Selection in a spatially structured population

Daniel Straulino

Department of Statistics

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2014

A mis padres.

Cada día que pasa solo agranda el cariño y la admiración que les tengo.

Acknowledgements

First of all, I would like to Professor Alison Etheridge. Without her guidance and encouragement I would have never made it this far. It was a pleasure to learn and work with her and I consider myself very lucky to have had her as my supervisor. I would also like to thank my examiners, Professors Anton Wakolbinger and Christina Goldschmidt, I appreciate the time and dedication they gave to revising my work, and I am grateful for their observations and suggestions. There have been many teachers that have guided and encouraged me to follow my passion for mathematics, and I would like to thank them all, from those in high school and the Mathematics Olympiads, to my Lecturers at ITAM. I would like to specially mention Juan Manuel Lozano, who supported my interest in the Mathematics Olympiads, and Manuel Mendoza, who directed my undergraduate dissertation.

My voyage through Graduate School wouldn't have been the same without the great officemates I had during my time in Oxford. Anna, Luke, Dominic and Sarah, I was very fortunate to have such good company, and I will miss our office very much. Whether it was to discuss mathematics or take way too many tea breaks, I could count on you. I also want to thank my College friends, too many to name, and specially the lunch gang. So many days it felt like lunch time was the only thing to look forward, and it never failed to deliver good laughs and surprising bits of information. I also want to thank someone whose presence I felt very close through these four years, Alex, the most special person I met at Oxford...

Last but not least, I want to thank my family, my mum, my dad and my brothers,

who have always supported me, and whom I feel closer than ever even though we have been living in different continents.

Selection in a spatially structured population

Abstract

Daniel Straulino Torre — Doctor of Philosophy

St Cross College — Trinity 2014

This thesis will focus on the effect that selection has on the ancestry of a spatially structured population. In the absence of selection, the ancestry of a sample from the population behaves as a system of random walks that coalesce upon meeting. Backwards in time, each ancestral lineage jumps, at the time of its birth, to the location of its parent, and whenever two ancestral lineages have the same parent they jump to the same location and coalesce. Introducing selective forces to the evolution of a population translates into branching when we follow ancestral lineages, a by-product of biased sampling forwards in time.

We study populations that evolve according to the Spatial Λ-Fleming-Viot process with selection. In order to assess whether the picture under selection differs from the neutral case we must consider the timescale dictated by the neutral mutation rate θ. Thus we look at the rescaled dual process with $n = 1/\theta$.

Our goal is to find a non-trivial rescaling limit for the system of branching and coalescing random walks that describe the ancestral process of a population. We show that the strength of selection (relative to the mutation rate) required to do so depends on the dimension; in one and two dimensions selection needs to be stronger in order to leave a detectable trace in the population.

The main results in this thesis can be summarised as follows. We define the selection coefficient to be

$$\mathbf{s_n} = \begin{cases} \frac{s}{\sqrt{n}} & if \quad d = 1 \\ \frac{s \log(n)}{n} & if \quad d = 2 \\ \frac{s}{n} & if \quad d \geqslant 3 \end{cases}$$

In two and higher dimensions the ancestral process of a sample of the population converges to branching Brownian motion. In one dimension, provided we do not allow ancestral lineages to jump over each other, the ancestral process converges to a subset of the Brownian net.

We also provide numerical results that show that the non-crossing restriction in one dimension cannot be lifted without a qualitative change in the behaviour of the process. Finally, through simulations, we study the rate of convergence in the two-dimensional case.

Statement of authorship

This dissertation is submitted to the University of Oxford for the degree of Doctor in Philosophy. I hereby detail which parts of my thesis are my own and which ones are the result of a collaboration effort.

1. Chapters 0 through 4 as well as chapter 6 and 7 are solely my work.

2. Chapter 5 is the result of a joint effort with my supervisor Professor Alison Etheridge, and Dr Nic Freeman (University of Bristol).

Date: 1st October 2014

Daniel Straulino

Contents

CHAPTER 0

Outline

The last few decades have seen substantial progress in the field of DNA sequencing, with more efficient techniques allowing scientists to collect vast amounts of genetic data. In parallel, the mathematical models of coalescence have been developed in order to provide useful tools for interpretation and inference on this data. In addition to this practical purpose, the coalescent models have become a topic of great interest in their own right.

For both mathematicians and biologist alike, one of the more interesting questions has been how to incorporate spatial information into the models in a way that both provides a tractable mathematical form, and is as faithful as possible to the biological principles that govern the dynamics of populations. In this spirit Etheridge [2008] proposed a model that is know as the

spatial Λ Fleming-Viot process. This model provides a very flexible framework in which to understand the evolution of a population that lives in a spatial continuum.

In this thesis we extend the existing results for the spatial Λ Fleming-Viot process by considering the case of a population with small neighbourhood size under the effects of selective pressure. Our aims are twofold. On the one hand, we address a question of interest in population genetics: when will the action of natural selection on a gene in a spatially structured population cause a detectable trace in the patterns of genetic variation observed in the contemporary population? On the other, we investigate some of the rich mathematical structure underlying mathematical models for spatially evolving populations and, in particular, the systems of interacting random walks that, as dual processes (corresponding to ancestral lineages of the model), describe the genetic relationships, across both time and space, between individuals sampled from those populations.

Here we provide an outline of the thesis.

Chapter 1. Population Genetics. The first chapter aims to provide a concise introduction to the classical theory of population genetics. We begin by presenting the Wright-Fisher model, one of the earliest attempts to provide a mathematical description of the evolution of a population. It provides a good starting point since it allows us to introduce terms common in the population genetics literature, such as fixation probability and effective population size. If instead of going forwards in time we consider what happens when looking back at the ancestors of individuals sampled from the population, we arrive at the ancestral process of the population, which takes us to the Kingman coalescent. We then introduce the Moran model, which allows us to consider the (moment) duality between the forwards and backwards in time models for a population. Finally, we consider some generalisations of the previous models that give rise to the Λ-coalescent and the generalised Fleming-Viot process. The aim of the chapter is to familiarise the reader with key ideas that will appear in our later work, and to serve as a brief guide for mathematicians that are interested in the field of population genetics but lack the biological background.

Chapter 2. The Spatial Λ Fleming-Viot process. Here we give an overview of the different ways in which mathematicians and biologists have attempted to introduce spatial structure into population models. The difficulties of considering a population evolving in a continuum motivates the introduction of the SΛFV model, which will be the starting point of our investigations. This model was introduced relatively recently by Etheridge [2008]. We proceed to rigorously define the model, and we finish the chapter with a short survey of the literature surrounding the SΛFV model. The first two chapters contain no original results, but we hope they provide background and motivation for the work to follow.

Chapter 3. Natural Selection This chapter begins by discussing selection, as well as some of the ways in which we might introduce selective forces to our population models. After visiting some classical constructions, we present the the SΛFV model with selection. We proceed to consider its dual (that is, the process that follows ancestral lineages backwards in time). The rest of the thesis will be devoted to the analysis of this dual process and its rescaling limits. In particular, we will consider a diffusive rescaling limit. The reason behind this choice is that, in the absence of selection, an ancestral lineage in isolation follows a symmetric random walk, and thus will converge to a Brownian motion under our rescaling.

The last two sections of this chapter have our first results concerning these limits, and will serve as a warm up for the two main results. Here we rescale the selective coefficient by n, that is, $s_n = s/n$. We then show that limiting object depends strongly on the dimension. While in dimensions three and higher we obtain a branching Brownian motion in the limit, in the two dimensional case we obtain a system of independent Brownian motions, and finally, in one dimension we obtain a system of independent Brownian motions that coalesce upon meeting. Since going backwards in time selection manifests itself in the branching of the ancestral lineages, we deduce that for lower dimensions selection is too weak to be detected and we just recover the same limit we would do in the neutral case.

Chapter 4. Dimension 2. This chapter shows that by taking $s_n = s\log(n)/n$, we can ensure that selection is still felt by the genealogical process in the limit. In fact, with this selection

coefficient we also obtain branching Brownian motion as the limit, just as we did for the high dimensional case. The crucial observation is that even though the branching rate explodes when $n \to \infty$, all but finitely many of the branches are annulled through coalescing. There is a considerable amount of work involved in showing that this picture produces a non degenerate limit. We begin by considering the interactions between the two lineages that arise from a selective event (a branch) and obtain some useful estimates on the time and the probability of coalescing. Then, we provide a construction of what we have termed the 'skeleton', which can be thought of as the collection of lineages that manage not to coalesce with their parents. Finally, we show that the skeleton provides a good approximation of the full process, and furthermore, that it converges to a branching Brownian motion.

Chapter 5. Dimension 1. In this chapter we prove that the (backwards in time) genealogical process converges to a subset of the Brownian net. The same reasoning that made us take $s_n = s\log(n)/n$ in two dimensions suggests we take $s_n = s/\sqrt{n}$ in the one dimensional case. Sun and Swart [2008] consider a discrete analogue of our dual process, with the further restriction that paths cannot cross each other, and prove that it converges to a limiting object they christened the 'Brownian net'. By considering our model with this extra restriction, we show that in our continuous setting the Brownian net still arises in the limit. Our proof owes a lot to the existing literature, but it is nevertheless not trivial to extend the previous results to our setting. Some of the more technical results that allow us to prove the convergence of the genealogical process to the Brownian net appear at the end in the form of an appendix.

Chapter 6. Simulations. In this chapter we summarise the results we obtained by running simulations of the process for both one and two dimensions. The motivation for these simulations is different in each case. As we mentioned above, in the one dimensional case we had to add a non-crossing condition in order to use the framework of the Brownian net. Here we lift that restriction and show that, in contrast to the neutral setting, it seems to have a qualitative effect on the behaviour of the genealogical process. Meanwhile, since our result in two dimensions only gives a logarithmic bound on the rate of convergence, we are interested in whether the process

might actually converge faster. We consider two summary statistics of our process, and show that simulations suggest that both of their empirical distributions seem unaffected by further increasing the parameter n once n is sufficiently large, but still relatively small.

Chapter 7. Conclusions. We conclude our thesis by providing a review of the results obtained, as well as some comments on their biological interpretation. Finally we explore potential directions for future research that arise naturally from our findings.

CHAPTER 1

Population Genetics

At the intersection between Mathematics and Biology lies the field of Population Genetics. This field aims to model the dynamics of the genetic composition of a population in order to understand its evolution. In this first chapter we aim to give a quick overview of the models and ideas that have shaped the field, and that motivate our research.

1.1 The Wright-Fisher Model

The Wright-Fisher model, developed independently by Sewall Wright and R.A. Fisher, is the simplest random model of allele frequencies in a population. It considers an idealized finite

population and the idea behind it is simple. Take a population with a fixed number of individuals, say N, each of which has a type k chosen from some type space K. Given a generation, construct the next one by letting each one of the new individuals choose a parent uniformly from among the N individuals of the previous generation. Each new individual inherits the type of its parent.

Of course this model makes a few assumptions apart from a fixed size population: the chances of any individual's reproductive success are independent of its type (*neutrality*) and every individual is equally likely to reproduce (*panmictic*). Nevertheless, it is a very useful model, which in spite of its simplicity captures important features of many populations. And it is also possible to introduce other forces to the model, such as mutation, selection, recombination and varying population size. For the moment though we will restrict our study to the elementary version we previously described.

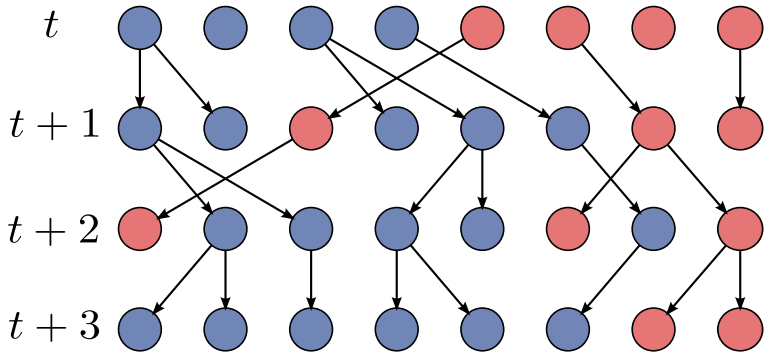

Figure 1.1: A graphical representation of the Wright-Fisher model. Here we see a how the Wright-Fisher model where we took $N = 8$ and 4 generations. The two colours represent two different types.

We notice that each individual has only one parent, so it does not look like a good model for species that reproduce sexually, but thinking about it differently allows us to bridge this gap. Instead of thinking about distinct individuals we can imagine ourselves following N genes. In the case of an *haploid* population, where each individual carries only one copy of each gene we can perfectly identify gene with individual and this distinction makes no difference. For a diploid population, even though each individual has a pair of each gene, we know that each gene was inherited from one of the parents in the previous generation, so the gene pool can still

be modelled in the same way, by sampling from the previous generation. Indeed it is common practice to use this model for diploid populations as well, by taking the population size to be $2N$ instead. One has to be careful when considering a sexed population, but we will address this when we speak about effective population size. Unless otherwise stated, we will use the terms gene and individual interchangeably.

For the moment we will stick to the haploid case. If we think that the gene we are looking at in the population has two alleles, say A and a, then the model already allows us to calculate some quantities of interest. Let us look at the process $X(t)$ that records the number of type A individuals at time t.

From the description of the Wright Fisher model, it follows that $X(t)$ is a Markov Chain on the space $\{0, 1, ..., N\}$. Being a Markov process means that given the present state of the process, the future states are independent of the past ones (in other words, all the information we have to predict the future is contained in the present state). Since each new gene chooses a parent uniformly from the previous generation, we have that the number of type A individuals in generation $t+1$ is a binomial $(N, \frac{X(t)}{N})$ and it is easy to see that the transition probabilities of the chain are given by

$$p_{i,j}(t) = \binom{N}{j} \left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j}.$$

Notice that we have two absorbing states for $X(t)$, namely 0 and N. We call these states *fixed states* since the population will be fixed with only the a or A allele. We can now calculate the *fixation probability* of the A allele. Let us call our fixation time τ;

$$\tau = \min\{t : X(t) = 0 \ \texttt{or} \ X(t) = N\}.$$

It is not hard to convince oneself that $P(\tau < \infty) = 1$. Indeed, no matter in what state we currently are, we always have a positive probability of fixation in the next generation, and this probability is bounded from below by $2\left(\frac{1}{N}\right)^N$. Then, we have the next intuitive result:

Theorem 1.1 *In the Wright Fisher model, the probability of fixation of the A allele is*

$$P(X(\tau) = N) = \frac{X(0)}{N}.$$

PROOF: It is immediate to check that the process $X(t)$ has the martingale property, using the fact that the number of A individuals in the generation is just a binomial random variable with parameters $X(t)/N$ and N. Since it is also bounded and we know that the time to fixation is almost surely finite, we can use the optional stopping theorem to deduce

$$E(X(\tau)) = (0)P(X(\tau) = 0) + (N)P(X(\tau) = N) = X(0).$$

$$\Rightarrow P(X(\tau) = N) = \frac{X(0)}{N}$$

■

We would also like to have a better idea of how long it takes for the population to fix, but arriving at an expression for $E(\tau)$ is not easy in our Markov Chain. Nevertheless, we can find a related quantity that gives us an idea of the time to fixation.

The *heterozygosity* of a population is the probability that if we take two genes without replacement, they are of different type

$$H(t) = \frac{2X(t)(N - X(t))}{N(N-1)}.$$

Of course, if we have achieved fixation, the heterozygosity will be 0.

Theorem 1.2 *Let h(t)=E(H(t)) be the expected value of the heterozygosity at time t. For the Wright-Fisher model we have that*

$$h(t) = \left(1 - \frac{1}{N}\right)^t H(0).$$

PROOF: If we sample two genes at time t, for them to be of different types they must descend from different genes in the original generation (otherwise they would share the type of their common ancestors). This gives us the first factor of the right hand side, since every generation

we trace backwards, we have a probability $1/N$ of sharing the parent. And if we avoided sharing a parent all the way to generation 0, the the probability that two individuals chosen at random from this generation have different types is just $H(0)$. ■

1.1.1 Effective population size

So far our model has not included selection, recombination, or any of the forces we can imagine shape the evolution of a population, but nevertheless we have seen that the allele frequency will vary and that we should expect one type to fix and the other to disappear. The variation in allele frequency that results from random reproduction (mating) in a finite population is called *genetic drift* and the Wright-Fisher model is the standard way of accounting for it.

Of course, there are many other models of genetic drift, and one way to relate all of them is the *effective population size*, which intuitively is the size of a Wright-Fisher population that would give the same rate of genetic drift as the model in question. This will also allow us to see whether certain added complexities and structure increase the rate of genetic drift compared to the one it has in the standard Wright-Fisher model.

There is no clear cut way to measure the rate of genetic drift, and thus the way we decide to account for it will be reflected in the resulting effective population size. The Wright-Fisher model has three main properties related to the genetic drift, namely:

1. The probability that two randomly chosen genes descend from the same parent in the previous generation is $1/N$.

2. The maximum non unit eigenvalue of the transition matrix is $1 - 1/N$.

3. If we let $p(t) = X(t)/N$, then $var(p(t+1)|p(t)) = p(t)(1-p(t))/N$.

So, if we have a model where the probability that two genes descend from the same parent is π_2, then we have that the *inbreeding* effective population size is simply $N_e^i = (\pi_2)^{-1}$. In a similar way, the *eigenvalue* effective population size can be calculated as $N_e^e = (1 - \lambda_{max})^{-1}$

where λ_{max} is the maximum non unit eigenvalue of the transition matrix. Finally, the *variance* effective population size is given by $N_e^v = p(t)(1-p(t))/var(p(t+1)|p(t))$. For the special case of exchangeable models with discrete generations, Ewens [2004] shows that all three definitions are equivalent, and he further considers other problems, like changing population size, larger family sizes and populations with two sexes, where the equivalence no longer holds.

To both illustrate effective population size, and to give a better picture of how to model a two-sex population, we will consider such a population. Suppose that in any generation there are N_m diploid males and N_f diploid females, with $N = N_f + N_m$ and that each diploid offspring inherits one gene chosen at random from the male population and a second one chosen at random from the female population. Since we have diploid individuals, our gene pool will have size $2N$.

Theorem 1.3 *For the two sex model, the inbreeding effective population size is*

$$N_e^i = \frac{4N_f N_m}{N}$$

PROOF: If we sample two genes from the population, for them to share a parent there are two possibilities: they both descend from a male or they both descend from a female (otherwise they would clearly descend from different individuals). The probability they both descend from a male is $1/4$, and given that, the probability they have the same parent is $1/2N_m$. In a similar way, the probability they both have the same mother is $1/8N_f$. Thus, the probability that both genes share a parent in the previous generation is

$$\pi_2 = \frac{1}{4}\frac{1}{2N_f} + \frac{1}{4}\frac{1}{2N_m} = \frac{1}{4}\left(\frac{1}{2N_f} + \frac{1}{2N_m}\right)$$
$$= \frac{1}{8}\left(\frac{N_m + N_f}{N_m N_f}\right) = \frac{1}{8}\left(\frac{N}{N_m N_f}\right),$$

which implies that the inbreeding effective population size is

$$N_e^i = (2\pi_2)^{-1} = \frac{4N_f N_m}{N}.$$

■

Nordborg and Krone [2002] propose a different effective population size. They define the *coalescent* effective population size to be the quantity by which one must rescale time in order to recover the Kingman coalescent as the genealogy of a sample from the population when we let the population size tend to infinity. Durrett [2008] and Sjödin *et al.* [2005] consider it the most satisfactory definition. What is meant by it will become clear after we introduce the coalescent.

1.2 The Coalescent

So far we have considered a population evolving forwards in time, but equally interesting is to consider what happens if we go backwards in time, that is, if we look at the genealogies of the individuals in our population. The way these genealogies behave will be dictated by the model we use for our population, so let us consider what happens under the haploid Wright-Fisher model.

Suppose that we pick two genes from our population. Then the probability that they share a parent in the previous generation is $1/N$. If they do not, then their parents have again a $1/N$ chance of having the same parent in the previous generation, and so on. Thus, we see that the number of generations we must go backwards before two random genes share an ancestor is geometric with parameter $1/N$.

The first individual they both encounter while tracing back their ancestral lineages is called the *most recent common ancestor* or MRCA, and will be an important concept hereafter. When we encounter the MRCA we say that the two lineages have coalesced, since now both genes will share all the same ancestors and instead of two lineages we find we need to follow only one.

In general, we will be concerned with large populations, where we hope our models can capture part of the big picture, and in that spirit we will consider what happens when we rescale time by a large factor. Consider the number of generations we must look back to find the MRCA of a sample of size 2. We know the expected number of generations is N, and this gives

us an indication of a sensible timescale to look at. Thus we will measure time in units of N generations.

Now consider a larger sample, of say size k, and follow the genealogies of the k genes backwards in time. The probability that three or more genes share a parent in the previous generation is of order $\mathcal{O}(N^{-2})$, and similarly, the probability that two different pairs of genes coalesce in the previous generation is of order $\mathcal{O}(N^{-2})$. So whereas the expected time for two lineages to coalesce is of order $\mathcal{O}(N)$, the expected time we need to wait before seeing one of these multi-lineage events is of order $\mathcal{O}(N^2)$. Thus for large N we expect that before we can see one of these events, all k lineages will have coalesced via pairwise events. In fact the probability that this is the case tends to 1 as $N \to \infty$.

For large N we thus expect that the genealogical process will be driven almost purely by pairwise mergers. If we are following k lineages then the probability that two of them coalesce in the previous generation is $\binom{k}{2}N^{-1} + \mathcal{O}(N^{-2})$, which means that for large N, in our rescaled model the waiting time to see such an event is approximated by an exponential random variable with mean $\binom{k}{2}$, at which time a random pair of genes coalesce while the rest remain unaffected. After that, we now follow in a similar fashion the $k-1$ genes till the next merger, till we end up with a single gene.

The process we have informally described above is called the k-coalescent, and we will now proceed to define it rigorously, which will allow us to show that indeed the genealogy of a sample of size k in the Wright-Fisher model converges to the k-coalescent. But first we need some notation. We will call the set of the first k integers $[k] := \{1, 2, ..., k\}$, and we denote the space of partitions of $[k]$ by $\mathcal{P}^k$.

Definition 1.4 *The k-coalescent is a continuous time strong Markov process $(\pi^k(t))_{t\geqslant 0}$ that takes values in the space $\mathcal{P}^k$, with initial condition $\pi^k(0) = \{\{1\}, \{2\}, ...\{k\}\}$, the trivial partition*

made only from singletons, and with transition rates given by:

$$q(\pi, \pi') = \begin{cases} 1 & \text{if } \pi' \text{ is obtained by merging two blocks from } \pi, \\ 0 & \text{otherwise.} \end{cases}$$

One of the nicest properties of the k-coalescent is consistency. By this we mean that if we take an l-coalescent with $l > k$ and then look at its restriction to $[k]$ we obtain a k-coalescent. In other words, if we call the restricted process $\pi^{l,k}$, we have that:

$$\pi^{l,k} \overset{(d)}{=} \pi^k.$$

The consistency holds because the size of the blocks do not affect the rate at which they merge, and so ignoring the elements larger than k does not alter the behaviour.

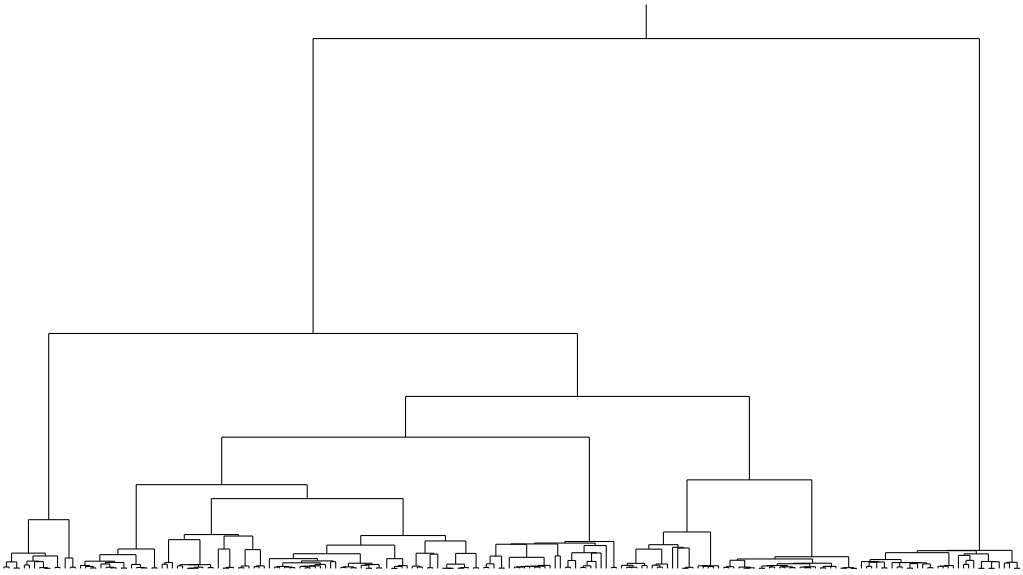

Figure 1.2: The Kingman Coalescent. This is a single realisation of the coalescent for n=200. Courtesy of Luke Miller.

The most important consequence of this property is that it allowed Kingman (1982) to use Kolmogorov's extension theorem to show that there exists a unique in law process $(\pi(t))_{t\geqslant 0}$ with values in $\mathcal{P}$ (the space of partitions of $\mathbb{N}$) such that its restriction to $[k]$ is a k-coalescent. The process π is called Kingman's coalescent. Although we won't concern ourselves at the moment

with proving it, it is worth mentioning one of the most beautiful properties that Kingman's coalescent has, namely, *coming down from infinity*. This property tells us that even though we start the coalescent with a countably infinite number of singletons at time 0, for any time $t \geqslant 0$, $\pi(t)$ consists of only finitely many blocks almost surely.

We are now ready to show that Kingman's coalescent is indeed the limit of the genealogical process of a population evolving according to the Wright-Fisher model. The proof is taken from Berestycki [2009]. Here we are considering that time goes backwards, and thus when we say time t we mean time t before the present.

Theorem 1.5 *Take a population of size N evolving (forwards in time) according to the Wright-Fisher model. Fix $k \geqslant 1$, and let $\pi_k^N(t)$ denote the ancestral partition at time t of k randomly chosen individuals from the population at time $t = 0$, where two individuals, i and j, belong to the same block of $\pi_k^N(t)$ if and only if they share the same ancestor at time t. Then, $\forall t \geqslant 0$,*

$$\pi_k^N(Nt) \xrightarrow{d} \pi_k(t).$$

PROOF: Start by considering two distinct randomly sampled genes x and y. We previously showed that the number of generations it takes them to coalesce is a geometric random variable with success probability $1/N$. Thus, when we rescale time by N and let this parameter go to infinity, the coalescing time converges to an exponential random variable with expectation 1.

Take now an arbitrary k, then for every pair the above still holds. We want to show that the rates $q_{WF}(\pi, \pi')$ of the genealogical process indeed converge to the ones of the k-coalescent. We have already shown that the probability of seeing multiple mergers vanishes as $N \to \infty$, thus, after rescaling we have:

$$q_{WF}(\pi, \pi') = \begin{cases} 1 + \mathcal{O}(N^{-1}) & \text{if } \pi' \text{ is obtained by merging two blocks from } \pi, \\ \mathcal{O}(N^{-1}) & \text{otherwise.} \end{cases}$$

This shows that, in the limit, the rates are precisely those of the k-coalescent. Karr [1975] showed that weak convergence of Markov Chains follows from the convergence of the transition

kernels, which concludes the proof. ■

This result has a certain *universality* to it, by which we mean that the result holds for populations that evolve according to different dynamics than the ones prescribed by the Wright-Fisher model. One can have overlapping generations, or different reproduction mechanisms, and still obtain the Kingman coalescent as the limiting object for the genealogies. What is required for this universality to hold is, roughly speaking, that the population is panmictic and that its size remains stable, that each individual has few offspring, and that there is no selection (see Berestycki [2009]). It is of course important to remember that we have rescaled time in order to obtain this limit, and different models require a different rescaling to arrive at Kingman's coalescent, a fact that we should bear in mind whenever we compare models or decide to do inference about the population.

When we couple the universality with the simplicity of the model, and the fact that it is rather straightforward to simulate, it should be no surprise that the coalescent has become a very popular model for genealogies, and been studied in depth, both by mathematicians and population geneticists. We conclude this first exploration of the coalescent process by calculating the time that it takes a sample of k genes to coalesce if they follow the dynamics of the Kingman coalescent.

Theorem 1.6 *Let W_k be the time it takes for a k-coalescent to merge into a single block. Then*

$$\mathbb{E}[W_k] = 2\left(1 - \frac{1}{k}\right).$$

Proof:

Clearly, the time to merge into a single block is the sum of the times that it takes to go from k to $k-1$ blocks plus the time it takes to go from $k-1$ to $k-2$ and so on. Let these times be T_i for $i = k, k-1, ..2$, since we know that T_i is distributed as an exponential random variable

with parameter $\binom{i}{2}$,

$$\mathbb{E}[W_k] = \mathbb{E}\left[\sum_{i=2}^{k} T_i\right] = \sum_{i=2}^{k} \mathbb{E}[T_i]$$
$$= \sum_{i=2}^{k} \left[\frac{2}{(i)(i-1)}\right] = 2\sum_{i=2}^{k}\left[\frac{1}{(i-1)} - \frac{1}{i}\right]$$
$$= 2\left(1 - \frac{1}{k}\right)$$

■

It is interesting to notice that the expected time it takes the last two blocks to coalesce is longer than the expected time it takes to go down to these final two blocks from any number of initial blocks. In general, we will spend most of the time following a small number of blocks, and thus should expect that adding more genes to our sample quickly stops increasing the information we obtain from the genealogy. Figure (1.2) shows a simulation of the Kingman coalescent for $k = 200$ that illustrates this last observation.

1.3 The Moran model

Now we will introduce a continuous time model for a population. The Wright-Fisher model we have previously discussed considers a population evolving in non overlapping generations, which might be accurate for certain plants (like seasonal crops and fruits) but is clearly not true about humans or flies. The Moran model (due to Moran [1958]) is a continuous time process which allows for overlapping of generations, in particular the times at which two given individuals reproduce are almost surely distinct.

Definition 1.7 *A population of size N evolves according to the Moran model if at exponential times with rate $\binom{N}{2}$ a pair of individuals is sampled uniformly at random from the population, one of which dies and the other splits in two, the two possibilities being equiprobable.*

We can see in Figure 1.3 how the Moran model works both forwards and backwards in time. The left diagram shows the evolution of the population, where an up or down arrow tells us that

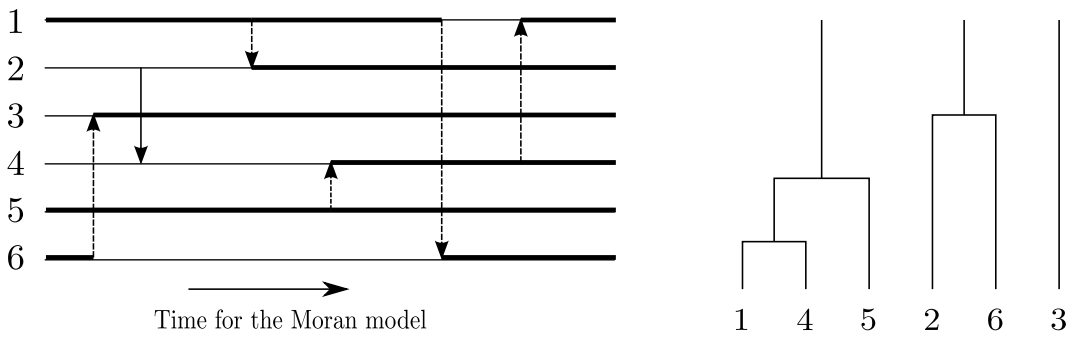

Figure 1.3: The Moran model. On the left we see the evolution of a population according to the Moran model. The arrow points to the individual who is replaced by a copy of the individual at the source of the arrow. On the right we see the resulting genealogy, obtained by tracing the individuals backwards in time.

an event occurred at that time, as well as which two individuals are involved, with the one at the origin of the arrow reproducing, while the one at the tip dies and is replaced by a copy of the reproducing individual. We construct the genealogy by following a sample of genes backwards in time, and whenever one of them meets the tip of an arrow, it coalesces with the individual at the origin of said arrow (which can be said to be its parent).

Here we followed Etheridge [2011] when choosing the rates at which events happen, since this way we have that the genealogical process of the population is determined by Kingman's coalescent without requiring any time change or rescaling. Since the genealogy of the Wright-Fisher model is well approximated by the coalescent, we can imagine that even though the dynamics of these two models are somewhat different, the big picture they convey is not too different.

1.3.1 Diffusion approximation and Duality

We will now show that, under the right scaling, the behaviour of both the Wright-Fisher and the Moran model is also very close forwards in time, as they both share a limiting object, namely, the Wright-Fisher diffusion. As the name suggests, this diffusion can be obtained by rescaling the Wright-Fisher model (see, for example, Etheridge [2011] or Durrett [2008]). We will assume this fact and proceed to prove that it is also the case for the Moran model. Remember we used

$p(t)$ to refer to the proportion of the population carrying the A allele.

Definition 1.8 *The Wright-Fisher diffusion is the solution of stochastic differential equation:*

$$dp(t) = \sqrt{p(t)(1-p(t))}dW(t); \;\; p(0) = p,$$

where W is a standard Brownian motion.

Equivalently, we can think of the Wright-Fisher diffusion as the strong Markov process with infinitesimal generator

$$\mathcal{L}f(p) = \frac{1}{2}p(1-p)\frac{d^2}{dp^2}f(p).$$

The well posedness of the stochastic differential equation is addressed in Sections 5.3 and 8.8 of Durrett [2008], where its shown that a weak solution exists. Though we won't go into any more details here, but we will use this fact to show that the rescaled version of the Moran model converges to this diffusion.

Theorem 1.9 *For a population of size N evolving according to the Moran model, let $(p^N(t))_{t\geqslant}$ be the fraction of individuals of type A, and suppose the initial state is $p^N(0) = p \in (0,1)$. Then*

$$\left(p^N(t)\right)_{t\geqslant 0} \xrightarrow{(d)} \left(p(t)\right)_{t\geqslant 0} \text{ as } N \to \infty.$$

Proof: Here we will follow the proof given by Berestycki [2009] We begin by looking at the infinitesimal behaviour of the Moran model. Suppose that $p(t) = p$, and let Δt be a small (infinitesimal) increment,

$$p(t+\Delta t) = \begin{cases} p + \frac{1}{N} & w.p. \quad \frac{N(N-1)}{2}p(1-p)\Delta t + o(\Delta t), \\ p - \frac{1}{N} & w.p. \quad \frac{N(N-1)}{2}p(1-p)\Delta t + o(\Delta t), \\ p & w.p. \quad 1 - N(N-1)p(1-p)\Delta t + o(\Delta t). \end{cases}$$

This is clear from our construction of the Moran model through exponential clocks, where at rate $\binom{N}{2}$ we select a pair and then split one of them while killing the other. Let $\Delta p(t) =$

$p(t + \Delta t) - p(t)$, it is immediate to obtain that the infinitesimal parameters are given by:

$$\mathbb{E}\left[\Delta p(t)|p(t)\right] = 0,$$

$$var\left[\Delta p(t)|p(t)\right] = \frac{N(N-1)}{N^2} p(t)(1-p(t))\Delta t + o(\Delta t).$$

It is clear that as $N \to \infty$, these parameters converge to the infinitesimal parameters of the Wright-Fisher diffusion, and thus, standard martingale problem arguments guarantee the convergence. ■

Now we can go back to some of the questions we were trying to answer earlier. We had asked about the time to fixation of a population $T := \min\{t : p(t) = 1 \text{ or } p(t) = 1\}$, that is, how long does it take for the allele A to either disappear or take over the whole population? Though this is a hard question to answer for populations of fixed size N, for the Wright-Fisher diffusion it is relatively simple to calculate the time it takes to hit either 0 or 1. Indeed, using the methods found in Karlin and Taylor [1981] we arrive at

$$\mathbb{E}_p[T] = 2\left(p\log(p) + (1-p)\log(1-p)\right),$$

where p is the initial fraction of the population with the A allele. If we wish to relate this to the number of generations in the Wright-Fisher model we obtain

$$2N\left(p\log(p) + (1-p)\log(1-p)\right)$$

as an approximation whenever we have a large population, where the factor of N accounts for the rescaling we did to obtain the limiting diffusion (remember we rescaled time by units of N generations). This showcases the applications and power of the diffusion approximation.

We have seen that the proportion of of a certain type in both the Moran and the Wright-Fisher model can be approximated forwards in time by a diffusion, and that both of their genealogies are well approximated by Kingman's coalescent. Now we will take a look at a very powerful tool when dealing with population models (and particle systems in general): duality.

