capacity zipper picture at the origin. This step, however, is somewhat technical.

6.3.1 The capacity zipper

In this section we prove Theorem 6.13. That is, we construct the stationary two-sided capacity zipper, using the coupling theorem, Theorem 6.1.

Let $\bar{h}_0$ be as in the original Theorem 6.1 (i.e. $\bar{h}_0$ has the distribution (6.1)), and let $\eta = \eta_0$ be an independent infinite SLE_κ curve from 0 to ∞. As in the coupling theorem, set $\bar{h}_t = \bar{h}_0 \circ \tilde{g}_t^{-1} + Q \log |(\tilde{g}_t^{-1})'|$, where $\tilde{g}_t$ is the centered Loewner map corresponding to $\eta_0([0,t])$ for each t, and let η_t be the image by $\tilde{g}_t$ of the initial infinite curve $\eta = \eta_0$. Then Theorem 6.1 says that $\bar{h}_t = \bar{h}_0$ in distribution, and in fact we can also see that the joint distribution $(\bar{h}_t, \eta_t)$ is identical to that of $(\bar{h}_0, \eta_0)$.

For $0 \leq t \leq T$, let $\bar{h}^t = \bar{h}_{T-t}$, and let $\eta^t = \eta_{T-t}$. Then it is an easy consequence of Theorem 6.1 that the following lemma holds:

Lemma 6.17. *The laws of the process $(\bar{h}^t, \eta^t)_{0 \leq t \leq T}$ (with values in $\bar{\mathcal{D}}'(\mathbb{H}) \times C([0,\infty))$ are consistent as T increases.*

By Lemma 6.17, and applying Kolmogorov's extension theorem, it is obvious that there is a well-defined process $(\bar{h}^t, \eta^t)_{0 \leq t < \infty}$ whose restriction to $[0,T]$ agrees with the process described above. Hence for $t > 0$, starting from $\bar{h}^0$ and an infinite curve η^0, there is a well-defined "welding" procedure giving rise to $(\bar{h}^t, \eta^t)$. The dynamic on the field is obtained by applying the change of coordinates formula to $\bar{h}^0$, with respect to a flow $(f_s)_{s \leq t}$ that has the *marginal* law of a reverse Loewner flow, but we stress that here the reverse Loewner flow is not independent of $\bar{h}^0$ (rather, it will end up being uniquely determined by $\bar{h}^0$, while $(f_s)_{s \leq t}$ will be independent of $\bar{h}^t$).

But we could also go in the other direction, cutting $\mathbb{H}$ along η^0, as in Theorem 6.1. Indeed we could define, for $t < 0$ this time, a field $\bar{h}^t$ by considering the centered Loewner flow $(\tilde{g}_{|t|})_{t<0}$ associated to the infinite curve η^0, and setting

$$\bar{h}^t = \bar{h}^0 \circ \tilde{g}_{|t|}^{-1} + Q \log |(\tilde{g}_{|t|}^{-1})'| \quad (t < 0).$$

We can also, of course, get a new curve η^t for $t < 0$ by pushing η^0 through the map $\tilde{g}_{|t|}$. This gives rise to the two-sided stationary process $(\bar{h}^t, \eta^t)_{t \in \mathbb{R}}$ of Theorem 6.15.

Remark 6.18. An equivalent way to define this process would be as follows. Start from the setup of Theorem 6.1: thus $\bar{h}_0$ is a field distributed as in (6.1), and η_0 an independent infinite SLE_κ curve. Set $\bar{h}_t = \bar{h}_0 \circ \tilde{g}_t^{-1} + Q \log |(\tilde{g}_t^{-1})'|$ as before, and $\eta_t = g_t(\eta^0 \setminus \eta_0[0,t])$. Then Theorem 6.1 tells us that $(\bar{h}_t, \eta_t)_{t \geq 0}$ is a stationary process, so we can consider the limit as $t_0 \to \infty$ of $(\bar{h}_{t_0+t}, \eta_{t_0+t})_{t \geq -t_0}$, which defines a two-sided process. The capacity zipper process $(\bar{h}^t, \eta^t)_{t \in \mathbb{R}}$ can then be defined as the image of this process under the time change $t \mapsto -t$.

6.3.2 The quantum zipper

⋆ *We recall the notation*

$$f(h) := h \circ f^{-1} + Q \log |(f^{-1})'|. \tag{6.9}$$

that will be used repeatedly in what follows.

In this section we show the existence and stationarity of the quantum zipper: Theorem 6.15. In what follows, we will usually take our quantum wedges to be in the **unit circle embedding** $(\mathbb{H}, h, 0, \infty)$ (recall that the law of $h - \alpha \log(1/|z|)$ restricted to the upper unit semi-disc is then just that of a Neumann GFF with additive constant fixed so that its average on the upper unit semi-circle is equal to zero).

The key to the proof of Theorem 6.15 is the following:

Proposition 6.19. *Let $(h, \zeta) = ((h, \mathbb{H}, 0, \infty), \zeta)$ be a $(\gamma - 2/\gamma)$ quantum wedge in the unit circle embedding, together with an independent SLE_κ. If ζ is parametrised by half-plane capacity, let σ be the smallest time such that the ν_h boundary length of the right hand side of $\zeta([0, \sigma])]$ exceeds 1*[10]*. Let g_σ be the centered Loewner map from $\mathbb{H} \setminus \zeta([0, \sigma]) \to \mathbb{H}$. Then $(g_\sigma(h), g_\sigma(\zeta))$ is equal in law to (h, ζ) as a surface-curve pair. That is, if ψ is the unique conformal map such that $(\psi \circ g_\sigma)(h)$ is in the unit circle embedding, then*

$$(\psi \circ g_\sigma(h), \psi \circ g_\sigma(\zeta)) \overset{(d)}{=} (h, \eta).$$

In words: if we start with a $(\gamma - 2/\gamma)$ quantum wedge and an independent SLE_κ, and "zip" down by one unit of right quantum boundary length, the law of the resulting quantum surface-curve pair does not change.

Proof of Theorem 6.15 given Proposition 6.19. Note that there is nothing special about the choice to zip down by quantum boundary length one in Proposition 6.19. Indeed we could replace one by any other $t > 0$ and would obtain the result. Then the existence and stationarity of the quantum zipper follows in the same way that Theorem 6.13 followed from Theorem 6.1 (see the previous section). □

The proof of Proposition 6.19 is quite tricky, and consists of several steps.

Step 1: Reweighting We write $\mathbb{P}$ for the law of $(\tilde{h}^t, \eta^t)_{t \geq 0}$, the capacity zipper as in Remark 6.14, where the constant for $\tilde{h}^0$ has been fixed so that its unit semicircle average around the point 10 is equal to 0 (this is fairly arbitrary, apart from the fact that the measure is supported a good distance away from the origin). We can extend this to define a law $\mathbf{P}$ on $\mathcal{D}'(\mathbb{H}) \times C([0, \infty), \mathbb{H}) \times [1, 2]$, by setting $\mathbf{P} := \mathbb{P} \times \mathrm{Leb}_{[1,2]}$ (so a sample from $\mathbf{P}$ consists

[10] Recall that to measure this boundary length, we map the right hand side of the curve down to an interval $[0, x]$ of the positive real line using the centred Loewner map. Then we take the quantum boundary length of $[0, x]$ with respect to the field defined by applying the change of coordinates formula to h with respect to this map.

of a capacity zipper $(\tilde{h}^t, \eta^t)_{t\geq 0}$ as just described, plus a point Z chosen independently from Lebesgue measure on $[1,2]$). Define

$$c(z) := \mathbb{E}_{\mathbf{P}}(e^{\frac{\gamma}{2}\tilde{h}^0_\delta(z)}\delta^{\gamma^2/4}) \text{ for } z \in [1,2],$$

which by Theorem 5.36 does not depend on $\delta > 0$ and is a smooth function on $z \in [1,2]$.

We want to study the joint law of the capacity zipper plus a quantum boundary length typical point (in $[1,2]$). In fact, this is much easier do if we reweight the law of the field $\tilde{h}^0$. To this end, we define a family of laws $(\mathbf{Q}_\varepsilon)_{\varepsilon>0}$ by setting

$$\frac{d\mathbf{Q}_\varepsilon}{d\mathbf{P}} = \frac{\mathrm{e}^{\frac{\gamma}{2}\tilde{h}^0_\varepsilon(Z)}\mathcal{E}^{\frac{\gamma^2}{4}}}{\int_{[1,2]} c(z)\,dz} =: \frac{\mathrm{e}^{\frac{\gamma}{2}\tilde{h}^0_\varepsilon(Z)}\mathcal{E}^{\frac{\gamma^2}{4}}}{c([1,2])} \tag{6.10}$$

for each ε.

Under $\mathbf{Q}_\varepsilon$, the marginal law of $(\tilde{h}^t, \eta^t)_{t\geq 0}$ is its $\mathbb{P}$ law weighted by

$$\frac{\nu_{\tilde{h}^0_\varepsilon}([1,2])}{c([1,2])}.$$

Moreover, given $(\tilde{h}^t, \eta^t)_{t\geq 0}$ the point Z is sampled from the ε-approximate measure $\nu_{\tilde{h}^0_\varepsilon}$ (restricted to $[1,2]$ and normalised to be a probability measure). Therefore, since $\nu_{\tilde{h}^0_\varepsilon}([1,2]) \to \nu_{\tilde{h}^0}([1,2])$ in $\mathcal{L}^1$ as $\varepsilon \to 0$ and the measure $\nu_{\tilde{h}^0_\varepsilon}$ converges weakly in probability to $\nu_{\tilde{h}^0}$, we can deduce that

$$\mathbf{Q}_\varepsilon \Rightarrow \mathbf{Q}$$

as $\varepsilon \to 0$ where $\mathbf{Q}$ is the unique measure satisfying (a) and (b) of Lemma 6.20 below.

This reweighting is analogous to the argument used to describe the GFF viewed from a Liouville typical point – see Theorem 2.4. As in this proof, we can reverse the order in which $(\tilde{h}^0, \eta^0)$ and Z are sampled, and this leads to the alternative description given by points (c) to (e) in following lemma.

Lemma 6.20. *Under* $\mathbf{Q}$*, the following is true:*

(a) the marginal law of $(\tilde{h}^t, \eta^t)_{t\in\mathbb{R}}$ *is given by* $\nu_{\tilde{h}^0}([1,2])/c([1,2])d\mathbb{P}$ *(and is therefore absolutely continuous with respect to* $\mathbb{P}$*);*

(b) conditionally on $(\tilde{h}^t, \eta^t)_{t\in\mathbb{R}}$*,* Z *is chosen uniformly from* $\nu_{\tilde{h}^0}$ *on* $[1,2]$*;*

(c) the marginal law of Z *on* $[1,2]$ *has density* $c(z)/c([1,2])$ *with respect to Lebesgue measure;*

(d) conditionally on Z*, for every* $0 \leq t \leq \tau_Z$ *(where* τ_Z *is the first time that* $f_t(Z) = 0$*) the law of* $\eta^t([0,t])$ *is that of a reverse* $SLE_\kappa(\kappa, -\kappa)$ *curve with force points* $(Z, 10)$*, run up to time* t*;*

(e) *conditionally on Z, for every $0 \le t \le \tau_Z$, we can write (as an equality of distributions modulo constants)*

$$\tilde{h}^t \overset{(d)}{=} \bar{h} + \frac{2}{\gamma}\log(|\cdot|) + \frac{\gamma}{2}G^{\mathbb{H}}(\cdot, f_t(Z)) - \frac{\gamma}{2}\int G^{\mathbb{H}}(\cdot, f_t(y))\rho_{10,1}(dy)$$

where $\bar{h}$ has the law of a Neumann GFF (modulo constants) that is independent of $(f_s)_{0\le s\le t}$. Here for $x \in \mathbb{R}, \delta > 0$, $\rho_{x,\delta}$ denotes uniform measure on the upper semicircle of radius δ around x.

Remark 6.21. The force point at 10 in (d) and the final term in the expression for $\tilde{h}^t$ in (e) make these descriptions look rather complicated. However, we will really be interested in taking $t = \tau_Z$ and looking at $(\tilde{h}^t, \eta^t)$ in small neighbourhoods of the origin. In such a setting, as we will soon see, these terms will have asymptotically negligible contribution to the behaviour. The only features in the descriptions (d) and (e) that are genuinely important, are the force point of weight κ at Z, and the function $(2/\gamma)\log(|\cdot|) + (\gamma/2)G(\cdot, f_t(Z))$.

Proof. (a) and (b) define the measure $\mathbf{Q}$ (see discussion above the lemma) and (c) follows since this is true under $\mathbf{Q}_\varepsilon$ for every $\varepsilon > 0$.

For (d), we first claim that for any $t \ge 0$ and for any measurable function F of $(f_s; s \le t)$ we have

$$\mathbf{Q}_\varepsilon(F(f_s; s \le t)\mathbf{1}_{\{t\le\tau_{Z-\varepsilon}\}}) = \mathbf{P}(F(f_s; s \le t)\mathbf{1}_{\{t\le\tau_{Z-\varepsilon}\}}\mathrm{e}^{\frac{\gamma}{2}(M_t(Z)-M_t(10))-\frac{\gamma^2}{8}[M(Z)-M(10)]_t}). \quad (6.11)$$

To see this, we note that by definition $\tilde{h}^0 = \tilde{h}^t \circ f_t + Q\log|(f_t)'|$ and, due to the normalisation we chose for $\tilde{h}^0$, $\tilde{h}^0_\varepsilon = (\tilde{h}^0, \bar{\rho}^\varepsilon_Z) := (\tilde{h}^0, \rho_{Z,\varepsilon} - \rho_{10,1})$. Therefore

$$\mathbf{P}(e^{\frac{\gamma}{2}\tilde{h}^\varepsilon_0(Z)} \mid Z, (f_s; s \le t)) = e^{\frac{\gamma}{2}(M_t, \rho_{Z,\varepsilon}-\rho_{10,1})}\mathbf{P}(e^{(\tilde{h}^t\circ f_t - (2/\gamma)\log(|\cdot|)\circ f_t, \rho_{Z,\varepsilon}-\rho_{10,1})} \mid Z, (f_s; s \le t))$$

where, because the average value of $\rho_{Z,\varepsilon}-\rho_{10,1}$ is equal to 0, $(\tilde{h}^t - (2/\gamma)\log(|\cdot|)\circ f_t, \rho_{Z,\varepsilon}-\rho_{10,1})$ depends only on the equivalence class modulo constants of $\tilde{h}^t - (2/\gamma)\log(|\cdot|)$. Moreover, by stationarity of the capacity zipper, this law is that of a Neumann GFF $\bar{h}$ (modulo constants) that is independent of $(f_s; s \le t)$. Thus we are reduced to doing a simple Gaussian computation. This is very similar what was carried out in the proof of Theorem 6.1 and yields that

$$\mathbf{P}(e^{(\tilde{h}^t - (2/\gamma)\log(|\cdot|)\circ f_t, \rho_{Z,\varepsilon}-\rho_{10,1})} \mid Z, (f_s; s \le t)) = e^{-\frac{\gamma^2}{8}[(M, \rho_{Z,\varepsilon}-\rho_{10,1})]_t}.$$

We may also note that when $t \le \tau_{Z-\varepsilon}$, M_t can be extended by Schwarz reflection to a harmonic function on a domain containing $B(Z, \varepsilon)$ and $B(10, 1)$, and so by the mean value theorem $(M_t, \bar{\rho}_{Z,\varepsilon}-\rho_{10,1}) = M_t(Z) - M_t(10)$. Similarly, on the event that $t \le \tau_{Z-\varepsilon}$, $[(M, \rho_{Z,\varepsilon}-\rho_{10,1})]_t = [M(Z) - M(10)]_t$. (6.11) then follows by definition of $\mathbf{Q}_\varepsilon$ and conditioning.

Next, recall from the proof of Lemma 6.5 that $dM^*_r = -(2/(\gamma f_r(z)))dW_r$ where W is the driving function of $(f_r)_r$ (and is a Brownian motion run at speed γ^2). Hence, by (6.11) and the Cameron-Martin-Girsanov theorem we have that (under $\mathbf{Q}_\varepsilon$, conditionally on Z and up to time $\tau_{Z-\varepsilon}$), $W_t - \frac{\gamma}{2}[W, M(Z) - M(10)]_t$ is a (speed γ^2) Brownian motion, or equivalently

$$dW_t = \gamma dB_t - \gamma^2\Re(\frac{1}{f_t(Z)})dt + \gamma^2\Re(\frac{1}{f_t(10)})dt.$$

Since this does not depend on ε, the same must hold under $\mathbf{Q}^Z = \mathbf{Q}(\cdot|Z)$, at least up to time $\tau_{Z-\varepsilon}$. However, as $\varepsilon > 0$ was arbitrary, it in fact holds until time τ_Z. Since this is exactly the equation satisfied by the driving function of an $\mathrm{SLE}_\kappa(\kappa, -\kappa)$ process with force points at $(Z, 10)$, we conclude the proof of (d).

