Brownian continuum objects:

the excursion, tree, ~~snake~~, ~~map~~, web, net, ~~and castle~~

J.M. Swart

April 27, 2022

Contents

Preface

The road to hell is paved with good intentions. The original idea of these lecture notes was to give an overview of a number of “Brownian” continuum objects that have been studied in recent years, namely the Brownian continuum random tree, the Brownian snake, the Brownian map, the Brownian web, the Brownian net, and the Brownian castle. Here the first three (the tree, snake, and map) are closely interrelated, and so are the second three (the web, the net, and the castle). The relations between these two groups are a bit more loose, but there are still sufficiently many similarities to justify treating them together. For example, there has been recent work viewing the Brownian web as a real-tree. There are also relations to older, by now classical topics. In particular, the theory of the Brownian continuum random tree makes use of Brownian excursion theory and the Brownian snake is closely related to super Brownian motion.

Introducing all these objects in one semester course was always going to be hard. There are some useful overview texts. For the Brownian continuum random tree and the Brownian snake, there is Le Gall’s excellent exposition [LeG05] (for the continuum random tree there is moreover [Eva08]) while for the Brownian web and net there is [SSS16]. Nevertheless, after being impressed by the technical difficulty of these texts, I decided at the very last moment to write my own lecture notes. Unsurprisingly, I managed to cover less material than I hoped for, so sadly, I did not find time to include the snake, map, and castle. The lecture notes are also still very much work in progress, but as a somewhat chaotic first draft, they are nevertheless written. Perhaps they will be improved and extended in due time.

Chapter 1

Topological prerequisites

1.1 Topological spaces

A *topological space* is a set $\mathcal{X}$ equipped with a collection $\mathcal{O}$ of subsets of $\mathcal{X}$ that are called *open* sets, such that

(i) If $(O_\gamma)_{\gamma\in\Gamma}$ is any collection of (possibly uncountably many) sets $O_\gamma \in \mathcal{O}$, then $\bigcup_{\gamma\in\Gamma} O_\gamma \in \mathcal{O}$.

(ii) If $O_1, O_2 \in \mathcal{O}$, then $O_1 \cap O_2 \in \mathcal{O}$.

(iii) $\emptyset, \mathcal{X} \in \mathcal{O}$.

Any such collection of sets is called a *topology*. It is fairly standard to also assume the *Hausdorff* property

(iv) For each $x_1, x_2 \in \mathcal{X}$, $x_1 \neq x_2$ $\exists O_1, O_2 \in \mathcal{O}$ s.t. $O_1 \cap O_2 = \emptyset$, $x_1 \in O_1$, $x_2 \in O_2$.

A set $V \subset \mathcal{X}$ is a *neighbourhood* of a point $x \in \mathcal{X}$ if $x \in O \subset V$ for some $O \in \mathcal{O}$. We let $\mathcal{V}_x$ denote the set of all neighbourhoods of x. A *fundamental system* of neighbourhoods of x is a set $\mathcal{V}'_x \subset \mathcal{V}_x$ such that

$$\forall V \in \mathcal{V}_x \ \exists V' \in \mathcal{V}'_x \text{ s.t. } V' \subset V.$$

For example, the set of all $O \in \mathcal{O}$ such that $x \in O$ is a fundamental system of neighbourhoods of x. A sequence of points $x_n \in \mathcal{X}$ converges to a limit x in a given topology $\mathcal{O}$ if for each $V \in \mathcal{V}_x$ there is an n such that $x_m \in V$ for all $m \geq n$. It suffices to check this condition for a fundamental system of neighbourhoods $\mathcal{V}'_x$. If the topology is Hausdorff, then limits are unique, i.e., $x_n \to x$ and $x_n \to x'$ implies $x = x'$.

If $(\mathcal{X},\mathcal{O})$ is a topological space (with $\mathcal{O}$ the collection of open subsets of $\mathcal{X}$) and $\mathcal{X}'\subset\mathcal{X}$ is any subset of $\mathcal{X}$, then $\mathcal{X}'$ is also naturally equipped with a topology given by the collection of open subsets $\mathcal{O}' := \{O\cap\mathcal{X}' : O\in\mathcal{O}\}$. This topology is called the *induced* topology from $\mathcal{X}$. If $x_n, x\in\mathcal{X}'$, then $x_n\to x$ in the induced topology on $\mathcal{X}'$ if and only if $x_n\to x$ in $\mathcal{X}$.

A *basis* of a topology is a subset $\mathcal{O}'\subset\mathcal{O}$ such that each element of $\mathcal{O}$ can be written as the union of (possibly uncountably many) elements of $\mathcal{O}'$. Equivalently, this says that

$$\mathcal{O} = \{O\subset\mathcal{X} : \forall x\in O\ \exists O'\in\mathcal{O}' \text{ s.t. } x\in O'\subset O\}.$$

If $\mathcal{O}'$ is a basis for $\mathcal{O}$, then $\mathcal{V}'_x := \{O\in\mathcal{O}' : x\in O\}$ is a fundamental system of neighbourhoods of x. A topology is *first countable* if every $x\in\mathcal{X}$ has a countable fundamental system of neighbourhoods. A topology is *second countable* if there exists a countable basis of the topology.

A set $C\subset\mathcal{X}$ is called *closed* if its complement is open. Because of property (i) in the definition of a topology, for each $A\subset\mathcal{X}$, the union of all open sets contained in A is itself an open set. We call this the *interior* of A, denoted as $\mathrm{int}(A) := \bigcup\{O : O\subset A,\ O \text{ open}\}$. Then clearly $\mathrm{int}(A)$ is the largest open set contained in A. Similarly, by taking complements, for each set $A\subset\mathcal{X}$ there exists a smallest closed set containing A. We call this the *closure* of A, denoted as $\overline{A} := \bigcap\{C : C\supset A,\ C \text{ closed}\}$. If the topology is first countable, then

$$\overline{A} = \{x\in\mathcal{X} : \exists x_n\in\mathcal{X} \text{ s.t. } x_n\to x\}, \tag{1.1}$$

i.e., $\overline{A}$ is the set of all limits of sequences in A. A similar statement holds for general topological spaces if we replace sequences by the more general concept of a *net*, that we will not discuss here. Since a set is closed if and only if it coincides with its closure, it follows from (1.1) that in a first countable topological space, knowing all convergent sequences and their limits uniquely determines the closed sets and their complements, the open sets, and hence the whole topology.

A topological space is called *separable* if there exists a countable set $D\subset\mathcal{X}$ such that D is dense in $\mathcal{X}$, where we say that a set $D\subset\mathcal{X}$ is *dense* if its closure is $\mathcal{X}$, or equivalently, if every nonempty open subset of $\mathcal{X}$ has a nonempty intersection with D.

A *metric* on a set $\mathcal{X}$ is a function $d:\mathcal{X}\times\mathcal{X}\to[0,\infty)$ such that for all $x,y,z\in\mathcal{X}$,

(i) $d(x,y) = d(y,x)$,

(ii) $d(x,z)\leq d(x,y)+d(y,z)$,

(iii) $d(x,y)=0$ implies $x=y$.

A *metric space* is a space with a metric defined on it. If d is a metric on $\mathcal{X}$, and $B_\varepsilon(x) := \{y \in \mathcal{X} : d(x,y) < \varepsilon\}$ denotes the open ball around x of radius ε, then

$$\mathcal{O} := \big\{O \subset \mathcal{X} : \forall x \in O\ \exists \varepsilon > 0 \text{ s.t. } B_\varepsilon(x) \subset O\big\}$$

defines a Hausdorff topology on $\mathcal{X}$ such that convergence $x_n \to x$ in this topology is equivalent to $d(x_n, x) \to 0$. Note that the open balls form a basis for this topology. Since open balls of radius $1/n$ around a point x form a fundamental system of neighbourhoods, metric spaces are first countable. We say that the metric d *generates* the topology $\mathcal{O}$. If for a given topology $\mathcal{O}$ there exists a metric d that generates $\mathcal{O}$, then we say that the topological space $(\mathcal{X},\mathcal{O})$ is *metrisable*. Such a metric, if it exist, can always be chosen such that it is bounded. For example, if d is any metric on $\mathcal{X}$, then $d'(x,y) := d(x,y) \wedge 1$ is a bounded metric that generates the same topology. A metrisable space is always first countable. It is second countable if and only if it is separable.

A sequence x_n in a metric space $(\mathcal{X}, d)$ is a *Cauchy sequence* if for all $\varepsilon > 0$ there is an n such that $d(x_k, x_l) \leq \varepsilon$ for all $k, l \geq n$. A metric space is *complete* if every Cauchy sequence converges. Every metric space $(\mathcal{X}, d)$ has a *completion*, i.e., there exists a complete metric space $(\overline{\mathcal{X}}, \overline{d})$ such that $\mathcal{X} \subset \overline{\mathcal{X}}$ is dense and the metric on $\mathcal{X}$ is the *induced metric* from $\overline{\mathcal{X}}$, i.e., $d(x,y) = \overline{d}(x,y)$ for all $x, y \in \mathcal{X}$. Such a completion is unique up to isometries.

A *Polish space* is a separable topological space $(\mathcal{X},\mathcal{O})$ such that there exists a metric d on $\mathcal{X}$ with the property that $(\mathcal{X}, d)$ is complete and d generates $\mathcal{O}$. *Warning:* there may be many different metrics on $\mathcal{X}$ that generate the same topology. It may even happen that $\mathcal{X}$ is not complete in some of these metrics, and complete in others (in which case $\mathcal{X}$ is still Polish). Example: $\mathbb{R}$ is separable and complete in the usual metric $d(x,y) = |x-y|$, and therefore $\mathbb{R}$ is a Polish space. But $d'(x,y) := |\arctan(x) - \arctan(y)|$ is another metric that generates the same topology, while $(\mathbb{R}, d')$ is not complete. (Indeed, the completion of $\mathbb{R}$ w.r.t. the metric d' is $[-\infty, \infty]$.)

1.2 Compactness

A subset K of a general topological space $\mathcal{X}$ (with collection of open sets $\mathcal{O}$) is called *compact* if every open cover has a finite subcover, i.e., if for any collection $(O_\gamma)_{\gamma \in \Gamma}$ of open subsets of $\mathcal{X}$ such that $\bigcup_{\gamma \in \Gamma} O_\gamma \supset K$, there exists a finite $\Delta \subset \Gamma$ such that $\bigcup_{\gamma \in \Delta} O_\gamma \supset K$. Using this definition, it is easy to see that the image of a compact set under a continuous function is again compact. Compact subsets of Hausdorff topological spaces are closed. A subset K of a metric space $\mathcal{X}$ is compact if and only if it is closed and *totally bounded*, which means that for every $\varepsilon > 0$ there exists a finite collection $\{B_\varepsilon(x_1), \ldots, B_\varepsilon(x_n)\}$ of open balls such that

$$B_\varepsilon(x_1) \cup \cdots \cup B_\varepsilon(x_n) \supset K.$$

From this, it is not hard to see that compact metrisable spaces are always separable. If $(x_n)_{n\in\mathbb{N}}$ is a sequence and $m : \mathbb{N} \to \mathbb{N}$ is a function such that $m(n) \to \infty$ as $n \to \infty$, then setting $x'_n := x_{m(n)}$ $(n \in \mathbb{N})$ defines a new sequence. Such a sequence is called a *subsequence* of the original sequence. A *cluster point* of a sequence is a limit of a subsequence.

Theorem 1.1 (Bolzano-Weierstrass) *Let $\mathcal{X}$ be a metrisable space and let $K \subset \mathcal{X}$. Then K is compact if and only if every sequence in K has a subsequence that converges to a limit in K.*

The Bolzano-Weierstrass theorem also holds for second countable spaces. (Note that metrisable spaces need in general not be second countable, and conversely, not every second countable space is metrisable.) There is also a version of the Bolzano-Weierstrass theorem that holds in general topological spaces but in this case one has to replace sequences by the more general nets. A set A is *precompact* if its closure is compact. In metrisable spaces, this is equivalent to the statement that each sequence of points $x_n \in A$ has a convergent subsequence. Note that in this case we do not require that the limit is an element of A. The following simple lemma is often useful.

Lemma 1.2 (Convergence and compactness) *Let $\mathcal{X}$ be a metrisable space and let $x, x_n \in \mathcal{X}$. Then $x_n \to x$ if and only if the following two conditions are satisfied.*

(i) *The set $\{x_n : n \in \mathbb{N}\}$ is precompact.*

(ii) *For every subsequence $x_{n(m)}$ such that $x_{n(m)} \underset{m\to\infty}{\longrightarrow} x'$ for some $x' \in \mathcal{X}$, one has $x' = x$.*

If $(\mathcal{X}, \mathcal{O})$ is a topological space, then a *compactification* of $\mathcal{X}$ is a compact topological space $\overline{\mathcal{X}}$ such that $\mathcal{X}$ is a dense subset of $\overline{\mathcal{X}}$ and the topology on $\mathcal{X}$ is the induced topology from $\overline{\mathcal{X}}$. If $\overline{\mathcal{X}}$ is metrisable, then we say that $\overline{\mathcal{X}}$ is a *metrisable compactification* of $\mathcal{X}$. It turns out that each separable metrisable space $\mathcal{X}$ has a metrisable compactification [Cho69, Theorem 6.3].

A topological space $\mathcal{X}$ is called *locally compact* if for every $x \in \mathcal{X}$ there exists a compact neighbourhood of x. We cite the following proposition from [Eng89, Thms 3.3.8 and 3.3.9].

Proposition 1.3 (Compactification of locally compact spaces) *Let $\mathcal{X}$ be a metrisable topological space. Then the following statements are equivalent.*

(i) *$\mathcal{X}$ is locally compact and separable.*

(ii) *There exists a metrisable compactification $\overline{\mathcal{X}}$ of $\mathcal{X}$ such that $\mathcal{X}$ is an open subset of $\overline{\mathcal{X}}$.*

(iii) *For each metrisable compactification $\overline{\mathcal{X}}$ of $\mathcal{X}$, $\mathcal{X}$ is an open subset of $\overline{\mathcal{X}}$.*

We note that if $\mathcal{X}$ satisfies the equivalent conditions of Proposition 1.3, then it is possible to find a metrisable compactification $\overline{\mathcal{X}}$ of $\mathcal{X}$ such that $\overline{\mathcal{X}}\backslash\mathcal{X}$ consists of just one point, usually denoted by ∞. In this case, $\overline{\mathcal{X}} = \mathcal{X} \cup \{\infty\}$ is called the *one-point compactification* of $\mathcal{X}$. The open sets of $\mathcal{X} \cup \{\infty\}$ are all open sets of $\mathcal{X}$ plus all sets of the form $\{\infty\} \cup O$ where $\mathcal{X}\backslash O$ is a compact subset of $\mathcal{X}$.

A subset $A \subset \mathcal{X}$ of a topological space $\mathcal{X}$ is called a *G_δ-set* if A is a countable intersection of open sets (i.e., there exist $O_i \in \mathcal{O}$ such that $A = \bigcap_{i=1}^\infty O_i$. If $\mathcal{X}$ is metrisable, then every closed set $A \subset \mathcal{X}$ is a G_δ-set, since it is the intersection of the open sets $\{x \in \mathcal{X} : d(x, A) < 1/n\}$. The following result can be found in [Bou58, §6 No. 1, Theorem. 1]. See also [Oxt80, Thms 12.1 and 12.3].

Proposition 1.4 (Compactification of Polish spaces) *Let $\mathcal{X}$ be a metrisable topological space. Then the following statements are equivalent.*

(i) *$\mathcal{X}$ is Polish.*

(ii) *There exists a metrisable compactification $\overline{\mathcal{X}}$ of $\mathcal{X}$ such that $\mathcal{X}$ is a G_δ-subset of $\overline{\mathcal{X}}$.*

(iii) *For each metrisable compactification $\overline{\mathcal{X}}$ of $\mathcal{X}$, $\mathcal{X}$ is a G_δ-subset of $\overline{\mathcal{X}}$.*

Moreover, a subset $\mathcal{Y} \subset \mathcal{X}$ of a Polish space $\mathcal{X}$ is Polish in the induced topology if and only if $\mathcal{Y}$ is a G_δ-subset of $\mathcal{X}$.

We note that if $\overline{\mathcal{X}}$ is a compactification of a Polish space $\mathcal{X}$, equipped with a concrete metric, then $\overline{\mathcal{X}}$ is also the completion of $\mathcal{X}$ in this metric. Thus, unless $\mathcal{X}$ is itself compact, it will never be complete in such a metric (even though, by the definition of a Polish space, there exists metrics generating the same topology with respect to which $\mathcal{X}$ is complete).

1.3 Decomposition of measures

Let $(\mathcal{X}, \mathcal{F})$ be a measurable space and let $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ be a Polish space, equipped with its Borel-σ-field. By definition, a *probability kernel* from $\mathcal{X}$ to $\mathcal{Y}$ is a measurable map

$$\mathcal{X} \ni x \mapsto K(x, \cdot\,) \in \mathcal{M}_1(\mathcal{Y}).$$

Since the Borel-σ-field on $\mathcal{M}_1(\mathcal{Y})$ is generated by the maps $\mu \mapsto \mu(A)$ with $A \in \mathcal{B}(\mathcal{Y})$, the measurability of K is equivalent to the statement that for each $A \in \mathcal{B}(\mathcal{Y})$, the function $K(\,\cdot\,, A)$ is a measurable real function on $\mathcal{X}$. More generally, if $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ is replaced by a general measurable space $(\mathcal{Y}, \mathcal{G})$, then we define a *probability kernel* from $\mathcal{X}$ to $\mathcal{Y}$ to be a map $K : \mathcal{X} \times \mathcal{G} \to [0, 1]$ such that $K(x, \cdot\,)$ is a probability measure on $\mathcal{Y}$ for each $x \in \mathcal{X}$ and $K(\,\cdot\,, A)$ is a measurable real function on $\mathcal{X}$ for each $A \in \mathcal{G}$. With these definitions, one has the following result.

Theorem 1.5 (Decomposition of probability measures) *Let $(\mathcal{X}, \mathcal{F})$ and $(\mathcal{Y}, \mathcal{G})$ be a measurable spaces. Let μ be a probability measure on $\mathcal{X}$ and let K be a probability kernel from $\mathcal{X}$ to $\mathcal{Y}$. Then there exists a unique probability measure ν on $\mathcal{X} \times \mathcal{Y}$, equipped with the product-σ-field, so that for any measurable function $f : \mathcal{X} \times \mathcal{Y} \to [0, \infty]$,*

$$\int f(x, y)\nu\big(\mathrm{d}(x, y)\big) = \int \mu(\mathrm{d}x) \int K(x, \mathrm{d}y) f(x, y), \tag{1.2}$$

where on the right-hand side, $x \mapsto \int K(x, \mathrm{d}y) f(x, y)$ is a measurable function on $\mathcal{X}$ that is integrated against μ.

Assume that moreover, $(\mathcal{Y}, \mathcal{G}) = (\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ is a Polish space equipped with its Borel-σ-field. Then conversely for each probability measure ν on $\mathcal{X} \times \mathcal{Y}$, there exist a probability measure μ on $\mathcal{X}$ and probability kernel K from $\mathcal{X}$ to $\mathcal{Y}$ such that (1.2) holds. If (1.2) holds, then μ is the first marginal of ν. Moreover, (1.2) determines the kernel K a.s. uniquely, i.e., if K, K' are probability kernels so that (1.2) holds both for K and K', then there exists a set $N \in \mathcal{F}$ with $\mu(N) = 0$ such that $K(x, \cdot\,) = K'(x, \cdot\,)$ for all $x \in \mathcal{X} \backslash N$.

We note that by subtracting a constant, we see that (1.2) holds more generally for functions f that are bounded from below.

The deep part of Theorem 1.5 is the existence of K given ν;. this may fail in general if the Polish space $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ is replaced by an arbitrary measurable space $(\mathcal{Y}, \mathcal{G})$. Formally, we may define a ‘measure’ ρ on $\mathcal{X}$ with values in $\mathcal{M}_1(\mathcal{Y})$ by $\rho(A) := \nu(A \times \,\cdot\,)$ $(A \in \mathcal{F})$. Letting μ denote the first marginal of ν, we obviously have $\rho(A) = 0$ whenever $\mu(A) = 0$, i.e., formally $\rho \ll \mu$. Now (1.2) says that

$$\rho(A) = \int_A \mu(\mathrm{d}x) K(x,\,\cdot\,),$$

which we can formally read as saying that ρ has a density with respect to μ, which is given by the function $x \mapsto K(x,\,\cdot\,)$. Thus, Theorem 1.5 amounts to proving something like a Radon-Nikodym theorem for functions and measures with values in $\mathcal{M}_1(\mathcal{Y})$. In fact, if we are just interested in $K(\,\cdot\,, B)$ for one fixed $B \in \mathcal{B}(\mathcal{Y})$, then (1.2) says that

$$\nu(A \times B) = \int_A \mu(\mathrm{d}x) K(x, B) \qquad (A \in \mathcal{F}). \tag{1.3}$$

Since $\nu(\,\cdot\, \times B) \ll \mu$ (where μ is the first marginal of ν), the usual Radon-Nikodym now tells us that for this fixed B, there exists an a.s. unique function $K(\,\cdot\,, B)$ such that (1.3) holds. This argument does not tell us, however, whether for fixed x, we can choose $K(x,\,\cdot\,)$ such that it is a probability measure. Theorem 1.5 tells us that that such a *regular version* of K exists.

Corollary 1.6 (Regular conditional probability) *Let Y be a random variable defined on an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in some Polish space $\mathcal{Y}$, and let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-field. Then there exists a $\mathcal{M}_1(\mathcal{Y})$-valued random variable $\mathbb{P}[Y \in \cdot \,|\, \mathcal{G}]$, which is unique up to a $\mathcal{G}$-measurable set of probability zero, such that*

(i) $\mathbb{P}[Y \in \cdot \,|\, \mathcal{G}]$ *is $\mathcal{G}$-measurable.*

(ii) $\mathbb{E}\big[1_A \mathbb{P}[Y \in B \,|\, \mathcal{G}]\big] = \mathbb{E}[1_A 1_{\{Y \in B\}}]$ *for all $A \in \mathcal{G}$, $B \in \mathcal{B}(\mathcal{Y})$.*

Proof It is not hard to see that there exists a unique probability measure on $\omega \times \mathcal{Y}$, equipped with the σ-field $\mathcal{G} \otimes \mathcal{B}(\mathcal{Y})$, such that

$$\int \nu\big(\mathrm{d}(\omega, y)\big) f(\omega, y) := \int \mathbb{P}(\mathrm{d}\omega)\, f(\omega, Y(\omega))$$

for all $f : \omega \times \mathcal{Y} \to [0, \infty]$ that are measurable w.r.t. $\mathcal{G} \otimes \mathcal{B}(\mathcal{Y})$. Applying Theorem 1.5 to ν, we obtain a $\mathcal{G}$-measurable, $\mathcal{M}_1(\mathcal{Y})$-valued random variable $\mathbb{P}[Y \in \cdot \,|\, \mathcal{G}]$ (i.e., a probability kernel from $(\Omega, \mathcal{G})$ to $\mathcal{Y}$), unique up to a $\mathcal{G}$-measurable set of probability zero, such that

$$\int \mathbb{P}(\mathrm{d}\omega)\, f(\omega, Y(\omega)) = \int \mathbb{P}(\mathrm{d}\omega) \int \mathbb{P}[Y \in \mathrm{d}y \,|\, \mathcal{G}](\omega)\, f(\omega, y)$$

By the uniqueness theorem (applied to ν), to verify this equation, it suffices to check it for functions f of the form $f = 1_{A\times B}$ with $A \in \mathcal{G}$ and $B \in \mathcal{B}(\mathcal{Y})$. Thus, $\mathbb{P}[Y \in \cdot \,|\, \mathcal{G}]$ is a.s. uniquely determined by the requirement that

$$\begin{aligned}\mathbb{E}\big[1_A \mathbb{P}[Y \in B \,|\, \mathcal{G}]\big] &= \int \mathbb{P}(\mathrm{d}\omega)\mathbb{P}[Y \in \mathrm{d}y \,|\, \mathcal{G}](\omega) 1_{A\times B}(\omega, y) \\ &= \nu(A \times B) = \mathbb{E}[1_A 1_{\{Y=B\}}].\end{aligned}$$

∎

1.4 Weak convergence

Let $\mathcal{X}$ be a metrisable space. We let $\mathcal{B}(\mathcal{X})$ denote *Borel-σ-field* on $\mathcal{X}$, i.e., the σ-field generated by the open sets. We let $\mathcal{C}(\mathcal{X})$ denote the space of all continuous functions $f : \mathcal{X} \to \mathbb{R}$. We let $B_{\mathrm{b}}(\mathcal{X})$ denote the space of all bounded Borel-measurable real functions on $\mathcal{X}$ and we let $\mathcal{C}_{\mathrm{b}}(\mathcal{X}) := \mathcal{C}(\mathcal{X}) \cap B_{\mathrm{b}}(\mathcal{X})$ denote the space of all bounded continuous real functions on $\mathcal{X}$. We equip with $\mathcal{C}_{\mathrm{b}}(\mathcal{X})$ with the *supremumnorm*

$$\|f\|_\infty := \sup_{x\in\mathcal{X}} |f(x)|.$$

With this norm, $\mathcal{C}_{\mathrm{b}}(\mathcal{X})$ is a Banach space [Dud02, Theorem 2.4.9]. We let $\mathcal{M}(\mathcal{X})$ denote the space of all finite measures on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ and write $\mathcal{M}_1(\mathcal{X})$ for the subspace of all probability measures. We cite the following well-known fact from [EK86, Theorems 3.1.7 and 3.3.1].

Proposition 1.7 (Weak convergence) *Let $\mathcal{X}$ be a separable metrisable space. Then it is possible to equip $\mathcal{M}_1(\mathcal{X})$ with a metric d_{P} such that*

(i) *$(\mathcal{M}_1(\mathcal{X}), d_{\mathrm{P}})$ is a separable metric space,*

(ii) *$d_{\mathrm{P}}(\mu_n, \mu) \to 0$ if and only if $\int f \mathrm{d}\mu_n \to \int f \mathrm{d}\mu$ for all $f \in \mathcal{C}_{\mathrm{b}}(\mathcal{X})$.*

If $\mathcal{X}$ is a Polish space, then d_{P} can be chosen such that $(\mathcal{M}_1(\mathcal{X}), d_{\mathrm{P}})$ is moreover complete.

In many applications, we are not interested in the precise choice of d_{P} (there are several canonical ways to define such a metric). Since a metrisable topology is uniquely characterized by its convergent sequences, property (ii) uniquely characterizes the topology generated by d_{P} in terms of the topology on $\mathcal{X}$. We call this topology the *topology of weak convergence* and denote convergence in this topology as

$$\mu_n \Rightarrow \mu.$$

Proposition 1.7 shows in particular that if $\mathcal{X}$ is a Polish space, then so is $\mathcal{M}_1(\mathcal{X})$, equipped with the topology of weak convergence.

One possible choice for a metric $d_{\rm P}$ as in Proposition 1.7 is the Prohorov metric. For each subset $A \subset \mathcal{X}$ and $\varepsilon > 0$, we set

$$A^\varepsilon := \big\{x \in \mathcal{X} : d(x, A) < \varepsilon\big\} \quad \text{with} \quad d(x, A) := \inf_{y\in A} d(x, y).$$

If $(\mathcal{X}, d)$ is a metric space, then the *Prohorov metric* is the metric $d_{\rm P}$ on $\mathcal{M}_1(\mathcal{X})$ defined as

$$d_{\rm P}(\mu, \nu) := \inf\{\varepsilon > 0 : \mu(A) \le \nu(A^\varepsilon) + \varepsilon \ \forall A \in \mathcal{B}(\mathcal{X})\big\}.$$

It follows from [EK86, Lemma 3.1.1] that $d_{\rm P}$ is a metric. It is possible to give an alternative characterisation of $d_{\rm P}$ in terms of coupling. Let $C(\mu, \nu)$ denote the space of all probability measures η on $X \times X$ whose first and second marginals are given by μ and ν, respectively. We cite the following lemma from [EK86, Thm 3.1.2].

Lemma 1.8 (Prohorov metric and coupling) *Let $(\mathcal{X}, d)$ be a separable metric space and let $\mu, \nu \in \mathcal{M}_1(\mathcal{X})$. Then*

$$d_{\rm P}(\mu, \nu) = \inf\big\{\varepsilon > 0 : \exists \eta \in C(\mu, \nu) \text{ s.t. } \eta(\{(x, y) \in \mathcal{X}^2 : d(x, y) \ge \varepsilon\}) \le \varepsilon\big\}. \tag{1.4}$$

In words, (1.4) says that $d_{\rm P}(\mu, \nu)$ is the infimum of all $\varepsilon > 0$ for which it is possible to couple random variables X, Y with laws μ, ν such that $\mathbb{P}[d(X, Y) \ge \varepsilon] \le \varepsilon$. We cite the following lemmas from [EK86, Thms 3.1.7 and 3.3.1].

Lemma 1.9 (Properties of Prohorov metric) *Let $(\mathcal{X}, d)$ be a separable metric space and let $d_{\rm P}$ be the Prohorov metric. Then $(\mathcal{M}_1(\mathcal{X}), d_{\rm P})$ is a separable metric space. If $(\mathcal{X}, d)$ is complete, then so is $(\mathcal{M}_1(\mathcal{X}), d_{\rm P})$.*

Lemma 1.10 (Prohorov metric and weak convergence) *Let $(\mathcal{X}, d)$ be a separable metric space and let $d_{\rm P}$ be the Prohorov metric. Then $\mu_n, \mu \in \mathcal{M}_1(\mathcal{X})$ satisfy $d_{\rm P}(\mu_n, \mu) \to 0$ if and only if $\int f \mathrm{d}\mu_n \to \int f \mathrm{d}\mu$ for all $f \in \mathcal{C}_{\rm b}(\mathcal{X})$.*

In particular, Lemmas 1.9 and 1.10 imply Proposition 1.7. The following well-known alternative characterisation of weak convergence [EK86, Theorem 3.3.1] is sometimes useful.

Lemma 1.11 (Characterization with open and closed sets) *Let μ_n and μ be probability measures on a metrisable space $\mathcal{X}$. Then the following statements are equivalent.*

(i) $\mu_n \Rightarrow \mu$.

(ii) $\limsup_{n\to\infty} \mu_n(C) \leq \mu(C)$ *for all closed* $C \subset \mathcal{X}$.

(iii) $\liminf_{n\to\infty} \mu_n(O) \geq \mu(O)$ *for all open* $O \subset \mathcal{X}$.

Exercise 1.12 (Measures concentrated on a subset) *Let $\mathcal{X}$ be a Polish space and let $\mathcal{X}' \subset \mathcal{X}$ be a G_δ-set, equipped with the induced topology. We naturally identify $\mathcal{M}_1(\mathcal{X}')$ with the subset of $\mathcal{M}_1(\mathcal{X})$ consisting of all $\mu \in \mathcal{M}_1(\mathcal{X})$ such that $\mu(\mathcal{X}') = 1$. Show that the topology on $\mathcal{M}_1(\mathcal{X}')$ coincides with the induced topology from its embedding in $\mathcal{M}_1(\mathcal{X})$. (Hint: Lemma 1.11.) Use this to conclude that $\mathcal{M}_1(\mathcal{X}')$ is a G_δ-subset of $\mathcal{M}_1(\mathcal{X})$. (Hint: Proposition 1.4).*

A very useful characterization of weak convergence in terms of coupling is given by the next theorem [EK86, Cor 3.1.6 and Thm 3.1.8].

Theorem 1.13 (Skorohod representation) *Let μ_n and μ be probability measures on a Polish space $\mathcal{X}$. Then $\mu_n \Rightarrow \mu$ if and only if it is possible to couple random variables X_n, X with laws μ_n, μ, respectively, in such a way that $X_n \to X$ a.s.*

The next result is known as Prohorov's theorem (see, e.g., [EK86, Theorem 3.2.2] or [Bil99, Theorems 5.1 and 5.2]).

Theorem 1.14 (Prohorov) *Let $\mathcal{X}$ be a Polish space. Let $\mathcal{M}_1(\mathcal{X})$ be equipped with the topology of weak convergence. Then a subset $\mathcal{C} \subset \mathcal{M}_1(\mathcal{X})$ is precompact if and only if $\mathcal{C}$ is* tight, *i.e.,*

$$\forall \varepsilon > 0 \ \exists K \subset \mathcal{X} \text{ compact, s.t. } \sup_{\mu\in\mathcal{C}} \mu(\mathcal{X}\backslash K) \leq \varepsilon.$$

1.5 Locally uniform convergence

Let E be a metric space and let $I \subset \mathbb{R}$ be a closed interval. We let $\mathcal{C}_I(E)$ denote the space of all continuous functions $w : I \to \mathbb{R}$.

Lemma 1.15 (Locally uniform convergence) *For $w_n, w \in \mathcal{C}_I(E)$, the following conditions are equivalent:*

(i) $\sup_{t\in C} d\big(w_n(t), w(t)\big) \underset{n\to\infty}{\longrightarrow} 0$ *for all compact* $C \subset I$,

(ii) $w_n(t_n) \underset{n\to\infty}{\longrightarrow} w(t)$ *for all* $t_n, t \in I$ *such that* $t_n \underset{n\to\infty}{\longrightarrow} t$.

Proof Assume (i) and let $t_n, t \in I$ satisfy $t_n \underset{n\to\infty}{\longrightarrow} t$. By Lemma 1.2 (i), there exists a compact set $C \subset I$ such that $t_n \in C$ for all n (and hence also $t \in C$). Then for each $\varepsilon > 0$, there exists an $N < \infty$ such that $d(w_n(t), w(t)) \leq \varepsilon$ for all $n \geq N$. Now

$$d\big(w_n(t_n), w(t)\big) \leq d\big(w_n(t_n), w(t_n)\big) + d\big(w(t_n), w(t)\big) \leq \varepsilon + d\big(w(t_n), w(t)\big)$$

for all $n \geq N$, and hence

$$\limsup_{n\to\infty} d\big(w_n(t_n), w(t)\big) \leq \varepsilon$$

by the continuity of w. Since $\varepsilon > 0$ is arbitrary, this shows that (i) implies (ii). On the other hand, if (i) fails for some compact $C \subset I$, then we can choose $t_n \in C$ and $\varepsilon > 0$ such that

$$d\big(w_n(t_n), w(t_n)\big) \geq \varepsilon \quad \forall n.$$

Since C is compact, by going to a subsequence, we can without loss of generality assume that $t_n \to t$ for some $t \in C$. Since

$$d\big(w_n(t_n), w(t)\big) \geq d\big(w_n(t_n), w(t_n)\big) - d\big(w(t_n), w(t)\big) \geq \varepsilon + d\big(w(t_n), w(t)\big),$$

using the continuity of w, we see that

$$\liminf_{n\to\infty} d\big(w_n(t_n), w(t)\big) \geq \varepsilon,$$

which contradicts (ii). ∎

There exists a metrisable topology on $\mathcal{C}_I(E)$ such that a $w_n \in \mathcal{C}_I(E)$ converges to a limit w if and only if the equivalent conditions of Lemma 1.15 are satisfied. Note that by (1.1) and the remarks below it, these conditions uniquely determine the topology. Note also that by condition (ii) of Lemma 1.15, the topology on $\mathcal{C}_I(E)$ depends only on the topology on E and not on the precise choice of the metric on E. A possible choice of a metric on $\mathcal{C}_I(E)$ is

$$\rho(v, w) := \sum_{n=1}^{\infty} 2^{-n} \sup_{t\in[0,n]} d\big(v(t), w(t)\big),$$

where d is a bounded metric that generates the topology on E. Such a metric can always be found: if d is any metric generating the topology on E, then $d'(x, y) := d(x, y) \wedge 1$ is a bounded metric that generates the same topology. Usually, we do not care about the precise choice of the metric on $\mathcal{C}_I(E)$; apart from ρ, there are many other possible choices. We call this the topology on $\mathcal{C}_I(E)$ the *topology of locally uniform convergence.*

1.6 The Hausdorff metric

Let (E,d) be a metric space, let $\mathcal{K}(E)$ be the space of all compact subsets of E and set $\mathcal{K}_+(E) := \{K \in \mathcal{K}(E) : K \neq \emptyset\}$. Then the *Hausdorff metric* d_{H} on $\mathcal{K}_+(E)$ is defined as

$$\begin{aligned} d_{\mathrm{H}}(K_1,K_2) &:= \sup_{x_1\in K_1} d(x_1,K_2) \vee \sup_{x_2\in K_2} d(x_2,K_1) \\ &= \inf\big\{\varepsilon>0 : K_1 \subset K_2^\varepsilon \text{ and } K_2 \subset K_1^\varepsilon\big\}, \end{aligned} \tag{1.5}$$

where as before $d(x,A) := \inf_{y\in A} d(x,y)$ denotes the distance between a point $x \in E$ and a set $A \subset E$ and $A^\varepsilon := \big\{x \in \mathcal{X} : d(x,A) < \varepsilon\big\}$. The corresponding topology is naturally called the *Hausdorff topology*. Note the subtle difference between "the Hausdorff topology" (the topology generated by the Hausdorff metric) and "a Hausdorff topology" (any topology satisfying condition (iv) of Section 1.1). We extend this topology to $\mathcal{K}(E)$ by adding $\emptyset$ as an isolated point.

A good source for the Hausdorff topology is [SSS14, Appendix B], where one can find the proofs of all the lemmas in this section. Some more information can be found in [BBI01, Chapter 7]. The first lemma of this section shows that the Hausdorff topology depends only on the topology on E, and not on the choice of the metric.

Lemma 1.16 (Convergence criterion) *Let $K_n, K \in \mathcal{K}_+(E)$ $(n \geq 1)$. Then $K_n \to K$ in the Hausdorff topology if and only if there exists a $C \in \mathcal{K}_+(E)$ such that $K_n \subset C$ for all $n \geq 1$ and*

$$\begin{aligned} K &= \{x \in E : \exists x_n \in K_n \text{ s.t. } x_n \to x\} \\ &= \{x \in E : \exists x_n \in K_n \text{ s.t. } x \text{ is a cluster point of } (x_n)_{n\in\mathbb{N}}\}. \end{aligned} \tag{1.6}$$

The following lemma shows that $\mathcal{K}(E)$ is Polish if E is.

Lemma 1.17 (Properties of the Hausdorff metric)

(a) *If (E,d) is separable, then so is $(\mathcal{K}_+(E), d_{\mathrm{H}})$.*

(b) *If (E,d) is complete, then so is $(\mathcal{K}_+(E), d_{\mathrm{H}})$.*

The following lemma shows in particular that $\mathcal{K}(E)$ is compact if E is compact.

Lemma 1.18 (Compactness in the Hausdorff topology) *A set $\mathcal{A} \subset \mathcal{K}(E)$ is precompact if and only if there exists a $C \in \mathcal{K}(E)$ such that $K \subset C$ for each $K \in \mathcal{A}$.*

The following lemma is useful when proving convergence of $\mathcal{K}(E)$-valued random variables.

Lemma 1.19 (Tightness criterion) *Assume that E is a Polish space and let K_n $(n \geq 1)$ be $\mathcal{K}(E)$-valued random variables. Then the collection of laws $\{\mathbb{P}[K_n \in \cdot\,] : n \geq 1\}$ is tight if and only if for each $\varepsilon > 0$ there exists a compact $C \subset E$ such that $\mathbb{P}[K_n \subset C] \geq 1 - \varepsilon$ for all $n \geq 1$.*

1.7 Squeezed space

Let (E, d) be a metric space, let $\{*\}$ be a set containing a single element called $*$, and let

$$\mathcal{R}(E) := \big(E \times \mathbb{R}\big) \cup \big\{(*, -\infty), (*, +\infty)\big\}. \tag{1.7}$$

We extend d to $E \cup \{*\}$ by setting $d(x, *) = d(*, x) := \infty$ if $x \neq *$ and $:= 0$ otherwise. Let $\overline{\mathbb{R}} := [-\infty, \infty]$ denote the usual two-point compactification of the real line. We fix a continuous function $\phi : \overline{\mathbb{R}} \to [0, \infty)$ such that $\phi(t) > 0$ for all $t \in \mathbb{R}$ and $\phi(\pm\infty) = 0$, we choose a metric $d_{\overline{\mathbb{R}}}$ that generates the topology on $\overline{\mathbb{R}}$, and we define $\rho : \mathcal{R}(E)^2 \to [0, \infty)$ by

$$\rho\big((x, s), (y, t)\big) := \big(\phi(s) \wedge \phi(t)\big)\big(d(x, y) \wedge 1\big) + \big|\phi(s) - \phi(t)\big| + d_{\overline{\mathbb{R}}}(s, t) \tag{1.8}$$

Lemma 1.20 (Metric on squeezed space) *The function ρ is a metric on $\mathcal{R}(E)$.*

Proof For brevity, we write $d'(x, y) := d(x, y) \wedge 1$. Then d' is a metric on E. The only nontrivial statement that we have to prove is the triangle inequality, and it suffices to prove this for the function

$$\rho'\big((x, s), (y, t)\big) := \big(\phi(s) \wedge \phi(t)\big)d'(x, y) + \big|\phi(s) - \phi(t)\big|.$$

We estimate

$$\rho'\big((x, s), (z, u)\big) \leq \big(\phi(s) \wedge \phi(u)\big)\big(d'(x, y) + d'(y, z)\big) + \big|\phi(s) - \phi(u)\big|. \tag{1.9}$$

If $\phi(t) \geq \phi(s) \wedge \phi(u)$, then $\phi(s) \wedge \phi(u)$ is less than $\phi(s) \wedge \phi(t)$ and also less than $\phi(t) \wedge \phi(u)$, so we can simply estimate the expression in (1.9) from above by

$$\big(\phi(s) \wedge \phi(t)\big)d'(x, y) + \big(\phi(t) \wedge \phi(u)\big)d'(y, z)\big) + \big|\phi(s) - \phi(t)\big| + \big|\phi(t) - \phi(u)\big|$$

and we are done. On the other hand, if $\phi(t) < \phi(s) \wedge \phi(u)$, then

$$\big|\phi(s) - \phi(t)\big| + \big|\phi(t) - \phi(u)\big| = \big|\phi(s) - \phi(u)\big| + 2\big(\phi(s) \wedge \phi(u) - \phi(t)\big).$$

Using the fact that $d' \leq 1$, we can now estimate the right-hand side of (1.9) from above by

$$\begin{aligned}
&\phi(t)\big(d'(x,y)+d'(y,z)\big)+2\big(\phi(s)\wedge\phi(u)-\phi(t)\big)+\big|\phi(s)-\phi(u)\big| \\
&\quad=\big(\phi(s)\wedge\phi(t)\big)d'(x,y)+\big(\phi(t)\wedge\phi(u)\big)d'(y,z) \\
&\qquad+\big|\phi(s)-\phi(t)\big|+\big|\phi(t)-\phi(u)\big|,
\end{aligned}$$

and again we are done. ∎

The following lemma shows that the topology generated by the metric ρ depends only on the topology on E and not on the choice of the metric on E. Recall that by (1.1), a metrisable topology is uniquely characterised by its convergent sequences, so the topology on $\mathcal{R}(E)$ is uniquely characterised by the conditions (i) and (ii) below.

Lemma 1.21 (Topology on squeezed space) *A sequence $(x_n,t_n)\in\mathcal{R}(E)$ converges to a limit (x,t) in the metric ρ defined in (1.8) if and only if the following two conditions are satisfied:*

(i) *$t_n\to t$ in the topology on $\overline{\mathbb{R}}$,*

(ii) *if $t\in\mathbb{R}$, then $x_n\to x$ in the topology on E.*

Proof This is immediate from the definition of ρ. ∎

We can think of the space $\mathcal{R}(E)$ as being obtained from $E\times\overline{\mathbb{R}}$ by squeezing the sets $E\times\{\pm\infty\}$ into the single points $(*,\pm\infty)$. For this reason, we call $\mathcal{R}(E)$ the *squeezed space*. In the special case that $E=\overline{\mathbb{R}}$, we can make a picture of $\mathcal{R}(\overline{\mathbb{R}})$ by mapping $\overline{\mathbb{R}}\times\mathbb{R}$ into the closed unit disc using the function

$$(x,t)\mapsto\big(\sqrt{1-\psi(t)^2}\,\psi(x),\psi(t)\big)\quad\text{with}\quad\psi(z):=\frac{z}{1+|z|}$$

(with $\psi(\pm\infty):=\pm1$), and mapping the points $(*,\pm\infty)$ to $(0,\pm1)$. The following lemma shows that $\mathcal{R}(E)$ is a Polish space if E is Polish.

Lemma 1.22 (Properties of squeezed space)

(a) *If (E,d) is separable, then so is $(\mathcal{R}(E),\rho)$.*

(b) *If (E,d) is complete, then so is $(\mathcal{R}(E),\rho)$.*

Proof If D is a countable dense subset of (E,d), then $D\times\mathbb{Q}$ is a countable dense subset of $(\mathcal{R}(E),\rho)$, proving (a).

To prove (b), let (x_n,t_n) be a Cauchy sequence in $(\mathcal{R}(E),\rho)$. Then by (1.8) t_n is a Cauchy sequence in $\overline{\mathbb{R}}$ and hence $t_n\to t$ for some $t\in\overline{\mathbb{R}}$. If $t\in\mathbb{R}$, then by (1.8) x_n is a Cauchy sequence in (E, d) so by the completeness of the latter, $x_n \to x$ for some $x \in E$. By Lemma 1.21, it follows that (x_n, t_n) converges, proving the completeness of $(\mathcal{R}(E), \rho)$. ∎

The following lemma identifies the compact subsets of $\mathcal{R}(E)$. In particular, the lemma shows that $\mathcal{R}(E)$ is compact if E is compact.