Definition 1.10 *We say that two processes X and Y are dual to each other if there exists a function $\psi(x, y)$ such that the following relation holds:*

$$\mathbb{E}_x^X[\psi(X(t), y)] = \mathbb{E}_y^Y[\psi(x, Y(t))]$$

for all x and y, where we use $\mathbb{P}_z^Z[\cdot]$ for the law of Z conditioned on $Z(0) = z$.

Notice that x and y belong to different spaces, since x varies in the state space of the process X, while y does so in the state space of Y, and these two spaces might be very different. For example, in our case, one of the processes, the forwards in time model, gives us the proportion of a certain type in our population, while the (backwards in time) coalescent takes values in the space of partitions. Still, if allowing y to vary in $\psi(x_0, y)$ gives place to a rich enough family of functions, we can characterize the law of $X(t)$ started from x by looking at the right hand side of the equation above. This approach has proven very successful in proving uniqueness and existence of processes by looking at its duals. For a more in depth discussion, see Etheridge [2000].

In our case there is a particularly intuitive interpretation of the duality relation, namely, the dual of the process that models the population forwards in time is precisely the one that models its genealogies backwards in time.

We now borrow an argument from Berestycki [2009] that gives good intuition as to why a duality relationship between the Wright-Fisher diffusion and Kingman's coalescent holds. Consider a population with an initial proportion p_0 of A alleles. At time t pick n individuals at random from the population. What is the probability of $E = \{n \text{ individuals are of type } A\}$? We could go back t units of time, when we would have $|\pi_t|$ different lineages, each of which has the right type with probability p_0, and only if all of them do will E hold:

$$\mathbb{P}[E] = \mathbb{E}_{|\pi_0|=n}[p_0^{|\pi_t|}].$$

But we could also use the Wright-Fisher diffusion for our calculations. At time t the proportion of individuals of type A is given by $p(t)$, so if we pick n of them at random, we have

$$\mathbb{P}[E] = \mathbb{E}_{p_0}[p(t)^n].$$

Putting both estimates together we conclude:

Theorem 1.11 *Let $\mathbb{P}^d$ and $\mathbb{P}^c$ be the laws of the Wright-Fisher diffusion and Kingman's coalescent respectively. Then, for all $0 < p < 1$, and for all $n \in \mathbb{N}$ we have:*

$$\mathbb{E}^d_{p_0}[p(t)^n] = \mathbb{E}^c_n[p_0^{|\pi_t|}].$$

1.4 Many types and big events

So far we have considered a population where there are only two types, but one would expect most populations to manifest a much larger variety. Fortunately, the results explained above can be easily generalised to the case where we consider k types.

We can take as the starting point either a Wright-Fisher model or a Moran model, but where individuals can be of k types. It is not hard to show that, after a suitable rescaling, the population can be modelled by a k-dimensional Wright-Fisher diffusion. If we take $p^i(t)$ to be the limiting proportion of the population that is of type i, we have that the vector

$$p(t) = \Big(p^1(t), p^2(t), ..., p^k(t)\Big) \in \{(p_1, p_2, ..., p_k) : p_i \geqslant 0, \sum_i p_i = 1\}$$

is a diffusion with generator

$$\mathcal{L}(f) = \frac{1}{2}\sum_{i=1}^{K-1} p_i(1-p_i)\frac{\partial^2 f}{\partial p_i^2} - \sum_{1\leqslant i<j\leqslant K-1} p_i p_j \frac{\partial^2 f}{\partial p_i \partial p_j}.$$

The proof has the same ingredients as the proof of Theorem 1.9; for a rigorous derivation we refer to Etheridge [2011].

One could go one step further and consider an infinite number of types. That was what

Ohta and Kimura [1973] did by taking a countable set of possible types when studying electrophoretically detectable alleles. Later Fleming and Viot [1978] would generalise the setting even further, allowing the type space to be a compact metric space. Before explaining the (generalised) Fleming-Viot model, we will make an observation.

In the models we have presented so far, as well as in the models mentioned in the previous paragraph, the number of offspring that any individual leaves is negligible compared to the population size, a fact that shows the moment we decide to trace back the ancestry of a sample. Since the ancestry in all these population models is approximated by Kingman's coalescent, we know that the probability that three or more individuals share a parent in the previous generation is very small, even when compared to the probability that two do.

There are many scenarios where one could consider this to be a bad approximation. There are species where a single individual might lay thousands of eggs, and thus it is likely that sometimes a sizeable fraction of the next generation comes from it. One can also consider the so called population bottlenecks; say that for some reason the population is severely reduced in numbers and only a few individuals survive, then each of them is likely to father a significant fraction of the next generation.

In any case, introducing big events, understood as events where one individual parents a sizeable proportion of the population, should be of interest. Notice that if we are following the ancestry of a sample, in contrast with the Kingman coalescent, we would actually expect that more than two lineages coalesce given a big event. We will come back to this idea later.

1.4.1 The Fleming-Viot process

Donnelly and Kurtz [1999] introduced a clever construction of the Fleming-Viot process that incorporates big events. Here we will give an overview of the so called modified lookdown construction. The qualifier 'modified' has been used to distinguish both the process and the construction from the classical (no big events) Fleming-Viot and the original lookdown construction that appeared on an earlier paper of the same authors (Donnelly and Kurtz [1996]);

we will drop it from here on. The construction is very general and allows for many different scenarios, for example a varying population size, mutation of types, etc, but we will restrict ourselves to a fixed population size without mutation since this is enough for our exposition.

Consider a countably infinite system of particles, each one identified by a level $j \in \mathbb{N}$, and to each level we assign types ζ_t^j from a type space E. We can think of the type space as being $\{1, 2, ..., k\}$, $[0, 1]$ or any other compact metric space that suits our purpose. We require the initial configuration of types to be exchangeable, so that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} \delta_{\zeta_0^j} = \mu$$

for some finite measure μ on E. For clarity we can consider the types to be i.i.d uniform random variables on $[0, 1]$.

The types are updated in two different ways, both of which preserve the exchangeability.

The first part resembles the Moran model. For $i < j \in \mathbb{N}$ let $L_{i,j}$ be independent Poisson processes with rate 1. Whenever the Poisson process $L_{i,j}$ jumps, the particle at level j changes its type to the type of the particle at level i, that is, if $\Delta L_{i,j}(t) = 1$ then

$$\zeta_t^j = \zeta^i(t-).$$

Observe that the updating is well defined since for any n there are only finitely many processes $L_{i,j}$, $i < j$ at whose jump times the type of ζ^j must be modified. The updating mechanism gives place to the lookdown moniker since it is always the individual in the higher level that takes the type of one at a lower level.

We now introduce the big events. Let Π be a Poisson point process on $\mathbb{R} \times (0, 1]$ with intensity $dt \times F(du)$ for a probability measure F on $(0, 1]$. For every point $(t_k, u_k) \in \Pi$, let Y_{kj}, $j \in \mathbb{N}$ be i.i.d uniform ($[0, 1]$) random variables. We select a proportion u_k of the levels by taking those for which $y_{kj} < u_k$, and we take the smallest such j, say level i. We can now update the types

in the following manner:

$$\zeta_t^j = \begin{cases} \zeta_t^i & \text{for } j \text{ such that } , y_{kj} < u_k \\ \zeta_{t-}^j & \text{otherwise,} \end{cases}$$

We can see that this mechanism is well defined. On one hand, the rate at which this events arrive is finite, and on the other, at any such event the type at level n is determined by $\{Y_{kj}\}_{1 \leqslant j \leqslant n}$.

Theorem 1.12 *There exists a Markov process $(Z_t)_{t \geqslant 0}$ that takes values in the space of probability measures on E such that:*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{j=1}^{N} \delta_{\zeta_t^j} = Z_t.$$

This process is known as the generalised or Λ-Fleming-Viot process.

The reason for calling the limiting process Λ-Fleming-Viot will be addressed in the next section.

1.5 Λ-Coalescents

The previous section introduced a more general model for populations, where infinitely many types can be considered, as well as large events where a particular individual might have a sizeable impact on the configuration of the population. Here we consider the genealogical process pertaining to this model. It is called the Λ-coalescent and was first introduced and studied by Donnelly and Kurtz [1999], Pitman [1999] and Sagitov [1999] who were interested in coalescents that would allow multiple collisions, that is, where more than two blocks might merge simultaneously. In order to define it we need to begin with a restricted version that starts from k blocks.

Definition 1.13 *A k-Λ-coalescent is a $\mathcal{P}^k$-valued Markov process $\pi^k = (\pi_t^k, t \geqslant 0)$ such that π_t^k is exchangeable for any $t \geqslant 0$ and consistent in the sense that the law of π^k restricted to $[m]$*

is that of π^m *for all* $1 \leqslant m \leqslant k$. *The rate at which* j *blocks of* π^k *merge given that there are currently* b *blocks in the partition is given by* $\lambda_{b,j}$, *independently of* k.

Given a family of consistent k-Λ-coalescents, in similar fashion to the extension of the Kingman coalescent, we know that there exists a process π with values in $\mathcal{P}$ whose restriction to $[k]$ has the law of π^k. We call this process the Λ-coalescent. It is not hard to see that in order to be consistent the rates must satisfy the following recursion:

$$\lambda_{b,j} = \lambda_{b+1,j} + \lambda_{b+1,j+1}.$$

Pitman [1999] showed that in order for the process to exist these rates must have very precise forms:

Theorem 1.14 *The* Λ*-coalescent exists if and only if there is a finite measure* Λ *on* $[0,1]$ *such that:*

$$\lambda_{b,j} = \int_0^1 u^{j-2}(1-u)^{b-j}\Lambda(du).$$

An interesting by-product of this result is that it allows a Poissonian construction of the process. Take π, we say that a u-merger happens if we choose a proportion u of the blocks by independent coin toss and merge all of them together. The Λ-coalescent can be then thought of as a succession of u-mergers plus a Kingman component:

Theorem 1.15 *Let* Λ *be a measure on* $[0,1]$ *and let* $\alpha := \Lambda(\{0\})$. *Let* $(t_i, u_i)_{i\geqslant 1}$ *be the points of a Poisson point process on* $\mathbb{R}^+ \times (0,1]$ *with intensity* $dt \times u^{-2}(\Lambda(du) - \alpha\delta_0)$. *The process* $(\pi_t,\, t \geqslant 0)$ *may be constructed by allowing every pair of blocks to merge at rate* α *and performing a* u_i*-merger at each time* t_i.

This construction has very similar ingredients to the construction of the Λ-Fleming Viot in §1.4.1, where the u-mergers are analogous to the big events, while the pairwise mergers relate to the lookdown events. In the absence of the Kingman component we see that the measure $F(du)$ in §1.4.1 is given by $F(du) = u^{-2}\Lambda(du)$, which motivates the naming of the process.

It is no surprise that one can state a duality relation between this pair of processes, in particular if one considers the case where $\Lambda(\{0\}) = 0$, and extends the Poisson point process to the whole real line, then the same Poisson point process and its time reversal drive the Λ-Fleming-Viot and the Λ-coalescent. This forwards and backwards in time duality can be stated as follows:

Theorem 1.16 *Let $\mathbb{E}^{\rightarrow}$ denote the expectation for the Λ-Fleming-Viot process $(Z_t,\, t \geqslant 0)$ and $\mathbb{E}^{\leftarrow}$ denote the expectation for the Λ-coalescent $(\pi_t,\, t \geqslant 0)$ restricted to $[k]$. Let $f : E^k \to \mathbb{R}$ be a bounded measurable function. Then, for $\Phi : \mathcal{M}_E \times \mathcal{P}^k \to \mathbb{R}$*

$$\Phi(Z, \pi) = \int \int Z(dx_1)...Z(dx_{|\pi|}) f(x^{\pi}),$$

we have that:

$$\mathbb{E}^{\rightarrow}[\Phi(Z_t, \pi_0)] = \mathbb{E}^{\leftarrow}[\Phi(Z_0, \pi_t)].$$

There are many generalisations to the models we have considered here, from varying population size to mutations, that can be introduced to the framework we have so far presented, but we now turn our attention to a different difficulty that arises when studying populations and that we have so far ignored, spatial structure.

CHAPTER 2

The Spatial Λ Fleming Viot Process

2.1 Introducing spatial structure

So far we have discussed models where we assume that the population is well mixed, and thus have ignored any spatial structure for the population. But in reality spatial distribution is a key feature of any population, since it can strongly affect which individuals interact with each other as well as the environment in which they develop, among other possible effects. Therefore is not surprising that a lot of effort has been made to introduce spatial structure to models of population genetics.

One of the very first attempts to do so was Wright's island model (1943). As the name

suggests, one could imagine a group of islands, each of which evolves according to the usual Wright-Fisher model, except for the fact that some migration is allowed between all islands. The population of each island constitutes what is called a *deme*.

More generally, we could take any (finite or countable) graph and let a deme sit at every vertex $i \in I$. Suppose that each deme evolves according to the Wright-Fisher model, but every generation, after reproduction we send $N_i m_{ij}$ individuals from deme i to deme j where N_i is the population size at deme i ($m_{ij} = 0$ whenever the two vertices are not connected in the graph). In order to preserve population size at each deme we ask that:

$$\sum_j N_j m_{ji} = \sum_j N_i m_{ij}.$$

The genealogical trees for this model are given by a structured coalescent process, where two lineages coalesce at rate $1/N_i$ if they are both in deme i, and each lineage jumps from deme i to deme j at rate $m_{ji}N_j/N_i$, which one can think of as the probability that it had a parent in the other deme in the previous generation.

If we rescale this model diffusively, as we did with the Moran model earlier (now we also let the migration rates go to 0 by taking $m^n_{ji} := m_{ji}/N$) , we arrive at the following system of equations:

$$dp_i = \sum_j m_{ji}(p_j - p_i)dt + \sqrt{\frac{1}{N_e^i} p_i(1-p_i)}dW_i, \tag{2.1}$$

where $\{W_i\}_{i\in I}$ are independent Brownian motions, N_e^i is the effective population size of deme i, and p_i is the proportion of type A alleles in deme i. This system of equations is known as the Kimura stepping stone model Kimura [1953], and it is usual to choose a lattice, say $\mathbb{Z}^2$, as the underlying graph, to mimic a 2-dimensional space. Although there have been many variations on this theme, the main features of most models are very similar to the ones of the stepping stone model, namely, a population is subdivided into demes, and these interact via a migration process.

The genealogical process one obtains from the stepping stone model is very similar to the

Kingman coalescent, and is referred to as a *structured coalescent*. Say we follow a finite sample of individuals backwards in time, and we start from a vector $(n_i(0))_{i\in I}$ where $n_i(t)$ is the number of individuals (or later blocks) in deme i at time t. The dynamics are given by:

1. for each $i \in I, n_i \to n_i - 1$ at rate $\frac{1}{N_i}\binom{n_i}{2}$,

2. for all $i \neq j \in I$ we have $n_i \to n_i - 1$ and $n_j \to n_j + 1$ at rate $n_i m_{ij}$.

In words, two blocks in the same deme coalesce at the rate of the Kingman coalescent, but any block can jump to a different deme according to the migration rates. So, if we start with individuals in different demes, we first have to wait for them to arrive in the same deme before they have a chance to coalesce. In a similar fashion to the non spatial case, we have a moment duality between the structured coalescent $\underline{n} := (n_i)_{i\in I}$ and the allele frequencies of the stepping stone model $\underline{p} := (p_i)_{i\in I}$. Let $f(\underline{p}, \underline{n}) = \prod_{i\in I} p_i^{n_i} =: \underline{p}^{\underline{n}}$, we have

Lemma 2.1 *Suppose that $\underline{p}(t)$ evolves according to the Kimura stepping stone model, and that $\underline{n}(t)$ is the corresponding structured coalescent, then we have that:*

$$\mathbb{E}^K[\underline{p}(t)^{\underline{n}(0)}] = \mathbb{E}^C[\underline{p}(0)^{\underline{n}(t)}].$$

An interesting feature of the genealogies in a structured population like this is that, under lenient conditions (see Wilkinson-Herbots [2003]), the expected coalescence time of two lineages chosen at random from the the same deme is the same as if we took a non structured population with the same total size (in this case, the number of demes times their population).

This is actually easy to show for the Wright island model with D demes. Let T_{11} be the expected coalescence time of two lineages sampled from the same island, and T_{12} be the coalescence time of two lineages sampled from different islands, and take m to be the migration rate across islands. If initially they are both in the same island, the rate at which either of them migrates or they coalesce is $2m + 1/N$, but whenever they are in different islands the only interesting event is either of them migrating which occurs at rate $2m$, and will send one of them

to the other's island with probability $1/(D-1)$. Thus:

$$T_{11} = \frac{1}{2m + 1/N} + \frac{2m}{2m + 1/N} T_{12},$$

$$T_{12} = \frac{1}{2m} + \frac{1}{D-1} T_{11} + \frac{D-2}{D-1} T_{12},$$

which yields $T_{11} = ND$, $T_{12} = (D-1)/2m + ND$. Notice that if $m \to \infty$, then the same result also holds for lineages sampled from different islands.

More generally, Nordborg and Krone [2002] coined the term *collapse of structure* to indicate that provided migration is fast enough, the genealogical tree of a sample can be well approximated by an unstructured Kingman coalescent. Roughly, fast migration rates result in an averaging procedure that ends up ignoring the spatial structure. The variance of the time to coalesce will differ from an unstructured population, but this difference will vanish if migration happens sufficiently fast. In similar fashion Wakeley [2001], Wakeley and Aliacar [2001] show that an analogous effect is achieved if the number of demes is allowed to go to infinity, but time must also be rescaled by a similar factor. In Theorem 2.4 we will show that a similar result holds for our model under certain regimes.

So far, the models we discussed introduce spatial structure by subdividing the population into demes. Though this might be a good approximation for certain situations, it is of interest to understand what happens with populations that are allowed to live in a continuum.

2.2 Trouble: modelling evolution in a spatial continuum

A natural approach to obtaining a model for a spatial continuum would be to take a diffusive rescaling of the stepping stone model. Indeed, if we are working in one dimension, we can rescale the system of ordinary differential equations (2.1) to obtain a single stochastic partial differential equation:

$$dp_t = \Delta p dt + \sqrt{p_t(1-p_t)} W(dt, dx), \tag{2.2}$$

where W is space-time white noise. The genealogy of a population converges to a system of coalescing Brownian motions, where the coalescence rate depends upon the local time they spend together. But this approach breaks down if we try to move to higher dimensions. For example, in dimensions two and higher the genealogies still converge to Brownian motions, but they never meet, and thus we see no coalescence. Moreover, as one should expect since coalescence corresponds to genetic drift forwards in time, the diffusive rescaling of the forwards in time model in two dimensions gives us a deterministic system of equations. If instead, we consider the 2-dimensional version of (2.2) we find that it has no solution.

To try and bridge these difficulties various ideas have been proposed. For example, although two dimensional Brownian motions never meet, they will come arbitrarily close to each other. So we could model genealogies through 2-dimensional Brownian motions that coalesce at a rate that is a function of their distance, with both of them jumping to a new location. One of the problems with this idea is that it lacks sampling consistency. That is, if we follow the genealogy of a sample of size k we will see jumps that we wouldn't if we ignored one of the individuals.

Malécot [1948] (and similarly Wright [1943]) pioneered the study of populations in a continuum. They assumed an infinitely many alleles model with a fixed mutation rate, where the spatial displacement was given by a Gaussian distribution, which indicated the position of the offspring relative to that of its parent. They were interested in calculating what they called the probability of identity by descent $F(x)$, that is, the probability that two individuals at a distance x carry identical alleles inherited from the same ancestor. Even though their work preceded the development of the coalescent theory by a few decades, the question they were asking is very closely related. Since in the infinitely many alleles model, two individuals will have the same allele only if they inherited it from the same ancestor and no mutation happened on that locus, it is equivalent to calculating the generating function of the time to the MRCA of the sample. Through a recursion, Malécot was able to obtain an approximation to $F(x)$ given by

$$F(x) \approx \frac{1}{\mathcal{N} + \log(\sigma/\kappa\sqrt{2\pi})} K_0\left(\sqrt{2\mu}\frac{|x|}{\sigma}\right), \quad |x| > \kappa, \tag{2.3}$$

where κ is a local scale, σ^2 is the variance of the Gaussian distribution that determines the location of the offspring, K_0 is the modified Bessel function of the second kind, and $\mathcal{N}$ is the neighbourhood sice, which can be thought of as the potential number of parents of an individual. The formula 2.3 is known as the *Wright-Malécot formula*.

Wright and Malécot's models shared three assumptions about the dynamics of a population evolving in a continuum, namely: 1) individuals are distributed randomly across space with constant expected density; 2) they reproduce independently, leaving a Poisson (with mean 1) number of offspring; 3) the offspring are scattered independently according to a Gaussian distribution.

Though this seems a very reasonable extension of the Wright-Fisher dynamics to a continuum, Felsenstein [1975] showed that these assumptions are inconsistent, since in one and two dimensions, a population that evolves according to a spatial branching process, as given by the last two assumptions, has no stationary distribution, and thus won't be uniformly dispersed but it will rather develop "clumps". Felsenstein tried to overcome this by working on a torus instead of $\mathbb{R}^2$, and to avoid extinction he decided to specify the population size exogenously, but could still not avoid the appearance of clumps, a problem he famously named *the pain in the torus*.

It is worth mentioning that although Wright and Malécot's assumptions were inconsistent, the formula 2.3 provides a very good approximation of what would happen in the stepping stone model. It is somewhat surprising that the Wright-Malécot formula, based on an inconsistent model, and a discrete space approximation both give remarkably similar estimates. In Barton *et al.* [2002] the authors extend Wright and Malécot's's formula to a class of locally regulated populations, but sadly there are very few explicit models in this class. In summary so far, we can either mimic a continuum by subdividing the population and taking a fine lattice, or we can work directly on $\mathbb{R}^2$ with Malécot's formula if we are willing to ignore the inconsistency of its assumptions.

More recently, models that involve branching rates that are locally regulated have been developed, with the explicit aim of overcoming clumping, but they soon become too complicated

to analyse, plus the genealogical trees are usually intractable [Barton *et al.*, 2013b].

2.2.1 Biological considerations

Before we move forward it is valuable to bring to the reader's attention some biological issues that have arisen with the arrival of new biological data. In addition to their previously discussed mathematical shortcomings, the patterns that are seen in the data are at odds with the predictions of the models. In two dimensional space the Wright-Malécot formula predicts, roughly, exponential decay in the probability of identity by descent as a function of the separation between individuals, while populations show correlation over much larger scales. At the same time the classical models assume that genes at unlinked loci evolve independently, while we can actually find patterns across loci, reflecting migration and other large demographic events, and giving rise to the field of phylogeography. We can also add to these the fact that the discrepancy between census and effective population size, meaning that the actual level of genetic diversity for large populations is orders of magnitude smaller than the theory suggests. Barton *et al.* [2010b] provide a much more in depth look into these biological considerations.

Finally, a feature of the classical models is that as we let N tend to infinity we are implicitly allowing our individuals to pick their parents uniformly from a very large pool. This is reflected as we trace genealogies backwards in that we obtain a (structured) Kingman coalescent as an accurate approximation to the family trees: since we have a large pool of parents, we expect that coalescence will only happen through pairwise mergers. The size of this pool is commonly referred to as *neighbourhood size* and for many populations it may not be sensible to allow it to go to infinity.

The Spatial Λ Fleming Viot (SΛFV) model which we will now discuss aims to overcome the mathematical problems discussed in § 2.2, and at the same time address the biological issues raised above.

2.3 The individual based model

So far we have given an overview of the classical approach to population genetics, introduced some important concepts and definitions, and recounted the problems that have been encountered when introducing spatial structure to these models, in particular whenever the goal has been to model populations evolving across a continuum. Here we will present a model first proposed by Etheridge [2008] and further developed in other papers (see for example Barton *et al.* [2010a], Berestycki *et al.* [2009], Berestycki *et al.* [2013], Etheridge and Véber [2012]) which aims to overcome the pain in the torus while providing a model for understanding evolution and a suitable framework for doing inference on the vast amounts of genetic data that have become available in recent years.

We begin our presentation with an individual based model. Just as in the case of the stepping stone model, where we first considered finite populations evolving according to the Wright-Fisher model plus an immigration process, it is this first model where we can more readily see the dynamics that govern the population, and at the same time it allows for some intuition of the limiting process. The model can be thought of as a spatial version of the Λ Fleming-Viot process, which we described earlier, and which we also obtained as the limit of individual based models.

Take a population evolving in $\mathbb{R}^d$, represented by a purely atomic measure where each atom gives the spatial location of an individual, and give each individual a type $k \in K$. The model will then be parametrized by a real number $\lambda > 0$ and a σ-finite measure on $(0, \infty) \times (0, 1]$

$$\xi(dr, du) = \mu(dr)\nu_r(du),$$

where we assume that

$$\int_0^\infty \int_0^1 ur^d(1 + r^d)\nu_r(du)\mu(dr) < \infty. \tag{2.4}$$

Take the initial configuration of the population to be a Poisson point process in $\mathbb{R}^d$ with intensity λ (the types can be taken to be uniform on K) and take Π to be a Poisson point process on

$\mathbb{R}^+ \times \mathbb{R}^d \times (0,\infty) \times (0,1]$ with intensity $dt \otimes dx \otimes \xi(dr, du)$. The dynamics of the process are as follows.

1. A point $(t, x, r, u) \in \Pi$ indicates that a reproduction event takes place in the ball $B_r(x)$, centred at $x \in \mathbb{R}^d$ and with radius r.

2. If the ball is empty do nothing. If not:

   - pick one of the individuals inside uniformly at random to be the parent;
   - kill each individual (including the parent) inside $B_r(x)$ independently with probability u;
   - throw down offspring of the same type as the parent inside the ball according to an independent Poisson point process with intensity $u\lambda dy$.

The first thing we notice is that now reproduction is not driven by individuals, but by events that affect only a particular region of the space. This can be thought of as a regulating mechanism, since whenever the event encompasses a crowded region, the probability that any particular individual reproduces is small. At the same time, an individual in a sparsely populated region has a good chance of leaving offspring whenever it is inside an event.

Notice as well that our choice of ξ allows us to model both 'small' and frequent events together with rare 'large' events. The rare events are introduced to model large demographic events, such as extinction/recolonization events, which would act as a sort of bottlenecks for the genealogies and might help explain the low rate of genetic diversity as well as the geographical correlation across loci.

Berestycki *et al.* [2009] look at this individual model, under the condition (2.4) which guarantees the existence of the process. They are able to prove that for λ large enough, if we start the population from a Poisson random measure with constant intensity, then the process survives with probability one. The proof relies on comparison to oriented percolation.

Following the genealogy of a sample backwards though soon becomes messy. Since regions of space can and will become empty, the presence of individuals in a certain region provides

information about the reproduction events that have happened there. Ideally we would like to reverse the Poisson point process and then follow the lineage of a gene backwards. But since the lineage can't jump to a region that was unoccupied, it is not possible to recreate the genealogy just by this reversal. But if λ is large enough, we expect to see very few empty regions, and thus can expect the genealogy to be well approximated by coalescing random walks driven by the reversed Poisson point process.

This reasoning brings us to consider the high density limit of the individual based model, namely, letting $\lambda \to \infty$. That way we wouldn't have to worry about empty regions, and can expect that the genealogies will be obtained by just reversing the Poisson point process. Interestingly, even though we will let the population density go to infinity, we will be able to retain the signature of a small neighbourhood size, which will become apparent when we see that multiple mergers of ancestral lineages are still possible when we follow the genealogical process.

It is possible to build an individual based model that overcomes the problem of empty regions. Barton *et al.* [2010b] consider a modification of the above dynamics, where instead of a ball, an event can affect the whole population, with the probability of it affecting a given individual decaying as a Gaussian distribution away from the centre of the event. This way, it is always possible, when going backwards, to find a parent to which the genealogical process can jump. Although closely related, the limiting versions of both individual based models are distinct; in the second scenario the probability that a particular individual is affected by an event decays according to its distance from the event's centre, whereas in the model we consider all individuals inside the event have the same probability of being affected.

2.4 Definition of the spatial Λ Fleming-Viot process

We will now introduce the Spatial Λ Fleming-Viot process. As before, we take the population to evolve in $\mathbb{R}^d$ and the types to come from some compact space K. Denote the space of of all probability measures on K by $\mathcal{M}_1(K)$. The dynamics are again dictated by the Poisson point

process Π, which takes values in $\mathbb{R}^+ \times \mathbb{R}^d \times (0,\infty) \times (0,1]$ and has intensity $dt \otimes dx \otimes \xi(dr, du)$.

Definition 2.2 *The spatial Λ-Fleming-Viot process* $\{\rho(t,x,\cdot), x \in \mathbb{R}^d, t \geqslant 0\}$ *taking values in the space of measurable functions* $\rho : \mathbb{R}^d \to \mathcal{M}_1(K)$ *specifies a probability measure on* K *for every* $t \geqslant 0$ *and for almost every* $x \in \mathbb{R}^d$. *The dynamics of the process are given driven by* Π. *At every point* $(t,x,r,u) \in \Pi$ *we choose a position* z *uniformly at random from with the ball* $B_r(x)$ *and sample a type* k *according to* $\rho(t-,z,\cdot)$. *Then, for all* $y \in B_r(x)$ *we update the process in the following manner:*

$$\rho(t,y,\cdot) = (1-u)\rho(t-,y,\cdot) + u\delta_k,$$

while for all y *outside the ball the process remains unaffected.*

As with the individual based model, conditions on the intensity of Π are required to guarantee the existence of the process. In Barton *et al.* [2010a] it is shown that is enough to assume that:

$$\Lambda(u) := \int_{(0,\infty)} u r^d \nu_r(du)\mu(dr) \tag{2.5}$$

defines a finite measure on $[0,1]$. Notice that this is a weaker condition than (2.4) which was required for the existence of the individual based model. Instead of describing the location and type of specific individuals, for every space-time point, this process specifies a measure $\rho(t,x,\cdot)$. The natural interpretation is that if we were to sample an individual from the point x at time t, its genetic type would be given by a random pick from this measure.

Though the SΛFV was proposed as a high density limit of the individual based model, a rigorous proof of the convergence has only recently been obtained by Etheridge and Kurtz [2014], in which the authors also weaken the conditions for existence of both the individual based model and the limiting model. As a corollary of their work, they have shown that the genealogies also converge so that the duality between the forwards in time allele frequencies and the coalescent is more robust than just the weak moment duality. The next section will introduce this dual process.

2.5 The genealogical process

In order to consider the ancestry we extend the Poisson point process Π to the whole real line so that it is time reversible. Now, if we sample an individual and follow its ancestry backwards in time, the ancestral lineage will be given by a jump process where the jumps are driven by Π. Whenever the lineage is inside one of the events (t, z, r, u) it has a probability u of being an offspring of the event and thus it jumps to the location of the parent, which is uniformly distributed in the ball $B_r(z)$. The rate at which the lineage jumps is given by:

$$dt \otimes \left(\int_{\mathbb{R}^d} \int_{(0,\infty)\times[0,1]} u \mathbf{1}_{\{0 \in B(x,r)\}} \xi(dr, du) dx \right).$$

The lineage can only jump from y to $y + x$ if the centre of an event lies inside $B_r(x) \cap B_r(y)$. Let V_r be the volume of a ball of radius r, and $L_r(x)$ be the volume of the intersection $B_r(x) \cap B_r(0)$, if we are interested in the intensity of the jumps from 0 t x it follows that:

$$\begin{aligned} m(x) : &= \int_{\mathbb{R}^d} \int_{(0,\infty)\times[0,1]} \frac{u}{V_r} \mathbf{1}_{\{z \in B_r(x) \cap B_r(0)\}} \xi(dr, du) dz \\ &= \int_{(0,\infty)\times[0,1]} \frac{u L_r(x)}{V_r} \xi(dr, du). \end{aligned}$$

In order to understand this last equation, suppose that the lineage is currently located at the origin, in order for it to jump to x, first both of them must be inside $B_r(z)$ for some $(t, z, r, u) \in \Pi$, so z must be in $B_r(x) \cap B_r(0)$ which has volume $L_r(x)$. And since the new location is chosen uniformly at random inside $B_r(z)$ and it must be in the portion of the population that was affected by the event, we obtain the factor u/V_r. For this jump intensity to correspond to a well-defined Lévy process, it is necessary that:

$$\int (1 \wedge |x|^2) m(x) dx < \infty,$$

which is guaranteed by our existence condition (2.5) for the SΛFV. In fact, (2.5) guarantees that the displacement of a lineage will be given by a compound Poisson point process.

So far we focused on the dynamics of a single lineage, and showed it follows a well-defined

jump process. But usually we are interested in tracing back the genealogy of a larger sample, in which case we need to follow the different lineages. We denote the genealogical process by $\mathcal{A}$. Starting from a finite sample of our population at time 0, for any given $t > 0$ (backwards in time) $\mathcal{A}_t$ indicates the position of all potential ancestors of our sample. It might be convenient then to think that $\mathcal{A}_t$ takes values in the space of finite integer point measures on $\mathbb{R}^d$. Just as it was the case forwards in time, the dynamics of the genealogy backwards in time are governed by Π. If $(t, x, r, u) \in \Pi$ we update the genealogical process at time t in the following manner:

1. We check if any of the ancestral lineages of $\mathcal{A}_{t-}$ are inside $B(x, r)$. If none of them are, we do nothing.

2. Otherwise every element of $\mathcal{A}_{t-}$ inside $B(x, r)$ is killed with probability u. If one or more lineages are killed, we add a new element to $\mathcal{A}_t$ with location z uniformly sampled from $B(x, r)$.

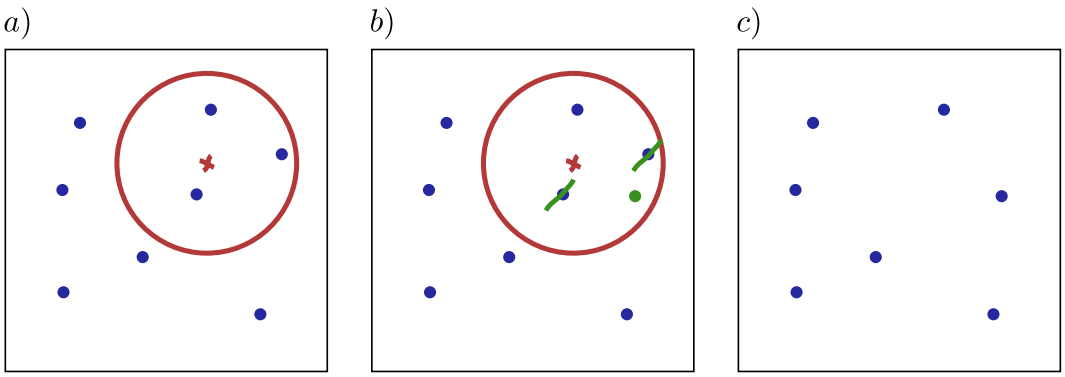

Figure 2.1: The updating procedure of the genealogical process $\mathcal{A}$. In panel a) we have the position of the lineages at time $t-$ in blue, and in red the centre and the border of the ball of an event $(t, x, r, u) \in \Pi$. Particles crossed out in panel b) are those that were killed in the event, while the green particle is the parent sampled from within the event. Panel c) shows the new configuration of particles.

We want to emphasize that the lineages inside an event are killed with probability u independently of the others, and thus even if only a single lineage is present, it might jump to a new location. This is in contrast to the model we mentioned in §2.2, where lineages behave as Brownian motion and can coalesce and jump to a new location only when they are close enough.

In this way, our model preserves sampling consistency, where the behaviour of k lineages has the same distribution as the behaviour of the restriction of $k+m$ lineages to just k of them.

We see that our lineages are moving according to a jump process, and that they might coalesce when close together. Let's look at the rate at which two lineages at a separation $z \in \mathbb{R}^d$ coalesce, which is given by:

$$\int_{(0,\infty)\times[0,1]} u^2 L_r(z)\xi(dr, du).$$

As was the case with the jump rate, condition (2.5) guarantees that this rate is finite, which in turns bounds the rate of all other possible coalescence events. Provided we have k lineages inside one of the events $(t, x, r, u) \in \Pi$, it is straightforward to notice that the probability that l of them coalesce is given simply by $u^l(1-u)^{k-l}$, which is closely related to the rates of a Λ-coalescent. The likelihood of seeing multiple mergers will strongly depend on our choice of parameters, as we will show in the next section, but this observation clarifies why the dual of the SΛFV is called the spatial Λ-coalescent. We will also give a formal statement of the duality between the the forwards and backwards in time models, where we will restrict ourselves to only 2 types, simplifying the required notation.