Finally, we deal with (e). For this, we use the same rewriting of $\tilde{h}_0^\varepsilon$ as above to see that

$$\mathbf{Q}^\varepsilon(F(\tilde{h}^t) \,|\, (f_s)_{0\le s\le t}, Z) = \frac{\mathbf{P}(F(\tilde{h}^t) e^{\frac{\gamma}{2}(\tilde{h}^t, (\rho_{Z,\varepsilon}-\rho_{10,1})\circ f_t^{-1})} \,|\, (f_s)_{0\le s\le t}, Z)}{\mathbf{P}(e^{\frac{\gamma}{2}(\tilde{h}^t, (\rho_{Z,\varepsilon}-\rho_{10,1})\circ f_t^{-1})} \,|\, (f_s)_{0\le s\le t}, Z)} \tag{6.12}$$

for any bounded measurable function F of $\tilde{h}^t$ modulo constants. On the other hand, recall that under $\mathbf{P}$, $\tilde{h}^t$ viewed modulo constants is independent of $(f_s)_{s\le t}$, and is distributed like a Neumann GFF plus the function $(2/\gamma)\log|\cdot|$ (modulo constants). Thus, by the Cameron–Martin–Girsanov theorem applied conditionally on $(Z, (f_s)_{s\le t})$, the law of $\tilde{h}^t$ under $\mathbf{Q}_\varepsilon$ and conditionally on $(Z, (f_s)_{s\le t})$, considered modulo constants, is that of a Neumann GFF (modulo constants) *plus* the function $(2/\gamma)\log|\cdot|$, *plus* the "drift distribution" that sends smooth g to $\mathbb{E}_{\mathbb{P}}((\tilde{h}^t, g)(\tilde{h}^t, \bar{\rho}_Z^\varepsilon \circ f_t^{-1}))$. Now, for any $t \le \tau_{Z-\varepsilon}$ and any $w \in \mathbb{H}\setminus f_t(B(Z,\varepsilon)\cap\mathbb{H})$ we have $\int G^{\mathbb{H}}(w,y)(\bar{\rho}_Z^\varepsilon \circ f_t^{-1})(dy) = G^{\mathbb{H}}(w, f_t(Z))$, and so on the set $\mathbb{H}\setminus f_t(B(Z,\varepsilon)\cap\mathbb{H})$ we can write (as distributions modulo constants)

$$\tilde{h}^t \overset{(d)}{=} \bar{h} + \frac{2}{\gamma}\log|\cdot| + \frac{\gamma}{2}G^{\mathbb{H}}(\cdot, f_t(Z)) - \frac{\gamma}{2}\int G^{\mathbb{H}}(\cdot, f_t(y))\rho_{10}^1(dy), \tag{6.13}$$

where the equality in distribution holds under $\mathbf{Q}_\varepsilon$ conditionally on Z and $(f_s; s \le t)$, and where $\bar{h}$ is as described in the statement of (e). Taking a limit as $\varepsilon \to 0$ we obtain the result. □

Corollary 6.22. *Setting $t = \tau_Z$ in the previous lemma, we see that under $\mathbf{Q}^Z = \mathbf{Q}(\cdot \,|\, Z)$, $(\tilde{h}^{\tau_Z}, \eta^{\tau_Z})$ can be described as follows:*

- *$\eta^{\tau_Z}([0,\tau_Z])$ has the law of a reverse $SLE_\kappa(\kappa,-\kappa)$, with force points at $(Z,10)$, and run until the point Z reaches 0;*
- *as an equality of distributions modulo constants $\tilde{h}^{\tau_Z} \overset{(d)}{=} \bar{h} + (\gamma - 2/\gamma)\log(1/|\cdot|) - \frac{\gamma}{2}\int G^{\mathbb{H}}(\cdot, f_{\tau_Z}(y))\rho_{10,1}(dy)$, where $\bar{h}$ has the law of a Neumann GFF that is independent of $(f_s)_{0\le s\le\tau_Z}$.*

We will use this to show that when we zoom in at this weighted capacity zipper at time τ_Z, we obtain a field and curve whose joint law is that in the statement of Proposition 6.19.

Step 2: Zooming in to get a wedge and an independent SLE Suppose that η is a simple curve from 0 to ∞ in $\mathbb{H}$, considered up to time reparametrisation, and that $K \subset \mathbb{H}$ is compact. In what follows, by η restricted to K, we mean the trace of η run up to the first time that it exits the set K (which does not depend on the choice of time parametrisation). If $h \in \mathcal{D}_0'(\mathbb{H})$, by h restricted to K, we mean the restriction in the standard sense of restriction of distributions.

Lemma 6.23. *Let $((\tilde{h}^t, \eta^t)_{0 \le t \le \tau_Z}, Z)$ be sampled from $\mathbf{Q}$. Let φ_C be the unique conformal map $\mathbb{H} \to \mathbb{H}$ such that $(\mathbb{H}, \varphi_C(\tilde{h}^{\tau_Z} + C), 0, \infty)$ is the unit circle embedding of $(\mathbb{H}, \tilde{h}^{\tau_Z} + C, 0, \infty)$. Then for any $K \subset \mathbb{H}$ compact, the law of $(\varphi_C(\tilde{h}^{\tau_Z} + C), \varphi_C(\eta^{\tau_Z}))$ restricted to K converges in total variation distance to the law of (h, ζ) restricted to K, where (h, ζ) is as in Proposition 6.19.*

Remark 6.24. Note that $\{(\varphi_C(\tilde{h}^{\tau_Z} + C), \varphi_C(\eta^{\tau_Z})) : C > 0\}$ is completely determined by $(\tilde{h}^{\tau_Z}, \eta^{\tau_Z}, Z)$.

For the proof, we define an auxiliary triple $(\tilde{h}, \tilde{\eta}, \tilde{Z})$ where:

- $\tilde{Z}$ has the (marginal) $\mathbf{Q}$ law of Z;
- conditionally on $\tilde{Z}$, $\tilde{\eta}$ is the segment of curve generated by a reverse $\mathrm{SLE}_\kappa(\kappa)$ flow $(\tilde{f}_t)_t$ with a force point at $\tilde{Z}$, and run up until the time $\tilde{\tau}_Z$ that $\tilde{Z}$ reaches 0.
- $\tilde{h} = h + (\gamma - 2/\gamma)\log(1/|\cdot|)$ where h is a Neumann GFF independent of $\tilde{Z}$ and $(\tilde{f}_t)_t$, with additive constant fixed so that its value on the upper unit semi circle is zero;

Also for $C > 0$, let $\tilde{\varphi}_C$ be the unique conformal map such that $(\mathbb{H}, \tilde{\varphi}_C(\tilde{h} + C), 0, \infty)$ is in the unit circle embedding.

From now on we let K be fixed. To start, let us imagine we are working with $(\tilde{h}, \tilde{\eta}, \tilde{Z})$ rather than $(\tilde{h}^{\tau_Z}, \eta^{\tau_Z}, Z)$. Then it is not too hard to see that:

Lemma 6.25. *The law of $(\tilde{\varphi}_C(\tilde{h} + C), \tilde{\varphi}_C(\tilde{\eta}))$ restricted to K converges in total variation distance to the law of (h, ζ) restricted to K.*

Proof of Lemma 6.25. Indeed, the fact that $\tilde{\varphi}_C(\tilde{h} + C)$ restricted to K converges in total variation to h restricted to K is exactly the content of Theorem 5.53. So we just need to see why, conditionally on $\tilde{h}$, the conditional law of $\tilde{\varphi}_C(\tilde{\eta})$ restricted to K converges in total variation distance to that of an SLE_κ restricted to K. For this, we use the time-reversal symmetry of $\mathrm{SLE}_\kappa(\rho)$ – Corollary A.20 – which tells us that $\tilde{\eta}$ has the law of an ordinary SLE_κ curve run until an a.s. positive time Λ. As we increase C and apply $\tilde{\varphi}_C$, which corresponds to zooming in at the curve near the origin by a random amount that is independent of $\tilde{\eta}$ and blows up as $C \to \infty$, the total variation distance between the law of $\tilde{\varphi}_C(\tilde{\eta})$ restricted to K and an infinite SLE_κ restricted to K, goes to 0. □

Unfortunately, $(\tilde{h}^{\tau_Z}, \eta^{\tau_Z}, Z)$ does not quite have the same law as $(\tilde{h}, \tilde{\eta}, \tilde{Z})$. However this is very close to being true, if we look at the field and curve near the origin (which is all we need to do when considering K fixed and C large). More precisely, we will prove the following.

Lemma 6.26. *$(\tilde{h}^{\tau_Z}, \eta^{\tau_Z}, Z)$ and $(\tilde{h}, \tilde{\eta}, \tilde{Z})$ can be coupled so that $Z = \tilde{Z}$, and with probability arbitrarily close to 1 as $\delta \downarrow 0$, the restrictions of $(\tilde{h}^{\tau_Z}, \eta^{\tau_Z})$ and $(\tilde{h}, \tilde{\eta})$ to $\overline{B(0, \delta) \cap \mathbb{H}}$ agree.*

Proof of Lemma 6.23 given Lemma 6.26. From here we can conclude the proof of Lemma 6.23, since we can then choose C_0 large enough such that on the event in Lemma 6.26, with as close to full (sub)probability as we like, the maps $\tilde{\varphi}_C$ are determined by the restriction of $\tilde{h}$ to $\overline{B(0,\delta)\cap\mathbb{H}}$ for all $C \geq C_0$. Hence, we can choose C large enough that $(\varphi_C(\tilde{h}^{\tau_Z}+C), \varphi_C(\eta^{\tau_Z}))$ and $(\tilde{\varphi}_C(\tilde{h}+C), \tilde{\varphi}_C(\tilde{\eta}))$ can be coupled so that their restrictions to K agree with arbitrarily high probability. This gives the result by Lemma 6.25. □

Proof of Lemma 6.26. Given $\varepsilon > 0$ fixed, we will show that for δ small enough we can construct a coupling as in the claim, so that the restrictions to $\overline{B(0,\delta)\cap\mathbb{H}}$ agree with probability greater than $1-\varepsilon$. The construction goes as follows.

- Pick ε' such that $(1-\varepsilon')^4 > 1-\varepsilon$.
- Sample $(Z, \tilde{h}^0, \eta^0)$ from $\mathbf{Q}$ and set $\tilde{Z} = Z$.
- Choose $a, \delta' > 0$ small enough that:
  - for any $R > a^{-1}$ the total variation distance between an $\mathrm{SLE}_\kappa(\kappa)$ with a force point at 1 and an $\mathrm{SLE}_\kappa(\kappa,-\kappa)$ with force points at $(1,R)$, both run up until the first time that 1 reaches 0, is less than ε'; and
  - for a reverse $\mathrm{SLE}_\kappa(\kappa)$ with a force point at a, with probability greater than $(1-\varepsilon')$, the image of $B(0,1)\cap\mathbb{H}$ under the flow at the first time that a hits 0, contains $B(0,\delta')\cap\mathbb{H}$.

  This is possible by Lemma A.15.
- Given Z (and $\tilde{Z} = Z$) sample (f_t) and $(\tilde{f}_t)$ independently until the respective times that $Z, \tilde{Z}$ reach a. Note that by Lemma A.15, the image of $\{w \in \bar{\mathbb{H}} : |w-10| = 1\}$ under f at this time lies outside of $B(0,1)\cap\mathbb{H}$. Couple the flows f and $\tilde{f}$ for the remaining time (until a is mapped to 0) so that they agree with (conditional) probability $(1-\varepsilon')$. This is possible by the choice of a, conditioning on the image of the point 10 under f, and scaling. Call this good event A_1, so A_1 has probability greater than $(1-\varepsilon')$. Define $\tilde{\eta}$ to be the curve generated by $\tilde{f}_{\tilde{\tau}_Z}$ and η^{τ_Z} to be the image of η^0 under f_{τ_Z}.
- Further write $A_2 \subset A_1$ for the event with probability greater than $(1-\varepsilon')^2$, that when a is mapped to 0 by this final bit of flow, the image of $B(0,1)\cap\mathbb{H}$ contains $B(0,\delta')\cap\mathbb{H}$.
- Now we claim that, uniformly on the event A_2, the total variation distance between
  - the conditional law of $\tilde{h}^{\tau_Z} = f_{\tau_Z}(\tilde{h}^0)$ restricted to $\overline{B(0,\delta)\cap\mathbb{H}}$ given $(f_t)_{t\leq\tau_Z}$, and
  - the law of $\tilde{h}$ (recall this is independent of $\tilde{f}$ and $\tilde{Z}$) restricted to $\overline{B(0,\delta)\cap\mathbb{H}}$

  tends to 0 as $\delta \to 0$. For this, note that the first law above is that of the function $(\gamma - 2/\gamma)\log(1/|\cdot|) - \frac{\gamma}{2}\int G^{\mathbb{H}}(\cdot, f_{\tau_Z}(y))\rho_{10,1}(dy)$, plus a Neumann GFF normalised to have zero average on the image of $\{w \in \bar{\mathbb{H}} : |w-10| = 1\}$ under f_{τ_Z}. The claim then follows by definition of A_2 and Lemma 5.19.

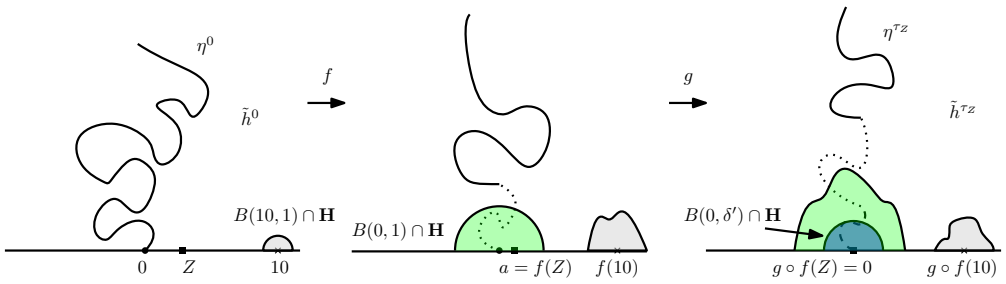


Figure 16: The idea behind the proof of Lemma 6.26. With the notation of the figure $f_{\tau_Z} = g \circ f$, and for δ very small, $\eta^{\tau_Z} \cap \overline{B(0,\delta)\cap\mathbb{H}}$ will only depend on the map g. Note that g is determined by a reverse $\mathrm{SLE}_\kappa(\kappa,-\kappa)$ flow with force points at $(a, f(10))$. But if a is small enough, this can be successfully coupled with a reverse $\mathrm{SLE}_\kappa(\kappa)$ flow with force point at a, because $f(10)$ will be proportionally far away with high probability. Furthermore, the conditional law of $\tilde{h}^{\tau_Z}$ given Z and f_{τ_Z} is that of a Neumann GFF + a $(\gamma-2/\gamma)$ log-singularity at the origin + a function that is very close to constant at the origin when $g\circ f(B(10,1)\cap\mathbb{H})$ is far away. The choice of normalising constant for $\tilde{h}^{\tau_Z}$ also only depends on the field close to the point $g\circ f(10)$. If a, δ' are small enough, the image of $B(10,1)\cap\mathbb{H}$ under $g\circ f$ will be distance δ' from the origin with high probability, and therefore the conditional law of $\tilde{h}^{\tau_Z}$ in $\overline{B(0,\delta)\cap\mathbb{H}}$, given f_{τ_Z} and Z, will have law very close to that of $\tilde{h}$.

- Thus on the event A_2, if δ is chosen small enough, $\tilde{\eta}$ and η^{τ_Z} will agree on $B(0,\delta)\cap\mathbb{H}$ with (conditional) probability $\geq (1-\varepsilon')$, and we can couple $\tilde{h}^{\tau_Z}$ and $\tilde{h}$ so that they agree on $\overline{B(0,\delta)\cap\mathbb{H}}$ with (conditional) probability $\geq (1-\varepsilon')$. Call A_3 this successful coupling event, so that A_3 has probability $> (1-\varepsilon')^4 > (1-\varepsilon)$.