Lemma 1.23 (Compactness criterion) *A set $A \subset \mathcal{R}(E)$ is precompact if and only if for each $T < \infty$, there exists a compact set $K \subset E$ such that $\{x \in E : (x, t) \in A,\ t \in [-T, T]\} \subset K$.*

Proof Assume that $A \subset \mathcal{R}(E)$ has the property that for each $T < \infty$, there exists a compact set $K \subset E$ such that $\{x \in E : (x, t) \in A,\ t \in [-T, T]\} \subset K$. To show that A is precompact, we will show that each sequence $(x_n, t_n) \in A$ has a convergent subsequence. By the compactness of $\overline{\mathbb{R}}$, we can select a subsequence (x'_n, t'_n) such that $t'_n \to t$ for some $t \in \overline{\mathbb{R}}$. If $t = \pm\infty$, then by Lemma 1.21 $(x'_n, t'_n) \to (*, \pm\infty)$ and we are done. Otherwise, there exists a $T < \infty$ such that $t'_n \in [-T, T]$ for all n large enough. By assumption, there then exists a compact set $K \subset E$ such that $x'_n \in K$ for all n large enough, so we can select a further subsequence such that (x''_n, t''_n) converges to a limit $(x, t) \in E \times \mathbb{R}$.

Assume, on the other hand, that $A \subset \mathcal{R}(E)$ has the property that for some $T < \infty$, there does not exist a compact set $K \subset E$ such that $\{x \in E : (x, t) \in A,\ t \in [-T, T]\} \subset K$. Set

$$B := \big\{x \in E : (x, t) \in A \text{ for some } t \in [-T, T]\big\}$$

The closure of B cannot be compact, since this would contradict our assumption. It follows that there exists a sequence $x_n \in B$ that does not contain a convergent subsequence, and there exist $t_n \in [-T, T]$ such that $(x_n.t_n) \in A$. But then, in view of Lemma 1.21, the sequence (x_n, t_n) cannot contain a convergent subsequence either, proving that A is not precompact. ∎

1.8 Path space

Let E be a metrisable space and let $\mathcal{R}(E)$ be the squeezed space defined in Section 1.7. By definition, a *path* in E is a nonempty compact subset $\pi \subset \mathcal{R}(E)$ such that $\{x \in E : (x, t) \in \pi\}$ has at most one element for each given $t \in \overline{\mathbb{R}}$. The set $\overline{I}_\pi := \big\{t \in \overline{\mathbb{R}} : \exists x \in E \text{ s.t. } (x, t) \in \pi\big\}$ is called the *domain* of π and

$$\sigma_\pi := \inf \overline{I}_\pi \quad \text{and} \quad \tau_\pi := \sup \overline{I}_\pi \tag{1.10}$$

the *starting time* and *final time* of the path π. For each $t \in \overline{I}_\pi$, we let $\{\pi(t)\} := \{x \in E : (x,t) \in \pi\}$ denote the unique point $\pi(t) \in E$ such that $(\pi(t),t) \in \pi$. Then $t \mapsto \pi(t)$ is a function from $\overline{I}_\pi$ to E. We let $\Pi(E)$ denote the set of all paths in E and set $I_\pi := \overline{I}_\pi \cap \mathbb{R}$.

Lemma 1.24 (Path viewed as a function) *The domain $\overline{I}_\pi$ of a path $\pi \in \Pi(E)$ is a closed subset of $\overline{\mathbb{R}}$, and $t \mapsto \pi(t)$ is a continuous function from I_π to E. Conversely, if $I \subset \mathbb{R}$ is closed and $t \mapsto f(t)$ is a continuous function from I to E, then there exists a path $\pi \in \Pi(E)$ such that $I_\pi = I$ and $\pi(t) = f(t)$ $(t \in I)$.*

Proof We first show that for each $\pi \in \Pi(E)$, the function $I_\pi \ni t \mapsto \pi(t)$ is continuous. Assume that $t_n, t \in I_\pi$ and $t_n \to t$. Since π is compact, the sequence $(\pi(t_n), t_n)$ is precompact. Since $\pi(t)$ is the only element of $\{x \in E : (x,t) \in \pi\}$, each subsequence of the $(\pi(t_n), t_n)$ must converge to $(\pi(t), t)$. By Lemma 1.2, we conclude that $(\pi(t_n), t_n) \to (\pi(t), t)$. Since $t \in \mathbb{R}$, by Lemma 1.21, we conclude that $\pi(t_n) \to \pi(t)$, which shows that $I_\pi \ni t \mapsto \pi(t)$ is continuous on I as claimed.

Let $I \subset \mathbb{R}$ be closed and let $f : I \to E$ be continuous. If I is nonempty, then let $\overline{I}$ be the closure of I in $\overline{\mathbb{R}}$, and set $\overline{I} := \{\infty\}$ otherwise. Extend f to $\overline{I}$ by setting $f(t) := *$ if $t = \pm\infty$. Let $\pi := \{(f(t), t) : t \in \overline{I}\}$. It follows from Lemma 1.21 and the continuity of f that the map

$$\overline{I} \ni t \mapsto \big(f(t), t\big) \in \mathcal{R}(E) \tag{1.11}$$

is continuous. Since $\overline{I}$ is compact and since π is the image of $\overline{I}$ under the continuous map (1.11), we conclude that π is compact. Clearly, $\{x \in E : (x,t) \in \pi\}$ has precisely one element for $t \in \overline{I}$, and is empty for $t \notin \overline{I}$. This shows that $\pi \in \Pi(E)$. ∎

In view of Lemma 1.24, we often view a path $\pi \in \Pi(E)$ as a continuous function defined on a closed domain $I_\pi \subset \mathbb{R}$. The correspondence between paths and continuous functions is almost one-to-one. The only ambiguity arises when $-\infty$ and/or $+\infty$ are not elements of the closure of I_π, and we have the choice whether to include them in $\overline{I}_\pi$ or not. If I_π is nonempty, then it is natural to include $\pm\infty$ only when they are elements of the closure of I_π. With this convention, if $I \subset \mathbb{R}$ is a closed nonempty interval, then we identify the space $\mathcal{C}_I(E)$ defined in Section 1.5 with the set $\{\pi \in \Pi(E) : \overline{I}_\pi = \overline{I}\}$, where $\overline{I}$ denotes the closure of I in $\overline{\mathbb{R}}$.

Let $\mathcal{K}(\mathcal{R}(E))$ be the set of compact subsets of the squeezed space $\mathcal{R}(E)$. We equip $\mathcal{K}(\mathcal{R}(E))$ with the Hausdorff topology. We observe that $\Pi(E)$ is a subset of $\mathcal{K}(\mathcal{R}(E))$. We naturally equip $\Pi(E)$ with the induced topology from its embedding in $\mathcal{K}(\mathcal{R}(E))$.

Lemma 1.25 (Paths with a fixed domain) *Let $I \subset \mathbb{R}$ be a closed nonempty interval. The induced topology on $\mathcal{C}_I(E)$ from its embedding in $\Pi(E)$ is the topology of locally uniform convergence.*

Proof Assume that $\pi_n, \pi \in \mathcal{C}_I(E)$, viewed as functions, satisfy $\pi_n \to \pi$ locally uniformly. We need to show that viewed as compact subsets of $\mathcal{R}(E)$, the sets π_n, π satisfy $\pi_n \to \pi$ in the Hausdorff topology on $\mathcal{K}(\mathcal{R}(E))$. Let $\overline{I}$ denote the closure of I in $\overline{\mathbb{R}}$. By Lemma 1.16, we need to show that $\bigcup_n \pi_n$ is precompact and

$$\begin{aligned}
&\pi \subset \big\{(x,t) \in \mathcal{R}(E) : \exists t_n \in \overline{I} \text{ s.t. } \big(\pi_n(t_n), t_n\big) \to (x,t)\big\}, \\
&\big\{(x,t) \in \mathcal{R}(E) : (x,t) \text{ is a cluster} \\
&\qquad\qquad \text{point of } \big(\pi_n(t_n), t_n\big) \text{ for some } t_n \in \overline{I}\big\} \subset \pi.
\end{aligned} \tag{1.12}$$

To see that $\bigcup_n \pi_n$ is precompact, we need to show that each sequence of the form $(\pi_{n(m)}(t_m), t_m)_{m\geq 1}$ has a convergent subsequence. If $n(m)$ infinitely often takes the same value n, then the claim is obvious from the compactness of π_n, so without loss of generality we may assume that $n(m) \to \infty$. Going to a subsequence if necessary, we may assume that $t_m \to t$ for some $t \in \overline{I}$. If $t = \pm\infty$, then the claim is again obvious so we may assume that $t \in I$. In this case Lemma 1.15 (ii) tells us that $\pi_{n(m)}(t_m) \to \pi(t)$ so we have found a convergent subsequence as required.

To prove the first inclusion in (1.12), let $(\pi(t), t) \in \pi$ and set $t_n := t$ for all n. If $t \in I$, then $\pi_n(t) \to \pi(t)$ since locally uniform convergence implies pointwise convergence, and if $t = \pm\infty$ then trivially $(*, t) \to (*, t)$ as $n \to \infty$. To prove the second inclusion, assume that $(\pi_{n(m)}(t_{n(m)}), t_{n(m)}) \to (x,t)$ as $m \to \infty$ for some $(x,t) \in \mathcal{R}(E)$, $t_n \in \overline{I}$, and $n(m) \to \infty$. If $t \in I$, then we can use Lemma 1.15 (ii) which tells us that $\pi_{n(m)}(t_{n(m)}) \to \pi(t)$ and hence $(x,t) = (\pi(t), t) \in \pi$. If $t = \pm\infty$, then trivially $x = *$ and $(*, t) \in \pi$.

Assume, conversely, that $\pi_n \to \pi$ in the Hausdorff topology on $\mathcal{K}(\mathcal{R}(E))$. We need to show that $\pi_n, \pi \in \mathcal{C}_I(E)$ and that $\pi_n \to \pi$ locally uniformly. Assume that $t_n, t \in I$ such that $t_n \to t$. By Lemma 1.15 (ii), it suffices to show that $\pi_n(t_n) \to \pi(t)$ for all such t_n, t. Equivalently, we may show that $(\pi_n(t_n), t_n) \to (\pi(t), t)$. By Lemma 1.2, it suffices to show that the set $\{(\pi_n(t_n), t_n) : n \in \mathbb{N}\}$ is precompact and $(\pi(t), t)$ is the only cluster point of the sequence $(\pi_n(t_n), t_n)$. By Lemma 1.16, there exists a compact set $C \subset \mathcal{R}(E)$ such that $\pi_n \subset C$ for all n, so $\{(\pi_n(t_n), t_n) : n \in \mathbb{N}\}$ is precompact as required. Let (x,t) be any cluster point. By Lemma 1.16 (ii), $(x,t) \in \pi$ and hence $x = \pi(t)$, which shows that $\pi_n(t_n) \to \pi(t)$ as required. ∎

Let $\pi \in \Pi(\mathbb{R}^d)$. Assume that $\overline{I}_\pi$ is the closure of I_π in $\overline{\mathbb{R}}$. Recall that σ_π and τ_π denote the starting time and final time of π. For each $t \in [\sigma_\pi, \tau_\pi] \cap \mathbb{R}$, let us write

$$\lfloor t \rfloor := \sup\{s \in I_\pi : s \leq t\} \quad \text{and} \quad \lceil t \rceil := \inf\{u \in I_\pi : t \leq u\}.$$

We define a *linearly interpolated* path $\hat{\pi}$ with domain $\overline{I}_{\hat{\pi}} := [\sigma_\pi, \tau_\pi]$ by $\hat{\pi}(t) := \pi(t)$ $(t \in \overline{I}_\pi)$ and

$$\hat{\pi}(t) := \frac{\lceil t \rceil - t}{\lceil t \rceil - \lfloor t \rfloor}\pi\big(\lfloor t \rfloor\big) + \frac{t - \lfloor t \rfloor}{\lceil t \rceil - \lfloor t \rfloor}\pi\big(\lceil t \rceil\big) \qquad (t \in [\sigma_\pi, \tau_\pi]\backslash\overline{I}_\pi).$$

Note that this is well-defined because our assumption that $\overline{I}_\pi$ is the closure of I_π in $\overline{\mathbb{R}}$ ensures that $\lfloor t \rfloor$ and $\lceil t \rceil$ are finite for all $t \in [\sigma_\pi, \tau_\pi]\backslash\overline{I}_\pi$. It often happens that a sequence of functions $f_n : \mathbb{N} \to \mathbb{R}^d$ converges, after a rescaling of time, to a continuous limit $f : [0, \infty) \to \mathbb{R}^d$. To formulate this properly, it is a common habit to linearly interpolate the functions f_n so that all functions are elements of the space $\mathcal{C}_{[0,\infty)}(\mathbb{R}^d)$. As the following exercise shows, when one uses the path space $\Pi(\mathbb{R}^d)$, no interpolation is necessary to formulate the result.

Exercise 1.26 (Convergence of interpolated paths) *Let that $\pi, \pi_n \in \Pi(\mathbb{R}^d)$. Assume that $\overline{I}_\pi$ is the closure of I_π and that $\overline{I}_{\pi_n}$ is the closure of I_{π_n} for each n. Show that $\pi_n \to \pi$ in the topology on $\Pi(\mathbb{R}^d)$ if and only if $\hat{\pi}_n \to \hat{\pi}$.*

Sometimes, when formulating convergence of a sequence of functions f_n to a limit f, one extrapolates with the aim of ensuring that all functions are defined on the same space. Let E be a metrisable space and for each $\pi \in \Pi(E)$, let π^+ denote the path with domain $\overline{I}_{\pi^+} := \overline{I}_\pi \cup [\tau_\pi, \infty]$ defined as $\pi^+(t) := \pi(t)$ if $t \in \overline{I}_\pi$ and

$$\pi^+(t) := \pi(\tau_\pi) \quad (\tau_\pi < t < \infty) \quad \text{and} \quad \pi^+(\infty) := *.$$

The next exercise shows that when one uses the path space $\Pi(E)$, no extrapolation is necessary.

Exercise 1.27 (Convergence of extrapolated paths) *Let $\pi_n, \pi \in \Pi(E)$. Show that the following conditions are equivalent:*

(i) $\pi_n \to \pi$

(ii) $\pi_n^+ \to \pi^+$ *and* $\tau_{\pi_n} \to \tau_\pi$.

Our next proposition says that the space of paths in E is Polish provided E has this property.

Proposition 1.28 (Polish space) *If E is a Polish space, then so is $\Pi(E)$.*

The proof of Proposition 1.28 needs some preparations. Let d be a metric generating the topology on E and let $\pi \in \Pi(E)$. For each $\pi \in \Pi(E)$, $\delta > 0$ and $T < \infty$, we define

$$m_{T,\delta}(\pi) := \sup\big\{d\big(\pi(s), \pi(t)\big) : s, t \in I_\pi,\ -T \leq s \leq t \leq T,\ t - s \leq \delta\big\}. \tag{1.13}$$

The quantity $m_{T,\delta}(\pi)$ is called the *modulus of continuity* of the path π. More generally, for any compact subset $K \subset \mathcal{R}(E)$, we can define

$$m_{T,\delta}(K) := \sup\big\{d\big(x, y\big) : (x, s), (y, t) \in K,\ -T \leq s \leq t \leq T,\ t - s \leq \delta\big\},$$

which coincides with our previous definition if π is a path.

Lemma 1.29 (Characterisation of paths) *A compact subset $\pi \subset \mathcal{R}(E)$ is an element of the path space $\Pi(E)$ if and only if $\lim_{\delta\to 0} m_{T,\delta}(\pi) = 0$ for all $T < \infty$.*

Proof Assume that $\pi \in \mathcal{K}(\mathcal{R}(E))$ and $\limsup_{\delta\to 0} m_{T,\delta}(\pi) > 0$ for some $T < \infty$. Then we can find $(x_n, s_n), (y_n, t_n) \in \pi$ and $\delta > 0$ with $d(x_n, y_n) \geq \delta$, $-T \leq s_n \leq t_n \leq T$, and $t_n - s_n \leq 1/n$. Since π is compact, by going to a subsequence, we can assume that $(x_n, s_n) \to (x, s)$ and $(y_n, t_n) \to (y, t)$ for some $(x, s), (y, t) \in \pi$ with $d(x, y) \geq \delta > 0$, $-T \leq s \leq t \leq T$, and $t - s = 0$. This shows that $\pi \notin \Pi(E)$.

Conversely, if $\pi \notin \Pi(E)$, then there exist $(x, t), (y, t) \in \pi$ with $x \neq y$. Since $(*, \pm\infty)$ are the only points in $\mathcal{R}(E)$ with time coordinate $\pm\infty$ we must have $t \in \mathbb{R}$. But then $m_{T,\delta}(\pi) \geq d(x, y) > 0$ for all $T \geq |t|$, which shows that $\limsup_{\delta\to 0} m_{T,\delta}(\pi) > 0$ for some $T < \infty$. ∎

Proof of Proposition 1.28 If E is a Polish space, then by Lemma 1.22 so is $\mathcal{R}(E)$ and hence by Lemma 1.17 so is $\mathcal{K}(\mathcal{R}(E))$. For each $\varepsilon, \delta > 0$ and $T < \infty$, the set

$$A_{T,\varepsilon,\delta} := \{K \in \mathcal{K}(\mathcal{R}(E)) : m_{T,\delta}(K) \geq \varepsilon\}$$

is a closed subset of $\mathcal{K}(\mathcal{R}(E))$ and hence its complement $A^{\rm c}_{T,\varepsilon,\delta}$ is open. By Lemma 1.29,

$$\Pi(E) = \bigcap_{n,m} \bigcup_k A^{\rm c}_{n,1/m,1/k},$$

which is a countable intersection of open sets, i.e., a G_δ-set. ∎

A set $\mathcal{A} \subset \Pi(E)$ is called *equicontinuous* if

$$\lim_{\delta\to 0} \sup_{\pi\in\mathcal{A}} m_{T,\delta}(\pi) = 0 \qquad (T < \infty).$$

The following theorem identifies the compact subsets of $\Pi(E)$. Condition (ii) is called the *compact containment* condition. If $I \subset \mathbb{R}$ is a closed nonempty interval, then $\mathcal{C}_I(E)$ is a closed subset of Π and hence the following theorem can also be used to identify the precompact subsets of $\mathcal{C}_I(E)$. In this context, the result is known as the *Arzela-Ascoli theorem.* Note that while the definition of equicontinuity depends (at least a priori) on the choice of the metric d on E, whether a set $\mathcal{A} \subset \Pi(E)$ is precompact only depends on the topology on E, so when verifying conditions (i) and (ii) below, we are free to choose any metric d that generates the topology on E.

Theorem 1.30 (Arzela-Ascoli) *A set $\mathcal{A} \subset \Pi(E)$ is precompact if and only if*

(i) *$\mathcal{A}$ is equicontinuous,*

(ii) *for each $T < \infty$, there exists a compact set $K \subset E$ such that $\big\{\pi(t) : \pi \in \mathcal{A},\ t \in I_\pi \cap [-T,T]\big\} \subset K$.*

Proof Let $\overline{\mathcal{A}}$ denote the closure of $\mathcal{A}$, viewed as a subset of the space $\mathcal{K}(\mathcal{R}(E))$, equipped with the Hausdorff topology. Then $\mathcal{A}$ is a precompact subset of $\Pi(E)$ if and only if $\overline{\mathcal{A}}$ is a compact subset of $\mathcal{K}(\mathcal{R}(E))$ and $\overline{\mathcal{A}} \subset \Pi(E)$. By Lemmas 1.18 and 1.23, $\overline{\mathcal{A}}$ is a compact subset of $\mathcal{K}(\mathcal{R}(E))$ if and only if condition (ii) holds. To complete the proof, it suffices to show that assuming that (ii) holds, one has $\overline{\mathcal{A}} \subset \Pi(E)$ if and only if (i) holds.

We first show that (i) implies $\overline{\mathcal{A}} \subset \Pi(E)$. Assume that $\pi_n \in \mathcal{A}$ converge in the Hausdorff topology to a compact subset $\pi \subset \mathcal{R}(E)$. To show that $\pi \in \Pi(E)$, will apply Lemma 1.29. If $(x,s),(y,t) \in \pi$, then by Lemma 1.16, there exist $(x_n,s_n),(y_n,t_n) \in \pi_n$ such that $(x_n,s_n) \to (x,s)$ and $(y_n,t_n) \to (y,t)$. If $s,t \in [-T,T]$ and $|t-s| \le \delta$, then for n large enough we have $s_n,t_n \in [-T-1,T+1]$ and $|t_n - s_n| \le 2\delta$. Since $d(x_n,y_n) \to d(x,y)$, it follows that

$$\limsup_{\delta \to 0} m_{T,\delta}(\pi) \le \limsup_{\delta \to 0} \sup_n m_{T+1,2\delta}(\pi_n) = 0 \qquad (\delta > 0,\ T < \infty),$$

which by Lemma 1.29 implies that $\pi \in \Pi(E)$.

Assume now that (ii) holds but (i) fails. Then there exist $T < \infty$ and $\varepsilon > 0$ such that for each $n \ge 1$, we can find $\pi_n \in \mathcal{A}$ with $m_{T,1/n}(\pi_n) \ge \varepsilon$. This means that there exist $-T \le s_n \le t_n \le T$ such that $d(\pi_n(s_n),\pi_n(t_n)) \ge \varepsilon$ and $t_n - s_n \le 1/n$. By (ii), $\overline{\mathcal{A}}$ is a compact subset of $\mathcal{K}(\mathcal{R}(E))$, so by going a subsequence we may assume that $\pi_n \to \pi \in \mathcal{K}(\mathcal{R}(E))$. By going to a further subsequence, we may assume that $s_n \to s$ and $t_n \to t$ for some $s,t \in [-T,T]$. But then $s = t$ since $t_n - s_n \le 1/n$. Let $x_n := \pi_n(s_n)$ and $y_n := \pi_n(t_n)$. By (ii), we can select a further subsequence such that $x_n \to x$ and $y_n \to y$ for some x, y with $d(x, y) \geq \varepsilon$. By Lemma 1.16, we have $(x, t), (y, t) \in \pi$ which shows that $\pi \notin \Pi(E)$ and hence $\overline{\mathcal{A}}$ is not a subset of $\Pi(E)$. ∎

For real-valued paths, the compact containment condition of Theorem 1.30 can be relaxed.

Theorem 1.31 (Arzela-Ascoli - real version) *Assume that $\pi_n \in \Pi(\mathbb{R})$ satisfy:*

(i) $\{\pi_n : n \in \mathbb{N}\}$ *is equicontinuous,*

(ii) *there exist $t_n \in I_{\pi_n}$ such that $\sup_n |t_n| < \infty$ and a compact set $K \subset \mathbb{R}$ such that $\pi_n(t_n) \in K$ for all n.*

Then $\{\pi_n : n \in \mathbb{N}\}$ is a precompact subset of $\Pi(\mathbb{R})$.

Proof For any set $A \subset \mathbb{R}$ and $r \geq 0$, we write $A^r := \{x \in \mathbb{R} : \inf_{y \in A} |x - y| \leq r\}$. Then A^r is a closed subset of $\mathbb{R}$. If A is compact, then so is A^r.

To prove the claim of the theorem, it suffices to check condition (ii) of Theorem 1.30. It suffices to check this for T sufficiently large, so without loss of generality, we can assume that $t_n \in [-T, T]$ for all n. Fix $\varepsilon > 0$. By equicontinuity, we can choose $\delta > 0$ such that $|\pi_n(s) - \pi_n(t)| \leq \varepsilon$ for all n and $s, t \in I_{\pi_n} \cap [-T, T]$ with $|s - t| \leq \delta$. Let K be the compact set from condition (ii) above. Then $\pi_n(t) \in K^\varepsilon$ for all $t \in I_{\pi_n}$ such that $|t - t_n| \leq \delta$, and by induction, for each $k \geq 1$, we obtain that $\pi_n(t) \in K^{k\varepsilon}$ for all $t \in I_{\pi_n}$ such that $|t - t_n| \leq k\delta$. Choosing k large enough such that $\delta k \geq 2T$, we see that $\big\{\pi_n(t) : n \in \mathbb{N},\ t \in I_{\pi_n} \cap [-T, T]\big\} \subset K^{k\delta}$, so condition (ii) of Theorem 1.30 is satisfied. ∎

Chapter 2

The excursion

2.1 Scaling limit of simple random walk

Let $(X_k)_{k\geq 1}$ be i.i.d. and uniformly distributed on $\{-1,+1\}$, and let

$$S_n := \sum_{k=1}^{n} X_k \qquad (n \geq 0),$$

with naturally $S_0 := 0$. Then $(S_n)_{n\geq 0}$ is a one-dimensional nearest-neighbour random walk. It will be convenient to interpolate linearly. We set

$$S_t := \big(\lceil t\rceil - t\big)S_{\lfloor t\rfloor} + \big(t - \lfloor t\rfloor\big)S_{\lceil t\rceil} \qquad (t \geq 0).$$

Then $S = (S_t)_{t\geq 0}$ is a random variable taking in the space

$$\mathcal{C}_0 := \big\{f \in \mathcal{C}_{[0,\infty)}(\mathbb{R}) : f_0 = 0\big\}. \tag{2.1}$$

Donsker's invariance principle says that S, diffusively rescaled, converges to Brownian motion. To formulate this properly, for $\lambda > 0$, let $\theta_\lambda : \mathbb{R}^2 \to \mathbb{R}^2$ be the *diffusive scaling* map defined as

$$\theta_\lambda(x,t) := (\lambda x, \lambda^2 t) \qquad \big((x,t) \in \mathbb{R}^2\big). \tag{2.2}$$

which we extend to a (clearly unique) continuous map $\theta_\lambda : \mathcal{R}(\overline{\mathbb{R}}) \to \mathcal{R}(\overline{\mathbb{R}})$. For any subset $A \subset \mathcal{R}(\overline{\mathbb{R}})$, we let $\theta_\lambda A$ denote the image of A under θ_λ. In particular, we can apply this to S, which we can view as an element of the path space $\Pi(\mathbb{R})$ and hence as a compact subset of $\mathcal{R}(\mathbb{R})$. For each $\varepsilon > 0$, the *diffusively rescaled* path

$$S^\varepsilon := \theta_\varepsilon S \tag{2.3}$$

is then the random variable taking values in the space $\mathcal{C}_0$ defined as

$$S^\varepsilon_t := \varepsilon S_{\varepsilon^{-2}t} \qquad (t \in \varepsilon\mathbb{N}).$$

The following fact is well-known. Below, we naturally identify the path $(B_t)_{t\geq 0}$ of a Brownian motion with an element of the path space $\Pi(\mathbb{R})$.

Theorem 2.1 (Donsker's invariance principle) *One has*

$$\mathbb{P}\big[(S^\varepsilon_t)_{t\geq 0} \in \cdot\,\big] \underset{\varepsilon\to 0}{\Longrightarrow} \mathbb{P}\big[(B_t)_{t\geq 0} \in \cdot\,\big], \tag{2.4}$$

where $(B_t)_{t\geq 0}$ is a standard Brownian motion and $\Rightarrow$ denotes weak convergence of probability measures on $\mathcal{C}_0$, equipped with the topology of locally uniform convergence.

As we have seen in Exercise 1.26, to formulate Theorem 2.1, it was in fact not necessary to interpolate linearly. Instead, we can also view S as an element of the path space $\Pi(\mathbb{R})$ with domain $I_S = \mathbb{N}$ and then formulate Theorem 2.1, as weak convergence in law of random variables with values in $\Pi(\mathbb{R})$. However, in what follows, the linear interpolation will turn out to be convenient for other purposes.

Note that combining Donsker's invariance principle with Skorohod's representation theorem (Theorem 1.13), one obtains that if ε_n are positive constants tending to zero, then the random variables $(S^{\varepsilon_n}_t)_{t\geq 0}$ for different values of n can be coupled to a Brownian motion $(B_t)_{t\geq 0}$ in such a way that

$$\sup_{t\in[0,T]} \big|S^{\varepsilon_n}_t - B_t\big| \underset{n\to\infty}{\longrightarrow} 0 \quad \text{a.s.} \quad \forall T < \infty. \tag{2.5}$$

We conclude this section with a well-known fact.

Lemma 2.2 (Brownian scaling) *If $B = (B_t)_{t\geq 0}$ is a Brownian motion and $\lambda > 0$, then the process $B^\lambda := \theta_\lambda B$ is equally distributed with B.*

Proof This is of course well-known, but it is interesting to observe that it actually follows from Theorem 2.1. Indeed, the latter says that if $\varepsilon_n > 0$ converge to zero, then the processes $\theta_{\varepsilon_n} S$ converge weakly in law to B. Since the map $\Pi(\mathbb{R}) \ni \pi \mapsto \theta_\lambda \pi \in \Pi(\mathbb{R})$ is continuous, it follows that the processes $\theta_{\lambda\varepsilon_n} S$ converge weakly in law to B^λ. On the other hand, since $\varepsilon'_n := \lambda\varepsilon_n$ are positive constants tending to zero, Theorem 2.1 also tells us that the processes $\theta_{\lambda\varepsilon_n} S$ converge weakly in law to B, so B^λ and B must be equal in law. This proof reveals a general fact: a probability law that arises as the scaling limit of other probability laws must itself be scale invariant. ∎

2.2 Brownian local time

Recall from (2.1) that $\mathcal{C}_0$ is the space of continuous functions $f : [0,\infty) \to \mathbb{R}$ that satisfy $f_0 = 0$. We let

$$m_t(f) := \inf_{0\leq s\leq t} f_s \qquad \big(t \geq 0,\ f \in \mathcal{C}_0\big) \tag{2.6}$$

denote the *running minimum* of the function f. We will be interested in

$$g_t := f_t - m_t(f) \qquad (t \geq 0).$$

We observe that $g_t \geq 0$ and $h_t := -m_t(f)$ is a nondecreasing function that increases only at times when $g_t = 0$. The following lemma says that these properties characterise g and h uniquely. Note that if $h \in \mathcal{C}_0$ is nondecreasing, then it is the distribution function of a measure on $[0,\infty)$, which we denote by $\mathrm{d}h$. Condition (iii) below says that this measure is concentrated on the set $\{t \in [0,\infty) : g(t) = 0\}$. This makes precise the intuitive concept that h increases only at times when $g_t = 0$.

Lemma 2.3 (Skorohod reflection) *For each $f \in \mathcal{C}_0$, there exist unique functions $g, h \in \mathcal{C}_0$ such that*

(i) $g_t = f_t + h_t$ $(t \geq 0)$,

(ii) $g \geq 0$ *and h is nondecreasing,*

(iii) $\displaystyle\int_0^\infty 1_{\{g(t)>0\}}\mathrm{d}h(t) = 0.$

These functions are given by

$$g_t = f_t - m_t(f) \quad \textit{and} \quad h_t = -m_t(f) \qquad (t \geq 0). \tag{2.7}$$

Proof (sketch) It is not hard to check that if we define g and h by (2.7), then (i)–(iii) are satisfied. To prove uniqueness, it suffices to show that if g, h and g', h' both solve (i)–(iii), then $g' \leq g$. Imagine that $g'_t > g_t$ for some $t > 0$. Let $s := \sup\{u \in [0,t] : g'_u = g_u\}$. Then $g'_u > g_u$ for all $s < u \leq t$. By (i) we have $h'_s = h_s$. Now

$$g'_t - g_t = \big(f_t + h'_t\big) - \big(f_t + h_t\big) = h'_t - h_t. \tag{2.8}$$

By (ii) we have $g \geq 0$ and hence $g'_u > g_u \geq 0$ for all $s < u \leq t$, which by (iii) implies that $h'_t = h'_s$. On the other hand, by (ii) h is nondecreasing and hence $h_t \geq h_s$. It follows that the right-hand side of (2.8) is $\leq h'_s - h_s = 0$, which contradicts $g'_t > g_t$. ∎

We will especially be interested in the case that the function f from Lemma 2.3 is Brownian motion. In this case, the function g is reflected Brownian motion, and h is its local time at the origin. To explain this in a bit more detail, we need to take a small detour.

If $(B_t)_{t\geq 0}$ is a d-dimensional Brownian motion, then we can define a stochastic process $(\ell_t)_{t\geq 0}$ taking values in the space $\mathcal{M}(\mathbb{R}^d)$ of finite measures on $\mathbb{R}^d$ by

$$\int_{\mathbb{R}^d} \ell_t(\mathrm{d}x)\, f(x) := \int_0^t \mathrm{d}s\, f(B_s) \qquad \big(t \geq 0,\ f \in B_{\mathrm{b}}(\mathbb{R}^d)\big).$$

The random measure ℓ_t is called the *occupation local measure* of the Brownian motion $(B_t)_{t\geq 0}$. In particular

$$\ell_t(A) = \int_0^t \mathrm{d}s\, 1_A(B_s) \qquad \big(A \in \mathcal{B}(\mathbb{R}^d)\big)$$

is the amount of time the Brownian motion has spent inside a measurable set A up to time t. In one dimension, it is well-known that ℓ_t has a density with respect to the Lebesgue measure. The following theorem is originally due to Trotter. The process $(L_t)_{t\geq 0}$ below is called *Brownian local time.*

Theorem 2.4 (Brownian local time) *Let $(B_t)_{t\geq 0}$ be a one-dimensional Brownian motion. Then almost surely, there exists a random continuous function $L : [0,\infty) \times \mathbb{R} \to [0,\infty)$ such that*

$$\int_{\mathbb{R}} \mathrm{d}x\, L_t(x) f(x) = \int_0^t \mathrm{d}s\, f(B_s) \qquad \big(t \geq 0,\ f \in B_{\mathrm{b}}(\mathbb{R}^d)\big).$$

Modern proofs of Theorem 2.4 are based on *Tanaka's formula*, which says that

$$|B_t| = \int_0^t \mathrm{sgn}(B_s)\, \mathrm{d}B_s + L_t(0) \qquad (t \geq 0), \tag{2.9}$$

where the integral is an Itô stochastic integral. Tanaka's formula can be used as a definition of Brownian local time, for which one then proves the properties described in Theorem 2.4. For details, we refer to [McK69, Mey76, RW87]. In fact, in the remainder of this chapter, we will mostly work with Tanaka's formula as the definition of $L_t(0)$ and do not really need its interpretation as local time in the sense of Theorem 2.4.

Proposition 2.5 (Reflected Brownian motion) *Let $B = (B_t)_{t\geq 0}$ be a standard Brownian motion and let $(L_t(0))_{t\geq 0}$ be its local time at 0. Let $W = (W_t)_{t\geq 0}$ be another standard Brownian motion and let*

$$A_t := W_t - m_t(W) \quad \textit{and} \quad L_t := -m_t(W) \qquad (t \geq 0). \tag{2.10}$$

Then

$$\mathbb{P}\big[\big(|B_t|, L_t(0)\big)_{t\geq 0} \in \cdot\,\big] = \mathbb{P}\big[(A_t, L_t)_{t\geq 0} \in \cdot\,\big].$$

Proof (sketch) Let $(B_t)_{t\geq 0}$ be a Brownian motion and let

$$W_t := -\int_0^t \mathrm{sgn}(B_s)\,\mathrm{d}B_s \qquad (t \geq 0).$$

It is not hard to show that $W = (W_t)_{t\geq 0}$ is a Brownian motion. We will show that $A_t = |B_t|$ and $L_t = L_t(0)$ $(t \geq 0)$. We apply Lemma 2.3. Tanaka's formula (2.9) says that $|B_t| = L_t(0) - W_t$ $(t \geq 0)$. Clearly $|B_t|$ is nonnegative and $L_t(0)$ is nondecreasing and increases only when $|B_t| = 0$. For the details, we refer to [KS91, Thm 3.6.17]. ∎

2.3 Scaling limit of reflected random walk

Let S be the simple random walk defined in Section 2.1 and let $(R_t, K_t)_{t\geq 0}$ be defined by

$$R_t := S_t - m_t(S) \quad \text{and} \quad K_t := -m_t(S) \qquad (t \geq 0). \tag{2.11}$$

It is easy to see that $(R_t, K_t)_{t\in[0,\infty)}$ is the linear interpolation of the discrete time process $(R_n, K_n)_{n\in\mathbb{N}}$. Moreover, $(R_n)_{n\in\mathbb{N}}$ is a Markov chain with state space $\mathbb{N}$ and transition probabilities

$$P(x,y) = \mathbb{P}\big[R_n = y \,\big|\, R_{n-1} = x\big] \qquad (x, y \in \mathbb{N})$$

given by

$$\left.\begin{array}{l} P(x,x+1) = \frac{1}{2}, \\ P(x,x-1) = \frac{1}{2}, \end{array}\right\} \quad (x > 0) \quad \text{and} \quad \begin{array}{l} P(0,1) = \frac{1}{2}, \\ P(0,0) = \frac{1}{2}. \end{array}$$

In words, in each step, the process R_n jumps up or down by one with equal probabilities, except when this would result in a negative value, in which case the chain stays in 0. We call $(R_n)_{n\geq 0}$ a random walk with *reflection* at zero. It is now easy to see that the process K from (2.11) is given by

$$K_n = \sum_{k=1}^{n} 1_{\{R_{k-1} = R_k = 0\}} \qquad (n \geq 0).$$

Informally, K_n counts the number of times the chain $(R_n)_{n\geq 0}$ has "attempted to jump below zero", but was reflected. The following theorem says that the process (R, K) has a diffusive scaling limit.

Theorem 2.6 (Scaling limit of reflected random walk) *Let (R, K) be defined in (2.11) and for each $\varepsilon > 0$, let $(R^\varepsilon, K^\varepsilon)$ denote the diffusively rescaled process*

$$(R^\varepsilon_t, K^\varepsilon_t) := \big(\varepsilon R_{\varepsilon^{-2}t}, \varepsilon K_{\varepsilon^{-2}t}\big) \qquad (t \geq 0). \tag{2.12}$$

Let $B = (B_t)_{t\geq 0}$ be a standard Brownian motion and let $(L_t(0))_{t\geq 0}$ be its local time at 0. Then

$$\mathbb{P}\big[(R^\varepsilon_t, K^\varepsilon_t)_{t\geq 0} \in \cdot\,\big] \underset{\varepsilon\to 0}{\Longrightarrow} \mathbb{P}\big[\big(|B_t|, L_t(0)\big)_{t\geq 0} \in \cdot\,\big],$$

where $\Rightarrow$ denotes weak convergence of probability measures on $\mathcal{C}_{[0,\infty)}(\mathbb{R}^2)$, equipped with the topology of locally uniform convergence.

Proof We observe that

$$R^\varepsilon_t := S^\varepsilon_t - m_t(S^\varepsilon) \quad \text{and} \quad K^\varepsilon_t := -m_t(S^\varepsilon) \qquad (t \geq 0),$$

where S^ε is the diffusively rescaled random walk defined in (2.3). It is straightforward to check that the map

$$\mathcal{C}_0 \ni f \mapsto (g, h) \in \mathcal{C}_{[0,\infty)}(\mathbb{R}^2)$$

defined in (2.7) is continuous with respect to the topology of locally uniform convergence. Therefore, Theorem 2.1 implies that

$$\mathbb{P}\big[(R^\varepsilon_t, K^\varepsilon_t)_{t\geq 0} \in \cdot\,\big] \underset{\varepsilon\to 0}{\Longrightarrow} \mathbb{P}\big[(A_t, L_t)_{t\geq 0} \in \cdot\,\big],$$

where $(A_t, L_t)_{t\geq 0}$ is the reflected Brownian motion defined in (2.10). The claim now follows from Proposition 2.5. ∎

Theorem 2.6 yields the following useful consequence.

Lemma 2.7 (Scale invariance) *Let $(B_t)_{t\geq 0}$ be a standard Brownian motion and let $(L_t(0))_{t\geq 0}$ is its local time at* 0*. Then*

$$\mathbb{P}\big[\big(|B_t|, L_t(0)\big)_{t\geq 0} \in \cdot\,\big] = \mathbb{P}\big[\big(\lambda|B_{\lambda^{-2}t}|, \lambda L_{\lambda^{-2}t}(0)\big)_{t\geq 0} \in \cdot\,\big] \qquad (\lambda > 0).$$

Proof The proof is very similar to the proof of Lemma 2.2. As we observed there, a probability law that arises as the scaling limit of other probability laws must itself be scale invariant. In the present setting, we can make this general principle precise as follows. Fix $\lambda > 0$ and let ε_n be positive constants tending to zero. By (2.12),

$$(R^{\lambda\varepsilon}_t, K^{\lambda\varepsilon}_t) = (\lambda R^{\varepsilon}_{\lambda^{-2}t}, \lambda K^{\varepsilon}_{\lambda^{-2}t}) \qquad (t \geq 0),$$

so Theorem 2.6 tells us that

$$\mathbb{P}\big[(R^{\lambda\varepsilon}_t, K^{\lambda\varepsilon}_t)_{t\geq 0} \in \cdot\,\big] \underset{\varepsilon\to 0}{\Longrightarrow} \mathbb{P}\big[\big(\lambda|B_{\lambda^{-2}t}|, \lambda L_{\lambda^{-2}t}(0)\big)_{t\geq 0} \in \cdot\,\big].$$

However, $\lambda\varepsilon_n$ are positive constants tending to zero, so Theorem 2.6 also tells us that

$$\mathbb{P}\big[(R^{\lambda\varepsilon}_t, K^{\lambda\varepsilon}_t)_{t\geq 0} \in \cdot\,\big] \underset{\varepsilon\to 0}{\Longrightarrow} \mathbb{P}\big[\big(|B_t|, L_t(0)\big)_{t\geq 0} \in \cdot\,\big].$$

∎

2.4 Excursion decomposition

We will be interested in the theory of Brownian excursions. Our exposition is loosely inspired by [Rog89]. Recall from (2.1) that $\mathcal{C}_0$ is the space of continuous functions $f : [0,\infty) \to \mathbb{R}$ that satisfy $f_0 = 0$. We let

$$\begin{aligned}\mathcal{R}_0 := \big\{(g,h) : g, h \in \mathcal{C}_0,\ g \geq 0,\ h \text{ is nondecreasing},\\ \text{and} \quad \textstyle\int_0^\infty 1_{\{g(t)>0\}}\mathrm{d}h(t) = 0\big\},\end{aligned} \tag{2.13}$$

denote the set of pairs of functions (g,h) that satisfy conditions (ii) and (iii) of Lemma 2.3. We view $\mathcal{R}_0$ as a subset of $\mathcal{C}_{[0,\infty)}(\mathbb{R}^2)$ and equip it with the topology of locally uniform convergence. In Lemma 2.3, we have seen that setting $g_t := f_t - m_t(f)$ and $h_t := -m_t(f)$ defines a bijection $f \mapsto (g,h)$ from $\mathcal{C}_0$ to $\mathcal{R}_0$.

We now want to go one step further, and decompose the function g in excursions away from zero. Recall that σ_π and τ_π, defined in (1.10), denote the starting time and final time of a path π. We define a *space of excursions* by

$$\mathcal{E} := \big\{\pi \in \Pi([0,\infty)) : 0 \leq \tau_\pi < \infty,\ \overline{I}_\pi = [0,\tau_\pi],\ \pi(0) = \pi(\tau_\pi) = 0\big\}. \tag{2.14}$$

We call the final time τ_π of an excursion $\pi \in \mathcal{E}$ the *duration* of π. We observe that

$$\mathcal{F} := \{\pi : \overline{I}_\pi = [0, \tau_\pi],\ \pi(0) = 0,\ \text{and } \pi(\tau_\pi) = 0 \text{ if } \tau_\pi < \infty\}$$

is a closed subset of $\Pi([0, \infty))$ and $\mathcal{E}$ is an open subset of $\mathcal{F}$, so using Proposition 1.4 we see that $\mathcal{E}$ is a Polish space. We set

$$\mathring{\mathcal{E}} := \big\{\pi \in \mathcal{E} : \tau_\pi > 0,\ \pi(t) > 0\ \forall 0 < t < \tau_\pi\big\}. \tag{2.15}$$

We call elements of $\mathring{\mathcal{E}}$ *proper excursions.*

Let $h \in \mathcal{C}_0$ be nondecreasing. By definition, a *plateau* of h is an open interval $\iota = (\iota^-, \iota^+)$ with $0 \leq \iota^- < \iota^+ < \infty$ such that $h_{\iota^-} = h_{\iota^+}$, and no strictly larger open subinterval of $[0, \infty)$ has this property. We set

$$\mathcal{I}(h) := \big\{\iota : \iota \text{ is a plateau of } h\big\}. \tag{2.16}$$

For brevity, we write

$$h_\iota := h_{\iota^-} = h_{\iota^+} \qquad (\iota \in \mathcal{I}(h)).$$

For each $(g, h) \in \mathcal{R}_0$ and $\iota \in \mathcal{I}(h)$, setting

$$\tau_\iota := \iota^+ - \iota^- \quad \text{and} \quad \pi^{g,\iota}_t := g_{t-\iota^-} \quad (0 \leq t \leq \tau_\iota)$$

defines an excursion $\pi^{g,\iota} \in \mathcal{E}$ with duration τ_ι. Given a function $f \in \mathcal{C}_0$ and functions $(g, h) \in \mathcal{R}_0$ defined as in (2.7), we set

$$\Xi(f) := \big\{\big(h_\iota, \pi^{g,\iota}\big) : \iota \in \mathcal{I}(h)\big\}. \tag{2.17}$$

We will especially be interested in the case that f is a (diffusively rescaled) simple random walk, or Brownian motion. In this case, g is a reflected random walk or Brownian motion and h is its reflection local time at the origin. The set Ξ records all excursions of the reflected random walk or Brownian motion away from the origin together with the reflection local time when such an excursion happens.