Remark 2.3 *The name spatial Λ-coalescent was previously used by Limic and Sturm [2006], but their version was constructed by considering a finite graph, where lineages migrate between demes, and whenever there is more than one lineage in a deme, they coalesce according to a Λ-coalescent.*

2.6 Some results about the SΛFV

We finish this chapter with an overview of some of the results that have been obtained regarding the spatial Λ Fleming-Viot process. We do not aim to provide a rigorous treatment of every result, but hope to give a brief outline of the main ideas behind them.

2.6.1 Large scale events and genealogies

Barton *et al.* [2010a] provide a proof of existence and uniqueness of the process, and they obtain some asymptotic results for the genealogies of individuals sampled from far apart. For the first part, they followed the ideas of Evans [1997] who uses a system of coalescing particles to construct a family of Markov processes that take values on the space of probability measures on $[0, 1]$. Although the particles in the work of Evans move independently, Barton et al. were able to adapt his arguments to work with the spatial Λ-coalescent we described above, and obtained existence and uniqueness of the SΛFV under condition (2.5).

Then they explicitly introduce a distinction between frequent small and rare large extinction / recolonization events, showing that the interplay between the two types of events dictates the asymptotic (in terms of the scaling) behaviour of the genealogies. They consider a torus $\mathbb{T} \subset \mathbb{R}^2$ of length L, and take small events to have bounded radius, while large events will affect a region of radius $\mathcal{O}(L^\alpha)$ for some $0 \leqslant \alpha \leqslant 1$. Now, if the small events arrive at some fixed rate, while large events arrive with a rate $1/\rho_L$ (which goes to 0 as L goes to infinity), what does the genealogy look like if we allow $L \to \infty$?

The answer is twofold. For $\alpha < 1$ they obtain a Kingman coalescent where the effective population size depends on the rate of the large events. If we allow large events to be commensurable with the length of the torus then the picture depends on the rate at which they arrive. A precise statement can be found in the paper, but here we outline the two main theorems.

Theorem 2.4 *(Thereom 3.3 in Barton et al. [2010a]) Suppose that $\alpha < 1$, then the law of the genealogical process for particles sampled uniformly at random from the torus of side L converges in law to the law of the Kingman coalescent.*

Depending on ρ_L the effective population size that will determine the timescale might depend on both small and large events.

Theorem 2.5 *(Thereom 3.7 in Barton et al. [2010a]) Take $\alpha = 1$, then we have three different cases according to the rate of the large events:*

1. *if* $\rho_L \approx L^2$*, on the timescale* ρ_L *the genealogical process converges to a spatial* Λ *coalescent where the lineages perform independent Brownian motions in between events.*

2. *If* $\rho_L \approx L^2 \log L$ *then the genealogy converges, on the timescale* ρ_L*, to a (non spatial)* Λ*-coalescent, possibly with a Kingman component.*

3. *If* $\rho_L \gg L^2 \log L$*, on the timescale* $L^2 \log L$ *the genealogy converges to a Kingman coalescent.*

Theorem 2.4 tells us that if large events are not large enough, in the limit they still won't capture more than two lineages at a time, which is why we arrive at the Kingman coalescent, but still they will affect the effective population size.

On the other hand, Theorem 2.5 shows that different regimes for the large events have qualitatively different effects on the genealogy. If they happen fast enough they will be the main source of coalescence, and in between events the lineages will move independently. In the intermediate regime, the time it takes for a large event to arrive allows mixing of the lineages' locations to be achieved, which means that the position of a lineage is uniform on the torus, and independent of the original location. Thus, the probability that any given lineage is affected by a large event is given by the fraction of the torus the event covers, and this way we recover a non spatial Λ-coalescent. Finally, if big events happen at a very slow rate, pairwise mergers dominate and we obtain a Kingman coalescent.

Notice that except for one case, we have that once again the spatial structure collapses and that the limiting genealogies do not have a spatial component.

2.6.2 The interface between two types

We will now restrict the type space to just two different alleles, say a and A, in order to describe the interplay between them on a large scale. By dividing the population into two, the aim is to understand the geometry that arises when different populations come together. Berestycki *et al.* [2013] work on the special case where the initial configuration is given by one half plane having only individuals of type a while the complementary half plane only has individuals of

type A. Considering only two types also has the advantage that we can present the moment duality of the SΛFV more concisely. The measure ρ can be replaced by the proportion of type A individuals in the following manner:

$$w(t,x) = \rho_t(x)(\{A\}) \quad \textit{Lebesgue a.e.}$$

We specify an initial state for our process by $w(0,\cdot) = w_0(\cdot)$, where w_0 is not random. We can now state the duality between the SΛFV and its genealogical process:

Theorem 2.6 *(Theorem 4.2 in Barton et al. [2010a]) Sample individuals from locations $x_1, ..., x_j$ and write $\xi_t^1, ..., \xi_t^{N_t}$ for the location of lineages evolving according to the genealogical process of § 2.5. Then for each $\psi \in C((\mathbb{R}^d)^j) \cap L^1(dx_1...dx_j)$ and every $t \geqslant 0$,*

$$\int_{(\mathbb{R}^d)^j} \psi(x_1, ...x_j)\mathbb{E}[w(t,x_1)...w(t,x_j)|w(0,\cdot) = w_0(\cdot)]dx_1...dx_j$$

$$= \int_{(\mathbb{R}^d)^j} \psi(x_1, ...x_j)\mathbb{E}[w_0(\xi_t^1)...w_0(\xi_t^{N_t})|N_0 = j, \xi_0^1 = x_1, ..., \xi_0^j = x_j]dx_1...dx_j.$$

When studying the interaction between two types, the authors consider two different cases, one where only small events happen, and a second case where large scale events are allowed to happen. Since in the first case a single lineage behaves as a random walk with bounded jumps, it seems natural to consider a diffusive rescaling

$$w_t^n(x) = w(nt, \sqrt{n}x).$$

For the next theorem, we start from the initial condition described above, that is, the half plane $\mathbb{H} = \{x \in \mathbb{R}^d : x_1 < 0\}$ has only type A individuals while $\mathbb{H}^c$ has only type a individuals. Also, we assume that the impact parameter u and the radius of the events r are fixed.

Theorem 2.7 *(Theorem 1.1 in Berestycki et al. [2013]) There exists a random space-time field $\{w_t^{(2)}(x), x \in \mathbb{R}^d, t \geqslant 0\}$ such that $w_t^n \to w_t^{(2)}$ as $n \to \infty$ in the sense of the finite dimensional distributions. Furthermore, at every $t \geqslant 0$, the local density $w_t^{(2)}(\cdot)$ of type A individuals can be*

described as follows. If W is a standard Brownian motion in d-dimensions and

$$p_2(t,x) := \mathbb{P}_x[W_{u\sigma^2 t} \in \mathbb{H}], \quad t \geqslant 0, x \in \mathbb{R}^d,$$

where $\sigma^2 := 1/d \int |z|^2 m(dz)$, then:

1. *if $d = 1$, $w^{(2)}$ is a random field of correlated Bernoulli random variables such that $\mathbb{E}[w_t^{(2)}(x)] = p_2(t,x)$ a.e., where the correlations can be obtained from Theorem 2.6. Furthermore, if started from the Heaviside function, the boundary between the two types follows a Brownian motion;*

2. *if $d \geqslant 2$, $w^{(2)}$ is deterministic and $w_t^{(2)} = p_2(t,x)$ a.e.*

We notice that in one dimension the types never coexist almost surely, since the proportions at every point are given by a Bernoulli random variable, while for higher dimensions the two types coexist everywhere. This is again related to the properties of Brownian motion. The ancestral lineages converge to independent Brownian motions that coalesce upon meeting, and since they can only meet in $d = 1$, we obtain the deterministic heat flow for higher dimensions.

In order to allow for large events, we take the radius to be random with $\mu(dr) = r^{-\alpha-d-1}\mathbf{1}_{\{r\geqslant 1\}}dr$ for a fixed $\alpha \in (1,2)$. The appropriate rescaling in this case will be

$$w_t^n(x) = w(nt, n^{1/\alpha}x).$$

Theorem 2.8 *(Theorem 1.5 in Berestycki et al. [2013]) There exists a random space-time field $\{w_t^{(\alpha)}(x), x \in \mathbb{R}^d, t \geqslant 0\}$ such that $w_t^n \to w_t^{(\alpha)}$ as $n \to \infty$ in the sense of finite-dimensional distributions. Furthermore, there exists a symmetric α-stable process X^α such that for all $t \geqslant 0$, $w^{(\alpha)}$ is a random field of correlated Bernoulli random variables with parameter $p^\alpha(x,t) := \mathbb{P}_x[X^\alpha_{u\sigma^2 t} \in \mathbb{H}]$ almost everywhere.*

The first thing we observe is that when large events are allowed to happen, the limiting process remains stochastic in all dimensions. The intuition is that when we look at the genealo-

gies, we would expect them to converge to symmetric stable processes. Though independent symmetric stable processes do not meet in higher dimensions, in this construction lineages do not evolve independently, and coalescence is still possible through the larger events that might capture lineages that are at arbitrarily large separations. If the ancestral lineages can coalesce, then forwards in time we expect to see random genetic drift.

2.6.3 Other results

The SΛFV framework was introduced relatively recently [Etheridge, 2008], but there are already a number of papers concerned with different aspects of the model. For example, Etheridge and Véber [2012] introduced recombination, that is, the possibility that different sections of the genome might be inherited from different parents. Berestycki *et al.* [2009] discuss an individual-based model that can be thought of as a prelimiting version of the SΛFV; later Véber and Wakolbinger [2013] would provide a lookdown construction of a particle system whose empirical distribution has the same law as the SΛFV; and more recently Etheridge and Kurtz [2014] provided an even more general construction of a particle system that has the SΛFV process as its high density limit.

There has been also interest in more applied aspects of the model, namely simulation. In Barton *et al.* [2013c] and Kelleher *et al.* [2014] an algorithm for efficiently simulating the coalescent process is introduced, and the software developed is publicly available (`https://pypi.python.org/pypi/discsim`). Work is under way to develop tools to perform statistical inference under the model, which would allow for genetic data to be incorporated into the model in order to better understand the evolutionary forces. In particular Barton *et al.* [2013a] outline approaches for inference on the neighbourhood size and the dispersal rate of a population based on allele frequencies.

CHAPTER 3

Natural Selection

The cornerstone of Darwin's evolutionary theory (Darwin [1859]) is the idea of natural selection, which can be summarized as stating that "heritable traits that increase reproductive success will become more common in a population" (Etheridge [2011]). This simple idea was revolutionary at the time and has been shaping the life sciences for well over a century. Although the idea slowly became accepted by the scientific community towards the end of the 19th century, it was not till the modern evolutionary synthesis led by R.A Fisher, S. Wright and J.B.S. Haldane that it was successfully reconciled with Mendelian genetics, which provided a consistent mechanism for the traits to be passed on to future generations as well as a source of variation upon which selection can act.

How to measure the impact that selection has on the genetic diversity of a population remains elusive though, and even among population geneticists we find very different points of view, from those who consider that selection plays a minor role compared to the random genetic drift (the neutral theory), to those who believe that it is behind even the smallest changes in phenotype.

3.1 Selection in the non spatial models

3.1.1 The Wright-Fisher model with selection

We will begin our study by incorporating natural selection into the Wright-Fisher model. Consider a panmictic haploid population of size N. Suppose that we have two types, a and A, and let s be the selection coefficient. If the current generation has k individuals of type A, we obtain the next generation by sampling N times from the following distribution:

$$\mathbb{P}(\text{ a is sampled}) = \frac{N-k}{N+sk} \qquad \mathbb{P}(\text{ A is sampled}) = \frac{(1+s)k}{N+sk},$$

which we notice is giving an extra weight of s to the type A individuals. We say that the relative fitness of alleles A and a is $1+s:1$. If s is positive then A is advantageous, while if it is negative the allele A is deleterious.

Before moving forward we will give a definition of fitness. The fitness of an individual is simply the number of offspring it has in the next generation. Similarly, the fitness of a gene is the number of copies it leaves after one generation, and the fitness of an allele is the average fitness of genes of that type. We see that $1+s$ is the mean fitness of the type A genes.

It is important to remember that fitness (and thus the selection coefficient) captures in a single parameter many factors that affect reproductive success. For example, if we have a population where the individuals need to mature before reproducing (we say they have a juvenile state), the fitness reflects the survival rate throughout this phase, as well as the number of offspring and the viability of said offspring. So we should be careful not to confuse fitness with any one isolated factor.

As we did in the neutral case, we can take a diffusive limit of p_t, the proportion of the A allele in the population at time t. In order to do so we will need to let the parameter s also go to zero. If we allow $s^N := s/N$ and let $N \to \infty$ we obtain a diffusion with parameters $\mu(p) = sp(1-p)$ and $\sigma^2(p) = p(1-p)$ which solves the stochastic differential equation:

$$dp_t = sp_t(1-p_t)dt + \sqrt{p_t(1-p_t)}dW_t.$$

This is called the weak selection limit of the Wright-Fisher model with selection, and the derivation is very similar to the one we did for the neutral case in §1.3.1. A rigorous proof can be found in section 5.2 of Etheridge [2011]. A natural question to ask is how different is the model with selection compared to the neutral case. Using the standard tools of diffusion theory it is not hard to obtain that, if $s > 0$ (i.e. the type A is advantageous), the probability of fixation of the A allele given that its initial proportion is p_0 satisfies:

$$\mathbb{P}_{p_0}(\text{ allele A becomes fixed }) = \frac{1-\exp(-2sp_0)}{1-\exp(-2s)}.$$

As expected, if $s > 0$ we have that the fixation probability of the advantageous allele is strictly larger than under the neutral model. It is clear that the introduction of selection has altered the dynamics of our population forwards in time, and we would like to now understand how the process is affected when we look backwards in time.

3.1.2 The Moran model with selection and the Ancestral Selection Graph

In order to understand the genealogical process it will be convenient to think about selection in the Moran model, instead of the Wright-Fisher model. Recall that the coalescent had a very natural interpretation for the Moran model, where the genealogy of a finite sample was given precisely by the Kingman coalescent. Similarly, we will see that the genealogies will have a very simple interpretation when we incorporate selection.

Once again suppose that the allele A has a selective advantage $s > 0$. We consider a population of size N, and as in the neutral case, at a rate $\binom{N}{2}$ we pick a pair of individuals at

random, one of which dies and is replaced by an offspring of the other one. But in addition to these neutral events, we have selective events happening at a rate $\binom{N}{2}s$, in which we again sample a pair of individuals, and if they have different types we replace the one of type a by a type A individual, otherwise we pick one of the individuals at random and replace it by an offspring of the other one. If we take a weak selection limit, with $s^N := s/N$ we also arrive at the Wright-Fisher diffusion with selection (see section 5.4 of Etheridge [2011]), so we have that for large populations it makes little difference if we consider either model.

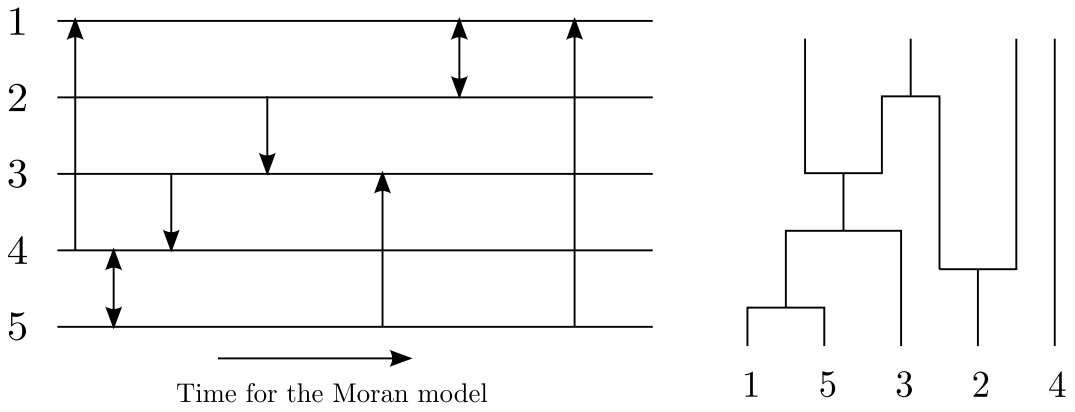

Figure 3.1: The Moran model with selection and the Ancestral Selection Graph. The picture on the left represents the Moran model, where now we find two headed arrows that indicate a selective event. When we trace the genealogy backwards, these events resolve into branching of the lineages, and instead of a coalescent, we arrive at the Ancestral Selection Graph.

We could have chosen a different way of incorporating selection into the Moran model, but this mechanism has two advantages: it allows for a useful graphical representation, and it also prefigures the construction we will use later on in our spatial model. In Figure 3.1 we illustrate how the Moran model with selection works. As in the neutral case, we have different levels that represent the individuals in our population, with time running rightwards. The arrows between levels indicate that the pair at either end of the arrow was involved in an event, with the individual on the level at the tip of the arrow is replaced by an offspring of the other one. But we also have now double headed arrows, which represent selective events. In order to know which individual is killed, we need to look at the types of both of them.

The dual can then be constructed from this graphical representation in the manner illustrated on the left side of Figure 3.1. If we have a sample from the population, we follow it backwards in time up to the first time we see an arrow. One ended arrows represent coalescence, where the lineage located at the end of the arrow jumps to the position at the origin of the arrow, which can be thought of as the parent of the event. Whenever we find two ended arrows, since we don't know the types of the lineages, we must now follow the two 'potential' ancestral lineages. This process was introduced by Krone and Neuhauser [1997], and it contains all the potential ancestors of a sample from a population evolving according to a Moran model with selection (or according to a Wright-Fisher diffusion with selection).

3.2 Introducing selection to the SΛFV

We will now introduce selection into our model. Just as we mentioned in §3.1, it is possible to think of different ways to account for selection, and once we take space into account more factors, such as local frequencies or a random environment, can be considered. Here we will mimic the classical approach of the Wright-Fisher model. In doing so, we will restrict our attention to a type space K with two types, one of which is advantageous. Although as with the classical models, this is a simplistic scenario, it already provides a qualitatively different behaviour from the neutral case.

We suppose that the population is evolving in a geographical space $\mathbb{R}^d$. It will be convenient to index time by the whole of $\mathbb{R}$. At each time t, the population will be represented by a random function $\{w_t(x),\, x \in \mathbb{R}^d\}$ defined, up to a Lebesgue null set of $\mathbb{R}^d$ by

$$w_t(x) := \text{ proportion of type } A \text{ at spatial position } x \text{ at time } t.$$

A construction of an appropriate state space for $x \mapsto w_t(x)$ can be found in Véber and Wakolbinger [2013]. Using the identification

$$\int_{\mathbb{R}^d \times \{A,a\}} f(x,\kappa) M(dx, d\kappa) = \int_{\mathbb{R}^d} \big\{ w(x) f(x,A) + (1 - w(x)) f(x,a) \big\}\, dx,$$

this state space is in one-to-one correspondence with the space $\mathcal{M}_\lambda$ of measures on $\mathbb{R}^d \times \{a, A\}$ with 'spatial marginal' Lebesgue measure, which we endow with the topology of vague convergence. By a slight abuse of notation, we also denote the state space of the process $(w_t)_{t\in\mathbb{R}}$ by $\mathcal{M}_\lambda$. Just as in the previous chapter, the process will be updated through reproductive events.

To introduce selection to the process, instead of sampling the parent uniformly at random from inside the event, our choice will be weighted according to the types. So, if just before the time of the event the proportion of the advantageous type A is $\bar{w}$, then the probability that the parent will be of type A is given by:

$$\frac{(1+s)\bar{w}}{1+s\bar{w}}.$$

So far, there is no requirement on the magnitude of s, so we are able to model both strong and weak selection following this idea. Nevertheless, in order to relate to the classical approach, we will study weak selection, and thus consider s to be small. For $s \ll 1$ we can approximate the probability of choosing a type A parent by:

$$\frac{(1+s)\bar{w}}{1+s\bar{w}} = (1-s)\bar{w} + s\bar{w}^2 + 2s\bar{w}(1-\bar{w}) + \mathcal{O}(s^2),$$

while the probability of choosing a type a individual as the parent can be approximated by

$$\frac{1-\bar{w}}{1+s\bar{w}} = (1-\bar{w}) - s\bar{w}(1-\bar{w}) + \mathcal{O}(s^2) = (1-s)(1-\bar{w}) + s(1-\bar{w})^2 + \mathcal{O}(s^2).$$

Thus, up to an error of order $\mathcal{O}(s^2)$, we can decompose our events into two kinds. Neutral events, with an intensity $(1-s)dt\otimes dx\otimes\xi(dr, du)$, behave just as the events in the neutral SΛFV. Meanwhile selective events have an intensity $sdt \otimes dx \otimes \xi(dr, du)$, and instead of sampling one parent uniformly at random, we pick two potential parents; if either of them is of type A then the offspring will also be type A, only if both of them are type a will the offspring inherit this type. This mirrors the selective mechanism in the Moran model, and will similarly result in a genealogical process where in addition to coalescence, we see branching. The SΛFV with selection will therefore look like the neutral case except that a proportion s of the events will be selective and more likely to pick an individual of type A as the parent.

Also, to simplify calculations going forward we will take a fixed $u \in (0,1)$, and thus we have that $\xi(dr, du)$ can be written as $\mu(dr) \times \delta_u$. We shall sometimes refer to the parameter u as the 'impact' of the event. In biological terms, it can be loosely thought of as inversely proportional to the local population density or, equivalently, to the 'neighbourhood size'. When fixing u (or more generally, by keeping it macroscopic) we are choosing to consider populations with small neighbourhood size. The results we present in this chapter generalize to the case where u is sampled at random from $(0,1)$ but for clarity we will keep it fixed.

Definition 3.1 (SΛFV with (fecundity) selection (SΛFVS)) *Fix $\mathcal{R} \in (0,\infty)$ and $u \in (0,1)$ and let μ be a finite measure on $(0,\mathcal{R}]$. Further, let Π be a Poisson point process on $\mathbb{R}^d \times \mathbb{R} \times (0,\infty)$ with intensity measure*

$$dx \otimes dt \otimes \mu(dr). \tag{3.1}$$

The spatial Λ-Fleming-Viot process with selection *(SLFVS) driven by (3.1) is the $\mathcal{M}_\lambda$-valued process $(w_t)_{t\in\mathbb{R}}$ with dynamics given as follows.*

If $(x,t,r) \in \Pi$, a reproduction event occurs at time t, within the closed ball of radius r centred on x, which we denote by $\mathcal{B}_r(x)$. With probability $1-\boldsymbol{s}$ the event (x,t,r) is neutral, in which case:

1. *Choose a parental location z uniformly at random within $\mathcal{B}_r(x)$, and a parental type, κ, according to $w_{t-}(z)$, that is $\kappa = A$ with probability $w_{t-}(z)$ and $\kappa = a$ with probability $1 - w_{t-}(z)$.*

2. *For every $y \in \mathcal{B}_r(x)$, set $w_t(y) = (1-u)w_{t-}(y) + u\mathbf{1}_{\{\kappa=A\}}$.*

With the complementary probability $\boldsymbol{s}$, (x,t,r) corresponds to a selective event within $\mathcal{B}_r(x)$ at time t, in which case:

1. *Choose two 'potential' parental locations z, z' independently and uniformly at random within $\mathcal{B}_r(x)$, and at each of these sites 'potential' parental types κ, κ', according to $w_{t-}(z), w_{t-}(z')$ respectively.*

2. *For every* $y \in \mathcal{B}_r(x)$ *set* $w_t(y) = (1-u)w_{t-}(y) + u\mathbf{1}_{\{\kappa=A \text{ or } \kappa'=A\}}$. *Declare the parental location to be* z *if* $\kappa = \kappa' = a$ *or* $\kappa = \kappa' = A$ *and to be* z *(resp.* z'*) if* $\kappa = A, \kappa' = a$ *(resp.* $\kappa = a, \kappa' = A$*).*

In fact this is a very special case of the SΛFVS introduced in Etheridge *et al.* [2014] and even more special than those constructed in Etheridge and Kurtz [2014] but, in the interests of tractability, we have refrained from working in a more general context.

3.3 The genealogical process

We are especially concerned with the dual process of the SΛFVS, a system of branching and coalescing paths that encodes all the *potential ancestors* of individuals in a sample from the population. When there is no selection, the dual contains only coalescing random walks and each such walk corresponds to the ancestral lineage ℓ of some individual at (x,t); meaning that ℓ traces out the locations in space-time occupied by the ancestors of that individual. However, if selection is present then, looking backwards in time, at a selective event we cannot determine the genetic types of the 'potential' parents without looking further into the past, just as in the ASG of §3.1.2. To avoid intractable non-Markovian dynamics, in this case we define a dual in which each (branching, coalescing) random walk corresponds to a *potential* ancestral lineage. It traces all the locations in space-time which *could* have contained ancestors of individuals in a sample S.

The dynamics of the dual are driven by the same Poisson point process of events, Π, that drove the forwards in time process. The distribution of this Poisson point process is invariant under time reversal and so we shall abuse notation by reversing the direction of time when discussing the dual.

We suppose that at time 0 (which we think of as 'the present'), we sample k individuals from locations $x_1, \ldots, x_k$ and we write $\xi_s^1, \ldots, \xi_s^{N_s}$ for the locations of the N_s *potential ancestors* that make up our dual at time s before the present.

Definition 3.2 (Branching and coalescing dual) *The branching and coalescing dual process $(\Xi_t)_{t\geqslant 0}$ is the $\bigcup_{n\geqslant 1}(\mathbb{R}^d)^n$-valued Markov process with dynamics defined as follows: At each event $(x,t,r)\in\Pi$, with probability $1-\boldsymbol{s}$,*

1. *For each $\xi^i_{t-}\in\mathcal{B}_r(x)$, independently mark the corresponding potential ancestral lineage with probability u;*

2. *if at least one lineage is marked, all marked lineages disappear and are replaced by a single potential ancestor, whose location is drawn uniformly at random from within $\mathcal{B}_r(x)$.*

With the complementary probability $\boldsymbol{s}$, the event is selective:

1. *For each $\xi^i_{t-}\in\mathcal{B}_r(x)$, independently mark the corresponding potential ancestral lineage with probability u;*

2. *if at least one lineage is marked, all marked lineages disappear and are replaced by two potential ancestors, whose locations are drawn independently and uniformly from within $\mathcal{B}_r(x)$.*

In both cases, if no particles are marked, then nothing happens.

Since we only consider finitely many initial individuals in the sample, the jump rate in this process is finite and so this description gives rise to a well-defined process.

This dual process is the analogue for the SΛFVS of the Ancestral Selection Graph (ASG). Duality of this type can be expressed in several different ways, but perhaps the simplest is so called *weak duality*, which is established in the case of the ASG by checking that the ASG is the moment dual to the Wright-Fisher diffusion with selection.

To establish the analogous weak duality for the SΛFVS, we would need to be able to identify $\mathbb{E}[\prod_{i=1}^n w_t(x_i)]$ for any choice of points $x_1,\dots,x_n\in\mathbb{R}^d$. The difficulty is that, just as in the neutral case, the SLFVS $w_t(x)$ is only defined at Lebesgue almost every point x and so we have to be satisfied with a 'weak' moment duality, very similar to the duality relationship for the neutral case that we showed in Theorem 2.6.

Proposition 3.3 *[Etheridge et al. [2014]] The spatial Λ-Fleming-Viot process with selection is dual to the process $(\Xi_t)_{t\geqslant 0}$ in the sense that for every $k \in \mathbb{N}$ and $\psi \in C((\mathbb{R}^d)^k) \cap L^1((\mathbb{R}^d)^k)$, we have*

$$\mathbb{E}_{w_0}\left[\int_{(\mathbb{R}^d)^k} \psi(x_1,\dots,x_k)\left\{\prod_{j=1}^{k} w_t(x_j)\right\} dx_1\dots dx_k\right]$$
$$= \int_{(\mathbb{R}^d)^k} \psi(x_1,\dots,x_k)\mathbb{E}_{\{x_1,\dots,x_k\}}\left[\prod_{j=1}^{N_t} w_0\left(\xi_t^j\right)\right] dx_1\dots dx_k. \tag{3.2}$$

In fact, a stronger form of this duality holds, in which the forwards in time process of allele frequencies and the process of potential ancestors of a sample are realised on the same probability space though a lookdown construction; see Etheridge and Kurtz [2014].

As we saw in section §2.5, if we had no selective events, a single lineage ξ would perform a random walk on $\mathbb{R}^d$ driven by the neutral events. That is, it would stay in its position until one of the events affected it, in which case it would jump to a new position uniformly distributed in the ball associated to the event. If we start with a sample of various lineages, each of them performs a random walk, almost independently of each other. But once two lineages are close enough, there is a positive probability that an event will affect both of them and thus they would coalesce. So we have a system of coalescing random walks.

Once we add to the mix the selective events we find that when we look at a single lineage it performs the same random walk as above until it is hit and affected by a selection event, in which case it branches into two lineages, each one in at independent uniformly distributed position inside the ball. If we start with a sample of more than one lineage then we have that each of the lineages performs a random walk till two of them come close enough to coalesce, or one of them is affected by a selection event leading to branching, or both. So we have a system of branching and coalescing random walks.

Just as in section §2.5, we have that the intensity of the jump process is:

$$m(dz) = \left(\int_{(0,\infty)} u\frac{L_r(z)}{V_r}\mu(dr)\right) dz,$$

where we write V_r for the volume of a ball of radius r, and $L_r(z)$ for the volume of the intersection $B(0,r) \cap B(z,r)$ (that is of two balls whose centres are at a distance $|z|$). We can also calculate the variance of the jump distribution to obtain

$$\sigma^2 := \int \frac{|z|^2}{d} m(dz) < \infty. \tag{3.3}$$

The branching rate of an individual lineage in isolation will simply be given by $us \int V_r dr$, the rate at which it encounters selective events times the probability it is actually hit.

3.4 First results

In a similar fashion to the results of Barton *et al.* [2010a] for the neutral case, we look for a rescaling that can provide us with a more tractable limiting process. It seems natural to consider the diffusive rescaling of the genealogical process, since the jumps performed by the lineages have all the same bounded distribution which suggest Brownian motion as a limit for the displacement of the lineages. This is precisely what we will attempt in the rest of this chapter.

Remark 3.4 *Notice that we are rescaling time and space, but we have kept the impact parameter fixed. In [Etheridge et al., 2012] the case where the impact parameters vanishes as $n \to \infty$ is considered. The difference between these two cases, from a biological point of view, has to do with the idea of neighbourhood size. By fixing u and v we are able to maintain the signature of a small neighbourhood size in our population, since every event replaces a positive fraction of the population inside the ball.*

Let us describe the scaling more precisely. Suppose that μ is a finite measure on $(0, \mathcal{R}]$. For each $n \in \mathbb{N}$ define the measure μ^n by $\mu^n(A) = \mu(n^{1/2}A)$, for all Borel subsets A of $(0, \mathcal{R}]$. At the nth stage of the rescaling, our rescaled dual is driven by the Poisson point process Π^n on $\mathbb{R}^d \times \mathbb{R} \times [0, \infty)$ with intensity

$$n^{1/2}\, dx \otimes n\, dt \otimes \mu^n(dr). \tag{3.4}$$

Each event of Π^n, independently, is neutral with probability $1-\mathbf{s}_n$ and selective with probability $\mathbf{s}_n$, where the sequence $(\mathbf{s}_n)$ is defined follows:

$$\mathbf{s}_n = \frac{\mathbf{s}}{n}.$$

The behaviour of the limiting process will be dictated by the dimension of the space where the population evolves, and we will show that branching vanishes in the one and two dimensional cases, a direct result of the (neighbourhood) recurrence of Brownian motion. Before considering the different cases we introduce a tool that will facilitate the calculations in this and the next chapter.

3.4.1 Embedding

The rules that govern the genealogical process are simple, with both the jumps and branching determined by the Poisson point processes, but since the random walks performed by the lineages are correlated the analysis of the process $(\Xi_t)_{t\geqslant 0}$ soon becomes rather involved. Nevertheless, it is possible to construct a generalized Skorokhod embedding for it, which allows us to readily use results from the Brownian motion literature to study the genealogical process. We will follow the construction given by Berestycki *et al.* [2013] which we outline here and will use throughout the rest of this chapter as well as in Chapter 4.

Consider the behaviour of a single rescaled lineage $(\xi^n_t)_{t\geqslant 0}$ and, for the moment, ignore the branching (which can be done simply by always picking only one of the two possible branches at random). It behaves as a jump process: it remains at a given location up to the time it is hit by an event, in which case it will jump to the location of the parent. The expected rate at which the lineage jumps is given by:

$$\rho_n := nu\mathbb{E}[V_r]$$

where V_r is the volume of a ball of radius r. Since the parents are picked uniformly at random from inside the ball, it follows that the jumps are rotationally invariant. And we notice that

the length of the jumps is bounded by $2\mathcal{R}_n := 2\mathcal{R}/\sqrt{n}$. For most of our calculations we will be interested in $\eta^n = \xi^{n,1} - \xi^{n,2}$, the difference between two lineages. Since two lineages evolve independently as long as they are further apart than $2\mathcal{R}_n$, it follows that, as long as $|\eta^n| > 2\mathcal{R}_n$, η^n will behave as a single lineage but sped up by a factor of 2.

Now, take a d-dimensional standard Brownian motion $\mathcal{W}$ started from x. Let $(R_j^n)_{j\geqslant 1}$ be a sequence of i.i.d. random variables distributed according to the length of the jumps of η^n, also independent of $\mathcal{W}$. We proceed to recursively define a sequence of random times by:

1. $s_0^n := 0$,

2. for every $j \geqslant 1$, s_j^n, is the first time greater than s_{j-1}^n at which the Brownian motion $\mathcal{W}$ exits the ball $\mathcal{B}_{R_j^n}(\mathcal{W}(s_{j-1}^n))$.

In words, we have a sequence of "jump lengths", and every time the Brownian motion has moved a distance equivalent to this length we record the time, and wait again for it to move a distance equivalent to the next jump, record the time, and continue in the same fashion. Since both Brownian motion and the jump process of a lineage share rotational symmetry, we are in fact mimicking the lineage by considering the Brownian motion only at certain times.

Indeed, conditional on the jump size being γ, the location of η^n after the jump is uniformly located over $\delta\mathcal{B}_\gamma(\eta_{t-}^n)$. Equivalently, conditioned on $R_j^n = \gamma$, the location of $\mathcal{W}(s_j^n)$ is uniformly distributed over $\mathcal{B}_\gamma(\mathcal{W}(s_{j-1}^n))$. Since the jumps of the lineage arrive according to a Poisson process of intensity ρ_n, if we take a Poisson point process $(j(n,t))_{t\geqslant 0}$ with intensity ρ_n, we have that the laws of η_t^n and $\mathcal{W}(s_{j(n,t)}^n)$ are identical.

Proposition 3.5 *Consider the process η^n started in $\mathcal{B}^c_{2\mathcal{R}_n}(0)$ and let $(s_j^n)_{j\in\mathbb{N}}$ be the random times defined above. If we take a Poisson process $j(n,t)$ with intensity ρ_n we have that $\forall t > 0$,*

$$\eta^n(t) \overset{d}{=} \mathcal{W}(s_{j(n,t)}^n)$$

*We shall refer to this construction as the **embedding**.*

Notice that $j(n,t)$ is simply counting how many jumps the process has done up to time t, while $s_j^n - s_{j-1}^n$ reflects the time elapsed between the consecutive jumps. The following Lemma follows by applying the tower property of conditional expectation to $s^n_{j(n,t)}$ while conditioning on $j(n,t)$.

Lemma 3.6 *There exists a constant $C > 0$ such that:*

$$\frac{\mathbb{E}[s^n_{j(n,t)}]}{t} \to C \text{ as } n \to \infty \ .$$

In addition, Chebyshev's inequality allows us to state the following useful estimate.

Lemma 3.7 *For every fixed $t > 0$:*

$$\mathbb{P}\left(|s^n_{j(n,t)} - Ct| > n^{-1/4}\right) \leqslant n^{-1/4}.$$

These lemmas give us some control on the clock of the embedding, and allow us to apply some well know estimates for the excursions of Brownian motion to our genealogical process.

Remark 3.8 *The embedding holds for an individual lineage, but fails whenever we follow two or more particles that come closer than $2\mathcal{R}_n$, since they no longer evolve independently. Nevertheless, we will only require it to calculate the duration of excursions of lineages away from each other, during which they do move independently.*

3.5 Dimensions 3 and higher

We will now consider the limiting behaviour of $(\Xi^n_t)_{t\geqslant 0}$ as $n \to \infty$, and we begin with the case $d \geqslant 3$. Our argument has two steps. First we show that after a selective event there is a positive probability $\rho_d \in (0,1)$ that the two branches never coalesce back together; we then show that the process $(\Xi^n_t)_{t\geqslant 0}$ converges to a Branching Brownian motion (BBM).