□

Step 3: Stationarity In Step 2 above, we have shown that if one zooms in at the capacity zipper with reweighted law $\mathbf{Q}$ at time τ_Z, then one obtains a field/curve pair having the distribution of (h,ζ) as in Proposition 6.19. In this section we will prove that the operation of "zipping down right quantum boundary length one" does not change this law, and hence prove Proposition 6.19.

Given a sample $((\tilde{h}^t,\eta^t)_{0\leq t\leq\tau_Z}, Z)$ from $\mathbf{Q}$, and $C>0$, let $Z_C \in [0,Z]$ be such that $\nu_{\tilde{h}^0}([Z_C,Z]) = \mathrm{e}^{-C\gamma/2}$. If this is not possible (ie. if $\nu_{\tilde{h}^0}([0,Z]) < \mathrm{e}^{-C\gamma/2}$), set $Z_C = 0$. Set $\tau_C = \tau_{Z_C}$ and let ϕ_C be the unique conformal map such that $(\mathbb{H}, \phi_C(\tilde{h}^{\tau_C}+C), 0, \infty)$ is the unit circle embedding of $(\mathbb{H}, \tilde{h}^{\tau_C}+C, 0, \infty)$.

Recall the notation g_σ, ψ from Proposition 6.19.

Lemma 6.27. *For any $K\subset\mathbb{H}$ compact, $(\phi_C(\tilde{h}^{\tau_C}+C), \phi_C(\eta^{\tau_C}))$ restricted to K converges in total variation distance to $(\psi\circ g_\sigma(h), \psi\circ g_\sigma(\zeta))$ restricted to K, as $C\to\infty$.*

Lemma 6.28. *For any $K\subset\mathbb{H}$ compact, $(\phi_C(\tilde{h}^{\tau_C}+C), \phi_C(\eta^{\tau_C}))$ restricted to K converges in total variation distance to (h,ζ) restricted to K, as $C\to\infty$.*

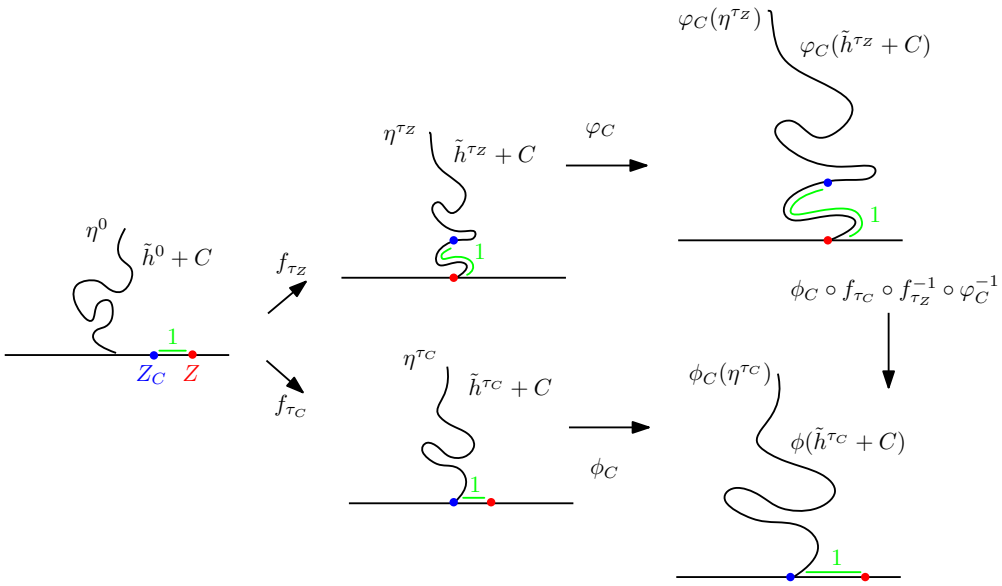


Figure 17: All the marked quantum boundary lengths (with respect to the field indicated on the relevant diagram) are equal to one. This is by definition of the conformal maps $f_{\tau_Z}, f_{\tau_C}, \phi_C$ and φ_C. Recall that $f(h)$ is obtained from h by applying the conformal change of coordinates formula which preserves quantum boundary length.

Proof of Proposition 6.19. Lemmas 6.27 and 6.28 tell us that for any K we can couple (h, ζ) and $(\psi \circ g_\sigma(h), \psi \circ g_\sigma(\zeta))$ together so that they agree when restricted to K with as high probability as we like. Thus their laws, when restricted to K, must agree. Since K was arbitrary, we can conclude. □

Proof of Lemma 6.27. Let $\varepsilon > 0$ be arbitrary. First observe that we can choose $K_\varepsilon \subset \mathbb{H}$ compact so that $K \subset g_\sigma(K_\varepsilon)$ with probability greater than $1 - \varepsilon$. By Lemma 6.23, for large enough C we can also couple $(\varphi_C(\tilde{h}^{\tau_Z} + C), \varphi_C(\eta^{\tau_Z}))$ and (h, ζ) such that with probability $> 1 - \varepsilon$ they are equal in K_ε. We may also (by taking C large enough) require that $Z_C \neq 0$ on this event. Then, since on this event we have that

$$(\phi_C(\tilde{h}^{\tau_C} + C), \phi_C(\eta^{\tau_C})) = (\psi \circ g_\sigma(h), \psi \circ g_\sigma(\zeta))$$

(this is clear since these pairs are obtained from $(\varphi_C(\tilde{h}^{\tau_Z} + C), \varphi_C(\eta^{\tau_Z}))$ and (h, ζ) respectively by zipping down 1 unit of right quantum boundary length and applying a conformal map so as to be in the unit circle parametrisation) the result follows. □

Proof of Lemma 6.28. For this, observe that if μ is the law of a uniform point in $[0, A]$ for $A > 0$, and ν is the law of $U - \varepsilon$ for $U \sim \mu$, then the total variation distance between ν and μ tends to 0 as $\varepsilon \to 0$. This means that we can couple the $\mathbf{Q}$ laws of $(Z, (\tilde{h}^t, \eta^t)_{t \geq 0})$ and $(Z_C, (\tilde{h}^t, \eta^t)_{t \geq 0})$ such that they are equal with probability tending to 1 as $C \to \infty$

(by Lemma 6.20 (b), definition of Z_C and the fact that the $\tilde{h}^0$ boundary length of $[1,2]$ is finite almost surely). Hence we can couple the $\mathbf{Q}$ laws of $(\varphi_C(\tilde{h}^{\tau_z}+C), \varphi_C(\eta^{\tau_z}))$ and $(\phi_C(\tilde{h}^{\tau_C}), \phi_C(\eta^{\tau_C}))$ so they are equal with probability tending to 1 as $C \to \infty$. Since the former law converges to that of (h,ζ) as $C \to \infty$ (Lemma 6.23), the same therefore holds for the latter. □

6.4 Uniqueness of the welding

Consider the **capacity zipper** $(\tilde{h}^t, \eta^t)_{t\in\mathbb{R}}$ of Remark 6.14 (where the additive constant for $\tilde{h}^0$ is fixed). The (reverse) Loewner flow associated to $(\eta^t)_{t\geq 0}$ has the property that it zips together intervals of $\mathbb{R}_+$ and $\mathbb{R}_-$ with the same $\nu_{\tilde{h}^0}$ quantum length by Theorem 6.9. It is natural to wonder if this actually determines the reverse flow. That is to ask: could there be any other Loewner flow with the property that intervals of identical quantum length on either side of zero are being zipped together?

We will now show that the answer to this question is no, and hence the Loewner flow for $t \geq 0$ is entirely determined by $\tilde{h}^0$.

Theorem 6.29. *Let $(\tilde{h}^t, \eta^t)_{t\in\mathbb{R}}$ be a capacity zipper as in Remark 6.14, with reverse Loewner flow $(f_t)_{t\geq 0}$. Then for $t > 0$ the following holds almost surely. If $\hat{f}_t : \mathbb{H} \to \hat{H}_t := \hat{f}_t(\mathbb{H})$ is a conformal map such that:*

- *$\hat{H}_t$ is the complement of a simple curve $\hat{\eta}^t$,*
- *$\hat{f}_t$ has the hydrodynamic normalisation $\lim_{z\to\infty} \hat{f}_t(z) - z = 0$;*
- *$\hat{f}_t$ has the property that $\hat{f}_t(z^-) = \hat{f}_t(z^+)$ as soon as $\nu_{\tilde{h}^0}([z^-, 0]) = \nu_{\tilde{h}^0}([0, z^+])$ and $f_t(z^-) \in \mathbb{H} \cup \{0\}$;*

then $\hat{f}_t = f_t$ and $\hat{\eta}^t = \eta^t$. In particular, the reverse Loewner flow $(f_t)_{t\geq 0}$ is determined by $\tilde{h}^0$ only (and hence $((\tilde{h}^t, \eta^t))_{t\geq 0}$ is entirely determined by $(\tilde{h}^0, \eta^0)$).

Proof. Before we start the proof, we recall from the definition of the capacity zipper in Theorem 6.13, that we only have defined the reverse Loewner flow as being coupled to $\tilde{h}^0$ in a certain way specified by the application of Kolmogorov's theorem. Usually, proving that objects coupled to a GFF are determined by it can be quite complicated (e.g., this is the case in the setup of imaginary geometry, or when making sense of level lines of the GFF).

Here the proof will turn out to be quite simple, given some classical results from the literature. Indeed consider

$$\phi = \hat{f}_t \circ f_t^{-1}.$$

A priori, ϕ is a conformal map on $f_t(\mathbb{H}) = H_t$, and its image is $\phi(H_t) = \hat{H}_t$. However, because of our assumptions on $\hat{f}_t$ (and the properties of f_t), the definition of ϕ can be extended unambiguously to all of $\mathbb{H}$. Moreover when we do so, the extended map is a homeomorphism of $\mathbb{H}$ onto $\mathbb{H}$, which is conformal off the curve $\eta^t([0,t])$. Thus the theorem will be proved if we can show that any such map must be the identity. In the terminology

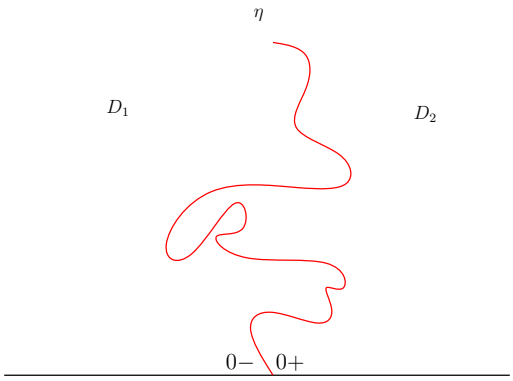


Figure 18: An independent SLE slices an $(\gamma - 2/\gamma)$-thick wedge into two independent γ-thick wedges.

of complex analysis, this is equivalent to asking that the curve $\eta^t([0,t])$ is a *removable* set. Now, by a result of Rohde and Schramm [RS05], the complement H_t of the curve is a.s. a Hölder domain for $\kappa < 4$ (or $\gamma < 2$), and by a result of Jones and Smirnov [JS00] it follows that $\eta^t([0,t])$ is a removable set. Hence the theorem follows. □

Remark 6.30. By the same argument, it also holds that for the quantum zipper $(h^t, \zeta^t)_{t\in\mathbb{R}}$ of Theorem 6.15 (h^t, ζ^t) is a.s. determined by (h^0, ζ^0) for any $t > 0$. In the language of conformal welding (h^t, ζ^t) is obtained from (h^0, ζ^0) by welding the interval on the left of 0 with h^0 quantum length t to the interval on the right of 0 with h^0 quantum length t (and pushing through ζ^0 by the resulting conformal map).

6.5 Slicing a wedge with an SLE

In this section we complement our previous discussion by the following remarkable theorem due to Sheffield [She16a]. This result is fundamental to the theory developed in [DMS14], where the main technical tool is a generalisation of the result below.

Suppose we are given a $(\gamma - 2/\gamma)$-quantum wedge $(\mathbb{H}, h, 0, \infty)$ in some embedding, and an independent SLE_κ curve η with $\kappa = \gamma^2 < 4$. Then the curve η slices the wedge into two surfaces (see picture). The result below says that *as quantum surfaces* these are independent, and that they are both γ-thick wedges. See Figure 18.

Theorem 6.31. *Suppose we are given an $(\gamma - 2/\gamma)$-quantum wedge $(\mathbb{H}, h, 0, \infty)$ in the unit circle embedding, and an independent SLE_κ curve η with $\kappa = \gamma^2 < 4$. Let D_1, D_2 be the two connected components of $\mathbb{H} \setminus \eta$, whose boundaries contain the negative and positive real lines respectively. Let $h_1 = h|_{D_1}$ and $h_2 = h|_{D_2}$. Then the two surfaces $(D_1, h_1, 0-, \infty)$ and $(D_2, h_2, 0+, \infty)$ are independent γ-thick wedges.*

Remark 6.32. This does *not* imply that the fields, or generalised functions, h_1 and h_2 are independent. It is a statement about the two doubly marked surfaces $(D_1, h_1, 0-, \infty)$

and $(D_2, h_2, 0+, \infty)$. So what it does say, for example, is that if $\tilde{h}_1$ and $\tilde{h}_2$ are the fields corresponding to the unit circle embeddings of these surfaces then $\tilde{h}_1$ and $\tilde{h}_2$ are independent.

Remark 6.33. By the same argument as in the previous subsection, the surfaces $(D_1, h_1, 0-, \infty)$ and $(D_2, h_2, 0+, \infty)$ determine h and η in the following sense. Suppose that $(\mathbb{H}, \tilde{h}_1, 0, \infty)$ and $(\mathbb{H}, \tilde{h}_2, 0, \infty)$ are the two unit circle embeddings of these surfaces, and that $(\hat{f}_1, \hat{f}_2, \hat{\eta})$ are such that:

- $\hat{\eta}$ is a simple curve from 0 to ∞;
- $\hat{f}_1$ (resp. $\hat{f}_2$) is a conformal map from $\mathbb{H}$ to the left-hand side (resp. right-hand side) of $\hat{\eta}$;
- $\hat{f}_1, \hat{f}_2$ extend to $\mathbb{R}$ in such a way that for any $x^{\pm} \in \mathbb{R}_{\pm}$ with $\nu_{\tilde{h}^1}([0, x^+]) = \nu_{\tilde{h}^2}([x^-, 0])$ we have $\hat{f}_1(x^+) = \hat{f}_2(x^-)$.

Then if $\hat{h}$ is defined by setting it equal to $\hat{f}_1(\tilde{h}_1)$ (resp. $\hat{f}_2(\tilde{h}_2)$) on the left-hand side (resp. right-hand side) of $\hat{\eta}$, we have that with probability one, $(\hat{h}, \hat{\eta}) = (\phi(h), \phi(\eta))$ for some simple scaling map $\phi : z \mapsto az$.

We also remark that the choice of embedding for the $(\gamma - 2/\gamma)$-wedge in Theorem 6.31 does not matter, which can be argued as follows. Suppose that $(\mathbb{H}, h, 0, \infty)$ is some parametrisation of a $(\gamma - 2/\gamma)$ quantum wedge and that η is an SLE_κ that is independent of h. Then there exists a scaling map $\varphi : \mathbb{H} \to \mathbb{H}$ such that $(\mathbb{H}, \varphi(h), 0, \infty)$ is the unit circle embedding of the quantum wedge. Since φ is independent of η and SLE is scale invariant, $\varphi(\eta)$ is an SLE_κ that is independent of $\varphi(h)$. Thus, applying Theorem 6.31, we see that the two quantum surfaces obtained by slicing $\varphi(h)$ along $\varphi(\eta)$ are two independent γ-quantum wedges. On the other hand, these surfaces are by definition equivalent to the two surfaces obtained by slicing h along η. This means that the latter pair also have the law (as doubly marked quantum surfaces) of two independent γ-quantum wedges.