It follows from the way we have defined plateaus that $h(\iota) \neq h(\iota')$ for each $\iota, \iota' \in \mathcal{I}(h)$ with $\iota \neq \iota'$. We use this to define a function $s \mapsto E^s$ from $[0, \infty)$ to $\mathcal{E}$ by

$$E^s = \begin{cases} \pi & \text{if } (s, \pi) \in \Xi \text{ for some } \pi \in \mathcal{E}, \\ o & \text{otherwise,} \end{cases} \tag{2.18}$$

where $o \in \mathcal{E}$ denotes the *trivial excursion* of duration $\tau_o := 0$.

The excursion set $\Xi(S)$ of the simple random walk S of Section 2.1 is easy to understand. Let $(R, K) = (R_t, K_t)_{t\geq 0}$ be the (linearly interpolated) reflected random walk defined in terms of S as in (2.11). We inductively define $\iota^\pm(k) \in \mathbb{N}$ ($k \in \mathbb{N}$) by $\iota^-(0) := 0$ and

$$\left.\begin{aligned} \iota^+(k) &:= \inf\big\{i \geq \iota^-(k) : K_{i+1} > K_i\big\}, \\ \iota^-(k+1) &:= \iota^+(k) + 1 \end{aligned}\right\} \qquad (k \geq 0).$$

For each $k \in \mathbb{N}$, we define an excursion E^k with duration τ_k by

$$\tau_k := \iota^+(k) - \iota^-(k) \quad \text{and} \quad E^k_t := R_{\iota^-(k)+t} \quad (0 \leq t \leq \tau_k). \tag{2.19}$$

Note that it may happen that $\tau_k = 0$, in which case $E^k = o$, the trivial excursion of duration zero. Then set of plateaus of the function K is

$$\mathcal{I}(K) = \big\{\big(\iota^-(k), \iota^+(k)\big) : k \geq 0,\ \iota^-(k) < \iota^+(k)\big\},$$

and the excursion set $\Xi(S)$ is given by

$$\Xi(S) = \big\{(k, E^k) : k \in \mathbb{N},\ E^k \neq o\big\}. \tag{2.20}$$

For the reflected random walk, not all excursions are proper excursions, since it may happen that $R_i = 0$ for some $\iota^-(k) < i < \iota^+(k)$. Since the process "starts anew" after each increase of K, it is easy to see that:

$$\text{The } \mathcal{E}\text{-valued random variables } (E^k)^{k\in\mathbb{N}} \text{ are i.i.d.} \tag{2.21}$$

For Brownian motion, the situation is more complex, since we can no longer enumerate the excursions by the time at which they occur. Nevertheless, something similar to the i.i.d. property of (2.21) still holds. The following theorem is due to Itô [Ito71]. The σ-finite measure ν below is called the *excursion measure.*

Theorem 2.8 (Poisson set of excursions) *There exists a σ-finite measure ν on $\mathcal{E}$ such that the set Ξ is a Poisson point set with intensity measure $\ell\otimes\nu$, where ℓ is the Lebesgue measure on $[0,\infty)$. The measure ν is concentrated on $\mathring{\mathcal{E}}$.*

As a preparation for the proof of Theorem 2.8, we make the following observation.

Lemma 2.9 (Only proper excursions) *The excursion set $\Xi(B)$ of a Brownian motion B is concentrated on the set of proper excursions $\mathring{\mathcal{E}}$.*

Proof To show that Ξ is concentrated on $\mathring{\mathcal{E}}$, one has to show that $L_t(0)$ increases each time B_t hits zero. By Proposition 2.5, one may equivalently show that if a Brownian motion $W = (W_t)_{t\geq 0}$ is started at some initial state $W_0 = x > 0$ and $\tau_0 := \inf\{t \geq 0 : W_t = 0\}$, then W immediately crosses the time axis, i.e., $\inf\{t \geq 0 : W_t < 0\} = \tau_0$. By the strong Markov property, it suffices to show that Brownian motion started in zero immediately crosses the time axis, which is well-known. ∎

Proof of Theorem 2.8 (crude sketch) In Section 2.6 we will give a proof of Theorem 2.8 based on finite approximation. Traditionally, there is a tendency to view such proofs as ugly.[1] Whether that is a good philosophy is questionable. Here, we sketch the outline of a classical proof using stochastic analysis.

The idea is to show that for each measurable $A \subset \mathcal{E}$, the process

$$N_s(A) := \Xi\big([0, s] \times A\big) \qquad (s \geq 0)$$

is stationary with independent increments, and moreover, if $A_1, \ldots, A_n$ are disjoint, then the processes $(N_s(A_1))_{s\geq 0}, \ldots, (N_s(A_n))_{s\geq 0}$ are independent. For each deterministic $s \geq 0$, the random time

$$\rho_s := \inf\big\{t \geq 0 : L_t(0) \geq s\big\}$$

is a stopping time for the Markov process $(|B_t|, L_t(0))_{t\geq 0}$. Using the strong Markov property for the stopping time ρ_s, one obtains that

$$\big(|B_{\rho_s+t}|, L_{\rho_s+t}(0) - L_{\rho_s}(0)\big)_{t\geq 0} \text{ is independent of } \big(|B_t|, L_t(0)\big)_{0\leq t\leq \rho_s},$$

and equally distributed with the original process $(|B_t|, L_t(0))_{t\geq 0}$. Using this, one obtains that for any $0 \leq s_1 \leq s_2$, the increment $N_{s_2}(A) - N_{s_1}(A)$ is independent of the restriction of Ξ to $[0, s_1] \times \mathcal{E}$ and equally distributed with $N_{s_2-s_1}(A)$, i.e., the process $(N_s(A))_{s\geq 0}$ is stationary with independent increments as claimed. Using also that it is a pure jump process one can now apply abstract results to conclude that Ξ must be a Poisson point set with the claimed properties. ∎

It is possible to "invert" the decomposition into excursions and reconstruct a reflected random walk or reflected Brownian motion from the set Ξ defined in (2.17). The construction is slightly different in the discrete and

[1] For example, it seems the main reason, apart from some minor inaccuracies, why the original proof of the Jordan curve theorem was not widely accepted, was that it used discrete approximation.

continuous cases. For the reflected random walk $(R_t, K_t)_{t\geq 0}$, we set

$$\begin{aligned}
\rho_s &:= s + \sum_{(u,\pi)\in\Xi:\, u<s} \tau_\pi && (s \geq 0),\\
K_t &:= \sup\{s \geq 0 : \rho_s \leq t\} && (t \geq 0),\\
R_t &:= E^{K_t}_{t-\rho_{K_t}} && (t \geq 0).
\end{aligned} \tag{2.22}$$

For the reflected Brownian motion $\big(|B_t|, L_t(0)\big)_{t\geq 0}$, we set

$$\begin{aligned}
\rho_s &:= \sum_{(u,\pi)\in\Xi:\, u<s} \tau_\pi && (s \geq 0),\\
L_t(0) &:= \sup\{s \geq 0 : \rho_s \leq t\} && (t \geq 0),\\
|B_t| &:= E^{(L_t)}_{t-\rho_{L_t}} && (n \geq 0).
\end{aligned} \tag{2.23}$$

The only difference between these formulas is in the definition of the function $(\rho_s)_{s\geq 0}$, which is the inverse of the reflection local times $(K_t)_{t\geq 0}$ and $(L_t(0))_{t\geq 0}$, respectively. In the discrete case, compared to the continuum case, we have to add a term $+s$ to to the definition of ρ_s. This has to do with the fact that K_t increases at speed one during the times when R_t is zero, while $L_t(0)$ increases at infinite speed during the times when $|B_t|$ is zero.

Formula (2.23) shows how to construct the absolute value of Brownian motion, i.e., the process $(|B_t|)_{t\geq 0}$, together with the local time at the origin of $(B_t)_{t\geq 0}$, from a Poisson set of excursions. In a similar way, one can also construct the Brownian motion $(B_t)_{t\geq 0}$ itself (instead of its absolute value). The idea is to assign signs to the excursions that are i.i.d. and uniformly distributed on $\{-1, +1\}$. It is also interesting to consider signs that are i.i.d. but not uniformly distributed on $\{-1, +1\}$. In this case, one obtains a Markov process known as *skew Brownian motion*.

The following proposition is a consequence of Brownian scaling. As before, we view paths as compact subsets of $\mathcal{R}(\mathbb{R})$ and we let $\theta_\lambda\pi$ denote the image of π under the diffusive scaling map θ_λ defined in (2.2). In this way, in (2.24) below, we naturally view θ_λ as a map from $\mathcal{E}$ to $\mathcal{E}$.

Proposition 2.10 (Diffusive scaling) *The excursion measure ν from Theorem 2.8 satisfies*

$$\nu \circ \theta_\lambda^{-1} = \lambda\nu \qquad (\lambda > 0). \tag{2.24}$$

Proof Let $(B_t)_{t\geq 0}$ be a standard Brownian motion and let $(L_t(0))_{t\geq 0}$ be its local time at 0. Fix $\lambda > 0$ and set

$$B^\lambda_t := \lambda B_{\lambda^{-2}t} \quad \text{and} \quad L^\lambda_t(0) := \lambda L_{\lambda^{-2}t}(0) \qquad (t \geq 0).$$

By Lemma 2.7, the processes $(|B_t|, L_t(0))_{t\geq 0}$ and $(|B^\lambda_t|, L^\lambda_t(0))_{t\geq 0}$ are equally distributed. Define Ξ and Ξ^λ as in (2.17) in terms of $(|B_t|, L_t(0))_{t\geq 0}$ and $(|B^\lambda_t|, L^\lambda_t(0))_{t\geq 0}$, respectively. Then

$$\Xi^\lambda = \big\{(\lambda s, \theta_\lambda \pi) : (s, \pi) \in \Xi\big\}.$$

Since both Ξ and Ξ^λ are Poisson point sets on $[0, \infty) \times \mathcal{E}$ with intensity measure $\ell \otimes \nu$, we see that the measure $\ell \otimes \nu$ is equal to its image under the map

$$(s, \pi) \mapsto (\lambda s, \theta_\lambda \pi).$$

In particular, for any measurable $A \subset \mathcal{E}$, we have

$$\begin{aligned}\lambda\nu(A) &= \ell \otimes \nu\big([0, \lambda] \times A\big)\\ &= \ell \otimes \nu\big([0, 1] \times \theta_\lambda^{-1}(A)\big) = \nu \circ \theta_\lambda^{-1}(A).\end{aligned}$$

∎

2.5 Standard excursions

We continue our study of the excursion measure ν from Theorem 2.8. We let

$$\mathcal{H}_r := \big\{\pi \in \mathring{\mathcal{E}} : \sup_{0\leq t\leq \tau_\pi} \pi(t) \geq r\big\} \qquad (r \geq 0) \tag{2.25}$$

denote the set of proper excursions that have height at least r. The next lemma determines $\nu(\mathcal{H}_r)$.

Lemma 2.11 (Height of the excursion) *Let ν be the excursion measure from Theorem 2.8. Then*

$$\nu(\mathcal{H}_r) = r^{-1} \qquad (r > 0). \tag{2.26}$$

Proof Let $(B_t)_{t\geq 0}$ be a standard Brownian motion and let $(L_t(0))_{t\geq 0}$ is its local time at 0. Let

$$\sigma_r := L_{\tau_r} \quad \text{with} \quad \tau_r := \inf\big\{t \geq 0 : |B_t| = r\big\}.$$

Then

$$\sigma_r := \inf\big\{s \geq 0 : \Xi \cap ([0, s] \times \mathcal{H}_r) \neq \emptyset\big\}.$$

By Theorem 2.8, σ_r is exponentially distributed with mean $1/\nu(\mathcal{H}_r)$. By Tanaka's formula (2.9), $|B_t| - L_t(0)$ is a martingale. By optional stopping, it follows that

$$\mathbb{E}\big[|B_{\tau_r \wedge t}|\big] = \mathbb{E}\big[L_{\tau_r \wedge t}(0)\big] \qquad (t \geq 0).$$

Letting $t \to \infty$, using the fact that $L_{\tau_r \wedge t}(0)$ increases to $L_{\tau_r} = \sigma_r$, and using dominated convergence for the left-hand side, together with $|B_{\tau_r}| = r$, we obtain that $\mathbb{E}[\sigma_r] = r$ and hence $\nu(\mathcal{H}_r) = 1/r$. ∎

Since $0 < \nu(\mathcal{H}_r) < \infty$ for each $r > 0$, we can define a conditional probability laws $\nu(\,\cdot\,|\mathcal{H}_r)$ on $\mathcal{E}$ by the usual formula

$$\nu(A|\mathcal{H}_r) := \frac{\nu(A \cap \mathcal{H}_r)}{\nu(\mathcal{H}_r)} \qquad \big(A \in \mathcal{B}(\mathcal{E})\big).$$

For each excursion $E \in \mathcal{H}_r$, we let

$$\sigma_{E,r} := \inf\{t \geq 0 : E_t = r\} \tag{2.27}$$

denote the first time the excursion E reaches the height r. As before, τ_E denotes the duration of E.

Lemma 2.12 (Conditional excursion law) *For each $r > 0$, under the conditional law $\nu(\,\cdot\,|\mathcal{H}_r)$, the process*

$$(E_{\sigma_{E,r}+t})_{0 \leq t \leq \tau_R - \sigma_{E,r}}$$

is distributed as a Brownian motion started at r and stopped at the first time it hits zero.

Proof (sketch) Let $(B_t)_{t \geq 0}$ be a standard Brownian motion. Let

$$\sigma_r^1 := \inf\{t \geq 0 : |B_t| = r\}, \quad \sigma_r^2 := \inf\{t \geq \sigma_r^1 : |B_t| = 0\},$$
$$\text{and} \quad \sigma_r^0 := \sup\{t < \sigma_r^1 : |B_t| = 0\},$$

and let $E \in \mathring{\mathcal{E}}$ be the excursion with duration $\tau_E := \sigma_r^2 - \sigma_r^0$ defined by

$$E_t := B_{\sigma_r^0 + t} \qquad (0 \leq t \leq \tau_E).$$

Then E_t is the first excursion in the Poisson point set Ξ of Theorem 2.8 that has height $\geq r$. Using the strong Markov property of Poisson point sets, one sees that E is distributed according to the conditional law $\nu(\,\cdot\,|\mathcal{H}_r)$. Using the strong Markov property of Brownian motion, one sees that

$$(B_{\sigma_r^1 + t})_{0 \leq t \leq \sigma_r^2 - \sigma_r^1}$$

is distributed as a Brownian motion started at r and stopped at the first time it hits zero. ∎

We let

$$\mathcal{D}_t := \big\{\pi \in \mathring{\mathcal{E}} : \tau_\pi > t\big\} \qquad (t \geq 0)$$

denote the set of excursions that have duration at least t. The next lemma determines $\nu(\mathcal{D}_t)$.

Lemma 2.13 (Duration of the excursion) *Let ν be the excursion measure from Theorem 2.8. Then*

$$\nu(\mathcal{D}_t)=\frac{2}{\sqrt{2\pi}}t^{-1/2}\qquad(t>0).\tag{2.28}$$

Proof of Lemma 2.13 We define $\mathcal{H}_r$ as in (2.25) and for each $E\in\mathcal{H}_r$ we define $\sigma_{E,r}$ as in (2.27). As before, τ_E denotes the duration of E. For each $t>0$, we set

$$\mathcal{H}_{r,t}:=\big\{E\in\mathcal{H}_r:\tau_E\geq\sigma_{E,r}+t\big\},$$

i.e., these are all excursions that reach the height r and after that live for at least time t. Lemma 2.12 implies that

$$\nu(\mathcal{H}_{r,t})=\nu(\mathcal{H}_r)\mathbb{P}\big[r+B_s>0\ \forall 0\leq s\leq t\big],$$

where $(B_t)_{t\geq 0}$ is a standard Brownian motion. It is a consequence of the reflection principle that

$$\mathbb{P}\big[r+B_s>0\ \forall 0\leq s\leq t\big]=\mathbb{P}\big[|B_t|\leq r\big]=\int_{-r}^{r}\frac{1}{\sqrt{2\pi t}}e^{-\frac{1}{2t}x^2}\mathrm{d}x.$$

Using Lemma 2.11, which tells us that $\nu(\mathcal{H}_r)=r^{-1}$, it follows that

$$\nu(\mathcal{H}_{r,t})=r^{-1}\Big[\frac{2rt^{-1/2}}{\sqrt{2\pi}}+O(r^2)\Big]\quad\text{as } r\to 0.$$

Letting $r\to 0$, using the fact that $\mathcal{H}_{r,t}$ increases to $\mathcal{D}_t$, the claim follows. ∎

We let

$$\mathcal{E}_1:=\big\{\pi\in\mathcal{E}:\tau_\pi=1\big\}\tag{2.29}$$

denote the space of excursions of duration one and set $\mathring{\mathcal{E}}_1:=\mathcal{E}_1\cap\mathring{\mathcal{E}}$. A random variable whose law is the probability measure ν_1 from Proposition 2.14 below is called a *standard Brownian excursion.*

Proposition 2.14 (Decomposition of the excursion measure) *Let ρ be the measure on $(0,\infty)$ defined as*

$$\rho(\mathrm{d}t):=\frac{1}{\sqrt{2\pi}}t^{-3/2}\mathrm{d}t.\tag{2.30}$$

There exists a probability measure ν_1 on $\mathring{\mathcal{E}}_1$ such that the excursion measure from Theorem 2.8 is the image of the measure $\rho\otimes\nu_1$ under the map

$$(0,\infty)\times\mathcal{E}_1\ni(t,f)\mapsto\theta_{\sqrt{t}}f\in\mathcal{E}.\tag{2.31}$$

Proof (sketch) The map in (2.31) is invertible. Its inverse is the map

$$\mathcal{E} \ni \pi \mapsto \big(\tau_\pi, \theta_{1/\sqrt{\tau_\pi}}\pi\big) \in (0,\infty) \otimes \mathcal{E}_1,$$

where as before τ_π denotes the duration of an excursion $\pi \in \mathcal{E}$. Let μ be the image of the excursion measure ν under this inverse map. Then Proposition 2.10 implies that

$$\mu \circ \psi_\lambda^{-1} = \lambda\mu \quad \text{where} \quad \psi_\lambda(s,f) := (\lambda^2 s, f) \qquad (\lambda, s > 0,\ f \in \mathcal{E}_1). \tag{2.32}$$

Using the fact that by Lemma 2.13, $\nu(\mathcal{D}_1)$ is finite, it follows that we can decompose μ as

$$\mu(\mathrm{d}\lambda, \mathrm{d}f) = \rho(\mathrm{d}\lambda)P(\lambda, \mathrm{d}f)$$

for some probability kernel P (compare Theorem 1.5). By (2.32), $P(\lambda,\,\cdot\,)$ does not depend on λ, so in fact $\mu = \rho \otimes \nu_1$ for some probability measure ν_1 on $\mathring{\mathcal{E}}_1$. The scaling relation (2.32) moreover implies that

$$\rho\big([\lambda^{-2}t, \infty)\big) = \lambda\rho\big([t,\infty)\big) \qquad (\lambda, t > 0),$$

which shows that there exists a constant $c > 0$ such that

$$\rho\big([t,\infty)\big) = ct^{-1/2} \qquad (\lambda > 0).$$

The correct formula for the constant c follows from Lemma 2.13. ∎

2.6 Scaling limits of excursions

In this section, we give a proof of Theorem 2.8 based on finite approximation. As a side result, we obtain that if S^ε are the diffusively rescaled simple random walks defined in (2.3) and let B is a standard Brownian motion, then the excursion sets $\Xi(S^\varepsilon)$ defined in (2.17) converge in an appropriate sense to $\Xi(B)$.

We first need a few definitions. By definition, a *local* subset of the set of excursions $\mathcal{E}$ is a measurable set $A \subset \mathcal{E}$ such that $o \notin \overline{A}$, where $\overline{A}$ denotes the closure of A and o denotes the trivial excursion of duration $\tau_o := 0$. Similarly, a *local* subset of $[0,\infty) \times \mathcal{E}$ is a measurable set $B \subset [0,\infty) \times \mathcal{E}$ such that

$$B \subset [0,S] \times A \quad \text{for some } S < \infty \text{ and local } A \subset \mathcal{E}.$$

We say that a measure ν on $\mathcal{E}$ is *locally finite*[2] if $\nu(A) < \infty$ for all local $A \subset \mathcal{E}$. Similarly, we say that a measure μ on $[0,\infty) \times \mathcal{E}$ is *locally finite* if

[2]We use this term in an unusual sense here. More usually, a measure μ on a locally compact space $\mathcal{X}$ is called *locally finite* if $\mu(K) < \infty$ for each compact $K \subset \mathcal{X}$. The space $\mathcal{E}$ is, however, not locally compact, so such a definition would not make much sense in our setting.

$\mu(B) < \infty$ for all local $B \subset [0, \infty) \times \mathcal{E}$. When $\mathcal{X} = \mathcal{E}$ or $= [0, \infty) \times \mathcal{E}$, we let $\mathcal{M}_{\rm loc}(\mathcal{X})$ denote the space of locally finite measures on $\mathcal{X}$.

The *support* of a measurable real function f defined on a topological space $\mathcal{X}$ is the set

$$\mathrm{supp}(f) := \overline{\{x \in \mathcal{X} : f(x) \neq 0\}},$$

where the overbar denotes closure. In the special case that $\mathcal{X} = \mathcal{E}$ or $= [0, \infty) \times \mathcal{E}$, we define a *local* function on $\mathcal{X}$ to be a measurable function $f : \mathcal{X} \to \mathbb{R}$ such that $\mathrm{supp}(f)$ is a local subset of $\mathcal{X}$. We let $\mathcal{C}_{\rm loc}(\mathcal{X})$ denote the space of all local bounded continuous functions on $\mathcal{X}$ and we let $\mathcal{M}_{\rm loc}(\mathcal{X})$ denote the space of all locally finite measures on $\mathcal{X}$.

Let $\mathcal{X}$ be a Polish space and let $(x_i)_{i \in I}$ be a countable collection of points in $\mathcal{X}$. Then

$$\xi := \sum_{i \in I} \delta_{x_i} \tag{2.33}$$

defines a counting measure on $\mathcal{X}$. In particular, if Ξ is a countable subset of $\mathcal{X}$, then Ξ defines a counting measure by

$$\xi_\Xi := \sum_{x \in \Xi} \delta_x. \tag{2.34}$$

Note that ξ_Ξ is *simple*, in the sense that $\xi_\Xi(\{x\}) \leq 1$ for all $x \in \mathcal{X}$. In general, counting measures of the form (2.33) need not be simple, since it may happen that $x_i = x_j$ for some $i \neq j$. We often tacitly identify countable subsets of $\mathcal{X}$ with their associated counting measures. In particular, we say that a countable subset Ξ of $\mathcal{X} = \mathcal{E}$ or $= [0, \infty) \times \mathcal{E}$ is *locally finite* if ξ_Ξ has this property.

Let $\phi : \mathcal{X} \to [0, 1]$ be measurable and let $(\chi_i)_{i \in I}$ be independent *Bernoulli* random variables (i.e., variables with values in $\{0, 1\}$) with $\mathbb{P}[\chi_i = 1] = \phi(x_i)$. Then the random counting measure

$$\xi' := \sum_{i \in I} \chi_i \delta_{x_i}$$

is called a *ϕ-thinning* of ξ. In the special case that $\mathcal{X}$ is either $\mathcal{E}$ or $[0, \infty) \times \mathcal{E}$, we let $\mathcal{N}_{\rm loc}(\mathcal{X})$ denote the space of all locally finite counting measures on $\mathcal{X}$. Then

$$K_\phi(\xi, \,\cdot\,) := \mathbb{P}\big[\xi' \in \cdot\,\big]$$

defines a probability kernel on $\mathcal{N}_{\rm loc}(\mathcal{X})$. Generalising our earlier definition of a thinning, when ξ and ξ' are random locally finite counting measures on $\mathcal{X}$, then we say that ξ' is a *ϕ-thinning* of ξ if

$$\mathbb{P}\big[\xi' \in \cdot \,\big|\, \xi\big] = K_\phi(\xi, \,\cdot\,).$$

For any counting measure ξ of the form (2.33) and measurable $\phi : \mathcal{X} \to [0,1]$, we introduce the notation

$$\phi^\xi := \prod_{i\in I} \phi(x_i) = e^{\int \xi(\mathrm{d}x)\,\log\phi(x)},$$

with the convention that $e^{-\infty} := 0$. If ξ' is a ϕ-thinning of ξ, then it is easy to see that

$$\mathbb{P}\big[\xi' = 0\big] = \mathbb{E}\big[(1-\phi)^\xi\big].$$

We say that μ on a measurable space $\mathcal{X}$ is *nonatomic* if $\mu(\{x\}) = 0$ for all $x \in \mathcal{X}$. Recall that a counting measure ξ is called *simple* if $\xi(\{x\}) \leq 1$ for all $x \in \mathcal{X}$. We need the following result.[3]

Theorem 2.15 (Poisson counting measure) *Let μ be a locally finite measure on $[0,\infty) \times \mathcal{E}$. Then there exists a random locally finite counting measure ξ on $[0,\infty) \times \mathcal{E}$ such that*

$$\mathbb{E}\big[(1-\phi)^\xi\big] = e^{-\int \phi\,\mathrm{d}\mu} \tag{2.35}$$

for each measurable $\phi : [0,\infty) \times \mathcal{E} \to [0,1]$. The law of ξ is uniquely determined by the requirement that (2.35) holds for all local continuous ϕ. If $B_1, \ldots, B_n$ are disjoint measurable local subsets of $[0,\infty) \times \mathcal{E}$, then

$$\xi(B_1), \ldots, \xi(B_n) \quad \textit{are Poisson distributed with mean} \quad \mu(B_1), \ldots, \mu(B_n).$$

If μ is nonatomic, then ξ is almost surely simple.

Formula (2.35) has an interpretation in terms of thinning. Let $\phi\mu$ denote the measure μ weighted with the density ϕ. If ξ' is a ϕ-thinning of ξ, then ξ' is a Poisson counting measure with intensity measure $\phi\mu$. In particular, if $\int \phi\,\mathrm{d}\mu < \infty$, then the number of points of ξ' is Poisson distributed with mean $\int \phi\,\mathrm{d}\mu < \infty$, and hence $\mathbb{P}[\xi' = 0] = \exp(-\int \phi\,\mathrm{d}\mu)$, which is formula (2.35).

We now turn our attention to the proof of Theorem 2.8. We will use discrete approximation. Let S be the simple random walk from Section 2.1, let S^ε be the diffusively rescaled random walk from (2.3), and let B be standard Brownian motion. We recall from (2.20) that the excursion set of S is given by

$$\Xi(S) = \big\{(k, E^k) : k \in \mathbb{N},\ E^k \neq o\big\},$$

[3] This is largely standard, but many sources such as [Kal97, Chapter 10] treat only locally compact spaces. Oir definition of local finiteness is also nonstandard.

where $(E^k)^{k\in\mathbb{N}}$ are the i.i.d. excursions from (2.21). It follows that

$$\Xi(S^\varepsilon) = \big\{(\varepsilon k, \theta_\varepsilon E^k) : k \in \mathbb{N},\ E^k \neq o\big\}. \tag{2.36}$$

Note that in view of (2.12), we have to rescale the reflection local time k by a factor ε and not by ε^2. We will prove Theorem 2.8 together with the following theorem, which describes the tail of the law of E^0, i.e., in the small probabilities of very large excursions.

Theorem 2.16 (Tail of the excursion law) *Let ν be the excursion measure from Theorem 2.8. One has*

$$\varepsilon^{-1}\mathbb{E}\big[g(\theta_\varepsilon E^0)\big] \underset{\varepsilon\to 0}{\longrightarrow} \int_{\mathcal{E}} g(\pi)\,\nu(\mathrm{d}\pi) \tag{2.37}$$

for each $g \in \mathcal{C}_{\rm loc}(\mathcal{E})$.

The proof of Theorem 2.8 depends on two technical results, the proofs of which will be postponed till the next section. Recall from (2.1) that $\mathcal{C}_0 := \big\{f \in \mathcal{C}_{[0,\infty)}(\mathbb{R}) : f_0 = 0\big\}$.

Lemma 2.17 (Locally finite excursion set) *For each $f \in \mathcal{C}_0$ such that $\liminf_{t\to\infty} f_t = -\infty$, the set $\Xi(f)$ defined in (2.17) is locally finite.*

We have seen in (2.5) that it is possible to couple diffusively rescaled random walks S^{ε_n} and a Brownian motion B such that almost surely $S^{\varepsilon_n} \to B$ locally uniformly. The following theorem says that then also the associated excursion sets converge. In (2.38) below, we identify the countable sets $\Xi(S^{\varepsilon_n})$ and $\Xi(B)$ with their associated counting measures as in (2.34).

Theorem 2.18 (Scaling limit of excursion sets) *Let ε_n be positive constants tending to zero, let S^{ε_n} be the diffusively rescaled simple random walk defined in (2.3) and let B be a standard Brownian motion. Assume that these random variables are coupled as in (2.5). Then the excursion sets $\Xi(S^{\varepsilon_n})$ and $\Xi(B)$ defined in (2.17) almost surely satisfy*

$$(1-\phi)^{\Xi(S^{\varepsilon_n})} \underset{n\to\infty}{\longrightarrow} (1-\phi)^{\Xi(B)} \tag{2.38}$$

for all local continuous $\phi : [0,\infty) \times \mathcal{E} \to [0,1]$.

Proof of Theorems 2.8 and 2.16 Let ε_n be positive constants tending to zero. We fix a local continuous function $g : \mathcal{E} \to [0,1]$ and a continuous compactly supported function $h : [0,\infty) \to [0,1]$. Then $\phi(s,\pi) := h(s)g(\pi)$

defines a local continuous function $\phi : [0,\infty) \times \mathcal{E} \to [0,1]$. Applying Theorem 2.18, using bounded pointwise convergence to interchange the integral and the limit, we see that

$$\mathbb{E}\big[(1-\phi)^{\Xi(S^{\varepsilon_n})}\big] \underset{n\to\infty}{\longrightarrow} \mathbb{E}\big[(1-\phi)^{\Xi(B)}\big]. \tag{2.39}$$

By (2.21) and (2.36), we can rewrite the left-hand side as

$$\mathbb{E}\big[(1-\phi)^{\Xi(S^{\varepsilon_n})}\big] = \prod_{k=0}^{\infty}\Big(1 - h(\varepsilon_n k)\mathbb{E}\big[g(\theta_{\varepsilon_n}E^0)\big]\Big).$$

By going to a subsequence, we can assume that

$$G_n := \varepsilon_n^{-1}\mathbb{E}\big[g(\theta_{\varepsilon_n}E^0)\big] \underset{n\to\infty}{\longrightarrow} G \in [0,\infty].$$

We claim that then

$$\mathbb{E}\big[(1-\phi)^{\Xi(S^{\varepsilon_n})}\big] \underset{n\to\infty}{\longrightarrow} e^{-G\int_0^\infty h(t)\mathrm{d}t}.$$

The claim is trivial if $h = 0$, so we assume $h \neq 0$ without loss of generality. We use the concavity of the logarithm and Riemman sum approximation of the integral to estimate

$$\begin{aligned}\log\mathbb{E}\big[(1-\phi)^{\Xi(S^{\varepsilon_n})}\big] &= \sum_{k=0}^{\infty}\log\Big(1-\varepsilon_n G_n h(\varepsilon_n k)\Big)\\ &\le -G_n\varepsilon_n\sum_{k=0}^{\infty}h(\varepsilon_n k) \underset{n\to\infty}{\longrightarrow} -G\int_0^\infty h(t)\mathrm{d}t.\end{aligned}$$

This already proves the statement when $G = \infty$, so it suffices to prove the other inequality under the assumption that $G < \infty$. Then $\varepsilon_n G_n \to 0$ while $h \le 1$, so

$$\log\Big(1-\varepsilon_n G_n h(\varepsilon_n k)\Big) = -\varepsilon_n G_n h(\varepsilon_n k) + O(\varepsilon_n^2).$$

Since h is compactly supported, only $O(\varepsilon_n^{-1})$ terms in the sum are nonzero, so the claim follows easily. Using (2.39), we now see that the limit G has to be the same for each subsequence, so for each local continuous function $g : \mathcal{E} \to [0,1]$, there exists a constant $\nu(g) \in [0,\infty]$ such that

$$\varepsilon_n^{-1}\mathbb{E}\big[g(\theta_{\varepsilon_n}E^0)\big] \underset{n\to\infty}{\longrightarrow} \nu(g). \tag{2.40}$$

Formula (2.39) moreover tells us that for any local continuous $g : \mathcal{E} \to [0,1]$ and continuous compactly supported $h : [0,\infty) \to [0,1]$,

$$\mathbb{E}\Big[\prod_{(s,\pi)\in\Xi(B)}\big(1-h(s)g(\pi)\big)\Big] = e^{-\nu(g)\int_0^\infty h(t)\mathrm{d}t}.$$

By Lemma 2.17, the set $\Xi(B)$ is a.s. locally finite, so only finitely many factors in the product are different from one. If $h \leq \frac{1}{2}$, then the product is with positive probability positive, which proves that $\nu(g) < \infty$ for each local continuous $g : \mathcal{E} \to [0, 1]$. Combining this with (2.40), we see that there must exist a locally finite measure ν on $\mathcal{E}$ such that

$$\nu(g) = \int_{\mathcal{E}} \nu(\mathrm{d}\pi)\, g(\pi),$$

and (2.40) holds more generally for bounded local continuous $g : \mathcal{E} \to \mathbb{R}$. To complete the proof, it suffices to prove that $\Xi(B)$ is a Poisson point set with intensity measure $\ell \otimes \nu$. By Theorem 2.15, it suffices to show that

$$\mathbb{E}\big[(1-\phi)^{\Xi(B)}\big] = e^{-\int_0^\infty \mathrm{d}s\, \int_{\mathcal{E}} \nu(\mathrm{d}\pi) \phi(s,\pi)}$$

for each local continuous $\phi : [0, \infty) \times \mathcal{E} \to [0, 1]$. Our arguments so far already show that this is true for ϕ of the form[4] $\phi(s, \pi) = h(s)g(\pi)$ with local continuous $g : \mathcal{E} \to [0, 1]$ and continuous compactly supported $h : [0, \infty) \to [0, 1]$. We again use (2.39) and setting $g_s(\pi) := \phi(s, \pi)$, we write

$$\log \mathbb{E}\big[(1-\phi)^{\Xi(S^{\varepsilon_n})}\big] = \sum_{k=0}^{\infty} \log\Big(1 - \mathbb{E}\big[g_{\varepsilon_n k}(\theta_{\varepsilon_n} E^0)\big]\Big),$$

where we can estimate

$$\log\Big(1 - \mathbb{E}\big[g_{\varepsilon_n k}(\theta_{\varepsilon_n} E^0)\big]\Big) = -\varepsilon_n \int \nu(\mathrm{d}\pi) g_{\varepsilon_n k}(\pi) + O(\varepsilon_n^2).$$

The claim then follows from Riemann sum approximation to the integral. ∎

2.7 Limits of excursion sets

In this section we provide the proofs of Lemma 2.17 and Theorem 2.18, which are still missing.

Proof of Lemma 2.17 and Theorem 2.18 The main work is the proof of Theorem 2.18. We will obtain Lemma 2.17 as a side result. If the map $\mathcal{C}_0 \ni f \mapsto \Xi(f)$ were continuous with respect to the sort of convergence we are considering, then the statement of Theorem 2.18 would be trivial. This is not true, but we will show that if B is a Brownian motion, then the map

[4] I actually do not know if this is already enough to conclude that $\Xi(B)$ is a Poisson point set with intensity measure $\ell \otimes \nu$.

$f \mapsto \Xi(f)$ is almost surely continuous in the point $B \in \mathcal{C}_0$, which is all we need.

We will prove the following statement. Assume that $f_n, f \in \mathcal{C}_0$ satisfy $f_n \to f$ locally uniformly, that $\liminf_{t\to\infty} f(t) = -\infty$, and that $\Xi(f)$ is concentrated on $[0,\infty) \times \mathring{\mathcal{E}}$. Then

$$(1-\phi)^{\Xi(f_n)} \underset{n\to\infty}{\longrightarrow} (1-\phi)^{\Xi(f)} \tag{2.41}$$

for all local continuous $\phi : [0,\infty)\times\mathcal{E} \to [0,1]$. Note that this is a deterministic statement: the only way randomness enters our proof is in the fact that if B is a Brownian motion, then almost surely $\liminf_{t\to\infty} B_t = -\infty$ and $\Xi(B)$ is concentrated on $[0,\infty) \times \mathring{\mathcal{E}}$, which follows from Theorem 2.8.

Assume, therefore, that $f_n, f \in \mathcal{C}_0$ satisfy $f_n \to f$ locally uniformly, that $\liminf_{t\to\infty} f(t) = -\infty$, and that $\Xi(f)$ is concentrated on $[0,\infty)\times\mathring{\mathcal{E}}$. Let (g_n, h_n) be defined in terms of f_n as in (2.7) and let (g, h) be similarly defined in terms of f. Let $\mathcal{I}(h)$ denote the set of plateaus of h, i.e., an open intervals of maximal length on which h is constant, and let $\mathcal{I}(h_n)$ be the plateaus of h_n.

Let $\iota = (\iota^-, \iota^+) \in \mathcal{I}(h)$. Our assumption that $\Xi(f)$ is concentrated on $[0,\infty) \times \mathring{\mathcal{E}}$ means that the function g is strictly positive on ι. The locally uniform convergence $g_n \to g$ then implies that for each $\varepsilon > 0$, the function g_n must be strictly positive on $(\iota^- + \varepsilon, \iota^+ - \varepsilon)$ for all n large enough. Since h_n increases only at times when g_n is zero, this then implies that h_n must be constant on $(\iota^- + \varepsilon, \iota^+ - \varepsilon)$.

On the other hand, since ι is a maximal interval on which h is constant, $h(t) < h(\iota^-)$ for all $t < \iota^-$ and $h(\iota^+) < h(t)$ for all $\iota^+ < t$. The locally uniform convergence $h_n \to h$ then implies that for each $\varepsilon > 0$, the function h_n is not constant on $(\iota^- - \varepsilon, \iota^+ + \varepsilon)$ for all n large enough. These arguments show that for each plateau $\iota \in \mathcal{I}(h)$ and for each $\varepsilon \le \varepsilon_0 := (\iota^+ - \iota^-)/3$, there exists an $m(\varepsilon)$ such that for all $n \ge m(\varepsilon)$, there exists a (clearly unique) plateau $\jmath \in \mathcal{I}(h_n)$ with $|\jmath^\pm - \iota^\pm| \le \varepsilon$. For $n \ge m(\varepsilon_0)$, we let $\phi_n(\iota) := \jmath$ denote this plateau, and we define $\phi_n(\iota)$ in an abritrary way for the remaining values of n. Then clearly the left and right boundaries of the plateau $\phi_n(\iota)$ satisfy

$$\phi_n(\iota)^\pm \underset{n\to\infty}{\longrightarrow} \iota^\pm. \tag{2.42}$$

Let $(s,\pi) \in \Xi(f)$ denote the excursion of g corresponding to the plateau ι, and let $\psi_n(s,\pi) \in \Xi(f_n)$ denote the excursion of g_n corresponding to the plateau $\phi_n(\iota)$. Using the fact that $g_n \to g$ and $h_n \to h$ locally uniformly, we see that

$$\psi_n(s,\pi) \underset{n\to\infty}{\longrightarrow} (s,\pi) \tag{2.43}$$

in the topology on $[0,\infty)\times\mathcal{E}$.

For each $S<\infty$ and $\delta>0$, let us set

$$\mathcal{I}_{S,\delta}(h) := \big\{\iota\in\mathcal{I}(h) : h(\iota^\pm)<S,\ \iota^+-\iota^->\delta\big\},$$

and define $\mathcal{I}_{S,\delta}(h_n)$ similarly. We claim that for large n, the map ϕ_n is a bijection from $\mathcal{I}_{S,\delta}(h)$ to $\mathcal{I}_{S,\delta}(h_n)$ and hence ψ_n is a bijection from $\Xi_{S,\delta}(f)$ to $\Xi_{S,\delta}(f_n)$. Let $T:=\sup\{t : h(t)<S\}$ and $T_n:=\sup\{t : h_n(t)<S\}$. Then $T<\infty$ by the assumption that $\liminf_{t\to\infty} f(t)=-\infty$ and $T_n\to T$ by the fact that $h_n\to h$ locally uniformly. Since all plateaus $\iota\in\mathcal{I}_{S,\delta}$ are contained in $[0,T]$, the set $\mathcal{I}_{S,\delta}(h)$ can contain at most T/δ elements and is therefore finite. It follows from (2.42) and (2.43) that for large enough n, the map ϕ_n maps the space $\mathcal{I}_{S,\delta}(h)$ into $\mathcal{I}_{S,\delta}(h_n)$. It follows immediately from our definition of ϕ_n that this map is also one-to-one for n large enough.

To see that it is moreover surjective for n large enough, assume that conversely, for infinitely many values of n, there exists a $\jmath_n\in\mathcal{I}_{S,\delta}(h_n)$ that is not the image under ϕ_n of some $\iota\in\mathcal{I}_{S,\delta}(h)$. Since $\jmath_n\subset[0,T_n]$, by going to a subsequence, we can assume that $\jmath_n^\pm\to\jmath^\pm$ for some interval $\jmath$. But then h has to be constant on $\jmath$, which implies that $\jmath\subset\iota$ for some $\iota\in\mathcal{I}(h)$. But this implies that $\jmath_n$ has nonempty intersection with $\phi_n(\iota)$ for all n large enough, which leads to a contradiction.

For $S<\infty$ and $\delta,\varepsilon>0$, let us set

$$\begin{aligned}\Xi_{S,\delta}(f) &:= \big\{(s,\pi)\in\Xi(f) : s<S,\ \tau_\pi>\delta\big\},\\ \Xi^\varepsilon_S(f) &:= \big\{(s,\pi)\in\Xi(f) : s<S,\ \sup_{0\le t\le\tau_\pi}\pi(t)>\varepsilon\big\}.\end{aligned}$$

There is a one-to-one correspondence between $\mathcal{I}_{S,\delta}(f)$ and $\Xi_{S,\delta}(f)$. We have just proved that the former is finite[5] for each S and δ, and hence the same is true for the latter. We claim that

$$\forall\varepsilon>0\ \exists\delta>0 \text{ s.t. } \Xi^\varepsilon_S(f)\subset\Xi_{S,\delta}(f)\qquad (S<\infty). \tag{2.44}$$

As before, let $T:=\sup\{t : h(t)<S\}$. Let $m_{T,\delta}(f)$ be the modulus of continuity defined in (1.13). If there exists a $(s,\pi)\in\Xi_{S,\varepsilon}$ such that $(s,\pi)\not\in\Xi_{S,\delta}(f)$, then $m_{T,\delta}(f)>\varepsilon$. Now (2.44) follows from the fact that by Lemma 1.29, for each $\varepsilon>0$, there exists a $\delta>0$ such that $m_{T,\delta}(f)\le\varepsilon$. By the same argument, using the equicontinuity of the functions f_n, which follows from

[5]This part of the argument used that $\liminf_{t\to\infty} f_t=-\infty$ and hence $T:=\sup\{t : h(t)<S\}$ is finite, but it did not need the assumption that $\Xi(f)$ is concentrated om $[0,\infty)\times\mathring{\mathcal{E}}$.

the fact that $f_n \to f$ and The Arzela-Ascoli theorem (Theorem 1.30), we see that

$$\forall \varepsilon > 0 \ \exists \delta > 0 \text{ s.t. } \Xi^{\varepsilon}_{S}(f_n) \subset \Xi_{S,\delta}(f_n) \qquad (n \geq 1,\ S < \infty). \tag{2.45}$$

We are finally ready to prove (2.41). Fix a local continuous function $\phi : [0,\infty) \times \mathcal{E} \to [0,1]$, and let $A := \mathrm{supp}(\phi)$ be its support. We claim that there exist $\delta, \varepsilon > 0$ such that

$$\forall \pi \in A \ \tau_\pi > \delta \text{ or } \sup_{0 \leq t \leq \tau_\pi} \pi_t > \varepsilon. \tag{2.46}$$

Indeed, if (2.46) does not hold, then there exist $\pi_n \in A$ with $\tau_\pi \leq n^{-1}$ and $\sup_{0 \leq t \leq \tau_\pi} \pi_t \leq n^{-1}$. But then $\pi_n \to o$, the trivial excursion, which contradicts the fact that A is closed with $o \notin A$. Using (2.44) and (2.45), we see that there exists a $\delta' > 0$ such that

$$\begin{aligned} \big\{(s,\pi) \in \Xi(f) : s \leq S,\ \pi \in A\big\} &\subset \Xi_{S,\delta'}(f), \\ \big\{(s,\pi) \in \Xi(f_n) : s \leq S,\ \pi \in A\big\} &\subset \Xi_{S,\delta'}(f_n) \qquad (n \geq 1). \end{aligned}$$

Using (2.43) and the fact that for large n, the map ψ_n is a bijection from the finite set $\Xi_{S,\delta'}(f)$ to $\Xi_{S,\delta'}(f_n)$, we see that (2.41) holds. This completes the proof of Theorem 2.18.

