Some Exactly Solvable Models And Their Asymptotics

Mark Rychnovsky

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2021

Abstract

Some exactly solvable models and their asymptotics

Mark Rychnovsky

In this thesis we present three projects studying exactly solvable models in the KPZ universality class and one project studying a generalization of the SIR model from epidemiology. The first chapter gives an overview of the results and how they fit into the study of KPZ universality when applicable. Each of the following 4 chapters corresponds to a published or submitted article.

In the first project we study an oriented first passage percolation model for the evolution of a river delta. We show that at any fixed positive time, the width of a river delta of length L approaches a constant times $L^{2/3}$ with Tracy-Widom GUE fluctuations of order $L^{4/9}$. This result can be rephrased in terms of a particle system generalizing pushTASEP. We introduce an exactly solvable particle system on the integer half line and show that after running the system for only finite time the particle positions have Tracy-Widom fluctuations.

In the second project we study n-point sticky Brownian motions: a family of n diffusions that evolve as independent Brownian motions when they are apart, and interact locally so that the set of coincidence times has positive Lebesgue measure with positive probability. These diffusions can also be seen as n random motions in a random environment whose distribution is given by so-called stochastic flows of kernels. For a specific type of sticky interaction, we prove exact formulas characterizing the stochastic flow and show that in the large deviations regime, the random fluctuations of these stochastic flows are Tracy-Widom GUE distributed. An equivalent

formulation of this result states that the extremal particle among n sticky Brownian motions has Tracy-Widom distributed fluctuations in the large n and large time limit. These results are proved by viewing sticky Brownian motions as a diffusive limit of the exactly solvable beta random walk in random environment.

In the third project we study a class of probability distributions on the six-vertex model, which originate from the higher spin vertex model. For these random six-vertex models we show that the behavior near their base is asymptotically described by the GUE-corners process.

In the fourth project we study a model for the spread of an epidemic. This model generalizes the classical SIR model to account for inhomogeneity in the infectiousness and susceptibility of individuals in the population. A first statement of this model is given in terms of infinitely many coupled differential equations. We show that solving these equations can be reduced to solving a one dimensional first order ODE, which is easy to solve numerically. We use the explicit form of this ODE to characterize the total number of people who are ever infected before the epidemic dies out. This model is not related to the KPZ universality class.

Table of Contents

Acknowledgements

I am deeply indebted to my advisor Ivan Corwin. He has been a source of constant support and of invaluable insights. He has constantly helped me to seek a broader view and a more intuitive grasp of my field of study.

I am also thankful to Guillaume Barraquand and Evgeni Dimitrov for great discussions and extremely detailed answers to my many questions. Both Guillaume Barraquand and Evgeni Dimitrov have been superb collaborators from which I have learned an enormous amount. I have also benefited greatly from my interactions with fellow graduate students at Columbia including Sayan Das, Promit Ghosal, Yier Lin, Shalin Parekh, Xuan Wu, and Weitao Zhu.

More broadly I would like to thank Maribel Bueno Cachadina for introducing me to proof based math at the beginning of my undergraduate studies. Thank you to Maribel and to Padraic Bartlett for giving me so many opportunities to thrive in math. Thank you Kyle Kawagoe for constantly getting me excited about new problems.

Finally I would like to thank my parents for all of their support.

Chapter 1: Introduction

1.1 KPZ universality

In 1986 Kardar, Parisi, and Zhang [116] studied the time evolution of randomly growing interfaces a prototypical model: the KPZ equation. The KPZ equation is a stochastic partial differential equation with one spatial dimension and one time dimension describing the height of an interface $h(t, x)$ at time $t \geq 0$ above position $x \in \mathbb{R}$. The equation reads

$$\partial_t h(t, x) = \partial_{xx} h + (\partial_x h(t, x))^2 + \zeta(t, x),$$

where $\zeta(t, x)$ is a space-time white noise. They predicted that in the long time t scaling limit many growing interfaces would share important behavior with the KPZ equation, particularly that the interface would have fluctuations of scale $t^{1/3}$ and that these fluctuations should decorrelate at spatial scale $t^{2/3}$. This prediction was quite prescient and a broad class of models with this scaling behavior are now called the KPZ universality class.

Though originally conceived as a class of growing interfaces, the KPZ unversality class also contains a wide range of interacting particle systems, random matrices, traffic models, directed polymers, and stochastic PDEs. Each of these models can be transformed to reveal a growing interface which contains:

1. Growth in the direction normal to the interface. In the KPZ equation this is given by the $(\partial_x h(t, x))^2$ term.

2. A smoothing force similar to surface tension, so that deep valleys and sharp peaks tend to shrink. In the KPZ equation this is given by the $\partial_{xx} h(t, x)$ term.

3. A random driving force with short scale correlations in space and time. In the KPZ equation this is given by the space-time white noise $\zeta(t, x)$.

In addition to shared scaling behavior, numerics and some theoretical results have demonstrated universality for the probability distributions describing the $t^{1/3}$ scale fluctuations. Several distributions from random matrix theory are possible depending on the initial conditions of the model, but of particular interest to us is the Tracy-Widom distribution which corresponds to an interface growing from a single point or "droplet".

Although many models lie in the KPZ universality class, much of our understanding comes from a few "exactly solvable" models whose algebraic structure allows for the derivation of exact formulas. These exact formulas give a way to study the universal limit behavior, and when the limit behavior of an exactly solvable model is proven, it suggests that other similar models have similar limit behavior.

Chapters 2,3, and 4 of this thesis will present projects which analyze exactly solvable models in the KPZ universality class.

1.1.1 The Tracy-Widom GUE distribution

The Tracy-Widom Gaussian unitary ensumble (GUE) distribution gives the fluctuations we expect to see when a growth model in the KPZ universality class starts from a single point. For example if a fire is started in the center of piece of paper, the fluctuations between the burnt and unburnt regions of the paper will be Tracy-Widom GUE distributed in the long time limit.

This distribution first appeared in the context of random matrix theory in [189]. We will introduce it in this context. Let A be an n by n matrix with independent identically distributed complex, mean 0, variance 1 Gaussian entrees. Then the Hermitian random matrix

$$M = \frac{A + A^*}{\sqrt{2}}$$

is a GUE distributed random matrix. Because this matrix is Hermitian, all of its eigenvalues are

real. This probability measure on matrices is called the Gaussian Unitary Ensemble (GUE).

As the size n of a GUE matrix goes to ∞, the largest eigenvalue has asymptotics

$$\lambda_1 = 2\sqrt{n} + n^{-1/6}\chi_{\mathrm{TW}},$$

where χ_{TW} has the Tracy-Widom GUE distribution.

A careful study of this limit can be used to arrive at a formula for the Tracy-Widom GUE distribution which we give now.

Definition 1.1.1. For any contour C and any measurable function $K : C \times C \to \mathbb{C}$, which we will call a kernel, the Fredholm determinant $\det(1 + K)_{L^2(C)}$ is defined by

$$\det(1 + K)_{L^2(C)} = 1 + \sum_{k=1}^{\infty} \frac{1}{k!} \int_{C^k} \det(K(x_i, x_j))_{1 \le i,j \le k} \prod_{i=1}^{k} dx_i, \tag{1.1}$$

provided the right hand side converges absolutely.

Consider the kernel

$$K(x, y) = \frac{Ai(x)Ai'(y) - Ai'(x)Ai(y)}{x - y},$$

where $Ai(x)$ is the Airy function

$$Ai(x) = \frac{1}{2\pi \mathbf{i}} \int_{\infty e^{-\mathbf{i}\pi/3}}^{\infty e^{\mathbf{i}\pi/3}} e^{z^3/3 - zx} dz. \tag{1.2}$$

The integration bounds in the definition of the Airy function are shorthand for integrating over a contour which goes from $\infty e^{-\mathbf{i}\pi/3}$ to 0, then from 0 to $\infty e^{\mathbf{i}\pi/3}$.

The Tracy-Widom distribution has distribution function

$$F_{\mathrm{GUE}}(y) = \det(1 - K_{\mathrm{Ai}})_{L^2(y,\infty)}. \tag{1.3}$$

A few years after its discovery in the context of random matrix theory the Tracy-Widom distribution was found to describe the limit behavior of the longest increasing subsequence of a permu-

tation in [16] and the current across an edge of an interacting particle system in [112]. These were the first major hints of the role of the Tracy-Widom distribution in the KPZ universality class and were quickly followed by many other appearances of the Tracy-Widom GUE distribution.

1.2 Directed polymers

Directed polymers were first introduced in [108] in the context of statistical physics and received their first rigorous mathematical treatment in [109]. A directed polymer in $(1+1)$ dimension is a random probability measure on up-right paths in $\mathbb{Z}^2_{\geq 0}$. There are two sources of randomness. The first is a random environment which comes from assigning random variables w_v to either each vertex (or w_e to each edge) in $\mathbb{Z}^2_{\geq 0}$. Once the environment is chosen, the weight of an up-right path is given by the product of the random variables assigned to each vertex (or to each edge) in the path raised to the power of the inverse temperature $\beta > 0$. In the final measure on paths, the probability of choosing any path π is proportional to the weight of that path. It is worth noting that in the zero temperature limit ($\beta \to \infty$) only the path of maximum weight will occur. If all the w_v (or w_e) are greater than 1 this zero temperature limit is a last passage percolation model with weights $\log(w_e)$, and if w_v (or w_e) are less than 1 the limit is a first passage percolation model with weights $-\log(w_e)$.

One prediction of KPZ universality is that the free energy of a wide class of directed polymers should have Tracy-Widom GUE fluctuations with scaling exponent $1/3$. The first positive temperature exactly solvable polymer model on the lattice was the (vertex weighted) log-gamma polymer introduced in [170]. The scale of the fluctuations of its free energy was confirmed to be $1/3$ in the same paper, and the fluctuations were shown to be Tracy-Widom distributed in [66, 42].

We are particularly interested in an exactly solvable (edge weighted) polymer called the beta random walk in random environment (RWRE). This model was introduced and its fluctuations were shown to be Tracy-Widom distributed with scaling exponent $1/3$ in [23].

1.2.1 The beta random walk in random environment

In this section we define the beta random walk in random environment, which is a probability measure on directed lattice paths in $\mathbb{Z} \times \mathbb{Z}_{\geq 0}$. We will be particularly interested in two limits of the beta random walk in random environment for which we prove KPZ type limit theorems.

Definition 1.2.1. The beta random walk in random environment (beta RWRE) depends on two parameters $\alpha > 0$ and $\beta > 0$. Let $\{w_{(x,t)}\}_{x \in \mathbb{Z}, t \in \mathbb{Z}_{\geq 0}}$ be iid beta distributed random variables with parameters α, β. Recall that a beta random variable w with parameters $\alpha, \beta > 0$ is defined by

$$\mathbb{P}(w \in dx) = \mathbb{1}_{x \in [0,1]} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} dx,$$

where $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$. We will call the values of the random variables $w_{(x,t)}$ for all $x \in \mathbb{Z}, t \in \mathbb{Z}_{\geq 0}$ the random environment.

After sampling the random environment, we begin k independent random walks $(X_1(t), ..., X_k(t))$ from position $\vec{x}_0$. Each random walker has jump distribution

$$\mathsf{P}(X(t+1) = x+1 | X(t) = x) = w_{(x,t)} \qquad \mathsf{P}(X(t+1) = x-1 | X(t) = x) = 1 - w_{(x,t)}.$$

We will use $\vec{X}^{\vec{x}}(t) = (X_1^{x_1}(t), ..., X_k^{x_k}(t))$ to refer to the position of k independent random walks started from $(x_1, ..., x_k)$ at time t. Unless another initial condition is specified, $\vec{X}(t) = (X_1(t), ..., X_k(t))$ will refer to the position of k random walkers started from the origin.

We will interpret this as a physical system where the random walks are trajectories of indistinguishable particles and the random environment accounts for microscopic fluctuations we cannot measure. Mathematically this means that we cannot tell what environment we are in, so we will average our probability measure on particle trajectories over all realizations of the random environment. Before averaging, the transition probabilities of our particles are called quenched, and after averaging the transition probabilities are called annealed.

We use the symbol P with bold font for the quenched probability measure on paths, which is

obtained by conditioning on the environment. The usual symbols $\mathbb{P}$ (resp. $\mathbb{E}$) will be used to denote the measure (resp. the expectation) of the environment.

Explicitly the annealed measure on particle trajectories says that when n particles are in the same position x at time t, then exactly k of these particles move to position $x+1$ (and the rest move to position $x-1$) at time $t+1$ with probability

$$\phi(k|n) = \mathbb{E}\left[\binom{n}{k} w_{x,t}^k (1-w_{x,t})^{n-k}\right] = \binom{n}{k} \frac{(\alpha)_k (\beta)_{n-k}}{(\alpha+\beta)_n}, \tag{1.4}$$

where $(\alpha)_k = \alpha(\alpha+1)...(\alpha+k-1)$ is the ascending Pochhammer symbol.

Note that any single annealed trajectory of the beta RWRE is just a simple random walk and the random environment has no effect. However, if we consider multiple annealed trajectories, then even though we have averaged out the environment, the paths are still correlated by the fact that they were run through the same environment pre-averaging, not through independent copies. In particular, they do not behave as simple random walks when they meet.

Central to the exact solvability of the beta RWRE is a non-commutative binomial formula related to the transition probabilities in (1.4).

Theorem 1.2.2. *If X and Y generate an associative algebra and satisfy the commutation relation*

$$YX = \frac{1}{\alpha+\beta+1} XX + \frac{\alpha+\beta-1}{\alpha+\beta+1} XY + \frac{1}{\alpha+\beta+1} YY,$$

then we have the following non-commutative binomial identity:

$$\left(\frac{\alpha}{\alpha+\beta} X + \frac{\beta}{\alpha+\beta} Y\right)^n = \sum_{k=0}^{n} \binom{n}{k} \frac{(\alpha)_k (\beta)_{n-k}}{(\alpha+\beta)_n} X^j Y^{n-j}. \tag{1.5}$$

The commutativity relation in this theorem is equivalent to the $n=2$ case of the identity.

Loosely speaking, thinking of X and Y as operators acting on a collection of n-particles, this identity allows one to factor the generator for the beta RWRE for n-particles (corresponding to RHS of (1.5)) into a product of n copies of the beta RWRE generator acting on a single particle

(corresponding to LHS of (1.5)) provided the one particle generators satisfy a boundary condition when two particles are in the same position (corresponding to the $n = 2$ case of (1.5)). This is part of a procedure called the coordinate Bethe ansatz which is one common method of producing exactly solvable models.

The beta RWRE was originally introduced in [23] where exact formulas, and KPZ limit theorems for the model were derived. They realized the identity (1.5) as a special case of a more general non-commutative binomial formula which [158] proved and used to develop a more general exactly solvable model called the q-Hahn TASEP. We will be particularly interested in two results from [23].

Consider the quenched probability $\mathsf{P}(X(t) > x)$ in the beta random walk in random environment, where $X(t)$ is the path of a single particle that starts from 0 at time 0. It satisfies the following formula which will be a starting point for producing exact formulas for two degenerations of the beta RWRE that we study.

Theorem 1.2.3 ([23, Theorem 1.13])**.** *For $u \in \mathbb{C} \setminus \mathbb{R}_{>0}$ and $\alpha, \beta > 0$, fix $t \in \mathbb{Z}_{\geq 0}$ and $x \in \{-t, ..., t\}$ with the same parity. Then*

$$\mathbb{E}[e^{u\mathsf{P}(X(t)>x)}] = \det(I - K_u^{\mathrm{RW}})_{\mathbb{L}^2(C_0)},$$

where C_0 is a small positively oriented contour that contains 0 and does not contain the points $-\alpha - \beta$ and -1, and $K_u^{\mathrm{RW}} : \mathbb{L}^2(C_0) \to \mathbb{L}^2(C_0)$ is defined in terms of its kernel

$$K_u^{\mathrm{RW}}(v, v') = \frac{1}{2\pi\mathbf{i}} \int_{\frac{1}{2}-\mathbf{i}\infty}^{\frac{1}{2}+\mathbf{i}\infty} \frac{\pi}{\sin(\pi s)} (-u)^s \frac{g^{\mathrm{RW}}(v)}{g^{\mathrm{RW}}(v+s)} \frac{ds}{s + v - v'},$$

where

$$g^{\mathrm{RW}}(v) = \left(\frac{\Gamma(v)}{\Gamma(\alpha + v)}\right)^{(t-x)/2} \left(\frac{\Gamma(\alpha + \beta + v)}{\Gamma(\alpha + v)}\right)^{(t+x)/2} \Gamma(v).$$

We also draw attention to the KPZ limit theorem.

Theorem 1.2.4. *[23, Theorem 1.15] For $0 < \theta < 1/2$ and $\alpha = \beta = 1$, we have that*

$$\lim_{t\to\infty} \mathbb{P}\left(\frac{\log\left(\mathsf{P}\big(X(t) > x(\theta)t\big)\right) + I\big(x(\theta)\big)t}{t^{1/3}\sigma\big(x(\theta)\big)} \leqslant y \right) = F_{\mathrm{GUE}}(y), \tag{1.6}$$

where

$$x(\theta) = \frac{1+2\theta}{\theta^2+(\theta+1)^2}, \qquad I(x(\theta)) = \frac{1}{\theta^2+(\theta+1)^2}, \qquad \sigma(\theta) = \left(\frac{2I(x(\theta))^2}{1-I(x(\theta))}\right)^{\frac{1}{3}}$$

In the following sections we will discuss two limits of the beta RWRE. Theorem 1.2.3 will give rise to exact formulas for both of these models. Much of our attention in the next two sections will focus on proving analogs of Theorem 1.2.4 for these limits of the beta RWRE.

1.2.2 Bernoulli-exponential first passage percolation

This section is an overview of the work in Chapter 2. We are interested in the behavior of a zero temperature limit of the beta random walk in random environment called Bernoulli-exponential first passage percolation (FPP). Bernoulli-exponential FPP models the growth of a river delta in $\mathbb{Z}^2_{\geq 0}$. The model depends on two parameters $a, b > 0$. At time 0, we begin with a river which starts from the origin and follows an up-right path in $\mathbb{Z}^2_{\geq 0}$ chosen so that at each step the river goes up with probability $a/(a+b)$ and right with probability $b/(a+b)$(thick black line in Figure 1.1). This means the initial river will have approximate slope a/b.

As time passes the river erodes its banks creating forks. At each vertex which the river leaves in the rightward (respectively upward) direction, it takes an amount of time distributed as an exponential random variable with rate a (resp. b) for the river to erode through its upward (resp. rightward) bank. Once the river erodes one of its banks at a vertex, the flow at this vertex branches to create a tributary (see gray paths in Figure 1.1). The path of the tributary is selected by the same rule as the path of the time 0 river, except that when the tributary meets an existing river it joins the river and follows the existing path. The full path of the tributary is added instantly when the river

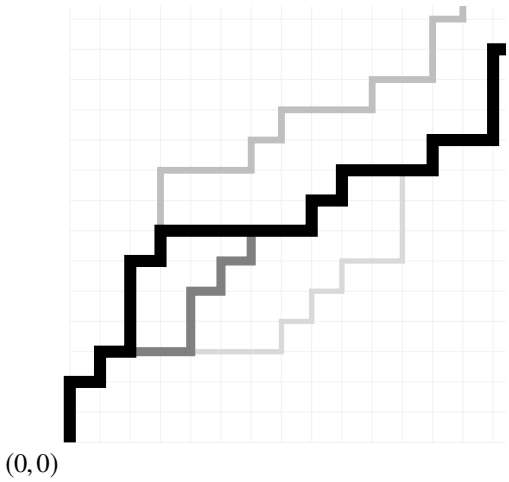

Figure 1.1: A sample of the river delta (Bernoulli-exponential FPP percolation cluster) near the origin. The thick black random walk path corresponds to the river (percolation cluster) at time 0. The other thinner and lighter paths correspond to tributaries added to the river delta (percolation cluster) at later times.

erodes its bank.

In this model the river is infinite, and the main object of study is the set of vertices included in the river at time t, i.e. the percolation cluster. We will also refer to the shape enclosed by the outermost tributaries at time t as the river delta (see Figure 1.2 for a large scale illustration of the river delta).

Now we give a more precise definition in terms of first passage percolation following [23].

Definition 1.2.5 (Bernoulli-exponential first passage percolation)**.** Let E_e be a family of independent exponential random variables indexed by the edges e of the lattice $\mathbb{Z}^2_{\geq 0}$. Each E_e is distributed as an exponential random variable with parameter a if e is a vertical edge, and with parameter b if e is a horizontal edge. Let $(\zeta_{i,j})$ be a family of independent Bernoulli random variables with parameter $b/(a+b)$. We define the passage time t_e of each edge e in the lattice $\mathbb{Z}^2_{\geq 0}$ by

$$t_e = \begin{cases} \zeta_{i,j} E_e & \text{if } e \text{ is the vertical edge } (i,j) \to (i,j+1), \\ (1-\zeta_{i,j}) E_e & \text{if } e \text{ is the horizontal edge } (i,j) \to (i+1,j). \end{cases}$$

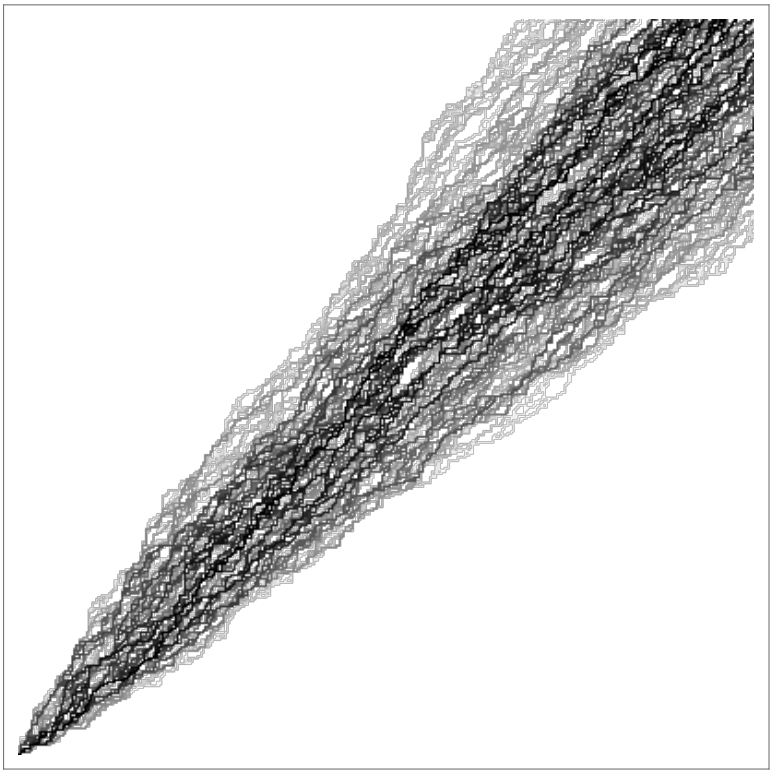

Figure 1.2: The percolation cluster for 400×400 Bernoulli-exponential FPP at time 1 with $a = b = 1$. Paths occurring earlier are shaded darker, so the darkest paths occur near $t = 0$ and the lightest paths occur near $t = 1$.

We define the point to point passage time $T^{\text{PP}}(n,m)$ by

$$T^{\text{PP}}(n,m) = \min_{\pi:(0,0)\to(n,m)} \sum_{e\in\pi} t_e.$$

where the minimum is taken over all up-right paths π from $(0,0)$ to (n,m). We define the percolation cluster $C(t)$, at time t, by

$$C(t) = \left\{(n,m) : T^{\text{PP}}(n,m) \le t\right\}.$$

At each time t, the percolation cluster $C(t)$ is the set of points visited by a collection of up-right random walks in the quadrant $\mathbb{Z}^2_{\ge 0}$. $C(t)$ evolves in time as follows:

- At time 0, the percolation cluster contains all points in the path of a directed random walk starting from $(0,0)$, because at any vertex (i,j) we have passage time 0 to either $(i,j+1)$ or $(i+1,j)$ according to the independent Bernoulli random variables $\zeta_{i,j}$.
- At each vertex (i,j) in the percolation cluster $C(t)$, with an upward (resp. rightward) neighbor outside the cluster, we add a random walk starting from (i,j) with an upward (resp. rightward) step to the percolation cluster with exponential rate a (resp. b). This random walk will almost surely hit the percolation cluster after finitely many steps, and we add to the percolation cluster only those points that are in the path of the walk before the first hitting point (see Figure 1.1).

Define the height function $H_t(n)$ by

$$H_t(n) = \sup\{m \in \mathbb{Z}_{\ge 0} | T^{\text{PP}}(n,m) \le t)\}, \tag{1.7}$$

so that $(n, H_t(n))$ is the upper border of $C(t)$.

Bernoulli-exponential first passage percolation is the zero temperature limit of the beta random walk in random environment in the following sense: Set $\alpha_\varepsilon = \varepsilon a$ and $\beta_\varepsilon = \varepsilon\beta$ and refer to the parameter $\varepsilon > 0$ as temperature. If $X(t)$ is a beta RWRE with parameters $\alpha_\varepsilon, \beta_\varepsilon$ and $H_t(n)$ is the

height function of a Bernoulli-exponential first passage percolation with parameters a, b, then for all $m, n \geq 0$, we have

$$-\varepsilon \log\left((\mathsf{P}(X(n+m) = m-n)\right) \xrightarrow{\varepsilon \to 0} T^{\mathrm{PP}}(n,m),$$

in distribution. Recall $\mathsf{P}(X(n+m) = m-n)$ is the probability that the the beta RWRE $X(t)$ is at position $m-n$ at time $n+m$, and that this probability is random as it depends on the random environment. The change of variables in this limit come from rotating the beta RWRE 45 degrees and scaling by $\frac{1}{\sqrt{2}}$ so that the walk trajectories become up-right paths in $\mathbb{Z}^2_{\geq 0}$.

Results

The study of the large scale behavior of passage times $T^{\mathrm{PP}}(n,m)$ was initiated in [23]. At large times, the upper border of the percolation cluster (described by the height function $H_t(n)$) has GUE Tracy-Widom fluctuations on the scale $n^{1/3}$.

Theorem 1.2.6 ([23, Theorem 1.19])**.** *Fix parameters $a, b > 0$. For any $\theta > 0$ and $x \in \mathbb{R}$,*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{H_{\tau(\theta)n} - \kappa(\theta)n}{\tilde{\rho}(\theta)n^{1/3}} \leq x\right) = F_{\mathrm{GUE}}(x), \tag{1.8}$$

where F_{GUE} is the GUE Tracy-Widom distribution (see Definition 2.2.3) and $\kappa(\theta)$, $\tau(\theta)$, $\tilde{\rho}(\theta) = \frac{\kappa'(\theta)}{\tau'(\theta)}\rho(\theta)$ are functions defined in [23] by

$$\begin{aligned}
\kappa(\theta) &:= \frac{\frac{1}{\theta^2} - \frac{1}{(a+\theta)^2}}{\frac{1}{(a+\theta)^2} - \frac{1}{(a+b+\theta)^2}},\\
\tau(\theta) &:= \frac{1}{a+\theta} - \frac{1}{\theta} + \kappa(\theta)\left(\frac{1}{a+\theta} - \frac{1}{a+b+\theta}\right) = \frac{a(a+b)}{\theta^2(2a+b+2\theta)},\\
\rho(\theta) &:= \left[\frac{1}{\theta^3} - \frac{1}{(a+\theta)^3} + \kappa(\theta)\left(\frac{1}{(a+b+\theta)^3} - \frac{1}{(a+\theta)^3}\right)\right]^{1/3}.
\end{aligned}$$

Note that as θ ranges from 0 to ∞, $\kappa(\theta)$ ranges from $+\infty$ to a/b and $\tau(\theta)$ ranges from $+\infty$ to 0.

Remark 1.2.7. In [23] the limit theorem is incorrectly stated as

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\min_{i\le n} T^{\mathrm{PP}}(i,\kappa(\theta)n) - \tau(\theta)n}{\rho(\theta)n^{1/3}} \le x\right) = F_{\mathrm{GUE}}(x),$$

but following the proof in [23, Section 6.1], we can see that the inequality and the sign of x should be reversed. Further, we have reinterpreted the limit theorem in terms of height function $H_t(n)$ instead of passage times $T^{\mathrm{PP}}(n,m)$ using the relation (1.7).

We are interested in the fluctuations of $H_t(n)$ for large n but fixed time t. Let us scale θ in (1.8) above as

$$\theta = \left(\frac{na(a+b)}{2t}\right)^{1/3},$$

so that

$$\tau(\theta)n = t + O(n^{-1/3}).$$

Let us introduce constants

$$\lambda = \left(\frac{a(a+b)}{2t}\right)^{1/3}, \quad d = \frac{3a(a+b)}{2b\lambda}, \quad \sigma = \left(\frac{3a(a+b)\lambda}{2b^3}\right)^{1/3}. \tag{1.9}$$

Then, we have the approximations

$$\kappa(\theta)n = \frac{a}{b}n + dn^{2/3} + o(n^{4/9}),$$
$$\tilde{\rho}(\theta)n^{1/3} = \sigma n^{4/9} + o(n^{4/9}).$$

Thus, formally letting θ and n go to infinity in (1.8) suggests that for a fixed time t, it is natural to scale the height function as

$$H_t(n) = \frac{a}{b}n + dn^{2/3} + \sigma n^{4/9}\chi_n,$$

and study the asymptotics of the sequence of random variables χ_n.

We find the following.

Theorem 1.2.8. *Fix parameters $a, b > 0$. For any $t > 0$ and $x \in \mathbb{R}$,*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{H_t(n) - \frac{a}{b}n - dn^{2/3}}{\sigma n^{4/9}} \leq x \right) = F_{GUE}(x),$$

where F_{GUE} is the GUE Tracy-Widom distribution.

Note that the heuristic argument presented above to guess the scaling exponents and the expression of constants d and σ is not rigorous, since Theorem 1.2.6 holds for fixed θ. Theorem 1.2.6 could be extended without much effort to a weak convergence uniform in θ for θ varying in a fixed compact subset of $(0, +\infty)$. However the case of θ and n simultaneously going to infinity requires more careful analysis. Indeed, for θ going to infinity very fast compared to n, Tracy-Widom fluctuations would certainly disappear as this would correspond to considering the height function at time $\tau(\theta)n \approx 0$, which corresponds to a simple random walk and gives Gaussian fluctuations on the $n^{1/2}$ scale.

The scaling exponents in Theorem 1.2.8 might seem unusual, but the preceding heuristic computation explains how they result from rescaling a model which has the usual KPZ scaling exponents. A similar situation occurs for scaling exponents of the height function of directed last passage percolation in thin rectangles [17, 33] and for the free energy of directed polymers [13] under the same limit.

1.2.3 Other interpretations of the model

There are several equivalent interpretations of Bernoulli-exponential first passage percolation. We will present the most interesting here.

A particle system on the integer line

The height function of the percolation cluster $H_t(n)$ is equivalent to the height function of an interacting particle system we call geometric jump pushTASEP, which generalizes pushTASEP (the $R = 0$ limit of PushASEP introduced in [44]) by allowing jumps of length greater than 1. This

model is similar to Hall-Littlewood pushTASEP introduced in [91], but has a slightly different particle interaction rule.

Definition 1.2.9 (Geometric jump pushTASEP)**.** Let Geom(q) denote a geometric random variable with $\mathbb{P}(\text{Geom}(q) = k) = q^k(1-q)$. Let $1 \leq p_1(t) < p_2(t) < \ldots < p_i(t) < \ldots$ be the positions of ordered particles in $\mathbb{Z}_{\geq 1}$. At time $t = 0$ the position $n \in \mathbb{Z}_{\geq 0}$ is occupied with probability $b/(a+b)$. Each particle has an independent exponential clock with parameter a, and when the clock corresponding to the particle at position p_i rings, we update each particle position p_j in increasing order of j with the following procedure. ($p_i(t-)$ denotes the position of particle i infinitesimally before time t.)

- If $j < i$, then p_j does not change.
- p_i jumps to the right so that the difference $p_i(t) - p_i(t-)$ is distributed as $1 + \text{Geom}(a/(a+b))$
- If $j > i$, then
  - If the update for $p_{j-1}(t)$ causes $p_{j-1}(t) \geq p_j(t-)$, then $p_j(t)$ jumps right so that $p_j(t) - p_{j-1}(t)$ is distributed as $1 + \text{Geom}(a/(a+b))$.
  - Otherwise p_j does not change.
  - All the geometric random variables in the update procedure are independent.

Another way to state the update rule is that each particle jumps with exponential rate a, and the jump distance is distributed as $1 + \text{Geom}(a/(a+b))$. When a jumping particle passes another particle, the passed particle is pushed a distance $1 + \text{Geom}(a/(a+b))$ past the jumping particle's ending location (see Figure 1.3).

The height function $\overline{H}_t(n)$ at position n and time t is the number of unoccupied sites weakly to the left of n. If we begin with the distribution of $(n, H_t(n))$ in our percolation model, and rotate the first quadrant clockwise 45 degrees, the resulting distribution is that of $(n, \overline{H}_t(n))$. The horizontal

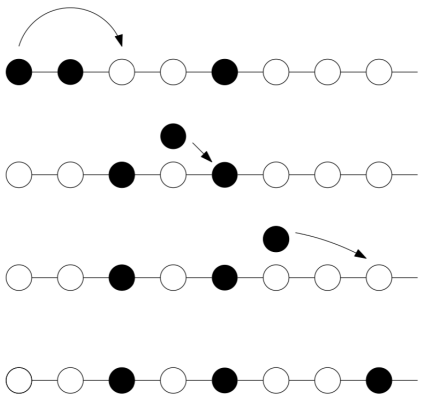

Figure 1.3: This figure illustrates a single update for geometric jump pushTASEP. The clock corresponding to the leftmost particle rings, activating the particle. The first particle jumps 2 steps pushing the next particle and activating it. This particle jumps 1 step pushing the rightmost particle and activating it. The rightmost particle jumps 3 steps, and all particles are now in their original order, so the update is complete.

segments in the upper border of the percolation cluster correspond to the particle positions, thus

$$H_t(n) = p_t(n) - n = \sup\{k : \overline{H}_t(n+k) \geq k\}.$$

A direct translation of Theorem 2.1.4 gives:

Corollary 1.2.10. *Fix parameters $a, b > 0$. For any $t > 0$ and $x \in \mathbb{R}$,*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{p_t(n) - \left(\frac{a+b}{b}\right) n - dn^{2/3}}{\sigma n^{4/9}} \leq x \right) = F_{\mathrm{GUE}}(x),$$

where $F_{\mathrm{GUE}}(x)$ is the Tracy-Widom GUE distribution.

To the authors knowledge Corollary 1.2.10 is the first result in interacting particle systems showing Tracy-Widom fluctuations for the position of a particle at finite time.

Degenerations

If we set $b = 1, t' = t/a$, and let $a \to 0$, then in the new time variable t' each particle performs a jump with rate 1 and with probability going to 1, each jump is distance 1, and each push is distance 1. This limit is pushTASEP on $\mathbb{Z}_{\geq 0}$ where every site is occupied by a particle at time 0. Recall that in pushTASEP, the dynamics of a particle are only affected by the (finitely many) particles to its left, so this initial data makes sense.

We can also take a continuous space degeneration. Let x be the spatial coordinate of geometric jump pushTASEP, and let $\exp(\lambda)$ denote an exponential random variable with rate λ. Choose a rate $\lambda > 0$, and set $b = \frac{\lambda}{n}, x' = x/n, a = \frac{n-\lambda}{n}$, and let $n \to \infty$. Then our particles have jump rate $\frac{n-\lambda}{n} \to 1$, jump distance $\frac{\text{Geom}(1-\lambda/n)}{n} \to \exp(\lambda)$, and push distance $\frac{\text{Geom}(1-\lambda/n)}{n} \to \exp(\lambda)$. This is a continuous space version of pushTASEP on $\mathbb{R}_{\geq 0}$ with random initial conditions such that the distance between each particle position p_i and its rightward neighbor p_{i+1} is an independent exponential random variable of rate λ. Each particle has an exponential clock, and when the clock corresponding to the particle at position p_i rings, an update occurs which is identical to the update for geometric jump pushTASEP except that each occurrence of the random variable $1 + \text{Geom}(a/(a+b))$ is replaced by the random variable $\exp(\lambda)$.

1.3 Sticky Brownian motions

This section is an overview of the work in Chapter 3. We will study the behavior of a diffusive limit of the beta RWRE called sticky Brownian motion. The definition of sticky Brownian motion is somewhat technical and we will introduce some background first.

1.3.1 Definitions

Recall that the local time of a Brownian motion B_t at the point a is defined by the almost-sure limit

$$\ell_t^a(B) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^t \mathbb{1}_{a-\varepsilon \leq B_s \leq a+\varepsilon} ds = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^t \mathbb{1}_{a \leq B_s \leq a+\varepsilon} ds.$$

For a continuous semimartingale X_t, the natural time scale is given by its quadratic variation $\langle X, X\rangle_t$ and we define the local time as the almost sure limit [164, Corollary 1.9, Chap. VI]

$$\ell_t^a(X) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^t \mathbb{1}_{a \le X_s \le a+\varepsilon} d\langle X, X\rangle_t.$$

Feller initiated the study of Brownian motions sticky at the origin in [77], while studying general boundary conditions for diffusions on the half line.

Definition 1.3.1. *Brownian motion sticky at the origin* can be defined as the weak solution to the system of stochastic differential equations

$$\begin{aligned} dX_t &= \mathbb{1}_{\{X_t \neq 0\}} dB_t, \\ \int_0^t \mathbb{1}_{X_s = 0} ds &= \frac{1}{2\lambda} \ell_t^0(X), \end{aligned} \tag{1.10}$$

where B_t is a Brownian motion. Reflected Brownian motion sticky at the origin can be defined as $Y_t = |X_t|$ where X_t is a Brownian motion sticky at the origin.

Remark 1.3.2 (Time change). Brownian motion sticky at the origin can be viewed as a time change of Brownian motion in a construction due to Ito and McKean [110]. Consider the Brownian motion B_t, and define the continuous increasing function $A(t) = t + \frac{1}{2\lambda}\ell_t^0(B)$. Let $T(t) = A^{-1}(t)$ and set $X_t = B_{T(t)}$. We see that X_t is a usual Brownian motion when $X_t \neq 0$, because the local time of B_t only increases when $B_t = 0$. When $X_t = 0$ time slows down. We know $\int_0^t \mathbb{1}_{X_s > 0} ds = T(t)$, so $\int_0^t \mathbb{1}_{X_s = 0} ds = t - T(t) = \frac{1}{2\lambda}\ell_{T(t)}^0(B) = \frac{1}{2\lambda}\ell_t^0(X)$. This type of time change can be used to produce many processes with sticky interactions.

Remark 1.3.3 (Discrete limit). Reflected Brownian motion sticky at the origin Y_t can also be viewed as the diffusive limit of a sequence of random walks which tend to stay at 0. For small $\varepsilon > 0$, let Z_t^ε be a discrete time random walk on $\mathbb{Z}_{\ge 0}$, which behaves as a simple symmetric random walk when it is not at the point 0. When Z_t^ε is at the point 0, at each time step it travels to 1 with probability ε and stays at 0 with probability $1 - \varepsilon$. The diffusive limit $\varepsilon Z^{2\lambda\varepsilon}_{\varepsilon^{-2}t}$ converges to Y_t weakly as $\varepsilon \to 0$.

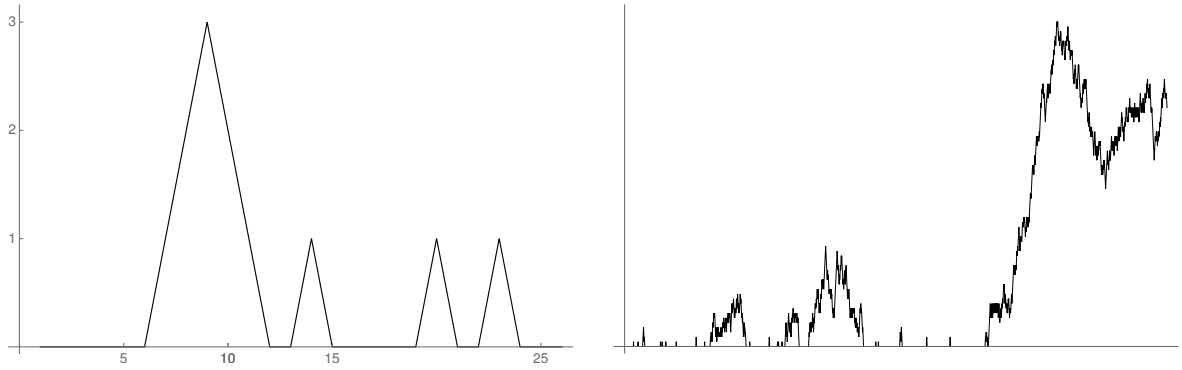

Figure 1.4: Left panel: Random walk $Z_t^{1/5}$ leaving 0 with probability $1/5$, up to time 25. Right panel: Reflected Brownian motion sticky at 0 obtained by the scaling limit of Z_t^{ε}.

To understand this convergence see equation (1.12), and note that the drift of the limiting motion at 0 is equal to 2λ because in each unit of time there are ε^{-2} opportunities to jump from 0 to ε and the proportion of these opportunities that is taken is approximately $2\lambda\varepsilon$. The analogous statement is also true for Brownian motion sticky at the origin. See Figure 1.4 where a simulation of $Z_t^{1/5}$ is shown alongside Y_t.

From Remark 1.3.2 and the Tanaka Formula for reflected Brownian motion it is easy to see that Y_t is a weak solution to the system of stochastic differential equations

$$\begin{aligned} dY_t &= \frac{1}{2} d\ell_t^0(Y) + \mathbb{1}_{\{Y_t>0\}} dB_t, \\ \mathbb{1}_{\{Y_t=0\}} dt &= \frac{1}{4\lambda} d\ell_t^0(Y), \end{aligned} \tag{1.11}$$

Equations (1.11) is equivalent to the single SDE

$$dY_t = 2\lambda \mathbb{1}_{\{Y_t=0\}} dt + \mathbb{1}_{\{Y_t>0\}} dB_t, \tag{1.12}$$

in the sense that a weak solution to one is a weak solution to the other [76]. Existence and uniqueness of weak solutions to (1.10) and (1.11) can be found in [76] and references therein.

Nonexistence of strong solutions to equations (1.10) and (1.11) was first shown in [57] and [197] (see also [76] for a more canonical arguments which would more easily generalize to other sticky processes). Several other works have been published on the existence of solutions to similar SDEs with indicator functions as the coefficient of dB_t or dt including [114, 29]. A more complete history of these SDEs can be found in [76].

We wish to study the evolution of n particles in one spatial dimension where the difference between any pair of particles is a Brownian motion sticky at the origin. First we do this for a pair of sticky Brownian motions.

Definition 1.3.4. The stochastic process $(X_1(t), X_2(t))$ is a pair of Brownian motions with sticky interaction if each X_i is marginally distributed as a Brownian motion and

$$\langle X_1, X_2 \rangle(t) = \int_0^t \mathbb{1}_{X_1(s)=X_2(s)} ds, \tag{1.13}$$

$$\int_0^t \mathbb{1}_{X_1(s)=X_2(s)} ds = \frac{1}{2\lambda} \ell_t^0(X_1 - X_2). \tag{1.14}$$

In other words $(X_1(t), X_2(t))$ are sticky Brownian motions if they evolve as independent Brownian motions when they are at different positions and their difference is a Brownian motion sticky at 0 (see a simulation in Fig. 1.5). The parameter λ can be understood as the rate (in a certain excursion theoretic sense) at which the two particles split when they are at the same position.

One can use Tanaka's formula to show that equation (1.14) is equivalent to saying

$$|X_1(t) - X_2(t)| - 2\lambda \int_0^t \mathbb{1}_{X_1(s)=X_2(s)} ds \tag{1.15}$$

is a martingale. Howitt and Warren [104] made this observation and generalized this martingale problem for a family of n particles with pairwise sticky interaction, which we call *n-point sticky Brownian motions*. In the most general case, the stickiness behaviour cannot be characterized uniquely by a single parameter λ. One needs to define for each $k, l \geq 1$ the "rate" at which a group of $k + l$ particles at the same position will split into two groups of respectively k and l coinciding

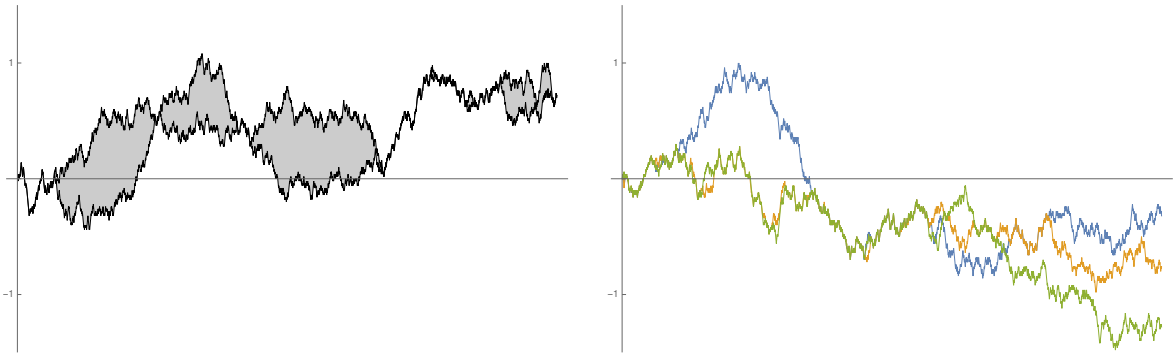

Figure 1.5: Left panel: Two Brownian motions with sticky interaction. Right panel: 3-point sticky Brownian motions. Not only do the paths stick pairwise, but sometimes all 3 paths may stick together. Both simulations are discretizations of sticky Brownian motions using the beta RWRE with $\varepsilon = 0.02$ (see Section 3.1.3).

particles. Following the notations in [104, 166, 169] this rate is denoted

$$\binom{k+l}{k}\theta(k,l).$$

Furthermore, we impose that the law of n-point sticky Brownian motions are consistent in the sense that any subsets of k particles for $k \leq n$ follow the law of the k-point sticky Brownian motions. This implies the relation $\theta(k+1,l) + \theta(k,l+1) = \theta(k,l)$. Under this relation, the family of nonnegative real numbers $\theta(k,l)$ can be equivalently (see [166, Lemma A.4]) characterized by a measure ν on $[0,1]$ such that

$$\int_0^1 x^{k-1}(1-x)^{l-1}\nu(dx) = \theta(k,l).$$

The following definition of n-point sticky Brownian motions from [169] is a reformulation of the Howitt-Warren martingale problem [104]. See Figure 1.5 and Figure 1.6 for simulations of n-point Brownian motions.

Definition 1.3.5 ([169, Theorem 5.3])**.** A stochastic process $\vec{X}(t) = (X_1(t), ..., X_n(t))$ started from $\vec{X}(0)$ will be called n-point sticky Brownian motions if it solves the following martingale problem called the Howitt-Warren martingale problem with drift β and characteristic measure ν.

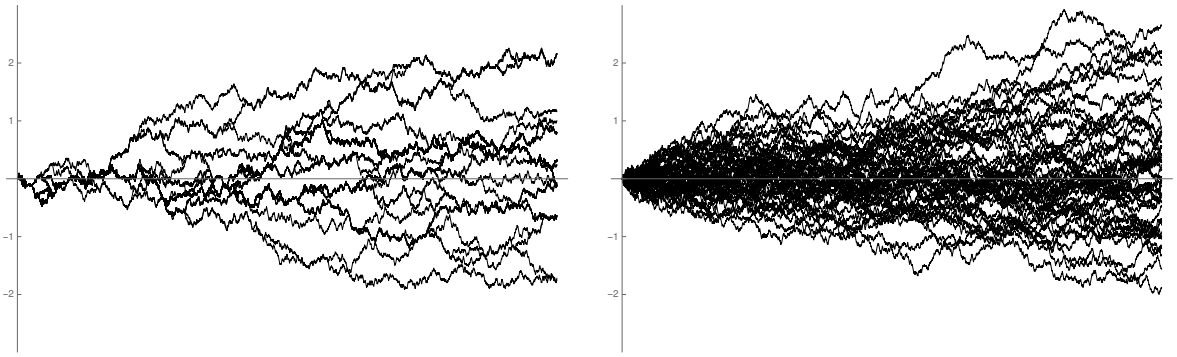

Figure 1.6: Left panel: 50 point-sticky Brownian motions using the same discretization as in Fig. 1.5. Because of the stickiness, the number of trajectories seems much smaller than 50. Right panel: 50 independent Brownian motions.

- (i) $\vec{X}$ is a continuous, square integrable martingale.
- (ii) The processes X_i and X_j have covariation process

$$\langle X_i, X_j \rangle(t) = \int_0^t \mathbb{1}_{X_i(s)=X_j(s)} ds, \qquad \text{for } t \geq 0, i, j = 1, ..., n.$$

- (iii) Consider any $\Delta \subset \{1, ..., n\}$. For $\vec{x} \in \mathbb{R}^n$, let

$$f_\Delta(\vec{x}) := \max_{i \in \Delta} \{x_i\} \qquad \text{and} \qquad g_\Delta(\vec{x}) := |\{i \in \Delta : x_i = f_\Delta(\vec{x})\}|,$$

  where $|S|$ denotes the number of elements in a set S. Then

$$f_\Delta(\vec{X}(t)) - \int_0^t \beta_+(g_\Delta(\vec{X}(t)) ds$$

  is a martingale with respect to the filtration generated by $\vec{X}$, where

$$\beta_+(1) := \beta \qquad \textit{and} \qquad \beta_+(m) := \beta + 2 \int \nu(dy) \sum_{k=0}^{m-2} (1-y)^k = \beta + 2 \sum_{k=1}^{m-1} \theta(1, k).$$

Remark 1.3.6. Definition 1.3.5 generalizes the definition of 2-point sticky Brownian motions because each particle marginally evolves as a Brownian motion, and the marginal distribution of any

pair of particles is that of a 2 point Brownian motion stickiness parameter $\lambda = \beta_+(2)$. Further, the consistency of the n-point motion is clear from property (iii).

We will be interested in a particular exactly solvable case of the Howitt-Warren Martingale problem.

Definition 1.3.7. An n-point stochastic process $(B_1(t), ..., B_n(t))$ will be called the *n-point uniform sticky Brownian motions* with stickiness λ if it solves the Howitt-Warren Martingale problem with drift $\beta = 0$ and characteristic measure

$$\nu(dx) = \mathbb{1}_{x\in[0,1]}\frac{\lambda}{2}dx.$$

This choice corresponds to choosing the fragmentation rates $\theta(k,l) = B(k,l)$, where $B(k,l) = \frac{\Gamma(k)\Gamma(l)}{\Gamma(k+l)}$ denotes the beta function.

In order to realize the n-point sticky Brownian motions as a family of independent random motions in a random environment, we need to introduce the notion of stochastic flows of kernels. Let $\mathcal{B}$ be the Borel σ-algebra of $\mathbb{R}$. For any $s \le t$, a *random probability kernel*, denoted $\mathsf{K}_{s,t}(x, A)$, for $x \in \mathbb{R}$ and $A \in \mathcal{B}$, is a measurable function defined on some underlying probability space Ω, such that, for each $(x, \omega) \in \mathbb{R} \times \Omega$, it defines a probability measure on $\mathbb{R}$. In order to interpret this as the random probability that a particle arrives in A at time t after starting in position x at time s, the kernel needs to satisfy the following additional hypotheses.

Definition 1.3.8 ([169, Definition 5.1])**.** A family of random probability kernels $(\mathsf{K}_{s,t})_{s\le t}$ on $\mathbb{R}$ is called a *stochastic flow of kernels* if the following conditions are satisfied.

(i) For any real $s \le t \le u$ and $x \in \mathbb{R}$, almost surely $\mathsf{K}_{s,s}(x, A) = \delta_x(A)$, and

$$\int_{\mathbb{R}} \mathsf{K}_{s,t}(x, dy)\mathsf{K}_{t,u}(y, A)dy = \mathsf{K}_{s,u}(x, A)$$

for all $A \in \mathcal{B}$.

(ii) For any $t_1 \leq t_2 \leq \ldots \leq t_k$, the random kernels $(\mathsf{K}_{t_i,t_{i+1}})_{i=1}^{k-1}$ are independent.

(iii) For any $s \leq u$ and t real, $\mathsf{K}_{s,u}$ and $\mathsf{K}_{s+t,u+t}$ have the same finite dimensional distributions.

Remark 1.3.9. Additional continuity hypotheses were given in the original definition of a stochastic flow of kernels in [133], but we will only be interested in Feller processes for which these hypotheses are automatically satisfied.

The n-point motion of a stochastic flow of kernels is a family of n stochastic processes $X_1, \ldots, X_n$ on $\mathbb{R}$ with transition probabilities given by

$$P(\vec{x}, d\vec{y}) = \mathbb{E}\left[\prod_{i=1}^{n} \mathsf{K}_{0,t}(x_i, dy_i)\right]. \tag{1.16}$$

Every consistent family of n-point motions that is Feller, is the n-point motion of some stochastic flow of kernels [133]. Any solution to the Howitt-Warren martingale problem is a consistent family as was noted after Definition 1.3.5, and is Feller by [104]. So any solution to the Howitt-Warren martingale problem is the n-point motion of some stochastic flow of kernels.

Definition 1.3.10. A stochastic flow of kernels whose n-point motions solve the Howitt-Warren martingale problem is called a *Howitt-Warren flow*. The stochastic flow corresponding to the special case of the Howitt-Warren martingale problem considered in Definition 1.3.7 (that we called the uniform Howitt-Warren martingale problem), is sometimes called the *Le Jan-Raimond flow*, after the paper [135], following the terminology used in [169, 166].

In condition (i) of Definition 1.3.8, if we assume that we can move the quantifier "almost surely" so it occurs before choosing s, t, u and x, then we can sample all $\mathsf{K}_{s,t}$ and almost surely these kernels define the transition kernels for some continuous space-time Markov process. Conditionally on the kernels we can describe the n-point motion as independent stochastic processes which evolve according to the transition kernels $\mathsf{K}_{s,t}$. Put simply the n-point motion can be seen as continuous space time random motions in a random environment which is given by the set of all transition kernels $\mathsf{K}_{s,t}$. In [166] (see also [169, Section 5]) it is shown that the change in quantifiers

in (i) necessary for this description can be done for Howitt-Warren flows. The random environment is explicitly constructed [166, Section 3] (see also [169, Section 5]) and consists of a Brownian web [1] plus a marked Poisson process at special points of the Brownian web [148]. The random motions in this environment essentially follow the Brownian web trajectories, except at these special points where they may turn left or right with a random probability. For Howitt-Warren flows such that $\int q(1-q)^{-1}\nu(dq) < \infty$ (which is not true for the Le Jan-Raimond flow), the random environment can also be constructed (see [166, Section 4]) using the Brownian net [179, 168].

Note that when starting from a set of particles on the real line and assuming that these particles will branch and coalesce following paths given by either the Brownian net or the Brownian web, the positions of the particles at a later time are given by a Pfaffian point process [85]. This type of evolution of Brownian particles is also related to random matrix theory, in particular the Ginibre evolution [191, 190, 192] (the evolution of real eigenvalues in a Ginibre matrix with Brownian coefficients), but these results do not seem directly related our results.

Results

Our first result is a Fredholm determinant formula for the Laplace transform of the uniform Howitt-Warren stochastic flow of kernels $\mathsf{K}_{0,t}(0,[x,\infty))$, or Le Jan-Raimond flow.

First recall the definition of the gamma function

$$\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx,$$

and the polygamma functions

$$\psi(\theta) = \partial_z \log\Gamma(z)|_{z=\theta}, \qquad \psi_i(\theta) = (\partial_z)^i\psi(z)|_{z=\theta}.$$

Theorem 1.3.11. *Let* $\mathsf{K}_{0,t}(0,[x,\infty))$ *denote the kernel of the uniform Howitt-Warren flow with stick-*

[1] The Brownian web was introduced in [11], see also [185].

iness parameter $\lambda > 0$. For $u \in \mathbb{C} \setminus \mathbb{R}_{>0}$, and $x > 0$, we have

$$\mathbb{E}[e^{u\mathsf{K}_{0,t}(0,[x,\infty))}] = \det(I - K_u)_{L^2(C)}, \tag{1.17}$$

(the R.H.S is a Fredholm determinant, see Definition 1.1.1), where

$$K_u(v, v') = \frac{1}{2\pi\mathbf{i}} \int_{1/2-\mathbf{i}\infty}^{1/2+\mathbf{i}\infty} \frac{\pi}{\sin(\pi s)} (-u)^s \frac{g(v)}{g(v+s)} \frac{ds}{s+v-v'},$$

and

$$g(v) = \Gamma(v) \exp\left(\lambda x \psi_0(v) + \frac{\lambda^2 t}{2} \psi_1(v)\right).$$

where C is a positively oriented circle with radius $1/4$ centered at $1/4$. (It is important that this contour passes through zero at the correct angle. The actual radius of the circle C does not matter.)

Remark 1.3.12. We use two very different notions of kernels, which are both denoted by the letter K. We will reserve the font K for stochastic flows of kernels, and the usual font K for the kernels of $\mathbb{L}^2$ operators arising in Fredholm determinants.

We reach Theorem 1.3.11 by taking a limit of a similar Fredholm determinant formula [23, Theorem 1.13] for the beta RWRE defined in Section 3.1.3. Theorem 1.3.11 is proved in Section 3.4.

We perform a rigorous saddle-point analysis of the Laplace transform formula (1.17) to obtain a quenched large deviation principle for the uniform Howitt-Warren stochastic flow.

Theorem 1.3.13. *Let $\lambda > 0$ and $x \geq 1.35$. Let $\mathsf{K}_{s,t}$ be the kernel of a uniform Howitt-Warren flow. Then we have the following convergence in probability*

$$\frac{1}{t} \log \mathsf{K}_{0,t}(0, [xt, \infty)) \xrightarrow[t\to\infty]{} -\lambda^2 J(x/\lambda), \tag{1.18}$$

where

$$J(x) = \max_{\theta \in \mathbb{R}_{>0}} \left\{ \frac{1}{2}\psi_2(\theta) + x\psi_1(\theta) \right\}. \tag{1.19}$$

The condition $x \geq 1.35$ is technical and is addressed in Remark 1.3.16. We expect that the limit holds almost surely. This should follow from subadditivity arguments, though we do not pursue this in the current work (see [163] for an almost sure quenched large deviation principle for discrete random walks). We emphasize that in Theorem 1.3.13, the rate function $J(x)$ is expressed explicitly using well-known special functions, which is in contrast with what one would obtain using subadditivity arguments. Another large deviation principle was shown in [69] for the empirical distribution of a certain class of n-point sticky Brownian motions, but this does not seem to be related to the present Theorem 1.3.13.

Remark 1.3.14. The annealed[2] analogue of this large deviation principle just describes the tail behavior of a standard Brownian motion. Indeed,

$$\frac{1}{t}\log \mathbb{E}[\mathsf{K}_{0,t}(0,[xt,\infty))] = -x^2/2.$$

It can be easily checked that $\lambda^2 J(x/\lambda) > x^2/2$ which, in the context of directed polymers, means that the model exhibits strong disorder. Note that the sign of the inequality is consistent with Jensen's inequality (assuming (1.18) holds in L^1). The inequality becomes an equality in the $\lambda \to \infty$ limit, which corresponds to Brownian motions with no stickiness.

When uniform sticky Brownian motions are viewed as random walks in a random environment, Theorem 1.3.13 gives a large deviation principle whose rate function is deterministic despite the randomness of the environment. The random variable $\log \mathsf{K}_{0,t}$ does depend on the environment, but its fluctuations are small enough that they are not detected by the large deviation principle. We prove that the model is in the KPZ universality class in the sense that the random lower order corrections to the large deviation principle, or equivalently the fluctuations of $\log \mathsf{K}_{0,t}$, are Tracy-Widom GUE distributed and are of order $t^{1/3}$.

Theorem 1.3.15. *Let $\mathsf{K}_{s,t}$ be the kernel of a uniform Howitt-Warren flow with stickiness parameter*

[2] In the context of random walks in random environment and directed polymers, the (limiting) quenched free energy or rate function is the limit obtained for almost every environment and the annealed analogues correspond to the same quantities for the averaged environment.

$\lambda > 0$. *Let* $0 < \theta < 1$. *We have*

$$\lim_{t\to\infty} \mathbb{P}\left(\frac{\log(\mathsf{K}_{0,t}(0,[x(\theta)t,\infty)) + \lambda^2 J(x(\theta)/\lambda)t}{t^{1/3}\sigma(\theta)} < y\right) = F_{\mathrm{GUE}}(y),$$

where $F_{\mathrm{GUE}}(y)$ *is the cumulative density function of the Tracy-Widom distribution (defined in 1.1.1), and*

$$x(\theta) = -\frac{\lambda}{2}\frac{\psi_3(\theta)}{\psi_2(\theta)}, \qquad \sigma(\theta) = \frac{\lambda^{2/3}}{2^{1/3}}\left(\frac{-1}{2}\psi_4(\theta) - \frac{x(\theta)}{\lambda}\psi_3(\theta)\right)^{\frac{1}{3}}. \tag{1.20}$$

Theorem 1.3.15 comes from applying a rigorous steep descent analysis to the Fredholm determinant in Theorem 1.3.11. The parametrization of functions J and σ arising in the limit theorem via the variable θ may appear unnatural at this point. It will appear more natural in the proof as θ is the location of the critical point used in the steep descent analysis. We expect that there should exist another interpretation of the parameter θ. It should naturally parametrize stationary measures associated with the uniform Howitt-Warren flow, and KPZ scaling theory [173, 126] would predict the expressions for $J(x)$ and $\sigma(\theta)$ given above. This approach would require to degenerate to the continuous limit the results from [18] and we leave this for future investigation (the analogue of parameter θ in the discrete setting is denoted $\lambda(\xi)$ in [18, Theorem 2.7]).

Remark 1.3.16. Note that $x(\theta)$ is a decreasing function of θ and the technical hypothesis $\theta < 1$ corresponds to approximately $1.35 \leq x(\theta)$. Similarly $J(x)$ is an increasing function of x and $\theta < 1$ corresponds approximately to $1.02 < J(x(\theta))$. We expect Theorem 1.3.15 to hold for all $\theta > 0$, and Theorem 1.3.13 to hold for all $x > 0$, however if $\theta \geq 1$ we pick up additional residues while deforming the contours of our Fredholm determinant during the asymptotic analysis which make the necessary justifications significantly more challenging.

More generally, we believe that the result of Theorem 1.3.11 should be universal and hold for more general Howitt-Warren flows under mild assumptions on the characteristic measure ν. This would be analogous to a conjecture that for discrete polymer models the fluctuations of the free energy are Tracy-Widom distributed as long as the weights of the polymer have finite fifth moments

[7, Conjecture 2.6]. Moreover, based on [162, Theorem 4.3], we expect that the random variable

$$\log \mathsf{K}_{0,t}(0, [xt, xt + a)),$$

for any $a > 0$, satisfies the same limit theorems as $\log \mathsf{K}_{0,t}(0, [xt, +\infty))$ in Theorem 1.3.13 and Theorem 1.3.15, with the same constants (the prediction that the constant $\sigma(\theta)$ should remain the same is suggested by the results of [183]).

Following [23] we can state a corollary of Theorem 1.3.15. In general, tail probability estimates provide information about the extremes of independent samples. In the present case, we obtain that the largest among n uniform sticky Brownian motions fluctuates asymptotically for large n according to the Tracy-Widom distribution. We will see that the result is very different from the case of n independent Brownian motions, as can be expected from the simulations in Figure 1.6.

Corollary 1.3.17. *Let $c \in [1.02, \infty)$, let x_0 be such that $\lambda^2 J(x_0/\lambda) = c$, let θ_0 be such that $x(\theta_0) = x_0$, and let $\{B_i(t)\}$ be uniform n-point sticky Brownian motions with stickiness parameter $\lambda > 0$ and scale n as $n = e^{ct}$, then*

$$\lim_{t\to\infty} \mathbb{P}\left(\frac{\max_{i=1,\dots,n}\{B_i(t)\} - t x_0}{t^{1/3}\sigma(\theta_0)/(\lambda^2 J'(x_0/\lambda))} \leq y \right) = F_{\mathrm{GUE}}(y). \tag{1.21}$$

The proof of Corollary 1.3.17 is very similar to the proof of [23, Corollary 5.8] and uses the fact that after conditioning on the environment we are dealing with independent motions along with our strong control of the random variable $\mathsf{K}_{0,t}(0, [xt, \infty))$ from Theorem 1.3.15. The details of the proof can be found at the end of Section 3.2.

1.4 Boundary-weighted stochastic six vertex model

This section is an overview of the work in Chapter 4. We will study a version of the six vertex model with a boundary condition leading to an interesting phase diagram.

Consider a family of six-vertex models on the half-infinite strip $D_n = \mathbb{Z}_{\geq 0} \times \{1, \ldots, n\}$ where $n \in \mathbb{N}$. Specifically, the state space of the models is the set $\mathcal{P}_n$ consisting of all collections of n up-right paths, with nearest neighbor steps in D_n with the paths starting from the points $\{(0, i) : 1 \leq i \leq n\}$ and exiting the top boundary. We add the additional condition, that no two paths can share a horizontal or vertical edge, see Figure 1.7.

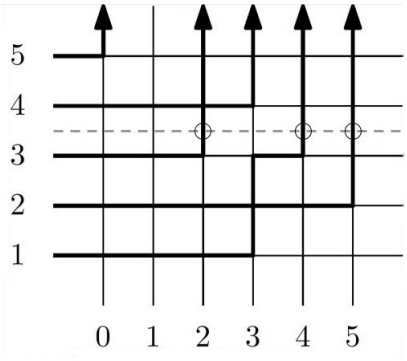

Figure 1.7: An example of a path collection π in $\mathcal{P}_5$. Here $\lambda_1^3(\pi) = 5$, $\lambda_2^3(\pi) = 4$, $\lambda_3^3(\pi) = 2$

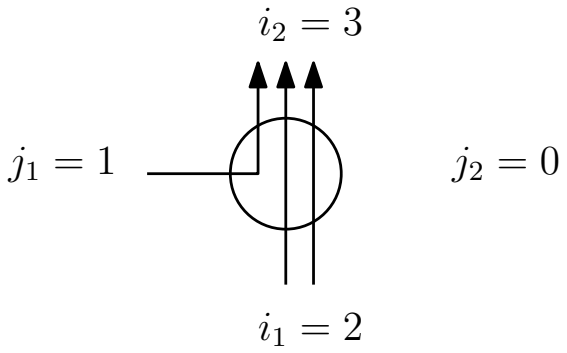

Figure 1.8: An example of a vertex of type $(i_1, j_1; i_2, j_2) = (2, 1; 3, 0)$

In the next few paragraphs we explain the types of probability measures we put on $\mathcal{P}_n$ (they are given in equation (1.23) below), but to accomplish this we need a bit of notation. A *signature* of length n is a nonincreasing sequence $\lambda = (\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n)$ with $\lambda_i \in \mathbb{Z}$. We use Sign_n to denote the set of all signatures of length n, and use Sign_n^+ for the set of such signatures with $\lambda_n \geq 0$. To each collection of n up-right paths $\pi \in \mathcal{P}_n$ one can identify a sequence of signatures $\lambda^i(\pi) \in \mathsf{Sign}_i^+$ for $i = 1, \ldots, n$, where $(\lambda_1^i(\pi), \lambda_2^i(\pi), \ldots, \lambda_i^i(\pi))$ are the ordered x-coordinates at which the paths in π intersect the horizontal line $y = i + 1/2$, see Figure 1.7.

Given an up-right path collection $\pi \in \mathcal{P}_n$, each vertex is given a *vertex type* based on four numbers $(i_1, j_1; i_2, j_2)$, where i_1 and j_1 denote the number paths entering the vertex vertically and horizontally respectively, while i_2 and j_2 denote the number of paths leaving the vertex vertically and horizontally respectively, see Figure 1.8. For complex parameters s and u we define the following vertex weights

$$
\begin{aligned}
w_1 &= w(0,0;0,0) = 1, & w_2 &= w(1,1;1,1) = \frac{u - s^{-1}}{1 - su} \\
w_2 &= w(1,0;1,0) = \frac{1 - s^{-1}u}{1 - su}, & w_4 &= w(0,1;0,1) = \frac{u - s}{1 - su}, \\
w_5 &= w(1,0;0,1) = \frac{(1 - s^2)u}{1 - su}, & w_6 &= w(0,1;1,0) = \frac{1 - s^{-2}}{1 - su}.
\end{aligned} \tag{1.22}
$$

This nonintuitive parametrization of weights by s and u comes from [46], where it is important in defining a higher spin generalization of the six-vertex model. Later in (4.8) we present the higher spin vertex weights, and one obtains the weights in (1.22) by setting $q = s^{-2}$ in (4.8).

For $\pi \in \mathcal{P}_n$ we let $\pi(i, j)$ denote the vertex type of the vertex at position (i, j) in the path collection π. Given complex numbers s and u, and a function $f : \mathsf{Sign}_n^+ \to \mathbb{C}$ we define the weight of a path collection $\pi \in \mathcal{P}_n$ by

$$
\mathcal{W}^f(\pi) = f(\lambda^n(\pi)) \prod_{i=1}^{\infty} \prod_{j=1}^{n} w(\pi(i, j)).
$$

All but finitely many $\pi(i, j)$ are equal to $(0, 0; 0, 0)$ and have weight 1 by (1.22), so the product is well defined. If one chooses u and s in $\mathbb{C}$ and the function f so that the weights $\mathcal{W}^f(\pi)$ are nonnegative, not all zero and summable then one can use the weights $W^f(\pi)$ to define a probability measure on $\mathcal{P}_n$ through

$$
\mathbb{P}^f(\pi) = (Z^f)^{-1} \cdot \mathcal{W}^f(\pi), \text{ where } Z^f := \sum_{\pi \in \mathcal{P}_n} \mathcal{W}^f(\pi). \tag{1.23}
$$

Equation (1.23) gives the general form of the measures we will study. In plain words $\mathbb{P}^f$ is the usual six-vertex measure except that the path collections π are reweighted based on their top boundary, namely $\lambda^n(\pi)$, through the boundary weight function f. See Figure 1.9

Remark 1.4.1. When we go to our main results we will take $u > s > 1$ above. In the usual weight parametrization of the six-vertex model we have that

$$
a_1 = 1, \quad a_2 = \frac{u - s^{-1}}{su - 1}, \quad b_1 = \frac{1 - s^{-1}u}{1 - su}, \quad b_2 = \frac{u - s}{su - 1}, \quad c_1 = \frac{(1 - s^2)u}{1 - su}, \text{ and } c_2 = \frac{1 - s^{-2}}{su - 1}.
$$

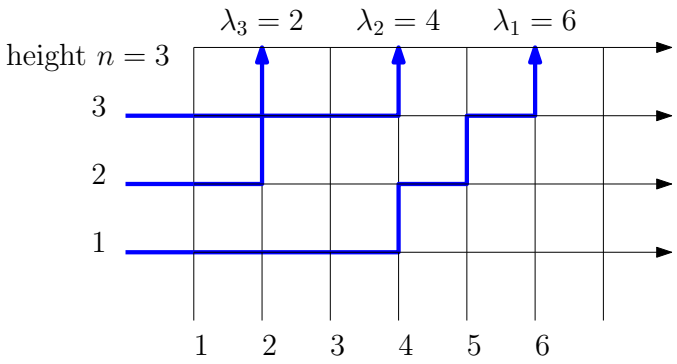

Figure 1.9: An example of a path configuration in the boundary-weighted six vertex model which demonstrates the boundary term $\lambda^3 = \lambda$.

We mention that the latter weights are the absolute values of those in (4.1), where ultimately the sign difference will be absorbed in the boundary weight function f of the model so that $\mathcal{W}^f(\pi) \geq 0$ for all $\pi \in \mathcal{P}_n$. Associated with the six weights is an *anisotropy parameter* Δ, given by

$$\Delta(a_1, a_2, b_1, b_2, c_1, c_2) = \frac{a_1 a_2 + b_1 b_2 - c_1 c_2}{2\sqrt{a_1 a_2 b_1 b_2}}, \tag{1.24}$$

which is believed to be directly related with the qualitative and quantitative properties of the model, see [153]. The choice of weights as in (4.1) with $u > s > 1$ corresponds to $\Delta > 1$, which is known as the *ferroelectric phase* of the six-vertex model.

There are many different choices of parameters and functions f that lead to meaningful measures in (1.23). For example, if $f(\lambda) = 0$ unless $\lambda_{n-i+1} = i - 1$ for $i = 1, \ldots, n$ the measure in $\mathbb{P}^f$ becomes the six-vertex model with *domain wall boundary condition* (DWBC), [123]. Another special case of the measures in (4.2) includes the case when $u > s > 1$ and

$$f(\lambda) = \mathsf{G}^c_\lambda(\rho) := (-1)^n \cdot \mathbb{1}_{m_0=0} \prod_{i=1}^{\infty} \mathbb{1}_{m_i \leq 1} \prod_{j=1}^{n} (-s)^{\lambda_j}, \tag{1.25}$$

where $\lambda = 0^{m_0} 1^{m_1} 2^{m_2} \ldots$. In the latter notation m_i is the number of times i appears in the list $(\lambda_1, \ldots, \lambda_n)$ and $\mathbb{1}_E$ is the indicator function of the set E. With this choice of parameters and function f, the measure $\mathbb{P}^f$ becomes what is known as stochastic six-vertex model, see e.g. [98],

[39], with parameters

$$b_1 = \frac{1 - s^{-1}u}{1 - su}, \qquad b_2 = \frac{s^2 - su}{1 - su}.$$

For a quick proof of the latter statement we refer the reader to [46, Section 6.5].

A prediction in [50], which has been very recently partially verified in [3], states that the pure states μ of the ferroelectric six-vertex model are parametrized by a *slope* $(s,t) \in [0,1]^2$, where s and t denote the probabilities that a given vertical and horizontal edge is occupied under μ. For a certain open lens-shaped set $\mathfrak{H} \subset [0,1]^2$ one has the following characterization of pure states for the ferroelectric six-vertex model (here $\overline{\mathfrak{H}} = \mathfrak{H} \cup \partial\mathfrak{H}$):

1. *Nonexistence*: If $(s,t) \in \mathfrak{H}$, then there are no pure states $\mu_{s,t}$ of slope (s,t).

2. *KPZ States*: If $(s,t) \in \partial\mathfrak{H}$, then $\mu_{s,t}$ should exhibit Kardar-Parisi-Zhang (KPZ) behavior.

3. *Liquid States*: If $(s,t) \in (0,1)^2 \backslash \overline{\mathfrak{H}}$, then $\mu_{s,t}$ should exhibit Gaussian free field (GFF) behavior.

4. *Frozen States*: If (s,t) is on the boundary of $[0,1]^2$, then $\mu_{s,t}$ should be frozen.

From the above conjectural classification, [3] established the nonexistence statement (1) and proved the existence and uniqueness of KPZ states (2) for all $(s,t) \in \partial\mathfrak{H}$. It is worth mentioning that the above classification sharply contrasts the one for dimer models. Specifically, the pure states in dimer models were classified in [171] and [120] and they come in three types. The first is *frozen*, where the associated height function is basically deterministic; the second is *gaseous*, where the variance of the height function is bounded but non-zero; the third is *liquid*, where the hight function fluctuations diverge logarithmically in the lattice size. In particular, for dimer models there are no Nonexistence or KPZ pure states.

Going back to our previous discussion, the stochastic six-vertex model considered in [39], which corresponds to f as in (1.25), was shown to asymptotically have a phase diagram that consists of two frozen regions, i.e. regions where the local behavior of the model is described by Frozen States, and a non-frozen region, where one observes solely KPZ States, see Figure

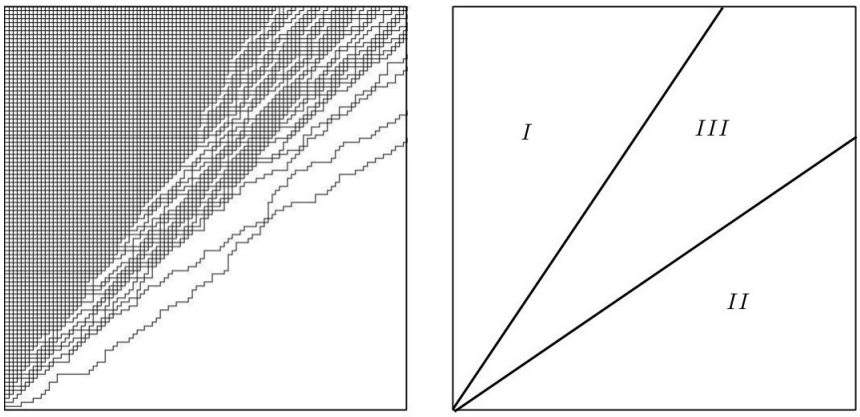

Figure 1.10: The left picture represents a sample of $\mathbb{P}^f$ with f as in (1.25) for the parameters $n = 100$, $u = 2$, $s^{-2} = 0.5$. The picture on the right side depicts the phase diagrams for these measures when n is large. The regions I and II correspond to Frozen States and region III corresponds to KPZ States

1.10. More specifically, in [39] it was shown that the one-point marginals of the height function $h(x, y)$, which at a location (x, y) counts the number of horizontal edges crossed by the vertical segment connecting $(x, 0)$ and (x, y) in the non-frozen region III of Figure 1.10 are asymptotically governed by the GUE Tracy-Widom distribution [189]. This type of behavior is characteristic of models in the KPZ universality class described in Section 1.1. For the DWBC six-vertex model a very different phase diagram is expected, although we emphasize that it has not been established rigorously. Specifically, for the DWBC it is expected that as n becomes large the model again develops macroscopic frozen regions that are separated by a non-frozen region where one observes solely Liquid States. The only instance where this has been rigorously established is when $\Delta = 0$, which is the free fermion point of the model, see [118], [119]. We emphasize that this is no longer in the ferroelectric phase so one should be cautious when comparing to Figure 1.10. When $\Delta = 0$ the six-vertex model falls into the framework of the dimer models, which is what enables its precise mathematical analysis. We mention; however, that there are non-rigorous physics works and numerical simulations that indicate that for $\Delta < 1$ the six vertex model with DWBC has solely Liquid States in the non-frozen region, and by analogy with the dimer models the fluctuations of those are no longer KPZ, but rather governed by a suitable pullback of the Gaussian free field, [96]. In the ferroelectric $\Delta > 1$ case similar heuristics suggest that one observes only frozen states [59,

180].

The above paragraphs explain that by picking different boundary weight functions f we can obtain qualitatively different phase diagrams for our six-vertex model. We will consider a very special class of boundary functions f. This class will be described fully in Chapter 4 as we prove these results. In the remainder of this section we explain the very high level motivations that have guided our choice of f.

First of all, our discussion above indicates that for the stochastic six-vertex model of [39] the non-frozen region consists entirely of KPZ States, while for the DWBC (at least conjecturally) it consists solely of Liquid States (or states with Gaussian statistics). A natural question is whether we can find a boundary weight function f for which both types of pure states co-exist in the non-frozen region of the model. A second point is that, for general functions f, the asymptotic analysis for $\mathbb{P}^f$ is prohibitively complicated – indeed even for the DWBC the phase diagram is largely conjectural, and so one is inclined to consider special boundary weight functions f for which the analysis of the model is tractable. Our choice of f is motivated by our desire that the resulting model satisfies these two properties.

We study a special case of (1.23) when the boundary weight function f is given by

$$f(\lambda) = \sum_{\mu \in \mathsf{Sign}_n^+} \mathsf{G}_\mu^c(\rho) \mathsf{G}_{\lambda/\mu}^c(v, \ldots, v). \tag{1.26}$$

In (1.26) the function $\mathsf{G}_\mu^c(\rho)$ is as in (1.25) and the functions $\mathsf{G}_{\lambda/\mu}^c$ are a remarkable class of symmetric rational functions, which were introduced in [34]. The definition of $\mathsf{G}_{\lambda/\mu}^c$ is given in Definition 4.2.1, and these functions depend on M complex variables $v_1, \ldots, v_M$ that have all been set to the same complex number v in (1.26). We mention that $\mathsf{G}_{\lambda/\mu}^c$ are one-parameter generalizations of the classical (skew) Hall-Littlewood symmetric functions [142] and carry the name of (skew) *spin Hall-Littlewood symmetric functions*, see [49].

One can check that if $v^{-1} > u > s > 1$ then the measure $\mathbb{P}^f$ from (1.23) with f as in (1.26) is a well-defined probability measure, see Section 4.2.2. We will denote this measure by $\mathbb{P}_{u,v}^{N,M}$.

Even though the choice of f in (1.26) seems complicated we emphasize that the resulting measure $\mathbb{P}^f$ enjoys many remarkable properties and its asymptotic structure appears to be rich and interesting. We elaborate on these points in the next few paragraphs, summarizing some results from [71] where this model was studied in detail.

First of all, the choice of f as in (1.26) makes the model integrable and the distribution $\mathbb{P}^f$ analogous to the *ascending Macdonald processes* of [36]. What plays the role of the (skew) Macdonald symmetric functions $P_{\lambda/\mu}$ and their duals $Q_{\lambda/\mu}$ is a class of symmetric rational functions $\mathsf{F}_{\lambda/\mu}$ and their duals $\mathsf{G}^c_{\lambda/\mu}$ that were mentioned above. The functions $\mathsf{F}_{\lambda/\mu}, \mathsf{G}^c_{\lambda/\mu}$ enjoy many of the same properties as the Macdonald symmetric functions, including branching rules, orthogonality relations, (skew) Cauchy identities and so on. One consequence of the integrability of the model that can be appreciated by readers unfamiliar with symmetric function theory is that the partition function Z^f for our choice of f in (1.26) takes the following extremely simple product form

$$Z^f = (s^{-2}; s^{-2})_n \left(\frac{1 - s^{-1}u}{1 - su}\right)^n \left(\frac{1 - s^{-2}uv}{1 - uv}\right)^{nM}, \text{ where } (a; q)_m = (1 - a)(1 - aq) \cdots (1 - aq^{m-1}).$$

The latter formula for the partition function is recalled in Section 4.2.2.

Another consequence of the integrability of the model is the fact that it is self-consistent in the following sense. Suppose that we sample a path collection π on $\mathcal{P}_n$ according to $\mathbb{P}^{n,M}_{u,v}$ and then project the path collection to the first k rows where $1 \leq k \leq n$. The resulting path collection is now a random element in $\mathcal{P}_k$ and its distribution is precisely $\mathbb{P}^{k,M}_{u,v}$ – we recall this in Lemma 4.2.12. This self-consistency of the measures $\mathbb{P}^{n,M}_{u,v}$ for $n \in \mathbb{N}$ allows us for example to define a measure on up-right paths on the whole of $\mathbb{Z}^2_{\geq 0}$ whose projection to the bottom n rows has law $\mathbb{P}^{n,M}_{u,v}$.

Yet another consequence of the integrability of the model is given by the fact that for fixed n and varying $m \in \mathbb{Z}_{\geq 0}$ the measures $\mathbb{P}^{n,M}_{u,v}$ can be stochastically linked as we next explain. One can interpret the distribution $\mathbb{P}^{n,m}_{u,v}$ as the time m distribution of a Markov chain $\{X_m\}_{m=0}^{\infty}$ taking values in $\mathcal{P}_n$ for each m. This Markov chain is started from the stochastic six-vertex model at time zero,

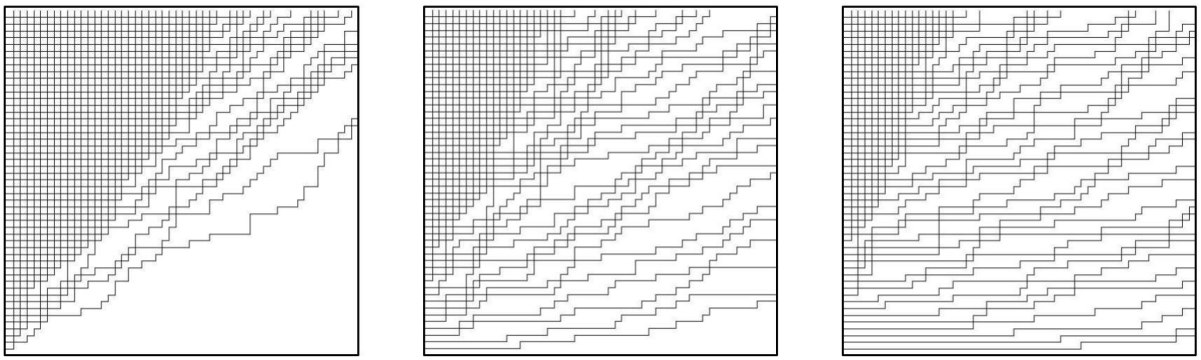

Figure 1.11: The pictures represent samples of the Markov chain $\{X_m\}_{m=0}^{\infty}$ when $n = 50$ at times $m = 0$, $m = 50$ and $m = 100$. The parameters of the process are $s^{-2} = 0.5$, $u = 2$ and all $v = 0.25$

and its dynamics are governed by sequential update rules. For more details and a precise formulation we refer the reader to [46, Section 6] as well as [71, Section 8] where an exact sampling algorithm of this process was developed. For a pictorial description of how the configurations X_m evolve as time increases see Figure 1.11. This interpretation is similar to known interpretations of the Schur process and Macdonald process as fixed time distributions of certain Markov processes, see [35, 36].

The above few paragraphs explained some of the structure and relationships between the measures $\mathbb{P}_{u,v}^{N,M}$ for varying $N, M \in \mathbb{N}$. These measures arise as degenerations of the higher-spin vertex models that were studied in [46], which is the origin of their integrability. For our purposes, the main consequence of the integrability of the model that is utilized is that one has formulas for the one-dimensional projections of $\mathbb{P}_{u,v}^{N,M}$ that are suitable for asymptotic analysis . This is what makes the analysis of the model tractable.

Our primary probabilistic interest in the measures $\mathbb{P}_{u,v}^{N,M}$ comes from the fact that as $N, M \to \infty$ the phase diagram of the model (at least conjecturally) exhibits all three types of pure states – Frozen, Liquid and KPZ. The presence of all three types of pure states is the second high-level motivation behind our choice of f as in (1.26) and we illustrate the phase diagram in Figure 1.12. The phase diagram in Figure 1.12, which corresponds to $\mathbb{P}_{u,v}^{N,M}$ when N and M are large, should be compared to the one in Figure 1.10, which corresponds to the stochastic six-vertex model or

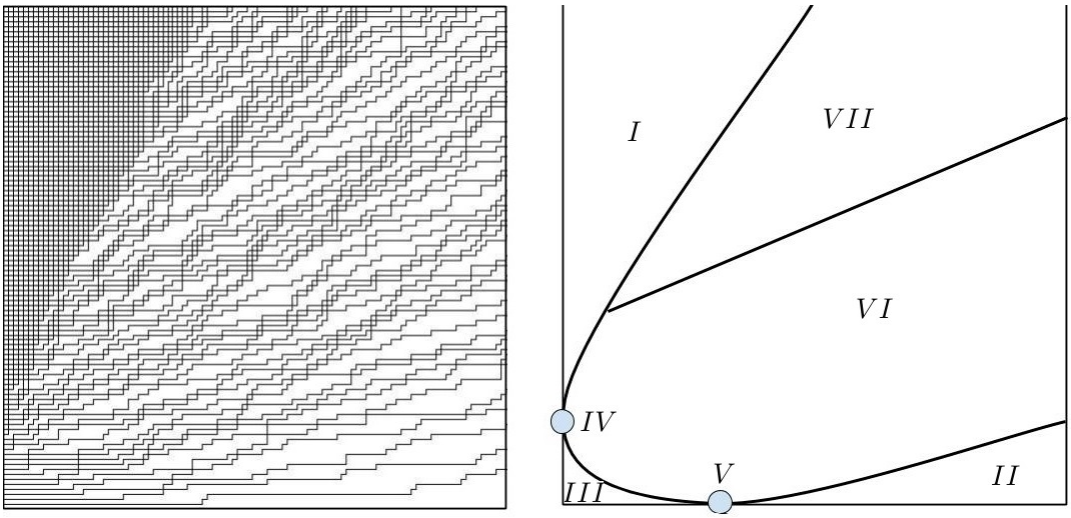

Figure 1.12: The picture on the left represent a sample of $\mathbb{P}_{u,v}^{N,M}$ with $N = M = 100$, $u = 2$, $s^{-2} = 0.5$, $v = 0.25$. The picture on the right represents the (conjectural) phase diagram of the model as $N, M \to \infty$

equivalently to the measure $\mathbb{P}_{u,v}^{N,0}$ (recall that the measures $\mathbb{P}_{u,v}^{N,m}$ were stochastically linked through a Markov chain with time zero distribution gave precisely by the stochastic six-vertex model). At least based on the simulations one observes that as the vertex model evolves in time $m = 0, \ldots, M$ the frozen regions I and II from the stochastic six-vertex model in Figure 1.10 begin to deform and a new frozen region, denoted by III in Figure 1.12 and consisting of vertices of type $(0, 1; 0, 1)$, is formed near the origin. With this new frozen region two new points IV and V are formed. These are sometimes referred to as *turning points* and they arise where two different frozen regions meet each other. Furthermore, our prediction is that, under the Markovian dynamics evolving the six-vertex model, the KPZ cone (i.e. region III in Figure 1.10) that is present at time $m = 0$ is translated away from the origin to region VII and a new GFF region (denoted by VI in Figure 1.12) takes its place. We mention here that the exact nature of the Markovian dynamics is not important for our analysis. The reason we mention it is to emphasize that the stochastic six-vertex model and the measures $\mathbb{P}_{u,v}^{N,M}$ we consider here are related to each other and the presence of the KPZ region VII in $\mathbb{P}_{u,v}^{N,M}$ can be traced back to the presence of region III in $\mathbb{P}_{u,v}^{N,0}$. If the same dynamics are run from a different initial configuration one may very well see a completely different phase diagram than the one in Figure 1.12.