Proof of Theorem 6.31. It is clear from the definition that $(D_1, h_1, 0-, \infty)$ and $(D_2, h_2, 0+, \infty)$ a.s. have finite LQG areas in neighbourhoods of $0-$ and $0+$ respectively, and infinite LQG areas in neighbourhoods of ∞. Therefore, we can define unique conformal maps $\phi_1 : D_1 \to \mathbb{H}$ sending $0- \to 0$ and $\infty \to \infty$ and $\phi_2 : D_2 \to \mathbb{H}$ sending $0+ \to 0$ and $\infty \to \infty$, so that $(\mathbb{H}, \phi_i(h_i), 0, \infty)$ gives LQG area one to the upper unit semidisc $B(0,1) \cap \mathbb{H}$ for $i = 1, 2$. Recall that we refer to $\phi_i(h_i)$ as the canonical description of the surface $(D_i, h_i, 0\pm, \infty)$, and we continue to use the "change of coordinate" notation (6.9) for conformal maps applied to fields. It clearly suffices to show that for any large semidisc $K \subset \mathbb{H}$, $(\phi_1(h_1)|_K, \phi_2(h_2)|_K)$ agrees in law with $(h_1^{\mathrm{wedge}}|_K, h_2^{\mathrm{wedge}}|_K)$ where h_1^{wedge} and h_2^{wedge} are independent, and each has the law of the canonical description of a γ-quantum wedge. (The reason we choose to work with the canonical description rather than the unit circle embedding here is simply to avoid any ambiguity concerning the a priori existence of the maps ϕ_1 and ϕ_2.)

To show this equality in law, we need to appeal to the results of the previous section: in particular Lemma 6.23 and Theorem 6.9. Consider the process $((\tilde{h}^t, \eta^t)_{t \geq 0}, Z)$ under the

law $\mathbf{Q}$ from Lemma 6.20, and in this set-up, let Y denote the point to the left of zero such that the $\tilde{h}^0$ boundary length of $[Y, 0]$ is equal to that of $[0, Z]$. Write h_Z^C for the canonical description of $(H_Z, \tilde{h}^0 + C, Z, \infty)$ and h_Y^C for the canonical description of $(H_Y, \tilde{h}^0 + C, Z, \infty)$ where H_Z and H_Y are the connected components of $\mathbb{H} \setminus \tilde{\eta}^0$ containing Z and Y respectively. Combining Lemma 6.23 and Theorem 6.9 gives that:

Claim 6.34. *We can couple pairs of fields with*

- *the joint law of* (h_Y^C, h_Z^C) *under* $\mathbf{Q}$*, and*
- *the joint law of* $(\phi_1(h_1), \phi_2(h_2))$ *described in the first paragraph,*

so that they agree when restricted to K*, with probability arbitrarily close to one as* $C \to \infty$*.*

Proof of claim. (See Figure 19). First we observe that one (slightly convoluted!) way to sample a pair with the law of (h_Y^C, h_Z^C) under $\mathbf{Q}$ is to:

(1) consider the "zipper" $((\tilde{h}^t, \eta^t)_{t \geq 0}, Z)$ under $\mathbf{Q}$ and apply the conformal map $f_{\tau_Z}^C$ that zips up Z to 0 and then scales $\mathbb{H}$ so that $f_{\tau_Z}^C(\tilde{h}^0)$ is in the unit circle embedding;

(2) then, restrict the field $f_{\tau_Z}^C(\tilde{h}^0 + C)$ to the left and right of $f_{\tau_Z}^C(\eta^0)$, and apply conformal maps from these left and right hand sides to $\mathbb{H}$, such that the resulting fields (under the change of coordinates formula) are the canonical descriptions of these two surfaces.

Here we are using the fact, due to Theorem 6.9, that Y is zipped up to 0 at exactly the same time as Z.

On the other hand, Lemma 6.23 says that we can couple $(f_{\tau_Z}^C(\tilde{h}^0 + C), f_{\tau_Z}^C(\eta^0))$ with (h, η) as in the statement of the present theorem, so that they agree in any large semidisc K', with probability arbitrarily close to one as $C \to \infty$. Consequently, if we restrict the field $f_{\tau_Z}^C(\tilde{h}^0 + C)$ to the left and right of $f_{\tau_Z}^C(\eta^0)$, and apply conformal maps as in the second step of the previous bullet point, then the resulting pair of fields can be coupled with $(\phi_1(h_1), \phi_2(h_2))$ so that they agree when restricted to K with arbitrarily high probability.

Combining these two paragraphs yields the claim. □

So, with the claim in hand, it actually suffices to show that we can couple (h_Y^C, h_Z^C) with $(h_1^{\text{wedge}}, h_2^{\text{wedge}})$ (recall that the latter are an pair of independent γ-wedge fields in their canonical descriptions) so that their restrictions to K agree with probability arbitrarily close to 1 as $C \to \infty$. The idea is that when C is very large, the restrictions of h_Y^C and h_Z^C to K will correspond to images - under the conformal change of coordinates (6.9) - of $\tilde{h}^0 + C$ restricted to very tiny neighbourhoods of Z and Y. Roughly speaking, these restrictions become independent in the limit as the size of the neighbourhoods goes to 0, and furthermore, the field near Z (and by symmetry near Y) converges to a γ-quantum wedge field.

To be more precise, let us consider a sample $(\tilde{h}^0, Z)$ from $\mathbf{Q}$, together with a field $\tilde{h} = \hat{h} + (\gamma - 2/\gamma) \log(|\cdot|^{-1})$, where $\hat{h}$ is a Neumann GFF normalised to have average 0 on the upper unit semicircle that is *independent* of $\tilde{h}^0$. Then we have the following:

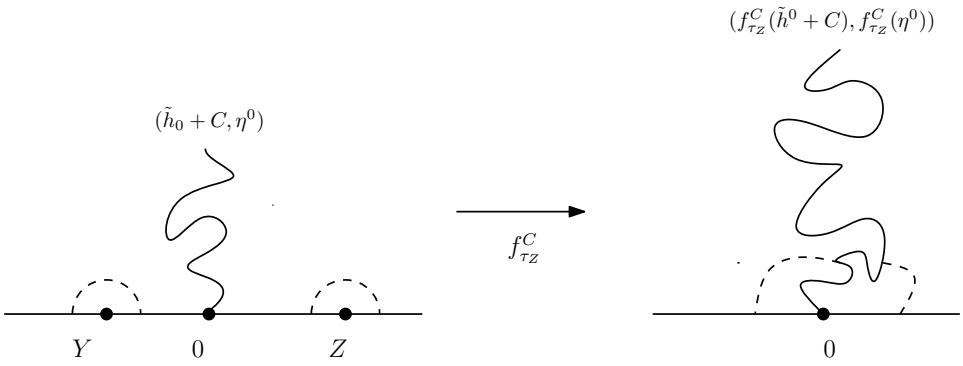


Figure 19: The surfaces to the left and right of η^0 on the left-hand picture (defined using the field $\tilde{h}^0 + C$ and marked at (Y, ∞) and (Z, ∞)) have canonical descriptions given by $(\mathbb{H}, h^C_Y, 0, \infty)$ and $(\mathbb{H}, h^C_Z, 0, \infty)$. So the same is true, by definition, for the surfaces to the left and right of the curve on the right-hand picture (defined using the field $f^C_{\tau_Z}(\tilde{h}^0 + C)$). But Lemma 6.23 says that for C large, the joint law of the field and curve on the right-hand picture is very close to that of (h, η) from the statement of Theorem 6.31. So, the law of the canonical descriptions of the surfaces to the left and right of the curve is very close to that of $(\mathbb{H}, \phi_1(h_1), 0, \infty), (\mathbb{H}, \phi_2(h_2), 0, \infty)$. Hence we can approximate the joint law of $(\phi_1(h_1), \phi_2(h_2))$ by that of (h^C_Y, h^C_Z) for C large.

Lemma 6.35. *As above, let $(\tilde{h}^0, Z)$ have their $\mathbf{Q}$-joint law, and let $\tilde{h} = \hat{h} + (\gamma - 2/\gamma)\log(|\cdot|^{-1})$, where $\hat{h}$ is a Neumann GFF normalised to have average 0 on the upper unit semicircle, that is independent of $\tilde{h}^0$. Then the total variation distance between*

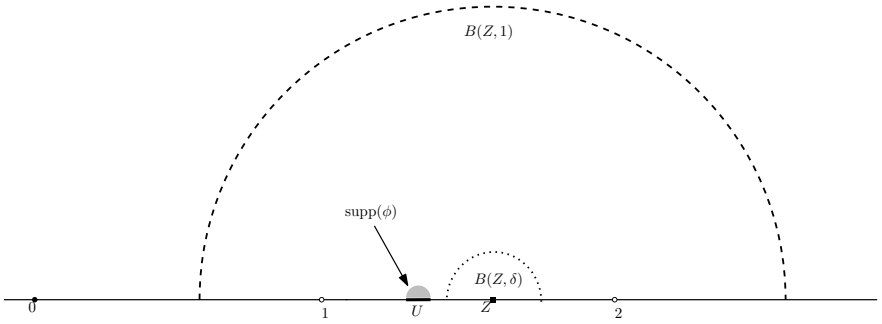
$$(\tilde{h}^0|_{\overline{B(Z,\varepsilon)\cap\mathbb{H}}}, \tilde{h}^0|_{\mathbb{H}\setminus B(Z,1)}, \nu_{\tilde{h}^0}[1, Z], \nu_{\tilde{h}^0}[Z, 2]) \text{ and } (\tilde{h}(\cdot - Z)|_{\overline{B(Z,\varepsilon)\cap\mathbb{H}}}, \tilde{h}^0|_{\mathbb{H}\setminus B(Z,1)}, \nu_{\tilde{h}^0}[1, Z], \nu_{\tilde{h}^0}[Z, 2])$$

converges to 0 as $\varepsilon \to 0$.

In words this says that *conditionally* on $\tilde{h}^0$ outside of $B(Z, 1)$ *and* on the boundary lengths $\nu_{\tilde{h}^0}[1, Z]$, $\nu_{\tilde{h}^0}[Z, 2]$, the law of $\tilde{h}^0$ restricted to $\overline{B(Z, \varepsilon) \cap \mathbb{H}}$ is very close in total variation distance to the field $\tilde{h}$ recentred at Z and restricted to $\overline{B(Z, \varepsilon) \cap \mathbb{H}}$.

Before proving the lemma, let us first see how it allows us to conclude the proof of the theorem. From now on, we assume that $K \subset \mathbb{H}$ is large, fixed semidisc. Consider a pair $(h^C_Y, \tilde{h}^C)$ where h^C_Y has its $\mathbf{Q}$-law, and $\tilde{h}^C$ is independent of h^C_Y having the law of the canonical description of $(\mathbb{H}, \tilde{h} + C, 0, \infty)$. The consequence of Lemma 6.35 is that by taking ε very small and then C sufficiently large, we can couple the joint law of (h^C_Y, h^C_Z) with that of the pair $(h^C_Y, \tilde{h}^C)$, so that the fields agree when restricted to K with probability arbitrarily close to one. Since the law of $\tilde{h}^C|_K$ converges in total variation distance to $h_2^{\text{wedge}}|_K$ as $C \to \infty$, see Corollary 5.55, this means that we can couple (h^C_Y, h^C_Z) with $(h^C_Y, h_2^{\text{wedge}})$ (where the latter pair are independent) so that they agree when restricted to K with arbitrarily high probability as $C \to \infty$.

To finish the proof, we observe that by symmetry, h^C_Y has the same law as h^C_Z for each C. Since the argument above clearly gives that $h^C_Z|_K \to h_2^{\text{wedge}}|_K$ in total variation distance as



$C \to \infty$, it must therefore also be the case that $h_Y^C|_K$ converges in total variation distance to $h_1^{\text{wedge}}|_K$ as $C \to \infty$. Thus $(h_Y^C, h_2^{\text{wedge}})$ can be coupled with $(h_1^{\text{wedge}}, h_2^{\text{wedge}})$ so that the fields agree when restricted to K with arbitrarily high probability as $C \to \infty$. Putting this together with the previous paragraph, we obtain the desired result. □

Proof of Lemma 6.35. We first claim that for any $\delta > 0$,

$$d_{TV}\left((\tilde{h}^0|_{\overline{B(Z,\varepsilon)\cap\mathbb{H}}}, \tilde{h}^0|_{\mathbb{H}\setminus B(Z,\delta)}),\ (\tilde{h}(\cdot - Z)|_{\overline{B(Z,\varepsilon)\cap\mathbb{H}}}, \tilde{h}^0|_{\mathbb{H}\setminus B(Z,\delta)})\right) \to 0 \tag{6.14}$$

as $\varepsilon \to 0$. Indeed, by Lemma 6.20, the $\mathbf{Q}^Z$ (i.e., $\mathbf{Q}(\cdot|Z)$) law of $\tilde{h}^0$ recentered at Z is that of $h' + (\gamma - 2/\gamma)\log(|\cdot|^{-1}) + \mathfrak{h}$, where h' is a Neumann GFF normalised to have average 0 on the upper unit semicircle centered at 10 and $\mathfrak{h}$ is a harmonic function that independent of h' and is deterministically bounded in $B(Z,1)$. Hence (6.14) follows from Lemma 5.34 and Remark 5.35.

We will now extend this in the following way. We are going to show that the law of $\nu_{\tilde{h}^0}([1,Z])$ is basically the same (when ε is small enough) whether we condition on $\tilde{h}^0$ restricted to $\mathbb{H}\setminus B(Z,1)$ *and* $\overline{\mathbb{H}\cap B(Z,\varepsilon)}$, or just restricted to $\mathbb{H}\setminus B(Z,1)$: see (6.15). The basic idea for the proof is that, given the restriction of $\tilde{h}^0$ to $\mathbb{H}\setminus B(Z,1)$, the restriction of $\tilde{h}^0$ to $\overline{\mathbb{H}\cap B(Z,\varepsilon)}$ has a very tiny influence on the boundary length of $[1,Z]$ when ε is small. On the other hand, there is quite a bit of variation in the boundary length coming from sources completely independent of $\tilde{h}^0|_{\overline{B(Z,\varepsilon)\cap\mathbb{H}}}$. To argue this rigorously, we will use the Fourier decomposition of the free field, similarly to the argument [She16a].

We take $\delta > 0$ small, and fix a function ϕ that is smooth, positive and supported in the upper unit semidisc of radius $\delta/4$ centered at $Z - 3\delta/2$, with $(\phi,\phi)_\nabla = 1$. Let us write $U := [Z - 7\delta/4, Z - 5\delta/4]$, $U^c = [1,Z]\setminus U$. This will be non-empty with arbitarily high probability if δ is small enough, so let us assume from now on that Z is such that this is the case. Then by Definition 5.2 and Definition 5.23, we can decompose $\tilde{h}^0 = X\phi + h$ where X is Gaussian and h is independent of X.

Next, we observe that due to the decomposition of $\tilde{h}^0$, the conditional law of $\nu_{\tilde{h}^0}(U)$ given $h|_{\mathbb{H}\setminus B(Z,\delta)}$ a.s. has smooth density $F^{h|_{\mathbb{H}\setminus B(Z,\delta)}}$ with respect to Lebesgue measure: indeed, given h restricted to $\mathbb{H}\setminus B(Z,\delta)$, $\nu_{\tilde{h}^0}(U)$ is a.s. smooth and increasing in X. In particular,

$$\text{the conditional law of } \nu_{\tilde{h}^0}([1,Z]) \text{ given } h \text{ has density } \propto F^{h|_{\mathbb{H}\setminus B(Z,\delta)}}(\cdot - \nu_h(U^c))$$

with respect to Lebesgue measure. Using the fact that F is smooth, that ν_h a.s. does not have an atom at Z, and (6.14) applied with $\delta' \ll \delta$, we may deduce from this that for any $x \in \mathbb{R}$, the quantity

$$\mathbb{E}(F^{h|_{\mathbb{H}\setminus B(Z,\delta)}}(x-\nu_h(U^c)) \,\big|\, h_{\mathbb{H}\setminus B(Z,\delta)}) - \mathbb{E}(F^{h|_{\mathbb{H}\setminus B(Z,\delta)}}(x-\nu_h(U^c)) \,\big|\, h_{\mathbb{H}\setminus B(Z,\delta)}, h|_{\overline{B(Z,\varepsilon)\cap\mathbb{H}}})$$

tends to 0 almost surely as $\varepsilon \to 0$. This is important because it means that

$$d_{TV}\left(\mathcal{L}(\nu_{\tilde{h}^0}([1,Z]) \,|\, h|_{\mathbb{H}\setminus B(Z,\delta)}), \mathcal{L}(\nu_{\tilde{h}^0}([1,Z]) \,|\, h|_{\mathbb{H}\setminus B(Z,\delta)}, h|_{\overline{B(Z,\varepsilon)\cap\mathbb{H}}})\right) \to 0 \tag{6.15}$$

in probability $\varepsilon \to 0$ (where $\mathcal{L}(Y_1|Y_2)$ denotes the law of Y_1 conditioned on Y_2). In fact, since

$$h|_{\overline{B(Z,\varepsilon)\cap\mathbb{H}}} = \tilde{h}^0|_{\overline{B(Z,\varepsilon)\cap\mathbb{H}}},$$

and by combining with (6.14) this actually means that

$$d_{TV}\left((\nu_{\tilde{h}^0}([1,Z]), \tilde{h}^0|_{\overline{B(Z,\varepsilon)\cap\mathbb{H}}}, h|_{\mathbb{H}\setminus B(Z,\delta)}),\ (\nu_{\tilde{h}^0}([1,Z]), \tilde{h}(\cdot - Z)|_{\overline{B(Z,\varepsilon)\cap\mathbb{H}}}, h|_{\mathbb{H}\setminus B(Z,\delta)})\right) \to 0$$

in probability $\varepsilon \to 0$. This is extends with exactly the same argument (but a little more notation) to the same statement with $\nu_{\tilde{h}^0}([1,Z]), \nu_{\tilde{h}^0}([Z,2])$ in place of just $\nu_{\tilde{h}^0}([1,Z])$. Putting this together with the fact that $h = h^0$ outside of $B(Z,1)$ completes the proof.