Along the way, we have established that if $f \in \mathcal{C}_0$ satisfies $\liminf_{t\to\infty} f_t = -\infty$, then the set

$$\big\{(s,\pi) \in \Xi(f) : s \leq S,\ \pi \in A\big\}$$

is finite for each $S < \infty$ and $A \subset \mathcal{E}$ that is closed with $o \notin A$, proving Lemma 2.17. ∎

2.8 Large random walk excursions

Let $(R,K) = (R_t, K_t)_{t \geq 0}$ be the (linearly interpolated) reflected random walk defined in (2.11), and let

$$\tau_0 := \inf\big\{i \in \mathbb{N} : K_{i+1} > K_i\big\} \quad \text{and} \quad E^0 := (R_t)_{0 \leq t \leq \tau_0}, \tag{2.47}$$

i.e., E^0 is the first of the i.i.d. excursions $(E^k)^{k\in\mathbb{N}}$ of (R,K) defined in (2.19). Note that since $R_0 = 0 = R_{\tau_0}$ and up to time τ_0, the reflected random walk R steps up or down in each time step, τ_0 is almost surely an even number. In Theorem 2.16, we have seen that the Brownian excursion measure ν describes the tail of the law of E^0, i.e., the small probabilities of very large excursions. In the present section, our aim is to prove the following theorem.

Theorem 2.19 (Scaling limit of large excursion) *Let* $\varepsilon_n := 1/\sqrt{2n}$. *Then*

$$\mathbb{P}\big[\theta_{\varepsilon_n} E^0 \in \cdot \,\big|\, \tau_0 = 2n\big] \underset{n\to\infty}{\Longrightarrow} \nu_1, \tag{2.48}$$

where ν_1 *is the law of the standard Brownian excursion, defined in Proposition 2.14.*

Despite its apparent simplicity, the proof of Theorem 2.19 is quite tricky and we will not completely prove it in this section. We will get quite close, however, and indicate what needs to be done to complete the proof. We want to use excursion theory to prove Theorem 2.19. This may seem natural, but apparently a proof using this approach has been published only fairly recently in [LeG10, Thm 6.1]. That paper is concerned with a class of discrete excursions that is more general than the one we consider, but also a bit different so that Theorem 2.19 is not formally included in [LeG10, Thm 6.1] although it is very similar.

The proof of Theorem 2.19 needs some preparations. We say that a measure ρ on $(0,\infty)$ is *locally finite* if $\rho([s,S]) < \infty$ for all $0 < s < S < \infty$. We say that a sequence of locally finite measures ρ_n on $(0,\infty)$ converges *vaguely* to a limit ρ if

$$\int_0^\infty \rho_n(\mathrm{d}t)\, h(t) \underset{n\to\infty}{\longrightarrow} \int_0^\infty \rho(\mathrm{d}t)\, h(t)$$

for all continuous compactly supported $h : (0,\infty) \to \mathbb{R}$. We postpone the proof of the following simple lemma till later.

Lemma 2.20 (Integrals along paths) *Let* $\pi_n, \pi \in \Pi(\mathbb{R})$ *be paths such that* $I_\pi = [0,\infty)$ *and* $I_{\pi_n} \subset [0,\infty)$ *for all* n. *Let* ρ_n, ρ *be locally finite measures on* $(0,\infty)$ *such that* ρ_n *is concentrated on* I_{π_n} *for each* n. *Assume that* $\pi_n \to \pi$ *in the topology on path space* $\Pi(\mathbb{R})$ *and that the* ρ_n *converge vaguely to* ρ. *Then*

$$\int_0^\infty \rho_n(\mathrm{d}t) h(t) \pi_n(t) \underset{n\to\infty}{\longrightarrow} \int_0^\infty \rho(\mathrm{d}t) h(t) \pi(t)$$

for each continuous compactly supported $h : (0,\infty) \to \mathbb{R}$.

For each $m \in 2\mathbb{N} := \{2n : n \in \mathbb{N}\}$, we let μ_m denote the conditional law

$$\mu_m := \mathbb{P}\big[\theta_{1/\sqrt{m}}\, E^0 \in \cdot \,\big|\tau_0 = m\big]. \tag{2.49}$$

For any bounded continuous funtion $g : \mathcal{E}_1 \to \mathbb{R}$, we write

$$\langle \mu_m, g \rangle := \int_{\mathcal{E}_1} \mu_m(\mathrm{d}\pi)\, g(\pi).$$

We will need the following technical result, that we will not prove in this chapter.

Proposition 2.21 (Equicontinuity of conditional laws) *Let $g : \mathcal{E}_1 \to \mathbb{R}$ be bounded and continuous. For each $\delta \in (0,1]$, let $\pi^\delta \in \Pi(\mathbb{R})$ be the path defined by*

$$I_{\pi^\delta} := 2\delta^2\mathbb{N} \cap [1,\infty) \quad \textit{and} \quad \pi^\delta(t) := \langle \mu_{\delta^{-2}t}, g\rangle \quad (t \in I_{\pi^\delta}). \tag{2.50}$$

Then the paths $\{\pi^\delta : \delta \in (0,1]\}$ are equicontinuous.

Proof of Theorem 2.19 Let $g : \mathcal{E}_1 \to \mathbb{R}$ be bounded and continuous and let $h : (0,\infty) \to \mathbb{R}$ be continuous and compactly supported. Define $f : \mathcal{E} \to \mathbb{R}$ by

$$f(\pi) := h(\tau_\pi) g(\theta_{1/\sqrt{\tau_\pi}}\pi) \qquad (\pi \neq o),$$

with $f(o) := 0$. Then f is bounded and continuous with $o \notin \mathrm{supp}(f)$, so Theorem 2.16 tells us that

$$\delta^{-1}\mathbb{E}\big[f(\theta_\delta E^0)\big] \underset{\delta\to 0}{\longrightarrow} \int_{\mathcal{E}} \nu(\mathrm{d}\pi)\, f(\pi). \tag{2.51}$$

By Proposition 2.14, we can rewrite the right-hand side of (2.51) as

$$\int_{\mathcal{E}} \nu(\mathrm{d}\pi)\, f(\pi) = \langle \nu_1, g\rangle \int_0^\infty \rho(\mathrm{d}t) h(t), \tag{2.52}$$

where ρ is the measure in (2.30). We rewrite the left-hand side of (2.51) as

$$\begin{aligned}
&\delta^{-1}\mathbb{E}\big[f(\theta_\delta E^0)\big] \\
&\quad = \delta^{-1} \sum_{m\in 2\mathbb{N}} \mathbb{P}\big[\tau_0 = m\big] h\big(\delta^2 m\big) \mathbb{E}\big[g(\theta_{1/\sqrt{m}}\, E^0) \,\big|\, \tau_0 = m\big] \\
&\quad = \delta^{-1} \sum_{m\in 2\mathbb{N}} \mathbb{P}\big[\tau_0 = m\big] h\big(\delta^2 m\big) \langle \mu_m, g\rangle,
\end{aligned}$$

where $\langle \mu_m, g\rangle$ denotes the integral of g with respect to the measure μ_m defined in (2.49). Using the definition

$$\rho_\delta := \delta^{-1}\sum_{n=1}^\infty \mathbb{P}\big[\tau_0 = 2n\big] \delta_{2\delta^2 n},$$

we can rewrite (2.51) as

$$\int_0^\infty \rho_\delta(\mathrm{d}t) h(t) \langle \mu_{\delta^{-2}t}, g\rangle \underset{\delta\to 0}{\longrightarrow} \langle \nu_1, g\rangle \int_0^\infty \rho(\mathrm{d}t) h(t). \tag{2.53}$$

Assume that $\delta_n \in (0,1]$ satisfy $\delta_n \to 0$. Applying (2.53) with g the function that is constantly one and general h, we see that the measures ρ_{δ_n} converge vaguely to ρ as $n \to \infty$. Let

$$t_n := \inf\big(2\delta_n^2\mathbb{N} \cap [1,\infty)\big)$$

and let $\pi_n \in \Pi(\mathbb{R})$ be the path defined by

$$I_{\pi_n} := 2\delta_n^2\mathbb{N} \quad \text{and} \quad \pi_n(t) := \begin{cases} \langle \mu_{\delta_n^{-2}t}, g\rangle & \text{if } t \geq 1, \\ \langle \mu_{\delta_n^{-2}t_n}, g\rangle & \text{if } t < 1. \end{cases}$$

By Proposition 2.21, the paths π_n are equicontinuous. Since g is bounded and $\mu_{\delta^{-2}t_n}$ is a probability measure, there exists a compact set $C \subset \mathbb{R}$ such that $\pi_n(t) \in C$ for all n and $t \in I_{\pi_n}$. Therefore, by the Arzela-Ascoli theorem (Theorem 1.30), $\{\pi_n : n \in \mathbb{N}\}$ is a precompact subset of $\Pi(\mathbb{R})$. As a consequence, by Lemma 1.2, to show that the paths π_n converge in the topology on $\Pi(\mathbb{R})$ to a limit π, it suffices to show that all cluster points of the sequence π_n are the same.

Assume that a subsequence $\pi_{n(m)}$ converges as $m \to \infty$ to a limit $\pi \in \Pi(\mathbb{R})$. Then clearly $I_\pi = [0,\infty)$. By Lemma 2.20,

$$\int_0^\infty \rho_{\delta_n}(\mathrm{d}t)h(t)\pi_n(t) \underset{n\to\infty}{\longrightarrow} \int_0^\infty \rho(\mathrm{d}t)h(t)\pi(t) \tag{2.54}$$

for each continuous compactly supported $h : (0,\infty) \to \mathbb{R}$. Since the paths π_n are constant on $[0,1]$, their limit π must have the same property. If $h : (0,\infty) \to \mathbb{R}$ is continuous and compactly supported with $\mathrm{supp}(h) \subset [1,\infty)$, then combining (2.53) with (2.54) we see that

$$\int_0^\infty \rho(\mathrm{d}t)h(t)\pi(t) = \langle \nu_1, g\rangle \int_0^\infty \rho(\mathrm{d}t)h(t).$$

Since $\pi : [0,\infty) \to \mathbb{R}$ is a continuous function, the measure ρ in (2.30) has a density with respect to the Lebesgue measure, and h is arbitrary, we conclude that $\pi(t) = \langle \nu_1, g\rangle$ for all $t \geq 1$. Since π is constant on $[0,1]$, this equality extends to $t \geq 0$. This proves that the only cluster point of the sequence π_n is the constant path

$$\pi(t) = \langle \nu_1, g\rangle \qquad (t \geq 0),$$

and hence by Lemma 1.2 $\pi_n \to \pi$ in the topology on $\Pi(\mathbb{R})$. This clearly implies (2.48), so the proof is complete. ∎

We conclude this chapter by providing the proof of Lemma 2.20.

Proof Lemma 2.20 We claim that

$$\sup_{t\in[0,T]\cap I_{\pi_n}} \big|\pi_n(t) - \pi(t)\big| \underset{n\to\infty}{\longrightarrow} 0 \qquad (T < \infty).$$

This can be proved directly by the same sort of arguments as used in the last paragraph of the proof of Lemma 1.25. Alternatively we can extend π_n to

$[0,\infty)$ by linear interpolation and constant extrapolation and use Exercises 1.26 and 1.27 to see that the extended paths also converge to π in the topology on $\Pi(\mathbb{R})$. Then Lemma 1.25 implies that the extended paths converge locally uniformly to π, which implies the claim.

Choosing T large enough such that $\mathrm{supp}(h) \subset [0,T]$ and setting

$$\varepsilon_n := \sup_{t\in[0,T]\cap I_{\pi_n}} \big|\pi_n(t) - \pi(t)\big|,$$

we can now estimate

$$\begin{aligned} &\Big|\int_0^\infty \rho_n(\mathrm{d}t)h(t)\pi_n(t) - \int_0^\infty \rho(\mathrm{d}t)h(t)\pi(t)\Big| \\ &\quad \leq \varepsilon_n \int_0^\infty \rho_n(\mathrm{d}t)\big|h(t)\big| + \Big|\int_0^\infty \rho_n(\mathrm{d}t)h(t)\pi(t) - \int_0^\infty \rho(\mathrm{d}t)h(t)\pi(t)\Big|. \end{aligned} \tag{2.55}$$

Here the second term on the right-hand side tends to zero since $t \mapsto h(t)\pi(t)$ is a continuous compactly supported function and $\rho_n \to \rho$ vaguely. Since $t \mapsto |h(t)|$ is also a continuous compactly supported function, we moreover have that

$$\int_0^\infty \rho_n(\mathrm{d}t)\big|h(t)\big| \underset{n\to\infty}{\longrightarrow} \int_0^\infty \rho(\mathrm{d}t)\big|h(t)\big|,$$

which shows in particular that

$$\limsup_{n\to\infty} \int_0^\infty \rho_n(\mathrm{d}t)\big|h(t)\big| < \infty$$

and hence the first term on the right-hand side of (2.55) tends to zero. ∎

Chapter 3

The tree

3.1 Graphs

By definition, a *graph* is a pair $G = (V, E)$ where V is a set and E is another set whose elements are subsets of V containing precisely two elements. A *finite* graph is a graph for which V (and hence also E) are finite. Elements of V are called *vertices* and elements of E are called *edges*. Two vertices v, w are called *adjacent* if $\{v, w\} \in E$. The number of vertices w that are adjacent to v is called the *degree* of the vertex v. A *graph isomorphism* between two graphs $G = (V, E)$ and $G' = (V', E')$ is a bijection $\psi : V \to V'$ such that $\{\psi(v), \psi(w)\} \in E'$ if and only if $\{v, w\} \in E$. If such an isomorphism exists, the graphs are called *isomorphic*. A *subgraph* of $G = (V, E)$ is a graph $G' = (V', E')$ such that $V' \subset V$ and $E' \subset E$.

Two vertices $v, w \in V$ are *disconnected* if there exists a subset $W \subset V$ such that $v \in V \backslash W$, $w \in W$, and $\{v', w'\} \notin E$ for all $v' \in V \backslash W$ and $w' \in W$. Two vertices that are not disconnected are called *connected*. We write $v \leftrightsquigarrow w$ if v is connected to w. It is easy to see that $\leftrightsquigarrow$ is an equivalence relation on V. The equivalence classes are called the *connected components* of G.

A *cycle* is a nonempty finite connected graph in which each vertex has degree precisely two. A *tree* is a nonempty connected graph G that does not contain cycles, i.e., there exists no subgraph G' of G that is a cycle. In a tree, vertices of degree one are called *leaves* and all other vertices are called *internal vertices*. A *binary* tree is a tree in which each vertex has degree 3 or 1. A *path* is a finite tree in which each vertex has degree at most two.

If $G = (V, E)$ is a path, then we can enumerate the elements of V as $V = \{v_0, \ldots, v_n\}$ with $n \geq 0$ and $v_k \neq v_l$ for all $0 \leq k < l \leq n$, in such a way that $E = \big\{\{v_{k-1}, v_k\} : 1 \leq k \leq n\big\}$. The integer n is called the *length* of the path and v_0 and v_n are called its *endvertices*. If $G = (V, E)$ is an arbitrary graph and $v, w \in V$, then a path *connecting* v and w is a subgraph G' of G such that G' is a path and v and w (which may coincide) are its endvertices.

A *walk* in a graph is an ordered sequence $(v_0, \ldots, v_n)$ of vertices with $n \geq 0$ such that $\{v_{k-1}, v_k\} \in E$ for all $1 \leq k \leq n$. Note that contrary to paths, walks can pass more than once through the same vertex. We call n the *length* and we call v_0 and v_n its *endvertices*. We also say that the walk *connects* v_0 and v_n.

If $G = (V, E)$ and $v, w \in V$, then one can check that the following conditions are equivalent:

(i) $v \leftrightsquigarrow w$,

(ii) there exists a path connecting v and w,

(iii) there exists a walk connecting v and w.

The *graph distance* $d(v, w)$ between two vertices $v, w \in V$ is the length of the shortest walk connecting v and w if such a walk exists, and $d(v, w) := \infty$ if there does not exist a walk connecting v and w. One can check that d is a metric on V. By our earlier remarks, $d(v, w) < \infty$ if and only if $v \leftrightsquigarrow w$. Each walk of length $d(v, w)$ connecting v and w is actually a path. One can check that a graph $G = (V, E)$ is a tree if and only if for each $v, w \in V$, there exists a unique path connecting v and w.

If $T = (V, E)$ is a tree, then for each $x_1, x_2, x_3 \in V$, there exists a unique point $c = c(x_1, x_2, x_3) \in V$ such that

$$d(x_i, x_j) = d(x_i, c) + d(c, x_j) \qquad \forall i, j \in \{1, 2, 3\},\ i \neq j.$$

The point $c(x_1, x_2, x_3)$ is called the *branch point* of x_1, x_2, x_3. Trees can be characterised entirely in terms of the branch point map $c : V^3 \to V$, which must satisfy certain axioms. This leads to the theory of *algebraic trees*, which we unfortunately have no time to elaborate on in these lecture notes.

3.2 Random trees

A *labeled* tree is a tree $T = (V, E)$ with a given vertex set V. We let

$$\mathcal{T}(V) := \big\{(V, E) : (V, E) \text{ is a tree}\big\}$$

denote the set of all trees with a given vertex set V. *Cayley's formula* says that

$$\big|\mathcal{T}(\{1, \ldots, n\})\big| = n^{n-2} \qquad (n \geq 1).$$

If L is a finite set, then a *cladogram* on L is a binary tree (V, E) that has L as its set of leaves. Two cladograms $T = (V, E)$ and $T' = (V', E')$ on the same set L are called *isomorphic* if there exists a graph isomorphism $\psi : V \to V'$ that preserves the leaves, i.e., $\psi(v) = v$ for all $v \in L$. We let $\mathcal{C}(L)$ denote the set of all cladograms (up to isomorphism) on a given set of leaves L. It is easy to see that a cladogram with $n \geq 2$ leaves has $2n - 3$ edges. We can create a cladogram with $n + 1$ leaves from one with n leaves by adding a vertex in the middle of an existing edge and then attaching a new leaf to this vertex. Using this, it is easy to prove the inductive formula

$$\big|\mathcal{C}(\{1, \ldots, n + 1\})\big| = (2n - 3)|\mathcal{C}(\{1, \ldots, n\})| \qquad (n \geq 2).$$

A *rooted* tree is a tree $T = (V, E)$ with one specially marked vertex $\varnothing \in V$, which is called the *root*. Two rooted trees $T = (V, E)$ and $T' = (V', E')$ are called *isomorphic* if there exists a graph isomorphism $\psi : V \to V'$ that preserves the root, i.e., $\psi(\varnothing) = \varnothing$. In a rooted tree, for each $\{v, w\} \in E$, either $d(\varnothing, v) = d(\varnothing, w) - 1$ or $d(\varnothing, v) = d(\varnothing, w) + 1$. In the first case, we say that w is a *child* of v and in the second case, we say that w is the *parent* of v. Note that parents are unique.

Recall that in an unrooted tree, vertices of degree one are called leaves. In rooted trees, the convention is slightly different and the word *leaf* is used for vertices without children. Vertices that are not leaves are called *internal vertices*. For rooted trees, the definition of a binary tree is also somewhat different. A rooted tree is *binary* if each internal vertex has precisely two children. When we make a picture of T, we draw the root at the bottom and we draw the children of a vertex above the vertex.[1] The children, together with all their children and their children, recursively, are called the *descendants* of a vertex. Similarly, the parent, the parent of the parent, and so on are collectively called the *ancestors* of a vertex.

A natural way of attaching labels to the vertices of a rooted tree is as follows. Let $\mathbb{T}$ denote the space of all finite words $\mathbf{i} = i_1 \cdots i_n$ $(n \in \mathbb{N})$ made up from the alphabet $\mathbb{N}_+ = \{1, 2, \ldots\}$. We denote the length of a word $\mathbf{i} = i_1 \cdots i_n$ by $|\mathbf{i}| := n$ and let $\varnothing$ denote the word of length zero. We define the concatenation $\mathbf{ij}$ of two words $\mathbf{i}, \mathbf{j} \in \mathbb{T}$ with $\mathbf{i} = i_1 \cdots i_m$ and $\mathbf{j} = j_1 \cdots j_n$ by $\mathbf{ij} := i_1 \cdots i_m j_1 \cdots j_n$. A *plane tree* is a nonempty subset $\mathbb{U} \subset \mathbb{T}$ with the following properties:

(i) if $i_1 \cdots i_n \in \mathbb{U}$ and $n \geq 1$, then $i_1 \cdots i_{n-1} \in \mathbb{U}$,

(ii) if $i_1 \cdots i_n \in \mathbb{U}$ and $i_n > 1$, then $i_1 \cdots i_{n-1}(i_n - 1) \in \mathbb{U}$.

[1]This is a difference between mathematics and computer science. In computer science, the root of a tree sits at the top and the leaves at the bottom of the tree.

For each word $\mathbf{i} = i_1 \cdots i_n \in \mathbb{T}$ with length $n \geq 1$, we write $\overleftarrow{\mathbf{i}} := i_1 \cdots i_{n-1}$. Then condition (i) says that $\overleftarrow{\mathbf{i}} \in \mathbb{U}$ for all $\mathbf{i} \in \mathbb{U} \backslash \{\varnothing\}$. Note that (i) implies that $\varnothing \in \mathbb{U}$. We view $\mathbb{U}$ as a rooted tree with root $\varnothing$ and set of edges

$$E := \big\{ \{\overleftarrow{\mathbf{i}}, \mathbf{i}\} : \mathbf{i} \in \mathbb{U} \backslash \{\varnothing\} \big\}. \tag{3.1}$$

Because of condition (ii), for each $\mathbf{i} \in \mathbb{U}$, there is a $\kappa_{\mathbf{i}} \in \mathbb{N}$ such that

$$\mathbf{i}j \in \mathbb{U} \quad \text{if and only if} \quad 1 \leq j \leq \kappa_{\mathbf{i}}. \tag{3.2}$$

Then $\overleftarrow{\mathbf{i}}$ is the parent of $\mathbf{i}$ and $\mathbf{i}1, \ldots, \mathbf{i}\kappa_{\mathbf{i}}$ are its children. When we make a picture of $\mathbb{U}$, above each vertex $\mathbf{i}$, we draw its children $\mathbf{i}1, \ldots, \mathbf{i}\kappa_{\mathbf{i}}$ ordered from left to right. Note that in general, when we draw a rooted tree, there is no prescribed order in which to draw the children of a vertex. Therefore, there are different ways of drawing the same rooted tree in the plane. There is (essentially) only one way of drawing a plane tree in the plane, which explains their name. We set

$$\mathcal{U}_n := \big\{ \mathbb{U} : \mathbb{U} \text{ is a plane tree with } n+1 \text{ vertices} \big\}. \tag{3.3}$$

We will be interested in random rooted trees. We let

$$\mathcal{T}_n := \mathcal{T}(\{0, \ldots, n\})$$

denote the set of all labeled trees with vertex set $\{0, \ldots, n\}$. A natural way to create a random rooted tree with $n+1$ vertices is to first pick a labeled tree at random according to the uniform distribution on the set $\mathcal{T}_n$, call 0 the root, and then forget about the labels of all vertices other than the root. Another way is to choose a random plane tree with $n+1$ vertices according to the uniform distribution on $\mathcal{U}_n$, and then again forget about all labels except for the label $\varnothing$ of the root. It is easy to check (for example for $n = 2, 3, 4$ where the calculations can still be done by hand) that these two procedures are not equivalent, i.e., they lead to different distributions on the set of all (non-isomorphic) rooted trees with $n+1$ vertices.

Branching processes also provide a natural way to construct random rooted trees. Let $\rho = (\rho_k)_{k \geq 0}$ be a probability distribution on $\mathbb{N}$, and let $(\kappa_{\mathbf{i}})_{\mathbf{i} \in \mathbb{T}}$ be i.i.d. with common law ρ. Then

$$\mathbb{U} := \big\{ i_1 \cdots i_n \in \mathbb{T} : i_{k+1} \leq \kappa_{i_1 \cdots i_k} \ \forall 0 \leq k < n \big\} \tag{3.4}$$

defines a random plane tree. We call this the *Galton-Watson tree* with *offspring distribution* $\rho = (\rho_k)_{k \geq 0}$. If $\mathbb{U}$ is such a Galton-Watson tree, then setting

$$X_n := \big| \{ \mathbf{i} \in \mathbb{U} : |\mathbf{i}| = n \} \big| \qquad (n \geq 0) \tag{3.5}$$

defines *Galton-Watson branching process* $(X_n)_{n\geq 0}$ with initial state $X_0 = 1$. Such a process describes a population in which each individual, independently of the others, produces a number of children with offspring distribution ρ. Let

$$\langle \rho \rangle := \sum_{k=0}^{\infty} \rho_k k \tag{3.6}$$

denote the mean of the offspring distribution. A branching process is called *subcritical* if $\langle \rho \rangle < 1$, *critical* if $\langle \rho \rangle = 1$, and *supercritical* if $\langle \rho \rangle > 1$. It is well-known that, excluding the trivial case that $\rho_1 = 1$, a Galton-Watson tree $\mathbb{U}$ is a.s. finite if and only if the branching process is subcritical or critical.

There is a convenient way of coding plane trees in terms of random walk excursions. By definition, a *discrete interval* is a set of the form

$$[l : r] := \{l, \ldots, r\} = \{k \in \mathbb{Z} : l \leq k \leq r\} \tag{3.7}$$

with $l, r \in \mathbb{Z}$. A *contour function* (also called *Dyck path*[2]) of length $2n$ is a function $f : [0 : 2n] \to \mathbb{N}$ such that

$$f(0) = f(2n) = 0 \quad \text{and} \quad \big|f(k) - f(k-1)\big| = 1 \quad (0 < k \leq 2n). \tag{3.8}$$

We set

$$\mathcal{D}_n := \big\{ f : f \text{ is a contour function of length } 2n \big\}. \tag{3.9}$$

Each $f \in \mathcal{D}_n$ defines a pseudo-metric d^f on $[0 : 2n]$ by

$$d^f(x, z) = f(x) + f(z) - 2 \inf_{x \leq y \leq z} f(y) \qquad (0 \leq x \leq z \leq 2n).$$

We write $x \sim^f z$ if $d^f(x, z) = 0$ and let $\overline{x} := \{z : x \sim^f z\}$ denote the equivalence class containing x. Then setting $d^f(\overline{x}, \overline{z}) := d^f(x, z)$ $(x, z \in [0 : 2n])$ defines a metric on the set of equivalence classes $V^f := \{\overline{x} : x \in [0 : 2n]\}$. It is not hard to check (picture!) that

d^f is the graph distance on a tree $T(f) = (V^f, E^f)$ with vertex set V^f.

The children of a vertex $\overline{x} \in V^f$ are naturally ordered from left to right, so we can naturally equip $T(f)$ with the structure of a plane tree. Let $\mathbb{U}(f)$ denote the resulting plane tree. It is not hard to see (picture!) that the map

$$\mathcal{D}_n \ni f \mapsto \mathbb{U}(f) \in \mathcal{U}_n \tag{3.10}$$

is a bijection, i.e., for each plane tree $\mathbb{U}$ with $n + 1$ vertices there exists a a unique contour function f of length $2n$ such that $\mathbb{U} = \mathbb{U}(f)$. We call f the *contour function* of $\mathbb{U}$.

[2]After the German mathematician Walther Franz Anton von Dyck.

Lemma 3.1 (Geometric offspring distribution) *Let* $0 < p < 1$ *and let* $\mathbb{U}$ *be a Galton-Watson tree with offspring distribution* $\rho_k = p^k(1-p)$ $(k \geq 0)$. *Then*

$$\mathbb{P}\big[\mathbb{U} \in \cdot \,\big|\, |\mathbb{U}| = n+1\big] \tag{3.11}$$

is the uniform distribution on $\mathcal{U}_n$.

Proof Let $\mathbb{U}$ be a Galton-Watson tree with offspring distribution $\rho_k = p^k(1-p)$ $(k \geq 0)$, and let $U \in \mathcal{U}_n$ be a fixed plane tree U with $n+1$ vertices. Let $\kappa_{\mathbf{i}}$ denote the number of children of $\mathbf{i} \in U$. Then using the fact that $\sum_{\mathbf{i} \in U} \kappa_{\mathbf{i}} = n$, we see that

$$\mathbb{P}\big[\mathbb{U} = U\big] = \prod_{\mathbf{i} \in U} p^{\kappa_{\mathbf{i}}}(1-p) = p^n(1-p)^{n+1} \qquad (n \geq 0,\ U \in \mathcal{U}_n). \tag{3.12}$$

Since the right-hand side depends only on n and not (otherwise) on U, we see that the conditional law of $\mathbb{U}$ given that $\mathbb{U} \in \mathcal{U}_n$ is the uniform distribution on $\mathcal{U}_n$. ∎

Remark Let $S = (S_k)_{k \geq 0}$ be a random walk on $\mathbb{Z}$ with $S_0 = 0$ and transition probabilities

$$\mathbb{P}\big[S_{n+1} = x+1 \,\big|\, S_n = x\big] = p, \quad \mathbb{P}\big[S_{n+1} = x-1 \,\big|\, S_n = x\big] = 1-p. \tag{3.13}$$

Define a random variable N with values in $\mathbb{N} \cup \{\infty\}$ by

$$2N + 1 := \inf\big\{k \geq 1 : S_k = -1\big\}. \tag{3.14}$$

On the event that $N < \infty$, let F be the random element of the space $\mathcal{D}_N$ of contour functions of length $2N$ defined by

$$F(k) := S_k \qquad (0 \leq k \leq 2N). \tag{3.15}$$

Then

$$\mathbb{P}\big[(N, F) = (n, f)\big] = p^n(1-p)^{n+1} \qquad (n \geq 0,\ f \in \mathcal{D}_n), \tag{3.16}$$

where we have used that up to the time $2N+1$, the random walk S makes N upward jumps and $N+1$ downward jumps. Comparing this with (3.12), we see that the planar tree $\mathbb{U}(F)$ associated with the random contour function F is a Galton-Watson tree with geometric offspring distribution. More precisely, if N is the random variable in (3.14) and $\mathbb{U}$ is a Galton-Watson tree with offspring distribution $\rho_k = p^k(1-p)$ $(k \geq 0)$, then

$$\mathbb{P}[\mathbb{U}(F) = U,\ N < \infty] = \mathbb{P}[\mathbb{U} = U]$$

for each finite planar tree U. In particular, if $p < 1/2$, then $N < \infty$ a.s. and $\mathbb{U}(F)$ and $\mathbb{U}$ are equal in law.

For labeled trees, a similar result to Lemma 3.1 is known to hold if the geometric distribution is replaced by a Poisson distribution with mean one.

For each $n \geq 0$, let $\mathbb{U}_n$ be a random plane tree with $n+1$ vertices, chosen according to the uniform distribution on $\mathcal{U}_n$. We will be interested in the shape of the tree $\mathbb{U}_n$ when n is large. In Section 3.7 below, we will see that the trees $\mathbb{U}_n$, properly rescaled, converge in distribution to a continuum random tree whose contour function is the standard Brownian excursion. To formulate this properly, in the next sections, we start studying continuum trees.

3.3 The Gromov-Hausdorff metric

A *homeomorphism* between two topological spaces A and B is a bijection $\psi : A \to B$ such that both ψ and ψ^{-1} are continuous. Two topological spaces are *homeomorphic* if such a homeomorphism exists. An *isometry* between two metric spaces (V, d) and (V', d') is a map $\psi : V \to V'$ such that

$$d'\big(\psi(x), \psi(y)\big) = d(x, y) \qquad (x, y \in V).$$

It is easy to see that if $\psi : V \to V'$ is an isometry and $\psi(V) := \{\psi(x) : x \in V\}$ is the image of V under ψ, then $\psi : V \to \psi(V)$ is a homeomorphism. Two metric spaces (V, d) and (V, d') are called *isometric* if there exists a surjective isometry from V to V'.

We will be interested in the set of all complete separable metric spaces up to isomorphism. We have to be a bit careful with our terminology here, since there is no such thing as the "set of all complete separable metric spaces", just as talking about the "set of all sets" entails the risk of running into paradoxes such as Russel's paradox. We will argue that nevertheless, it is possible to define a set that we can effectively interpret as "the set of all complete separable metric spaces up to isomorphism".

By definition, the *Hilbert cube* is the set $[0, 1]^{\mathbb{N}}$, equipped with the product topology. By Tychonoff's theorem, $[0, 1]^{\mathbb{N}}$ is compact. It is easy to see it is also metrisable. A possible choice for the metric is

$$d(x, y) := \sum_{n=0}^{\infty} 2^{-n}|x_n - y_n|.$$

If $A \subset [0, 1]^{\mathbb{N}}$ is a subset of the Hilbert cube, then we equip A with the induced topology from $[0, 1]^{\mathbb{N}}$. Then A is a metrisable space. It is moreover second countable and hence separable. The following well-known lemma says that each separable metrisable space is homeomorphic to a space A of this form.

Lemma 3.2 (Set of all separable metrisable spaces) *Each subset $A \subset [0,1]^{\mathbb{N}}$, equipped with the induced topology, is a separable metrisable topological space. Moreover, for each separable metrisable topological space B, there exists a subset $A \subset [0,1]^{\mathbb{N}}$ such that A is homeomorphic to B.*

Proof (sketch) A subset of $[0,1]^{\mathbb{N}}$, equipped with the induced topology, is clearly metrisable and first countable, which implies that it must be separable too. Now let B be an arbitrary separable metrisable topological space and let d be a metric generating the topology on B. Since B is separable, there exists a countable dense set $D \subset B$. For each $z \in D$ and $n \geq 1$, let $f_{z,n} : B \to [0,1]$ be the function $f_{z,n}(y) := (1 - nd(x,z)) \vee 0$. Since $D \times \mathbb{N}_+$ is countable, we can enumerate its elements as

$$D \times \mathbb{N}_+ = \big\{(z_i, n_i) : i \in \mathbb{N}\big\}$$

Let $A \subset [0,1]^{\mathbb{N}}$ be the image of B under the map $\psi : B \to [0,1]^{\mathbb{N}}$ defined as

$$\psi(x) := \big(f_{z_i,n_i}(x)\big)_{i \in \mathbb{N}}.$$

Then one can check that ψ is a homeomorphism from B to A. ∎

By Proposition 1.4, if A is a subset of the Hilbert cube $[0,1]^{\mathbb{N}}$, then there exists a complete metric d on A that generates the induced topology from $[0,1]^{\mathbb{N}}$ if and only if A is G_δ-set. With this in mind, we define M to be the set of all pairs (A,d) such that A is a G_δ-subset of $[0,1]^{\mathbb{N}}$ and d is a complete metric on A that generates the induced topology from $[0,1]^{\mathbb{N}}$. As an immediate consequence of Lemma 3.2, we obtain:

Lemma 3.3 (Set of complete separable metric spaces) *Each complete separable metric space (V,d) is isometric to an element $(A,d') \in M$.*

Let us write $(V,d) \sim (V',d')$ to indicate that the complete separable metric spaces (V,d) and (V',d') are isometric. Then $\sim$ is an equivalence relation on M. We let $\mathbb{M}$ denote the set of equivalence classes. For each complete separable metric space (V,d), we define $[V,d] \in \mathbb{M}$ by

$$[V,d] := \big\{(A,d') \in M : (V,d) \sim (A,d')\big\}.$$

We say that (V,d) is a *representant* of the equivalence class $[V,d] \in \mathbb{M}$. We can view $\mathbb{M}$ as the set of all complete separable metric spaces, up to isomorphisms, while avoiding the paradoxes of naive set theory.

Let (V_i, d_i) $(i = 1, 2)$ and (V, d) be three metric spaces. By definition, a *joint isometric embedding* of V_1 and V_2 in V is a pair $\vec{\psi} = (\psi_1, \psi_2)$ of isometries $\psi_i : V_i \to V$ $(i = 1, 2)$. We let

$$\begin{aligned}\mathcal{J}(V_1, V_2) := \big\{(A, d, \psi) : (A, d) \in M,\ \vec{\psi} \text{ is a joint}\\ \text{isometric embedding of } V_1 \text{ and } V_2 \text{ in } A\big\}.\end{aligned} \tag{3.17}$$

We recall that each compact metric space is complete and separable. With this in mind, we let

$$\mathbb{M}_{\rm c} := \big\{[V, d] \in \mathbb{M} : (V, d) \text{ is compact}\big\}$$

denote the set of all compact metric spaces, up to isometry. The *Gromov-Hausdorff metric* is the metric $d_{\rm GH}$ on $\mathbb{M}_{\rm c}$ defined as

$$d_{\rm GH}(V_1, V_2) := \inf_{(A,d,\vec{\psi}) \in \mathcal{J}(V_1,V_2)} d_{\rm H}\big(\psi_1(V_1), \psi_2(V_2)\big), \tag{3.18}$$

where $d_{\rm H}$ is the Hausdorff metric on the space of compact subsets of (A, d), defined in (1.5). The fact that $d_{\rm H}$ is a metric on $\mathbb{M}_{\rm c}$ is proved in [BBI01, Thm 7.3.30]. The Gromov-Hausdorff metric was invented by Edwards in [Edw75] and independently by Gromov in [Gro81]. A good source of information about the Gromov-Hausdorff metric is [BBI01, Chapter 7]. See also [Eva08].

By definition, a *correspondence* between two sets V and $\mathcal{Y}$ is a set $R \subset V \times \mathcal{Y}$ such that for each $x \in V$, there exists at least one $y \in \mathcal{Y}$ such that $(x, y) \in R$, and likewise, for each $y \in \mathcal{Y}$, there exists at least one $x \in V$ such that $(x, y) \in R$. If (V_i, d_i) $(i = 1, 2)$ are metric spaces and R is a correspondence between V_1 and V_2, then the *distortion* of R is defined as

$$\mathrm{dis}(R) := \sup\big\{\big|d_1(x_1, y_1) - d(x_2, y_2)\big| : (x_1, x_2), (y_1, y_2) \in R\big\}.$$

In words, this is how much the distance between two points x_1, y_1 in V_1 can maximally change when we replace them by corresponding points x_2, y_2 in V_2. It is clear that if there exists a correspondence between two metric spaces that has a small distortion, then these metric spaces are "similar". The following result, which we cite from [Eva08, Thm 4.11], relates this to the Gromov-Hausdorff metric.

Proposition 3.4 (Distortions and the Gromov-Hausdorff metric) *For any $[V_1, d_1], [V_2, d_2] \in \mathbb{M}_{\rm c}$, one has*

$$d_{\rm GH}(V_1, V_2) = \tfrac{1}{2} \inf_{R \in \mathcal{R}(V_1,V_2)} \mathrm{dis}(R), \tag{3.19}$$

where $\mathcal{R}(V_1, V_2)$ denotes the set of all correspondences between V_1 and V_2.

Formula (3.19) is often easier to work with as a definition of the Gromov-Hausdorff metric than (3.18), since it is often easier to construct a correspondence between two metric spaces than a joint isometric embedding in a third space.

3.4 The Gromov-weak topology

The *support* of a finite measure μ on a Polish space V, denoted $\mathrm{supp}(\mu)$, is defined as

$$\mathrm{supp}(\mu) := \bigcap \big\{A \subset V : A \text{ is closed and } \mu(V\backslash A) = 0\big\}.$$

It is well-known that $\mu(V\backslash\mathrm{supp}(\mu)) = 0$, so $\mathrm{supp}(\mu)$ is the smallest closed subset of V such that μ is concentrated on it. Moreover, if V' is another Polish space and $\psi : V \to V'$ is a continuous map, then

$$\mathrm{supp}\big(\mu \circ \psi^{-1}\big) = \psi\big(\mathrm{supp}(\mu)\big),$$

i.e., the support of the image measure $\mu \circ \psi^{-1}$ is the image under ψ of the support of μ.

By definition, a *metric measure space* (*mm-space*) is a triple $\mathcal{V} = (V, d, \mu)$ where (V, d) is a complete separable metric space and μ is a probability measure on V (equipped with the Borel-σ-algebra). Two mm-spaces $\mathcal{V} = (V, d, \mu)$ and $\mathcal{V}' = (V', d', \mu')$ are *isomorphic* if there exists a map $\psi : \mathrm{supp}(\mu) \to \mathrm{supp}(\mu')$ such that

$$\mu' = \mu \circ \psi^{-1} \quad \text{and} \quad d'\big(\psi(x), \psi(y)\big) = d(x, y) \quad \forall x, y \in \mathrm{supp}(\mu).$$

We call such a map ψ an *isomorphism* of mm-spaces. In words, the first property says that $\psi : V \to V'$ is *measure-preserving*. The second property says that ψ is an isometry from $\mathrm{supp}(\mu)$ to its image under ψ. Combining this with the measure-preserving property, it is easy to see that $\psi : \mathrm{supp}(\mu) \to \mathrm{supp}(\mu')$ must be surjective, so ψ is an isometry between the metric spaces $(\mathrm{supp}(\mu), d)$ and $(\mathrm{supp}(\mu'), d')$.

It follows from this definition that if (V, d, μ) is an arbitrary mm-space and we set $V' := \mathrm{supp}(\mu)$ and choose for d' and μ' the restrictions of d and μ to $\mathrm{supp}(\mu)$, then (V, d, μ) and (V', d', μ') are isomorphic. Thus, if we are only interested in mm-spaces up to isomorphisms, then we can without loss of generality assume that $\mathrm{supp}(\mu) = V$. There are, nevertheless, sometimes reasons to allow for the case that V is strictly larger than $\mathrm{supp}(\mu)$. For example, it may happen that the metric space (V, d) has a certain additional structure (such as the structure of a real-tree, to be discussed below) that the smaller metric space $(\mathrm{supp}(\mu), d)$ does not have.

Recall that M is the space of all pairs (A, d) such that A is a G_δ-subset of the Hilbert cube $[0,1]^{\mathbb{N}}$ and d is a complete metric on A that generates the induced topology from $[0,1]^{\mathbb{N}}$. In the previous section, we defined an equivalence relation on M by setting $(A, d) \sim (A', d')$ if (A, d) and (A', d') are isometric, and we showed that the resulting set of equivalence classes $\mathbb{M}$ could be interpreted as the set of all complete separable metric spaces, up to isometry.

For metric measure space, we can carry out a similar construction. We let MM denote the space of all triples (A, d, μ) with $(A, d) \in M$ and μ a probability measure on A. We write $(V, d, \mu) \sim (V', d', \mu')$ to indicate that the mm-spaces (V, d, μ) and (V', d', μ') are isomorphic. Then $\sim$ is an equivalence relation on MM. We let $\mathbb{MM}$ denote the set of equivalence classes. For each mm-space (V, d, μ), we define $[V, d, \mu] \in \mathbb{MM}$ by

$$[V, d, \mu] := \big\{(A, d', \mu') \in MM : (V, d, \mu) \sim (A, d', \mu')\big\}.$$

We say that (V, d, μ) is a *representant* of the equivalence class $[V, d, \mu] \in \mathbb{MM}$. Informally, we can view $\mathbb{MM}$ as the set of all mm-spaces, up to isomorphisms.

Let $\mathcal{V} = (V, d, \mu)$ be an mm-space and let $X_1, \ldots, X_m$ be i.i.d. V-valued random variables with common law μ. Then setting

$$D_m[\mathcal{V}](i, j) := d(X_i, X_j) \qquad (1 \leq i, j \leq m)$$

defines a random metric on the finite set $\{1, \ldots, m\}$. We view $D_m[\mathcal{V}]$ as a random variable with values in $\mathbb{R}^{m^2}$, the space of all real functions on $\{1, \ldots, m\}^2$. We cite the following theorem from [GPW09, Thm 1].

Theorem 3.5 (The Gromov-weak topology) *Let $\mathbb{MM}$ be the set of all mm-spaces, up to isomorphisms. Then it is possible to equip $\mathbb{MM}$ with a metric d such that*

(i) $(\mathbb{MM}, d)$ *is a complete separable metric space,*

(ii) $d(\mathcal{V}_n, \mathcal{V}) \to 0$ *if and only if* $\mathbb{P}\big[D_m[\mathcal{V}_n] \in \cdot\,\big] \underset{n\to\infty}{\Longrightarrow} \mathbb{P}\big[D_m[\mathcal{V}] \in \cdot\,\big]$ *for each* $m \geq 1$.

Since a metrisable topology is uniquely characterised by its convergent sequences, property (ii) uniquely characterises a topology on $\mathbb{MM}$. We call this the *Gromov-weak topology*. Property (i) says that $\mathbb{MM}$, equipped with the Gromov-weak topology, is a Polish space. Note that Theorem 3.5 implies that if $\mathcal{V} = (V, d, \mu)$ and $\mathcal{V}' = (V', d', \mu')$ are mm-spaces such that $D_m[\mathcal{V}]$ and

$D_m[\mathcal{V}']$ are equal in law for all $m \geq 1$, then $\mathcal{V}$ and $\mathcal{V}'$ are isomorphic, which in itself is already a nontrivial statement.

There are several possible choices for a metric d on $\mathbb{MM}$ with properties as described in Theorem 3.5. Two possible choices are the Gromov-Prohorov metric and the Gromov-Wasserstein metric; see [GPW09, Prop. 10.5], which moreover lists two further metrics that generate the same topology but are not complete. Another metric, originally introduced in [Gro01, Chapter $3\frac{1}{2}$], was shown to be equivalent to the Gromov-Prohorov metric in [Loh13]. For brevity, we describe only one of these metrics, the Gromov-Prohorov metric.