As can be seen from Figure 1.12 the asymptotics of $\mathbb{P}_{u,v}^{N,M}$ as $N, M \to \infty$ appear to be quite complex. A long term program, initiated in [71], is to rigorously establish the phase diagram in Figure 1.12. So far only the asymptotics near the point IV have been understood. Specifically, in [71] it was shown that near IV a certain configuration of empty edges converges to the GUE-corners process, we define the latter here. Recall that the Gaussian Unitary Ensemble (GUE) is a measure on Hermitian matrices $\{X_{ij}\}_{i,j=1}^{k}$ with density proportional to $e^{-Tr(X^2)/2}$ with respect to Lebesgue measure. For $1 \leq r \leq k$, let $\lambda_1^r \leq \lambda_2^r \leq \cdots \leq \lambda_r^r$ denote the ordered eigenvalues of the submatrix $\{X_{ij}\}_{i,j=1}^{r}$ of X. The joint law of the eigenvalues $\{\lambda_i^j\}_{1\leq i\leq j\leq k}$ is called the *GUE-corners process* of rank k (or the GUE-minors process). The appearance of the GUE-corners process has been established in related contexts for random lozenge tilings in [113, 149, 154] and the uniform six-vertex model with domain-wall boundary conditions [95]. It is believed to be a universal scaling limit near points separating two different frozen regions such as the point IV.

This work, is a continuation of the program initiated in [71] of establishing the phase diagram in Figure 1.12. Specifically, in Figure 1.12 the point V is another turning point we will show that the statistics of the model $\mathbb{P}_{u,v}^{N,M}$ near this point are also described by the GUE-corners process. Before we state our main result we give our choice of parameters and some notation.

Definition 1.4.2. We assume that $v, u, s \in (0, \infty)$ satisfy $v^{-1} > u > s > 1$. With this choice of parameters we define the constants

$$\begin{aligned} a &= \frac{v\left(u - s^{-1}\right)\left(s^{-1}u - 1\right)}{(1-uv)(1-s^{-2}uv)}, \qquad b = \frac{(s^2-1)}{(u-s)(1-su)} \\ c &= \frac{1}{2}\left(a\left(\frac{1}{(u-s)^2} - \frac{s^2}{(1-su)^2}\right) - \frac{s^{-4}v^2}{(1-s^{-2}uv)^2} + \frac{v^2}{(1-uv)^2}\right), \qquad d = \frac{-\sqrt{2c}}{b}. \end{aligned} \tag{1.27}$$

If $v^{-1} > u > s > 1$ one observes that

$$a > 0, \qquad b < 0, \qquad c > 0, \qquad d > 0.$$

See Lemma 4.5.1 in the main text for a verification of this fact.

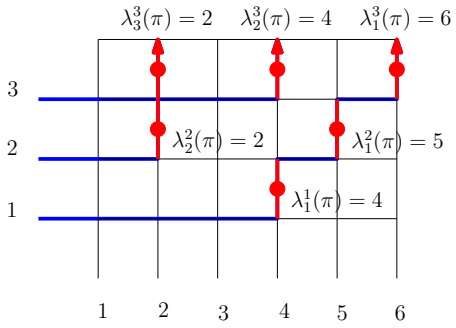

Figure 1.13: A demonstration of the particle positions $\lambda_j^i(\pi)$. Theorem 1.4.4 says that these particle positions after appropriate recentering and rescaling converge to the GUE-corners process.

Remark 1.4.3. The choice $v^{-1} > u > s > 1$ in Definition 1.4.2 corresponds to choosing $b_2 > b_1$. This condition is essential because if $b_1 > b_2$ then all the up-right paths will end up packed against the left boundary with high probability.

The main result of Chapter 4 is as follows.

Theorem 1.4.4. *Suppose that u, v, s, a, d are as in Definition 1.27 and $k \in \mathbb{N}$ is given. Suppose that $N(M)$ is a sequence of integers such that $N(M) \geq k$ for all M and let $\mathbb{P}_{u,v}^{N,M}$ be the measure on collections of paths $\pi \in \mathcal{P}_N$ as earlier in the section. Define the random vector $Y(N, M; k)$ through*

$$Y_i^j(N, M; k) = \frac{\lambda_{j-i+1}^j(\pi) - aM}{d\sqrt{M}} \text{ for } 1 \leq i \leq j \leq k, \tag{1.28}$$

See Figure 1.13. Then the sequence $Y(N, M; k)$ converges weakly to the GUE-corners process of rank k as $M \to \infty$.

Remark 1.4.5. In (1.28) we reverse the order of λ_i^j because the usual convention for signatures $\lambda = (\lambda_1, \ldots, \lambda_n)$ demands that λ_i be sorted in decreasing order, while for the eigenvalues of a random matrix the usual convention is that they are sorted in increasing order.

We mention here that while the asymptotic behaviors near *IV* and *V* are similar, the arguments used to establish the two results are quite different. The arguments in [71] rely on a remarkable class of difference operators, which can be used to extract averages of observables for $\mathbb{P}_{u,v}^{N,M}$ near

the left boundary of the model. These observables become useless for accessing the asymptotic behavior of the base of the model and consequently our approach in Part 4 is completely different, and arguably more direct as we explain here. In the remainder of this section we give an outline of our approach to proving Theorem 1.4.4. The discussion below will involve certain expressions that will be properly introduced in Chapter 4, and which should be treated as black boxes for the purposes of the outline.

Using the integrability of the model we obtain the formula

$$\mathbb{P}_{u,v}^{N,M}(\lambda_1^k(\pi) = \mu_1, \cdots, \lambda_1^k(\pi) = \mu_k) \propto \mathsf{F}_\mu([u]^k) \cdot f(\mu; [v]^M, \rho),$$

for any $\mu = (\mu_1, \ldots, \mu_k) \in \mathsf{Sign}_k^+$ for a certain function $\mathsf{F}_\mu(\cdot)$ defined in section 4.2. A generalization of this fact appears as Lemma 4.2.12 in Chapter 4. We then derive certain combinatorial estimates for $\mathsf{F}_\mu([u]^k)$ and a contour integral formula for $f(\mu; [v]^M, \rho)$ in Section 4.3, which are both suitable for studying the $M \to \infty$ limit of these expressions (for the function $\mathsf{F}_\mu([u]^k)$ the dependence on M is reflected in the scaling of the signature μ). The limit of the contour integral formula for $f(\mu; [v]^M, \rho)$ is derived in Section 4.5 using a steepest descent argument, while the combinatorial estimates for $\mathsf{F}_\mu([u]^k)$ prove sufficient for taking its limit. Combining our two asymptotic results for $\mathsf{F}_\mu([u]^k)$ and $f(\mu; [v]^M, \rho)$ we can prove that the sequence of random vectors in $\mathbb{R}^k$, given by $Y^k(N, M) = \left(Y_1^k(N, M; k), \ldots, Y_k^k(N, M; k)\right)$ with $Y(N, M; k)$ as in Theorem 1.4.4 converges to the measure of the ordered eigenvalues of a random GUE matrix $\mu_{\mathrm{GUE}}^k(dx_1, ..., dx_k)$, given by

$$\mu_{\mathrm{GUE}}^k(dx_1, ..., dx_k) = \mathbf{1}\{x_k > x_{k-1} > \cdots > x_1\} \left(\frac{1}{\sqrt{2\pi}}\right)^k \cdot \frac{1}{\prod_{i=1}^{k-1} i!} \cdot \prod_{1 \le i < j \le k} (x_i - x_j)^2 \prod_{i=1}^{k} e^{-\frac{x_i}{2}} dx_i.$$

The last statement appears as Proposition 4.4.3 in the text.

The above paragraph explains how we show that the top row of $Y(N, M; k)$ converges to the top row of the GUE-corners process of rank k. To obtain the full convergence statement we combine our top row convergence statement with the general formalism, introduced in [71], involving Gibbs

measures on interlacing arrays. In more detail, the top-row convergence of $Y(N, M; k)$ and the interlacing conditions

$$\lambda_1^{i+1}(\pi) \geq \lambda_1^i(\pi) \geq \lambda_2^{i+1}(\pi) \geq \lambda_2^i(\pi) \geq \cdots \geq \lambda_i^i(\pi) \geq \lambda_{i+1}^{i+1}(\pi),$$

for $i = 1, \ldots, k-1$ are enough to conclude the tightness of the full vector $Y(N, M; k)$ as $M \to \infty$. For each N, M the measures $\mathbb{P}_{u,v}^{N,M}$ satisfy what we call the *six-vertex Gibbs property* and in the $M \to \infty$ limit this property becomes what is known as the *continuous Gibbs property*, see Section 4.4.2. Combining the latter statements, one can conclude that any subsequential limit of $Y(N, M; k)$ as $M \to \infty$ has top row distribution $\mu_{\mathrm{GUE}}^k(dx_1, ..., dx_k)$ and satisfies the continuous Gibbs property, and these two characteristics are enough to identify this limit with the GUE-corners process of rank k. As the sequence $Y(N, M; k)$ is tight and all subsequential limits are the same and equal to the GUE-corners of rank k, we conclude the weak convergence of $Y(N, M; k)$. This argument is given in Section 4.4.2.

1.5 The method of steepest descent

Much of our work in Chapters 2,3 and 4 goes into taking asymptotics of exact integral formulas for observables of our models. An important piece of each of these asymptotic analyses is the method of steepest descent (also called saddle point approximation).

The method steepest descent is a procedure for finding the asymptotics of an integral of the form

$$I_M = \int_C e^{Mf(z)} dz,$$

as $M \to \infty$, where f is a holomorphic function and C is an integration contour in the complex plane. The technique is to find a critical point z_0 of f and deform the contour C so that it passes through z_0 in such a way that $\mathfrak{Re}[f(z)]$ decays quickly as z moves along the contour C away from z_0. In this situation $e^{Mf(z_0)}/e^{Mf(z)}$ has exponential decay in M. We use this, along with specific information about our f and C, to argue that the integral can be localized at z_0, i.e. the

asymptotics of $\int_{C\cap B_\varepsilon(z_0)} e^{Mf(z)}dz$ are the same as those of I_M ($B_\varepsilon(z_0)$ is a ball of radius ε around 0 in the complex plane). Then we Taylor expand f near z_0 and show that sufficiently high order terms do not contribute to the asymptotics. This converts the first term of the asymptotics of I_M into a simpler integral that we can often evaluate. The most difficult step in this procedure is usually finding a deformation of the contour C on which you can prove that $\mathfrak{Re}[f(z)]$ has a unique maximum at the point z_0.

In Chapters 2 and 3 the steepest descent argument begins with Fredholm determinant formulas $\det(1-K)_{L^2(C)}$ (see Definition 1.1.1) where the kernel $K(u,v)$ is written as a single integral over some contour $\mathcal{D}$. Loosely speaking we perform a steepest descent argument on both the contour C and the contour $\mathcal{D}$ to arrive at a recognizable Fredholm determinant formula (1.3) for the cumulative density function of the Tracy-Widom GUE distribution. Both these steepest descent arguments involve a double critical point of $f(z)$ so we end up with integrals of $e^{g(z)}$ where g is the third order Taylor expansion of f. This eventually leads to the appearance of the Airy function (1.2) in the asymptotics.

In Chapter 4 the steepest descent argument begins with a k fold integral over a single contour. We apply the same steepest descent argument in each variable to obtain the probability density function for the eigenvalues of a matrix in the Gaussian unitary ensemble. This steepest descent argument involves a single critical point z_0 for the function $f(z)$, so we end up with integrals of $e^{g(z)}$ where g is a second order Talyor expansion of f around z_0. These are Gaussian integrals and upon evaluating them we retain a Gaussian probability density function in each variable.

1.6 An Epidemiology model for inhomogeneous populations

The final chapter of this thesis will study a model for the spread of a disease. This model is not related to the KPZ universality class. The SIR (Susceptible Infected Recovered) model, introduced in [121], is one of the simplest models for the growth of an epidemic. It involves dividing the population into three compartments: those who are susceptible, those who are infected, and those who are recovered, and writing differential equations for the sizes of these compartments over

time. When S is the size of the susceptible compartment, I is the size of the infected compartment, and R is the size of the recovered component the equations are

$$\begin{aligned}\partial_t S_t &= -\beta S_t I_t,\\ \partial_t I_t &= \beta S_t I_t - \gamma I_t,\\ \partial_t R_t &= \gamma I_t,\end{aligned}$$

and $S + I + R = 1$, where $\beta > 0$ is the infection rate and $\gamma > 0$ is the recovery rate.

A standard short time approximation for this model is given by assuming that $S_t \sim 1$ for small t. This turns the equation for I_t into the simpler equation $\partial_t I_t \sim (\beta - \gamma) I_t$, with solution $I_t \sim I_0 e^{t(\beta-\gamma)}$. This approximation leads to an important quantity called the basic reproduction rate $R_0 = \frac{\beta}{\gamma}$ for the classical SIR model. R_0 is the typical number of secondary infections caused by a single infection. If $R_0 > 1$ the infection will spread and if $R_0 < 1$ the infection will decay exponentially until it dies off in the classical SIR model.

Features like vaccines, incubation time, loss of immunity, births, and deaths can be added into the model quite naturally through simple modifications of the differential equations or through adding additional compartments. However the SIR model involves a few essential assumptions which are more difficult to change. First, the model is mean field, meaning that it assumes the population is fully mixed so the amount of interaction between any two people in the population is the same. Second, the model is deterministic. A fully realistic model for diseases should include randomness coming from factors like how much a given infected person interacts with others while they are infected, and which of the people they interact with end up getting infected. The SIR model assumes that this randomness entirely washes out for a sufficiently large population which at least early in an epidemic will not be true. Third, the SIR model assumes that the population is homogeneous, i.e. there is no variability in how susceptible or infectious different people in the population are. None of these conditions can be relaxed without significantly changing the framework of the model.

Relaxing the assumption of homogeneity in the population is our focus in Chapter 5. We will

set the infection rate β in the SIR model to be the product of a susceptibility parameter s and an infectivity parameter σ. Then we let the susceptibility s and the infectivity σ be random variables whose joint law $p(\sigma, s)d\sigma ds$ we will specify. Instead of dividing the entire population into three compartments S, I, and R, for each choice of s and σ we divide the population with the given values of s and σ into three compartments $S(\sigma, s)$, $I(\sigma, s)$ and $R(\sigma, s)$. We then write a set of infinitely many coupled differential equations for these components.

$$
\begin{aligned}
&S_t(\sigma, s)+I_t(\sigma, s) + R_t(\sigma, s) = p(\sigma, s),\\
&\partial_t I_t(\sigma, s) = S_t(\sigma, s)s \int \sigma' I_t(\sigma', s') \,\mathrm{d}\sigma' \,\mathrm{d}s' - \gamma I_t(\sigma, s),\\
&\partial_t R_t(\sigma, s) = \gamma I_t(\sigma, s),
\end{aligned}
$$

where $S(\sigma, s) + I(\sigma, s) + R(\sigma, s) = p(\sigma, s)$ for all σ and s.

This inhomogeneous version of the SIR model is not entirely new. The case where susceptibility varies but infectivity does not appeared in [93, 129]. Our main contribution is finding the (unique) exact solution for these infinitely many coupled differential equations up to a time change that solves an explicit ODE. The solution takes the form

$$
I_t(\sigma, s) + R_t(\sigma, s) = p(\sigma, s) - S_0(\sigma, s)e^{-s\mathbb{E}[\sigma]\nu(t)},
$$

$$
I_t(\sigma, s) = -S_0(\sigma, s)\,\mathrm{e}^{-s\mathbb{E}[\sigma]\nu(t)} + \mathrm{e}^{-\gamma t}\big(p(\sigma, s) - R_0(\sigma, s)\big) + \gamma S_0(\sigma, s)\int_0^t \mathrm{d}t'\mathrm{e}^{-\gamma(t-t')-s\mathbb{E}[\sigma]\nu(t')}. \tag{1.29}
$$

where the time change $\nu(t)$ is the unique solution to the equation

$$
\partial_t \nu(t) = (1 - \gamma\nu(t)) - \frac{1}{\mathbb{E}[\sigma]}\int \left(\sigma S_0(\sigma, s)\mathrm{e}^{-s\mathbb{E}[\sigma]\nu(t)}\right) \mathrm{d}\sigma \,\mathrm{d}s,
$$

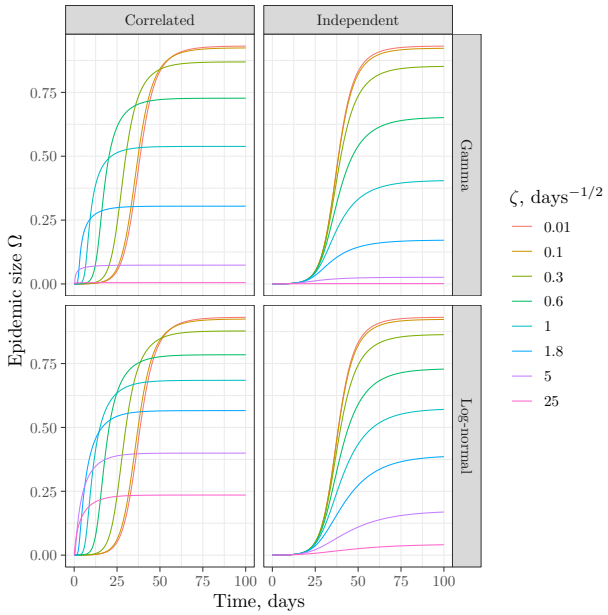

Figure 1.14: Comparison of epidemic spread for log-normal and Gamma distributions of infectivity and susceptibility with standard deviation ζ. The cases of independent or completely correlated σ and s are shown. We take $\mathbb{E}[s] = \mathbb{E}[\sigma] = 0.6$ and $\gamma = 1/8$ with initial conditions $I_0(\sigma, s) = 10^-4 p(\sigma, s)$.

with $\nu(0) = 0$. Numerically solving for $\nu(t)$ is quite easy and several numerical solutions for the total number of people who have been infected up to time t are shown in Figure 1.14.

We use this exact solution to show that in the limit where the initial infected population is taken close to zero in an appropriate sense, the total number of people who are ever infected before the disease dies out (called the final epidemic size Ω_∞) is given by

$$\Omega_\infty = 1 - \mathbb{E}[e^{-s\mathbb{E}[\sigma]L}] \tag{1.30}$$

where L is the unique positive root of

$$L - \frac{1}{\gamma} + \frac{1}{\gamma\mathbb{E}[\sigma]}\mathbb{E}[\sigma e^{-s\mathbb{E}[\sigma]L}] = 0, \tag{1.31}$$

if such a root exists. If no positive root exists then the infection begins with exponential decay rather than exponential growth, and as the beginning infected population is taken to zero, $\Omega_\infty = 0$.

We also give a short time approximation of the solution to the inhomogeneous SIR model

started with $R_0(\sigma, s) = 0$ and $I_0(\sigma, s) \ll p(\sigma, s)$. This approximate solution is

$$I_t(\sigma, s) \sim Cse^{(\mathbb{E}[\sigma s]-\gamma)t}.$$

An important upshot is that if a sufficiently small proportion of the population begins infected then the evolution of the disease depends very little on the initial conditions. In this short time approximation the initial distribution I_0 contributes only to the constant C. From this short time time approximation we see that the correct generalization for the reproduction rate for this inhomogeneous SIR model is $\mathcal{R}_0 = \frac{\mathbb{E}[\sigma s]}{\gamma}$, and sure enough when analyzing (1.31) one can see that a positive solution exists if and only if $\mathcal{R}_0 > 1$.

It is worth mentioning that the form of (1.30) and (1.31) allows us to show that if the marginal distributions of s and ω are fixed, then the joint law that maximizes the final epidemic size Ω_∞ is given by the percentile coupling where the nth most susceptible person is also the nth most infectious person. Similarly if s and σ are independent with $\mathbb{E}[\sigma]$ and $\mathbb{E}[s]$ fixed, then the final epidemic size Ω_∞ is maximized when s and σ are delta masses. This case where the distribution of susceptibility and infectiousness in the population are given by delta masses is just the classical SIR model.

Chapter 2: Tracy-Widom asymptotics for a river delta model

This chapter is based on the article [26] written by myself and Guillaume Barraquand.

2.1 Model and results

First passage percolation was introduced in 1965 to study a fluid spreading through a random environment [99]. This model has motivated many tools in modern probability, most notably Kingman's sub-additive ergodic theorem (see the review [14] and references therein); it has attracted attention from mathematicians and physicists alike due to the simplicity of its definition, and the ease with which fascinating conjectures can be stated.

The Kardar-Parisi-Zhang (KPZ) universality class has also become a central object of study in recent years [63]. Originally proposed to explain the behavior of growing interfaces in 1986 [115], it has grown to include many types of models including random matrices, directed polymers, interacting particle systems, percolation models, and traffic models. Much of the success in studying these has come from the detailed analysis of a few exactly solvable models of each type.

We study an exactly solvable model at the intersection of percolation theory and KPZ universality: Bernoulli-exponential first passage percolation (FPP). Here is a brief description (see Definition 2.1.1 for a more precise definition). Bernoulli-exponential FPP models the growth of a river delta beginning at the origin in $\mathbb{Z}^2_{\geq 0}$ and growing depending on two parameters $a, b > 0$. At time 0, the river is a single up-right path beginning from the origin chosen by the rule that whenever the river reaches a new vertex it travels north with probability $a/(a+b)$ and travels east with probability $b/(a+b)$ (thick black line in Figure 2.1). The line with slope a/b can be thought of as giving the direction in which the expected elevation of our random terrain decreases fastest.

As time passes the river erodes its banks creating forks. At each vertex which the river leaves

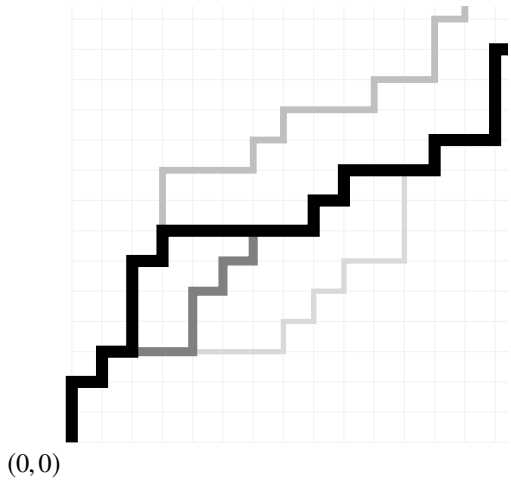

Figure 2.1: A sample of the river delta (Bernoulli-exponential FPP percolation cluster) near the origin. The thick black random walk path corresponds to the river (percolation cluster) at time 0. The other thinner and lighter paths correspond to tributaries added to the river delta (percolation cluster) at later times.

in the rightward (respectively upward) direction, it takes an amount of time distributed as an exponential random variable with rate a (resp. b) for the river to erode through its upward (resp. rightward) bank. Once the river erodes one of its banks at a vertex, the flow at this vertex branches to create a tributary (see gray paths in Figure 2.1). The path of the tributary is selected by the same rule as the path of the time 0 river, except that when the tributary meets an existing river it joins the river and follows the existing path. The full path of the tributary is added instantly when the river erodes its bank.

In this model the river is infinite, and the main object of study is the set of vertices included in the river at time t, i.e. the percolation cluster. We will also refer to the shape enclosed by the outermost tributaries at time t as the river delta (see Figure 2.2 for a large scale illustration of the river delta).

The model defined above can also be seen as the low temperature limit of the beta random walk in random environment (RWRE) model [23], an exactly solvable model in the KPZ universality class. Bernoulli-exponential FPP is particularly amenable to study because an exact formula for the distribution of the percolation cluster's upper border (Theorem 2.1.5 below) can be extracted

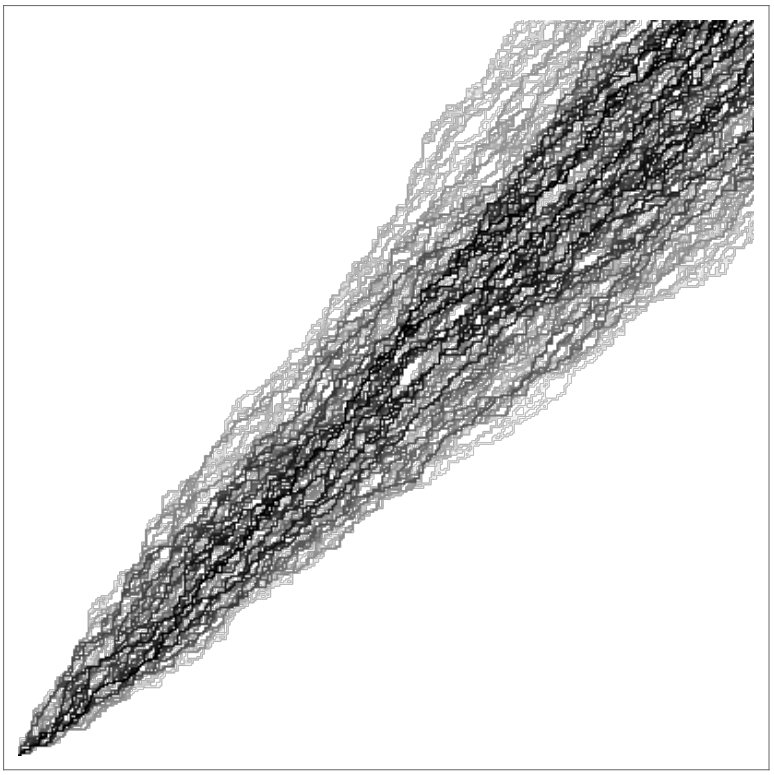

Figure 2.2: The percolation cluster for 400×400 Bernoulli-exponential FPP at time 1 with $a = b = 1$. Paths occurring earlier are shaded darker, so the darkest paths occur near $t = 0$ and the lightest paths occur near $t = 1$.

from an exact formula for the beta RWRE [23]. We perform an asymptotic analysis on this formula to prove that at any fixed time, the width of the river delta satisfies a law of large numbers type result with fluctuations converging weakly to the Tracy-Widom GUE distribution (see Theorem 2.1.4). Our law of large numbers result was predicted in [23] by taking a heuristic limit of [23, Theorem 1.19]; we present this non-rigorous computation in Section 2.1.3. We also give other interpretations of this result. In Section 2.1.5 we introduce an exactly solvable particle system and show that the position of a particle at finite time has Tracy-Widom fluctuations.

2.1.1 Definition of the model

We now define the model more precisely in terms of first passage percolation following [23].

Definition 2.1.1 (Bernoulli-exponential first passage percolation)**.** Let E_e be a family of indepen-

dent exponential random variables indexed by the edges e of the lattice $\mathbb{Z}^2_{\geq 0}$. Each E_e is distributed as an exponential random variable with parameter a if e is a vertical edge, and with parameter b if e is a horizontal edge. Let $(\zeta_{i,j})$ be a family of independent Bernoulli random variables with parameter $b/(a+b)$. We define the passage time t_e of each edge e in the lattice $\mathbb{Z}^2_{\geq 0}$ by

$$t_e = \begin{cases} \zeta_{i,j} E_e & \text{if } e \text{ is the vertical edge } (i,j) \to (i,j+1), \\ (1-\zeta_{i,j}) E_e & \text{if } e \text{ is the horizontal edge } (i,j) \to (i+1,j). \end{cases}$$

We define the point to point passage time $T^{\mathrm{PP}}(n,m)$ by

$$T^{\mathrm{PP}}(n,m) = \min_{\pi:(0,0)\to(n,m)} \sum_{e\in\pi} t_e.$$

where the minimum is taken over all up-right paths from $(0,0)$ to (n,m). We define the percolation cluster $C(t)$, at time t, by

$$C(t) = \left\{(n,m) : T^{\mathrm{PP}}(n,m) \leq t\right\}.$$

At each time t, the percolation cluster $C(t)$ is the set of points visited by a collection of up-right random walks in the quadrant $\mathbb{Z}^2_{\geq 0}$. $C(t)$ evolves in time as follows:

- At time 0, the percolation cluster contains all points in the path of a directed random walk starting from $(0,0)$, because at any vertex (i,j) we have passage time 0 to either $(i,j+1)$ or $(i+1,j)$ according to the independent Bernoulli random variables $\zeta_{i,j}$.
- At each vertex (i,j) in the percolation cluster $C(t)$, with an upward (resp. rightward) neighbor outside the cluster, we add a random walk starting from (i,j) with an upward (resp. rightward) step to the percolation cluster with exponential rate (a) (resp. b). This random walk will almost surely hit the percolation cluster after finitely many steps, and we add to the percolation cluster only those points that are in the path of the walk before the first hitting point (see Figure 2.1).

Define the height function $H_t(n)$ by

$$H_t(n) = \sup\{m \in \mathbb{Z}_{\geq 0} | T^{\text{PP}}(n, m) \leq t)\}, \tag{2.1}$$

so that $(n, H_t(n))$ is the upper border of $C(t)$.

2.1.2 History of the model and related results

Bernoulli-exponential FPP was first introduced in [23], which introduced an exactly solvable model called the beta random walk in random environment (RWRE) and studied Bernoulli-exponential FPP as a low temperature limit of this model (see also the physics works [184, 183] further studying the Beta RWRE and some variants). The beta RWRE was shown to be exactly solvable in [23] by viewing it as a limit of q-Hahn TASEP, a Bethe ansatz solvable particle system introduced in [158]. The q-Hahn TASEP was further analyzed in [41, 64, 193], and was recently realized as a degeneration of the higher spin stochastic six vertex model [5, 34, 46, 67], so that Bernoulli-exponential FPP fits as well in the framework of stochastic spin models.

Tracy-Widom GUE fluctuations were shown in [23] for Bernoulli-exponential FPP (see Theorem 2.1.2) and for Beta RWRE. In the Beta RWRE these fluctuations occur in the quenched large deviation principle satisfied by the random walk and for the maximum of many random walkers in the same environment.

The connection to KPZ universality was strengthened in subsequent works. In [65] it was shown that the heat kernel for the time reversed Beta RWRE converges to the stochastic heat equation with multiplicative noise. In [18] it was shown using a stationary version of the model that a Beta RWRE conditioned to have atypical velocity has wandering exponent $2/3$ (see also [54]), as expected in general for directed polymers in $1 + 1$ dimensions. The stationary structure of Bernoulli-exponential FPP was computed in [182] (In [182] Bernoulli-exponential FPP is referred to as the Bernoulli-exponential polymer).

The first occurrence of the Tracy-Widom distribution in the KPZ universality class dates back to

the work of Baik, Deift and Johansson on longest increasing subsequences of random permutations [16] (the connection to KPZ class was explained in e.g. [159]) and the work of Johansson on TASEP [112]. In the past ten years, following Tracy and Widom's work on ASEP [188, 186, 187] and Borodin and Corwin's Macdonald processes [36], a number of exactly solvable 1 + 1 dimensional models in the KPZ universality class have been analyzed asymptotically. Most of them can be realized as more or less direct degenerations of the higher-spin stochastic six-vertex model. This includes particle systems such as exclusion processes (q-TASEP [43, 21, 157, 156] and other models [24, 15, 91, 193]), directed polymers ([37, 42, 38, 68, 125, 151]), and the stochastic six-vertex model [6, 4, 22, 39, 45].

2.1.3 Main result

The study of the large scale behavior of passage times $T^{\mathrm{PP}}(n,m)$ was initiated in [23]. At large times, the fluctuations of the upper border of the percolation cluster (described by the height function $H_t(n)$) has GUE Tracy-Widom fluctuations on the scale $n^{1/3}$.

Theorem 2.1.2 ([23, Theorem 1.19])**.** *Fix parameters $a, b > 0$. For any $\theta > 0$ and $x \in \mathbb{R}$,*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{H_{\tau(\theta)n} - \kappa(\theta)n}{\tilde{\rho}(\theta)n^{1/3}} \le x\right) = F_{\mathrm{GUE}}(x), \tag{2.2}$$

where F_{GUE} is the GUE Tracy-Widom distribution (see Definition 2.2.3) and $\kappa(\theta)$, $\tau(\theta)$, $\tilde{\rho}(\theta) = \frac{\kappa'(\theta)}{\tau'(\theta)}\rho(\theta)$ are functions defined in [23] by

$$\begin{aligned}
\kappa(\theta) &:= \frac{\frac{1}{\theta^2} - \frac{1}{(a+\theta)^2}}{\frac{1}{(a+\theta)^2} - \frac{1}{(a+b+\theta)^2}},\\
\tau(\theta) &:= \frac{1}{a+\theta} - \frac{1}{\theta} + \kappa(\theta)\left(\frac{1}{a+\theta} - \frac{1}{a+b+\theta}\right) = \frac{a(a+b)}{\theta^2(2a+b+2\theta)},\\
\rho(\theta) &:= \left[\frac{1}{\theta^3} - \frac{1}{(a+\theta)^3} + \kappa(\theta)\left(\frac{1}{(a+b+\theta)^3} - \frac{1}{(a+\theta)^3}\right)\right]^{1/3}.
\end{aligned}$$

Note that as θ ranges from 0 to ∞, $\kappa(\theta)$ ranges from $+\infty$ to a/b and $\tau(\theta)$ ranges from $+\infty$ to 0.

Remark 2.1.3. In [23] the limit theorem is incorrectly stated as

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\min_{i\le n} T^{\mathrm{PP}}(i,\kappa(\theta)n) - \tau(\theta)n}{\rho(\theta)n^{1/3}} \le x\right) = F_{\mathrm{GUE}}(x),$$

but following the proof in [23, Section 6.1], we can see that the inequality and the sign of x should be reversed. Further, we have reinterpreted the limit theorem in terms of height function $H_t(n)$ instead of passage times $T^{\mathrm{PP}}(n,m)$ using the relation (2.1).

In this Chapter, we are interested in the fluctuations of $H_t(n)$ for large n but fixed time t. Let us scale θ in (2.2) above as

$$\theta = \left(\frac{na(a+b)}{2t}\right)^{1/3},$$

so that

$$\tau(\theta)n = t + O(n^{-1/3}).$$

Let us introduce constants

$$\lambda = \left(\frac{a(a+b)}{2t}\right)^{1/3}, \quad d = \frac{3a(a+b)}{2b\lambda}, \quad \sigma = \left(\frac{3a(a+b)\lambda}{2b^3}\right)^{1/3}. \tag{2.3}$$

Then, we have the approximations

$$\kappa(\theta)n = \frac{a}{b}n + dn^{2/3} + o(n^{4/9}),$$
$$\tilde{\rho}(\theta)n^{1/3} = \sigma n^{4/9} + o(n^{4/9}).$$

Thus, formally letting θ and n go to infinity in (2.2) suggests that for a fixed time t, it is natural to scale the height function as

$$H_t(n) = \frac{a}{b}n + dn^{2/3} + \sigma n^{4/9}\chi_n,$$

and study the asymptotics of the sequence of random variables χ_n.

Our main result is the following.

Theorem 2.1.4. *Fix parameters $a, b > 0$. For any $t > 0$ and $x \in \mathbb{R}$,*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{H_t(n) - \frac{a}{b}n - dn^{2/3}}{\sigma n^{4/9}} \leq x\right) = F_{GUE}(x),$$

where F_{GUE} is the GUE Tracy-Widom distribution.

Note that the heuristic argument presented above to guess the scaling exponents and the expression of constants d and σ is not rigorous, since Theorem 2.1.2 holds for fixed θ. Theorem 2.1.2 could be extended without much effort to a weak convergence uniform in θ for θ varying in a fixed compact subset of $(0, +\infty)$. However the case of θ and n simultaneously going to infinity requires more careful analysis. Indeed, for θ going to infinity very fast compared to n, Tracy-Widom fluctuations would certainly disappear as this would correspond to considering the height function at time $\tau(\theta)n \approx 0$, that is a simple random walk having Gaussian fluctuations on the $n^{1/2}$ scale. We explain in the next section how we shall prove Theorem 2.1.4.

The scaling exponents in Theorem 2 might seem unusual, although the preceding heuristic computation explains how they result from rescaling a model which has the usual KPZ scaling exponents. A similar situation occurs for scaling exponents of the height function of directed last passage percolation in thin rectangles [17, 33] and for the free energy of directed polymers [13] under the same limit.

2.1.4 Outline of the Proof

Recall that given an integral kernel $\mathsf{K} : \mathbb{C}^2 \to \mathbb{C}$, its Fredholm determinant is defined as

$$\det(1 + \mathsf{K})_{L^2(C)} := \frac{1}{2\pi\mathbf{i}} \sum_{n=0}^{\infty} \frac{1}{n!} \int_{C^n} \det[\mathsf{K}(x_i, x_j)]_{i,j=1}^{n} dx_1 \ldots dx_n.$$

To prove Theorem 2.1.4 we begin with the following Fredholm determinant formula for $\mathbb{P}(H_t(n) < m)$, and perform a saddle point analysis.

Theorem 2.1.5 ([23, Theorem 1.18])**.**

$$\mathbb{P}(H_t(n) < m) = \det(I - K_n)_{\mathbb{L}^2(C_0)},$$

where C_0 is a small positively oriented circle containing 0 but not $-a-b$, and $K_n : \mathbb{L}^2(C_0) \to \mathbb{L}^2(C_0)$ is defined by its integral kernel

$$K_n(u,u') = \frac{1}{2\pi\mathbf{i}} \int_{1/2-i\infty}^{1/2+i\infty} \frac{e^{ts}}{s} \frac{g(u)}{g(s+u)} \frac{ds}{s+u-u'}, \quad \text{where} \tag{2.4}$$

$$g(u) = \left(\frac{a+u}{u}\right)^n \left(\frac{a+u}{a+b+u}\right)^m \frac{1}{u}. \tag{2.5}$$

Remark 2.1.6. Note that [23, Theorem 1.18] actually states $\mathbb{P}(H_t(n) < m) = \det(I + K_n)_{\mathbb{L}^2(C_0)}$, instead of $\det(I - K_{t,n})_{\mathbb{L}^2(C_0)}$ due to a sign mistake.

This result was proved in [23] by taking a zero-temperature limit of a similar formula for the Beta RWRE obtained using the Bethe ansatz solvability of q-Hahn TASEP and techniques from [36, 43]. The integral (2.4) above is oscillatory and does not converge absolutely, but we may deform the contour so that it does. We will justify this deformation in Section 2.2.

Theorem 2.1.4 is proven in Section 2 by applying steep descent analysis to $\det(1-K_n)$, however the proofs of several key lemmas are deferred to later sections. The main challenge in proving Theorem 2.1.4 comes from the fact that, after a necessary change of variables $\omega = n^{-1/3}u$, the contours of the Fredholm determinant are being pinched between poles of the kernel K_n at $\omega = 0$ and $\omega = \frac{-a-b}{n^{1/3}}$ as $n \to \infty$. In order to show that the integral over the contour near 0 does not affect the asymptotics, we prove bounds for K_n near 0, and carefully choose a family of contours C_n on which we can control the kernel. This quite technical step is the main goal of Section 3. Section 4 is devoted to bounding the Fredholm determinant expansion of $\det(1 - K_n)_{L^2(C_n)}$, in order to justify the use of dominated convergence in Section 2.

2.1.5 Other interpretations of the model

There are several equivalent interpretations of Bernoulli-exponential first passage percolation. We will present the most interesting here.

A particle system on the integer line

The height function of the percolation cluster $H_t(n)$ is equivalent to the height function of an interacting particle system we call geometric jump pushTASEP, which generalizes pushTASEP (the $R = 0$ limit of PushASEP introduced in [44]) by allowing jumps of length greater than 1. This model is similar to Hall-Littlewood pushTASEP introduced in [91], but has a slightly different particle interaction rule.

Definition 2.1.7 (Geometric jump pushTASEP)**.** Let Geom(q) denote a geometric random variable with $\mathbb{P}(\text{Geom}(q) = k) = q^k(1-q)$. Let $1 \leq p_1(t) < p_2(t) < \ldots < p_i(t) < \ldots$ be the positions of ordered particles in $\mathbb{Z}_{\geq 1}$. At time $t = 0$ the position $n \in \mathbb{Z}_{\geq 0}$ is occupied with probability $b/(a+b)$. Each particle has an independent exponential clock with parameter a, and when the clock corresponding to the particle at position p_i rings, we update each particle position p_j in increasing order of j with the following procedure. ($p_i(t-)$ denotes the position of particle i infinitesimally before time t.)

- If $j < i$, then p_j does not change.
- p_i jumps to the right so that the difference $p_i(t) - p_i(t-)$ is distributed as $1 + \text{Geom}(a/(a+b))$
- If $j > i$, then
  - If the update for $p_{j-1}(t)$ causes $p_{j-1}(t) \geq p_j(t-)$, then $p_j(t)$ jumps right so that $p_j(t) - p_{j-1}(t)$ is distributed as $1 + \text{Geom}(a/(a+b))$.
  - Otherwise p_j does not change.
  - All the geometric random variables in the update procedure are independent.

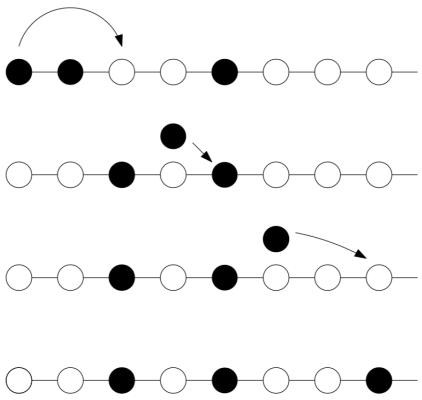

Figure 2.3: This figure illustrates a single update for geometric jump pushTASEP. The clock corresponding to the leftmost particle rings, activating the particle. The first particle jumps 2 steps pushing the next particle and activating it. This particle jumps 1 step pushing the rightmost particle and activating it. The rightmost particle jumps 3 steps, and all particles are now in their original order, so the update is complete.

Another way to state the update rule is that each particle jumps with exponential rate a, and the jump distance is distributed as $1 + \text{Geom}(a/(a+b))$. When a jumping particle passes another particle, the passed particle is pushed a distance $1 + \text{Geom}(a/(a+b))$ past the jumping particle's ending location (see Figure 2.3).

The height function $\overline{H}_t(n)$ at position n and time t is the number of unoccupied sites weakly to the left of n. If we begin with the distribution of $(n, H_t(n))$ in our percolation model, and rotate the first quadrant clockwise 45 degrees, the resulting distribution is that of $(n, \overline{H}_t(n))$. The horizontal segments in the upper border of the percolation cluster correspond to the particle positions, thus

$$H_t(n) = p_t(n) - n = \sup\{k : \overline{H}_t(n+k) \geq k\}.$$

A direct translation of Theorem 2.1.4 gives:

Corollary 2.1.8. *Fix parameters $a, b > 0$. For any $t > 0$ and $x \in \mathbb{R}$,*

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{p_t(n) - \left(\frac{a+b}{b}\right) n - dn^{2/3}}{\sigma n^{4/9}} \leq x \right) = F_{\text{GUE}}(x),$$

where $F_{\mathrm{GUE}}(x)$ is the Tracy-Widom GUE distribution.

To the authors knowledge Corollary 2.1.8 is the first result in interacting particle systems showing Tracy-Widom fluctuations for the position of a particle at finite time.

Degenerations

If we set $b = 1, t' = t/a$, and $a \to 0$, then in the new time variable t' each particle performs a jump with rate 1 and with probability going to 1, each jump is distance 1, and each push is distance 1. This limit is pushTASEP on $\mathbb{Z}_{\geq 0}$ where every site is occupied by a particle at time 0. Recall that in pushTASEP, the dynamics of a particle are only affected by the (finitely many) particles to its left, so this initial data makes sense.

We can also take a continuous space degeneration. Let x be the spatial coordinate of geometric jump pushTASEP, and let $\exp(\lambda)$ denote an exponential random variable with rate λ. Choose a rate $\lambda > 0$, and set $b = \frac{\lambda}{n}, x' = x/n, a = \frac{n-\lambda}{n}$, and let $n \to \infty$. Then our particles have jump rate $\frac{n-\lambda}{n} \to 1$, jump distance $\frac{\mathrm{Geom}(1-\lambda/n)}{n} \to \exp(\lambda)$, and push distance $\frac{\mathrm{Geom}(1-\lambda/n)}{n} \to \exp(\lambda)$. This is a continuous space version of pushTASEP on $\mathbb{R}_{\geq 0}$ with random initial conditions such that the distance between each particle position p_i and its rightward neighbor p_{i+1} is an independent exponential random variable of rate λ. Each particle has an exponential clock, and when the clock corresponding to the particle at position p_i rings, an update occurs which is identical to the update for geometric jump pushTASEP except that each occurrence of the random variable $1 + \mathrm{Geom}(a/(a+b))$ is replaced by the random variable $\exp(\lambda)$.

A benchmark model for travel times in a square grid city

The first passage times of Bernoulli-exponential FPP can also be interpreted as the minimum amount of time a walker must wait at streetlights while navigating a city [61]. Consider a city, whose streets form a grid, and whose stoplights have i.i.d exponential clocks. The first passage time of a point (n, m) in our model has the same distribution as the minimum amount of time a walker in the city has to wait at stoplights while walking n streets east and m streets north. Indeed

at each intersection the walker encounters one green stoplight with zero passage time and one red stoplight at which they must wait for an exponential time. Note that while the first passage time is equal to the waiting time at stoplights along the best path, the joint distribution of waiting times of walkers along several paths is different from the joint passage times along several paths in Bernoulli-exponential FPP.

2.1.6 Further directions

Bernoulli-exponential FPP has several features that merit further investigation. From the perspective of percolation theory, it would be interesting to study how long it takes for the percolation cluster to contain all vertices in a given region, or how geodesics from the origin coalesce as two points move together.

From the perspective of KPZ universality, it is natural to ask: what is the correlation length of the upper border of the percolation kernel, and what is the joint law of the topmost few paths.

Under diffusive scaling limit, the set of coalescing simple directed random walks originating from every point of $\mathbb{Z}^2$ converges to the Brownian web [80, 81]. Hence the set of all possible tributaries in our model converges to the Brownian web. One may define a more involved set of coalescing and branching random walks which converges to a continuous object called the Brownian net ([148], [179], see also the review [167]). Thus, it is plausible that there exist a continuous limit of Bernoulli-Exponential FPP where tributaries follow Brownian web paths and branch at a certain rate at special points of the Brownian web used in the construction of the Brownian net.

After seeing Tracy-Widom fluctuations for the edge statistics it is natural to ask whether the density of vertices inside the river along a cross section is also connected to random matrix eigenvalues and whether a statistic of this model converges to the positions of the second, third, etc. eigenvalues of the Airy point process.

2.1.7 Notation and conventions

We will use the following notation and conventions.

- $B_\varepsilon(x)$ will denote the open ball of radius $\varepsilon > 0$ around the point x.
- $\mathfrak{Re}[x]$ will denote the real part of a complex number x, and $\mathfrak{Im}[x]$ denotes the imaginary part.
- C and γ with any upper or lower indices will always denote an integration contour in the complex plane. K with any upper or lower indices will always represent an integral kernel. A lower index like γ_r, C_n, or K_n will usually index a family of contours or kernels. An upper index such as γ^ε, C^ε, or K^ε will indicate that we are intersecting our contour with a ball of radius ε, or that the integral defining the kernel is being restricted to a ball of radius ε.

Acknowledgements

The authors thank Ivan Corwin for many helpful discussions and for useful comments on an earlier draft of the Chapter. The authors thank an anonymous reviewer for detailed and helpful comments on the manuscript. G. B. was partially supported by the NSF grant DMS:1664650. M. R. was partially supported by the Fernholz Foundation's "Summer Minerva Fellow" program, and also received summer support from Ivan Corwin's NSF grant DMS:1811143.

2.2 Asymptotics

2.2.1 Setup

The steep descent method is a method for finding the asymptotics of an integral of the form

$$I_M = \int_C e^{Mf(z)} dz,$$

as $M \to \infty$, where f is a holomorphic function and C is an integration contour in the complex plane. The technique is to find a critical point z_0 of f, deform the contour C so that it passes through

z_0 and $\mathfrak{Re}[f(z)]$ decays quickly as z moves along the contour C away from z_0. In this situation $e^{Mf(z_0)}/e^{Mf(z)}$ has exponential decay in M. We use this along with specific information about our f and C, to argue that the integral can be localized at z_0, i.e. the asymptotics of $\int_{C\cap B_\varepsilon(z_0)} e^{Mf(z)}dz$ are the same as those of I_M. Then we Taylor expand f near z_0 and show that sufficiently high order terms do not contribute to the asymptotics. This converts the first term of the asymptotics of I_M into a simpler integral that we can often evaluate.

In Section 2.1 we will manipulate our formula for $\mathbb{P}(h(n) < m)$, and find a function f_1 so that the kernel K_n can be approximated by an integral of the form $\int_{\lambda+\mathbf{i}\mathbb{R}} e^{n^{1/3}[f_1(z)-f_1(\omega)]}dz$. Approximating K_n in this way will allow us to apply the steep descent method to both the integral defining K_n and the integrals over C_0 in the Fredholm determinant expansion.

For the remainder of the Chapter we fix a time $t > 0$, and parameters $a, b > 0$. All constants arising in the analysis below depend on those parameters t, a, b, though we will not recall this dependency explicitly for simplicity of notation.

We also fix henceforth

$$m = \left\lfloor \frac{a}{b}n + dn^{2/3} + n^{4/9}\sigma x \right\rfloor. \tag{2.6}$$

We consider K_n and change variables setting $\tilde{z} = s + u$, $d\tilde{z} = ds$ to obtain

$$\tilde{K}_n(u,u') = \frac{1}{2\pi\mathbf{i}} \int_{1/2+u-\mathbf{i}\infty}^{1/2+u+\mathbf{i}\infty} \frac{e^{t(\tilde{z}-u)}}{(\tilde{z}-u)(\tilde{z}-u')} \frac{g(u)}{g(\tilde{z})} d\tilde{z}.$$

In the following lemma, we change our contour of integration in the $\tilde{z}$ variable so that it does not depend on u.

Lemma 2.2.1. *For every fixed n,*

$$\tilde{K}_n(u,u') = \frac{1}{2\pi\mathbf{i}} \int_{n^{1/3}\lambda+\mathbf{i}\mathbb{R}} \frac{e^{t(\tilde{z}-u)}}{(\tilde{z}-u)(\tilde{z}-u')} \frac{g(u)}{g(\tilde{z})} d\tilde{z}.$$

Proof. Choose the contour C_0 to have radius $0 < r < \min[1/4, \lambda]$. This choice of r means that we do not cross C_0 when deforming the contour $1/2 + u + \mathbf{i}\mathbb{R}$ to $\lambda + \mathbf{i}\mathbb{R}$. In this region K is a

holomorphic function, so this deformation does not change the integral provided that for M real,

$$\frac{1}{2\pi\mathbf{i}}\int_{1/2+u+\mathbf{i}M}^{n^{1/3}\lambda+\mathbf{i}M}\frac{e^{t(\tilde{z}-u)}}{(\tilde{z}-u)(\tilde{z}-u')}\frac{g(u)}{g(\tilde{z})}d\tilde{z}\xrightarrow[M\to\pm\infty]{}0.$$

This integral converges to 0 because for all $\tilde{z}\in[n^{1/3}\lambda-\mathbf{i}M,1/2+u-\mathbf{i}M]\cup[n^{1/3}\lambda+\mathbf{i}M,1/2+u+\mathbf{i}M]$ we have

$$\left|\frac{1}{(\tilde{z}-u)(\tilde{z}-u')g(\tilde{z})}\right|\sim\frac{1}{M},$$

as $M\to\infty$.

□

Set

$$\tilde{h}_n(z)=-n\log\left(\frac{a+z}{z}\right)-m\log\left(\frac{a+z}{a+b+z}\right),\qquad\text{so that}\qquad e^{\tilde{h}_n(z)}=\frac{z}{g(z)}.$$

Then

$$K_n(u,u')=\frac{1}{2\pi\mathbf{i}}\int_{n^{1/3}\lambda+\mathbf{i}\mathbb{R}}\frac{e^{t\tilde{z}+\tilde{h}_n(\tilde{z})}}{e^{tu+\tilde{h}_n(u)}}\frac{\tilde{z}}{u}\frac{d\tilde{z}}{(\tilde{z}-u)(\tilde{z}-u')}.$$

Now perform the change of variables

$$z=n^{-1/3}\tilde{z},\omega=n^{-1/3}u,\omega'=n^{-1/3}u'.$$

If we view our change of variables as occuring in the Fredholm determinant expansion, then due to the $d\omega_i$s, we see that scaling all variables by the same constant does not change the Fredholm determinant $\det(1-\mathsf{K}_n)_{L^2(C)}$. Thus our change of variables gives

$$K_n(\omega,\omega')=\frac{1}{2\pi\mathbf{i}}\int_{\lambda+\mathbf{i}\mathbb{R}}\frac{e^{n^{1/3}t(z-\omega)}}{(z-\omega)(z-\omega')}e^{h_n(z)-h_n(\omega)}\frac{z}{\omega}dz$$

where

$$h_n(z)=\tilde{h}_n(n^{1/3}z)=-n\log\left(\frac{a+n^{1/3}z}{n^{1/3}z}\right)-m\log\left(\frac{a+n^{1/3}z}{a+b+n^{1/3}z}\right).$$

Remark 2.2.2. The contour for ω, ω' becomes $n^{-1/3}C_0$ after the change of variables, but $\mathsf{K}_n(\omega,\omega')$ is holomorphic in most of the complex plane. Examining of the poles of the integrand for $\mathsf{K}_n(\omega,\omega')$, we see that we can deform the contour for ω,ω' in any way that does not cross the line $\lambda + \mathbf{i}\mathbb{R}$, the pole at $-(a+b)/n^{1/3}$, or the pole at 0, without changing the Fredholm determinant $\det(I - \mathsf{K}_n)_{L^2(n^{-1/3}C_0)}$.

Taylor expanding the logarithm in the variable n gives

$$h_n(z) = -n^{1/3}\left(\frac{a(a+b)}{2z^2} - \frac{bd}{z}\right) - n^{1/9}\left(\frac{-b\sigma x}{z}\right) + r_n(z).$$

Here $r_n(z) = O(1)$ in a sense that we make precise in Lemma 2.2.7. The kernel can be rewritten as

$$K_n(\omega,\omega') = \frac{1}{2\pi\mathbf{i}}\int_{\lambda+\mathbf{i}\mathbb{R}} \frac{\exp(n^{1/3}(f_1(z)-f_1(w)) + n^{1/9}(f_2(z)-f_2(\omega)) + (r_n(z)-r_n(\omega)))}{(z-\omega)(z-\omega')}\frac{z}{\omega}dz$$

where

$$f_1(z) = tz - \frac{a(a+b)}{2z^2} + \frac{bd}{z}, \qquad f_2(z) = \frac{b\sigma x}{z}. \tag{2.7}$$

We have approximated the kernel as an integral of the form $\int e^{n^{1/3}[f_1(z)-f_1(\omega)]}dz$. To apply the steep-descent method, we want to understand the critical points of the function f_1. We have

$$f_1'(z) = t + \frac{a(a+b)}{z^3} - \frac{db}{z^2}, \qquad f_1''(z) = -\frac{3a(a+b)}{z^4} + \frac{2bd}{z^3}, \qquad f_1'''(z) = \frac{12a(a+b)}{z^5} - \frac{6bd}{z^4}. \tag{2.8}$$

Where a, b are the parameters associated to the model. Let the constant λ be as defined in (4.51), then $0 = f_1'(\lambda) = f_1''(\lambda) = 0$, and

$$f_1'''(\lambda) = \frac{3a(a+b)}{\lambda^5} = 2\left(\frac{b\sigma}{\lambda^2}\right)^3 = 2\left(\frac{-f_2'(\lambda)}{x}\right)^3,$$

is a positive real number. σ is defined in equation (4.51).

Recall the definition of the Tracy-Widom GUE distribution, which governs the largest eigen-

value of a gaussian hermitian random matrix.

Definition 2.2.3. The Tracy-Widom distribution's distribution function is defined as $F_{\mathrm{GUE}}(x) = \det(1 - K_{\mathrm{Ai}})_{L^2(x,\infty)}$, where K_{Ai} is the Airy kernel,

$$K_{\mathrm{Ai}}(s,s') = \frac{1}{2\pi\mathbf{i}} \int_{e^{-2\pi i/3}\infty}^{e^{2\pi\mathbf{i}/3}\infty} d\omega \frac{1}{2\pi\mathbf{i}} \int_{e^{-\pi\mathbf{i}/3}\infty}^{e^{\pi\mathbf{i}/3}\infty} dz \frac{e^{z^3/3 - zs}}{e^{\omega^3/3 - \omega s'}} \frac{1}{(z-\omega)}.$$

In the above integral the two contours do not intersect. We can think of the inner integral following the contour $(e^{-\pi\mathbf{i}/3}\infty, 1] \cup (1, e^{\pi\mathbf{i}/3}\infty)$, and the outer integral following the contour $(e^{-2\pi\mathbf{i}/3}\infty, 0] \cup (0, e^{2\pi\mathbf{i}/3}\infty)$. Our goal through the rest of the Chapter is to show that the Fredholm determinant $\det(I - \mathsf{K}_n)$ converges to the Tracy-Widom distribution as $n \to \infty$.

2.2.2 Steep descent contours

Definition 2.2.4. We say that a path $\gamma : [a,b] \to \mathbb{C}$ is steep descent with respect to the function f at the point $x = \gamma(0)$ if $\frac{d}{dt}\Re\mathfrak{e}[f(\gamma(t))] > 0$ when $t > 0$, and $\frac{d}{dt}\Re\mathfrak{e}[f(\gamma(t))] < 0$ when $t < 0$.

We say that a contour C is steep descent with respect to a function f at a point x, if the contour can be parametrized as a path satisfy the above definition. Intuitively this statement means that as we move along the contour C away from the point x, the function f is strictly decreasing.

In this section we will find a family of contours γ_r for the variable z and so that γ_r is steep descent with respect to $\Re\mathfrak{e}[f_1(z)]$ at the point λ, and study the behavior of $\Re\mathfrak{e}[f_1]$. The contours C_n for ω are constructed in Section 2.3.

Lemma 2.2.5. *The contour $\lambda + \mathbf{i}\mathbb{R}$ is steep descent with respect to the function $\Re\mathfrak{e}[f_1]$ at the point λ.*

Proof. We have that

$$\frac{d}{dy}\Re\mathfrak{e}[f_1(\lambda + \mathbf{i}y)] = -\Im\mathfrak{m}[f_1'(\lambda + \mathbf{i}y)] = -\Im\mathfrak{m}\left[t + \frac{a(a+b)}{(\lambda + \mathbf{i}y)^3} - \frac{bd}{\lambda + \mathbf{i}y}\right].$$

Now using the relation $2bd\lambda = 3a(a+b)$ and computing gives

$$\frac{d}{dy}\Re\mathrm{e}[f_1(\lambda + \mathbf{i}y)] = \frac{-4a(a+b)y^3}{(\lambda^2 + y^2)^3}.$$

This derivative is negative when $y > 0$ and positive when $y < 0$.

□

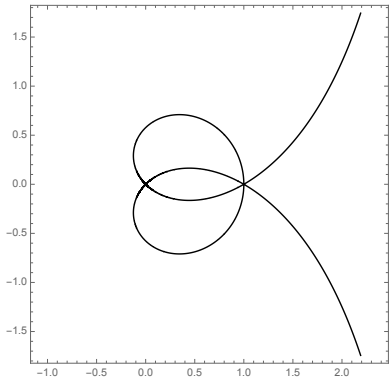

Figure 2.4: The level lines of the function $\Re\mathrm{e}[f_1(z)]$ at value $\Re\mathrm{e}[f_1(\lambda)]$. In this image we take $a = b = t = 1$.

Now we describe the contour lines of $\Re\mathrm{e}[f_1(z)]$ seen in Figure 2.4. $\Re\mathrm{e}[f_1]$ is the real part of a holomorphic function, so its level lines are constrained by its singularities, and because the singularities are not too complicated, we can describe its level lines. The contour lines of the real part of a holomorphic function intersect only at critical points and poles and the number of contour lines that intersect will be equal to the degree of the critical point or pole. We can see from the Taylor expansion of f_1 at λ, that there will be 3 level lines intersecting at λ with angles $\pi/6, \pi/2$, and $5\pi/6$. From the form of f_1, we see that there will be 2 level lines intersecting at 0 at angles $\pi/4$ and $3\pi/4$, and that a pair of contour lines will approach $\mathbf{i}\infty$ and $-\mathbf{i}\infty$ respectively with $\Re\mathrm{e}[z]$ approaching $f_1(\lambda)/t$. This shows that, up to a noncrossing continuous deformation of paths, the lines in Figure 2.4 are the contour lines $\Re\mathrm{e}[f_1(z)] = f_1(\lambda)$. We can also see that on the right side of the figure, tz will be the largest term of $\Re\mathrm{e}[f_1(z)]$, so our function will be positive. This determines the sign of $\Re\mathrm{e}[f_1(z)]$ in the other regions.

Our contour $\lambda + \mathbf{i}\mathbb{R}$ is already steep descent, but we will deform the tails, so that we can use

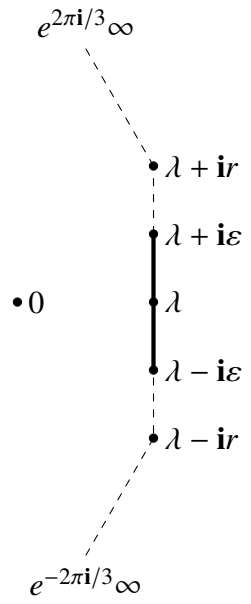

Figure 2.5: The contour γ_r is the infinite piecewise linear curve formed by the union of the vertical segment and the two semi infinite rays, oriented from bottom to top. The bold portion of this contour near λ is γ_r^ε.

dominated convergence in the next section.

Definition 2.2.6. For any $r > 0$, define the contour $\gamma_r = (e^{-2\pi\mathbf{i}/3}\infty, \lambda - r\mathbf{i}) \cup [\lambda - r\mathbf{i}, \lambda + r\mathbf{i}] \cup (\lambda + r\mathbf{i}, e^{2\pi\mathbf{i}/3}\infty)$ and $\gamma_r^\varepsilon = \gamma_r \cap B_\varepsilon(\lambda)$. These contours appear in Figure 2.5.

Because for any fixed n, we have $e^{h_n(z)} \to 1$ as $|z| \to \infty$, $\frac{z}{\omega(z-\omega)(z-\omega')}$ has linear decay in z, and $e^{n^{1/3}t(z-\omega)}$ has exponential decay in z, we can deform the vertical contour $\lambda + \mathbf{i}\mathbb{R}$ to the contour γ_r. Thus

$$K_n(\omega, \omega') = \int_{\gamma_r} \frac{e^{n^{1/3}t(z-\omega)}}{(z-\omega)(z-\omega')} e^{h_n(z)-h_n(\omega)} \frac{z}{\omega} dz.$$

The function $\mathfrak{Re}[f_1]$ is still steep descent on the contour γ_r with respect to the point λ. Lemma 2.2.5 shows that $\mathfrak{Re}[f_1]$ is steep descent on the segment $[\lambda - r\mathbf{i}, \lambda + r\mathbf{i}]$, and on $(e^{-2\pi\mathbf{i}/3}\infty, \lambda - r\mathbf{i}) \cup (\lambda + r\mathbf{i}, e^{2\pi\mathbf{i}/3}\infty)$ we inspect $f_1'(z)$ and note that for z sufficiently large, the constant term t dominates the other terms. Because our paths are moving in a direction with negative real component the contour γ_r is steep descent.

Up to this point we have been concerned with contours being steep descent with respect to

$\mathfrak{Re}[f_1]$, but the true function in our kernel is $\exp(n^{1/3}t(z-\omega) + h_n(z) - h_n(\omega))$. To show that γ_r is steep descent with respect to this function, we will need to control the error term $n^{1/3}tz + h_n(z) - n^{1/3}f_1(z) = n^{1/9}f_2(z) + r_n(z)$. The following lemma gives bounds on this error term away from $z = 0$.

Lemma 2.2.7. *For any $N, \varepsilon > 0$ there is a constant C depending only on ε, N such that*

$$|f_2(\omega)| \leq C \text{ and } |r_n(\omega)| \leq C, \tag{2.9}$$

for all $n \geq N$, and $\omega \geq \frac{|a+b|+\varepsilon}{N^{1/3}}$.
Similarly for any $\delta > 0$, there exists N_δ and C' depending only on δ, such that

$$|f_2'(\omega)| \leq C' \text{ and } |r_n'(\omega)| \leq C', \tag{2.10}$$

for all $n \geq N_\delta$, and ω satisfying $|\omega| \geq \delta$.

Lemma 2.2.7 is proved in Section 2.3.

At this point we have a contour γ_r for the variable z, which is steep descent with respect to $\mathfrak{Re}[f_1]$. We want to find a suitable contour for ω. The following lemma shows the existence of such a contour C_n, where property (c) below takes the place of being steep descent. This lemma is fairly technical and its proof is the main goal of Section 2.3. To see why observe that the function $n^{1/3}f_1(\omega)$ does not approximate $n^{1/3}t\omega - h_n(\omega)$ well when ω is near 0. The fact that the contribution near 0 is negligible is nontrivial because the function $n^{1/3}t\omega - h_n(\omega)$ has poles at 0 and $\frac{-a-b}{n^{1/3}}$, and our contour C_n is being pinched between them; we will use Lemma 2.2.8 to show that the asymptotics of $\det(1 - \mathsf{K}_n)_{L^2(C_n)}$ are not affected by these poles

Lemma 2.2.8. *There exists a sequence of contours $\{C_n\}_{n \geq N}$ such that:*

(a) For all n, the contour C_n encircles 0 counterclockwise, but does not encircle $(-a-b)n^{-1/3}$.

(b) C_n intersects the point λ at angles $-\pi/3$ and $-2\pi/3$.

(c) *For all $\varepsilon > 0$, there exists $\eta, N_\varepsilon > 0$ such that for all $n > N_\varepsilon$, $\omega \in C_n \setminus C_n^\varepsilon$ and $z \in \gamma_r$, we have*

$$\Re[n^{1/3}t(z-\omega) + h_n(z) - h_n(\omega)] \leq -n^{1/3}\eta,$$

where $C_n^\varepsilon = C_n \cap B_\varepsilon(\lambda)$.

(d) *There is a constant C such that for all $\omega \in C_n$,*

$$\Re[n^{1/3}t(\lambda-\omega) + h_n(\lambda) - h_n(\omega)] \leq n^{1/9}C.$$

The next lemma allows us to control $\Re[n^{1/3}tz + h_n(z)]$ on the contour γ_r.

Lemma 2.2.9. *For all $\varepsilon > 0$, and for sufficiently large r, there exists $C, N_\varepsilon > 0$, such that for all $\omega \in C_n$, and $z \in \gamma_r \setminus \gamma_r^\varepsilon$, then*

$$\Re[h_n(z) - h_n(\omega) + n^{1/3}t(z-\omega)] \leq -n^{-1/3}C.$$

Proof. We have already shown that γ_r is steep descent with respect to $f_1(z)$.

By Lemma 2.2.7, $|r_n| \leq C, |f_2| \leq Cn^{1/9}$ away from 0. We have

$$\begin{aligned} h_n(z) - h_n(\omega) + n^{1/3}t(z-\omega) =& n^{1/3}(f_1(z) - f_1(\omega)) + n^{1/9}(f_2(z) - f_2(\omega)) + (r_n(z) - r_n(\omega)) \\ \leq& n^{1/3}(f_1(z) - f_1(\omega)) + n^{1/9}C + C \leq n^{1/3}(f_1(z) - f_1(\omega) + \delta), \end{aligned}$$

for any sufficiently small $\delta > 0$. Because $f_1(z)$ is decreasing as we move away from λ, we have

$$n^{1/3}tz + h_n(z) < n^{1/3}t\lambda + h_n(\lambda) + Cn^{1/9}.$$

Thus by 2.2.7, we have that for all $\varepsilon > 0$ there exists C such that for $z \in \gamma_r \setminus \gamma_r^{\varepsilon}$,

$$\mathfrak{Re}[h_n(z) - h_n(\lambda) + n^{1/3}t(z-\lambda)] \leq -n^{1/3}C.$$

By Lemma 2.2.8 (d), we have

$$\mathfrak{Re}[h_n(\lambda) - h_n(\omega) + n^{1/3}t(\lambda-\omega)] \leq n^{1/9}C,$$

for $\omega \in C_n$. This completes the proof □

2.2.3 Localizing the integral

In this section we will use Lemma 2.2.8 and Lemma 2.2.9 to show that the asymptotics of $\det(1-K_n)_{L^2(C_n)}$ do not change if we replace C_n with $C_n^{\varepsilon} = C_n \cap B_{\varepsilon}(\lambda)$, and replace the contour γ_r defining K_n with the contour $\gamma_r^{\varepsilon} = \gamma_r \cap B_{\varepsilon}(0)$.

First we change variables setting $z = \lambda + n^{-1/9}\overline{z}, \omega = \lambda + n^{-1/9}\overline{\omega}$, and $\omega' = \lambda + n^{-1/9}\overline{z}$.

Definition 2.2.10. Define the contours $\mathcal{D}_0 = [-\mathbf{i}\infty, \mathbf{i}\infty]$, and $\mathcal{D}_0^{\delta} = \mathcal{D}_0 \cap B_{\delta}(0)$. (We will often use $\delta = n^{1/9}\varepsilon$.)

Our change of variables applied to the kernel $\mathsf{K}_n^{\varepsilon}$ gives

$$\overline{K}_n^{\varepsilon}(\overline{\omega},\overline{\omega}') = \frac{1}{2\pi\mathbf{i}}\int_{\mathcal{D}_0^{n^{1/9}\varepsilon}} \frac{1}{(\overline{z}-\overline{\omega})(\overline{z}-\overline{\omega}')}\frac{(\lambda+n^{-1/9}\overline{z})}{(\lambda+n^{-1/9}\overline{\omega})} e^{n^{1/3}f_1(\lambda+n^{-1/9}\overline{z})-f_1(\lambda+n^{-1/9}\overline{\omega})} \times e^{n^{1/9}f_2(\lambda+n^{-1/9}\overline{z})-f_2(\lambda+n^{-1/9}\overline{\omega})} e^{r_n(\lambda+n^{-1/9}\overline{z})-r_n(\lambda+n^{-1/9}\overline{\omega})} d\overline{z}. \quad (2.11)$$

Definition 2.2.11. The contours C_{-1} and C_{-1}^{ε} are defined as $C_{-1} = (e^{-2\pi\mathbf{i}/3}\infty, -1) \cup [-1, e^{2\pi\mathbf{i}/3}\infty)$ and $C_{-1}^{\varepsilon} = C_{-1} \cap B_{n^{1/9}\varepsilon}(-1)$.

By changing variables, for each m we have

$$\int_{(C_n^{\varepsilon})^m} \det(K_n^{\varepsilon}(\omega_i,\omega_j))_{i,j=1}^m d\omega_1 \ldots d\omega_m = \int_{(C_{-1}^{n^{1/9}\varepsilon})^m} \det(\overline{K}_n^{\varepsilon}(\overline{\omega}_i,\overline{\omega}_j))_{i,j=1}^m d\overline{\omega}_1 \ldots d\overline{\omega}_m.$$

This equality follows, because after rescaling the contour C_n^{ε}, we can deform it to the contour $C_{-1}^{n^{1/9}\varepsilon}$ without changing its endpoints. The previous equality implies

$$\det(1-K_n^{\varepsilon})_{L^2(C_{\varepsilon}^{\varepsilon})} = \det(1-\overline{K}_n^{\varepsilon})_{L^2(C_{-1}^{n^{1/9}\varepsilon})}.$$

We will make this change of variables often in the following arguments. Given a contour such as C_n or γ_r, we denote the contour after the change of variables by $\overline{C}_n$ or $\overline{\gamma}_r$. Now we are ready to localize our integrals.

Proposition 2.2.12. *For any sufficiently small $\varepsilon > 0$,*

$$\lim_{n\to\infty} \det(1-K_n(\omega,\omega'))_{L^2(C)} = \lim_{n\to\infty} \det(1-K_n^{\varepsilon}(\omega,\omega'))_{L^2(C_n^{\varepsilon})},$$

where

$$K_n^{\varepsilon} = \frac{1}{2\pi\mathbf{i}} \int_{\gamma_r^{\varepsilon}} \frac{e^{n^{1/3}t(z-\omega)+h_n(z)-h_n(\omega)}}{(z-\omega)(z-\omega')} \frac{z}{w} dz.$$

Proof. The proof will have two steps, and will use several lemmas that are proved in Section 4. In the first step we localize the integral in the z variable and show that $\lim_{n\to\infty} \det(1-K_n)_{L^2(C^{\varepsilon})} = \lim_{n\to\infty} \det(1-K_n^{\varepsilon})_{L^2(C^{\varepsilon})}$ using dominated convergence. In order to prove this, we appeal to Lemmas 2.4.1 and 2.4.2 to show that the Fredholm series expansions are indeed dominated. In the second step we localize the integral in the ω, ω' variables by using Lemma 2.4.3 to find an upper bound for $\det(1+K_n)_{L^2(C_n)} - \det(1+K_n)_{L^2(C_n^{\varepsilon})}$. Then we appeal to Lemma 2.4.4 to show that this upper bound converges to 0 as $n \to \infty$.

Step 1: By Lemma 2.2.9, for any $\varepsilon > 0$, there exists a $C', N > 0$ such that if $\omega \in C_n$ and $z \in \gamma_r \setminus \gamma_r^{\varepsilon}$, then for all $n > N$,

$$\mathfrak{Re}[h_n(z) - h_n(\omega) + n^{1/3}t(z-\omega)] \leq -n^{1/3}C'.$$

We bound our integrand on $\gamma_r \setminus \gamma_r^{\varepsilon}$, $\omega, \omega' \in C_n^{\varepsilon}$,

$$\left| \frac{e^{h_n(z)-h_n(\omega)+n^{1/3}t(z-\omega)}}{(z-\omega)(z-\omega')} \frac{z}{\omega} \right| \leq \frac{C}{\delta^2} z e^{-n^{1/3}C'} \xrightarrow[n\to\infty]{pointwise} 0.$$

(the δ^2 comes from the fact that $|z-\omega| \geq \delta$). By Lemma 2.2.7, there exists a $\eta > 0$ such that for sufficiently large n,

$$\left| \frac{e^{h_n(z)-h_n(\omega)+n^{1/3}t(z-\omega)}}{(z-\omega)(z-\omega')} \frac{z}{\omega} \right| < \left| \frac{e^{n^{1/3}(f_1(z)-f_1(\omega)+\eta)}}{(z-\omega)(z-\omega')} \frac{z}{\omega} \right|.$$

The linear term of $f_1(z)$ in (2.7) implies

$$\frac{1}{2\pi \mathbf{i}} \int_{\gamma_r} \left| \frac{e^{n^{1/3}(f_1(z)-f_1(\omega)+\eta)}}{(z-\omega)(z-\omega')} \frac{z}{\omega} \right| dz < \infty.$$

In the previous inequality we should write $|dz|$ instead of dz. We will often omit the absolute value in the $d\omega$ portion of the complex integral when the integrand is a positive real valued function.

So for each ω, ω', by dominated convergence

$$\frac{1}{2\pi \mathbf{i}} \int_{\gamma_r \setminus \gamma_r^{\varepsilon}} \frac{e^{h_n(z)-h_n(\omega)+n^{1/3}t(z-\omega)}}{(z-\omega)(z-\omega')} \frac{z}{\omega} dz \to 0 \quad \text{as} \quad n \to \infty,$$

So $\lim_{n\to\infty} K_n^{\varepsilon}(\omega, \omega') = \lim_{n\to\infty} K_n(\omega, \omega')$.

Now by Lemma 2.4.1, and 2.4.2, both Fredholm determinant expansions $\det(1-K_n)_{L^2(C^{\varepsilon})}$ and $\det(1-K_n^{\varepsilon})_{L^2(C^{\varepsilon})}$, are absolutely bounded uniformly in n. Thus we can apply dominated convergence to get

$$\lim_{n\to\infty} \det(1-K_n)_{L^2(C^{\varepsilon})} = \lim_{n\to\infty} \det(1-K_n^{\varepsilon})_{L^2(C^{\varepsilon})}. \tag{2.12}$$

Step 2: In the expansion

$$\det(1-K_n)_{L^2(C_n)} = \sum_{m=0}^{\infty} \frac{1}{m!} \int_{(C_n)^m} \det(K_n(\omega_i,\omega_j'))_{i,j=1}^n d\omega_1,...,d\omega_m.$$

The mth term can be decomposed as the sum

$$\int_{(C_n^\varepsilon)^m} \det(K_n(\omega_i,\omega_j))_{i,j=1}^n d\omega_1...d\omega_m + \int_{C_n^m\setminus(C_n^\varepsilon)^m} \det(K_n(\omega_i,\omega_j))_{i,j=1}^n d\omega_1...d\omega_m.$$

Lemma 2.4.3 along with Hadamard's bound on the determinant of a matrix in terms of it's row norms, implies that when $\omega_1 \in C_n \setminus C_n^\varepsilon$ and $\omega_2,...,\omega_m \in C^n$,

$$|\det(\overline{K}_n(\omega_i,\omega_j))_{i,j=1}^m| \le m^{m/2} M^{m-1/2} L_4 n^{4/9} e^{-n^{1/3}\eta} \to 0 \text{ as } n\to\infty. \tag{2.13}$$

Now let R be the maximum length of the paths C_n. The rescaled paths $\overline{C_n}$ will always have length less than $n^{1/9}R$. We have

$$\begin{aligned}
&\int_{C_n^m\setminus(C_n^\varepsilon)^m} |\det(K_n(\omega_i,\omega_j))_{i,j=1}^m| d\omega_1...d\omega_m \\
&\qquad \le m\int_{C_n\setminus C_n^\varepsilon} d\omega_1 \int_{C_n^{m-1}} |\det(K_n(\omega_i,\omega_j))_{i,j=1}^m| d\omega_2...d\omega_m \\
&\qquad \le m\int_{\overline{C}_n\setminus\overline{C}_n^\varepsilon} d\overline{\omega}_1 \int_{\overline{C}_n^{m-1}} |\det(\overline{K}_n(\overline{\omega}_i,\overline{\omega}_j))_{i,j=1}^m| d\overline{\omega}_2...d\overline{\omega}_m \\
&\qquad \le \int_{\overline{C}_n\setminus\overline{C}_n^\varepsilon} d\overline{\omega}_1 \int_{\overline{C}_n^{m-1}} m^{m/2} M^{(m-1)/2} L_4 n^{4/9} e^{-n^{1/3}\eta} d\overline{\omega}_2...d\overline{\omega}_m \\
&\qquad \le m(n^{1/9}R)^m m^{m/2} M^{(m-1)/2} L_4 n^{4/9} e^{-n^{1/3}\eta} \\
&\qquad \le e^{-n^{1/3}\eta}(n^{1/9})^m m^{1+m/2}(MR)^m n^{4/9}.
\end{aligned} \tag{2.14}$$

The first inequality follows from symmetry of the integrand in the ω_i. In the second inequality, we change variables from ω_i to $\overline{\omega}_i$. In the third inequality we use the first inequality of (2.13). In the fourth inequality, we use the fact that the total volume of our multiple integral is less than

$(n^{1/9}R)^m$. In the fifth inequality we rewrite and use $M^m > M^{(m-1)/2}$.

So we have

$$\sum_{m=1}^{\infty} \frac{1}{m!} \int_{C_n^m \setminus (C_n^\varepsilon)^m} |\det(K_n(\omega_i, \omega_j))_{i,j=1}^m| d\omega_1 \ldots d\omega_m$$

$$\leq \sum_{m=1}^{\infty} \frac{1}{m!} e^{-n^{1/3}\eta} (n^{1/9})^m m^{1+m/2} (MR)^m n^{4/9}$$

$$= n^{4/9} e^{-n^{1/3}\eta} \sum_{m=1}^{\infty} \frac{1}{m!} (MRn^{1/9})^m m^{1+m/2} \quad (2.15)$$

Applying Lemma 2.4.4 with $C = MRn^{1/9}$ gives.

$$n^{4/9} e^{-n^{1/3}\eta} \sum_{m=1}^{\infty} \frac{1}{m!} (MRn^{1/9})^m m^{1+m/2} \leq n^{4/9} e^{-n^{1/3}} 16 (MRn^{1/9})^4 e^{2(MR)^2 n^{2/9}} \xrightarrow[n\to\infty]{} 0.$$

Thus

$$\lim_{n\to\infty} \det(1 - K_n)_{L^2(C_n)} = \lim_{n\to\infty} \det(1 - K_n)_{L^2(C_n^\varepsilon)}. \quad (2.16)$$

Combining (2.12) and (2.16) concludes the proof of Proposition 2.11.

□

2.2.4 Convergence of the kernel

In this section we approximate $h_n(z) - h_n(\omega) + n^{1/3}t(z - \omega)$ by its Taylor expansion near λ, and show that this does not change the asymptotics of our Fredholm determinant.

Proposition 2.2.13. *For sufficiently small $\varepsilon > 0$,*

$$\lim_{n\to\infty} \det(1 - K_n^\varepsilon)_{L^2(C_\varepsilon^\varepsilon)} = \lim_{n\to\infty} \det(1 - K_{(x)})_{L^2(C_{-1})},$$

where

$$K_{(x)}(\overline{u},\overline{u}') = \frac{1}{2\pi\mathbf{i}}\int_{D'}\frac{e^{s^3/3-xs}}{e^{u^3-xu}}\frac{dz}{(z-u)(z-u')},$$

and

$$D' = (e^{-\pi\mathbf{i}/3}\infty, 0)\cup[0, e^{\pi\mathbf{i}/3}\infty).$$

Proof. Let

$$K(\overline{\omega},\overline{\omega}') = \frac{1}{2\pi\mathbf{i}}\int_{D'}\frac{d\overline{z}}{(\overline{z}-\overline{\omega})(\overline{z}-\overline{\omega}')}e^{f_1'''(\lambda)(\overline{z}^3-\overline{\omega}^3)/6+f_2'(\lambda)(\overline{z}-\overline{\omega})}, \tag{2.17}$$

We have seen in Section 2.2.3 that

$$\det(1-K_n^{\varepsilon}(\omega,\omega'))_{L^2(C_{\varepsilon}^{\varepsilon})} = \det(1-\overline{K}_n^{\varepsilon}(\overline{\omega},\overline{\omega}'))_{L^2(C_{-1}^{n^{1/9}\varepsilon})}.$$

The proof will have two main steps. In the first step we use dominated convergence to show that

$$\lim_{n\to\infty}\det(1-\overline{K}_n^{\varepsilon}(\overline{\omega},\overline{\omega}'))_{L^2(C_{-1}^{n^{1/9}\varepsilon})} = \lim_{n\to\infty}\det(1-\overline{K}_{(x)}(\overline{\omega},\overline{\omega}'))_{L^2(C_{-1}^{n^{1/9}\varepsilon})}.$$

In the second step we control the tail of the Fredholm determinant expansion to show that

$$\lim_{n\to\infty}\det(1-\overline{K}_{(x)}(\overline{\omega},\overline{\omega}'))_{L^2(C_{-1}^{n^{1/9}\varepsilon})} = \det(1-\overline{K}_{(x)}(\overline{\omega},\overline{\omega}'))_{L^2(C_{-1})}.$$

In step 1 we will use Lemma 2.4.1 to establish dominated convergence.