□

A Appendix

A.1 SLE

The aim of this appendix is to collect some relevant background material on Schramm–Loewner evolutions (SLE), primarily to accompany Chapter 6. For a much more detailed and pedagogical exposition, the reader is referred to [BN11, Kem17, Law05]. The presentation here most closely follows [BN11].

Complex analysis basics. First, we fix some basic notation and terminology.

- $K \subset \mathbb{H}$ is said to be a *complex* $\mathbb{H}$*-hull* if it is bounded and $H := \mathbb{H} \setminus K$ is a simply connected domain.
- For any such hull, by the Riemann Mapping Theorem, one can choose a conformal map $g_K : H \to \mathbb{H}$ such that $g_K(z) - z \to 0$ as $z \to \infty$. In fact, one can prove that for this g_K, the expansion $g_K(z) = z + \frac{a_K}{z} + O(|z|^{-2})$ holds as $z \to \infty$ for some $a_K \geq 0$. We call g_K the *Loewner map* of K.
- a_K is known as the *half plane capacity* of K and denoted by $\mathrm{hcap}(K)$.
- In some sense, the half-plane capacity measures the size of the hull K, when "viewed from infinity". In particular, the half plane capacity increases as a hull increases: if $K \subset K'$ are two complex $\mathbb{H}$-hulls, then $\mathrm{hcap}(K) \leq \mathrm{hcap}(K')$.

Loewner Chains. A Loewner chain is a family $(K_t)_{t\geq 0}$ of *increasing* ($K_s \subsetneq K_t$ for $s \leq t$) complex $\mathbb{H}$-hulls which satisfy a *local growth property*: for any $T \geq 0$,

$$\sup_{s,t\in[0,T],|s-t|\leq h} \mathrm{rad}\left(g_{K_s}(K_t \setminus K_s)\right) \to 0 \text{ as } h \to 0.$$

Here the radius of a hull means the radius of the smallest semicircle in which it can be inscribed. For such a chain one can show that the half plane capacity is a strictly increasing bijection from $[0,\infty) \to [0,\infty)$, so we can always assume (by convention) that time is parameterised so that $\mathrm{hcap}(K_t) = 2t$ for all t.

Theorem A.1 (Loewner's theorem)**.** *Loewner discovered that such chains (parameterised by half-plane capacity) are in bijection with continuous real valued functions via the following correspondence.*

- *Given $(K_t)_{t\geq 0}$ a Loewner chain, there is a unique point $\xi_t \in \overline{\cap_{h>0} g_{K_t}(K_{t+h} \setminus K_t)}$ for each $t \geq 0$. $(\xi_t)_{t\geq 0}$ is a continuous real-valued function called the* driving function of *$(K_t)_{t\geq 0}$.*
- *Given $(\xi_t)_{t\geq 0}$ a continuous real valued function, define, for each $z \in \mathbb{H}$, $g_t(z)$ to be the maximal solution to the* Loewner equation

$$\frac{\partial g_t(z)}{\partial t} = \frac{2}{g_t(z) - \xi_t}, \quad g_0(z) = z \tag{A.1}$$

Figure 20: A Loewner chain drawn up to two times: on the left, a time before $\zeta(z)$, and on the right, just after $\zeta(z)$.

which exists on some time interval $[0, \zeta(z)]$ *by classical ODE theory. Let* $K_t = \{z \in \mathbb{H} : \zeta(z) \leq t\}$. *Then* $(K_t)_{t\geq 0}$ *is a Loewner chain with driving function* ξ_t. *Moreover,* $g_t = g_{K_t}$ *for all* t.

We call $(g_t)_{t\geq 0}$ the (forward) *Loewner flow*. $\zeta(z)$ is the time that the growing hull K_t "swallows" the point z. See Figure 20.

Remark A.2. Continuous curves $(\gamma(t))_{t\geq 0} =: (\gamma_t)_{t\geq 0}$ in $\mathbb{H}$ which do not cross themselves and have $|\gamma_t| \to \infty$ as $t \to \infty$ provide examples of Loewner chains. More precisely, when one defines $H_t = \mathbb{H} \setminus K_t$ for each t to be the connected component of $\mathbb{H} \setminus \gamma([0, t])$ containing ∞. In this case the map g_t sends the tip of the curve, γ_t, to the point ξ_t (where g_t is extended by continuity).

Chordal SLE: definition. Chordal SLE_κ processes, for $\kappa > 0$, were introduced by Oded Schramm [Sch00] as a family of potential scaling limits for interfaces in critical statistical physics models. As we will soon see, they satisfy two very natural properties that make them appropriate candidates for such limits: conformal invariance and a certain domain Markov property.

It turns out ([Sch00]) that these two properties actually *characterise* SLE_κ as a one parameter family, which means that there really can be no other candidates. On the other hand, proving convergence of discrete interface models to SLE is typically very challenging. To date it has been verified for just a few special values of κ; for example, critical percolation interfaces, [Smi01], and the loop-erased random walk, [LSW04].

Definition A.3 (Chordal SLE in $\mathbb{H}$ from $0 \to \infty$)**.** *For* $\kappa > 0$, SLE_κ *in* $\mathbb{H}$ *from* 0 *to* ∞ *is defined to be the Loewner chain driven by* $\xi_t = \sqrt{\kappa} B_t$ *where* B_t *is a standard Brownian motion.*

One of the first things to note about SLE is that, due to the scaling property of Brownian motion (B_t has the same law as $\sqrt{t} B_1$ for any t), SLE is itself scale invariant. That is, for any $r \geq 0$ if $(K_t)_{t\geq 0}$ is an SLE_κ process, then the rescaled process $(r^{-1/2} K_{rt})_{t\geq 0}$ also has the law of an SLE_κ. This says that SLE is invariant under conformal maps of $\mathbb{H}$ that fix 0 and ∞. This allows us to define SLE, by conformal invariance, in any simply connected domain and between any two marked boundary points.

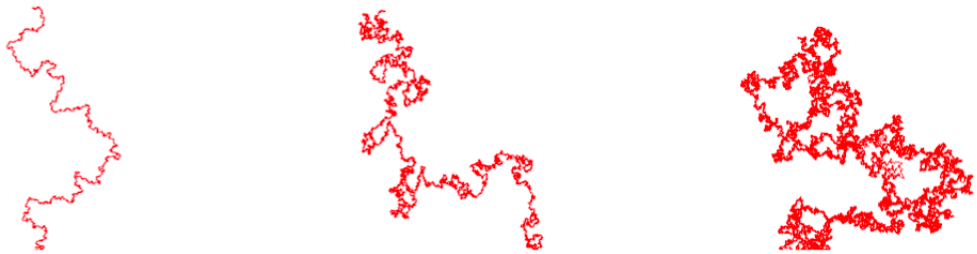


Figure 21: From left to right - SLE_2, SLE_4, SLE_6. Simulations by Tom Kennedy.

Definition A.4 (Chordal SLE)**.** *SLE_κ is a collection $(\mu_{D,a,b})_{D,a,b}$ of laws on Loewner chains, indexed by triples (D, a, b) where D is a simply connected domain and a and b are two marked boundary points. The law $\mu_{\mathbb{H},0,\infty}$ is that given by Definition A.3. For any other triple (D, a, b), $\mu_{D,a,b}$ is defined to be the image of $\mu_{\mathbb{H},0,\infty}$ under the (unique) conformal isomorphism sending $\mathbb{H}$ to D, 0 to a and ∞ to b.*

Chordal SLE: properties.

- Chordal SLE_κ is generated by a curve γ (in the sense of Remark A.2) for every $\kappa > 0$: due to [RS05] for $\kappa \neq 8$, and [LSW04] for $\kappa = 8$.
- *Conformal invariance:* if γ is an SLE_κ in D from a to b and $\psi : D \to D'$ is a conformal map with $\psi(a) = a'$ and $\psi(b) = b'$, then $\psi(\gamma)$ (after reparameterisation) has the law of an SLE_κ in D' from a' to b'.
- *Domain Markov property:* if γ is an SLE_κ from a to b in D and T is a bounded stopping time that is measurable with respect to γ, then conditionally on $\gamma([0, T])$, writing D_T for the connected component containing b of $D \setminus \gamma([0, T])$, $\gamma([T, \infty))$ has the law of an SLE_κ from $\gamma(T)$ to b in D_T.
- It has three distinct *phases*: for $\kappa \in [0, 4]$ SLE_κ is a.s. generated by a simple (non self-touching and non boundary-touching) curve; for $\kappa \in (4, 8)$ it a.s. hits (but doesn't cross) itself and the boundary of the domain; and for $\kappa \geq 8$ it is a.s. space filling. See Figure 21.

$\mathrm{SLE}_\kappa(\underline{\rho})$ processes. It is also very natural to consider random curves which only have a domain Markov property when additional marked points on the boundary of the domain are kept track of. Such curves (should) arise when considering scaling limits of discrete interface models with more complicated boundary behaviour. For instance, when there are boundary conditions that change at multiple points.

Such curves are described by variants of SLE_κ, which have an additional attraction or repulsion from certain *marked points* (also sometimes known as force points) in the domain or on its boundary. These are known as $\mathrm{SLE}_\kappa(\underline{\rho})$ and first appeared in [LSW03]; see also

[SW05]. The vector $\underline{\rho}$ encodes how strong this attraction or repulsion is, and in which direction.

Here we will consider the case of finitely many marked points $V^1, ..., V^m$ on the boundary, with corresponding weights $\rho^1, ..., \rho^m \in \mathbb{R}$ such that

$$\sum_{i \in S} \rho^i \geq -2 \text{ for every } S \subset \{1, ..., M\} \tag{A.2}$$

In this setting, $\mathrm{SLE}_\kappa(\underline{\rho}) = \mathrm{SLE}_\kappa(\rho^1, ..., \rho^m)$ is a Loewner chain that is a.s. generated by a curve and can be defined for all time. In fact, this holds under a weaker assumption on $\underline{\rho}$ (see for example [MSW17]) but some additional complications arise that we will not address in this brief overview.

By Loewner's theorem, $\mathrm{SLE}_\kappa(\underline{\rho})$ is defined by specifying its driving function. As for ordinary SLE_κ, this driving function is a random function closely related to Brownian motion. However, the Brownian motion now comes with a *drift*. This drift will be influenced by the behaviour of processes $(V_t^1, ..., V_t^m)_{t \geq 0}$, which describe the evolution of the marked points $(V^1, ..., V^m) = (V_0^1, ..., V_0^m)$ under the Loewner flow.

Definition A.5 ($\mathrm{SLE}_\kappa(\rho^1, ..., \rho^m)$ in $\mathbb{H}$ from 0 to ∞)**.** *Suppose that $V^1, ..., V^m \in (\mathbb{R} \backslash 0) \cup \{\infty\}$ are distinct and $\rho^1, ..., \rho^m$ satisfy the condition (A.2). $SLE_\kappa(\rho^1, ..., \rho^m)$ with marked points at $V^1, ..., V^m$ is the Loewner chain with driving function $(\xi_t)_{t \geq 0}$ satisfying the following system of SDEs:*

$$\begin{aligned} \xi_t &= \sqrt{\kappa} B_t + \sum_i \int_0^t \frac{\rho^i}{\xi_s - V_s^i} \, ds \\ V_t^i &= V^i + \int_0^t \frac{2}{V_s^i - \xi_s} \, ds \text{ for } 1 \leq i \leq M. \end{aligned} \tag{A.3}$$

Remark A.6. In the case of one marked point, $m = 1$, the process $(V_t^1 - \xi_t)_{t \geq 0}$ describing the distance between the driving function and the evolution of the marked point, is $\sqrt{\kappa}$ times a Bessel process. When $\underline{\rho} = \rho^1 = \rho$, the dimension of the Bessel process is

$$\delta = 1 + \frac{2(\rho + 2)}{\kappa}. \tag{A.4}$$

This formalises the notion that $\mathrm{SLE}_\kappa(\underline{\rho})$ processes have an additional attraction/repulsion from the marked points.

In fact, for a Loewner chain to satisfy a conformal Markov property with an extra marked point (that is, the property that for any stopping time σ, the future evolution after applying the Loewner map at time σ has the same law as the original process, with the marked point now located at the image of the original marked point) one finds that the difference between the driving function and the evolution of the marked point must be a continuous Markov process satisfying Brownian scaling. This implies that it actually has to be a Bessel process of some dimension. One can take this an explanation for the form of the SDEs (A.3).

Remark A.7. The definition can also be extended to the case where there are marked points located infinitesimally to the left and/or right of 0 (denoted 0^- and 0^+). This is done by taking a limit in law (with respect to the Carathéodory topology on Loewner chains) as one of the marked points approaches 0 from the left and/or one of the marked points approaches 0 from the right. Again this gives rise to unique laws on Loewner chains that are defined for all time.

When there is just one marked point, this boils down to starting a Bessel process of positive dimension from zero; in fact by (A.4) the dimension of this Bessel process is greater than 1 when $\rho > -2$. The reason why we assume the dimension to be greater than 1 (and so ρ to be > -2) is that this makes the integral in (A.3) convergent. When $\rho < -2$, assigning a meaning to this integral is less straightforward, though there are known procedures, including e.g. a principal value correction, see [She09].

Due to the scaling property of Brownian motion, it follows easily that $\mathrm{SLE}_\kappa(\underline{\rho})$ from 0 to ∞ in $\mathbb{H}$ also satisfies a form of scale invariance. More precisely, if $(K_t)_{t\geq 0}$ is an $\mathrm{SLE}_\kappa(\underline{\rho})$ process with marked points at $V^1, ..., V^m$, then the rescaled process $(r^{-1/2}K_{rt})_{t\geq 0}$ has the law of an $\mathrm{SLE}_\kappa(\underline{\rho})$ process with marked points at $r^{-1/2}V^1, ..., r^{-1/2}V^m$ for any $r > 0$. This allows us to extend the definition of $\mathrm{SLE}_\kappa(\underline{\rho})$ to arbitrary domains with finitely many marked boundary points.

Definition A.8 ($\mathrm{SLE}_\kappa(\rho^1, ..., \rho^m)$ in D from a to b)**.** *Suppose that $\rho^1, ..., \rho^m$ are as in Definition A.5 and $(D, a, b, V^1, ..., V^m)$ is a given domain with $(m+2)$-marked boundary points. Let $\psi : \mathbb{H} \to D$ be a conformal map sending a to 0 and b to ∞.*

$SLE_\kappa(\rho^1, ..., \rho^m)$ from a to b in D with marked points at $V^1, ..., V^m$ is defined to be the image under ψ of $SLE_\kappa(\rho^1, ..., \rho^m)$ from 0 to ∞ in $\mathbb{H}$, with marked points at $\psi^{-1}(V^1), ..., \psi^{-1}(V^m)$.

Remark A.9 (Properties)**.** $\mathrm{SLE}_\kappa(\underline{\rho})$ possesses many properties similar to those of SLE_κ, along with some additional features.