3.5.1 Escape probability

After a selective event it is possible that the two resulting potential lineages quickly coalesce back together. Indeed, since they are at a distance smaller than $2\mathcal{R}_n$, there is a positive probability that the next event to hit either of them actually hits them both in which case they would coalesce. It of course could happen after a few jumps, but in any case, if the time it takes them to coalesce is of $o(1)$ then the branch would not appear in the limit.

In order to see that branching occurs at a macroscopic scale we will find a positive lower bound for the probability that the two branches never coalesce. If two lineages never coalesce we will still be able to distinguish between them after taking the limit.

To show that in every selective event the two lineages have a positive probability of never coalescing we will make use of the embedding introduced in §3.4.1. This construction only holds as long as the distance between the lineages, $\eta^n = \xi^{n,1} - \xi^{n,2}$, is larger than $2\mathcal{R}_n$, so first we will estimate the probability that two lineages arising from a selective event move apart. Notice that this probability is not affected by the parameter n, since we are rescaling the PPP as well as the location of the lineages.

Lemma 3.9 *Consider $\eta_t^n := \xi_t^{n,1} - \xi_t^{n,2}$, the difference of two lineages that arise immediately after a selective event. Let τ^{out} be the first time $|\eta_t^n| > 3\mathcal{R}_n$, and similarly, τ^c the first time $|\eta_t^n| = 0$. We have that for all $d \geqslant 3$,*

$$p^{out} := \mathbb{P}[\tau^{out} < \tau^c] > 0.$$

Proof: For all $d \in \mathbb{N}$ the probability that two lineages sampled uniformly from inside the ball $\mathcal{B}_{\mathcal{R}_n}(0)$ are at a distance at least $\mathcal{R}_n$ apart is strictly positive. If two lineages are at a distance $\mathcal{R}_n$, the probability that the next event pushes them further away by $\mathcal{R}_n/2$ is also strictly positive. So the probability that the distance between them surpasses $3\mathcal{R}_n$ after the first 4 events that affect either of them is strictly positive as well, which proves the lemma.

■

Once the lineages are apart, we can use the embedding constructed in §3.4.1 in order to calculate the probability that the two lineages never coalesce.

Lemma 3.10 *Take two lineages started at a distance $|\eta_0| = 3\mathcal{R}$, then, for all $d \geqslant 3$, the probability they never coalesce is bounded from below by*

$$\mathbb{P}_{3\mathcal{R}}[\tau^c = \infty] \geqslant \left(1 - \left(\frac{2}{3}\right)^{d-2}\right).$$

PROOF: From the embedding construction, it is immediate that it suffices to consider the probability that a Brownian motion started from $|x| = 3\mathcal{R}$ visits the ball $\mathcal{B}_{2\mathcal{R}}$. This probability is precisely $1 - \left(\frac{2}{3}\right)^{d-2}$, and the result follows. ■

Proposition 3.11 *The probability ρ_d that a pair of lineages which have branched never coalesces is bounded away from 0 for all d.*

PROOF: From the previous two propositions it follows easily that

$$\rho_d \geqslant p^{out} \times \left(1 - \left(\frac{2}{3}\right)^{d-2}\right) > 0.$$

■

As mentioned earlier, these estimates are invariant under the rescaling we have chosen, and thus they hold uniformly in n.

3.5.2 Estimates

We proceed to obtain several estimates on the behaviour of two lineages. The reasoning behind these estimates will become apparent in the next section. So far we considered two lineages started close together. We can imagine that if lineages start far enough apart, they will be very unlikely to coalesce. The following Lemma provides a more rigorous way of stating this idea.

Lemma 3.12 *Take $\xi^{n,1}$, $\xi^{n,2}$ two lineages in the genealogical process such that $|\eta_0^n| =: y > n^{-1/3}$. Then, as $n \to \infty$,*

$$\mathbb{P}_y[\tau^0 < \infty] = o(1).$$

PROOF: Since the lineages are further apart than $2\mathcal{R}_n$, the embedding allows us to bound this probability

$$\mathbb{P}_y[\tau^0 < \infty] \leqslant \mathbb{P}_y[\hat{\tau}_{2\mathcal{R}_n} < \infty],$$

where $\hat{\tau}_{2\mathcal{R}_n}$ is the hitting time of $\mathcal{B}_{2\mathcal{R}_n}$ of a Brownian motion. Since we are considering dimensions three and higher,

$$\mathbb{P}_y[\hat{\tau}_{2\mathcal{R}_n} < \infty] \leqslant \frac{2n^{-1/2}}{n^{-1/3}} = \mathcal{O}(n^{-1/6}),$$

where the last estimate is the probability that a Brownian motion started at a point $n^{-1/3}$ away from 0 comes closer than $n^{-1/2}$ from the origin. This is a well known result for transient Brownian motion, see for example Durrett [1996] . ■

Now consider two lineages born at the same selective branching event. Since they move independently whenever they are not close, if they do not coalesce, we expect they will soon drift apart.

Lemma 3.13 *Take $\xi^{n,1}$, $\xi^{n,2}$ two lineages in the genealogical process such that $y = |\eta_0^n|$ satisfies $4\mathcal{R}_n < |y| < 6\mathcal{R}_n$, let $\tau_{2\mathcal{R}_n}$ be first time they are closer than $2\mathcal{R}_n$, and let $\tau_{n^{-1/3}}$ be the first time they manage to get to a distance of at least $n^{-1/3}$,*

$$\mathbb{E}_y[\tau_{n^{-1/3}} | \tau_{n^{-1/3}} < \tau_{2\mathcal{R}_n}] = \mathcal{O}(n^{-2/3})$$

PROOF: By conditioning on the event $\{\tau_{n^{-1/3}} < \tau_{2R_n}\}$ and its complement, it is easy to see that:

$$\mathbb{E}_y[\tau_{n^{-1/3}} | \tau_{n^{-1/3}} < \tau_{2\mathcal{R}_n}] \leqslant \frac{\mathbb{E}_y[\tau_{n^{-1/3}}]}{\mathbb{P}_y[\tau_{n^{-1/3}} < \tau_{2\mathcal{R}_n}]}.$$

Once again, the embedding allows us to estimate the expectation and the probability on the right hand side. The probability that appears as the denominator is $\Theta(1)$, since Brownian motion is

transient, and the expectation on the numerator is $\mathcal{O}(n^{-2/3})$, so

$$\frac{\mathbb{E}_y[\tau_{n^{-1/3}}]}{\mathbb{P}_y[\tau_{n^{-1/3}} < \tau_{2\mathcal{R}_n}]} = \mathcal{O}(n^{-2/3})$$

■

From the above Lemma we can conclude that:

$$\mathbb{P}_y[\tau_{n^{-1/3}} > n^{-1/2} | \tau_{n^{-1/3}} < \tau_{2R_n}] = o(1). \tag{3.5}$$

3.5.3 Convergence to BBM

In Proposition 3.11 we showed that the probability ρ_d that the two lineages that arise from a selective event never coalesce with each other is strictly positive. For convenience let us consider one of the lineages in a branch the 'parent'. The next lemma shows that if a lineages does not coalesce with its parent, with high probability, it will not coalesce with any of the other lineages either.

Lemma 3.14 *Take the process Ξ^n started from a single lineage at the origin. At each selective event we deem one of the particles to be the 'parent' branch and the other one to be the 'new' branch. Then, for any fixed time period $[0,T]$, the probability that one of the new branches coalesces with a particle present at the time of its birth that is not its parent tends to 0 as $n \to \infty$.*

Proof: The maximal number of particles in the system up to time T, say M, is stochastically bounded by a Yule process with branching rate sV_r, so in particular we have that $\mathbb{E}[M] < \infty$. The Markov inequality allows us to conclude that $\mathbb{P}[M > m] = \mathcal{O}(m^{-1})$. Let $A(t)$ be the event that a particle coalesces with an older particle that is not its parent before time t.

$$\begin{aligned}\mathbb{P}[A(T)] &= \mathbb{P}[A(T)|M > m]\mathbb{P}[M > m] + \mathbb{P}[A(T)|M < m]\mathbb{P}[M < m] \\ &\leqslant \mathbb{P}[M > m] + \mathbb{P}[A(T)|M < m].\end{aligned}$$

We are interested in obtaining a bound for the second term of the last line. Conditional on $\{M < m\}$, the rate at which new particles arrive is at most $sV_r m$, and thus we have that the time that elapsed between the $(i-2)$-th and the $(i-1)$-th branching events is stochastically bounded from below by an exponential random variable $\mathcal{T}$ with rate $sV_r m$, and so, $\mathbb{E}[\mathcal{T}|M < m] = \mathcal{O}(m^{-1})$. As long as $m \ll n^{1/2}$, we have that with high probability, all the particles are at least $n^{-1/3}$ apart (provided $m = \mathcal{O}(n^{1/8})$,if two particles do not coalesce, the probability that they are further apart than $n^{-1/3}$ after time $\mathcal{T}$ is $\mathcal{O}(n^{-1/6})$). We will use C for the event that some lineage is close to the parent at the ime of the $(i-1)$-th branching event.

$$\begin{aligned}\mathbb{P}[A(T)|M < m] &= \mathbb{P}[A(T)|M < m, C]\mathbb{P}[C|M < m] + \mathbb{P}[A(T)|M < m, C^c]\mathbb{P}[C^c|M < m] \\ &\leqslant \mathbb{P}[C|M < m] + \mathbb{P}[A(T)|M < m, C^c] \\ &\leqslant \mathcal{O}(mn^{-1/6}) + \mathcal{O}(mn^{-1/6}),\end{aligned}$$

holds as long as $m = \mathcal{O}(n^{1/8})$. By choosing $m = n^{1/10}$ we have

$$\begin{aligned}\mathbb{P}[A(T)] &\leqslant \mathbb{P}[M > m] + \mathcal{O}(mn^{-1/6}) \\ &\leqslant \mathcal{O}(n^{-1/10}) + \mathcal{O}(n^{-1/15}),\end{aligned}$$

and the result follows.

■

We have included the qualifier 'present at the time of the birth' because the new particle might coalesce with one of its offspring. Notice however that the Lemma ensures that lineages can only coalesce with their parent particle. If a 'new' branch coalesced with a particle different from its parent, the older one would be violating the premise that it can't coalesce with any particles alive at the time of its birth.

We have thus shown that the probability that a lineage coalesces with a lineage that its not its parent or its offspring tends to 0 as $n \to \infty$, and we also showed earlier that the probability that the two lineages arising from a branch do not coalesce is bounded away from 0. This suggest

a BBM with branching rate ρ_d as the natural candidate for the scaling limit of the process Ξ^n.

In order to finish our argument, we will use the following two definitions.

Definition 3.15 *If a selective event hits a potential ancestor in the genealogical process at time t_0, then we will call the time interval $[t_0, t_0 + t_n]$ the **deciding period** of the branch, where $t_n = n^{-1/2}$.*

Definition 3.16 *Take the difference between the two lineages in a branch $\eta_t^n := \xi^{n,1} - \xi^{n,2}$, we say the branch is **disjoint** if $|\eta_{t_n}| \geqslant n^{-1/3}$.*

Since η^n can only visit the ball $\mathcal{B}_{2\mathcal{R}_n}(0)$ a finite number of times before coalescing, equation (3.5) implies that the probability that a branch that hasn't coalesced by the end of the deciding period is also not disjoint is only $o(1)$. It follows that the probability a branch is disjoint is $\rho_d + o(1)$.

Theorem 3.17 *As $n \to \infty$, the genealogical process $(\Xi_t^n)_{0 \leqslant t \leqslant T}$ converges weakly to a branching Brownian motion with rate $u\boldsymbol{s}V_{\mathcal{R}}\rho_d$.*

Proof: We first observe that if their respective deciding periods do not overlap, the events that two different branches are disjoint are independent. On the other hand, since the process Ξ^n branches at a finite rate, we can control the number of branches M, just as we did in Lemma 3.14. Conditioning on $\{M < m\}$, the time between successive branches is bounded from below by an exponential random variable with parameter $\mathcal{O}(m)$. It follows that, conditional on $\{M < m\}$, the probability of two deciding periods overlapping is bounded by the probability that the minimum of m exponential random variables (with rate $\mathcal{O}(m)$). This probability is $\mathcal{O}(1 - \exp(m^2 n^{-1/2})$. By conveniently choosing m (say $m = n^{1/10}$, as before), we show that both this probability and $\mathbb{P}[M > m]$ vanish, and it follows the arrival of disjoint branches is given by a thinning of Π^n.

If we only consider disjoint branches, by Lemma 3.14 we have that after their deciding period, with probability tending to one, the lineages arising from disjoint branches evolve independently

after the deciding period.

Since, with probability tending to one, disjoint branches arrive at rate $\rho_d + o(1)$ and the resulting lineages evolve independently of each other and the other lineages after the deciding period, we have that the system that follows the lineages arising from disjoint branches converges to BBM. Since none disjoint branches have coalesced with probability $1 - o(1)$, it follows that the process Ξ^n converges to a BBM with branching rate $u\mathbf{s}V_{\mathcal{R}}\rho_d$.

■

3.6 The lower dimensional case

So far we have shown that if the population evolves in 3 or more dimensions, the genealogical process Ξ^n converges to a BBM, which indicates that selection, in the way we introduced it to the model, has a non negligible effect on the genealogy even in the limit. From a biological perspective though, higher dimensions are not that interesting, as most natural populations evolve across a 2, and sometimes 1, dimensional region.

Unfortunately, the approach we have followed so far doesn't give interesting results in low dimensions. As it should be clear from the previous section, branching in the limit is closely related to the transience of Brownian motion in higher dimensions, and we know that it is actually (neighbourhood) recurrent in dimensions 1 and 2. Effectively, this means that the branches that appear through a selective event are cancelled by coalescing much faster than our time scale, and so do not appear in the limit. We won't dwell too much in the calculations involved, since they have all already been done by Barton *et al.* [2010a] and Berestycki *et al.* [2013], but we will provide with an overview of the results necessary to show that the limiting object in dimensions 1 and 2 present no branching.

The first step is to show that coalescence happens very quickly whenever two rescaled lineages $\xi^{n,1}$ and $\xi^{n,2}$ are close to each other.

Lemma 3.18 *Take $\xi^{n,1}$ and $\xi^{n,2}$ and suppose that $|\eta_0^n| := |\xi_0^{n,1} - \xi_0^{n,2}| \leqslant 4\mathcal{R}^n$. Then, for all*

$t > 0$,

$$\mathbb{P}[\tau^c > t] \to 0 \text{ as } n \to \infty.$$

This lemma is proven in both Barton *et al.* [2010a] and Berestycki *et al.* [2013]. The key idea is to consider the un-rescaled lineages and notice that the process η will take a time of order $\mathcal{O}(1)$ to visit the ball $\mathcal{B}_{2\mathcal{R}}(0)$, where it has a positive probability of hitting 0, which would mean the lineages coalesced. Either way, whether the lineages coalesce or move apart again, the time spent in $\mathcal{B}_{2\mathcal{R}}(0)\backslash\{0\}$ is also of order $\mathcal{O}(1)$. If it exits the ball, it once again will take a time of order $\mathcal{O}(1)$ before visiting it again, and so on. Since every time it visits the ball $\mathcal{B}_{2\mathcal{R}}(0)$ the lineages might coalesce, it is possible to bound the number of excursions away from it by a geometric random variable, and each of the excursions and incursions lasts a time of order $\mathcal{O}(1)$. The total time it takes the un-rescaled lineages to coalesce is also of order $\mathcal{O}(1)$, so going back to the rescaled version gives the desired result.

Proposition 3.19 *Let $\Xi_0^n = \delta_x$ and $\mathcal{W}$ be a standard Brownian motion started from x. In dimensions 1 and 2 we have:*

$$\lim_{n\to\infty} \Xi_t^n \overset{(d)}{=} \mathcal{W}(\sigma^2 t),$$

where σ^2 is the variance of the jump distribution given in (3.3).

This result is clearly true in the absence of selection. We can consider one main potential line by picking one of the two possible lineages at every branching event. This lineage will clearly converge to $\mathcal{W}(\sigma^2 t)$ and lemma 3.18 implies that none of the branches that appear will last long enough to appear in the limit. So, we have argued that for lower dimensions, a lineage in the selective case will asymptotically behave like a neutral one. The final step is to consider a sample of finitely many lineages.

Theorem 3.20 *Let $(\Xi_t^n)_{t\geqslant 0}$ be the diffusively rescaled genealogical process of k particles originally started at distances of $O(\sqrt{n})$. We have that as $n \to \infty$:*

1. *In dimension 1,* $(\Xi_t^n)_{t\geqslant 0}$ *converges to a system of independent Brownian motions with clock speed* σ^2 *which coalesce instantly upon meeting;*

2. *in dimension 2,* $(\Xi_t^n)_{t\geqslant 0}$ *converges to a system of independent Brownian motions with clock speed* σ^2*;*

where the convergence in the sense of the finite dimensional distributions.

This theorem appears as Lemmas 4.1 and 4.2 in Berestycki *et al.* [2013], albeit for the neutral case. The selective case follows from it by virtue of Lemma 3.18. We can choose to follow only one of the potential ancestries of each of the original lineages by choosing one potential parent at random whenever we encounter a selective event. This is effectively a neutral genealogy, and so the convergence holds. Since the process is consistent, adding the paths that we ignored previously doesn't alter the behaviour of the original lineages, and given that every path discarded coalesces with its parental lineage in a negligible time, none of them should appear in the final picture.

In summary, it appears as if selection has a negligible effect on the genealogies when we consider populations in 1 and 2 dimensions. Nevertheless, real populations seem to be profoundly affected by selective forces, and we would like to have a model that reflects this fact. A natural course of action is to consider a different rescaling for the genealogical process, one that might have a non trivial limiting object, by which we mean, one that differs from the neutral case. From a biological perspective, we are strengthening selection, although we are still modelling somewhat weak selection.

CHAPTER 4

Dimension 2

In the previous chapter we showed that selection, in the manner we introduced it to the model, has a vanishing effect in lower dimensions, with the limiting process presenting no branching. Since we are considering a diffusive rescaling, in order for us to distinguish between two potential ancestors that arose through a selective event, it is necessary that they achieve a (non-rescaled) separation of $\mathcal{O}(\sqrt{n})$ before potentially coalescing. While for higher dimensions this event has a probability of $\mathcal{O}(1)$ (a by-product of the transience of the random walk), the probability is of order $1/\log(n)$ and $1/\sqrt{n}$ for $d = 2$ and $d = 1$ respectively. Thus, in $d = 2$ we must have roughly $\log(n)$ branching events before we can expect to find a branch that will persist. This suggests speeding up the branching rate by a factor of $\log(n)$, and this chapter is devoted to showing

that this correction indeed provides us with a non trivial limit, and that this limit is, just as in the higher dimensional case, a BBM.

The flavour of the arguments will be very similar to the previous chapter, although more care is needed to control the branching rate, which will tend to infinity as $n \to \infty$. In Section 4.1 we reintroduce the model and provide some useful definitions for the rest of the chapter. Section 4.2 will provide some estimates that illustrate what we have called 'separation of branches', by which we mean that a typical branch won't live very long and thus will be unlikely to interfere with any successive branches. Section 4.3 introduces the 'skeleton' of the process, and finally Section 4.4 proves the convergence of the genealogical process to a branching Brownian motion using the skeleton construction.

4.1 Definitions

Once again we are concerned with the rescaled version of the genealogical process $(\Xi_t^n)_{t\geqslant 0}$. Recall from the previous chapter that for each n we have a Poisson point process Π^n with intensity $n\,dt \otimes n^{1/2}dx \otimes \mu^n(dr)$ that dictates the arrival of the events, with μ^n a probability measure on $[0, \mathcal{R}_n]$ providing the radius of the event. With probability $1 - s_n$ the event is neutral, and with probability s_n it will be selective. As before, we have a parameter $u \in (0, 1)$ that determines the impact probability of the events. But now we make a different choice for the probability of an event being selective:

$$s_n = \frac{s\log(n)}{n}.$$

We are interested in the behaviour of the process $\eta^n = \xi^{n,1} - \xi^{n,2}$, which describes the separation between two lineages, and will make use of the embedding introduced in §3.4.1. We will be particularly interested in whether the size of $|\eta^n|$ exceeds

$$\gamma_n = \frac{1}{\log^{5/2}(n)}.$$

This choice of γ_n has been done for convenience, there is nothing special about 5/2 apart from simplifying some of our calculations. We use η^n to model the distance between two lineages that arose via a selective event. Earlier we mentioned that our argument relies on the fact that for most selective events, the two branches that arise coalesce back together rather quickly. This will be made more precise with the following definitions.

Definition 4.1 *If a selective event hits a potential ancestor in the genealogical process at time t_0, then we will call the time interval $[t_0, t_0 + t_n]$ the **deciding period** of the branching event, where $t_n = 1/\log^{5/2}(n)$ (which coincidentally is the same as γ_n).*

We want to differentiate between the different behaviours that η^n might show after the deciding period. Let τ_s be the first time η^n exits the ball $\mathcal{B}_{\gamma_n}(0)$, and τ_0 be the time at which the two lineages coalesce. Since our process is time and space homogeneous, without loss of generality we assume that the branching occurs at time 0, and we will use this convention here and in the following section. We will also drop the index n whenever we can do so without risk of confusion.

Definition 4.2 *We say that the branching event that gives place to η^n*

1. *is successful if $\tau_s < \tau_0 \wedge t_n$,*
2. *coalesces if $\tau_0 < \tau_s \wedge t_n$,*
3. *hesitates if $t_n < \tau_0 \wedge \tau_s$.*

Our aim is to be able to predict whether the two lineages will separate enough to appear in the limit at an early stage, before either lineage branches again, and this classification will allow us to do precisely that.

4.2 Separation of branches

We know that the lineages move independently whenever they are further apart than $2\mathcal{R}_n$. If a branch coalesces we know that $|\eta_n|$ must remain smaller than γ_n, and if it managed to get to a distance larger than $2\mathcal{R}_n$ then it must perform what we will call an 'excursion', wandering between $2\mathcal{R}_n$ and γ_n and coming back to $\mathcal{B}_{2\mathcal{R}_n}(0)$. The following lemma will provide, among other things, a bound for the duration of such an excursion. For our calculations it will be convenient to actually be further away than just $2\mathcal{R}_n$ in order to ensure independence and to make use of the embedding.

Remark 4.3 *It is possible to show, in a similar fashion to §3.5.1, that the probability that the two lineages started close together reach a separation $5\mathcal{R}_n$ before coalescing is uniformly bounded away from 0, and that the probability that they coalesce first instead is bounded away from 0 as well, both of them being independent of n. So, there exists $c > 0$ such that for all $y \in \mathcal{B}_{4\mathcal{R}_n}(0)$:*

$$\mathbb{P}_y[\tau_{5\mathcal{R}_n} < \tau_0] \geqslant c,$$

$$\mathbb{P}_y[\tau_0 < \tau_{5\mathcal{R}_n}] \geqslant c.$$

It should also be clear, although it is more cumbersome to show, that for all $y \in \mathcal{B}_{4\mathcal{R}_n}(0)$

$$\mathbb{E}_y[\tau_0 \wedge \tau_{5\mathcal{R}_n}] \leqslant \frac{\hat{c}}{n}$$

for some $\hat{c} < \infty$ (To see this, note that $\mathbb{E}_y[\tau_0 \wedge \tau_{5\mathcal{R}_n}] = \mathbb{E}_y[\tau_0 | \tau_0 < \tau_{5\mathcal{R}_n}]\mathbb{P}_y[\tau_0 < \tau_{5\mathcal{R}_n}] + \mathbb{E}_y[\tau_{5\mathcal{R}_n} | \tau_0 > \tau_{5\mathcal{R}_n}]\mathbb{P}_y[\tau_{5\mathcal{R}_n} < \tau_0]$, and notice that the first expectation is the expected time to coalesce conditional on both lineages remaining close, while the second one is the expected time to exit the ball conditioned on not coalescing, both of which must be bounded by $\hat{c}/n$ for some constant $\hat{c}$). We can therefore describe the behaviour of η^n up to time $\tau_0 \wedge \tau_s$ in the following terms. After branching, the branch will coalesce or separate to a distance $5\mathcal{R}_n$ in a short time, and then, starting from $|\eta^n| \in [5R_n, 7R_n]$ we will either re-enter the ball $\mathcal{B}_{4\mathcal{R}_n}(0)$ or exit $\mathcal{B}_{\gamma_n}(0)$. If it is the second one then we are done, otherwise, once inside $\mathcal{B}_{4\mathcal{R}_n}(0)$ we will either coalesce

or separate further than $5R_n$ again, starting a new excursion. We can apply the Strong Markov property to actually de-construct the process in the manner explained. Since every time we enter $\mathcal{B}_{4\mathcal{R}_n}(0)$ we have a probability at least c of coalescing, it follows we can only perform a geometric number of excursions before the lineages coalesce back together.

Lemma 4.4 *Let τ_{ex} be the hitting time of the ball $\mathcal{B}_{4\mathcal{R}_n}(0)$, then, for all $x : |x| \in [5\mathcal{R}_n, 7\mathcal{R}_n]$ we have:*

$$\mathbb{E}_x[\tau_{ex}|\tau_{ex} < \tau_s] = \mathcal{O}\left(\frac{1}{\log^5(n)}\right), \tag{4.1}$$

$$\mathbb{E}_x[\tau_s|\tau_s < \tau_{ex}] = \mathcal{O}\left(\frac{1}{\log^4(n)}\right), \tag{4.2}$$

$$\mathbb{P}_x[\tau_s < \tau_{ex}] = \Theta\left(\frac{1}{\log(n)}\right). \tag{4.3}$$

Proof: We will begin by considering (4.1). Notice that:

$$\mathbb{E}_x[\tau_{ex} \wedge \tau_s] = \mathbb{P}_x[\tau_{ex} < \tau_s]\mathbb{E}_x[\tau_{ex}|\tau_{ex} < \tau_s] + \mathbb{P}_x[\tau_{ex} > \tau_s]\mathbb{E}_x[\tau_s|\tau_{ex} > \tau_s].$$

Since $\mathbb{E}_x[\tau_{ex} \wedge \tau_s] < \infty$, and $\mathbb{P}_x[\tau_{ex} < \tau_s]$, $\mathbb{P}_x[\tau_{ex} > \tau_s]$ are positive, we can rearrange to obtain:

$$\mathbb{E}_x[\tau_{ex}|\tau_{ex} < \tau_s] \leqslant \frac{\mathbb{E}_x[\tau_s]}{\mathbb{P}_x[\tau_{ex} < \tau_s]}.$$

We know that when we have two lineages at a distance larger than $2\mathcal{R}_n$ they move independently of each other, since the radius of the events that drive their dynamics is bounded by $\mathcal{R}_n$. So, as long as $|\eta_t| > 2\mathcal{R}_n$, it behaves as a single lineage but sped up by a factor of 2, and we can use the embedding of Proposition 3.5 which states that $\eta^n(t) \stackrel{d}{=} \mathcal{W}(s^n_{j(n,t)})$. The embedding construction guarantees that the Brownian motion never gets further away than $2R_n$ from the jump process. We will use $\tilde{\tau}$ for stopping times of the Brownian motion and $\tilde{\mathbb{P}}$ for its law. We have that:

$$\mathbb{E}_x[\tau_s] \leqslant \tilde{\mathbb{E}}_x[\tilde{\tau}_{\gamma_n + 2\mathcal{R}_n}], \tag{4.4}$$

$$\mathbb{P}_x[\tau_{ex} < \tau_s] \geqslant \tilde{\mathbb{P}}_x[\tilde{\tau}_{2\mathcal{R}_n} < \tilde{\tau}_{\gamma_n}], \tag{4.5}$$

where $\tilde{\tau}_{\gamma_n+2\mathcal{R}_n}$, $\tilde{\tau}_{\gamma_n+2\mathcal{R}_n}$ and $\tilde{\tau}_{2\mathcal{R}_n}$ are the embedding Brownian motion's hitting times of $\mathcal{B}^c_{\gamma_n+2R_n}(0)$, $\mathcal{B}^c_{\gamma_n}(0)$ and $\mathcal{B}_{2R_n}(0)$ respectively. It is well known that for two dimensional Brownian motion $\tilde{\mathbb{P}}_x[\tilde{\tau}_m < \tilde{\tau}_M]$ can be easily calculated (see for example Durrett [1996] page 98), and is given by the formula $(\log M - \log|x|)/(\log M - \log m)$. It follows that:

$$\begin{aligned}\tilde{\mathbb{P}}_x[\tilde{\tau}_{2\mathcal{R}_n} < \tilde{\tau}_{\gamma_n}] &= \frac{\log(\gamma_n) - \log(x)}{\log(\gamma_n) - \log(2\mathcal{R}_n)} \\ &\geqslant 1 - \frac{\log(7/2)}{1/2\log(n) - 5/2\log(\log(n))} = 1 - \mathcal{O}\left(\frac{1}{\log(n)}\right).\end{aligned}$$

It is well known that for a Brownian motion we have that $\tilde{\mathbb{E}}_0[\tilde{\tau}_a] = a^2$ (see for example Durrett [1996] page 107) thus $\tilde{\mathbb{E}}_x[\tilde{\tau}_{\gamma_n+2\mathcal{R}_n}] \leqslant (\gamma_n + 2\mathcal{R}_n)^2$. We are now able to conclude that

$$\tilde{\mathbb{E}}_x[\tilde{\tau}_{ex}|\tilde{\tau}_{ex} < \tilde{\tau}_s] \leqslant \frac{\tilde{\mathbb{E}}_x[\tilde{\tau}_{\gamma_n+2\mathcal{R}_n}]}{\tilde{\mathbb{P}}_x[\tilde{\tau}_{2\mathcal{R}_n} < \tilde{\tau}_{\gamma_n}]} \leqslant \mathcal{O}\left(\frac{1}{\log^5(n)}\right).$$

To show that (4.3) holds, we just need a lower bound, since we already showed that $\mathbb{P}_x[\tau_s < \tau_{ex}] \leqslant \mathcal{O}(1/\log(n))$. Clearly $\mathbb{P}_x[\tau_s < \tau_{ex}] \geqslant \tilde{\mathbb{P}}_x[\tilde{\tau}_{\gamma_n+2\mathcal{R}_n} < \tilde{\tau}_{4\mathcal{R}_n}]$, and thus we have the desired lower bound. Equipped with it, the proof of (4.2) is entirely analogous to the proof of (4.1). ■

Lemma 4.5 (corollary to Lemma 4.4) *For all $y \in \mathcal{B}_{2\mathcal{R}_n}(0)$, the probability that a branch eventually separates satisfies:*

$$\mathbb{P}_y[\tau_s < \tau_0] = \Theta\left(\frac{1}{\log(n)}\right).$$

Proof: Clearly the probability that it eventually separates is greater than the probability it achieves it on its very first excursion:

$$\mathbb{P}_y[\tau_s < \tau_0] \geqslant \mathbb{P}_y[\tau_s < \tau_{4\mathcal{R}_n}]\mathbb{P}_y[\tau_{5\mathcal{R}_n} < \tau_0] = \Theta\left(\frac{1}{\log(n)}\right).$$

We now need to find an upper bound. We know that the behaviour of the branch can be described as a succession of excursions, and after each one of them we have at least a probability c of coalescing. So we can stochastically bound the number of excursions that η^n performs by a geometric random variable with parameter c. The probability that a given excursion manages

to separate the lineages further than γ_n is bounded above by some $p_n = \mathcal{O}(1/\log(n))$. Let X^e be a geometric with parameter p_n and X^c a geometric random variable with parameter q_c, we can bound the probability we see a successful excursion:

$$\begin{aligned}\mathbb{P}_y[\tau_s < \tau_0] \leqslant \mathbb{P}[X^e < X^c] &= \sum_{n=1}^{\infty} \mathbb{P}[X^e < X^c | X^c = n]\mathcal{P}[X^c = n] \\ &= \sum_{n=1}^{\infty} \left(1 - (1-p_n)^n\right)\left(c(1-c)^{n-1}\right) \\ &= 1 - c\sum_{n=1}^{\infty}(1-p_n)^n(1-c)^{n-1} \\ &= \frac{p_n}{c + p_n - cp_n} = \mathcal{O}(1/\log(n)).\end{aligned}$$

■

Recall from Definition 4.2 that in order for us to consider a branch successful, it must reach a distance γ_n by time t_n after branching. In the previous lemma we showed that the probability we reach this distance before coalescing is $\Theta(1/\log(n))$, now we want to see how likely it is to do so before the deciding period ends. To do so, first we will obtain a bound on the expected time to separate (conditioned on eventually separating).

Lemma 4.6 *For all* $y \in \mathcal{B}_{2\mathcal{R}_n}(0)$

$$\mathbb{E}_y[\tau_s | \tau_s < \tau_0] = \mathcal{O}\left(\frac{1}{\log^4(n)}\right).$$

PROOF: Since the branch eventually separates, its path must consist of a random number $\mathcal{E}$ of excursions that go back to $\mathcal{B}_{4\mathcal{R}_n}(0)$, before a last excursion that escapes. Let τ_1^{in} be the first time that η^n started from y exits $\mathcal{B}_{7\mathcal{R}_n}(0)$, and define recursively $\tau_i^{ex} = \inf\{t > \tau_i^{in} | \eta^n(t) \in \mathcal{B}_{4\mathcal{R}_n}(0)\}$, $\tau_i^{in} = \inf\{t > \tau_{i-1}^{in} | \eta^n(t) \in \mathcal{B}^c_{7\mathcal{R}_n}(0)\}$, which we can think as the inner and outer components of the excursions that η^n performs. We can decompose the time that it takes the lineages to separate into the excursions away from $\mathcal{B}_{4\mathcal{R}_n}(0)$ plus the last, successful, excursion.

$$\mathbb{E}_y[\tau_s|\tau_s < \tau_0] \leqslant \mathbb{E}[\sum_{i=1}^{\mathcal{E}}(\tau_i^{in} + \tau_i^{ex})|\tau_s < \tau_0] + \mathbb{E}[\tau_s|\tau_s < \tau_{ex}] \tag{4.6}$$

$$\leqslant \left(\frac{\hat{c}}{n} + \mathcal{O}\left(\frac{1}{\log^5(n)}\right)\right)\mathbb{E}[\mathcal{E}|\tau_s < \tau_0] + \mathcal{O}\left(\frac{1}{\log^4(n)}\right) = \mathcal{O}\left(\frac{1}{\log^4(n)}\right), \tag{4.7}$$

where we have used Remark 4.3 to bound $\mathbb{E}[\tau_i^{in}]$ by $\hat{c}/n$, and Lemma 4.4 to obtain $\mathbb{E}[\tau_i^{ex}|\tau_s < \tau_0] = \mathcal{O}\left(1/\log^5(n)\right)$ and $\mathbb{E}[\tau_s|\tau_s < \tau_{ex}] = \mathcal{O}\left(\frac{1}{\log^4(n)}\right)$. ■

We can now show that the probability of being successful, which means that the branch separates before the deciding period is over, is of the same order as the probability that it eventually separates. In other words, branches that separate do so before time t_n with high probability.

Lemma 4.7 *The success probability satisfies*

$$\mathbb{P}_y[\tau_s < \tau_0 \wedge t_n] = \Theta\left(\frac{1}{\log(n)}\right)$$

uniformly for all $y \in \mathcal{B}_{2\mathcal{R}_n}(0)\backslash\{0\}$.