Step 1: We have the following pointwise convengences

$$\frac{\lambda+n^{-1/9}\overline{z}}{\lambda+n^{-1/9}\overline{\omega}}\to 1,$$

and for $z=\lambda+n^{-1/9}\overline{z}, \omega=\lambda+n^{-1/9}\overline{\omega}$,

$$n^{1/3}(f_1(z)-f_1(\omega))+n^{1/9}(f_2(z)-f_2(\omega))+r_n(z)-r_n(\omega)\to\frac{1}{6}f_1'''(\lambda)(\overline{z}^3-\overline{\omega}^3)+f_2'(\lambda)(\overline{z}-\overline{\omega}). \tag{2.18}$$

Because z is purely imaginary, for each $\overline{\omega},\overline{\omega}'$, the exponentiating the right hand side of (2.18) gives

a bounded function of $\overline{z}$ and $z/\omega \leq \frac{|\lambda+\varepsilon|}{|\lambda-\varepsilon|}$. The left hand side of (2.18) can be chosen to be within $\delta/n^{1/9}$ of the right hand side by choosing ε small by Taylor's theorem, because all the functions on the left hand side are holomorphic in $B_\varepsilon(\lambda)$. Thanks to the quadratic denominator $\frac{1}{(\overline{z}-\overline{\omega})(\overline{z}-\overline{\omega}')}$, we can apply dominated convergence to get

$$\overline{K}_n^{\varepsilon}(\overline{\omega},\overline{\omega}') \xrightarrow[n\to\infty]{pointwise} \frac{1}{2\pi\mathbf{i}} \int_{\mathbf{i}\mathbb{R}} \frac{d\overline{z}}{(\overline{z}-\overline{\omega})(\overline{z}-\overline{\omega}')} e^{f_1'''(\lambda)(\overline{z}^3-\overline{\omega}^3)/6+f_2'(\lambda)(\overline{z}-\overline{\omega})}. \tag{2.19}$$

Because the integrand on the right hand side of (2.19) has quadratic decay in $\overline{z}$, we can deform the contour from γ_0 to D' without changing the integral, so the right hand side is equal to $K(\overline{\omega},\overline{\omega}')$ from 2.17. Now by Lemma 2.4.1 we can apply dominated convergence to the expansion of the Fredholm determinant $\det(1-\overline{K}_n^{\varepsilon})_{L^2(C_{-1}^{n^{1/9}\varepsilon})}$, to get

$$\lim_{n\to\infty} \det(1-\overline{K}_n^{\varepsilon})_{L^2(C_{-1}^{n^{1/9}\varepsilon})} = \lim_{n\to\infty} \det(1-K)_{L^2(C_{-1}^{n^{1/9}\varepsilon})}.$$

Step 2: Now we make the change of variables $s = -(f_2'(\lambda)/x)\overline{z}$, $u = -(f_2'(\lambda)/x)\overline{\omega}$, and $u' = -(f_2'(\lambda)/x)\overline{\omega}'$. Keeping in mind that $-2(f_2'(\lambda)/x)^3 = f_1'''(\lambda)$, we get

$$K(\overline{\omega},\overline{\omega}') = K_{(x)}(u,u') = \frac{1}{2\pi\mathbf{i}} \int_{D'} \frac{e^{s^3/3-xs}}{e^{u^3/3-xu}} \frac{ds}{(s-u)(s-u')}.$$

Recall the expansion:

$$\det(1-K_{(x)})_{L^2(C_{-1}^{\varepsilon})} = \sum_{m=0}^{\infty} \frac{(-1)^m}{m!} \int_{C_{-1}^m} \det(K_{(x)}(\omega_i,\omega_j))_{i,j=1}^m d\omega_1 \ldots d\omega_m,$$

where $C_{-1} = (e^{-2\pi\mathbf{i}/3}\infty, 1] \cup (1, e^{2\pi\mathbf{i}/3}\infty)$, and C_{-1}^m is a product of m copies of C_{-1}.

$$|\det(1-K_{(x)})_{L^2(C_{-1})} - \det(1-K_{(x)})_{L^2(C_{-1}^{\varepsilon})}| \leq$$
$$\sum_{m=0}^{\infty} \frac{(-1)^m}{m!} \int_{C_{-1}^m \setminus (C_{-1}^{n^{1/9}\varepsilon})^m} |\det(K_{(x)}(\omega_i,\omega_j))_{i,j=1}^m| d\omega_1 \ldots d\omega_m,$$

so to conclude the proof of the proposition, we are left with showing that

$$\sum_{m=0}^{\infty} \frac{1}{m!} \int_{C_{-1}^m \setminus (C_{-1}^{n^{1/9}\varepsilon})^m} |\det(K_{(x)}(\omega_i, \omega_j))_{i,j=1}^m| d\omega_1 ... d\omega_m \xrightarrow[n\to\infty]{} 0 \tag{2.20}$$

Note that

$$\int_{C_{-1}^m \setminus (C_{-1}^{n^{1/9}\varepsilon})^m} |\det(K_{(x)}(\omega_i, \omega_j))_{i,j=1}^m| d\omega_1 ... d\omega_m \leq$$
$$m \int_{C_{-1} \setminus C_{-1}^{n^{1/9}\varepsilon}} \int_{C_{-1}^{m-1}} |\det(K_{(x)}(\omega_i, \omega_j))_{i,j=1}^m| d\omega_1 ... d\omega_m.$$

Set

$$M_1 = \int_{D'} |\overline{z} e^{f_1'''(\lambda)\overline{z}^3/6 + f_2'(\lambda)\overline{z}}| d\overline{z} < \infty.$$

Then $K_{(x)}(\omega, \omega') \leq M_1 e^{-|\omega|^3 - x|\omega|}$, and Hadamard's bound gives

$$|\det(K_{(x)}(\omega_i, \omega_j))_{i,j=1}^m| \leq m^{m/2} M_1^m \prod_{i=1}^m |e^{-\omega_i^3/3 + x\omega_i}|.$$

We have

$$\begin{aligned} &\int_{C_{-1} \setminus C_{-1}^{n^{1/9}\varepsilon}} \int_{C_{-1}^{m-1}} |\det(K_{(x)}(\omega_i, \omega_j))_{i,j=1}^m| d\omega_1 ... d\omega_m \\ &\qquad \leq M_1 \int_{C_{-1} \setminus C_{-1}^{n^{1/9}\varepsilon}} \int_{C_{-1}^{m-1}} \prod_{i=1}^m |e^{-\omega_i^3/3 + x\omega_i}| d\omega_1 ... d\omega_m \\ &\qquad \leq m^{1+m/2} M_1^m M_2^{m-1} \int_{C_{-1} \setminus C_{-1}^{n^{1/9}\varepsilon}} |e^{-\omega_1^3 + x\omega_1}| d\omega_1, \end{aligned} \tag{2.21}$$

where $M_2 = \int_{C_{-1}} |e^{-\omega^3 - x\omega}| d\omega < \infty$ because $-\omega^3$ lies on the negative real axis. (2.21) goes to zero because $n^{1/9}\varepsilon \to \infty$. So

$$\int_{C_{-1} \setminus C_{-1}^{n^{1/9}\varepsilon}} \int_{C_{-1}^{m-1}} \left| \det(K_{(x)}(\omega_i, \omega_j))_{i,j=1}^m \right| d\omega_1 ... d\omega_m \xrightarrow[n\to\infty]{} 0.$$

Note also that

$$\int_{C_{-1}^m \setminus (C_{-1}^{n^{1/9}\varepsilon})^m} \left|\det(K_{(x)}(\omega_i,\omega_j))_{i,j=1}^m\right| d\omega_1 \ldots d\omega_m \leq \int_{C_{-1}^m} |\det(K_{(x)}(\omega_i,\omega_j))_{i,j=1}^m| d\omega_1 \ldots d\omega_m$$
$$\leq m^{1+m/2} M_1 M_2^m.$$

By Stirling's approximation

$$\sum_{m=0}^{\infty} \frac{1}{m!} m^{1+m/2} M_1^m M_2^m < \infty.$$

So by dominated convergence (2.20) holds which concludes the proof of Proposition 2.2.13. □

2.2.5 Reformulation of the kernel

Now we use the standard $\det(1+AB) = \det(1+BA)$ trick [37, Lemma 8.6] to identify $\det(1 - K_{(x)})_{L^2(C_{-1})}$ with the Tracy-Widom cumulative distribution function.

Lemma 2.2.14. *For $x \in \mathbb{R}$,*

$$\det(1 - K_{(x)})_{L^2(C_{-1})} = \det(1 - K_{\mathrm{Ai}})_{L^2(x,\infty)}.$$

Proof. First note that because $\mathfrak{Re}[z-\omega] > 0$ along the contours we have chosen, we can write

$$\frac{1}{z-\omega} = \int_{\mathbb{R}_+} e^{-\lambda(z-\omega)} d\lambda.$$

Now let $A : L^2(C_{-1}) \to L^2(\mathbb{R}_+)$, and $B : L^2(\mathbb{R}_+) \to L^2(C_{-1})$ be defined by the kernels

$$A(\omega,\lambda) = e^{-\omega^3/3+\omega(x+\lambda)}, \tag{2.22}$$

$$B(\lambda,\omega') = \int_{e^{-\pi \mathbf{i}/3}\infty}^{e^{\pi \mathbf{i}/3}\infty} \frac{dz}{2\pi \mathbf{i}} \frac{e^{z^3/3-z(x+\lambda)}}{z-\omega'}. \tag{2.23}$$

We compute

$$\begin{aligned} AB(\omega,\omega') &= \int_{\mathbb{R}_+} e^{-\omega^3/3+\omega(x+\lambda)} \int_{e^{-\pi\mathbf{i}/3}\infty}^{e^{\pi\mathbf{i}/3}\infty} \frac{dz}{2\pi\mathbf{i}} \frac{e^{z^3/3-z(x+\lambda)}}{z-\omega'} \\ &= \frac{1}{2\pi\mathbf{i}} \int_{e^{-\pi\mathbf{i}/3}\infty}^{e^{\pi\mathbf{i}/3}\infty} \frac{e^{z^3/3-zx}}{e^{\omega^3/3-\omega x}} \frac{dz}{(z-\omega)(z-\omega')} \\ &= K_{(x)}(\omega,\omega'). \end{aligned}$$

Similarly,

$$BA(s,s') = \frac{1}{2\pi\mathbf{i}} \int_{e^{-2\pi\mathbf{i}/3}\infty}^{e^{2\pi\mathbf{i}/3}\infty} d\omega \frac{1}{2\pi\mathbf{i}} \int_{e^{-\pi\mathbf{i}/3}\infty}^{e^{\pi\mathbf{i}/3}\infty} dz \frac{e^{z^3/3-z(x+s)}}{e^{\omega^3/3-\omega(x+s')}} \frac{1}{(z-\omega)} = K_{\mathrm{Ai}}(x+s,x+s').$$

Because both A and B are Hilbert-Schmidt operators, we have

$$\begin{aligned} \det(1-K_{(x)})_{L^2(C)} = \det(1-AB)_{L^2(\mathbb{R}_+)} = \det(1-BA)_{L^2(\mathbb{R}_+)} \\ = \det(1-K_{\mathrm{Ai}})_{L^2(x,\infty)} = F_{\mathrm{GUE}}(x). \end{aligned}$$

□

2.3 Constructing the contour C_n

This section is devoted to constructing the contours C_n and proving Lemma 2.2.8. We will prove several estimates for $n^{1/3}\omega + h_n(\omega)$; then we will construct the contour C_n, and prove it satisfies the properties of Lemma 2.2.8. We begin by proving that we can approximate $n^{1/3}\omega + h_n(\omega)$ by $n^{1/3} f_1(\omega)$ away from 0.

2.3.1 Estimates away from 0: proof of Lemma 2.2.7

Both inequalities for $|f_2| = \frac{b\sigma x}{\omega}$ follow from the fact that f_2 and f_2' are bounded on $\mathbb{C} \setminus B_\varepsilon(0)$. Let $y = 1/\omega$, and let $m = n^{-1/9}$. Define the function $g(y,m) = r_n(\omega)$. First we prove (2.9). Note that $h_n(\omega)$ is holomorphic in y and m except when $n = \infty$, $n^{1/3}\omega = 0, -a-b$. By Taylor expanding

$h_n(\omega)$, we see that $r_n(\omega) = g(y,m)$ is holomorphic in y and m, except at points (y,m) such that $n^{1/3}\omega = 0, -a-b$, in particular there is no longer a pole when $n = \infty$. Thus for any N, $g(y,m)$ is holomorphic with variables y and m, in the region $U = \{(y,m) : n > N, \omega > |a+b|/N^{1/3}\}$, because in this region $n^{1/3}\omega > |a+b|$. The region $U_\varepsilon = \{(y,m) : n > N, \omega \geq \frac{|a+b|+\varepsilon}{N^{1/3}}\}$ is compact in the variables y and m, and because $U_\varepsilon \subset U$, the function $g(y,m)$ is holomorphic in the region U_ε. Thus $g(y,m) = r_n(\omega)$ is bounded by a constant C in the region U_ε.

Now we prove (2.10). For any δ, pick an arbitrary ε and an N_δ large enough that $\frac{|a+b|+\varepsilon}{N_\delta^{1/3}} \leq \delta$. Because $g(y,m) = r_n(\omega)$ is holomorphic in the variables y and m in the compact set U_ε, the function $\frac{\partial}{\partial y} g(y,m) = -\omega^2 r_n'(\omega)$, is also holomorphic in y, m. So $|\omega^2 r_n'(\omega)| \leq C$ on U_ε. We rewrite as $|r_n'(\omega)| \leq C/|\omega|^2$, and this gives $|r_n'(\omega)| \leq \frac{C}{|\delta|^2} \leq C'$, on the set $U_\varepsilon \cap (\mathbb{N} \times B_\delta(0)^c)$. But by our choice of N_δ, we have $U_\varepsilon \cap (\mathbb{N} \times B_\delta(0)^c)$ is just the set $\{(y,m) : n \geq N_\delta, |\omega| \geq \delta\}$.

2.3.2 Estimates near 0

The function $n^{1/3} f_1(\omega)$ only approximates $-n^{1/3}t\omega - h_n(\omega)$ well away from 0. In this section we give two estimates for $-n^{1/3}t\omega - h_n(\omega)$: one in Lemma 2.3.1 when ω is of order $n^{-1/3}$ and one in Lemma 2.3.3 when ω is of order $n^{\delta-1/3}$ for $\delta \in (0, 1/3)$. Together with Lemma 2.2.7 which gives an estimate when ω is of order 1, this will give us the tools we need to control $-n^{1/3}t\omega - h_n(\omega)$ along C_n. First to prove the bound in Lemma 2.3.1, we choose a path which crosses the real axis at $-a$, between the poles at 0 and $-a-b$ before rescaling $\tilde{h}_n$ to h_n. We show that after the rescaling, we can bound $\mathfrak{Re}[-n^{-1/3}\omega - h_n(\omega)]$ on this path for small ω.

Lemma 2.3.1. *Fix any $c_0 > 1$ and let $s = c_0(a+b)$. For $C = \log\left(\sqrt{s^2+a^2}\right) - \log(s) > 0$, we have*

$$\limsup_{n\to\infty} \frac{1}{n} \sup_{y\in[-s,s]} \mathfrak{Re}[h_n(\lambda) - h_n(\mathbf{i}n^{-1/3}y - n^{-1/3}a)] < -C.$$

Proof. Let $y \in [-s,s]$ and expand $e^{\mathfrak{Re}[h_n(\lambda) - h_n(iy - an^{-1/3})]}$ to get

$$\left(\frac{y}{\sqrt{y^2+a^2}}\right)^n \left(\frac{y}{\sqrt{y^2+b^2}}\right)^m \left(\frac{n^{1/3}\lambda}{n^{1/3}\lambda + a}\right)^n \left(\frac{a+b+n^{1/3}\lambda}{n^{1/3}\lambda + a}\right)^m.$$

The third factor is always less than 1. For sufficiently large n, the second factor times the fourth factor is less than 1, because $|y| \leq |s|$ while $n^{1/3}\lambda \to \infty$. We can bound the first factor by

$$\left|\frac{y}{\sqrt{y^2+a^2}}\right|^n \leq \left(\frac{s}{\sqrt{s^2+a^2}}\right)^n = e^{-nC},$$

with $C = \log\left(\sqrt{(s^2+a^2)}\right) - \log(s)$. □

Next we will prove the estimate for ω of order $n^{\delta-1/3}$. In this proof we will consider ω of the form $\omega = -n^{-1/3}a + \mathbf{i}n^{\delta-1/3}c(a+b)$, choose c sufficiently large, then let $n \to \infty$. The largest term in the expansion of $-n^{-1/3}\omega - h_n(\omega)$ will be of order $\frac{n^{1-2\delta}}{c^2}$. We introduce the following definition to let us ignore the terms which are negligible compared to $\frac{n^{1-2\delta}}{c^2}$ uniformly in δ.

Definition 2.3.2. Let A and B be functions depending on n and c, we say $A \sim_\delta B$ or A is δ-equivalent to B, if for sufficiently large c and n,

$$|A-B| \leq \frac{n^{2/3-2\delta}}{c^2}M_1 + \frac{n^{1-3\delta}}{c^3}M_2 + \frac{n^{4/9-\delta}}{c}M_3.$$

for some constants M_1, M_2, M_3 independent of c and n.

Now we prove the estimate.

Lemma 2.3.3. *For all $\delta \in (0,1/3)$, setting $\omega = -n^{-1/3}a + \mathbf{i}n^{\delta-1/3}c(a+b)$, gives*

$$\mathfrak{Re}[n^{1/3}t\omega + h_n(\omega)] \sim_\delta \mathfrak{Re}[n^{1/3}f_1(\omega)] \sim_\delta M\frac{n^{1-2\delta}}{c^2},$$

where $\sim_\delta$ is defined in Definition 8.

The proof of this Lemma 2.3.3 comes from Taylor expanding h_n and keeping track of the order of different terms with respect to n and c.

Proof. Recall that

$$h_n(\omega) = -n\log\left(1+\frac{a}{n^{1/3}\omega}\right) + m\log\left(1+\frac{b}{a+n^{1/3}\omega}\right). \tag{2.24}$$

For $|n^{1/3}\omega| > a$ and $|a + n^{1/3}\omega| > b$, we can Taylor expand in $n^{1/3}\omega$ to get

$$h_n(\omega) = -n \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \left(\frac{a}{n^{1/3}\omega}\right)^k + m \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \left(\frac{b}{a + n^{1/3}\omega}\right)^k.$$

Let $\omega = -n^{-1/3}a + \mathbf{i}n^{\delta-1/3}c(a+b)$ for $\delta \in (0, 1/3)$, so $|n^{1/3}\omega|, |a+n^{1/3}\omega| > n^{\delta}c(a+b) > c(a+b)$, for a constant c to be determined later. If $c > 2$, we have

$$\sum_{k=1}^{\infty} \left|\left(\frac{a}{n^{1/3}\omega}\right)\right|^k \leq \sum_{k=1}^{\infty} \left(\frac{b}{n^{\delta}c(a+b)}\right)^k \leq \frac{a}{n^{\delta}c(a+b)} \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^k \leq \frac{2a}{n^{\delta}c(a+b)} = \frac{n^{-\delta}}{c}M, \qquad (2.25)$$

and

$$\sum_{k=1}^{\infty} \left|\left(\frac{b}{a + n^{1/3}\omega}\right)\right|^k \leq \sum_{k=1}^{\infty} \left(\frac{a}{n^{\delta}c(a+b)}\right)^k \leq \frac{a}{n^{\delta}c(a+b)} \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^k = \frac{2a}{n^{\delta}c(a+b)} = \frac{n^{-\delta}}{c}M. \quad (2.26)$$

In what follows, we will use (2.25) or (2.26) when we say that an infinite sum is δ-equivalent to its first term.

We examine the first term in (2.24).

$$-n \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \left(\frac{a}{n^{1/3}\omega}\right)^k = -\left(\frac{a}{n^{1/3}\omega}\right) + \frac{1}{2}\left(\frac{a}{n^{1/3}\omega}\right)^2 - n \sum_{k=3}^{\infty} \frac{(-1)^{k+1}}{k} \left(\frac{a}{n^{1/3}\omega}\right)^k,$$
$$\sim_{\delta} -\left(\frac{a}{n^{1/3}\omega}\right) + \frac{1}{2}\left(\frac{a}{n^{1/3}\omega}\right)^2.$$

where the $\delta-$equivalence follows because $\left| n \sum_{k=3}^{\infty} \frac{(-1)^{k+1}}{k} \left(\frac{a}{n^{1/3}\omega}\right)^k \right| \leq \frac{n^{1-3\delta}}{c^3}M$ for some M by (2.25).

Recall that

$$m \sum_{k=1}^{\infty} \left(\frac{b}{a + n^{1/3}\omega}\right)^k = \left[\left(\frac{a}{b}\right)n + dn^{2/3} + \sigma x n^{4/9}\right] \sum_{k=1}^{\infty} \left(\frac{b}{a + n^{1/3}\omega}\right)^k.$$

We decompose this series as three sums. First the $\left(\frac{a}{b}\right) n$ term gives

$$\frac{a}{b} n \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \left(\frac{b}{a+n^{1/3}\omega}\right)^k =$$
$$n\left(\frac{a}{b}\right)\left(\frac{b}{a+n^{1/3}\omega}\right) - \frac{n}{2}\left(\frac{a}{b}\right)\left(\frac{b}{a+n^{1/3}\omega}\right)^2 + \frac{a}{b} n \sum_{k=3}^{\infty} \frac{(-1)^{k+1}}{k} \left(\frac{b}{a+n^{1/3}\omega}\right)^k$$
$$\sim_\delta n\left(\frac{a}{b}\right)\left(\frac{b}{a+n^{1/3}\omega}\right) - \frac{n}{2}\left(\frac{b}{a+n^{1/3}\omega}\right)^2,$$

because $\left|-\frac{a}{b} n \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \left(\frac{b}{a+n^{1/3}\omega}\right)^k\right| \leq M n^{1-3\delta}/c^3$ for some M. The second term is

$$dn^{2/3} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \left(\frac{b}{a+n^{1/3}\omega}\right)^k = dn^{2/3}\left(\frac{b}{a+n^{1/3}\omega}\right) - dn^{2/3} \sum_{k=2}^{\infty} \frac{(-1)^{k+1}}{k} \left(\frac{b}{a+n^{1/3}\omega}\right)^k$$
$$\sim_\delta dn^{2/3}\left(\frac{b}{a+n^{1/3}\omega}\right)$$

because $\left|dn^{2/3} \sum_{k=2}^{\infty} \frac{(-1)^{k+1}}{k} \left(\frac{b}{a+n^{1/3}\omega}\right)^k\right| \leq M n^{2/3-2\delta}/c^2$ for some M. The third term is

$$n^{4/9}\sigma x \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \left(\frac{b}{a+n^{1/3}\omega}\right)^k \sim_\delta 0,$$

because the full sum $\left|n^{4/9}\sigma x \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \left(\frac{b}{a+n^{1/3}\omega}\right)^k\right| \leq \frac{M n^{4/9-\delta}}{c}$ for some M. Now we have shown

$$-n \log\left(1 + \frac{a}{n^{1/3}\omega}\right) \sim_\delta -n^{2/3}\frac{a}{\omega} + n^{1/3}\frac{a^2}{2\omega^2}, \tag{2.27}$$

$$m \log\left(1 + \frac{b}{a+n^{1/3}\omega}\right) \sim_\delta$$
$$n\left(\frac{a}{b}\right)\left(\frac{b}{a+n^{1/3}\omega}\right) - n\left(\frac{a}{2b}\right)\left(\frac{b}{a+n^{1/3}\omega}\right)^2 + dn^{2/3}\left(\frac{b}{a+n^{1/3}\omega}\right). \tag{2.28}$$

Adding (2.27) and (2.28) together yields

$$h_n(\omega) \sim_\delta -n^{2/3}\frac{a}{\omega} + n^{1/3}\frac{a^2}{2\omega^2} + n\left(\frac{a}{b}\right)\left(\frac{b}{a+n^{1/3}\omega}\right) - n\left(\frac{a}{2b}\right)\left(\frac{b}{a+n^{1/3}\omega}\right)^2 + dn^{2/3}\left(\frac{b}{a+n^{1/3}\omega}\right). \quad (2.29)$$

Adding the first and third terms from (2.29) gives the following cancellation.

$$-n^{2/3}\frac{a}{\omega} + n\left(\frac{a}{b}\right)\left(\frac{b}{a+n^{1/3}\omega}\right) = -n^{2/3}\frac{a}{\omega} + n^{2/3}\frac{a}{\omega}\left[1 - \frac{a}{n^{1/3}\omega} + \sum_{k=2}^{\infty}(-1)^k\left(\frac{a}{n^{1/3}\omega}\right)^k\right] \sim_\delta -n^{1/3}\frac{a^2}{\omega^2},$$

thus

$$h_n(\omega) \sim_\delta -n^{1/3}\left(\frac{a^2}{2\omega^2}\right) - n\left(\frac{a}{2b}\right)\left(\frac{b}{a+n^{1/3}\omega}\right)^2 + dn^{2/3}\left(\frac{b}{a+n^{1/3}\omega}\right).$$

When we expand $\frac{b}{a+n^{1/3}\omega} = \frac{b}{n^{1/3}\omega} + \left(\frac{b}{n^{1/3}\omega}\right)\sum_{k=1}^{\infty}\left(\frac{-a}{n^{1/3}\omega}\right)^k$, we see that because $n^{1/3}\omega \sim_\delta n^\delta \mathbf{i}c(a+b)$, the sum is of order $1/c$ times the first term. So we can take only the first terms in our expansion, just as when we Taylor expand. This approximation leads the $n^{2/3}$ terms to cancel giving

$$h_n(\omega) \sim_\delta -n^{1/3}\left(\frac{a^2+ab}{2\omega^2}\right) + dn^{1/3}\left(\frac{b}{\omega}\right) \sim_\delta n^{1/3}\left(f_1(\omega) - t\omega\right).$$

This implies that $\mathfrak{Re}[n^{1/3}t\omega + h_n(\omega)] \sim_\delta \mathfrak{Re}[n^{1/3}f_1(\omega)]$. Completing the first δ-equivalence in the statement of Lemma 2.3.3.

Now observe that in

$$\mathfrak{Re}[n^{1/3}f_1(\omega)] = \mathfrak{Re}\left[n^{1/3}\left(t\omega - \frac{a(a+b)}{2\omega^2} + \frac{bd}{\omega}\right)\right],$$

we can bound the first term $|\mathfrak{Re}[n^{1/3}t\omega]| \leq n^\delta M$. We can bound the third term by $\mathfrak{Re}\left[n^{1/3}\frac{bd}{\omega}\right] \leq$

$M\frac{n^{2/3-\delta}}{c}$. For the second term, we have $\left|\frac{a(a+b)}{2\omega^2}\right| \sim_\delta \left(\frac{a(a+b)}{2}\right)\left(\frac{n^{1-2\delta}}{c}\right)$. Thus

$$\Re\mathfrak{e}[n^{1/3}f_1(\omega)] \sim_\delta \left(\frac{a(a+b)}{2}\right)\left(\frac{n^{1-2\delta}}{c}\right).$$

This gives the second δ-equivalence in the statement of Lemma 2.3.3, and completes the proof. □

2.3.3 Construction of the contour C_n

To construct the contour C_n we will start with lines departing from λ at angles $e^{\pm 2\pi\mathbf{i}/3}$, and with a vertical line $-n^{1/3}a + \mathbf{i}\mathbb{R}$. We will cut both these infinite contours off at specific values q and p respectively which allow us to use our estimates from the previous section on these contours. We will then connect these contours using the level set $\{z : Re[-f_1(z)] = -f_1(\lambda) - \varepsilon\}$. The rest of this section is devoted to finding the values p and q, showing that our explanation above actually produces a contour, and controlling the derivative of f_1 on the vertical segment near 0.

We note

$$f_1(\lambda) = 3t^{2/3}\left(\frac{a(a+b)}{2}\right)^{1/3} > 0, \tag{2.30}$$

and let

$$p = \sqrt{\frac{1}{3}\left(\frac{a(a+b)}{2t}\right)^{2/3}} > 0. \tag{2.31}$$

By simple algebra, we see that $\Re\mathfrak{e}[-f_1(\pm\mathbf{i}y)] < \Re\mathfrak{e}[-f_1(\lambda)] < 0$, when $y < p$, with equality at $y = p$.

Lemma 2.3.4. $\frac{d}{dy}\Re\mathfrak{e}[-f_1(n^{-1/3}a + \mathbf{i}y)]$ *is positive for* $y \in [n^{-1/3}|a+b|, p]$*, and negative for* $y \in [-n^{-1/3}|a+b|, -p]$.

Proof. We compute

$$\frac{d}{dy}\Re\mathfrak{e}[f_1(n^{-1/3}a + \mathbf{i}y)] = -\,\mathfrak{Im}(\Re\mathfrak{e}[f_1(n^{-1/3}a + \mathbf{i}y)]) \tag{2.32}$$

$$= -\frac{y^3 a(a+b)}{|n^{-1/3}a + \mathbf{i}y|^6} + \frac{a^2(a+b)n^{-2/3}y}{|n^{-1/3}a + \mathbf{i}y|^6} + \frac{3a^2(a+b)bn^{-1/3}y}{2b\lambda|n^{-1/3}a + \mathbf{i}y|^4}. \tag{2.33}$$

Note that for $y \in [n^{-1/3}|a+b|, p] \cup [-n^{-1/3}|a+b|, -p]$, we have $|n^{-1/3}a + \mathbf{i}y| \sim |y|$, so the first term of (2.33) is of order y^{-3} and the third term of (2.33) is of order $y^{-3}n^{-1/3}$. So for large enough n, the third term of (2.33) is very small compared to the first term. For $y = \pm n^{-1/3}|a+b|$, we have $|n^{-1}a(a+b)^4| = |y^3 a(a+b)| > |a(a+b)n^{-2/3}ay| = |a^2(a+b)^2 n^{-1/3}|$, and the derivative of $y^3 a(a+b)$ is larger than the derivative of $a(a+b)n^{-2/3}ay$ for $y \in [n^{-1/3}|a+b|, p] \cup [-n^{-1/3}|a+b|, -p]$, so the first term of (2.33) has larger norm than the second term for $y \in [n^{-1/3}|a+b|, p] \cup [-n^{-1/3}|a+b|, -p]$. Thus the sign $\frac{d}{dy}\Re\mathfrak{e}[-f_1(n^{-1/3}a + \mathbf{i}y)]$ is determined by the first term of (2.33) in these intervals. □

Now we can define the contour C_n. We will give the definition, and then justify that it gives a well defined contour.

Definition 2.3.5. Let $q > 0$ be a fixed real number such that for $0 < y \leq q$, $\frac{d}{dy}\Re\mathfrak{e}[-f_1(\lambda \pm ye^{\pm 2\pi\mathbf{i}/3})] < 0$. Let

$$s = \max\Big\{\Re\mathfrak{e}[-f_1(\lambda + qe^{-2\pi\mathbf{i}/3})], \Re\mathfrak{e}[-f_1(\lambda + qe^{2\pi\mathbf{i}/3})], \\ \Re\mathfrak{e}[-f_1(n^{-1/3}(a - \mathbf{i}|a+b|))], \Re\mathfrak{e}[-f_1(n^{-1/3}(a + \mathbf{i}|a+b|))]\Big\}. \quad (2.34)$$

Let α be the contourline $\alpha = \{\omega : \Re\mathfrak{e}[-f_1(\omega)] = s\}$, and define the set

$$S_n = \{\lambda + ye^{\pm 2\pi\mathbf{i}/3} : 0 \leq y \leq q\} \cup \alpha \cup [-an^{-1/3} - \mathbf{i}p, -an^{-1/3} + \mathbf{i}p].$$

For sufficiently large n, define the path C_n to begin where α intersects $\{\lambda + ye^{-2\pi\mathbf{i}/3} : 0 \leq y \leq q\}$, follow the path $\{\lambda + ye^{-2\pi\mathbf{i}/3} : 0 \leq y \leq q\}$ toward $y = 0$, then follow the path $\{\lambda + ye^{2\pi\mathbf{i}/3} : 0 \leq y \leq q\}$ until it intersects α. C_n then follows α in either direction (pick one arbitrarily) until it intersects $[-an^{-1/3} - \mathbf{i}p, -an^{-1/3} + \mathbf{i}p]$ in the upper half plane. C_n then follows the path $[-an^{-1/3} - \mathbf{i}p, -an^{-1/3} + \mathbf{i}p]$ toward $-an^{-1/3} - \mathbf{i}p$ until it intersects α in the negative half plane. Then C_n follows α in either direction (pick one arbitrarily) until it reaches its starting point where it intersects $\{\lambda + ye^{-2\pi\mathbf{i}/3} : 0 \leq y \leq q\}$. See Figure 2.6

We see that the q in Definition 2.3.5 exists by applying Taylor's theorem along with the fact that

$f_1'''(\lambda) > 0$, and the $f_1'(\lambda) = f_1''(\lambda) = 0$.

Lemma 2.3.6. *The sets $\{\lambda + ye^{2\pi\mathbf{i}/3} : 0 \le y \le q\}$ and $\{\lambda + ye^{-2\pi\mathbf{i}/3} : 0 \le y \le q\}$ both intersect α at exactly one point. Lemma 2.3.7 and Lemma 2.3.6 will show that C_n is a well defined contour.*

This follows from the definition of q and s.

Lemma 2.3.7. *There exists $N > 0$ such that for all $n > N$, the sets $[n^{-1/3} + \mathbf{i}n^{-1/3}|a+b|, n^{-1/3}a + p]$ and $[-an^{-1/3} - n^{-1/3}|a + b|, -an^{-1/3} - p]$ both intersect α exactly once.*

Proof. This is true because

$$\Re\mathfrak{e}[-f_1(-n^{-1/3}(a \pm \mathbf{i}|a + b|))] < \Re\mathfrak{e}[-f_1(\lambda)]. \tag{2.35}$$

by the contour lines in Figure 2.4. This in addition to Lemma 2.3.4, and (2.30) implies the lemma.

□

2.3.4 Properties of the contour C_n: proof of Lemma 2.2.8

Most of the work is used to prove part (c). The idea of this proof is to patch together the different estimates from the beginning of Section 2.3. Away from 0 we use Lemma 2.2.7 and the fact that the contour is steep descent near λ. Very near 0 on the scale $n^{-1/3}$ we use Lemma 2.3.1. Moderately near 0 we use Lemma 2.3.3, and our control of the derivative of f_1 on the vertical strip of C_n near 0. This last argument allows us to get bounds uniform in $\delta \in (0, 1/3)$ when ω is on the scale $n^{1/3-\delta}$.

Proof of Lemma 2.2.8. (a) and (b) follow from the definition of C_n. By a slight modification of the proof of Lemma 2.8, we see that for $z \in \gamma_r$,

$$\Re\mathfrak{e}[h_n(z) - h_n(\lambda) + n^{1/3}t(z - \lambda) \le n^{1/9}C, \tag{2.36}$$

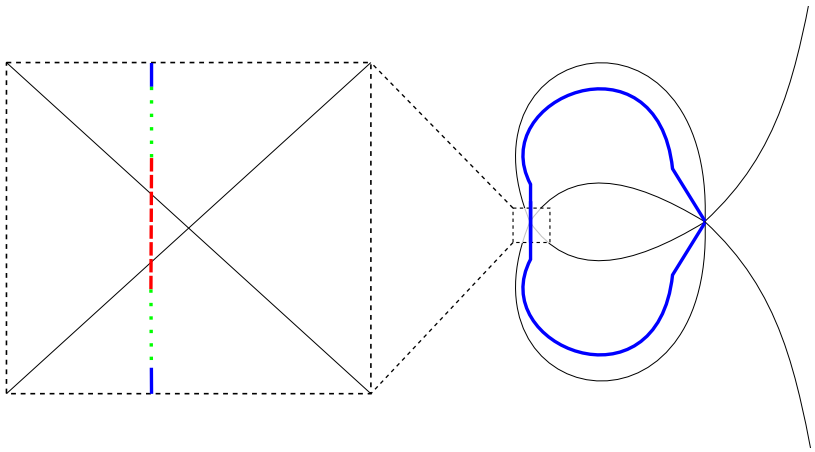

Figure 2.6: C_n is the thick, colored piecewise smooth curve, the contour lines $\{z : \mathfrak{Re}[-f_1(z)] = f_1(\lambda)\}$ are the thin black curves. On the right side of the image we see C_n as a thick blue curve sandwiched between the contour lines. On the left we zoom in near 0 and see C_n pass the real axis as a dotted line to the left of zero. The contour lines meet at the point 0 on the left and λ on the right. We will now describe what section of the proof of Theorem 2.2.8 bounds $h_n(z)-h_n(\omega)+nt^{1/3}(z-\omega)$ on different portions of C_n. The diagonal segments of C_n near λ are bounded in (ii). The curved segments in the right image, and the solid dark blue vertical segments at the top and bottom of the left image are bounded in (i). The dark red dashed segment that crosses the real axis in the left image is distance $O(n^{-1/3})$ from 0 and is bounded in (iii). The green dotted segments in the left image are distance $O(n^{\delta-1/3})$ from 0 for $\delta \in (0,1)$ and are bounded in (iv).

so to show (c) it suffices to show that for $\omega \in C_n \setminus C_n^{\varepsilon}$, we have

$$\Re\mathfrak{e}[h_n(\lambda) - h_n(\omega) + n^{1/3}t(\lambda - \omega)] \leq -n^{-1/3}\eta. \tag{2.37}$$

Below we split the contour into 4 pieces and bound each separately. See Figure 2.6.

(i) By Lemma 2.3.4 and the construction of C_n, we have $\Re\mathfrak{e}[-f_1(\omega)] \leq s < \Re\mathfrak{e}[-f_1(\lambda)]$ for $\omega \in C_n \setminus (\{\lambda + ye^{\pm 2\pi\mathbf{i}/3} : 0 \leq y \leq q\} \cup [n^{-1/3}(-a - \mathbf{i}|a+b|), n^{-1/3}(-a + \mathbf{i}|a+b|)])$. So we can apply Lemma 2.2.7 and the fact that f_2 is bounded outside a neighborhood of 0 to show that for any $c_1 < 0$, we have $\Re\mathfrak{e}[h_n(z) - h_n(\lambda) + n^{1/3}t(z-\lambda)] \leq -n^{-1/3}\eta$ for $\omega \in C_n \setminus (\{\lambda + ye^{\pm 2\pi\mathbf{i}/3} : 0 \leq y \leq q\} \cup [-n^{-1/3}a - \mathbf{i}c_1|a+b|, -n^{-1/3}a + \mathbf{i}c_1|a+b|])$.

(ii) By the definition of q, The contour $\{\lambda + ye^{\pm 2\pi\mathbf{i}/3} : 0 \leq y \leq q\}$ is steep descent with respect to the function f_1 at the point λ, so we can apply Lemma 2.2.7 and the fact that f_2 is bounded outside a neighborhood of 0 to show $\Re\mathfrak{e}[h_n(z) - h_n(\lambda) + n^{1/3}t(z-\lambda)] \leq -n^{-1/3}\eta$ for $\omega \in \{\lambda + ye^{\pm 2\pi\mathbf{i}/3} : 0 \leq y \leq q\} \setminus B_{\varepsilon}(\lambda)$.

(iii) By Lemma 2.3.1, for any c_0, we have $\Re\mathfrak{e}[h_n(z) - h_n(\lambda) + n^{1/3}t(z-\lambda)] \leq -n^{-1/3}\eta$ for all $\omega \in [n^{-1/3}(-a - \mathbf{i}c_0|a+b|), n^{-1/3}(-a - \mathbf{i}c_0|a+b|)]$.

(iv) Now we bound the $\Re\mathfrak{e}[h_n(z) - h_n(\lambda) + n^{1/3}t(z-\lambda)]$ on the last piece of our contour $[n^{-1/3}(-a - \mathbf{i}c_0|a+b|), -n^{-1/3}a + \mathbf{i}c_1|a+b|] \cup [-n^{-1/3}a - \mathbf{i}c_1|a+b|, n^{-1/3}(-a - \mathbf{i}c_0|a+b|)]$. We will do this by fixing a constant $c > c_1$, and bounding the function on $\omega = n^{-1/3}a + \mathbf{i}n^{\delta - 1/3}c(a+b)$ for all pairs $n > N, \delta \in (0, 1/3)$ such that $n^{1/3} \leq c_1/c$.

By Lemma 2.3.3, we have that when $\omega = n^{-1/3}a + \mathbf{i}n^{\delta-1/3}c(a+b)$, there exist constants M_1, M_2, M_3, such that

$$\Re\mathfrak{e}[n^{1/3}t\omega + h_n(\omega) - n^{1/3}f_1(\omega)] \leq \frac{n^{2/3-2\delta}}{c^2}M_1 + \frac{n^{1-3\delta}}{c^3}M_2 + \frac{n^{4/9-\delta}}{c}M_3,$$

and

$$f_1(\omega) \sim_\delta M\frac{n^{1-2\delta}}{c^2}.$$

First we consider the case when $\delta \in (0, 1/3 - \varepsilon)$. In this case, for any $r > 0$ we can choose c and N_r large enough that for all $n > N_r$,

$$\frac{\frac{n^{2/3-2\delta}}{c^2}M_1 + \frac{n^{1-3\delta}}{c^3}M_2 + \frac{n^{4/9-\delta}}{c}M_3}{\Re\mathrm{e}[n^{1/3}f_1(\omega)]} < r/2,$$

uniformly for all $\delta \in (0, 1/3 - \varepsilon)$. In this case we also have that, by Lemma 2.2.7,

$$|\Re\mathrm{e}[n^{1/3}tz + h_n(z)]| \leq n^{1/3}f_1(\lambda) + n^{1/9}f_2(\lambda) + C.$$

By potentially increasing N_r, we have that for all $n > N_r$

$$\frac{|\Re\mathrm{e}[n^{1/3}tz + h_n(z)]|}{\Re\mathrm{e}[n^{1/3}f_1(\omega)]} \leq r/2.$$

By Lemma 2.3.4 and (2.35), for all pairs n, δ such that $n^{\delta-1/3} < c/c_1$, there is an $\eta > 0$ such that

$$\Re\mathrm{e}[-f_1(\omega)] \leq \Re\mathrm{e}[-f_1(\lambda)] - 2\eta < -2\eta.$$

setting $r = 1/2$ gives

$$\Re\mathrm{e}[n^{1/3}t(z-\omega) + h_n(z) - h_n(\omega)] \leq \Re\mathrm{e}[-n^{1/3}f_1(\omega)] + \frac{1}{2}\Re\mathrm{e}[n^{-1/3}f_1(\omega)] < -\eta n^{1/3}.$$

Now we prove the case $\delta \in (1/3 - \varepsilon, 1/3)$. Note that in the expression

$$\Re\mathrm{e}[n^{1/3}t\omega + h_n(\omega) - n^{1/3}f_1(\omega)] \leq \frac{n^{2/3-2\delta}}{c^2}M_1 + \frac{n^{1-3\delta}}{c^3}M_2 + \frac{n^{4/9-\delta}}{c}M_3,$$

when n is sufficiently large, we can bound the right hand side by $(M_1 + M_2)n^{3\varepsilon} \leq (r/2)n^{1/3}$

for any $r > 0$. We also have

$$|\Re\mathrm{e}[n^{1/3}t\lambda - h_n(\lambda) - n^{1/3}f_1(\lambda)]| \leq n^{1/9}f_1(\lambda) + C \leq (r/2)n^{1/3}.$$

The first inequality comes from Lemma 2.2.7, and the second holds for large enough n. By Lemma 2.3.4 and (2.35), for all pairs n, δ such that $n^{\delta-1/3} < c/c_1$, there is an $\eta > 0$ such that

$$\Re\mathrm{e}[-f_1(\omega)] \leq \Re\mathrm{e}[-f_1(\lambda)] - 2\eta < -2\eta.$$

Setting $r = \eta$ gives

$$\Re\mathrm{e}[n^{1/3}t(\lambda - \omega) + h_n(\lambda) - h_n(\omega)] \leq n^{1/3}\Re\mathrm{e}[f_1(\lambda) - f_1(\omega)] + n^{1/3}\eta \leq -\eta n^{1/3}.$$

The c_1 in part (i) can be chosen as small as desired, the c in part (iv) has already been chosen, and the c_0 in part (iv) can be chosen as large as desired. Choose $c_1 < c < c_0$ to complete the proof of (c).

Given inequalities (2.36) and (2.37), part (d) follows if we can show

$$\Re\mathrm{e}[n^{1/3}t(\lambda - \omega) + h_n(\lambda) - h_n(\omega)],$$

for $\omega \in C_n^{\varepsilon}$. Indeed this follows from Lemma 2.2.7 and the fact that the contour $\{\lambda + ye^{\pm 2\pi\mathbf{i}/3} : 0 \leq y \leq q\}$ is steep descent with respect to the function $\Re\mathrm{e}[-f_1]$ at the point λ.

□

2.4 Dominated convergence

In this section we carefully prove that the series expansion for $\det(1 - \mathsf{K}_n)_{L^2(C_n^{\varepsilon})}$ gives an absolutely convergent series of integrals bounded uniformly in n. This allows us to use dominated convergence when we localize the integral in Proposition 2.2.12, and again when we approximate

the kernel by its Taylor expansion in Proposition 2.2.13. First we zoom in on a ball of radius epsilon and show that we can absolutely bound $\det(1-\mathsf{K}_n^{\varepsilon})_{L^2(C_n^{\varepsilon})}$ uniformly in n.

Lemma 2.4.1. *For any sufficiently small $\varepsilon > 0$, and sufficiently large r, there exists a function $\overline{F}(\overline{\omega},\overline{\omega}')$, such that for all $\overline{\omega},\overline{\omega}' \in C_{-1}^{n^{1/9}\varepsilon}$, $z \in \mathcal{D}_0^{n^{1/9}\varepsilon}$, $n > N$ the integrand of $\overline{K}_n^{\varepsilon}(\overline{\omega},\overline{\omega}')$ in equation (2.11) is absolutely bounded by $\overline{F}(\overline{\omega},\overline{\omega}',\overline{z})$, and*

$$\sum_{m=0}^{\infty}\int_{(C_{-1}^{n^{1/9}\varepsilon})^m}\left|\det\left(\int_{\mathcal{D}_0^{n^{1/9}\varepsilon}}\overline{F}(\overline{\omega}_i,\overline{\omega}_j,\overline{z})d\overline{z}\right)_{i,j=1}^{m}\right|d\overline{\omega}_1\ldots d\overline{\omega}_m<\infty. \tag{2.38}$$

Proof. For $\overline{\omega},\overline{\omega}' \in C_{-1}^{\varepsilon}$, and $\overline{z} \in \mathcal{D}_0^{\varepsilon}$, we have

$$\left|\frac{\lambda+n^{-1/9}\overline{z}}{\lambda+n^{-1/9}\overline{\omega}}\right| \le \left|\frac{\lambda+\varepsilon}{\lambda-\varepsilon}\right|,$$

and by Taylor approximation, we have the additional bounds

$$n^{1/3}(f_1(\lambda+n^{-1/9}\overline{z})-f_1(\lambda+n^{-1/9}\overline{\omega})) \le (f_1'''(\lambda)+\delta_1)(\overline{z}^3-\overline{\omega}^3), \tag{2.39}$$

$$n^{1/9}(f_2(\lambda+n^{-1/9}\overline{z})-f_2(\lambda+n^{-1/9}(\overline{\omega}))) \le (f_2'(\lambda)+\delta_2)(\overline{z}-\overline{\omega}), \tag{2.40}$$

$$r_n(\lambda+n^{-1/9}\overline{z})-r_n(\lambda+n^{-1/9}\overline{\omega}) \le Cn^{-1/9}(\overline{z}-\overline{\omega}) \le C\varepsilon \le \delta_3. \tag{2.41}$$

Note that in these bounds we can make $\delta_1,\delta_2,\delta_3$ as small as desired by choosing ε small. Equations (2.39) and (2.40) follow from the fact that f_1, and f_2 are holomorphic in the compact set $\overline{B}_{\varepsilon}(\lambda)$. And equation (2.41) follows from Lemma 2.2.7. Note that along $\mathcal{D}_0$, z is purely imaginary, so (2.39),(2.40), and (2.41) show that the full exponential in the integrand in (2.11) is bounded above by

$$e^{2\delta_3}e^{-(f_1'''(\lambda)-\delta_1)\overline{\omega}^3-(f_2'(\lambda)-\delta_2)\overline{\omega}}. \tag{2.42}$$

We choose ε small enough that $\delta_1 < f_1'''(\lambda)$, so that (2.42) has exponential decay as ω goes to ∞

in directions $e^{\pm 2\pi \mathbf{i}/3}$. Set

$$\overline{F}(\overline{\omega}, \overline{\omega}', \overline{z}) = \left| \left(\frac{\lambda + \varepsilon}{\lambda - \varepsilon} \right) e^{2\delta_3} e^{-(f_1'''(\lambda) - \delta_1)\overline{\omega}^3 - (f_2'(\lambda) - \delta_2)} \frac{1}{(\overline{z}+1)(\overline{z}+1)} \right|.$$

By the sentence preceeding (2.42) $\overline{F}$ absolutely bounds the integrand of $\overline{K}_n^{\varepsilon}$. Now set $L_1 = \frac{|\lambda+\varepsilon|}{|\lambda-\varepsilon|} e^{2\delta_3} \int_{\mathcal{D}_0} \frac{1}{(\overline{z}+1)(\overline{z}+1)} d\overline{z}$ so that $2e^{2\delta_3} \int_{\mathcal{D}_0} \frac{1}{(\overline{z}-\overline{\omega})(\overline{z}-\overline{\omega}')} d\overline{z} \leq L_1$. Then

$$\int_{\mathcal{D}_0^{\varepsilon}} \overline{F}(\overline{\omega}, \overline{\omega}', \overline{z}) \leq L_1 \left| e^{-(f_1'''(\lambda) - \delta_1)\overline{\omega}^3 - (f_2'(\lambda) - \delta_2)} \right|, \tag{2.43}$$

By Hadamard's bound

$$\left| \det \left(\int_{\mathcal{D}_0^{n^{1/9}\varepsilon}} \overline{F}(\overline{\omega}_i, \overline{\omega}_j', \overline{z}) d\overline{z} \right)_{i,j=1}^{m} \right| \leq m^{m/2} L_1^m \prod_{i=1}^{m} \left| e^{-(f_1'''(\lambda) - \delta)\overline{\omega}^3 - (f_2'(\lambda) - \delta)\overline{\omega}} \right|.$$

Now because $\delta_1 < f_1'''(\lambda)$, we can set

$$S = \int_{C_{-1}^{n^{1/9}\varepsilon}} \left| e^{-(f_1'''(\lambda) - \delta)\overline{\omega}^3 - (f_2'(\lambda) - \delta)\overline{\omega}} \right| d\overline{\omega} < \infty.$$

Then we have the bound,

$$\int_{(C_{-1}^{n^{1/9}\varepsilon})^m} \left| \det \left(\int_{\mathcal{D}_0^{n^{1/9}\varepsilon}} \overline{F}(\overline{\omega}_i, \overline{\omega}_j', \overline{z}) d\overline{z} \right)_{i,j=1}^{m} \right| d\overline{\omega}_1 \ldots d\overline{\omega}_m \leq m^{m/2} (SL_1)^m.$$

So by Stirling's approximation

$$\sum_{m=0}^{\infty} \int_{(C_{-1}^{n^{1/9}\varepsilon})^m} \left| \det \left(\int_{\mathcal{D}_0^{n^{1/9}\varepsilon}} \overline{F}(\overline{\omega}_i, \overline{\omega}_j, \overline{z}) d\overline{z} \right)_{i,j=1}^{m} \right| d\overline{\omega}_1 \ldots d\overline{\omega}_m < \infty.$$

□

The next lemma completes our dominated convergence argument, by controlling the contribution to $\det(I - K_n)_{L^2(C_n^{\varepsilon})}$ of $z \in \gamma_r \setminus \gamma_r^{\varepsilon}$.

Lemma 2.4.2. *For any sufficiently small $\varepsilon > 0$, and sufficiently large r, there is a function*

$\overline{G}(\overline{\omega},\overline{\omega}',\overline{z})$, *and a natural number* N, *such that for all* $\overline{\omega},\overline{\omega}' \in \overline{C}_n^{\varepsilon}$ *and* $\overline{z} \in \overline{\gamma}_r$, $n > N$, *the integrand of* $\overline{K}_n(\overline{\omega},\overline{\omega}')$ *is absolutely bounded by* $\overline{G}(\overline{\omega},\overline{\omega}',\overline{z})$, *and*

$$\sum_{m=0}^{\infty} \frac{1}{m!} \int_{(\overline{C}^{\varepsilon})^m} \left| \det\left(\int_{\overline{\gamma}_r} \overline{G}(\overline{\omega}_i,\overline{\omega}_j,\overline{z}) dz \right)_{i,j=1}^{m} \right| d\overline{\omega}_i \ldots d\overline{\omega}_j < \infty, \tag{2.44}$$

where $\overline{\gamma}_r$ *and* $\overline{C}_n^{\varepsilon}$ *are the rescaled contours of* γ_r *and* C_n^{ε} *respectively.*

Proof. Let $\overline{G} = \overline{F}$ for $z \in \gamma_r^{\varepsilon}$. We decompose the integral along γ_r in three parts: the integral along γ_r^{ε}, the integral along $(e^{-2\pi \mathbf{i}/3}\infty, -r) \cup (r, e^{2\pi \mathbf{i}/3}\infty)$ and the integral along $[-r,-\varepsilon] \cup [\varepsilon, r]$. For $z \in \gamma_r \setminus \gamma_r^{\varepsilon}$ we have the following bounds

$$\begin{aligned} |e^{n^{1/3}t(z-\omega)+h_n(z)-h_n(\omega)}| &\leq |e^{n^{1/3}(f_1(z)-f_1(\omega))+n^{1/9}C_2+C_3}| \\ &\leq |e^{n^{1/3}(f_1(z)-f_1(\omega)+\delta)}| \\ &\leq |e^{n^{1/3}(f_1(z)-f_1(\lambda)+\delta)}||e^{n^{1/3}(f_1(\lambda)-f_1(\omega))}|. \end{aligned} \tag{2.45}$$

Where the first inequality follows from Lemma 2.2.7. If we choose $\delta < \eta/2$, and recall that if $z \in \gamma_r \setminus \gamma_r^{\varepsilon}$, then $f_1(z) - f_1(\lambda) < -\eta$, so $f_1(z) - f_1(\lambda) + \delta < -\eta/2 < 0$. So if we wish we can bound (2.45) by either of the following expressions

$$|e^{n^{1/3}(f_1(\lambda)-f_1(\omega))}| \tag{2.46}$$

$$|e^{n^{1/9}(-tz+t\lambda)}||e^{n^{1/3}(f_1(\lambda)-f_1(\omega))}| \tag{2.47}$$

The bound (2.47) follows from the fact that we can choose r large enough so that $|f_1(z)+tz| \leq \delta$ outside $B_r(0)$. Then because the exponent in the first factor of (2.45) is negative, for large enough n we can remove the constant δ in return for reducing $n^{1/3}$ to $n^{1/9}$.

Now for $z \in [-r,-\varepsilon] \cup [\varepsilon, r]$, we have

$$\left|\frac{z}{\omega}\right| \leq \left|\frac{r+\lambda}{\lambda-\varepsilon}\right|, \qquad \left|\frac{1}{(\overline{z}-\overline{\omega})(\overline{z}-\overline{\omega}')}\right| \leq 1.$$

So for $z \in [-r,-\varepsilon] \cup [\varepsilon, r]$, we set

$$\overline{G}(\overline{\omega},\overline{\omega}',\overline{z}) = \left|\frac{r+\lambda}{\lambda-\varepsilon}\right| \left|\frac{1}{(\overline{z}-\overline{\omega})(\overline{z}-\overline{\omega}')}\right| \left|e^{n^{1/3}(f_1(\lambda)-f_1(\omega))}\right|.$$

Using the above bounds and (2.46) we see that the integrand of $\overline{K}_n$ is absolutely bounded by $\overline{G}$ in this region. Set $L_2 = \int_{i\mathbb{R}} \frac{r+\lambda}{\lambda-\varepsilon} \frac{1}{(\overline{z}+1)(\overline{z}+1)} d\overline{z}$ so that the integral of $\overline{G}$ on the rescaled contour of $[-r,-\varepsilon] \cup [\varepsilon, r]$ is bounded by $L_2|e^{n^{1/3}(f_1(\lambda)-f_1(\omega))}|$.

For $z \in (e^{-2\pi\mathbf{i}/3}\infty, -r) \cup (r, e^{2\pi\mathbf{i}/3}\infty)$, we have

$$\left|\frac{1}{(\overline{z}-\overline{\omega})(\overline{z}-\overline{\omega}')}\right| \leq 1.$$

So for $z \in (e^{-2\pi\mathbf{i}/3}\infty, -r) \cup (r, e^{2\pi\mathbf{i}/3}\infty)$, we set

$$\overline{G}(\overline{\omega},\overline{\omega}',\overline{z}) = \left|\frac{z}{\omega}\right| \left|e^{t(\lambda-\overline{z})}\right| \left|e^{(-f_1'''(\lambda)+\delta)\overline{\omega}}\right|.$$

Thus by (2.47), we can see that the integrand of $\overline{K_n}$ is absolutely bounded by $\overline{G}$ in this region. Now let $L_3 = \int_{(e^{-2\pi\mathbf{i}/3}\infty, -r] \cup [r, e^{2\pi\mathbf{i}/3}\infty)} \left|\frac{\lambda+\overline{z}}{\lambda-\varepsilon}\right| |e^{t(\lambda-\overline{z})}| d\overline{z}$. For all n, the integral of $\overline{G}$ over the rescaled contour $(e^{-2\pi\mathbf{i}/3}\infty, -r] \cup [r, e^{2\pi\mathbf{i}/3}\infty)$ is bounded above by $L_3|e^{(-f_1'''(\lambda)+\delta)\overline{\omega}^3}|$.

Let $\overline{\gamma}_r$ be the rescaled contour γ_r in the variable $\overline{z}$

$$\int_{\overline{\gamma}_r} \overline{G} d\overline{z} \leq (L_1 + L_2 + L_3) e^{(-f_1'''(\lambda)+\delta)\overline{\omega}^3} \leq L e^{(-f_1'''(\lambda)+\delta)\overline{\omega}^3}, \tag{2.48}$$

where the constant L comes from (2.43). Thus we have bounded $\int_{\overline{\gamma}_r} \overline{G} d\overline{z}$ by a constant times a term which has exponential decay as $\overline{\omega} \to e^{\pm 2\pi\mathbf{i}/3}\infty$. The same argument as in Lemma 2.4.1 shows that

$$\sum_{m=0}^{\infty} \frac{1}{m!} \int_{(C^{\varepsilon})^m} \left| \det \left(\int_{\gamma_r^{\varepsilon}} G(\omega_i, \omega_j, z) dz \right)_{i,j=1}^{m} \right| d\omega_i ... d\omega_j < \infty.$$

□

Lemma 2.4.3. *Let $\omega_1 \in C_n \setminus C_n^{\varepsilon}$ and $\omega_2, .., \omega_m \in C^n$. There exist positive constants $M, L_4, \eta > 0$ so that for sufficiently large n, we have*

$$|\overline{K}_n(\overline{\omega}_i, \overline{\omega}_j)| \leq M$$

and

$$|\overline{K}_n(\overline{\omega}_1, \overline{\omega}_i)| \leq L_4 n^{4/9} e^{-n^{1/3}\eta},$$

for all i, j.

Proof. By Lemma 2.2.8, for any $\varepsilon > 0$, there exists a $N, C > 0$, such that if $v \in C_n \setminus C_n^{\varepsilon}$, and $z \in \gamma_r$, then for all sufficiently large n, we have

$$\Re\mathfrak{e}[h_n(z) - h_n(\omega) + n^{1/3} t(z - \omega)] \leq -n^{1/3}\eta.$$

For $z \in \gamma_r$ and $\omega, \omega' \in C_n \setminus C_n^{\varepsilon}$, $n > N$ we have the following bounds:

$$\frac{1}{(z-\omega)(z-\omega')} \leq \left(\frac{2}{\varepsilon}\right)^2, \qquad \frac{1}{\omega} \leq \frac{n^{1/3}}{a},$$

and

$$|e^{n^{1/3}t(z-\omega)+h_n(z)-h_n(\omega)}| \leq |e^{n^{1/3}(f_1(z)-f_1(\omega)+\delta)}| \tag{2.49}$$

$$\leq |e^{n^{1/3}(f_1(z)-f_1(\lambda)}||e^{n^{1/3}(f_1(\lambda)-f_1(\omega)+\delta)}| \tag{2.50}$$

where (2.49) follows from (2.2.7) and the fact that f_2 is bounded away from 0. Note that for $z \in \gamma_r$,

$|f_1(z) - f_1(\lambda)| \leq 0$, and for $\omega, \omega' \in C_n \setminus C_n^{\varepsilon}$, $f_1(\lambda) - f_1(\omega) + \delta < -\eta$, so (2.50) is bounded above by

$$|e^{(f_1(z)-f_1(\lambda)}||e^{-n^{1/3}\eta}|.$$

Thus if we set $L_4 = \frac{2^2}{a\varepsilon^2}\int_{\gamma_r} |z||e^{f_1(z)-f_1(\lambda)}|dz < \infty$, we get

$$|K_n(\omega, \omega')| \leq L_4 n^{1/3} e^{-n^{1/3}\eta}.$$

So if we change the variable of integration to $d\overline{z} = n^{1/9}dz$ gives.

$$|\overline{K}_n(\overline{\omega}, \overline{\omega}')| \leq L_4 n^{4/9} e^{-n^{1/3}\eta} \qquad \text{for } \omega, \omega' \in C_n \setminus C_n^{\varepsilon} \tag{2.51}$$

Let $\omega_1 \in C_n \setminus C_n^{\varepsilon}$ and $\omega_2, .., \omega_m \in C^n$, then for $i \neq 1$,

$$\begin{aligned} &|\overline{K}_n(\overline{\omega}_1, \overline{\omega}_i)| \leq L_4 n^{4/9} e^{-n^{1/3}\eta}, \\ &|\overline{K}_n(\overline{\omega}_i, \overline{\omega}_j)| \leq \max[Le^{(-f_1'''(\lambda)+\delta)\overline{\omega}^3}, L_4 n^{4/9} e^{-n^{1/3}\eta}] \leq M. \end{aligned} \tag{2.52}$$

The first equality follows from (2.48) and the second inequality holds for large n, when we set $M = \max[L_4, L]$ because $-f_1'''(\lambda) + \delta < 0$. □

The last thing we need to complete the proof of Theorem 2.1.4 is to bound (2.15) from Proposition (2.2.3). We do so in the following lemma.

Lemma 2.4.4. *For any $C > 1$, we have*

$$\sum_{m=1}^{\infty} \frac{1}{m!} C^m m^{1+m/2} \leq 16C^4 e^{2C^2}.$$

Proof. We have

$$\frac{m^{1+m/2}}{m!} \leq \frac{m2^{m/2}}{(\lfloor m/2 \rfloor)!},$$

so that

$$\begin{aligned}\sum_{m=1}^{\infty} \frac{1}{m!} C^m m^{1+m/2} &\leq \sum_{m=1}^{\infty} \frac{m}{(\lfloor m/2 \rfloor)!} (2C^2)^{m/2} \\ &\leq \sum_{k=1}^{\infty} \frac{2k(2C^2)^k}{k!} + \sum_{k=1}^{\infty} \frac{(2k+1)(2C^2)^{k+1}}{k!} \\ &\leq 16C^4 e^{2C^2}.\end{aligned}$$

□

Chapter 3: Large deviations for sticky Brownian motions

This chapter is based on the article [25] written by myself and Guillaume Barraquand.

3.1 Introduction and main results

Families of interacting Brownian motions have been related to random matrix theory in a number of works. For instance at any fixed time nonintersecting Brownian motions have the same distribution as the eigenvalues of a matrix from the Gaussian unitary ensemble (GUE) [74]. Certain statistics of families of Brownian motions with asymmetric reflections also have Tracy-Widom GUE distributed fluctuations [195] as the number of particles goes to $+\infty$. There are many other examples (see for instance [28, 97, 152, 174, 78, 150, 36, 37]), and the ubiquitous occurrence of the GUE can be understood in the framework of the Kardar-Parisi-Zhang (KPZ) universality class. This framework predicts that in spatial dimension 1, many growth models, interacting particle systems and directed polymer models have Tracy-Widom fluctuations in the cube-root time scale, for appropriate initial data. This class is extremely broad and is not yet clearly delineated. In particular one may expect that many families of interacting Brownian motions fall in the KPZ universality class and are related to random matrix theoretic distributions. The examples cited above all deal with families of Brownian motions with repulsive interaction; in this chapter we study a family of Brownian motions with attractive interaction called sticky Brownian motions.

In 1952 Feller introduced a reflected Brownian motion sticky at the origin which evolves as a Brownian motion everywhere except at origin, and has its reflection off the origin slowed down so that the total time its trajectory spends at the origin has positive Lebesgue measure [77]. This motion's law can be characterized by a single stickiness parameter which determines how much time it spends at the origin. More recently, using stochastic flows and Dirichlet forms [133, 135]

or through a martingale problem [104, 105], several authors have defined families of n-particle diffusions where the distance between each pair of particles is a reflected Brownian motion sticky at the origin.

These n-point sticky Brownian motions describe the evolution of mesoscopic particles with attractive interaction at a scale smaller than their radius; this situation is common in the study of colloids [103]. Sticky Brownian motions are the diffusive scaling limit of various models: discrete random walks in random environment [105, 9], certain families of exclusion processes with a tunable interaction [160], and storage processes [100]. Using the language of stochastic flows of kernels, sticky Brownians motion can be described as independent motions in a space-time i.i.d. random environment [135, 132, 168, 169].

In this chapter we restrict our attention to a specific one-parameter family of sticky Brownian motions which we will call uniform sticky Brownian motions where the multiparticle interactions are completely determined by the two particle interactions. Within this restricted class, we prove a quenched large deviation principle (Theorem 3.1.13) for the random heat kernel (referred to below as the uniform Howitt-Warren stochastic flow of kernels). We then prove that the random lower order corrections to the large deviation principle, which come from the random environment, are Tracy-Widom GUE distributed in the large time limit (Theorem 3.1.15). This gives a positive answer, in the case of uniform sticky Brownian motions, to a question posed in [169, Section 8.3 (4)]. Our results can be rephrased to say that as time and the number of particles n are simultaneously sent to infinity, the position of the extremal particle of n uniform sticky Brownian motions has Tracy-Widom GUE distributed fluctuations (Corollary 3.1.17).

We prove these results by viewing uniform sticky Brownian motions as the limit of a discrete exactly solvable model: the beta random walk in random environment (RWRE). Using exact formulas for the latter, we prove a Fredholm determinant formula for the Laplace transform of the random heat kernel associated to sticky Brownian motions. We then perform rigorous saddle point asymptotics to prove the Tracy-Widom GUE limit theorem. We also provide mixed moment formulas for the stochastic flow of kernels, which yield concise formulas for the probability

distribution at time t of the maximum of n-point sticky Brownian motions started from arbitrary particle positions (Proposition 3.1.22). Though we uncover the integrability of the model by degenerating earlier results, this allows us to bring the techniques of integrable probability to bear on sticky Brownian motions and stochastic flows, which occur as the scaling limit of many stochastic processes. On a more technical side the asymptotic analysis of the Fredholm determinant formula for the beta RWRE was challenging and could only be performed for a very specific choice of parameters; we overcome some of these challenges in Section 3 through a careful analysis of the level lines of a meromorphic function with infinitely many poles.

We also describe intriguing connections (see Remark 3.1.24) between the uniform Howitt-Warren (or Le Jan-Raimond) stochastic flow of kernels and the a priori ill-posed diffusion (considered in physics [131])

$$dX_t = \xi(t, X_t)dt + dB_t,$$

where ξ is a space time white noise independent from the driving Brownian motion B or of the stochastic PDE

$$\partial_t v = \frac{1}{2}\partial_{xx} v + \xi \partial_x v,$$

associated to the above diffusion via the Kolmogorov backward equation.

3.1.1 Definitions

Before stating our main results, we need to introduce the notions of sticky Brownian motions and stochastic flows of kernels. Recall that the local time of a Brownian motion B_t at the point a is defined by the almost-sure limit

$$\ell_t^a(B) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^t \mathbb{1}_{a-\varepsilon \leq B_s \leq a+\varepsilon} ds = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^t \mathbb{1}_{a \leq B_s \leq a+\varepsilon} ds.$$

For a continuous semimartingale X_t, the natural time scale is given by its quadratic variation $\langle X, X\rangle_t$ and we define the local time as the almost sure limit [164, Corollary 1.9, Chap. VI]

$$\ell_t^a(X) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^t \mathbb{1}_{a \le X_s \le a+\varepsilon} d\langle X, X\rangle_t.$$

Feller initiated the study of Brownian motions sticky at the origin in [77], while studying general boundary conditions for diffusions on the half line.

Definition 3.1.1. *Brownian motion sticky at the origin* can be defined as the weak solution to the system of stochastic differential equations

$$\begin{aligned} dX_t &= \mathbb{1}_{\{X_t \neq 0\}} dB_t, \\ \int_0^t \mathbb{1}_{X_s = 0} ds &= \frac{1}{2\lambda} \ell_t^0(X), \end{aligned} \tag{3.1}$$

where B_t is a Brownian motion. Reflected Brownian motion sticky at the origin can be defined as $Y_t = |X_t|$ where X_t is a Brownian motion sticky at the origin.

Remark 3.1.2 (Time change). Brownian motion sticky at the origin can be viewed as a time change of Brownian motion in a construction due to Ito and McKean [110]. Consider the Brownian motion B_t, and define the continuous increasing function $A(t) = t + \frac{1}{2\lambda}\ell_t^0(B)$. Let $T(t) = A^{-1}(t)$ and set $X_t = B_{T(t)}$. We see that X_t is a usual Brownian motion when $X_t \neq 0$, because the local time of B_t only increases when $B_t = 0$. When $X_t = 0$ time slows down. We know $\int_0^t \mathbb{1}_{X_s > 0} ds = T(t)$, so $\int_0^t \mathbb{1}_{X_s = 0} ds = t - T(t) = \frac{1}{2\lambda}\ell_{T(t)}^0(B) = \frac{1}{2\lambda}\ell_t^0(X)$. This type of time change can be used to produce many processes with sticky interactions.

Remark 3.1.3 (Discrete limit). Reflected Brownian motion sticky at the origin Y_t can also be viewed as the diffusive limit of a sequence of random walks which tend to stay at 0. For small $\varepsilon > 0$, let Z_t^ε be a discrete time random walk on $\mathbb{Z}_{\ge 0}$, which behaves as a simple symmetric random walk when it is not at the point 0. When Z_t^ε is at the point 0, at each time step it travels to 1 with probability ε and stays at 0 with probability $1 - \varepsilon$. The diffusive limit $\varepsilon Z^{2\lambda\varepsilon}_{\varepsilon^{-2}t}$ converges to Y_t weakly as $\varepsilon \to 0$.

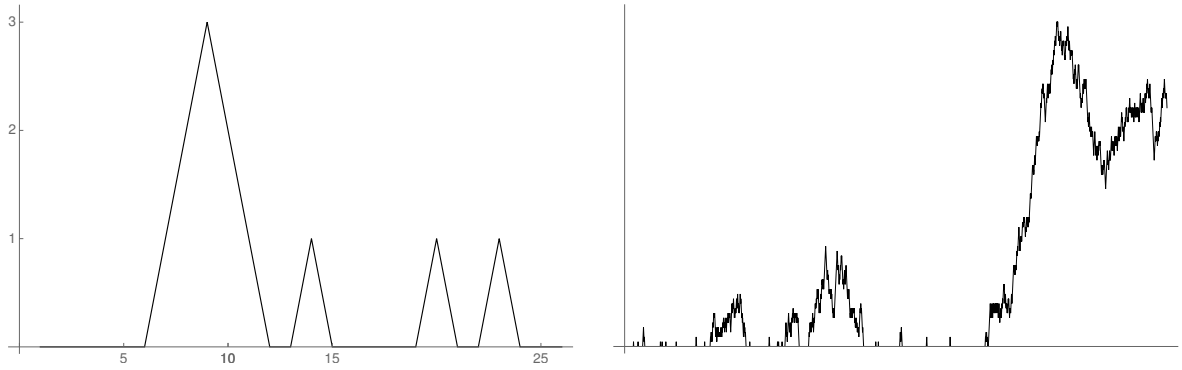

Figure 3.1: Left panel: Random walk $Z_t^{1/5}$ leaving 0 with probability $1/5$, up to time 25. Right panel: Reflected Brownian motion sticky at 0 obtained by the scaling limit of Z_t^{ε}.

To understand this convergence see equation (3.3), and note that the drift of the limiting motion at 0 is equal to 2λ because in each unit of time there are ε^{-2} opportunities to jump from 0 to ε and the proportion of these opportunities that is taken is approximately $2\lambda\varepsilon$. The analogous statement is also true for Brownian motion sticky at the origin. See Figure 3.1 where a simulation of $Z_t^{1/5}$ is shown alongside Y_t.

From Remark 3.1.2 and the Tanaka Formula for reflected Brownian motion it is easy to see that Y_t is a weak solution to the system of stochastic differential equations

$$\begin{aligned} dY_t &= \frac{1}{2} d\ell_t^0(Y) + \mathbb{1}_{\{Y_t>0\}} dB_t, \\ \mathbb{1}_{\{Y_t=0\}} &= \frac{1}{4\lambda} d\ell_t^0(Y), \end{aligned} \tag{3.2}$$

Equations (3.2) is equivalent to the single SDE

$$dY_t = 2\lambda \mathbb{1}_{\{Y_t=0\}} dt + \mathbb{1}_{\{Y_t>0\}} dB_t, \tag{3.3}$$

in the sense that a weak solution to one is a weak solution to the other [76]. Existence and uniqueness of weak solutions to (3.1) and (3.2) can be found in [76] and references therein.

Nonexistence of strong solutions to equations (3.1) and (3.2) was first shown in [57] and [197] (see also [76] for a more canonical arguments which would more easily generalize to other sticky processes). Several other works have been published on the existence of solutions to similar SDEs with indicator functions as the coefficient of dB_t or dt including [114, 29]. A more complete history of these SDEs can be found in [76].

We wish to study the evolution of n particles in one spatial dimension where the difference between any pair of particles is a Brownian motion sticky at the origin. First we do this for a pair of sticky Brownian motions.

Definition 3.1.4. The stochastic process $(X_1(t), X_2(t))$ is a pair of Brownian motions with sticky interaction if each X_i is marginally distributed as a Brownian motion and

$$\langle X_1, X_2 \rangle(t) = \int_0^t \mathbb{1}_{X_1(s)=X_2(s)} ds, \tag{3.4}$$

$$\int_0^t \mathbb{1}_{X_1(s)=X_2(s)} ds = \frac{1}{2\lambda} \ell_t^0(X_1 - X_2). \tag{3.5}$$

In other words $(X_1(t), X_2(t))$ are sticky Brownian motions if they evolve as independent Brownian motions when they are at different positions and their difference is a Brownian motion sticky at 0 (see a simulation in Fig. 3.2). The parameter λ can be understood as the rate (in a certain excursion theoretic sense) at which the two particles split when they are at the same position.

One can use Tanaka's formula to show that equation (3.5) is equivalent to saying

$$|X_1(t) - X_2(t)| - 2\lambda \int_0^t \mathbb{1}_{X_1(s)=X_2(s)} ds \tag{3.6}$$

is a martingale. Howitt and Warren [104] made this observation and generalized this martingale problem for a family of n particles with pairwise sticky interaction, which we call *n-point sticky Brownian motions*. In the most general case, the stickiness behaviour cannot be characterized uniquely by a single parameter λ. One needs to define for each $k, l \geq 1$ the "rate" at which a group of $k + l$ particles at the same position will split into two groups of respectively k and l coinciding

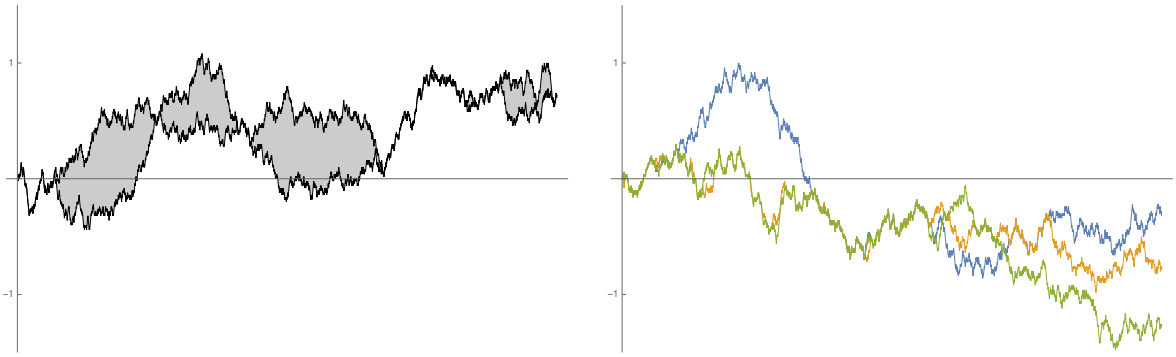

Figure 3.2: Left panel: Two Brownian motions with sticky interaction. Right panel: 3-point sticky Brownian motions. Not only do the paths stick pairwise, but sometimes all 3 paths may stick together. Both simulations are discretizations of sticky Brownian motions using the beta RWRE with $\varepsilon = 0.02$ (see Section 3.1.3).

particles. Following the notations in [104, 166, 169] this rate is denoted

$$\binom{k+l}{k}\theta(k,l).$$

Furthermore, we impose that the law of n-point sticky Brownian motions are consistent in the sense that any subsets of k particles for $k \leq n$ follow the law of the k-point sticky Brownian motions. This implies the relation $\theta(k+1,l) + \theta(k,l+1) = \theta(k,l)$. Under this relation, the family of nonnegative real numbers $\theta(k,l)$ can be equivalently (see [166, Lemma A.4]) characterized by a measure ν on $[0,1]$ such that

$$\int_0^1 x^{k-1}(1-x)^{l-1}\nu(dx) = \theta(k,l).$$

The following definition of n-point sticky Brownian motions from [169] is a reformulation of the Howitt-Warren martingale problem [104]. See Figure 3.2 and Figure 3.3 for simulations of n-point Brownian motions.

Definition 3.1.5 ([169, Theorem 5.3])**.** A stochastic process $\vec{X}(t) = (X_1(t), ..., X_n(t))$ started from $\vec{X}(0)$ will be called n-point sticky Brownian motions if it solves the following martingale problem called the Howitt-Warren martingale problem with drift β and characteristic measure ν.

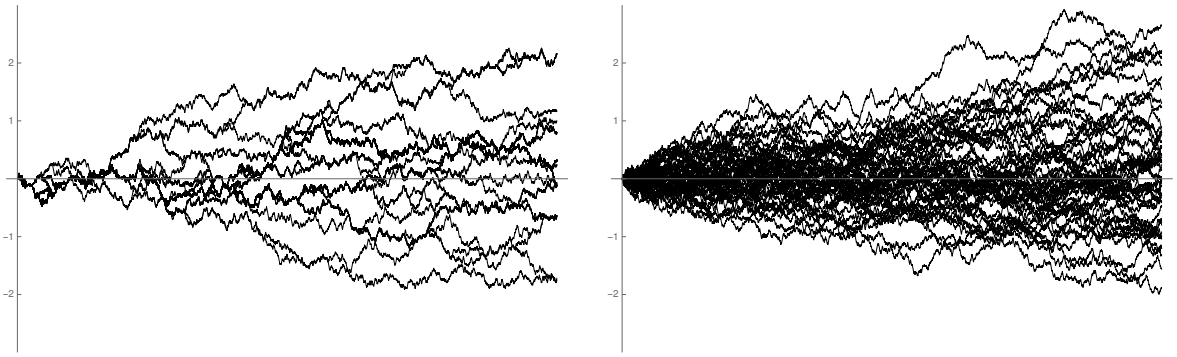

Figure 3.3: Left panel: 50 point-sticky Brownian motions using the same discretization as in Fig. 3.2. Because of the stickiness, the number of trajectories seems much smaller than 50. Right panel: 50 independent Brownian motions.

- (i) $\vec{X}$ is a continuous, square integrable martingale.

- (ii) The processes X_i and X_j have covariation process

$$\langle X_i, X_j \rangle(t) = \int_0^t \mathbb{1}_{X_i(s)=X_j(s)} ds, \qquad \text{for } t \geq 0, i, j = 1, ..., n.$$

- (iii) Consider any $\Delta \subset \{1, ..., n\}$. For $\vec{x} \in \mathbb{R}^n$, let

$$f_\Delta(\vec{x}) := \max_{i \in \Delta} \{x_i\} \qquad \text{and} \qquad g_\Delta(\vec{x}) := |\{i \in \Delta : x_i = f_\Delta(\vec{x})\}|,$$

  where $|S|$ denotes the number of elements in a set S. Then

$$f_\Delta(\vec{X}(t)) - \int_0^t \beta_+(g_\Delta(\vec{X}(t)) ds$$

  is a martingale with respect to the filtration generated by $\vec{X}$, where

$$\beta_+(1) := \beta \qquad \textit{and} \qquad \beta_+(m) := \beta + 2 \int \nu(dy) \sum_{k=0}^{m-2} (1-y)^k = \beta + 2 \sum_{k=1}^{m-1} \theta(1, k).$$

Remark 3.1.6. Definition 3.1.5 generalizes the definition of 2-point sticky Brownian motions because each particle marginally evolves as a Brownian motion, and the marginal distribution of any

pair of particles is that of a 2 point Brownian motion stickiness parameter $\lambda = \beta_+(2)$. Further, the consistency of the n-point motion is clear from property (iii).

We will be interested in a particular exactly solvable case of the Howitt-Warren Martingale problem.

Definition 3.1.7. An n-point stochastic process $(B_1(t), ..., B_n(t))$ will be called the *n-point uniform sticky Brownian motions* with stickiness λ if it solves the Howitt-Warren Martingale problem with drift $\beta = 0$ and characteristic measure

$$\nu(dx) = \mathbb{1}_{x\in[0,1]}\frac{\lambda}{2}dx.$$

This choice corresponds to choosing the fragmentation rates $\theta(k,l) = B(k,l)$, where $B(k,l) = \frac{\Gamma(k)\Gamma(l)}{\Gamma(k+l)}$ denotes the beta function. We explain below in Section 3.1.3 why this case is exactly solvable.

In order to realize the n-point sticky Brownian motions as a family of independent random motions in a random environment, we need to introduce the notion of stochastic flows of kernels. Let $\mathcal{B}$ be the Borel σ-algebra of $\mathbb{R}$. For any $s \le t$, a *random probability kernel*, denoted $\mathsf{K}_{s,t}(x, A)$, for $x \in \mathbb{R}$ and $A \in \mathcal{B}$, is a measurable function defined on some underlying probability space Ω, such that it defines for each $(x, \omega) \in \mathbb{R} \times \Omega$ a probability measure on $\mathbb{R}$. In order to interpret this as the random probability to arrive in A at time t starting at x at time s, the kernel needs to satisfy the following additional hypotheses.

Definition 3.1.8 ([169, Definition 5.1])**.** A family of random probability kernels $(\mathsf{K}_{s,t})_{s\le t}$ on $\mathbb{R}$ is called a *stochastic flow of kernels* if the following conditions are satisfied.

(i) For any real $s \le t \le u$ and $x \in \mathbb{R}$, almost surely $\mathsf{K}_{s,s}(x, A) = \delta_x(A)$, and

$$\int_{\mathbb{R}} \mathsf{K}_{s,t}(x, dy)\mathsf{K}_{t,u}(y, A)dy = \mathsf{K}_{s,u}(x, A)$$

for all $A \in \mathcal{B}$.

(ii) For any $t_1 \leq t_2 \leq \ldots \leq t_k$, the random kernels $(\mathsf{K}_{t_i,t_{i+1}})_{i=1}^{k-1}$ are independent.

(iii) For any $s \leq u$ and t real, $\mathsf{K}_{s,u}$ and $\mathsf{K}_{s+t,u+t}$ have the same finite dimensional distributions.

Remark 3.1.9. Additional continuity hypotheses were given in the original definition of a stochastic flow of kernels in [133], but we will only be interested in Feller processes for which these hypotheses are automatically satisfied.

The n-point motion of a stochastic flow of kernels is a family of n stochastic processes $X_1, \ldots, X_n$ on $\mathbb{R}$ with transition probabilities given by

$$P(\vec{x}, d\vec{y}) = \mathbb{E}\left[\prod_{i=1}^{n} \mathsf{K}_{0,t}(x_i, dy_i)\right]. \tag{3.7}$$

Every consistent family of n-point motions that is Feller, is the n-point motion of some stochastic flow of kernels [133]. Any solution to the Howitt-Warren martingale problem is a consistent family as was noted after Definition 3.1.5, and is Feller by [104]. So any solution to the Howitt-Warren martingale problem is the n-point motion of some stochastic flow of kernels.

Definition 3.1.10. A stochastic flow of kernels whose n-point motions solve the Howitt-Warren martingale problem is called a *Howitt-Warren flow*. The stochastic flow corresponding to the special case of the Howitt-Warren martingale problem considered in Definition 3.1.7 (that we called the uniform Howitt-Warren martingale problem), is sometimes called the *Le Jan-Raimond flow*, after the paper [135], following the terminology used in [169, 166].

In condition (i) of Definition 3.1.8, if we assume that we can move the almost surely so it occurs before choosing s, t, u and x, then we can sample all $\mathsf{K}_{s,t}$ and almost surely these kernels define the transition kernels for some continuous space-time markov process. Conditionally on the kernels we can describe the n-point motion as independent stochastic processes which evolve according to the transition kernels $\mathsf{K}_{s,t}$. Put simply the n-point motion can be seen as continuous space time random motions in a random environment which is given by the set of all transition kernels $\mathsf{K}_{s,t}$. In [166] (see also [169, Section 5]) it is shown that the change in quantifiers in (i)

necessary for this description can be done for Howitt-Warren flows. The random environment is explicitly constructed [166, Section 3] (see also [169, Section 5]) and consists of a Brownian web [1] plus a marked Poisson process at special points of the Brownian web [148]. The random motions in this environment essentially follow the Brownian web trajectories, except at these special points where they may turn left or right with a random probability. For Howitt-Warren flows such that $\int q(1-q)^{-1}\nu(dq) < \infty$ (which is not true for the Le Jan-Raimond flow), the random environment can also be constructed (see [166, Section 4]) using the Brownian net [179, 168].

Note that when starting from a set of particles on the real line and assuming that these particles will branch and coalesce following paths given by either the Brownian net or the Brownian web, the positions of the particles at a later time are given by a Pfaffian point process [85]. This type of evolution of Brownian particles is also related to random matrix theory, in particular the Ginibre evolution [191, 190, 192] (the evolution of real eigenvalues in a Ginibre matrix with Brownian coefficients), but these results do not seem to be directly related to the present chapter.

Following [166], we define a measure valued Markov process called the Howitt-Warren process by

$$\rho_t(dy) = \int \rho_0(dx)\mathsf{K}_{0,t}(x, dy).$$

It describes how a measure on the real line is transported by the Howitt-Warren flow. We also define a function valued Markov process called the dual smoothing process by

$$\zeta_t(x) = \int \mathsf{K}_{-t,0}(x, dy)\zeta_0(y). \tag{3.8}$$

This is a continuous analogue of the random average process [19]. For any fixed t, the processes ρ_t and ζ_t are related via the equality in distribution (called duality in [166])

$$\int \zeta_0(x)\rho_t(dx) = \int \zeta_t(x)\rho_0(dx).$$

[1] The Brownian web was introduced in [11], see also [185].

Note that a different and stronger form of Markov (self-) duality was investigated in [52] and applied to characterize the distribution of 2-point sticky Brownian motions. The result was restricted to 2-point motions and it is not clear if it translates in terms of stochastic flows of kernels.

The dual smoothing process was shown to lie in the Edwards-Wilkinson universality class [199], in the sense that for any fixed $x_0 \in \mathbb{R}$,

$$\mathcal{Z}_n(t,r) := \frac{1}{n^{1/4}} \zeta_{nt}(nx_0 + r\sqrt{n})$$

weakly converges as n goes to infinity – in the sense of finite dimensional marginals – to an explicit Gaussian process related to the stochastic heat equation with additive noise. This result holds under the assumption that at time $t = 0$, $\mathcal{Z}_n(0,x)$ converges to a smooth profile[2] (to which one may add some Brownian noise). An analogous statement in the discrete setting was proved in [19].

In the sequel, we will study the distribution of the dual smoothing process when $\zeta_0(y) = \mathbb{1}_{y>0}$ under a different scaling and we will see that the results are very different: instead of lying in the Edwards-Wilkinson universality class, the model lies in the Kardar-Parisi-Zhang universality class.

3.1.2 Results

Our first result is a Fredholm determinant formula for the Laplace transform of the uniform Howitt-Warren stochastic flow of kernels $\mathsf{K}_{0,t}(0,[x,\infty))$, or Le Jan-Raimond flow. In terms of the dual smoothing process, this corresponds to considering $\zeta_t(-x)$ with the initial condition $\zeta_0(y) = \mathbb{1}_{y>0}$.

First recall the definition of the gamma function

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx,$$

[2] the deterministic part of the initial profile needs to be C^1, and [199] assumes further that its derivative is bounded and Hölder $1/2 + \varepsilon$.

and the polygamma functions

$$\psi(\theta) = \partial_z \log \Gamma(z)|_{z=\theta}, \qquad \psi_i(\theta) = (\partial_z)^i \psi(z)|_{z=\theta}.$$

Theorem 3.1.11. *Let $\mathsf{K}_{0,t}(0,[x,\infty))$ denote the kernel of the uniform Howitt-Warren flow with stickiness parameter $\lambda > 0$. For $u \in \mathbb{C} \setminus \mathbb{R}_{>0}$, and $x > 0$, we have*

$$\mathbb{E}[e^{u\mathsf{K}_{0,t}(0,[x,\infty))}] = \det(I - \mathsf{K}_u)_{L^2(C)}, \tag{3.9}$$

(the R.H.S is a Fredholm determinant, see Definition 3.2.1 below), where

$$K_u(v,v') = \frac{1}{2\pi\mathbf{i}} \int_{1/2-\mathbf{i}\infty}^{1/2+\mathbf{i}\infty} \frac{\pi}{\sin(\pi s)} (-u)^s \frac{g(v)}{g(v+s)} \frac{ds}{s+v-v'},$$

and

$$g(v) = \Gamma(v) \exp\left(\lambda x \psi_0(v) + \frac{\lambda^2 t}{2} \psi_1(v)\right).$$

where C is a positively oriented circle with radius $1/4$ centered at $1/4$. (It is important that this contour passes through zero at the correct angle. The actual radius of the circle C does not matter.)

Remark 3.1.12. We use two very different notions of kernels, which are both denoted by the letter K. We will reserve the font K for stochastic flows of kernels, and the usual font K for the kernels of $\mathbb{L}^2$ operators arising in Fredholm determinants.

We reach Theorem 3.1.11 by taking a limit of a similar Fredholm determinant formula [23, Theorem 1.13] for the beta RWRE defined in Section 3.1.3. Theorem 3.1.11 is proved in Section 3.4.

We perform a rigorous saddle-point analysis of the Laplace transform formula (3.9) to obtain a quenched large deviation principle for the uniform Howitt-Warren stochastic flow.

Theorem 3.1.13. *Let $\lambda > 0$ and $x \geq 1.35$. Let $\mathsf{K}_{s,t}$ be the kernel of a uniform Howitt-Warren flow.*

Then we have the following convergence in probability

$$\frac{1}{t}\log \mathsf{K}_{0,t}(0,[xt,\infty)) \xrightarrow[t\to\infty]{} -\lambda^2 J(x/\lambda), \tag{3.10}$$

where

$$J(x) = \max_{\theta\in\mathbb{R}_{>0}} \left\{ \frac{1}{2}\psi_2(\theta) + x\psi_1(\theta) \right\}. \tag{3.11}$$

The condition $x \geq 1.35$ is technical and is addressed in Remark 3.1.16. We expect that the limit holds almost surely. This should follow from subadditivity arguments, though we do not pursue this in the present chapter (see [163] for an almost sure quenched large deviation principle for discrete random walks). We emphasize that in Theorem 3.1.13, the rate function $J(x)$ is expressed explicitly using well-known special functions, which is in contrast with what one would obtain using subadditivity arguments. Another large deviation principle was shown in [69] for the empirical distribution of a certain class of n-point sticky Brownian motions, but this does not seem to be related to the present Theorem 3.1.13.