Recall from (3.17) that $\mathcal{J}(V_1, V_2)$ denotes the set of all joint isometric embeddings $(A, d, \vec{\psi})$ of the metric spaces (V_i, d_i) $(i = 1, 2)$ in a metric space $(A, d) \in M$. By definition, the *Gromov-Prohorov metric* is the metric on $\mathbb{MM}$ defined as

$$d_{\mathrm{GP}}(\mathcal{V}_1, \mathcal{V}_2) := \inf_{(A,d,\vec{\psi}) \in \mathcal{J}(V_1,V_2)} d_{\mathrm{P}}\big(\mu_1 \circ \psi_1^{-1}, \mu_2 \circ \psi_2^{-1}\big), \tag{3.20}$$

where d_{P} is the Prohorov metric on $\mathcal{M}_1(A)$ defined in Section 1.4. The fact that d_{GP} is a metric on $\mathbb{MM}$ is proved in [GPW09, Lemma 5.4], [GPW09, Prop. 5.6] says that the space $(\mathbb{MM}, d_{\mathrm{GP}})$ is complete and separable, and [GPW09, Thm 5] says that $d(\mathcal{V}_n, \mathcal{V}) \to 0$ if and only if $\mathbb{P}\big[D_m[\mathcal{V}_n] \in \cdot\,\big] \Rightarrow \mathbb{P}\big[D_m[\mathcal{V}] \in \cdot\,\big]$ as $n \to \infty$ for each $m \geq 1$. Together, these results imply Theorem 3.5. In particular, they show that d_{GP} generates the Gromov-weak topology.

We will be interested in weak convergence of probability measures on $\mathbb{MM}$. Let $\mathcal{V} = [V, d, \mu]$ be an $\mathbb{MM}$-valued random variable and conditional on $\mathcal{V}$, let $X_1, \ldots, X_m$ be i.i.d. V-valued random variables with common law μ. Then setting

$$D_m[\mathcal{V}](i,j) := d(X_i, X_j) \qquad (1 \leq i, j \leq m)$$

defines a random metric on the finite set $\{1, \ldots, m\}$.

Lemma 3.6 (Convergence in law of random mm-spaces) *Let $\mathcal{V}_n, \mathcal{V}$ be random variables with values in $\mathbb{MM}$. Then the following statements are equivalent:*

(i) $\mathbb{P}\big[\mathcal{V}_n \in \cdot\,\big] \underset{n\to\infty}{\Longrightarrow} \mathbb{P}\big[\mathcal{V} \in \cdot\,\big]$*, where $\Rightarrow$ denotes weak convergence of probability measures on $\mathbb{MM}$, equipped with the Gromov-weak topology.*

(ii) $\mathbb{P}\big[D_m[\mathcal{V}_n] \in \cdot\,\big] \underset{n\to\infty}{\Longrightarrow} \mathbb{P}\big[D_m[\mathcal{V}] \in \cdot\,\big]$ *for each $m \geq 1$.*

Proof (sketch) Let $[V,d,\mu]\in\mathbb{MM}$ and let $\mu^m:=\mu\otimes\cdots\otimes\mu$ denote the product measure on V^m. Each bounded continuous function $\phi:\mathbb{R}^{m^2}\to\mathbb{R}$ defines a function $\Phi^\phi:\mathbb{MM}\to\mathbb{R}$ by

$$\Phi^\phi(V,d,\mu):=\int_{V^n}\mu^m(\mathrm{d}x)\phi\big(d(x_i,x_j)\big)_{1\leq i,j\leq m}$$

Then

$$\Phi^\phi(\mathcal{V})=\mathbb{E}\big[\phi(D_m[\mathcal{V}])\,|\,\mathcal{V}\big],$$

and hence (ii) is equivalent to the statement that

$$\mathbb{E}\big[\Phi^\phi(\mathcal{V}_n)\big]\underset{n\to\infty}{\longrightarrow}\mathbb{E}\big[\Phi^\phi(\mathcal{V})\big]$$

for each $m\geq 1$ and bounded continuous function $\phi:\mathbb{R}^{m^2}\to\mathbb{R}$. Based on Theorem 3.5, one can show that the Gromov-weak topology on $\mathbb{MM}$ is the weakest topology that makes the functions $\Phi^\phi:\mathbb{MM}\to\mathbb{R}$ continuous for all bounded continuous function $\phi:\mathbb{R}^{m^2}\to\mathbb{R}$. In particular, the functions Φ^ϕ are bounded and continuous, so the implication (i)⇒(ii) is trivial.

To prove the converse, one needs to prove that the class of functions of the form Φ^ϕ is *convergence determining*. It is not hard to see that they are *closed under multiplication*, i.e., if $\phi_i:\mathbb{R}^{m_i^2}\to\mathbb{R}$ $(i=1,2)$ are bounded and continuous, then we can find a bounded continuous $\psi:\mathbb{R}^{(m_1+m_2)^2}\to\mathbb{R}$ such that

$$\Phi^{\phi_1}(\mathcal{V})\Phi^{\phi_2}(\mathcal{V})=\Phi^\psi(\mathcal{V}).$$

Since the functions Φ^ϕ also generate the topology on $\mathbb{MM}$, one can now apply an old result of Le Cam [Cam57] to conclude they are convergence determining. For the details, we refer to [Loh13, Cor. 2.8]. It is interesting that the functions Φ^ϕ are not dense in the space of all bounded continuous functions $\Phi:\mathbb{MM}\to\mathbb{R}$, see [Loh13, Remark 2.6]. Very often, one proves that a class of continuous functions is distribution determining by showing that it is dense, but in this case, this approach does not work, even though the function class of interest is distribution determining and even convergence determining. ∎

Remark Let $\mathcal{V}_i=(V_i,d_i,\mu_i)$ $(i=1,2)$ be mm-spaces. Recall from Section 1.4 that $C(\mu_1,\mu_2)$ denotes the space of all *couplings* of μ_1 and μ_2, i.e., $C(\mu_1,\mu_2)$ is the space of all probability measures η on $V_1\times V_2$ whose first and second marginals are μ_1 and μ_2, respectively. For each $\varepsilon>0$, let us define $D_\varepsilon\subset(V_1\times V_2)^2$ by

$$D_\varepsilon:=\big\{\big((x_1,y_1),(x_2,y_2)\big):\big|d_1(x_1,y_1)-d_2(x_2,y_2)\big|\geq\varepsilon\big\}.$$

Then we can define the *distortion* of a coupling $\eta \in C(\mu_1, \mu_2)$ as

$$\mathrm{dis}(\eta) := \inf \big\{ \varepsilon > 0 : \eta \otimes \eta \, (D_\varepsilon) \leq \varepsilon \big\},$$

where $\eta \otimes \eta$ denotes the product measure on $(V_1 \times V_2)^2$. Here are some questions that I do not know the answer to. Does setting

$$d(\mathcal{V}_1, \mathcal{V}_2) := \inf_{\eta \in C(\mu_1, \mu_2)} \mathrm{dis}(\eta) \tag{3.21}$$

define a metric on the space $\mathbb{MM}$? Is $d(\mathcal{V}_k, \mathcal{V}) \to 0$ equivalent to convergence in the Gromov-weak topology? An affirmative answer to these questions would allow one to characterise Gromov-weak in terms of distortions, similar to Proposition 3.4. Anita Winter [personal communication] thinks that, probably, the answer to these questions is positive, and it should not be hard to prove so.

3.5 The four-point condition

In Section 3.1 we defined when a graph $G = (V, E)$ is a tree. We also saw that the vertex set V of a graph, equipped with the graph distance d, forms a metric space (V, d). Since $E = \big\{ \{v, w\} : d(v, w) = 1 \big\}$, all information about the graph $G = (V, E)$ is contained in the metric space (V, d).

In this section, we generalise the concept of a "tree" to more general metric spaces. Let (V, d) be a metric space such that

$$d(x, y) < \infty \quad \forall x, y \in V.$$

We will be interested in the following conditions on (V, d).

(i) *Four-point condition* $d(x_1, x_2) + d(x_3, x_4) \leq \big(d(x_1, x_3) + d(x_2, x_4)\big) \vee \big(d(x_1, x_4) + d(x_2, x_3)\big)$ for all $x_1, x_2, x_3, x_4 \in V$.

(ii) *Branch point condition* For each $x_1, x_2, x_3 \in V$, there exists a $c \in V$ such that $d(x_i, x_j) = d(x_i, c) + d(c, x_j)$ for all $i, j \in \{1, 2, 3\},\ i \neq j$.

As we will see in a moment, the four-point condition (i) already goes a long way towards saying that the metric space (V, d) is, in some way, a tree. By definition, a *weighted graph* is a triple (V, E, ℓ) where $G = (V, E)$ is a graph and $\ell : E \to (0, \infty)$ is a function that assigns to each edge $e \in E$ a positive *length* $\ell(e)$. If $(v_0, \ldots, v_n)$ is a walk in G, then we call

$$\sum_{k=1}^{n} \ell\big(\{v_{k-1}, v_k\}\big)$$

the *length* of the walk. The *length distance* $d(v,w)$ between two vertices $v,w \in V$ is the length of the shortest walk connecting v and w, if such a walk exists, and $d(v,w) := \infty$ otherwise. Note that in the special case that $\ell(e) = 1$ for all $e \in E$, this yields the usual graph distance. One can check that the length distance d is a metric on V and $d(v,w) < \infty$ if and only if v and w are connected. A *weighted tree* is a weighted graph (V,E,ℓ) such that (V,E) is a tree. The following theorem says that finite metric spaces satisfying the conditions (i) and (ii) are, basically, weighted trees.

Theorem 3.7 (Length distance) *Let (V,d) be a finite metric space. Then (V,d) satisfies the four-point condition (i) if and only if there exists a weighted tree $T = (W,E,\ell)$ such that $V \subset W$ and d is the length distance on T. The metric space (V,d) moreover satisfies the branch point condition (ii) if and only if $T = (W,E,\ell)$ can be chosen such that $V = W$.*

Proof Conditions (i) and (ii) are inspired by [ALW17, Def. 1.1], where some further information may be found. See also [Eva08, Theorem 3.38].

To get a better understanding of the four-point condition, let $x_1,\dots,x_4$ be four elements of V. There are three ways of partitioning $\{x_1,\dots,x_4\}$ into two sets of cardinality two. Let us set

$$\begin{aligned} A &:= d(x_1,x_2) + d(x_3,x_4), \\ B &:= d(x_1,x_3) + d(x_2,x_4), \\ C &:= d(x_1,x_4) + d(x_2,x_3). \end{aligned}$$

The four-point condition gives us

$$A \le B \vee C, \quad B \le A \vee C, \quad \text{and} \quad C \le A \vee B. \tag{3.22}$$

By symmetry, we can without loss of generality assume that $A \le B \le C$. Then it is easy to see that (3.22) is equivalent to $B = C$. If $x_1,\dots,x_4 \subset V$ are all different from each other, then we write $\{x_1,x_2\}|\{x_3,x_4\}$ if

$$d(x_1,x_2) + d(x_3,x_4) \le d(x_1,x_3) + d(x_2,x_4) = d(x_1,x_4) + d(x_2,x_3). \tag{3.23}$$

Then the four-point condition says that each set $A \subset V$ with $|A| = 4$ can be partitioned as $A = \{x_1,x_2\} \cup \{x_3,x_4\}$ in such a way that $\{x_1,x_2\}|\{x_3,x_4\}$. If the inequality in (3.23) is strict, then such a partition is unique.

By definition, a *weighted cladogram* on a given set of leaves V is a triple (W,E,ℓ) where (W,E) is a binary tree with set of leaves V and and $\ell : E \to [0,\infty)$ is a function that assigns to each edge $e \in E$ a nonnegative *length* $\ell(e)$. The *length distance* $d(x,y)$ between two vertices $x,y \in W$ in a weighted cladogram is defined just as in the case of weighted graphs. In the present setting, this is only a pseudo-metric since we allow for edges of length zero. For any finite (pseudo-) metric space (V, d), we say that the (pseudo-) metric of V is *generated by a weighted cladogram* if there exists a weighted cladogram (W, E, ℓ) with set of leaves V such that d corresponds to the length distance. In case d is a metric, it is not hard to see (picture!) that this is equivalent to the statement that there exists a weighted tree (W', E', ℓ') such that $V \subset W'$ and d corresponds to the length distance on (W', E', ℓ'). In particular, if the metric of V is generated by a weighted cladogram, then we can always contract all vertices at distance zero from each other in the cladogram to obtain a weighted tree such that the metric on V corresponds to the length distance in this tree.

It is easy to see (picture!) that if a (pseudo-) metric on a finite set V is generated by a weighted cladogram, then it satisfies the four-point condition. We need to prove the converse. If (V, d) is a metric space, then for any $x_1, x_2, x_3 \in V$ we define

$$d(x_1|x_2, x_3) := \tfrac{1}{2}\big[d(x_1, x_2) + d(x_1, x_3) - d(x_2, x_3)\big],$$

which is nonnegative by the triangle inequality. Clearly

$$d(x_1, x_2) = d(x_1|x_2, x_3) + d(x_2|x_1, x_3).$$

If the pseudo-metric on (V, d) is generated by a weighted cladogram (W, E, ℓ), then

$$d(x_1, c) = d(x_1|x_2, x_3), \quad d(x_2, c) = d(x_2|x_1, x_3), \quad \text{and} \quad d(x_3, c) = d(x_3|x_1, x_2),$$

where $c = c(x_1, x_2, x_3)$ denotes the branch point of x_1, x_2, x_3, defined in Section 3.1.

By definition, a *cherry* of a cladogram (W, E) is a pair of leaves $\{c_1, c_2\}$ such that the branch point $c(c_1, c_2, x)$ does not depend on $x \in V\backslash C$. If all edges have positive length, then it is not hard to see that this is equivalent to the statement that the length distance satisfies

$$\{c_1, c_2\}|\{x_1, x_2\} \quad \forall x_1, x_2 \in V\backslash C,\ x_1 \neq x_2. \tag{3.24}$$

Generalising, in any pseudo-metric space (V, d) that satisfies the four-point condition, we take (3.24) as the definition of a cherry.

We can *pick* a cherry. Let (V, d) be a pseudo-metric space that satisfies the four-point condition and let $C = \{c_1, c_2\} \subset V$ be a cherry. We set $V' := (V\backslash C) \cup \{c\}$, where c is an element not contained in $V\backslash C$, and define a symmetric function $d' : V' \times V' \to [0, \infty)$ by

$$d'(c, c) := 0, \quad d'(x, c) := d(x|c_1, c_2), \quad d'(x, y) := d(x, y) \qquad (x, y \in V\backslash C).$$

We call (V', d') the *reduced space* obtained by picking the cherry C. For any finite pseudo-metric space (V, d) with $|V| \geq 3$ that satisfies the four-point condition, we will prove the following claims:

I. (V, d) contains a cherry.

II. If (V', d') is the reduced space obtained by picking a cherry from (V, d), then (V', d') is a pseudo-metric space that satisfies the four-point condition.

We first show how these statements imply what we want, namely, that if a finite pseudo-metric space (V, d) satisfies the four-point condition, then its pseudo-metric is generated by a weighted tree. The proof is by induction on the number of elements of V. If V contains just two elements x_1, x_2, then we connect these by an edge of length $d(x_1, x_2)$ and we are done. Now assume that the statement is proved for all spaces with at most n elements. Let (V, d) be a pseudo-metric space with $n + 1$ elements that satisfies the four-point condition. Then by I, (V, d) contains a cherry $C = \{c_1, c_2\} \subset V$. Let (V', d') with $V' = (V \backslash C) \cup \{c\}$ be the reduced space obtained by picking the cherry. By II, (V', d') is a pseudo-metric space that satisfies the four-point condition. By the induction hypothesis, the pseudo-metric on V'' is generated by a weighted cladogram. We extend this cladogram by connecting c_i $(i = 1, 2)$ to c by edges of length

$$d(c_i, c) := d(c_i, x) - d(x|c_1, c_2) \qquad (i = 1, 2), \tag{3.25}$$

where x is any element of $V \backslash C$. We need to show that our definition does not depend on the choice of x. By symmetry, it suffices to prove the claim for $i = 1$. Filling in the definition of $d(x|c_1, c_2)$, we see that

$$d(c_1, c) = \tfrac{1}{2}\big[d(x, c_1) - d(x, c_2) + d(c_1, c_2)\big].$$

To see that this does not depend on the choice of $x \in V \backslash C$, it suffices to show that for each $x_1, x_2 \in V \backslash C$,

$$d(x_1, c_1) - d(x_1, c_2) = d(x_2, c_1) - d(x_2, c_2),$$

which holds by (3.23) and (3.24). Recalling the definition of the metric d' on V', we see from (3.25) that $d(c_i, x) = d(c_i, c) + d'(c, x)$ $(x \in V \backslash C)$, from which we see that the metric on (V, d) corresponds to the length distance on the exended tree. This completes the induction step.

It remains to prove I and II. We start with the proof of I, which needs some preparations. The four-point condition implies that $\{x_1, x_2\}|\{x_3, x_4\}$ is equivalent to

$$d(x_1, x_3) + d(x_2, x_4) = d(x_1, x_4) + d(x_2, x_3).$$

As a result, it is easy to see that

$$\{a_1, a_2\}|\{b_1, b_2\} \quad \text{and} \quad \{a_2, a_3\}|\{b_1, b_2\} \quad \text{imply} \quad \{a_1, a_3\}|\{b_1, b_2\}. \tag{3.26}$$

Indeed, subtracting the equalities

$$\begin{aligned} & d(a_1, b_1) + d(a_2, b_2) = d(a_1, b_2) + d(a_2, b_1) \\ \text{and} \qquad & d(a_2, b_2) + d(a_3, b_1) = d(a_2, b_1) + d(a_3, b_2) \end{aligned}$$

we obtain

$$\begin{aligned} & d(a_1, b_1) - d(a_3, b_1) = d(a_1, b_2) - d(a_3, b_2) \\ & \quad \Leftrightarrow d(a_1, b_1) + d(a_3, b_2) = d(a_1, b_2) + d(a_3, b_1), \end{aligned}$$

proving (3.26). For any $x_1, \ldots, x_4$, we write

$$d(x_1, x_2|x_3, x_4) := \big[d(x_1, x_3) + d(x_2, x_4)\big] - \big[d(x_1, x_2) + d(x_3, x_4)\big].$$

Using the four-point condition, we make the following simple observations:

$$\begin{aligned} & d(x_1, x_2|x_3, x_4) > 0 \quad \text{implies} \quad \{x_1, x_2\}|\{x_3, x_4\}, \\ & \{x_1, x_2\}|\{x_3, x_4\} \quad \text{implies} \quad d(x_1, x_2|x_3, x_4) \geq 0 \end{aligned} \tag{3.27}$$

Moreover:

$$\begin{aligned} & d(x_1, x_2|x_3, x_4) = d(x_2, x_1|x_4, x_3), \quad \text{and} \\ & \{x_1, x_2\}|\{x_3, x_4\} \quad \text{implies} \quad d(x_1, x_2|x_3, x_4) = d(x_2, x_1|x_3, x_4) \end{aligned} \tag{3.28}$$

We are now ready to prove I. If $|V| = 3$, then by (3.24), trivially every subset $C \subset V$ with $|C| = 2$ is a cherry, so we can without loss of generality assume that $|V| \geq 4$. Since V is finite, we can find $x_1, \ldots, x_4$ that maximise $d(x_1, x_2|x_3, x_4)$. Using (3.27) it is easy to see that we can without loss of generality assume that $\{x_1, x_2\}|\{x_3, x_4\}$. We will show that $C := \{x_1, x_2\}$ is a cherry. By the maximality of $d(x_1, x_2|x_3, x_4)$, for any $x'_3 \in V \backslash C$, we have

$$d(x_1, x_2|x_3, x_4) \geq d(x_1, x'_3|x_3, x_4).$$

This says that

$$\begin{aligned} & \big[d(x_1, x_3) + d(x_2, x_4)\big] - \big[d(x_1, x_2) + d(x_3, x_4)\big] \\ & \quad \geq \big[d(x_1, x_3) + d(x'_3, x_4)\big] - \big[d(x_1, x'_3) + d(x_3, x_4)\big], \end{aligned}$$

which can be simplified to

$$d(x_1, x'_3) + d(x_2, x_4) \geq d(x_1, x_2) + d(x'_3, x_4).$$

By (3.28), we have $d(x_1, x_2|x_3, x_4) = d(x_2, x_1|x_3, x_4)$. Now by the argument we have already seen $d(x_2, x_1|x_3, x_4) \geq d(x_2, x_3'|x_3, x_4)$ implies

$$d(x_2, x_3') + d(x_1, x_4) \geq d(x_2, x_2) + d(x_3', x_4).$$

Combining this with our previous formula, using the four-point condition, it follows that

$$\{x_1, x_2\}|\{x_3', x_4\} \quad \forall x_3' \in V\backslash C.$$

In the same way, we obtain that

$$\{x_1, x_2\}|\{x_3, x_4'\} \quad \forall x_4' \in V\backslash C.$$

Using also $\{x_1, x_2\}|\{x_3, x_4\}$ and (3.26), it follows that

$$\{x_1, x_2\}|\{x_3', x_4'\} \quad \forall x_3', x_4' \in V\backslash C,\ x_3' \neq x_4',$$

which proves that C is a cherry.

It remains to prove II. We need to show that d' satisfies the triangle inequality and the four-point condition. We start with the triangle inequality. For $x_1, x_2, x_3 \in V'$, we need to show that

$$d'(x_1, x_3) \leq d'(x_1, x_2) + d'(x_2, x_3). \tag{3.29}$$

The statement is trivial if $c \notin \{x_1, x_2, x_3\}$. If $c = x_3$, then we need to show that

$$d(x_1|c_1, c_2) \leq d(x_1, x_2) + d(x_2|c_1, c_2). \tag{3.30}$$

We have shown in (3.25) that

$$d(x, c_1, c_2) = d(x, c_1) - d(c, c_1) \qquad (x \in V\backslash C),$$

where $d(c, c_1)$ does not depend on the choice of $x \in V\backslash C$. Using this, we can rewrite (3.30) as

$$d(x_1, c_1) - d(c, c_1) \leq d(x_1, x_2) + d(x_2, c_1) - d(c, c_1),$$

which holds since d satisfies the triangle inequality. By symmetry, the case that $c = x_1$ in (3.29) is the same so it remains to treat the case $c = x_2$. In this case,

$$\begin{aligned} &d'(x_1, x) + d'(x, x_2) = d(x_1|c_1, c_2) + d(x_2|c_1, c_2) \\ &\quad = \tfrac{1}{2}\big[d(x_1, c_1) + d(x_1, c_2) - d(c_1, c_2)\big] + \tfrac{1}{2}\big[d(x_2, c_1) + d(x_2, c_2) - d(c_1, c_2)\big] \\ &\quad = d(x_1, c_1) + d(x_2, c_2) - d(c_1, c_2) \geq d(x_1, x_2), \end{aligned}$$

where in the last two steps we have used that $\{c_1, c_2\}|\{x_1, x_2\}$. This completes the proof of the triangle inequality. To prove also the four-point condition, let $x_1, \ldots, x_4$ be points in V' of which precisely one is the point c. By symmetry, we may assume that $x_1 = c$. Then

$$\begin{aligned} d'(c, x_2) + d'(x_3, x_4) &= d(c_1, x_2) + d(x_3, x_4) - d(c_1, c), \\ d'(c, x_3) + d'(x_2, x_4) &= d(c_1, x_3) + d(x_2, x_4) - d(c_1, c), \\ d'(c, x_4) + d'(x_2, x_3) &= d(c_1, x_4) + d(x_2, x_3) - d(c_1, c). \end{aligned}$$

To check the four-point condition, we must check that two of these expressions are equal while the third one is at most as large as the other two. Since we subtract the same constant in each case, this follows from the fact that d satisfies the four-point condition.

This completes the proof that a finite pseudo-metric space (V, d) satisfies the four-point condition if and only if the pseudo-metric is generated by a weighted cladogram. If (V, d) moreover satisfies the branch-point condition (ii) defined at the beginning of the section, then for each internal vertex x of the cladogram, there is a leaf $x' \in V$ such that $d(x, x') = 0$. From this, the statements of the theorem follow easily. ∎

3.6 Continuum trees

We now turn our attention to true continuum trees. A topological space V is *connected* if there do not exist disjoint open sets O_1, O_2 such that $V = O_1 \cup O_2$. A sufficient condition for this is that V is *path-connected*, which means that for each $x, y \in V$ there exists a continuous map $\alpha : [0, 1] \to V$ such that $\alpha(0) = x$ and $\alpha(1) = y$. Let (V, d) be a metric space such that $d(x, y) < \infty$ for all $x, y \in V$. A *geodesic*[3] in (V, d) is an isometry γ from a compact real interval $[0, T]$ into V. We say that γ is a geodesic from $\gamma(0)$ to $\gamma(T)$. A *segment* is a set of the form

$$[x, y] := \big\{\gamma(t) : t \in [0, T]\big\} \quad \text{where } \gamma \text{ is a geodesic from } x \text{ to } y.$$

In general metric spaces, the notation $[x, y]$ is of course ambiguous, but many metric spaces, such as $\mathbb{R}^d$ or the real-trees that we are about to define, have the property that for each $x, y \in V$, there exists a unique geodesic from x

[3]Sometimes in the literature one finds a weaker definition of a geodesic, which says that for all $t \in [0, T]$, there exists an $\varepsilon > 0$ such that $d\big(\gamma(t_1), \gamma(t_2)\big) = |t_2 - t_1|$ for all $t_1, t_2 \in (t - \varepsilon, t + \varepsilon) \cap [0, T]$. Note that this condition says that γ is "locally" a geodesic, according to our definition.

to y, and in such spaces $[x, y]$ is of course good notation. The *unit circle* is the set $S_1 := \{x \in \mathbb{R}^2 : |x| = 1\}$, equipped with the induced topology from $\mathbb{R}^2$. We will be interested in the following conditions on (V, d), the first two of which are the four-point condition and branch point condition from the previous section.

(i) $d(x_1, x_2) + d(x_3, x_4) \leq \big(d(x_1, x_3) + d(x_2, x_4)\big) \vee \big(d(x_1, x_4) + d(x_2, x_3)\big)$ for all $x_1, x_2, x_3, x_4 \in V$.

(ii) For each $x_1, x_2, x_3 \in V$, there exists a $y \in V$ such that $d(x_i, x_j) = d(x_i, y) + d(y, x_j)$ for all $i, j \in \{1, 2, 3\},\ i \neq j$.

(iii) (V, d) is connected as a topological space.

(iv) For each $x, y \in V$, there exists a geodesic from x to y.

(v) For each $x, z \in V$, there exists a unique geodesic from x to z.

(vi) If $[x, y]$ and $[y, z]$ are segments such that $[x, y] \cap [y, z] = \{y\}$, then $[x, y] \cup [y, z]$ is a segment.

(vii) If $\alpha : [0, T] \to V$ is continuous with $x = \alpha(0)$ and $y = \alpha(T)$, then $[x, y] \subset \{\alpha(t) : t \in [0, T]\}$.

(viii) If $\alpha : [0, T] \to V$ is continuous and injective with $x = \alpha(0)$ and $y = \alpha(T)$, then $[x, y] = \{\alpha(t) : t \in [0, T]\}$.

(ix) There exists no compact subset $C \subset V$ that is homeomorphic to the unit circle.

If (v) does not hold (or we do not yet know that it holds), then property (vi) should be interpreted in the sense that if there exist geodesics $\gamma : [0, T] \to V$ and $\gamma' : [0, T'] \to V$ such that

$$\{\gamma(t) : t \in [0, T]\} \cap \{\gamma'(t) : t \in [0, T']\} = \{\gamma(T)\} = \{\gamma'(0)\},$$

then there exists a geodesic $\gamma'' : [0, T''] \to V$ such that

$$\{\gamma(t) : t \in [0, T]\} \cup \{\gamma'(t) : t \in [0, T']\} = \{\gamma''(t) : t \in [0, T'']\}.$$

Theorem 3.8 (Real-trees) *Let (V, d) be a metric space such that $d(x, y) < \infty$ for all $x, y \in V$. Then conditions (i) and (iii) are equivalent to (iv) and (vi). Moreover, these conditions imply all the other conditions (ii), (v), and (vii)–(ix).*

Metric spaces satisfying the conditions (i)–(ix) are called *real-trees* or $\mathbb{R}$-*trees*. In [LeG05, Def 2.1] and [AG15, Def. 1.1], real-trees are defined by conditions (v) and (viii), which presumably also imply the other conditions.[4] I would not be surprised if also (iv) and (ix) imply all the other conditions. As the example of finite weighted trees shows, conditions (i) and (ii) do not imply (iii). Metric spaces that only satisfy conditions (i) and (ii) are sometimes called *metric trees* [ALW17].

Proof of Theorem 3.8 We will not really prove the theorem but only derive it from the literature. Good references for real-trees are [Chi01, Eva08]. It is shown in [Eva08, Lemma 3.12] that the four-point condition is equivalent to a condition known as 0-hyperbolicity. In [Eva08, Def. 3.15], real-trees are defined by conditions (iv) and (vi). With this in mind, [Eva08, Theorem 3.40] proves that conditions (i) and (iii) are equivalent to (iv) and (vi). The fact that real-trees satisfy (ii) and (v) is now proved in [Eva08, Lemma 3.20]. Property (vii) is proved in [Eva08, Lemma 3.26].

To prove (viii), we first observe that if $\gamma : [0, T] \to V$ is a geodesic and $0 \leq s < u \leq T$, then the restriction of γ to $[s, u]$ is also a geodesic. As a consequence, by (v), if $[x, y]$ is a segment and $x', y' \in [x, y]$, then $[x', y'] \subset [x, y]$. We next observe that if $[x, y]$ is a segment and $\alpha : [0, T] \to V$ is continuous with $\alpha(0), \alpha(T) \in [x, y]$ and $\alpha(t) \notin [x, y]$ for all $0 < t < T$, then we must have $\alpha(0) = \alpha(T)$. Indeed, if we would have $x' := \alpha(0) \neq \alpha(T) =: y'$, then by (vii) we would have $[x', y'] \subset \{\alpha(t) : t \in [0, T]\}$, which contradicts the assumption that $\alpha(t) \notin [x, y]$ for all $0 < t < T$. In words, this says that a continuous curve that leaves a segment and later enters it again must enter the segment in the same point where it left it. Together with (vii), this implies (viii).

To prove (ix), finally, we observe that the unit circle is homeomorphic to the subset $S_1 := \{e^{it} : 0 \leq t < 2\pi\}$ of the complex plane. Each continuous map $\alpha : S_1 \to V$ corresponds to a continuous function $\alpha' : [0, 2\pi] \to V$ such that $\alpha'(0) = \alpha'(2\pi)$. By (viii), we must have

$$\{\alpha'(t) : t \in [0, \pi]\} = \{\alpha'(t) : t \in [\pi, 2\pi]\},$$

which shows that α cannot be one-to-one. ∎

We let

$$\mathbb{T}_c := \big\{[V, d] \in \mathbb{M}_c : (V, d) \text{ is a compact real-tree}\big\}.$$

We cite the following result from [Eva08, Thm 4.23]. I do not know if the analogue statement for the larger space $\mathbb{M}_c$ also holds.

[4] Le Gall [LeG05, Def 2.1] only considers compact real-trees.

Proposition 3.9 (Space of compact real-trees is Polish) *The space $\mathbb{T}_c$, equipped with the Gromov-Hausdorff metric, is a complete separable metric space.*

3.7 Convergence to the CRT

In Section 3.2, we have seen how a contour function f of length $2n$ can be used to define a plane tree. For continuum trees, we can use exactly the same construction. Recall from (2.29) that $\mathcal{E}_1$ denotes the space of all continuous functions $f : [0,1] \to [0,\infty)$ such that $f(0) = f(1) = 0$. Each $f \in \mathcal{E}_1$ defines a pseudo-metric d^f on $[0,1]$ by

$$d^f(x,z) = f(x) + f(z) - 2 \inf_{x \leq y \leq z} f(y) \qquad (0 \leq x \leq z \leq 1).$$

We write $x \sim^f z$ if $d^f(x,z) = 0$ and let $\overline{x} := \{z : x \sim^f z\}$ denote the equivalence class containing x. Then setting $d^f(\overline{x},\overline{z}) := d^f(x,z)$ $(x,z \in [0,1])$ defines a metric on the set of equivalence classes $V^f := \{\overline{x} : x \in [0,1]\}$.

Lemma 3.10 (Real-tree defined by an excursion) *For each $f \in \mathcal{E}_1$, the metric space (V^f, d^f) is a compact real-tree. Moreover, the map $\psi : [0,1] \to V^f$ defined as $\psi(x) := \overline{x}$ $(x \in [0,1])$ is continuous.*

Proof We start by showing that the map ψ is continuous. Let $x \in [0,1]$. We recall from Lemma 1.29 that each continuous function $f : [0,1] \to [0,\infty)$ is uniformly continuous, i.e., for each $\varepsilon > 0$, there exists a $\delta > 0$ such that $|f(y) - f(x)| \leq \varepsilon$ for all $x,y \in [0,1]$ such that $|y - x| \leq \delta$. It follows that

$$d^f(x,z) = \sup_{x \leq y \leq z} \big[\big(f(x) - f(y)\big) + \big(f(z) - f(y)\big)\big] \leq 2\varepsilon$$

for all $x,z \in [0,1]$ such that $|z - x| \leq \delta$, which shows that $d^f(\overline{x}_n, \overline{x}) \to 0$ whenever $x_n \to x$, proving the continuity of ψ.

Since the continuous image of a compact set is compact, it follows that the metric space (V^f, d^f) is compact. For each $0 \leq x < z \leq 1$, setting $\alpha(y) := \overline{y}$ $(x \leq y \leq z)$ defines a continuous function $\alpha : [x,z] \to V^f$ that starts in $\overline{x}$ and ends in $\overline{z}$, proving that V^f is path-connected and hence in particular connected. Therefore, by Theorem 3.8, to prove that (V^f, d^f) is a real-tree it suffices to check the four-point condition (i) from Section 3.6. Let $V := \{\overline{x}_1, \overline{x}_2, \overline{x}_3, \overline{x}_4\} \subset V^f$. By Theorem 3.7, it suffices to show that there exists a weighted tree $T = (W, E, \ell)$ such that $V \subset W$ and the metric d^f on V corresponds to the length distance on T. This is easily verified in a picture, by drawing a tree below the excursion f. ∎

Lemma 3.11 (Map from excursions to real-trees) *The map from $\mathcal{E}_1$ to $\mathbb{T}_\mathrm{c}$ that assigns to an excursion $f \in \mathcal{E}_1$ a compact real-tree $[V^f, d^f] \in \mathbb{T}_\mathrm{c}$ is continuous with respect to the topology of uniform convergence on $\mathcal{E}_1$ and the Gromov-Hausdorff metric on $\mathbb{T}_\mathrm{c}$.*

Proof We apply Proposition 3.4. For each $f \in \mathcal{E}_1$ and $x, z \in [0,1]$, we write $x \sim^f z$ if $d^f(x,z) = 0$ and we let $\overline{x}^f := \{z : x \sim^f z\}$ denote the equivalence class containing x. Given $f, g \in \mathcal{E}_1$, we define a correspondence $R_{f,g}$ between the sets V^f and V^g by

$$R_{f,g} := \big\{(\overline{x}^f, \overline{x}^g) : x \in [0,1]\big\}.$$

The distortion of $R_{f,g}$ is given by

$$\begin{aligned}\mathrm{dis}(R_{f,g}) &= \sup\big\{\big|d^f(\overline{x}^f, \overline{z}^f) - d^g(\overline{x}^g, \overline{z}^g)\big| : x, z \in [0,1]\big\}\\ &= \sup\big\{\big|d^f(x,z) - d^g(x,z)\big| : 0 \le x \le z \le 1\big\}.\end{aligned}$$

Recalling the definition of d^f, we can estimate

$$\begin{aligned}&\big|d^f(x,z) - d^g(x,z)\big|\\ &\quad= \big|f(x) - g(x) + f(z) - g(z) - 2\inf_{x\le y\le z} f(y) + 2\inf_{x\le y\le z} g(y)\big|\\ &\quad\le \big|f(x) - g(x)\big| + \big|f(z) - g(z)\big| + 2\sup_{x\le y\le z}\big|f(y) - g(y)\big|\\ &\quad\le 4\sup_{x\in[0,1]}\big|f(x) - g(x)\big|,\end{aligned}$$

where in the first inequality we have used that

$$\inf_{y\in[x,z]} f(y) - \inf_{y\in[x,z]} g(y) = \sup_{y\in[x,z]}\big[\inf_{y'\in[x,z]} f(y') - g(y)\big] \le \sup_{y\in[x,z]}\big[f(y) - g(y)\big],$$

and similarly with the roles of f and g interchanged. It follows that if $f_n, f \in \mathcal{E}_1$ satisfy $f_n \to f$ uniformly, then by Proposition 3.4

$$d_{\mathrm{GH}}(V^{f_n}, V^f) \le 2\sup_{x\in[0,1]}\big|f_n(x) - f(x)\big| \underset{n\to\infty}{\longrightarrow} 0,$$

which proves the continuity of the map $f \mapsto [V^f, d^f]$. ∎

Instead of the Gromov-Hausdorff metric, we can also use the Gromov-weak topology. For each $f \in \mathcal{E}_1$, let (V^f, d^f) be the compact real-tree defined before. Let $\psi_f : [0,1] \to V^f$ be the map that assigns to each element $x \in [0,1]$ the corresponding equivalence class $\overline{x}^f \in V^f$, let ℓ be the Lebesgue measure on $[0,1]$, and let

$$\mu^f := \ell \circ \psi_f^{-1}$$

denote the image of ℓ under the map ψ_f. Then (V^f, d^f, μ^f) is a metric measure space.

Lemma 3.12 (Map from excursions to metric measure spaces) *The map from $\mathcal{E}_1$ to $\mathbb{MM}$ that assigns to an excursion $f \in \mathcal{E}_1$ the metric measure space $[V^f, d^f, \mu^f] \in \mathbb{MM}$ is continuous with respect to the topology of uniform convergence on $\mathcal{E}_1$ and the Gromov-weak topology on $\mathbb{MM}$.*

Proof Assume that $f_k, f \in \mathcal{E}_1$ satisfy $f_k \to f$ uniformly. Let

$$\mathcal{V}_k := (V^{f_k}, d^{f_k}, \mu^{f_k}) \quad \text{and} \quad \mathcal{V} := (V^f, d^f, \mu^f).$$

We have to show that

$$\mathbb{P}\big[D_m[\mathcal{V}_k] \in \cdot\,\big] \underset{k\to\infty}{\Longrightarrow} \mathbb{P}\big[D_m[\mathcal{V}] \in \cdot\,\big] \qquad (m \geq 1).$$

where $D_m(\mathcal{V})$ is the random metric defined on $\{1, \ldots, m\}$ defined in Section 3.4. Let $X_1, \ldots, X_m$ be i.i.d. uniformly distributed $[0, 1]$-valued random variables. It follows immediately from our definition of μ^f that setting

$$D_m[\mathcal{V}](i, j) := d^f(X_i, X_j) \qquad (1 \leq i, j \leq m)$$

defines a random metric on $\{1, \ldots, m\}$ with the right distribution, and we can define $D_m[\mathcal{V}_k]$ similarly, with f replaced by f_k. By precisely the same estimates as in the proof of Lemma 3.11, we then see that

$$D_m[\mathcal{V}_k](i, j) \underset{k\to\infty}{\longrightarrow} D_m[\mathcal{V}](i, j) \quad \text{a.s.} \qquad (1 \leq i, j \leq m).$$

Since almost sure convergence implies weak convergence in law, this completes the proof. ∎

The *Brownian Continuum Random Tree*, also called *Brownian CRT* or simply *CRT*, is the random compact real-tree $[V, d] \in \mathbb{T}_\mathrm{c}$ defined by

$$(V, d) := (V^\pi, d^\pi),$$

where π is a standard Brownian excursion, i.e., an $\mathcal{E}_1$-valued random variable with law ν_1 as defined in Proposition 2.14. Alternatively, we can also view the CRT as the random metric measure space $[V, d, \mu]$ defined by

$$(V, d, \mu) := (V^\pi, d^\pi, \mu^\pi).$$

The CRT was introduced by David Aldous in [Ald91a, Ald91b, Ald93].

Theorem 3.13 (Convergence to the CRT) *For each $n \geq 0$, let V_n be the vertex set of a random plane tree, chosen according to the uniform distribution on the set $\mathcal{U}_n$ of all plane trees with $n+1$ vertices. Let d_n denote the* *graph distance on* V_n *and let* μ_n *denote the uniform distribution on* $V_n\backslash\{\varnothing\}$*. Then one has*

$$\mathbb{P}\big[[V_n, \tfrac{1}{\sqrt{2n}}d_n] \in \cdot\,\big] \underset{n\to\infty}{\Longrightarrow} \mathbb{P}\big[[V, d] \in \cdot\,\big],$$

where $[V, d]$ *is the Brownian CRT and* $\Rightarrow$ *denotes weak convergence of probability measures on* $\mathbb{M}_\mathrm{c}$ *with respect to the topology generated by the Gromov-Hausdorff metric. Also,*

$$\mathbb{P}\big[[V_n, \tfrac{1}{\sqrt{2n}}d_n, \mu_n] \in \cdot\,\big] \underset{n\to\infty}{\Longrightarrow} \mathbb{P}\big[[V, d, \mu] \in \cdot\,\big],$$

where $[V, d, \mu]$ *is the CRT, viewed as a random metric measure space, and* $\Rightarrow$ *denotes weak convergence of probability measures on* $\mathbb{MM}$ *with respect to the Gromov-weak topology.*

Proof (sketch) We have seen in Section 3.2 that there is a bijection between the set $\mathcal{D}_n$ of all contour functions of length $2n$ and the set $\mathcal{U}_n$ of all plane trees with $n+1$ vertices. Let $F_n : [0:2n] \to \mathbb{N}$ be the contour function of the random plane tree V_n. We let $\overline{F}_n : [0, 2n] \to [0, \infty)$ denote the function F_n, linearly interpolated between integer times, and we let π_n denote the $\mathcal{E}_1$-valued random variable defined as

$$\pi_n(t) := \frac{1}{\sqrt{2n}}\overline{F}_n(2nt) \qquad \big(t \in [0, 1]\big).$$

Then Theorem 2.19 tells us that π_n converges weakly in law to the standard Brownian excursion π. Since $\mathcal{E}_1$ is a Polish space, we can apply Skorohod's representation theorem (Theorem 1.13) to couple the random variables π_n, π such that $\pi_n \to \pi$ a.s. with respect to the topology on $\mathcal{E}_1$, which is the topology of uniform convergence.

Let $(\overline{V}_n, \overline{d}_n) := (V^{\pi_n}, d^{\pi_n})$ be the random compact real-tree defined by the random excursion π_n. Then $(\overline{V}_n, \overline{d}_n)$ is a "linearly interpolated" and rescaled version of (V_n, d_n), where first neighbouring vertices have been connected by a segment of length one, and then all distances have been rescaled by a factor $1/\sqrt{2n}$. Since for our coupling $\pi_n \to \pi$ a.s., we can use Lemma 3.11 to conclude that

$$[\overline{V}_n, \overline{d}_n] \underset{n\to\infty}{\longrightarrow} [X, d] \quad \text{a.s.}$$

with respect to the Gromov-Hausdorff distance. Since a.s. convergence implies convergence in law, this shows in particular that the random variables $[\overline{V}_n, \overline{d}_n]$ converge weakly in law to $[[X, d]$ with respect to the Gromov-Hausdorff distance. Similarly, let $(\overline{V}_n, \overline{d}_n, \overline{\mu}_n) := (V^{\pi_n}, d^{\pi_n}, \mu^{\pi_n})$ be the random metric measure space defined by the random excursion π_n. Then Lemma 3.12 implies that

$$[\overline{V}_n, \overline{d}_n, \overline{\mu}_n] \underset{n\to\infty}{\longrightarrow} [X, d, \mu] \quad \text{a.s.}$$

with respect to the Gromov-weak topology, which again implies convergence in law.