Proof: Let S_n be the event that the branching event is successful. Then

$$\mathbb{P}_y[S_n] := \mathbb{P}_y[\tau_s < \tau_0 \wedge t_n] = \mathbb{P}_y[\tau_s < t_n|\tau_s < \tau_0]\mathbb{P}_y[\tau_s < \tau_0] \tag{4.8}$$

$$= \mathbb{P}_y[\tau_s < t_n|\tau_s < \tau_0]\Theta(1/\log(n)). \tag{4.9}$$

To estimate the probability on the last line, consider the complementary probability,

$$\mathbb{P}_y[\tau_s > t_n|\tau_s < \tau_0] := \mathbb{P}_y\left[\tau_s > \frac{1}{\log^{5/2}(n)} \,\middle|\, \tau_s < \tau_0\right] \leqslant \frac{\mathbb{E}_y\left[\tau_s|\tau_s < \tau_0\right]}{1/(\log^{5/2}(n))} = \mathcal{O}\left(\frac{1}{\log^{3/2}(n)}\right).$$

Substituting in Equation (4.9) gives us the desired result. ■

We proceed to calculate the time that it takes for an **unsuccessful** branch to coalesce. We will show that its coalescing time is negligible compared to t_n. This allows us to claim that consecutive branches do not overlap since, with high probability, by the time a new branch

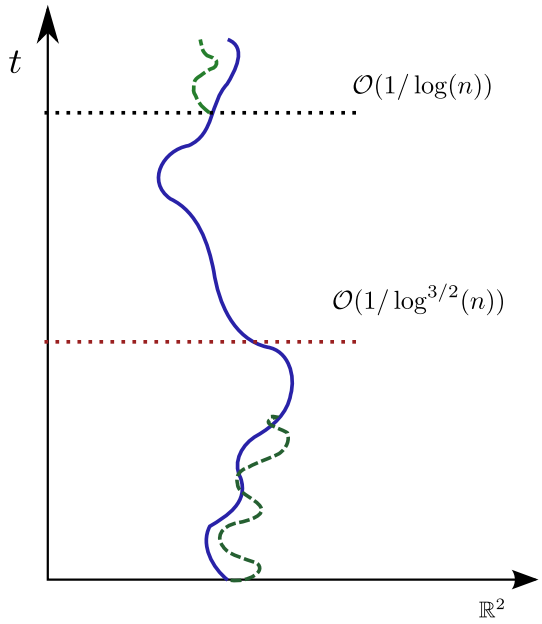

Figure 4.1: Here we illustrate how typically the coalescence of a branch will happen on a much shorter time-scale than the arrival of the next selective event, and thus an unsuccessful branch is unlikely to interfere with the next one.

appears the previous ones will have already either coalesced or escaped, and thus they are unlikely to interact with each other. We illustrate this in Figure 4.1.

Lemma 4.8 *The probability that a branch will have coalesced by time t_n after appearing is* $1 - \mathcal{O}(1/\log(n))$.

PROOF: Let $\mathcal{C}_n := \{\tau_0 < t_n\}$ be the event that the branch has coalesced by time t_n. We know that $\mathbb{P}_y[S_n] = \Theta(1/\log(n))$ for $y \in \mathcal{B}_{2\mathcal{R}_n}(0)\backslash\{0\}$.

$$\begin{aligned}\mathbb{P}_y[\mathcal{C}_n] &= \mathbb{P}_y[\mathcal{C}_n|S_n]\mathbb{P}_y[S_n] + \mathbb{P}_y[\mathcal{C}_n|S_n^c]\mathbb{P}_y[S_n^c] \\ &= \mathbb{P}_y[\mathcal{C}_n|S_n]\Theta(1/\log(n)) + \mathbb{P}_y[\mathcal{C}_n|S_n^c](1 - \Theta(1/\log(n))) \\ &\geqslant \mathbb{P}_y[\mathcal{C}_n|S_n^c] - \mathcal{O}(1/\log(n)).\end{aligned}$$

We need to show that $\mathbb{P}[\mathcal{C}_n|S_n^c] \geqslant 1-\mathcal{O}(1/\log(n))$. An identical argument to the one in Lemma 4.6 gives us that $\mathbb{E}_y[\tau_0|\tau_0 < \tau_s] = \mathcal{O}\left(\frac{1}{\log^4(n)}\right)$. So,

$$\mathbb{P}[\mathcal{C}_n|S_n^c] = \mathbb{P}_y[\tau_0 < t_n|\tau_0 < \tau_s] \geqslant 1 - \frac{\mathbb{E}_y[\tau_0|\tau_0 < \tau_s]}{1/\log^{5/2}(n)} = 1 - \mathcal{O}(1/\log^{3/2}(n)).$$

■

So far we have shown that the probability a branching event is successful is of order $\Theta(1/\log(n)$, and that the probability that it coalesces is $1-\mathcal{O}(1/\log(n))$. It could, of course, be the case that the branches 'hesitate' and neither coalesce nor separate. The following lemma shows that the probability of this happening is negligible for our purposes.

Lemma 4.9 *The probability that η^n hesitates satisfies :* $\mathbb{P}[t_n < \tau_0 \wedge \tau_s] = o(1/\log(n))$.

PROOF: Recall that we have chosen S_n to denote the event that a branching event is successful and C_n for the event that the branch is coalesced. We will use H_n for the event that the branch hesitates.

$$\mathbb{P}[H_n] = \mathbb{P}[S_n^c \cap C_n^c] = \mathbb{P}[S_n^c|C_n^c]\mathbb{P}[C_n^c].$$

Lemma 4.8 implies that $\mathbb{P}[C_n^c] = \mathcal{O}(1/\log(n))$, and thus we just need to show that $\mathbb{P}[S_n^c|C_n^c] = o(1)$. One would expect this to be the case since if the lineages ξ^1, ξ^2 were independent, η_{t_n} would be a long distance away from 0 with high probability, and conditioning on not coalescing should push it even further. We again use $\mathcal{E}$ to indicate the number of excursions that the process η performs before coalescing or escaping.

$$\mathbb{P}[S_n^c|C_n^c] = \mathbb{P}[S_n^c|C_n^c, \mathcal{E} \leqslant m]\mathbb{P}_y[\mathcal{E} \leqslant m|C_n^c] + \mathbb{P}[S_n^c|C_n^c, \mathcal{E} > m]\mathbb{P}[\mathcal{E} > m|C_n^c]. \tag{4.10}$$

We begin by considering the second summand, in particular

$$\mathbb{P}[\mathcal{E} > m|C_n^c] = \frac{\mathbb{P}[C_n^c|\mathcal{E} > m]\mathbb{P}[\mathcal{E} > m]}{\mathbb{P}_y[C_n^c]} \leqslant \frac{\mathbb{P}[C_n^c|\mathcal{E} > m]}{\mathbb{P}[C_n^c]} = \Theta(\log(n))\mathbb{P}[\mathcal{E} > m].$$

But $\mathbb{P}[\mathcal{E} > m]$ is the probability that a geometric random variable exceeds m, which can be bounded by $\exp(-\lambda m)$ for some $\lambda > 0$ (notice that $\mathcal{E}$, the number of times η^n visits $\mathcal{B}_{2\nabla_n}(0)$ before coalescing, is independent of n, and thus λ won't depend on n). Now, for the first summand in Equation (4.10) we focus on $\mathbb{P}[S_n^c | C_n^c, \mathcal{E} \leqslant m]$ which is the probability of remaining unsuccessful conditional on not coalescing and performing fewer than m excursions.

$$\mathbb{P}[S_n^c | C_n^c, \mathcal{E} \leqslant m] \leqslant \mathbb{P}\left[\sum_{i=1}^{\mathcal{E}} (\tau_i^{ex} + \tau_i^{in}) > t_n \,|\mathcal{E} \leqslant m\right] \leqslant \mathbb{P}\left[\tau^{ex} + \tau^{in} > \frac{t_n}{m}\right]$$

We can now use Markov's inequality to obtain:

$$\begin{aligned}\mathbb{P}\left[\tau^{ex} + \tau^{in} > \frac{\log^{-5/2}(n)}{m}\right] &\leqslant \mathbb{E}[\tau^{ex} + \tau^{in}] \times m \log^{5/2}(n) \\ &\leqslant m \log^{5/2}(n) \left(\frac{k_1}{\log^4(n)} + \frac{\hat{c}}{n}\right).\end{aligned}$$

We want to choose $m(n)$ so that both this last expression and $\log(n)\exp(-\lambda m(n))$ vanish in the limit as $n \to \infty$. We can achieve this by letting $m(n) = \log(n)$, and so we have shown that indeed $\mathbb{P}[S_n^c | C_n^c] = o(1)$ as required. ■

4.3 The skeleton

In the previous sections we made the distinction between successful and unsuccessful branching events, and showed that with high probability, there are only two cases we need to consider: either branches are successful or the lineages have coalesced by the end of the deciding period. Here we will focus on the successful lineages which we can think of as the 'skeleton' of the process of potential ancestral lineages. We will perform a sideways construction of the process started from a single lineage located at the origin. For convenience, at every selective event where the process branches we will deem one potential ancestor, chosen at random, to be the 'main branch' and the other one to be the 'secondary branch'.

4.3.1 Sideways construction

Fix $T > 0$, then for all $n \in \mathbb{N}$, we can construct the process $(\Xi_t^n)_{0 \leqslant t \leqslant T}$ of potential ancestral lineages in the following manner. We follow the lineage started at $(0, 0)$ up to the first time it is hit by a selective event. We keep record of the coordinates of the secondary branch in order to come back to it later, but follow the main branch until the next event. Once again we pick the main branch and add the secondary branch to our list of not yet visited branches. We continue in this manner until time T. This first ancestral path can be thought of as the spine of the process.

Once we hit time T, we go to the first secondary branch in our list, treating it as a main branch, and repeat the process until time T, or until we rejoin a path we have already added (which implies we coalesce with a path already added). For finite n we have a.s. only finitely many branches and thus our algorithm will halt in a finite number of steps almost surely. Since for every branch we meet we follow both the 'main branch' (the first time we arrive at the branch) and the 'secondary branch' (since we have added it to the list of paths to follow and we only stop when this list is empty), it follows that this algorithm will track down all of the paths of $(\Xi_t^n)_{0 \leqslant t \leqslant T}$.

The results in §4.2 imply that most secondary branches will quickly coalesce back with their corresponding main ones. We also know that unsuccessful branches won't appear in our rescaling limit, since only points that are at a distance of order $\mathcal{O}(1)$ will be distinguishable. Only successful branches will be able to reach such a separation. This leads us to the idea of a skeleton. Remember that successful branches are those for which $\tau_s < \tau_0 \wedge t_n$, so it only takes a time t_n after branching to know whether a branch is successful or not.

4.3.2 Constructing the skeleton

We now consider the skeleton. Once again we follow a lineage up to time T while keeping track of the branches not yet visited. But this time, whenever we start following a secondary branch from our list of 'yet to visit' branches, we follow it for a period t_n (the 'deciding period'

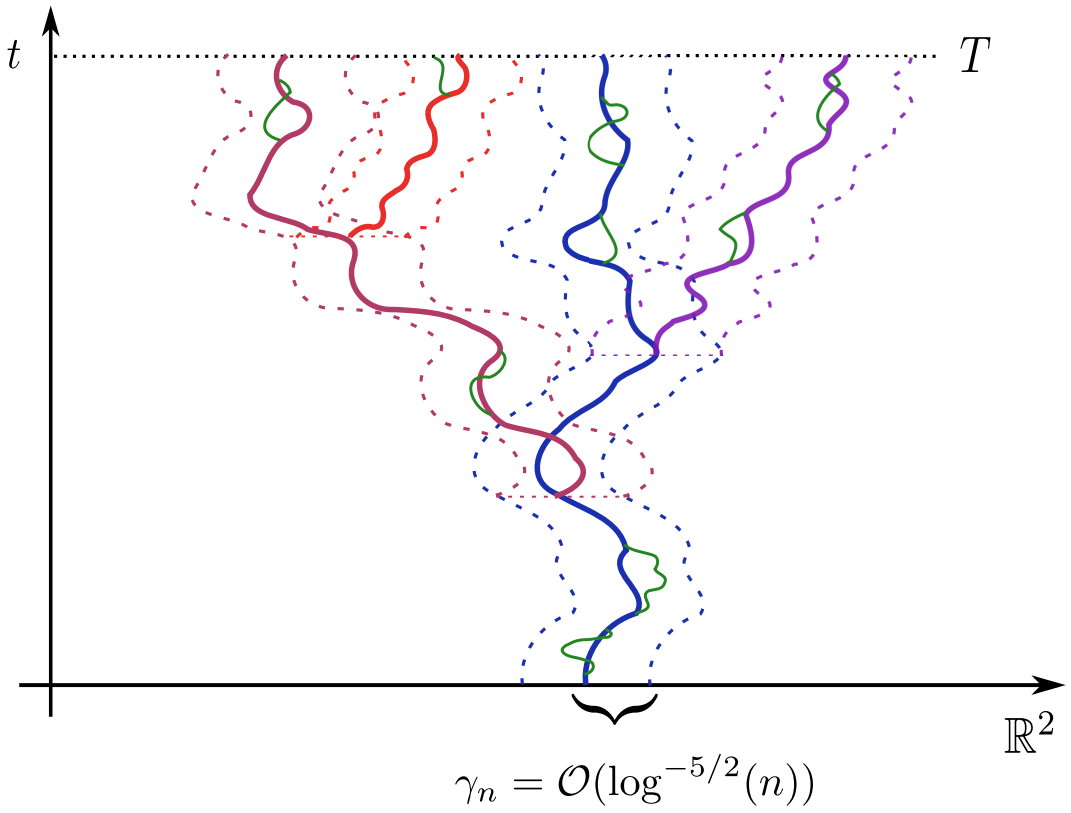

Figure 4.2: How to construct the skeleton. Here the blue line represents the spine, the first path we follow, and the dashed blue line is its envelope of size $\mathcal{O}(1/\log^{-5/2}(n))$. The paths in red, orange and violet are successful particles, with their own envelopes, while the green paths are unsuccessful particles which we discard.

introduced earlier in 4.1), and only if it is successful do we keep following it and paste it to the skeleton. Otherwise we simply discard it (together with any secondary branches that it might have).

Definition 4.10 *Take a fixed time interval* $[0,T]$*, the* ***n-skeleton*** $\mathcal{S}^n$ *is the collection of paths that is obtained when we only paste successful lineages in this way.*

Figure 4.2 illustrates the construction of the skeleton. We know that the probability of a branch being successful is $\mathcal{O}(1/\log(n))$, so it seems reasonable to expect that the number of branches in the skeleton is $\mathcal{O}(1)$. The following lemma gives us a bound on the rate at which branches are added to the skeleton.

Lemma 4.11 *There exists $\beta < \infty$ such that, for all $n \in \mathbb{N}$, $\mathbb{E}[|\mathcal{S}_T^n|] \leqslant \exp(\beta T)$.*

PROOF: Suppose that $|\mathcal{S}_t^n| = k$. Then the instantaneous rate at which $\mathcal{S}^n$ is hit by selective events is bounded by $kuV_{\mathcal{R}} n\mathbf{s_n} = \mathcal{O}(\log(n))$, since if lineages are further apart than $2\mathcal{R}_n$ they will branch independently, and if they are not, they might be hit by the same selective event, reducing the total rate. When a branch appears, the probability that it is successful is $\mathcal{O}(1/\log(n))$, in which case it should be pasted onto the skeleton. It follows that we can stochastically dominate $|\mathcal{S}_T^n|$ by the number of descendants at time T in a (continuous time) Galton-Watson branching process with branching rate

$$kuV_{\mathcal{R}} n\mathbf{s_n}\mathcal{O}(1/\log(n)) = \mathcal{O}(1),$$

and since $|\mathcal{S}_0^n| = 1$, the result follows. ■

The way we have constructed the skeleton, it is possible to miss successful paths, since the paths we chose not to paste might have branched during the deciding period (in which case we followed the 'main branch', and it could very well be that the 'secondary branch happens to be successful). The following Lemma shows that this possibility vanishes as n tends to infinity.

Lemma 4.12 *Fix $T > 0$ and construct the skeleton $\mathcal{S}^n$ of Ξ^n on $[0, T]$. The probability that there is a branch of Ξ^n along any of the paths discarded in the construction of $\mathcal{S}^n$ is of $o(1)$.*

PROOF: Since the expected number of lineages in the skeleton is bounded (Lemma 4.11), the expected number of selective events along it is of order $\mathcal{O}(\log(n))$. By Lemma 4.9 the probability that a branch 'hesitates' is $o(1/\log(n))$ and thus, with probability tending to 1, we only discard 'coalesced' branches. In Lemma 4.8 we showed that $\mathbb{E}[\tau_0 | \tau_0 < \tau_s] = \mathcal{O}(1/\log^3(n))$. It follows that the expected total duration of the unsuccessful branches is bounded by $\mathcal{O}(1/\log^2(n))$, and since selective events arrive at rate $\mathcal{O}(\log(n))$ it follows that the probability that any one of the unsuccessful branches is hit by a selective event is $o(1)$. ■

4.4 Convergence to BBM

Equipped with the skeleton and the estimates of section §4.2, we now proceed to show that the genealogical process converges to a BBM as n tends to infinity.

First we show, in a sense to be made more precise below, that the $\mathcal{S}^n$ and Ξ^n are actually very similar and in particular must converge to the same limit as $n \to \infty$. We then prove that the n-skeleton converges to a BBM with a branching rate of order $\mathcal{O}(1)$.

4.4.1 The skeleton and the genealogical process

It shouldn't be surprising to see that very little is lost if we are to consider the skeleton instead of the full process Ξ^n. Since, with probability tending to one, the skeleton contains all the paths starting from successful branches, all that is missing are the paths associated with unsuccessful branches. The expected total lifetime of the unsuccessful branches is $o(1)$, and whenever no unsuccessful branch is alive in the system the skeleton and the genealogical process are identical (with probability tending to one).

Notice that we can identify both $\Xi^n(s)$ and $\mathcal{S}^n(s)$ with a discrete measure that represents the locations of the corresponding lineages at time s. We abuse notation by using the same symbols to denote the discrete measure valued processes, and we write $\|\nu\|_{TV}$ for the total variation distance.

Proposition 4.13 *Fix an interval $[0,T]$, and take $(\Xi^n_s)_{0\leqslant s\leqslant T}$ be the diffusively rescaled genealogical process and its skeleton $(\mathcal{S}^n_s)_{0\leqslant s\leqslant T}$ thought of as measure valued processes, both started from $\delta_x = \mathcal{S}^n_s = \Xi^n_0$. We then have that for any finite sequence $0 \leqslant s_1 < s_2 < ... < s_k \leqslant T$, and for all $\epsilon > 0$:*

$$\lim_{n\to\infty} \mathbb{P}\left[\|\Xi^n_{s_1} - \mathcal{S}^n_{s_1}\|_{TV} < \epsilon, \|\Xi^n_{s_2} - \mathcal{S}^n_{s_2}\|_{TV} < \epsilon, ..., \|\Xi^n_{s_k} - \mathcal{S}^n_{s_k}\|_{TV} < \epsilon\right] = 1.$$

PROOF: Recall that we have constructed the skeleton on the same probability space as Ξ^n. Moreover, from Lemmas 4.9 and 4.12 we have that, with probability tending to one, no branch

'hesitates' and all successful branches are in the skeleton. Thus we see immediately that we can couple the two processes in such a way that, with probability tending to one, outside of the 'deciding periods' they must be identical.

It remains to show that the probability that one or more of the times $0 \leqslant s_1 < s_2 < ... < s_k \leqslant T$ falls within a deciding period is vanishingly small. Let us call this event $\mathcal{I}$. The probability that s_i falls within a deciding period is the probability that a selective event hits the skeleton in the time interval $(s_i - t_n, s_i]$ (because we only see branches along the skeleton).

By Lemma 4.11, the number of particles in the skeleton has finite expectation. Thus, as $m \to \infty$, $\mathbb{P}[|\mathcal{S}_T^n| > m] \to 0$. Given that there are at most m branches, the probability that one of them ocurrs the time interval $(s_i - t_n, s_i]$ for some i is $\mathcal{O}(m \log^{-5/2}(n))$. By choosing $m = \log^{1/2}(n)$ we show that $\mathbb{P}[|\mathcal{S}_T^n| > m] + \mathbb{P}[\mathcal{I}||\mathcal{S}_T^n| < m]$ vanishes and the result follows. ■

Remark 4.14 *We have shown that the probability that for any given $0 \leqslant s_1 < s_2 < ... < s_k \leqslant T$, the probability that Ξ^n and $\mathcal{S}^n$ are identical at these times converges to one, and so in particular if one of the processes converge weakly, in the sense of the finite dimensional distributions, they must both converge to the same limit point.*

4.4.2 BBM and the skeleton

It remains to check that the skeleton converges to a binary BBM. A binary BBM is characterised by the following three properties:

1. During its lifetime, each particle moves in $\mathbb{R}^2$ according to a Brownian motion independently of all other lineages.

2. Each particle lives for an exponentially distributed time.

3. When a particle dies it leaves two offspring at precisely the location of its death. Conditional on their time and place of birth, offspring evolve independently of each other, in the same manner as their parent.

We will show that the in the limit, the skeleton follows precisely these dynamics, and thus that it converges to a BBM. Before continuing we need one more definition.

Definition 4.15 *If θ_r is the distribution of the distance between two points sampled at random from the ball of radius r, we define $\mathbf{p_n}$ to be:*

$$\mathbf{p_n} := \int\int \mathbb{P}_x[S_n]\theta_r(dx)\mu^n(r),$$

*where $\mathbb{P}_x[S_n]$ is the probability that a branch η^n started from x is successful. We call $\mathbf{p_n}$ the **true success probability**. Notice that $\mathbf{p_n} = \Theta(1/\log(n))$ just like $\mathbb{P}_x[S_n]$.*

Proposition 4.16 *Let ξ be a path of the skeleton. With probability tending to one as $n \to \infty$ the arrival of successful branches is given by a Poisson process with finite rate.*

Proof: If all the success of branches along the path were independent with probability $\mathbf{p_n}$ (the success probability), then it would follow that the arrival of successful branches would be given by a thinning of the original Poisson point process, giving the desired result.

The way we have defined a successful branch it follows that if the deciding periods of two branches do not overlap, then their respective success is independent (the process Ξ^n has sampling consistency, so whether an extra lineage meanders around a branch doesn't alter its dynamics). It should be clear that if deciding periods along the path overlap, at least two consecutive branches are involved in the overlap, so it suffices to consider the pairs of consecutive branches. Two consecutive branches will have overlapping deciding periods if and only if either of the lineages in the earlier branch splits through a selective event before the deciding period is over.

We can bound from above the probability of this happening. Since the branching rate of the two lineages is at most twice the branching rate of a single one, letting $\mathcal{T}_n$ be an exponential random variable with parameter $2usV_R\log(n)$, the probability that we see an overlap is bounded by:

$$\mathbb{P}[\mathcal{T}_n < t_n] = 1 - \exp\left(2usV_R\log(n)t_n\right) = \mathcal{O}(\log^{-3/2}(n)).$$

Since the expected number of branches along the spine is $\mathcal{O}(\log(n))$, it follows that the probability of seeing deciding periods overlap is $o(1)$. Let N_t^s be the arrival process of successful branches, and N_t be a Poisson process with parameter $usV_{\mathcal{R}} \log(n)\mathbf{p_n}$, both defined on the interval $[0, T]$, then:

$$\mathbb{P}[N_t^s \overset{(d)}{=} N_t] = 1 - o(1).$$

■

By thinking of a branching event as the death of a lineage and the birth of two offspring, we can think of the time to the arrival of the first (successful) branch as the 'lifetime' of a particle in the skeleton. Proposition 4.16 guarantees that as $n \to \infty$ the lifetime of distinct particles converges to independent exponential random variables. Moreover, since paths arising from a branch start from a distance $\mathcal{O}(1/\sqrt{n})$ from each other and their 'parent', it follows (because this distance converges to 0) that as $n \to \infty$ we can think of the parent as splitting into two.

Theorem 4.17 *Fix $T > 0$, then, as $n \to \infty$, the skeleton $(\mathcal{S}_t^n)_{0 \leqslant t \leqslant T}$ converges weakly to a BBM with finite branching rate .*

Proof: It is left to show that the successful paths converge to independent Brownian motions. If we take two paths from a successful branch it is clear that they are actually not independent up to time t_n, since they are conditioned to move far apart. After the deciding period though, they will evolve independently unless they manage to get to a distance smaller than $2\mathcal{R}_n$.

Applying the strong Markov property at time τ_s, and using the embedding of Proposition 3.5, the probability that the lineages come within distance $2\mathcal{R}_n$ before reaching separation M is bounded by the probability that two Brownian motions started at a separation at least γ_n come within distance $2\mathcal{R}_n$ before reaching separation M, which is $\mathcal{O}(\log\log(n)/\log(n))$ for any fixed M.

Since the number of branches in the skeleton can be controlled uniformly in n, we can choose M arbitrarily large so that the probability that any branch coalesces with its siblings is smaller than any prescribed $\delta > 0$.

Of course we must also control the probability that a newly created branch coalesces with a branch in the skeleton other than its sibling. We can achieve this in the same manner we did in Lemma 3.14. It suffices to to show that at a birth event, the branches other than the newly created 'siblings' are sufficiently far apart. Again, since the number of branches in the skeleton is controlled independently of n, a simple induction shows that it is enough that two siblings are sufficiently far apart by the first time one of them branches. But on the event that they do not coalesce, this distance converges to that of two independent Brownian motions run for an exponential time with rate $\mathcal{O}(1)$, so the result follows.

■

Theorem 4.18 *Fix $T > 0$, then, as $n \to \infty$, the ancestral process $(\Xi_t^n)_{0 \leqslant t \leqslant T}$ converges weakly to a BBM with finite branching rate .*

Proof: It follows immediately from Theorem 4.2 and Proposition 4.13. ■

CHAPTER 5

Dimension 1

We now turn our attention to the case $d = 1$. As we mentioned in the beginning of §4, in the one-dimensional setting we expect that it will take $\mathcal{O}(\sqrt{n})$ branches before we might see a 'successful' branch, that is, one where the two lineages manage to separate to a macroscopic distance. Thus, in this chapter we shall take $\mathbf{s}_n = \mathbf{s}/\sqrt{n}$ in the hope of obtaining a non trivial limit for our genealogical process.

In fact, it turns out that a very similar problem has been tackled by Sun and Swart [2008], who consider a system of discrete branching and coalescing random walks, and obtain a limiting object they have called the *Brownian net*. In this chapter we will show that the scaling limit of the genealogical process also converges to this object. Although this result owes a lot to

the existing literature, the continuum setting provides some new technical challenges, and it is interesting to show that the Brownian net appears as the a limiting object for a different class of systems than that initially considered. In other words, we show that the process we consider belongs to the universality class of the Brownian net.

The first section of this chapter will provide a brief overview of the *Brownian net* as well as the *Brownian web*, two key objects in our study, and will state our main theorem in the one-dimensional case, namely, the convergence of the SΛFV dual process to the Brownian Net. The rest of the chapter will be dedicated to its proof.

5.1 The Brownian Web and the Brownian Net

To further motivate the one-dimensional result, let us consider first what happens in the absence of selection. The dual process then reduces to a system of coalescing random walks and, as proved in Berestycki *et al.* [2013], after a diffusive rescaling one recovers a system of coalescing Brownian motions. In fact, using for example the construction of Véber and Wakolbinger [2013], we can construct the coalescing dual started from any countable set of space-time points in $\mathbb{R}^2$. In the special case when $u = 1$, if we were to take the centre of the event, rather than a randomly chosen point, as the location of the parent, then this corresponds to the process of 'trajectoires d'exploration' of Micaux [2007], who constructs a stochastic flow of maps by considering the dual started from every space-time point in the plane. If we specialise still further so that all events have radius 1 and we start 'exploration paths' only from the centre of each reproduction event then we recover a càdlàg version of the Poisson trees of Ferrari *et al.* [2004]. By interpolation we recover the Poisson trees themselves. In Ferrari *et al.* [2005] it is shown that under a diffusive rescaling (and in a suitable topology) the Poisson trees converge to the Brownian web. The Brownian web is the random network consisting of the paths of coalescing one-dimensional Brownian motions starting from every space-time point in $\mathbb{R} \times \mathbb{R}$

We recall only the results that we need for the Brownian web. An excellent introduction,

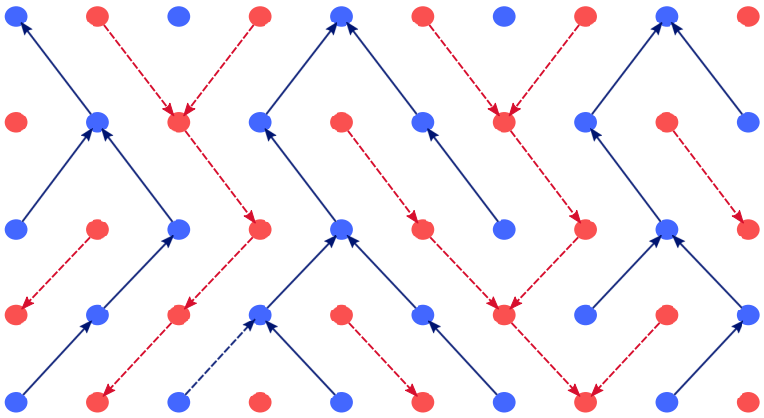

Figure 5.1: Web duality. The blue arrows represent the forwards in time pre-limiting coalescing random walks, while the red arrows represent the backwards in time dual.

some of which we now borrow, can be found in Sun and Swart [2008]. Ferrari *et al.* [2005] rests heavily on Fontes *et al.* [2004], who constructed an appropriate metric space, which we shall denote $(\widetilde{M}, \tilde{d})$ and describe in more detail in Remark 5.6, in which to consider convergence of sets of continuous paths and, using this space, established convergence of coalescing simple random walks to the Brownian web. We write $\mathcal{K}(\widetilde{M})$ for the space of all compact subsets of $\widetilde{M}$. Furthermore, for any $K \in \mathcal{K}(\widetilde{M})$ and $A \subset \mathbb{R}^2$, we let $K(A)$ denote the collection of all paths in K with starting points in A, and for $z \in \mathbb{R}^2$ we write $K(z) := K(\{z\})$.

Arratia [1979] was the first to observe that the Brownian web exhibits a self-duality. As illustrated in Figure 5.1, if one takes the graphical representation of the pre-limiting system of coalescing random walks, that is one represents paths as a series of arrows representing the jump made by the path out of each point of $\mathbb{Z}$ at each time $t \in \mathbb{Z}$, then there is a natural dual system of arrows, pointing in the opposite direction of time, which 'fills out the gaps' between the walkers forwards in time. Under the diffusive rescaling, we obtain joint convergence of the forwards and backwards systems to a pair $(\mathcal{W}, \widehat{\mathcal{W}})$, where $\mathcal{W}$ is the Brownian web as before and the dual web $\widehat{\mathcal{W}}$ has the same law as $\mathcal{W}$ rotated through 180 degrees. We write $\mathcal{K}(\widehat{M})$ for the state space of $\widehat{\mathcal{W}}$. The Brownian web and its dual a.s. uniquely determine one another. They can be characterised as follows.

Theorem 5.1 (Schertzer *et al.* [2014], Proposition 3.1) *For each $\beta \in \mathbb{R}$, there exists a $\mathcal{K}(\widetilde{M}) \times \mathcal{K}(\widehat{M})$-valued random variable $(\mathcal{W}, \widehat{\mathcal{W}})$, called the double Brownian web with drift β, whose distribution is uniquely determined by the following properties:*

1. *For each deterministic $z \in \mathbb{R}^2$, $\mathcal{W}(z)$ consists a.s. of a single path $\mathcal{W}(z) = \{\pi_z\}$ and $\widehat{\mathcal{W}}(z)$ consists a.s. of a unique dual path $\hat{\pi}_z$.*

2. *For any finite deterministic set of points $z_1, \dots, z_k \in \mathbb{R}^2$, $(\pi_{z_1}, \dots, \pi_{z_k})$ is distributed as a system of coalescing Brownian motions, each with drift β, starting at space-time points $z_1, \dots z_k$.*

3. *For any deterministic countable dense set $\mathcal{D} \subset \mathbb{R}^2$, a.s. $\mathcal{W}$ is the closure of $\mathcal{W}(\mathcal{D})$ in $\widetilde{M}$ and $\widehat{\mathcal{W}}$ is the closure of $\widehat{\mathcal{W}}(D)$ in $\widehat{M}$.*

4. *For any deterministic $z \in \mathbb{R}^2$, the dual path $\hat{\pi}_z$ is the a.s. unique path in $\widehat{M}$ started from z that does not cross any path in $\mathcal{W}$.*

Remark 5.2 *Fontes* et al. *[2004] defined the Brownian web with $\beta = 0$, but the more general form considered by Schertzer* et al. *[2014] is what we shall need in what follows.*

Working in (essentially) the same state space, Sun and Swart [2008] showed how to obtain an analogue of the Brownian web, which they dubbed the *Brownian net*, as the scaling limit of a system of branching and coalescing simple random walks. In order to obtain a non-trivial limit, the branching rate is, of course, scaled exactly as in our dual to the SΛFVS. Not surprisingly then, our result in $d = 1$ will concern convergence to the Brownian net.

In contrast to the Brownian web, the Brownian net will have a multitude of paths coming out of each space-time point. The key to its characterisation is that it has a well-defined left-most and right-most path, which we denote l_z and r_z respectively, emanating from each point $z = (x, t) \in \mathbb{R}^2$ and these determine what is called a *left-right Brownian web*. It turns out that a Brownian net is completely determined by the corresponding left-right Brownian web and so it will suffice to study this object.

For any deterministic points $z_1, \ldots, z_k, z'_1, \ldots, z'_{k'}$, the paths $l_{z_1}, \ldots, l_{z_k}$ are coalescing Brownian motions with drift ζ to the left, and $r_{z'_1}, \ldots r_{z'_{k'}}$ are coalescing Brownian motions with drift ζ to the right. Before we can describe the left-right web, we must explain how a left-most path $l_z = l_{(x,s)}$ and a right-most path $r_{z'} = r_{(x',s')}$ interact. Their joint evolution after time $s \vee s'$ is the unique weak solution to the left-right stochastic differential equation

$$\begin{aligned} dL_t &= \xi \mathbf{1}_{\{L_t \neq R_t\}} dB^l_t + \xi \mathbf{1}_{\{L_t = R_t\}} dB^c_t - \zeta dt, \\ dR_t &= \xi \mathbf{1}_{\{L_t \neq R_t\}} dB^r_t + \xi \mathbf{1}_{\{L_t = R_t\}} dB^c_t + \zeta dt, \end{aligned} \tag{5.1}$$

where B^l_t, B^r_t and B^c_t are independent standard Brownian motions. Sun and Swart [2008] proved (weak) existence and uniqueness of the solution to this system. These equations can be understood as modelling the interaction of two Brownian motions with drift, that coalesce instantly upon meeting but are immediately untangled by their opposite drifts, although they do manage to accumulate local time at the coalesced state.

Remark 5.3 *In Sun and Swart [2008] the drift parameter ζ of the left-right stochastic differential equation used to construct the Brownian net is allowed to vary but the diffusion constant, ξ^2, that is the 'clock', of the Brownian motions is not. Applying a linear time change to their construction yields this extension and we will use such webs and nets (and results from elsewhere extended trivially to apply to them) without further comment.*

This is enough to specify the joint distribution of any finite collection of left-right paths, which we will call left-right coalescing Brownian motions. The characterisation of the left-right Brownian web mirrors that of the Brownian web in Theorem 5.1.

Theorem 5.4 (Sun and Swart [2008], Theorem 1.5) *There exists a $(\mathcal{K}(\widetilde{M}))^2$-valued random variable $(\mathcal{W}^l, \mathcal{W}^r)$, the (standard) left-right Brownian web, whose distribution is uniquely determined by the following properties:*

1. *For each deterministic $z \in \mathbb{R}^2$, a.s. $\mathcal{W}^l(z)$ and $\mathcal{W}^r(z)$ each contain a single path $\mathcal{W}^l(z) = \{l_z\}$ and $\mathcal{W}^r(z) = \{r_z\}$.*

2. *For any finite deterministic set of points* $z_1, \ldots, z_k, z'_1, \ldots, z'_{k'} \in \mathbb{R}^2$*, the collection of paths* $(l_{z_1}, \ldots, l_{z_k}; r_{z'_1}, \ldots, r_{z'_{k'}})$ *is distributed as a collection of left-right coalescing Brownian motions.*

3. *For any deterministic countable dense sets* $\mathcal{D}^l, \mathcal{D}^r \subset \mathbb{R}^2$*,* $\mathcal{W}^l = \overline{\{l_z : z \in \mathcal{D}^l\}}$ *and* $\mathcal{W}^r = \overline{\{r_z : z \in \mathcal{D}^r\}}$ *a.s..*

Sun and Swart [2008] establish a self-duality for the left-right Brownian web which characterises the Brownian net in an analogous way to the characterisation of Theorem 5.1 for the Brownian web. It is this self-duality which underpins the proof of convergence to the Brownian net.

Since the left (right)-most web a.s. uniquely determines its dual, we can consider the pair of dual webs $(\widehat{\mathcal{W}}^l, \widehat{\mathcal{W}}^r)$. Sun and Swart [2008] show that $-(\widehat{\mathcal{W}}^l, \widehat{\mathcal{W}}^r)$ has the same distribution as $(\mathcal{W}^l, \mathcal{W}^r)$, and that it is possible to fully characterise the Brownian net associated to $(\mathcal{W}^l, \mathcal{W}^r)$ by the dual left-right Brownian web. In order to state the theorem we need to introduce the idea of a wedge. Take $\hat{l} \in \widehat{\mathcal{W}}^l$ and $\hat{r} \in \widehat{\mathcal{W}}^r$ such that $\hat{r}(s) < \hat{l}(s)$, where time s denotes the birth time of the younger path (notice that time is running backwards for the dual left-right web). Let $T := \sup\{t < s : \hat{r}(t) = \hat{l}(t)\}$, the first time time the paths meet. We call the open set

$$W(\hat{r}, \hat{l}) := \{(x, u) \in \mathbb{R}^2 : T < u < s, \hat{r}(u) < x < \hat{l}(u)\}$$

a wedge (see Figure 5.2). We say that a path π started at time σ_π enters W from outside if there exists $\sigma_\pi < u < t$ (notice the strict inequalities) such that $\pi(u) \notin W$ and $\pi(t) \in W$.