Remark 3.1.14. The annealed[3] analogue of this large deviation principle just describes the tail behavior of a standard Brownian motion. Indeed,

$$\frac{1}{t}\log \mathbb{E}[\mathsf{K}_{0,t}(0,[xt,\infty))] = -x^2/2.$$

It can be easily checked that $\lambda^2 J(x/\lambda) > x^2/2$ which, in the context of directed polymers, means that the model exhibits strong disorder. Note that the sign of the inequality is consistent with Jensen's inequality (assuming (3.10) holds in L^1). The inequality becomes an equality in the $\lambda \to \infty$ limit, which corresponds to Brownian motions with no stickiness.

When uniform sticky Brownian motions are viewed as random walks in a random environment, Theorem 3.1.13 gives a large deviation principle whose rate function is deterministic despite the

[3] In the context of random walks in random environment and directed polymers, the (limiting) quenched free energy or rate function is the limit obtained for almost every environment and the annealed analogues correspond to the same quantities for the averaged environment.

randomness of the environment. The random variable $\log \mathsf{K}_{0,t}$ does depend on the environment, but its fluctuations are small enough that they are not detected by the large deviation principle. We prove that the model is in the KPZ universality class in the sense that the random lower order corrections to the large deviation principle, or equivalently the fluctuations of $\log \mathsf{K}_{0,t}$, are Tracy-Widom GUE distributed on the $t^{1/3}$ scale.

Theorem 3.1.15. *Let $\mathsf{K}_{s,t}$ be the kernel of a uniform Howitt-Warren flow with stickiness parameter $\lambda > 0$. Let $0 < \theta < 1$. We have*

$$\lim_{t\to\infty} \mathbb{P}\left(\frac{\log(\mathsf{K}_{0,t}(0,[x(\theta)t,\infty)) + \lambda^2 J(x(\theta)/\lambda)t}{t^{1/3}\sigma(\theta)} < y\right) = F_{\mathrm{GUE}}(y),$$

where $F_{\mathrm{GUE}}(y)$ is the cumulative density function of the Tracy-Widom distribution (defined below in (3.22)*), and*

$$x(\theta) = -\frac{\lambda}{2}\frac{\psi_3(\theta)}{\psi_2(\theta)}, \qquad \sigma(\theta) = \frac{\lambda^{2/3}}{2^{1/3}}\left(\frac{-1}{2}\psi_4(\theta) - \frac{x(\theta)}{\lambda}\psi_3(\theta)\right)^{\frac{1}{3}}. \tag{3.12}$$

Theorem 3.1.15 comes from applying a rigorous steep descent analysis to the Fredholm determinant in Theorem 3.1.11. The proof is given in Section 3.2 with some technical challenges deferred to Section 3.3 and Appendix B. The parametrization of functions J and σ arising in the limit theorem via the variable θ may appear unnatural at this point. It will appear more natural in the proof as θ is the location of the critical point used in the steep descent analysis. We expect that there should exist another interpretation of the parameter θ. It should naturally parametrize stationary measures associated with the uniform Howitt-Warren flow, and KPZ scaling theory [173, 126] would predict the expressions for $J(x)$ and $\sigma(\theta)$ given above. This approach would require to degenerate to the continuous limit the results from [18] and we leave this for future investigation (the analogue of parameter θ in the discrete setting is denoted $\lambda(\xi)$ in [18, Theorem 2.7]).

Remark 3.1.16. Note that $x(\theta)$ is a decreasing function of θ and the technical hypothesis $\theta < 1$ corresponds to approximately $1.35 \leq x(\theta)$. Similarly $J(x)$ is an increasing function of x and $\theta < 1$

corresponds approximately to $1.02 < J(x(\theta))$. We expect Theorem 3.1.15 to hold for all $\theta > 0$, and Theorem 3.1.13 to hold for all $x > 0$, however if $\theta \geq 1$ we pick up additional residues while deforming the contours of our Fredholm determinant during the asymptotic analysis which make the necessary justifications significantly more challenging.

More generally, we believe that the result of Theorem 3.1.11 should be universal and hold for more general Howitt-Warren flows under mild assumptions on the characteristic measure ν. This would be analogous to a conjecture that for discrete polymer models the fluctuations of the free energy are Tracy-Widom distributed as long as the weights of the polymer have finite fifth moments [7, Conjecture 2.6]. Moreover, based on [162, Theorem 4.3], we expect that the random variable

$$\log \mathsf{K}_{0,t}(0, [xt, xt + a)),$$

for any $a > 0$, satisfies the same limit theorems as $\log \mathsf{K}_{0,t}(0, [xt, +\infty))$ in Theorem 3.1.13 and Theorem 3.1.15, with the same constants (the prediction that the constant $\sigma(\theta)$ should remain the same is suggested by the results of [183]).

Following [23] we can state a corollary of Theorem 3.1.15. In general, tail probability estimates provide information about the extremes of independent samples. In the present case, we obtain that the largest among n uniform sticky Brownian motions fluctuates asymptotically for large n according to the Tracy-Widom distribution. We will see that the result is very different from the case of n independent Brownian motions, as can be expected from the simulations in Figure 3.3.

Corollary 3.1.17. *Let $c \in [1.02, \infty)$, let x_0 be such that $\lambda^2 J(x_0/\lambda) = c$, let θ_0 be such that $x(\theta_0) = x_0$, and let $\{B_i(t)\}$ be uniform n-point sticky Brownian motions with stickiness parameter $\lambda > 0$ and scale n as $n = e^{ct}$, then*

$$\lim_{t\to\infty} \mathbb{P}\left(\frac{\max_{i=1,\dots,n}\{B_i(t)\} - tx_0}{t^{1/3}\sigma(\theta_0)/(\lambda^2 J'(x_0/\lambda))} \leq y\right) = F_{\mathrm{GUE}}(y). \tag{3.13}$$

The proof of Corollary 3.1.17 is very similar to the proof of [23, Corollary 5.8] and uses the fact that after conditioning on the environment we are dealing with independent motions along with our strong control of the random variable $\mathsf{K}_{0,t}(0,[xt,\infty))$ from Theorem 3.1.15. The details of the proof can be found at the end of Section 3.2.

3.1.3 Integrability for n-point uniform sticky Brownian motions

In 2013 Povolotsky [158] introduced the q-Hahn Boson, a three parameter family of Bethe ansatz solvable discrete zero range processes, computed the Bethe ansatz eigenfunctions, and conjectured their completeness. The q-Hahn Boson and its eigenfunctions were further studied in [64] where a Markov duality with the so-called q-Hahn TASEP, an interacting particle system closely related to the q-Hahn Boson, was used to compute integral formulas for the q-moments and the q-Laplace transform of the particle positions. The q-Hahn Boson eigenfunctions were also further studied in [34, 40] where the completeness of eigenfunctions was proved and their Plancherel theory was developed. In [23] a model of random walks in a one dimensional random environment, called the beta RWRE, was introduced as the $q \to 1$ limit of the q-Hahn TASEP. All features of the integrability of the model survive in the scaling limit. Uniform sticky Brownian motions are a limit of the beta RWRE and we show in the present article that it inherits as well all the integrability of the q-Hahn Boson. Note that the q-Hahn Boson fits into the more general framework of stochastic higher spin 6 vertex models [34, 46, 67], so uniform sticky Brownian motions are also a limit of a stochastic vertex model.

Definition 3.1.18. The beta random walk in random environment (beta RWRE) depends on two parameters $\alpha > 0$ and $\beta > 0$. Let $\{w_{(x,t)}\}_{x\in\mathbb{Z},t\in\mathbb{Z}_{\geq 0}}$ be iid beta distributed random variables with parameters α,β. Recall that a beta random variable w with parameters $\alpha,\beta > 0$ is defined by

$$\mathbb{P}(w \in dx) = \mathbb{1}_{x\in[0,1]}\frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}dx,$$

where $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$. We will call the values of the random variables $w_{(x,t)}$ for all $x \in \mathbb{Z}, t \in \mathbb{Z}_{\geq 0}$

the random environment.

Given a random environment, we begin k independent random walks $(X_1(t), ..., X_k(t))$ from position $\vec{x}_0$. Each random walker has jump distribution

$$\mathsf{P}(X(t+1) = x+1 | X(t) = x) = w_{(x,t)} \qquad \mathsf{P}(X(t+1) = x-1 | X(t) = x) = 1 - w_{(x,t)}.$$

We will use $\vec{X}^{\vec{x}}(t) = (X_1^{x_1}(t), ..., X_k^{x_k}(t))$ to refer to the position of k independent random walks started from $(x_1, ..., x_k)$ at time t. Unless another initial condition is specified, $\vec{X}(t) = (X_1(t), ..., X_k(t))$ will refer to the position of k random walkers started from the origin.

We use the symbol P with bold font for the quenched probability measure on paths, which is obtained by conditioning on the environment. Similarly we used the same fonts for the quenched probability kernels K which describe transition probabilities after conditioning on the environment. The usual symbols $\mathbb{P}$ (resp. $\mathbb{E}$) will be used to denote the measure (resp. the expectation) on the environment.

Note that any single trajectory of the beta RWRE is just a simple random walk and the random environment has no effect. However, if we consider multiple paths on the same environment, they are correlated by the environment. In particular, they do not behave as simple random walks when they meet.

We consider now the continuous limit of the model. If we simply rescale space and time diffusively, trajectories become Brownian motions $\mathbb{P}$-almost-surely [161]. Moreover, $\vec{X}(t)$ converges to a family of independent Brownian motions and the effect of the environment has vanished in the limit. In order to keep a dependence on the environment, we need to rescale the weights $w_{(x,t)}$ so that two paths at the same location have a high probability of staying together. This will be the case if $w_{(x,t)}$ is close to either 0 or 1 with high probability, which, for a beta distributed random variable, happens when both parameters go to 0. More precisely, choose a positive parameter λ and set $\alpha_\varepsilon = \beta_\varepsilon = \lambda\varepsilon$. We will be interested in the process $\vec{X}_\varepsilon(t) = (X_{1,\varepsilon}(t), ..., X_{k,\varepsilon}(t))$, which is obtained as the particle positions at time t of k random walkers in a beta distributed random

environment with parameters $\alpha_\varepsilon, \beta_\varepsilon$ started from the origin.

Lemma 3.1.19. *As $\varepsilon \to 0$, the n-point beta random walk in random environment $\left(\varepsilon \vec{X}_\varepsilon(\varepsilon^{-2}t)\right)_{t\geq 0}$ with parameters $\alpha_\varepsilon = \beta_\varepsilon = \lambda\varepsilon$ weakly converges to an n-point uniform sticky Brownian motions with stickiness parameter λ in the space of continuous functions equipped with the topology of uniform convergence on compact sets.*

Proof. We apply [169, Theorem 5.3] with drift $\beta = 0$, and $\nu(dx) = \frac{\lambda}{2}\mathbb{1}_{[0,1]}dx$. □

In fact random walks in a beta distributed random environment were the first random walk in random environment shown to converge to sticky Brownian motions in [132], though this result was shown on a torus. After reformulating sticky Brownian motions as a martingale problem, Howitt and Warren extended this convergence to random walks in any random environment provided the random variables defining the environment have certain scaling limits [104, 105]. This theorem was reformulated in [166, 169] to arrive at the form used above.

Now we quote a formula for the quenched probability $\mathsf{P}(X(t) > x)$ in the beta random walk in random environment, where $X(t)$ is the path of a single particle that starts from 0 at time 0. This quantity is the analogue of $K_{0,t}(0, [x, \infty))$ in the case of the beta random walk in random environment. It satisfies the following formula

Theorem 3.1.20 ([23, Theorem 1.13])**.** *For $u \in \mathbb{C}\setminus\mathbb{R}_{>0}$ and $\alpha, \beta > 0$, fix $t \in \mathbb{Z}_{\geq 0}$ and $x \in \{-t, ..., t\}$ with the same parity. Then*

$$\mathbb{E}[e^{u\mathsf{P}(X(t)>x)}] = \det(I - \mathsf{K}_u^{\mathrm{RW}})_{\mathbb{L}^2(C_0)},$$

where C_0 is a small positively oriented contour that contains 0 and does not contain the points $-\alpha - \beta$ and -1, and $\mathsf{K}_u^{\mathrm{RW}} : \mathbb{L}^2(C_0) \to \mathbb{L}^2(C_0)$ is defined in terms of its kernel

$$\mathsf{K}_u^{\mathrm{RW}}(v, v') = \frac{1}{2\pi\mathbf{i}} \int_{\frac{1}{2}-\mathbf{i}\infty}^{\frac{1}{2}+\mathbf{i}\infty} \frac{\pi}{\sin(\pi s)}(-u)^s \frac{g^{\mathrm{RW}}(v)}{g^{\mathrm{RW}}(v+s)} \frac{ds}{s+v-v'},$$

where

$$g^{\mathrm{RW}}(v) = \left(\frac{\Gamma(v)}{\Gamma(\alpha+v)}\right)^{(t-x)/2} \left(\frac{\Gamma(\alpha+\beta+v)}{\Gamma(\alpha+v)}\right)^{(t+x)/2} \Gamma(v).$$

Theorem 3.1.20 is the starting point for our study of the uniform sticky Brownian motion in this chapter, in particular Theorem 3.1.11 is derived as a limit of this formula.

Remark 3.1.21. There is a sign mistake in [23, Theorem 1.13]. It reads $\mathbb{E}[e^{u\mathsf{P}(X(t)\geq x)}] = \det(I + K_u^{\mathrm{RW}})_{\mathbb{L}^2(C_0)}$, but the right hand side should be $\det(1 - K_u^{\mathrm{RW}})_{\mathbb{L}^2(C_0)}$. This is corrected in Theorem 3.1.20.

As we have already mentioned, the crucial tool underlying the exact solvability of the beta RWRE is the Bethe ansatz. We will describe now the sense in which n-point uniform sticky Brownian motions are also amenable to Bethe ansatz diagonalization. This could lead to another proof of Theorem 3.1.11, though we do not provide, in this chapter, the necessary justifications to make this alternative proof complete.

Let K be the kernel of a uniform Howitt-Warren flow, and let $\vec{x} \in \mathbb{R}^k$. We define the function

$$\Phi_t^{(k)}(x_1, \ldots, x_k) := \mathbb{E}\Big[\mathsf{K}_{-t,0}(x_1, [0,+\infty)) \ldots \mathsf{K}_{-t,0}(x_k, (0,+\infty))\Big].$$

Note that since the random variables $\mathsf{K}_{-t,0}(x, (0,+\infty))$ are bounded between 0 and 1, so are the mixed moments $\Phi_t^{(k)}(x_1, ..., x_k)$. In particular the knowledge of $\Phi^{(k)}$ uniquely determines their distribution. For instance, we have for any $u \in \mathbb{C}$

$$\mathbb{E}\left[e^{u\mathsf{K}_{-t,0}(x,(0,+\infty))}\right] = \sum_{k=0}^{\infty} \frac{u^k}{k!} \Phi_t^{(k)}(x, \ldots, x). \tag{3.14}$$

where there are k occurrences of the variable x in the argument above.

Proposition 3.1.22. *For $x_1 \geq \cdots \geq x_k$, and $t > 0$,*

$$\Phi_t^{(k)}(x_1, \ldots, x_k) = \int_{\alpha_1 + \mathbf{i}\mathbb{R}} \frac{\mathrm{d}w_1}{2\mathbf{i}\pi} \cdots \int_{\alpha_k + \mathbf{i}\mathbb{R}} \frac{\mathrm{d}w_k}{2\mathbf{i}\pi} \prod_{1 \leqslant A < B \leqslant k} \frac{w_B - w_A}{w_B - w_A - w_A w_B} \prod_{j=1}^{k} \exp\left(\frac{t\lambda^2 w_j^2}{2} + \lambda x_j w_j\right) \frac{1}{w_j}, \quad (3.15)$$

where for $i < j$, $0 < \alpha_i < \frac{\alpha_j}{1+\alpha_j}$. The value at $t = 0$ should be understood as

$$\phi_0(x_1, ..., x_k) = \lim_{t \to 0^+} \phi_t(x_1, ..., x_k).$$

Proposition 3.1.22 is proved in Section 3.5. We also show in Section 3.5.2 that $\Phi_t^{(k)}(\vec{x})$ converges, under appropriate scaling, to the moments of the stochastic heat equation with multiplicative noise. This suggests that Howitt-Warren stochastic flows weakly converge in the weak noise limit ($\lambda \to +\infty$ with time and space rescaled) to the solution to the KPZ equation.

One may observe that (see details in Section 3.5.3) the right hand side of (3.15) satisfies the following heat equation subject to boundary conditions

$$\begin{cases} \partial_t u(t, \vec{x}) = \frac{1}{2} \Delta u(t, \vec{x}), \quad t \geq 0, \vec{x} \in \mathbb{R}, \\ (\partial_i \partial_{i+1} + \lambda(\partial_i - \partial_{i+1})) u(t, \vec{x})|_{x_i = x_{i+1}} = 0. \end{cases} \quad (3.16)$$

Proposition 3.1.22 shows that (3.16) can be solved using coordinate Bethe ansatz, at least for certain initial conditions. We refer to [62, Section 3.4.1] or [40] for background on coordinate Bethe ansatz in a similar context. In general, Bethe ansatz eigenfunctions corresponding to this problem can be parametrized by k complex numbers $z_1, \ldots, z_k$ and written as

$$\Psi_{\vec{z}}(\vec{x}) = \sum_{\sigma \in S_k} \prod_{1 \leq i < j \leq k} \frac{z_{\sigma(i)} - z_{\sigma(j)} - 1}{z_{\sigma(i)} - z_{\sigma(j)}} \prod_{j=1}^{k} e^{-\frac{\lambda x_j}{z_j}}. \quad (3.17)$$

Remark 3.1.23. It is natural (see Section 3.5.4) to associate to (3.16) the following Schrödinger

type equation on $\mathbb{R}^k$ with point interactions

$$\partial_t v(t,\vec{x}) = \frac{1}{2}\Delta v(t,\vec{x}) + \frac{1}{2\lambda}\sum_{i\neq j}\delta(x_i - x_j)\partial_{x_i}\partial_{x_j} v(t,\vec{x}). \tag{3.18}$$

We expect the operator $\frac{1}{2}\Delta + \frac{1}{2\lambda}\sum_{i\neq j}\delta(x_i - x_j)\partial_{x_i}\partial_{x_j}$ to be the generator of the n-point uniform sticky Brownian motions, though we do not address in the present chapter the details necessary to make rigorous sense of this statement. Note that similar operators appear in the study of turbulence, in particular in Kraichnan's model of passive scalar [32] and connections to sticky Brownian motions have been noticed in the physics literature [87].

Remark 3.1.24. Using $\mathbb{E}[\xi(s,x)\xi(t,y)] = \delta(t-s)\delta(y-x)$ for a space-time white noise ξ, the Schrödinger equation (3.18) is formally satisfied by the moments of the following stochastic PDE (assuming the existence of such an object, see more details in Section 3.5.4)

$$\begin{cases} \partial_t q(t,x) = \frac{1}{2}\partial_{xx} q(t,x) + \frac{1}{\sqrt{\lambda}}\xi(t,x)\partial_x q(t,x), \\ q(0,t) = q_0(x). \end{cases} \tag{3.19}$$

If ξ was a smooth and Lipschitz potential, the Kolmogorov backward equation would provide a representation of the solution as

$$q(x,t) = \mathsf{E}[q_0(X_0)|X_{-t} = x],$$

where X_t is the random diffusion

$$dX_t = \frac{1}{\sqrt{\lambda}}\xi(X_t,t)dt + dB_t, \tag{3.20}$$

where the Brownian motion B is independent from ξ, and E denotes the expectation with respect to B, conditionally on the environment ξ. For a space-time white noise drift, we have not found any rigorous construction in the literature, and the fact that ξ is not smooth introduces three problems.

First, when ξ is a white noise equation (3.20) is ill-defined. Second, If ξ were regularized to be smooth in space but white in time equation (3.20) would be incorrect (This case is studied in [196]). The final problem is explained in Remark 3.1.25.

Note that the same diffusion (3.20) is considered in the physics paper [131, Equation (2)] by Le Doussal and Thiery and our results are consistent with some of their predictions (if we identify the solution $q(t,x)$ of (3.19) with the dual smoothing process (defined in (3.8)) of the Le Jan-Raimond flow $\zeta_t(-x)$). Moreover, if we interpret ξ as a velocity field, (3.19) can be seen as an advection-diffusion equation as in Kraichnan's model [124], a model of turbulent flow designed to explain anomalous exponents not predicted by Kolmogorov theory of turbulence, we refer to the review articles [172] for physics background or [128] for a more mathematical exposition. Note that the series of physics works [55, 88, 32, 89, 90] on Kraichnan's model were part of the motivation for the work of Le Jan and Raimond [134, 133] on stochastic flows.

Remark 3.1.25. Despite the previous remark, one should not define the solution $q(t,x)$ of the stochastic PDE (3.19) as the dual smoothing process $\zeta_t(x)$ of the Le Jan-Raimond flow (defined in (3.8)), even though the moments of both quantities satisfy the same evolution equation (see more details in Section 3.5.4). Indeed, it was proved by Le Jan and Lemaire [136, 135] that the noise generated by the Le Jan-Raimond flow of kernels is black, which implies that, if ξ is a space-time white noise, there cannot be a probability space on which $\zeta_t(x)$ is a strong solution to (3.19).

Remark 3.1.26. We expect the Bethe ansatz eigenfunctions $\Psi_{\vec{z}}(\vec{x})$ (3.17) to be orthogonal with respect to a simple inner product and to form a basis of a large subspace of functions on $\mathbb{R}^k$. These properties would in principle allow to solve (3.16) for a large class of initial data, although we expect concise integral formulas such as (3.73) only in a handful cases. Proofs of such statements would likely come from degenerating the Plancherel theory [40, 41] for the q-Hahn Boson Bethe ansatz eigenfunctions.

3.1.4 Outline of the proofs

In Section 3.2 we begin with a Fredholm determinant formula for the Laplace transform of the random kernel for a uniform Howitt-Warren flow, then apply a rigorous saddle point analysis to show that the large deviation principle for this random kernel has Tracy-Widom corrections. For readability we will delay some details of the arguments to Section 3.3 and Appendix B. Section 3.3 is devoted to constructing a contour which is needed for the saddle point analysis in the previous section. This is one of the main challenges in our saddle point analysis and involves a study of the level set of the real part of a certain meromorphic function. Appendix B provides the bounds necessary to apply dominated convergence to our Fredholm determinant expansions in order to make the saddle point analysis in Section 3.2 rigorous.

In Section 3.4 we derive the Fredholm determinant formula for the Laplace transform of the point to half line probability for uniform sticky Brownian motions used in Section 3.2 as the limit of a similar formula for the beta RWRE. The argument is straightforward but requires technical bounds based on known asymptotics for the Gamma and PolyGamma functions. The proof is divided into three steps and the idea of the argument can be understood after reading the first step of the proof. The necessary bounds are provided in the latter two steps.

Section 3.5 is independent from the other sections and provides a proof of the mixed moment formulas for the uniform sticky Brownian motions by taking a limit of similar formulas for the beta RWRE. We also explain the relation between this moment formula and Bethe ansatz, the KPZ equation and the diffusion (3.20)

Appendix A gives precise bounds on the Gamma and Polygamma function which are necessary for the construction of the contours in our saddle point analysis.

3.1.5 Acknowledgements

We are greatly indebted to Rongfeng Sun for telling us about the convergence of random walks in random environment to sticky Brownian motions, and asking if one could study large deviation tails of stochastic flows via similar techniques as in [23]. G.B. and M.R. thank Ivan Corwin for

many useful discussions at all stages of this project. We also thank Yves Le Jan and Jon Warren for helpful comments on an initial version of the manuscript. G.B. also thanks Emmanuel Schertzer for enlightening explanations regarding sticky Brownian motions and stochastic flows and Denis Bernard and Pierre Le Doussal for useful discussions.

G.B. and M.R. were partially supported by the NSF grant DMS:1664650. M. R. was partially supported by the Fernholz Foundation's "Summer Minerva Fellow" program, and also received summer support from Ivan Corwin's NSF grant DMS:1811143.

3.2 Asymptotic analysis of the Fredholm determinant

The overall goal of this section is to show that for large time, the fluctuations of the log of the kernel of a uniform Howitt-Warren flow converges to the Tracy-Widom distribution (Theorem 3.1.15). We first use a trick from [36] to access the large time distribution of $\mathsf{K}_{0,t}(0,[x,\infty))$ from its Laplace transform without using Laplace inversion formula. Then we apply the method of steep descent to the Fredholm determinant from Theorem 3.1.11 and prove that, in the appropriate scaling limit, it converges to the cumulative density function of the Tracy-Widom distribution.

We first recall the definition of a Fredholm determinant.

Definition 3.2.1. For any contour C and any measurable function $K : C \times C \to \mathbb{C}$, which we will call a kernel, the Fredholm determinant $\det(1+K)_{L^2(C)}$ is defined by

$$\det(1+K)_{L^2(C)} = 1 + \sum_{k=1}^{\infty} \frac{1}{k!} \int_{C^k} \det(K(x_i,x_j))_{1\le i,j\le k} \prod_{i=1}^{k} dx_i, \tag{3.21}$$

provided the right hand side converges absolutely.

The Tracy-Widom distribution is defined by its cumulative density function

$$F_{\mathrm{GUE}}(x) = \det(I - K_{\mathrm{Ai}})_{L^2(x,\infty)}, \tag{3.22}$$

where the Airy kernel K_{Ai} is defined as

$$K_{\mathrm{Ai}}(x,y) = \frac{1}{2\pi\mathbf{i}} \int_{e^{-\frac{2\pi\mathbf{i}}{3}}\infty}^{e^{\frac{2\pi\mathbf{i}}{3}}\infty} d\omega \int_{e^{-\frac{\pi\mathbf{i}}{3}}\infty}^{e^{\frac{\pi\mathbf{i}}{3}}\infty} dz \frac{e^{\frac{z^3}{3}-zx}}{e^{\frac{\omega^3}{3}-\omega y}} \frac{1}{(z-\omega)}.$$

In this integral the contours for z and ω do not intersect. We may think of the integrating z over the contour $(e^{-\frac{\pi\mathbf{i}}{3}}\infty, 1] \cup (1, e^{\frac{\pi\mathbf{i}}{3}}\infty)$ and the integral w over the contour $(e^{-\frac{2\pi\mathbf{i}}{3}}\infty, 0] \cup (0, e^{\frac{2\pi\mathbf{i}}{3}}\infty)$.

Instead of inverting the Laplace transform in Theorem 3.1.11, we use a standard trick appearing as Lemma 4.1.39 in [36] and take a limit of the Laplace transform to obtain the following formula for the point to half line probability of sticky Brownian motions.

Proposition 3.2.2. *Let $K_u(v,v')$ be as defined in Theorem 3.1.11. For $\lambda > 0$, $\theta > 0$, $t > 0$, and arbitrary constants $x(\theta)$, $J(x(\theta))$, $\sigma(\theta)$ depending on θ, if $\lim_{t\to\infty} \det(I - K_{u_t(y)})_{L^2(C)}$ is the continuous cumulative density function of a random variable, then*

$$\lim_{t\to\infty} \mathbb{P}\left(\frac{\log(\mathsf{K}_{0,t}(0,[x(\theta)t,\infty))) + \lambda^2 J(x(\theta)/\lambda)t}{t^{1/3}\sigma(\theta)} < y \right) = \lim_{t\to\infty} \det(I - K_{u_t(y)})_{L^2(C)},$$

where $u_t(y) = -e^{t\lambda^2 J(x(\theta)/\lambda) - t^{1/3}\sigma(\theta)y}$

Proof of Proposition 3.2.2. Set $x = x(\theta)t$. Then

$$e^{u_t(y)\mathsf{K}_{0,t}(0,[x,\infty))} = \exp\left(-e^{t^{1/3}\sigma(\theta)\left(\frac{t\lambda^2 J(x(\theta)/\lambda) + \log(\mathsf{K}_{0,t}(0,[x(\theta)t,\infty))}{t^{1/3}\sigma(\theta)} - y \right)} \right).$$

Considering the function $f_t(x) = \exp(-e^{t^{1/3}\sigma(\theta)x})$ and keeping in mind that $\sigma(\theta) > 0$, we see that $f_t(x)$ is strictly decreasing in x, it approaches 0 as $x \to \infty$ and it approaches 1 as $x \to -\infty$. We also see that as $t \to \infty$ this function converges to $\mathbb{1}_{x<0}$ uniformly on the interval $\mathbb{R} \setminus [-\delta, \delta]$ for any choice of $\delta > 0$.

If we define the r shift $f_t^r(x) = f_t(x - r)$, then

$$\mathbb{E}[e^{u_t(r)\mathsf{K}_{0,t}(0,[x,\infty))}] = \mathbb{E}\left[f_t^r\left(\frac{t\lambda^2 J(x(\theta)/\lambda) + \log(\mathsf{K}_{0,t}(0,[x(\theta)t,\infty))}{t^{1/3}\sigma(\theta)} \right) \right].$$

By Theorem 3.1.11, $\lim_{t\to\infty} \mathbb{E}[e^{u_t(-y)\mathsf{K}_{0,t}(0,[x,\infty))}] = \lim_{t\to\infty} \det(I - K_{u_t(y)})_{L^2(C)}$, and by assumption, this is the continuous cumulative density function of a random variable. Using [36, Lemma 4.1.39], completes the proof. □

3.2.1 Setup

Most of this Section 3.2 will be devoted to proving the following Proposition 3.2.3. Together with Proposition 3.2.2 it proves Theorem 3.1.15.

Proposition 3.2.3. *For $\lambda > 0$, $t > 0$, $x > 0$, and constants $x(\theta), J(x(\theta)), \sigma(\theta)$ from* (3.12)*, we have*

$$\lim_{t\to\infty} \det(I - K_{u_t(y)})_{L^2(C)} = F_{\mathrm{GUE}}(y).$$

First we rewrite $K_{u_t(y)}$ in order to apply the method of steep descent. Performing the change of variables $z = s + v$ gives

$$K_{u_t(y)}(v, v') = \frac{1}{2\pi\mathbf{i}} \int_{1/2+\mathbf{i}\mathbb{R}} \frac{\pi}{\sin(\pi(z-v))} e^{(z-v)(t\lambda^2 J(x(\theta)/\lambda) - t^{1/3}\sigma(\theta)y)} \frac{g(v)}{g(z)} \frac{dz}{z-v'}.$$

Here we have used the fact that the contour for v can be made arbitrarily small so that the contour for z can be deformed from $1/2 + v + \mathbf{i}\mathbb{R}$ to $1/2 + \mathbf{i}\mathbb{R}$ without crossing poles of $\frac{\pi}{\sin(\pi(z-v))}$. Recall that

$$g(v) = \exp\left(\frac{\lambda^2 t}{2}\psi_1(v) + \lambda x \psi(v)\right)\Gamma(v),$$

so replacing x by xt gives

$$K_{u_t(y)}(v, v') = \frac{1}{2\pi\mathbf{i}} \int_{1/2+\mathbf{i}\mathbb{R}} \frac{\pi}{\sin(\pi(z-v))} e^{t(h(z)-h(v)) - t^{1/3}\sigma(\theta)y(z-v)} \frac{\Gamma(v)}{\Gamma(z)} \frac{dz}{z-v'},$$

where

$$h(z) := \lambda^2 J(x(\theta)/\lambda)z - \frac{\lambda^2}{2}\psi_1(z) - \lambda x(\theta)\psi(z) = \lambda^2/2\left[(\psi_2(\theta)z - \psi_1(z)) - \frac{\psi_3(\theta)}{\psi_2(\theta)}(\psi_1(\theta)z - \psi(z))\right].$$

The definitions of $x(\theta)$, $\sigma(\theta)$ and $J(x)$ in (3.12), (3.11) are tailored precisely so that

$$h'(\theta) = h''(\theta) = 0.$$

This will allow us to perform a critical point analysis at θ. Recall (3.12) and note that $\frac{1}{2}\psi_2(\theta) + x\psi_1(\theta)$, is maximized at $x(\theta)/\lambda$, so that we may alternatively define $J(x(\theta)/\lambda)$ by

$$J(x(\theta)/\lambda) = \frac{1}{2}\psi_2(\theta) + \frac{x(\theta)}{\lambda}\psi_1(\theta).$$

Then

$$\begin{aligned} h'(z) &= J(x(\theta)/\lambda) - \frac{\lambda^2}{2}\psi_2(z) - \lambda x(\theta)\psi_1(z), \\ h''(z) &= -\frac{\lambda^2}{2}\psi_3(z) - \lambda x(\theta)\psi_2(z), \end{aligned}$$

and one can immediately check that $h'(\theta) = h''(\theta) = 0$. We also have

$$h'''(z) = -\frac{\lambda^2}{2}\psi_4(z) - \lambda x(\theta)\psi_3(z) = -\frac{\lambda^2}{2}\left(\psi_4(z) - \frac{\psi_3(z)}{\psi_2(z)}\psi_3(z)\right),$$

which means that $2\sigma(\theta)^3 = h'''(\theta)$. To control the sign of $h'''(\theta)$, we need the following lemma.

Lemma 3.2.4. *For any $z > 0$,*

$$\psi_m(z)^2 < \psi_{m+1}(z)\psi_{m-1}(z).$$

Proof. We adapt the proof of [23, Lemma 5.3]. The integral representation for polygamma functions gives

$$\begin{aligned} \psi_m(z)^2 &= \int_0^\infty \int_0^\infty \frac{e^{-zt-zu}}{(1-e^{-t})(1-e^{-u})} u^m t^m du dt, \\ \psi_{m-1}(z)\psi_{m+1}(z) &= \int_0^\infty \int_0^\infty \frac{e^{-zt-zu}}{(1-e^{-t})(1-e^{-u})} u^{m-1} t^{m+1} du dt. \end{aligned}$$

Symmetrizing the second formula in u and t gives

$$\psi_{m-1}(z)\psi_{m+1}(z) = \int_0^\infty \int_0^\infty \frac{e^{-zt-zu}}{(1-e^{-t})(1-e^{-u})} u^{m-1}t^{m-1}\frac{u^2+t^2}{2}dudt.$$

comparing the integrands and using $ab \leq \frac{a^2+b^2}{2}$ gives the result. □

Lemma 3.2.5. *For all $\theta > 0$, $h'''(\theta) > 0$.*

Proof. We have $\psi_2(\theta) < 0$ for all $\theta > 0$, this reduces the positivity of $h'''(\theta)$ to the fact that $\psi_4(z)\psi_2(z) > \psi_3(z)^2$, which follows from Lemma 3.2.4. □

3.2.2 Outline of the steep descent argument

Before going further we provide a brief outline of the steep descent argument that the rest of this section will make precise. In this outline we will only describe pointwise convergence of the integrand of K_{u_t} to that of K_{Ai} without justifying convergence for the Kernel itself or for the Fredholm determinant. We will also ignore the contours of the Fredholm determinant $\det(I - K_{u_t})_{L^2(C)}$ and of the integral which defines the kernels. Consider

$$K_{u_t(y)}(v,v') = \frac{1}{2\pi\mathbf{i}}\int_{\mathcal{D}} \frac{\pi}{\sin(\pi(z-v))} e^{t(h(z)-h(v))-t^{1/3}\sigma(\theta)y(z-v)}\frac{\Gamma(v)}{\Gamma(z)}\frac{dz}{z-v'},$$

and assume that we can deform the contours C and $\mathcal{D}$ to $\mathscr{C}$ and $\mathscr{D}$ respectively so they pass through θ at appropriate angles. Perform the change of variables $v = \theta + \sigma(\theta)^{-1}t^{-1/3}\tilde{v}, v' = \theta + \sigma(\theta)^{-1}t^{-1/3}\tilde{v}', z = \theta + \sigma(\theta)^{-1}t^{-1/3}\tilde{z}$. We know that h has a double critical point at θ, as $h'(\theta) = h''(\theta) = 0$ so we Taylor expand and use the large t approximations

$$h(\theta + t^{-1/3}\omega z) \to h(\theta) + \frac{\tilde{z}^3}{3}, \qquad \frac{t^{-1/3}\pi}{\sin(\pi(z-v))} \to \frac{1}{\tilde{z}-\tilde{v}}, \qquad \frac{\Gamma(v)}{\Gamma(z)} \to 1.$$

So our kernel becomes

$$K_{(y)}(\tilde{v},\tilde{v}') = \frac{1}{2\pi\mathbf{i}}\int_{e^{-\frac{\pi\mathbf{i}}{3}}\infty}^{e^{\frac{\pi\mathbf{i}}{3}}\infty} \frac{e^{\tilde{z}^3/3-y\tilde{z}}}{e^{\tilde{v}^3/3-y\tilde{v}}}\frac{d\tilde{z}}{(\tilde{z}-\tilde{v})(\tilde{z}-\tilde{v}')}.$$

The Fredholm determinant of this kernel is then reformulated as the Fredholm determinant of the Airy kernel on $L^2(\mathbb{R})$ using the identity $\det(1 + AB) = \det(1 + BA)$ in Lemma 3.2.17.

This completes the brief formal critical point analysis. The main technical challenge is finding contours $\mathscr{C}$ and $\mathscr{D}$ such that the integrals along these contours have (asymptotically as $t \to \infty$) all of their mass near θ (see Section 3.2.3 and Section 3.3). This is made more difficult in our case because h is a function with infinitely many poles and it is difficult to explicitly enumerate its critical points. Once such contours are found, a careful argument is necessary to produce the bounds needed to apply dominated convergence to the integral over $\mathscr{D}$ and to the Fredholm determinant expansion (see Section 3.2.4 and Appendix B).

3.2.3 Steep descent contours

In order to perform our asymptotic analysis on $\det(I - K_{u_t(y)})_{L^2(C)}$, we need to find contours, such that the real part of h (and therefore the norm of the integrand of $K_{u_t(y)}(v, v')$) can be bounded above. In this section we find such contours for the z variable. The contour for the v, v' variables is more elaborate and will be constructed in Section 3.3.

Without loss of generality we may restrict our attention to $\lambda = 1$ in most of the remainder of the chapter due to the fact that $h(z)/\lambda^2$ does not depend on λ.

Lemma 3.2.6. *The curve $\mathcal{D} = \theta + \mathbf{i}\mathbb{R}$ is steep descent at the point θ with respect to the function $h(z)$. In other words $\partial_y \Re\mathfrak{e}[\theta + \mathbf{i}y] < 0$ for $y > 0$ and $\partial_y \Re\mathfrak{e}[\theta + \mathbf{i}y] > 0$ for $y < 0$.*

Proof. By definition,

$$\begin{aligned} h(z) &= \lambda^2/2 \left[(\psi_2(\theta)z - \psi_1(z)) - \frac{\psi_3(\theta)}{\psi_2(\theta)}(\psi_1(\theta)z - \psi(z)) \right] \\ &= \lambda^2/2 \left[(\psi_2(\theta) - \frac{\psi_3(\theta)}{\psi_2(\theta)}\psi_1(\theta))z - (\psi_1(z) - \frac{\psi_3(\theta)}{\psi_2(\theta)}\psi(z)) \right], \end{aligned}$$

and

$$h'(z) = \lambda^2/2 \left[(\psi_2(\theta) - \frac{\psi_3(\theta)}{\psi_2(\theta)}\psi_1(\theta)) - (\psi_2(z) - \frac{\psi_3(\theta)}{\psi_2(\theta)}\psi_1(z)) \right].$$

Note that $\partial_y \Re\mathfrak{e}[h(\theta + \mathbf{i}y)] = -\Im\mathfrak{m}[h'(\theta + \mathbf{i}y)]$. $(\psi_2(\theta) - \frac{\psi_3(\theta)}{\psi_2(\theta)}\psi_1(\theta))$ is a positive real by Lemma 3.2.4, and $-\psi_2(\theta)$ is positive, so we have

$$A := -\psi_2(\theta)\Im\mathfrak{m}[h'(\theta + \mathbf{i}y)] = \Im\mathfrak{m}[\psi_2(\theta)\psi_2(\theta + \mathbf{i}y) - \psi_3(\theta)\psi_1(\theta + \mathbf{i}y)] > 0 \qquad \text{for } y > 0,$$
$$< 0 \qquad \text{for } y < 0.$$

These two statements are equivalent because the function is odd in y. Below we assume $y > 0$. For $n \geq 1$, we will use the Polygamma series expansion (A.1). First we note that

$$\Im\mathfrak{m}[\psi_2(\theta + iy)] = -2\sum_{k=0}^{\infty} \frac{-3(t+k)^2 y + y^3}{((t+k)^2 + y^2)^3},$$
$$\Im\mathfrak{m}[\psi_1(\theta + iy)] = \sum_{k=0}^{\infty} \frac{-2(t+k)y}{((t+k)^2 + y^2)^2}.$$

Using the series expansion,

$$\begin{aligned} A &= 4\sum_{m,n=0}^{\infty} \frac{1}{(n+\theta)^3} \frac{-3(m+\theta)^2 y + y^3}{((m+\theta)^2 + y^2)^3} - 6\sum_{m,n=0}^{\infty} \frac{1}{(n+\theta)^4} \frac{-2(m+\theta)y}{((m+\theta)^2 + y^2)^2} \\ &= \sum_{m,n=0}^{\infty} \frac{1}{(n+\theta)^3} \frac{-12(m+\theta)^2 y + 4y^3}{((m+\theta)^2 + y^2)^3} + \frac{1}{(n+\theta)^4} \frac{12(m+\theta)y}{((m+\theta)^2 + y^2)^2} \\ &\geq \sum_{m,n=0}^{\infty} \frac{1}{(n+\theta)^3} \frac{-12(m+\theta)^2 y}{((m+\theta)^2 + y^2)^3} + \frac{1}{(n+\theta)^4} \frac{12(m+\theta)y}{((m+\theta)^2 + y^2)^2} = B. \end{aligned}$$

We will show that $B > 0$. set

$$T_{n,m} = \frac{1}{(n+\theta)^3} \frac{-12(m+\theta)^2 y}{((m+\theta)^2 + y^2)^3} + \frac{1}{(n+\theta)^4} \frac{12(m+\theta)y}{((m+\theta)^2 + y^2)^2},$$

so $B = \sum_{n,m=0}^{\infty} T_{n,m}$. We will prove the following claims for arbitrary $y > 0$ and $\theta > 0$:

1. For $0 \leq n \leq m$, $T_{n,m} > 0$.
2. For $0 \leq n \leq m$, then either $\frac{T_{n,m}}{T_{m,n}}$ is positive, or $\left|\frac{T_{n,m}}{T_{m,n}}\right| \geq 1$.

Together these claims imply that if $n \leq m$, then $T_{n,m} + T_{m,n} > 0$, thus B is positive.

In the following two arguments we assume $0 \leq n \leq m$.

- Proof of claim (1): Let

$$a = \frac{1}{(n+\theta)^4} \frac{12(m+\theta)y}{((m+\theta)^2+y^2)^2}, \quad b = \frac{1}{(n+\theta)^3} \frac{-12(m+\theta)^2 y}{((m+\theta)^2+y^2)^3},$$

so that $T_{n,m} = a + b$. a is positive and b is negative, so we need only show that $\left|\frac{a}{b}\right| > 1$. We have

$$\left|\frac{a}{b}\right| = \frac{(m+\theta)^2+y^2}{(m+\theta)(n+\theta)} > \frac{(m+\theta)^2}{(m+\theta)(n+\theta)} = \frac{m+\theta}{n+\theta} \geq 1.$$

which is true because we made the hypothesis that $n \leqslant m$.

- Proof of claim (2): Setting $m = n + k$ for $k \geq 0$ and simplifying gives

$$\frac{T_{n,m}}{T_{m,n}} = \frac{(n+k+\theta)^5((n+\theta)^2+y^2)^3(k(n+k+\theta)+y^2))}{(n+\theta)^5((n+k+\theta)^2+y^2)^3(-k(n+\theta)+y^2)}. \tag{3.23}$$

Note that

$$\frac{(n+k+\theta)^2}{(n+k)^2} \geq \frac{(n+k+\theta)^2+y^2}{(n+k)^2+y^2}. \tag{3.24}$$

In the case that $-k(n+\theta)+y^2 > 0$, $\frac{T_{n,m}}{T_{m,n}}$ is positive so there is nothing to show. If $-k(n+\theta)+y^2 \leq 0$, then we have

$$\left|-\frac{(k(n+k+\theta)+y^2)}{(k(n+\theta)-y^2)}\right| = \frac{(n+k+\theta)+y^2/k}{(n+\theta)-y^2/k} \geq \frac{(n+k+\theta)}{(n+\theta)}. \tag{3.25}$$

Then (3.23) and (3.25) give

$$\left|\frac{T_{n,m}}{T_{m,n}}\right| \geq \frac{(n+k+\theta)^6}{(n+\theta)^6} \frac{((n+\theta)^2+y^2)^3}{((n+k+\theta)^2+y^2)^3} \geq 1.$$

where the last inequality follows from (3.24). This completes the proof.

□

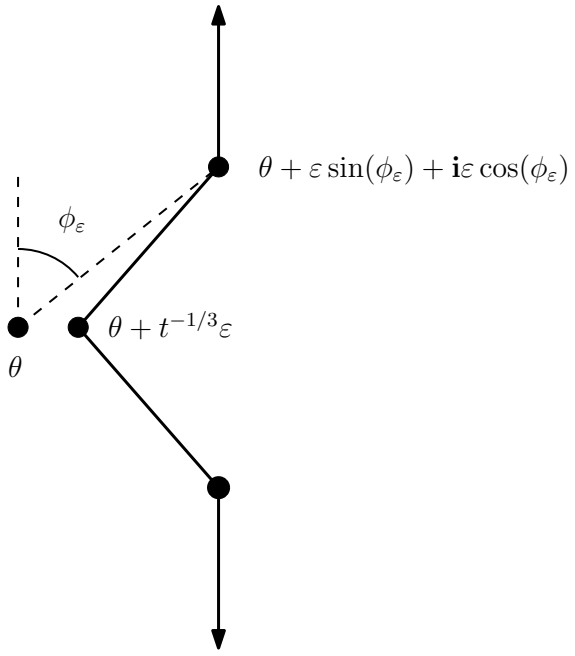

Figure 3.4: The contour $D_\varepsilon(\phi_\varepsilon)$ is shown in bold, and is oriented in the $+\mathbf{i}$ direction. The dotted line connecting θ and $\theta + \varepsilon \sin(\phi_\varepsilon) + \mathbf{i}\varepsilon \cos(\phi_\varepsilon)$ has length ε.

Lemma 3.2.6 will allow us to show that as $t \to \infty$, the kernel $K_{u_t(y)}(v, v')$, which is defined as an integral over $\theta + \mathbf{i}\mathbb{R}$ is the same as the limit as $t \to \infty$ of the same integral restricted to $[\theta - \mathbf{i}\varepsilon, \theta + \mathbf{i}\varepsilon]$. This is formalized in Lemma 3.2.15.

We will actually use a slight deformation of the contour $\mathcal{D}$.

Definition 3.2.7. In the following ε is positive, and ϕ_ε is a small positive angle. Let $\mathcal{D}_\varepsilon(\phi_\varepsilon)$ be the the union of the diagonal line segment $[\theta + t^{-1/3}\varepsilon, \theta + \varepsilon e^{\mathbf{i}(\pi - \phi_\varepsilon)})$, and the vertical line $[\theta + \varepsilon \sin(\phi_\varepsilon) + \mathbf{i} \cos(\phi_\varepsilon), \theta + \varepsilon \sin(\phi_\varepsilon) + \mathbf{i}\infty)$ along with both their reflections over the real axis, directed from $-\mathbf{i}\infty$ to $\mathbf{i}\infty$. See Figure 3.4.

Lemma 3.2.8. *For sufficiently small ε and ϕ_ε, there is an $\eta > 0$ such that for any $z \in \mathcal{D}_{\varepsilon,t}(\phi_\varepsilon) \setminus \mathcal{D}^{\varepsilon}_{\varepsilon,t}(\phi_\varepsilon)$, $\mathfrak{Re}[h(z) - h(\theta)] < -\eta$.*

Proof of Lemma 3.2.8. Because $h(\omega z) = \omega h(z)$, it is enough to prove the result in the upper half plane. The idea of this argument is that because h is a holomorphic function in a neighborhood of the contour $\mathcal{D}$, Taylor expanding and choosing ε small allows us to bound the difference between h' on $\mathcal{D}$ and h' on $\mathcal{D}_{\varepsilon,t}(\phi_\varepsilon)$ in a large bounded set. We control the difference outside this large

ball around 0 using a rigorous version of Stirling's approximation to control $h'(z)$ for $|\Im\mathfrak{m}[z]|$ very large.

First we control $h'(z)$ for large $|\Im\mathfrak{m}[z]|$. As $y \to +\infty$, the only term of $h(\theta + \mathbf{i}y)$ that does not go to 0 is the term containing $\psi(\theta+\mathbf{i}y)$. Lemma A.0.3 allows us to approximate $\psi(\theta+\mathbf{i}y)$ and gives $h(\theta + \mathbf{i}y) \sim -c\psi(\theta + \mathbf{i}y) \sim -c\log(\theta + \mathbf{i}y)$, where $c = \frac{-\psi_3(\theta)}{\psi_2(\theta)}$ is positive. Thus as $\Im\mathfrak{m}[z] \to +\infty$, $h(z) \to -\infty$ uniformly for $Re[z]$ in a compact set. Thus there is a large M such that for $\Im\mathfrak{m}[z] > M$, $z \in \mathcal{D}_{\varepsilon,t}(\phi_\varepsilon) \setminus \mathcal{D}^{\varepsilon}_{\varepsilon,t}(\phi_\varepsilon)$, $\Re\mathfrak{e}[h(z) - h(\theta)] < -\eta$.

Now we will control h' on a bounded set. By Lemma 3.2.6 $\partial_y \Re\mathfrak{e}[h(\theta + \mathbf{i}y)] < 0$ for $y > 0$. Thus for some large M, on the compact set $y \in [\cos(\phi_\varepsilon)\varepsilon, M]$, $\partial_y \Re\mathfrak{e}[h(\theta + \mathbf{i}y)]$ has some negative minimum. $h''(z)$ is analytic in the compact rectangle with corners $\theta + \mathbf{i}\varepsilon\cos(\phi_\varepsilon)$, $\theta + \varepsilon e^{\mathbf{i}\phi}$, $\theta + \mathbf{i}M$, $\theta+\varepsilon\sin(\phi_\varepsilon)+\mathbf{i}M$. Thus $|h''(z)|$ is bounded above by some R in this rectangle. Note that R depends only on $\varepsilon\cos(\phi_\varepsilon)$ and M, and R is increasing in $\cos(\phi_\varepsilon)$. We can choose ε and ϕ_ε so that $\varepsilon\cos(\phi_\varepsilon)$ remains fixed, and $\varepsilon\sin(\phi_\varepsilon)$ becomes arbitrarily small. Choosing so that $\varepsilon\sin(\phi_\varepsilon)R < \eta$ guarantees that $\partial_y Re[h(\theta + \sin(\phi_\varepsilon)\varepsilon + \mathbf{i}y)] > 0$ for $y \in [\cos(\phi_\varepsilon)\varepsilon, M]$. Because R is increasing ϕ_ε any smaller choice of $\phi_\varepsilon > 0$ also works.

Similarly by analyticity of h, we can uniformly bound $h'(z)$ on the line segment $[\theta+\mathbf{i}\varepsilon\cos(\phi_\varepsilon), \theta+\varepsilon\sin(\phi_\varepsilon) + \mathbf{i}\cos(\phi_\varepsilon)]$, and by Lemma 3.2.6 we know that $\Re\mathfrak{e}[h(\theta + \mathbf{i}\varepsilon\cos(\phi_\varepsilon)) - h(\theta)] < 0$. Thus for small enough $\varepsilon\sin(\phi_\varepsilon)$, $\Re\mathfrak{e}[h(\theta + \varepsilon\sin(\phi_\varepsilon) + \mathbf{i}\varepsilon\cos(\phi_\varepsilon)) - h(\theta)] < -\eta$. Again for a particular choice of $\varepsilon, \phi_\varepsilon$, any smaller ϕ_ε also works. □

Note that the kernel $K_{u_t(y)}$ is equal to

$$K_{u_t(y)} = \frac{1}{2\pi\mathbf{i}} \int_{\mathcal{D}_\varepsilon(\phi_\varepsilon)} \frac{\pi}{\sin(\pi(z-v))} e^{t(h(z)-h(v))-t^{1/3}\sigma(\theta)y(z-v)} \frac{\Gamma(v)}{\Gamma(z)} \frac{dz}{z-v'} \tag{3.26}$$

by Cauchy's theorem and the decay of the integrand as $\Im\mathfrak{m}[z] \to \pm\infty$.

Proposition 3.2.9. *There exists a closed contour C passing through θ and 0 , such that for any*

$\varepsilon > 0$, there exists $\eta > 0$, such that for all $v \in C \setminus B_\varepsilon(\theta)$,

$$\Re\mathfrak{e}[h(\theta) - h(v)] < \eta.$$

The proof of Proposition 3.2.9 requires a detailed understanding of the level set $\Re\mathfrak{e}[h(z)] = h(\theta)$. We will defer this proof to Section 3.3.

In the limit $\lim_{t\to\infty} \det(I - K_{u_t(y)})_{L^2(C)}$, Proposition 3.2.9 will allow us to restrict all contour integrals over C in the Fredholm determinant expansion to integrals over $C \cap B_\varepsilon(\theta)$.

3.2.4 Localizing the integrals

We perform the change of variables $v = \theta + t^{-1/3}\omega v, v' = \theta + t^{-1/3}\omega v', z = \theta + t^{-1/3}\omega z$. For every complex contour $\mathcal{M}$ we will define $\omega\mathcal{M} = \{z : \theta + t^{-1/3}z \in \mathcal{M}\}$. We will also define the kernel ωK_{u_t} by

$$\omega K_{u_t}(\omega v, \omega v') = t^{-1/3} K_{u_t}(\theta + t^{-1/3}v, \theta + t^{-1/3}v'),$$

so that

$$\det(I - \omega K_{u_t})_{L^2(\omega\mathcal{M})} = \det(I - K_{u_t})_{L^2(\mathcal{M})}.$$

For any contour $\mathcal{M}$ we define $\mathcal{M}^\varepsilon$ to be $\mathcal{M} \cap B_\varepsilon(\theta)$. Let $K^\varepsilon_{u_t(y)}(v, v')$ be defined as the right hand side of (3.26) with the contour of integration $\mathcal{D}_\varepsilon(\phi_\varepsilon)$ replaced by the cut off contour $\mathcal{D}_\varepsilon(\phi_\varepsilon)^\varepsilon$.

In this section we will use our control of the norm of the integrand of $K_{u_t(y)}(v, v')$ to show that

$$\lim_{t\to\infty} \det(I - K_{u_t(y)}(v, v'))_{L^2(C)} = \lim_{t\to\infty} \det(I - K^\varepsilon_{u_t(y)}(v, v'))_{L^2(C^\varepsilon)}.$$

In this and the next section we will need several bounds in order to apply dominated convergence to the kernel $K_{u_t}(v, v')$ and the Fredholm determinant expansion $\det(I - K_{u_t})_{L^2(C)}$. We give these bounds now, but defer most of their proofs to Appendix B.

Lemma 3.2.10. *For ε sufficiently small, t sufficiently large, and $v, v' \in C \setminus C^\varepsilon$, there are constants*

$R_2, \eta > 0$ depending on ε such that

$$|K_{u_t}(v, v')| \leq R_2 e^{-t\eta/4}. \tag{3.27}$$

For ε sufficiently small, and t sufficiently large, $v \in C \setminus C^\varepsilon$, $v' \in C$, for the same constants R_2 and η, we have

$$|\omega K_{u_t}(v, v')| \leq R_2 e^{-t\eta/4}. \tag{3.28}$$

This property of the contour C stated in Proposition 3.2.9 is the main tool necessary to prove Lemma 3.2.10. We defer the proof of Lemma 3.2.10 to Appendix B.

Lemma 3.2.11. *For $t > 1$, and for all sufficiently small $\varepsilon > 0$, there exists a constant $C_1 > 0$ such that for $v \in C^\varepsilon$ and $z \in \mathcal{D}^\varepsilon_{\varepsilon,t}(\phi_\varepsilon)$, the integrand of $\omega K^\varepsilon_{u_t}(\omega v, \omega v')$ is bounded above by a positive function of $\omega z, \omega v, \omega v'$ which does not depend on t and whose integral over $\mathcal{D}^\varepsilon_{\varepsilon,t}(\phi_\varepsilon)$ is finite. We also have*

$$\omega K^\varepsilon_{u_t}(\omega v, \omega v') \leq C_1 e^{-t\frac{h'''(\theta)}{24}\omega v^3}.$$

Lemma 3.2.12. *For all sufficiently small ε, for $v, v' \in C^\varepsilon$,*

$$\lim_{t\to\infty}(K_{u_t}(v, v') - K^\varepsilon_{u_t}(v, v')) \to 0.$$

The property of the contour $\mathcal{D}^\varepsilon_{\varepsilon,t}(\phi_\varepsilon)$ stated in Lemma 3.2.8 is the main tool in the proofs of Lemma 3.2.11 and Lemma 3.2.12. We will defer the proofs to Appendix B.

Lemma 3.2.13. *For sufficiently small ε and $t > 1$, there exists a function $\omega H_m(\omega v, \omega v')$ not depending on t such that for all $v \in C^\varepsilon$, $\omega H_m(\omega v, \omega v') \geq \left|\det(\omega K^\varepsilon_{u_t}(\omega v_i, \omega v_j)^m_{i,j=1})\right|$ and $\omega H_m(\omega v, \omega v') \geq \left|\det(\omega K_{u_t}(\omega v_i, \omega v_j)^m_{i,j=1})\right|$, and*

$$1 + \sum_{m=1}^{\infty} \frac{1}{m!} \int_{(\omega C^\varepsilon)^m} \omega H_m(\omega v, \omega v') \leq 1 + \sum_{m=1}^{\infty} \frac{1}{m!} \int_{(C_0)^m} \omega H_m(\omega v, \omega v') < \infty.$$

The Proof of Lemma 3.2.13 uses Lemma 3.2.11 and Lemma 3.2.12. We defer the proof to

Appendix B.

Lemma 3.2.14. *For any $t > 0$ and ε sufficiently small,*

$$\lim_{t\to\infty} \det(I - K_{u_t})_{L^2(C)} = \lim_{t\to\infty} \det(I - K_{u_t})_{L^2(C^\varepsilon)}.$$

Proof.

$$\begin{aligned}\det(I - K_{u_t})_{L^2(C)} - \det(I - K_{u_t})_{L^2(C^\varepsilon)} &= \sum_{m=1}^{\infty} \frac{1}{m!} \int_{C^m \setminus (C^\varepsilon)^m} \det\left(K_{u_t}(v_i, v_j)\right)_{i,j=1}^m \prod_{i=1}^m dv_i \\ &\leq \sum_{m=1}^{\infty} \frac{1}{m!} \int_{C^m \setminus (C^\varepsilon)^m} \left|\det\left(K_{u_t}(v_i, v_j)\right)_{i,j=1}^m\right| \prod_{i=1}^m dv_i. \quad (3.29)\end{aligned}$$

By Lemma 3.2.10, for $v_i \in C \setminus C^\varepsilon$,

$$\omega K_{u_t}(\omega v_i, \omega v_j) \leq R_2 e^{-t\eta/4}.$$

By similar reasoning we can allow $v_j \in C \setminus C^\varepsilon$ without changing the bounds provided by Lemma 3.2.12 and 3.2.11. Thus for $v_i \in C^\varepsilon$, $v_j \in C$, we have

$$\omega K_{u_t}(\omega v_i, \omega v_j) \leq C_1 e^{-t\frac{h'''(\theta)}{24}\omega v^3} + \eta \leq C_1 + \varepsilon.$$

Set $R_3 = \max[R_2, C_1 + \varepsilon]$. Then for all $v_i, v_j \in C$,

$$\omega K_{u_t}(\omega v_i, \omega v_j) \leq R_3.$$

Using Hadamard's bound with respect to the rows of $|\det(\omega K_{u_t}(\omega v_i, \omega v_j))_{i,j=1}^m|$ with $\omega v_1 \in \omega C \setminus \omega C^\varepsilon$, and $\omega v_j \in \omega C$ for all $j > 1$ we obtain

$$|\det(\omega K_{u_t}(\omega v_i, \omega v_j))_{i,j=1}^m| \leq m^{m/2} R_3^m e^{-t\eta/4}. \quad (3.30)$$

Indeed, because $|\det\left(K_{u_t}(v_i,v_j)_{i,j=1}^m\right)|$ is positive, and unchanged by permuting the $v_1,...,v_m$, we have

$$\begin{aligned}\int_{C^m\setminus(C^\varepsilon)^m}\left|\det\left(K_{u_t}(v_i,v_j)\right)_{i,j=1}^m\right|\prod_{i=1}^m dv_i &\le \int_{C\setminus C^\varepsilon}\left(\int_{C^{m-1}}\left|\det\left(K_{u_t}(v_i,v_j)\right)_{i,j=1}^m\right|\prod_{i=1}^{m-1}dv_i\right)dv_1\\ &\le \int_{\omega C\setminus\omega C^\varepsilon}\left(\int_{\omega C^{m-1}}\left|\det\left(\omega K_{u_t}(\omega v_i,\omega v_j)\right)_{i,j=1}^m\right|\prod_{i=1}^{m-1}d\omega v_i\right)d\omega v_1\\ &\le \int_{\omega C\setminus\omega C^\varepsilon}\left(\int_{\omega C^{m-1}} m^{m/2}R_3^m e^{-t\eta/4}\prod_{i=1}^{m-1}d\omega v_i\right)d\omega v_1\\ &\le m^{m/2}(t^{1/3}LR_3)^m e^{-t\eta/4}. \end{aligned}\tag{3.31}$$

In the first inequality we are strictly increasing the set on which we are integrating. In the second inequality we have changed variables from v_i to ωv_i. In the third inequality we have used (3.30). And in the last equality we have used that C has a finite length L, so ωC has length $t^{1/3}L$.

Thus

$$\begin{aligned}\sum_{m=1}^\infty\frac{1}{m!}\int_{C^m\setminus(C^\varepsilon)^m}\left|\det\left(K_{u_t}(v_i,v_j)\right)_{i,j=1}^m\right|\prod_{i=1}^m dv_i &\le \sum_{m=1}^\infty m^{m/2}(t^{1/3}LR_3)^m e^{-t\eta/4}\\ &\le e^{-t\eta/4}\sum_{m=1}^\infty m^{1+m/2}(t^{1/3}LR_3)^m \le e^{-t\eta/4}(16t^{1/3}LR_3)^4 e^{2t^{2/3}(LR_3)^2}\to 0.\end{aligned}\tag{3.32}$$

In the first inequality we used (3.31). In the second inequality we multiplied each term of the sum by m. In the third inequality, we use [27, Lemma 4.4] with $C = (t^{1/3}LR_3)$. Together (3.29) and (3.32) complete the proof. □

Lemma 3.2.15. *For $t > 0$ and ε sufficiently small,*

$$\lim_{t\to\infty}\det(I-K_{u_t})_{L^2(C^\varepsilon)} = \lim_{t\to\infty}\det(I-K_{u_t}^\varepsilon)_{L^2(C^\varepsilon)}.$$

Proof. First use Lemma 3.2.12 to obtain $\lim_{t\to\infty}K_{u_t}^\varepsilon(v,v') = \lim_{t\to\infty}K_{u_t}(v,v')$, then Lemma 3.2.13 allows us to apply dominated convergence to the Fredholm determinant expansion. □

3.2.5 Convergence to Tracy-Widom GUE distribution

Now we conclude the proof of Theorem 3.1.15 by identifying the limit of the Fredholm determinant over localized contours from the previous section with the Fredholm determinant expansion of $F_{\mathrm{GUE}}(x)$.

Proposition 3.2.16. *For $t > 0$ and ε sufficiently small,*

$$\lim_{t\to\infty} \det(I - K^{\varepsilon}_{u_t})_{L^2(C^{\varepsilon})} = \det(I - K_{(y)})_{L^2(C_0)}$$

where

$$K_{(y)}(u,u') = \frac{1}{2\pi\mathbf{i}} \int_{\mathcal{D}_0} \frac{e^{s^3/3-ys}}{e^{u^3/3-yu}} \frac{ds}{(s-u)(s-u')},$$

and the contours are defined as

$$\mathcal{D}_0 = (e^{-\pi\mathbf{i}/3}\infty, 1) \cup [1, e^{\pi\mathbf{i}/3}\infty), \;\; C_0 = (e^{-2\pi\mathbf{i}/3}\infty, 0) \cup [0, e^{2\pi\mathbf{i}/3}\infty).$$

Proof. First recall that $\det(I - K^{\varepsilon}_{u_t})_{L^2(C^{\varepsilon})} = \det(I - \omega K^{\varepsilon}_{u_t})_{L^2(\omega C^{\varepsilon})}$. We have the following pointwise limits in $\omega v, \omega v', \omega z$:

$$\frac{t^{-1/3}\pi}{\sin(\pi(t^{-1/3}(\omega z - \omega v))} \xrightarrow{t\to\infty} \frac{1}{\omega z - \omega v}, \tag{3.33}$$

$$e^{t(h(z)-h(v))} \xrightarrow{t\to\infty} e^{\frac{h'''(\theta)}{6}\omega z^3 - \frac{h'''(\theta)}{6}\omega v^3}, \tag{3.34}$$

$$\frac{\Gamma(\theta + t^{-1/3}\omega v)}{\Gamma(\theta + t^{-1/3}\omega z)} \xrightarrow{t\to\infty} 0. \tag{3.35}$$

Thus

$$\lim_{t\to\infty} \frac{t^{-1/3}\pi\Gamma(\theta + t^{-1/3}\omega v)e^{t(h(z)-h(v))-\sigma(\theta)y(\omega z-\omega v)}}{\sin(\pi(t^{-1/3}(\omega z - \omega v))\Gamma(\theta + t^{-1/3}\omega z)(\omega z - \omega v')} = \frac{e^{\frac{h'''(\theta)}{6}\omega z^3 - \sigma(\theta)y\omega z}}{e^{\frac{h'''(\theta)}{6}\omega v^3 - \sigma(\theta)y\omega v}} \frac{d\omega z}{(\omega z - \omega v)(\omega z - \omega v')}.$$

The left hand side is the integrand of $\omega K^{\varepsilon}_{u_t}(\omega v, \omega v')$. Lemma 3.2.11 allows us to use dominated

convergence to get

$$\lim_{t\to\infty} \omega K_{u_t}(\omega v, \omega v') = K'_{(y)}(\omega v, \omega v'), \tag{3.36}$$

where

$$K'_{(y)}(\omega v, \omega v') = \int_{\mathcal{D}_0(\phi_\varepsilon)} \frac{e^{\frac{h'''(\theta)}{6}\omega z^3 - \sigma(\theta) y \omega z}}{e^{\frac{h'''(\theta)}{6}\omega v^3 - \sigma(\theta) y \omega v}} \frac{d\omega z}{(\omega z - \omega v)(\omega z - \omega v')},$$

$$\mathcal{D}_0(\phi_\varepsilon) = (e^{(-\frac{\pi}{2}+\phi_\varepsilon)\mathbf{i}}\infty, \varepsilon) \cup [\varepsilon, e^{(\frac{\pi}{2}-\phi_\varepsilon)\mathbf{i}}\infty).$$

The real part of $\frac{h'''(\theta)}{6}\omega z^3$ is negative when $z = e^{i\phi}$ with $\phi \in [\frac{\pi}{2} - \phi_\varepsilon, \pi/3] \cup [-(\frac{\pi}{2} - \phi_\varepsilon, -\pi/3]$, so we can deform the contour $\mathcal{D}_0(\phi_\varepsilon)$ to the contour $\mathcal{D}_0$ without changing the value of $K'_{(y)}$. After performing this change of contour and the change of variables $s = \sigma(\theta)\omega z, u = \sigma(\theta)\omega v, u' = \sigma(\theta)\omega v'$, where $\sigma(\theta) = (h'''(\theta)/2)^{1/3}$, we have

$$K'_{(y)}(\omega v, \omega v') = \sigma(\theta) K_{(y)}(u, u'). \tag{3.37}$$

Note that

$$\det(I - \omega K^{\varepsilon}_{u_t})_{L^2(\omega C^\varepsilon)} = \det(I - \mathbb{1}_{\omega v \le t^{1/3}\varepsilon}(\omega v) K^{\varepsilon}_{u_t}(\omega v, \omega v') \mathbb{1}_{\omega v' \le t^{1/3}\varepsilon}(\omega v'))_{L^2(C_0)}. \tag{3.38}$$

By Lemma 3.2.13 we can apply dominated convergence to the Fredholm determinant expansion on the right hand side of (3.38). Along with (3.36) and (3.37) we have

$$\lim_{t\to\infty} \det(I - K^{\varepsilon}_{u_t})_{L^2(C^\varepsilon)} = \det(I - K_{(y)})_{L^2(C_0)}.$$

□

Lemma 3.2.17. *For all $y \in \mathbb{R}$,*

$$\det(I - K_{(y)})_{L^2(C_0)} = \det(I - K_{\mathrm{Ai}})_{L^2(y,+\infty)}.$$

where K_{Ai} is defined in Definition 3.2.1.

Proof. We apply [37, Lemma 8.6]. □

This reformulation is common in asymptotic analyses of Fredholm determinants. We are now able to conclude.

Proof of Proposition 3.2.3. Together Lemma 3.2.14, Lemma 3.2.15, Proposition 3.2.16, and Lemma 3.2.17 yield

$$\lim_{t\to\infty} \det(I - K_{u_t})_{L^2(C)} = \lim_{t\to\infty} \det(I - K_{u_t})_{L^2(C^\varepsilon)} = \lim_{t\to\infty} \det(I - K^{\varepsilon}_{u_t})_{L^2(C^\varepsilon)} = \det(I - K_{(y)})_{L^2(C_0)} = \det(I - K_{\mathrm{Ai}})_{L^2(y,+\infty)}. \quad (3.39)$$

□

The proof of Corollary 3.1.17 is almost identical to the argument used to obtain [23, Corollary 5.8] from [23, Theorem 1.15]. We include it here for completeness.

Proof of Corollary 3.1.17. Observe that we can sample e^{ct} uniform sticky Brownian motions $\{B_i(t)\}$ by first sampling the kernels $\mathsf{K}_{s,t}$ and then sampling e^{ct} iid continuous random walks with these kernels as transition probabilities. For any given kernel, the probability that none of the uniform sticky Brownian motions is greater than r is given by

$$\mathsf{P}\left(\max_{i=1,\dots,\lfloor e^{ct}\rfloor} B_i(t) \le r\right) = \left(1 - \mathsf{K}_{0,t}(0,[r,\infty))\right)^{\lfloor e^{ct}\rfloor} = \exp\left(\lfloor e^{ct}\rfloor \log(1 - \mathsf{K}_{0,t}(0,[r,\infty)))\right). \quad (3.40)$$

We set $r = xt = tx_0 + \frac{t^{1/3}\sigma(\theta_0)y}{\lambda^2 J'(x_0/\lambda)}$, and let θ_r be defined so that $x(\theta_r) = x$.. Because these motions are independent after conditioning on the environment,

$$\mathbb{P}\left(\max_{i=1,\dots,\lfloor e^{ct}\rfloor} B_i(t) \le r\right) = \mathbb{P}\left(\frac{\max_{i=1,\dots,\lfloor e^{ct}\rfloor} B_i(t) - tx_0}{t^{1/3}\sigma(\theta_0)/(\lambda^2 J'(x_0/\lambda))} \le y\right) \quad (3.41)$$

Use Theorem 3.1.15 to approximate

$$\log(\mathsf{K}_{0,t}(0,[r,\infty))) = -t\lambda^2 J(x/\lambda) + t^{1/3}\sigma(\theta_r)\chi_t, \quad (3.42)$$

with χ_t converging weakly a GUE Tracy-Widom distributed random variable as $t \to \infty$. Now Taylor expand

$$\lambda^2 J(x/\lambda) = \lambda^2 J(x_0/\lambda) + \sigma(\theta_0) t^{-2/3} y + O(t^{-\frac{4}{3}}).$$

We can take the derivative of $x(\theta)$ and apply Lemma 3.2.4 to see that x is a decreasing continuous surjective function of θ from $\mathbb{R}_{>0} \to \mathbb{R}_{>0}$. Thus we can define the inverse map $\theta(x)$ on $\mathbb{R}_{>0}$, and Taylor expand

$$\sigma(\theta_r) = \sigma(\theta_0) + \frac{t^{-2/3}\sigma'(\theta_0)\theta'(x_0)\sigma(\theta_0) y}{\lambda^2 J'(x_0/\lambda)} + O(t^{-\frac{4}{3}}).$$

We can now expand the right hand side of (3.42) as

$$-t\lambda^2 J(x_0/\lambda) + t^{1/3}\sigma(\theta)\chi_t = -t\lambda^2 J(x_0/\lambda) + \sigma(\theta_0) t^{1/3}(\chi_t - y) + O(t^{-1/3}) + O(t^{-1/3}\chi_t) \quad (3.43)$$

Choosing x_0 so that $\lambda^2 J(x_0/\lambda) = c$ gives

$$\mathbb{P}\left(\max_{i=1,\dots,\lfloor e^{ct}\rfloor} B_i(t) \le r\right) = \mathbb{E}\exp\left(\lfloor e^{ct}\rfloor \log(1 - \mathsf{K}_{0,t}(0,[r,\infty))\right) \quad (3.44)$$

$$= \mathbb{E}\exp\left(-\lfloor e^{ct}\rfloor \mathsf{K}_{0,t}(0,[r,\infty)) + O(e^{ct}\mathsf{K}_{0,t}(0,[r,\infty))^2)\right) \quad (3.45)$$

$$= \mathbb{E}\exp\left(-e^{t^{1/3}\sigma(\theta_0)(\chi_t - y) + O(t^{-1/3}(1+\chi_t)) + O(K_{0,t}(0,[r,\infty)) + O(e^{ct}\mathsf{K}_{0,t}(0,[r,\infty))^2)}\right) \quad (3.46)$$

The second equality is obtained by Taylor expanding the logarithm around 1. The third equality is obtained by combining (3.42) and (3.43).

Now we control the error terms. The random variable χ_t converges in distribution, so by Slutsky's theorem, $t^{-1/3}(1 + \chi_t) \to 0$ in probability. Recall that $\lambda^2 J(x_0/\lambda) = c$ to obtain

$$e^{ct} K_{0,t}(0,[r,\infty))^2 = e^{ct + 2\log K_{0,t}(0,[r,\infty))} = e^{-ct + O(t^{1/3}\chi_t)} = e^{-ct + t^{2/3}O(t^{-1/3}\chi_t)},$$

$$K_{0,t}(0,[r,\infty)) = e^{\log K_{0,t}(0,[r,\infty))} = e^{-ct + O(t^{1/3}\chi_t)} = e^{-c_t + t^{2/3}O(t^{-1/3}\chi_t)}.$$

Since $O(t^{-1/3}\chi_t) \to 0$ in probability, so do both $O(e^{ct}K_{0,t}(0,[r,\infty))^2)$ and $O(K_{0,t}(0,[r,\infty)))$. Combining this with (3.41) and the fact that $\exp(-e^{t^{1/3}x}) \xrightarrow[t\to\infty]{} \mathbb{1}_{x<0}$, and using bounded convergence completes the proof. □

3.3 Construction of steep descent contours

This section is devoted to constructing the contour C whose existence is stated in Proposition 3.2.9, and which is used in the asymptotic analysis of Section 2. The goal is first to study the level set $\mathfrak{Re}[-h(z)] = h(\theta)$, show that it contains well behaved paths from θ to 0 in the complex plane, and second to take the slight deformation $\mathfrak{Re}[-h(z)] = h(\theta) - \varepsilon$ and add small segments to a path in this set to arrive at a contour from θ to 0 on which we can bound $\mathfrak{Re}[-h(z)]$. The first step is the main difficulty.

Arguments of this type are often performed in cases where the function corresponding to our h is a rational function or the log of a rational function and thus has a finite explicit set of critical points and poles [39, 47, 36]. We will see that the infinite set of poles of h', and the fact that we do not explicitly know all zeros of h' both lead to challenges that we overcome through careful use of conservation of the number of paths in the level set of $\mathfrak{Re}[h]$ and $\mathfrak{Re}[h']$ which enter and leave a any compact set K.

Before studying the level sets, we will need some bounds. Rather than requiring very careful bounds on $\mathfrak{Re}[h(z)]$, we instead only find the sign of the derivative of $\mathfrak{Re}[h(z)]$ along the real and imaginary axis.

Lemma 3.3.1. *For all $y > 0$, $\mathfrak{Im}[\psi_2(iy)] < 0$.*

Proof. We split the proof into 2 cases. For case 1 assume $y > \frac{1}{\sqrt{5}}$. Applying Lemma A.0.1

$$\begin{aligned}\mathfrak{Im}[\psi_2(\mathbf{i}y)] &= -2\mathfrak{Im}\left[\frac{1}{2(\mathbf{i}y)^2} + \frac{1}{2(\mathbf{i}y^3)} + \frac{3}{6(\mathbf{i}y)^4} + R_m^3(\mathbf{i}y)\right] \\ &= -2\left(\frac{1}{2y^3} + \mathfrak{Im}[R_m^3(\mathbf{i}y)]\right) \leq \frac{-1}{y^3} + \frac{1}{5y^5}.\end{aligned}$$

Since $y > \frac{1}{\sqrt{5}}$, we have $\mathfrak{Im}[\psi_2(\mathbf{i}y)] < 0$. as desired.

For case 2 assume $y \leq \frac{1}{\sqrt{5}}$. Using the zeroth order Laurent expansion of ψ_2 around 0 gives

$$\mathfrak{Im}[\psi_2(\mathbf{i}y)] = \mathfrak{Im}\left[\frac{-2\mathbf{i}}{y^3} - 2\zeta(3) + R_2^0(\mathbf{i}y)\right] = \frac{-2}{y^3} + \mathfrak{Im}\left[R_2^0(\mathbf{i}y)\right] \leq \frac{-2}{y^3} + 3!\zeta(4)y.$$

Where $\zeta(\cdot)$ is the Riemann zeta function. $\frac{-2}{y^3} + 3!\zeta(4)y$ has the same sign as $\frac{-2}{y^4} + 3!\zeta(4)$, and when $y \leq \frac{1}{\sqrt{5}}$,

$$\frac{-2}{y^4} + 3!\zeta(4) \leq -50 + 6\zeta(4) < 0,$$

where we have used $\zeta(4) < 2$. Thus we have $\mathfrak{Im}[\psi_2(\mathbf{i}y)] \leq 0$. as well in case 2. □

Lemma 3.3.2. *We have* $\mathfrak{Im}[h'(\mathbf{i}y)] < 0$ *for* $y > 0$*, and* $\mathfrak{Im}[h'(\mathbf{i}y)] > 0$ *for* $y < 0$

Proof. Because $h'(\overline{z}) = \overline{h'(z)}$, the two statements in this lemma are equivalent; we will prove the first. Because $\psi_2(\theta) < 0$, this is equivalent to showing

$$A = A(\theta, y) = \mathfrak{Im}[\psi_2(\theta)\psi_2(\mathbf{i}y) - \psi_3(\theta)\psi_1(\mathbf{i}y)] > 0.$$

For $\theta > 0$, $\psi_3(\theta)$ is positive and $\mathfrak{Im}[\psi_1(\mathbf{i}y)]$ is negative, so the second term is positive. $\psi_2(\theta)$ is positive, and by Lemma 3.3.1 $\mathfrak{Im}[\psi_2(\mathbf{i}y)]$ is negative, so the first term is positive □

Let

$$p_\theta(a) = \psi_2(a) - \frac{\psi_3(\theta)}{\psi_2(\theta)}\psi_1(a). \tag{3.47}$$

So that $h'(a) = p(\theta) - p(a)$. We will often omit the θ subscript and simply write $p(a)$.

Lemma 3.3.3. *The function* p *satisfies* $p'(a) > 0$ *for all* $a < \theta$*, and* $p'(a) < 0$ *for all* $a > \theta$*.*

Proof. By Lemma 3.2.4 we have $\psi_4(\theta)\psi_2(\theta) - \psi_3(\theta)^2 > 0$ for all $\theta > 0$. After dividing by $\psi_2^2(\theta)$, we get

$$\partial_\theta \frac{\psi_3(\theta)}{\psi_2(\theta)} = \frac{1}{\psi_2(\theta)}\left(\psi_4(\theta) - \frac{\psi_3(\theta)^2}{\psi_2(\theta)}\right) > 0,$$

so $\frac{\psi_3(\theta)}{\psi_2(\theta)}$ is increasing in θ. This implies

$$\frac{\psi_3(a)}{\psi_2(a)} - \frac{\psi_3(\theta)}{\psi_2(\theta)} \quad \text{is} \quad \begin{cases} < 0 & \text{for } a < \theta, \\ > 0 & \text{for } a > \theta. \end{cases}$$

Multiplying by the negative term $\psi_2(a)$ gives

$$f'(a) = \psi_3(a) - \frac{\psi_3(\theta)}{\psi_2(\theta)}\psi_2(a) \quad \text{is} \quad \begin{cases} > 0 & \text{for } a < \theta, \\ < 0 & \text{for } a > \theta. \end{cases}$$

□

Lemma 3.3.4. *The function $a \mapsto \mathfrak{Re}[-h(a)]$ is increasing for $a < t$ and decreasing in a for $a > t$.*

Proof. $h'(a)$ and $h(a)$ are real for $a \in \mathbb{R}$, so $\partial_a \mathfrak{Re}[-h(a)] = -h'(a)$. From (3.47), we see that $-h'(a) = p(a) - p(\theta)$. Together with Claim 3.3.3 this gives that $h'(a)$ is negative for $a < t$ and positive for $a > t$. This completes the proof. □

3.3.1 Contour curves and Contour paths

Now using the sign of the derivatives of $\mathfrak{Re}[h(z)]$ along the real and imaginary axis, we begin a more careful study of the level sets of $\mathfrak{Re}[h(z)]$. First we introduce a helpful way to think about the level set of the real or imaginary part of an arbitrary meromorphic function by defining contour curves and contour paths.

Let f be a meromorphic function on the complex plane. Let $\gamma = \{z \in \mathbb{C} : \mathfrak{Im}[f(z)] = 0\}$. Then γ can be decomposed as a (potentially infinite) collection of differentiable curves which meet only at critical points and poles of f.

Definition 3.3.5. We will call a maximal connected subset of the level set γ that does not contain a critical point or a pole a *contour curve* of $\mathfrak{Im}[f(z)] = 0$. Contour curves will be differentiable paths with a critical point, a pole, or the point ∞ at either end. We assign an orientation to each

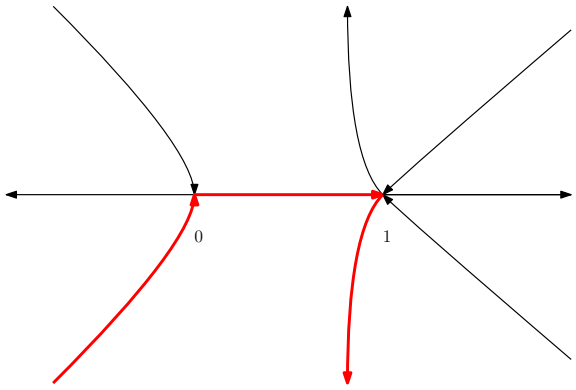

Figure 3.5: An image of the levelset $\Im\mathfrak{m}[f(z)] = 0$ for $f(z) = \frac{z^2}{2} - \frac{2z^3}{3} + \frac{z^4}{4}$. Because $f'(z) = z(z-1)^2$ we see a critical point at 0 and a double critical point at 1. On each contour curve we have drawn an arrow indicating the direction in which $\Re\mathfrak{e}[f(z)]$ is increasing. In thick red we show the contour path that starts at the point ∞, and exits its starting point along the contour curve in the lower half plane that connects ∞ to 0. As indicated in the Definition 3.3.1 at each critical point the next contour curve in the path is immediately counterclockwise to the previous contour curve.

contour curve so that $\Re\mathfrak{e}[f(z)]$ is an increasing function as we traverse the curve in the positive direction. We will say that a contour curve exits one of its endpoints and enters the other based on this orientation

We also define a notion of *contour path*, which connects a pole or the point infinity to another pole or to the point ∞ and on which $\Re\mathfrak{e}[f(z)]$ goes from $-\infty$ to ∞. To do this we need to make an arbitrary choice of what to do at critical points.

Definition 3.3.6. A contour path of $\Im\mathfrak{m}[f(z)] = 0$ is a subset of γ, which is also a path consisting of a union of contour curves and critical points constructed by the following procedure. Choose a pole or the point ∞ to be the starting point of the contourpath. Select one contour curve which exits the starting point. If this curve hits a critical point, then select the critical point and the contour curve leaving the critical point immediately counterclockwise to the previous contour curve. Repeat this step until you reach a pole or until you travel along a contour curve which is unbound in which case we say you reach the point ∞ (If a pole or ∞ is never reached then repeat this step infinitely many times). The contour path is the union of all contour curves and critical points selected by this procedure. See Figure 3.5.

Note that every contour path is a piecewise-differentiable path with endpoints either at a pole or

at the point ∞. Note also that if two contour-paths do not contain exactly the same set of contour-curves, then they have no contour-curves in common. This is true because each outgoing contour curve of a critical point has only one incoming curve immediately clockwise from it, and each incoming contour curve has only one outgoing curve immediately counter-clockwise from it.

We choose an orientation on the level set $\Im\mathfrak{m}[f(z)] = 0$ so that all contour curves and contour paths are directed so that $\Re\mathfrak{e}[f(z)]$ is an increasing function in the chosen direction. Such an orientation exists because we chose each contour path to exit a critical point along a contour curve neighboring the contour curve at which they entered the critical point.

With these definitions in place we would intuitively like to say that for any bounded set that does not contain a pole of f, the number of directed contour paths of γ entering the set is equal to the number of directed contour paths leaving the set. We give a more precise definition of "entering a set" then we state this conservation rigorously in Lemma 3.3.8.

Definition 3.3.7. We say the contour path $\gamma^i(t)$ (parametrized at unit speed in the positive direction) enters a set K at the point a if there is a t_a and $\varepsilon > 0$ such that for $t \in (t_a - \varepsilon, t_a]$, $\gamma_i(t) \notin \text{Int}(K)$, and for $t \in (t_a, t_a + \varepsilon)$, $\gamma_i(t) \in \text{Int}(K)$. We say a contour path $\gamma^i(t)$ exits K at the point b if there is a t_b and $\varepsilon > 0$ such that for $t \in (t_b - \varepsilon, t_b]$, $\gamma_i(t) \in K$, and for $t \in (t_b, t_b + \varepsilon)$, $\gamma_i(t) \notin K$. Let $[\gamma, K]_{\text{in}}$ be the multiset all of points at which a contour path in γ enters K (a point occurs n times in $[\gamma, K]_{\text{in}}$ if n contour paths enter at that point). Let $[\gamma, K]_{\text{out}}$ be the multiset of all points at which a contour path in γ exits K, similarly counted with multiplicity.

Lemma 3.3.8. *Let f be a meromorphic function, and let K be a connected compact set, so that no pole of f lies in K. If $[\gamma, K]_{in}$ consists of n points $a_1, ..., a_n$, then $[\gamma, K]_{out}$ consists of n points $b_1, ..., b_n$, so that there is a contour path in the set γ from a_i to b_i, and $\Re\mathfrak{e}[f(a_i)] \leq \Re\mathfrak{e}[f(b_i)]$ for all i. Note the a_is are not distinct if a critical point is on the boundary of K, and similarly for the b_i's.*

Proof. If K contains infinitely many critical points of f, then the derivative of f is 0, in which case the lemma is trivial.

Assume $K \cap \gamma$ has either no critical points, or 1 critical point of order r. At each critical point of order r there are r incoming contour curves of γ and r outgoing contour curves of γ.