To complete the proof, we need to show that the metric space $[\overline{V}_n, \overline{d}_n]$ is "close" to the metric space $[V_n, \frac{1}{\sqrt{2n}} d_n]$, and the mm-space $[\overline{V}_n, \overline{d}_n, \overline{\mu}_n]$ is "close" to teh mm-space $[V_n, \frac{1}{\sqrt{2n}} d_n, \mu_n]$. For each n, we define $\psi_n : [0,1] \to [0 : 2n]$ by

$$\psi_n(t) := \begin{cases} \lceil 2nt \rceil & \text{if } \pi_n(\lceil 2nt \rceil) \geq \pi_n(\lfloor 2nt \rfloor), \\ \lfloor 2nt \rfloor & \text{if } \pi_n(\lceil 2nt \rceil) \leq \pi_n(\lfloor 2nt \rfloor), \end{cases}$$

and we unambiguously define $\overline{\psi}_n : \overline{V}_n \to V_n$ by

$$\overline{\psi}_n(\overline{x}) := \overline{\psi_n(x)} \qquad \big(x \in [0,1]\big),$$

where $\overline{x} := \{x' : d^{\pi_n}(x, x') = 0\} \in \overline{V}_n = V^{\pi_n}$ denotes the equivalence class containing x, and likewise $\overline{\psi_n(x)}$ denotes the equivalence class containing $\psi_n(x)$. It is not hard to see (picture!) that $\psi_n : \overline{V}_n \to V_n$ maps a point in the interpolated tree $\overline{V}_n$ to the nearest point in V_n that lies above it. As a consequence,

$$\mu_n = \overline{\mu}_n \circ \overline{\psi}_n^{-1}$$

is the uniform distribution on $V_n \backslash \{\varnothing\}$. We can use the map $\overline{\psi}_n$ to define a correspondence R_n between $\overline{V}_n$ and V_n by

$$R_n := \big\{ \big(\overline{x}, \overline{\psi}_n(\overline{x})\big) : \overline{x} \in \overline{V}_n \big\}.$$

The distortion of R_n is $2/\sqrt{2n}$, so using Proposition 3.4, we see that

$$\begin{aligned} & d_{\mathrm{GH}}\big([V_n, \tfrac{1}{\sqrt{2n}} d_n], [X, d]\big) \\ & \quad \leq d_{\mathrm{GH}}\big([V_n, \tfrac{1}{\sqrt{2n}} d_n], [\overline{V}_n, \overline{d}_n]\big) + d_{\mathrm{GH}}\big([\overline{V}_n, \overline{d}_n], [X, d]\big) \underset{n\to\infty}{\longrightarrow} 0 \quad \text{a.s.}, \end{aligned}$$

which implies convergence in law with respect to the topology generated by the Gromov-Hausdorff metric. For the Gromov-weak topology, we argue similarly. We can use the map $\overline{\psi}_n$ to define a coupling η_n between $\overline{\mu}_n$ and μ_n by

$$\eta_n(A) := \overline{\mu}_n\big(\big\{\overline{x} \in \overline{V}_n : \big(\overline{x}, \overline{\psi}_n(\overline{x})\big) \in A\big\}\big),$$

i.e., η_n is the image of $\overline{\mu}_n$ under the map $\overline{x} \mapsto \big(\overline{x}, \overline{\psi}_n(\overline{x})\big)$. The distortion of these couplings clearly tends to zero as $n \to \infty$, so if (3.21) defines a metric that generates the Gromov-weak topology, then we can argue in the same way as for the Gromov-Hausdorff metric. Without (3.21), the argument can also be completed but the technical details are a bit messier. ∎

3.8 Distances in the CRT

Theorem 3.13 says that the Brownian CRT, often simply called the CRT, is the scaling limit of large plane trees, chosen according to the uniform distribution on the set of all plane trees with a given number of vertices. Here the convergence is weak convergence in law, both in the space $\mathbb{M}_c$ of all compact metric spaces (up to isometry), equipped with the Gromov-Hausdorff metric, and in the space $\mathbb{MM}$ of all metric measure spaces (up to isomorphism), equipped with the Gromov-weak topology. By Lemma 3.6, a sequence $\mathcal{V}_n$ of $\mathbb{MM}$-valued random variables converges weakly in law to a limit $\mathcal{V}$ if and only if

$$\mathbb{P}\big[D_m[\mathcal{V}_n] \in \cdot\,\big] \underset{n\to\infty}{\Longrightarrow} \mathbb{P}\big[D_m[\mathcal{V}] \in \cdot\,\big] \qquad (m \geq 1), \tag{3.31}$$

where we recall that if $\mathcal{V} = [V, d, \mu]$ is a random metric measure space (mm-space), then $D_m(\mathcal{V})$ is the random metric on $\{1, \ldots, m\}$ defined by

$$D_m[\mathcal{V}](i,j) := d(X_i, X_j) \qquad (1 \leq i, j \leq m),$$

where $X_1, \ldots, X_m$ are V-valued random variables whose conditional law given $\mathcal{V}$ is the product law

$$\mathbb{P}\big[(X_1, \ldots, X_m) \in \cdot \,\big|\, \mathcal{V}\big] = \underbrace{\mu \otimes \cdots \otimes \mu}_{m \text{ times}}.$$

We can view (3.31) as a sort of "convergence of finite dimensional distributions" for random mm-spaces.

For the Brownian CRT, we can actually give an elegant description of these finite dimensional distributions. Recall that $\mathcal{C}(L)$ denotes the set of all cladograms (up to isomorphism) with a given set of leaves L. Elements of $\mathcal{C}(L)$ are (equivalence classes of) binary trees (V, E) so that $L \subset V$ is the set of leaves of V. In the proof of Theorem 3.7 we also defined *weighted cladograms*, which are triples (V, E, ℓ) such that (V, E) is a cladogram and $\ell : E \to [0, \infty)$ is a function, and we showed that a pseudo-metric d on L satisfies the four-point condition if and only if there exists a weighted cladogram (V, E, ℓ) with set of leaves L such that d corresponds to the length distance on (V, E, ℓ).

Theorem 3.14 (Finite dimensional distributions of the CRT) *For each $m \geq 2$, let (V_m, E_m) be a random cladogram, chosen according to the uniform distribution on $\mathcal{C}(\{1, \ldots, m\})$, and conditional on (V_m, E_m), let ℓ_m be a random variable taking values in $[0, \infty)^E$, whose law has a density with respect to the Lebesgue measure given by*

$$\Big(\sum_{e\in E} l(e)\Big) \prod_{e\in E} e^{-\frac{1}{2}l(e)^2} \qquad \big(l \in [0, \infty)^E\big).$$

Let d_m denote the length distance on (V_m, E_m, ℓ) and let $[\mathcal{V}, d, \mu]$ denote the CRT, viewed as a random metric measure space. Then

$$\mathbb{P}\big[\big(D_m(\mathcal{V})(i,j)\big)_{1\leq i,j\leq m} \in \cdot\,\big] = \mathbb{P}\big[\big(d_m(i,j)\big)_{1\leq i,j\leq m} \in \cdot\,\big].$$

Proof See [LeG05, Section 2.6]. There is actually not a perfect agreement in the literature according to the definition of the Brownian CRT. The CRT as originally introduced by Aldous in [Ald91a, Ald91b, Ald93] uses a different normalisation than Le Gall uses in [LeG05]. I believe tthe theorem above refers to Aldous' normalisation. ∎

Chapter 4

The web

4.1 Arrow configurations

By definition, we call

$$\mathbb{Z}^2_{\text{even}} := \big\{(x,t) \in \mathbb{Z}^2 : x+t \text{ is even}\big\}$$

the *even sublattice* of $\mathbb{Z}^2$. Let $\omega = (\omega_z)_{z\in\mathbb{Z}^2_{\text{even}}}$ be an i.i.d. collection of random variables that are uniformly distributed on $\{-1,+1\}$. We can use ω to define a random directed graph with vertex set $\mathbb{Z}^2_{\text{even}}$ and set of oriented edges

$$\vec{E} := \big\{\big((x,t),(x+\omega_{(x,t)},t+1)\big) : (x,t) \in \mathbb{Z}^2_{\text{even}}\big\}.$$

We call the random directed graph $(\mathbb{Z}^2_{\text{even}}, \vec{E})$ an *arrow configuration.* See Figure 4.1 for a picture.

In Section 1.8, for any metrisable space $\mathcal{X}$, we gave a definition of the path space $\Pi(\mathcal{X})$. Recall that $\overline{I}_\pi$ denotes the domain of a path $\pi \in \Pi(\mathcal{X})$ and that $I_\pi := \overline{I}_\pi \cap \mathbb{R}$. We will especially be interested in the case that the metrisable space $\mathcal{X}$ is $\overline{\mathbb{R}} := [-\infty,\infty]$, the extended real line. Recall that σ_π denotes the starting time of a path π. We let

$$\Pi^\uparrow := \big\{\pi \in \Pi(\overline{\mathbb{R}}) : \overline{I}_\pi = [\sigma_\pi,\infty]\big\}.$$

We call $\Pi^\uparrow$ the space of all *upward paths.* In view of Lemma 1.24, elements of $\Pi^\uparrow$ correspond to continuous functions $\pi : I_\pi \to \overline{\mathbb{R}}$, where I_π is an interval of the form $[\sigma_\pi,\infty)$ if the starting time σ_π is finite, and

$$I_\pi = \mathbb{R} \text{ if } \sigma_\pi = -\infty \quad \text{and} \quad I_\pi = \emptyset \text{ if } \sigma_\pi = +\infty.$$

We will call the point

$$z_\pi := \big(\pi(\sigma_\pi), \sigma_\pi\big)$$

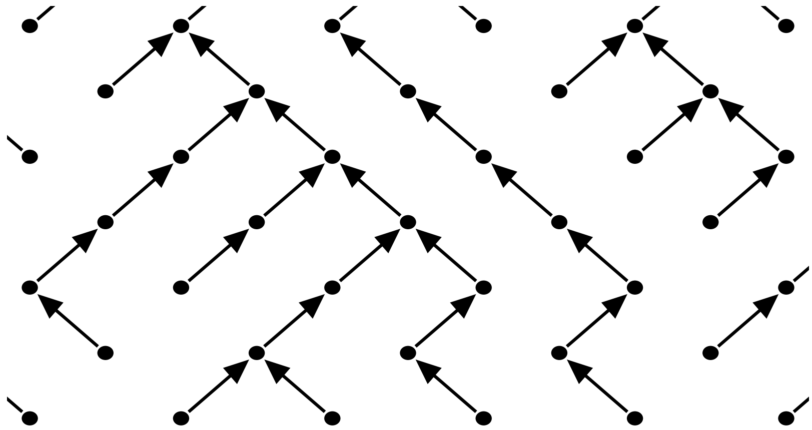

Figure 4.1: An arrow configuration.

the *starting point* of the path π. Note that in general z_π is an element of $\mathcal{R}(\overline{\mathbb{R}})$, the squeezed space defined in Section 1.7. By definition, a *path in the arrow configuration* $(\mathbb{Z}^2_{\rm even}, \vec{E})$, or simply a *path in* ω, is a path $\pi \in \Pi^\uparrow$ with the following properties:

(i) $\big(\pi(\sigma_\pi), \sigma_\pi\big) \in \mathbb{Z}^2_{\rm even}$,

(ii) $\pi(t+1) = \pi(t) + \omega_{(\pi(t),t)}$ $(t \in \mathbb{Z},\ t \geq \sigma_\pi)$,

(iii) $\pi(t+s) = (1-s)\pi(t) + s\pi(t+1)$ $(0 \leq s \leq 1,\ t \in \mathbb{Z},\ t \geq \sigma_\pi)$.

In words, these are upward paths that start at a point in the even sublattice and follow the arrows, with linear interpolation between integer times. We let

$$\mathcal{U} = \mathcal{U}(\omega) := \big\{\pi \in \Pi^\uparrow : \pi \text{ is a path in } \omega\big\}. \tag{4.1}$$

We let $\overline{\mathcal{U}}$ denote the closure of $\mathcal{U}$ in the topology on $\Pi^\uparrow$. The following proposition says that $\overline{\mathcal{U}}$ is a.s. compact and compared to $\mathcal{U}$ only contains a few extra trivial paths. Below, we use the notation $\overline{\mathbb{Z}} := \mathbb{Z} \cup \{-\infty, \infty\}$, i.e., this is the closure of $\mathbb{Z}$ in $\overline{\mathbb{R}}$.

Proposition 4.1 (Compact set of paths) *The closure $\overline{\mathcal{U}}$ of the random set of upward paths $\mathcal{U}$ defined in (4.1) is almost surely a compact subset of $\Pi^\uparrow$. Moreover, almost surely, the set $\overline{\mathcal{U}} \backslash \mathcal{U}$ consists of all paths $\pi \in \Pi^\uparrow$ with $\sigma_\pi \in \overline{\mathbb{Z}}$ and either $\pi(t) = -\infty$ for all $t \in I_\pi$ or $\pi(t) = +\infty$ for all $t \in I_\pi$.*

We postpone the proof of Proposition 4.1 till the end of Section 4.8 and turn to what we are mainly interested in, which is the diffusive scaling limit of arrow configurations. For each $\varepsilon > 0$, we define a diffusive scaling map $\theta_\varepsilon : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$\theta_\varepsilon(x,t) := (\varepsilon x, \varepsilon^2 t) \qquad \big((x,t) \in \mathbb{R}^2\big).$$

Let $\mathcal{R}(\overline{\mathbb{R}})$ be the squeezed space defined in Section 1.7. We extend θ_ε continuously to $\mathcal{R}(\overline{\mathbb{R}})$ in the obvious way, by setting

$$\theta_\varepsilon(\pm\infty,t) := (\pm\infty, \varepsilon^2 t) \quad (t \in \mathbb{R}) \quad \text{and} \quad \theta_\varepsilon(*,\pm\infty) := (*,\pm\infty).$$

For any subset $A \subset \mathcal{R}(\overline{\mathbb{R}})$, we let

$$\theta_\varepsilon(A) := \big\{\theta_\varepsilon(z) : z \in A\big\}$$

denote the image of A under θ_ε. In particular, this notation applies to paths $\pi \in \Pi(\overline{\mathbb{R}})$, which according to their defininition in Section 1.8 correspond to compact subsets of $\mathcal{R}(\overline{\mathbb{R}})$. It is easy to see that $\theta_\varepsilon(\pi) \in \Pi^\uparrow$ for all $\pi \in \Pi^\uparrow$, so the diffusive scaling map $\theta_\varepsilon : \mathcal{R}(\overline{\mathbb{R}}) \to \mathcal{R}(\overline{\mathbb{R}})$ naturally gives rise to a diffusive scaling map from $\Pi^\uparrow$ to $\Pi^\uparrow$ which by a slight abuse of notation we also denote by θ_ε. Going one step further, for any subset $\mathcal{A} \subset \Pi^\uparrow$, we let

$$\theta_\varepsilon(\mathcal{A}) := \big\{\theta_\varepsilon(\pi) : \pi \in \mathcal{A}\big\}$$

denote the image of $\mathcal{A}$ under this map.

In Section 1.6, we equipped the space $\mathcal{K}(\mathcal{X})$ of all compact subsets of a metrisable topological space $\mathcal{X}$ with the Hausdorff topology. We make a simple observation.

Lemma 4.2 (Map acting on compact sets) *Let $\mathcal{X}$ be a metrisable topological space and let $\mathcal{K}(\mathcal{X})$ be the set of all compact subsets of $\mathcal{X}$. Let $\psi : \mathcal{X} \to \mathcal{X}$ be a continuous map and let*

$$\hat\psi(K) := \big\{\psi(x) : x \in K\big\} \qquad \big(K \in \mathcal{K}(\mathcal{X})\big).$$

Then $\psi(K) \in \mathcal{K}(\mathcal{X})$ for all $K \in \mathcal{K}(\mathcal{X})$, and the map $\hat\psi : \mathcal{K}(\mathcal{X}) \to \mathcal{K}(\mathcal{X})$ is continuous with respect to the Hausdorff topology.

Proof The well-known fact that the continuous image of a compact set is itself a compact set has already been mentioned at the beginning of Section 1.2. To see that $\hat\psi : \mathcal{K}(\mathcal{X}) \to \mathcal{K}(\mathcal{X})$ is continuous, assume that $K_n \to K$. Then by Lemma 1.16,

$$\exists C \in \mathcal{K}(\mathcal{X}) \text{ s.t. } K_n \subset C \ \forall n \geq 1 \tag{4.2}$$

and

$$\begin{aligned} K &= \{x \in \mathcal{X} : \exists x_n \in K_n \text{ s.t. } x_n \to x\} \\ &= \{x \in \mathcal{X} : \exists x_n \in K_n \text{ s.t. } x \text{ is a cluster point of } (x_n)_{n\in\mathbb{N}}\}. \end{aligned} \tag{4.3}$$

Since $\hat{\psi}(C)$ is compact and $\hat{\psi}(K_n) \subset \hat{\psi}(C)$ for all $n \geq 1$, by Lemma 1.16, to prove that $\hat{\psi}(K_n) \to \hat{\psi}(K)$, it suffices to show that

$$\begin{aligned} \hat{\psi}(K) &= \{y \in \mathcal{X} : \exists y_n \in \hat{\psi}(K_n) \text{ s.t. } y_n \to y\} \\ &= \{y \in \mathcal{X} : \exists y_n \in \hat{\psi}(K_n) \text{ s.t. } y \text{ is a cluster point of } (y_n)_{n\in\mathbb{N}}\}. \end{aligned}$$

The latter condition can be rewritten as

$$\begin{aligned} \big\{\psi(x) : x \in K\big\} &= \{y \in \mathcal{X} : \exists x_n \in K_n \text{ s.t. } \psi(x_n) \to y\} \\ &= \{y \in \mathcal{X} : \exists x_n \in K_n \text{ s.t. } y \text{ is a cluster point of } \big(\psi(x_n)\big)_{n\in\mathbb{N}}\}. \end{aligned}$$

It therefore suffices to prove that

(i) $\big\{\psi(x) : x \in K\big\} \subset \{y \in \mathcal{X} : \exists x_n \in K_n \text{ s.t. } \psi(x_n) \to y\}$,

(ii) $\{y \in \mathcal{X} : \exists x_n \in K_n \text{ s.t. } y \text{ is a cluster point of } \big(\psi(x_n)\big)_{n\in\mathbb{N}}\} \subset \big\{\psi(x) : x \in K\big\}$.

To prove (i), we use that by (4.3), for each $x \in K$ there exist $x_n \in K_n$ such that $x_n \to x$, and hence $\psi(x_n) \to \psi(x)$ by the continuity of ψ. To prove (ii), assume that $x_n \in K_n$ ($n \in \mathbb{N}$) and there exists a sequence $(n(m))_{m\geq 1}$ with $\lim_{m\to\infty} n(m) = \infty$ such that $y = \lim_{m\to\infty} \psi(x_{n(m)})$. By (4.2) and the compactness of C, by going to a further subsequence if necessary, we can assume without loss of generality that $\lim_{m\to\infty} x_{n(m)} = x$ for some $x \in C$. Then $x \in K$ by (4.3) and $\lim_{m\to\infty} \psi(x_{n(m)}) = \psi(x)$ by the continuity of ψ which shows that $y = \psi(x)$. ∎

As an immediate consequence of Lemma 4.2, we obtain:

Lemma 4.3 (Scaling of paths) *For each $\varepsilon > 0$, the map $\theta_\varepsilon : \Pi^\uparrow \to \Pi^\uparrow$ is continuous.*

Proof Immediate from Lemma 4.2, the continuity of the map $\theta_\varepsilon : \mathcal{R}(\overline{\mathbb{R}}) \to \mathcal{R}(\overline{\mathbb{R}})$, and the fact that in Section 1.8 we viewed the path space $\Pi(\overline{\mathbb{R}})$ as a subset of $\mathcal{K}(\mathcal{R}(\overline{\mathbb{R}}))$ and equipped it with the induced topology from this embedding. ∎

Let $\mathcal{U}$ be the set of all paths in an arrow configuration and let $\overline{\mathcal{U}}$ be its closure, which by Proposition 4.1 is a random compact subset of $\Pi^\uparrow$. Then, since the continuous image of a compact set is compact, by Lemma 4.3, for each $\varepsilon > 0$, the diffusively rescaled set of paths $\theta_\varepsilon(\overline{\mathcal{U}})$ is a random compact subset of $\Pi^\uparrow$. Our aim is to prove that

$$\mathbb{P}\big[\theta_\varepsilon(\overline{\mathcal{U}}) \in \cdot\,\big] \underset{\varepsilon\to 0}{\Longrightarrow} \mathbb{P}\big[\mathcal{W} \in \cdot\,\big] \tag{4.4}$$

where $\Rightarrow$ denotes weak convergence of probability laws on the space $\mathcal{K}(\Pi^\uparrow)$, equipped with the Hausdorff topology, and $\mathcal{W}$ is a random compact subset of $\Pi^\uparrow$ that will be called the *Brownian web*.

4.2 Coalescing Brownian motions

As a first step towards proving (4.4), we start by proving something like convergence of finite dimensional distributions. More precisely, for each $\varepsilon > 0$, we choose finitely many points $z_1^\varepsilon, \ldots, z_n^\varepsilon$ in the diffusively rescaled lattice $\theta_\varepsilon(\mathbb{Z}^2_{\rm even})$, in such a way that

$$(z_1^\varepsilon, \ldots, z_n^\varepsilon) \underset{\varepsilon\to 0}{\longrightarrow} (z_1, \ldots, z_n)$$

for some $z_1, \ldots, z_n \in \mathbb{R}^2$. Letting $\pi_1^\varepsilon, \ldots, \pi_n^\varepsilon$ denote the paths in $\mathcal{U}$ with starting points $z_1^\varepsilon, \ldots, z_n^\varepsilon$, we will argue that $(\pi_1^\varepsilon, \ldots, \pi_n^\varepsilon)$ converges in distribution to a collection of coalescing Brownian motions.

Let $B^1 = (B^1_t)_{t\geq 0}$ and $B^2 = (B^2_t)_{t\geq 0}$ be two independent standard one-dimensional Brownian motions started from initial states $B^i_0 = x_i$ $(i = 1, 2)$, and let

$$\tau := \inf\{t \geq 0 : B^1_t = B^2_t\},$$

which is a.s. finite since $(B^1_t - B^2_t)_{t\geq 0}$ is a Brownian motion (with double the quadratic variation of a standard Brownian motion), and one-dimensional Brownian motion is point recurrent. Let $\tilde{B}^2 = (\tilde{B}^2_t)_{t\geq 0}$ be defined by

$$\tilde{B}^2_t := \begin{cases} B^2_t & \text{if } t \leq \tau, \\ B^1_t & \text{if } \tau \leq t. \end{cases}$$

Then it is easy to check that $\tilde{B}^2$ is a standard Brownian motion. However, B^1 and $\tilde{B}^2$ are of course not independent. The process $(B^1_t, B^2_t)_{t\geq 0}$ is a Markov process that is known as *coalescing Brownian motions.* Although this is not completely immediate from our definition (at least if one wants to give a formal proof), our definition is symmetric in the sense that $(B^2_t, B^1_t)_{t\geq 0}$ is a Markov process with the same transition probabilities as $(B^1_t, B^2_t)_{t\geq 0}$.

We can carry out the same construction for any finite number of Brownian motions, that can moreover start at different times. Let $z_1,\ldots,z_n\in\mathbb{R}^2$ with $z_i=(x_i,s_i)$ $(i=1,\ldots,n)$, and let $B^1,\ldots,B^n$ be independent Brownian motions such that $B^i=(B^i_t)_{t\geq s_i}$ starts at time s_i in $B^i_{s_i}=x_i$. We set $\tau_1:=\infty$, $A_1:=\{(B^1_t,t):s_1\leq t<\infty\}$ and define inductively for $j=2,\ldots,n$

$$\begin{aligned}\tau_j &:= \inf\big\{t\geq s_i:(B^j_t,t)\in A^1\cup\cdots\cup A^{j-1}\big\},\\ A_j &:= \big\{(B^j_t,t):s_j\leq t<\tau_j\big\}.\end{aligned}$$

By the recurrence of one-dimensional Brownian motion, almost surely $\tau_j<\infty$ for all $2\leq j\leq n$. Note that the sets $A_1,\ldots,A_n$ are disjoint. In view of this, we can uniquely define $i(j)\in\{1,\ldots,j-1\}$ by the requirement that

$$(B^j_{\tau_j},\tau_j)\in A^{i(j)}.$$

Using this, we define inductively $\tilde B^1:=B^1$ and

$$\tilde B^j_t:=\begin{cases} B^j_t & \text{if } s_i\leq t\leq\tau_j,\\ \tilde B^{i(j)}_t & \text{if } \tau_j\leq t.\end{cases}$$

We call $\tilde B^1,\ldots,\tilde B^n$ *coalescing Brownian motions* starting from the space-time points $z_1,\ldots,z_n\in\mathbb{R}^2$.

We are now ready to formulate a result about the convergence in law of finitely many paths in an arrow configuration. We have already become used (hopefully!) to the slight abuse of notation by which θ_ε can denote both a diffusive scaling map acting on space-time points, or on sets of space-time points such as paths, or even sets of paths. Taking this one step further, we also denote

$$\theta_\varepsilon(z_1,\ldots,z_n):=\big(\theta_\varepsilon(z_1),\ldots,\theta_\varepsilon(z_n)\big),\quad \theta_\varepsilon(\pi_1,\ldots,\pi_n):=\big(\theta_\varepsilon(\pi_1),\ldots,\theta_\varepsilon(\pi_n)\big)$$

when $z_1,\ldots,z_n$ are space-time points and $\pi_1,\ldots,\pi_n$ are paths.

Proposition 4.4 (Convergence of finite dimensional distributions) *Let $\varepsilon_k>0$ satisfy $\varepsilon_k\to0$. Fix $n\geq1$ and for each k, let $z^k_1,\ldots,z^k_n\in\mathbb{Z}^2_{\rm even}$. Assume that*

$$\theta_{\varepsilon_k}(z^k_1,\ldots,z^k_n)\underset{k\to\infty}{\longrightarrow}(z_1,\ldots,z_n)\in(\mathbb{R}^2)^n.$$

Fix an arrow configuration and for each k, let $\pi^k_1,\ldots,\pi^k_n$ be the unique paths in the arrow configuration with starting points $z^k_1,\ldots,z^k_n$. Then

$$\mathbb{P}\big[\theta_{\varepsilon_k}(\pi^k_1,\ldots,\pi^k_n)\in\cdot\,\big]\underset{k\to\infty}{\Longrightarrow}\mathbb{P}\big[(\pi_1,\ldots,\pi_n)\in\cdot\,\big],$$

where $\Rightarrow$ denotes weak convergence of probability measures on $(\Pi^\uparrow)^n$ and $\pi_1,\ldots,\pi_n$ are coalescing Brownian motions starting from $z_1,\ldots,z_n$.

Proof Our definition of coalescing Brownian motions involved a procedure that started with n independent Brownian motions $(B^1, \ldots, B^n)$ and used them to construct n coalescing Brownian motions $(\tilde{B}^1, \ldots, \tilde{B}^n)$. More formally, we can view $(\tilde{B}^1, \ldots, \tilde{B}^n)$ as the image of $(B^1, \ldots, B^n)$ under a map

$$(\pi_1, \ldots, \pi_n) \mapsto (\tilde{\pi}_1, \ldots, \tilde{\pi}_n) \tag{4.5}$$

that takes n paths $\pi_1, \ldots, \pi_n$ in $\Pi^\uparrow$ with starting points in $\mathbb{R}^2$ and maps them into n new paths $\tilde{\pi}_1, \ldots, \tilde{\pi}_n$ with the same starting points.

For each k, let $(R^{k,1}, \ldots, R^{k,n})$ be a collection of independent random walks started from $z^k_1, \ldots, z^k_n$, and let $(\tilde{R}^{k,1}, \ldots, \tilde{R}^{k,n})$ be its image under the map from (4.5). Then $(\tilde{R}^{k,1}, \ldots, \tilde{R}^{k,n})$ are coalescing random walks. It is easy to see that they are equal in law with $(\pi^k_1, \ldots, \pi^k_n)$. We want to show that

$$\mathbb{P}\big[\theta_{\varepsilon_k}(\tilde{R}^{k,1}, \ldots, \tilde{R}^{k,n}) \in \cdot\,\big] \underset{k\to\infty}{\Longrightarrow} \mathbb{P}\big[(\tilde{B}^1, \ldots, \tilde{B}^n) \in \cdot\,\big].$$

It is easy to see that the diffusive scaling map commutes with the map in (4.5), i.e., the random variable in the left-hand side of our equation is the same as what we would obtain if we first diffusively rescale the independent random walk paths and then apply the map from (4.5).

Weak convergence in law of diffusively rescaled independent random walks to independent Brownian motions follows from Donsker's invariance principle (Theorem 2.1). Using Skorohod's representation theorem (Theorem 1.13), we can couple our random variables such that

$$\theta_{\varepsilon_k}(R^{k,1}, \ldots, R^{k,n}) \underset{k\to\infty}{\longrightarrow} (B^1, \ldots, B^n) \quad \text{a.s.}$$

in the topology on $(\Pi^\uparrow)^n$. If the map in (4.5) would be continuous with respect to the topology on $(\Pi^\uparrow)^n$, then the rest of the proof would now be easy, since we would just apply this map to both sides of our last equation and we would be done.

Things are not quite so simple, however, since it is easy to check (even for $n = 2$) that the map in (4.5) is not continuous with respect to the topology on $(\Pi^\uparrow)^n$. It turns out, however, that $(B^1, \ldots, B^n)$ is almost surely a point of continuity of this map, which is just as good. Here, with a point of continuity of the map in (4.5) we mean, of course, a collection of paths $(\pi_1, \ldots, \pi_n)$ with the property that for each $(\pi^k_1, \ldots, \pi^k_n)$ such that

$$(\pi^k_1, \ldots, \pi^k_n) \underset{k\to\infty}{\longrightarrow} (\pi_1, \ldots, \pi_n),$$

one also has

$$(\tilde{\pi}^k_1, \ldots, \tilde{\pi}^k_n) \underset{k\to\infty}{\longrightarrow} (\tilde{\pi}_1, \ldots, \tilde{\pi}_n).$$

That $(B^1, \ldots, B^n)$ is almost surely a point of continuity follows quite easily from our definitions and from Lemma 4.5 and Exercise 4.6 below. We leave the details to the reader. ∎

Lemma 4.5 (Brownian paths cross when they meet) *Let $B^i = (B^i_t)_{t \geq s_i}$ $(i = 1, 2)$ be independent Brownian motions started from deterministic space-time points $z_i = (x_i, s_i)$ $(i = 1, 2)$, respectively, and let*

$$\tau := \inf\{t \geq s_1 \vee s_2 : B^1_t = B^2_t\}.$$

Then almost surely, for each $\varepsilon > 0$, there exist times $t_-, t_+ \in [\tau - \varepsilon, \tau + \varepsilon] \cap [s_1 \vee s_2, \infty)$ such that

$$B^1_{t_-} < B^2_{t_-} \quad \textit{and} \quad B^1_{t_+} > B^2_{t_+}.$$

Proof By the strong Markov property, $(B^1_{\tau+t} - B^2_{\tau+t})_{t \geq 0}$ is a Brownian motion, so it suffices to prove that a Brownian motion B started in zero changes sign infinitely often in each positive time interval. By symmetry, it suffices to show that $\mathbb{P}[B_t \geq 0 \ \forall 0 \leq t \leq \varepsilon] = 0$ for each $\varepsilon > 0$, which is of course well-known. ∎

Exercise 4.6 (Convergence of meeting times) *Let $\pi_1, \pi_2 \in \Pi^\uparrow$ have starting points $z_i = (x_i, s_i)$ $(i = 1, 2)$, respectively, and assume that their first meeting time*

$$\tau := \inf\{t \geq s_1 \vee s_2 : \pi_1(t) = \pi_2(t)\}$$

satisfies $\tau < \infty$. Assume moreover that for each $\varepsilon > 0$, there exist times $t_-, t_+ \in [\tau - \varepsilon, \tau + \varepsilon] \cap [s_1 \vee s_2, \infty)$ such that

$$\pi_1(t_-) < \pi_2(t_-) \quad \textit{and} \quad \pi_1(t_+) > \pi_2(t_+).$$

Let $\pi^k_1, \pi^k_2 \in \Pi^\uparrow$ satisfy $\pi^k_i \to \pi_i$ $(i = 1, 2)$. Then the first meeting times τ_k of π^k_1 and π^k_2 satisfy $\tau_k \to \tau$. Hint: *First show that generally $\tau \leq \liminf_{k\to\infty} \tau_k$. Then use the assumption about crossing to prove that $\limsup_{k\to\infty} \tau_k \leq \tau_k$.*

4.3 Tightness

As a preparation for the proof of (4.4), in the present section, we derive a criterion for a random subset of the path space $\Pi^\uparrow$ to be compact, and a criterion for tightness of a sequence of probability laws on the space $\mathcal{K}(\Pi^\uparrow)$ of compact sets of $\Pi^\uparrow$.

Lemma 4.7 (Precompactness) *Let $\mathcal{A}$ be a subset of $\Pi^{\uparrow}$. Then $\mathcal{A}$ is precompact if and only if for all $T < \infty$ and $\varepsilon > 0$, there exists a $\delta > 0$ such that*

$$\begin{gathered}\big|\pi(u) - \pi(t)\big| \leq \varepsilon \textit{ for all } \pi \in \mathcal{A} \textit{ and } \sigma_\pi \leq t \leq u \\ \textit{s.t. } (\pi(t), t), (\pi(u), u) \in [-T, T]^2,\ u - t \leq \delta.\end{gathered}$$

Proof We recall that $\Pi^{\uparrow} \subset \Pi(\overline{\mathbb{R}})$, with the induced topology. Let $\phi : \overline{\mathbb{R}} \to [-1, 1]$ be strictly increasing and continuous with $\phi(\pm\infty) = \pm 1$. Then

$$d(x, y) := \big|\phi(x) - \phi(y)\big| \qquad (x, y \in \overline{\mathbb{R}}).$$

is a metric generating the topology on $\overline{\mathbb{R}}$. Since $\overline{\mathbb{R}}$ is compact, by the Arzela-Ascoli theorem (Theorem 1.30), $\mathcal{A}$ is precompact if and only if it is equicontinuous, i.e.,

$$\begin{aligned}\sup \big\{ d\big(\pi(t), \pi(u)\big) : \pi \in \mathcal{A},\ \sigma_\pi \leq t \leq u,& \\ t, u \in [-T, T],\ u - t \leq \delta \big\} \underset{\delta \to 0}{\longrightarrow} 0 &\qquad \forall T < \infty.\end{aligned}$$

In other words, $\mathcal{A}$ is *not* precompact if and only if

$$\begin{gathered}\exists T < \infty \text{ and } \varepsilon > 0 \text{ s.t. } \forall \delta > 0\ \exists \pi \in \mathcal{A} \text{ and } \sigma_\pi \leq t \leq u \\ \text{s.t. } t, u \in [-T, T],\ u - t \leq \delta \text{ and } d\big(\pi(t), \pi(u)\big) > \varepsilon.\end{gathered} \tag{4.6}$$

We claim that this is equivalent to

$$\begin{gathered}\exists S, T < \infty \text{ and } \varepsilon > 0 \text{ s.t. } \forall \delta > 0\ \exists \pi \in \mathcal{A} \text{ and } \sigma_\pi \leq t \leq u \text{ s.t.} \\ t, u \in [-T, T],\ \pi(t), \pi(u) \in [-S, S],\ u - t \leq \delta \text{ and } d\big(\pi(t), \pi(u)\big) > \varepsilon/2.\end{gathered} \tag{4.7}$$

The implication (4.7)⇒(4.6) is trivial. To prove the converse, assume that (4.6) holds. Making ε smaller if necessary, we can without loss of generality assume that $0 < \varepsilon < 1$. In view of our choice of the metric d, we can define $S > 0$ by $d(\pm S, \pm\infty) = \varepsilon/2$. If $\pi(t), \pi(u) \in [-S, S]$ already holds we are done. Otherwise, we are in one of the following cases: 1. $\pi(t) \in (S, \infty]$, 2. $\pi(t) \in (S, \infty]$, 3. $\pi(u) \in (S, \infty]$, 4. $\pi(u) \in (S, \infty]$. Let us assume that we are in case 1. Then we must have $\pi(u) \in [-\infty, S)$ and $d(\pi(u), S) > \varepsilon/2$. By continuity, there must be some $t' \in [t, u]$ such that $\pi(t') = S$. Replacing t by t', we see that (4.7) holds. The other three cases are similar.

Replacing S and T by $S \vee T$ if necessary, we see (4.7) is equivalent to

$$\begin{gathered}\exists T < \infty \text{ and } \varepsilon > 0 \text{ s.t. } \forall \delta > 0\ \exists \pi \in \mathcal{A} \text{ and } \sigma_\pi \leq t \leq u \text{ s.t.} \\ (\pi(t), t), (\pi(u), u) \in [-T, T]^2,\ u - t \leq \delta \text{ and } d\big(\pi(t), \pi(u)\big) > \varepsilon/2.\end{gathered}$$

Using the fact that for each $T < \infty$, there exist constants $0 < c < C < \infty$ such that

$$c|x - y| \leq d(x,y) \leq C|x - y| \qquad \big(x, y \in [-T, T]\big),$$

the claim of the lemma now follows. ∎

Proposition 4.8 (Almost sure precompactness) *Let $\mathcal{A}$ be a random subset of $\Pi^{\uparrow}$. Then $\mathcal{A}$ is almost surely a precompact subset of $\Pi^{\uparrow}$ if and only if*

$$\begin{aligned}\mathbb{P}\big[\, \big|\pi(u) - \pi(t)\big| \geq \varepsilon \textit{ for some } \pi \in \mathcal{A} \textit{ and } \sigma_\pi \leq t \leq u \\ \textit{s.t. } (\pi(t), t) \in [-T, T]^2,\ u - t \leq \delta\big] \underset{\delta \to 0}{\longrightarrow} 0 \quad \forall T < \infty,\ \varepsilon > 0.\end{aligned}$$

Proof Let $A^{\delta}_{T,\varepsilon}$ denote the event that

$$\begin{aligned}\big|\pi(u) - \pi(t)\big| \geq \varepsilon \text{ for some } \pi \in \mathcal{A} \text{ and } \sigma_\pi \leq t \leq u \\ \text{s.t. } (\pi(t), t) \in [-T, T]^2,\ u - t \leq \delta.\end{aligned}$$

Then $\delta \leq \delta'$ implies $A^{\delta}_{T,\varepsilon} \subset A^{\delta'}_{T,\varepsilon}$ and $A_{T,\varepsilon} := \bigcap_{\delta > 0} A^{\delta}_{T,\varepsilon}$ is the event that

$$\begin{aligned}\forall \delta > 0\ \exists \pi \in \mathcal{A} \text{ and } \sigma_\pi \leq t \leq u \\ \text{s.t. } (\pi(t), t) \in [-T, T]^2,\ u - t \leq \delta, \text{ and } \big|\pi(u) - \pi(t)\big| \geq \varepsilon.\end{aligned}$$

The assumption of the proposition implies that $P(A_{T,\varepsilon}) = 0$ for all $T < \infty$ and $\varepsilon > 0$. In particular, if $(T_n)_{n \geq 1}$ and $(\varepsilon_m)_{m \geq 1}$ are sequences of positive constants such that $T_n \to \infty$ and $\varepsilon_n \to 0$, then

$$\mathbb{P}\big(\bigcup_{n \geq 1} \bigcup_{m \geq 1} A_{T_n, \varepsilon_m}\big) = 0,$$

which shows that almost surely, for all $n \geq 1$ and $m \geq 1$, there exists a $\delta > 0$ such that

$$\begin{aligned}\big|\pi(u) - \pi(t)\big| < \varepsilon_m \text{ for all } \pi \in \mathcal{A} \text{ and } \sigma_\pi \leq t \leq u \\ \text{s.t. } (\pi(t), t) \in [-T_n, T_n]^2,\ u - t \leq \delta.\end{aligned}$$

By Lemma 4.7, it follows that $\mathcal{A}$ is almost surely precompact.

On the other hand, if the assumption of the proposition does not hold, then the event $A_{T,\varepsilon}$ has positive probability for some $T < \infty$ and $\varepsilon > 0$, which by Lemma 4.7 implies that $\mathcal{A}$ is with positive probability not precompact. ∎

Proposition 4.9 (Tightness of random compact sets of paths) *Let $\mathcal{K}(\Pi^\uparrow)$ be the set of compact subsets of $\Pi^\uparrow$, equipped with the Hausdorff topology. Let $(\mathcal{A}_n)_{n\geq 1}$ be a sequence of random variables with values in $\mathcal{K}(\Pi^\uparrow)$. Then the probability laws $\big(\mathbb{P}[\mathcal{A}_n \in \cdot\,]\big)_{n\geq 1}$ are tight if and only if*

$$\sup_{n\geq 1} \mathbb{P}\big[\,\big|\pi(u)-\pi(t)\big| \geq \varepsilon \text{ for some } \pi \in \mathcal{A}_n \text{ and } \sigma_\pi \leq t \leq u \\ \text{s.t. } (\pi(t),t) \in [-T,T]^2,\ u-t \leq \delta\big] \underset{\delta\to 0}{\longrightarrow} 0 \quad \forall T<\infty,\ \varepsilon>0.$$

Proof For brevity, we skip the proof. ∎

4.4 The Brownian web

Let $\mathcal{D} \subset \mathbb{R}^2$ be countable. Since $\mathcal{D}$ is countable, we can enumerate it as $\mathcal{D} := \{z_i : i \geq 1\}$ where $(z_i)_{i\geq 1}$ be a sequence of space-time points $z_i \in \mathbb{R}^2$. Then for each $n \geq 1$, we can construct a collection of random paths $(\pi_1, \ldots, \pi_n)$ that are distributed as coalescing Brownian motions starting from $(z_1, \ldots, z_n)$. Since these laws are consistent, by Kolmogorov's extension theorem, we can construct a random collection of paths $(\pi_z)_{z\in\mathcal{D}}$ such that for each finite set $\Delta \subset \mathcal{D}$, the paths $(\pi_z)_{z\in\Delta}$ that are distributed as coalescing Brownian motions starting from the points in Δ. We call $(\pi_z)_{z\in\mathcal{D}}$ a *collection of coalescing Brownian motions* started from the countable set $\mathcal{D}$.

Proposition 4.10 (Precompactness) *Let $(\pi_z)_{z\in\mathcal{D}}$ be a collection of coalescing Brownian motions started from a countable set $\mathcal{D} \subset \mathbb{R}^2$. Then $\{\pi_z : z \in \mathcal{D}\}$ is almost surely a precompact subset of $\Pi^\uparrow$.*

Proof (sketch) We apply Proposition 4.8 to $\mathcal{A} := \{\pi_z : z \in \mathcal{D}\}$. Fix $T < \infty$ and $\varepsilon, \delta > 0$ and consider the grid

$$\mathcal{G}_{\varepsilon,\delta} := \big\{(\tfrac{1}{3}k\varepsilon, l\delta) : k, l \in \mathbb{Z}\big\}.$$

Since $\mathcal{G}_{\varepsilon,\delta}$ is countable, we can add coalescing Brownian motions $\{\pi_z : z \in \mathcal{G}_{\varepsilon,\delta}\}$ starting from any point in $\mathcal{G}_{\varepsilon,\delta}$. Since paths in $\mathcal{A}$ cannot cross paths in $\{\pi_z : z \in \mathcal{G}_{\varepsilon,\delta}\}$, it is not hard to see that almost surely on the event

$$\big|\pi(u)-\pi(t)\big| \geq \varepsilon \text{ for some } \pi \in \mathcal{A} \text{ and } \sigma_\pi \leq t \leq u \\ \text{s.t. } (\pi(t),t) \in [-T,T]^2,\ u-t \leq \delta$$

one has that

$$\big|\pi_{(x,s)}(s+r) - x\big| \geq \tfrac{1}{3}\varepsilon \text{ for some } (x,s) \in \mathcal{G}_{\varepsilon,\delta} \cap [-T,T]^2 \text{ and } r \in [0,2\delta]. \tag{4.8}$$

Using the reflection principle, one can show that if B is a standard Brownian motion, then

$$\mathbb{P}\big[\sup_{r\in[0,2\delta]}|B_r|\geq\tfrac{1}{3}\varepsilon\big]\leq C\,e^{-c\varepsilon^2/\delta},$$

for some $C<\infty$ and $c>0$. A simple union bound then tells us that the probability of the event in (4.8) can be estimated from above by

$$C_T\varepsilon^{-1}\delta^{-1}e^{-c\varepsilon^2/\delta}$$

for some $C_T<\infty$ and $c>0$. This quantity goes to zero as $\delta\to 0$ for fixed $T<\infty$ and $\varepsilon>0$, so by Proposition 4.8 we conclude that $\{\pi_z : z\in\mathcal{D}\}$ is almost surely precompact. ∎

Let $(\pi_z)_{z\in\mathcal{D}}$ be a collection of coalescing Brownian motions started from a countable dense set $\mathcal{D}\subset\mathbb{R}^2$. By Proposition 4.10, we can define a random compact subset $\mathcal{W}\subset\Pi^\uparrow$ by setting

$$\mathcal{W}:=\overline{\{\pi_z : z\in\mathcal{D}\}},\tag{4.9}$$

where $\overline{\mathcal{A}}$ denotes the closure of a set $\mathcal{A}\subset\Pi^\uparrow$. We will later see that this definition does not depend on the choice of the countable dense set $\mathcal{D}\subset\mathbb{R}^2$. We will call $\mathcal{W}$ the *Brownian web*. Our aim is to show that $\mathcal{W}$ is the limiting object in (4.4). For that, we need the following proposition.

Proposition 4.11 (Tightness of rescaled arrow configurations) *Let $\mathcal{U}$ be the set of all paths in an arrow configurations and let $\overline{\mathcal{U}}$ be its closure. Let $\varepsilon_n>0$ be positive constants such that $\varepsilon_n\to 0$. The the probability laws*

$$\big(\mathbb{P}[\theta_{\varepsilon_n}(\overline{\mathcal{U}})\in\cdot\,]\big)_{n\geq 1}$$

on $\mathcal{K}(\Pi^\uparrow)$ are tight.

Proof (crude sketch) One needs to check the tightness criterion of Proposition 4.9. This is very similar to the proof of Proposition 4.10. One uses convergence of finite dimensional distributions (Proposition 4.4 and then uses a grid as in the proof of Proposition 4.10 to estimate the event in Proposition 4.9. We refer to [FINR04, Prop. B2] and [SSS16, Prop. 6.6.4] for details. ∎

We let

$$\Pi^\uparrow_{\rm triv}:=\big\{\pi\in\Pi^\uparrow:\pi(t)=-\infty\ \forall t\in I_\pi\big\}\cup\big\{\pi\in\Pi^\uparrow:\pi(t)=+\infty\ \forall t\in I_\pi\big\}$$

denote the set of trivial paths (with arbitrary starting times $\sigma_\pi\in\overline{\mathbb{R}}$ that are constantly $-\infty$ or $+\infty$.