Theorem 5.5 (Sun and Swart [2008], Theorem 1.10) *Let* $(\mathcal{W}^l, \mathcal{W}^r, \widehat{\mathcal{W}}^l, \widehat{\mathcal{W}}^r)$ *be a left-right Brownian Web and its dual. Then almost surely,*

$$\mathcal{N} = \Big\{\pi \in \Pi : \pi \text{ does not enter any wedge of } (\widehat{\mathcal{W}}^l, \widehat{\mathcal{W}}^r) \text{ from outside}\Big\} \tag{5.2}$$

is the Brownian net associated with $(\mathcal{W}^l, \mathcal{W}^r)$*.*

In all the work described above, lineages coalesce instantly on meeting and, in particular, they

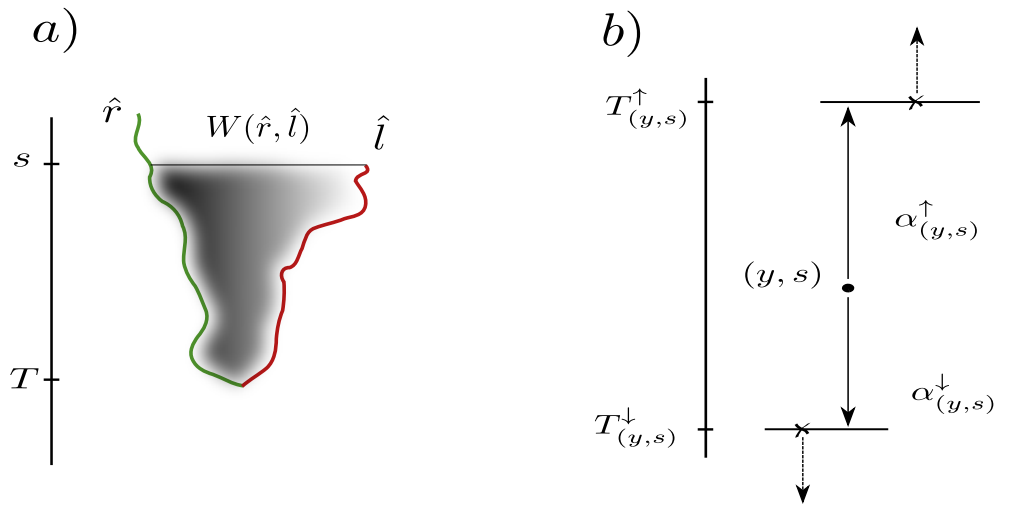

Figure 5.2: a) Wedges. On the left we see a represenation of a wedge $W(\hat{r},\hat{l})$, which is the area shadowed with grey. **b) Paths and arrows.** The second picture illustrates how to build paths from arrows. We give the forward and backwards arrow of the point (y,s), which are constant functions up to the time of the next event. One we see an event, we follow the arrow originating from the parent of said event.

cannot 'jump over' one another. Our one-dimensional results also require this property which is achieved by setting $u = 1$ in the SΛFV process. In this regime, the entire local population is replaced during a reproduction event. We further impose the condition that, for any t and x, $w_t(x) \in \{0,1\}$ (and note that this is consistent with taking $u = 1$).

In the absence of selection, it is shown in Berestycki *et al.* [2013] that if instead we fix $u \in (0,1)$, the scaling limit that we obtain is a simple timechange of that for $u = 1$. However, in Chapter 6 we present simulations which suggest that things are a bit more complicated once selection is incorporated.

Nevertheless, even after the simplification $u = 1$, our prelimiting process is considerably more complex than that considered in Sun and Swart [2008]. For each point $p = (x,t) \in \mathbb{R}^2$, we think of an individual living at (x,t) and construct the set $\mathscr{A}_n(p)$ of its possible ancestral lineages by looking forwards in (coalescent) time and tracing out the locations in space-time of successive possible ancestors of p. When lineages are covered by the same neutral reproduction event, they coalesce. In particular, more than two lineages can coalesce in a single event. At

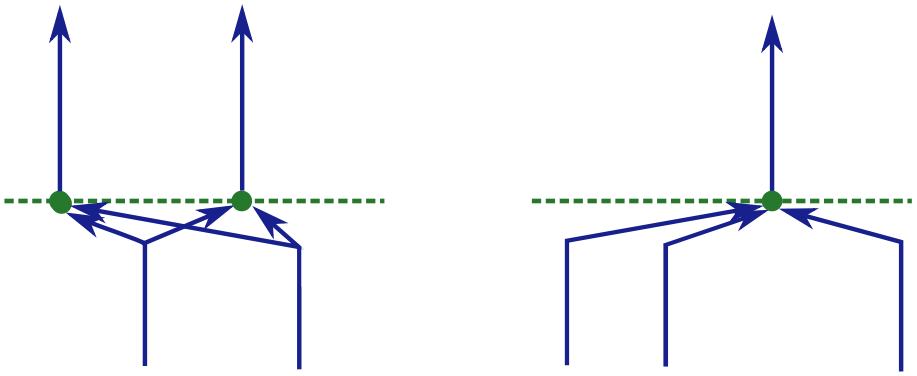

Figure 5.3: Complications. The diagram to the left illustrates how paths might both coalesce and branch through the same event. The second diagram presents a case of multiple collisions.

selective events, when we must trace two potential parents, we can see either just branching or, if more than one lineage lies in the region affected by the event, a combination of branching and coalescence (see Figure 5.3).

Further complications compared to systems of branching and coalescing simple random walks arise since (a) our ancestral lineages jump at random times and the displacement caused by such jumps is random; and (b) the motion of distinct ancestral lineages becomes dependent when they are within distance $2\mathcal{R}_n$ of each other.

In spite of the additional complexity, it still makes sense to talk about left-most and right-most paths and this will be the key to our analysis. In fact (b) can be handled through elementary arguments. It turns out that the times in which ancestral lineages are 'nearby but not coalesced' are too brief to affect the limit. In order to overcome (a), we must identify the dual system of branching and coalescing lineages which will give rise to the dual left-right web. At first sight, it is far from obvious that such a dual exists, and still less obvious that it can be constructed in such a way as to have the same distribution as the forwards system. The required dual, which is defined in § 5.2 is shown in Figure 5.4.

To formulate our result in $d = 1$, we first need to define the space in which the convergence will take place. This is a slight modification of that used by Sun and Swart [2008], to take account of the fact that our branching and coalescing ancestral lineages are only càdlàg, not

continuous. For $s \in [-\infty, \infty]$, we set

$$D[s] = \{f : [s, \infty] \to [-\infty, \infty]\,;\, f \text{ is càdlàg}\}.$$

For $f \in D[s]$, it will be convenient to define $\sigma_f = s$ to be the first time at which it is defined. In similar fashion to Sun and Swart [2008], we will compactify the space we are working on in order to use compactness when proving convergence.

For each $s \in [-\infty, \infty]$ and $f \in D[s]$ we define a function $\bar{f}$ as follows. Let

$$\kappa(t) = \kappa_t = \tanh^{-1}(t)$$

and note that κ is an order preserving homeomorphism between $[-1, 1]$ and $[-\infty, \infty]$. (We use the symbol κ in place of $\tanh^{-1}$ to denote a change of time rather than rescaling of space.)

Then if $f \in D[-\infty]$ we define

$$\bar{f}(t) = \frac{\tanh(f(\kappa_t))}{1 + |\kappa_t|} \tag{5.3}$$

for $t \in [-1, 1]$. Similarly, if $s \in (-\infty, \infty)$ and $f \in D[s]$ then for $t \in [\kappa^{-1}(\sigma_f), 1]$ we define $\bar{f}$ as above and for $t \in [-1, \kappa^{-1}(\sigma_f)]$ we define $\bar{f}(t) = \bar{f}(\kappa^{-1}(\sigma_f))$. Finally, for $f \in D[\infty]$ we define $\bar{f}(\infty) = 0$.

For each $s \in [-\infty, \infty]$ and $f \in D[s]$, $\bar{f}$ is an element of the space

$$G = \{g : [-1, 1] \to [-1, 1]\,;\, g \text{ is càdlàg}\}.$$

We equip G with its usual Skorohod metric ρ. Let

$$d(f_1, f_2) = \rho(\bar{f}_1, \bar{f}_2) \vee |\tanh(\sigma_{f_1}) - \tanh(\sigma_{f_2})| \tag{5.4}$$

where $f_i \in D[t_i]$. It is easily seen that d is a pseudo-metric on $\bigcup_{s\in[-\infty,\infty]} D[s]$. We define M to be the set of equivalence classes of $\bigcup_{s\in[-\infty,\infty]} D[s]$ under d and note that d naturally induces a metric on M.

Remark 5.6 *If we restrict ourselves to continuous paths, and replace the Skorohod metric with*

the usual L^∞ distance (that is, we replace d by the pseudometric $\widetilde{d}(f_1, f_2) = \left\|\bar{f}_1 - \bar{f}_2\right\|_\infty \vee |tanh(\sigma(f_1) - tanh(\sigma(f_2)|$), then we recover the space $(\widetilde{M}, \widetilde{d})$ introduced by Fontes et al. [2004] with two cosmetic differences:

1. *We have identified the values of functions at $\{-\infty, \infty\}$ using equivalence classes of $\widetilde{M}$ instead of using a particular compactification of $[-\infty, \infty]^2$ (see the appendix of Sun and Swart 2008).*

2. *Using $\bar{f}$, we extended $f \in D[\sigma_f]$ to all time and so the Skorohod metric in (5.4) is over the whole interval of (κ^{-1}-rescaled) time $[-1, 1]$ instead of just $[\kappa^{-1}(\sigma_f), 1]$. This makes no difference for continuous paths but is better suited to càdlàg paths.*

We define also the set $\mathcal{K}(M)$ of compact subsets of M, equipped with the Hausdorff metric m and including the empty set $\varnothing$ as an isolated point. We show (in the Appendix, A.3) that M is complete and separable; the space $\mathcal{K}(M)$ inherits these properties.

The reason for using the metric (5.4) is captured in the following lemma (which we prove in the Appendix, A.2).

Lemma 5.7 *Let $f_n \in M$ for $n \in \mathbb{N}$ and let $f \in M$. Extend f_n and f to functions on $[-\infty, \infty]$ by setting $f_\cdot(t) = f(\sigma_f)$ for $t < \sigma_f$. Then $f_n \to f$ in (M, d) if and only if both:*

1. *for all $T' \in (0, \infty)$, $f_n \to f$ in $D_{\mathbb{R}}[-T', T']$ (càdlàg paths with the usual Skorohod metric),*

2. *$\sigma(f_n) \to \sigma(f)$ in $[-\infty, \infty]$.*

The topology for the backwards paths will be the same except rotated through 180 degrees.

Let $\mathscr{A}_n(p)$ denote the set of all possible ancestral lineages started from the space-time point p. We define

$$\mathcal{D}_n = \{(kn^{-1/2}, jn^{-1/2})\,;\, j, k \in \mathbb{Z}\} \tag{5.5}$$

and

$$\mathscr{A}_n = \bigcup_{p \in \mathcal{D}_n} \mathscr{A}_n(p). \tag{5.6}$$

Loosely speaking, Theorem 5.8 says that $\mathscr{A}_n$ converges in law in $\mathcal{K}(M)$ to the Brownian net. In order to make this statement precise we shall need a rather more involved description of $\mathscr{A}_n(p)$, which can be found in Section 5.2, and we shall augment $\mathscr{A}_n$ with 'boundary paths' in order to exploit the compactness of the state space. The corresponding collection of paths is defined in §5.2.3 and will be denoted $\mathcal{P}_n^{\uparrow}(\mathcal{D}_n)$. Our main result in dimension $d = 1$ is then:

Theorem 5.8 *Let $\mathcal{N}$ denote the Brownian net. As $n \to \infty$, $\mathcal{P}_n^{\uparrow}(\mathcal{D}_n)$ converges weakly to $\mathcal{N}$ in $\mathcal{K}(M)$.*

The proof will be object of the rest of this Chapter. Since we will mirror the approach of Sun and Swart [2008] we will attempt, as far as possible, to adhere to their notation. Before proceeding we will introduce one extra piece of terminology.

Definition 5.9 *In a selective event we refer to both the potential parental locations as parents of the event. In a neutral event, we call the (sole) parental location the parent of the event.*

5.2 Paths and arrows

In order to discuss the self-dual systems of branching and coalescing lineages that converge to the Brownian net, we must be precise about what we mean by 'branching-coalescing paths' and, in particular, about the direction of time. We shall follow Fontes *et al.* [2004] in using segments of paths called *arrows*. Loosely speaking, paths are formed by concatenating arrows. A path (or an arrow) is an $\mathbb{R}$-valued function whose domain is a subinterval of $\mathbb{R}$. If a path/arrow is *forwards* (resp. *backwards*), then 'moving along it' means moving along the image of the path forwards (resp. backwards) with respect to the usual ordering on the time domain. We shall use $\uparrow$ to denote forwards and $\downarrow$ to denote backwards paths.

When we are following a backwards path or arrow we interchange left and right, in the same way as left and right interchange if we reverse the direction in which we walk. For clarity, we reserve the terms *north, south, east and west* for global directions associated to the plane $\mathbb{R}^2$

and use the words *right* and *left* for local directions whose frame of reference depends on the direction in which we are travelling.

5.2.1 Constructing paths from arrows

We refer to each $(x,t,r) \in \Pi^n$ as an event affecting the set $\mathcal{B}_r(x) \times \{t\} \subseteq \mathbb{R}^2$ or, equivalently, affecting each point $y \in \mathcal{B}_r(x)$ at time t. The east- and west-most points of this event are $(x+r,t)$ and $(x-r,t)$ respectively. To each $(y,s) \in (-\infty,\infty) \times (-\infty,\infty)$ we associate a unique forwards arrow and a unique backwards arrow, defined as follows. Let

$$T^{\uparrow}_{y,s} = \inf\{t\,;\, \exists (x,t,r) \in \Pi^n, y \in \mathcal{B}_r(x), t \geqslant s\}$$
$$T^{\downarrow}_{y,s} = \sup\{t\,;\, \exists (x,t,r) \in \Pi^n, y \in \mathcal{B}_r(x), t \leqslant s\}.$$

Let $\star \in \{\uparrow,\downarrow\}$. An arrow starting at (y,s) is a path $\alpha^{\star}_{y,s} : [s, T^{\star}_{y,s}) \to \mathbb{R}$ defined by $\alpha^{\star}_{y,s}(v) = y$ $\forall v \in [s, T^{\star}_{y,s})$. Forwards arrows point due north, backwards ones due south. In words, $T^{\uparrow}_{y,s}$ is the time of the first event (non-strictly) northwards of time s that affects the point y and $T^{\downarrow}_{y,s}$ is the time of the first event (non-strictly) southwards of time s that affects y. We illustrate these concepts in Figure 5.2. The corresponding event $(x,t,r) \in \Pi^n$ will be called the *finishing event* of $\alpha_{y,s}$. For each $\star \in \{\uparrow,\downarrow\}$, we can now associate two important sets of paths to each point (y,s).

Let us first consider the forwards paths. The set $\mathcal{P}^{\uparrow}_n(y,s)$ is best described in words; it is the set of paths that are obtained by following the arrow $\alpha^{\uparrow}_{y,s}$ out of (y,s) and then, every time we finish an arrow, following a new arrow that starts from one of the parents of the finishing event of $\alpha_{y,s}$. The forwards paths from (y,s) correspond precisely to the set of potential ancestral lineages of an individual who lived at the point y at time s.

Remark 5.10 *If parents of events were always at the centre of the event, then a forwards path would be precisely an exploration path in the sense of Micaux [2007]. These still differ from the paths in the Poisson trees of Ferrari et al. [2005] which are continuous (they interpolate between*

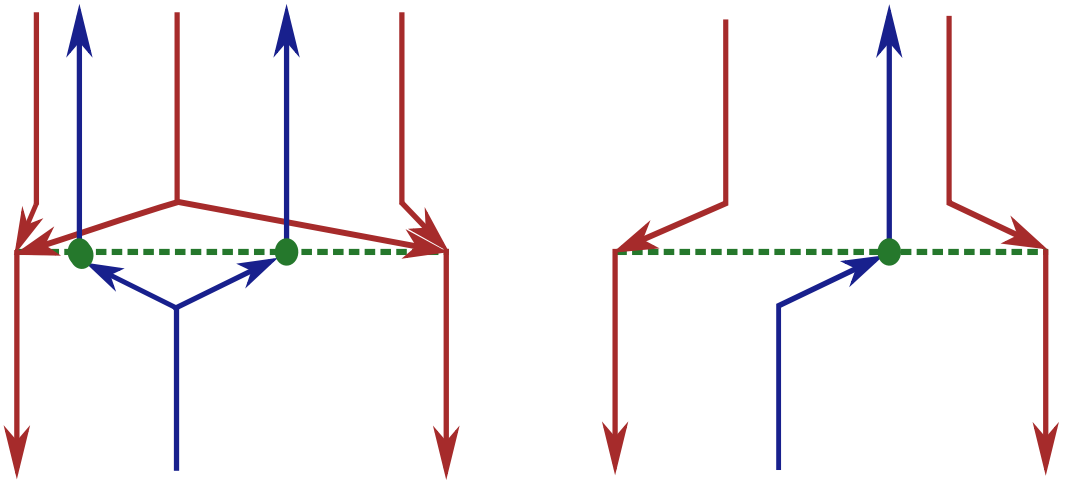

Figure 5.4: The movement of forwards and backwards paths (illustrated as interpolated arrows) about a selective event (left) and a neutral event (right). The events are shown as green dotted horizontal lines and the parent(s) as small circles. Forwards paths (in blue) travel northwards and backwards paths (in red) travel southwards.

the centres of events).

The set $\mathcal{P}_n^{\downarrow}(y,s)$ of backwards paths is also best described in words. It is the set of paths obtained by following the arrow $\alpha_{y,s}^{\downarrow}$ out of (y,s) and then, every time we finish an arrow $\alpha_{y',s'}^{\downarrow}$:

1. If the finishing event of $\alpha_{y',s'}$ is neutral with parent v then
   (a) if $y' \leqslant v$, follow the arrow out of the west-most point of the finishing event of $\alpha_{y',s'}$,
   (b) if $y' > v$, follow the arrow out of the east-most point of the finishing event of $\alpha_{y',s'}$.
2. If the finishing event of $\alpha_{y',s'}$ is selective with parents $v < v'$ then
   (a) if $y' < v$, follow the arrow out of the west-most point of the finishing event of $\alpha_{y',s'}$,
   (b) if $y' > v'$, follow the arrow out of the east-most point of the finishing event of $\alpha_{y',s'}$,
   (c) if $y' \in [v, v']$, a path can follow either one of the arrows out of the east-most/west-most points of the finishing event of $\alpha_{y',s'}$.

See Figure 5.4 for an illustration of the forwards and backwards paths.

In keeping with our previous notation, for each path we write $\sigma(\alpha) = \sigma_\alpha = s$, so that σ associates to each forwards/backwards path the time at which it starts.

5.2.2 Interpolated paths and arrows

We wish to exploit the existing theory of Brownian webs and nets, which was developed in a setting restricted to continuous paths, and so we will approximate the systems of (càdlàg) forwards and backwards paths of the last subsection by corresponding systems in which the jumps have been interpolated. Ferrari *et al.* [2005] achieve this by simply taking paths that interpolate between the starting points of arrows. However, in our situation this would result in arrows which cross each other and, worse, pass through events that did not previously affect them. Instead, we adopt a 'just in time' approach to our interpolation: we find small non-overlapping intervals of time and space about each event in which to interpolate. The following lemma makes this precise.

Lemma 5.11 *Let $n \in \mathbb{N}$. For each $p = (x, t, r) \in \Pi^n$ and each $\epsilon > 0$ define the set*

$$\mathscr{B}_\epsilon(x, t, r) = \{(y, s) \in \mathbb{R}^2 \,;\, |x - y| \leqslant r, |t - s| \leqslant \epsilon\}.$$

Almost surely, there exists a map $\Upsilon : \Pi^n \to (0, \infty)$ such that the sets $(\mathscr{B}_{\Upsilon(p)}(p))_{p \in \Pi_n}$ are distinct.

Proof: This follows immediately since Π^n has finite intensity and so the points of Π^n have no point of accumulation. ■

Let $f^\uparrow \in \mathcal{P}^\uparrow(y, s)$ and let $\alpha^\uparrow_{y,s} : [s, T^\uparrow_{y,s}) \to \mathbb{R}$ be the first of the forwards arrows that make up $f^\uparrow$. Suppose that α' is the next arrow in $f^\uparrow$ after $\alpha^\uparrow_{y,s}$. Let $p = (x, T^\uparrow_{y,s}, r)$ denote the finishing event of $\alpha^\uparrow_{y,s}$ and let $z = \alpha'(T^\uparrow_{y,s})$ be the starting point of the next arrow. We say that $\widetilde{\alpha}^\uparrow_{y,s} : [s, t) \to \mathbb{R}$ is the interpolated arrow of $\alpha^\uparrow_{y,s}$ if both

1. $\widetilde{\alpha}^\uparrow_{y,s}(v) = \alpha^\uparrow_{y,s}(v)$ for all $v \leqslant T^\uparrow_{y,s} - \Upsilon(p)$,
2. $\widetilde{\alpha}^\uparrow_{y,s}(v)$ is linear on $[T^\uparrow_{y,s} - \Upsilon(p), t)$ and $\lim_{v \uparrow t} \widetilde{\alpha}^\uparrow_{y,s}(v) = z$.

Note that the interpolation of an arrow depends on the path f in which it is contained.

Given a forwards path $f \in \mathcal{P}_n^{\uparrow}(y,s)$, we define the continuous path $\tilde{f}$ to be the concatenation of the interpolations of the arrows within f and we define $\widetilde{\mathcal{P}}_n^{\uparrow}(y,s) = \{\tilde{f}\,;\, f \in \mathcal{P}_n^{\uparrow}(y,s)\}$. The interpolated backwards paths $\widetilde{\mathcal{P}}_n^{\downarrow}(y,s)$ are defined in an entirely analogous way. Of course, interpolated paths are close to their equivalent non-interpolated paths.

Lemma 5.12 *Let $(y,s) \in \mathbb{R}^2$ and let $f \in \mathcal{P}_n^{\uparrow}(y,s)$. Then*

$$\sup_{t \in (\sigma_f, \infty)} |f(t) - \widetilde{f}(t)| < 2\mathcal{R}n^{-1/2}$$

and the analogous estimate holds for backwards paths.

Proof: Note that $\sigma_f = \sigma_{\widetilde{f}}$. By definition, $f(u) = \widetilde{f}(u)$ unless u is such that $(f(u),u) \in \mathscr{B}(x,t,r)$ for some $(x,t,r) \in \Pi^n$. By definition of $\widetilde{f}$, when $(f(u),u) \in \mathscr{B}(x,t,r)$ we have $|f(u) - \widetilde{f}(u)| \leqslant 2r$. By definition of μ^n we have $r \leqslant \mathcal{R}n^{-1/2}$, which completes the proof. ■

5.2.3 Left-most and right-most paths

We now associate to each (y,s) four special paths.

Definition 5.13 *Left-most and right-most forward and backward paths are defined as follows.*

1. *The left-most forward path from (y,s) is the element of $\mathcal{P}_n^{\uparrow}(y,s)$ obtained by choosing the (forwards) arrow with the west-most pre-parent, whenever a choice is available.*

2. *The right-most forward path from (y,s) is the element of $\mathcal{P}_n^{\uparrow}(y,s)$ obtained by choosing the (forwards) arrow with the east-most pre-parent, whenever a choice is available.*

3. *The left-most backward path from (y,s) is the element of $\mathcal{P}_n^{\downarrow}(y,s)$ obtained by choosing the (backwards) arrow from the east-most point of the finishing event whenever a choice is available.*

4. *The right-most backward path from (y,s) is the element of $\mathcal{P}_n^{\downarrow}(y,s)$ obtained by choosing the (backwards) arrow from the west-most point of the finishing event, whenever a choice is available.*

We will sometimes shorten 'left-most' and 'right-most' to l-most and r-most.

For each $\star \in \{\uparrow,\downarrow\}$, $\dagger \in \{l,r\}$ and $D \subseteq \mathbb{R}^2$ we define

$$\mathcal{P}_n^{\star,\dagger}(D) = \{f\,;\, f = f^{\star}_{y,s} \text{ is the } \dagger\text{-most path of some } (y,s) \in D\},$$

and $\widetilde{\mathcal{P}}_n^{\star,\dagger}(D) = \{\widetilde{f}\,;\, f \in \mathcal{P}^{\star,\dagger}(D)\}$ to be the corresponding set of interpolated paths. Additionally, we include the boundary paths

$$\mathcal{B} = \{f : f(x) = -\infty, x \in [\sigma_f,\infty], \sigma_f \in [-\infty,\infty]\} \cup \{f : f(x) = \infty, x \in [\sigma_f,\infty], \sigma_f \in [-\infty,\infty]\}$$

in both $\mathcal{P}^{\star,\dagger}(D)$ and $\widetilde{\mathcal{P}}^{\star,\dagger}(D)$. Since the usefulness of our state space relies on the compactness of $[-\infty,\infty]$, it is convenient to include ∞ in the domain of each $f \in \mathcal{P}^{\star,\dagger}(D)$ by defining $f(\infty) = 0$.

Definition 5.14 *Let f and f' be paths. We say f and f' cross at t if there exists $\epsilon > 0$ such that $f(s) \leqslant f'(s)$ for $s \in (t-\epsilon, t)$ and $f(s) > f'(s)$ for $s \in (t, t+\epsilon)$.*

Lemma 5.15 *Let D be any subset of $\mathbb{R}^2$ and let $n \in \mathbb{N}$. Let $\dagger \in \{l,r\}$. Then, almost surely, for all $f^{\uparrow} \in \mathcal{P}_n^{\uparrow,\dagger}(D)$ and $f^{\downarrow} \in \mathcal{P}_n^{\downarrow,\dagger}(D)$, the paths $f^{\uparrow}$ and $f^{\downarrow}$ do not cross.*

Further, the same result holds for interpolated paths $f^{\uparrow} \in \widetilde{\mathcal{P}}_n^{\uparrow,\dagger}(D)$ and $f^{\downarrow} \in \widetilde{\mathcal{P}}_n^{\downarrow,\dagger}(D)$.

Proof: Evidently paths can only cross if both are affected by a common event and then the conclusion is an easy consequence of the definitions (or Figure 5.4). ■

Remark 5.16 *A forwards left-most path and backwards right-most path can cross each other, as can a forwards right-most and a backwards left-most path. Similarly, a forwards left-most path can cross a forwards right-most path. These three statements also hold when forwards and backwards are interchanged.*

Although not immediately obvious from the definition, an important feature of our construction is that the backwards paths, when rotated through 180 degrees, have the same distribution as the forwards ones.

Lemma 5.17 *A forwards left- (right-) most path has the same distribution as a backwards left- (right-) most path which has been rotated through 180 degrees.*

PROOF: First observe that the rate at which an event falls on (an arrow in) a path has the same distribution whether we look forwards or backwards in time and, when an event falls on a (forwards or backwards) path, the spatial position of the path will be uniformly distributed over the region affected by the event. Let us denote that position by V. It will be convenient to measure spatial position relative to the west-most point of the event. Thus if the event corresponds to $p = (x, t, r)$, then V is uniformly distributed on $[0, 2r]$.

Consider first a left-most forwards path. If a neutral event falls on it, then the path jumps to the position of the pre-parent, which we denote by U, which is also uniformly distributed on the event. Thus, by symmetry, the path is equally likely to jump left or right and the length of the jump is the distance between U and V which, again by symmetry, has the same distribution as $\min(U, V)$ (on $[0, 2r]$).

Now consider a selective event. The two pre-parents are sampled uniformly from the event. We denote their positions by $U_1 < U_2$. We now have three uniformly distributed random variables on $[0, 2r]$. Let us write them in ascending order as $U^{(1)}$, $U^{(2)}$, $U^{(3)}$.

The left-most forward path jumps to U_1. Let us write J for the jump that it has made. Each of the following events has probability $1/3$:

1. $V = U^{(1)}$, in which case, necessarily, $U_1 = U^{(2)}$ and $J = U^{(2)} - U^{(1)} \stackrel{d}{=} U^{(1)}$;

2. $V = U^{(2)}$, in which case $J = U^{(1)} - U^{(2)} \stackrel{d}{=} -U^{(1)}$;

3. $V = U^{(3)}$, in which case $J = U^{(1)} - U^{(3)} \stackrel{d}{=} -U^{(2)}$.

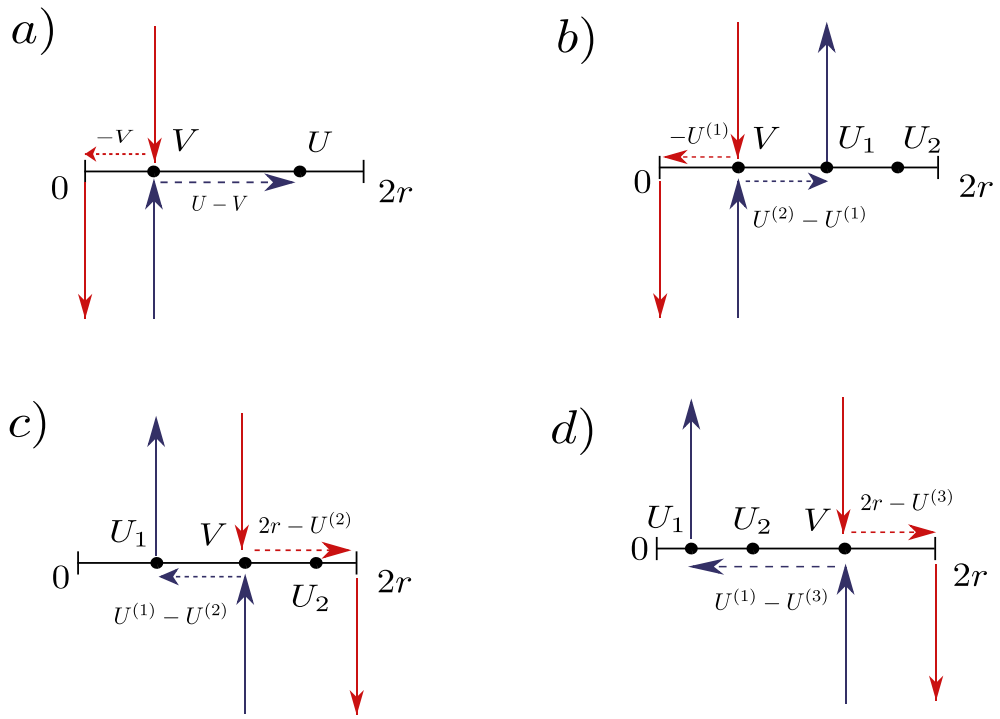

Figure 5.5: Jump distribution of forward and backwards paths. In a) we illustrate the situation in a neutral event. If $V < U$, the left-most forwards path jumps $U - V \stackrel{d}{=} \min(U, V)$ while left-most backwards path jumps $-V \stackrel{d}{=} -\min(U, V)$, which has the same distribution after rotating through 180 degrees. Pictures b), c) and d) deal with a selective event. Each one illustrates one of the three possibilities: the position of the path is to the left of both parents, in between, or to the right. Each is equally likely, and in the case of the left-most forward path we see that it jumps $U^{(2)} - U^{(1)} \stackrel{d}{=} U^{(1)}$, $U^{(1)} - U^{(2)} \stackrel{d}{=} -U^{(1)}$ or $U^{(1)} - U^{(3)} \stackrel{d}{=} U^{(2)}$, while the left-most backwards path jumps are given by $-U^{(1)}$, $2r - U^{(1)} \stackrel{d}{=} U^{(2)}$ and $2r - U^{(3)} \stackrel{d}{=} U^{(1)}$, and so we see that the backwards path has the same jump distribution, but rotated thorugh 180 degrees.

Now consider a left-most backwards path. Retaining the notation above, at a neutral event, with probability $1/2$, $V < U$ and the path jumps to the west-most endpoint, which, once rotated through 180 degrees becomes a positive jump of size $\min(U, V)$. On the other hand if $V > U$, which also has probability $1/2$, the path jumps to the east-most endpoint, which upon rotation becomes a negative jump with (by symmetry) the same distribution as $\min(U, V)$.

At a selective event, again each with probability $1/3$,

1. $V = U^{(1)}$ and the path jumps to the west-most endpoint of the interval, a jump whose rotation through 180 degrees has distribution $U^{(1)}$;

2. $V = U^{(2)}$, in which case the left-most path jumps to the east-most point of the interval, a jump which when rotated has distribution $-U^{(2)}$;

3. $V = U^{(3)}$ in which case the path jumps to the east-most point of the interval, a jump which when rotated has distribution equal to that of $-U^{(1)}$ (by symmetry).

Thus the left-most forwards path and the rotated left-most backwards paths have the same distribution and a symmetric argument yields the same result for right-most paths. See Figure 5.5. ■

5.3 Convergence to the left-right web

Most of the work that lies ahead is in proving convergence of our spaces of left-most and right-most paths to a left-right Brownian web. To show convergence of the spaces of left-most paths (and, by symmetry right-most paths) to a Brownian web, we shall apply a result due to Fontes *et al.* [2004], restated in convenient form in Ferrari *et al.* [2005]. To state this result, we first need some notation.

For a set X of paths in $\widetilde{M}$, let $\eta^X(t_0, t, a, b)$ denote the number of distinct points in $\{t_0+t\}\times\mathbb{R}$ that are touched by paths in X which also pass through $\{t_0\} \times [a, b]$ (Figure 5.6).

We denote the Borel σ-algebra of $\mathcal{K}(\widetilde{M})$ by $\mathcal{F}_\mathcal{K}$.

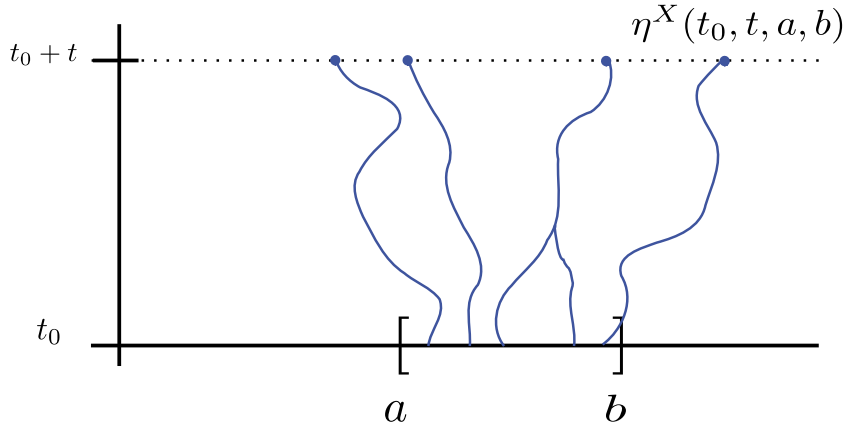

Figure 5.6: $\eta^X(t_0, t, a, b)$.

Theorem 5.18 (Ferrari *et al.* [2005], Theorem B) *Let $\mathcal{D}$ be a countable dense set of points in $\mathbb{R}^2$ (time-space). Suppose that $X_1, X_2, \ldots$ are $(\mathcal{K}(\widetilde{M}), \mathcal{F}_{\mathcal{K}})$-valued random variables with non-crossing paths. If, in addition, the following three conditions hold, the distribution of X_n converges to the distribution of the standard Brownian web.*

1. *There exist $\theta_n^y \in X_n$ such that for any deterministic $y_1, \ldots, y_m \in \mathcal{D}$, $\theta_n^{y_1}, \ldots, \theta_n^{y_m}$ converge in distribution as $n \to \infty$ to coalescing Brownian motions starting at $y_1, \ldots, y_m$.*

2. $$\limsup_{n\to\infty} \sup_{(a,t_0)\in\mathbb{R}^2} \mathbb{P}[\eta^{X_n}(t_0, t; a, a+\epsilon) \geqslant 2] \to 0 \text{ as } \epsilon \downarrow 0;$$

3. $$\frac{1}{\epsilon} \limsup_{n\to\infty} \sup_{(a,t_0)\in\mathbb{R}^2} \mathbb{P}[\eta^{X_n}(t_0, t; a, a+\epsilon) \geqslant 3] \to 0 \text{ as } \epsilon \downarrow 0.$$

Condition 1 is an easy modification of Lemma 4.1 of Berestycki *et al.* [2013] and we do not verify it here. Conditions 2 and 3 will be verified for our left-most paths in §5.3.3. Compactness of $\widetilde{\mathcal{P}}^{\uparrow,l}(\mathcal{D}_n)$, so that $\widetilde{\mathcal{P}}^{\uparrow,l}(\mathcal{D}_n) \in (\mathcal{K}(\widetilde{M}), \mathcal{F}_{\mathcal{K}})$, is deferred to §5.3.4, where we combine these results to deduce convergence to the left-right web, first for our interpolated paths and then for our original system.