Enumerate all contour paths γ_i entering K and pair them so that γ_i enters K at the point a_i. We define the parametrization of γ_i by $|\gamma_i'(t)| = \max[\frac{1}{(\partial_z \Re\mathfrak{e}[f])(\gamma_i(t))}, 1]$ so that $\partial_t \Re\mathfrak{e}[f(\gamma_i(t))] \geq 1$. $\Re\mathfrak{e}[f]$ is bounded in K, so the path $\gamma_i(t)$ eventually leaves K. Set $t_i = \inf\{t|\gamma_i(t) \notin K\}$ and set $b_i = \gamma_i(t_i)$, then b_i is the point at which $\gamma_i(t)$ exits K. Thus there are at least n exit points $b_1, ..., b_n$, and we have traversed $\gamma_i(t)$ in the positive direction to get from a_i to b_i, so $\Re\mathfrak{e}[f(a_i)] \leq \Re\mathfrak{e}[f(b_i)]$. To show that there are only n points at which γ exits K we can follow the paths in reverse direction (i.e. apply the same argument to $-f$). To prove the lemma for m critical points in $\gamma \cap K$, we proceed by induction dividing K into one set containing $m-1$ critical points, for which the lemma holds, and one containing 1 critical point, for which the above argument yields the lemma, then delete all entry and exit points along the shared boundary between the two sets.

□

Now we are in a position to see why, for a rational function g with a finite explicit set of critical points and poles, we can find a contour curve in $\{z : \Re\mathfrak{e}[g(z)] = 0\}$ from θ to ∞. Up to homotopy there is a finite number of contour curve configurations so that each critical point or pole has the correct number of incident contour curves (twice its order). This means if our sets of critical points and poles are small we can rule out a few possible configurations by controlling $\Re\mathfrak{e}[g(z)]$ or its derivatives until the only remaining configurations have the desired curve.

We will follow the same general plan for our function h, however we will have to address the fact that we are dealing with a nonexplicit set of critical points and an infinite set of poles. The more difficult problem of critical points is addressed in Lemma 3.3.9 by examining level sets of $h'(z)$ using our conservation property for contour paths and our control of the sign of $\Re\mathfrak{e}[h(z)]$ along the real and imaginary axis.

Lemma 3.3.9. *The only critical point of $-h$ with nonnegative real part is at θ.*

Proof. Recall $-h'(z) = p(z) - p(\theta)$, and $p(\theta) > 0$. Thus if a is a critical point of $-h$, then

$\mathfrak{Im}[p(a)] = 0$. We will examine the level set $\mathfrak{Im}[p(z)] = 0$ in the right half plane. $p(z)$ differs from $h'(z)$ by a real number, so by Lemma 3.3.2 the level set $\mathfrak{Im}[p(z)] = 0$ does not intersect the imaginary axis. As $z \to \infty$, in the right half plane, $p(z) \to 0$. $\mathfrak{Re}[p(z)]$ is increasing along contour paths of p, so no contour path of p can travel from ∞ to ∞. Thus every contour path for $\mathfrak{Im}[p(z)] = 0$ must start or end at a pole, and the only pole of $p(z)$ in the right half plane is at 0. This pole has highest order term $1/z^3$ near 0, so there are at most 3 contour paths of $\mathfrak{Im}[p(z)]$ in the right half plane. One contour path begins at $-\infty$ and travels along the real line (directed away from 0], the other two contour paths are directed toward 0 with one above the real line and one below the real line.

The point θ is a zero of p and a critical point of p with negative second derivative (because $h'''(0) > 0$)), so p is equivalent to $-(z-\theta)^2$ near θ. Thus p has contour curves entering θ along the positive and negative real line, and has contour curves leaving parallel to the positive and negative imaginary axis. By Lemma 3.3.3, there exists a contour curve directed from 0 to θ along the real axis, and a contour curve directed from ∞ to θ along the real axis.

We have $p(\overline{z}) = \overline{p(z)}$, so it is enough to consider the level set of $\mathfrak{Im}[p(z)] = 0$ restricted to the upper right quarter plane. In the upper right quarter plane, $p(z) \to 0$ as $|z| \to \infty$ uniformly in $|z|$. Let $\overline{D}_\theta$ be a disk centered at 0 intersected with the upper right quarter plane, with the disk chosen large enough that $\mathfrak{Re}[p(z)] < p(\theta) > 0$ for all $z \notin D_\theta$ in the upper right quarter plane. Let D_θ be the set $\overline{D}_\theta$ with an arbitrarily small circle around 0 removed, so D_θ contains no poles.

By Lemma 3.3.8, the contour path entering D_θ at θ must exit D_θ at a point b such that $\mathfrak{Re}[p(b)] > \mathfrak{Re}[p(\theta)]$. By our choice of D_θ this contour path cannot exit D_θ toward ∞, by Lemma 3.3.2 it cannot exit along the imaginary axis, and by Claim 3.3.3 it cannot exit along the real axis, because the real axis is contained in the level set of $\mathfrak{Im}[p(z)] = 0$, and θ is the only critical point along the real axis. Thus the contour path entering at θ must exit toward the pole at 0, so there is a contour path $\alpha(t)$ of $\mathfrak{Im}[p(z)] = 0$ from θ to 0.

Furthermore the contour path $\alpha(t)$ connecting θ to 0 contains no critical points of h' other than θ. We prove this by contradiction. Assume $\alpha(t)$ has a critical point besides θ, then it has finitely

many critical points because $\alpha(t)$ is contained in the compact set D_θ. Let z_c be the critical point for which $\mathfrak{Re}[\alpha(z)]$ is smallest. Let $\overline{A}$ be the compact set enclosed between $\alpha(t)$ and the line segment $[0,t]$. Let A be $\overline{A}$ with an arbitrarily small circle around 0 removed. One contour line exits A at z_c, so by Lemma 3.3.8 there must be a contour line entering A at a point z_b with $\mathfrak{Re}[\alpha(z_b)] < \mathfrak{Re}[\alpha(z_c)]$. Because c minimizes $\mathfrak{Re}[\alpha(z)]$ over all critical points of $\alpha(t)$, and no critical point occurs along the real axis, we arrive at a contradiction.

We have classified the contour curves of $\mathfrak{Im}[p(z)] = 0$ in the right half plane as: one contour curve with real part of $p(z)$ increasing from θ to 0 along the real line, one contour curve with real part of $p(z)$ decreasing from θ to ∞ along the real line, one contour curve with real part of $p(z)$ increasing from θ to 0 above the real line, one contour curve with real part of $p(z)$ increasing from θ to 0 below the real line. Any critical point of $-h$ must have $\mathfrak{Im}[p(z)] = 0$ and $\mathfrak{Re}[p(z)] = p(\theta)$. Thus any critical point must be on one of the four contour lines described above or the critical point θ, but every point z on these contour curves has been specified to have $\mathfrak{Re}[p(z)]$ either strictly greater than, or strictly less than $p(\theta)$. So θ is the only critical point of $-h$.

□

Now we can address the simpler problem that h has an infinite number of poles using the conservation of contour paths and the the sign of the derivative of $\mathfrak{Re}[h(z)]$ along the imaginary axis. We do so in Lemma 3.3.10 and prove the existence of a contour curve in $\{z : \mathfrak{Re}[h(z)] = h(\theta)\}$ with the desired properties.

Lemma 3.3.10. *The contour curve γ_1 for $\mathfrak{Re}[h(z)] = h(\theta)$ which exits θ at angle $\frac{5\pi}{6}$ enters 0 at angle $\pi/4$, and the contour curve γ_2 for $\mathfrak{Re}[h(z)] = h(\theta)$ which exits θ at angle $\frac{\pi}{2}$ crosses the positive imaginary axis.*

Proof. Lemma A.0.1 shows that $\lim_{y\to\infty} \mathfrak{Im}[\mathbf{i}h(x+\mathbf{i}y)] = \lim_{y\to\infty} \mathfrak{Re}[-\psi(x+\mathbf{i}y)] = -\infty$, and that this convergence is uniform with respect to x for $x \in [0,\theta]$. Let C be large enough that for all $y > C$, $\mathfrak{Im}[\mathbf{i}h(x+\mathbf{i}y)] < \mathfrak{Im}[\mathbf{i}h(x+\mathbf{i}y)] < h(\theta)$, and consider the rectangle $\overline{S} = [0,\theta] \times [0,\mathbf{i}C]$ in the complex plane. Let S be $\overline{S}$ with an arbitrarily small open circle around 0 removed. Neither γ_1

nor γ_2 can cross the line $[\mathbf{i}C, t + \mathbf{i}C]$ because for $z \in [\mathbf{i}C, t + \mathbf{i}C]$, $\Re\mathfrak{e}[h(z)] < h(\theta)$. Multiplying h by $\mathbf{i}$ and applying Lemma 3.3.8 tells us that the contour curve γ_1 enters S at θ, and must exit S at a point b with $\Im\mathfrak{m}[h(b)] > \Im\mathfrak{m}[h(\theta)] = 0$. It cannot exit S along $[0, \theta]$, because $\Im\mathfrak{m}[h(t)] = 0$ for all $t \in \mathbb{R}$.

Examining the critical point at θ shows that if we follow γ_1 away from θ, then $\Re\mathfrak{e}[-h(z)]$ is positive for z immediately to the left of γ_1 and negative immediately to the right. Thus if this contour curve were to cross the imaginary axis, $\Re\mathfrak{e}[-h]$ would be decreasing in a neighborhood of the intersection. This contradicts Lemma 3.3.2 so γ_1 cannot cross the imaginary axis.

The contour curve γ_2, is left of the line $\theta + \mathbf{i}\mathbb{R}$. Examining the critical point at θ shows that if we follow this new contour away from θ, then $\Re\mathfrak{e}[h(z)]$ is positive for z immediately to its right, and negative for z immediately to its left. If this contour were to cross the line $\theta + \mathbf{i}\mathbb{R}$ then $\Re\mathfrak{e}[h(z)]$ would be increasing on this line in a neighborhood of the intersection. This contradicts Lemma 3.2.6, so γ_2 cannot cross the line $\theta + \mathbf{i}\mathbb{R}$. By Lemma 3.3.9, θ is the only critical point of h in the right half plane, thus the contour line γ_1 cannot cross γ_2 to exit S on the right. Thus the only possible place for γ_1 to exit S is to the pole at 0. γ_2 cannot cross γ_1 to reach $(0, t]$, we have already shown that it does not cross $[\mathbf{i}C, t + \mathbf{i}C]$ or $\theta + \mathbf{i}\mathbb{R}$, and no other contour lines leave 0 into the upper right half plane, so γ_2 must cross the positive imaginary axis. □

Now we are prepared to prove Proposition 3.2.9 by deforming the contour curve found in Lemma 3.3.10 so that it lies in the level set $\{z : \Re\mathfrak{e}[h(z)] = h(\theta) - \varepsilon\}$.

Proof of Proposition 3.2.9. Because $h(\omega z) = \omega h(z)$, it is enough to prove the lemma in the upper half plane. As $z \to \infty$, only one term of h becomes infinite, so $h(z) \sim -c\psi(z) \sim -c\log(z)$ by Lemma A.0.3, and as $\Im\mathfrak{m}[z] \to +\infty$, $h(z) \to -\infty$ uniformly for $\Re\mathfrak{e}[z]$ in a compact set. Thus there exists a large M such that for $\Im\mathfrak{m}[z] > M$, $z \in \mathcal{D}_{\varepsilon,t}(\phi_\varepsilon) \setminus \mathcal{D}^{\varepsilon}_{\varepsilon,t}(\phi_\varepsilon)$, $\Re\mathfrak{e}[h(z) - h(\theta)] < -\eta$.

By Lemma 3.2.6 $\partial_y \Re\mathfrak{e}[h(\theta + \mathbf{i}y)] < 0$ for $y > 0$. Thus for some large M, on the compact set $y \in [\cos(\phi_\varepsilon)\varepsilon, M]$, $\partial_y \Re\mathfrak{e}[h(\theta + \mathbf{i}y)]$ has some negative minimum. The function $h''(z)$ is analytic in the compact rectangle with corners $\theta + \mathbf{i}\varepsilon\cos(\phi_\varepsilon)$, $\theta + \varepsilon e^{\mathbf{i}\phi}$, $\theta + \mathbf{i}M$, $\theta + \varepsilon\sin(\phi_\varepsilon) + \mathbf{i}M$. Thus $|h''(z)|$ is bounded above by some R in this rectangle. Note that R depends only on $\varepsilon\cos(\phi_\varepsilon)$ and M, and

R is increasing in $\cos(\phi_\varepsilon)$. We can choose ε and ϕ so that $\varepsilon\cos(\phi_\varepsilon)$ remains fixed, and $\varepsilon\sin(\phi_\varepsilon)$ becomes arbitrarily small. Choosing so that $\varepsilon\sin(\phi_\varepsilon)R < \eta$ guarantees that $\partial_y \mathfrak{Re}[h(\theta + \sin(\phi_\varepsilon)\varepsilon + \mathbf{i}y)] > 0$ for $y \in [\cos(\phi_\varepsilon)\varepsilon, M]$. Because R is increasing ϕ_ε any smaller choice of $\phi_\varepsilon > 0$ also works.

Similarly by analyticity of h, we can uniformly bound $h'(z)$ on the line segment $[\theta+\mathbf{i}\varepsilon\cos(\phi_\varepsilon), \theta+\varepsilon\sin(\phi_\varepsilon) + \mathbf{i}\cos(\phi_\varepsilon)]$, and by Lemma 3.2.6 we know that $\mathfrak{Re}[h(\theta + \mathbf{i}\varepsilon\cos(\phi_\varepsilon)) - h(\theta)] < 0$. Thus for small enough $\varepsilon\sin(\phi_\varepsilon)$, $\mathfrak{Re}[h(\theta + \varepsilon\sin(\phi_\varepsilon) + \mathbf{i}\varepsilon\cos(\phi_\varepsilon)) - h(\theta)] < -\eta$. Again for a particular choice of $\varepsilon, \phi_\varepsilon$, any smaller ϕ_ε also works. □

3.4 Proof of the Fredholm determinant formula

In this section we will degenerate the Fredholm determinant formula in Theorem 3.1.20 for the Laplace transform of the quenched point to half line probability of a beta RWRE to arrive at the Fredholm determinant formula for the Laplace transform of $\mathsf{K}_{0,t}(0, [x, \infty))$ given in Theorem 3.1.11.

Proof of Theorem 3.1.11. Let $X_\varepsilon(t)$ be $X(t)$ be as in Definition 3.1.18 with parameters $\alpha = \beta = \varepsilon\lambda$. By Lemma 3.1.19, we have $\mathsf{K}_{0,t}(0, [x, \infty)) = \lim_{\varepsilon\to 0} \mathsf{P}(\varepsilon X_\varepsilon(\varepsilon^{-2}t) \geq x)$. Note that in the expression for $K_u^{RW}(v, v')$, the only place where any of x, t, α, β appear is in the definition of g^{RW}.

By Theorem 3.1.20 we can write

$$\mathbb{E}\left[e^{u\mathsf{P}(\varepsilon X_\varepsilon(\varepsilon^{-2}t)\geq x)}\right] = \det(I - K_{u,\varepsilon}^{RW})_{L^2(C_0)}. \tag{3.48}$$

with

$$K_{u,\varepsilon}^{RW}(v, v') = \frac{1}{2\pi\mathbf{i}} \int_{1/2-\mathbf{i}\infty}^{1/2+\mathbf{i}\infty} \frac{\pi}{\sin(\pi s)}(-u)^s \frac{g_\varepsilon^{RW}(v)}{g_\varepsilon^{RW}(v+s)} \frac{ds}{s + v - v'},$$

$$g_\varepsilon^{RW}(v) = \left(\frac{\Gamma(v)}{\Gamma(\varepsilon a + v)}\right)^{(\varepsilon^{-2}t-\varepsilon^{-1}x)/2} \left(\frac{\Gamma(\varepsilon(a+b)+v)}{\Gamma(\varepsilon a + v)}\right)^{(\varepsilon^{-2}t+\varepsilon^{-1}x)/2} \Gamma(v),$$

and C_0 is a positively oriented circle around $1/2$ with radius $1/2$.

We will take the limit of (3.48) as $\varepsilon \to 0$. The expression $e^{\mathsf{P}(\varepsilon X_\varepsilon(\varepsilon^{-2}t)\geq x)}$ is bounded above by e, so in the left hand side we can pass the limit through the expectation to get

$$\lim_{\varepsilon\to 0} \mathbb{E}[e^{\mathsf{P}(\varepsilon X_\varepsilon(\varepsilon^{-2}t)\geq x)}] = \mathbb{E}[e^{u\mathsf{K}_{0,t}(0,[x,\infty))}].$$

Thus to complete the proof, we only need to show that

$$\lim_{\varepsilon\to 0} \det(I - K^{RW}_{u,\varepsilon})_{L^2(C_0)} = \det(I - K_u)_{L^2(C)}. \tag{3.49}$$

We prove (3.49) in three steps; step 1 gives the reason why this convergence should hold, while steps 2 and 3 provide the bounds necessary to make the argument rigorous.

Step 1: First for fixed v, v', s we show the integrand of $K^{RW}_{u,\varepsilon}(v, v')$ converges to the integrand of $K_u(v, v')$ as $\varepsilon \to 0$.

By Taylor expanding in ε, and setting $a = b = \lambda$, we have

$$g^{RW}_\varepsilon(v) = \Big(1+\varepsilon^2\lambda^2\psi_1(v)+O(\varepsilon^3)\Big)^{\varepsilon^{-2}t/2}\Big(1-\varepsilon\lambda\psi(v)+O(\varepsilon^2)\Big)^{-\varepsilon^{-1}x/2}\Big(1+\varepsilon\lambda\psi(v)+O(\varepsilon^2)\Big)^{\varepsilon^{-1}x/2}\Gamma(v).$$

Taking the limit as $\varepsilon \to 0$ gives

$$\lim_{\varepsilon\to 0} g^{RW}_\varepsilon(v) = g(v), \qquad \text{for } v \in \mathbb{C} \setminus \mathbb{Z}_{\leq 0}. \tag{3.50}$$

The limit (3.50) shows pointwise convergence of the integrand of $K^{RW}_{u,\varepsilon}(v, v')$ to the integrand of $K_u(v, v')$.

Additionally if K is a compact set which is separated from all poles of the Gamma function, then the convergence in (3.50) is uniform for $v \in K$. This follows from the fact that the Lagrange form of the remainder in the taylor expansions is bounded for $v \in K$, because $\Gamma''(v)$ is bounded for $v \in K$. So we have shown that integrand of $K^{RW}_{u,\varepsilon}(v, v')$ converges to the integrand of $K_u(v, v')$.

uniformly for v in a compact set K that does not contain poles of the Gamma function.

Step 2: Now we prove that for fixed v, v', the kernel $K^{RW}_{u,\varepsilon}(v, v') \to K_u(v, v')$ as $\varepsilon \to 0$. We do this by proving bounds on the integrand of $K^{RW}_{u,\varepsilon}(v, v')$ in order to apply dominated convergence to the pointwise convergence of the integrand in step 1.

For $s \in 3/4 + \mathbf{i}\mathbb{R}$, $v \in B_{1/8}(0)$, we have the following bounds

$$\left|\frac{\pi}{\sin(\pi s)\Gamma(s+v)}\right| \leq \frac{2\pi}{e^{\pi|\Im\mathfrak{m}[s]|}} e^{\frac{\pi}{2}|\Im\mathfrak{m}[s+v]|-C+(\frac{1}{4}+\Re\mathfrak{e}[v])\log|\Im\mathfrak{m}[s]|}, \tag{3.51}$$

$$|(-u)^s| \leq |u|^{3/4} \tag{3.52}$$

Equation (3.51) follows from Lemma A.0.4 and Lemma A.0.7. Equation (3.52) follows from the fact that $u \in \mathbb{R}$. Note that for $s = 3/4 + \mathbf{i}y$, $|y| > M$,

$$\left|\left(\frac{\Gamma(v+s+\lambda\varepsilon)^2}{\Gamma(v+s)\Gamma(v+s+2\lambda\varepsilon)}\right)^{\varepsilon^{-2}t/2}\right| = \exp(\log(\Gamma(v+s+\lambda\varepsilon) - \log(\Gamma(v+s)$$
$$+ \log(\Gamma(v+s+\lambda\varepsilon) - \log(\Gamma(v+s+2\lambda\varepsilon)) \leq 1. \tag{3.53}$$

The last inequality can be seen by applying Stirling's approximation in a precise way. For details see Lemma A.0.5. Similarly for $|y| > M$,

$$\left|\left(\frac{\Gamma(v)}{\Gamma(v+s+2\lambda\varepsilon)}\right)^{\varepsilon^{-1}x/2}\right| = \exp(\log(\Gamma(v+s)) - \Gamma(v+s+2\lambda\varepsilon)) < 1. \tag{3.54}$$

The last inequality follows from an approximation of the Gamma function which is similar to Stirling's approximation. See Lemma A.0.6 for details. For the final s dependent term of the integrand, there is a constant $C > 0$ such that

$$\left|\frac{\Gamma(v+s)}{g^{RW}(v+s)}\right| = \left|\left(\frac{\Gamma(v+s+\lambda\varepsilon)^2}{\Gamma(v+s)\Gamma(v+s+2\lambda\varepsilon)}\right)^{\varepsilon^{-2}t/2}\left(\frac{\Gamma(v+s)}{\Gamma(v+s+2\lambda\varepsilon)}\right)^{\varepsilon^{-1}x/2}\right| \leq C \tag{3.55}$$

When $y > M$, (3.55) follows from $x, t \geq 0$ along with (3.53) and (3.54). When $y \leq M$, (3.55)

follows from uniform convergence for $s \in [3/4 - \mathbf{i}M, 3/4 + \mathbf{i}M]$ of $\frac{\Gamma(v+s)}{g^{RW}(v+s)}$ to $e^{-\lambda x \psi_0(v) - \frac{\lambda^2 t}{2}\psi_1(v)}$. By (3.51), (3.52), and (3.55) we see that for $s \in 3/4 + \mathbf{i}\mathbb{R}$ the integrand of $K^{RW}_{u,\varepsilon}(v, v')$ is bounded, and has exponential decay coming from (3.52) as $\Im\mathfrak{m}[s] \to \infty$. Thus we can apply dominated convergence to show that

$$\lim_{\varepsilon \to 0} K^{RW}_{u,\varepsilon}(v, v') = K_u(v, v'). \tag{3.56}$$

Step 3: Now we complete the proof of (3.49) by bounding the full Fredholm determinant expansion of $\det(I - K^{RW}_{u,\varepsilon})_{L^2(C_0)}$ in order to apply dominated convergence to the pointwise convergence of kernels proved in step 2.

Let $\mathcal{A}_\varepsilon$ be a rectangle with corners at $\frac{1}{8} + \mathbf{i}\frac{1}{8}$, $\frac{1}{8} - \mathbf{i}\frac{1}{8}$, $-\lambda\varepsilon + \mathbf{i}\frac{1}{8}$, $-\lambda\varepsilon - \mathbf{i}\frac{1}{8}$, oriented in the counterclockwise direction. The convergence in (3.50) is uniform on $A_\varepsilon \setminus B_\delta(0)$, so for sufficiently small $\varepsilon > 0$, and $v \in A_\varepsilon \setminus B_\delta(0)$ there is a constant C such that

$$g^{RW}_\varepsilon(v) \leq C.$$

Now setting $v = \mathbf{i}y + \lambda\varepsilon$, we need to control

$$g^{RW}_\varepsilon(-\lambda\varepsilon + \mathbf{i}y) = \Gamma(\mathbf{i}y - \lambda\varepsilon)\left(\frac{\Gamma(\mathbf{i}y + \lambda\varepsilon)\Gamma(\mathbf{i}y - \varepsilon)}{\Gamma(\mathbf{i}y)^2}\right)^{\varepsilon^{-2}t/2}\left(\frac{\Gamma(\mathbf{i}y + \lambda\varepsilon)}{\Gamma(\mathbf{i}y - \lambda\varepsilon)}\right)^{\varepsilon^{-1}x/2}, \tag{3.57}$$

for $\varepsilon, y \leq \delta$. Let $R(z) = \Gamma[z] - 1/z$ and note that $R(z)$ is holomorphic in a neighborhood of 0. By Taylor's theorem,

$$R(\mathbf{i}y + 1 + \varepsilon) = R(\mathbf{i}y + 1) + R'(\mathbf{i}y + 1)\varepsilon + \mathrm{Rem}(\mathbf{i}y + 1, \varepsilon)\varepsilon^2 \tag{3.58}$$

$$R(\mathbf{i}y + 1 - \varepsilon) = R(\mathbf{i}y + 1) - R'(\mathbf{i}y + 1)\varepsilon + \mathrm{Rem}(\mathbf{i}y + 1, -\varepsilon)\varepsilon^2, \tag{3.59}$$

where $R(\mathbf{i}y + 1)$, $R'(\mathbf{i}y + 1)$, $\mathrm{Rem}(\mathbf{i}y + 1, \varepsilon)$, and $\mathrm{Rem}(\mathbf{i}y + 1, -\varepsilon)$ are bounded uniformly for $y \in (-\delta, \delta)$, $\varepsilon \in (0, \delta)$.

$$\left(\frac{\Gamma(\mathbf{i}y+\lambda\varepsilon)\Gamma(\mathbf{i}y-\varepsilon)}{\Gamma(\mathbf{i}y)^2}\right) = \frac{\frac{1}{\mathbf{i}y+\varepsilon}+R(\mathbf{i}y+\varepsilon+1)}{\frac{1}{\mathbf{i}y}+R(\mathbf{i}y+1)}\frac{\frac{1}{\mathbf{i}y-\varepsilon}+R(\mathbf{i}y-\varepsilon+1)}{\frac{1}{\mathbf{i}y}+\mathbb{R}(\mathbf{i}y+1)} \tag{3.60}$$
$$= \left(\frac{(\mathbf{i}y)^2}{(\mathbf{i}y+\varepsilon)(\mathbf{i}y-\varepsilon)}\right)\left(\frac{\left(1+(\mathbf{i}y+\varepsilon)\left(R(\mathbf{i}y+1)+R'(\mathbf{i}y+1)\varepsilon+\mathrm{Rem}(\mathbf{i}y+1,\varepsilon)\varepsilon^2\right)\right)}{(1+\mathbf{i}yR(\mathbf{i}y+1))^2}\right)$$
$$\times\left(1+(\mathbf{i}y-\varepsilon)\left(R(\mathbf{i}y+1)-R'(\mathbf{i}y+1)\varepsilon+\mathrm{Rem}(\mathbf{i}y+1,-\varepsilon)\varepsilon^2\right)\right)$$
$$= \left(\frac{1}{1+\frac{\varepsilon^2}{y^2}}\right)\left(\frac{(1+\mathbf{i}yR(\mathbf{i}y+1))^2+\varepsilon^2\mathrm{Rem}_1(\mathbf{i}y+1,\varepsilon)}{(1+\mathbf{i}yR(\mathbf{i}y+1))^2}\right),$$

where $\mathrm{Rem}_1(\mathbf{i}y+1,\varepsilon)$ is bounded uniformly for $y\in(-\delta,\delta),\varepsilon\in(0,\delta)$. The first equality follows from the definition of R and the second follows from (3.58) and (3.59). The third equality follows expanding

$$\left(1+(\mathbf{i}y+\varepsilon)\left(R(\mathbf{i}y+1)+R'(\mathbf{i}y+1)\varepsilon+\mathrm{Rem}(\mathbf{i}y+1,\varepsilon)\varepsilon^2\right)\right) \tag{3.61}$$
$$\times\left(1+(\mathbf{i}y-\varepsilon)\left(R(\mathbf{i}y+1)-R'(\mathbf{i}y+1)\varepsilon+\mathrm{Rem}(\mathbf{i}y+1,-\varepsilon)\varepsilon^2\right)\right), \tag{3.62}$$

and noting that the coefficient of ε^0 is $(1+\mathbf{i}yR(\mathbf{i}y+1))^2$, the coefficient of ε^1 is 0. The fact that $\mathrm{Rem}_1(\mathbf{i}y+1,\varepsilon)$ is bounded comes from the fact that every coefficient of ε^k in the two terms of (3.61) is bounded uniformly in y,ε.

Define

$$\mathrm{Rem}_2(\mathbf{i}y+1,\varepsilon) = \frac{\mathrm{Rem}_1(\mathbf{i}y+1,\varepsilon)}{(1+\mathbf{i}yR(\mathbf{i}y+1))^2}. \tag{3.63}$$

We have that for $x \in (0, \delta)$, $y \in (-\delta, \delta)$,

$$\begin{aligned}\left|\left(\frac{\Gamma(\mathbf{i}y + \lambda\varepsilon)\Gamma(\mathbf{i}y - \varepsilon)}{\Gamma(\mathbf{i}y)^2}\right)^{-1}\right| &= \left|\left(1 + \frac{\varepsilon^2}{y^2}\right)\left(\frac{1}{1 + \varepsilon^2 \mathrm{Rem}_2(\mathbf{i}y + 1, \varepsilon)}\right)\right| \\ &\geq \left|\left(1 + \frac{\varepsilon^2}{y^2}\right)(1 - \varepsilon^2 \mathrm{Rem}_2(\mathbf{i}y + 1, \varepsilon))\right| \\ &\geq \left|1 + \varepsilon^2\left(\frac{1}{y^2} - \mathrm{Rem}_2(\mathbf{i}y + 1, \varepsilon) - \varepsilon^2 \frac{\mathrm{Rem}_2(\mathbf{i}y + 1, \varepsilon)}{y^2}\right)\right| \\ &\geq \left(1 + \varepsilon^2 \frac{3}{4y^2}\right). \end{aligned} \tag{3.64}$$

The first equality follows from (3.60). The first inequality follows from the fact that for any $0 < x < 1$, $\left|\frac{1}{1+x}\right| \geq |1 - x|$,. The final inequality may require us to choose a still smaller $\delta > 0$ and follows from the fact that $\mathrm{Rem}_2(\mathbf{i}y + 1, \varepsilon)$ is bounded.

By Laurent expanding the Gamma function around 0, we can see that

$$\Gamma(-\lambda\varepsilon + \mathbf{i}y) \leq \frac{1}{\sqrt{y^2 + \varepsilon^2}} + C \leq \frac{1}{y} + C, \tag{3.65}$$

for $0 < \varepsilon < \delta$ and $y \in (-\delta, \delta)$. We also have

$$\begin{aligned}\left(\left(\frac{\Gamma(\mathbf{i}y + \lambda\varepsilon)\Gamma(\mathbf{i}y - \varepsilon)}{\Gamma(\mathbf{i}y)^2}\right)^{\varepsilon^{-2}t/2}\right)^{-1} &\geq \left(1 + \varepsilon^2 \frac{3}{4y^2}\right)^{\frac{\varepsilon^{-2}t}{2}} \geq 1 + \frac{3t}{8y^2} \\ &\geq \left(\frac{1}{y} + C\right) \geq \Gamma(\mathbf{i}y - \varepsilon\lambda), \end{aligned} \tag{3.66}$$

for y sufficiently small. The first inequality follows from (3.64), the equality is Newton's generalized binomial theorem, the second inequality uses Bernoulli's inequality. The third inequality is true if y large in particular $1/y > 1 + 8C/3t$. The fourth inequality follows (3.65).

Equation (3.66) implies

$$\Gamma(\mathbf{i}y - \lambda\varepsilon)\left(\frac{\Gamma(\mathbf{i}y + \lambda\varepsilon)\Gamma(\mathbf{i}y - \varepsilon)}{\Gamma(\mathbf{i}y)^2}\right)^{\varepsilon^{-2}t/2} \leq 1. \tag{3.67}$$

By Taylor's theorem, there exists a function $\mathrm{Rem}_3(\mathbf{i}y, \varepsilon)$ which is bounded for $\varepsilon \in (0, \delta)$, $y \in (-\delta, \delta)$ satisfying

$$\Gamma(\mathbf{i}y + \varepsilon) = \frac{1}{\mathbf{i}y + \varepsilon} + \mathrm{Rem}_3(\mathbf{i}y, \varepsilon)\varepsilon \quad \text{and} \quad \Gamma(\mathbf{i}y - \varepsilon) = \frac{1}{\mathbf{i}y - \varepsilon} + \mathrm{Rem}_3(\mathbf{i}y, -\varepsilon)\varepsilon.$$

Thus

$$\frac{\Gamma(\mathbf{i}y + \varepsilon)}{\Gamma(\mathbf{i}y - \varepsilon)} = \frac{\frac{1}{\mathbf{i}y+\varepsilon} + \mathrm{Rem}_3(\mathbf{i}y, \varepsilon)\varepsilon}{\frac{1}{\mathbf{i}y-\varepsilon} + \mathrm{Rem}_3(\mathbf{i}y, -\varepsilon)\varepsilon} = \left(\frac{\mathbf{i}y - \varepsilon}{\mathbf{i}y + \varepsilon}\right)\left(\frac{1 + (\mathbf{i}y + \varepsilon)\mathrm{Rem}_3(\mathbf{i}y, \varepsilon)\varepsilon}{1 + (\mathbf{i}y - \varepsilon)\mathrm{Rem}_3(\mathbf{i}y, -\varepsilon)\varepsilon}\right) = \left(\frac{\mathbf{i}y - \varepsilon}{\mathbf{i}y + \varepsilon}\right)(1+C_{\varepsilon,y}\varepsilon),$$

where for any η we can choose δ small enough that $C_{\varepsilon,y} \leq \eta$. Thus for all $\varepsilon \in (0, \delta), y \in (-\delta, \delta)$,

$$(1 - \eta\varepsilon) \leq \left|\frac{\Gamma(\mathbf{i}y + \varepsilon)}{\Gamma(\mathbf{i}y - \varepsilon)}\right| \leq (1 + \eta\varepsilon). \tag{3.68}$$

This implies

$$\left|\left(\frac{\Gamma(\mathbf{i}y + \lambda\varepsilon)}{\Gamma(\mathbf{i}y - \lambda\varepsilon)}\right)^{\varepsilon^{-1}x/2}\right| \leq (1 + \eta\varepsilon)^{e^{-1}x/2} \leq e^{\eta x/2}. \tag{3.69}$$

Together (3.57), (3.67), and (3.69) imply that for δ small, $\varepsilon \in (0, \delta), y \in (-\delta, \delta)$,

$$|g_\varepsilon^{RW}(-\lambda\varepsilon + \mathbf{i}y)| \leq e^{\eta x/2}.$$

Thus $K_{u,\varepsilon}^{RW}(v, v')$ is bounded by some C on the contour $\mathcal{A}_\varepsilon$. Hadamard's bound implies that

$$\frac{\det[K_{u,\varepsilon}(x_i, x_j)]_{i,j=1}^k}{k!} \leq \frac{C^k k^{k/2}}{k!},$$

where the right hand side decays at rate $\frac{C^k}{e^{k\log(k)/2}}$ by Stirling's formula. Together with (3.56), and the fact that the contours A_ε are finite volume, this allows us to apply dominated convergence to the Fredholm determinant expansion $\det(I - K_{u,\varepsilon}^{RW}(v, v'))_{L^2(\mathcal{A}_\varepsilon)}$, to get

$$\lim_{\varepsilon \to 0} \det(I - K_{u,\varepsilon}^{RW})_{L^2(\mathcal{A}_\varepsilon)} = \det(I - K_u)_{L^2(\mathcal{A}_\varepsilon)}. \tag{3.70}$$

We can deform the contour $\mathcal{A}_\varepsilon$ to C without crossing any poles of $K_u(v,v')$ and we can deform $\mathcal{A}_\varepsilon$ to C_0 without crossing any poles of $K^{RW}_{u,\varepsilon}(v,v')$, so by (3.70)

$$\lim_{\varepsilon\to 0} \det(I - K^{RW}_{u,\varepsilon})_{L^2(C_0)}) = \det(I - K_u)_{L^2(C)}.$$

□

3.5 Moment formulas and Bethe ansatz

3.5.1 Proof of the moment formula Proposition 3.1.22

In this section we find moment formulas for kernels of uniform Howitt-Warren flows, by taking the diffusive limit of [23, Proposition 3.4]. In order to state precisely how we use results from [23], we first explain the connection between the beta RWRE and another model called the beta polymer, which was also introduced in [23].

Definition 3.5.1 (beta polymer)**.** The beta polymer is a probability measure on oriented lattice paths constructed as follows. We consider paths in $\mathbb{Z}^2$ with allowed edges of the form $(i,j) \to (i+1,j)$ and $(i,j) \to (i+1,j+1)$. In other terms, we allow paths to make either right or up-right steps. The measure depends on two parameters $\nu > \mu > 0$. Let $\{B_{(i,j)}\}_{i,j\mathbb{Z}^2}$ be a family of iid random variables distributed according to the beta distribution with parameters $\nu, \nu-\mu$. To each horizontal edge $e = (i-1,j) \to (i,j)$ we assign the Boltzmann weight $w_e = B_{i,j}$, and to each diagonal edge $e = (i-1,j-1) \to (i,j)$ we associate the Boltzmann weight $w_e = (1 - B_{i,j})$.

For fixed points $S,T \in \mathbb{Z}^2$, the beta polymer is a measure on paths $\pi : S \to T$ such that the probability of a path π is proportional to $\prod_{e\in\pi} w_e$. In this chapter, we are mostly interested in paths between the half-line $D := \{(0,i) : i > 0\}$ and any point of coordinates (t,n) for $t \geqslant 0$. The associated partition function is defined as

$$\mathsf{Z}(t,n) = \sum_{i=1}^{n} \sum_{\pi:(0,i)\to(t,n)} \prod_{e\in\pi} w_e.$$

By the definition of our Boltzmann weights, for $t \geq 0, n \in \mathbb{Z}$, the partition function $\mathsf{Z}(t,n)$ is characterized by the following recurrence relation.

$$\begin{cases} \mathsf{Z}(t,n) = B_{t,n}\mathsf{Z}(t-1,n) + (1-B_{t,n})\mathsf{Z}(t-1,n-1), & \text{if } t > 0 \\ \mathsf{Z}(0,n) = \mathbb{1}_{n>0}. \end{cases}$$

Note that this half line to point partition function is the same as the partition function $Z(t,n)$ defined in [23, Definition 1.2].

Now we rephrase the relation between the beta RWRE and the beta polymer from [23, Proposition 1.6].

Proposition 3.5.2. *Consider the beta RWRE with parameters $\alpha, \beta > 0$ (see Definition 3.1.18) and the beta polymer with parameters $\mu = \alpha$, $\nu = \alpha + \beta$. For $t \geq 0$ and $n_1, \ldots, n_k \in \mathbb{Z}$, we have the following equality in distribution,*

$$(\mathsf{Z}(t,n_1), ..., \mathsf{Z}(t,n_k)) = (\mathsf{P}(X_1^{x_1}(t) \geq 0), ..., \mathsf{P}(X_k^{x_k}(t) \geq 0)) \qquad \text{for } x_i = 2n_i - 2 - t,$$

and

$$\mathbb{E}\left[\prod_{i=1}^{k} \mathsf{Z}(t,n_i)\right] = \mathbb{E}\left[\prod_{i=1}^{k} \mathsf{P}(X_i^{x_i}(t) \geq 0)\right],$$

where these expectations are taken over the random environments of the beta polymer and the beta RWRE respectively.

Proof. First note that although the beta RWRE was defined for positive time, we can apply a spatial shift to our variables so that it is defined for all $t > -L$ for any $L \in \mathbb{Z}$. We will use this interpretation when we describe a particle trajectory in the beta RWRE starting from a point with a negative time coordinate. Consider the change of coordinates $x = 2n - 2 - t$ and rewrite $\mathsf{Z}(t,n)$ in terms of (t,x). This corresponds to transforming horizontal edges into diagonal down-right edges. Then, reversing the time direction allows us to identify paths from D to (t,n) in the beta polymer with space-time trajectories in the beta RWRE from the point x at time $-t$ to the half line $[0,+\infty)$ at time 0, such

that the Boltzmann weight of the beta polymer path is equal in distribution to the probability of the beta RWRE trajectory. This equality in distribution also holds jointly for arbitrary collections of paths. Finally, shifting all paths forward in time by t steps in the beta RWRE does not change their law, thus we have the desired equality in distribution. □

Now we can prove the mixed moments formula (Proposition 3.1.22).

Proof of Proposition 3.1.22. We begin with the moment formula [23, Proposition 3.4], Using Proposition 3.5.2 to rewrite $\mathsf{Z}(t,n)$ in terms of $\mathsf{P}(X^x(t) \geq 0)$ gives, for $x_1 \geq \cdots \geq x_k$,

$$\mathbb{E}[\mathsf{P}(X_1^{x_1}(t) \geq 0)...\mathsf{P}(X_k^{x_k}(t) \geq 0)] = \frac{1}{(2\pi\mathbf{i})^k} \int_{\gamma_1} ... \int_{\gamma_k} \prod_{A<B} \frac{z_A - z_B}{z_A - z_B - 1} \prod_{j=1}^{k} \left(\frac{\nu + z_j}{z_j}\right)^{\frac{t+x}{2}-1} \left(\frac{\mu + z_j}{\nu + z_j}\right)^t \frac{dz_j}{z_j + \nu}. \quad (3.71)$$

Where γ_k is a small contour around 0 and γ_i contains $1 + \gamma_j$ for $i < j$, and all contours exclude $-\nu$. To choose the γ_i precisely, fix a small $a_k > 0$ and define the contour $\gamma_k = \gamma_k^\varepsilon$ to be a short vertical line segment $\{-\lambda\varepsilon + \mathbf{i}y : y \in [-a_k, a_k]\}$ union a half circle a $\{-\lambda\varepsilon + a_k e^{\mathbf{i}\theta} : \theta \in [-\pi/2, \pi/2]\}$. Let construct $\gamma_i = \gamma_i^\varepsilon$ in the same way with a_i replacing a_k and choose each a_i large enough that $1 + \gamma_{i+1}^\varepsilon$ is contained in γ_i^ε. Recalling Lemma 3.1.19 and taking $\varepsilon \to 0$ in (3.71) gives

$$\mathbb{E}[\mathsf{K}_{-t,0}(x_1, [0, +\infty))...\mathsf{K}_{-t,0}(x_k, [0, +\infty))] = \lim_{\varepsilon \to 0} \frac{1}{(2\pi\mathbf{i})^k} \int_{\gamma_1^\varepsilon} ... \int_{\gamma_k^\varepsilon} \prod_{A<B} \frac{z_A - z_B}{z_A - z_B - 1} \prod_{j=1}^{k} \left(\frac{2\lambda\varepsilon + z_j}{z_j}\right)^{\frac{\varepsilon^{-2}t+\varepsilon^{-1}x}{2}} \left(\frac{\lambda\varepsilon + z_j}{2\lambda\varepsilon + z_j}\right)^{\varepsilon^{-2}t} \frac{dz_j}{z_j}.$$

We simplify the product

$$\prod_{j=1}^{k} \left(\frac{2\lambda\varepsilon + z_j}{z_j}\right)^{\frac{\varepsilon^{-2}t+\varepsilon^{-1}x}{2}} \left(\frac{\lambda\varepsilon + z_j}{2\lambda\varepsilon + z_j}\right)^{\varepsilon^{-2}t} \frac{dz_j}{z_j} = \prod_{j=1}^{k} \left(1 + \frac{2\lambda\varepsilon}{z_j}\right)^{\frac{\varepsilon^{-1}x}{2}} \left(\frac{(\lambda\varepsilon + z_j)^2}{z_j(2\lambda\varepsilon + z_j)}\right)^{\frac{\varepsilon^{-2}t}{2}} \frac{dz_j}{z_j}. \quad (3.72)$$

Taking the pointwise limit of the integrand suggests that

$$\mathbb{E}[\mathsf{K}_{-t,0}(x_1,[0,+\infty))\ldots\mathsf{K}_{-t,0}(x_k,[0,+\infty))] = \int_{\gamma_1^0}\frac{dz_1}{2\pi\mathbf{i}}\ldots\int_{\gamma_k^0}\frac{dz_k}{2\pi\mathbf{i}}\prod_{A<B}\frac{z_A-z_B}{z_A-z_B-1}\prod_{j=1}^{k}\exp\left(\frac{\lambda^2 t}{2z_j^2}+\frac{\lambda x_j}{z_j}\right)\frac{1}{z_j}, \quad (3.73)$$

where now the contours γ_k^0,... γ_1^0 all pass through 0 in the vertical direction and γ_i^0 contains $1+\gamma_j^0$ for all $i<j$. We will justify this limit by applying dominated convergence at the end of the proof.

The condition $\alpha_i < \frac{\alpha_j}{1+\alpha_j}$ for all $i<j$ implies that if $\bar{\gamma}_i$ is the circle centered at $\alpha_i^{-1}/2$ with radius $\alpha_i^{-1}/2$ oriented in the counterclockwise direction, then $1+\bar{\gamma}_j$ is contained in $\bar{\gamma}_i$ for all $i<j$. Thus Cauchy's theorem allows us to deform the integration contours γ_i to $\bar{\gamma}_i$ in (3.73) without collecting any residues.

We perform a change of variables $w_j = 1/z_j$ on (3.73) and use the fact that the pointwise inverse in the complex plane of a circle with center $\alpha^{-1}/2$ and radius $\alpha^{-1}/2$ is the line $\alpha+\mathbf{i}\mathbb{R}$. We obtain

$$\mathbb{E}[\mathsf{K}_{-t,0}(x_1,[0,+\infty))\ldots\mathsf{K}_{-t,0}(x_k,[0,+\infty))] = \int_{\alpha_1+\mathbf{i}\mathbb{R}}\frac{\mathrm{d}w_1}{2\mathbf{i}\pi}\cdots\int_{\alpha_k+\mathbf{i}\mathbb{R}}\frac{\mathrm{d}w_k}{2\mathbf{i}\pi}\prod_{1\leqslant A<B\leqslant k}\frac{w_B-w_A}{w_B-w_A-w_Aw_B}\prod_{j=1}^{k}\exp\left(\frac{t\lambda^2 w_j^2}{2}+\lambda x_j w_j\right)\frac{1}{w_j}. \quad (3.74)$$

Now use dominated convergence to justify the the $\varepsilon\to 0$ limit which gave (3.73). The contours $\gamma_i(\varepsilon)$ depend on ε and in order to apply dominated convergence, we perform the change of variables $z_i = \bar{z}_i - \lambda\varepsilon$ in (3.72) so that our contours of integration change from $\gamma_i[\varepsilon]$ to $\gamma_i[0]$, and set $\gamma_i' = \gamma_i[0]$.

Now that all our contours of integration do not depend on ε, all we need to do is bound the integrand along these contours. The argument which allows us to apply dominated convergence to get (3.73) is a simplified form of the argument which allows us to apply dominated convergence in

the proof of Theorem 3.1.11. Taylor expanding shows that

$$(1+\varepsilon a)^{\varepsilon^{-1}} \xrightarrow{\varepsilon\to 0} e^a, \qquad \text{uniformly in } a \text{ for } |a| < R.$$

Thus, uniformly for z_j outside a neighborhood of 0,

$$\left(1+\frac{2\lambda\varepsilon}{z_j}\right)^{\frac{\varepsilon^{-1}x}{2}} \xrightarrow[\varepsilon\to 0]{} e^{\frac{\lambda x}{z_j}}, \tag{3.75}$$

and

$$\left(\frac{(\lambda\varepsilon+z_j)^2}{z_j(2\lambda\varepsilon+z_j)}\right)^{\frac{\varepsilon^{-2}t}{2}} = \left(1+\frac{\varepsilon^2\lambda^2}{z_j^2+2\lambda z\varepsilon}\right)^{\frac{\varepsilon^{-2}t}{2}} \xrightarrow[\varepsilon\to 0]{} e^{\frac{\lambda^2 t}{2z_j^2}}. \tag{3.76}$$

Now we bound the integrands along $\gamma_i[\varepsilon] \cap B_\delta(0)$. Near 0 we have $z_i = -\lambda\varepsilon + \mathbf{i}y$, so

$$\left|\left(1+\frac{2\lambda\varepsilon}{z_j}\right)^{\frac{\varepsilon^{-1}x}{2}}\right| = \left|\frac{\lambda\varepsilon+\mathbf{i}y}{-\lambda\varepsilon+\mathbf{i}y}\right|^{\frac{\varepsilon^{-1}x}{2}} = 1, \tag{3.77}$$

and

$$\left|\left(\frac{(\lambda\varepsilon+z_j)^2}{z_j(2\lambda\varepsilon+z_j)}\right)^{\frac{\varepsilon^{-2}t}{2}}\right| = \left|\frac{-y^2}{-y^2-\lambda^2\varepsilon^2}\right|^{\frac{\varepsilon^{-2}t}{2}} < 1. \tag{3.78}$$

Together (3.75), (3.76), (3.77), and (3.78), and the fact that γ_k^ε has uniformly bounded length, allow us to apply dominated convergence to (3.72) to obtain (3.73). This completes the proof. □

3.5.2 Limit to the KPZ equation

In this Section, we show that the moment formula from Proposition 3.1.22 converges to the moments of the solution to the multiplicative noise stochastic heat equation with delta initial data, which suggests that Howitt-Warren stochastic flows of kernels converge to the KPZ equation.

Consider $Z(t,x)$ the solution to the multiplicative noise stochastic heat equation

$$\partial_t Z(t,x) = \frac{1}{2}\partial_{xx} Z(t,x) + \sqrt{\kappa}\xi(t,x)Z(t,x),\ t>0, x\in\mathbb{R},$$

where ξ is a space time white noise and $\kappa > 0$ is a parameter controlling the noise strength. This stochastic PDE has attracted much attention recently because the solution to the KPZ equation

$$\partial_t h(t,x) = \frac{1}{2}\partial_{xx} h(t,x) + \frac{1}{2}\left(\partial_x h(t,x)\right)^2 + \sqrt{\kappa}\xi(t,x)$$

is given by $h(t,x) = \log Z(t,x)$. It is expected that models in the KPZ class which depend on a tunable parameter controlling noise or asymmetry converge to the KPZ equation in the weak asymmetry/noise scaling limit. We refer to [63] for background on these scalings and stochastic PDEs.

Let

$$u_\kappa(t,\vec{x}) = \mathbb{E}\left[Z(t,x_1)\dots Z(t,x_k)\right].$$

It was shown in [36, Section 6.2] (see also [92]) that for Dirac delta initial data $u(0,\cdot) = \delta_0(\cdot)$, the function u_κ can be written for $x_1 \leqslant \dots \leqslant x_k$ as

$$u_\kappa(t,\vec{x}) = \int_{r_1+\mathbf{i}\mathbb{R}} \frac{dz_1}{2\mathbf{i}\pi} \cdots \int_{r_k+\mathbf{i}\mathbb{R}} \frac{dz_k}{2\mathbf{i}\pi} \prod_{1\leq A<B\leq k} \frac{z_A - z_B}{z_A - z_B - \kappa} \prod_{j=1}^{k} e^{x_j z_j + \frac{t}{2} z_j^2}, \tag{3.79}$$

where the contours are such that $r_i > r_{i+1} + \kappa$ for all $1 \leqslant i \leqslant k$.

Recall the moments of the uniform Howitt-Warren flow

$$\Phi_t^{(k)}(x_1,\dots,x_k) = \mathbb{E}\Big[\mathsf{K}_{-t,0}(x_1,[0,+\infty))\dots\mathsf{K}_{-t,0}(x_k,[0,+\infty))\Big],$$

and recall that they depend on a noise parameter λ.

Proposition 3.5.3. *Let $\gamma > 0$ and consider the scalings*

$$T = \lambda^2 t,\ X_i = \lambda^2 t\gamma + \lambda x_i. \tag{3.80}$$

Let $\mathsf{K}_t(x,\cdot)$ be the kernel of the uniform Howitt-Warren stochastic flow with stickiness parameter

λ. We have that for fixed $t > 0$ and $x_1 \leqslant \cdots \leqslant x_k$,

$$\lim_{\lambda\to\infty} (\lambda\gamma)^k \exp\left(\frac{k}{2}t\lambda^2\gamma^2 + \lambda\gamma\sum_{j=1}^{k} x_j\right)\Phi_T^{(k)}(-\vec{X}) = u_{\gamma^2}(t,\vec{x}).$$

Remark 3.5.4. Proposition 3.5.3 suggests that under the scalings (3.80),

$$Z_\lambda(t,x) := \gamma\lambda e^{t\lambda^2\gamma^2/2+\lambda\gamma x}\mathsf{K}_{-T}(-X,[0,+\infty))$$

weakly converges as λ goes to $+\infty$ (in the space of continuous time space trajectories) to the solution of the multiplicative noise stochastic heat equation $Z(t,x)$ with Dirac delta initial data and $\kappa = \gamma^2$. Equivalently, $\log Z_\lambda(t,x)$ would converge weakly to the solution to the KPZ equation with narrow wedge initial data. The analogous statement for discrete random walks in space-time iid random environment is proved in [65].

Proof. Consider (3.15) and perform the change of variables $w_j = \frac{\gamma}{\lambda} + \frac{z_i}{\lambda^2}$. For large enough λ, the contour for z_i may be chosen as $r_i + \mathbf{i}\mathbb{R}$ where $r_{i+1} > r_i + \gamma^2$ for all $1 \leqslant i \leqslant k$. Under the scalings (3.80), we have (dropping unnecessary indices)

$$\frac{T}{2}\lambda^2 w^2 - \lambda X w = \frac{t}{2}z^2 - xz - \gamma x\lambda - \frac{t}{2}\gamma^2\lambda^2,$$

and we have the pointwise convergences

$$\frac{w_b - w_a}{w_b - w_a - w_a w_b} \xrightarrow[\lambda\to+\infty]{} \frac{z_b - z_a}{z_b - z_a - \gamma^2}, \quad \frac{1}{\lambda w_i} \xrightarrow[\lambda\to+\infty]{} \frac{1}{\gamma}.$$

Moreover, it is easy to see that the ratios stay bounded for z_a, z_b, z_i belonging to their fixed vertical

contours. Thus, by dominated convergence,

$$(\lambda\gamma)^k e^{\sum_{j=1}^k \frac{t}{2}\gamma^2\lambda^2+\gamma x_i\lambda}\Phi_T^{(k)}(-X_1,\dots,-X_k) \xrightarrow[\lambda\to+\infty]{} \int_{r_1+\mathbf{i}\mathbb{R}}\frac{dz_1}{2\mathbf{i}\pi}\cdots\int_{r_k+\mathbf{i}\mathbb{R}}\frac{dz_k}{2\mathbf{i}\pi}\prod_{1\le A<B\le k}\frac{z_B-z_A}{z_B-z_A-\gamma^2}\prod_{j=1}^k e^{\frac{t}{2}z_j^2-x_jz_j}.$$

We finally obtain (3.79) by the change of variables $z_i = -\tilde{z}_i$. □

3.5.3 Bethe Ansatz solvability of n-point uniform sticky Brownian motions

For $x \in \mathbb{R}^k$ and $t \ge 0$, let $u(t,\vec{x})$ be the right hand side of (3.74). We claim that u satisfies

$$\partial_t u = \frac{1}{2}\Delta u, \tag{3.81}$$

$$(\partial_i\partial_{i+1} + \lambda(\partial_i - \partial_{i+1}))u|_{x_i=x_{i+1}} = 0. \tag{3.82}$$

Indeed, for any $w \in \mathbb{C}$, the function $\exp\left(\frac{t\lambda^2w^2}{2} + \lambda x w\right)$ is clearly a solution to (3.81). This equation is linear and hence any superposition of solutions satisfies it, so does $u(t,\vec{x})$.

Regarding the boundary condition (3.82), let us apply the operator $\partial_i\partial_{i+1}+\lambda(\partial_i-\partial_{i+1})$ to $u(t,\vec{x})$. The operator can be brought inside the integrals in (3.74) and yields a multiplicative factor

$$\lambda^2 w_i w_{i+1} + \lambda(\lambda w_i - \lambda w_{i+1}).$$

This factor cancels the denominator of

$$\frac{w_B - w_A}{w_B - w_A - w_A w_B}$$

when $A = i$, $B = i+1$, so that the integral in w_{i+1} does not have a pole anymore at $w_{i+1} = w_i/(1+w_i)$. Thus, by Cauchy's theorem, one can shift the w_{i+1} contour from $\alpha_{i+1} + \mathbf{i}\mathbb{R}$ to $\alpha_i + \mathbf{i}\mathbb{R}$. Now that variables w_i and w_{i+1} are integrated on the same contour, we notice that for $x_{i+1} = x_i$, the integrand is antisymmetric with respect to exchanging w_i and w_{i+1} (because of the factor $w_i - w_{i+1}$), and

hence the integral is zero. Thus $u(t,\vec{x})$ satisfies (3.82).

More generally, the function

$$\Psi_{\vec{z}}(\vec{x}) = \sum_{\sigma} \prod_{i<j} \frac{z_{\sigma(i)} - z_{\sigma(j)} - 1}{z_{\sigma(i)} - z_{\sigma(j)}} \prod_{j=1}^{k} e^{\frac{\lambda x_j}{z_j}}, \tag{3.83}$$

satisfies (3.81), (3.82) for any $\vec{z} \in (\mathbb{C} \setminus \{0\})^k$.

The function $u(0,\vec{x})$ is a linear superposition of $\Psi_{\vec{z}}(\vec{x})$ which additionally satisfies the initial condition for $x_1 > \cdots > x_k$ that

$$u(0,\vec{x}) = \prod_{i=1}^{k} \mathbb{1}_{x_i>0}.$$

Note that the function $\Phi_t^{(k)}(\vec{x}) := \mathbb{E}[\mathsf{K}_{-t,0}(x_1,[0,+\infty))...\mathsf{K}_{-t,0}(x_k,[0,+\infty))]$ satisfies the same initial condition.

The discrete analogue of $\Phi_t^{(k)}(\vec{x})$ is $\mathbb{E}[\mathsf{P}(X^{x_1}(t) > 0)...\mathsf{P}(X^{x_k}(t) > 0)]$ (in the sense of Lemma 3.1.19). It was shown in [23, Section 3.1] using simple probabilistic considerations that the latter quantity satisfies discretizations of (3.81), (3.82).

It would be interesting to provide a probabilistic explanation for why $\Phi_t^{(k)}(\vec{x})$ must satisfy (3.81), (3.82). Note that $\Phi_t^{(k)}(\vec{x})$ is symmetric in the x_i's so that we need to understand it only in the Weyl chamber $\mathbb{W}_k := \{x \in \mathbb{R}^k : x_1 \geq \cdots \geq x_k\}$. Then (3.81), (3.82) should be regarded as Kolmogorov's backward equation for k-point uniform sticky Brownian motions. Inside the open sector $x_1 > \cdots > x_k$, it is clear that the generator should be given by the Laplacian (since k-point sticky Brownian motions evolve as k independent Brownian motions), hence the heat equation (3.81). However, we have not found in the literature a rigorous definition of the generator for n-point uniform sticky Brownian motions and we are unable to deduce the boundary condition (3.82) directly from the definition of uniform sticky Brownian motions. After the posting of the manuscript on arXiv, we have learned from Jon Warren that it is possible to derive (3.81), (3.82) directly from the martingale problem characterizing sticky Brownian motions, and this will be explained in the forthcoming paper [48].

3.5.4 A formal relation to diffusions with white noise drift

By analogy with the Lieb-Liniger model (we refer the reader to the book [86, Chap. 4] for background on the Lieb Liniger model, or [36, Section 6] for its relation to the KPZ equation), it is natural from the physics point of view to associate to the equation (3.81) with boundary condition (3.82) the following PDE on $\mathbb{R}^k$

$$\partial_t v(t,\vec{x}) = \frac{1}{2}\Delta v(t,\vec{x}) + \frac{1}{2\lambda}\sum_{i\neq j}\delta(x_i - x_j)\partial_{x_i}\partial_{x_j} v(t,\vec{x}). \tag{3.84}$$

In order to see that (3.84) satisfies the boundary condition (3.82), integrate the equation over the variable $y = x_{i+1} - x_i$ in a neighborhood of 0, and use the fact that $v(t,\vec{x})$ is symmetric in the x_i's for symmetric initial condition. Assuming uniqueness of solutions to (3.81)+(3.82) and (3.84), their restrictions to the Weyl chamber $\mathbb{W}_k := \{x \in \mathbb{R}^k : x_1 \geq \cdots \geq x_k\}$ must coincide, provided the initial conditions coincide on $\mathbb{W}_k$.

Consider now the stochastic PDE

$$\begin{cases} \partial_t q(t,x) = \frac{1}{2}\partial_{xx} q(t,x) + \frac{1}{\sqrt{\lambda}}\xi(t,x)\partial_x q(t,x), \\ q(0,t) = q_0(x). \end{cases} \tag{3.85}$$

It is not clear to us if a solution theory is available when ξ is a space-time white noise, although this is the case we are ultimately interested in. However, if ξ is a smooth and Lipschitz potential, the Kolmogorov backward equation provides a representation of the solution as

$$q(t,x) = \mathsf{E}[q_0(X_0)|X_{-t} = x],$$

where X_t is the random diffusion [131] (see also [32, Eq. (2.9)])

$$dX_t = \frac{1}{\sqrt{\lambda}}\xi(t,X_t)dt + dB_t, \tag{3.86}$$

where the Brownian motion B is independent from ξ, and E denotes the expectation with respect to B, conditionally on the environment ξ. For a white noise potential ξ depending only on space, such diffusion can be constructed rigorously [181]. Note that q does not satisfy (3.85) for ξ white in time and smooth in space due to Ito corrections in the derivation of (3.85) from (3.86).

Let

$$\tilde{v}(t,\vec{x}) := \mathbb{E}\left[q(t,x_1)\dots q(t,x_k)\right], \tag{3.87}$$

where q solves (3.85). We claim that $\tilde{v}(t,\vec{x})$ satisfies (3.84) in the following formal sense. The following arguments are non rigorous, as we will discard many analytic difficulties such as exchanging derivatives with expectation without justification and we implicitly assume existence and uniqueness of solutions of (3.85) when ξ is a space time white noise.

By definition, a solution to (3.85) satisfies

$$q(t,x) = p_t * q_0(x) + \frac{1}{\sqrt{\lambda}}\int_0^t ds \int_{\mathbb{R}} dy p_{t-s}(x-y)\xi(s,y)\partial_y q(s,y),$$

where $*$ denotes convolution in space, and $p_t(x) = \frac{1}{\sqrt{2\pi t}}e^{-x^2/2t}$ denotes the heat kernel. Note that when $\xi(s,y)$ is not smooth in space, the integral against $\xi(s,y)\partial_y q(s,y)$ is not well defined even using Ito calculus. Let us assume for the moment that the covariance of the environment ξ is given by

$$\mathbb{E}\left[\xi(t,x)\xi(x,y)\right] = \delta(t-s)R(x-y), \tag{3.88}$$

where R is a smooth and compactly supported function. Considering the case $k=2$ for simplicity, we may write

$$\begin{aligned}
&\tilde{v}(t,x_1,x_2) = \\
&\frac{1}{\lambda}\mathbb{E}\left[\int_{\mathbb{R}} dy_1 \int_{\mathbb{R}} dy_2 \int_0^t ds_1 \int_0^t ds_2 p_{t-s_1}(x_1-y_1)p_{t-s_2}(x_2-y_2)\xi(s_1,y_1)\xi(s_2,y_2)\partial_{y_1}q(s_1,y_1)\partial_{y_2}q(s_2,y_2)\right] \\
&+\frac{1}{\sqrt{\lambda}}p_t * q_0(x_1)\ \mathbb{E}\int_{\mathbb{R}} dy_2 \int_0^t ds_2 p_{t-s_2}(x_2-y_2)\xi(s_2,y_2)\partial_{y_2}q(s_2,y_2) + 1 \leftrightarrow 2 + p_t * q_0(x_1)p_t * q_0(x_2),
\end{aligned}$$

where $1 \leftrightarrow 2$ denotes the previous term after exchanging indices 1 and 2. In the sequel we will discard the terms depending on $p_t * q_0$ which play no role in the following computation, because they solve the homogeneous heat equation. Using (3.88), we obtain

$$\tilde{v}(t, x_1, x_2) = \frac{1}{\lambda} \int_{\mathbb{R}} dy_1 \int_{\mathbb{R}} dy_2 R(y_1 - y_2) \int_0^t ds\, p_{t-s}(x_1 - y_1) p_{t-s}(x_2 - y_2) \mathbb{E}\left[\partial_{y_1} q(s, y_1) \partial_{y_2} q(s, y_2)\right] + \text{ terms depending on } p_t * q_0.$$

Thus, using that $p_t(x)$ solves the heat equation and $p_{t-s}(\cdot) \Rightarrow \delta_0(\cdot)$ as $s \to t$, we obtain

$$\partial_t \tilde{v}(t, x_1, x_2) = \frac{1}{\lambda} R(x_1 - x_2) \mathbb{E}\left[\partial_{x_1} q(t, x_1) \partial_{x_2} q(t, x_2)\right] + \frac{1}{2}(\partial_{x_1}^2 + \partial_{x_2}^2) \mathbb{E}\left[q(t, x_1) q(t, x_2)\right].$$

Finally, if R converges to a delta function, the noise ξ becomes a space time white noise, and assuming one can exchange the derivatives $\partial_{x_1}, \partial_{x_2}$ with the expectation, we obtain that $\tilde{v}(t, x_1, x_2)$ satisfies (3.84).

Remark 3.5.5. The fact that the function $\tilde{v}(t, \vec{x})$ defined in (3.87) satisfies the evolution (3.84) was essentially known in the physics literature. Indeed, the operator in the right hand side of (3.84) appears in [32, Eq. (2.17)] where it was shown to be related to the moments of a stochastic PDE [32, Eq. (2.2)] which has the same form as (3.85).

Chapter 4: GUE corners process in boundary-weighted six-vertex models

This chapter is based on the article [72] written by myself and Evgeni Dimitrov.

4.1 introduction

The six-vertex model is a well studied exactly solvable model in statistical mechanics. Linus Pauling introduced the model in 1935 to describe the residual entropy of ice crystals. In addition to its original purpose of describing ice, the six-vertex model has been useful in understanding other physical phenomena such as phase transitions in magnetism [30, 138].

In this chapter we consider a family of six-vertex models on the half-infinite strip $D_n = \mathbb{Z}_{\geq 0} \times \{1, \ldots, n\}$ where $n \in \mathbb{N}$. Specifically, the state space of the models is the set $\mathcal{P}_n$ consisting of all collections of n up-right paths, with nearest neighbor steps in D_n with the paths starting from the points $\{(0, i) : 1 \leq i \leq n\}$ and exiting the top boundary. We add the additional condition, that no two paths can share a horizontal or vertical edge, see Figure 4.1.

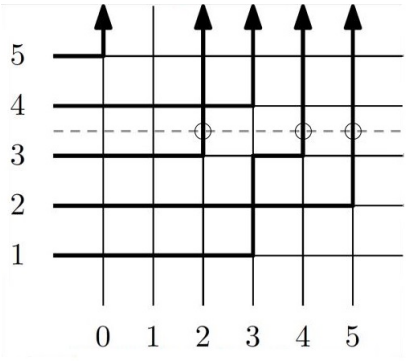

Figure 4.1: An example of a path collection π in $\mathcal{P}_5$. Here $\lambda_1^3(\pi) = 5$, $\lambda_2^3(\pi) = 4$, $\lambda_3^3(\pi) = 2$

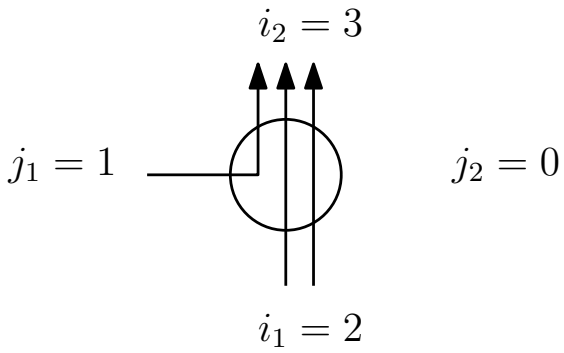

Figure 4.2: An example of a vertex of type $(i_1, j_1; i_2, j_2) = (2, 1; 3, 0)$

In the next few paragraphs we explain the types of probability measures we put on $\mathcal{P}_n$ (they are given in equation (4.2) below), but to accomplish this we need a bit of notation. A *signature*

of length n is a nonincreasing sequence $\lambda = (\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n)$ with $\lambda_i \in \mathbb{Z}$. We use Sign_n to denote the set of all signatures of length n, and use Sign_n^+ for the set of such signatures with $\lambda_n \geq 0$. To each collection of n up-right paths $\pi \in \mathcal{P}_n$ one can identify a sequence of signatures $\lambda^i(\pi) \in \mathsf{Sign}_i^+$ for $i = 1, \ldots, n$, where $(\lambda_1^i(\pi), \lambda_2^i(\pi), \ldots, \lambda_i^i(\pi))$ are the ordered x-coordinates at which the paths in π intersect the horizontal line $y = i + 1/2$, see Figure 4.1.

Given an up-right path collection $\pi \in \mathcal{P}_n$, each vertex is given a *vertex type* based on four numbers $(i_1, j_1; i_2, j_2)$, where i_1 and j_1 denote the number paths entering the vertex vertically and horizontally respectively, while i_2 and j_2 denote the number of paths leaving the vertex vertically and horizontally respectively, see Figure 4.2. For complex parameters s and u we define the following vertex weights

$$\begin{aligned} w_1 &= w(0,0;0,0) = 1, & w_2 &= w(1,1;1,1) = \frac{u - s^{-1}}{1 - su} \\ w_2 &= w(1,0;1,0) = \frac{1 - s^{-1}u}{1 - su}, & w_4 &= w(0,1;0,1) = \frac{u - s}{1 - su}, \\ w_5 &= w(1,0;0,1) = \frac{(1 - s^2)u}{1 - su}, & w_6 &= w(0,1;1,0) = \frac{1 - s^{-2}}{1 - su}. \end{aligned} \tag{4.1}$$

This nonintuitive parametrization of weights by s and u comes from [46], where it is important in defining a higher spin generalization of the six-vertex model. Later in (4.8) we present the higher spin vertex weights, and one obtains the weights in (4.1) by setting $q = s^{-2}$ in (4.8).

For $\pi \in \mathcal{P}_n$ we let $\pi(i, j)$ denote the vertex type of the vertex at position (i, j) in the path collection π. Given complex numbers s and u, and a function $f : \mathsf{Sign}_n^+ \to \mathbb{C}$ we define the weight of a path collection $\pi \in \mathcal{P}_n$ by

$$\mathcal{W}^f(\pi) = f(\lambda^n(\pi)) \prod_{i=1}^{\infty} \prod_{j=1}^{n} w(\pi(i, j)).$$

All but finitely many $\pi(i, j)$ are equal to $(0, 0; 0, 0)$ and have weight 1 by (4.1), so the product is well defined. If one chooses u and s in $\mathbb{C}$ and the function f so that the weights $\mathcal{W}^f(\pi)$ are nonnegative, not all zero and summable then one can use the weights $W^f(\pi)$ to define a probability measure on

$\mathcal{P}_n$ through

$$\mathbb{P}^f(\pi) = (Z^f)^{-1} \cdot \mathcal{W}^f(\pi), \text{ where } Z^f := \sum_{\pi \in \mathcal{P}_n} \mathcal{W}^f(\pi). \tag{4.2}$$

Equation (4.2) gives the general form of the measures we study in this chapter. In plain words $\mathbb{P}^f$ is the usual six-vertex measure except that the path collections π are reweighed based on their top boundary, namely $\lambda^n(\pi)$, through the boundary weight function f.

Remark 4.1.1. When we go to our main results we will take $u > s > 1$ above. In the usual weight parametrization of the six-vertex model we have that

$$a_1 = 1, \quad a_2 = \frac{u - s^{-1}}{su - 1}, \quad b_1 = \frac{1 - s^{-1}u}{1 - su}, \quad b_2 = \frac{u - s}{su - 1}, \quad c_1 = \frac{(1 - s^2)u}{1 - su}, \text{ and } c_2 = \frac{1 - s^{-2}}{su - 1}.$$

We mention that the latter weights are the absolute values of those in (4.1), where ultimately the sign difference will be absorbed in the boundary weight function f of the model so that $\mathcal{W}^f(\pi) \geq 0$ for all $\pi \in \mathcal{P}_n$. Associated with the six weights is an *anisotropy parameter* Δ, given by

$$\Delta(a_1, a_2, b_1, b_2, c_1, c_2) = \frac{a_1 a_2 + b_1 b_2 - c_1 c_2}{2\sqrt{a_1 a_2 b_1 b_2}}, \tag{4.3}$$

which is believed to be directly related with the qualitative and quantitative properties of the model, see [153]. The choice of weights as in (4.1) with $u > s > 1$ corresponds to $\Delta > 1$, which is known as the *ferroelectric phase* of the six-vertex model.

There are many different choices of parameters and functions f that lead to meaningful measures in (4.2). For example, if $f(\lambda) = 0$ unless $\lambda_{n-i+1} = i - 1$ for $i = 1, \ldots, n$ the measure in $\mathbb{P}^f$ becomes the six-vertex model with *domain wall boundary condition* (DWBC), [123]. Another special case of the measures in (4.2) includes the case when $u > s > 1$ and

$$f(\lambda) = \mathsf{G}^c_\lambda(\rho) := (-1)^n \cdot \mathbb{1}_{m_0 = 0} \prod_{i=1}^{\infty} \mathbb{1}_{m_i \leq 1} \prod_{j=1}^{n} (-s)^{\lambda_j}, \tag{4.4}$$

where $\lambda = 0^{m_0} 1^{m_1} 2^{m_2} \ldots$. In the latter notation m_i is the number of times i appears in the list

$(\lambda_1, \ldots, \lambda_n)$ and $\mathbb{1}_E$ is the indicator function of the set E. With this choice of parameters and function f, the measure $\mathbb{P}^f$ becomes what is known as stochastic six-vertex model, see e.g. [98], [39], with parameters

$$b_1 = \frac{1 - s^{-1}u}{1 - su}, \quad b_2 = \frac{s^2 - su}{1 - su}.$$

For a quick proof of the latter statement we refer the reader to [46, Section 6.5].

Different choices of the boundary weight function f lead to qualitatively different behavior of the measures $\mathbb{P}^f$ in (4.2). We illustrate this point by comparing the DWBC and the stochastic six-vertex model we just introduced. In order to begin understanding the qualitative differences between these two models we need to discuss the *pure states* (or the ergodic, translation-invariant Gibbs measures) of the six-vertex model. For this we follow [3], see also [60, Section 1.2.1].

A prediction in [50], which has been very recently partially verified in [3], states that the pure states μ of the ferroelectric six-vertex model are parametrized by a *slope* $(s,t) \in [0,1]^2$, where s and t denote the probabilities that a given vertical and horizontal edge is occupied under μ. For a certain open lens-shaped set $\mathfrak{H} \subset [0,1]^2$ one has the following characterization of pure states for the ferroelectric six-vertex model (here $\overline{\mathfrak{H}} = \mathfrak{H} \cup \partial\mathfrak{H}$):

1. *Nonexistence*: If $(s,t) \in \mathfrak{H}$, then there are no pure states $\mu_{s,t}$ of slope (s,t).

2. *KPZ States*: If $(s,t) \in \partial\mathfrak{H}$, then $\mu_{s,t}$ should exhibit Kardar-Parisi-Zhang (KPZ) behavior.

3. *Liquid States*: If $(s,t) \in (0,1)^2 \backslash \overline{\mathfrak{H}}$, then $\mu_{s,t}$ should exhibit Gaussian free field (GFF) behavior.

4. *Frozen States*: If (s,t) is on the boundary of $[0,1]^2$, then $\mu_{s,t}$ should be frozen.

From the above conjectural classification, [3] established the nonexistence statement (1) and proved the existence and uniqueness of KPZ states (2) for all $(s,t) \in \partial\mathfrak{H}$. It is worth mentioning that the above classification sharply contrasts the one for dimer models. Specifically, the pure states in dimer models were classified in [171] and [120] and they come in three types. The first is *frozen*, where the associated height function is basically deterministic; the second is *gaseous*, where the

variance of the height function is bounded but non-zero; the third is *liquid*, where the hight function fluctuations diverge logarithmically in the lattice size. In particular, for dimer models there are no Nonexistence or KPZ pure states.

Going back to our previous discussion, the stochastic six-vertex model considered in [39], which corresponds to f as in (4.4), was shown to asymptotically have a phase diagram that consists of two frozen regions, i.e. regions where the local behavior of the model is described by Frozen States, and a non-frozen region, where one observes solely KPZ States, see Figure 4.3. More specifically, in [39] it was shown that the one-point marginals of the height function $h(x, y)$, which at a location (x, y) counts the number of horizontal edges crossed by the vertical segment connecting $(x, 0)$ and (x, y) in the non-frozen region III of Figure 4.3 are asymptotically governed by the GUE Tracy-Widom distribution [189]. This type of behavior is characteristic of models in the KPZ universality class (for more background on this class we refer to the excellent survey [63]). For the DWBC six-vertex model a very different phase diagram is expected, although we emphasize that it has not been established rigorously. Specifically, for the DWBC it is expected that as n becomes large the model again develops macroscopic frozen regions that are separated by a non-frozen region where one observes solely Liquid States. The only instance where this has been rigorously established is when $\Delta = 0$, which is the free fermion point of the model, see [118], [119]. When $\Delta = 0$ the six-vertex model falls into the framework of the dimer models, which is what enables its precise mathematical analysis. We mention; however, that there are non-rigorous physics works and numerical simulations that indicate that for $\Delta < 1$ the six vertex model with DWBC has solely Liquid States in the non-frozen region, and by analogy with the dimer models the fluctuations of those are no longer KPZ, but rather governed by a suitable pullback of the Gaussian free field, [96]. In the ferroelectric $\Delta > 1$ case similar heuristics suggest that one observes only frozen states [59, 180].

The above paragraphs explain that by picking different boundary weight functions f we can obtain qualitatively different phase diagrams for our six-vertex model. In this chapter we consider

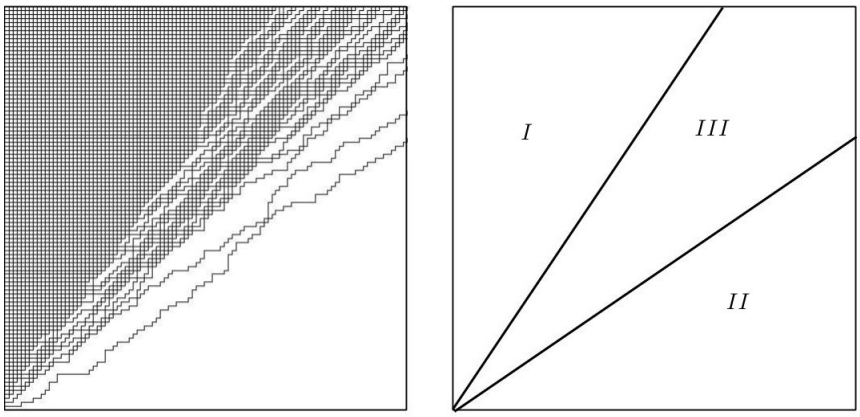

Figure 4.3: The left picture represent a sample of $\mathbb{P}^f$ with f as in (4.4) for the parameters $n = 100$, $u = 2$, $s^{-2} = 0.5$. The picture on the right side depicts the phase diagrams for these measures when n is large. The regions I and II correspond to Frozen States and region III corresponds to KPZ States

a very special class of boundary functions f. This class will be described in the next section, where the definition of the models, some of their structural properties and main result we prove for them are presented. In the remainder of this section we explain the very high level motivations that have guided our choice of f.

First of all, our discussion above indicates that for the stochastic six-vertex model of [39] the non-frozen region consists entirely of KPZ States, while for the DWBC (at least conjecturally) it consists solely of Liquid States (or states with Gaussian statistics). A natural question is whether we can find a boundary weight function f for which both types of pure states co-exist in the non-frozen region of the model. A second point is that, for general functions f, the asymptotic analysis for $\mathbb{P}^f$ is prohibitively complicated – indeed even for the DWBC the phase diagram is largely conjectural, and so one is inclined to consider special boundary weight functions f for which the analysis of the model is tractable. Our choice of f is motivated by our desire that the resulting model satisfies these two properties.

4.1.1 Model and results

In this chapter we study a special case of (4.2) when the boundary weight function f is given by

$$f(\lambda) = \sum_{\mu \in \mathsf{Sign}_n^+} \mathsf{G}^c_\mu(\rho)\mathsf{G}^c_{\lambda/\mu}(v, \dots, v). \tag{4.5}$$

In (4.5) the function $\mathsf{G}^c_\mu(\rho)$ is as in (4.4) and the functions $\mathsf{G}^c_{\lambda/\mu}$ are a remarkable class of symmetric rational functions, which were introduced in [34]. In this chapter one can find the definition of $\mathsf{G}^c_{\lambda/\mu}$ in Definition 4.2.1, and these functions depend on M complex variables $v_1, \dots, v_M$ that have all been set to the same complex number v in (4.5). We mention that $\mathsf{G}^c_{\lambda/\mu}$ are one-parameter generalizations of the classical (skew) Hall-Littlewood symmetric functions [142] and carry the name of (skew) *spin Hall-Littlewood symmetric functions*, see [49].

One can check that if $v^{-1} > u > s > 1$ then the measure $\mathbb{P}^f$ from (4.2) with f as in (4.5) is a well-defined probability measure, see Section 4.2.2. We will denote this measure by $\mathbb{P}^{N,M}_{u,v}$.

Even though the choice of f in (4.5) seems complicated we emphasize that the resulting measure $\mathbb{P}^f$ enjoys many remarkable properties and its asymptotic structure appears to be rich and interesting. We elaborate on these points in the next few paragraphs, summarizing some results from [71] where this model was studied in detail.

First of all, the choice of f as in (4.5) makes the model integrable and the distribution $\mathbb{P}^f$ analogous to the *ascending Macdonald processes* of [36]. What plays the role of the (skew) Macdonald symmetric functions $P_{\lambda/\mu}$ and their duals $Q_{\lambda/\mu}$ is a class of symmetric rational functions $\mathsf{F}_{\lambda/\mu}$ and their duals $\mathsf{G}^c_{\lambda/\mu}$ that were mentioned above. The functions $\mathsf{F}_{\lambda/\mu}, \mathsf{G}^c_{\lambda/\mu}$ enjoy many of the same properties as the Macdonald symmetric functions, including branching rules, orthogonality relations, (skew) Cauchy identities and so on. One consequence of the integrability of the model that can be appreciated by readers unfamiliar with symmetric function theory is that the partition

function Z^f for our choice of f in (4.5) takes the following extremely simple product form

$$Z^f = (s^{-2}; s^{-2})_n \left(\frac{1 - s^{-1}u}{1 - su}\right)^n \left(\frac{1 - s^{-2}uv}{1 - uv}\right)^{nM}, \text{ where } (a; q)_m = (1 - a)(1 - aq) \cdots (1 - aq^{m-1}).$$

The latter formula for the partition function is recalled in Section 4.2.2 in this chapter.

Another consequence of the integrability of the model is the fact that it is self-consistent in the following sense. Suppose that we sample a path collection π on $\mathcal{P}_n$ according to $\mathbb{P}^{n,M}_{u,v}$ and then project the path collection to the first k rows where $1 \leq k \leq n$. The resulting path collection is now a random element in $\mathcal{P}_k$ and its distribution is precisely $\mathbb{P}^{k,M}_{u,v}$ – we recall this in Lemma 4.2.12. This self-consistency of the measures $\mathbb{P}^{n,M}_{u,v}$ for $n \in \mathbb{N}$ allows us for example to define a measure on up-right paths on the whole of $\mathbb{Z}^2_{\geq 0}$ whose projection to the bottom n rows has law $\mathbb{P}^{n,M}_{u,v}$.

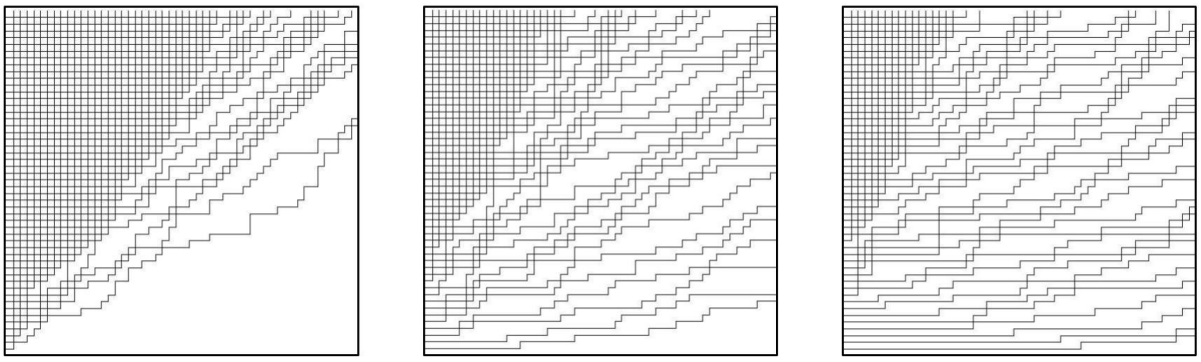

Figure 4.4: The pictures represent samples of the Markov chain $\{X_m\}_{m=0}^{\infty}$ when $n = 50$ at times $m = 0$, $m = 50$ and $m = 100$. The parameters of the process are $s^{-2} = 0.5$, $u = 2$ and all $v = 0.25$

Yet another consequence of the integrability of the model is given by the fact that for fixed n and varying $m \in \mathbb{Z}_{\geq 0}$ the measures $\mathbb{P}^{n,M}_{u,v}$ can be stochastically linked as we next explain. One can interpret the distribution $\mathbb{P}^{n,m}_{u,v}$ as the time m distribution of a Markov chain $\{X_m\}_{m=0}^{\infty}$ taking values in $\mathcal{P}_n$ for each m. This Markov chain is started from the stochastic six-vertex model at time zero, and its dynamics are governed by sequential update rules. For more details and a precise formulation we refer the reader to [46, Section 6] as well as [71, Section 8] where an exact sampling algorithm of this process was developed by one of the authors. For a pictorial description of how the configurations X_m evolve as time increases see Figure 4.4. This interpretation is similar to

known interpretations of the Schur process and Macdonald process as fixed time distributions of certain Markov processes, see [35, 36].

The above few paragraphs explained some of the structure and relationships between the measures $\mathbb{P}_{u,v}^{N,M}$ for varying $N, M \in \mathbb{N}$. These measures arise as degenerations of the higher-spin vertex models that were studied in [46], which is the origin of their integrability. For the purposes of this chapter, the main consequence of the integrability of the model that is utilized is that one has suitable for asymptotic analysis formulas for the one-dimensional projections of $\mathbb{P}_{u,v}^{N,M}$. This is what makes the analysis of the model tractable, which as we recall from the end of the previous Section is one of our desired properties.

Our primary probabilistic interest in the measures $\mathbb{P}_{u,v}^{N,M}$ comes from the fact that as $N, M \to \infty$ the phase diagram of the model (at least conjecturally) exhibits all three types of pure states – Frozen, Liquid and KPZ. The presence of all three types of pure states is the second high-level motivation behind our choice of f as in (4.5) and we illustrate the phase diagram in Figure 4.5. The phase diagram in Figure 4.5, which corresponds to $\mathbb{P}_{u,v}^{N,M}$ when N and M are large, should

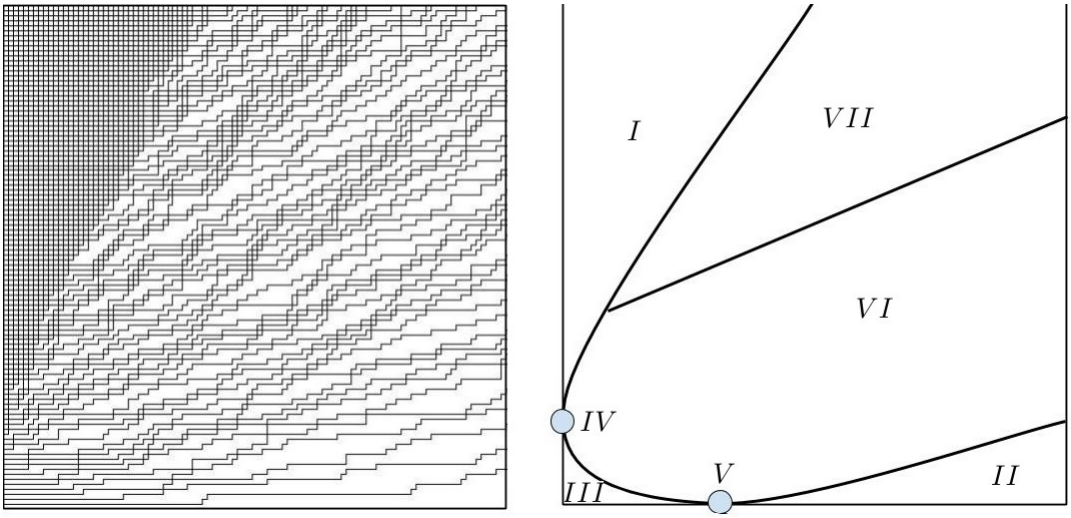

Figure 4.5: The picture on the left represent a sample of $\mathbb{P}_{u,v}^{N,M}$ with $N = M = 100$, $u = 2$, $s^{-2} = 0.5$, $v = 0.25$. The picture on the right represents the (conjectural) phase diagram of the model as $N, M \to \infty$

be compared to the one in Figure 4.3, which corresponds to the stochastic six-vertex model or

equivalently to the measure $\mathbb{P}_{u,v}^{N,0}$ (recall that the measures $\mathbb{P}_{u,v}^{N,m}$ were stochastically linked through a Markov chain with time zero distribution given precisely by the stochastic six-vertex model). At least based on the simulations one observes that as the vertex model evolves in time $m = 0, \ldots, M$ the frozen regions I and II from the stochastic six-vertex model in Figure 4.3 begin to deform and a new frozen region, denoted by III in Figure 4.5 and consisting of vertices of type $(0, 1; 0, 1)$, is formed near the origin. With this new frozen region two new points IV and V are formed. These are sometimes referred to as *turning points* and they arise where two different frozen regions meet each other. Furthermore, our prediction is that, under the Markovian dynamics evolving the six-vertex model, the KPZ cone (i.e. region III in Figure 4.3) that is present at time $m = 0$ is translated away from the origin to region VII and a new GFF region (denoted by VI in Figure 4.5) takes its place. We mention here that the exact nature of the Markovian dynamics is not important for this chapter. The reason we mention it is to emphasize that the stochastic six-vertex model and the measures $\mathbb{P}_{u,v}^{N,M}$ we consider here are related to each other and the presence of the KPZ region VII in $\mathbb{P}_{u,v}^{N,M}$ can be traced back to the presence of region III in $\mathbb{P}_{u,v}^{N,0}$. If the same dynamics are run from a different initial configuration one may very well see a completely different phase diagram than the one in Figure 4.5.