Lemma 4.12 (Trivial paths) *Let $(\pi_z)_{z\in\mathcal{D}}$ be a collection of coalescing Brownian motions started from a countable dense set $\mathcal{D}\subset\mathbb{R}^2$ and let $\mathcal{W}:=\overline{\{\pi_z : z\in\mathcal{D}\}}$. Then $\Pi^{\uparrow}_{\rm triv}\subset\mathcal{W}$ and each $\pi\in\mathcal{W}\backslash\Pi^{\uparrow}_{\rm triv}$ satisfies $\pi(t)\in\mathbb{R}$ for all $t\in I_\pi$.*

Proof Fix $s\in\mathbb{R}$ and choose $z_n\in\mathcal{D}$ such that $z_n\to(s,-\infty)$ as $n\to\infty$. By Proposition 4.10, by going to a subsequence if necessary, we can assume that $\pi_{z_n}\to\pi$ for some $\pi\in\Pi^{\uparrow}$ with starting time $\sigma_\pi=s$. Then $\pi_{z_n}(t)\to\pi(t)$ for all $t>s$. Since π_{z_n} is a Brownian motion starting from z_n and $z_n\to(s,-\infty)$, we have that the law of $\pi_{z_n}(t)$ converges weakly to the delta-measure on $-\infty$, for each $t>s$. It follows that π is the trivial path defined by $\sigma_\pi=s$ and $\pi(t)=-\infty$ for all $t\in[s,\infty)$. In the same way, we see that $\mathcal{W}$ contains all trivial paths π with $\sigma_\pi\in\mathbb{R}$ and $\pi(t)=-\infty$ for all $t\in I_\pi$. Since $\mathcal{W}$ is closed, it also contains all limits of such paths, so letting $s\to\infty$ or $s\to-\infty$ we see that $\mathcal{W}$ also contains all trivial paths with $\sigma_\pi=-\infty$ and either $\pi(t)=-\infty$ for all $t\in\mathbb{R}$ or $\pi(t)=+\infty$ for all $t\in\mathbb{R}$, as well as the trivial path with $\sigma_\pi=+\infty$. This completes the proof that $\Pi^{\uparrow}_{\rm triv}\subset\mathcal{W}$.

To complete the proof, we must show that if $\pi\in\mathcal{W}$ satisfies $\pi(t)\in\mathbb{R}$ for some $t\in I_\pi$, then $\pi(t)\in\mathbb{R}$ for all $t\in I_\pi$. Assume that $\pi\in\mathcal{W}$ satisfies $\pi(t)\in\mathbb{R}$ for some $t\in I_\pi$. Choose $z_n=(x_n,s_n)\in\mathcal{D}$ with $s_n<t$ such that $z_n\to(\infty,s)$ for some $s\in\mathbb{R}$. Then π_{z_n} is a Brownian motion started from z_n. By our previous arguments, $\pi_{z_n}(t)\to\infty$ a.s. so $\pi(t)<\pi_{z_n}(t)$ for all n large enough. Since coalescing Brownian motions cannot cross each other, it follows that $\pi(u)\leq\pi_{z_n}(u)<\infty$ for all $u\geq s_n\vee\sigma_\pi$ and for all n large enough. Since s is arbitrary, it follows that $\pi(t)<\infty$ for all $t\in I_\pi$ and by a symmetric argument also $-\infty<\pi(t)$ for all $t\in I_\pi$. ∎

4.5 Dual arrow configurations

By definition, we call

$$\mathbb{Z}^2_{\rm odd}:=\big\{(x,t)\in\mathbb{Z}^2 : x+t\text{ is odd}\big\}$$

the *odd sublattice* of $\mathbb{Z}^2$. In Section 4.1, we showed how an i.i.d. collection $\omega=(\omega_z)_{z\in\mathbb{Z}^2_{\rm even}}$ of uniformly distributed $\{-1,+1\}$-valued random variables defines a random directed graph $(\mathbb{Z}^2_{\rm even},\vec{E})$ that we called an arrow configuration. Given ω, we define $\hat\omega=(\hat\omega_z)_{z\in\mathbb{Z}^2_{\rm odd}}$ by

$$\hat\omega_{(x,t+1)}=\omega_{(x,t)}\qquad\big((x,t)\in\mathbb{Z}^2_{\rm even}\big).$$

We can use $\hat{\omega}$ to define a random directed graph with vertex set $\mathbb{Z}^2_{\text{odd}}$ and set of oriented edges

$$\vec{F} := \Big\{ \big((x,t), (x - \omega_{(x,t)}, t-1) \big) : (x,t) \in \mathbb{Z}^2_{\text{odd}} \Big\}.$$

We call the random directed graph $(\mathbb{Z}^2_{\text{odd}}, \vec{F})$ the *dual arrow configuration* associated with the original ("forward") arrow configuration $(\mathbb{Z}^2_{\text{even}}, \vec{E})$. The dual arrows are uniquely characterised in terms of the forward arrows by the property that dual arrows and forward arrows do not cross. See Figure 4.2 for a picture.

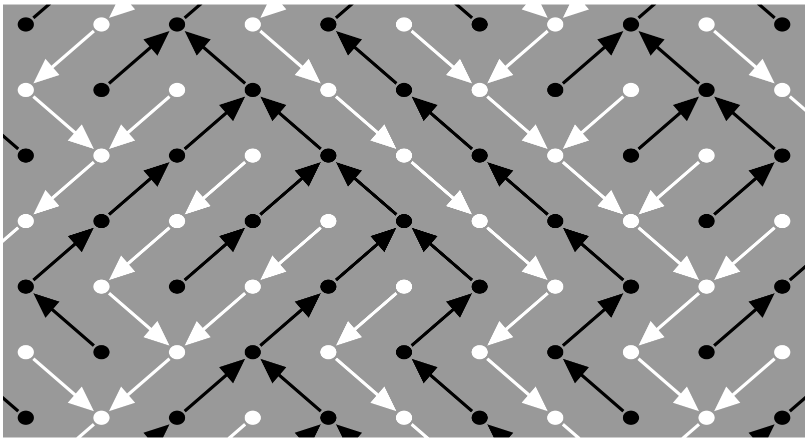

Figure 4.2: An arrow configuration (black) and its dual (white).

Recall that in general, σ_π and τ_π denote the starting and final time of a path $\pi \in \Pi(\overline{\mathbb{R}})$. In particular, we define

$$\Pi^{\downarrow} := \big\{ \pi \in \Pi(\overline{\mathbb{R}}) : \overline{I}_\pi = [-\infty, \tau_\pi] \big\}.$$

We call $\Pi^{\downarrow}$ the space of all *downward paths.* Clearly, $\Pi^{\downarrow}$ is equal to $\Pi^{\uparrow}$ after a rotation over 180 degrees. When no confusion can arrive,[1] we will call the point

$$z_\pi := \big(\pi(\sigma_\pi), \sigma_\pi \big)$$

the *starting point* of a downward path $\pi \in \Pi^{\downarrow}$. We define a *downward path in the dual arrow configuration* $(\mathbb{Z}^2_{\text{odd}}, \vec{F})$, or simply a *path in* $\hat{\omega}$ in exactly the

[1]We have to be careful since the intersection of $\Pi^{\uparrow}$ and $\Pi^{\downarrow}$ is not empty, but consists of all bi-infinite paths for which $\sigma_\pi = -\infty$ and $\tau_\pi = \infty$. As we will see in a moment, however, there are no nontrivial bi-infinite paths in an arrow configuration.

same way as we defined upward paths in the forward arrow configuration. We let

$$\mathcal{U}' = \mathcal{U}'(\hat{\omega}) := \big\{\pi \in \Pi^{\downarrow} : \pi \text{ is a path in } \hat{\omega}\big\} \tag{4.10}$$

denote the set of all downward paths in the dual arrow configuration and we let $\overline{\mathcal{U}'}$ denote the closure of $\mathcal{U}'$ in the topology on $\Pi^{\downarrow}$.

4.6 The dual Brownian web

We have already introduced notation for the diffusive scaling map θ_ε which may be applied to points $z = (x,t)$ in space-time $\mathcal{R}(\overline{\mathbb{R}})$, to subsets of space-time such as paths, and even to sets of paths. We will use similar notation for the map

$$\mathcal{R}(\overline{\mathbb{R}}) \ni (x,t) \mapsto -(x,t) = (-x,-t) \in \mathcal{R}(\overline{\mathbb{R}}).$$

Thus, for any set $A \subset \mathcal{R}(\overline{\mathbb{R}})$, we set $-A := \{-z : z \in A\}$. In particular, this applies to the case that $A = \pi \in \Pi^{\uparrow}$. Then $\Pi^{\uparrow} \ni \pi \mapsto -\pi \in \Pi^{\downarrow}$ is a bijection from $\Pi^{\uparrow}$ to $\Pi^{\downarrow}$. Also, if $\mathcal{A} \subset \Pi^{\uparrow}$ is a sets whose elements are paths, then we set $-\mathcal{A} := \{-\pi : \pi \in \mathcal{A}\}$. Using this notation, we say that $\hat{\pi}_1, \ldots, \hat{\pi}_n$ are *downward* coalescing Brownian motions starting from space-time points $z_1, \ldots, z_n$ if $-\hat{\pi}_1, \ldots, -\hat{\pi}_n$ are (usual, forward) coalescing Brownian motions starting from space-time points $-z_1, \ldots, -z_n$. In the same way, we define countable collections of downward coalescing Brownian motions.

Let $\hat{\pi}_1, \hat{\pi}_2 \in \Pi^{\downarrow}$ be two downward paths started from space-time points $(x_i, s_i) \in \mathbb{R}^2$ $(i = 1,2)$, and let

$$\tau = \tau(\hat{\pi}_1, \hat{\pi}_2) := \sup\big\{t \leq s_1 \wedge s_2 : \hat{\pi}_1(t) = \hat{\pi}_2(t)\big\}$$

be their first meeting time (in the downward direction), which may be $-\infty$. The open set

$$W(\hat{\pi}_1, \hat{\pi}_2) := \big\{(x,t) : \tau < t < s_1 \wedge s_2 : \hat{\pi}_1(t) < x < \hat{\pi}_2(t)\big\}$$

is called the *wedge* defined by $\hat{\pi}_1, \hat{\pi}_2$. We say that a (forward) path $\pi \in \Pi^{\uparrow}$ *enters* the wedge $W(\hat{\pi}_1, \hat{\pi}_2)$ if there exist times $\sigma_\pi < s < t$ such that

$$\big(\pi(s), s\big) \notin \overline{W}(\hat{\pi}_1, \hat{\pi}_2) \quad \text{and} \quad \big(\pi(t), t\big) \in W(\hat{\pi}_1, \hat{\pi}_2),$$

where $\overline{W}(\hat{\pi}_1, \hat{\pi}_2)$ denotes the closure of $W(\hat{\pi}_1, \hat{\pi}_2)$. In a completely analogous way, we define the first meeting time of two forward paths, the wedge defined by two forward paths, and what it means for a downward path to enter such a wedge. We make the following simple observation.

Lemma 4.13 (Limits of wedges) *Let $(\hat\pi^n_i)_{n\geq 1}$ $(i = 1,2)$ be sequences of downward paths and let $(\pi^n)_{n\geq 1}$ be a sequence of forward paths. Assume that there exist $\hat\pi_i \in \Pi^\downarrow$ $(i = 1,2)$ and $\pi \in \Pi^\uparrow$ such that*

$$\hat\pi^n_i \underset{n\to\infty}{\longrightarrow} \hat\pi_i \quad (i = 1,2) \quad \textit{and} \quad \pi^n \underset{n\to\infty}{\longrightarrow} \pi$$

in the topologies on $\Pi^\downarrow$ and $\Pi^\uparrow$, and that moreover

$$\tau(\hat\pi^n_1, \hat\pi^n_2) \underset{n\to\infty}{\longrightarrow} \tau(\hat\pi_1, \hat\pi_2).$$

Assume that for each n, the path π^n does not enter the wedge $W(\pi^n_1, \pi^n_2)$. Then the path π does not enter the wedge $W(\pi_1, \pi_2)$.

Proof By definition, if π enters the wedge $W(\pi_1, \pi_2)$, then there exist times $\sigma_\pi < s < t$ such that

$$\big(\pi(s), s\big) \not\in \overline{W}(\hat\pi_1, \hat\pi_2) \quad \text{and} \quad \big(\pi(t), t\big) \in W(\hat\pi_1, \hat\pi_2).$$

But then our assumptions imply that for n sufficiently large, $\sigma_{\pi_n} < s < t$ and

$$\big(\pi^n(s), s\big) \not\in \overline{W}(\hat\pi^n_1, \hat\pi^n_2) \quad \text{and} \quad \big(\pi^n(t), t\big) \in W(\hat\pi^n_1, \hat\pi^n_2),$$

which contradicts the assumption that π^n does not enter $W(\pi^n_1, \pi^n_2)$. ∎

Proposition 4.14 (Dual coalescing Brownian motions) *Let $\mathcal{D}, \hat{\mathcal{D}}$ be countable dense subsets of $\mathbb{R}^2$. Then it is possible to construct a collection $(\pi_z)_{z\in\mathcal{D}}$ of coalescing Brownian motions together with a collection $(\hat\pi_z)_{z\in\hat{\mathcal{D}}}$ of downward coalescing Brownian motions in such a way that:*

- *For each $z \in \mathcal{D}$ and $z_1, z_2 \in \hat{\mathcal{D}}$, the path π_z does not enter the wedge $W(\hat\pi_{z_1}, \hat\pi_{z_2})$.*
- *For each $z \in \hat{\mathcal{D}}$ and $z_1, z_2 \in \mathcal{D}$, the downward path $\hat\pi_z$ does not enter the wedge $W(\pi_{z_1}, \pi_{z_2})$.*

The proof of Proposition 4.14 makes use of the following simple lemma.

Lemma 4.15 (Tightness of joint law) *Let $\mathcal{X}, \mathcal{Y}$ be Polish spaces, let $(X_n, Y_n)_{n\geq 1}$ be a sequence of random variables with values in $\mathcal{X} \times \mathcal{Y}$, and let X and Y be random variables with values in $\mathcal{X}$ and $\mathcal{Y}$, respectively. Assume that*

$$\mathbb{P}[X_n \in \cdot\,] \underset{n\to\infty}{\Longrightarrow} \mathbb{P}[X \in \cdot\,] \quad \textit{and} \quad \mathbb{P}[Y_n \in \cdot\,] \underset{n\to\infty}{\Longrightarrow} \mathbb{P}[Y \in \cdot\,]$$

Then the probability laws

$$\Big(\mathbb{P}\big[(X_n, Y_n) \in \cdot\,\big]\Big)_{n\geq 1}$$

are tight.

Proof The convergence of the marginal laws implies that the probability laws

$$\big(\mathbb{P}[X_n \in \cdot\,]\big)_{n\geq 1} \quad \text{and} \quad \big(\mathbb{P}[Y_n \in \cdot\,]\big)_{n\geq 1}$$

are tight, so for each $\varepsilon > 0$, there exist compact sets $C \subset \mathcal{X}$ and $K \subset \mathcal{Y}$ such that

$$\sup_{n\geq 1} \mathbb{P}[X_n \notin C] \leq \varepsilon \quad \text{and} \quad \sup_{n\geq 1} \mathbb{P}[Y_n \notin K] \leq \varepsilon$$

Then $C \times K$ is compact and

$$\sup_{n\geq 1} \mathbb{P}\big[(X_n, Y_n) \notin C \times K\big] \leq 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, it follows that the laws of (X_n, Y_n) are tight. ∎

Proof of Proposition 4.14 (sketch) Let $\mathcal{U}$ be the collection of paths in an arrow configuration and let $\mathcal{U}'$ be the collection of downward paths in the associated dual arrow configuration. Let ε_n be positive constants tending to zero. For each $z \in \mathcal{D}$, choose $z_n \in \mathbb{Z}^2_{\text{even}}$ such that $\theta_{\varepsilon_n}(z_n) \to z$, and for each $z \in \hat{\mathcal{D}}$, choose $z^n \in \mathbb{Z}^2_{\text{odd}}$ such that $\theta_{\varepsilon_n}(z^n) \to z$. For each $z \in \mathcal{D}$ and $n \geq 1$, let $R^n_z \in \mathcal{U}$ be the unique forward path starting at z_n, let $\hat{R}^n_z \in \mathcal{U}'$ be the unique downward path starting at z^n, and let

$$\pi^n_z := \theta_{\varepsilon_n}(R^n_z) \quad \text{and} \quad \hat{\pi}^n_z := \theta_{\varepsilon_n}(\hat{R}^n_z)$$

denote the associated diffusively rescaled paths. We claim that

$$\begin{aligned}
\mathbb{P}\big[(\pi^n_z)_{z\in\mathcal{D}} \in \cdot\,\big] &\underset{n\to\infty}{\Longrightarrow} \mathbb{P}\big[(\pi_z)_{z\in\mathcal{D}} \in \cdot\,\big],\\
\mathbb{P}\big[(\hat{\pi}^n_z)_{z\in\hat{\mathcal{D}}} \in \cdot\,\big] &\underset{n\to\infty}{\Longrightarrow} \mathbb{P}\big[(\hat{\pi}_z)_{z\in\mathcal{D}} \in \cdot\,\big],
\end{aligned}$$

where $\Rightarrow$ denotes weak convergence of probability laws on the spaces $(\Pi^{\uparrow})^{\mathcal{D}}$ and $(\Pi^{\downarrow})^{\mathcal{D}}$, respectively, which are equipped with the product topology, and $(\pi_z)_{z\in\mathcal{D}}$ is a collection of coalescing Brownian motions while $(\hat{\pi}_z)_{z\in\hat{\mathcal{D}}}$ is a collection of downward coalescing Brownian motions. Indeed, to prove this, by the definition of the product topology, it suffices to prove convergence of finite dimensional distribitions. But this has already been done in Proposition 4.4.

In fact, using Exercise 4.6, we can strengthen our previous claim in a sense that also includes convergence of meeting times. More precisely, one can show that

$$\begin{aligned}
&\mathbb{P}\big[\big((\pi^n_z)_{z\in\mathcal{D}}, (\tau(\pi^n_{z_1}, \pi^n_{z_2}))_{(z_1,z_2)\in\mathcal{D}^2}\big) \in \cdot\,\big]\\
&\quad \underset{n\to\infty}{\Longrightarrow} \mathbb{P}\big[\big((\pi_z)_{z\in\mathcal{D}}, (\tau(\pi_{z_1}, \pi_{z_2}))_{(z_1,z_2)\in\mathcal{D}^2}\big) \in \cdot\,\big],
\end{aligned}$$

and similarly for the collection of downward paths.

By Lemma 4.15, going to a subsequence if necessary, we can assume that the joint law of the random variables

$$(\pi^n_z)_{z\in\mathcal{D}}, \quad (\tau(\pi^n_{z_1},\pi^n_{z_2}))_{(z_1,z_2)\in\mathcal{D}^2}, \quad (\hat\pi^n_z)_{z\in\mathcal{D}}, \quad (\tau(\hat\pi^n_{z_1},\hat\pi^n_{z_2}))_{(z_1,z_2)\in\mathcal{D}^2}$$

converges weakly. Then we can use Skorohod's representation theorem (Theorem 1.13) to couple our random variables so that the convergence is almost sure, i.e., we can find a coupling such that

$$\pi^n_z \underset{n\to\infty}{\longrightarrow} \pi_z \text{ a.s.} \quad \text{and} \quad \tau(\pi^n_{z_1},\pi^n_{z_2}) \underset{n\to\infty}{\longrightarrow} \tau(\pi_{z_1},\pi_{z_2}) \text{ a.s.}$$

for all $z, z_1, z_2 \in \mathcal{D}$, and likewise for downward paths. Since paths of $\mathcal{U}$ do not enter wedges of $\mathcal{U}'$ and vice versa, we can use Lemma 4.13 to conclude that the same is true for the limit object. ∎

Theorem 4.16 (Wedge characterisation of the Brownian web) *Let $\mathcal{D}, \hat{\mathcal{D}}$ be countable dense subsets of $\mathbb{R}^2$, let $(\pi_z)_{z\in\mathcal{D}}$ be a collection of coalescing Brownian motions started from $\mathcal{D}$, and let $(\hat\pi_z)_{z\in\hat{\mathcal{D}}}$ be a collection of downward coalescing Brownian motions started from $\hat{\mathcal{D}}$. Assume that paths in $(\pi_z)_{z\in\mathcal{D}}$ do not enter wedges of $(\hat\pi_z)_{z\in\hat{\mathcal{D}}}$. Let*

$$\begin{aligned}\mathcal{W}_- &:= \overline{\{\pi_z : z \in \mathcal{D}\}},\\ \mathcal{W}_+ &:= \big\{\pi \in \Pi^\uparrow : \pi \textit{ does not enter wedges of } (\hat\pi_z)_{z\in\hat{\mathcal{D}}}\big\}.\end{aligned}$$

Then $\mathcal{W}_- = \mathcal{W}_+$.

Proof (sketch) To prove the inclusion $\mathcal{W}_- \subset \mathcal{W}_+$, let $\pi \in \mathcal{W}_-$. Then there exists $z_n \in \mathcal{D}$ such that $\pi_{z_n} \to \pi$ as $n \to \infty$. Let $z^1, z^2 \in \hat{\mathcal{D}}$. By assumption, π_{z_n} does not enter the wedge $W(\hat\pi_{z^1}, \hat\pi_{z^2})$ for any $n \geq 1$. By Lemma 4.13, it follows that π does not enter $W(\hat\pi_{z^1}, \hat\pi_{z^2})$. This completes the proof that $\mathcal{W}_- \subset \mathcal{W}_+$.

Before we continue, we note that our assumptions imply that the forward paths do not cross downward paths, in the sense that if $z = (x,s) \in \mathcal{D}$ and $z' = (y,u) \in \hat{\mathcal{D}}$ satisfy $s < u$, then $\pi_z(s) < \hat\pi_{z'}(s)$ implies $\pi_z(t) \leq \hat\pi_{z'}(t)$ for all $t \in [s,u]$. Indeed, we can always choose some $z'' = (y', u') \in \hat{\mathcal{D}}$ with $u \leq u'$ such that $\hat\pi_{z'}(u) < \hat\pi_{z''}(u)$ and the meeting time $\tau(\hat\pi_{z'}, \hat\pi_{z''})$ is less than s. Then $\pi_z(t) > \hat\pi_{z'}(t)$ for some $t \in (s,u]$ would imply that π_z enters the wedge $W(\hat\pi_{z'}, \hat\pi_{z''})$, contradicting our assumptions.

We now prove that $\mathcal{W}_+ \subset \mathcal{W}_-$. Let $\pi \in \mathcal{W}_+$. By Lemma 4.12 we can without loss of generality assume that $\pi(t) \in \mathbb{R}$ for all $t \in I_\pi$. Fix $\sigma_\pi < t_1 < \cdots < t_m$ and $\varepsilon > 0$. We claim that there exists a $z = (x,s) \in \mathcal{D}$

such that $\sigma_\pi < s < t_1$ and $|\pi_z(t_i) - \pi(t)| \leq \varepsilon$ for all $i = 1, \ldots, m$. To see this, for each $i = 1, \ldots, m$, we choose $z^i_\pm = (x^i_\pm, t^i_\pm) \in \hat{\mathcal{D}}$ such that $t^i_\pm > t_i$ and

$$\pi(t_i) - \varepsilon < \hat{\pi}_{z_-}(t_i) < \pi(t_i) < \hat{\pi}_{z_+}(t_i) < \pi(t_i) + \varepsilon.$$

Since π does not enter the wedge $W(\hat{\pi}_{z^i_-}, \hat{\pi}_{z^i_+})$, the meeting time of $\hat{\pi}_{z^i_-}$ and $\hat{\pi}_{z^i_+}$ must satisfy

$$\tau(\hat{\pi}_{z^i_-}, \hat{\pi}_{z^i_+}) \leq \sigma_\pi,$$

and we have $\hat{\pi}_{z^i_-}(t) \leq \pi(t) \leq \hat{\pi}_{z^i_+}(t)$ for all $t \in [\sigma_\pi, t_i]$. We can now choose $z = (x, s) \in \mathcal{D}$ such that $\sigma_\pi < s < t_1$ and

$$\sup_{1 \leq i \leq m} \hat{\pi}_{z^i_-}(t_1) < \pi_z(t_1) < \inf_{1 \leq i \leq m} \hat{\pi}_{z^i_+}(t_1).$$

Since the path π_z cannot cross any of the downward paths $\hat{\pi}_{z^i_\pm}$, we must have

$$\hat{\pi}_{z_-}(t_i) < \pi_z(t_i) < \hat{\pi}_{z_+}(t_i) \qquad (1 \leq i \leq m)$$

and hence $|\pi_z(t_i) - \pi(t)| \leq \varepsilon$ for all $i = 1, \ldots, m$, proving our claim.

Now let $\varepsilon_n > 0$ satisfy $\varepsilon_n \to 0$ and let $\sigma_\pi < t_1 < \cdots < t_m$. By what we have just proved, for each n there exists a $z_n \in \mathcal{D}$ such that $|\pi_{z_n}(t_i) - \pi(t)| \leq \varepsilon$ for all $i = 1, \ldots, m$. By Proposition 4.10, the closure of $\{\pi_z : z \in \mathcal{D}\}$ is compact, so we can find a convergent subsequence. It follows that there exists a $\pi' \in \mathcal{W}_-$ such that $\pi'(t_i) = \pi(t_i)$ for all $i = 1, \ldots, m$. Now let $\{t_i : i \in \mathbb{N}\} \subset (\sigma_\pi, \infty)$ be countable and dense. By what we have just proved, for each m, there exists a $\pi_m \in \mathcal{W}_-$ such that $\pi_m(t_i) = \pi(t_i)$ for all $i = 1, \ldots, m$. Since $\mathcal{W}_-$ is compact, we can find a convergent subsequence, the limit of which must be the path π. This proves that $\mathcal{W}_+ \subset \mathcal{W}_-$. ∎

4.7 Convergence to the Brownian web

Let $\mathcal{D}, \hat{\mathcal{D}}$ be countable dense subsets of $\mathbb{R}^2$. By Proposition 4.14, we can construct such a collection $(\pi_z)_{z \in \mathcal{D}}$ of coalescing Brownian motions starting from $\mathcal{D}$ and a collection $(\hat{\pi}_z)_{z \in \hat{\mathcal{D}}}$ of downward coalescing Brownian motions starting from $\hat{\mathcal{D}}$ such that paths in $(\pi_z)_{z \in \mathcal{D}}$ do not enter wedges of $(\hat{\pi}_z)_{z \in \hat{\mathcal{D}}}$ and vice versa. We call the pair $(\mathcal{W}, \hat{\mathcal{W}})$ defined as

$$\mathcal{W} := \overline{\{\pi_z : z \in \mathcal{D}\}} \quad \text{and} \quad \hat{\mathcal{W}} := \overline{\{\hat{\pi}_z : z \in \hat{\mathcal{D}}\}} \tag{4.11}$$

the *double Brownian web*.

Lemma 4.17 (Double Brownian web) *The law of the random variable $(\mathcal{W},\hat{\mathcal{W}})$ does not depend on the choice of the countable dense sets $\mathcal{D},\hat{\mathcal{D}}\subset\mathbb{R}^2$.*

Proof Let $\mathcal{D},\mathcal{D}',\hat{\mathcal{D}}$ be countable dense subsets of $\mathbb{R}^2$. Let $(\pi_z)_{z\in\mathcal{D}}$ be a collection of coalescing Brownian motions starting from $\mathcal{D}$, let $(\pi'_z)_{z\in\mathcal{D}'}$ be a collection of coalescing Brownian motions starting from $\mathcal{D}'$, and let $(\hat{\pi}_z)_{z\in\hat{\mathcal{D}}}$ be a collection $(\hat{\pi}_z)_{z\in\hat{\mathcal{D}}}$ of downward coalescing Brownian motions starting from $\hat{\mathcal{D}}$. By Proposition 4.14, we can couple $(\pi_z)_{z\in\mathcal{D}}$ to $(\hat{\pi}_z)_{z\in\hat{\mathcal{D}}}$ in such a way that paths in $(\pi_z)_{z\in\mathcal{D}}$ do not enter wedges of $(\hat{\pi}_z)_{z\in\hat{\mathcal{D}}}$ and vice versa. Similarly, we can couple $(\pi'_z)_{z\in\mathcal{D}'}$ to $(\hat{\pi}_z)_{z\in\hat{\mathcal{D}}}$ in such a way that paths in $(\pi'_z)_{z\in\mathcal{D}'}$ do not enter wedges of $(\hat{\pi}_z)_{z\in\hat{\mathcal{D}}}$ and vice versa. We can then couple all three collections $(\pi_z)_{z\in\mathcal{D}}$, $(\pi'_z)_{z\in\mathcal{D}'}$, and $(\hat{\pi}_z)_{z\in\hat{\mathcal{D}}}$ in such a way that the joint law of $(\pi_z)_{z\in\mathcal{D}}$ and $(\hat{\pi}_z)_{z\in\hat{\mathcal{D}}}$ is as before and the joint law of $(\pi'_z)_{z\in\mathcal{D}'}$ and $(\hat{\pi}_z)_{z\in\hat{\mathcal{D}}}$ is also as before. For example, this can be achieved by making $(\pi_z)_{z\in\mathcal{D}}$ and $(\pi'_z)_{z\in\mathcal{D}'}$ conditionally indepenent given $(\hat{\pi}_z)_{z\in\hat{\mathcal{D}}}$, and with the same conditional laws as before.

For this coupling, let $(\mathcal{W},\hat{\mathcal{W}})$ be defined using $\mathcal{D},\hat{\mathcal{D}}$ and let $(\mathcal{W}',\hat{\mathcal{W}})$ be defined using $\mathcal{D}',\hat{\mathcal{D}}$. Then Theorem 4.16 tells us that

$$\mathcal{W}=\big\{\pi\in\Pi^{\uparrow}:\pi\text{ does not enter wedges of }(\hat{\pi}_z)_{z\in\hat{\mathcal{D}}}\big\}=\mathcal{W}'\quad\text{a.s.}$$

It follows that the joint law of $(\mathcal{W},\hat{\mathcal{W}})$ is the same as the joint law of $(\mathcal{W}',\hat{\mathcal{W}})$. In the same way, we can also replace $\hat{\mathcal{D}}$ by another countable dense subset of $\mathbb{R}^2$ without changing the law of the double Brownian web. ∎

The following theorem, which is the main result of this chapter, implies in particular the convergence in (4.4).

Theorem 4.18 (Approximation of the double Brownian web) *Let $\mathcal{U}$ be the set of paths in an arrow configuration and let $\mathcal{U}'$ be the set of downward paths in the associated dual arrow configuration. Then*

$$\mathbb{P}\big[\theta_\varepsilon(\overline{\mathcal{U}},\overline{\mathcal{U}}')\in\cdot\,\big]\underset{\varepsilon\to0}{\Longrightarrow}\mathbb{P}\big[(\mathcal{W},\hat{\mathcal{W}})\in\cdot\,\big],\tag{4.12}$$

where $\Rightarrow$ denotes weak convergence of probability laws on the space $\mathcal{K}(\Pi^{\uparrow})\times\mathcal{K}(\Pi^{\downarrow})$, and $(\mathcal{W},\hat{\mathcal{W}})$ is the double Brownian web.

Proof Fix countable dense sets $\mathcal{D},\hat{\mathcal{D}}\subset\mathbb{R}^2$ and define $(\mathcal{W},\hat{\mathcal{W}})$ as in (4.11). It suffices to prove convergence along any sequence ε_n of positive constants tending to zero. It follows from Proposition 4.11 (compare Lemma 4.15) that the laws

$$\big(\mathbb{P}\big[\theta_{\varepsilon_n}(\overline{\mathcal{U}},\overline{\mathcal{U}}')\in\cdot\,\big]\big)_{n\geq1}$$

are tight, so by going to a subsequence, we may assume that they converge to some limit law $\mathbb{P}[(\mathcal{V},\hat{\mathcal{V}})\in\cdot\,]$. By Lemma 1.2, it suffices to show that each such subsequential limit is equal to $\mathbb{P}[(\mathcal{W},\hat{\mathcal{W}})\in\cdot\,]$.

As in the proof of Proposition 4.14, for each $z\in\mathcal{D}$, we choose $z_n\in\mathbb{Z}^2_{\rm even}$ such that $\theta_{\varepsilon_n}(z_n)\to z$, and for each $z\in\hat{\mathcal{D}}$, we choose $z^n\in\mathbb{Z}^2_{\rm odd}$ such that $\theta_{\varepsilon_n}(z^n)\to z$. For each $z\in\mathcal{D}$ and $n\geq 1$, we let $R^n_z\in\mathcal{U}$ be the unique forward path starting at z_n, we let $\hat{R}^n_z\in\mathcal{U}'$ be the unique downward path starting at z^n, and we let

$$\pi^n_z:=\theta_{\varepsilon_n}(R^n_z)\quad\text{and}\quad\hat{\pi}^n_z:=\theta_{\varepsilon_n}(\hat{R}^n_z)$$

denote the associated diffusively rescaled paths. In the proof of Proposition 4.14, we have shown that

$$\begin{aligned}&\mathbb{P}\big[\big((\pi^n_z)_{z\in\mathcal{D}},(\tau(\pi^n_{z_1},\pi^n_{z_2}))_{(z_1,z_2)\in\mathcal{D}^2}\big)\in\cdot\,\big]\\&\quad\underset{n\to\infty}{\Longrightarrow}\mathbb{P}\big[\big((\pi_z)_{z\in\mathcal{D}},(\tau(\pi_{z_1},\pi_{z_2}))_{(z_1,z_2)\in\mathcal{D}^2}\big)\in\cdot\,\big],\end{aligned}$$

and similarly for the collection of downward paths. We argued there that going to a subsequence if necessary and using Skorohod's representation theorem, we can couple our random variables such that

$$\pi^n_z\underset{n\to\infty}{\longrightarrow}\pi_z\text{ a.s.}\quad\text{and}\quad\tau(\pi^n_{z_1},\pi^n_{z_2})\underset{n\to\infty}{\longrightarrow}\tau(\pi_{z_1},\pi_{z_2})\text{ a.s.}$$

for all $z,z_1,z_2\in\mathcal{D}$, and likewise for downward paths. We can extend this argument to obtain that moreover

$$\theta_{\varepsilon_n}(\overline{\mathcal{U}},\overline{\mathcal{U}}')\underset{n\to\infty}{\longrightarrow}(\mathcal{V},\hat{\mathcal{V}})\quad\text{a.s.}$$

in the topology on $\mathcal{K}(\Pi^\uparrow)\times\mathcal{K}(\Pi^\downarrow)$ for some random compact sets $\mathcal{V}\subset\Pi^\uparrow$ and $\hat{\mathcal{V}}\subset\Pi^\downarrow$. We will show that for this particular coupling, $(\mathcal{V},\hat{\mathcal{V}})=(\mathcal{W},\hat{\mathcal{W}})$ a.s., where the latter is defined in terms of $(\pi_z)_{z\in\mathcal{D}}$ and $(\hat{\pi}_z)_{z\in\hat{\mathcal{D}}}$. This shows that all subsequential limit laws are the same and hence by Lemma 1.2 that the original sequence converges.

By symmetry between forward and dual webs, it suffices to prove that $\mathcal{V}=\mathcal{W}$. We will prove that $\mathcal{W}_-\subset\mathcal{V}\subset\mathcal{W}_+$, where $\mathcal{W}_-$ and $\mathcal{W}_+$ are defined as in Theorem 4.16. Since $\mathcal{W}=\mathcal{W}_-=\mathcal{W}_+$, the claim then follows.

Since $\mathcal{V}$ is closed, to prove that $\mathcal{W}_-\subset\mathcal{V}$, it suffices to prove that $\pi_z\in\mathcal{V}$ for all $z\in\mathcal{D}$. Recalling Lemma 1.16, this is obvious since $\pi^n_z\in\theta_{\varepsilon_n}(\mathcal{U})$ for all n while $\pi^n_z\to\pi_z$ a.s. and $\theta_{\varepsilon_n}(\mathcal{U})\to\mathcal{V}$ a.s.

To prove that $\mathcal{V}\subset\mathcal{W}_+$, we need to show that paths $\pi\in\mathcal{V}$ do not enter wedges of $(\hat{\pi}_z)_{z\in\hat{\mathcal{D}}}$. By Lemma 1.16, for each $\pi\in\mathcal{V}$, there exist $\pi^n\in\theta_{\varepsilon_n}(\mathcal{U})$

such that $\pi^n \to \pi$. To see that π does not enter any wedge $W(\hat\pi_{z_1}, \hat\pi_{z_2})$ of $(\hat\pi_z)_{z\in\hat{\mathcal{D}}}$, we use that for each n, the path π^n does not enter the wedge $W(\hat\pi^n_{z_1}, \hat\pi^n_{z_2})$. By our assumptions, the discrete paths $\hat\pi^n_{z_i}$ $(i = 1, 2)$ converge a.s. to $\hat\pi_{z_i}$ $(i = 1, 2)$ and moreover their meeting times converge a.s., so we can use Lemma 4.13 to conclude that π does not enter $W(\hat\pi_{z_1}, \hat\pi_{z_2})$. ∎

4.8 The coalescing point set

Let $\mathcal{W}$ be a Brownian web. For each closed set $A \subset \mathbb{R}$, we define a process $(\xi^A_t)_{t\geq 0}$ by

$$\xi^A_t := \big\{\pi(t) : \pi \in \mathcal{W},\ \sigma_\pi = 0,\ \pi(0) \in A\big\} \qquad (t \geq 0).$$

One can check that $(\xi^A_t)_{t\geq 0}$ is a Markov process taking values in the space of closed subsets of $\mathbb{R}$. We will see in a moment that in fact, for each $t > 0$, the set ξ^A_t is already locally finite. Since clearly, $A \subset B$ implies $\xi^A_t \subset \xi^B_t$, it suffices to prove the statement for $\xi^{\mathbb{R}}_t$. Roughly speaking, the following result says that if we start particles performing coalescing Brownian motions from each point on the real line, then at each positive time there are only locally finitely many particles left.

Proposition 4.19 (Density of the coalescing point set) *One has*

$$\mathbb{E}\big[\big|\xi^{\mathbb{R}}_t \cap [a, b]\big|\big] = \frac{b-a}{\sqrt{\pi t}} \qquad (a < b,\ t > 0).$$

Proof We first calculate the probability that $\xi^{\mathbb{R}}_t \cap [a, b] \neq \emptyset$. We construct $(\mathcal{W}, \hat{\mathcal{W}})$ from collections $(\pi_z)_{z\in\mathcal{D}}$ and $(\hat\pi_z)_{z\in\hat{\mathcal{D}}}$ of forward and downward coalescing Brownian motions, so that paths in $(\pi_z)_{z\in\mathcal{D}}$ do not enter wedges of $(\hat\pi_z)_{z\in\hat{\mathcal{D}}}$ and vice versa. We choose $\hat{\mathcal{D}}$ such that $(a, t), (b, t) \in \hat{\mathcal{D}}$. Let

$$\tau_{a,b} = \tau(\hat\pi_{(a,t)}, \hat\pi_{(b,t)})$$

be the first meeting time of the downward paths started at (a, t) and (b, t). We claim that

$$\xi^{\mathbb{R}}_t \cap (a, b) \neq \emptyset \quad \text{implies} \quad \tau_{a,b} \leq 0 \quad \text{implies} \quad \xi^{\mathbb{R}}_t \cap [a, b] \neq \emptyset.$$

Indeed, if $\tau_{a,b} > 0$, then the paths $\hat\pi_{(a,t)}$ and $\hat\pi_{(b,t)}$ form a wedge that prevents paths in Wi starting at time zero from passing between (a, t) and (b, t), proving the first implication. On the other hand, if $\tau_{a,b} \leq 0$, then for each time $s > 0$ we can find some x such that $\hat\pi_{(a,t)}(s) < x < \hat\pi_{(b,t)}(s)$. The web $\mathcal{W}$

must contain a path π starting at (x,s) and since such a path cannot cross the downward paths $\hat{\pi}_{(a,t)}$ and $\hat{\pi}_{(b,t)}$, it must satisfy $a \leq \pi(t) \leq b$. We can construct such a path π^s with starting time s for each $s > 0$, so using the compactness of $\mathcal{W}$, we see that $\mathcal{W}$ must also contain a path π^0 starting at time zero such that $a \leq \pi(t) \leq b$, proving the second equality.

The difference $(B_1(s) - B_2(s))_{s\geq 0}$ of two Brownian motions is equally distributed with $(\sqrt{2}B(s))_{s\geq 0}$, where $(B(s))_{t\geq 0}$ is a single Brownian motion. Therefore, using the reflection principle,

$$\begin{aligned}\mathbb{P}\big[\tau_{a,b} \leq 0\big] &= \mathbb{P}\big[\sup_{0\leq s\leq t}\big(B_2(s) - B_1(s)\big) \leq b-a\big]\\ &= \mathbb{P}\big[\sup_{0\leq s\leq t} B(s) \leq \frac{b-a}{\sqrt{2}}\big] = \frac{1}{\sqrt{2\pi t}}\int_{-\frac{b-a}{\sqrt{2}}}^{\frac{b-a}{\sqrt{2}}} e^{-x^2/2t}\,\mathrm{d}x.\end{aligned}$$

In particular, this implies that

$$\mathbb{P}[x \in \xi^{\mathbb{R}}_t] = \lim_{\varepsilon\to 0}\mathbb{P}\big[\xi^{\mathbb{R}}_t \cap (x-\varepsilon, x+\varepsilon) \neq \emptyset\big] = 0 \qquad (x \in \mathbb{R},\ t > 0),$$

and hence

$$\mathbb{P}\big[\xi^{\mathbb{R}}_t \cap (a,b) \neq \emptyset\big] = \mathbb{P}\big[\xi^{\mathbb{R}}_t \cap [a,b] \neq \emptyset\big] = \mathbb{P}[\tau_{a,b} \leq 0].$$

Now

$$\begin{aligned}\mathbb{E}\big[\big|\xi^{\mathbb{R}}_t \cap [0,1]\big|\big] &= \lim_{n\to\infty}\sum_{i=1}^{2^n}\mathbb{P}\big[\xi^{\mathbb{R}}_t \cap [(i-1)2^{-n}, i2^{-n}] \neq \emptyset\big]\\ &= \lim_{\varepsilon\to 0}\varepsilon^{-1}\frac{1}{\sqrt{2\pi t}}\int_{-\varepsilon/\sqrt{2}}^{\varepsilon/\sqrt{2}} e^{-x^2/2t}\,\mathrm{d}x = \frac{1}{\sqrt{\pi t}}.\end{aligned}$$

A similar formula holds for the expectation of $\big|\xi^{\mathbb{R}}_t \cap [0,r]\big|$ for any $r > 0$ and the general result follows by translation invariance. ∎

Proof of Proposition 4.1 Equicontinuity is obvious so $\mathcal{U}$ is precompact by Proposition 4.8. If $\pi \in \overline{\mathcal{U}}$ has a starting point in $\mathbb{R}^2$, then clearly $\pi \in \mathcal{U}$. Also, clearly, each path $\pi \in \overline{\mathcal{U}}$ has a starting time $\sigma_\pi \in \overline{\mathbb{Z}}$. Let $\tilde{\Pi}^{\uparrow}_{\mathrm{triv}} := \{\pi \in \Pi^{\uparrow}_{\mathrm{triv}} : \sigma\pi \in \overline{\mathbb{Z}}\}$ By the same arguments as those used in the proof of Lemma 4.12 we see that $\tilde{\Pi}^{\uparrow}_{\mathrm{triv}} \subset \overline{\mathcal{U}}$ and each $\pi \in \overline{\mathcal{U}}\backslash\tilde{\Pi}^{\uparrow}_{\mathrm{triv}}$ satisfies $\pi(t) \in \mathbb{R}$ for all $t \in I_\pi$. To complete the proof, we must show that $\overline{\mathcal{U}}$ does not contain paths π with $\sigma_\pi = -\infty$ and $\pi(t) \in \mathbb{R}$ for all $t \in I_\pi$.

By translation invariance, it suffices to show that $\overline{\mathcal{U}}$ does not contain paths π with $\sigma_\pi = -\infty$ and $\pi(0) = 0$. If such a path exists, then for each $s \in \mathbb{Z}$ with $s \leq 0$, the restriction of π to $[s,\infty]$ would have to be a path in $\mathcal{U}$. Thus, it suffices to show that

$$\mathbb{P}\big[\pi(0) = 0 \text{ for some } \pi \in \mathcal{U} \text{ with } \sigma_\pi = s\big]$$

tends to zero as $s \to -\infty$. This is very similar to the proof of Proposition 4.19. Letting $\hat{\pi}_{(-1,0)}$ and $\hat{\pi}_{(1,0)}$ denote the paths in the dua arrow configuration starting from $(\pm 1, 0)$, and letting

$$\tau_{-1,1} := \tau\big(\hat{\pi}_{(-1,0)}, \hat{\pi}_{(1,0)}\big)$$

denote their meeting time, we have that the probability in our previous formula is equal to $\mathbb{P}[\tau_{-1,1} < s]$, which by the recurrence of one-dimensional random walk tends to zero as $s \to -\infty$. ∎

4.9 Special points

We have defined the Brownian web $\mathcal{W}$ as the closure of $\{\pi_z : z \in \mathcal{D}\}$, where $(\pi_z)_{z\in\mathcal{D}}$ is a collection of coalescing Brownian motions started from a countable dense set $\mathcal{D} \subset \mathbb{R}^2$. Here $\{\pi_z : z \in \mathcal{D}\}$ is precompact by Proposition 4.10 and hence $\mathcal{W}$ is a compact subset of $\Pi^\uparrow$. Using compactness and the fact that $\mathcal{D}$ is dense, we see that for each $z \in \mathbb{R}^2$, there exists at least one path $\pi \in \mathcal{W}$ that starts at z. For each $z \in \mathbb{R}^2$, we let

$$m_{\rm out}(z) := \big|\{\pi \in \mathcal{W} : \pi \text{ starts at } z\}\big|$$

denote the number of paths in $\mathcal{W}$ that start at z. In this section, we will prove that almost surely, there exist points z with $m_{\rm out}(z) = 2$ and even $m_{\rm out}(z) = 3$, but a deterministic point $z \in \mathbb{R}^2$ has almost surely $m_{\rm out}(z) = 1$. The key to understanding this is (again) duality.