In §5.3.1-§5.3.2 we shall verify that finite collections of our left and right-most paths will satsify Condition 2 of Theorem 5.4.

5.3.1 Convergence of pairs of paths

To prove convergence of spaces of left/right-most paths to the left-right web, we need to verify Condition 2 of Theorem 5.4. As a first step, in this subsection we take the limit of a pair of paths, comprising one left-most path and one right-most path started at some time s (which, since Π^n is homogeneous in both space and time, we may, without loss of generality, take to be zero) and show that it satisfies the system (5.1). With this in mind, let L^n and R^n denote respectively the left-most and right-most forwards paths associated to the points $(y^{n,l}, 0)$ and $(y^{n,r}, 0)$. We assume that the sequences of starting points converge to $(y^l, 0)$ and $(y^r, 0)$ respectively.

Following Sun and Swart [2008], and using their Lemma 2.2, there is a one-to-one correspondence between weak solutions of (5.1) and solutions of the system

$$dL_s = \xi dB^l_{S_s} + \xi dB^c_{C_s} - \zeta ds, \tag{5.7}$$

$$dR_s = \xi dB^r_{S_s} + \xi dB^c_{C_s} + \zeta ds, \tag{5.8}$$

$$s = S_s + C_s, \tag{5.9}$$

$$0 = \int_0^s \mathbb{1}\{L_s < R_s\} dC_s, \tag{5.10}$$

where B^l, B^r and B^c are independent standard one dimensional Brownian motions. The positive constants ξ and ζ depend on μ and will be specified below in (5.14) and (5.15) respectively. The solution (L_s, R_s) to this system is a $C_{\mathbb{R}^2}[0,\infty)$ valued process.

Proposition 5.19 *Let $T \in (0,\infty)$. As $n \to \infty$, $(L^n_s, R^n_s)_{s\in[0,T]}$ converges weakly to $(L_s, R_s)_{s\in[0,T]})$ in the sense of $D_{\mathbb{R}}[0,T]$ valued processes.*

In fact we shall need the analogue of this result for the corresponding interpolated paths, which we denote by $(\widetilde{L}^n, \widetilde{R}^n)$:

Corollary 5.20 *Let $T \in (0,\infty)$. As $n \to \infty$, $(\widetilde{L}^n_s, \widetilde{R}^n_s)_{s\in[0,T]}$ converges weakly to $(L_s, R_s)_{s\in[0,T]})$ in the sense of $C_{\mathbb{R}}[0,T]$ valued processes with the supremum norm.*

Proof: By Lemma 5.12, the weak convergence of Proposition 5.19 also holds (in $D_{\mathbb{R}}[0,T]$)

when (L^n, R^n) are replace by $(\widetilde{L}^n, \widetilde{R}^n)$. Since the supremum topology on continuous paths is continuously embedded in the Skorohod topology, it follows that this same convergence holds in $C_{\mathbb{R}}[0,T]$. ■

The remainder of Subsection 5.3.1 is devoted to the proof of Proposition 5.19. We begin by breaking down the evolution of the pair (L^n_s, R^n_s) into several different pieces. At time $s \geqslant 0$, we say L^n_s and R^n_s are

$$\begin{aligned} \textit{coalesced} \ \text{ if } L^n_s &= R^n_s, \\ \textit{nearby} \ \text{ if } L^n_s \neq R^n_s \text{ and } |L^n_s - R^n_s| &\leqslant \frac{2\mathcal{R}}{n^{1/2}}, \\ \textit{separated} \ \text{ if } |L^n_s - R^n_s| &> \frac{2\mathcal{R}}{n^{1/2}}. \end{aligned}$$

For $s \geqslant 0$ we set

$$\begin{aligned} C^n_s &= \int_0^s \mathbb{1}\{L^n_u, R^n_u \text{ are coalesced}\} du, \\ N^n_s &= \int_0^s \mathbb{1}\{L^n_u, R^n_u \text{ are nearby}\} du, \\ S^n_s &= \int_0^s \mathbb{1}\{L^n_u, R^n_u \text{ are separated}\} du, \end{aligned} \tag{5.11}$$

and we note that $C^n_s + N^n_s + S^n_s = s$.

Remark 5.21 *Here the symbols C, N, S correspond respectively to the terms coalesced, nearby and separated. Note that this does not match the notation used in Sun and Swart [2008].*

We define sequences of stopping times to track the changes of state of (L^n, R^n) during $[0,T]$. Firstly,

$$\begin{aligned} \tau^{n,C}_1 &= \inf\{s \geqslant 0\,;\, (L^n_s, R^n_s) \text{ are coalesced}\}; \\ \tau^{n,C}_k &= \inf\{s \geqslant \tau^{n,C}_{k-1}\,;\, (L^n_s, R^n_s) \text{ are coalesced, } (L^n_{s-}, R^n_{s-}) \text{ are not coalesced}\}. \end{aligned}$$

Similarly, we define sequences $\tau^{n,N}_k$ and $\tau^{n,S}_k$ for 're-entrance' times of (L^n_r, R^n_s) to the states of 'nearby' and 'separated' respectively. It is easily seen that each $\tau^{n,C}_k, \tau^{n,N}_k, \tau^{n,S}_k$ is a stopping

time and (L^n, R^n) is strong Markov.

Each jump of (L^n, R^n) is caused by one or both lineages being affected by a single event of Π^n. If (L^n, R^n) is coalesced immediately before this event then the event affects both L^n and R^n, whereas if they are separated the event affects only one of the two. The motion is more complicated when (L^n, R^n) is in the nearby state, when events can affect one or both lineages, but we shall see that the time spent in that state is negligible as we pass to the limit.

In order to identify the limiting objects, it is convenient to isolate the parts of the motion that contribute to the drift from those that contribute to the martingale terms in (5.7) and (5.8). The decomposition we make is not unique. Our particular choice highlights the fact that the martingale part of the motion of lineages is driven by neutral events, while the drift can be attributed to selection.

First we are going to define three random walks, from which we can build (L^n, R^n) when we are in the coalesced or separated states. To see what they should be, first suppose that a lineage is hit by a neutral event. When this happens, the position, y, of the lineage is uniformly distributed on the region affected by the event and it will jump to the position z of the parent, which is also uniformly distributed on the region. Neutral events fall according to a Poisson Point Process with intensity

$$n^{1/2}dx \otimes n(1-\mathbf{s}_n)dt \otimes \mu^n(dr),$$

so they hit y at rate

$$\begin{aligned} K_n &= n(1-\mathbf{s}_n)\int_{-\infty}^{\infty}\int_0^R\int_{-r}^{r} \mathbb{1}\{y \in \mathcal{B}_r(x)\}\, dx\, \mu^n(dr)\, n^{1/2}dx \\ &= 2n(1-\mathbf{s}_n)\int_0^R r\mu(dr). \end{aligned} \tag{5.12}$$

Thus we define V^n to be a symmetric random walk driven by a Poisson Point Process with intensity

$$n(1-\mathbf{s}_n)dt \otimes 2r\mu(dr).$$

At an event (t, r), the walk jumps with displacement $J_1/\sqrt{n}$ where

$$\mathbb{P}[J_1 \in A] = \mathbb{P}[Z_r - U_r \in A], \tag{5.13}$$

where U_r and Z_r are independent uniform random variables on $[-r, r]$.

Now consider the motion due to selective events. If the pair is coalesced immediately before the event, then their position is uniformly distributed on the affected region and the left-most path will jump with displacement $z_1 - y$ and the right-most path jumps with displacement $z_2 - y$ where $z_1 < z_2$ are the (uniformly distributed) positions of the two pre-parents of the event. If the pair (L^n, R^n) is separated, then only one of them will be affected by any given event. Selective events fall with intensity

$$n^{1/2} dx \otimes n\mathbf{s}_n dt \otimes \mu^n(dr)$$

and so we define the random walk $(D^{n,-}, D^{n,+})$, whose jumps are driven by a Poisson Point Process with intensity

$$n\mathbf{s}_n dt \otimes 2r\mu(dr).$$

At an event (t, r), $D^{n,-}$ jumps with displacement $(Z_1 - Y)/\sqrt{n}$ and $D^{n,+}$ jumps with displacement $(Z_2 - Y)/\sqrt{n}$ where $Z_1 = \min\{U_1, U_2\}$ and $Z_2 = \max\{U_1, U_2\}$ with U_1, U_2 and Y independent uniformly distributed random variables on $[0, 2r]$.

Lemma 5.22 *As $n \to \infty$, V^n converges weakly to ξB where B is a standard Brownian motion and*

$$\xi^2 = \frac{4}{9} \int_0^{\mathcal{R}} r^3 \mu(dr). \tag{5.14}$$

PROOF: Evidently J_1 has mean zero and, conditional on r its variance is $4r^2$ times that of the minimum of two independent uniform random variables on $[0, 1]$, which is $2r^2/9$. The lemma now follows from the Functional Central Limit Theorem (see, for example, Ethier and Kurtz [1986], Section 7.1). ■

Now consider $D^{n,\pm}$.

Lemma 5.23 *Let $T > 0$. As $n \to \infty$, $(D^{n,-}, D^{n,+})$ converges weakly to the deterministic process $s \mapsto (-\zeta s, \zeta s)$ where*

$$\zeta = \frac{2}{3} \int_0^{\mathcal{R}} r^2 \mu(dr). \tag{5.15}$$

PROOF: These walks experience jumps of size $\mathcal{O}(1/\sqrt{n})$ at rate

$$2n\mathbf{s}_n \int_0^{\mathcal{R}} r\mu(dr) = 2\sqrt{n} \int_0^{\mathcal{R}} r\mu(dr), \tag{5.16}$$

which is proportional to $\sqrt{n}$. In the notation above, conditional on r,

$$\mathbb{E}[Z_2 - Y] = -\mathbb{E}[Z_1 - Y] = \frac{r}{3}.$$

By the functional law of large numbers (see for example Rackauskas and Suquet [2013]) as $n \to \infty$, $(D^{n,-}, D^{n,+})$ converges weakly to the deterministic process

$$s \mapsto (-\zeta s, \zeta s),$$

with ζ as in the statement of the lemma. ■

When (L^n, R^n) is coalesced, its jumps have the same distribution as $(V^n + D^{n,-}, V^n + D^{n+})$. When (L^n, R^n) is separated, its jumps have the same distribution as $(V^{n,l} + D^{n,l,-}, V^{n,r} + D^{n,r,+})$ where $V^{n,l}, V^{n,r}$ are independent copies of V^n and $D^{n,l,-}$, $D^{n,r,+}$ are independent and with the same distribution as $D^{n,-}$, $D^{n,+}$ respectively. When L^n and R^n are nearby the evolution is more complicated; in fact in this case the joint jump distribution depends on $|L^n - R^n|$. Happily, because of Lemma 5.26 (see below) we will not need to describe the evolution in this case explicitly and we will denote it simply by $(\mathcal{N}_s^{n,l}, \mathcal{N}_s^{n,r})$.

Since (L^n, R^n) is always in exactly one of the states 'coalesced', 'nearby', and 'separated', and using spatial and temporal homogeneity of Π^n, it follows from the above that we can represent the dynamics of (L^n, R^n) in terms of three independent copies of the triple $(V^n, D^{n\pm})$ which we denote $(V^{n,\alpha}, D^{n,\alpha,\pm})$ with $\alpha \in \{c, l, r\}$:

$$L_s^n = L_0^n + V_{S_s^n}^{n,l} + D_{S_s^n}^{n,l,-} + \mathcal{N}_{N_s^n}^{n,l} + V_{C_s^n}^{n,c} + D_{C_s^n}^{n,c,-} \tag{5.17}$$

$$R_s^n = R_0^n + V_{S_s^n}^{n,r} + D_{S_s^n}^{n,r,+} + \mathcal{N}_{N_s^n}^{n,r} + V_{C_s^n}^{n,c} + D_{C_s^n}^{n,c,+} \tag{5.18}$$

$$s = C_s^n + N_s^n + S_s^n \tag{5.19}$$

$$0 = \int_0^s \mathbb{1}\{R_u^n > L_u^n\} dC_u^n. \tag{5.20}$$

Of course the 'time variables' (C^n, S^n, N^n) are coupled with the random walks $V^{n,\alpha}$ and $D^{n,\alpha,\pm}$.

Our next task is to prove that the time spent in the 'nearby' state is negligible. We require two preliminary estimates on the time between changes of state. The first says that each visit to the nearby state lasts at most $\mathcal{O}(1/n)$ units of time.

The second estimate says that visits to the coalesced state last $\mathcal{O}(1/\sqrt{n})$ units of time. This will limit the possible number of such visits in the time interval $[0, T]$ to be $\mathcal{O}(\sqrt{n})$ and since, moreover, the number of visits to the nearby state before the pair visits the coalesced state is $\mathcal{O}(1)$, this in turn allows us to control the number of visits to the nearby state.

Lemma 5.24 *Let $k \in \mathbb{N}$ and let $\tau'_{k'}$ be the next state change after $\tau_k^{n,N}$. Then the random variables $(\tau'_{k'} - \tau_k^{n,N})_{k\in\mathbb{N}}$ are an i.i.d. sequence and there exists $A \in (0,\infty)$, not dependent on k, such that*

$$\mathbb{E}\left[\tau'_{k'} - \tau_k^{n,N}\right] \leqslant \frac{A}{n}.$$

Further, there exists $q \in (0,1)$, not dependent on k, such that the probability that (L^n, R^n) is coalesced at $\tau'_{k'}$ is greater than q.

Proof: If (L^n, R^n) are nearby they must either be at a distance smaller than $\frac{3}{2}\mathcal{R}/n^{1/2}$, or at a distance between $\frac{3}{2}\mathcal{R}/n^{1/2}$ and $2\mathcal{R}/n^{1/2}$. In the first scenario, the probability that they coalesce through the the very next event that affects either of them is bounded away from 0. In the second scenario, the probability that the next event that affects either of them brings them closer than $\frac{3}{2}\mathcal{R}/n^{1/2}$ is also bounded away from 0. Since the rate at which events arrive is $\mathcal{O}(n)$ the result follows.

■

Lemma 5.25 *Let $k \in \mathbb{N}$ and let $\tau''_{k''}$ be the next state change after $\tau^{n,C}_k$. Then the random variables $(\tau''_{k''} - \tau^{n,C}_k)_{k\in\mathbb{N}}$ are an i.i.d. sequence and there exists $A' \in (0,\infty)$, not dependent on k, such that*

$$\mathbb{E}\left[\tau''_{k''} - \tau^{n,C}_k\right] \geqslant \frac{A'}{\sqrt{n}}.$$

PROOF: This is trivial, since any jump out of the coalesced state is due to a selective event and the rate at which these occur is given by (5.16). ■

Lemma 5.26 *Fix $T > 0$ and let N^n_T denote the total time spent in the nearby state up to time T. Then $N^n_T \to 0$ in probability as $n \to \infty$.*

PROOF: The idea is simple. Since $C_T \leqslant T$, using Lemma 5.25, the number of visits to the coalesced state in $[0,T]$ has mean at most $T\sqrt{n}/A'$. But by Lemma 5.24, the expected number of visits to the coalesced state is at least q times the expected number of visits to the nearby state. Thus the expected number of visits to the nearby state is at most $T\sqrt{n}/(qA')$ and since, again by Lemma 5.24, each has expected duration at most A/n, $\mathbb{E}[N_T] \leqslant TA/(qA'\sqrt{n})$ and the result is proved on applying Markov's inequality. ■

Since L^n and R^n evolve as $V^n + D^{n,-}$ and $V^n + D^{n,+}$ respectively, both converge individually and so their joint law is tight. Moreover, just as in Sun and Swart [2008], since C^n, N^n and S^n are continuous increasing processes, with rate of increase bounded by one, their joint law is also tight. Evidently we now have that

$$(L^n, R^n, V^{n,l}, V^{n,r}, V^{n,c}, D^{n,l,-}, D^{n,r,+}, D^{n,c,-}, D^{n,c,+}, C^n, N^n, S^n)$$

is tight and by passing to a subsequence we may assume that it converges weakly to some limiting process

$$(L_s, R_s, \xi B^l_s, \xi B^r_s, \xi B^c_s, -\zeta s, \zeta s, -\zeta s, \zeta s, C_s, 0, S_s)_{s\geqslant 0},$$

where B^l, B^r and B^c are independent (by construction). By Skorohod's Representation Theorem, by passing to a further subsequence if necessary, we can assume that the convergence is

almost sure. We claim that the limit (L_s, R_s, C_s, S_s) then satsifies (5.7-5.10). Indeed, letting $n \to \infty$ in (5.17), (5.18) and (5.19) we obtain precisely (5.7), (5.8) and (5.9).

Obtaining (5.10) from (5.20) requires a little more work (because the function $x \mapsto \mathbb{1}\{x > 0\}$ is not continuous), but again we need only adapt the approach of Sun and Swart [2008]. For each $\delta > 0$ let ρ_δ be a continuous non-decreasing function such that $\rho_\delta(u) = 0$ for $u \in [0, \delta]$ and $\rho_\delta(u) = 1$ for $u \in [\delta, \infty)$. Using (5.20) we have

$$\begin{aligned} 0 = \int_0^T \mathbb{1}\{R_s^n > L_s^n\} dC_s^n &= \int_0^T \mathbb{1}\left\{R_s^n - L_s^n > \frac{2\mathcal{R}}{n^{1/2}}\right\} dC_s^n + \int_0^T \mathbb{1}\left\{R_s^n - L_s^n \in \left(0, \frac{2\mathcal{R}}{n^{1/2}}\right]\right\} dC_s^n \\ &\geqslant \int_0^T \rho_\delta(R_s^n - L_s^n) dC_s^n \geqslant 0, \end{aligned}$$

provided that $\delta \geqslant 2\mathcal{R}/\sqrt{n}$. For such n we thus have $\int_0^T \rho_\delta(R_s^n - L_s^n) dC_s^n = 0$ and letting $n \to \infty$ we obtain $\int_0^T \rho_\delta(R_s - L_s) dC_s = 0$ for all $\delta > 0$. Letting $\delta \to 0$ we obtain $\int_0^T \mathbb{1}\{R_s > L_s\} dC_s = 0$ which is (5.10).

This completes the proof of Proposition 5.19. An entirely analogous proof gives convergence of a pair of backwards right and left-most paths. That the constants ξ and ζ are unchanged follows from Lemma 5.17.

5.3.2 Multiple left/right paths

In this subsection we extend Proposition 5.19 to larger collections of left and right-most paths. Let $N \in \mathbb{N}$. Given a set

$$D = \{(y_i, s_i) \in \mathbb{R}^2 \,;\, i = 1, \ldots, N\} \tag{5.21}$$

of distinct points in $\mathbb{R}^2$ and a function

$$O : \{1, \ldots, N\} \to \{l, r\} \tag{5.22}$$

we can define the set $\mathcal{P}_n^\uparrow(D, O) = \{f_1^\uparrow, \ldots, f_N^\uparrow\}$ where $f_i^\uparrow$ is the $O(i)$-most forward path from (y_i, s_i) driven by events in Π^n.

Sun and Swart [2008], Section 2.2, construct a system of left-right coalescing Brownian motions started from the points of D. (Recall that two left-most paths coalesce on meeting.) We write

$$\mathcal{P}^{\uparrow}(D,O) = \{B^{\uparrow,(y_i,s_i)}\,;\, i=1,\ldots,N\} \tag{5.23}$$

which is a system of left-right coalescing Brownian motions, while $\mathcal{P}_n^{\uparrow}(D,O)$ is a system of (pre-limiting) paths. Note that $\mathcal{P}^{\uparrow}(D,O)$ and $\mathcal{P}_n^{\uparrow}(D,O)$ are random elements of the product space M^N, where N is the number of paths we are considering.

For each $i=1,\ldots,N$ let $d_i^n = (y_i^n, s_i^n) \in \mathbb{R}^2$ be such that $d_i^n \to d_i = (y_i, s_i) \in \mathbb{R}^2$. Set $D_n = \{d_i^n\,;\, i=1,\ldots,N\}$ and $D = \{d_i\,;\, i=1,\ldots,N\}$.

Lemma 5.27 *Let $N \in \mathbb{N}$ and let D_n, D, O be as above. Then $\mathcal{P}_n^{\uparrow}(D_n,O)$ converges weakly in M^N to $\mathcal{P}^{\uparrow}(D,O)$ as $n\to\infty$. Further, $\widetilde{\mathcal{P}}_n^{\uparrow}(D_n)$ converges weakly in $\widetilde{M}^N$ to $\mathcal{P}^{\uparrow}(D,O)$ as $n\to\infty$.*

Proof: The argument is very similar to the proof of Proposition 5.2 of Sun and Swart [2008] so we will only sketch it here. The construction of $\mathcal{P}^{\uparrow}(D,O)$ in Section 2.2 of Sun and Swart [2008] is an inductive construction that views $\mathcal{P}^{\uparrow}(D,O)$ as made up of several independent pieces consisting of segments of either single left-most paths, single right-most paths or a pair of left/right paths. The same inductive construction breaks down $\mathcal{P}_n^{\uparrow}(D,O)$ into corresponding pieces. The stopping times used in this construction are continuous functionals on M^N with respect to the law of independent evolutions of paths within each such piece, so the first part of the lemma follows from Proposition 5.19 and Lemma 5.7.

Lemma 5.12 combined with the argument in the proof of Corollary 5.20 allows us to show that $\widetilde{\mathcal{P}}_n^{\uparrow}(D,O)$ converges weakly to $\widetilde{\mathcal{P}}^{\uparrow}(D,O)$. ∎

5.3.3 Coalescence of paths

In order to apply Theorem 5.18, we need to control the rate of coalescence of left-most (resp. right-most) paths. By symmetry, we restrict ourselves to forwards left-most paths.

To ease notation, we shall write $\eta_n^l(t_0, t, a, b)$ for the number of distinct points in $\{t_0 + t\} \times \mathbb{R}$ that are touched by paths in $\mathcal{P}_n^{\uparrow,l}(\mathcal{D}_n)$ which also pass through $\{t_0\} \times [a, b]$. The key estimates that we require are contained in the following two Lemmas. The proofs are adaptations of analogous results in Ferrari *et al.* [2003], who consider the neutral case with μ a point-mass δ_r and the (single) parent z always at the centre x of the event (x, t, r).

Remark 5.28 *In fact, our proof of convergence to the Brownian net will first work with our continuous interpolated paths, and so we shall need these results for paths in $\widetilde{\mathcal{P}}_n^{\uparrow,l}(\mathcal{D}_n)$. However, from Lemma 5.12, those follow on replacing the a and b by $a - \mathcal{R}/\sqrt{n}$ and $b + \mathcal{R}/\sqrt{n}$ in our estimates.*

Lemma 5.29 *Let $t_0, t, a \in \mathbb{R}$. Then* $\lim_{\epsilon \downarrow 0} \limsup_{n \uparrow \infty} \mathbb{P}\left[\eta_n^l(t_0, t, a, a + \epsilon) \geqslant 2\right] = 0.$

Proof: Without loss of generality, since the system is homogeneous in both space and time, it will suffice to take $(a, t_0) = (0, 0)$. We write f_n^x for the left-most path constructed from Π^n, starting from the point $(x, 0)$.

Consider the leftmost paths f_n^0 and f_n^ϵ. Let $T_n^\epsilon = \inf\{t > 0\,;\, f_n^0(t) = f_n^\epsilon(t)\}$, so that $\{\eta_n^l(0, t, 0, \epsilon) \geqslant 2\} = \{T_n^\epsilon > t\}$. As $n \to \infty$ $(f_n^0, f_n^\epsilon, T_n^\epsilon)$ converges in law in $C_{[0,T]}(\mathbb{R})$ to $(Z^0, Z^\epsilon, T^\epsilon)$ where (Z^0, Z^ϵ) are a pair of coalescing Brownian motions with constant drift to the left (independent before coalescence) and $T^\epsilon = \inf\{t > 0\,;\, Z^0(t) = Z^\epsilon(t)\}$. In particular, $Z^\epsilon - Z^0$ is a Brownian motion (with diffusion constant $2\xi^2$) absorbed at 0 and so, by the reflection principle,

$$\begin{aligned}\limsup_{n \to \infty} \mathbb{P}\left[f_n^0(t) < f_n^\epsilon(t)\right] &\leqslant 1 - 2\int_\epsilon^\infty \frac{1}{\sqrt{4\pi\xi^2 t}} e^{-x^2/(4\xi^2 t)} dx \\ &\leqslant \sqrt{\frac{1}{t\xi^2\pi}}\epsilon. \end{aligned} \tag{5.24}$$

The result follows immediately. ■

Lemma 5.30 *Let $t_0, t, a \in \mathbb{R}$. Then* $\lim_{\epsilon \downarrow 0} \limsup_{n \uparrow \infty} \epsilon^{-1}\mathbb{P}\left[\eta_n^l(t_0, t, a, a + \epsilon) \geqslant 3\right] = 0.$

The proof of this result is more involved. As a preliminary step, we will need the following lemma. We retain the notation from the proof above.

Lemma 5.31 *Let $n \in \mathbb{N}$ and $0 \leqslant r_n \leqslant n^{-1/2}$. There exists $c_2 \in (0, \infty)$ such that for all $n \in \mathbb{N}$,*

$$\mathbb{P}\left[f_n^0(t) < f_n^{r_n}(t)\right] \leqslant \frac{c_2}{\sqrt{nt}}.$$

PROOF: We adapt the proof of Lemma 2.9 in Ferrari *et al.* [2003]. Using our previous terminology, whilst f_n^0 and $f_n^{r_n}$ are separated, they evolve independently of one another. Adapting the notation of Section 5.3.1, write $\tau^n = \tau_{k+1}^{n,N} - \tau_k^{n,N}$ for the random time between visits of the pair to the nearby state. Each can be bounded above by the time taken to enter the nearby state if we start from separation $2\mathcal{R}/\sqrt{n}$ and then the proof of Lemma 2.9 of Ferrari *et al.* [2005] (which is presented in full on Ferrari *et al.* [2003]) tells us (when we take into account our scaling of space and time) that

$$\mathbb{P}[\tau^n > t] \leqslant \frac{c}{\sqrt{nt}},$$

for a constant c.

The number of visits to the nearby state before coalescence can be bounded above by a geometric random variable, which we denote by K, with success probability q (a proof of this can be found in Lemma 5.24). Then

$$\begin{aligned}
\mathbb{P}\left[f_n^0(t) < f_n^{r_n}(t)\right] &\leqslant \sum_k \mathbb{P}\left[f_n^0(t) < f_n^{r_n}(t) | K = k\right] \mathbb{P}[K = k] \\
&\leqslant \sum_k k\mathbb{P}[\tau^n > t/k]\mathbb{P}[K = k] \\
&\leqslant \sum_k k\frac{c\sqrt{k}}{\sqrt{nt}}\mathbb{P}[K = k] = \frac{c_2}{\sqrt{nt}}.
\end{aligned}$$

■

The idea of the proof of Lemma 5.30 is to reduce its statement to one about pairs of paths, using an FKG inequality. Again we can adapt arguments from Ferrari *et al.* [2005]. Define a

partial order $\preceq$ on $D_{\mathbb{R}}[0,T]$ by

$$f \preceq g \text{ if and only if } f(t) - f(s) \leqslant g(t) - g(s) \text{ for all } t \geqslant s \geqslant 0 \tag{5.25}$$

and define increasing/decreasing events and functions with respect to this partial order in the usual way. We shall write μ_L^n for the law of the left-most path, L^n, started from $(0,0)$. The proof of Lemma 2.8 in Ferrari *et al.* [2003] immediately yields:

Lemma 5.32 *The law μ_L^n satisfies the FKG inequality. That is for any increasing events $A, B \subseteq D_{\mathbb{R}}[0,T]$,*

$$\mu_L^n(A \cap B) \geqslant \mu_L^n(A)\mu_L(B).$$

The next step is to find a notation for, and a bound on the size of, the set of all points in $[0,\epsilon] \times \{0\}$ through which at least one path of $P_n^{\uparrow,l}(\mathcal{D}_n)$ passes. It is the coalescence of the left-most paths from these points that we must control.

Let

$$I'_n = \{(x,0)\,;\, x \in [0,\epsilon], \exists f \in \mathcal{P}_n^{\uparrow,l}(\mathcal{D}_n), \sigma_f < 0, f(0) = x\}$$

be the set of points in $[0,\epsilon]\times\{0\}$ through which at least one path of $\mathcal{P}_n^{\uparrow,l}(\mathcal{D}_n)$ which starts before time 0 passes, and set

$$I_n = I'_n \cup \{(k/n^{-1/2}, 0)\,;\, k \in \mathbb{N}, 0 \leqslant k \leqslant \epsilon n^{1/2}\}.$$

Lemma 5.33 *In the notation above,*

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}} \mathbb{E}\left[|I_n|\right] \leqslant c_3\epsilon.$$

Proof: Evidently it suffices to control the number of points in $\widetilde{I}_n := I_n \backslash \{(k/n^{-1/2}, 0)\,;\, k \in \mathbb{N}, 0 \leqslant k \leqslant \epsilon n^{1/2}\}$. All these points correspond to paths f that have been hit by at least one event in the time interval $(\sigma_f, 0)$ and $f(0)$ is the position of the (left-most) parent in the most recent such event.

We define a tesselation of the interval $[0,\epsilon]$ by tracing due south from each point of the

interval until we first encounter an event of Π^n. Points 'descended' from the same event are in the same cell of the tesselation. This tesselation arises from the equilibrium distribution of what, in two dimensions, is known as the 'dead leaves tesselation', introduced in Matheron [1968] (The dead leaves process is defined as the sequential superposition of random sets, and takes its name from the fact that it resembles the process by which leaves will fall on and cover an area. For a more recent overview we refer to Bordenave *et al.* [2006]). In that language, we are concerned with the expected number of cells to intersect the interval $[0, \epsilon]$ (when the tesselation is at stationarity). Although a great deal is known about the tesselation, we were unable to find a reference for exactly the result needed here. However, under our assumptions there is a very simple proof.

Consider an interval of length r_0, where $\mu(3r_0, \mathcal{R}) > \delta$ for some $\delta > 0$. Tracing backwards in time, the time T_{r_0} that we must wait until an event of Π^1 completely covers this interval is an exponentially distributed random variable with rate bounded below by $5r_0\delta$. Conditional on T_{r_0}, the expected number of events of Π^1 which fall in $[-T_{r_0}, 0]$ and for which the region affected intersects our interval of length r_0 is bounded above by $T_{r_0} r_0 \int_0^{\mathcal{R}} r\mu(dr)$. This is certainly an upper bound on the number of cells of the tesselation that intersect the interval, which, on averaging over T_{r_0}, we see has mean bounded above by some constant A, say.

By scaling, we are interested in the number of cells of the partition corresponding to Π^1 that intersect an interval of length $\epsilon\sqrt{n}$. This is at most

$$\left(\frac{1}{r_0}\epsilon\sqrt{n} + 1\right) A \leqslant c_3\epsilon\sqrt{n}, \qquad \text{as } n \to \infty.$$

■

Proof: [Of Lemma 5.30]

We proceed exactly as in Ferrari *et al.* [2005]. Suppose that $I_n = \{(\xi_i^n, 0)\}_{i=1}^N$ and simplify notation by writing $f_n^i = f_n^{\xi_i}$ for the left-most path starting at $(\xi_i, 0)$. We write μ_i^n for the law of f_n^i, which is of course just a spatial translation of μ_L^n.

If $\eta_{I_n}^l(0, t, 0, \epsilon) \geqslant 3$ then there must be some $j \in \{1 \dots N-1\}$ such that $f_n^j(t) < f_n^{j+1}(t) <$

$f_n^N(t)$. Hence,

$$\begin{aligned}
&\mathbb{P}_{I_n}\left[\eta^l_{I_n}(0,t,0,\epsilon) \geqslant 3\right] \qquad (5.26)\\
&\leqslant \sum_{i=1}^{N-2} \mathbb{P}\left[f_n^i(t) < f_n^{i+1}(t) < f_n^N(t)\right]\\
&= \sum_{i=1}^{N-2} \int \mathbb{P}_{I_n}\left[f_n^i(t) < f_n^{i+1}(t) < f_n^N(t) \mid f_n^{i+1} = g\right] \mu_{i+1}^n(dg)\\
&\leqslant \sum_{i=1}^{N-2} \int \mathbb{P}_{I_n}\left[f_n^i(t) < f_n^{i+1}(t) \mid f_n^{i+1} = g\right] \mathbb{P}_{I_n}\left[f_n^{i+1}(t) < f_n^N(t) \mid f_n^{i+1} = g\right] \mu_{i+1}^n(dg) \qquad (5.27)\\
&\leqslant \sum_{i=1}^{N-2} \left(\int \mathbb{P}_{I_n}\left[f_n^i(t) < f_n^{i+1}(t) \mid f_n^{i+1} = g\right] \mu_{i+1}^n(dg)\right)\\
&\qquad\qquad \times \left(\int \mathbb{P}_{I_n}\left[f_n^{i+1}(t) < f_n^N(t) \mid f_n^{i+1} = g\right] \mu_i(dg)\right)\\
&= \sum_{i=1}^{N-2} \mathbb{P}_{I_n}\left[f_n^i(t) < f_n^{i+1}(t)\right] \mathbb{P}_{I_n}\left[f_n^{i+1}(t) < f_n^N(t)\right]\\
&\leqslant (N-2)\, \mathbb{P}_{I_n}\left[f_n^0(t) < f_n^{r_n}(t)\right] \mathbb{P}_{I_n}\left[f_n^0(t) < f_n^{\epsilon}(t)\right]
\end{aligned}$$

where $r_n \leqslant 1/\sqrt{n}$. To see the inequality in (5.27), notice that in contrast to the setting of Ferrari *et al.* [2005], knowing f_n^{i+1} tells us the times of the events that have affected the path, but gives only partial information about their left and right-most points. However, consider one such event, occurring at time τ, say. Partitioning over the size of the event, it is evident that the events $\{f_n^i(\tau) = f_n^{i+1}(\tau)\}$ and $\{f_n^{i+1}(\tau) = f_n^N(\tau)\}$ are negatively correlated, which implies the inequality.

To pass from the third line to the fourth we have applied the FKG inequality.

Now, applying Lemma 5.31, equation (5.24) and Lemma 5.33, we have

$$\begin{aligned}
\mathbb{P}_{I_n}\left[\eta^l_{I_n}(0,t,0,\epsilon) \geqslant 3\right] &\leqslant \mathbb{E}[N] \frac{c_2}{\sqrt{nt}} \frac{1}{\sqrt{\pi t \xi}} \epsilon\\
&\leqslant c_4 \epsilon^2,
\end{aligned}$$

for some constant c_4, and the result follows. ■

5.3.4 Compactness

Before we can apply Theorem 5.18, we must verify compactness.