As can be seen from Figure 4.5 the asymptotics of $\mathbb{P}_{u,v}^{N,M}$ as $N, M \to \infty$ appear to be quite complex. A long term program, initiated in [71], is to rigorously establish the phase diagram in Figure 4.5. So far only the asymptotics near the point IV have been understood. Specifically, in [71] one of the authors showed that near IV a certain configuration of empty edges converges to the GUE-corners process, we define the latter here. Recall that the Gaussian Unitary Ensemble (GUE) is a measure on Hermitian matrices $\{X_{ij}\}_{i,j=1}^{k}$ with density proportional to $e^{-Tr(X^2)/2}$ with respect to Lebesgue measure. For $1 \leq r \leq k$, let $\lambda_1^r \leq \lambda_2^r \leq \cdots \leq \lambda_r^r$ denote the ordered eigenvalues of the submatrix $\{X_{ij}\}_{i,j=1}^{r}$ of X. The joint law of the eigenvalues $\{\lambda_i^j\}_{1 \leq i \leq j \leq k}$ is called the *GUE-corners process* of rank k (or the GUE-minors process). The appearance of the GUE-corners process has been established in related contexts for random lozenge tilings in [113, 149, 154] and the uniform six-vertex model with domain-wall boundary conditions [95]. It is believed

to be a universal scaling limit near points separating two different frozen regions such as the point IV.

This chapter, is a continuation of the program initiated in [71] of establishing the phase diagram in Figure 4.5. Specifically, in Figure 4.5 the point V is another turning point, and in this chapter, we show that the statistics of the model $\mathbb{P}_{u,v}^{N,M}$ near this point are also described by the GUE-corners process. Before we state our main result we give our choice of parameters and some notation.

Definition 4.1.2. We assume that $v, u, s \in (0, \infty)$ satisfy $v^{-1} > u > s > 1$. With this choice of parameters we define the constants

$$\begin{aligned} a &= \frac{v\left(u - s^{-1}\right)\left(s^{-1}u - 1\right)}{(1-uv)(1-s^{-2}uv)}, \qquad b = \frac{(s^2-1)}{(u-s)(1-su)} \\ c &= \frac{1}{2}\left(a\left(\frac{1}{(u-s)^2} - \frac{s^2}{(1-su)^2}\right) - \frac{s^{-4}v^2}{(1-s^{-2}uv)^2} + \frac{v^2}{(1-uv)^2}\right), \qquad d = \frac{-\sqrt{2c}}{b}. \end{aligned} \tag{4.6}$$

If $v^{-1} > u > s > 1$ one observes that

$$a > 0, \qquad b < 0, \qquad c > 0, \qquad d > 0.$$

See Lemma 4.5.1 in the main text for a verification of this fact.

The main result of the chapter is as follows.

Theorem 4.1.3. *Suppose that u, v, s, a, d are as in Definition 4.6 and $k \in \mathbb{N}$ is given. Suppose that $N(M)$ is a sequence of integers such that $N(M) \geq k$ for all M and let $\mathbb{P}_{u,v}^{N,M}$ be the measure on collections of paths $\pi \in \mathcal{P}_N$ as earlier in the section. Define the random vector $Y(N, M; k)$ through*

$$Y_i^j(N, M; k) = \frac{\lambda_{j-i+1}^j(\pi) - aM}{d\sqrt{M}} \text{ for } 1 \leq i \leq j \leq k. \tag{4.7}$$

Then the sequence $Y(N, M; k)$ converges weakly to the GUE-corners process of rank k as $M \to \infty$.

Remark 4.1.4. In (4.7) we reverse the order of λ_i^j because the usual convention for signatures $\lambda = (\lambda_1, \ldots, \lambda_n)$ demands that λ_i be sorted in decreasing order, while for the eigenvalues of a

random matrix the usual convention is that they are sorted in increasing order.

We mention here that while the asymptotic behaviors near IV and V are similar, the arguments used to establish the two results are quite different. The arguments in [71] rely on a remarkable class of difference operators, which can be used to extract averages of observables for $\mathbb{P}_{u,v}^{N,M}$ near the left boundary of the model. These observables become useless for accessing the asymptotic behavior of the base of the model and consequently our approach in this chapter is completely different, and arguably more direct as we explain here. In the remainder of this section we give an outline of our approach to proving Theorem 4.1.3. The discussion below will involve certain expressions that will be properly introduced in the main text, and which should be treated as black boxes for the purposes of the outline.

Using the integrability of the model we obtain the formula

$$\mathbb{P}_{u,v}^{N,M}(\lambda_1^k(\pi) = \mu_1, \cdots, \lambda_1^k(\pi) = \mu_k) \propto \mathsf{F}_\mu([u]^k) \cdot f(\mu; [v]^M, \rho),$$

for any $\mu = (\mu_1, \ldots, \mu_k) \in \mathsf{Sign}_k^+$. A generalization of this fact appears as Lemma 4.2.12 in the main text. We then derive certain combinatorial estimates for $\mathsf{F}_\mu([u]^k)$ and a contour integral formula for $f(\mu; [v]^M, \rho)$ in Section 4.3, which are both suitable for studying the $M \to \infty$ limit of these expressions (for the function $\mathsf{F}_\mu([u]^k)$ the dependence on M is reflected in the scaling of the signature μ). The limit of the contour integral formula for $f(\mu; [v]^M, \rho)$ is derived in Section 4.5 using a steepest descent argument, while the combinatorial estimates for $\mathsf{F}_\mu([u]^k)$ prove sufficient for taking its limit. Combining our two asymptotic results for $\mathsf{F}_\mu([u]^k)$ and $f(\mu; [v]^M, \rho)$ we can prove that the sequence of random vectors in $\mathbb{R}^k$, given by $Y^k(N, M) = \left(Y_1^k(N, M; k), \ldots, Y_k^k(N, M; k)\right)$ with $Y(N, M; k)$ as in Theorem 4.1.3 converges to the measure of the ordered eigenvalues of a random GUE matrix $\mu_{\mathrm{GUE}}^k(dx_1, ..., dx_k)$, given by

$$\mu_{\mathrm{GUE}}^k(dx_1, ..., dx_k) = \mathbf{1}\{x_k > x_{k-1} > \cdots > x_1\} \left(\frac{1}{\sqrt{2\pi}}\right)^k \cdot \frac{1}{\prod_{i=1}^{k-1} i!} \cdot \prod_{1 \le i < j \le k} (x_i - x_j)^2 \prod_{i=1}^{k} e^{-\frac{x_i}{2}} dx_i.$$

The last statement appears as Proposition 4.4.3 in the text.

The above paragraph explains how we show that the top row of $Y(N, M; k)$ converges to the top row of the GUE-corners process of rank k. To obtain the full convergence statement we combine our top row convergence statement with the general formalism, introduced in [71], involving Gibbs measures on interlacing arrays. In more detail, the top-row convergence of $Y(N, M; k)$ and the interlacing conditions

$$\lambda_1^{i+1}(\pi) \geq \lambda_1^i(\pi) \geq \lambda_2^{i+1}(\pi) \geq \lambda_2^i(\pi) \geq \cdots \geq \lambda_i^i(\pi) \geq \lambda_{i+1}^{i+1}(\pi),$$

for $i = 1, \ldots, k-1$ are enough to conclude the tightness of the full vector $Y(N, M; k)$ as $M \to \infty$. For each N, M the measures $\mathbb{P}_{u,v}^{N,M}$ satisfy what we call the *six-vertex Gibbs property* and in the $M \to \infty$ limit this property becomes what is known as the *continuous Gibbs property*, see Section 4.4.2. Combining the latter statements, one can conclude that any subsequential limit of $Y(N, M; k)$ as $M \to \infty$ has top row distribution $\mu_{\mathrm{GUE}}^k(dx_1, ..., dx_k)$ and satisfies the continuous Gibbs property, and these two characteristics are enough to identify this limit with the GUE-corners process of rank k. As the sequence $Y(N, M; k)$ is tight and all subsequential limits are the same and equal to the GUE-corners of rank k, we conclude the weak convergence of $Y(N, M; k)$. This argument is given in Section 4.4.2.

4.1.2 Outline of the chapter

In Section 4.2 we define and state some facts about the functions $\mathsf{F}_{\lambda/\mu}$ and $\mathsf{G}^c_{\lambda/\mu}$, we also define inhomogeneous versions of $\mathbb{P}_{u,v}^{N,M}$ called $\mathbb{P}_{\mathbf{u},\mathbf{v}}$ and give formulas for their projections. In Section 4.3 we derive certain combinatorial estimates for $\mathsf{F}_\mu([u]^k)$ and a contour integral formula for $f(\mu; \mathbf{v}, \rho)$, where the latter appear in our projection formulas for $\mathbb{P}_{\mathbf{u},\mathbf{v}}$ from Section 4.2. In Section 4.4 we prove Theorem 4.1.3 modulo Lemma 4.4.5 , which is proved in Section 4.5.

4.1.3 Acknowledgements

The authors would like to thank Ivan Corwin and Ioannis Karatzas for many helpful conversations. M. R. is partially supported by the NSF grant DMS:1664650. E.D. is partially supported by the Minerva Foundation Fellowship.

4.2 Measures on up-right paths

In this section we provide some results about a certain class of measures $\mathbb{P}_{\mathbf{u},\mathbf{v}}$ that are inhomogeneous analogues of the measures $\mathbb{P}^{N,M}_{u,v}$ from Section 4.1.1. For the most part, this section summarizes the results in [71, Section 2].

4.2.1 Symmetric rational functions

In this section we introduce some necessary notation from [46] and summarize some of the results from the same paper. A *signature* of length N is a sequence $\lambda = (\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N)$ with $\lambda_i \in \mathbb{Z}$ for $i = 1, \ldots, N$. A signature λ will sometimes be represented by $0^{m_0(\lambda)}1^{m_1(\lambda)}2^{m_2(\lambda)}\ldots$ where $m_i(\lambda) := |\{j : \lambda_j = i\}|$ is the number of times i appears in the list $(\lambda_1, \ldots, \lambda_N)$. We denote by Sign_N the set of all signatures of length N and by Sign^+_N the set of signatures of length N with $\lambda_N \geq 0$. We also denote by $\mathsf{Sign}^+ := \sqcup_{N\geq 0}\mathsf{Sign}^+_N$ the set of all non-negative signatures. We recall for later use the q-Pochhammer symbol $(a;q)_n := (1-a)(1-qa)\cdots(1-q^{n-1}a)$.

In what follows, we define the *weight* of a finite collection of up-right paths in some region D of $\mathbb{Z}^2$, which is equal to the product of the weights of all vertices that belong to the path collection. Throughout we will always assume that the weight of an empty vertex is 1 and so alternatively the weight of a path configuration can be defined through the product of the weights of all vertices in D. Figure 4.6 gives examples of collections of up-right paths.

The path configuration at a vertex is determined by four non-negative integers $(i_1, j_1; i_2, j_2)$, where i_1 (resp. i_2) is the number of incoming (resp. outgoing) vertical paths, and j_1 (resp. j_2) is the number of incoming (resp. outgoing) horizontal paths, see Figure 4.2. If the path configuration of

a vertex is $(i_1, j_1; i_2, j_2)$ we will also say that the vertex is of type $(i_1, j_1; i_2, j_2)$. Vertex weights will be given as a function of these four numbers. We require the number of paths entering and leaving each vertex to be the same, i.e. $i_1 + j_1 = i_2 + j_2$, and we will constrain the horizontal number of paths by $j_1, j_2 \in \{0, 1\}$ (the weight of any vertex that does not satisfy these two conditions is 0).

We consider two sets of vertex weights depending on parameters s and q (these are fixed throughout this section), and a *spectral parameter* u. The first set of vertex weights is given by

$$\begin{aligned} w_u(g,0;g,0) &= \frac{1-sq^g u}{1-su}, & w_u(g+1,0;g,1) &= \frac{(1-s^2q^g)u}{1-su} \\ w_u(g,1;g,1) &= \frac{u-sq^g}{1-su}, & w_u(g,1;g+1,0) &= \frac{1-q^{g+1}}{1-su}, \end{aligned} \tag{4.8}$$

where g is any nonnegative integer and all other weights are assumed to be 0. The second set of weights, called *conjugated* weights, are defined by

$$\begin{aligned} w^c_u(g,0;g,0) &= \frac{1-sq^g u}{1-su}, & w^c_u(g+1,0;g,1) &= \frac{(1-q^{g+1})u}{1-su} \\ w^c_u(g,1;g,1) &= \frac{u-sq^g}{1-su}, & w^c_u(g,1;g+1,0) &= \frac{1-s^2q^g}{1-su}, \end{aligned} \tag{4.9}$$

where as before $g \in \mathbb{Z}_{\geq 0}$ and all other weights are assumed to be 0. For more background and motivation for this particular choice of weights we refer the reader to [46, Section 2].

Let us fix $n \in \mathbb{N}$, n indeterminates $u_1, \ldots, u_n$ and the region $D_n = \mathbb{Z}_{\geq 0} \times \{1, \ldots, n\}$. Let π be a finite collection of up-right paths in D_n, which end in the top boundary, but are allowed to start from the left or bottom boundary of D_n. We denote the path configuration at vertex $(i, j) \in D_n$ by $\pi(i, j)$. Then the weight of π with respect to the two sets of weights above is defined through

$$\mathcal{W}(\pi) = \prod_{i=0}^{\infty} \prod_{j=1}^{n} w_{u_j}(\pi(i,j)), \qquad \mathcal{W}^c(\pi) = \prod_{i=0}^{\infty} \prod_{j=1}^{n} w^c_{u_j}(\pi(i,j)). \tag{4.10}$$

Note that from (4.8) and (4.9) we have that $w_u(0,0;0,0) = 1 = w^c_u(0,0;0,0)$ and so the above products are in fact finite. With the above notation we define the following partition functions.

Definition 4.2.1. Let $N, n \in \mathbb{Z}_{\geq 0}$, $\lambda, \mu \in \mathsf{Sign}^+_N$ and $u_1, \ldots, u_n \in \mathbb{C}$ be given. Let $\mathcal{P}^c_{\lambda/\mu}$ be the collection of up-right paths π on D_n, which

- begin with N vertical edges $(\mu_i, 0) \to (\mu_i, 1)$ for $i = 1, \ldots, N$ along the horizontal axis;
- end with N vertical edges $(\lambda_i, n) \to (\lambda_i, n+1)$ for $i = 1, \ldots, N$.

Then we define

$$\mathsf{G}^c_{\lambda/\mu}(u_1, \ldots, u_n) = \sum_{\pi \in \mathcal{P}^c_{\lambda/\mu}} \mathcal{W}^c(\pi). \tag{4.11}$$

We also use the abbreviation G^c_λ for $\mathsf{G}^c_{\lambda/(0,0,\ldots,0)}$.

For the second set of weights we have an analogous definition.

Definition 4.2.2. Let $N, n \in \mathbb{Z}_{\geq 0}$, $\mu \in \mathsf{Sign}^+_N$, $\lambda \in \mathsf{Sign}^+_{N+n}$ and $u_1, \ldots, u_n \in \mathbb{C}$ be given. Let $\mathcal{P}_{\lambda/\mu}$ be the collection of up-right paths π on D_n, which

- begin with edges $(\mu_i, 0) \to (\mu_i, 1)$ for $i = 1, \ldots, N$ along the bottom boundary of D_n and with edges $(-1, y) \to (0, y)$ for $y = 1, \ldots, n$ along the left boundary;
- end with $N + n$ vertical edges $(\lambda_i, k) \to (\lambda_i, n+1)$ for $i = 1, \ldots, N+n$.

Then we define

$$\mathsf{F}_{\lambda/\mu}(u_1, \ldots, u_n) = \sum_{\pi \in \mathcal{P}_{\lambda/\mu}} \mathcal{W}(\pi). \tag{4.12}$$

We also use the abbreviation $\mathsf{F}_\lambda = \mathsf{F}_{\lambda/\varnothing}$.

Path configurations that belong to $\mathcal{P}_{\lambda/\mu}$ and $\mathcal{P}^c_{\lambda/\mu}$ are depicted in Figure 4.6. In the definitions above we define the weight of a collection of paths to be 1 if it has no interior vertices. Also, the weight of an empty collection of paths is 0.

Below we summarize some of the properties of the functions $\mathsf{G}^c_{\lambda/\mu}$ and $\mathsf{F}_{\lambda/\mu}$ in a sequence of propositions; we refer the reader to [46, Section 4] for the proofs. We mention here that the statements we write below for $\mathsf{G}^c_{\lambda/\mu}$ appear in [46, Section 4] for a slightly different but related

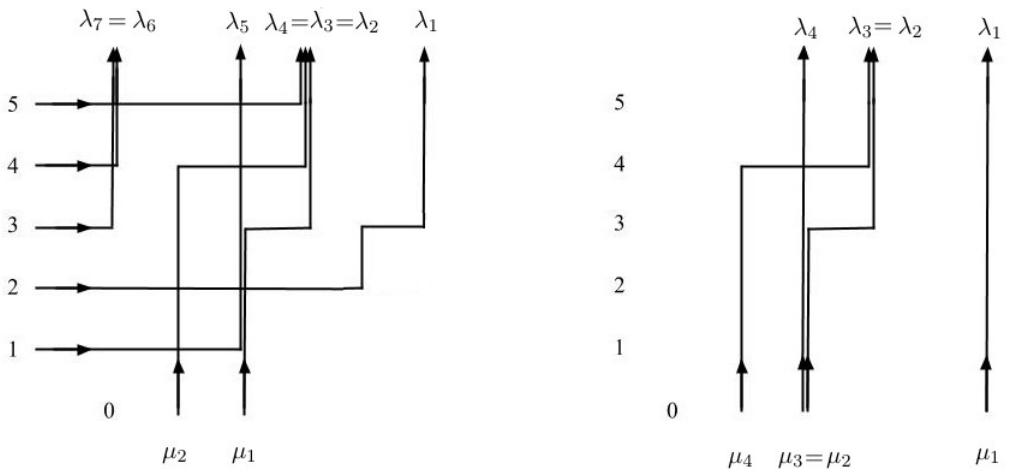

Figure 4.6: Path collections belonging to $\mathcal{P}_{\lambda/\mu}$ (left) and $\mathcal{P}^c_{\lambda/\mu}$ (right).

function $\mathsf{G}_{\lambda/\mu}$. The function $\mathsf{G}_{\lambda/\mu}$ has the same definition as $\mathsf{G}^c_{\lambda/\mu}$ except that one uses the vertex weights (4.8) rather than the conjugated weights (4.9). One directly checks that the two sets of weights are related through the equation

$$w^c_u(i_1, j_1; i_2, j_2) = \frac{(q;q)_{i_1}(s^2;q)_{i_2}}{(q;q)_{i_2}(s^2;q)_{i_1}} \cdot w_u(i_1, j_1; i_2, j_2),$$

which results in the relation

$$\mathsf{G}^c_{\lambda/\mu} = \frac{c(\lambda)}{c(\mu)} \cdot \mathsf{G}_{\lambda/\mu}, \text{ where } c(\lambda) = \prod_{k=0}^{\infty} \frac{(s^2;q)_{n_k}}{(q;q)_{n_k}} \text{ for } \lambda = 0^{n_0}1^{n_1}2^{n_2}\ldots. \tag{4.13}$$

Proposition 4.2.3. *[46, Proposition 4.5] The functions $\mathsf{F}_{\lambda/\mu}(u_1, \ldots, u_n)$ and $\mathsf{G}^c_{\lambda/\mu}(u_1, \ldots, u_n)$ defined above are rational symmetric functions in the variables $u_1, \ldots, u_n$.*

Proposition 4.2.4. *[46, Proposition 4.6]* **1.** *For any $N, n_1, n_2 \in \mathbb{Z}_{\geq 0}$, $\mu \in \mathsf{Sign}^+_N$ and $\lambda \in \mathsf{Sign}^+_{N+n_1+n_2}$ one has*

$$\mathsf{F}_{\lambda/\mu}(u_1, \ldots, u_{n_1+n_2}) = \sum_{\kappa \in \mathsf{Sign}^+_{N+n_1}} \mathsf{F}_{\lambda/\kappa}(u_{n_1+1}, \ldots, u_{n_1+n_2})\mathsf{F}_{\kappa/\mu}(u_1, \ldots, u_{n_1}). \tag{4.14}$$

2. *For any $N, n_1, n_2 \in \mathbb{Z}_{\geq 0}$ and $\lambda, \mu \in \mathsf{Sign}^+_N$, one has*

$$\mathsf{G}^c_{\lambda/\mu}(u_1, \ldots, u_{n_1+n_2}) = \sum_{\kappa \in \mathsf{Sign}^+_N} \mathsf{G}^c_{\lambda/\kappa}(u_{n_1+1}, \ldots, u_{n_1+n_2}) \mathsf{G}^c_{\kappa/\mu}(u_1, \ldots, u_{n_1}). \tag{4.15}$$

The properties of the last proposition are known as *branching rules*.

Definition 4.2.5. We say that two complex numbers $u, v \in \mathbb{C}$ are *admissible* with respect to the parameter s if $\left|\frac{u-s}{1-su} \cdot \frac{v-s}{1-sv}\right| < 1$.

Proposition 4.2.6. *[46, Corollary 4.10] Let $u_1, \ldots, u_N$ and $v_1, \ldots, v_K$ be complex numbers such that u_i, v_j are admissible for all $i = 1, \ldots, N$ and $j = 1, \ldots, K$. Then for any $\lambda, \nu \in \mathsf{Sign}^+$*

$$\begin{aligned} &\sum_{\kappa \in \mathsf{Sign}^+} \mathsf{G}^c_{\kappa/\lambda}(v_1, \ldots, v_K) \mathsf{F}_{\kappa/\nu}(u_1, \ldots, u_N) = \\ &\prod_{i=1}^{N} \prod_{j=1}^{K} \frac{1 - q u_i v_j}{1 - u_i v_j} \sum_{\mu \in \mathsf{Sign}^+} \mathsf{F}_{\lambda/\mu}(u_1, \ldots, u_N) \mathsf{G}^c_{\nu/\mu}(v_1, \ldots, v_K). \end{aligned} \tag{4.16}$$

Remark 4.2.7. Equation (4.16) is called the *skew Cauchy identity* for the functions $\mathsf{F}_{\lambda/\mu}$ and $\mathsf{G}^c_{\lambda/\mu}$ because of its similarity with the skew Cauchy identities for Schur, Hall-Littlewood, or Macdonald symmetric functions [142]. The sum on the right-hand side (RHS) of (4.16) has finitely many non-zero terms and is thus well-defined. The left-hand side (LHS) can potentially have infinitely many non-zero terms, but part of the statement of the proposition is that if the variables are admissible, then this sum is absolutely convergent and numerically equals the right side.

A special case of (4.16), when $\lambda = \varnothing$ and $\nu = (0, 0, \ldots, 0)$ leads us to the *Cauchy identity*

$$\sum_{\nu \in \mathsf{Sign}^+_N} \mathsf{F}_\nu(u_1, \ldots, u_N) \mathsf{G}^c_\nu(v_1, \ldots, v_K) = (q; q)_N \prod_{i=1}^{N} \left(\frac{1}{1 - s u_i} \prod_{j=1}^{K} \frac{1 - q u_i v_j}{1 - u_i v_j} \right). \tag{4.17}$$

We end this section with the *symmetrization formulas* for G^c_ν and F_μ and also formulas for the functions when the variable set forms a geometric progression with parameter q.

Proposition 4.2.8. *[46, Theorem 4.14]* **1.** *For any $N \in \mathbb{Z}_{\geq 0}$, $\mu \in \mathsf{Sign}_N^+$ and $u_1, \ldots, u_N \in \mathbb{C}$*

$$\mathsf{F}_\mu(u_1, \ldots, u_N) = \frac{(1-q)^N}{\prod_{i=1}^N (1-su_i)} \sum_{\sigma \in S_N} \sigma \left(\prod_{1 \leq \alpha < \beta \leq N} \frac{u_\alpha - q u_\beta}{u_\alpha - u_\beta} \left(\frac{u_i - s}{1 - s u_i} \right)^{\mu_i} \right). \tag{4.18}$$

2. *Let $n \geq 0$ and $\mathsf{Sign}_n^+ \ni \nu = 0^{n_0} 1^{n_1} 2^{n_2} \cdots$. Then for any $N \geq n - n_0$ and $u_1, \ldots, u_N \in \mathbb{C}$ we have*

$$\begin{aligned} \mathsf{G}_\nu^c(u_1, \ldots, u_N) = \frac{(1-q)^N (q;q)_n}{\prod_{i=1}^N (1-su_i)(q;q)_{N-n+n_0}(q;q)_{n_0}} \prod_{k=1}^\infty \frac{(s^2;q)_{n_k}}{(q;q)_{n_k}} \times \\ \sum_{\sigma \in S_N} \sigma \left(\prod_{1 \leq \alpha < \beta \leq N} \frac{u_\alpha - q u_\beta}{u_\alpha - u_\beta} \prod_{i=1}^n \left(\frac{u_i - s}{1 - s u_i} \right)^{\nu_i} \prod_{i=1}^{n-n_0} \frac{u_i}{u_i - s} \prod_{j=n-n_0+1}^N (1 - s q^{n_0} u_j) \right). \end{aligned} \tag{4.19}$$

In both equations above, S_N denotes the permutation group on $\{1, \ldots, N\}$ and an element $\sigma \in S_N$ acts on the expression by permuting the variable set to $u_{\sigma(1)}, \ldots, u_{\sigma(N)}$. If $N < n - n_0$, then $\mathsf{G}_\nu^c(u_1, \ldots, u_N)$ is equal to 0.

Remark 4.2.9. We mention here that the formulas in Proposition 4.2.8 a priori make sense when $u_\alpha \neq u_\beta$ for $\alpha \neq \beta$ because the factors $u_\alpha - u_\beta$ appear in the denominator on the right sides of (4.18) and (4.19). However, the formulas can be extended continuously to all $(u_1, \ldots, u_N)$ such that $u_i \notin \{s, s^{-1}\}$ for all $i = 1, \ldots, N$. One observes this after putting all summands under the same denominator and realizing that the numerator is a skew-symmetric polynomial in $(u_1, \ldots, u_N)$, which is thus divisible by the Vandermonde determinant $\prod_{1 \leq \alpha < \beta \leq N} (u_\alpha - u_\beta)$.

Proposition 4.2.10. **1.** *For any $N \in \mathbb{Z}_{\geq 0}$, $\mu \in \mathsf{Sign}_N^+$ and $u \in \mathbb{C}$, one has*

$$\mathsf{F}_\mu(u, qu, \ldots, q^{N-1}u) = (q;q)_N \prod_{i=1}^N \left(\frac{1}{1 - s q^{i-1} u} \left(\frac{u q^{i-1} - s}{1 - s q^{i-1} u} \right)^{\mu_i} \right). \tag{4.20}$$

2. *Let $n \geq 0$ and $\mathsf{Sign}_n^+ \ni \nu = 0^{n_0} 1^{n_1} 2^{n_2} \cdots$. Then for any $N \geq n - n_0$ and $u \in \mathbb{C}$ we have*

$$\mathsf{G}_\nu^c(u, qu, \ldots, q^{N-1}u) = \prod_{k=1}^\infty \frac{(s^2;q)_{n_k}}{(q;q)_{n_k}} \frac{(q;q)_N (su;q)_{N+n_0} (q;q)_n \prod_{i=1}^N \left(\frac{1}{1-sq^{i-1}u} \left(\frac{q^{i-1}u - s}{1 - s q^{i-1} u} \right)^{\nu_j} \right)}{(q;q)_{N-n+n_0} (su;q)_n (q;q)_{n_0} (su^{-1};q^{-1})_{n-n_0}}, \tag{4.21}$$

where we agree that $v_j = 0$ *if* $j > n$.

4.2.2 The measure $\mathbb{P}_{\mathbf{u},\mathbf{v}}$

In this section we briefly explain how to construct the measure $\mathbb{P}_{\mathbf{u},\mathbf{v}}$ and summarize some of its basic properties. For a more detailed derivation of this measure we refer the reader to [71, Sections 2.2 and 2.3].

Let us briefly explain the main steps of the construction of $\mathbb{P}_{\mathbf{u},\mathbf{v}}$. We begin by considering the bigger space of all up-right paths in the half-infinite strip that share no horizontal piece but are allowed to share vertical pieces. For each such collection of paths we define its weight using the functions from Section 4.2.1. Afterwards we specialize $s = q^{-1/2}$ in those functions and perform a limit transition for some of the other parameters. This procedure has the effect of killing the weight of those path configurations that share a vertical piece. Consequently, we are left with weights that are non-zero only for six-vertex configurations, are absolutely summable and their sum has a product form. We check that each weight is non-negative, and define $\mathbb{P}_{\mathbf{u},\mathbf{v}}$ as the quotient of these weights with the partition function. We explain this in more detail below.

We fix positive integers N, M, J, and $K = M + J$, as well as real numbers $q \in (0,1)$ and $s > 1$. In addition, we suppose $\mathbf{u} = (u_1, \dots, u_N)$ and $\mathbf{w} = (w_1, \dots, w_K)$ are positive real numbers, such that $\max_{i,j} u_i w_j < 1$ and $u := \min_i u_i > s$. One readily verifies that the latter conditions ensure that u_i, w_j are admissible with respect to s for $i = 1, \dots, N$ and $j = 1, \dots, K$.

Let us go back to the setup of Section 4.1.1. We let $\mathcal{P}'_N$ be the collection of N up-right paths drawn in the sector $D_N = \mathbb{Z}_{\geq 0} \times \{1, \dots, N\}$ of the square lattice, with all paths starting from a left-to-right edge entering each of the points $\{(0,m) : 1 \leq m \leq N\}$ on the left boundary and all paths exiting from the top boundary of D_N. We still assume that no two paths share a horizontal piece, but sharing vertical pieces is allowed. As in Section 4.1.1 we let $\mathcal{P}_N \subset \mathcal{P}'_N$ be those collections of paths that do not share vertical pieces. For $\pi \in \mathcal{P}'_N$ and $k = 1, \dots, N$ we let $\lambda^k(\pi) \in \mathsf{Sign}^+_k$ denote the ordered x-coordinates of the intersection of π with the horizontal line $y = k + 1/2$. We denote

by $\pi(i,j)$ the type of the vertex at $(i,j) \in D_N$. We also let $f : \mathsf{Sign}_N^+ \to \mathbb{R}$ be given by

$$f(\lambda;\mathbf{w}) := \mathsf{G}^c_\mu(w_1,\dots,w_J,w_{J+1},\dots,w_K) = \sum_{\mu\in\mathsf{Sign}_N^+} \mathsf{G}^c_\mu(w_1,\dots,w_J)\mathsf{G}^c_{\lambda/\mu}(w_{J+1},\dots,w_K),$$

where the equality above follows from Proposition 4.2.4. With the above data, we define the weight of a collection of paths π by

$$\mathcal{W}^f_{\mathbf{u},\mathbf{w}}(\pi) = \prod_{i=1}^{N}\prod_{j=1}^{\infty} w_{u_i}(\pi(i,j)) \times f(\lambda^N(\pi);\mathbf{w}).$$

Using the branching relations, Proposition 4.2.4, and the Cauchy identity (4.17) we get

$$\sum_{\pi\in\mathcal{P}'_N} \mathcal{W}^f_{\mathbf{u},\mathbf{w}}(\pi) = (q;q)_N \prod_{i=1}^{N}\left(\frac{1}{1-su_i}\prod_{j=1}^{K}\frac{1-qu_iw_j}{1-u_iw_j}\right) =: Z^f(\mathbf{u};\mathbf{w}). \tag{4.22}$$

In view of the admissability conditions satisfied by $\mathbf{u}$ and $\mathbf{w}$, the above sum is in fact absolutely convergent.

We next fix $s = q^{-1/2}$, set $w_i = q^{i-1}\varepsilon$ for $i = 1,\dots,J$ and put $v_j = w_{j+J}$ for $j = 1,\dots,M$. Here $\varepsilon > 0$ is chosen sufficiently small so that the admissability condition is maintained. One shows that with the above specialization of parameters f becomes a function of $\lambda, \mathbf{v}, \varepsilon$ and q^J and we denote it by $f_\varepsilon(\lambda;\mathbf{v},q^J)$. Specifically, if $q^J = X$ one has

$$\begin{aligned} f_\varepsilon(\lambda;\mathbf{v},X) = \sum_{\nu\in\mathsf{Sign}_N^+} &\frac{(q;q)_N(-q)^{n_0-N}(s\varepsilon;q)_{N-n_0}}{(q;q)_{n_0}(s\varepsilon;q)_N(s\varepsilon^{-1};q^{-1})_{N-n_0}}(Xq^{-N+n_0+1};q)_{N-n_0}(s\varepsilon X;q)_{n_0}\times \\ &\prod_{i=1}^{\infty}\mathbf{1}_{\{n_i\le 1\}}\prod_{j=1}^{N-n_0}\left(\frac{1}{1-s\varepsilon q^{j-1}}\left(\frac{\varepsilon q^{j-1}-s}{1-s\varepsilon q^{j-1}}\right)^{\nu_j}\right)\mathsf{G}^c_{\lambda/\nu}(v_1,\dots,v_M), \end{aligned} \tag{4.23}$$

where $\nu = 0^{n_0}1^{n_1}2^{n_2}\cdots$ and $\mathbf{1}_E$ is the indicator function of a set E. In addition, specializing our w

variables in $Z^f(\mathbf{u};\mathbf{w})$ and replacing q^J with X, we get that $Z^f(\mathbf{u};\mathbf{w})$ becomes

$$(q;q)_N \prod_{i=1}^{N}\left(\frac{1}{1-su_i}\frac{1-X\varepsilon u_i}{1-\varepsilon u_i}\prod_{j=1}^{M}\frac{1-qu_iv_j}{1-u_iv_j}\right) =: Z^{f_\varepsilon}(\mathbf{u};\mathbf{v},X).$$

We subsitute $X=(s\varepsilon)^{-1}$ and take the limit as $\varepsilon\to 0$. Under this limit transition we have

$$\begin{aligned} &f(\lambda;\mathbf{v},\rho) := \lim_{\varepsilon\to 0} f_\varepsilon(\lambda;\mathbf{v},(s\varepsilon)^{-1}) = \\ &(-1)^N(q;q)_N \sum_{\nu\in \mathsf{Sign}^+_N} \mathbf{1}_{\{n_0=0\}}\prod_{i=1}^{\infty}\mathbf{1}_{\{n_i\le 1\}}\prod_{j=1}^{N}(-s)^{\nu_j}\,\mathsf{G}^c_{\lambda/\nu}(v_1,\dots,v_M), \end{aligned} \tag{4.24}$$

and

$$Z^f(\mathbf{u}) := \lim_{\varepsilon\to 0} Z^{f_\varepsilon}(\mathbf{u};\mathbf{v},(s\varepsilon)^{-1}) = (q;q)_N \prod_{i=1}^{N}\left(\frac{1-s^{-1}u_i}{1-su_i}\prod_{j=1}^{M}\frac{1-qu_iv_j}{1-u_iv_j}\right).$$

The above formulas imply that $f(\lambda;\mathbf{v},\rho)=0$ if $\lambda_N=0$ or $\lambda_i=\lambda_j$ for $i\neq j$.

Equation (4.22) can now be analytically extended in X (both sides become polynomials in X), and after specializing $X=(s\varepsilon)^{-1}$ and taking the limit as $\varepsilon\to 0^+$ we get

$$\sum_{\pi\in\mathcal{P}'_N}\prod_{i=1}^{N}\prod_{j=1}^{\infty} w_{u_i}(\pi(i,j))\times f(\lambda^N(\pi);\mathbf{v},\rho) = Z^f(\mathbf{u}), \tag{4.25}$$

where again the right side can be shown to be absolutely convergent. With $f(\lambda;\mathbf{v},\rho)$ as above we define the following weight of a collection of paths in $\mathcal{P}'_N$

$$\mathcal{W}^f_{\mathbf{u},\mathbf{v}}(\pi) = \prod_{i=1}^{N}\prod_{j=1}^{\infty} w_{u_i}(\pi(i,j))\times f(\lambda^N(\pi);\mathbf{v},\rho). \tag{4.26}$$

One can check that $\mathcal{W}^f_{\mathbf{u},\mathbf{v}}(\pi)\ge 0$, $\mathcal{W}^f_{\mathbf{u},\mathbf{v}}(\pi)=0$ for all $\pi\in\mathcal{P}'_N\setminus\mathcal{P}_N$. As weights are non-negative and the partition function $Z^f(\mathbf{u})$ is positive and finite, we see from (4.25) that

$$\mathbb{P}_{\mathbf{u},\mathbf{v}}(\pi) := \frac{\mathcal{W}^f_{\mathbf{u},\mathbf{v}}(\pi)}{Z^f(\mathbf{u})},$$

defines an honest probability measure on $\mathcal{P}_N$. For future reference we summarize the parameter choices we have made in the following definition.

Definition 4.2.11. Let $N, M \in \mathbb{N}$. We fix $q \in (0,1)$ and $s = q^{-1/2}$, $\mathbf{u} = (u_1, \dots, u_N)$ with $u_i > s$ and $\mathbf{v} = (v_1, \dots, v_M)$ with $v_j > 0$, and $\max_{i,j} u_i v_j < 1$. With these parameters, we denote $\mathbb{P}_{\mathbf{u},\mathbf{v}}$ to be the probability measure on $\mathcal{P}_N$, defined above.

We end this section with the following result that provides a formula for the finite-dimensional projections of $\mathbb{P}_{\mathbf{u},\mathbf{v}}$.

Lemma 4.2.12. *Let $N, M \in \mathbb{N}$. Fix $q \in (0,1)$ and $s = q^{-1/2}$, $\mathbf{u} = (u_1, \dots, u_N)$ with $u_i > s$ and $\mathbf{v} = (v_1, \dots, v_M)$ with $v_j > 0$, and $\max_{i,j} u_i v_j < 1$. With these parameters let $\mathbb{P}_{\mathbf{u},\mathbf{v}}$ be as in Definition 4.2.11. Let us fix $k \in \mathbb{N}$, $1 \le m_1 < m_2 < \cdots < m_k \le N$ and $\mu^{m_i} \in \mathsf{Sign}^+_{m_i}$. Then*

$$\begin{aligned} &\mathbb{P}_{\mathbf{u},\mathbf{v}}\left(\lambda^{m_1}(\pi) = \mu^{m_1}, \dots, \lambda^{m_k}(\pi) = \mu^{m_k}\right) = \frac{\prod_{r=0}^{k-1} \mathsf{F}_{\mu^{m_{r+1}}/\mu^{m_r}}(u_{m_r+1}, \dots, u_{m_{r+1}}) f(\mu^{m_k}; \mathbf{v}, \rho)}{Z^f(\mathbf{u}, \mathbf{v}; m_k)}, \\ &\text{where } Z^f(\mathbf{u}, \mathbf{v}; m_k) = (q;q)_{m_k} \prod_{i=1}^{m_k} \left(\frac{1 - s^{-1}u_i}{1 - s u_i} \prod_{j=1}^{M} \frac{1 - q u_i v_j}{1 - u_i v_j} \right). \end{aligned} \tag{4.27}$$

Remark 4.2.13. If $k \le N$ and $m_i = i$ for $i = 1, \dots, k$ then (4.27) implies that the projection of $\mathbb{P}_{\mathbf{u},\mathbf{v}}$ to D_k has law $\mathbb{P}_{\mathbf{u}_k,\mathbf{v}}$, where $\mathbf{u}_k = (u_1, \dots, u_k)$. In particular, the measures $\mathbb{P}_{\mathbf{u},\mathbf{v}}$ are consistent with each other and can be used to define a measure on up-right paths on the entire region $\mathbb{Z}^2_{\ge 0}$.

Proof. Equation (4.27) can be found as [71, Equation (29)] and we refer the interested reader to Section 2.3 in that paper for the proof. □

4.3 Estimates for $f(\lambda; \mathbf{v}, \rho)$ and F_λ

In this section we give a contour integral formula for the functions $f(\lambda; \mathbf{v}, \rho)$, and a combinatorial estimate for the function F_λ from Definition 4.2.2. The results we derive in this section will be used in Sections 4.4 and 4.5 to prove Theorem 4.1.3.

4.3.1 Integral formulas for $f(\lambda; \mathbf{v}, \rho)$

The purpose of this section is to derive a contour integral formula for the function $f(\lambda; \mathbf{v}, \rho)$ from Section 4.2.2. We accomplish this in Lemma 4.3.3 after we derive a contour integral formula for the functions G^c_λ from Definition 4.2.1 in Lemma 4.3.1. In the remainder of the chapter we denote by ι the root $\sqrt{-1}$ that lies in the complex upper half-plane.

Lemma 4.3.1. *Suppose that $k, N \in \mathbb{N}$ satisfy $N \geq k$, $q \in (0,1)$, $s > 1$ and $v_1, \ldots, v_N$ are complex numbers such that $|v_i| < s^{-1}$ for all $i = 1, \ldots, N$. Then for any $\lambda \in \mathsf{Sign}^+_k$ with $\lambda_k \geq 1$ we have*

$$\mathsf{G}^c_\lambda(v_1, \ldots, v_N) = c(\lambda)(q;q)_k \cdot \oint_\gamma \cdots \oint_\gamma \prod_{1 \leq \alpha < \beta \leq k} \frac{u_\alpha - u_\beta}{u_\alpha - q u_\beta} \times \prod_{i=1}^k \frac{1}{(1-su_i)(u_i - s)} \left(\frac{1 - su_i}{u_i - s} \right)^{\lambda_i} \prod_{i=1}^k \prod_{j=1}^N \frac{1 - q u_i v_j}{1 - u_i v_j} \prod_{i=1}^k \frac{du_i}{2\pi\iota}. \tag{4.28}$$

In (4.28) the constant $c(\lambda)$ is as in (4.13) and the contour γ is a zero-centered positively oriented circle of radius $R \in \left(s, \min_{1,\ldots,N} |v_i|^{-1}\right)$, where the latter set is non-empty by our assumption on v_i's.

Remark 4.3.2. We mention here that a similar result to the above lemma was proved in [46, Corollary 7.16] with several important differences. First of all, the formula in [46, Corollary 7.16] is for the functions G_λ rather than G^c_λ, but in view of (4.13) this difference is inessential. A more significant difference is that the authors in that paper assumed that $s \in (-1, 0)$ unlike our case of $s > 1$ – this difference is also minor and can be overcome by an analytic continuation argument in the parameter s. A crucial difference is that the contour integral formula in [46, Corollary 7.16] is based on small contours that encircle s while the contours in Lemma 4.3.1 above are large contours. In particular, the formulas in Lemma 4.3.1 are different and cannot be obtained by a direct application of Cauchy's theorem from the ones in [46, Corollary 7.16]. That being said, we mention that the existence of both small and large contour formulas is known in the related context of Macdonald processes, see [36, Section 3.2.3] and the derivation of both types of formulas is similar in spirit.

Proof. The proof is a standard computation of residues for the integrals on the right side of (4.28),

but for clarity we split the proof into two steps.

Step 1. We claim that (4.28) holds when $v_1, \dots, v_N \in (0, s^{-1})$ are such that $v_i \neq v_j$ for $i \neq j$. We prove this statement in the second step. In the present step we assume its validity and conclude the proof of the lemma.

Let Ω denote the open disc of radius s^{-1}, centered at the origin in $\mathbb{C}$. Observe that by Definition 4.2.1 and (4.9) we know that $\mathsf{G}^c_\lambda(v_1, \dots, v_N)$ is a finite sum of rational functions in $v_1, \dots, v_N$ that are analytic in Ω^N (here we used that the possible poles of G^c_λ come from the zeros of the denominators of $w_v(i_1, j_1; i_2, j_2)$ which are all located at $v = s^{-1}$). This means that for each $i = 1, \dots, N$ and $v_j \in \Omega$ for $j \neq i$ the left side of (4.28) as a function of v_i is analytic in Ω. Since γ has radius bigger than s by assumption, we see that for each $i = 1, \dots, N$ and $v_j \in \Omega$ for $j \neq i$ the integrand on the right side of (4.28) is also analytic on Ω as a function of v_i. It follows by [176, Theorem 5.4] that for each $i = 1, \dots, N$ and $v_j \in \Omega$ for $j \neq i$ the right side of (4.28) is analytic on Ω as a function of v_i.

We claim for each $k = 0, \dots, N$ that (4.28) holds if $v_1, \dots, v_{N-k} \in (0, s^{-1})$ are such that $v_i \neq v_j$ for $1 \leq i \neq j \leq N - k$ and $v_{N-k+1}, \dots v_N \in \Omega$. We prove this statement by induction on k with base case $k = 0$ being true by our claim in the beginning of the step. Let us assume this result for k and prove it for $k + 1$. We fix $v_{N-k+1}, \dots, v_N \in \Omega$ and $v_1, \dots, v_{N-k-1}$ in $(0, s^{-1})$ with $v_i \neq v_j$ for $1 \leq i \leq j \leq N - k - 1$. Put $m = \max(v_1, \dots, v_{N-k-1})$ and observe that by our discussion in the previous paragraph both sides of (4.28) are analytic functions of v_{N-k} in Ω and by induction hypothesis these two functions are equal when $v_{N-k} \in (m, s^{-1})$. Since the latter set is contained in Ω and has a limit point in that set we conclude by [176, Corollary 4.9] that both sides of (4.28) agree for all $v_{N-k} \in \Omega$. This proves the desired result for $k + 1$ and we conclude by induction that the result holds when $k = N$, which is precisely the statement of the lemma.

Step 2. In this step we prove that (4.28) holds when $v_1, \dots, v_N \in (0, s^{-1})$ are such that $v_i \neq v_j$

for $i \neq j$. We proceed to sequentially compute the integral with respect to u_i for $i = 1, 2, \ldots, k$ in this order as a sum of residues. Observe that after we have evaluated the (minus) residues outside of γ for u_j with $j = 1, \ldots, i-1$ the integrand only has simple poles when $u_i = v_{m_i}^{-1}$ (there are no poles at infinity since the integrand is $\sim u_i^2$ as $|u_i| \to \infty$, and also no new poles are introduced after evaluating the residues at $u_j = v_{m_j}^{-1}$ for $j = 1, \ldots, i-1$). Furthermore, the Vandermonde determinant $\prod_{1 \leq \alpha < \beta \leq k}(u_\alpha - u_\beta)$ in the integrand implies that we only get a non-trivial contribution from the residues when $m_1, \ldots, m_k$ are all distinct. Putting this all together, we conclude by the Residue theorem that the right side of (4.28) is equal to

$$c(\lambda)(q;q)_k \sum_I \prod_{1 \leq \alpha < \beta \leq k} \frac{v_{m_\alpha}^{-1} - v_{m_\beta}^{-1}}{v_{m_\alpha}^{-1} - q v_{m_\beta}^{-1}} \prod_{i=1}^{k} \frac{1}{(1 - s v_{m_i}^{-1})(v_{m_i}^{-1} - s)} \left(\frac{1 - s v_{m_i}^{-1}}{v_{m_i}^{-1} - s} \right)^{\lambda_i} \times \frac{\prod_{i=1}^{k} \prod_{j=1}^{N} (v_j^{-1} - q v_{m_i}^{-1})}{\prod_{i=1}^{k} \prod_{j=1, j \neq m_i}^{N} (v_j^{-1} - v_{m_i}^{-1})},$$

where the sum is over injective functions $I : \{1, \ldots, k\} \to \{1, \ldots, N\}$ and we have denoted $I(r) = m_r$. Performing some simplifications and rearrangements we conclude that

$$\begin{aligned} &c(\lambda)(q;q)_k (1-q)^k \sum_I \prod_{1 \leq \alpha < \beta \leq k} \frac{v_{m_\alpha} - q v_{m_\beta}}{v_{m_\alpha} - v_{m_\beta}} \prod_{i=1}^{k} \frac{v_{m_i}}{(v_{m_i} - s)(1 - s v_{m_i})} \left(\frac{v_{m_i} - s}{1 - s v_{m_i}} \right)^{\lambda_i} \times \\ &\prod_{i=1}^{k} \prod_{j \in J} \frac{(v_{m_i} - q v_j)}{(v_{m_i} - v_j)} = \text{RHS of (4.28)}, \end{aligned} \tag{4.29}$$

where $J = \{1, \ldots, N\} \setminus \{m_1, \ldots, m_k\}$.

On the other hand, by (4.19), we have that the left side of (4.28) is equal to

$$\frac{c(\lambda)(1-q)^N (q;q)_k}{(q;q)_{N-k}} \sum_{\sigma \in S_N} \sigma \left(\prod_{1 \leq \alpha < \beta \leq N} \frac{v_\alpha - q v_\beta}{v_\alpha - v_\beta} \prod_{i=1}^{k} \frac{v_i}{(v_i - s)(1 - s v_i)} \left(\frac{v_i - s}{1 - s v_i} \right)^{\lambda_i} \right),$$

where $\lambda = 0^{n_0} 1^{n_1} 2^{n_2} \cdots$. We next split the latter sum over the possible values of $\sigma(1), \ldots, \sigma(k)$

and rewrite the above as

$$\frac{c(\lambda)(1-q)^N(q;q)_k}{(q;q)_{N-k}}\sum_I \prod_{1\le\alpha<\beta\le k}\frac{v_{m_\alpha}-qv_{m_\beta}}{v_{m_\alpha}-v_{m_\beta}}\cdot\prod_{i=1}^k\prod_{j\in J}\frac{(v_{m_i}-qv_j)}{(v_{m_i}-v_j)}$$
$$\prod_{i=1}^k\frac{v_{m_i}}{(v_{m_i}-s)(1-sv_{m_i})}\left(\frac{v_{m_i}-s}{1-sv_{m_i}}\right)^{\lambda_i}\cdot\sum_{\tau\in S_{N-k}}\tau\left(\prod_{1\le\alpha<\beta\le N-k}\frac{v_{j_{\tau(\alpha)}}-qv_{j_{\tau(\beta)}}}{v_{j_{\tau(\alpha)}}-v_{j_{\tau(\beta)}}}\right), \tag{4.30}$$

where as before the sum is over injective maps $I:\{1,\dots,k\}\to\{1,\dots,N\}$, $m_r=I(r)$ for $r=1,\dots,k$ and $J=\{1,\dots,N\}\setminus\{m_1,\dots,m_k\}$. The inner sum is over pemutations τ of $\{1,\dots,N-k\}$ and $j_1,\dots,j_{N-k}$ denote the elements of J in increasing order (the particular order does not matter). We know from [142, Chapter III, (1.4)] that

$$\sum_{\tau\in S_{N-k}}\tau\left(\prod_{1\le\alpha<\beta\le N-k}\frac{v_{j_{\tau(\alpha)}}-qv_{j_{\tau(\beta)}}}{v_{j_{\tau(\alpha)}}-v_{j_{\tau(\beta)}}}\right)=\frac{(q;q)_{N-k}}{(1-q)^{N-k}}.$$

Substituting this in (4.30) we conclude that

$$c(\lambda)(q;q)_k(1-q)^k\sum_I \prod_{1\le\alpha<\beta\le k}\frac{v_{m_\alpha}-qv_{m_\beta}}{v_{m_\alpha}-v_{m_\beta}}\cdot\prod_{i=1}^k\prod_{j\in J}\frac{(v_{m_i}-qv_j)}{(v_{m_i}-v_j)}$$
$$\prod_{i=1}^k\frac{v_{m_i}}{(v_{m_i}-s)(1-sv_{m_i})}\left(\frac{v_{m_i}-s}{1-sv_{m_i}}\right)^{\lambda_i}=\text{ LHS of (4.28).}$$

Comparing the last equation with (4.29) we conclude that the left and right sides of (4.28) agree when $v_1,\dots,v_N\in(0,s^{-1})$ are such that $v_i\neq v_j$ for $i\neq j$. This suffices for the proof. □

The next lemma provides a contour integral formula for $f(\lambda;\mathbf{v},\rho)$ from (4.24).

Lemma 4.3.3. *Suppose that $k,M\in\mathbb{N}$, $q\in(0,1)$, $s=q^{-1/2}$ and $v_1,\dots,v_M\in(0,s^{-1})$. Then for any $\lambda\in\mathsf{Sign}_k^+$ with $\lambda_k\ge 1$ we have*

$$f(\lambda;\mathbf{v},\rho)=c(\lambda)(q;q)_k\cdot\oint_\gamma\cdots\oint_\gamma\prod_{1\le\alpha<\beta\le k}\frac{u_\alpha-u_\beta}{u_\alpha-qu_\beta}\times$$
$$\prod_{i=1}^k\frac{1}{-s(1-su_i)}\left(\frac{1-su_i}{u_i-s}\right)^{\lambda_i}\prod_{i=1}^k\prod_{j=1}^M\frac{1-qu_iv_j}{1-u_iv_j}\prod_{i=1}^k\frac{du_i}{2\pi\iota}. \tag{4.31}$$

In (4.31) the constant $c(\lambda)$ is as in (4.13) and the contour γ is a zero-centered positively oriented circle of radius $R \in \left(s, \min_{1,\dots,M} |v_i|^{-1}\right)$, where the latter set is non-empty by our assumption on v_i's.

Proof. We start from (4.28) with $N = M + J$, where $J \in \mathbb{N}$ and variables $w_1, \dots, w_N$ in place of $v_1, \dots, v_N$. We then set $w_i = q^{i-1}\varepsilon$ for $i = 1, \dots, J$ and $w_{J+i} = v_i$ for $i = 1, \dots, M$. Here $\varepsilon \in (0, s^{-1})$. This gives

$$\mathsf{G}^c_\lambda(\varepsilon, q\varepsilon, \cdots, q^{J-1}\varepsilon, v_1, \dots, v_M) = c(\lambda)(q;q)_k \cdot \oint_\gamma \cdots \oint_\gamma \prod_{1 \le \alpha < \beta \le k} \frac{u_\alpha - u_\beta}{u_\alpha - q u_\beta} \times$$
$$\prod_{i=1}^k \frac{1}{(1-su_i)(u_i - s)} \left(\frac{1-su_i}{u_i - s}\right)^{\lambda_i} \prod_{i=1}^k \prod_{j=1}^M \frac{1 - q u_i v_j}{1 - u_i v_j} \cdot \prod_{i=1}^k \frac{1 - q^J \varepsilon u_i}{1 - u_i \varepsilon} \prod_{i=1}^k \frac{du_i}{2\pi\iota}.$$

In particular, we see that if $f_\varepsilon(\lambda; \mathbf{v}, X)$ is as in (4.23) we have

$$f_\varepsilon(\lambda; \mathbf{v}, X) = c(\lambda)(q;q)_k \cdot \oint_\gamma \cdots \oint_\gamma \prod_{1 \le \alpha < \beta \le k} \frac{u_\alpha - u_\beta}{u_\alpha - q u_\beta} \times$$
$$\prod_{i=1}^k \frac{1}{(1-su_i)(u_i - s)} \left(\frac{1-su_i}{u_i - s}\right)^{\lambda_i} \prod_{i=1}^k \prod_{j=1}^M \frac{1 - q u_i v_j}{1 - u_i v_j} \cdot \prod_{i=1}^k \frac{1 - X \varepsilon u_i}{1 - u_i \varepsilon} \prod_{i=1}^k \frac{du_i}{2\pi\iota}, \tag{4.32}$$

whenever $X = q^J$ for any $J \ge 1$. In view of (4.23) we see that both sides of (4.32) are degree k polynomials in X and since they agree for infinitely many points $X = q^J, J \ge 1$ they must agree for all X. Then if we set $X = (s\varepsilon)^{-1}$ and let $\varepsilon \to 0$ we conclude in view of (4.24) that

$$f(\lambda; \mathbf{v}, \rho) = \lim_{\varepsilon \to 0} c(\lambda)(q;q)_k \cdot \oint_\gamma \cdots \oint_\gamma \prod_{1 \le \alpha < \beta \le k} \frac{u_\alpha - u_\beta}{u_\alpha - q u_\beta} \times$$
$$\prod_{i=1}^k \frac{1}{(1-su_i)(u_i - s)} \left(\frac{1-su_i}{u_i - s}\right)^{\lambda_i} \prod_{i=1}^k \prod_{j=1}^M \frac{1 - q u_i v_j}{1 - u_i v_j} \cdot \prod_{i=1}^k \frac{1 - s^{-1} u_i}{1 - u_i \varepsilon} \prod_{i=1}^k \frac{du_i}{2\pi\iota},$$

which clearly implies (4.31) by the bounded convergence theorem. □

4.3.2 Combinatorial estimates for F_λ

We continue to use the notation from Section 4.2. In this section we estimate the function F_λ when $q \in (0,1)$, $s = q^{-1/2}$, $\lambda \in \mathsf{Sign}_k^+$ and $u_1, \ldots, u_k$ are all equal to the same parameter $u > s$. We denote this function by $\mathsf{F}_\lambda([u]^k)$. For $\lambda \in \mathsf{Sign}_k^+$ we denote $|\lambda| = \lambda_1 + \cdots + \lambda_k$. The purpose of this section is to establish the following result.

Lemma 4.3.4. *Fix $k \in \mathbb{N}$, $q \in (0,1)$, $s = q^{-1/2}$ and $u > s$. Then there exists a constant $C > 0$ depending on k, q, u such that for all $\lambda \in \mathsf{Sign}_k^+$ with $\lambda_1 > \lambda_2 > \cdots > \lambda_k$ we have*

$$\prod_{1 \le i < j \le k} \frac{\lambda_i - \lambda_j + j - i}{j - i} - C \cdot (\lambda_1 - \lambda_k + k)^{\binom{k}{2}-1} \le \mathsf{F}_\lambda([u]^k) \left(\frac{1-q}{1-su}\right)^{-\binom{k+1}{2}} \times$$
$$\left(\frac{(1-q^{-1})u}{1-su}\right)^{-\binom{k}{2}} \left(\frac{u-s}{1-su}\right)^{-|\lambda|+\binom{k}{2}} \le \prod_{1 \le i < j \le k} \frac{\lambda_i - \lambda_j + j - i}{j - i} + C \cdot (\lambda_1 - \lambda_k + k)^{\binom{k}{2}-1}. \tag{4.33}$$

We give the proof of Lemma 4.3.4 in the end of the section. The general idea of the proof is as follows. From Definition 4.2.2 the function $\mathsf{F}_\lambda([u]^k)$ is equal to a sum of weights $\mathcal{W}(\pi)$. For the majority of path collections π, which we call typical – see Definition 4.3.6 below, we have that the weight $\mathcal{W}(\pi)$ is equal to

$$\mathcal{W}_{typ} = \left(\frac{1-q}{1-su}\right)^{\binom{k+1}{2}} \left(\frac{(1-q^{-1})u}{1-su}\right)^{\binom{k}{2}} \cdot \left(\frac{u-s}{1-su}\right)^{|\lambda|-\binom{k}{2}}.$$

We prove this in Lemma 4.3.7 below. We show that the weights $\mathcal{W}(\pi)$ for all path collections π are within a constant multiple of the above weight – we do this in Lemma 4.3.5 below. Combining these two statements one deduces that $\mathsf{F}_\lambda([u]^k) \approx \mathcal{W}_{typ} \times K$ where K is the number of typical path collections. By a counting argument one can show that $K \approx \prod_{1 \le i < j \le k} \frac{\lambda_i - \lambda_j + j - i}{j-i}$. Combining these three statements one obtains Lemma 4.3.4. We now turn to filling in the details of the above outline.

Lemma 4.3.5. *Fix $k \in \mathbb{N}$, $q \in (0,1)$, $s = q^{-1/2}$ and $u > s$. Let $\lambda \in \mathsf{Sign}_k^+$ and $\pi \in \mathcal{P}_{\lambda/\varnothing}$. Then*

there is a constant $\tilde{C}$ that depends on k, q, u such that

$$|\mathcal{W}(\pi)| \leq \tilde{C} \left(\frac{u-s}{su-1} \right)^{|\lambda|}, \tag{4.34}$$

where $|\lambda| = \lambda_1 + \cdots + \lambda_k$ and $\mathcal{W}(\pi)$ is as in (4.10) for $u_1 = \cdots = u_k = u$.

Proof. From the definition of $\mathcal{P}_{\lambda/\varnothing}$ we know that a path collection π has $|\lambda|$ horizontal edges and $\binom{k+1}{2}$ vertical edges in $\mathbb{Z}^2_{\geq 0}$. Each edge borders two vertices except the top k vertical edges whose upper vertex is not counted. Due to this there are at most $2\binom{k+1}{2} - k = k^2$ vertices adjacent to a vertical edge. If we associate to each horizontal edge its left vertex and to each vertical edge its bottom vertex we obtain a surjective map from the set of edges to the set of vertices in π, whose path configuration is not $(0,0;0,0)$. Consequently, there are at most $|\lambda| + \binom{k+1}{2}$ nontrivial (i.e. not type $(0,0;0,0)$) vertices in π. Also the above mapping from horizontal edges to their left vertices contains all vertices of type $(0,1;0,1)$ in its range and the pre-image of each such vertex contains exactly one element. This implies that the number of vertices of type $(0,1;0,1)$ is at least $|\lambda| - k^2$.

We now recall from (4.8) that

$$w_u(0,0;0,0) = 1 \text{ and } w_u(0,1;0,1) = \frac{u-s}{1-su}.$$

Let $C \geq 1$ be a constant such that

$$|w_u(i_1, j_1; i_2; j_2)| \leq C$$

for all $i_1, i_2 \in \{0, \ldots, k\}$ and $j_1, j_2 \in \{0, 1\}$. The existence of C is ensured by (4.8) and it depends on k, u, q (here $s = q^{-1/2}$). Then our work from the previous paragraph and (4.10) suggest that

$$|\mathcal{W}(\pi)| \leq C^{\binom{k+1}{2}+k^2} \cdot \left(\frac{u-s}{su-1} \right)^{|\lambda|-k^2},$$

which clearly implies (4.34). □

Definition 4.3.6. Let $k \in \mathbb{N}$ and $\lambda \in \mathsf{Sign}^+_k$ be such that $\lambda_1 > \lambda_2 > \cdots > \lambda_k$. We say that a

path collection $\pi \in \mathcal{P}_{\lambda/\varnothing}$ is a *typical* path collection of $\mathcal{P}_{\lambda/\varnothing}$ if it only contains vertices of type $(0,0;0,0),(0,1;0,1),(0,1;1,0)$ and $(1,0;0,1)$. We denote the set of all typical path collections by $\mathcal{P}^{typ}_{\lambda/\varnothing}$. See Figure 4.7.

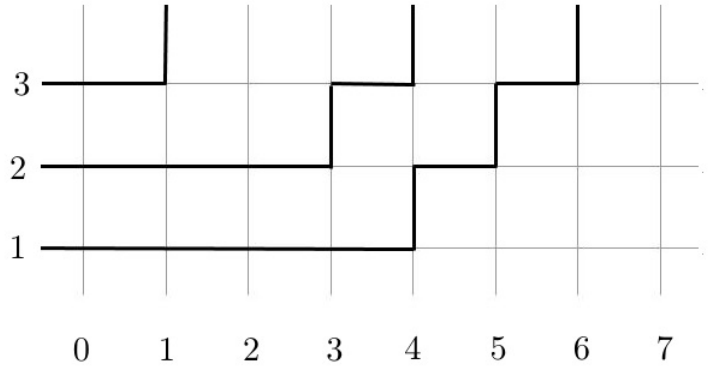

Figure 4.7: Example of a path collection belonging to $\mathcal{P}^{typ}_{\lambda/\varnothing}$ where $\lambda = (6,3,1)$.

Lemma 4.3.7. *Fix $k \in \mathbb{N}$, $q \in (0,1)$, $s = q^{-1/2}$ and $u > s$. Let $\lambda \in \mathsf{Sign}^+_k$ be such that $\lambda_1 > \lambda_2 > \cdots > \lambda_k$. If π is a typical path collection of $\mathcal{P}_{\lambda/\varnothing}$, then*

$$\mathcal{W}(\pi) = \left(\frac{1-q}{1-su}\right)^{\binom{k+1}{2}} \left(\frac{(1-q^{-1})u}{1-su}\right)^{\binom{k}{2}} \cdot \left(\frac{u-s}{1-su}\right)^{|\lambda|-\binom{k}{2}}, \tag{4.35}$$

where $|\lambda| = \lambda_1 + \cdots + \lambda_k$ and $\mathcal{W}(\pi)$ is as in (4.10) for $u_1 = \cdots = u_k = u$.

Proof. Since π is typical we know that it only contains vertices of type $(0,0;0,0)$, $(0,1;0,1)$, $(0,1;1,0)$ and $(1,0;0,1)$. Furthermore, we have from (4.8) that

$$w_u(0,0;0,0) = 1,\ \ w_u(0,1;0,1) = \frac{u-s}{1-su},\ \ w_u(0,1;1,0) = \frac{1-q}{1-su},\ \ w_u(1,0;0,1) = \frac{(1-q^{-1})u}{1-su},$$

where we used that $s^2 = q^{-1}$. If A, B, C denote the number of vertices in π with path configuration $(0,1;0,1)$, $(0,1;1,0)$ and $(1,0;0,1)$ respectively then by (4.10) we know that

$$\mathcal{W}(\pi) = \left(\frac{1-q}{1-su}\right)^{B} \left(\frac{(1-q^{-1})u}{1-su}\right)^{C} \cdot \left(\frac{u-s}{1-su}\right)^{A}.$$

Consequently, it suffices to show that if π is typical then $A = |\lambda| - \binom{k}{2}$, $B = \binom{k+1}{2}$, $C = \binom{k}{2}$.

We now proceed to simply count the the number of vertices of each type in a typical path collection. Notice that between row i and row $i+1$ there are precisey i vertical edges. The bottom vertex of each such edge has type $(0,1;1,0)$ and the top vertex of each such edge has type $(1,0;0,1)$. All other vertices in π have type $(0,0;0,0)$ or $(0,1;0,1)$. We conclude from this that $C = 1 + 2 + \cdots + (k-1) = \binom{k}{2}$ and $B = 1 + 2 + \cdots + k = \binom{k+1}{2}$ (notice that the top vertex of the edges connecting row k and $k+1$ are not included in the product defining $\mathcal{W}(\pi)$, while the bottom ones are). What we are left with is computing A.

From the definition of $\mathcal{P}_{\lambda/\varnothing}$ we know that a path collection π has $|\lambda|$ horizontal edges in $\mathbb{Z}^2_{\geq 0}$. The map that sends a horizontal edge to its left vertex endpoint maps the set of horizontal edges bijectively to the vertices of type $(0,1;0,1)$ and $(0,1;1,0)$ in π and so $A + C = |\lambda|$. We conclude that $A = |\lambda| - \binom{k}{2}$ as desired. □

Proof. (Lemma 4.3.4) Combining Lemmas 4.3.5 and 4.3.7 we know that there is a constant C_1 that depends on k, q, u such that

$$\begin{aligned} &|\mathcal{P}^{typ}_{\lambda/\varnothing}| - C_1\left(\left|\mathcal{P}_{\lambda/\varnothing}\right| - \left|\mathcal{P}^{typ}_{\lambda/\varnothing}\right|\right) \leq \mathsf{F}_\lambda([u]^k)\left(\frac{1-q}{1-su}\right)^{-\binom{k+1}{2}} \times \\ &\left(\frac{(1-q^{-1})u}{1-su}\right)^{-\binom{k}{2}}\left(\frac{u-s}{1-su}\right)^{-|\lambda|+\binom{k}{2}} \leq |\mathcal{P}^{typ}_{\lambda/\varnothing}| + C_1\left(\left|\mathcal{P}_{\lambda/\varnothing}\right| - \left|\mathcal{P}^{typ}_{\lambda/\varnothing}\right|\right). \end{aligned} \tag{4.36}$$

From [71, Equation (85)] we know that

$$\left|\mathcal{P}_{\lambda/\varnothing}\right| = \prod_{1\leq i<j\leq k} \frac{\lambda_i - \lambda_j + j - i}{j-i} \tag{4.37}$$

and from [71, Equation (86)] we know that

$$\left|\mathcal{P}^{typ}_{\lambda/\varnothing}\right| \geq \prod_{1\leq i<j\leq k} \frac{\lambda_i - \lambda_j - j + i}{j-i}. \tag{4.38}$$

In particular, the equations (4.36), (4.37) and (4.38) imply that

$$\prod_{1\le i<j\le k}\frac{\lambda_i-\lambda_j+j-i}{j-i}-[C_1+1]\left(\left|\mathcal{P}_{\lambda/\varnothing}\right|-\left|\mathcal{P}^{typ}_{\lambda/\varnothing}\right|\right)\le \mathsf{F}_\lambda([u]^k)\left(\frac{1-q}{1-su}\right)^{-\binom{k+1}{2}}\times$$
$$\left(\frac{(1-q^{-1})u}{1-su}\right)^{-\binom{k}{2}}\left(\frac{u-s}{1-su}\right)^{-|\lambda|+\binom{k}{2}}\le \prod_{1\le i<j\le k}\frac{\lambda_i-\lambda_j+j-i}{j-i}+C_1\left(\left|\mathcal{P}_{\lambda/\varnothing}\right|-\left|\mathcal{P}^{typ}_{\lambda/\varnothing}\right|\right).$$

The latter equation now clearly implies (4.33) since

$$0\le \left|\mathcal{P}_{\lambda/\varnothing}\right|-\left|\mathcal{P}^{typ}_{\lambda/\varnothing}\right|\le \prod_{1\le i<j\le k}\frac{\lambda_i-\lambda_j+j-i}{j-i}-\prod_{1\le i<j\le k}\frac{\lambda_i-\lambda_j-j+i}{j-i}\le C_2(\lambda_1-\lambda_k+k)^{\binom{k}{2}-1},$$

for some sufficiently large constant $C_2>0$ depending on k alone. □

4.4 Proof of Theorem 4.1.3

In this section we prove Theorem 4.1.3. We accomplish this in two steps. In the first step we prove that the random vectors $\left(Y^k_1(N,M;k),\dots,Y^k_k(N,M;k)\right)$ (i.e. the projections of the random vectors $Y(N,M;k)$ from Theorem 4.1.3 to their top row) weakly converge to the Hermite ensemble. In the second step we combine the convergence of $\left(Y^k_1(N,M;k),\dots,Y^k_k(N,M;k)\right)$ to the Hermite ensemble, with the fact that our model satisfies the six-vertex Gibbs property from [71, Section 6] to conclude the convergence of $Y(N,M;k)$ to the GUE-corners process of rank k.

4.4.1 Convergence to the Hermite ensemble

We begin by recalling the joint distribution of the eigenvalues $\lambda_1\le\cdots\le\lambda_k$ of a $k\times k$ matrix from the GUE (recall that these were random Hermitian $k\times k$ matrices with density proportional to $e^{-Tr(X^2)/2}$). Specifically, from [10, Equation (2.5.3)] we have the following formula.

Definition 4.4.1. If μ^k_{GUE} denotes the joint distribution of the ordered eigenvalues of a random

$k \times k$ GUE matrix, then μ^k_{GUE} has the following density with respect to Lebesgue measure

$$\mathbf{1}\{x_k > x_{k-1} > \cdots > x_1\} \left(\frac{1}{\sqrt{2\pi}}\right)^k \cdot \frac{1}{\prod_{i=1}^{k-1} i!} \cdot \prod_{1 \le i < j \le k} (x_i - x_j)^2 \prod_{i=1}^{k} e^{-\frac{x_i}{2}}. \tag{4.39}$$

Remark 4.4.2. In the literature, the measure (4.39) is sometimes referred to as the *Hermite ensemble* due to its connection to Hermite orthogonal polynomials.

The main result of this section is as follows.

Proposition 4.4.3. *Under the same assumptions as in Theorem 4.1.3 we have that the random vectors* $Y^k(N,M) = \left(Y^k_1(N,M;k), \ldots, Y^k_k(N,M;k)\right)$ *converge weakly to* μ^k_{GUE} *as* $M \to \infty$.

The starting point of our proof of Proposition 4.4.3 is Lemma 4.2.12, from which we know that

$$\begin{aligned}
&\mathbb{P}^{N,M}_{u,v}(\lambda^k(\pi) = \mu) = A_M(\mu) \cdot B_M(\mu), \text{ where} \\
&A_M(\mu) = \mathsf{F}_\mu([u]^k) \cdot M^{-\binom{k}{2}\cdot(1/2)} \cdot \left(\frac{1-q}{1-su}\right)^{-\binom{k+1}{2}} \left(\frac{(1-q^{-1})u}{1-su}\right)^{-\binom{k}{2}} \cdot \left(\frac{u-s}{1-su}\right)^{-|\mu|+\binom{k}{2}} \\
&B_M(\mu) = f(\mu;[v]^M,\rho) \cdot M^{\binom{k}{2}\cdot(1/2)} \cdot \left(\frac{1-q}{1-su}\right)^{\binom{k+1}{2}} \left(\frac{(1-q^{-1})u}{1-su}\right)^{\binom{k}{2}} \cdot \left(\frac{u-s}{1-su}\right)^{|\mu|-\binom{k}{2}} Z_M^{-1}, \\
&\text{with } Z_M = (q;q)_k \cdot \left(\frac{1-s^{-1}u}{1-su}\right)^k \cdot \left(\frac{1-quv}{1-uv}\right)^{kM}.
\end{aligned} \tag{4.40}$$

We recall that $\mathsf{F}_\mu([u]^k)$ stands for F_μ with $u_1 = \cdots = u_k = u$ and also $f(\mu;[v]^M,\rho)$ stands for $f(\mu;\mathbf{v},\rho)$ with $v_1 = \cdots = v_M = v$. We also recall that $|\mu| = \mu_1 + \cdots + \mu_k$.

The following lemma details the asymptotics of $A_M(\lambda)$ using the combinatorial estimates for $\mathsf{F}_\lambda([u]^k)$ from Lemma 4.3.4.

Lemma 4.4.4. *Suppose that* u,q,s *satisfy* $q \in (0,1)$, $s = q^{-1/2}$, $u > s$. *Fix* $a, A > 0$ *and suppose that* $x_1, \ldots, x_k \in \mathbb{R}$ *satisfy* $A \ge x_k > x_{k-1} > \cdots > x_1 \ge -A$. *Let* $M_0(a,A) \ge 1$ *be sufficiently large so that* $aM_0 - A\sqrt{M_0} \ge 1$. *For all* $M \ge M_0$ *we define* $\lambda(M) \in \mathsf{Sign}^+_k$ *through* $\lambda_i(M) =$

$\lfloor aM + \sqrt{M}x_{k-i+1} \rfloor$ *for* $i = 1, \ldots, k$. *Then we have*

$$\lim_{M\to\infty} A_M(\lambda(M)) = \prod_{1\le i<j\le k} \frac{x_j - x_i}{j-i} = \frac{1}{\prod_{i=1}^{k-1} i!} \cdot \prod_{1\le i<j\le k} (x_j - x_i). \tag{4.41}$$

Moreoever, there is a constant $C > 0$ *(it depends on* k, a, A, u, q*) such that for all* $M \ge M_0$ *we have*

$$|A_M(\lambda(M))| \le C. \tag{4.42}$$

Proof. We first prove (4.42). From Lemma 4.3.5 and Definition 4.2.2 we know that

$$|A_M(\lambda(M))| \le \tilde{C} M^{-\binom{k}{2}\cdot(1/2)} \cdot |\mathcal{P}_{\lambda(M)/\varnothing}| = \tilde{C} \cdot \prod_{1\le i<j\le k} \frac{\lambda_i(M) - \lambda_j(M) + j - i}{M^{1/2}(j-i)},$$

where in the last equality we used (4.37) and $\tilde{C}$ is as in Lemma 4.3.5 . Plugging in the definition of $\lambda_i(M)$ we see that for $M \ge M_0$ we have

$$|A_M(\lambda(M))| \le \tilde{C} \prod_{1\le i<j\le k} \frac{x_j - x_i + 2kM^{-1/2}}{j-i} \le \tilde{C}[2A + 2k]^{\binom{k}{2}},$$

which clearly implies (4.42).

In the remainder of the proof we establish (4.41). By Lemma 4.3.4 we know that there is a constant C that depends on k, u, q such that for all large enough M we have

$$\left| A_M(\lambda(M)) - \prod_{1\le i<j\le k} \frac{\lambda_i(M) - \lambda_j(M) + j - i}{M^{1/2}(j-i)} \right| \le C \cdot M^{-\binom{k}{2}\cdot(1/2)} \cdot [2AM^{1/2} + 1 + k]^{\binom{k}{2}-1}.$$

Using that

$$\lim_{M\to\infty} \prod_{1\le i<j\le k} \left(\lambda_i(M) - \lambda_j(M) + j - i\right) M^{-1/2} = \prod_{1\le i<j\le k} (x_j - x_i),$$

we see that the above equation implies (4.41). □

The following lemma details the asymptotics of $B_M(\lambda)$.

Lemma 4.4.5. *Suppose that v, u, q, s, a, d are as in Definition 4.1.2 and $k \in \mathbb{N}$. Fix $A > 0$ and suppose that $x_1, \ldots, x_k \in \mathbb{R}$ satisfy $A \geq x_k > x_{k-1} > \cdots > x_1 \geq -A$. Let $M_0(a, A) \geq 1$ be sufficiently sufficiently large so that $aM_0 - A\sqrt{M_0} \geq 1$. For all $M \geq M_0$ we define $\lambda(M) \in \mathsf{Sign}_k^+$ through $\lambda_i(M) = \lfloor aM + d\sqrt{M} x_{k-i+1} \rfloor$ for $i = 1, \ldots, k$. Then we have*

$$\lim_{M \to \infty} d^k M^{k/2} B_M(\lambda(M)) = d^{-\binom{k}{2}} \cdot (\sqrt{2\pi})^{-k} \prod_{1 \leq i < j \leq k} (x_j - x_i) \cdot \prod_{i=1}^{k} e^{-x_i^2/2}. \tag{4.43}$$

Moreoever, there is a constant $C > 0$ (it depends on k, a, A, u, v, q) such that for all $M \geq M_0$

$$|d^k M^{k/2} B_M(\lambda(M))| \leq C. \tag{4.44}$$

Lemma 4.4.5 is the main technical result we need in the proof of Theorem 4.1.3. The proof of this lemma is postponed until Section 4.5, and relies on a careful steepest descend analysis using the contour integral formula for $f(\mu; [v]^M, \rho)$ afforded by Lemma 4.3.3.

In the remainder of this section we prove Proposition 4.4.3

Proof. (Proposition 4.4.3) For clarity we split the proof into two steps.

Step 1. Let $\mathbb{W}_k^o$ denote the open Weyl chamber in $\mathbb{R}^k$, i.e.

$$\mathbb{W}_k^o := \{(x_1, \ldots, x_k) \in \mathbb{R}^k : x_k > x_{k-1} > \cdots > x_1\}.$$

Suppose that $R = [a_1, b_1] \times \cdots \times [a_k, b_k]$ is a closed rectangle such that $R \subset \mathbb{W}_k^o$. The purpose of this step is to establish the following statement

$$\lim_{M \to \infty} \mathbb{P}\left(Y^k(N, M) \in R\right) = \int_R \mu_{GUE}^k(dx_1, \ldots, dx_k). \tag{4.45}$$

Let A be sufficiently large so that $A \geq 1 + \max_{1 \leq i \leq k} |a_i| + \max_{1 \leq i \leq k} |b_i|$. In addition if $M \in \mathbb{N}$

is given and $\mu \in \mathsf{Sign}_k^+$ we denote by Q_μ the cube

$$Q_\mu = [\mu_k, \mu_k + 1) \times \cdots \times [\mu_1, \mu_1 + 1)\,.$$

We also write $L_i(M) = \lceil a_i d\sqrt{M} + aM \rceil$ and $U_i(M) = \lfloor b_1 d\sqrt{M} + aM \rfloor$ for $i = 1, \ldots, k$.

We first observe that for all sufficiently large M we have

$$\mathbb{P}\left(Y^k(N,M) \in R\right) = \sum_{\lambda_1 = L_k(M)}^{U_k(M)} \cdots \sum_{\lambda_k = L_1(M)}^{U_1(M)} \mathbb{P}_{u,v}^{N,M}\left(\lambda_i^k(\pi) = \lambda_i \text{ for } i = 1, \ldots, k\right) = \int_{[-A,A]^k} f_M(x_1, \ldots, x_k) dx_1 \cdots dx_k, \tag{4.46}$$

where $f_M(x)$ is a step function that is given by $d^k M^{k/2} A_M(\mu) B_M(\mu)$ if $x d\sqrt{M} + \mathbf{1}_k aM \in Q_\mu$ for some $\mu = (\mu_1, \ldots, \mu_k) \in \mathsf{Sign}_k^+$ such that $L_i(M) \le \mu_{k-i+1} \le U_i(M)$ for $i = 1, \ldots, k$; and $f_M(x) = 0$ otherwise. In the latter formula $\mathbf{1}_k$ is the vector in $\mathbb{R}^k$ with all coordinates equal to 1.

By Lemmas 4.4.4 and 4.4.5 we know that for almost every $x \in [-A, A]^k$ we have

$$\lim_{M \to \infty} f_M(x_1, \ldots, x_k) \to \mathbf{1}_R \cdot \left(\frac{1}{\sqrt{2\pi}}\right)^k \cdot \frac{1}{\prod_{i=1}^{k-1} i!} \cdot \prod_{1 \le i < j \le k} (x_i - x_j)^2 \prod_{i=1}^{k} e^{-\frac{x_i}{2}}$$

and $|f_M(x)| \le C$ for some C that depends on A, u, q, v, k alone. Consequently, by the bounded convergence theorem we see that the $M \to \infty$ limit of (4.46) implies (4.45).

Step 2. The main goal of this step is to prove the following statement. For any open set U with $U \subset \mathbb{W}_k^o$ we have that

$$\liminf_{M \to \infty} \mathbb{P}\left(Y^k(N,M) \in U\right) \ge \int_U \mu_{GUE}^k(dx_1, \ldots, dx_k). \tag{4.47}$$

If we assume the validity of (4.47) then we have that for any open set $O \subset \mathbb{R}^k$,

$$\liminf_{M \to \infty} \mathbb{P}\left(Y^k(N,M) \in O\right) \ge \liminf_{M \to \infty} \mathbb{P}\left(Y^k(N,M) \in O \cap \mathbb{W}_k^o\right) \ge$$

$$\int_{O\cap\mathbb{W}_k^o} \mu_{GUE}^k(dx_1,\ldots,dx_k) = \int_O \mu_{GUE}^k(dx_1,\ldots,dx_k),$$

where in the last equality we used that the density of μ_{GUE}^k is zero outside of $\mathbb{W}_k^o$. The latter inequality and [73, Theorem 3.2.11] imply the weak convergence of $Y_k(N,M)$ to μ_{GUE}^k. Thus it suffices to prove (4.47).

Let U be an open subset of $\mathbb{W}_k^o$. Then by [177, Chapter 1, Theorem 1.4] we know that $U = \cup_{i=1}^\infty R_i$ where R_i are closed rectangles with disjoint interiors. Let $n \in \mathbb{N}$ and $\varepsilon > 0$ be given. For $i = 1,\ldots,n$ we let

$$R_i^\varepsilon = [a_1^i + \varepsilon, b_1^i - \varepsilon] \times \cdots \times [a_k^i + \varepsilon, b_k^i - \varepsilon] \text{ where } R_i = [a_1^i, b_1^i] \times \cdots \times [a_k^i, b_k^i].$$

Using our result from Step 1 we know that

$$\liminf_{M\to\infty} \mathbb{P}\left(Y^k(N,M) \in U\right) \geq \liminf_{M\to\infty} \mathbb{P}\left(Y^k(N,M) \in \cup_{i=1}^n R_i^\varepsilon\right) =$$

$$\liminf_{M\to\infty} \sum_{i=1}^n \mathbb{P}\left(Y^k(N,M) \in R_i^\varepsilon\right) = \sum_{i=1}^n \int_{R_i^\varepsilon} \mu_{GUE}^k(dx_1,\ldots,dx_k).$$

Letting $\varepsilon \to 0$ and applying the dominated convergence theorem with dominating function

$$\mathbf{1}\{x_k > x_{k-1} > \cdots > x_1\} \left(\frac{1}{\sqrt{2\pi}}\right)^k \frac{1}{\prod_{i=1}^{k-1} i!} \cdot \prod_{1\leq i<j\leq k} (x_i - x_j)^2 \prod_{i=1}^k e^{-\frac{x_i}{2}}$$

we conclude that

$$\liminf_{M\to\infty} \mathbb{P}\left(Y^k(N,M) \in U\right) \geq \sum_{i=1}^n \int_{R_i} \mu_{GUE}^k(dx_1,\ldots,dx_k).$$

Letting $n \to \infty$ and using the monotone convergence theorem we conclude that (4.47) holds. □

4.4.2 Gibbs properties

In this section we give the proof of Theorem 4.1.3. The proof will be an easy consequence of Proposition 4.4.3 and the fact that $\mathbb{P}^{N,M}_{u,v}$ satisfies what is known as the six-vertex Gibbs property, while the GUE-corners process satisfies what is known as the continuous Gibbs property. We start by explaining the latter two Gibbs properties. Our discussion will be brief, and we refer the interested reader to [71, Sections 5 and 6] for a more detailed exposition.

We define several important concepts, adopting some of the notation from [95]. Let GT_k denote the set of k-tuples of *distinct* integers

$$\mathsf{GT}_n = \{\lambda \in \mathbb{Z}^n : \lambda_1 < \lambda_2 < \cdots < \lambda_k\}.$$

We let GT^+_k be the subset of GT_k with $\lambda_1 \geq 0$. We say that $\lambda \in \mathsf{GT}_k$ and $\mu \in \mathsf{GT}_{k-1}$ *interlace* and write $\mu \preceq \lambda$ if

$$\lambda_1 \leq \mu_1 \leq \lambda_2 \leq \cdots \leq \mu_{k-1} \leq \lambda_k.$$

Let GT^k denote the set of sequences

$$\mu^1 \preceq \mu^2 \preceq \cdots \preceq \mu^k, \quad \mu^i \in \mathsf{GT}_i, \ 1 \leq i \leq k.$$

We call elements of GT^k *half-strict* Gelfand-Tsetlin patterns (they are also known as monotonous triangles, cf. [144]). We also let GT^{k+} be the subset of GT^k with $\mu^k \in \mathsf{GT}^+_k$. For $\lambda \in \mathsf{GT}_k$ we let $\mathsf{GT}_\lambda \subset \mathsf{GT}^k$ denote the set of half-strict Gelfand-Tsetlin patterns $\mu^1 \preceq \cdots \preceq \mu^k$ such that $\mu^k = \lambda$.