- For any $\kappa > 0$ and $\underline{\rho}$ satisfying (A.2), $\mathrm{SLE}_\kappa(\underline{\rho})$ is a.s. generated by a continuous curve γ, with $\gamma(0) = a$ and $\gamma(t) \to b$ as $t \to \infty$: see [MS16a].
- By definition, if $\psi : D \to D'$ is a conformal map sending $(a, b, V^1, ..., V^m)$ on the boundary of D to $(a', b', (V^1)', ..., (V^m)')$, then the image of $\mathrm{SLE}_\kappa(\underline{\rho})$ from a to b in D with marked points at $V^1, \cdots, V^m$ has the law of $\mathrm{SLE}_\kappa(\underline{\rho})$ from a' to b' in D' with marked points at $(V^1)', ..., (V^m)'$.
- Going back to the set up in the upper half plane, the processes $(V_t^i)_{t\geq 0}$ from (A.3) describe the evolution of the marked points V^i under the Loewner flow. More precisely, for each i and until the first time that V^i is swallowed, V_t^i is equal to $g_t(V^i)$ (where g_t is extended to the boundary by continuity). After this time, if $V^i \in \{0^+\} \cup (0, \infty)$ (resp. $\in \{0^-\} \cup (-\infty, 0)$), then V_t^i is equal to the image under g_t of the furthest right (resp. furthest left) point on the real line that has been swallowed at time t. That is, $V_t^i = g_t(V^i) \vee g_t(\sup\{x \in (0, \infty) : \zeta(x) \leq t\})$ in the first case, and $V_t^i = g_t(V^i) \wedge g_t(\inf\{x \in (-\infty, 0) : \zeta(x) \leq t\})$ in the second. See Figure 22.

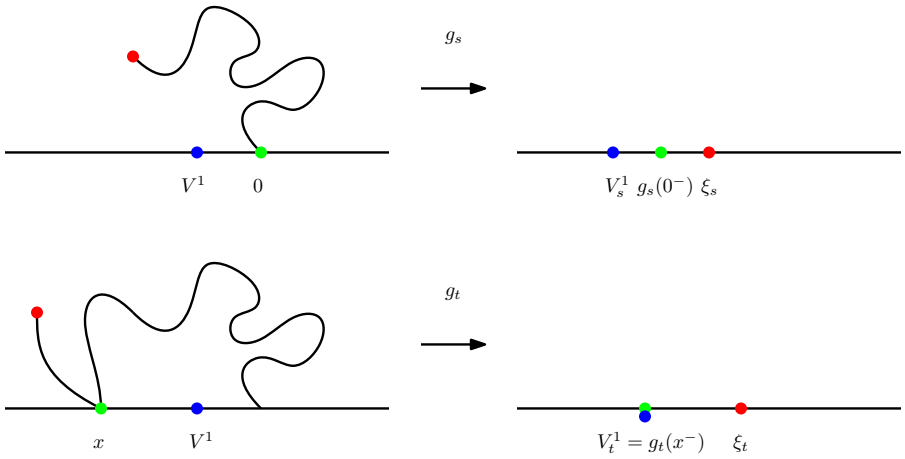


Figure 22: A schematic picture of an $\mathrm{SLE}_\kappa(\rho)$ with one marked point, drawn up to two times s, t with $s < t$. At time s the marked point has not been swallowed, but at time t it has. After time t the evolution of V_1^t coincides (by definition) with the evolution under the Loewner flow of the point infinitesimally to the left of x.

- As we have mentioned already, $\mathrm{SLE}_\kappa(\underline{\rho})$ satisfies a domain Markov property, that now keeps track of the marked points. To state this precisely, suppose that γ is an $\mathrm{SLE}_\kappa(\underline{\rho})$ from a to b in D with marked points initially at $V^1, ..., V^m$ and that T is a bounded stopping time for γ. Write D_T for the connected component of $D \setminus \gamma([0,T])$ containing b. Then conditionally on $\gamma([0,T])$, $\gamma([T,\infty))$ has the law of an $\mathrm{SLE}_\kappa(\underline{\rho})$ from $\gamma(T)$ to b in D_T, with marked points at $V_T^1, ..., V_T^m$.

Reverse Loewner flow. Until now this appendix has focused on standard Loewner evolutions, describing increasing families of compact hulls: in nice cases, growing curves. However, these should really be referred to as *forward* Loewner evolutions, because they also have a counterpart: *reverse Loewner evolutions*. A reverse Loewner evolution is no longer a family of hulls that increases in time, but rather a family of hulls where in each infinitesimal increment of time, an infinitesimal new piece of hull is added "at the root". The whole of the previous hull is then conformally mapped to something slightly different (one might envisage the new piece of hull as "pushing" the existing one further into the domain). See Figure 23. Note that one can never therefore speak of a "single curve" associated to a reverse Loewner evolution.

In the following, we will only ever speak about *centered* reverse Loewner evolutions. Informally, this means that new pieces of curve are always added at the origin.

Definition A.10 (Reverse Loewner evolution in $\mathbb{H}$)**.** *Let $(\xi_t)_{t\geq 0}$ be a continuous real valued function with $\xi_0 = 0$. The solution $(f_t(z))_{t\geq 0, z\in\mathbb{H}}$ to the family of equations*

$$\frac{\partial(f_t(z)+\xi_t)}{\partial t} = \frac{-2}{f_t(z)}, \quad f_0(z) = z \ ; \ z \in \mathbb{H} \tag{A.5}$$

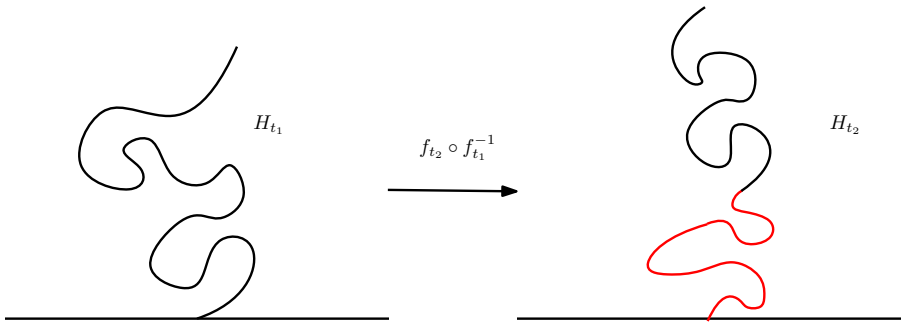


Figure 23: A reverse Loewner evolution at two times t_1, t_2 with $t_1 < t_2$. In both cases H_{t_i} is the complement of the curve. One can see that between the two times, a new piece of curve (drawn in red) is added "at the root", and the existing curve (black) is conformally mapped into the domain formed by the complement of the red curve. In contrast, under the forward Loewner flow, new pieces of curve are always added "at the tip" of the existing curve.

is called the reverse Loewner flow driven by $(\xi_t)_{t\geq 0}$*. In contrast to the forward case,* $f_t(z)$ *is defined for all* $t \geq 0$ *and* $z \in \mathbb{H}$*. This means that* f_t *defines a conformal map from* $\mathbb{H}$ *to some domain* H_t *for all* t *(and one can check that* $f_t(z) \sim z$ *as* $z \to \infty$ *for each* t*).* $(\mathbb{H} \setminus H_t)_{t\geq 0}$ *is called the reverse Loewner evolution driven by* $(\xi_t)_{t\geq 0}$*.*

We will now discuss the (deterministic) relation between forward and reverse Loewner evolutions. For this, it is helpful to consider the centered forward Loewner maps $\tilde{g}_t := g_t - \xi_t$ and associated with a given driving function $(\xi_t)_{t\geq 0}$.

Lemma A.11 (Forward/Reverse flow)**.** *Suppose that* $(\tilde{g}_t)_{t\geq 0}$ *is the centered forward Loewner flow with driving function* $(\xi_t)_{t\geq 0}$*. Fix* $T > 0$ *and write* $\hat{\xi}_t = \xi_{T-t} - \xi_T$ *for* $0 \leq t \leq T$*. Let* $(\hat{f}_t)_{0\leq t\leq T}$ *be the centered reverse Loewner flow with driving function* $(\hat{\xi}_t)_{0\leq t\leq T}$*. Then*

$$\hat{f}_t(z) := \tilde{g}_{T-t} \circ \tilde{g}_T^{-1}(z) \; ; \; t \in [0, T] \, , \, z \in \mathbb{H}.$$

In particular, $\tilde{f}_T \equiv \tilde{g}_T^{-1}$*.*

Proof. Since $\tilde{g}_t = g_t - \xi_t$ by definition, the forward Loewner equation (A.1) and then the substitution $t \mapsto T - t$ yields

$$d(\tilde{g}_t(z)) = \frac{2}{\tilde{g}_t(z)} dt + d\xi_t \; ; \; d(\tilde{g}_{T-t}(z)) = -\frac{2}{\tilde{g}_{T-t}(z)} - d\hat{\xi}_t$$

for every z. Replacing z with $\tilde{g}_T^{-1}(z)$, we may deduce that $\hat{f}_t(z)$ satisfies the reverse Loewner equation (A.5) with driving function $\hat{\xi}$. □

Reverse SLE. Now we have defined reverse Loewner evolutions, reverse SLE_κ is simply defined in the analogous way to forward SLE_κ.

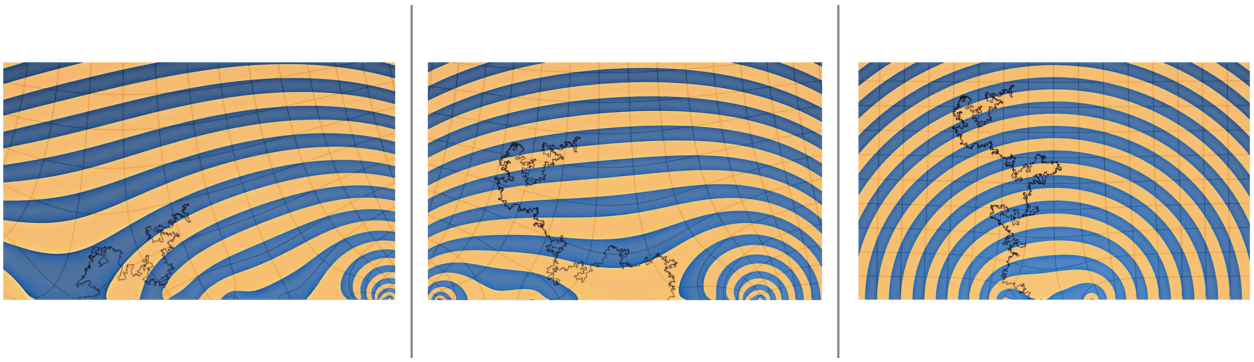


Figure 24: A reverse SLE_4 at three increasing times, simulation due to Henry Jackson. The background shows the deformation of the upper half plane under the reverse Loewner flow.

Definition A.12 (Reverse SLE_κ)**.** *Reverse SLE_κ for $\kappa > 0$ is the centered reverse Loewner evolution driven by a Brownian motion with diffusivity κ. That is, with driving function $(\xi_t)_{t\geq 0} = (\sqrt{\kappa}B_t)_{t\geq 0}$ where B is a standard Brownian motion.*

Definition A.13 (Reverse $\mathrm{SLE}_\kappa(\underline{\rho})$ [She16a])**.** *Suppose that $V^1, ..., V^m \in \bar{\mathbb{H}}$ and $\rho^1, ..., \rho^m$ are real numbers. Reverse $SLE_\kappa(\underline{\rho}; V^1, ..., V^m)$ with marked points at $V^1, ..., V^m$ is the reverse (centered) Loewner evolution with driving function $(\xi_t)_{t\geq 0}$ satisfying:*

$$\xi_t = \sqrt{\kappa}B_t - \sum_i \int_0^t \Re\left(\frac{\rho^i}{f_s(V^i)}\right) ds \tag{A.6}$$

It is immediate that this has a unique solution in law, at least until the first time that $f_t(V^i) = 0$ for some i. We will only consider the reverse $SLE_\kappa(\underline{\rho})$ up until this time.

Remark A.14. In the case $m = 1$ and $\rho^1 = \rho$, a straightforward calculation shows that $f_t(V^1)$ is $\sqrt{\kappa}$ times a Bessel process of dimension

$$\delta = 1 + \frac{2(\rho - 2)}{\kappa}.$$

Note the difference with Remark A.6. Roughly speaking, this is because the reverse $\mathrm{SLE}_\kappa(\underline{\rho})$ generally pulls points towards the origin, while the forward version will be pushes them away (for intuition, consider the case $\rho = 0$ and the way that the flow is defined).

The following properties of reverse $\mathrm{SLE}_\kappa(\underline{\rho})$ will be needed for a technical discussion in **??** of these notes (and as such, will probably seem rather specific/unmotivated). A reader simply wishing to learn about SLE would be safe to skip this.

Lemma A.15. *(1) Let $(f_t)_{t\leq\tau_1}$ be a reverse $SLE_\kappa(\kappa)$ flow, with a force point at $1 \in \mathbb{R}$ and τ_1 the first time that $f_t(1) = 0$. Then as $R \to \infty$, the probability that $f_{\tau_1}(B(0, R)) \supset B(0, 1)$ tends to 1.*

(2) *Let $(\tilde{f}_t)_{t\le\tilde{\tau}_1}$ be a reverse $SLE_\kappa(\kappa,-\kappa)$ flow with force points at $(1,R)$ and $\tilde{\tau}_1$ the first time that $\tilde{f}_t(1)=0$. Then the total variation distance between $(f_t)_{t\le\tau_1}$ and $(\tilde{f}_t)_{t\le\tilde{\tau}_1}$ tends to 0 as $R\to\infty$.*

(3) *Let $(\tilde{f}_t)_t$ be a reverse $SLE_\kappa(\kappa,-\kappa)$ flow with force points at $(z,10)$, and $z\in[1,2]$. For $a\in(0,1]$ let $\tilde{\tau}_a$ be the first time that $\tilde{f}_t(z)=a$. Then with probability one, $f_{\tilde{\tau}_a}(\{w\in\bar{\mathbb{H}}:|w-10|=1\})\subset\bar{\mathbb{H}}\setminus B(0,1))$.*

Proof. (1) Note that $f_t(1)$ is $\sqrt{\kappa}$ times a Bessel process of dimension $3-(4/\kappa)<2$ started from 1, and so the time τ_1 is almost surely finite. Moreover, the driving function is continuous up to and including time τ_1, because the integral $\int_0^{\tau_1} f_t(1)^{-1}\,dt$ converges almost surely. This implies that $f_{\tau_1}(z)\to\infty$ as $z\to\infty$ (see Definition A.10), which is the same thing as (1).

(2) For this we compute the Radon–Nikodym derivative between $(\tilde{f}_t)_{t\le\tilde{\tau}_1}$ and $(f_t)_{t\le\tau_1}$ explicitly using Girsanov's theorem. Let us write ξ_t and $\tilde{\xi}_t$ for their respective driving functions. Then

$$d\xi_t=\sqrt{\kappa}dB_t-\frac{\kappa}{f_t(1)}\,dt$$

where B_t is a standard Brownian motion (and for $z\in\mathbb{H}$ we have $df_t(z)=-(2/f_t(z))\,dt-d\xi_t$.) Let us consider the process $Z_t:=-\sqrt{\kappa}\log f_t(R)$, which is adapted to the filtration generated by B. Then by Itô's formula:

$$dZ_t=\frac{\kappa}{f_t(R)}dB_t+\left(\frac{\sqrt{\kappa}(2+\kappa)}{f_t(R)^2}-\frac{\kappa^{3/2}}{f_t(1)f_t(R)}\right)dt\ \ ;\ \ d[Z]_t=\frac{\kappa^2}{f_t(R)^2}\,dt.$$

If we set

$$M_t:=\exp(Z_t-[Z]_t/2)=f_t(R)^{-\sqrt{\kappa}}e^{-\kappa\int_0^t f_s(R)^{-2}ds}$$

then because $(d/dt)(f_t(R)-f_t(1))=(2/f_t(1))-(2/f_t(R))>0$ for all $t\le\tau_1$, M_t is bounded above by $(R-1)^{-\sqrt{\kappa}}$ for all $t\le\tau_1$. Thus $M_{t\wedge\tau_1}$ is a positive, bounded martingale. Since $d[Z,B]_t=(\kappa/f_t(R))\,dt$, Girsanov's theorem tells us that if we change measure using the martingale $(M_t)_{t\le\tau_1}$, the process $\tilde{B}_t=B_t-\int_0^t(\kappa/f_s(R))\,ds$ will be a Brownian motion under the new measure. Rewriting the expression for $d\xi_t$ in terms of $\tilde{B}$ we get $d\xi_t=\sqrt{\kappa}d\tilde{B}_t-(\kappa/f_t(1))dt+(\kappa/f_t(R))dt$, and we see that under this new measure, ξ_t satisfies the same SDE as $\tilde{\xi}_t$. Hence, the Radon–Nikodym derivative between $(\tilde{\xi}_t)_{t\le\tilde{\tau}_1}$ and $(\xi_t)_{t\le\tau_1}$ (equivalently between $(\tilde{f}_t)_{t\le\tilde{\tau}_1}$ and $(f_t)_{t\le\tau_1}$) is equal to

$$f_{\tau_1}(R)^{-\sqrt{\kappa}}\exp(-\kappa\int_0^{\tau_1}f_t(R)^{-2}\,dt),$$

which is deterministically bounded above by $(R-1)^{-\sqrt{\kappa}}$. Since this goes to 0 as $R\to\infty$, we obtain the desired convergence in total variation distance.