We say that a path $\pi \in \mathcal{W}$ *enters* a point $z = (x,t) \in \mathbb{R}^2$ if $\sigma_\pi < t$ and $\pi(t) = x$. We call two paths π, π' entering z *equivalent* if there exists a $\sigma_\pi \vee \sigma_{\pi'} \leq s < t$ such that $\pi(r) = \pi'(r)$ for all $s \leq r \leq t$. This obviously defines an equivalence relation on the set of all paths $\pi \in \mathcal{W}$ entering z. We let $m_{\rm in}(z)$ denote the number of equivalence classes of paths in $\mathcal{W}$ entering z. We call $(m_{\rm in}(z), m_{\rm out}(z))$ the *type* of a point $z \in \mathbb{R}^2$.

Theorem 4.20 (Special points of the Brownian web) *Let $\mathcal{W}$ be a Brownian web. Then almost surely, all points in $\mathbb{R}^2$ are of one of the following types:*

$$(0,1), \quad (0,2), \quad (0,3), \quad (1,1), \quad (1,2), \quad (2,1),$$

and all these types occur. For each deterministic $t \in \mathbb{R}$, almost surely, all points in $\mathbb{R} \times \{t\}$ are of one of the following types:

$$(0,1), \quad (0,2), \quad (1,1),$$

and all these types occur. A deterministic point $(x,t) \in \mathbb{R}^2$ is almost surely of type $(0,1)$.

The proof of Theorem 4.20 is based on the following lemma, which is of independent interest.

Lemma 4.21 (Types of points in dual web) *Let* $(\hat{m}_{\rm in}(z), \hat{m}_{\rm out}(z))$ *denote the type of a point* $z \in \mathbb{R}^2$ *in the dual Brownian web* $\hat{\mathcal{W}}$. *Then for each* $z \in \mathbb{R}^2$,

$$m_{\rm out}(z) = \hat{m}_{\rm in}(z) + 1 \quad \textit{and} \quad \hat{m}_{\rm out}(z) = m_{\rm in}(z) + 1.$$

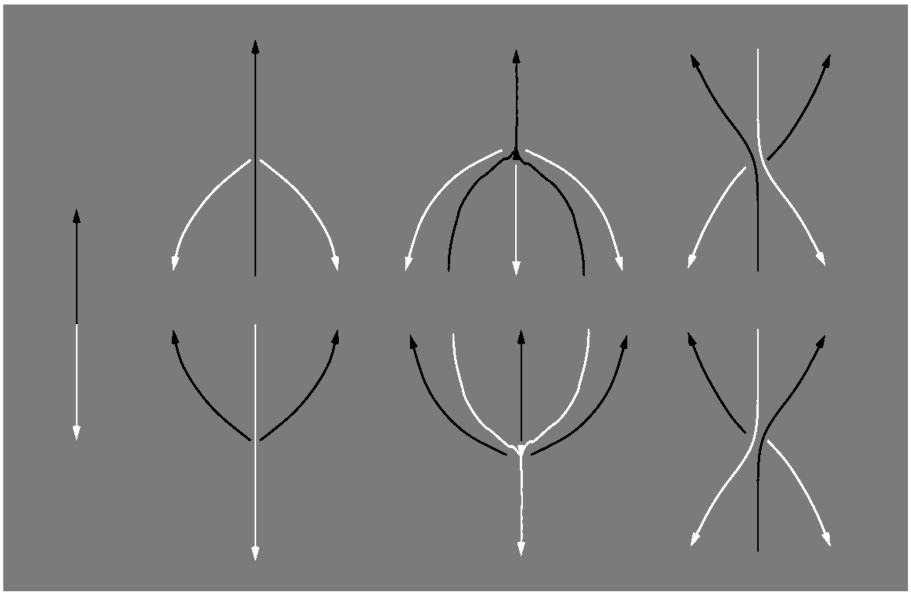

Figure 4.3: Possibe types of points in the Brownian web and its dual.

Proof (crude sketch) By symmetry, it suffices to prove that $m_{\rm out}(z) = \hat{m}_{\rm in}(z) + 1$. If there is an incoming path in $\hat{\mathcal{W}}$ at z, then forward paths started on either side of such a dual path cannot coalesce until the starting time of teh dual path, since otherwise the dual path would enter the wedge define by these forward paths. As a result, since the incoming paths divide the area just above z into $\hat{m}_{\rm in}(z) + 1$ regions, approaching the point z from different directions, using the compactness of $\mathcal{W}$, we see that there are at least $\hat{m}_{\rm in}(z) + 1$ distinct paths in $\mathcal{W}$ starting at z. On the other hand, if there are two outgoing paths in $\mathcal{W}$ at z, then any dual path that is started between these paths must stay between these forward paths and pass through z. Therefore, $\hat{m}_{\rm in} \geq m_{\rm out} - 1$. Together with our earlier claim that $m_{\rm out}(z) \geq \hat{m}_{\rm in}(z) + 1$, this proves the claim.

We have left out quite a bit of detail here. We have defined $\mathcal{W}$ and $\hat{\mathcal{W}}$ as the closures of $\{\pi_z : z \in \mathcal{D}\}$ and $\{\hat{\pi}_z : z \in \hat{\mathcal{D}}\}$, where $(\pi_z)_{z \in \mathcal{D}}$ and $(\hat{\pi}_z)_{z \in \hat{\mathcal{D}}}$ are collections of forward and dual coalescing Brownian motions, started from countable dense sets $\mathcal{D}, \hat{\mathcal{D}} \subset \mathbb{R}^2$. We have seen that paths in $\{\pi_z : z \in \mathcal{D}\}$ coalesce as soon as they meet and that paths in $\{\pi_z : z \in \mathcal{D}\}$ cannot enter wedges of $\{\hat{\pi}_z : z \in \hat{\mathcal{D}}\}$. These statements can be extended to all paths in $\mathcal{W}$ and $\hat{\mathcal{W}}$, where because of the possibility that multiple paths start at the same time, we must define the first meeting time of two paths $\pi, \pi' \in \mathcal{W}$ as

$$\tau(\pi, \pi') := \inf\big\{t > \sigma_\pi \vee \sigma_{\pi'} : \pi(t) = \pi'(t)\big\}.$$

In particular, we need $t > \sigma_\pi \vee \sigma_{\pi'}$ in this definition if we want the statement that $\pi(t) = \pi'(t)$ for all $t \geq \tau(\pi, \pi')$ to be true for all $\pi, \pi' \in \mathcal{W}$. In the proof of such statements, Proposition 4.19 plays an important role, which can be used to show that for each $\pi \in \mathcal{W}$ and $t > \sigma_\pi$, there exists a $z = (x, s) \in \mathcal{D}$ with $s < t$ such that $\pi(u) = \pi_z(u)$ for all $u \geq t$. ∎

Proof of Theorem 4.20 (crude sketch) It is clear that there exist points z with $\hat{m}_{\rm in}(z) = 1$ and $\hat{m}_{\rm in}(z) = 2$. On the other hand, using the fact mentioned at the end of the proof of Lemma 4.21 that for each $\pi \in \mathcal{W}$ and $t > \sigma_\pi$, there exists a $z = (x, s) \in \mathcal{D}$ with $s < t$ such that $\pi(u) = \pi_z(u)$ for all $u \geq t$, it is easy to see that a deterministic point z almost surely has $\hat{m}_{\rm in}(z) = 0$. Using the same fact, one moreover obtains that there are ony countably many points z with $\hat{m}_{\rm in}(z) = 2$ and it is not too hard to show that these points have $m_{\rm in}(z) = 0$.

To see that there exist points with $m_{\rm in}(z) = 1 = \hat{m}_{\rm in}(z)$, we observe that in an arrow configuration, disjoint parts of space-time are independent. This property carries over to the limit which has the consequence that dual paths do not “see” forward paths until they hit them. In fact, it is known that dual paths are reflected off forward paths by Skorohod reflection. At deterministic times, however, we do not see such points since two Brownian motions started in the forward and downward directions have zero probability to be at a deterministic time at the same position.

These arguments show that all the mentioned types of points exist, and no other types of points can exist. ∎

We could have proved the fact that deterministic points in $\mathbb{R}^2$ are a.s. of type $(0,1)$ earlier, by more elementary methods. Using this fact, one can prove the classical, “textbook” definition of the Brownian web $\mathcal{W}$, which says that $\mathcal{W}$ is a random compact subset of $\Pi^\uparrow$ that is uniquely characterised by the following properties:

(i) For each deterministic $z \in \mathbb{R}^2$, there a.s. exists a unique path $\pi_z \in \mathcal{W}$ with starting point z.

(ii) For deterministic $z_1, \ldots, z_n$, the paths $\pi_{z_1}, \ldots, \pi_{z_n}$ are distributed as coalescing Brownian motions.

(iii) For each deterministic countable dense $\mathcal{D} \subset \mathbb{R}^2$, one has almost surely $\mathcal{W} = \overline{\{\pi_z : z \in \mathcal{D}\}}$.

4.10 Some historical notes

The Brownian web originated from Arratia's PhD thesis [Arr79] and a subsequential unfinished manuscript [Arr81]. The topic remained dormant until the work of Tóth and Werner [TW98] who used the Brownian web to study a form of one-dimensional self-repellent random walk. They classified all types of special points. Together with Soucaliuc [STW00] they also proved that forward and dual paths interact through Skorohod reflection. Fontes, Isopi, Newman and Stein got interested in the Brownian web motivated by a one-dimensional model in mathematical physics [FINS01], which led Fontes, Isopi, Newman and Ravishankar [FINR04] to study this object in more detail. In particular, they were the first to give the Brownian web its name, view it as a compact set of paths, and prove convergence with respect to the Hausdorff topology. Wedges were first introduced in the concept of the Brownian net in [SS08]. A more detailed account of the history of the Brownian web can be found in [SSS16].

Chapter 5

The net

5.1 Adding branching and deaths

As in Chapter 4, we let $\mathbb{Z}^2_{\rm even}$ and $\mathbb{Z}^2_{\rm odd}$ denote the even and odd sublattices of $\mathbb{Z}^2$. Generalising the set-up of Chapter 4, let $\omega = (\omega_z)_{z\in\mathbb{Z}^2_{\rm even}}$ be an i.i.d. collection of random variables that take values in the subsets of $\{-1,+1\}$. We can use ω to define a random directed graph with vertex set $\mathbb{Z}^2_{\rm even}$ and set of oriented edges

$$\vec{E} := \big\{\big((x,t),(x+y,t+1)\big) : (x,t)\in\mathbb{Z}^2_{\rm even},\ y\in\omega_{(x,t)}\big\}.$$

We call the random directed graph $(\mathbb{Z}^2_{\rm even},\vec{E})$ an *arrow configuration*. In particular, when ω_z takes the values $\{-1\}$ and $\{+1\}$ with equal probabilities, this is an arrow configuration as defined in Section 4.1. In the present chapter, we look at sequences ω^n of arrow configurations where $\omega^n = (\omega^n_z)_{z\in\mathbb{Z}^2_{\rm even}}$, for each $n\geq 1$, is a an i.i.d. collection with common law

$$\begin{aligned}\mathbb{P}\big[\omega^n_z=\{-1\}\big]&=l_n, \quad \mathbb{P}\big[\omega^n_z=\{+1\}\big]=r_n,\\ \mathbb{P}\big[\omega^n_z=\{-1,+1\}\big]&=b_n, \quad \mathbb{P}\big[\omega^n_z=\emptyset\big]=d_n.\end{aligned} \tag{5.1}$$

Here l_n is the probability that at a given point $z\in\mathbb{Z}^2_{\rm even}$, there starts (only) an arrow to the left, r_n is the probability of an arrow to the right, b_n is the branching probability, i.e., the probability that both arrows are present, and d_n is the death probability, i.e., the probability that no arrows are present.

Recall that σ_π and τ_π denote the starting time and final time of a path $\pi\in\Pi(\overline{\mathbb{R}})$. We let

$$\Pi^{|} := \big\{\pi\in\Pi(\overline{\mathbb{R}}) : \overline{I}_\pi = [\sigma_\pi,\tau_\pi]\big\},$$

and we equip $\Pi^{|}$ with the induced topology from $\Pi(\overline{\mathbb{R}})$. By definition, a *path in the arrow configuration* ω^n, is a path $\pi\in\Pi^{|}$ with the following properties:

(i) $\sigma_\pi, \tau_\pi \in \overline{\mathbb{Z}}$,

(ii) $\pi(t+1) = \pi(t) + \omega_{(\pi(t),t)}$ $(t \in \mathbb{Z},\ t \geq \sigma_\pi)$,

(iii) $\pi(t+s) = (1-s)\pi(t) + s\pi(t+1)$ $(0 \leq s \leq 1,\ t \in \mathbb{Z},\ t \geq \sigma_\pi)$.

We let $\mathcal{V}_n$ denote the set of all paths in ω^n. Note that even in the special case when $l_n = r_n = \frac{1}{2}$ and $b_n = d_n = 0$, this is not quite the same object as the set $\mathcal{U}$ defined in Section 4.1, since we allow paths to end at some final time $\tau_\pi < \infty$. We have $\mathcal{U} = \mathcal{V}_n \cap \Pi^\uparrow$ and conversely $\mathcal{V}_n$ can be obtained from $\mathcal{U}$ by adding all shortened paths, that are cut off at an arbitrary time in $\mathbb{Z}$.

We let $\overline{\mathcal{V}}_n$ denote the closure of $\mathcal{V}_n$ in $\Pi^|$. One can again prove that $\overline{\mathcal{V}}_n$ is a compact subset of $\Pi^|$ and that $\overline{\mathcal{V}}_n \backslash \mathcal{V}_n$ only contains trivial paths, that are constantly $-\infty$ or $+\infty$. In general, when the death probability d_n is nonzero, it is not so easy to determine whether $\overline{\mathcal{V}}_n$ contains *all* trivial paths of this form, but this does not bother us.

As before, we consider a sequence ε_n of positive constants tending to zero and ask whether the diffusively rescaled set of paths converges in law, i.e., whether there exists a random compact set $\mathcal{K} \subset \Pi^|$ such that

$$\mathbb{P}\big[\theta_{\varepsilon_n}(\mathcal{V}_n) \in \cdot\,\big] \underset{n\to\infty}{\Longrightarrow} \mathbb{P}\big[\mathcal{K} \in \cdot\,\big].$$

The answer is known: it turns out the limit exists provided that

$$\varepsilon_n^{-1}(r_n - l_n) \underset{n\to\infty}{\longrightarrow} \alpha, \quad \varepsilon_n^{-1} b_n \underset{n\to\infty}{\longrightarrow} \beta, \quad \text{and} \quad \varepsilon_n^{-2} d_n \underset{n\to\infty}{\longrightarrow} \delta$$

for some constants $\alpha \in \mathbb{R}$ and $\beta, \delta \in [0,\infty)$. In the special case that $\alpha = \beta = \delta = 0$, the limit is the Brownian web, or more precisely, $\mathcal{W} := \mathcal{K} \cap \Pi^\uparrow$ is a Brownian web and $\mathcal{K}$ consists of all paths that can be obtained by cutting off paths in $\mathcal{W}$ at arbitrary times in $\overline{\mathbb{R}}$. If $\alpha \neq 0$ and $\beta = \delta = 0$, then the limit is still essentially a Brownian web, except that the coalescing Brownian motions that define it now have a drift α. If $\beta > 0$ and $\delta = 0$, then the limit is a *Brownian net* and if also $\delta > 0$, the limit object is known as a *Brownian net with killing*. Once we know how to construct a Brownian net, we will see that it is not very hard to also construct a Brownian net with killing. In view of this, in this chapter, we mostly concentrate on the case that $\alpha = 0$, $\beta = 1$, and $\delta = 0$, which corresponds to the *standard Brownian net*.

5.2 Left and right paths

We consider a sequence ω^n of arrow configurations as in the previous section with $d_n = 0$ (no deaths) and

$$\varepsilon_n^{-1}(r_n - l_n) \underset{n\to\infty}{\longrightarrow} 0 \quad \text{and} \quad \varepsilon_n^{-1} b_n \underset{n\to\infty}{\longrightarrow} 1.$$

We define $\mathcal{V}_n$ as in the previous section and set $\mathcal{U}_n := \mathcal{V}_n \cap \Pi^\uparrow$. Since the death probability is zero, $\mathcal{V}_n$ can simply be recovered from $\mathcal{U}_n$ by adding all shortened paths, that are cut off at an arbitrary time in $\mathbb{Z}$. Thus, all information is contained in the set $\mathcal{U}_n$ and we can continue to work with the space $\Pi^\uparrow$ that we are used to from the previous chapter.

By definition, a *left path* in ω^n is a path $\pi \in \mathcal{U}_n$ that satisfies

$$\pi(t+1) = \pi(t) - 1 \quad \text{if } \omega_{(\pi(t),t)} = \{-1, +1\},$$

i.e., left paths always turn left at branching points. Similarly, we define right paths as paths in $\mathcal{U}_n$ that always turn right at branching points. We let $\mathcal{U}_n^{\rm l}$ and $\mathcal{U}_n^{\rm r}$ denote the collections of all left and right paths in $\mathcal{U}_n$, respectively. We claim that

$$\begin{aligned}\mathbb{P}\big[\theta_{\varepsilon_n}(\mathcal{U}_n^{\rm l}) \in \cdot\,\big] &\underset{n\to\infty}{\Longrightarrow} \mathbb{P}\big[\mathcal{W}^{\rm l} \in \cdot\,\big],\\ \mathbb{P}\big[\theta_{\varepsilon_n}(\mathcal{U}_n^{\rm r}) \in \cdot\,\big] &\underset{n\to\infty}{\Longrightarrow} \mathbb{P}\big[\mathcal{W}^{\rm r} \in \cdot\,\big],\end{aligned}$$

where $\mathcal{W}^{\rm l}$ and $\mathcal{W}^{\rm r}$ are Brownian webs with drift -1 and $+1$, respectively, i.e., these are objects that are defined exactly in the same way as the Brownian web, except that the coalescing standard Brownian motions are replaced by coalescing Brownian motions with drift -1 and $+1$, respectively.

Indeed, letting L^n_z and R^n_z denote the unique left and right paths in $\mathcal{U}_n$ starting from a point $z \in \mathbb{Z}^2_{\rm even}$, we observe that

$$\begin{aligned}\mathbb{E}\big[L^n_z(t+1) - L^n_z(t)\big] &= r_n - l_n - b_n \sim -\varepsilon_n,\\ \mathbb{E}\big[R^n_z(t+1) - L^n_z(t)\big] &= r_n - l_n + b_n \sim +\varepsilon_n\end{aligned}$$

as $n \to \infty$, which is easily seen to imply that L^n_z and R^n_z converge after diffusive rescaling to Brownian motions with drift -1 and $+1$, respectively.

A more tricky question is how to describe the scaling limit of the joint law of L^n_z and R^n_z. In [SS08], it is shown that the interaction between left and right paths is in the limit described by the stochastic differential equation (SDE)

$$\begin{aligned}\mathrm{d}L_t &= 1_{\{L_t \neq R_t\}}\mathrm{d}B^{\rm l}_t + 1_{\{L_t = R_t\}}\mathrm{d}B^{\rm s}_t - \mathrm{d}t,\\ \mathrm{d}R_t &= 1_{\{L_t \neq R_t\}}\mathrm{d}B^{\rm r}_t + 1_{\{L_t = R_t\}}\mathrm{d}B^{\rm s}_t + \mathrm{d}t,\end{aligned} \tag{5.2}$$

where $B^{\rm l}, B^{\rm r}, B^{\rm s}$ are independent Brownian motions and solutions to (5.2) are subject to the condition that $L_t \leq R_t$ for all $t \geq \tau := \inf\{t \geq 0 : L_t = R_t\}$. Subject to this condition, it turns out that (5.2) has a weak solution for each initial state $(L_0, R_0) \in \mathbb{R}^2$, and such a solution is unique in law. In words, the SDE (5.2) says that L_t and R_t are Brownian motions with drift -1 and $+1$, that evolve independently when they are apart but are driven by the same Brownian motion when they are on the same position. It turns out that the interacting between L and R is a form of *sticky interaction*: conditional on $\tau < \infty$, the set $\{t \in \mathbb{R} : L_t = R_t\}$ has positive Lebesgue measure but is nowhere dense, i.e., each open time interval contains smaller open intervals on which $L_t < R_t$.

In [SS08], it is shown that

$$\mathbb{P}\big[\theta_{\varepsilon_n}(\mathcal{U}^{\rm l}_n, \mathcal{U}^{\rm r}_n) \in \cdot\,\big] \underset{n\to\infty}{\Longrightarrow} \mathbb{P}\big[(\mathcal{W}^{\rm l}, \mathcal{W}^{\rm r}) \in \cdot\,\big], \tag{5.3}$$

where $(\mathcal{W}^{\rm l}, \mathcal{W}^{\rm r})$ is a pair of Brownian webs with drift -1 and $+1$, respectively, whose joint law is described by the SDE (5.2). More precisely, if $\mathcal{D} \subset \mathbb{R}^2$ is a countable dense set, and for each $z \in \mathcal{D}$, we let $\pi^{\rm l}_z$ and $\pi^{\rm r}_z$ denote the almost surely unique paths in $\mathcal{W}^{\rm l}$ and $\mathcal{W}^{\rm r}$ starting from z, then $(\pi^{\rm l}_z)_{z\in\mathcal{D}}$ are coalescing Brownian motions with drift -1, $(\pi^{\rm r}_z)_{z\in\mathcal{D}}$ are coalescing Brownian motions with drift $+1$, the interacting between left and right paths is described by (5.2) and paths that are on different positions evolve independently. For brevity, we skip the details. We call $(\mathcal{W}^{\rm l}, \mathcal{W}^{\rm r})$ a *left-right Brownian web.*

5.3 The hopping and wedge constructions

Ultimately, we are not interested in left and right paths only, but in the scaling limit of the sets $\mathcal{U}_n$ of all paths in the arrow configurations ω^n. Let $\pi^{\rm l}_1, \pi^{\rm r}_2, \pi^{\rm l}_3, \ldots$ be a finite sequence of paths that are alternatively taken from $\mathcal{W}^{\rm l}$ and $\mathcal{W}^{\rm r}$, such that

$$\sigma_{\pi^{\rm l}_1} < \sigma_{\pi^{\rm r}_2} < \sigma_{\pi^{\rm l}_3} < \cdots$$

and

$$\pi^{\rm r}_2(\sigma_{\pi^{\rm r}_2}) < \pi^{\rm l}_1(\sigma_{\pi^{\rm r}_2}), \quad \pi^{\rm l}_1(\sigma_{\pi^{\rm l}_3}) < \pi^{\rm r}_2(\sigma_{\pi^{\rm l}_3}), \ldots$$

i.e., the second path, which is a right path, is started on the left of the first path, which is a left path, and then the third path, which is a left path, is started on the right of the second path and so on. Let us also assume that

$$\tau(\sigma_{\pi^{\rm l}_1}, \sigma_{\pi^{\rm r}_2}) < \sigma_{\pi^{\rm l}_3}, \quad \tau(\sigma_{\pi^{\rm r}_2}, \sigma_{\pi^{\rm l}_3}) < \sigma_{\pi^{\rm r}_4}, \ldots$$

i.e., we start the third path only after the first meeting time of the first two paths and so on. Then we can define a path π with starting time $\sigma_\pi := \sigma_{\pi^{\rm l}_1}$ by

$$\pi(t) := \begin{cases} \pi^{\rm l}_1(t) & \big(\sigma_{\pi^{\rm l}_1} \leq t \leq \tau(\sigma_{\pi^{\rm l}_1}, \sigma_{\pi^{\rm r}_2})\big), \\ \pi^{\rm l}_2(t) & \big(\tau(\sigma_{\pi^{\rm l}_1}, \sigma_{\pi^{\rm r}_2}) \leq t \leq \tau(\sigma_{\pi^{\rm r}_2}, \sigma_{\pi^{\rm l}_3})\big), \\ \pi^{\rm r}_3(t) & \big(\tau(\sigma_{\pi^{\rm r}_2}, \sigma_{\pi^{\rm l}_3}) \leq t \leq \tau(\sigma_{\pi^{\rm l}_3}, \sigma_{\pi^{\rm r}_4})\big), \end{cases}$$

and so on, i.e., we start by following the path π_1^{l}, then "hop" onto the path π_2^{r} at the first time when π_1^{l} meets π_2^{r}, and so on, until we arrive at the last path in our finite sequence, which we follow till time $+\infty$. We fix a countable dense set $\mathcal{D} \subset \mathbb{R}^2$ and let

$$\begin{aligned}\mathcal{N}_- := \text{the closure of } \big\{\pi : \pi \text{ is obtained by hopping}\\ \text{between paths in } (\pi_z^{\mathrm{l}})_{z\in\mathcal{D}} \text{ and } (\pi_z^{\mathrm{r}})_{z\in\mathcal{D}}\big\}.\end{aligned}$$

The set $\mathcal{N}_-$ will play the role of the lower bound in the proof of convergence to the Brownian net, similar to the set $\mathcal{W}_-$ in the proof of Theorem 4.18.

We also need an upper bound. This will again involve wedges and be very similar to what we did for the Brownian web. The left and right Brownian webs $\mathcal{W}^{\mathrm{l}}$ and $\mathcal{W}^{\mathrm{r}}$ each have a dual $\hat{\mathcal{W}}^{\mathrm{l}}$ and $\hat{\mathcal{W}}^{\mathrm{r}}$, where $-\hat{\mathcal{W}}^{\mathrm{l}}$ is equally distributed with $\mathcal{W}^{\mathrm{l}}$ (both are Brownian webs with drift -1) and $-\hat{\mathcal{W}}^{\mathrm{r}}$ is equally distributed with $\mathcal{W}^{\mathrm{r}}$. In fact, one can check (this is most easily seen using finite approximation) that $(-\hat{\mathcal{W}}^{\mathrm{l}}, -\hat{\mathcal{W}}^{\mathrm{r}})$ is equally distributed with $(\mathcal{W}^{\mathrm{l}}, \mathcal{W}^{\mathrm{r}})$. We now define

$$\begin{aligned}\mathcal{N}_+ := \big\{\pi \in \Pi^{\uparrow} : \pi \text{ does not enter wedges}\\ \text{of the form } W(\hat{\pi}_z^{\mathrm{r}}, \hat{\pi}_z^{\mathrm{l}}) \text{ with } z \in \mathcal{D}\big\}.\end{aligned}$$

Note that here the left boundary of the wedge is formed by a dual right path and the right boundary is a dual left path. Because of the drift, these paths may fail to meet so the wedge may be infinite in size. In particular, the fact that paths do not enter wedges of this form implies that paths in $\mathcal{N}_+$ do not cross dual left paths from right to left, or dual right paths from left to right. The following theorem is similar to Theorem 4.16 (and in fact historically predates it). We call the compact set $\mathcal{N} := \mathcal{N}_- = \mathcal{N}_+$ from the following theorem the *Brownian net.*

Theorem 5.1 (Wedge characterisation of the Brownian net) *Let $\mathcal{D}$ be a countable dense subset of $\mathbb{R}^2$ and let $\mathcal{N}_-$ and $\mathcal{N}_+$ be defined in terms of a left-right Brownian web $(\mathcal{W}^{\mathrm{l}}, \mathcal{W}^{\mathrm{r}})$ and its dual as above. Then $\mathcal{N}_- = \mathcal{N}_+$.*

Proof (sketch) The first step is to prove that paths in $\mathcal{W}^{\mathrm{l}}$ or $\mathcal{W}^{\mathrm{r}}$ cannot enter wedges of the form $W(\hat{\pi}_z^{\mathrm{r}}, \hat{\pi}_z^{\mathrm{l}})$. One way to see this is to use finite approximation and (5.3). The same is then true for paths that are constructed by hopping between left and right paths at their first meeting times, from which we conclude that $\mathcal{N}_- \subset \mathcal{N}_+$. The fact that left or right paths cannot enter wedges also implies that forward left paths cannot cross dual right paths from left to right, and forward right paths cannot cross dual left paths from right to left.

The next step is similar to the proof of Theorem 4.16. We fix $\pi \in \mathcal{N}_+$, $\sigma_\pi < t_1 < \cdots < t_m$, and $\varepsilon > 0$. We claim that we can construct a path π^{hop}

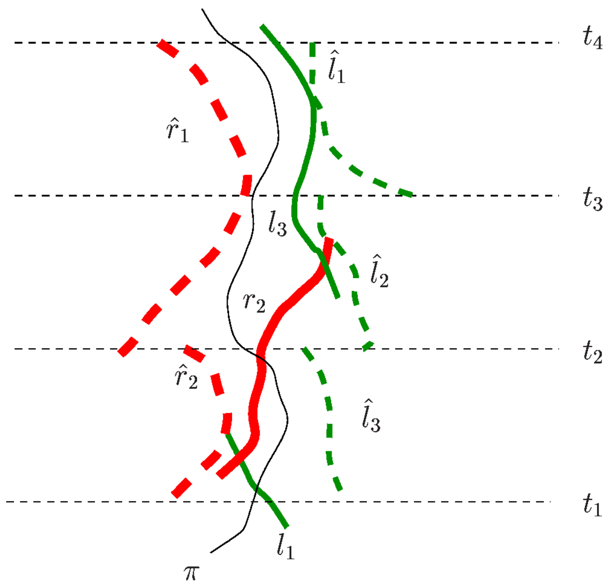

Figure 5.1: Proof of Theorem 5.1. A path π that does not enter wedges is approximated by a path that is constructed by hopping between left and right paths l_1, r_2, and l_3.

by hopping finitely often between paths in $(\pi^{\rm l}_z)_{z\in\mathcal{D}}$ and $(\pi^{\rm r}_z)_{z\in\mathcal{D}}$, such that $\sigma_\pi < \sigma_{\pi^{\rm hop}} < t_1$ and $|\pi^{\rm hop}(t_i) - \pi(t)| \leq \varepsilon$ for all $i = 1,\ldots,m$. To see this, for each $i = 1,\ldots,m$, we choose $z^i_\pm = (x^i_\pm, t^i_\pm) \in \mathcal{D}$ such that $t^i_\pm > t_i$ and

$$\pi(t_i) - \varepsilon < \hat{\pi}^{\rm r}_{z_-}(t_i) < \pi(t_i) < \hat{\pi}^{\rm l}_{z_+}(t_i) < \pi(t_i) + \varepsilon,$$

see Figure 5.1. Since π does not enter the wedge $W(\hat{\pi}^{\rm r}_{z^i_-}, \hat{\pi}^{\rm l}_{z^i_+})$, the meeting time of $\hat{\pi}^{\rm r}_{z^i_-}$ and $\hat{\pi}^{\rm l}_{z^i_+}$ must satisfy

$$\tau(\hat{\pi}^{\rm r}_{z^i_-}, \hat{\pi}^{\rm l}_{z^i_+}) \leq \sigma_\pi,$$

and we have $\hat{\pi}^{\rm r}_{z^i_-}(t) \leq \pi(t) \leq \hat{\pi}^{\rm l}_{z^i_+}(t)$ for all $t \in [\sigma_\pi, t_i]$. We can now choose $z = (x, s) \in \mathcal{D}$ such that $\sigma_\pi < s < t_1$ and

$$\sup_{1\leq i\leq m} \hat{\pi}^{\rm r}_{z^i_-}(t_1) < \pi^{\rm l}_z(t_1) < \inf_{1\leq i\leq m} \hat{\pi}^{\rm l}_{z^i_+}(t_1).$$

The forward left path $\pi^{\rm l}_z$ cannot cross any of the left downward paths $\hat{\pi}^{\rm l}_{z^i_+}$, but it can cross the right downward paths $\hat{\pi}^{\rm r}_{z^i_-}$. Just before it does so, however, we can hop onto a cleverly chosen forward right path and continue until it threatens to cross one of the left downward paths $\hat{\pi}^{\mathrm{l}}_{z^i_+}$. Just before it does, we can again hop onto a left path, and so on. Using the equicontinuity of $\mathcal{W}^{\mathrm{l}}$ and $\mathcal{W}^{\mathrm{r}}$, one can prove that with a finite number of hoppings, one can steer the hopping path so that it stays between the bounding dual right and left paths and hence satisfies $|\pi^{\mathrm{hop}}(t_i) - \pi(t)| \leq \varepsilon$ for all $i = 1, \ldots, m$.

The rest of the proof is now the same as in the proof of Theorem 4.16. ∎

Theorem 5.2 (Convergence to the Brownian net) *Let ε_n be positive constants tending to zero and let $\mathcal{U}_n$ be the set of paths in arrow configurations ω_n for which the probabilities in (5.1) satisfy $d_n = 0$, $\varepsilon_n^{-1}(r_n - l_n) \to 0$, and $\varepsilon_n^{-1} b_n \to 1$. Then*

$$\mathbb{P}\big[\theta_{\varepsilon_n}(\mathcal{U}_n) \in \cdot\,\big] \underset{n\to\infty}{\Longrightarrow} \mathbb{P}\big[\mathcal{N} \in \cdot\,\big],$$

where $\mathcal{N} := \mathcal{N}_- = \mathcal{N}_+$ is defined as in Theorem 4.16.

Proof This is very similar to the proof of Theorem 4.18. One first proves tightness, which in fact follows easily from the convergence of the collections of left and right paths and the fact that all paths starting from a point must stay between the left and right path starting from such a point. It then suffices to prove that all subsequential limit laws are equal. By going to a subsequence, we can assume that all left and right paths and dual left and right paths starting some countable dense set $\mathcal{D}$ also converge in law, and also their meeting times. We can then use Skorohods representation theorem to construct a coupling such that the convergence is almost sure. We then use the paths starting from the set $\mathcal{D}$ to construct sets $\mathcal{N}_-$ and $\mathcal{N}_+$. The proof then consists of showing that the limit $\mathcal{N}$ of the set of all paths satisfies $\mathcal{N}_- \subset \mathcal{N} \subset \mathcal{N}_+$. The lower bound follows from the fact that if discrete left and right paths converge to left and right paths that cross, then the approximating discrete paths must for n large enough also cross. The discrete path constructed by hopping between these paths is then certainly an element of $\mathcal{U}_n$, and hence $\mathcal{N}_- \subset \mathcal{N}$. The upper bound follows in the same way as in the proof of Theorem 4.18, by showing that if the approximating discrete paths do not enter wedges, then this property must be preserved in the limit. ∎

5.4 The branching-coalescing point set

Let $\mathcal{N}$ be a Brownian net. For each closed set $A \subset \mathbb{R}$, we define a process $(\xi^A_t)_{t\geq 0}$ by

$$\xi^A_t := \big\{\pi(t) : \pi \in \mathcal{N},\ \sigma_\pi = 0,\ \pi(0) \in A\big\} \qquad (t \geq 0).$$

One can check that $(\xi_t^A)_{t\geq 0}$ is a Markov process taking values in the space of closed subsets of $\mathbb{R}$. We call $(\xi_t^A)_{t\geq 0}$ the *branching-coalescing point set*. It is proved in [SS08, Prop. 1.12] that for all $a<b$,

$$\mathbb{E}\big[\big|\xi_t^{\mathbb{R}}\cap[a,b]\big|\big]=(b-a)\cdot\Big(\frac{e^{-t}}{\sqrt{\pi t}}+2\Phi(\sqrt{2t})\Big), \tag{5.4}$$

where $\Phi(x):=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-y^y/2}\mathrm{d}y$ is the distribution function of the normal distribution. The proof is similar to the proof of Proposition 4.19, except that the calculation now involves calculating the distribution of the supremum up to time t of a Brownian motion with a negative drift.

Formula (5.4) says that at deterministic times $t>0$, the random set $\xi_t^{\mathbb{R}}$ is a locally finite point set. Contrary to what we had for the coalescing point set of Section 4.8, it is however not true that $\xi_t^{\mathbb{R}}$ is a locally finite point set at all $t>0$. In fact, it is shown in [SSS09, Prop. 3.14] that there exists a dense set of times when $\xi_t^{\mathbb{R}}$ does not contain isolated points.

Intuitively, we can think about $(\xi_t^A)_{t\geq 0}$ as describing branching and coalescing Brownian motions. However, this heuristic description is a bit too simplistic, since the branching rate is, in a sense, infinite. Indeed, in Theorem 5.2 we assumed that the branching probability satisfies $\varepsilon_n^{-1}b_n\to 1$ while we rescaled time by a factor ε_n^2. This means that the expected number of branching events per unit time along a rescaled discrete path is ε_n^{-1}, which tends to infinity. Intuitively, we need infinite branching rate to counter the coalescence.

If we take the limit $t\to\infty$ in (5.4) then the right-hand side converges to $2(b-a)$. It can be shown (see [SS08, Prop. 1.15]) that the law of a Poisson point process with intensity 2 is a reversible invariant law for the Markov process $(\xi_t^A)_{t\geq 0}$, and the limit law for the process started in any nonempty initial state.

5.5 The Brownian net with killing

We have seen so far how it is possible to add branching to the Brownian web. The resuling object is called the Brownian net. As already alluded to in the introductory section of this chapter, it is possible to add deaths too. The resulting object is then called the *Brownian net with killing* and has been studied in [NRS15]. Adding deaths is somewhat easier than adding branching. We saw that to obtain a nontrivial limit, we had to assume that the branching probabilities b_n satisfy $\varepsilon_n^{-1}b_n\to 1$ while we rescale time by a factor ε_n^2, which results in a effective branching rate of ε_n^{-1}, which tends to infinity. By contrast, for the death probabilities d_n, to get a nontrivial limit we assume that $\varepsilon^{-2}d_n \to \delta$ for some $\delta > 0$, which means that the expected number of death events per unit time along a rescaled discrete path is ε_n^{-1} tends to a finite limit.

More formally, one can take the left Brownian web $\mathcal{W}^{\rm l}$ (or the right Brownian web $\mathcal{W}^{\rm r}$, which one we choose does not matter) and view it as a real-tree. One can then equip it with the length measure, which assings to each segment its length. Finally, one can construct a Poisson point set that has this length measure, multiplied with the death rate δ, as its intensity measure, and "cut" the Brownian net at the points of this Poisson point set. The result is the Brownian net with killing. For the details, we refer to [NRS15].

Bibliography

[AG15] M. Albenque and C. Goldschmidt. The Brownian continuum random tree as the unique solution to a fixed point equation. *Electron. Commun. Probab.* 20 (2015), paper no. 61.

[Ald91a] D. Aldous. The continuum random tree. I. *Ann. Probab.* 19(1) (1991), 1–28.

[Ald91b] D. Aldous. The continuum random tree. II: An overview. *Stochastic analysis, Proc. Symp., Durham/UK 1990*, Lond. Math. Soc. Lect. Note Ser. 167, 23–70, 1991.

[Ald93] D. Aldous. The continuum random tree. III. *Ann. Probab.* 21(1) (1993), 248–289.

[ALW17] S. Athreya, W. Löhr, and A. Winter. Invariance principle for variable speed random walks on trees. *Ann. Probab.* 45(2) (2017), 625–667.

[Arr79] R. Arratia. Coalescing Brownian motions on the line. Ph.D. Thesis, University of Wisconsin, Madison, 1979.

[Arr81] R. Arratia. Coalescing Brownian motions and the voter model on $\mathbb{Z}$. Unpublished partial manuscript. Available from rarratia@math.usc.edu.

[BBI01] D. Burago, Y. Burago, and S. Ivanov. *A Course in Metric Geometry.* Graduate Studies in Mathematics 33. AMS, Providence, 2001.

[Bil99] P. Billingsley. *Convergence of Probability Measures. 2nd ed.* Wiley, New York, 1999.

[Bou58] N. Bourbaki. *Éléments de Mathématique. VIII. Part. 1: Les Structures Fondamentales de l'Analyse. Livre III: Topologie Générale.*

Chap. 9: Utilisation des Nombres Réels en Topologie Générale. 2iéme éd. Actualités Scientifiques et Industrielles 1045. Hermann & Cie, Paris, 1958.

[Chi01] I. Chiswell. *Introduction to Λ-trees.* World Scientific Publishing Co. Inc., River Edge, NJ, 2001.

[Cam57] L. Le Cam. Convergence in distribution of stochastic processes. *University of California Publications in Statistics* 2 (1957), 207–236.

[Cho69] G. Choquet. *Lectures on Analysis. Volume I. Integration and Topological Vector Spaces.* Benjamin, London, 1969.

[Dud02] R.M. Dudley. *Real Analysis and Probability.* Reprint of the 1989 edition. Camebridge University Press, Camebridge, 2002.

[Edw75] D.A. Edwards. The structure of superspace. *Stud. Topol., Proc. Conf. Charlotte, N.C., 1974*, 121–133 (1975).

[EK86] S.N. Ethier and T.G. Kurtz. *Markov Processes; Characterization and Convergence.* John Wiley & Sons, New York, 1986.

[Eng89] R. Engelking. *General Topology.* Heldermann, Berlin, 1989.

[Eva08] S.N. Evans. *Probability and real trees. Ecole d'Eté de Probabilités de Saint-Flour XXXV – 2005.* Lecture given at the Saint-Flour probability summer school, July 6–23, 2005. Lecture Notes in Mathematics 1920. Springer, Berlin, 2008.

[FINS01] L.R.G. Fontes, M. Isopi, C.M. Newman, and D. Stein. Aging in 1D discrete spin models and equivalent systems. *Phys. Rev. Lett.* 87 (2001), 110201.

[FINR04] L.R.G. Fontes, M. Isopi, C.M. Newman and K. Ravishankar. The Brownian web: characterization and convergence. *Ann. Probab.* 32(4) (2004), 2857–2883.

[GPW09] A. Greven, P. Pfaffelhuber, and A. Winter. Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees). *Probab. Theory Relat. Fields* 145(1-2) (2009), 285–322.

[Gro81] M. Gromov. Groups of polynomial growth and expanding maps. Appendix by Jacques Tits. *Publ. Math., Inst. Hautes Étud. Sci.* 53 (1981), 53–78.

[Gro01] M. Gromov. *Metric structures for Riemannian and non-Riemannian spaces.* With appendices by M. Katz, P. Pansu, and S. Semmes. English translation by Sean Michael Bates. Reprint of the 2001 Edition. Birkhäuser, Boston, 2007.

[Ito71] K. Itô. Poisson point processes attached to Markov processes. *Proc. Sixth Berkeley Sympos. Math. Statist. Probab.* 3 (1971), 225–240.

[Kal97] O. Kallenberg. *Foundations of modern probability.* Springer, New York, 1997.

[KS91] I. Karatzas and S.E. Shreve. *Brownian motion and stochastic calculus. 2nd ed.* Graduate Texts in Mathematics 113. Springer, New York, 1991.

[LeG05] J.-F. Le Gall. Random trees and applications. *Probability Surveys* 2 (2005), 245–311.

[LeG10] J.-F. Le Gall. Itô's excursion theory and random trees. *Stochastic Processes Appl.* 120(5) (2010), 721–749.

[Loh13] W. Löhr. Equivalence of Gromov-Prohorov- and Gromov's $\underline{\square}_\lambda$-metric on the space of metric measure spaces. *Electron. Commun. Probab.* 18 (2013), paper no. 17.

[McK69] H.P. McKean. *Stochastic integrals.* Academic Press, New York, 1969.

[Mey76] P.A. Meyer. Un cours sur lest intégrales stochastiques. Pages 245–400 in: *Sém. de Probabilités X*, Lecture Notes in Math. 511. Springer, Berlin, 1976.

[NRS15] C.M. Newman, K. Ravishankar, and E. Schertzer. The Brownian net with killing. *Stochastic Processes Appl.* 125 (2015), 1148–1194.

[Rog89] L.C.G. Rogers. A guided tour through excursions. *Bull. Lond. Math. Soc.* 21(4) (1989), 305–341.

[RW87] L.C.G. Rogers and D. Williams. *Diffusions, Markoc processes, and martingales*, Vol. 2. Wiley, Chichester, 1987.

[SS08] R. Sun and J.M. Swart. The Brownian net. *Ann. Probab.* 36(3), 1153-1208, 2008.

[SSS09] E. Schertzer, R. Sun, and J.M. Swart. Special points of the Brownian net. *Electron. J. Probab.* 14 (2009), paper No. 30, 805–864.

[SSS14] E. Schertzer, R. Sun, and J.M. Swart. Stochastic flows in the Brownian web and net. *Mem. Am. Math. Soc.* Vol. 227 (2014), Nr. 1065.

[SSS16] E. Schertzer, R. Sun, and J.M. Swart. The Brownian Web, the Brownian Net, and their Universality. Pages 270-368 in: P. Contucci and C. Giardinà (Eds.) *Advances in Disordered Systems, Random Processes and Some Applications*, Cambridge University Press, 2016.

[STW00] F. Soucaliuc, B. Tóth, W. Werner. Reflection and coalescence between one-dimensional Brownian paths. *Ann. Inst. Henri Poincaré Probab. Statist.* 36 (2000), 509–536.

[TW98] B. Tóth and W. Werner. The true self-repelling motion. *Probab. Theory Related Fields* 111 (1998), 375–452.

[Oxt80] J.C. Oxtoby. *Measure and Category. Second Edition.* Springer, New York, 1980.

Index