We turn back to the notation from Section 4.1.1 and consider $\pi \in \mathcal{P}_N$. For $k = 1, \ldots, N$ we have that if we define $\mu^k_i(\pi) = \lambda^k_{k-i+1}(\pi)$ for $i = 1, \ldots, k$ then $\mu^k \in \mathsf{GT}^+_k$. In addition, $\mu^{k+1} \succeq \mu^k$ for $k = 1, \ldots, N-1$. Consequently, the sequence $\mu^1, \ldots, \mu^k$ defines an element of GT^{k+}. It is easy to see that the map $h : \mathcal{P}_k \to \mathsf{GT}^{k+}$, given by $h(\pi) = \mu^1(\pi) \preceq \cdots \preceq \mu^k(\pi)$, is a bijection. For

$\lambda \in \mathsf{GT}_k^+$ we let

$$\mathcal{P}_k^\lambda = \{\pi \in \mathcal{P}_k : \lambda_i^k(\pi) = \lambda_{k-i+1} \text{ for } i = 1, \dots, k\}.$$

One observes that by restriction, the map h is a bijection between GT_λ and $\mathcal{P}_k^\lambda$. Given $\pi \in \mathcal{P}_k^\lambda$ and a vertex path configuration $(i_1, j_1; i_2, j_2)$ we let $N_{\pi,\lambda}(i_1, j_1; i_2, j_2)$ denote the number of vertices $(x, y) \in [1, \lambda_k] \times [1, k] \cap \mathbb{Z}^2$ with arrow configuration $(i_1, j_1; i_2, j_2)$. We abbreviate $N_1 = N_{\pi,\lambda}(0, 0; 0, 0)$, $N_2 = N_{\pi,\lambda}(1, 1; 1, 1)$, $N_3 = N_{\pi,\lambda}(1, 0; 1, 0)$, $N_4 = N_{\pi,\lambda}(0, 1; 0, 1)$, $N_5 = N_{\pi,\lambda}(1, 0; 0, 1)$, and $N_6 = N_{\pi,\lambda}(0, 1; 1, 0)$.

With the above notation we make the following definition.

Definition 4.4.6. Fix $w_1, w_2, w_3, w_4, w_5, w_6 > 0$. A probability distribution ρ on GT^{k+} is said to satisfy the *six-vertex Gibbs property* (with weights $(w_1, w_2, w_3, w_4, w_5, w_6)$) if the following holds. For any $\lambda \in \mathsf{GT}_k^+$ such that

$$\sum_{(\mu^1, \dots, \mu^k) \in \mathsf{GT}^{k+} : \mu^k = \lambda} \rho\left(\mu^1, \dots, \mu^k\right) > 0$$

we have that the measure ν on $\mathcal{P}_k^\lambda$ defined through

$$\nu(h^{-1}(\omega)) = \rho(\omega | \mu^k = \lambda)$$

satisfies the condition

$$\nu(h^{-1}(\omega)) \propto w_1^{N_1} w_2^{N_2} w_3^{N_3} w_4^{N_4} w_5^{N_5} w_6^{N_6}.$$

In the above $\rho(\cdot | \mu^k = \lambda)$ stands for the measure ρ conditioned on $\mu^k = \lambda$ and the numbers $N_1, \dots, N_6$ are defined with respect to λ and the path collection $\pi = h^{-1}(\omega)$.

Remark 4.4.7. In simple terms, Definition 4.4.6, states that a probability measure on GT^{k+} satisfies the six-vertex Gibbs property if it can be realized from a measure of the type (4.2) with vertex weights $w_1, \dots, w_6$ for the six types of vertices under the bijection h.

One readily observes by the definition of $\mathbb{P}_{u,v}^{N,M}$ that if ω is $\mathbb{P}_{u,v}^{N,M}$-distributed and we define

$\mu_i^j(\pi) = \lambda_{j-i+1}^j(\pi)$ for $1 \le i \le j \le k$ then the law of $\left(\mu_i^j\right)_{1\le i\le j\le k}$ satisfies the six-vertex Gibbs property with weights

$$(w_1, w_2, w_3, w_4, w_5, w_6) = \left(1, \frac{u - s^{-1}}{us - 1}, \frac{us^{-1} - 1}{us - 1}, \frac{u - s}{us - 1}, \frac{u(s^2 - 1)}{us - 1}, \frac{1 - s^{-2}}{us - 1}\right). \tag{4.48}$$

The change of sign above compared to (4.1) is made so that the above weights are positive (recall $u > s > 1$ in our case).

We next explain the continuous Gibbs property. We start by introducing some terminology from [70] and [95]. Let C_n be the *Weyl chamber* in $\mathbb{R}^n$ i.e.

$$C_n := \{(x_1, ..., x_n) \in \mathbb{R}^n : x_1 \le x_2 \le \cdots \le x_n\}.$$

For $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^{n-1}$ we write $x \succeq y$ to mean that

$$x_1 \le y_1 \le x_2 \le y_2 \le \cdots \le x_{n-1} \le y_{n-1} \le x_n.$$

For $x = (x_1, ..., x_n) \in C_n$ we define the *Gelfand-Tsetlin polytope* to be

$$GT_n(x) := \{(x^1, ..., x^n) : x^n = x, x^k \in \mathbb{R}^k, x^k \succeq x^{k-1}, 2 \le k \le n\}.$$

We define the *Gelfand-Tsetlin cone* GT^n to be

$$GT^n = \{y \in \mathbb{R}^{n(n+1)/2} : y_i^{j+1} \le y_i^j \le y_{i+1}^{j+1},\ 1 \le i \le j \le n-1\}.$$

We make the following definition after [95].

Definition 4.4.8. A probability measure μ on GT^n is said to satisfy the *continuous Gibbs property* if conditioned on y^n the distribution of $(y^1, ..., y^{n-1})$ under μ is uniform on $GT_n(y^n)$.

Remark 4.4.9. We refer the reader to [71, Section 5] for a detailed discussion of the definition of the uniform measure on $GT_n(y)$, but in words the latter is a compact affine surface of finite dimension, which carries a natural uniform measure that is proportional to the Lebesgue measure on the affine space spanned by this surface.

With the above notation we are finally ready to give the proof of Theorem 4.1.3.

Proof. (Theorem 4.1.3) By Proposition 4.4.3 we know that $Y^k(N,M) = \left(Y_1^k(N,M;k),\dots,Y_k^k(N,M;k)\right)$ converge weakly to μ^k_{GUE} as $M \to \infty$. Observe that by the interlacing conditions $\lambda^i(\pi) \preceq \lambda^{i+1}(\pi)$, for all $1 \le i \le k-1$ we have that

$$Y_1^k(N,M;k) \le Y_i^j(N,M;k) \le Y_k^k(N,M;k) \text{ for all } 1 \le i \le j \le k.$$

Since $Y_1^k(N,M;k)$ and $Y_k^k(N,M;k)$ weakly converge we conclude from the last inequality that the random vectors $Y(N,M;k)$ are tight.

Let $Y(\infty) = (Y_i^j(\infty) : 1 \le i \le j \le k)$ denote any subsequential limit of $Y(N(M),M;k)$, and let $Y(N(M_n),M_n;k)$ be a subsequence converging weakly to $Y(\infty)$. In view of Proposition 4.4.3 we know that the joint distribution of $(Y_1^k(\infty),\dots,Y_k^k(\infty))$ is μ^k_{GUE}. Furthermore, from our discussion earlier in the section, we know that the distribution of $\mu_i^j(\pi) = \lambda^j_{j-i+1}(\pi)$ for $1 \le i \le j \le k$, where π has distribution $\mathbb{P}_{u,v}^{N(M_n),M_n}$ satisfies the six-vertex Gibbs property with weights $w_1,\dots,w_6$ as in (4.48). We may now apply [71, Proposition 6.7] and conclude that $Y(\infty)$ satisfies the continuous Gibbs property. We remark that in [71, Proposition 6.7] the roles of n and k are swapped compared to our present notation and one should take $b(n) = d\sqrt{M_n}$ and $a(n) = aM_n$ in that proposition.

Since $Y(\infty)$ satisfies the continuous Gibbs property and its top row $(Y_1^k(\infty),\dots,Y_k^k(\infty))$ has law μ^k_{GUE}, we conclude that $Y(\infty)$ is the GUE-corners process of rank k. Since the sequence $Y(N,M;k)$ is tight and all weak subsequential limits are given by the GUE-corners process we conclude that $Y(N,M;k)$ converges weakly to the GUE-corners process of rank k as desired. □

4.5 Asymptotic analysis

In this section we prove Lemma 4.4.5. We accomplish this in Section 4.5.2 after we introduce some useful notation for the proof in Section 4.5.1.

4.5.1 Setup

Recall from Definition 4.1.2 that our parameters q, u, v, s satisfy

$$q \in (0,1), \qquad q = s^{-2}, \qquad 1 < s < u < v^{-1}, \tag{4.49}$$

which we assume in what follows. If we assume the same notation as in Lemma 4.4.5 then in view of (4.40) and Lemma 4.3.3 we have for $M \geq M_0$ that

$$\begin{aligned} &d^k M^{k/2} B_M(\lambda(M)) = d^k M^{\binom{k+1}{2}\cdot(1/2)} \cdot \left(\frac{1-q}{1-su}\right)^{\binom{k+1}{2}} \left(\frac{(1-q^{-1})u}{1-su}\right)^{\binom{k}{2}} \cdot \left(\frac{u-s}{1-su}\right)^{-\binom{k}{2}} \cdot \\ &\oint_\gamma \cdots \oint_\gamma \prod_{1\leq\alpha<\beta\leq k} \frac{u_\alpha - u_\beta}{u_\alpha - q u_\beta} \cdot \prod_{i=1}^k \frac{s(1-su)}{(1-su_i)(1-s^{-1}u)} \left(\frac{1-su_i}{u_i - s} \cdot \frac{u-s}{1-su}\right)^{\lambda_i(M)} \times \\ &\prod_{i=1}^k \prod_{j=1}^M \left(\frac{1-qu_iv_j}{1-u_iv_j} \cdot \frac{1-uv_j}{1-quv_j}\right) \prod_{i=1}^k \frac{du_i}{2\pi\iota}. \end{aligned} \tag{4.50}$$

Above, we can take γ to be a zero-centered positively oriented circle of radius u and we recall that $\iota = \sqrt{-1}$.

Recall from Definition 4.1.2 the constants

$$\begin{aligned} &a = \frac{v\left(u - s^{-1}\right)\left(s^{-1}u - 1\right)}{(1-uv)(1-s^{-2}uv)}, \qquad b = \frac{(s^2-1)}{(u-s)(1-su)} \\ &c = \frac{1}{2}\left(a\left(\frac{1}{(u-s)^2} - \frac{s^2}{(1-su)^2}\right) - \frac{s^{-4}v^2}{(1-s^{-2}uv)^2} + \frac{v^2}{(1-uv)^2}\right), \qquad d = \frac{-\sqrt{2c}}{b}. \end{aligned} \tag{4.51}$$

We establish the following statement about the constants in (4.51).

Lemma 4.5.1. *For u, v, s, q satisfying the conditions from* (4.49)*, we have the following inequalities*

$$a > 0, \qquad b < 0, \qquad c > 0, \qquad d > 0.$$

Proof. Since every factor in $a = \frac{v(u-s^{-1})(s^{-1}u-1)}{(1-uv)(1-quv)}$ is positive we conclude that $a > 0$. Examining the factors of $b = \frac{(s^2-1)}{(u-s)(1-su)}$ shows that $(1 - su)$ is negative and the other factors are positive so $b < 0$. Once we show that c is positive we will conclude that $d = -\frac{\sqrt{2c}}{b}$ is also positive. Showing c is positive requires a short argument that we present below.

Simplifying c gives

$$c = \frac{v(1-q)(1-s^{-1}v)T}{2(s^{-1}-u)(s^{-1}u-1)(1-uv)^2(1-quv)^2},$$

where

$$T = 1 + s^{-2} - 2s^{-2}uv + s^{-3}u^2v + s^{-1}u^2v - 2s^{-1}u.$$

From the above factorization formula for c, we see that to show that $c > 0$ it suffices to prove that $T < 0$. Let us put $v = yu^{-1}$ and $u = rs$ so that (4.49) becomes the condition $r > 1$ and $0 < y < 1$. In these variables we have

$$T(r, y) = 1 + q - 2qy + qyr + ry - 2r = r(qy + y - 2) + (1 + q - 2qy).$$

The latter is a linear function in r with a leading negative coefficient. Thus its maximum on $[1, \infty)$ is attained when $r = 1$ and then $T(1, y) = -(1 - y)(1 - q) < 0$. We conclude that $T(r, y) < 0$ for all $r > 1$ and $y \in (0, 1)$, which proves that $c > 0$ as desired. □

Definition 4.5.2. If $z \in \mathbb{C} \setminus \{0\}$ we define $\log(z) = \log|z| + \iota\phi$ where $z = |z|e^{\iota\phi}$ with $\phi \in (-\pi, \pi]$ (i.e. we take the principal branch of the logarithm). For $u_1, \ldots, u_k \in \mathbb{C}$ such that $u_i \neq qu_j$ and

$u_i \neq s$ we define

$$p(u_1, \ldots, u_k) = p(\vec{u}) = \prod_{1 \leq \alpha < \beta \leq k} \frac{u_\alpha - u_\beta}{u_\alpha - q u_\beta} \cdot \prod_{i=1}^{k} \frac{s(1-su)}{(1-su_i)(1-s^{-1}u)}. \tag{4.52}$$

We also define the functions

$$G(z) = a \cdot \log\left(\frac{1-sz}{z-s}\right) + \log\left(\frac{1-qzv}{1-zv}\right) - a \cdot \log\left(\frac{1-su}{u-s}\right) + \log\left(\frac{1-quv}{1-uv}\right), \tag{4.53}$$

$$g(z) = \log\left(\frac{1-sz}{z-s}\right) - \log\left(\frac{1-su}{u-s}\right), \tag{4.54}$$

and for $x \in \mathbb{R}$ we let $h_M(x)$ be the unique element of $(-1, 0]$ so that $aM + dx\sqrt{M} + h_M(x)$ is an integer. In the latter equations q, u, v, s are as in (4.49) and a, d are as in (4.51). Finally, we define

$$A_k = d^k \cdot \left(\frac{1-q}{1-su}\right)^{\binom{k+1}{2}} \left(\frac{(1-q^{-1})u}{1-su}\right)^{\binom{k}{2}} \cdot \left(\frac{u-s}{1-su}\right)^{-\binom{k}{2}}. \tag{4.55}$$

Definition 4.5.3. We let C denote the positively oriented contour that goes from $u - 2\iota u$ straight up to $u + 2\iota u$ and then follows the half-cirlce of radius $2u$ centered at u, see Figure 4.8. For $\varepsilon \in (0, 1)$ we also denote by C^ε the contour that goes from $u - \iota\varepsilon$ straight up to $u + \iota\varepsilon$.

We may deform the γ contours in (4.50) to the contour C from Definition 4.5.3 without crossing any poles of the integrals, which by Cauchy's theorem does not change the value of the integral. After doing this contour deformation and utilizing the notation from Definition 4.5.2 we see that if $M \geq M_0$ we have

$$\begin{aligned} d^{-k} M^{k/2} B_M(\lambda(M)) = A_k \cdot M^{\binom{k+1}{2} \cdot (1/2)} \cdot \oint_C \cdots \oint_C p(\vec{u}) \cdot \\ \exp\left(\sum_{i=1}^{k} MG(u_i) + \sqrt{M} dx_i g(u_i) + h_M(x_i) g(u_i)\right) \prod_{i=1}^{k} \frac{du_i}{2\pi\iota}. \end{aligned} \tag{4.56}$$

Our asymptotic analysis in the next section depends on a careful study of the functions G and g along the contour C. We establish several useful properties in the following lemma.

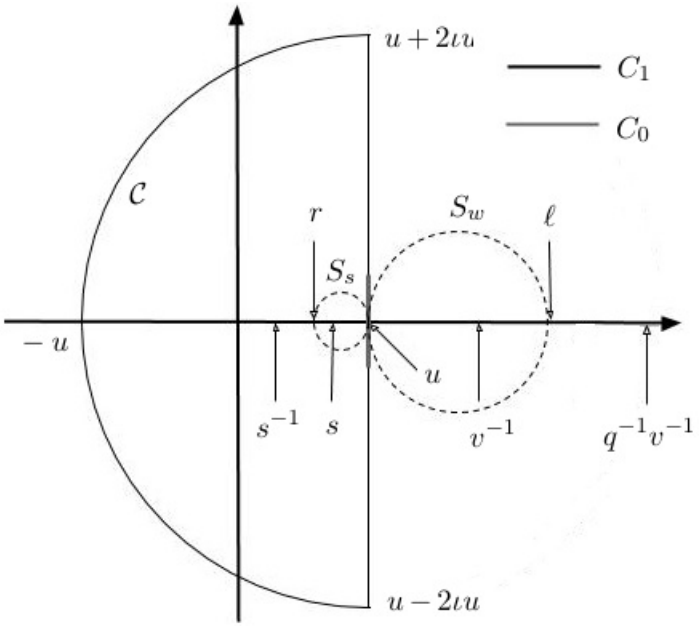

Figure 4.8: The figure represents the contour C from Definition 4.5.3, in addition to S_s, S_w as in the proof of Lemma 4.5.4 and the contours C_0, C_1 as in the proof of Lemma 4.4.5 in Section 4.5.2

Lemma 4.5.4. *Suppose that G, g are as in Definition 4.5.2 and C, C^ε is as in Definition 4.5.3. We have*

$$G(u) = g(u) = G'(u) = 0, \quad G''(u) = 2c, \quad g'(u) = b. \tag{4.57}$$

For any $z \in C$ we have that

$$Re[G(z)] \leq 0. \tag{4.58}$$

Moreover, for any $\varepsilon \in (0,1)$ there exists $\delta > 0$ such that if $z \in C \setminus C^\varepsilon$

$$Re[G(z)] \leq -\delta. \tag{4.59}$$

There exists $\varepsilon_1 \in (0,1)$ and $C_1 > 0$ such that if $z \in C^{\varepsilon_1}$ we have that

$$\left|G(z) - c(z-u)^2\right| \leq C_1|z-u|^3, \quad 2C_1\varepsilon_1 < c, \quad |g(z) - b(z-u)| \leq C_1|z-u|^2. \tag{4.60}$$

Proof. The fact that $G(u) = g(u) = 0$ is immediate from the definition. Next we have by a direct

computation that

$$G'(z) = a \cdot \frac{q^{-1}-1}{(1-sz)(z-s)} - \frac{v(1-q^{-1})}{(q^{-1}-vz)(1-vz)},$$

from which one checks directly (using the definition of a) that $G'(u) = 0$. Similar direct computations show that $G''(u) = 2c$ and $g'(u) = b$.

By definition, we have that

$$Re[G(z)] = a \log\left|\frac{z-s^{-1}}{z-s}\right| + \log\left|\frac{z-q^{-1}v^{-1}}{z-v^{-1}}\right| - a\log\left|\frac{u-s^{-1}}{u-s}\right| - \log\left|\frac{u-q^{-1}v^{-1}}{u-v^{-1}}\right|.$$

Let ℓ denote the unique point in the segment $[s^{-1}, s]$ such that $\frac{\ell - s^{-1}}{s-\ell} = \frac{u-s^{-1}}{u-s}$, and r be the unique point in the segment $[v^{-1}, q^{-v}v^{-1}]$ such that $\frac{q^{-1}v^{-1}-r}{r-v^{-1}} = \frac{q^{-1}v^{-1}-u}{v^{-1}-u}$. We also denote by S_s the circle, whose diameter is given by the segment $[\ell, u]$ and by S_w the circle whose diameter is given by the segment $[u, r]$, see Figure 4.8.

The circles S_s and S_w are sometimes called Apollonius circles and they satisfy the properties

$$\left|\frac{z-s^{-1}}{z-s}\right| \leq \frac{u-s^{-1}}{u-s} \text{ if } z \text{ lies outside of } S_s \text{ and } \left|\frac{z-s^{-1}}{z-s}\right| \geq \frac{u-s^{-1}}{u-s} \text{ if } z \text{ lies inside } S_s;$$

$$\left|\frac{z-q^{-1}v^{-1}}{z-v^{-1}}\right| \leq \frac{q^{-1}v^{-1}-u}{v^{-1}-u} \text{ if } z \text{ lies outside of } S_w \text{ and } \left|\frac{z-q^{-1}v^{-1}}{z-v^{-1}}\right| \geq \frac{q^{-1}v^{-1}-u}{v^{-1}-u},$$

if z lies inside S_w. Since C lies outside of $S_w \cup S_s$ except for the point u and $a > 0$ we conclude that for all $z \in C$ we have $Re[G(z)] \leq Re[G(u)] = 0$, while for any $z \in C \setminus \{u\}$ we have $Re[G(u)] < Re[G(u)] = 0$. This proves (4.58) and by continuity of G on C we also see that for any $\varepsilon > 0$ there is a $\delta > 0$ such that (4.59) holds.

Finally, from our work above we know that in a neighborhood of u we have

$$G(z) = c(z-u)^2 + O(|z-u|^3) \text{ and } g(z) = b(z-u) + O(|z-u|^2).$$

We can thus find $\varepsilon_0 \in (0,1)$ and $C_1 > 0$ such that if $|z-u| \le \varepsilon_0$ we have

$$\left|G(z) - c(z-u)^2\right| \le C_1|z-u|^3, \quad |g(z) - b(z-u)| \le C_1|z-u|^2.$$

Finally, since $c > 0$ we can pick $\varepsilon_1 < \varepsilon_0$ sufficiently small so that $2C_1\varepsilon_1 < c$ and then all the inequalities in (4.60) hold. This suffices for the proof. □

4.5.2 The steepest descent argument

In this section we prove Lemma 4.4.5.

Proof. (Lemma 4.4.5) We follow the same notation as in Lemma 4.4.5 and Section 4.5.1 above. For clarity we split the proof into four steps.

Step 1. Let $\varepsilon_1 \in (0,1)$ be as in the statement of Lemma 4.5.4. We also let $\delta_1 > 0$ be as in Lemma 4.5.4 for $\varepsilon = \varepsilon_1$. We denote by C_0 the contour C^{ε_1} and by C_1 the contour $C \setminus C^{\varepsilon_1}$, see Figure 4.8. We have that $C = C_0 \cup C_1$ and C_0 is a small piece near u while C_1 is the part of C away from u. In view of (4.56) we have that if $M \ge M_0$ we have

$$\begin{aligned} d^{-k}M^{k/2}B_M(\lambda(M)) &= A_k \cdot M^{\binom{k+1}{2}\cdot(1/2)} \cdot \sum_{\sigma_1,\ldots,\sigma_k \in \{0,1\}} B(\sigma_1,\ldots,\sigma_k), \text{ where} \\ B(\sigma_1,\ldots,\sigma_k) &= \oint_{C_{\sigma_1}} \cdots \oint_{C_{\sigma_k}} p(\vec{u}) \exp\left(\sum_{i=1}^{k} MG(u_i) + \sqrt{M}dx_i g(u_i) + h_M(x_i)g(u_i)\right) \prod_{i=1}^{k} \frac{du_i}{2\pi\iota}. \end{aligned} \tag{4.61}$$

In this step we prove that if $\sigma_1,\ldots,\sigma_k \in \{0,1\}$ are such that $|\sigma| = \sigma_1 + \cdots + \sigma_k \ge 1$ we have that

$$B(\sigma_1,\ldots,\sigma_k) = O\left(e^{-(\delta_1/2)M}\right), \tag{4.62}$$

where the constant in the big O notation depends on k, a, A, u, v, q.

Let $K_1, K_2 > 0$ be such that if $u_1, \ldots, u_k, z \in C$ we have

$$|g(z)| \leq K_1 \text{ and } |p(u_1, \ldots, u_k)| \leq K_2.$$

Then in view of the definition of δ_1, and equations (4.58), (4.59) we have that if $u_i \in C_{\sigma_i}$ for $i = 1, \ldots, k$ we have

$$\left| p(\vec{u}) \exp\left(\sum_{i=1}^{k} MG(u_i) + \sqrt{M} dx_i g(u_i) + h_M(x_i) g(u_i) \right) \right| \leq K_2 \exp\left(-M|\sigma|\delta_1 + \sqrt{M} k K_1 [Ad + 1] \right).$$

In deriving the above equation we used that $|e^z| \leq e^{|z|}$ for any complex z. The above equation now clearly implies (4.62).

Step 2. In view of (4.61) and (4.62) we see that to prove the lemma it suffices to show that

$$\lim_{M \to \infty} A_k \cdot M^{\binom{k+1}{2} \cdot (1/2)} B(0, \ldots, 0) = d^{-\binom{k}{2}} \cdot (\sqrt{2\pi})^{-k} \prod_{1 \leq i < j \leq k} (x_j - x_i) \cdot \prod_{i=1}^{k} e^{-x_i^2/2}, \tag{4.63}$$

and that there is a constant $C > 0$ depending on k, a, A, u, v, q such that

$$|A_k \cdot M^{\binom{k+1}{2} \cdot (1/2)} B(0, \ldots, 0)| \leq C. \tag{4.64}$$

In this step we prove (4.64). The proof of (4.63) is given in the next steps.

We perform a change of variables $u_i = u + \iota \cdot M^{-1/2} \cdot y_i$ for $i = 1, \ldots, k$. This gives the formula

$$A_k \cdot M^{\binom{k+1}{2} \cdot (1/2)} B(0, \ldots, 0) = A_k \int_{\mathbb{R}^k} \hat{p}_M(\vec{y}) \exp\left(\sum_{i=1}^{k} H_M(y_i) \right) \prod_{i=1}^{k} \mathbf{1}\{|y_i| \leq \varepsilon_1 M^{1/2}\} \frac{dy_i}{2\pi}, \tag{4.65}$$

where

$$H_M(y) = G(u + \iota \cdot M^{-1/2}y) + \sqrt{M}dx_i g(u + \iota \cdot M^{-1/2}y) + h_M(x_i)g(u + \iota \cdot M^{-1/2}y), \text{ and}$$
$$\hat{p}_M(\vec{y}) = \prod_{1\le\alpha<\beta\le k} \frac{\iota y_\alpha - \iota y_\beta}{(1-q)u + \iota y_\alpha M^{-1/2} - q\iota y_\beta M^{-1/2}} \prod_{i=1}^{k} \frac{s(1-su)}{(1-su-s\iota M^{-1/2}y_i)(1-s^{-1}u)}. \tag{4.66}$$

We see from (4.66) and (4.60) that there are constants $c_1, c_2 > 0$ that depend on k, a, A, u, v, q such that for all $M \in \mathbb{N}$ and $y_1, \ldots, y_k \in \mathbb{R}$ we have

$$\left| \hat{p}_M(\vec{y}) \exp\left(\sum_{i=1}^{k} H_M(y_i) \right) \prod_{i=1}^{k} \mathbf{1}\{|y_i| \le \varepsilon_1 M^{1/2}\} \right| \le h(\vec{y}),$$

where

$$h(\vec{y}) = c_1 \cdot \prod_{1\le\alpha<\beta\le k} |y_\alpha - y_\beta| \cdot \exp\left(-(c/2) \sum_{i=1}^{k} y_i^2 + c_2 \cdot \sum_{i=1}^{k} |y_i| \right).$$

Combining the last inequality and (4.65) we conclude that for all $M \ge M_0$ we have

$$\left| A_k \cdot M^{\binom{k+1}{2}\cdot(1/2)} B(0, \ldots, 0) \right| \le A_k \int_{\mathbb{R}^k} h(\vec{y}) \prod_{i=1}^{k} \frac{dy_i}{2\pi},$$

which implies (4.64).

Step 3. In this step we prove (4.63). From our work in the previous step we know that $h(\vec{y})$ is a dominating function for the functions

$$\hat{p}_M(\vec{y}) \exp\left(\sum_{i=1}^{k} H_M(y_i) \right) \prod_{i=1}^{k} \mathbf{1}\{|y_i| \le \varepsilon_1 M^{1/2}\},$$

which in view of (4.60) and (4.66) converge pointwise to

$$\prod_{1\le\alpha<\beta\le k} \frac{\iota y_\alpha - \iota y_\beta}{(1-q)u} \prod_{i=1}^{k} \frac{s e^{-c y_i^2 + \iota d x_i y_i}}{1 - s^{-1}u}.$$

Consequently, by the dominated convergence theorem, we conclude that

$$\lim_{M\to\infty} A_k \cdot M^{\binom{k+1}{2}\cdot(1/2)} B(0,\dots,0) = A_k \cdot ((1-q)u)^{-\binom{k}{2}} \cdot \left(\frac{s}{1-s^{-1}u}\right)^k \times \int_{\mathbb{R}^k} \prod_{1\le\alpha<\beta\le k} (\iota y_\alpha - \iota y_\beta) \prod_{i=1}^k e^{-cy_i^2 + \iota d b x_i y_i} \frac{dy_i}{2\pi}. \tag{4.67}$$

Substituting A_k from (4.55) and performing the change of variables $z_i = \sqrt{2c}y_i$ (recall that $d = \frac{-\sqrt{2c}}{b}$) we obtain

$$\lim_{M\to\infty} A_k \cdot M^{\binom{k+1}{2}\cdot(1/2)} B(0,\dots,0) = d^{-\binom{k}{2}} \int_{\mathbb{R}^k} \prod_{1\le\alpha<\beta\le k} (\iota z_\alpha - \iota z_\beta) \prod_{i=1}^k e^{-z_i^2/2 - \iota x_i z_i} \frac{dz_i}{2\pi}.$$

We next use the formula for the Vandermonde determinant

$$\prod_{1\le\alpha<\beta\le k} (\iota z_\alpha - \iota z_\beta) = (\iota)^{\binom{k}{2}} \det\left[z_i^{k-j}\right]_{i,j=1}^k,$$

and the linearity of the determinant to conclude that

$$\lim_{M\to\infty} A_k \cdot M^{\binom{k+1}{2}\cdot(1/2)} B(0,\dots,0) = d^{-\binom{k}{2}} (\iota)^{\binom{k}{2}} \det\left[\psi_{k-j}(x_i)\right]_{i,j=1}^k, \text{ where} \tag{4.68}$$

$$\psi_{k-j}(x) = \int_{\mathbb{R}} z^{k-j} e^{-z^2/2 - \iota x z} \frac{dz}{2\pi}.$$

We claim that

$$(\iota)^{\binom{k}{2}} \det\left[\psi_{k-j}(x_i)\right]_{i,j=1}^k = (\sqrt{2\pi})^{-k} \prod_{1\le i<j\le k} (x_j - x_i) \cdot \prod_{i=1}^k e^{-x_i^2/2}. \tag{4.69}$$

Notice that (4.68) and (4.69) together imply (4.63). We have thus reduced the proof of the lemma to establishing (4.69), which we do in the next and final step.

Step 4. In this step we prove (4.69). Let $h_n(x)$ stands for the n-th Hermite polynomial, i.e.

$$h_n(x) = (-1)^n e^{\frac{x^2}{2}} \partial_x^n e^{-\frac{x^2}{2}}, \tag{4.70}$$

see e.g. [10, Section 3.2.1] for the definition and basic properties of these polynomials. Our first observation is that for $n \in \mathbb{Z}_{\geq 0}$ we have

$$\psi_n(x) = (-\iota)^n (\sqrt{2\pi})^{-1} e^{-\frac{x^2}{2}} h_n(x). \tag{4.71}$$

We argue this by induction on n with base case $n = 0$ being true in view of

$$\psi_0(x) = \int_{\mathbb{R}} e^{-z^2/2 - \iota x z} \frac{dz}{2\pi} = (\sqrt{2\pi})^{-1} \cdot e^{-x^2/2},$$

where we used the formula for the characteristic function of a standard normal random variable. Suppose we know that (4.71) holds for n and differentiate both sides with respect to x. For the right side we have using (4.70) that

$$\partial_x \left((-\iota)^n (\sqrt{2\pi})^{-1} e^{-\frac{x^2}{2}} h_n(x) \right) = -(-\iota)^{-n} (\sqrt{2\pi})^{-1} e^{-\frac{x^2}{2}} h_{n+1}(x) = (-\iota)^{n+2} (\sqrt{2\pi})^{-1} e^{-\frac{x^2}{2}} h_{n+1}(x),$$

while for the left side we have

$$\partial_x \psi_n(x) = \int_{\mathbb{R}} z^n \partial_x e^{-z^2/2 - \iota x z} \frac{dz}{2\pi} = (-\iota) \int_{\mathbb{R}} z^{n+1} e^{-z^2/2 - \iota x z} \frac{dz}{2\pi} = (-\iota) \psi_{n+1}(x),$$

where we can differentiate under the integral by the rapid decay of the integrand near infinity. The last two equations imply (4.71) for $n + 1$ and so we conclude that (4.71) holds for all $n \in \mathbb{Z}_{\geq 0}$ by induction.

In view of (4.71) and the linearity of the determinant we see that to prove (4.69) it suffices to

show that

$$\det\left[h_{k-j}(x_i)\right]_{i,j=1}^{k} = \prod_{1\le i<j\le k}(x_i - x_j).$$

The latter is now clear since $h_n(x)$ is a monic polynomial of degree n, cf. [10, (3.2.3)], and so

$$\det\left[h_{k-j}(x_i)\right]_{i,j=1}^{k} = \det\left[x_i^{k-j}\right]_{i,j=1}^{k} = \prod_{1\le i<j\le k}(x_i - x_j),$$

by the Vandermonde determinant formula. This suffices for the proof. □

Chapter 5: Epidemic dynamics in inhomogeneous populations

This chapter is based on the physics article [117]. Along with myself the article has authors Kyle Kawagoe, Serina Chang, Greg Huber, Lucy Li, Jonathan Miller, Reuven Pnini, Boris Veytsman, and Yllanes David. Kyle Kawagoe and I are joint first authors.

5.1 Introduction

A strong temptation in modeling a system consisting of many similar parts is to make the assumption that these parts have identical properties. Accordingly, the classical models in epidemiology assume (often implicitly) that everyone has the same propensity to be infected and, if infected, the same propensity to infect others [101]. This assumption may be justified when differences in the salient parameters are small. However, one of the interesting features of the current COVID-19 pandemic is the huge variation in infectivity: small numbers of infectious events or individuals seem to be responsible for a large number of cases [82, 127, 2, 143, 75, 53]. This feature seems to be present in other coronavirus epidemics including SARS [102, 194, 137] and MERS [145, 122, 58]. One can point to different explanations for this phenomenon: individual variations in viral shedding [165], in droplet production (see the review in [79]), in contact networks [8], and differences in the features of ventilation systems at certain events and venues [141, 56]. Inhomogeneity seems to have played an important role for other epidemics as well [83, 139, 111], leading to the rule of thumb that "20% of patients produce 80% of infections" [178]. However, it seems that for coronavirus-related infections the variability is even higher than that heuristic.

There are two related, but distinct, notions of superspreading in this literature, namely, *superspreading events* and *superspreading individuals*. Superspreading events are events that produce many infections. Superspreading individuals (*superspreaders*) are specific people that produce

many infections (such as Typhoid Mary in the early 1900s). As one might imagine, in reality, some combination of these two processes is present. In this paper, however, we set our sights on the latter phenomenon: a superspreader is always an individual, rather than an event.

It is reasonable to assume that a variability in infectivity is accompanied by a variability in susceptibility. Common explanations of variability in individual infectivity — increased shedding, increased exposure period, and increased personal contacts — all suggest that increased infectivity may correlate with increased susceptibility. Thus superspreaders might be more prominent at the early stages of an epidemic. During the course of an epidemic, the fraction of superspreaders will typically decrease with time. This would lead to a change in the apparent value of the average transmission rate, which could make it difficult to evaluate the effectiveness of mitigation measures. This effect might be quite large and is not captured by many standard models. John Cardy has observed that some models seem to be unaware that the mean of an exponential growth is not the exponential of the mean [51]. Understanding the effect of inhomogeneity would increase the fidelity of models based on real-world data, and lead to more effective public policy.

Several recent works (see, e.g., [94, 130, 147]) have addressed the issue of heterogeneity in the population, but they either concentrate on specific distributions or treat the variability in infectivity and susceptibility separately, without considering the effect of a possible correlation between the two.

In this work we discuss the epidemic dynamics for a population with variable infectivity potential accompanied by variable individual susceptibility. We obtain the results for the general case of an arbitrary distribution of susceptibility and infectivity. We also give a nonintuitive calculation of $\mathcal{R}_0$ that quantifies the effect of superspreaders on the early growth rate of the epidemic and find that it depends strongly on the correlation between susceptibility and infectivity.

Moreover, one of the distributions holds a special interest. If we assume that the main driver of inhomogeneity is diversity in the number of social contacts for an individual, then data [140] on the distribution of these contacts suggests a very wide distribution of infectivity and susceptibility.

An important question for modeling the inhomogeneity is whether the result depends only on

the moments of the distribution (mean, variance, skewness, ...) or on the behavior of the tails of the distribution. The answer to this question could inform the construction of realistic predictive models in the future. We discuss both the cases of fat tails and skinny tails, and the transition between these regimes.

The rest of the paper is organized as follows: In Section 5.2, we give a mathematical description of the dynamics of our model. In Section 5.3, we reduce our model to a one dimensional integro-differential equation, analyze the long time dynamics, and describe an early time criterion for epidemic outbreak. In Section 5.4, we compare the results of our model for different distributions of population attributes, including an empirical one from anonymized cell phone data. We end with our discussion and conclusions in Section 5.5. In the Appendices, we provide derivations which are relevant to the main text and we discuss some of the methodological aspects of our empirical data.

5.2 The Model

Classic SIR models [101] divide the population into three compartments: susceptible S, infected I, and recovered (or dead) R. The rate of new infections in this model is proportional to the number of encounters of susceptible persons with the infected persons, while the rate of recovery is proportional to the number of infected persons. This gives us the well-known SIR equations

$$\begin{aligned} \dot{I} &= \beta SI - \gamma I, \\ \dot{R} &= \gamma I, \end{aligned} \tag{5.1}$$

where S, I, and R are the fractions of susceptible, infected, and recovered persons to the constant population size, dot means the time derivative, β and γ are non-negative constants, and we use the fact that, with our normalization, the fraction of susceptible persons S satisfies the equation

$$S + I + R = 1. \tag{5.2}$$

We use the simplest version of the model, which accounts neither for additional births and deaths, nor for population migration. Additionally, we do not allow for the possibility of recovered individuals being reinfected.

We now allow the parameters to be different for different individuals. Namely, let the infection rate β in equation (5.1) be the product of individual susceptibility s and infectivity σ. To obtain the rate of infection, we integrate over the values of s for susceptible individuals and over the values of σ for infected individuals. Note that in our model the values of s, σ, and γ are fixed for each person and do not change with time.

Let $p(\sigma, s, \gamma)\,\mathrm{d}\sigma\,\mathrm{d}s\,\mathrm{d}\gamma$ be the probability that a person selected uniformly at random from the population has susceptibility s, and, when infected, has infectivity σ and recovery rate γ. Note that p does not change with time in our model. We will have reason to make repeated use of the averaging operator $\mathbb{E}$: for any function $f(\sigma, s, \gamma)$, we define

$$\mathbb{E}[f] \equiv \int f(\sigma, s, \gamma) p(\sigma, s, \gamma)\,\mathrm{d}\sigma\,\mathrm{d}s\,\mathrm{d}\gamma. \tag{5.3}$$

Equations (5.1) and (5.2) should now be rewritten, because I, R and S are not just functions of time t, but also depend on s, σ, and γ. Namely, let $I(\sigma, s, \gamma, t)\,d\sigma\,ds\,d\gamma$ be the probability that a person selected uniformly from the entire population at time t is infected and has (initial) susceptibility s, infectivity σ and recovery rate γ. Similarly we introduce $S(\sigma, s, \gamma, t)$ and $R(\sigma, s, \gamma, t)$. Then equation (5.2) becomes

$$S(\sigma, s, \gamma, t) + I(\sigma, s, \gamma, t) + R(\sigma, s, \gamma, t) = p(\sigma, s, \gamma), \tag{5.4}$$

and equations (5.1) become

$$\dot{I}(\sigma, s, \gamma, t) = S(\sigma, s, \gamma, t) s \int \eta I(\eta, q, \kappa, t)\,\mathrm{d}q\,\mathrm{d}\kappa\,\mathrm{d}\eta - \gamma I(\sigma, s, \gamma, t), \tag{5.5}$$

$$\dot{R}(\sigma, s, \gamma, t) = \gamma I(\sigma, s, \gamma, t). \tag{5.6}$$

When the proportion of infected individuals is small, $S(\sigma, s, \gamma, t)$ in equation (5.5) is close to $p(\sigma, s, \gamma)$, giving a linear approximation of equation (5.5). For distributions where is γ a constant, it can be shown (Section 5.7) that the early behavior of an epidemic is determined by $\mathcal{R}_0 = \mathbb{E}[\sigma s]/\gamma$.

The total fraction $\Omega(t)$ of persons who have ever been infected at time t is the sum of currently infected and recovered individuals. If we stratify Ω by s, σ, and γ, we can write down

$$\Omega(t) = \int T(\sigma, s, \gamma, t)\, \mathrm{d}\sigma\, \mathrm{d}s\, \mathrm{d}\gamma \tag{5.7}$$

with

$$T(\sigma, s, \gamma, t) = I(\sigma, s, \gamma, t) + R(\sigma, s, \gamma, t). \tag{5.8}$$

The final epidemic size is

$$\Omega_\infty = \lim_{t\to\infty} \int T(\sigma, s, \gamma, t)\, \mathrm{d}\sigma\, \mathrm{d}s\, \mathrm{d}\gamma. \tag{5.9}$$

We will use index 0 for the initial conditions in equations (5.5) and (5.6), so $I_0(\sigma, s, \gamma) = I(\sigma, s, \gamma, 0)$ etc.

5.3 Analytic results

In this section we discuss the general properties of our model. We assume that the distribution of infectivity and susceptibility is such that the moments $\mathbb{E}[\sigma]$, $\mathbb{E}[s]$, and $\mathbb{E}[\sigma s]$ as defined in equation (5.3) exist. If the distribution is so heavy tailed that these moments do not exist then important integrals in our analysis will not converge. This is not a merely technical restriction. For instance the short time behavior of the model should be quite different if $\mathbb{E}[\sigma s]$ is infinite.

Let us introduce the notation:

$$\phi(t) = \frac{1}{\mathbb{E}[\sigma]} \int \sigma I(\sigma, s, \gamma, t)\, \mathrm{d}\sigma\, \mathrm{d}s\, \mathrm{d}\gamma, \tag{5.10}$$

$$\psi(t) = \frac{1}{t} \int_0^t \phi(t')\mathrm{d}t'. \tag{5.11}$$

An individual has infectivity σ if infected and 0 if not. Therefore $\mathbb{E}[\sigma]$ is the maximal average infectivity (when everyone is infected simultaneously), and $\phi(t)$ is the ratio of the current average infectivity and the maximal one. Further, $\psi(t)$ is the historical average of $\phi(t)$. Both these quantities are thus between zero and one. In our model (without births or immigration and no persons with zero recovery rate) there are no infected persons at $t \to \infty$, so in this limit

$$\lim_{t\to\infty} \phi(t) = 0, \quad \lim_{t\to\infty} \psi(t) = 0. \tag{5.12}$$

It is shown in Section 5.6 that the stratified fraction of people who ever have been infected at time t [see equations (5.7) and (5.8)] is

$$T(\sigma, s, \gamma, t) = p(\sigma, s, \gamma) - S_0(\sigma, s, \gamma)\, \mathrm{e}^{-s\mathbb{E}[\sigma]\psi(t)t}. \tag{5.13}$$

For outbreaks started with a small number of infected persons, almost all remaining individuals are susceptible, so $S_0 \approx p$. The number of currently infected individuals is

$$\begin{aligned} I(\sigma, s, \gamma, t) = &- S_0(\sigma, s, \gamma)\, \mathrm{e}^{-s\mathbb{E}[\sigma]\psi(t)t} \\ &+ \mathrm{e}^{-\gamma t}\big(p(\sigma, s, \gamma) - R_0(\sigma, s, \gamma)\big) \\ &+ \gamma S_0(\sigma, s, \gamma) \int_0^t \mathrm{d}t' \mathrm{e}^{-\gamma(t-t') - s\mathbb{E}[\sigma]\psi(t')t'}. \end{aligned} \tag{5.14}$$

Therefore, if we know $\psi(t)$, then we know the full solution. It is shown in Section 5.6 that $\psi(t)$ is a solution of the equation

$$\mathbb{E}[\sigma]\frac{\mathrm{d}(t\psi(t))}{\mathrm{d}t} = \int \sigma \left[I_0(\sigma, s, \gamma)\mathrm{e}^{-\gamma t} - S_0(\sigma, s, \gamma) \int_0^t \mathrm{d}t' \mathrm{e}^{-\gamma(t-t')} \frac{\mathrm{d}}{\mathrm{d}t'} \left(\mathrm{e}^{-s\mathbb{E}[\sigma]\psi(t')t'} \right) \right] \mathrm{d}\sigma\, \mathrm{d}s\, \mathrm{d}\gamma. \tag{5.15}$$

To study the behavior of equation (5.15) we will make several simplifying assumptions. First,

we assume a constant recovery rate across the population:

$$p(\sigma, s, \gamma') = p(\sigma, s)\delta(\gamma - \gamma'). \tag{5.16}$$

This means that the other variables (S, I, R) are also proportional to $\delta(\gamma - \gamma')$; we will use the same notation for them as functions of σ and s.

Second, we assume the initial number of recovered individuals is zero,

$$R_0(\sigma, s) = 0. \tag{5.17}$$

Third, we assume that the initial distribution of infected persons is proportional to $p(\sigma, s)$, and is small:

$$\begin{aligned} I_0(\sigma, s) &= \varepsilon p(\sigma, s), \\ S_0(\sigma, s) &= (1 - \varepsilon)p(\sigma, s), \\ 0 < \varepsilon &\ll 1. \end{aligned} \tag{5.18}$$

To see why any other initial distribution I_0 that is small should behave similarly see Section 5.7.

With these assumptions equation (5.15) can be further transformed from an integro-differential equation to a first-order differential equation

$$\begin{aligned} \mathbb{E}[\sigma](\dot{\nu} + \gamma\nu) = &- (1 - \varepsilon) \int_0^\infty \mathrm{d}s \int_0^\infty \mathrm{d}\sigma \sigma p(\sigma, s) \mathrm{e}^{-s\mathbb{E}[\sigma]\nu(t)} \\ &+ \mathbb{E}[\sigma], \end{aligned} \tag{5.19}$$

for the function $\nu(t) = \psi(t)t$ (See equation (5.47)).

To numerically solve equation (5.15) it is convenient to rewrite it as two first-order differential equations (See Section 5.10). In the rest of this section we discuss the properties of the solution of this equation.

Let us start with the final epidemic size [equation (5.9)]. It can be shown (Section 5.6) that at

$t \to \infty$ the function $\psi(t)$ in equation (5.11) behaves as $1/t$. Choose L so that at large t,

$$\psi(t) \approx \frac{L}{t}, \quad L \geq 0. \tag{5.20}$$

Then equation (5.9) with T from equation (5.13) becomes (see Section 5.6)

$$\Omega_\infty(\varepsilon) = 1 - (1-\varepsilon)\mathbb{E}\left[\mathrm{e}^{-s\mathbb{E}[\sigma]L}\right], \tag{5.21}$$

where L is the unique nonnegative root of the equation

$$F(L) = L - \frac{1}{\gamma} + \frac{1-\varepsilon}{\gamma\mathbb{E}[\sigma]}\mathbb{E}\left[\sigma\,\mathrm{e}^{-s\mathbb{E}[\sigma]L}\right] = 0. \tag{5.22}$$

We are interested in an infection started with a small number of initial cases, which corresponds to $\varepsilon \to 0$. If in this limit equation (5.22) has a strictly positive root, the final epidemic size

$$\Omega_\infty = \Omega_\infty(0) \tag{5.23}$$

is non-zero, and does not depend on ε: in other words, the epidemic takes off. If the limit does not have a strictly positive root then the infection immediately dies out and the final epidemic size is 0. In this $\varepsilon \to 0$ limit $F(0) = 0$ and $F(1/\gamma) > 0$, so equation (5.22) has a positive (non-zero) root if $\mathrm{d}F(0)/\mathrm{d}L < 0$. Taking the derivative, we see that a non-zero root corresponds to the condition

$$\mathcal{R}_0 = \frac{\mathbb{E}[\sigma s]}{\gamma} \geq 1. \tag{5.24}$$

Given this result, we take a brief detour from our discussion of $t \to \infty$. Another way to look at epidemic spread is to study the short term behavior of the solution. Our analysis (Section 5.7) shows that the initial small infection spreads with exponential rate $\mathcal{R}_0 = \mathbb{E}[\sigma s]/\gamma$ determined by equation (5.24). The upshot is that the growth rate of the epidemic is highly dependent on how correlated the infectivity and susceptibility are.

One naive generalization of $\mathcal{R}_0$ from the SIR model, i.e., the average number of secondary infections produced by a typical infection would be $\mathcal{R}_0^{'} = \mathbb{E}[\sigma]\mathbb{E}[s]/\gamma$. To explain why $\mathcal{R}_0$, rather than $\mathcal{R}_0^{'}$, determines the exponential growth rate of the infected population we will illustrate what the two quantities measure. If we choose a person from the *entire* population uniformly at random and infect them, then the average number of secondary infections would be $\mathcal{R}_0^{'}$. For instance if a cruise ship travels somewhere and almost everyone is infected, then when they return home the expected number of secondary infections each person produces will be $\mathcal{R}_0^{'}$. On the other hand a person who was infected via community spread (early in the epidemic) will cause on average $\mathcal{R}_0$ secondary infections. The difference between these cases is that in the first case almost all travelers are infected so the fact that someone is infected tells us little about their susceptibility, whereas in the second case people are infected via community spread which occurs with a probability proportional to their susceptibility early in the epidemic. See Section 5.7 for details.

We will now continue our discussion of the final epidemic size with some limiting cases. As mentioned above, for an epidemic to spread, it is necessary that $\mathcal{R}_0^{'} = \mathbb{E}[\sigma s]/\gamma \geq 1$. Near this transition, where $\mathcal{R}_0^{'} \approx 1$, we may write down an approximation for L. Again, we will be interested in the limit of small initial epidemic size $\varepsilon \to 0$, although it is not difficult to generalize the following result for non-zero ε. Let $\mathcal{R}_0 > 1$. Assuming that L is small, and that $p(\sigma, s)$ falls off quickly enough for large s, we may approximate equation (5.22) as

$$F_0(L) \approx \gamma\mathbb{E}[\sigma]L - \mathbb{E}[\sigma] + \ + \int \sigma p(\sigma, s)\left(1 - s\mathbb{E}[\sigma]L + \frac{(s\mathbb{E}[\sigma]L)^2}{2}\right) \mathrm{d}s\, \mathrm{d}\sigma. \tag{5.25}$$

Therefore, if we get close enough to the transition where $\mathbb{E}[\sigma s] - \gamma$ is small

$$L \approx \frac{2}{\mathbb{E}[\sigma]\mathbb{E}[\sigma s^2]}(\mathbb{E}[\sigma s] - \gamma). \tag{5.26}$$

In this regime equation (5.21) gives the total epidemic size as

$$\Omega_\infty \approx \int p(\sigma, s)(1 - \mathrm{e}^{-2s(\mathbb{E}[\sigma s]-\gamma)/\mathbb{E}[\sigma s^2]})\, \mathrm{d}s\, \mathrm{d}\sigma. \tag{5.27}$$

Let us now briefly discuss the opposite limit. Instead of γ being so large that the epidemic almost doesn't start, we study γ so small that the epidemic infects almost everyone. It is expected that if $\gamma = 0$, then the entire population will eventually become infected; that is, $\Omega_\infty = 1$. Equation (5.21) shows that in this case $L \to \infty$. It is easy to show that for small γ, $L \approx 1/\gamma$, and equation (5.21) predicts an exponentially small number of individuals not infected.

This framework allows one to make predictions for a number of specific distributions discussed in the next section. We conclude the general discussion with one very interesting case: when the distribution has a very small number of "superspreaders", individuals with anomalously high infectivity. (Here very small means small enough to not appreciably change $\mathbb{E}[\sigma s]$.) A relevant question is whether these individuals have an oversized contribution in the epidemic. Equations (5.21) and (5.22) show that this is *not* the case, and the contribution of superspreaders is limited by the linear term in the average value of $\mathbb{E}[\sigma s]$ (see Section 5.9). Therefore, while superspreaders still contribute to the dynamics, they are only a primary driver of infection in our model when they significantly change $\mathcal{R}_0$. That being said, increasing the number of superspreaders in a population will increase $\mathcal{R}_0$, which will cause the epidemic to spread faster, and will also cause a larger final epidemic size.

5.4 Results for different distributions of infectivity and susceptibility

Let us further illustrate the general results using specific distributions for s and σ. First, consider an N-component SIR model. That is, there are N different types of individuals who have parameters σ_i, s_i, γ_i and represent a portion of the population p_i, and

$$p(\sigma, s, \gamma) = \sum_{i=1}^{N} p_i \delta(\sigma - \sigma_i)\delta(s - s_i)\delta(\gamma - \gamma_i), \tag{5.28}$$

$\delta(x)$ being Dirac's delta-function. In the case where $N = 1$, this reduces to the standard SIR model. We see in Section 5.7 that this model is a limiting case of the model presented in this paper [1].

[1] In Section 5.7, we set $\gamma_i = \gamma$, but our claim that the N-component SIR model is a special case of our model does not rely on this assumption

Another useful distribution to study is the Gamma distribution with $\sigma = s$. In particular, we are interested in the distribution

$$p(\sigma, s, \gamma') = p(s)\delta(\sigma - s)\delta(\gamma' - \gamma) \tag{5.29}$$

where

$$p(s) = \frac{\beta^\alpha s^{\alpha-1} \mathrm{e}^{-\beta s}}{\Gamma(\alpha)} \tag{5.30}$$

and α, β are positive constants. This system is interesting to study because the integrals involved in solving for L are analytically tractable. In the case where $\alpha = 1$ we recover the exponential distribution and we can find Ω_∞ exactly (equation (5.104)). We analyze the case of the Gamma distribution in Section 5.11.

We further illustrate the dynamics of epidemics using several special cases of distributions of infectivity σ and susceptibility s with the assumption of constant recovery rate γ. (See Section 5.8 for an analysis of which distributions lead to the worst outcomes for the final epidemic size.)

Even with constant γ the answer depends on the probability distribution $p(\sigma, s)$. We discuss three limiting cases: (i) completely independent σ and s, with $p(\sigma, s) = p_\sigma(\sigma)p_s(s)$; (ii) completely positively correlated σ and s with $\sigma \propto s$; and (iii) positively correlated σ and s with a correlation coefficient ρ.

Note that since only the product σs enters the equations, we always can multiply σ by a constant factor f, and s by the factor $1/f$. We choose this factor to ensure that $\mathbb{E}[\sigma] = \mathbb{E}[s]$. In the numerical calculations in this section we used the following parameters roughly following [20, 107, 106]

$$\begin{aligned} \mathbb{E}[\sigma] = \mathbb{E}[s] &= 0.6\,\mathrm{day}^{-1/2}, \\ \gamma &= 0.125\,\mathrm{day}^{-1}, \\ \varepsilon &= 10^{-4}. \end{aligned} \tag{5.31}$$

At present, our understanding of variability in individual suceptibility and infectivity is far from complete. While the consensus is that they have a wide distribution (see the discussion in the

Introduction), the shape of this distribution is not known, and most studies assume a convenient one for their calculations. Since we want to explore the dependence of the dynamics on the distribution itself, rather than on its parameters, we compare two reasonable *a priori* assumptions: a log-normal distribution with the parameters μ and $\tilde{\sigma}$, and a Gamma distribution with the parameters α and β. Another approach is to suggest some mechanism for the variability and choose a distribution that follows this mechanism. One such mechanism is the variability of individual contacts: the more contacts has a person, the higher is their s and σ. It is important to note that in this model s is completely correlated with σ because they are caused by the same mechanism.

We are fortunate to be able to use empirical data about the number of contacts from the "path-crossing" network described in Looi et al. [140]. Their network is constructed from the mobility data provided by SafeGraph, a company that aggregates and anonymizes geolocation data from cell phone applications. SafeGraph collects GPS location pings for millions of adult smartphone users in the United States, where each ping represents the latitude and longitude of one user at one timestamp. Looi et al. [140] transform the set of location pings into a dynamic network, where users are represented as nodes, and edges indicate the number of times two users crossed their paths (see Section 5.12 for the details). We use the number of path crossings as a proxy for the number of users' social contacts, which is in its turn a proxy for susceptibility and infectivity. Due to the number of assumptions here one should be careful with the interpretation of the results. We do not claim that the SafeGraph data provide *the* distribution of σ and s. Rather we think they suggest features of the real distribution.

An interesting feature of the SafeGraph distribution is that it is very wide. The average number of contacts per user is 0.342×10^3, while the standard deviation is 1.04×10^3. We can try to approximate the empirical distribution of contacts using a theoretical distribution. On Figure 5.1 we show log-normal and gamma approximations together with the empirical distribution with the same mean and variance.

In the remainder of this section we discuss the numerical solutions of the model equations for the log-normal, Gamma, and empirical distributions obtained with the approach discussed in

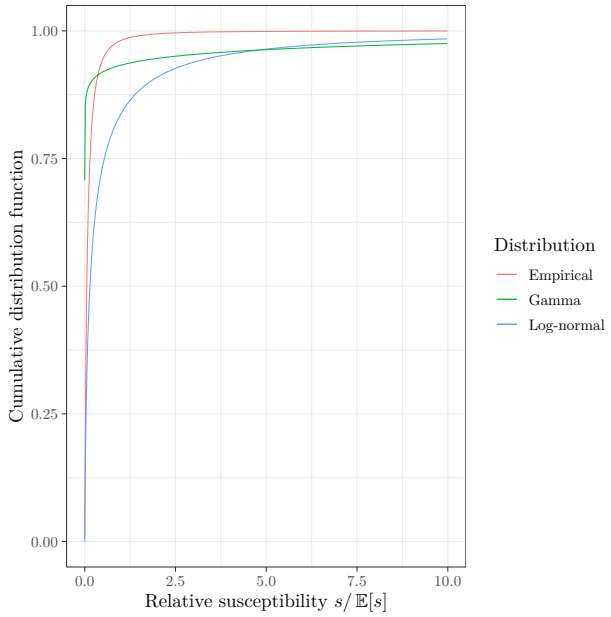

Figure 5.1: Comparison of empirical, log-normal and gamma distributions with the same average infectivity $\mathbb{E}[s] = 0.6\,\mathrm{day}^{-1/2}$ and variance ζ^2 with $\zeta = 4.16\,\mathrm{day}^{-1/2}$.

Section 5.10. See Section 5.11 for analytical solutions in special cases.

In Figure 5.2 we compare the epidemic's progression for log-normal and Gamma distributions with the same mean s and varying distribution widths. We see that a wider distribution leads to a lower epidemic size. When the width of the distribution decreases, the curve goes to the one for the classical SIR model. An interesting feature is that a wide correlated distribution of s and σ leads to an earlier start of the epidemics instead of the S-like curve of the standard SIR model.

In Figure 5.3 we study the influence of the positive correlation between infectivity and susceptibility. For simplicity we show just the final size Ω_∞. As demonstrated by this figure, the more correlated these parameters are, the higher the size is, as predicted by the analysis in the previous section.

For another comparison we take the empirical number of contacts between the individuals (Section 5.12) as a proxy for both s and σ. We renormalize the number of contacts to obtain the average infectivity $\mathbb{E}[s]$ in equation (5.31). This leads to variance $\zeta^2 = 17.27\,\mathrm{day}^{-1}$ ($\zeta = 4.16\,\mathrm{day}^{-1/2}$). Then we fit the parameters of log-normal and Gamma distributions to get the same $\mathbb{E}[s]$ and ζ. All three distributions are shown on Figure 5.1.

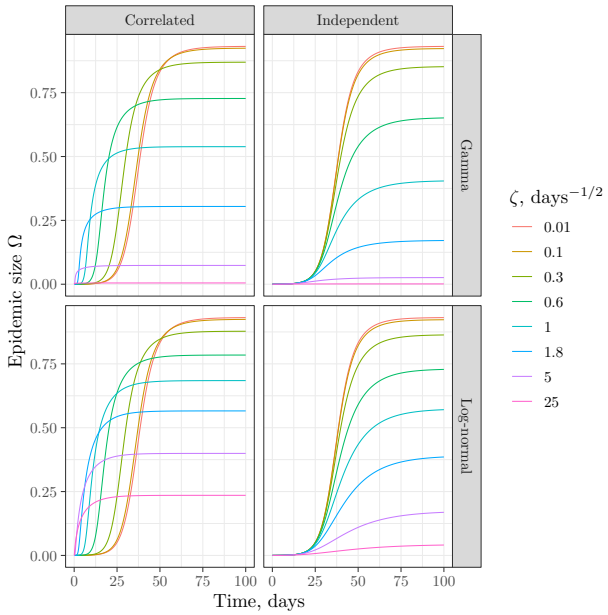

Figure 5.2: Comparison of epidemic spread for log-normal and Gamma distributions of infectivity and susceptibility with standard deviation ζ and parameters in equation (5.31). The cases of independent or completely correlated σ and s are shown.

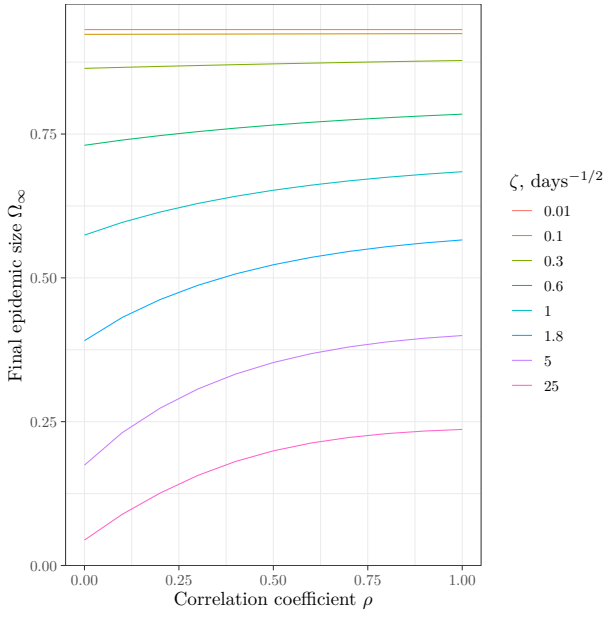

Figure 5.3: Dependence of final epidemic size Ω_∞ on ρ where $(\log(s), \log(\sigma))$ is a Gaussian vector with mean $\mathbb{E}[s] = \mathbb{E}[\sigma] = 0.6\ \text{day}^{-1/2}$ and covariances $\text{Var}(s) = \text{Var}(\sigma) = \zeta^2$, $\text{Cov}(s, \sigma) = \rho\zeta^2$. Note that ρ is the correlation coefficient for $\log(s)$ and $\log(\sigma)$ rather than for s and σ.

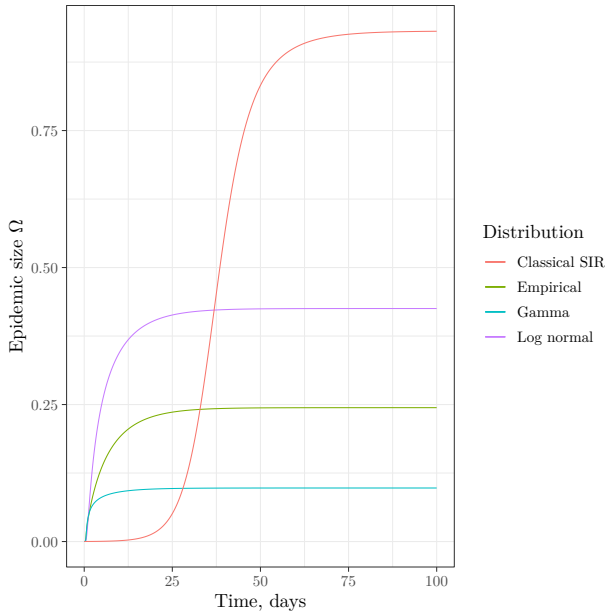

Figure 5.4: Epidemics progression for the distributions shown on Figure 5.1 with parameters in equation (5.31). A classical SIR solution for the same susceptibility and infectivity is also shown.

The results are shown in Figure 5.4 together with the solution for the classical SIR model with the infectivity and susceptibility equal to the averages $\mathbb{E}[s]$ and $\mathbb{E}[\sigma]$.

The figures suggest that, generally speaking, variability in susceptibility and infectivity lowers the final epidemic size, and the correlation between them increases it. Important special cases of this statement are proven in Section 5.8, and based on the figures, we expect it to hold more generally.

Of special interest is the question of whether individuals with high infectivity ("superspreaders") influence the epidemic dynamics and final epidemic size. To model the effect of superspreaders we can discuss a special bimodal distribution of infectivity,

$$p(\sigma) = (1 - \lambda)p_{\mathrm{n}}(\sigma) + \lambda p_{\mathrm{s}}(\sigma), \tag{5.32}$$

where p_{n} describes "normal" persons with low σ, and p_{s} describes superspreaders with high σ. In

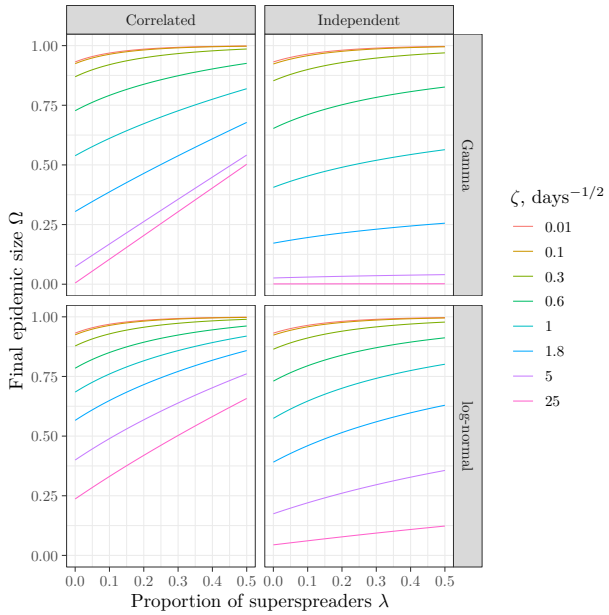

Figure 5.5: Final epidemic size for a mix of normal individuals (same distribution as on Figure 5.2) and superspreaders described by equation (5.33). The effect of superspreaders is at most linear in their proportion.

our numerical experiments we modeled superspreaders using a power-law distribution

$$p_{\mathrm{s}}(\sigma) = \begin{cases} 0, & \sigma < b, \\ (a-1)b^{a-1}\sigma^{-a}, & \sigma \geq b \end{cases} \tag{5.33}$$

with the parameters $a = 4$, $b = 1.2\,\mathrm{day}^{-1/2}$. With these parameters the average infectivity of superspreaders is $1.8\,\mathrm{day}^{-1/2}$, i.e., three times the average infectivity in our simulations. The results are shown on Figure 5.5. We see that the influence of superspreaders is at most linear in their proportion λ. This is not coincidental: as shown in Section 5.9, the effect of superspreaders is at most linear.

5.5 Discussion and conclusions

The aim of any idealized model is to provide insights about the "real world". We believe our model provides several important insights beyond the assumptions involved in its derivation and

treatment.

First, the variation in individual susceptibility and infectivity does matter. All examples studied in Section 5.4 have the same average susceptibility and infectivity—but the outcomes greatly differ. Generally wider distribution lead to lower final epidemic size, and, in the case of correlated infectivity and susceptibility, faster initial outbreak.

Second, the correlation between infectivity and susceptibility is important: the higher the correlation, the larger the epidemic size.

Third, the average and the width of infectivity and susceptibility are *not* enough to predict the outcome: the actual shape of the distribution matters too. The comparisons of log-normal and Gamma distributions in Figure 5.2, and of three different distributions having the same first and second moments in Figure 5.4, demonstrate this clearly.

This conclusion shows that a prediction of the epidemic's spread is a hard task from the practical point of view. Indeed, we never know the exact shape of the distribution, since it involves the measurement of individual infectivity and susceptibility of many people. The sensitivity to the shape of the distribution beyond a couple of moments is bad news for precise predictions.

Having said this, we still need to answer the question of which features of the distribution are the most salient for predictions. There were a number of works stressing the importance of superspreaders: individuals or events with anomalously high potential for spreading (see the Introduction). Our model suggests a more nuanced view. On one hand, because the susceptibility and infectiousness of individuals are correlated through how many people someone interacts with, increasing the number of superspreaders in a way that does not change the average infectivity or susceptibility will increase $\mathcal{R}_0 = \mathbb{E}[\sigma s]/\sigma$, which greatly increases how fast the infection takes off and somewhat increases the final epidemic size. In the unrealistic case where we add pairs of one superspreader and one unusually careful person so that the variance increases and $\mathcal{R}_0$ is unchanged, adding both these people will actually tend to decrease the final epidemic size. This can be seen in equations (5.21) and (5.22) where we have exponentials suppressing the contribution of individuals with anomalously high susceptibility (or high infectivity if these parameters are correlated). This

can also be seen in Section 5.8 and in Figure 5.4. The final result is determined by the average $\mathbb{E}[\sigma s]$ *and* the distribution shape at low to moderate susceptibilities. It should be noted, that for wide distributions median s and mean s are quite different, and our conclusion concerns mean, rather than median, susceptibility.

These conclusions rely on the fact that in our model a recovered person can never be infected again. If we allow for the reinfection of recovered individuals, such as in an SIRS model, we would expect superspreaders to have a much greater impact on the course of the epidemic. This is because their removal from the system at early times is now only temporary. This is an important limitation of our model that would be useful to examine in future investigations.

Perhaps the following analogy may help to understand the meaning of this result. In comic books the outcome of a war is determined by a handful of superheroes and supervillains. In reality it is determined by the combined effort of many people at the lowest rungs of the military hierarchy: privates, petty and junior officers, and so forth. Our conclusion is that epidemic spread is like the "real war" rather than the "comic-book one". This has an essential implication for public health policy. While the prevention of superspreading is important (it changes the exponential growth rate $\mathcal{R}_0 = \mathbb{E}[\sigma s]/\gamma$ and drives down the averages in equations (5.21) and (5.22)), it is the mundane everyday efforts that matter most.

Lastly, we provide a simple, but efficient mathematical apparatus to calculate the epidemic dynamics for a population with variable infectivity and susceptibility, and cast it in a form suitable for numerical estimates. We hope this apparatus might turn out to be useful beyond the insights formulated in this paper.

5.6 Derivation of main equations

This Section is dedicated to the derivation of the main equation and the results of the general analysis in Section 5.3.

First, we derive equation (5.13). Let us add equations (5.5) and (5.6) and use the definitions of

$T(\sigma, s, \gamma, t)$ to obtain

$$\dot{T}(\sigma, s, \gamma, t) = [p(\sigma, s, \gamma) - T(\sigma, s, \gamma, t)]s\mathbb{E}[\sigma]\phi(t) \tag{5.34}$$

By inspection, we may verify that Eq.5.13 is a solution to this differential equation. We see that this solution satisfies the initial conditions

$$T(\sigma, s, \gamma, 0) = p(\sigma, s, \gamma) - S_0(\sigma, s, \gamma) \tag{5.35}$$

We now turn to the derivation of equation (5.14). First, we use equation (5.5) and the definitions of $T(\sigma, s, \gamma, t)$ and $\phi(t)$ to write down

$$\dot{I}(\sigma, s, \gamma, t) = \big(p(\sigma, s, \gamma) - T(\sigma, s, \gamma, t)\big)s\mathbb{E}[\sigma]\phi(t) - \gamma I(\sigma, s, \gamma, t). \tag{5.36}$$

Substituting this expression into equation (5.34), we arrive at

$$\dot{I}(\sigma, s, \gamma, t) = \dot{T}(\sigma, s, \gamma, t) - \gamma I(\sigma, s, \gamma, t), \tag{5.37}$$

or, equivalently

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(\mathrm{e}^{\gamma t} I(\sigma, s, \gamma, t)\big) = e^{\gamma t}\dot{T}(\sigma, s, \gamma, t). \tag{5.38}$$

This differential equation admits a solution

$$I(\sigma, s, \gamma, t) = \mathrm{e}^{-\gamma t}\left(\int_0^t \mathrm{d}t'\,\mathrm{e}^{\gamma t'}\dot{T}(\sigma, s, \gamma, t') + I_0(\sigma, s, \gamma)\right). \tag{5.39}$$

We now integrate the integral in the above equation by parts. In the second step and the second-to-

last step, we will use our solution for $T(\sigma, s, \gamma, t)$ from equation (5.13).

$$e^{\gamma t} I(\sigma, s, \gamma, t) = e^{\gamma t} T(\sigma, s, \gamma, t) - T_0(\sigma, s, \gamma) - \int_0^t dt' \gamma\, e^{\gamma t'} T(\sigma, s, \gamma, t) + I_0(\sigma, s, \gamma) =$$

$$e^{\gamma t} T(\sigma, s, \gamma, t) - (T_0(\sigma, s, \gamma) - I_0(\sigma, s, \gamma)) - \int_0^t dt'\, \gamma e^{\gamma t'} (p(\sigma, s, \gamma) - S_0(\sigma, s, \gamma)) e^{-s\mathbb{E}[\sigma]\psi(t')t'} =$$

$$e^{\gamma t} T(\sigma, s, \gamma, t) - R_0(\sigma, s, \gamma) - \left(e^{\gamma t} - 1\right) p(\sigma, s, \gamma) + S_0(\sigma, s, \gamma)\gamma \int_0^t dt'\, e^{\gamma t'} e^{-s\mathbb{E}[\sigma]\psi(t')t'} =$$

$$e^{\gamma t} (T(\sigma, s, \gamma, t) - p(\sigma, s, \gamma)) + (p(\sigma, s, \gamma) - R_0(\sigma, s, \gamma)) + S_0(\sigma, s, \gamma)\gamma \int_0^t dt' e^{\gamma t'} e^{-s\mathbb{E}[\sigma]\psi(t')t'} =$$

$$- e^{\gamma t} S_0(\sigma, s, \gamma) e^{-s\mathbb{E}[\sigma]\psi(t)t} + (p(\sigma, s, \gamma) - R_0(\sigma, s, \gamma)) + S_0(\sigma, s, \gamma)\gamma \int_0^t dt'\, e^{\gamma t'} e^{-s\mathbb{E}[\sigma]\psi(t')t'}, \quad (5.40)$$

and therefore

$$I(\sigma, s, \gamma, t) =$$

$$- S_0(\sigma, s, \gamma) e^{-s\mathbb{E}[\sigma]\psi(t)t} + e^{-\gamma t} (p(\sigma, s, \gamma) - R_0(\sigma, s, \gamma)) + S_0(\sigma, s, \gamma)\gamma \int_0^t dt'\, e^{-\gamma(t-t')} e^{-s\mathbb{E}[\sigma]\psi(t')t'}. \quad (5.41)$$

This final line matches Eq. (5.14).

Finally, we derive the equations of motion for $t\psi(t)$ as written in (5.15). We begin by substituting in our solution for $I(\sigma, s, \gamma, t)$ into the definition of $\phi(t)$ in equation (5.10):

$$\mathbb{E}[\sigma]\phi(t) = \int \sigma I(\sigma, s, \gamma, t) d\sigma\, ds\, d\gamma = \int \left(\sigma\gamma S_0(\sigma, s, \gamma) \int_0^t dt'\, e^{-\gamma(t-t') - s\mathbb{E}[\sigma]\psi(t')t'} \right) d\sigma\, ds\, d\gamma +$$
$$\int \sigma \left\{ e^{-\gamma t} \left[p(\sigma, s, \gamma) - R_0(\sigma, s, \gamma) \right] - S_0(\sigma, s, \gamma) e^{-s\mathbb{E}[\sigma]\psi(t)t} \right\} d\sigma\, ds\, d\gamma. \quad (5.42)$$

Noticing that $\phi(t) = \mathrm{d}(\psi(t)t)/\mathrm{d}t$, we get

$$\mathbb{E}[\sigma]\frac{\mathrm{d}(t\psi(t))}{\mathrm{d}t} = \int \left(\sigma\gamma S_0(\sigma,s,\gamma)\int_0^t \mathrm{d}t'\, \mathrm{e}^{-\gamma(t-t')-s\mathbb{E}[\sigma]\psi(t')t'}\right)\mathrm{d}\sigma\,\mathrm{d}s\,\mathrm{d}\gamma + \int \sigma\left\{\mathrm{e}^{-\gamma t}\left[p(\sigma,s,\gamma) - R_0(\sigma,s,\gamma)\right] - S_0(\sigma,s,\gamma)\mathrm{e}^{-s\mathbb{E}[\sigma]\psi(t)t}\right\}\mathrm{d}\sigma\,\mathrm{d}s\,\mathrm{d}\gamma. \quad (5.43)$$

Integrating this equation by parts, we get

$$\mathbb{E}[\sigma]\frac{\mathrm{d}(t\psi(t))}{\mathrm{d}t} = \int \left(\left(\sigma S_0(\sigma,s,\gamma)\left(\mathrm{e}^{-s\mathbb{E}[\sigma]\psi(t)t} - \mathrm{e}^{-\gamma t} - \int_0^t \mathrm{d}t'\,\mathrm{e}^{-\gamma(t-t')}\frac{\mathrm{d}}{\mathrm{d}t'}\left(\mathrm{e}^{-s\mathbb{E}[\sigma]\psi(t')t'}\right)\right)\right) + \int \sigma\left\{\mathrm{e}^{-\gamma t}\left[I_0(\sigma,s,\gamma) + S_0(\sigma,s,\gamma)\right] - S_0(\sigma,s,\gamma)\mathrm{e}^{-s\mathbb{E}[\sigma]\psi(t)t}\right\}\right)\mathrm{d}\sigma\,\mathrm{d}s\,\mathrm{d}\gamma = \int \sigma\left[I_0(\sigma,s,\gamma)\mathrm{e}^{-\gamma t} - S_0(\sigma,s,\gamma)\int_0^t \mathrm{d}t'\mathrm{e}^{-\gamma(t-t')}\frac{\mathrm{d}}{\mathrm{d}t'}\left(\mathrm{e}^{-s\mathbb{E}[\sigma]\psi(t')t'}\right)\right]\mathrm{d}\sigma\,\mathrm{d}s\,\mathrm{d}\gamma, \quad (5.44)$$

which matches equation (5.15).

Let us now derive equation (5.22) and propose an iterative algorithm for its numerical solution.

Assuming constant γ (equation (5.16)), we multiply both sides of equation (5.15) by $\mathrm{e}^{\gamma t}$ and take a time derivative of both sides:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\mathbb{E}[\sigma]\frac{\mathrm{d}(\psi(t)t)}{\mathrm{d}t}\mathrm{e}^{\gamma t}\right) = -\int\left[\sigma S_0(\sigma,s)\mathrm{e}^{\gamma t}\frac{\mathrm{d}}{\mathrm{d}t}\left(\mathrm{e}^{-s\mathbb{E}[\sigma]\psi(t)t}\right)\right]\mathrm{d}\sigma\,\mathrm{d}s. \quad (5.45)$$

Taking the derivative of the left hand side, multiplying by $\mathrm{e}^{-\gamma t}$ and integrating over time, we get

$$\mathbb{E}[\sigma](\dot{\psi}t + \psi + \gamma\psi t) = -\int\left(\sigma S_0(\sigma,s)\mathrm{e}^{-s\mathbb{E}[\sigma]\psi(t)t}\right)\mathrm{d}\sigma\,\mathrm{d}s + C, \quad (5.46)$$

where C is a constant based on initial conditions. With the initial conditions (5.18), we get $C = \mathbb{E}[\sigma]$. As an aside, we may alternatively write equation (5.46) as a first-order, time-independent equation using $\nu(t) = \psi(t)t$.

$$\mathbb{E}[\sigma](\dot{\nu} + \gamma\nu) = -\int\left(\sigma S_0(\sigma,s)\mathrm{e}^{-s\mathbb{E}[\sigma]\nu}\right)\mathrm{d}\sigma\,\mathrm{d}s + \mathbb{E}[\sigma], \quad (5.47)$$

We rewrite as

$$\dot{\nu} = (1 - \gamma\nu) - \frac{1}{\mathbb{E}[\sigma]} \int \left(\sigma S_0(\sigma, s) \mathrm{e}^{-s\mathbb{E}[\sigma]\nu} \right) \mathrm{d}\sigma \, \mathrm{d}s. \tag{5.48}$$

It is not hard to see that the right hand side is Lipshitz in ν, so the solution exists and is unique on $\mathbb{R}_{\geq 0}$ by a standard application of the Picard-Lindelof theorem. In fact we have a bijection between solutions to the system (5.4), (5.5), (5.6) and solutions to (5.48) given by

$$\nu(t) = \int_0^t \left(\int \sigma I(\sigma, s, \gamma, t') ds d\sigma \right) dt' \tag{5.49}$$

in one direction and by equations (5.13) and (5.14) in the other. Thus existence and uniqueness of solutions to (5.48) implies the existence and uniqueness of solutions to the system (5.4), (5.5), (5.6).

We already know that $\lim_{t\to\infty} \psi(t) = 0$ (equation (5.12)). Suppose that $\int_0^\infty \phi(t)\, dt$ converges, and thus the following limit exists:

$$\lim_{t\to\infty} \psi(t) t = L. \tag{5.50}$$

In this case we obtain

$$\Omega_\infty = 1 - \int S_0(\sigma, s) \mathrm{e}^{-s\mathbb{E}[\sigma]L} d\sigma ds, \tag{5.51}$$

where L is the unique nonnegative root of the equation

$$F(L) = 1 - \gamma L - \frac{1}{\mathbb{E}[\sigma]} \int S_0(\sigma, s) \sigma \, \mathrm{e}^{-s\mathbb{E}[\sigma]L} d\sigma ds = 0. \tag{5.52}$$

As an aside, observe that in the case $S_0(\sigma, s) = (1 - \varepsilon)p(\sigma, s)$ we obtain equation (5.22).

If we take any sequence of initial conditions $\{S_0^n(\sigma, s)\}_{n=0}^\infty$ so that $S_0^n(\sigma, s)$ converges weakly to $p(\sigma, s)$ and $\sigma S_0^n(\sigma, s)$ converges weakly to $\sigma p(\sigma, s)$, then Ω_∞ converges to $1 - \mathbb{E}[e^{-s\mathbb{E}[\sigma]L}]$ uniformly in L, and $F(L)$ converges to $1 - \gamma L - \mathbb{E}[\sigma e^{-s\mathbb{E}[\sigma]L}]$ uniformly in L. This implies that in

this limit the final epidemic toll converges to

$$1 - \mathbb{E}[e^{-s\mathbb{E}[\sigma]L}], \tag{5.53}$$

where L is the unique nonnegative root of

$$1 - \gamma L - \mathbb{E}[\sigma e^{-s\mathbb{E}[\sigma]L}] = 0, \tag{5.54}$$

if such a root exists. It is not hard to see that a nonnegative root exists if and only if $\mathbb{E}[\sigma s]/\gamma > 1$. This parameter $\mathbb{E}[\sigma s]/\gamma$ turns out to be the correct generalization for the basic reproduction rate $\mathcal{R}_0$ in this inhomogeneous SIR model, see the next section for details.

To justify the assumption (5.50) we construct an algorithm to calculate L and prove it converges to a non-negative root of equation (5.22). We use the following iterations We will find the solution using the following iterations:

$$L_0 = \frac{1}{\gamma}, \tag{5.55}$$

$$L_i = \frac{1}{\gamma} - \frac{1-\varepsilon}{\gamma\mathbb{E}[\sigma]}\mathbb{E}\left[\sigma e^{-s\mathbb{E}[\sigma]L_{i-1}}\right], \quad i = 1, 2, \ldots. \tag{5.56}$$

Below we will prove that the sequence L_i converges to the relevant root.

Lemma 5.6.1. *Suppose equation* (5.22) *has non-negative roots, and $\tilde{L}$ is the largest root. Then the sequence $L_0, L_1, \ldots$ converges to $\tilde{L}$.*

Proof. We will prove that for all i

$$\tilde{L} \leq L_i \leq L_{i-1}. \tag{5.57}$$

Then the sequence $L_0, L_1, \ldots$ is bounded and non-increasing, and therefore converges. The limit of this sequence is a root of equation (5.22), and due to inequality (5.57) and the fact that $\tilde{L}$ is the largest root, it converges to $\tilde{L}$.

First, note that from equations (5.22) and (5.55) follows that $\tilde{L} \leq 1/\gamma = L_0$.

For $i = 1$ we have from the iteration equation (5.56) $L_1 \leq L_0$ and, since $\tilde{L} \leq L_0$,

$$L_1 \geq \frac{1}{\gamma} - \frac{1-\varepsilon}{\gamma \mathbb{E}[\sigma]} \mathbb{E}\left[\sigma e^{-s\mathbb{E}[\sigma]\tilde{L}}\right] = \tilde{L}, \tag{5.58}$$

so inequality (5.57) is true.

Suppose this inequality is true for $i-1$, i.e.

$$\tilde{L} \leq L_{i-1} \leq L_{i-2}. \tag{5.59}$$

Then we will prove it for i. Indeed,

$$L_i = \frac{1}{\gamma} - \frac{1-\varepsilon}{\gamma \mathbb{E}[\sigma]} \mathbb{E}\left[\sigma e^{-s\mathbb{E}[\sigma]L_{i-1}}\right] \leq \frac{1}{\gamma} - \frac{1-\varepsilon}{\gamma \mathbb{E}[\sigma]} \mathbb{E}\left[\sigma e^{-s\mathbb{E}[\sigma]L_{i-2}}\right] = L_{i-1} \tag{5.60}$$

and

$$L_i \geq \frac{1}{\gamma} - \frac{1-\varepsilon}{\gamma \mathbb{E}[\sigma]} \mathbb{E}\left[\sigma e^{-s\mathbb{E}[\sigma]\tilde{L}}\right] = \tilde{L} \tag{5.61}$$

In other words if the inequality is true for $i-1$, it is true for i, so it is true for all i. □

Lemma 5.6.2. *Equation* (5.22) *always has a non-negative root no smaller than* ε/γ.

Proof. Similarly to the proof of Lemma 5.6.1 we can prove the inequality

$$\frac{\varepsilon}{\gamma} \leq L_i \leq L_{i-1}. \tag{5.62}$$

Indeed, for any i we can iteratively prove that

$$L_i \geq \frac{1}{\gamma} - \frac{1-\varepsilon}{\gamma \mathbb{E}[\sigma]} \mathbb{E}\left[\sigma e^{-s\mathbb{E}[\sigma]\cdot 0}\right] = \frac{\varepsilon}{\gamma}. \tag{5.63}$$

Therefore the sequence $L_0, L_1, \ldots$ converges to a number no smaller than ε/γ. This number is a root of equation (5.22), which, according to Lemma 5.6.1 is the largest root. □

The last lemma shows that the assumed behavior of $\psi(t)$ at large t is indeed $\psi \approx L/t$.

5.7 Short-time behavior and initial conditions

In this section we show that in a mixed population the parameter that determines whether an infection grows exponentially or dies out is

$$\mathcal{R}_0 = \frac{\mathbb{E}[\sigma s]}{\gamma}.$$

We also show that the long term behavior of the epidemic does not depend on the initial conditions.

At early time, when the proportion of the population infected, and the proportion of the population recovered are very small, equations (5.5) and (5.6) can be linearized as

$$\dot{I}(\sigma, s, \gamma, t) = p(\sigma, s, \gamma)s \int \eta I(\eta, q, \kappa, t) \mathrm{d}\eta \mathrm{d}q \mathrm{d}\kappa - \gamma I(\sigma, s, \gamma, t) \tag{5.64}$$

and

$$\dot{R}(\sigma, s, \gamma, t) = \gamma I(\sigma, s, \gamma, t). \tag{5.65}$$

We consider the case where γ is fixed for the entire population, and the distribution $p(\sigma, s) = \sum_{i=1}^{n} p_i \delta_{\sigma_i, s_i}(\sigma, s)$ is a finite combination of delta functions. With the notation $I_i(t) = I(\sigma_i, s_i, t)$, equations (5.64) and (5.65) become a finite-dimensional system of equations

$$\frac{\mathrm{d}I_i(t)}{\mathrm{d}t} = p_i s_i \left(\sum_{j=1}^{n} \sigma_j I_j(t) \right) - \gamma I_i. \tag{5.66}$$

We rewrite this as

$$\frac{\mathrm{d}I}{\mathrm{d}t} = AI, \tag{5.67}$$

with $I = \big(I_1(t), \dots I_n(t)\big)^T$ and $A_{ij} = p_i s_i \sigma_j - \gamma \mathbb{1}_{i=j}$. Let

$$\sigma = \begin{bmatrix} \sigma_1 \\ \vdots \\ \sigma_n \end{bmatrix}, \qquad (sp) = \begin{bmatrix} s_1 p_1 \\ \vdots \\ s_n p_n \end{bmatrix}, \tag{5.68}$$

Let

$$A_{ij} = |(sp)\rangle\langle\sigma| - \gamma I. \tag{5.69}$$

From this we see that the largest eigenvalue of A is $\mathbb{E}[\sigma s] - \gamma = \langle\sigma|(sp)\rangle - \gamma = \sum_{i=1}^n s_i \sigma_i p_i - \gamma$ with the associated eigenvector $|(sp)\rangle$, and that all other eigenvectors are perpendicular to σ and have eigenvalue $-\gamma$.