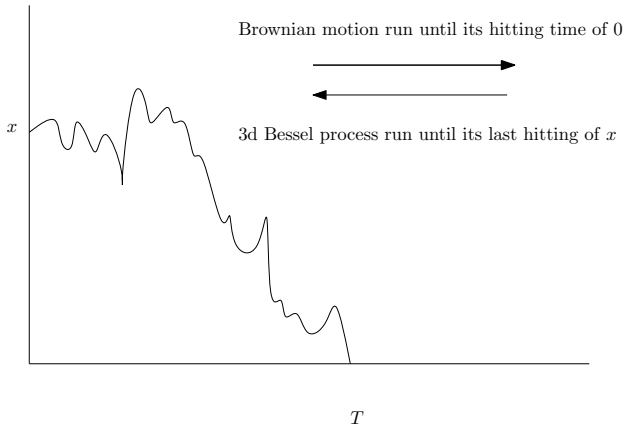


Figure 25: Illustration of Williams' path decomposition theorem. The classical result says that if X is a Brownian motion started from $x > 0$ and T is its hitting time of zero, then its time-reversal $\hat{X} = (X_{T-t})_{0 \le t \le T}$ is distributed as three-dimensional Bessel process, run until its last visit Λ to x.

(3) For this, we claim that for $w \in \mathbb{H}$ with $\Re(w) > 2$ and $z \in [1, 2]$, the process $\Re(\tilde{f}_t(w)) - \tilde{f}_t(z)$ is increasing for $t \le \tilde{\tau}_a$ (which clearly implies the result). To see the claim, observe that by definition of the reverse flow

$$\frac{\partial(\Re(\tilde{f}_t(w)) - \tilde{f}_t(z))}{\partial t} = \frac{2}{\tilde{f}_t(z)} - \Re(\frac{2}{\tilde{f}_t(w)}) = \frac{2}{\tilde{f}_t(z)} - \frac{2\Re(\tilde{f}_t(w))}{|\tilde{f}_t(w)|^2},$$

which is positive as long as $\Re(\tilde{f}_t(w)) > \tilde{f}_t(z) > 0$. Since this is true at time 0 for w with $\Re(w) > 2$, it is therefore positive for all $t \le \tilde{\tau}_a$, and the process $\Re(\tilde{f}_t(w)) - \tilde{f}_t(z)$ is increasing for this range of t.

□

Symmetries in law for forward/reverse SLE_κ and $\mathrm{SLE}_\kappa(\rho)$. Now, because Brownian motion has time reversal symmetry, the relationship Lemma A.11 between forward and reverse Loewner evolutions has particularly nice consequences for SLE.

More specifically, if $T > 0$ is fixed and $(\xi_t)_{0 \le t \le T}$ is $\sqrt{\kappa}$ times a Brownian motion, then $(\hat{\xi}_t)_{0 \le t \le T} = (\xi_T - \xi_{T-t})_{0 \le t \le T}$ also has the law of $\sqrt{\kappa}$ times a Brownian motion. Consequently:

Lemma A.16. *For any fixed $T > 0$ the curve generated by a reverse SLE_κ run up to time T and the curve generated by a forward SLE_κ run up to time T are equal in law.*

Mind that the *processes* of the previous lemma, defined for all times $t \in [0, T]$, are *not* the same in law. Indeed, we have seen that forward and reverse Loewner evolutions generate hulls via a completely different dynamic. Nonetheless, it is a very useful property that at any fixed time, the laws of the generated hulls are equal.

There are similar consequences for $\mathrm{SLE}_\kappa(\underline{\rho})$ processes, but the reversibility properties of solutions to (A.3) are somewhat more complicated. We will explain now what happens in the simplest case of one marked point. Due to remarks Remarks A.6 and A.14, this requires understanding how Bessel processes behave under time reversal.

Remark A.17 (Bessel process properties)**.** Recall that the dimension δ of a Bessel process determines how often it returns to 0: if $\delta \geq 2$, then the Bessel process will a.s. be strictly positive for all positive times; while if $\delta < 2$ then from any starting point it will return to 0 in finite time a.s.

The following is an extension of a classical result about Brownian motion, due to Williams (see e.g. Corollary (4.6) in Chapter VII of [RY99] and Figure 25).

Lemma A.18 (Time reversal of Bessel processes)**.** *Suppose that X is a Bessel process of dimension $\delta \in (0,2)$ started from $x > 0$, run until its first hitting time T of zero. Then its time-reversal $\hat{X} = (X_{T-t}, 0 \leq t \leq T)$ is a Bessel process of dimension $\hat{\delta} = 4 - \delta \in (2,4)$, run until its last visit Λ to x.*

The proof of this will boil down to an analogous result for Brownian motion with drift, that we state and prove first.

Lemma A.19. *Let $\mu > 0$. Then the time reversal of a Brownian motion with drift μ, started from 0 and stopped at its last hitting time of $y > 0$, has the law of a Brownian motion with drift $-\mu$, started from y and run up to its last hitting time of 0.*

Proof. Let $(X_t)_{t\in\mathbb{R}} = (B_t + \mu t)_{t\in\mathbb{R}}$, where B_t is a standard two-sided Brownian motion with $B_0 = 0$. Then $(\hat{X}_t)_{t\in\mathbb{R}} := (X_{-t})_{t\in\mathbb{R}}$ is equal in law to $(B_t - \mu t)_{t\in\mathbb{R}}$. Define $\tau_0 := \{\inf : s \leq 0 : X_s = 0\}$ and $\tau_y = \sup\{s \geq 0 : X_t = y\}$. Then by the strong Markov property at time τ_0, $(X_{\tau_0+s})_{0\leq s\leq \tau_y-\tau_0}$ has the law of a Brownian motion with drift μ, started from 0 and stopped at its last hitting time of y. So, we need to show that the time reversal $(X_{\tau_y-s})_{0\leq s\leq\tau_y-\tau_0}$ has the law of a Brownian motion with drift $-\mu$, started from y and run up to its last hitting time of 0.

For this, we use the fact that, by definition of $\hat{X}$,

$$(X_{\tau_y-s})_{0\leq s\leq\tau_y-\tau_0} = (\hat{X}_{s-\tau_y})_{0\leq s\leq\tau_y-\tau_0} = (\hat{X}_{s+\hat{\tau}_y})_{0\leq s\leq\hat{\tau}_0},$$

where $\hat{\tau}_y$ is the first time before 0 that $\hat{X}$ hits y, and $\hat{\tau}_0$ is the last time that $(\hat{X}_{t+\hat{\tau}_y})_{t\geq 0}$ hits 0. Since $\hat{X}$ is equal in law to a two-sided Brownian motion with drift $-\mu$, the law of the process on the right hand side above is (by the strong Markov property again, but this time for $\hat{X}$) indeed that of a Brownian motion with drift $-\mu$, started from y and run up to its last hitting time of 0. This concludes the proof. □

Proof of Lemma A.18. ([DMS14, Proposition 3.5]) We will make use of the following fact, which can be found in [RY99, Chapter XI, Exercise 1.28], and is easy to verify using basic stochastic calculus:

- Let $\tau(t) = \inf\{s > 0 : [\log(X)]_t > t\}$ and let $Z_t = \log(X_{\tau(t)})$ (recall that $[M]_t$ denotes the quadratic variation of the continuous semimartingale M). Note that because $\delta \in (0, 2)$, $\tau(t) \uparrow T$ as $t \uparrow \infty$. Then

$$(Z_t)_{t\geq 0} \overset{\text{(law)}}{=} (B_t + \frac{\delta - 2}{2}t)_{t\geq 0}, \tag{A.7}$$

  where B is a standard Brownian motion with $B_0 = \log x$.

We now want to use this, along with the time reversal symmetry of Brownian motion, to draw a conclusion similar to (A.7) about the time reversal $\hat{X}$ of X, but with the opposite drift (corresponding to a dimension $\hat{\delta} = 4 - \delta$, as claimed). However, there is a slight technical complication that arises, since $\hat{X}_0 = 0$ and so $\log(\hat{X}_0) = -\infty$.

To get around this, we also define for any $\varepsilon < x$, T_ε to be the last time before T that $(X_t)_{t\geq 0}$ hits ε. Then $(Z_t)_{t\in[0,[\log X]_{T_\varepsilon}]}$ is a Brownian motion with drift as in (A.7), started from $\log x$ and stopped at its last hitting time of $\log(\varepsilon)$. This implies (by Lemma A.19) that the time reversal of Z with respect to this time interval is a Brownian motion with drift $-(\delta - 2)/2 = (\hat{\delta} - 2)/2$, started from $\log \varepsilon$ and run up to its last hitting time of $\log x$.

Reversing the argument for (A.7) (i.e., taking the exponential and reparameterising by quadratic variation), this implies that the time reversal of $(X_t)_{t\in[0,T_\varepsilon]}$ is a Bessel process of dimension $(4 - \delta)$, started from ε and run up to its last hitting time of x. Taking a limit as $\varepsilon \to 0$ provides the result. □

As a consequence of this and Remarks A.6 and A.14, we obtain the following:

Corollary A.20 (Symmetries for forward and reverse $\mathrm{SLE}_\kappa(\rho)$)**.** *Suppose that $(f_t)_{t\geq 0}$ is the reverse flow for a centered, reverse $SLE_\kappa(\rho)$ process with a single marked point at $x > 0$ of weight $\rho < \kappa/2 + 2$. Consider the first time τ that $f_t(x) = 0$. Then $H_\tau = f_\tau(\mathbb{H})$ has the same law as $\mathbb{H} \setminus \eta([0, \sigma])$, where η is a forward $SLE_\kappa(\kappa - \rho)$ curve with a marked point at 0^+, run until the* last time Λ *that the centered forward Loewner flow for η sends 0^+ to x.*

References

[ABBS13] Elie Aïdékon, Julien Berestycki, Éric Brunet, and Zhan Shi. Branching Brownian motion seen from its tip. *Probab. Theory Related Fields*, 157(1-2):405–451, 2013.

[ABK13] Louis-Pierre Arguin, Anton Bovier, and Nicola Kistler. The extremal process of branching Brownian motion. *Probab. Theory Related Fields*, 157(3-4):535–574, 2013.

[AC15] Omer Angel and Nicolas Curien. Percolations on random maps I: Half-plane models. *Ann. Inst. Henri Poincaré Probab. Stat.*, 51(2):405–431, 2015.

[Ald93] David Aldous. The continuum random tree. III. *Ann. Probab.*, 21(1):248–289, 1993.

[Ang03] Omer Angel. Growth and percolation on the uniform infinite planar triangulation. *Geom. Funct. Anal.*, 13(5):935–974, 2003.

[Aru15] Juhan Aru. KPZ relation does not hold for the level lines and SLE_κ flow lines of the Gaussian free field. *Probab. Theory Related Fields*, 163(3-4):465–526, 2015.

[Aru17] Juhan Aru. Gaussian multiplicative chaos through the lens of the 2D Gaussian free field. *To appear in IRS 2017 special issue in Markov. Process Relat. Fields*, 2017.

[AS03] Omer Angel and Oded Schramm. Uniform infinite planar triangulations. *Comm. Math. Phys.*, 241(2-3):191–213, 2003.

[Bax00] Rodney J. Baxter. Equivalence of the two results for the free energy of the chiral Potts model. *J. Statist. Phys.*, 98(3-4):513–535, 2000.

[BBG12a] Gaëtan Borot, Jérémie Bouttier, and Emmanuel Guitter. Loop models on random maps via nested loops: the case of domain symmetry breaking and application to the Potts model. *J. Phys. A*, 45(49):494017, 35, 2012.

[BBG12b] Gaëtan. Borot, Jérémie. Bouttier, and Emmanuel Guitter. A recursive approach to the $O(n)$ model on random maps via nested loops. *J. Phys. A*, 45(4):045002, 38, 2012.

[BBM11] Olivier Bernardi and Mireille Bousquet-Mélou. Counting colored planar maps: algebraicity results. *J. Combin. Theory Ser. B*, 101(5):315–377, 2011.

[BC13] Itai Benjamini and Nicolas Curien. Simple random walk on the uniform infinite planar quadrangulation: subdiffusivity via pioneer points. *Geom. Funct. Anal.*, 23(2):501–531, 2013.

[Bef08] Vincent Beffara. The dimension of the SLE curves. *Ann. Probab.*, 36(4):1421–1452, 2008.

[Ber07] Olivier Bernardi. Bijective counting of tree-rooted maps and shuffles of parenthesis systems. *Electron. J. Combin.*, 14(1):Research Paper 9, 36, 2007.

[Ber08] Olivier Bernardi. A characterization of the Tutte polynomial via combinatorial embeddings. *Ann. Comb.*, 12(2):139–153, 2008.

[Ber17] Nathanaël Berestycki. An elementary approach to Gaussian multiplicative chaos. *Electron. Commun. Probab.*, 22:Paper No. 27, 12, 2017.

[BGRV16] Nathanaël Berestycki, Christophe Garban, Rémi Rhodes, and Vincent Vargas. KPZ formula derived from Liouville heat kernel. *J. Lond. Math. Soc. (2)*, 94(1):186–208, 2016.

[BH19] Stéphane Benoist and Clément Hongler. The scaling limit of critical Ising interfaces is CLE_3. *Ann. Probab.*, 47(4):2049–2086, 2019.

[BHS17] Olivier Bernardi, Nina Holden, and Xin Sun. Percolation on triangulations: a bijective path to Liouville quantum gravity. *Preprint arXiv:1705.03573*, 2017.

[BJM14] Jérémie Bettinelli, Emmanuel Jacob, and Grégory Miermont. The scaling limit of uniform random plane maps, *via* the Ambjørn-Budd bijection. *Electron. J. Probab.*, 19:no. 74, 16, 2014.

[BLR17] Nathanaël Berestycki, Benoît Laslier, and Gourab Ray. Critical exponents on Fortuin-Kasteleyn weighted planar maps. *Comm. Math. Phys.*, 355(2):427–462, 2017.

[BM03] Emmanuel Bacry and Jean-Francois Muzy. Log-infinitely divisible multifractal processes. *Comm. Math. Phys.*, 236(3):449–475, 2003.

[BMR19] Erich Baur, Grégory Miermont, and Gourab Ray. Classification of scaling limits of uniform quadrangulations with a boundary. *Ann. Probab.*, 47(6):3397–3477, 2019.

[BN11] Nathanaël Berestycki and James R. Norris. Lectures on Schramm–Loewner Evolution. available on the authors' websites, 2011.

[BPR20a] Nathanaël Berestycki, Ellen Powell, and Gourab Ray. $(1+\varepsilon)$-moments suffice to characterise the Gaussian free field. *Preprint arXiv:2005.02349*, 2020.

[BPR20b] Nathanaël Berestycki, Ellen Powell, and Gourab Ray. A characterisation of the Gaussian free field. *Probab. Theory Related Fields*, 176(3-4):1259–1301, 2020.

[BSS14] Nathanaël Berestycki, Scott Sheffield, and Xin Sun. Equivalence of Liouville measure and Gaussian free field. *Preprint arXiv:1410.5407*, 2014.

[CDCH+14] Dmitry Chelkak, Hugo Duminil-Copin, Clément Hongler, Antti Kemppainen, and Stanislav Smirnov. Convergence of Ising interfaces to Schramm's SLE curves. *C. R. Math. Acad. Sci. Paris*, 352(2):157–161, 2014.

[Cha84] Isaac Chavel. *Eigenvalues in Riemannian geometry*, volume 115 of *Pure and Applied Mathematics*. Academic Press, Inc., Orlando, FL, 1984. Including a chapter by Burton Randol, With an appendix by Jozef Dodziuk.

[Cha06] Pierre Chainais. Multidimensional infinitely divisible cascades. *The European Physical Journal B-Condensed Matter and Complex Systems*, 51(2):229–243, 2006.

[Che04] Mu-Fa Chen. *From Markov chains to non-equilibrium particle systems*. World Scientific Publishing Co., Inc., River Edge, NJ, second edition, 2004.

[Che17] Linxiao Chen. Basic properties of the infinite critical-FK random map. *Ann. Inst. Henri Poincaré D*, 4(3):245–271, 2017.

[CK15] Nicolas Curien and Igor Kortchemski. Percolation on random triangulations and stable looptrees. *Probab. Theory Related Fields*, 163(1-2):303–337, 2015.

[CN08] Federico Camia and Charles M. Newman. SLE_6 and CLE_6 from critical percolation. In *Probability, geometry and integrable systems*, volume 55 of *Math. Sci. Res. Inst. Publ.*, pages 103–130. Cambridge Univ. Press, Cambridge, 2008.

[DDDF19] Jian Ding, Julien Dubédat, Alexander Dunlap, and Hugo Falconet. Tightness of Liouville first passage percolation for $\gamma \in (0, 2)$. *Preprint arXiv:1904.08021*, 2019.

[DG19] Jian Ding and Subhajit Goswami. Upper bounds on Liouville first-passage percolation and Watabiki's prediction. *Comm. Pure Appl. Math.*, 72(11):2331–2384, 2019.

[DG20] Jian Ding and Ewain Gwynne. The fractal dimension of Liouville quantum gravity: universality, monotonicity, and bounds. *Comm. Math. Phys.*, 374(3):1877–1934, 2020.

[DIM77] Richard T. Durrett, Donald L. Iglehart, and Douglas R. Miller. Weak convergence to Brownian meander and Brownian excursion. *Ann. Probability*, 5(1):117–129, 1977.

[DMS14] Bertrand Duplantier, Jason Miller, and Scott Sheffield. Liouville quantum gravity as a mating of trees. *Preprint arXiv:1409.7055*, 2014.

[DRSV14a] Bertrand Duplantier, Rémi Rhodes, Scott Sheffield, and Vincent Vargas. Critical Gaussian multiplicative chaos: convergence of the derivative martingale. *Ann. Probab.*, 42(5):1769–1808, 2014.

[DRSV14b] Bertrand Duplantier, Rémi Rhodes, Scott Sheffield, and Vincent Vargas. Renormalization of critical Gaussian multiplicative chaos and KPZ relation. *Comm. Math. Phys.*, 330(1):283–330, 2014.

[DS11] Bertrand Duplantier and Scott Sheffield. Liouville quantum gravity and KPZ. *Invent. Math.*, 185(2):333–393, 2011.

[Dub09] Julien Dubédat. SLE and the free field: partition functions and couplings. *J. Amer. Math. Soc.*, 22(4):995–1054, 2009.

[EK95] Bertrand Eynard and Charlotte Kristjansen. Exact solution of the O(n) model on a random lattice. *Nuclear Phys. B*, 455(3):577–618, 1995.

[Eva10] Lawrence C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2010.

[Fal14] Kenneth Falconer. *Fractal geometry*. John Wiley & Sons, Ltd., Chichester, third edition, 2014. Mathematical foundations and applications.

[FB08] Yan V. Fyodorov and Jean-Phillipe Bouchaud. Freezing and extreme value statistics in a random energy model with logarithmically correlated potential. *J. Phys. A: Math. Theor.*, 41(37), 2008.

[GH20] Ewain Gwynne and Tom Hutchcroft. Anomalous diffusion of random walk on random planar maps. *To appear in Probab. Theory Relat. Fields*, 2020.

[GHM20] Ewain Gwynne, Nina Holden, and Jason Miller. An almost sure KPZ relation for SLE and Brownian motion. *Ann. Probab.*, 48(2):527–573, 2020.

[GHSS18] Christophe Garban, Nina Holden, Avelio Sepúlveda, and Xin Sun. Negative moments for Gaussian multiplicative chaos on fractal sets. *Electron. Commun. Probab.*, 23:Paper No. 100, 10, 2018.

[GJSZJ12] Alice Guionnet, Vaughan F. R. Jones, Dimitri Shlyakhtenko, and Paul Zinn-Justin. Loop models, random matrices and planar algebras. *Comm. Math. Phys.*, 316(1):45–97, 2012.

[GKMW18] Ewain Gwynne, Adrien Kassel, Jason Miller, and David B. Wilson. Active spanning trees with bending energy on planar maps and SLE-decorated Liouville quantum gravity for $\kappa > 8$. *Comm. Math. Phys.*, 358(3):1065–1115, 2018.

[GM17a] Ewain Gwynne and Jason Miller. Random walk on random planar maps: spectral dimension, resistance, and displacement. *Preprint arXiv:1711.00836*, 2017.

[GM17b] Ewain Gwynne and Jason Miller. Scaling limit of the uniform infinite half-plane quadrangulation in the Gromov-Hausdorff-Prokhorov-uniform topology. *Electron. J. Probab.*, 22:Paper No. 84, 47, 2017.

[GM19a] Ewain Gwynne and Jason Miller. Conformal covariance of the Liouville quantum gravity metric for $\gamma \in (0, 2)$. *Preprint arXiv:1905.00384*, 2019.

[GM19b] Ewain Gwynne and Jason Miller. Existence and uniqueness of the Liouville quantum gravity metric for $\gamma \in (0, 2)$. *Preprint arXiv:1905.00383*, 2019.

[GMS19] Ewain Gwynne, Cheng Mao, and Xin Sun. Scaling limits for the critical Fortuin-Kasteleyn model on a random planar map I: Cone times. *Ann. Inst. Henri Poincaré Probab. Stat.*, 55(1):1–60, 2019.

[GN16] Evarist Giné and Richard Nickl. *Mathematical foundations of infinite-dimensional statistical models*. Cambridge Series in Statistical and Probabilistic Mathematics, [40]. Cambridge University Press, New York, 2016.

[GP19] Ewain Gwynne and Joshua Pfeffer. KPZ formulas for the Liouville quantum gravity metric. *Preprint arXiv:1905.11790*, 2019.

[GP20] Ewain Gwynne and Josh Pfeffer. Bounds for distances and geodesic dimension in liouville first passage percolation. *Electronic Journal of Probability*, to appear, 2020+.

[GS15] Ewain Gwynne and Xin Sun. Scaling limits for the critical Fortuin-Kastelyn model on a random planar map III: finite volume case. *Manuscript in preparation*, 2015.

[GS17] Ewain Gwynne and Xin Sun. Scaling limits for the critical Fortuin-Kasteleyn model on a random planar map II: local estimates and empty reduced word exponent. *Electron. J. Probab.*, 22:Paper No. 45, 56, 2017.

[HK71] Raphael Høegh-Krohn. A general class of quantum fields without cut-offs in two space-time dimensions. *Comm. Math. Phys.*, 21:244–255, 1971.

[HMP10] Xiaoyu Hu, Jason Miller, and Yuval Peres. Thick points of the Gaussian free field. *Ann. Probab.*, 38(2):896–926, 2010.

[HRV18] Yichao Huang, Rémi Rhodes, and Vincent Vargas. Liouville quantum gravity on the unit disk. *Ann. Inst. Henri Poincaré Probab. Stat.*, 54(3):1694–1730, 2018.

[HS19] Nina Holden and Xin Sun. Convergence of uniform triangulations under the Cardy embedding. *Preprint arXiv:1905.13207*, 2019.

[Jan97] Svante Janson. *Gaussian Hilbert spaces*, volume 129 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1997.

[Jos02] Jürgen Jost. *Partial differential equations*, volume 214 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002. Translated and revised from the 1998 German original by the author.

[JS00] Peter W. Jones and Stanislav K. Smirnov. Removability theorems for Sobolev functions and quasiconformal maps. *Ark. Mat.*, 38(2):263–279, 2000.

[JS17] Janne Junnila and Eero Saksman. Uniqueness of critical Gaussian chaos. *Electron. J. Probab.*, 22:Paper No. 11, 31, 2017.

[Kah85] Jean-Pierre Kahane. Sur le chaos multiplicatif. *Ann. Sci. Math. Québec*, 9(2):105–150, 1985.

[Kah86] Jean-Pierre Kahane. Une inégalité du type de Slepian et Gordon sur les processus gaussiens. *Israel J. Math.*, 55(1):109–110, 1986.

[Kem17] Antti Kemppainen. *Schramm-Loewner evolution*, volume 24 of *SpringerBriefs in Mathematical Physics*. Springer, Cham, 2017.

[Ken00] Richard Kenyon. The asymptotic determinant of the discrete Laplacian. *Acta Math.*, 185(2):239–286, 2000.

[Ken01] Richard Kenyon. Dominos and the Gaussian free field. *Ann. Probab.*, 29(3):1128–1137, 2001.

[KMSW19] Richard Kenyon, Jason Miller, Scott Sheffield, and David B. Wilson. Bipolar orientations on planar maps and SLE_{12}. *Ann. Probab.*, 47(3):1240–1269, 2019.

[KS16] Antti Kemppainen and Stanislav Smirnov. Conformal invariance in random cluster models. II. full scaling limit as a branching SLE. *Preprint arXiv:1609.08527*, 2016.

[Law05] Gregory F. Lawler. *Conformally invariant processes in the plane*, volume 114 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2005.

[LG05] Jean-François Le Gall. Random trees and applications. *Probab. Surv.*, 2:245–311, 2005.

[LG13] Jean-François Le Gall. Uniqueness and universality of the Brownian map. *Ann. Probab.*, 41(4):2880–2960, 2013.

[LP16] Russell Lyons and Yuval Peres. *Probability on trees and networks*, volume 42 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, New York, 2016.

[LSW01] Gregory F. Lawler, Oded Schramm, and Wendelin Werner. The dimension of the planar Brownian frontier is 4/3. *Math. Res. Lett.*, 8(4):401–411, 2001.

[LSW03] Gregory F. Lawler, Oded Schramm, and Wendelin Werner. Conformal restriction: the chordal case. *J. Amer. Math. Soc.*, 16(4):917–955, 2003.

[LSW04] Gregory F. Lawler, Oded Schramm, and Wendelin Werner. Conformal invariance of planar loop-erased random walks and uniform spanning trees. *Ann. Probab.*, 32(1B):939–995, 2004.

[LSW17] Yiting Li, Xin Sun, and Samuel S. Watson. Schnyder woods, SLE(16), and Liouville quantum gravity. *Preprint arXiv:1705.03573*, 2017.

[Mie13] Grégory Miermont. The Brownian map is the scaling limit of uniform random plane quadrangulations. *Acta Math.*, 210(2):319–401, 2013.

[Mol96] George M. Molchan. Scaling exponents and multifractal dimensions for independent random cascades. *Comm. Math. Phys.*, 179(3):681–702, 1996.

[MP10] Peter Mörters and Yuval Peres. *Brownian motion*, volume 30 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2010. With an appendix by Oded Schramm and Wendelin Werner.

[MS16a] Jason Miller and Scott Sheffield. Imaginary geometry I: interacting SLEs. *Probab. Theory Related Fields*, 164(3-4):553–705, 2016.

[MS16b] Jason Miller and Scott Sheffield. Imaginary geometry II: reversibility of $\mathrm{SLE}_\kappa(\rho_1;\rho_2)$ for $\kappa \in (0,4)$. *Ann. Probab.*, 44(3):1647–1722, 2016.

[MS16c] Jason Miller and Scott Sheffield. Imaginary geometry III: reversibility of SLE_κ for $\kappa \in (4,8)$. *Ann. of Math. (2)*, 184(2):455–486, 2016.

[MS16d] Jason Miller and Scott Sheffield. Liouville quantum gravity and the Brownian map III: the conformal structure is determined. *Preprint arXiv:1608.05391*, 2016.

[MS17] Jason Miller and Scott Sheffield. Imaginary geometry IV: interior rays, whole-plane reversibility, and space-filling trees. *Probab. Theory Related Fields*, 169(3-4):729–869, 2017.

[MSW17] Jason Miller, Scott Sheffield, and Wendelin Werner. CLE percolations. *Forum Math. Pi*, 5:e4, 102, 2017.

[Mul67] Ronald C. Mullin. On the enumeration of tree-rooted maps. *Canadian J. Math.*, 19:174–183, 1967.

[Pem91] Robin Pemantle. Choosing a spanning tree for the integer lattice uniformly. *Ann. Probab.*, 19(4):1559–1574, 1991.

[Pit82] Loren D. Pitt. Positively correlated normal variables are associated. *Ann. Probab.*, 10(2):496–499, 1982.

[Pow18] Ellen Powell. Critical Gaussian chaos: convergence and uniqueness in the derivative normalisation. *Electron. J. Probab.*, 23:31(26p), 2018.

[Rem20] Guillaume Remy. The Fyodorov-Bouchaud formula and Liouville conformal field theory. *Duke Math. J.*, 169(1):177–211, 2020.

[RS05] Steffen Rohde and Oded Schramm. Basic properties of SLE. *Ann. of Math. (2)*, 161(2):883–924, 2005.

[RV10a] Rémi Rhodes and Vincent Vargas. Multidimensional multifractal random measures. *Electron. J. Probab.*, 15:no. 9, 241–258, 2010.

[RV10b] Raoul Robert and Vincent Vargas. Gaussian multiplicative chaos revisited. *Ann. Probab.*, 38(2):605–631, 2010.

[RV11] Rémi Rhodes and Vincent Vargas. KPZ formula for log-infinitely divisible multifractal random measures. *ESAIM Probab. Stat.*, 15:358–371, 2011.

[RV19] Rémi Rhodes and Vincent Vargas. The tail expansion of Gaussian multiplicative chaos and the Liouville reflection coefficient. *Ann. Probab.*, 47(5):3082–3107, 2019.

[RY99] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.

[Sch00] Oded Schramm. Scaling limits of loop-erased random walks and uniform spanning trees. *Israel J. Math.*, 118:221–288, 2000.

[Sha16] Alexander Shamov. On Gaussian multiplicative chaos. *J. Funct. Anal.*, 270(9):3224–3261, 2016.

[She07] Scott Sheffield. Gaussian free fields for mathematicians. *Probab. Theory Related Fields*, 139(3-4):521–541, 2007.

[She09] Scott Sheffield. Exploration trees and conformal loop ensembles. *Duke Math. J.*, 147(1):79–129, 2009.

[She16a] Scott Sheffield. Conformal weldings of random surfaces: SLE and the quantum gravity zipper. *Ann. Probab.*, 44(5):3474–3545, 2016.

[She16b] Scott Sheffield. Quantum gravity and inventory accumulation. *Ann. Probab.*, 44(6):3804–3848, 2016.

[Smi01] Stanislav K. Smirnov. Critical percolation in the plane: conformal invariance, Cardy's formula, scaling limits. *C. R. Acad. Sci. Paris Sér. I Math.*, 333(3):239–244, 2001.

[SS13] Oded Schramm and Scott Sheffield. A contour line of the continuum Gaussian free field. *Probab. Theory Related Fields*, 157(1-2):47–80, 2013.

[Ste05] Kenneth Stephenson. *Introduction to circle packing.* Cambridge University Press, Cambridge, 2005. The theory of discrete analytic functions.

[SW05] Oded Schramm and David B. Wilson. SLE coordinate changes. *New York J. Math.*, 11:659–669, 2005.

[SW16] Scott Sheffield and Menglu Wang. Field-measure correspondence in liouville quantum gravity almost surely commutes with all conformal maps simultaneously. *Preprint arXiv:1605.06171*, 2016.

[Wat93] Yoshiyuki Watabiki. Analytic study of fractal structure of quantized surface in two-dimensional quantum gravity. *Progress of Theoretical Physics Supplement*, 114:1–17, 1993.

[Wil96] David Bruce Wilson. Generating random spanning trees more quickly than the cover time. In *Proceedings of the Twenty-eighth Annual ACM Symposium on the Theory of Computing (Philadelphia, PA, 1996)*, pages 296–303. ACM, New York, 1996.

[Won19] Mo Dick Wong. Universal tail profile of Gaussian multiplicative chaos. *Preprint arXiv:1902.04054*, 2019.

[WP20] Wendelin Werner and Ellen Powell. Lecture notes on the Gaussian free field. *Preprint arXiv:2004.04720*, 2020.

[Zei15] Ofer Zeitouni. Gaussian fields (notes for lectures). available on the author's website, 2015.

Index