Now a general distribution $p(\sigma, s)$ can be approximated by a sum of delta masses, to conclude that the linear equations (5.64) and (5.65) have the largest eigenvalue

$$\lambda = \mathbb{E}[s\sigma] - \gamma \tag{5.70}$$

with corresponding eigenvector $I(\sigma, s) = sp(\sigma, s)$ and all other eigenvectors negative.

If $p(\sigma, s)$ is a compactly supported distribution we conclude that if a small enough proportion of the total population is infected at time zero, then until the proportion of the population that is susceptible drops appreciably below 1, we have

$$I_t(\sigma, s) \sim C \mathrm{e}^{t(\mathbb{E}[s\sigma]-\gamma)} sp(\sigma, s), \tag{5.71}$$

where

$$C = \frac{\int sp(\sigma, s) I_0(\sigma, s) d\sigma ds}{\int s^2 p(\sigma, s)^2 d\sigma ds}.$$

The quantity $\mathcal{R}_0$ is also what epidemiologists measure when they measure the number of secondary infections produced by a typical infection in the very early stages of the epidemic. The key to understanding why this number is $\mathbb{E}[s\sigma]/\gamma$ instead of $\mathbb{E}[s]\mathbb{E}[\sigma]/\gamma$ comes from the word

"typical." Based on equation (5.71), early in the epidemic the probability $q(\sigma, s)$ that a person with infectivity σ and susceptibility s is infected is proportional to $sp(\sigma, s)$, so

$$q(\sigma, s) = \frac{sp(\sigma, s)}{\int sp(\sigma, s) ds d\sigma} = \frac{sp(\sigma, s)}{\mathbb{E}[s]} \tag{5.72}$$

To find the number secondary infections per unit time this "typical infection" produces, we take this person's infectivity and multiply by the average susceptibility in the population to get $\sigma_{\text{typical}}\mathbb{E}[s]$. Averaging σ_{typical} over the measure $q(\sigma, s)$ gives

$$\mathbb{E}[\sigma_{\text{typical}}]\mathbb{E}[s] = \int s\sigma p(\sigma, s) ds d\sigma \frac{\mathbb{E}[s]}{\mathbb{E}[s]} = \mathbb{E}[s\sigma]. \tag{5.73}$$

Multiplying by the typical recovery time $\frac{1}{\gamma}$ gives the expected number of secondary infections.

As with the usual SIR model, if $\mathcal{R}_0 > 1$ the infection will spread and if $\mathcal{R}_0 < 1$ the infection will die out. This allows us to see that the growth rate of an epidemic is highly dependent on how correlated s and σ are, with higher correlation leading to a higher growth rate. In a true population we expect a persons infectivity σ and susceptibility s to be highly correlated through factors like how many people someone interacts with. In particular superspreaders have an outsize effect on the early growth of the epidemic in the most realistic case where s and σ are highly correlated, because in this case $\mathcal{R}_0$ grows like the second moment $\mathbb{E}[\sigma^2]$ of the infectivity rather than the first moment.

The second takeaway is that if the proportion of the population that is infected at time 0 is small enough, there is essentially only one possible initial condition for the system (5.5), (5.6). This can be seen by writing the initial profile of infected $I_0(\sigma, s)$ as a sum of eigenvectors for equations (5.64) and (5.65),

$$I_0(\sigma, s) = Csp(\sigma, s) + I_0'(\sigma, s), \tag{5.74}$$

and comparing with (5.71) to see that $I_0'(\sigma, s)$ has minimal effect, and the long term solution is

almost identical to the solution starting from initial condition

$$I_0(\sigma, s) = Csp(\sigma, s). \tag{5.75}$$

5.8 Worst-case distributions

In this section we discuss which distributions provide the highest possible epidemic size Ω_∞ (the "worst-case scenarios").

We prove two statements

1. *Variability is good.* If s and σ are independent, then the final epidemic size is less than or equal to the final epidemic size of the classical SIR model with $s_0 = \mathbb{E}[s]$, $\sigma_0 = \mathbb{E}[\sigma]$.

2. *Strong positive correlation is bad.* If the marginal distributions of s and σ are known, then the joint distribution $p(\sigma, s)$ that maximizes the final epidemic size is given by the "percentile coupling", where the nth most infectious person is also the nth most susceptible person.

Both these statements follow from the following lemma:

Lemma 5.8.1. *Let μ and ν be two possible joint distributions for (s, σ). Let $\mathbb{E}^\mu$ and $\mathbb{E}^\nu$ denote the expectation with respect to μ and ν respectively, and similarly for final epidemic sizes Ω_∞^μ and Ω_∞^ν. If*

$$\mathbb{E}^\mu[\sigma] \geq \mathbb{E}^\nu[\sigma], \tag{5.76}$$

and for all $c > 0$,

$$\mathbb{E}^\mu[e^{-cs}] \leq \mathbb{E}^\nu[e^{-cs}], \tag{5.77}$$

and also

$$\mathbb{E}^\mu[\sigma e^{-sc}] \leq \mathbb{E}^\nu[\sigma e^{-sc}], \tag{5.78}$$

then

$$\Omega_\infty^\mu \geq \Omega_\infty^\nu. \tag{5.79}$$

Proof. Using equations (5.77) and (5.76) together with equation (5.22) we see that for any $L > 0$,

$$F^{\mu}(L) = L - \frac{1}{\gamma} + \frac{1}{\gamma \mathbb{E}^{\mu}[\sigma]} \mathbb{E}^{\mu}[\sigma \mathrm{e}^{-s\mathbb{E}^{\mu}[\sigma]L}] \leq L - \frac{1}{\gamma} + \frac{1}{\gamma \mathbb{E}^{\nu}[\sigma]} \mathbb{E}^{\nu}[\sigma \mathrm{e}^{-s\mathbb{E}^{\nu}[\sigma]L}] = F^{\nu}(L). \quad (5.80)$$

Let L^{μ} be the unique positive zero of $F^{\mu}(L)$ if such a zero exists, and otherwise let $L^{\mu} = 0$. Now $F^{\mu}(0) = F^{\nu}(0) = 0$ and both are convex functions of L, which together with equation (5.80) gives $L^{\mu} \geq L^{\nu}$.

Then from equations (5.78) and (5.76) we obtain

$$\Omega_{\infty}^{\mu} = 1 - \mathbb{E}^{\mu}[\mathrm{e}^{-s\mathbb{E}^{\mu}[\sigma]L^{\mu}}] \geq 1 - \mathbb{E}^{\nu}[\mathrm{e}^{-s\mathbb{E}^{\nu}[\sigma]L^{\nu}}] = \Omega_{\infty}^{\nu}. \quad (5.81)$$

□

To prove (1) let us take a distribution ν with independent σ and s, and let $\mu = \delta(\sigma - \mathbb{E}^{\nu}[\sigma])\delta(s - \mathbb{E}^{\nu}[s])$. We have $\mathbb{E}^{\mu}[\sigma] = \mathbb{E}^{\nu}[\sigma]$ by definition. From Jensen's ineqality [155, §1.7(iv)]

$$E^{\mu}[\mathrm{e}^{-cs}] = \mathrm{e}^{-c\mathbb{E}^{\nu}[s]} \leq \mathbb{E}^{\nu}[\mathrm{e}^{-cs}], \quad (5.82)$$

and from Jensen's inequality and independence of s and σ under distribution ν we have

$$E^{\mu}[\sigma \mathrm{e}^{-cs}] = \mathbb{E}^{\nu}[\sigma]\mathrm{e}^{-c\mathbb{E}^{\nu}[s]} \leq \mathbb{E}^{\nu}[\sigma \mathrm{e}^{-cs}]. \quad (5.83)$$

Thus the final epidemic size for our arbitrary distribution with independent s and σ is not greater than the final epidemic size of a delta mass with the same mean.

To prove (2) let ν be an arbitrary measure with the correct marginal distributions, and let μ be the percentile coupling: the most susceptible person is the most infectious, the second most susceptible person is the second most infectious and so on. In particular if we sample twice from μ and obtain (s_1, σ_1) and (s_2, σ_2), then with probability 1, the statement $s_1 \geq s_2$ implies $\sigma_1 \geq \sigma_2$. This property implies that if f is an arbitrary decreasing function, and g is an arbitrary increasing

function, then the percentile coupling is the coupling that minimizes the expectation $\mathbb{E}[f(s)g(\sigma)]$ for the given marginal distributions of s and σ. In particular this distribution minimizes $E[\sigma e^{-sc}]$ for all $c > 0$, so it satisfies equation (5.78). It also has the same marginals as the other measure ν, thus inequalities (5.76) and (5.77) are satisfied. Thus for the given marginal distributions of s and σ the percentile coupling is the worst possible joint law in that it maximizes the final epidemic size of the infection.

5.9 The effect of superspreaders

In this Section we discuss the effect of a superspreaders: a small subpopulation of people with anomalously high infectivity.

Consider the distribution of infectivity σ and susceptibility s as a sum of the "normal" distribution p_n and the susperspreaders p_s with the latter having support at $\sigma > \sigma_s$ with large σ_s, as shown in equation (5.32).

The short term behavior is determined by the value of $\mathbb{E}[s\sigma]$, which can be represented as

$$\mathbb{E}[\sigma s] = (1 - \lambda)\mathbb{E}_n[\sigma s] + \lambda \mathbb{E}_s[\sigma s], \tag{5.84}$$

where subscripts n and s denote averaging with the distributions p_n and p_s correspondingly. This equation shows that (i) the only way superspreaders come into short term behavior is the renormalization of average σs, and (ii) their influence is linear in the proporion of superspreaders λ.

Let us discuss the case where the number of superspreaders is low enough, so the contribution of superspreaders to the averages is small, i.e.

$$\lambda \mathbb{E}_s[\sigma s] \ll E_n[\sigma s]. \tag{5.85}$$

In this case the contribution of superspreaders into the short term dynamics is small according to equation (5.3). We are going to show that there is no anomalous contribution to the long term dynamics either.

We are looking into the final epidemic size, which is determined by equations (5.22) and (5.21).

First, consider the case where superspreaders have the same susceptibility distribution as the other individuals. In this case s and σ are independent, and our equations become

$$L - \frac{1}{\gamma} + \frac{1}{\gamma}\mathbb{E}_n\left[\mathrm{e}^{-s\mathbb{E}[\sigma]L}\right] = 0, \tag{5.86}$$

$$\Omega_\infty = 1 - \mathbb{E}_n\left[\mathrm{e}^{-s\mathbb{E}[\sigma]L}\right]. \tag{5.87}$$

We see that in this case the only way superspreaders contribute is the changing of $\mathbb{E}[\sigma]$.

Now consider the case where superspreaders have anomalous suceptibility s, and higher σ corresponds to higher s. Then the contribution of superspreaders is asymptotically small in both equations (5.22) and (5.21), i.e., again no worse than linear in the number of superspreaders.

5.10 Numerically solvable equations

In this Section we will recast equation (5.15) into a set of differential equations suitable for numerical analysis.

With the constant γ assumption (5.16) and initial conditions (5.17) and (5.18), we can write down equation (5.15) as

$$\mathbb{E}[\sigma]\frac{\mathrm{d}(t\psi(t))}{\mathrm{d}t} = \int\left(\int_0^t \mathrm{e}^{-\gamma(t-t')-s\mathbb{E}[\sigma]\psi(t')t'}\,\mathrm{d}t'\gamma(1-\varepsilon) + \mathrm{e}^{-\gamma t} - (1-\varepsilon)\mathrm{e}^{-s\mathbb{E}[\sigma]\psi(t)t}\right)p(\sigma,s)\sigma\,\mathrm{d}\sigma\,\mathrm{d}s, \tag{5.88}$$

with

$$I(\sigma,s,t) = \left(\int_0^t \mathrm{e}^{-\gamma(t-t')-s\mathbb{E}[\sigma]\psi(t')t'}\,\mathrm{d}t'\gamma(1-\varepsilon) + \mathrm{e}^{-\gamma t} - (1-\varepsilon)\mathrm{e}^{-s\mathbb{E}[\sigma]\psi(t)t}\right)p(\sigma,s), \tag{5.89}$$

and

$$T(\sigma,s,t) = p(\sigma,s)\left(1-(1-\varepsilon)\mathrm{e}^{-s\mathbb{E}[\sigma]\psi(t)t}\right). \tag{5.90}$$

The initial condition is

$$\psi(0) = \varepsilon. \tag{5.91}$$

We introduce the function $\nu(t)$:

$$\nu(t) = t\psi(t). \tag{5.92}$$

We multiply both parts of equation (5.88) by $e^{\gamma t}$ and divide by $\mathbb{E}[\sigma]$:

$$e^{\gamma t}\frac{d\nu(t)}{dt} = 1 + \frac{\gamma(1-\varepsilon)}{\mathbb{E}[\sigma]}\int_0^t \left(\int e^{\gamma t' - s\mathbb{E}[\sigma]\nu(t')}\sigma p(\sigma, s)\,d\sigma\,ds\right) dt \tag{5.93}$$

$$- \frac{1-\varepsilon}{\mathbb{E}[\sigma]}\int e^{\gamma t - s\mathbb{E}[\sigma]\nu(t)}\sigma p(s,\sigma)\,d\sigma\,ds. \tag{5.94}$$

We differentiate this equation with respect to t and multiply by $e^{-\gamma t}$:

$$\ddot{\nu} + \gamma\dot{\nu} = (1-\varepsilon)\dot{\nu}\int e^{-s\mathbb{E}[\sigma]\nu}s\sigma p(\sigma, s)\,d\sigma\,ds. \tag{5.95}$$

Let us introduce a new variable

$$\xi = \dot{\nu}, \tag{5.96}$$

then we can write down equation (5.95) as

$$\begin{aligned} \dot{\xi} &= \xi\left[(1-\varepsilon)\int e^{-s\mathbb{E}[\sigma]\nu}s\sigma p(\sigma, s)\,d\sigma\,ds - \gamma\right], \\ \dot{\nu} &= \xi. \end{aligned} \tag{5.97}$$

We need initial conditions for equations (5.97). By definition (5.92), $\nu(0) = 0$. From equations (5.96), (5.92) and (5.91) we get $\xi(0) = \psi(0) = \varepsilon$, so we can write initial conditions as

$$\nu(0) = 0, \quad \xi(0) = \varepsilon. \tag{5.98}$$

Equations (5.97) with the initial conditions (5.98) depend at any moment t on $\xi(t)$ and $\nu(t)$ only, and therefore can be solved by any suitable method for differential equations.

5.11 Special distributions

For several important distributions we can provide analytical results. These results can be used for more sophisticated models, so we provide them below. We are particularly interested in the low-γ limit, where outbreaks are large and not easily controlled.

We discuss the completely correlated case when $\sigma(s)$ is a monotonic function. Since we always can rescale them keeping σs constant, let us assume $\sigma = s$, so

$$p(\sigma, s, \gamma') = p(s)\delta(\sigma - s)\delta(\gamma' - \gamma). \tag{5.99}$$

5.11.1 The Gamma distribution

Consider a Gamma distribution with fixed γ and $s = \sigma$, so $p(s)$ in equation (5.99) becomes

$$p(s) = \frac{\beta^\alpha s^{\alpha-1} \mathrm{e}^{-\beta s}}{\Gamma(\alpha)} \tag{5.100}$$

α and β being positive constants. First, let us calculate L, the root of equation (5.22). In our case we have

$$0 = \mathbb{E}[\sigma](\gamma L - 1) + \int_0^\infty \frac{\beta^\alpha s^\alpha \mathrm{e}^{-\beta s}}{\Gamma(\alpha)} \mathrm{e}^{-s\mathbb{E}[\sigma]L} \, \mathrm{d}s, \tag{5.101}$$

where $\mathbb{E}[\sigma] = \alpha/\beta$. This gives for L the equation

$$0 = \gamma L - 1 + \left(1 + \frac{\alpha L}{\beta^2}\right)^{-\alpha-1} \tag{5.102}$$

which can be easily solved numerically. The final epidemic size is given by equation 5.21, and may be written as

$$\Omega_\infty = \int_0^\infty p(s) \left(1 - \mathrm{e}^{-s\mathbb{E}[\sigma]L}\right) \mathrm{d}s = 1 - \left(1 + \frac{\alpha L}{\beta^2}\right)^{-\alpha} = \\ 1 - (\gamma L - 1)(1 + \alpha L/\beta^2) = \left(\frac{\alpha}{\beta^2} - \gamma\right) L + \frac{\alpha\gamma}{\beta^2} L^2 \tag{5.103}$$

In the case of the exponential distribution (i.e., $\alpha = 1$) equation (5.103) becomes

$$\Omega_\infty = \frac{(3 - 4\beta^2\gamma + \sqrt{1 + 4\beta^2\gamma})(1 - 2\beta^2\gamma + \sqrt{1 + 4\beta^2\gamma})}{4\beta^2\gamma}, \tag{5.104}$$

when $\mathcal{R}_0 > 1$. We emphasize that (5.103) and (5.104) are exact formulas.

In the low γ limit we may approximate L by $L = 1/\gamma - f(\gamma)$ (See equation (5.56) and Lemma 5.6.1), where the second term can be written as

$$f(\gamma) \approx \frac{\beta}{\alpha\gamma} \int_0^\infty \frac{\beta^\alpha s^\alpha \mathrm{e}^{-\beta s}}{\Gamma(\alpha)} \mathrm{e}^{-s\alpha/(\beta\gamma)} \, \mathrm{d}s \approx \frac{1}{\gamma} \left(1 + \frac{\alpha}{\gamma\beta^2}\right)^{-\alpha-1} \approx \left(\frac{\beta^2}{\alpha}\right)^{\alpha+1} \gamma^\alpha \tag{5.105}$$

Since $\alpha > 0$, $f(\gamma)$ is well defined near $\gamma = 0$ and the approximation is well-controlled.

5.11.2 Low-recovery-rate limit for the log-normal distribution

Let us discuss a log-normal distribution with fixed γ and $s = \sigma$, where $p(s)$ in equation (5.99) becomes:

$$p(s) = \frac{1}{\tau s\sqrt{2\pi}} \exp\left(-\frac{(\log s - \mu)^2}{2\tau^2}\right) \tag{5.106}$$

with the constants $\tau > 0$ and μ. Note that due to equation (5.99),

$$\mathbb{E}[s] = \mathbb{E}[\sigma] = \exp(\mu + \tau^2/2). \tag{5.107}$$

Using equations (5.56) and (5.107), we obtain the iterative equation for $\varepsilon \to 0$:

$$\begin{aligned} L_i &= \frac{1}{\gamma} - \frac{1}{\gamma\mathbb{E}[\sigma]} \int_0^\infty s\mathrm{e}^{-s\mathbb{E}[\sigma]L_{i-1}} \frac{1}{\tau s\sqrt{2\pi}} \mathrm{e}^{-\frac{(\log s-\mu)^2}{2\tau^2}} \, \mathrm{d}s \\ &= \frac{1}{\gamma} - \frac{\mathrm{e}^\mu}{\gamma\tau\mathbb{E}[\sigma]\sqrt{2\pi}} \int_0^\infty \mathrm{d}s \mathrm{e}^{-s\mathrm{e}^\mu\mathbb{E}[\sigma]L_{i-1} - \frac{(\log s)^2}{2\tau^2}} \\ &= \frac{1}{\gamma} - \frac{\mathrm{e}^\mu}{\gamma\tau\mathbb{E}[\sigma]\sqrt{2\pi}} J_\tau(\mathrm{e}^\mu\mathbb{E}[\sigma]L_{i-1}) \end{aligned} \tag{5.108}$$

where we have defined

$$J_\tau(a) \equiv \int_{-\infty}^{\infty} \mathrm{e}^{-a\mathrm{e}^y + y - \frac{y^2}{2\tau^2}}\,\mathrm{d}y \tag{5.109}$$

In principle, these equations are enough to construct an iterative solution for L. However, we may take this a step further for the low γ (large L) limit. In particular, if L is large, then so is each L_i. For $a \equiv \mathrm{e}^{\mu}\mathbb{E}[s]L \gg 1$, Eqn. (5.109) can be evaluated by a standard saddle point approximation[146, 12]. Setting $y' = y\sqrt{a} + \tau^2$ and expanding around $y_{\text{s.p.}}\sqrt{a} = \tau^2 + W(a\tau^2\mathrm{e}^{\tau^2})$ gives

$$J_\tau(a) \simeq \frac{\mathrm{e}^{\tau^2/2}\tau\sqrt{2\pi}}{\sqrt{W(a\tau^2\mathrm{e}^{\tau^2})+1}}\mathrm{e}^{-\frac{1}{2\tau^2}[2W(a\tau^2\mathrm{e}^{\tau^2})+W(a\tau^2\mathrm{e}^{\tau^2})^2]} \tag{5.110}$$

where $W(a\tau^2\mathrm{e}^{\tau^2})$ is the principal branch of the Lambert W-function, satisfying $W(\rho)\exp W(\rho) = \rho$. This expression is valid up to a small correction of order $O(\tau^2/W) \sim O(\tau^2/\log(a)) \ll 1$.

Returning to our iterative solution for L in Eqn. (5.109), we will now plug in the previous expression. Note that $\mathbb{E}[\sigma] = \mathrm{e}^{\mu+\tau^2/2}$

$$L_i = \frac{1}{\gamma} - \frac{1}{\gamma}\mathrm{e}^{-\frac{1}{\tau^2}[W(\tau^2\mathrm{e}^{2\mu-\tau^2/2}L_{i-1})+W(\tau^2\mathrm{e}^{2\mu-\tau^2/2}L_{i-1})^2]} \tag{5.111}$$

In particular,

$$L_1 = \frac{1}{\gamma} - \frac{1}{\gamma}\mathrm{e}^{-\frac{1}{\tau^2}[W(\tau^2\mathrm{e}^{2\mu-\tau^2/2}/\gamma)+W(\tau^2\mathrm{e}^{2\mu-\tau^2/2}/\gamma)^2]} \tag{5.112}$$

One may continue this iteration procedure to arbitrary precision.

5.12 SafeGraph Data

In this Section we describe the approach by Looi et al. [140] to transform the set of location pings into a dynamic network. In this network users are represented as nodes, and an edge (u, v, t) indicates that user u crossed paths with user v at time t. A path crossing is defined to occur when

two users have pings which are separated by less than 50 meters and less than 5 minutes. It should be stressed that a path here is the same as a world line in relativity theory: it encompasses spatial and temporal dimensions, so the users cross paths if they are at the same place at the same time.

To ensure that users are represented accurately, various filters are applied; for example, excluding users with fewer than 500 pings or removing duplicate users, which could potentially occur if a single person carries multiple mobile devices. To compute the path crossings efficiently, the authors apply a sliding time window, and, within each time slice, use a k-d tree to identify all pairs of points within 50 meters of each other. We refer the reader to the original paper for details of the network-construction methodology. The constructed network captures 1 613 884 111 path crossings between 9 451 697 users across three evenly spaced months in 2017 (March, July, and November). The network provides an estimate of the true contact network, where each user's number of contacts represents how many people they could possibly transmit the virus to or from. Thus, we can use each user's degree in the path crossing network to estimate their susceptibility and infectivity.

Previous analyses of SafeGraph data have shown that it is representative of the US population, in that it does not systematically over-represent users from certain income levels, racial demographics, degrees of educational attainment, or geographic regions [175]. Recently, their mobility patterns have been instrumental in helping researchers study responses to the COVID-19 pandemic and to model the role of mobility in the spread of disease [53, 84, 31, 198]. Even so, there are caveats to the data that we use. Most notably, the path-crossing network covers three months in 2017, but individuals' mobility patterns may have changed substantially following the onset of the pandemic. Furthermore, different types of noise may affect an individual's number of observed crossings; for example, the frequency with which their phone pings. Filtering for only well-represented users can help to mitigate this issue.

Appendix A: Approximating Gamma and PolyGamma functions

For $n \geq 1$ the polygamma functions have a series representation

$$\psi_n(t) = (-1)^{n+1} n! \sum_{k=0}^{\infty} \frac{1}{(t+k)^{n+1}}, \tag{A.1}$$

Lemma A.0.1. *For all $m \geq 1$, $z \in \mathbb{C} \setminus [0, -\infty)$,*

$$\psi_m(z) = (-1)^{m+1} m! \left[\frac{1}{m z^m} + \frac{1}{2 z^{m+1}} + \frac{m+1}{6 z^{m+2}} + \int_0^{\infty} \frac{(m+1)(m+2)(m+3)}{(x+z)^{m+4}} \frac{P_3(x)}{6} dx \right],$$

where $P_3(x)$ is the third order Bernoulli polynomial with period 1, and

$$\left| \int_0^{\infty} \frac{(m+1)(m+2)(m+3)}{(x+z)^{m+4}} \frac{P_3(x)}{6} dx \right| \leq \left| \frac{(m+1)(m+2)}{120} \frac{1}{z^{m+3}} \right|.$$

Proof. The first statement is proved by applying the Euler-Maclaurin formula to the series expansion (A.1) of $\psi_m(z)$. The inequality follows from the fact that $\sup_x |P_3(x)| \leq 1/20$. □

Lemma A.0.2. *For $|z| < 1$, $m \geq 0$,*

$$\psi_m(z) = (-1)^{m+1} m! z^{-(m+1)} + \sum_{k=0}^{\infty} (-1)^{k+m+1} \zeta(k+m+1)(k+1)_m z^k.$$

We also have

$$\psi_m(z) = (-1)^{m+1} m! z^{-(m+1)} + \sum_{k=0}^{n} (-1)^{k+m+1} \zeta(k+m+1)(k+1)_m z^k + R_m^n(z),$$

where

$$|\mathfrak{Re}[R_m^n(z)]|, |\mathfrak{Im}[R_m^n(z)]| \leq \frac{(n+m+1)!}{(n+1)!} \zeta(n+m+2) |z|^{n+1}.$$

Proof. The first equation is the Laurent expansion of $\psi_m(z)$ around 0. The bound on the remainder comes from Taylor's theorem. □

Lemma A.0.3. *for arg(z) strictly inside* $(-\pi, \pi)$*, as* $|z| \to \infty$*,*

$$\log \Gamma(z) = \left(z - \frac{1}{2}\right) \log(z) - z + \frac{1}{2} \log(2\pi) + O\left(\frac{1}{z}\right).$$

and

$$\psi(z) = \log(z) - \frac{1}{2z} + O\left(\frac{1}{z^2}\right).$$

These are special cases of [1, equations 6.1.42 and 6.4.11]

Lemma A.0.4. *for each* $\theta > 0$*, there exist constants C and D, such that for all y,*

$$e^{-\frac{\pi}{2}|y| + C + (\theta - \frac{1}{2}) \log(|y|)} \leq |\Gamma(\theta + \mathbf{i}y)|,$$

and for each $\varepsilon, \theta > 0$*, there exists M such that for all* $y > M$*,*

$$e^{(-\frac{\pi}{2} - \varepsilon)|y|} \leq |\Gamma(\theta + \mathbf{i}y)| \leq e^{(-\frac{\pi}{2} + \varepsilon)|y|}.$$

Proof. The first statement follows from applying the Euler-Maclaurin formula to the series expansion of $\log(\Gamma(z))$ and simplifying. The second statement follows from the first order Stirling approximation of $\Gamma(z)$. □

Lemma A.0.5. *For any* $\varepsilon > 0$*, there exists an* $M > 0$ *such that if* $y > M$*,* $t \geq \frac{1}{2} + \varepsilon$*, then* $\Re\mathfrak{e}[\psi_1(t + \mathbf{i}y)] > 0$

Proof. By Lemma A.0.1, we have

$$\Re\mathfrak{e}[\psi_1(t + \mathbf{i}y)] \geq \Re\mathfrak{e}\left[\frac{1}{t + \mathbf{i}y} + \frac{1}{2(t + \mathbf{i}y)^2} + \frac{1}{3(t + \mathbf{i}y)^3}\right] - \left|\frac{1}{20(t + \mathbf{i}y)^4}\right|. \tag{A.2}$$

We expand the first two summands of (A.2)

$$\Re\mathfrak{e}\left[\frac{1}{t+\mathbf{i}y}+\frac{1}{2(t+\mathbf{i}y)^2}\right]=\frac{t}{t^2+y^2}+\frac{t^2-y^2}{2(t^2+y^2)^2}\geq\frac{t-\frac{1}{2}}{t^2+y^2}>\frac{\varepsilon}{t^2+y^2}.$$

The third and fourth summands of (A.2) are bounded above by $\frac{1}{3(t^2+y^2)^{3/2}}$ and $\frac{1}{20((t^2+y^2)^2}$ respectively, so we can choose an M large enough that $\Re\mathfrak{e}[\psi_1(t+\mathbf{i}y)]>0$. □

Lemma A.0.6. *There exists $M\in\mathbb{R}$ such that for any $t\in[0,1]$, $|y|>M$, $\Re\mathfrak{e}[\psi(t+\mathbf{i}y)]>0$*

Proof. Lemma A.0.3 implies that as $y\to\infty$ are

$$\psi(t+\mathbf{i}y)\sim 1/2\log((t-1)^2+y^2)+\mathbf{i}\arctan\left(\frac{y}{t-1}\right)+\frac{1}{2(t+\mathbf{i}y)}.$$

Thus as $|y|\to\infty$, $\Re\mathfrak{e}[\psi(t+\mathbf{i}y)]\to+\infty$. □

Lemma A.0.7. *For all $\theta\in\mathbb{R}$ and $|y|\geq 1$, we have*

$$\frac{2\pi}{e^{\pi|y|}+1}\leq\frac{\pi}{|\sin(\pi(\theta+\mathbf{i}y))|}\leq\frac{2\pi}{e^{\pi|y|}-1}.$$

For all $\theta,y\in\mathbb{R}$, we have

$$\frac{\pi}{|\sin(\pi(\theta+\mathbf{i}y))}\leq\frac{\pi}{|\sin(\pi\theta)|}.$$

Proof. The inequalities are straightforward to prove using $\sin(z)=\frac{e^{\mathbf{i}z}-e^{-\mathbf{i}z}}{2\mathbf{i}}$. □

Appendix B: Bounds for dominated convergence of sticky Brownian motion

In this appendix we will complete the proofs of Lemma 3.2.10 and Lemma 3.2.11, 3.2.12, and 3.2.13.

Proof of Lemma 3.2.10. We first prove (3.27). For $z \in \mathcal{D}_\varepsilon(\phi_\varepsilon)$, and $v', v \in C \setminus C^\varepsilon$, the expression $\left|\frac{1}{z-v'}\right|$ is bounded, and

$$\left|\frac{\pi}{\sin(\pi(z-v))}\frac{1}{\Gamma(z)}\right| \leq \frac{2\pi e^{\frac{\pi}{2}|\Im\mathfrak{m}[z]|-C-Re[z-1/2]\log[\Im\mathfrak{m}[z]]}}{e^{\pi|\Im\mathfrak{m}[z-v]|-1}}. \tag{B.1}$$

by Lemma A.0.4 and Lemma A.0.7. Because $\theta < 1$, for small enough ε, $1/2\varepsilon \leq Re[z-v] \leq 1-\delta$, so that $|\sin(\pi(z-v))|$ is bounded below by a constant c by Lemma A.0.7, and $\frac{1}{|\Gamma(z)|}$ is bounded above on $\mathcal{D}_{\varepsilon,t}(\phi_\varepsilon)$ by Lemma A.0.4. Thus

$$\left|\frac{\pi}{\sin(\pi(z-v))}\frac{1}{\Gamma(z)}\right| \leq C. \tag{B.2}$$

for some constant C.

The function $\Gamma(v)$ has a pole at 0, and $h(v)$ has a pole of order 2 at 0. For small enough δ and $t > 1$, when $v \in C \cap B_\delta(0)$. We know $\Gamma(z)$ is well approximated by $\frac{1}{z}$ near 0 and $h(\theta) - h(v)$ is well approximated by $\frac{1}{z^2}$ near 0. For any constant $\eta > 0$, we can choose an ε, such that for all $y \in (-\varepsilon, \varepsilon)$,

$$\left|\frac{1}{\mathbf{i}y}e^{\frac{1}{(\mathbf{i}y)^2}}\right| \leq \left|\frac{1}{\varepsilon}e^{-\frac{1}{\varepsilon^2}}\right| < \eta.$$

The contour C crosses 0 along the imaginary axis, so we can use the above bound with η as small as desired to control $e^{t\frac{(h(\theta)-h(v))}{2}}$, and for any $v \in C \setminus B_\delta(0)$, $\Gamma(v)$ is holomorphic and thus bounded, so

$$\left|\Gamma(v)e^{t\frac{(h(\theta)-h(v))}{2}}\right| \leq C', \tag{B.3}$$

for some constant C'.

For all z, we have

$$|e^{t(h(z)-h(\theta))}| \leq e^{\frac{h'''(\theta)}{4}\varepsilon^3} \leq C'''. \tag{B.4}$$

For all $v \in C \setminus C^{\varepsilon}$,

$$|e^{t\frac{(h(\theta)-h(v))}{4}}| \leq e^{-t\eta/4},$$

by Proposition 3.2.9. Thus there exists $T > 0$ such that for all $t > T$,

$$|e^{t\frac{(h(\theta)-h(v))}{4}}|e^{t^{-1/3}\sigma(\theta)y(z-v)}| \leq |e^{-t\eta/4}e^{-t^{1/3}\sigma(\theta)y}| < 1. \tag{B.5}$$

The last inequality comes from choosing t sufficiently large.

Altogether (B.1), (B.2), (B.3), (B.4), (B.5) imply that for all $z \in \mathcal{D}_{\varepsilon}(\phi_{\varepsilon})$, $v \in C^{\varepsilon}$,

$$\begin{aligned}\left|\frac{\pi}{\sin(\pi(z-v))}\frac{\Gamma(v)}{\Gamma(z)}\frac{e^{t(h(z)-h(v))-t^{1/3}\sigma(\theta)yRe[z-v]}}{z-v'}\right| \\ \leq \frac{2C'''}{\varepsilon}e^{-t\eta/4}\max[C',C'']\min\left[C, \frac{2\pi e^{\frac{\pi}{2}|\Im\mathfrak{m}[z]|-C-Re[z-1/2]\log(\Im\mathfrak{m}[z])}}{e^{\pi|\Im\mathfrak{m}[z-v]|-1}}\right].\end{aligned} \tag{B.6}$$

The left hand side of (B.6) is the integrand of $K_{u_t}(v,v')$, so we can set $G(z,v,v')$ equal to the right hand side of (B.6). Observe that $\min\left[C, \frac{2\pi e^{\frac{\pi}{2}|\Im\mathfrak{m}[z]|-C-Re[z-1/2]\log[\Im\mathfrak{m}[z]]}}{e^{\pi|\Im\mathfrak{m}[z-v]|-1}}\right]$ is bounded above by a constant and has exponential decay in $\Im\mathfrak{m}[z]$ for $\Im\mathfrak{m}[z] \to +\infty$, thus we can set

$$R_1 = \int_{\mathcal{D}_{\varepsilon,t}(\phi_{\varepsilon})} \min\left[C, \frac{2\pi e^{\frac{\pi}{2}|\Im\mathfrak{m}[z]|-C-Re[z-1/2]\log[\Im\mathfrak{m}[z]]}}{e^{\pi|\Im\mathfrak{m}[z-v]|-1}}\right]dz < \infty.$$

Then

$$|K_{u_t}(v,v')| \leq \int_{\mathcal{D}_{\varepsilon,t}(\phi_{\varepsilon})} G(z,v,v')dz \leq R_1\frac{2C'''}{\varepsilon}e^{-t\eta/4}\max[C',C''] \leq R_2e^{-t\eta/4},$$

where $R_2 = R_1 \frac{2C'''}{\varepsilon} \max[C', C'']$.

Note that (3.28) follows from (3.27), because $K_{u_t}(v, v')$ depends on v' only through the factor $\frac{1}{z-v'}$ in the integrand. Thus we can apply the same argument where $\frac{1}{z-v'}$ is multiplied by $t^{-1/3}$ to get

$$|K_{u_t}(v, v')| \leq R_2 t^{-1/3} e^{-t\eta/4}. \tag{B.7}$$

Now (3.28) follows from (B.7) and the definition of ωK. □

Proof of Lemma 3.2.11. For $z \in \mathcal{D}_\varepsilon^\varepsilon(\phi_\varepsilon)$ and $v \in C^\varepsilon$ the function $\left|\frac{1}{\omega z - \omega v'}\right|$ is bounded and

$$\left| \frac{t^{-1/3}\pi}{\sin(\pi t^{-1/3}(\omega z - \omega v))} \frac{\Gamma(\theta + t^{-1/3}\omega v)}{\Gamma(\theta + t^{-1/3}\omega z)} \right| \leq c \frac{t^{-1/3}}{t^{-1/3}\varepsilon} \leq \frac{c}{\varepsilon}, \tag{B.8}$$

The second inequality is true because Γ is holomorphic in a neighborhood of θ, and $\sin(\theta + \mathbf{i}y) \geq \sin(\theta)$ for all $y \neq 0$. Set $r = \max_{\omega z \in B_{3\varepsilon}(0)} Re[\omega z]^3$. we then have

$$|e^{t[h(z)-h(v)]}| \leq e^{(\frac{h'''(\theta)}{6}+\eta)r + t(\frac{h'''(\theta)}{6}+\eta)(Re[\omega z^3]-r) + t(-\frac{h'''(\theta)}{6}+\eta)\omega v^3} \leq e^{(\frac{h'''(\theta)}{4})r + t(\frac{h'''(\theta)}{12})(Re[\omega z^3]-r) - t(\frac{h'''(\theta)}{12})\omega v^3}, \tag{B.9}$$

$$|e^{-\sigma(\theta)y(\omega z - \omega v)}| \leq e^{-\sigma(\theta)y(\varepsilon - \omega v)} \tag{B.10}$$

The first inequality follows from Taylor expanding $h(z)$ around θ, setting $\eta = \frac{h'''(\theta)}{12}$. The second inequality is true because $\Re\mathfrak{e}[z] = \varepsilon$.

Inequalities (B.9) and (B.10) together yield

$$\begin{aligned} &|\exp\left(t[h(z) - h(v)] - \sigma(\theta)y(\omega z - \omega v)\right)| \\ &\qquad \leq \exp\left(\frac{h'''(\theta)}{4} r + t\frac{h'''(\theta)}{12}(Re[\omega z^3] - r) - t\frac{h'''(\theta)}{12}\omega v^3 - \sigma(\theta)y(\varepsilon - \omega v) \right) \end{aligned} \tag{B.11}$$

and

$$\text{(B.11)} \le \exp\left(\frac{h'''(\theta)}{4}r + t\frac{h'''(\theta)}{12}(Re[\omega z^3] - r) - t\frac{h'''(\theta)}{24}\omega v^3 - \sigma(\theta)y\left(\varepsilon - \frac{\sqrt{24\sigma(\theta)y}}{h'''(\theta)}\right)\right)$$
$$\le c' \exp\left(t\frac{h'''(\theta)}{12}(Re[\omega z^3] - r) - t\frac{h'''(\theta)}{24}\omega v^3\right).$$

The last inequality is true because $\Re\mathfrak{e}[t(h'''(\theta)/24) + \omega v^4\sigma(\theta)y]$ achieves its maximum at $\Re\mathfrak{e}[\omega v] = \sqrt{24\sigma(\theta)y}h'''(\theta)$.

Let

$$\omega f(\omega v, \omega v', \omega z) = \frac{C}{\varepsilon^2} e^{t\frac{h'''(\theta)}{12}(Re[\omega z^3]-r)} e^{-t\frac{h'''(\theta)}{24}},$$

where $C = cc'$. Altogether (B.8), and (B.11) yield

$$\left|\frac{1}{\omega z - \omega v'}\frac{t^{-1/3}\pi}{\sin(\pi t^{-1/3}(\omega z - \omega v))}\frac{\Gamma(\theta + t^{-1/3}\omega v)}{\Gamma(\theta + t^{-1/3}\omega z)} e^{t[h(z)-h(v)]-\sigma(\theta)y(\omega z - \omega v)}\right| \le \omega f(\omega v, \omega v', \omega z).$$

where the left hand side is the integrand of $\omega K^{\varepsilon}_{u_t}(\omega v, \omega v')$, so the integrand is bounded above by $\omega f(\omega v, \omega v', \omega z)$. Note that ωf is decreasing in t, so setting $t = 1$ gives that the integrand of $\omega K_{u_t}(\omega v, \omega v')$ is less than or equal to $\frac{C}{2\varepsilon}\exp\left(\frac{h'''(\theta)}{12}(Re[\omega z^3] - r) - \frac{h'''(\theta)}{24}\omega v^3\right)$. This function is independent of t and has exponential decay in $\cos(3\phi_\varepsilon)|z|^3$ so integrating it over $\mathcal{D}^{\varepsilon}_{\varepsilon,t}(\phi_\varepsilon)$ gives a finite result, so we have proven the first claim.

Set $\ell = \int_{\varepsilon}^{e^{i(\pi-\phi)}\infty} e^{\frac{h'''(\theta)}{12}(Re[\omega z^3]-r)} d\omega z < \infty$, then

$$\omega K^{\varepsilon}_{u_t}(\omega v, \omega v') \le \int_{\mathcal{D}^{\varepsilon}_{\varepsilon,t}(\phi_\varepsilon)} \omega f(\omega v, \omega v', \omega z) d\omega z \le \frac{\ell C}{\varepsilon^2} e^{-\frac{h'''(\theta)}{24}\omega v^3} = C_1 e^{-\frac{h'''(\theta)}{24}\omega v^3},$$

where $C_1 = \frac{\ell C}{2\varepsilon^2}$. This completes the proof. □

Proof of Lemma 3.2.12. For $v, v' \in C^{\varepsilon}$, $z \in \mathcal{D}_{\varepsilon,t}(\phi_\varepsilon) \setminus \mathcal{D}^{\varepsilon}_{\varepsilon,t}(\phi_\varepsilon)$, the function $\left|\frac{1}{z-v}\right|$ is bounded and

$$\left|\frac{\pi}{\sin(\pi(z-v))}\frac{\Gamma(v)}{\Gamma(z)}\right| \le \frac{ce^{\frac{\pi}{2}|\Im\mathfrak{m}[z]| - C - (\theta - \frac{1}{2})\log(|\Im\mathfrak{m}[z]|)}}{e^{\pi|\Im\mathfrak{m}[(z-v)]|} - 1} \tag{B.12}$$

This inequality follows from Lemma A.0.4. As long as $\phi_\varepsilon < \frac{\pi}{6}$, $|\Im\mathfrak{m}[z-v]| > \delta$ for some δ, so the right hand side of (B.12) is bounded for $\Im\mathfrak{m}[z] \in \mathbb{R}$, and when $\Im\mathfrak{m}[z]$ is large it has exponential decay of order $e^{-\pi/2\Im\mathfrak{m}[z]}$. Also

$$|e^{-t^{1/3}\sigma(\theta)y(z-v)}| = |e^{-t^{1/3}\sigma(\theta)y(\Re\mathfrak{e}[z-v])}| \leq |e^{-t^{1/3}\sigma(\theta)y(\varepsilon\sin(\phi_\varepsilon)-\omega v)}|, \tag{B.13}$$

and by Lemma 3.2.8 there exists $\eta > 0$, such that

$$|e^{t[h(z)-h(v)]}| = |e^{h(z)-h(\theta)}||e^{h(\theta)-h(v)}| \leq |e^{-t\eta}e^{-\frac{h'''(\theta)}{12}\omega v^3}|. \tag{B.14}$$

The last inequality follows from Taylor expanding the v variable term, and applying Lemma 3.2.8 to the z variable term. There exists a constant $T > 0$ such that for all $t > T$, $t\eta/2 \geq t^{1/3}\sigma(\theta)y\varepsilon\sin(\phi_\varepsilon)$. This inequality together with (B.13) and (B.14) implies that for all $t > T$,

$$\begin{aligned}|e^{t[h(z)-h(v)-t^{1/3}\sigma(\theta)y(z-v)}| &\leq |e^{-t\eta-\frac{h'''(\theta)}{12}\omega v^3-t^{1/3}\sigma(\theta)y(\varepsilon\sin(\phi_\varepsilon)-\omega v)}| \leq |e^{-t\eta/2-\frac{h'''(\theta)}{24}\omega v^3+(-\frac{h'''(\theta)}{24}\omega v^3-t^{1/3}\sigma(\theta)y\omega v)} \\ &\leq e^{-t\eta/2-t\frac{h'''(\theta)}{24}\omega v^3-\sigma(\theta)y(\varepsilon-\frac{\sqrt{24\sigma(\theta)y}}{h'''(\theta)})} \leq c'e^{-t\eta/2-t\frac{h'''(\theta)}{24}\omega v^3}.\end{aligned} \tag{B.15}$$

The first inequality follows from our choice of T, and the second inequality follows from the fact that $\Re\mathfrak{e}[t(h'''(\theta)/24) - \omega v^3\sigma(\theta)y(-\omega v)]$ achieves its maximum at $\Re\mathfrak{e}[v] = \sqrt{24\sigma(\theta)y}h'''(\theta)$.

Set

$$g(z,\omega v,\omega v') = \frac{10}{\varepsilon}c'e^{-t\eta/2-t\frac{h'''(\theta)}{24}\omega v^3}\left(\frac{ce^{\frac{\pi}{2}|\Im\mathfrak{m}[z]|-C-(\theta-\frac{1}{2})\log(|\Im\mathfrak{m}[z]|)}}{e^{\pi|\Im\mathfrak{m}[z-v]|}}\right).$$

Together (B.12) and (B.11) imply

$$\left|\frac{1}{z-v}e^{t[h(z)-h(v)-t^{1/3}\sigma(\theta)y(z-v)}\frac{\pi}{\sin(\pi(z-v))}\frac{\Gamma(v)}{\Gamma(z)}\right| \leq g(z,\omega v,\omega v'). \tag{B.16}$$

The left hand side of (B.16) is the integrand of $K_{u_t}(v,v')$. The expression $\left(\frac{ce^{\frac{\pi}{2}|\Im\mathfrak{m}[z]|-C-(\theta-\frac{1}{2})\log(|\Im\mathfrak{m}[z]|)}}{e^{\pi|\Im\mathfrak{m}[(z-v)]|}}\right)$

is bounded as $\Im\mathfrak{m}[z]$ varies in $(-\infty, +\infty)$, and has exponential decay in $\Im\mathfrak{m}[z]$ for large $\Im\mathfrak{m}[z]$, so we can set

$$S = \int_{\mathcal{D}_{\varepsilon,t}(\phi_\varepsilon)\setminus\mathcal{D}^{\varepsilon}_{\varepsilon,t}(\phi_\varepsilon)} \left(\frac{c e^{\frac{\pi}{2}|\Im\mathfrak{m}[z]| - C - (\theta-\frac{1}{2})\log(|\Im\mathfrak{m}[z]|)}}{e^{\pi|\Im\mathfrak{m}[z-v]|}} \right) dz < \infty.$$

Then

$$\begin{aligned} &\left|K_t(\theta+\omega v, \theta+\omega v') - K^{\varepsilon}_t(\theta+\omega v, \theta+\omega v')\right| \\ &\qquad \leq \int_{\mathcal{D}_{\varepsilon,t}(\phi_\varepsilon)\setminus\mathcal{D}^{\varepsilon}_{\varepsilon,t}(\phi_\varepsilon)} g(z, \omega v, \omega v') dz \leq \frac{10Sc'}{\varepsilon} e^{-t\eta/2} e^{t(-h'''(\theta)/24)\omega v^3} \xrightarrow{t\to\infty} 0. \quad \text{(B.17)} \end{aligned}$$

□

Proof of Lemma 3.2.13. We have the following inequalities,

$$\omega K^{\varepsilon}_{u_t}(\omega v_i, \omega v_j) \leq C_1 e^{-t\frac{h'''(\theta)}{24}\omega v_i^3}, \tag{B.18}$$

$$\begin{aligned} \omega K_{u_t}(\omega v_i, \omega v_j) &= \omega K^{\varepsilon}_{u_t}(\omega v_i, \omega v_j) \quad \text{(B.19)} \\ &\quad + \left(t^{-1/3} K_{u_t}(\theta + t^{-1/3}\omega v_i, \theta + t^{-1/3}\omega v_j) - t^{-1/3} K^{\varepsilon}_{u_t}(\theta + t^{-1/3}\omega v_i, \theta + t^{-1/3}\omega v_j) \right) \\ &\leq C_2 e^{-t\eta/2} e^{t(-h'''(\theta)/24)\omega v^3} + C_1 e^{-t\frac{h'''(\theta)}{24}\omega v_i^3} \leq C_3 e^{t(-h'''(\theta)/24)\omega v^3}, \quad \text{(B.20)} \end{aligned}$$

where $C_3 = C_1 + C_2 e^{-\eta/2}$. Inequality (B.18) follows from Lemma 3.2.11. The first inequality of (B.20) comes from Lemma 3.2.11, Lemma 3.2.12 and the fact that $t > 1$. Hadamard's bound implies

$$|\det(\omega K^{\varepsilon}_{u_t}(\omega v_i, \omega v_j))_{i,j=1}^m| \leq m^{m/2} C_3^{m/2} \prod_{i=1}^m e^{-t\frac{h'''(\theta)}{24}\omega v_i^3}.$$

Set $\omega H_m(\omega v, \omega v') = m^{m/2} C_3^{m/2} \prod_{i=1}^m e^{-t\frac{h'''(\theta)}{24}\Re\mathfrak{e}[\omega v_i]^3}$, and set $L = \int_0^\infty e^{-\frac{h'''(\theta)}{24}x^3} dx < \infty$. Then

$$\int_{(\omega C^{\varepsilon})^m} \omega H_m(\omega v, \omega v') \leq \int_{(C_0)^m} \omega H_m(\omega v, \omega v') \leq m^{m/2} C_3^{m/2} L^m.$$

Thus

$$1 + \sum_{m=1}^{\infty} \frac{1}{m!} \int_{(C_0)^m} \omega H_m(\omega v, \omega v') \leq 1 + \sum_{m=1}^{\infty} \frac{m^{m/2} C_3^{m/2} L^m}{m!}.$$

So because $m! \geq \sqrt{\frac{2\pi}{m}} \left(\frac{m}{e}\right)^m$, we have

$$1 + \sum_{m=1}^{\infty} \frac{1}{m!} \int_{(\omega C^{\varepsilon})^m} \omega H_m(\omega v, \omega v') \leq 1 + \sum_{m=1}^{\infty} \frac{1}{m!} \int_{(C_0)^m} \omega H_m(\omega v, \omega v') < \infty.$$

□

References

[1] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Courier Corporation, 1965, vol. 55.

[2] D. Adam, P. Wu, J. Wong, E. Lau, T. Tsang, S. Cauchemez, G. Leung, and B. Cowling, *Clustering and superspreading potential of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections in Hong Kong*, Preprint (Version 1) available at Research Square, 2020.

[3] A. Aggarwal, "Nonexistence and uniqueness of pure states of ferroelectric six-vertex models", *arXiv preprint arXiv:2004.13272v2*, pp. 1–35, 2020.

[4] A. Aggarwal, "Current fluctuations of the stationary ASEP and six-vertex model", *Duke Math. J., arXiv:1608.04726*, 2016.

[5] ——, "Dynamical stochastic higher spin vertex models", *Selecta Math.*, pp. 1–77, 2017.

[6] A. Aggarwal and A. Borodin, "Phase transitions in the ASEP and stochastic six-vertex model", *Ann. Probab.*, 2016.

[7] T. Alberts, K. Khanin, and J. Quastel, "The intermediate disorder regime for directed polymers in dimension 1 + 1", *Ann. Probab.*, vol. 42, no. 3, pp. 1212–1256, 2014.

[8] A. Allard, C. Moore, S. V. Scarpino, B. M. Althouse, and L. Hébert-Dufresne, *The role of directionality, heterogeneity and correlations in epidemic risk and spread*, 2020. arXiv: 2005.11283 [physics.soc-ph].

[9] M. Amir, "Sticky Brownian motion as the strong limit of a sequence of random walks", *Stoch. Proc. Appl.*, vol. 39, no. 2, pp. 221–237, 1991.

[10] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*. Cambridge university press, 2010, vol. 118.

[11] R. A. Arratia, "Coalescing brownian motions on the line.", 1980.

[12] S. Asmussen, J. L. Jensen, and L. Rojas-Nandayapa, "On the laplace transform of the lognormal distribution", *Methodol. Comput. Appl. Probab.*, vol. 18, no. 2, pp. 441–458, 2016.

[13] A. Auffinger, J. Baik, and I. Corwin, "Universality for directed polymers in thin rectangles", *arXiv preprint arXiv:1204.4445*, 2012.

[14] A. Auffinger, M. Damron, and J. Hanson, *50 Years of First-Passage Percolation*. American Mathematical Soc., 2017, vol. 68.

[15] J. Baik, G. Barraquand, I. Corwin, and T. Suidan, "Facilitated exclusion process", *The Abel Symposium*, 2016.

[16] J. Baik, P. Deift, and K. Johansson, "On the distribution of the length of the longest increasing subsequence of random permutations", *Amer. Math. Soc.*, vol. 12, no. 4, pp. 1119–1178, 1999.

[17] J. Baik and T. M. Suidan, "A GUE central limit theorem and universality of directed first and last passage site percolation", *Int. Math. Res. Not.*, vol. 2005, no. 6, pp. 325–337, 2005.

[18] M. Balázs, F. Rassoul-Agha, and T. Seppäläinen, "Large deviations and wandering exponent for random walk in a dynamic beta environment", *Ann. Probab.*, 2019.

[19] ——, "The random average process and random walk in a space-time random environment in one dimension", *Comm. Math. Phys.*, vol. 266, no. 2, pp. 499–545, 2006.

[20] Y. M. Bar-On, A. Flamholz, R. Phillips, and R. Milo, "SARS-CoV-2 (COVID-19) by the numbers", *eLife*, vol. 9, M. B. Eisen, Ed., e57309, 2020.

[21] G. Barraquand, "A phase transition for q-TASEP with a few slower particles", *Stochastic Process. Appl.*, vol. 125, no. 7, pp. 2674 –2699, 2015.

[22] G. Barraquand, A. Borodin, I. Corwin, and M. Wheeler, "Stochastic six-vertex model in a half-quadrant and half-line open asep", *Duke Math. J.*, 2018.

[23] G. Barraquand and I. Corwin, "Random-walk in Beta-distributed random environment", *Prob. Theory Rel. Fields*, vol. 167, no. 3-4, pp. 1057–1116, 2017.

[24] ——, "The q-Hahn asymmetric exclusion process", *Ann. Appl. Probab.*, vol. 26, no. 4, pp. 2304–2356, 2016.

[25] G. Barraquand and M. Rychnovsky, "Large deviations for sticky brownian motions", *Electron. J. Probab.*, vol. 25, 2020.

[26] ——, "Tracy-widom asymptotics for a river delta model", in *International workshop on Stochastic Dynamics out of Equilibrium*, Springer, 2017, pp. 483–522.

[27] ——, "Tracy-Widom asymptotics for a river delta model", in *International workshop on Stochastic Dynamics out of Equilibrium*, Springer, 2017, pp. 483–522.

[28] Y. Baryshnikov, "GUEs and queues", *Probab. Theory Rel. Fields*, vol. 119, no. 2, pp. 256–274, 2001.

[29] R. Bass, "A stochastic differential equation with a sticky point", *Electron. J. Probab.*, vol. 19, 2014.

[30] R. J. Baxter, *Exactly solved models in statistical mechanics*. Elsevier, 2016.

[31] S. G. Benzell, A. Collis, and C. Nicolaides, "Rationing social contact during the COVID-19 pandemic: Transmission risk and social benefits of US locations", *Proceedings of the National Academy of Sciences*, 2020.

[32] D. Bernard, K. Gawedzki, and A. Kupiainen, "Slow modes in passive advection", *J. Stat. Phys.*, vol. 90, no. 3-4, pp. 519–569, 1998.

[33] T. Bodineau and J. Martin, "A universality property for last-passage percolation paths close to the axis", *Electron. Commun. Probab.*, vol. 10, pp. 105–112, 2005.

[34] A. Borodin, "On a family of symmetric rational functions", *Adv. Math.*, vol. 306, pp. 973–1018, 2017.

[35] ——, "Schur dynamics of the schur processes", *arXiv preprint arXiv:1001.3442*, 2010.

[36] A. Borodin and I. Corwin, "Macdonald processes", *Probab. Theory and Related Fields*, vol. 158, no. 1-2, pp. 225–400, 2014.

[37] A. Borodin, I. Corwin, and P. Ferrari, "Free energy fluctuations for directed polymers in random media in 1+1 dimension", *Comm. Pure Appl. Math.*, vol. 67, no. 7, pp. 1129–1214, 2014.

[38] A. Borodin, I. Corwin, P. Ferrari, and B. Vető, "Height fluctuations for the stationary KPZ equation", *Math. Phys. Anal. Geom.*, vol. 18, no. 1, Art. 20, 95, 2015.

[39] A. Borodin, I. Corwin, and V. Gorin, "Stochastic six-vertex model", *Duke Math. J.*, vol. 165, no. 3, pp. 563–624, 2016.

[40] A. Borodin, I. Corwin, L. Petrov, and T. Sasamoto, "Spectral theory for interacting particle systems solvable by coordinate Bethe ansatz", *Comm. Math. Phys.*, vol. 339, no. 3, pp. 1167–1245, 2015.

[41] ——, "Spectral theory for the q-boson particle system", *Compositio Mathematica*, vol. 151, no. 1, pp. 1–67, 2015.

[42] A. Borodin, I. Corwin, and D. Remenik, "Log-gamma polymer free energy fluctuations via a fredholm determinant identity", *Comm. Math. Phys.*, vol. 324, no. 1, pp. 215–232, 2013.

[43] A. Borodin, I. Corwin, and T. Sasamoto, "From duality to determinants for q-TASEP and ASEP", *Ann. Probab.*, vol. 42, no. 6, pp. 2314–2382, 2014.

[44] A. Borodin and P. Ferrari, "Large time asymptotics of growth models on space-like paths I: PushASEP", *Electron. J. Probab.*, vol. 13, pp. 1380–1418, 2008.

[45] A. Borodin and G. Olshanski, "The ASEP and determinantal point processes", *Commun. Math, Phys.*, vol. 353, no. 2, pp. 853–903, 2017.

[46] A. Borodin and L. Petrov, "Higher spin six vertex model and symmetric rational functions", *Selecta Math.*, vol. 24, no. 2, pp. 751–874, 2018.

[47] ——, "Integrable probability: From representation theory to Macdonald processes", *Probab. Surv.*, vol. 11, pp. 1–58, 2014.

[48] D. Brockington and J. Warren, "The Bethe ansatz for sticky Brownian motions", *in preparation*,

[49] A. Bufetov and L. Petrov, "Yang-Baxter field for spin Hall-Littlewood symmetric functions", *Forum Math. Sigma*, vol. 7, 2019.

[50] J. Bukman and J. Shore, "The conical point in the ferroelectric six-vertex model", *J. Stat. Phys.*, vol. 78, pp. 1277–1309, 1995.

[51] J. Cardy, *Private communication*, 2020.

[52] G. Carinci, C. Giardina, and F. Redig, "Exact formulas for two interacting particles and applications in particle systems with duality", *Ann. Appl. Probab.*, 2017.

[53] S. Chang, E. Pierson, P. W. Koh, J. Gerardin, B. Redbird, D. Grusky, and J. Leskovec, "Mobility network models of covid-19 explain inequities and inform reopening", *Nature*, 2020.

[54] H. Chaumont and C. Noack, "Fluctuation exponents for stationary exactly solvable lattice polymer models via a Mellin transform framework", *arXiv preprint arXiv:1711.08432*, 2017.

[55] M Chertkov, G Falkovich, I Kolokolov, and V Lebedev, "Normal and anomalous scaling of the fourth-order correlation function of a randomly advected passive scalar", *Phys. Rev. E*, vol. 52, no. 5, p. 4924, 1995.

[56] P. Y. Chia, K. K. Coleman, Y. K. Tan, S. W. X. Ong, M. Gum, S. K. Lau, X. F. Lim, A. S. Lim, S. Sutjipto, P. H. Lee, T. T. Son, B. E. Young, D. K. Milton, G. C. Gray, S. Schuster, T. Barkham, P. P. De, S. Vasoo, M. Chan, B. S. P. Ang, B. H. Tan, Y.-S. Leo, O.-T. Ng, M. S. Y. Wong, and K. Marimuthu, "Detection of air and surface contamination by SARS-CoV-2 in hospital rooms of infected patients", *Nat. Commun.*, vol. 11, 2020.

[57] R. J. Chitashvili, "On the nonexistence of a strong solution in the boundary problem for a sticky Brownian motion", *Proc. A. Razmadze Math. Inst.*, vol. 112, pp. 17–31, 1997.

[58] S. Y. Cho, J.-M. Kang, Y. E. Ha, G. E. Park, J. Y. Lee, J.-H. Ko, J. Y. Lee, J. M. Kim, C.-I. Kang, I. J. Jo, J. G. Ryu, J. R. Choi, S. Kim, H. J. Huh, C.-S. Ki, E.-S. Kang, K. R. Peck, H.-J. Dhong, J.-H. Song, D. R. Chung, and Y.-J. Kim, "MERS-CoV outbreak following a single patient exposure in an emergency room in South Korea: An epidemiological outbreak study", *Lancet*, vol. 388, pp. 994–1001, 2016.

[59] F. Colomo and A. Sportiello, "Arctic curves of the six-vertex model on generic domains: The tangent method", *J. Stat. Phys.*, vol. 164, no. 6, pp. 1488–1523, 2016.

[60] I. Corwin, P. Ghosal, H. Shen, and L.-C. Tsai, "Stochastic pde limit of the six vertex model", *Comm. Math. Phys.*, pp. 1–35, 2020.

[61] I. Corwin, "Kardar-Parisi-Zhang universality", *Not. AMS*, 2016.

[62] ——, "Macdonald processes, quantum integrable systems and the Kardar-Parisi-Zhang universality class", *Proc. ICM 2014, arXiv:1403.6877*, 2014.

[63] ——, "The Kardar–Parisi–Zhang equation and universality class", *Rand. Mat.: Theor. Appl.*, vol. 1, no. 01, p. 1 130 001, 2012.

[64] ——, "The q-Hahn Boson process and q-Hahn TASEP", *Int. Math. Res. Not.*, vol. 2015, no. 14, pp. 5577–5603, 2014.

[65] I. Corwin and Y. Gu, "Kardar–Parisi–Zhang Equation and Large Deviations for Random Walks in Weak Random Environments", *J. Stat. Phys.*, vol. 166, no. 1, pp. 150–168, 2017.

[66] I. Corwin, N. O'Connell, T. Seppäläinen, and N. Zygouras, "Tropical combinatorics and whittaker functions", *Duke Math. J.*, vol. 163, no. 3, pp. 513–563, 2014.

[67] I. Corwin and L. Petrov, "Stochastic higher spin vertex models on the line", *Comm. Math. Phys.*, vol. 343, no. 2, pp. 651–700, 2016.

[68] I. Corwin, T. Seppäläinen, and H. Shen, "The strict-weak lattice polymer", *J. Stat. Phys.*, pp. 1–27, 2015.

[69] H. Dai and Y. Q. Zhao, "Multidimensional Sticky Brownian Motions: Tail Behaviour of the Joint Stationary Distribution", *arXiv preprint arXiv:1901.07529*, 2019.

[70] M. Defosseux, "Orbit measures, random matrix theory and interlaced determinantal processes", *Ann. Inst. Henri Poincaré Probab. Stat.*, vol. 46, pp. 209–249, 2010.

[71] E. Dimitrov, "Six-vertex models and the GUE-corners process", *arXiv preprint arXiv:1610.06893*, 2016.

[72] E. Dimitrov and M. Rychnovsky, "Gue corners process in boundary-weighed six-vertex models", *to appear in Ann. Henri Poincaré arXiv:2005.06836*, 2020.

[73] R. Durrett, *Probability: theory and examples*. Cambridge University Press, Cambridge, 2019, vol. 49.

[74] F. J. Dyson, "A Brownian-motion model for the eigenvalues of a random matrix", *J. Math. Phys.*, vol. 3, no. 6, pp. 1191–1198, 1962.

[75] A. Endo, S. Abbott, A. J. Kucharski, and S. Funk, "Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China", *Wellcome Open Research*, vol. 5, no. 67, 2020.

[76] H.-J. Engelbert and G. Peskir, "Stochastic differential equations for sticky Brownian motion", *Stochastics*, vol. 86, no. 6, pp. 993–1021, 2014.

[77] W. Feller, "The parabolic differential equations and the associated semi-groups of transformations", *Ann. Math.*, pp. 468–519, 1952.

[78] P. L. Ferrari, H. Spohn, and T. Weiss, "Scaling limit for Brownian motions with one-sided collisions", *Ann. App. Probab.*, vol. 25, no. 3, pp. 1349–1382, 2015.

[79] J. Fiegel, R. Clarke, and D. A. Edwards, "Airborne infectious disease and the suppression of pulmonary bioaerosols", *Drug Discovery Today*, vol. 11, no. 1, pp. 51–57, 2006.

[80] L. Fontes, M. Isopi, C. Newman, and K. Ravishankar, "The Brownian web", *Proc. Nat. Acad. Sci.*, vol. 99, no. 25, pp. 15 888–15 893, 2002.

[81] ——, "The Brownian web: Characterization and convergence", *Ann. Probab.*, vol. 32, no. 4, pp. 2857–2883, 2004.

[82] T. R. Frieden and C. T. Lee, "Identifying and interrupting superspreading events-implications for control of Severe Acute Respiratory Syndrome Coronavirus 2", *Emerging Infecious Diseases*, 2020.

[83] A. P. Galvani and R. M. May, "Dimensions of superspreading", *Nature*, vol. 438, pp. 293–295, 2005.

[84] S. Gao, J. Rao, Y. Kang, Y. Liang, and J. Kruse, "Mapping county-level mobility pattern changes in the united states in response to covid-19", *SIGSPATIAL Special*, vol. 12, no. 1, pp. 16–26, 2020.

[85] B. Garrod, R. Tribe, and O. Zaboronski, "Examples of interacting particle systems on z as pfaffian point processes: Coalescing–branching random walks and annihilating random walks with immigration", in *Ann. Inst. H. Poincaré*, Springer, vol. 21, 2020, pp. 885–908.

[86] M. Gaudin, *The Bethe Wavefunction*. Cambridge University Press, 2014.

[87] K. Gawedzki and P. Horvai, "Sticky behavior of fluid particles in the compressible Kraichnan model", *J. Stat. Phys.*, vol. 116, no. 5-6, pp. 1247–1300, 2004.

[88] K. Gawedzki and A. Kupiainen, "Anomalous scaling of the passive scalar", *Phys. Rev. Lett.*, vol. 75, no. 21, p. 3834, 1995.

[89] ——, "University in turbulence: An exactly solvable model", in *Low-dimensional models in statistical physics and quantum field theory*, Springer, 1996, pp. 71–105.

[90] K. Gawedzki and M. Vergassola, "Phase transition in the passive scalar advection", *Phys. D: Nonlinear Phenomena*, vol. 138, no. 1-2, pp. 63–90, 2000.

[91] P. Ghosal, "Hall-Littlewood-pushTASEP and its KPZ limit", *arXiv preprint arXiv:1701.07308*, 2017.

[92] ——, "Moments of the SHE under delta initial measure", *arXiv preprint arXiv:1808.04353*, 2018.

[93] M. G. M. Gomes, R. Aguas, R. M. Corder, J. G. King, K. E. Langwig, C. Souto-Maior, J. Carneiro, M. U. Ferreira, and C. Penha-Goncalves, "Individual variation in susceptibility or exposure to sars-cov-2 lowers the herd immunity threshold", *MedRxiv*, 2020.

[94] M. G. M. Gomes, R. M. Corder, J. G. King, K. E. Langwig, C. Souto-Maior, J. Carneiro, G. Goncalves, C. Penha-Goncalves, M. U. Ferreira, and R. Aguas, "Individual variation in susceptibility or exposure to SARS-CoV-2 lowers the herd immunity threshold", *medRxiv*, 2020.

[95] V. Gorin, "From alternating sign matrices to the gaussian unitary ensemble", *Comm. Math. Phys.*, vol. 332, no. 1, pp. 437–447, 2014.

[96] E. Granet, L. Budzynski, J. Dubail, and J. L. Jacobsen, "Inhomogeneous Gaussian free field inside the interacting arctic curve", *J. Stat. Mech.*, no. 1, :013102, 2019.

[97] J. Gravner, C. A. Tracy, and H. Widom, "Limit theorems for height fluctuations in a class of discrete space and time growth models", *J. Stat. Phys.*, vol. 102, no. 5-6, pp. 1085–1132, 2001.

[98] L.-H. Gwa and H. Spohn, "Six-vertex model", *Phys. Rev. Lett.*, vol. 68, pp. 725–728, 1992.

[99] J. Hammersley and D. Welsh, "First-passage percolation, subadditive processes, stochastic networks, and generalized renewal theory", in *Bernoulli 1713, Bayes 1763, Laplace 1813*, Springer, 1965, pp. 61–110.

[100] J. M. Harrison and A. J. Lemoine, "Sticky Brownian motion as the limit of storage processes", *J. Appl. Probab.*, vol. 18, no. 1, pp. 216–226, 1981.

[101] H. W. Hethcote, "The mathematics of infectious diseases", *SIAM Rev.*, vol. 42, no. 4, pp. 599–653, 2020.

[102] M.-S. Ho, "Severe acute respiratory syndrome (SARS)", in *Tropical Infectious Diseases: Principles, Pathogens and Practice*, R. L. Guerrant, D. H. Walker, and P. F. Weller, Eds., Third Edition, Edinburgh: W.B. Saunders, 2011, pp. 392–397, ISBN: 978-0-7020-3935-5.

[103] M. Holmes-Cerfon, "Sticky-sphere clusters", *Ann. Rev. Condens. Matter Phys.*, vol. 8, pp. 77–98, 2017.

[104] C. Howitt and J. Warren, "Consistent families of Brownian motions and stochastic flows of kernels", *Ann. Probab.*, vol. 37, no. 4, pp. 1237–1272, 2009.

[105] ——, "Dynamics for the Brownian web and the erosion flow", *Stochastic Process. Appl.*, vol. 119, no. 6, pp. 2028–2051, 2009.

[106] G. Huber, M. Kamb, K. Kawagoe, L. M. Li, A. McGeever, J. Miller, B. Veytsman, and D. Zigmond, "A minimal model for household-based testing and tracing in epidemics", *medRxiv*, 2020.

[107] G. Huber, M. Kamb, K. Kawagoe, L. M. Li, B. Veytsman, D. Yllanes, and D. Zigmond, "A minimal model for household effects in epidemics", *Physical Biology*, vol. 17, no. 6, p. 065 010, Oct. 2020.

[108] D. A. Huse and C. L. Henley, "Pinning and roughening of domain walls in ising systems due to random impurities", *Phys. rev. lett.*, vol. 54, no. 25, p. 2708, 1985.

[109] J. Z. Imbrie and T. Spencer, "Diffusion of directed polymers in a random environment", *J. stat. Phys.*, vol. 52, no. 3, pp. 609–626, 1988.

[110] K. Itô and H. P. McKean, "Brownian motions on a half line", *Illinois J. Math.*, vol. 7, no. 2, pp. 181–231, 1963.

[111] A. James, J. W. Pitchford, and M. J. Plank, "An event-based model of superspreading in epidemics", *Roy. Soc. Proc. Biol. Sci.*, vol. 274, no. 1610, pp. 741–747, 2007.

[112] K. Johansson, "Shape fluctuations and random matrices", *Comm. Math. Phys.*, vol. 209, no. 2, pp. 437–476, 2000.

[113] K. Johansson and E. Nordenstam, "Eigenvalues of gue minors", *Electron. J. Probab.*, vol. 11, pp. 1342–1371, 2006.

[114] I. Karatzas, A. Shiryaev, and M. Shkolnikov, "On the one-sided Tanaka equation with drift", *Electron. Commun. Probab.*, vol. 16, pp. 664–677, 2011.

[115] M. Kardar, G. Parisi, and Y. Zhang, "Dynamic Scaling of Growing Interfaces", *Phys. Rev. Lett.*, vol. 56, pp. 889–892, 9 1986.

[116] M. Kardar, G. Parisi, and Y.-C. Zhang, "Dynamic scaling of growing interfaces", *Phys. Rev. Lett.*, vol. 56, no. 9, p. 889, 1986.

[117] K. Kawagoe, M. Rychnovsky, S. Y. Chang, G. Huber, L. M. Li, J. Miller, R. Pnini, B. Veytsman, and D. Yllanes, "How much do superspreaders matter? epidemic dynamics in inhomogeneous populations", *medRxiv*, 2021.

[118] R. Kenyon, "Conformal invariance of domino tilings", *Ann. Probab.*, vol. 28, pp. 759–795, 2000.

[119] ——, "Dominos and the Gaussian Free Field", *Ann. Probab.*, vol. 29, pp. 1128–1137, 2001.

[120] R. Kenyon, A. Okounkov, and S. Sheffield, "Dimers and amoebae", *Ann. Math.*, vol. 163, pp. 1019–1056, 2006.

[121] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics", *Proc. R. Soc. London. Series A*, vol. 115, no. 772, pp. 700–721, 1927.

[122] K. H. Kim, T. E. Tandi, J. W. Choi, J. M. Moon, and M. S. Kim, "Middle East Respiratory Syndrome Coronavirus (MERS-CoV) outbreak in South Korea, 2015: Epidemiology, characteristics and public health implications", *J. Hosp. Infect.*, vol. 95, no. 2, pp. 207–213, 2017.

[123] V. E. Korepin, "Calculations of norms of Bethe wave functions", *Commun. Math. Phys.*, vol. 86, pp. 391–418, 3 1982.

[124] R. H. Kraichnan, "Small-scale structure of a scalar field convected by turbulence", *The Phys. of Fluids*, vol. 11, no. 5, pp. 945–953, 1968.

[125] A. Krishnan and J. Quastel, "Tracy-Widom fluctuations for perturbations of the log-gamma polymer in intermediate disorder", *Ann. Appl. Probab.*, 2016.

[126] J. Krug, P. Meakin, and T. Halpin-Healy, "Amplitude universality for driven interfaces and directed polymers in random media", *Phys. Rev. A*, vol. 45, no. 2, p. 638, 1992.

[127] K. Kupferschmidt, “Why do some COVID-19 patients infect many others, whereas most don’t spread the virus at all?”, *Science*, 2020.

[128] A. Kupiainen, “Lessons for turbulence”, in *Visions in Math.* Springer, 2010, pp. 316–333.

[129] M. Lachiany and Y. Louzoun, “Effects of distribution of infection rate on epidemic models”, *Phys. Rev. E*, vol. 94, no. 2, p. 022 409, 2016.

[130] ——, “Effects of distribution of infection rate on epidemic models”, *Phys. Rev. E*, vol. 94, p. 022 409, 2 Aug. 2016.

[131] P. Le Doussal and T. Thiery, “Diffusion in time-dependent random media and the Kardar-Parisi-Zhang equation”, *Phys. Rev. E*, vol. 96, no. 1, p. 010 102, 2017.

[132] Y. Le Jan and S. Lemaire, “Products of Beta matrices and sticky flows”, *Probab. Theory Rel. Fields*, vol. 130, no. 1, pp. 109–134, 2004.

[133] Y. Le Jan and O. Raimond, “Flows, coalescence and noise”, *Ann. Probab.*, vol. 32, no. 2, pp. 1247–1315, 2004.

[134] ——, “Integration of Brownian vector fields”, *Ann. Probab.*, vol. 30, no. 2, pp. 826–873, 2002.

[135] ——, “Sticky flows on the circle and their noises”, *Probab. Theor. Rel. Fields*, vol. 129, no. 1, pp. 63–82, 2004.

[136] ——, “The noise of a Brownian sticky flow is black”, *arXiv preprint math/0212269*, 2002.

[137] Y. Li, I. T. S. Yu, P. Xu, J. H. W. Lee, T. W. Wong, P. L. Ooi, and A. C. Sleigh, “Predicting super spreading events during the 2003 severe acute respiratory syndrome epidemics in Hong Kong and Singapore”, *American Journal of Epidemiology*, vol. 160, no. 8, pp. 719–728, Oct. 2004.

[138] E. H. Lieb and F. Y. Wu, “Two-dimensional ferroelectric models”, 1980.

[139] J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz, “Superspreading and the effect of individual variation on disease emergence”, *Nature*, vol. 438, no. 7066, pp. 355–359, 2005.

[140] W. Looi, E. Pierson, B. Redbird, B. Villanueva, N. Fishman, Y. Chen, J. Sholar, J. Leskovec, and D. Grusky, “The segregation of interaction”, *Working paper*, 2020. eprint: `medRxiv`.

[141] J. Lu, J. Gu, K. Li, C. Xu, W. Su, Z. Lai, D. Zhou, C. Yu, B. Xu, and Z. Yang, “COVID-19 outbreak associated with air conditioning in restaurant, Guangzhou, China, 2020”, *Emerg. Infect. Dis.*, 2020.

[142] I. G. Macdonald, *Symmetric functions and Hall polynomials*, 2nd ed. Oxford University Press Inc., New York, 1995.

[143] D. Miller, M. A. Martin, N. Harel, T. Kustin, O. Tirosh, M. Meir, N. Sorek, S. Gefen-Halevi, S. Amit, O. Vorontsov, D. Wolf, A. Peretz, Y. Shemer-Avni, D. Roif-Kaminsky, N. Kopelman, A. Huppert, K. Koelle, and A. Stern, "Full genome viral sequences inform patterns of SARS-CoV-2 spread into and within Israel", *medRxiv*, 2020.

[144] W. H. Mills, D. P. Robbins, and H. Rumsey, "Alternating sign matrices and descending plane partitions", *J. Comb. Theory. Series A.*, vol. 34, pp. 340–359, 3 1983.

[145] C.-K. Min, S. Cheon, N.-Y. Ha, K. M. Sohn, Y. Kim, A. Aigerim, H. M. Shin, J.-Y. Choi, K.-S. Inn, J.-H. Kim, J. Y. Moon, M.-S. Choi, N.-H. Cho, and Y.-S. Kim, "Comparative and kinetic analysis of viral shedding and immunological responses in MERS patients representing a broad spectrum of disease severity", *Scientific Reports*, vol. 6, 25359, 2016.

[146] L. R. Nandayapa, "Risk probabilities: Asymptotics and simulation", PhD thesis, University of Aarhus, 2008.

[147] J. Neipel, J. Bauermann, S. Bo, T. Harmon, and F. Jülicher, "Power-law population heterogeneity governs epidemic waves", *PLOS One*, vol. 15, e0239678, 2020.

[148] C. M. Newman, K Ravishankar, and E Schertzer, "Marking (1, 2) points of the Brownian web and applications", vol. 46, no. 2, pp. 537–574, 2010.

[149] E. Nordenstam, "Interlaced particles in tilings and random matrices", PhD thesis, KTH, 2009.

[150] N. O'Connell, "Directed polymers and the quantum Toda lattice", *Ann. Probab.*, vol. 40, no. 2, pp. 437–458, 2012.

[151] N. O'Connell and J. Ortmann, "Tracy-Widom asymptotics for a random polymer model with gamma-distributed weights", *Electron. J. Probab.*, vol. 20, no. 25, 1–18, 2015.

[152] N. O'Connell and M. Yor, "A representation for non-colliding random walks", *Electron. Commun. Probab.*, vol. 7, pp. 1–12, 2002.

[153] A. Okounkov and N. Reshetikhin, "Lectures on the integrability of the six-vertex model", *Exact methods in low-dimensional statistical physics and quantum computing*, pp. 197–266, 2010.

[154] A. Y. Okounkov and N. Y. Reshetikhin, "The birth of a random matrix", *Moscow Mathematical Journal*, vol. 6, no. 3, pp. 553–566, 2006.

[155] F. W. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, *NIST Handbook of Mathematical Functions*, 1st. USA: Cambridge University Press, 2010, ISBN: 0521140633.

[156] D. Orr and L. Petrov, "Stochastic higher spin six vertex model and q-TASEPs", *Adv. Math.*, vol. 317, pp. 473–525, 2017.

[157] F. Patrick and V. Balint, "Tracy Widom asymptotics for q-TASEP", in *Ann. Inst. Henri Poincaré Probab. Stat.*, Institut Henri Poincaré, vol. 51, 2015, pp. 1465–1485.

[158] A. M. Povolotsky, "On the integrability of zero-range chipping models with factorized steady states", *J. Phys. A*, vol. 46, no. 46, p. 465 205, 2013.

[159] M. Prähofer and H. Spohn, "Universal distributions for growth processes in 1+ 1 dimensions and random matrices", *Phys. rev. lett.*, vol. 84, no. 21, p. 4882, 2000.

[160] M. Z. Rácz and M. Shkolnikov, "Multidimensional sticky Brownian motions as limits of exclusion processes", *Ann. Appl. Probab.*, vol. 25, no. 3, pp. 1155–1188, 2015.

[161] F. Rassoul-Agha and T. Seppäläinen, "An almost sure invariance principle for random walks in a space-time random environment", *Probab. Theory Related Fields*, vol. 133, no. 3, pp. 299–314, 2005.

[162] ——, "Quenched point-to-point free energy for random walks in random potentials", *Probab. Theory Related Fields*, vol. 158, no. 3-4, pp. 711–750, 2014.

[163] F. Rassoul-Agha, T. Seppäläinen, and A. Yilmaz, "Quenched free energy and large deviations for random walks in random potentials", *Comm. Pure Appl. Math.*, vol. 66, no. 2, pp. 202–244, 2013.

[164] D. Revuz and M. Yor, *Continuous martingales and Brownian motion*. Springer Science & Business Media, 2013, vol. 293.

[165] J. L. Santarpia, D. N. Rivera, V. Herrera, M. J. Morwitzer, H. Creager, G. W. Santarpia, K. K. Crown, D. Brett-Major, E. Schnaubelt, M. J. Broadhurst, J. V. Lawler, S. P. Reid, and J. J. Lowe, "Transmission potential of SARS-CoV-2 in viral shedding observed at the University of Nebraska Medical Center", *medRxiv*, 2020.

[166] E Schertzer, R Sun, and J. M. Swart, "Stochastic flows in the Brownian web and net", *Memoirs Amer. Math. Soc.*, vol. 227, no. 1065, pp. 1–172, 2014.

[167] E. Schertzer, R. Sun, and J. Swart, "The Brownian web, the Brownian net, and their universality", *Advances in Disordered Systems, Random Processes and Some Applications*, pp. 270–368, 2015.

[168] E. Schertzer, R. Sun, and J. M. Swart, "Special points of the Brownian net", *Electr. J. Probab.*, vol. 14, pp. 805–864, 2009.

[169] E. Schertzer, R. Sun, and J. M. Swart, "The Brownian web, the Brownian net, and their universality", *Advances in Disordered Systems, Random Processes and Some Applications*, pp. 270–368, 2015.

[170] T. Seppäläinen, "Scaling for a one-dimensional directed polymer with boundary conditions", *Ann. Probab.*, vol. 40, no. 1, pp. 19–73, 2012.

[171] S. Sheffield, "Random surfaces", *Astérisque*, vol. 304, 2005.

[172] B. I. Shraiman and E. D. Siggia, "Scalar turbulence", *Nature*, vol. 405, no. 6787, p. 639, 2000.

[173] H. Spohn, "KPZ scaling theory and the semidiscrete directed polymer model", *Random Matrix Theory, Interacting Particle Systems and Integrable Systems*, vol. 65, 2012.

[174] H. Spohn and T. Sasamoto, "Point-interacting Brownian motions in the KPZ universality class", *Electron. J. Probab.*, vol. 20, 2015.

[175] R. F. Squire, "What about bias in the SafeGraph dataset?", 2019, Available at https://safegraph.com/blog/what-about-bias-in-the-safegraph-dataset.

[176] E. M. Stein and R. Shakarchi, *Complex analysis*. Princeton University Press, 2010, vol. 2.

[177] E. Stein and R. Shakarchi, *Real analysis: measure theory, integration, and Hilbert spaces*. Princeton University Press, Princeton, 2009.

[178] R. A. Stein, "Super-spreaders in infectious diseases", *International Journal of Infectious Diseases*, vol. 15, no. 8, e510–e513, 2011.

[179] R. Sun and J. M. Swart, "The Brownian net", *Ann. Probab.*, vol. 36, no. 3, pp. 1153–1208, 2008.

[180] O. F. Syljuåsen and M. Zvonarev, "Directed-loop monte carlo simulations of vertex models", *Phys. Rev. E*, vol. 70, no. 1, p. 016 118, 2004.

[181] H. Tanaka, "Limit theorems for a Brownian motion with drift in a white noise environment", *Chaos Solitons Fractals*, vol. 8, pp. 1807–1816, 1997.

[182] T. Thiery, "Stationary measures for two dual families of finite and zero temperature models of directed polymers on the square lattice", *J. Stat. Phys.*, vol. 165, no. 1, pp. 44–85, 2016.

[183] T. Thiery and P. Le Doussal, "Exact solution for a random walk in a time-dependent 1D random environment: the point-to-point Beta polymer", *J. Phys. Al*, vol. 50, no. 4, p. 045 001, 2016.

[184] ——, "On integrable directed polymer models on the square lattice", *J. Phys. A*, vol. 48, no. 46, p. 465 001, 2015.

[185] B. Tóth and W. Werner, "The true self-repelling motion", *Probab. Theor. Rel. Fields*, vol. 111, no. 3, pp. 375–452, 1998.

[186] C. A. Tracy and H. Widom, "A Fredholm determinant representation in ASEP", *J. Stat. Phys.*, vol. 132, no. 2, pp. 291–300, 2008.

[187] ——, "Asymptotics in ASEP with step initial condition", *Comm. Math. Phys.*, vol. 290, no. 1, pp. 129–154, 2009.

[188] ——, "Integral formulas for the asymmetric simple exclusion process", *Comm. Math. Phys.*, vol. 279, no. 3, pp. 815–844, 2008.

[189] C. A. Tracy and H. Widom, "Level-spacing distributions and the Airy kernel", *Comm. Math. Phys.*, vol. 159, no. 1, pp. 151–174, 1994.

[190] R. Tribe, J. Yip, and O. Zaboronski, "One dimensional annihilating and coalescing particle systems as extended pfaffian point processes", *Electron. Commun. Probab.*, vol. 17, 2012.

[191] R. Tribe and O. Zaboronski, "Pfaffian formulae for one dimensional coalescing and annihilating systems", *Electron. J. Probab.*, vol. 16, pp. 2080–2103, 2011.

[192] ——, "The ginibre evolution in the large-n limit", *J. Math. Phys*, vol. 55, no. 6, p. 063 304, 2014.

[193] B. Vető, "Tracy-Widom limit of q-Hahn TASEP", *Electron. J. Probab.*, vol. 20, 2015.

[194] S. X. Wang, Y. M. Li, B. C. Sun, S. W. Zhang, W. H. Zhao, M. T. Wei, K. X. Chen, X. L. Zhao, Z. L. Zhang, M. Krahn, A. C. Cheung, and P. P. Wang, "The SARS outbreak in a general hospital in Tianjin, China—the case of super-spreader", *Epidemiol. Infect.*, vol. 134, no. 4, pp. 786–791, 2006.

[195] J. Warren, "Dyson's Brownian motions, intertwining and interlacing", *Electr. J. Probab.*, vol. 12, pp. 573–590, 2007.

[196] ——, "Sticky particles and stochastic flows", in *In Memoriam Marc Yor-Séminaire de Probabilités XLVII*, Springer, 2015, pp. 17–35.

[197] J. Warren, “Branching processes, the Ray-Knight theorem, and sticky Brownian motion”, in *Séminaire de Probabilités XXXI*, Springer, 1997, pp. 1–15.

[198] W. Yang, S. Kandula, M. Huynh, S. K. Greene, G. Van Wye, and W. Li, “Estimating the infection-fatality risk of SARS-CoV-2 in New York City during the spring 2020 pandemic wave: a model-based analysis”, *The Lancet Infectious Diseases*, 2020.

[199] J. Yu, “Edwards–Wilkinson fluctuations in the Howitt–Warren flows”, *Stoch. Proc. Appl.*, vol. 126, no. 3, pp. 948–982, 2016.