Lemma 5.34 *Let $\dagger \in \{l, r\}$ and $n \in \mathbb{N}$. Then $\widetilde{\mathcal{P}}_n^{\uparrow,\dagger}(\mathcal{D}_n)$ and $-\widetilde{\mathcal{P}}_n^{\downarrow,\dagger}(\mathcal{D}_n)$ are compact subsets of $\widetilde{M}$.*

PROOF: As usual, without loss of generality we may assume $\star = \uparrow$ and $\dagger = l$. It follows easily from Lemma A.9 that $\widetilde{\mathcal{P}}_n^{\uparrow,l}(\mathcal{D}_n)$ is relatively compact, so it remains only to prove that $\widetilde{\mathcal{P}}_n^{\uparrow,l}(\mathcal{D}_n)$ is closed. If (f_m) is a sequence of elements of $\widetilde{\mathcal{P}}_n^{\uparrow,l}(\mathcal{D})$ converging to some $f \in \widetilde{M}$ then by (5.6) we have $f_m(\sigma_{f_m}) \to f(\sigma_f)$ in $[-\infty, \infty]$. If $\sigma_f = \infty$ then it is immediate that (the equivalence class of) f is an element of $\widetilde{M}$ so without loss of generality we may assume $\sigma_f < \infty$. Since $(f_m(\sigma_{f_m}), \sigma_{f_m}) \in \mathcal{D}_n$ and $\mathcal{D}_n$ is a set of isolated points we must have $f(\sigma_f) \in \{\infty, -\infty\}$; without loss of generality we assume that $f(\sigma_f) = \infty$. ■

Proposition 5.35 *Let $\star \in \{\uparrow, \downarrow\}$ and let $\dagger \in \{l, r\}$. Then as $n \to \infty$, $\widetilde{P}_n^{\star,\dagger}(\mathcal{D}_n)$ converge weakly in $\widetilde{\mathcal{K}}$ to a $\dagger$-$\star$-wards Brownian web which we denote by $\mathcal{W}^{\star,\dagger}$.*

PROOF: We have now checked the conditions of Theorem 5.18 and the result follows. ■

Proposition 5.36 *Let $\star \in \{\uparrow, \downarrow\}$. Then $(\mathcal{W}^{\star,l}, \mathcal{W}^{\star,r})$ is a left-right $\star$-wards Brownian web.*

PROOF: Without loss of generality it suffices to prove the case $\star = \uparrow$. We verify the conditions of Theorem 5.4. Conditions 1 and 3 follow from Proposition 5.35 and Condition 2 follows from Lemma 5.27. ■

5.4 Convergence to the Brownian Net

We are finally in a position to prove Theorem 5.8. Recall the set $\mathcal{D}_n$ defined in (5.5). Our goal is to prove convergence of $\widetilde{\mathcal{P}}_n^{\uparrow}(\mathcal{D}_n)$ (and consequently $\mathcal{P}_n^{\uparrow}(\mathcal{D}_n)$) to the Brownian net, weakly in $\mathcal{K}(\widetilde{M})$ (resp. in $\mathcal{K}(M)$). First we check that $\widetilde{\mathcal{P}}_n^{\uparrow}(\mathcal{D}_n)$ is a compact subset of $\widetilde{M}$.

Lemma 5.37 *Let $\star \in \{\uparrow, \downarrow\}$ and $n \in \mathbb{N}$. Then $\widetilde{\mathcal{P}}_n^{\uparrow}(\mathcal{D}_n)$ and $-\widetilde{\mathcal{P}}_n^{\downarrow}(\mathcal{D}_n)$ are compact subsets of $\widetilde{M}$.*

PROOF: Using Lemma A.9 from the appendix, a set $\widetilde{P}$ in $\widetilde{M}$ is precompact if and only if its modulus of continuity, defined by

$$m_{\widetilde{P}}(\delta) = \sup\left\{\left|\frac{\tanh \pi(t)}{1+|t|} - \frac{\tanh \pi(s)}{1+|s|}\right| : \pi \in \widetilde{P}; s, t \geqslant \sigma(\pi), |\tanh t - \tanh s| < \delta\right\}$$

tends to zero as $\delta \downarrow 0$. By the definition of interpolated arrows,

$$m_{\widetilde{\mathcal{P}}_n^{\uparrow}(\mathcal{D}_n)}(\delta) \leqslant \max\left(m_{\widetilde{\mathcal{P}}_n^{\uparrow,l}(\mathcal{D}_n)}(\delta), m_{\widetilde{\mathcal{P}}_n^{\uparrow,r}(\mathcal{D}_n)}(\delta)\right) \tag{5.28}$$

so $\widetilde{\mathcal{P}}^{\uparrow}(\mathcal{D}_n)$ is relatively compact by Lemma 5.34. The same argument as in the proof of Lemma 5.34 shows that $\widetilde{\mathcal{P}}^{\uparrow}(\mathcal{D}_n)$ is closed, hence it is compact. An analogous argument applies to $-\widetilde{\mathcal{P}}_n^{\downarrow}(\mathcal{D}_n)$. ■

5.4.1 Webs and nets

By Proposition 5.35, the sequence of quadruplets

$$\left(\widetilde{\mathcal{P}}_n^{\uparrow,l}(\mathcal{D}_n), -\widetilde{\mathcal{P}}_n^{\downarrow,l}(\mathcal{D}_n), \widetilde{\mathcal{P}}_n^{\uparrow,r}(\mathcal{D}_n), -\widetilde{\mathcal{P}}_n^{\downarrow,r}(\mathcal{D}_n)\right) \tag{5.29}$$

is tight in $\mathcal{K}(\widetilde{M})^4$. Let $(\mathcal{W}^{\uparrow,l}, -\mathcal{W}^{\downarrow,l}, \mathcal{W}^{\uparrow,r}, -\mathcal{W}^{\downarrow,r})$ be a quadruplet of $\mathcal{K}(\widetilde{M})$-valued random variables that are a weak limit point of this sequence. Note in particular that, by Lemma 5.35 this means the $(\mathcal{W}^{\star,\dagger})$ are coupled random variables and that the marginal distribution of $W^{\star,\dagger}$ is a $\dagger$-$\star$-wards Brownian web.

Lemma 5.38 *Let $\dagger \in \{l, r\}$. Then $\mathcal{W}^{\uparrow,\dagger}$ and $\mathcal{W}^{\downarrow,\dagger}$ are dual Brownian webs.*

PROOF: Without loss of generality we assume that $\dagger = l$. The result will follow once we have checked Condition 4 of Theorem 5.1. For any given $(y, s) \in \mathbb{R}^2$ there is almost surely a unique forwards path $g \in \mathcal{W}^{\uparrow,l}$ and we must prove that (almost surely) this is the unique element of M which starts at (y, s) and does not cross any path in $\mathcal{W}^{\downarrow,l}$.

Our first step towards this is to prove that, almost surely, any pair of paths $f^{\uparrow} \in \mathcal{W}^{\uparrow,l}$ and $f^{\downarrow} \in \mathcal{W}^{\downarrow,l}$ do not cross. To see this, suppose, for a contradiction, that $f^{\uparrow}$ and $f^{\downarrow}$ did cross. Without loss of generality, suppose that $f^{\uparrow}$ was west of $f^{\downarrow}$ immediately southwards of the crossing. Then there would exist $\epsilon > 0$, $t \in \mathbb{R}$ such that $f^{\uparrow}(s) < f^{\downarrow}(s)$ for $s \in [t-\epsilon, t)$ and $f^{\uparrow}(s) > f^{\downarrow}(s)$ for $s \in (t, t+\epsilon]$. Set

$$\delta = \frac{1}{3} \min \left\{ f^{\downarrow}(t-\epsilon) - f^{\uparrow}(t-\epsilon), f^{\uparrow}(t+\epsilon) - f^{\downarrow}(t+\epsilon) \right\}.$$

By choosing n sufficiently large, there exists paths $f_n^{\uparrow} \in \widetilde{\mathcal{P}}^{\uparrow,l}(\mathcal{D}_n, O^l)$ and $f_n^{\downarrow} \in \widetilde{\mathcal{P}}^{\downarrow,l}(\mathcal{D}_n, O^l)$ such that $\widetilde{d}(f^{\uparrow}, f_n^{\uparrow}) < \delta$ and $\widetilde{d}(f^{\downarrow}, f_n^{\downarrow}) < \delta$. It follows that $f_n^{\downarrow}(t-\epsilon) < f_n^{\uparrow}(t-\epsilon)$ and $f^{\uparrow}(t+\epsilon) > f_n^{\downarrow}(t+\epsilon)$, which implies that $f_n^{\uparrow}$ and $f_n^{\downarrow}$ cross; this contradicts Lemma 5.15.

Now, let $g^{\uparrow} \in M$ start at (y, s) and suppose that $g^{\uparrow}$ does not cross any path of $\mathcal{W}^{\downarrow,l}$. Let $f^{\uparrow} \in M$ be the (almost surely unique) path in $\mathcal{W}^{\uparrow,l}$ that starts at (y, s). Let $\epsilon > 0$, $m \in \mathbb{N}$ and let $(t_i)_{i=0}^m$ be a strictly increasing sequence of times with $t_0 = s$.

For each i, let $L_i^{\downarrow}$ be the (almost surely unique) path in $\mathcal{W}^{\downarrow,l}$ which starts at $(f^{\uparrow}(t_i) - \epsilon, t_i)$ and let $R_i^{\downarrow}$ be the (almost surely unique) path in $\mathcal{W}^{\downarrow,l}$ which starts at $(f^{\uparrow,l}(t_i) + \epsilon, t_i)$. Since g does not cross any of the paths $L_i^{\downarrow}, R_i^{\downarrow}$, for each $i \geqslant 1$ and all $t \in (t_{i-1}, t_i]$ we have

$$L_i^{\downarrow}(t) \leqslant g^{\uparrow}(t) \leqslant R_i^{\downarrow}(t), \tag{5.30}$$

(in Sun and Swart 2008 this type of construction is called a *fish-trap*).

Since (by the first part of this proof) $f^{\uparrow}$ also does not cross any of the $L_i^{\downarrow}, R_i^{\downarrow}$ we have (5.30) with $f^{\uparrow}$ in place of $g^{\uparrow}$. In particular, for all $i \geqslant 1$ we have

$$|f^{\uparrow}(t_i) - g^{\uparrow}(t_i)| \leqslant |L_i^{\downarrow}(t_i) - R_i^{\downarrow}(t_i)| = 2\epsilon.$$

Note also that trivially $f^{\uparrow}(t_0) = g^{\uparrow}(t_0)$. Since the sequence $(t_i)_0^m$ and $\epsilon > 0$ were arbitrary, by continuity of f and g we have $f = g$ (almost surely). ■

By Theorem 1.9 of Sun and Swart [2008] a Brownian web uniquely determines its dual (in the pathwise sense as well as in distribution) and, by Theorem 5.4 the distribution of a pair

of left-right webs is unique. Hence, by Lemmas 5.38 and 5.36 in fact the distribution of the $\mathcal{K}(M)^4$ valued random variable $(\mathcal{W}^{\uparrow,l}, -\mathcal{W}^{\downarrow,l}, \mathcal{W}^{\uparrow,r}, -\mathcal{W}^{\downarrow,r})$ does not depend on subsequence of (5.29) it represents; therefore (5.29) has a unique weak limit point. We abuse notation slightly by continuing to denote the (unique) weak limit point of (5.29) as

$$(\mathcal{W}^{\uparrow,l}, -\mathcal{W}^{\downarrow,l}, \mathcal{W}^{\uparrow,r}, -\mathcal{W}^{\downarrow,r}) \tag{5.31}$$

and noting that Lemmas 5.35, 5.38 and 5.36 apply to $\mathcal{W}^{\star,\dagger}$.

Let $\mathcal{N}$ denote the Brownian net associated to $(\mathcal{W}^{\uparrow,l}, \mathcal{W}^{\uparrow,r})$. We are now ready to show that $\widetilde{\mathcal{P}}_n^{\uparrow}(\mathcal{D}_n)$ converges weakly to $\mathcal{N}$ in $\mathcal{K}(\widetilde{M})$. We will then complete the argument by using Lemma 5.12 to show that this implies $\mathcal{P}_n^{\uparrow}(\mathcal{D}_n)$ converges to $\mathcal{N}$ in $\mathcal{K}(M)$.

Lemma 5.39 *The sequence $\widetilde{\mathcal{P}}_n^{\uparrow}(\mathcal{D}_n)$ converges weakly to $\mathcal{N}$ in $\widetilde{\mathcal{K}}$.*

PROOF: The argument is essentially the same as the proof of Theorem 5.4 of Sun and Swart [2008] and we will explain only the steps which are (slightly) different.

Note that tightness of $\widetilde{\mathcal{P}}_n^{\uparrow}(\mathcal{D}_n)$ in $\mathcal{K}(\widetilde{M})$ follows from (5.28) and tightness of $(\widetilde{\mathcal{P}}^{\uparrow,l}(\mathcal{D}_n), \widetilde{\mathcal{P}}^{\uparrow,r}(\mathcal{D}_n))$, which was proved in Lemma 5.35. Let $\mathcal{N}^*$ denote some weak limit point of this sequence.

Let $\mathcal{D}^l$ and $\mathcal{D}^r$ be dense countable subsets of $\mathbb{R}^2$. For each $z^l \in \mathcal{D}^l$ and $z^r \in \mathcal{D}^r$ we choose sequences $(z_n^l), (z_n^r) \subseteq \mathbb{R}^2$ such that $z_n^\dagger \in \mathcal{D}_n$ and $z_n^l \to z^l$, $z_n^r \to z^r$. Let l_n^z (resp. r_n^z) denote the leftmost (resp. rightmost) path, made up of arrows corresponding to events in Π^n, from z_n^l (from z_n^r). Given any two paths π, π' we define

$$\tau(\pi, \pi') = \sup\{t < \max(\sigma_\pi, \sigma_{\pi'})\,;\, \pi(t) = \pi'(t)\}.$$

By (5.31), and the Skorohod Representation Theorem we can pass to a subsequence and assume that

$$\begin{aligned}\Big(\widetilde{\mathcal{P}}_n^{\uparrow}(\mathcal{D}_n), \widetilde{\mathcal{P}}_n^{\uparrow,l}(\mathcal{D}_n), -\widetilde{\mathcal{P}}_n^{\downarrow,l}(\mathcal{D}_n), \widetilde{\mathcal{P}}_n^{\uparrow,r}(\mathcal{D}_n), -\widetilde{\mathcal{P}}_n^{\downarrow,r}(\mathcal{D}_n), (\tau(l_z^n, r_{z'}^n))_{z\in\mathcal{D}^l, z'\in\mathcal{D}^r}\Big) \\ \to \Big(\mathcal{N}^*, \mathcal{W}^{\uparrow,l}, -\mathcal{W}^{\downarrow,l}, \mathcal{W}^{\uparrow,r}, -\mathcal{W}^{\downarrow,r}, (\tau(l_z, r_{z'}))_{z\in\mathcal{D}^l, z'\in\mathcal{D}^r}\Big)\end{aligned} \tag{5.32}$$

almost surely. By Lemma 5.38, for each $\dagger \in \{l, r\}$ we have that $(\mathcal{W}^{\uparrow,\dagger}, \mathcal{W}^{\downarrow,\dagger})$ is a double Brownian web. By Lemma 5.36 we have that for each $\star \in \{\uparrow, \downarrow\}$, $(\mathcal{W}^{\star,l}, \mathcal{W}^{\star,r})$ is a left-right Brownian web. From this point on, noting that (5.32) is the equivalent in our setting of equation (5.16) in Sun and Swart [2008], the proof is identical to the proof of Theorem 5.4 of Sun and Swart [2008] and establishes that $\mathcal{N}^*$ is equal to $\mathcal{N}$ in distribution. ■

Finally, we deduce Theorem 5.8, that is $\mathcal{P}_n^{\uparrow}(\mathcal{D}_n)$ converges weakly to $\mathcal{N}$ in $\mathcal{K}(M)$.

PROOF: [of Theorem 5.8] The result follows from Lemma 5.39 by noticing that there is a one to one correspondence of paths in $\mathcal{P}_n^{\uparrow}(\mathcal{D}_n)$ to paths in $\widetilde{\mathcal{P}}_n^{\uparrow}(\mathcal{D}_n)$, and thus we can bound the Hausdorff distance between $\mathcal{P}_n^{\uparrow}(\mathcal{D}_n)$ and $\widetilde{\mathcal{P}}_n^{\uparrow}(\mathcal{D}_n)$ by $2\mathcal{R}/\sqrt{n}$ (by Lemma 5.12). ■

CHAPTER 6

Simulations

In the previous two chapters we obtained convergence results for the genealogical process $(\Xi_t^n)_{t\geqslant 0}$ in one and two dimensions. In the case of two-dimensional space, we showed that the (potential) ancestral lineages converge to a system of branching Brownian motions, while in the one-dimensional case we showed that the process is embedded in the Brownian net, and the genealogy of a sample behaves as what Sun and Swart [2008] term the 'branching and coalescing point set'. But each of these results came with a caveat. In the case of two dimensions, the rate of convergence we obtained is rather slow, on the order of $\mathcal{O}(\log(n))$. This is a by-product of the way we approached the problem, which although it seems rather natural, might not give the best bound on this rate. In the one-dimensional case, we only considered what happens when

the impact parameter u is 1, which allowed us to make use of the machinery of the Brownian net. Here we will address these issues by running simulations of the genealogical process in order to better understand its behaviour. It will be a rather informal approach, but it should be considered an exploratory analysis first and foremost. Before proceeding to the numerical results and analysis, we want to give a quick description of the algorithms employed for our simulations.

6.1 Algorithms

The updating mechanism of our process is fully given by the Poisson point process Π^n, and so in order to simulate the process our first concern will be to simulate Π^n properly. When simulating $(\Xi_t^n)_{t\geqslant 0}$, we could begin by considering some large region $A \in \mathbb{R}^d$ and then simulating the restriction of Π^n to A, but if the region is very large we would incur a very large cost by sampling many events that do not affect our sample. On the other hand, if A is too small we run the risk of losing track of potential ancestors that jump outside of the region.

Since we know that only events that fall within $2\mathcal{R}_n$ of any of the lineages present in the sample have a chance of hitting the lineages, it makes sense to restrict our attention to this region, which of course will also evolve with time. If $\Xi_t^n = \sum_{i=1}^{\mathcal{N}_t} \delta_{\xi_t^i}$ we can identify this region as:

$$A_t := \bigcup_{0\leqslant i\leqslant \mathcal{N}_t} \mathcal{B}_{2\mathcal{R}_n}(\delta_{\xi_t^i}).$$

Although it is simple to obtain and keep track of this region in 1 dimension, as soon as we consider two or more dimensions it becomes a very complicated problem.

One way around it, which we will use to simulate the two dimensional case, is to use what Kelleher *et al.* [2014] termed 'pixels'. The idea is rather simple. Divide the space into unit pixels of length 1 (they will be squares, cubes or hypercubes depending on the dimension). For every lineage in the sample, we can keep track of its neighbouring pixels - that is, those pixels close enough that an event that lands on them might affect the lineage. We can restrict ourselves

to sampling events from within the pixels, and since they do not intersect, the volume of the collection of pixels is simply its number. Of course, we will still sample events that cannot affect the lineages, but the number of such events will be considerably smaller than if we were sampling from a very large region. The algorithm becomes considerably more efficient.

We would like to point out that though very similar to the algorithm of Kelleher *et al.* [2014], our code for the two dimensional case differs from it in two meaningful ways. The first one is that we incorporate selective events, allowing the genealogy to branch. The second one is that instead of considering a bounded domain and constructing a sparse matrix of pixels, we only keep a list of occupied pixels. This way we do not need to define a priori the region where the process evolves. Since we are interested in rescaling limits, this is rather useful since we want the process to evolve for very long times and be able to visit distant regions.

Algorithm 1 *To simulate $(\Xi_t^n)_{t\geqslant 0}$ we will keep track of two lists, one is the sample, which just includes the location of the potential lineages, and the second one encodes its neighbourhood. In one dimension the neighbourhood is a collection of intervals, while for dimension 2 it is a collection of occupied pixels. Remember from the definition of the SΛFV process that events are given by a Poisson point process, and the rate λ at which they arrive in a fixed bounded region is proportional to the volume of said region.*

1. *(Initialisation) We set* $\mathbf{sample} \leftarrow (x_1, x_2, ..., x_k)$, *construct* $\mathbf{nb} \leftarrow neighbourhood(\mathbf{sample})$ *and we also set* $t \leftarrow 0$.

2. *(Sampling event) We obtain the volume of* **nb**, *which allows us to calculate the rate λ at which events fall in the neighbourhood. We can thus update the time by $t \leftarrow t + e$ where e is drawn from an exponential random variable with rate λ. We sample a location y uniformly from* **nb** *to be the centre of the event. Finally, we draw U from a uniform distribution $U([0,1])$, if U is smaller than $\mathbf{s_n}$ we tag the event as selective, otherwise as neutral.*

3. *(Updating the sample) We identify the elements of* **sample** *that are closer than $2\mathcal{R}_n$.*

Each one is dropped from the list with probability u. We sample one or two new locations uniformly from $\mathcal{B}_{2\mathcal{R}_n}(y)$ according to whether the event's tag is neutral or selective.

4. *(Updating the neighbourhood) The neighbourhood* **nb** *is updated to reflect the loss and addition of elements to the sample. In the two-dimensional case we do so by considering the pixel we drew the event from as well as its eight neighbouring pixels. For each one of them we verify if any element of the sample is close enough to keep it, otherwise we drop them. For the one dimensional case, given the updated sample we construct the exact neighbourhood If $t < T$, go back to 2. Otherwise, halt.*

In the one dimensional case we will restrict our algorithm to the fixed radius case. This is simply to speed up the algorithm. In two dimensions we do sample a radius uniformly from $[0, \mathcal{R}_n]$.

6.2 The one dimensional case

Earlier we pointed out that in §5 we only considered the case $u = 1$. Allowing u to be smaller than 1 would mean that we no longer have a strict ordering of the left-most and right-most paths. Indeed, while for $u = 1$ a left-most path $\mathscr{P}^l(x_1, 0)$ can never be to the right of a left-most path $\mathscr{P}^l(x_2, 0)$ for $x_2 > x_1$, if we take $u < 1$ the left-most path started from x_1 could potentially jump to the right of $\mathscr{P}^l(x_2, 0)$ if both of them are affected by the same event and $\mathscr{P}^l(x_2, 0)$ doesn't jump. To our knowledge, only non-crossing branching and coalescing walks have been shown to converge to the Brownian net; nevertheless Sun [2005] showed that, in the case of the Brownian web, one can relax this condition and still arrive at the same limit starting with crossing paths.

We could imagine that in the case $u < 1$ we could still define left-most and right-most paths, where the drift would be reduced by u, and potentially construct left-most right-most webs just as we did earlier. But as we will show now, the picture is more complicated than that. In order to illustrate this, Figure 6.1 shows some simulations of Ξ_t^n.

In the figure we have presented four typical realisations of the genealogical process with

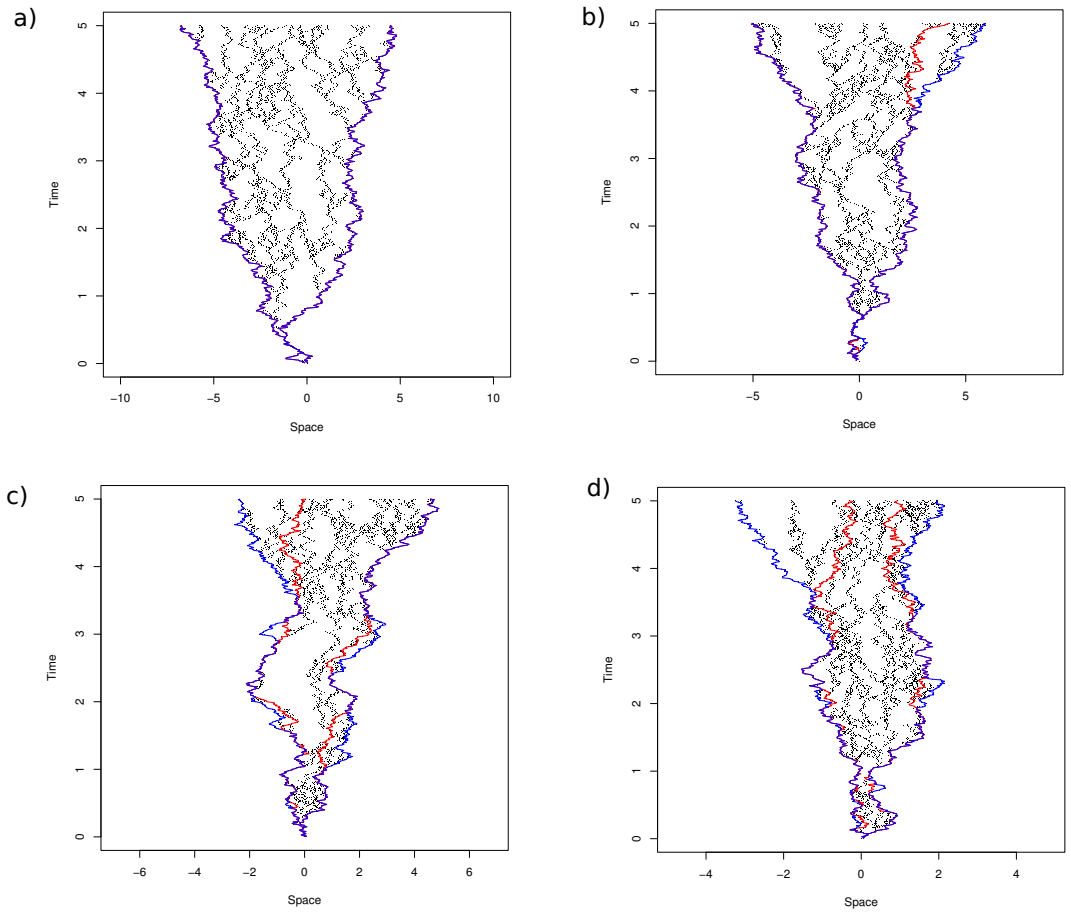

Figure 6.1: Naive and true right/left-most paths. We present realisations of the genealogical process Ξ_t^n for different values of u: *a*) $u = 1$, *b*) $u = 2/3$, *c*) $u = 1/2$, *d*) $u = 1/4$. All of these simulations were done for $n = 5000$ and run from time 0 to time 5. The black lines represent all the ancestral paths, while the blue ones are the left/right-most ones. The red paths are 'naive' left/right-most paths.

initial condition $\Xi_0^n = \delta_0$ (a single lineage started at the origin), each one for a different value of u, where the black dots indicate the location of potential ancestors at a give time. We have also recorded what we call 'naive' left/right-most paths. The 'naive' left-most path is obtained by following the original particle until the first selective event, upon which we jump to the left-most of the two possible locations, and follow this lineage until the next selective event, where we again jump to the left-most potential ancestor, and so on. It is easy to check that the path we follow in this manner converges to a Brownian motion with drift, whose drift is $-2u/3$. In the

same fashion, we can construct the naive right-most path whose drift is $+2u/3$. These paths are coupled with the process Ξ^n and shown in red in Figure 6.1. We also have coloured in blue the actual left-most and right-most paths.

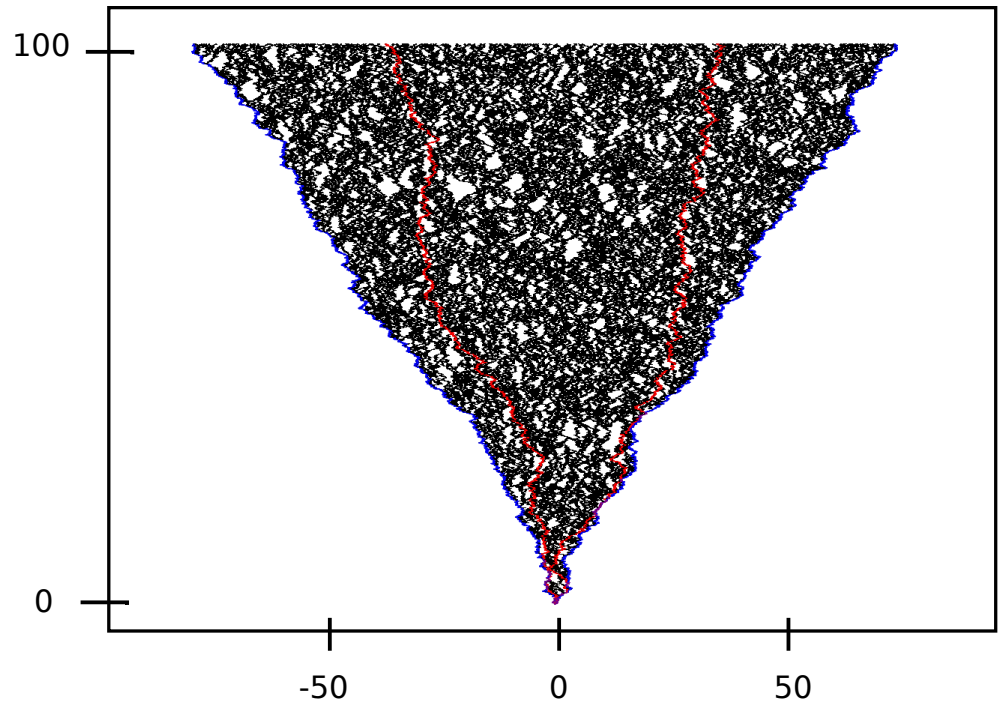

Figure 6.2: Long view of Ξ^n. Here we see a simulation run until time 100, for $u = 1/2$.

Clearly, when $u = 1$ both the naive and true left/right-most paths are identical, but in the other three cases we can see that this is no longer the case. The blue and red paths seem to follow a 'sticky' behaviour, so whenever they coalesce they remain that way for a while, but the distance between them might grow significantly. From these pictures is hard to decide whether the 'naive' and true paths remain close together or if they diverge, but Figure 6.2 gives a stronger indication that the latter might hold.

The examples we have shown suggest that, when $u < 1$, the left/right-most paths won't

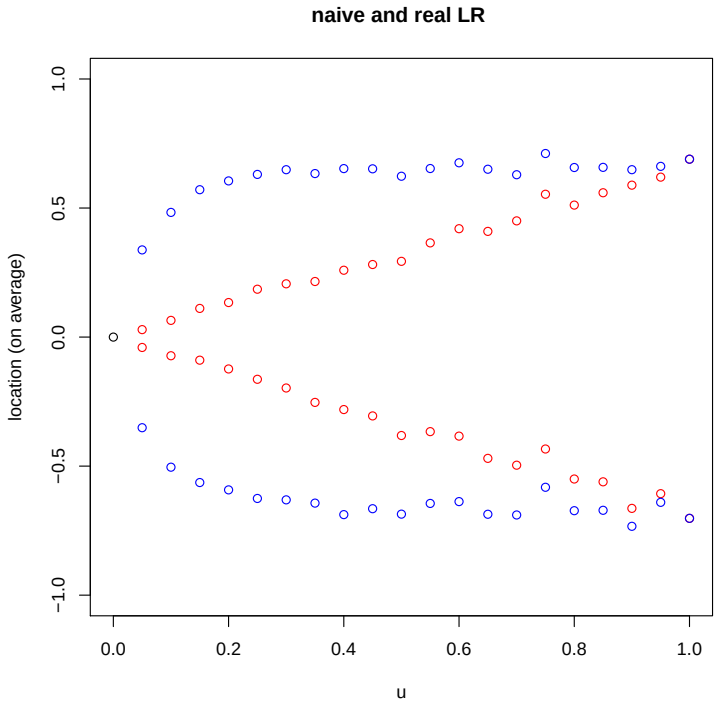

Figure 6.3: Naive and true paths comparison. Taking $n = 5000$ we have ran 2000 simulations for each $u = i/20, i = 1...20$. In red we see the average location of the left/right-most paths at time 1, and in blue the average location of the actual left/right-most lineages.

converge to Brownian motions with drift, or at least not with the drift we might have expected. It looks like the ability of paths to cross each other has a significant impact on the range of Ξ^n. In order to test this further we have considered 20 different values of u and run 2000 simulations of the process for each of them, with all of the simulations taking $n = 5000$. By taking the expected location at time 1 of the right/left-most paths, as well as of the coupled 'naive' paths, we can better assess the discrepancy between them. We summarize our results in Figure 6.3.

We can see in Figure 6.3 that the naive paths behave exactly as expected, their average location is given by $\pm(2/3)u$, the drift. But the location of the actual left/right-most paths, pictured in blue, do not seem to behave linearly in u, and it is not clear what the relationship might be. To rule out the possibility that the divergence between the real and naive paths is compounding with time, we ran similar tests for $t = 3, 5$. As we can see in Figure 6.4, the shape

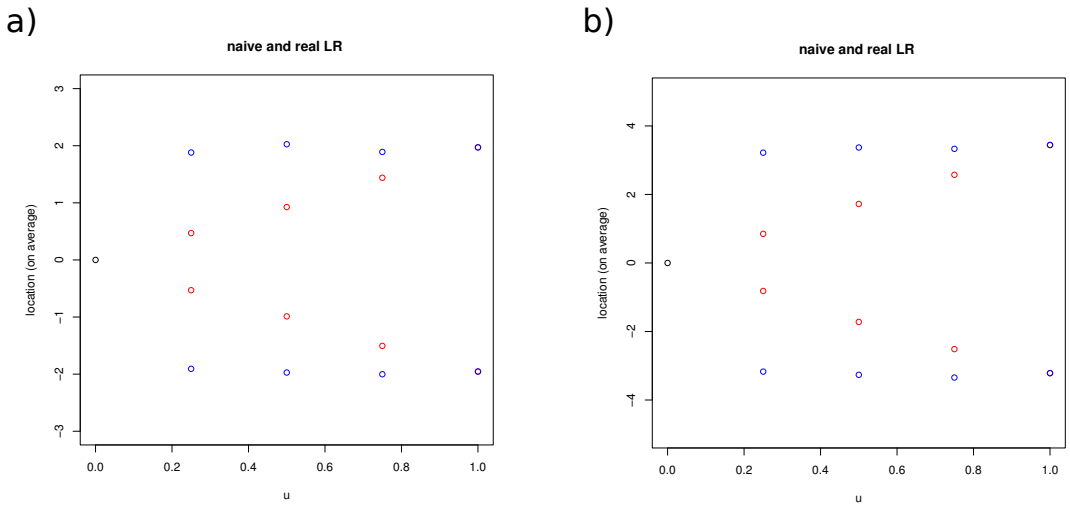

Figure 6.4: Looking at other times. These two graphs mimic Figure 6.3, with $n = 3000$ and 500 simulations; in a) we run it until time $t = 3$ and in b) until $t = 5$.

of the curve described by the blue dots seems to remain consistent through time, so although we cannot rule it out, it seems reasonable to consider for the moment that the position of the left/right-most lineages is roughly linear in t.

Let us try to understand what is going on in the pictures. Even though u directly impacts the speed at which lineages jump and branch, reducing this parameter seems to have little effect on how far the process is able to reach. For values over $u = 1/4$ the curve turns rather flat, which seems surprising, since one might expect that the parameter u would behave somewhat like a clock of the process. Indeed, if we were in the neutral case, the genealogy would converge to a system of independent Brownian motions that coalesce upon meeting, and where the clock-speed of the Brownian motions is proportional to u.

With this idea on mind, it seems reasonable to consider the process stopped at time $1/u$ instead of simply at 1. In this way we hope to get a picture that is more comparable across different values of u. This is precisely what we have done in Figure 6.5. Once again, the 'naive' paths behave precisely as expected, all at roughly the same height. The average displacement of the true paths left/right-most lineages increases very rapidly as one approaches $u = 0$, which is of course not defined (if $u = 0$, the process is trivial since the lineages never jump). From this

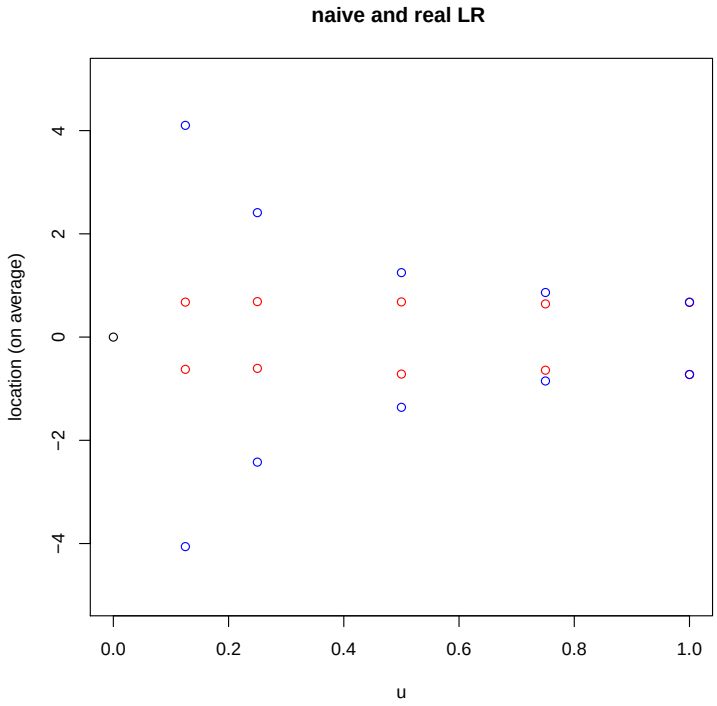

Figure 6.5: Simulating up to time 1/u. Here we have taken $n = 1000$ and ran 1000 simulations for each value of u. The plot shows the average position of the true and naive left/right-most paths at time $1/u$.

it is clear that the likelihood of lineages jumping over each other has a qualitative effect on the process.

Notice that if we were to consider $u \ll 1$, the probability that lineages coalesce becomes negligible, and so we would be mostly considering a system of branching random walks that could be approximated by a BBM with branching rate $\mathcal{O}(\sqrt{n})$. Since for any fixed u the branching rate still explodes, this could perhaps explain the behaviour we see in Figure 6.5.

The position of the right-most particle of BBM has been studied for a long time (Kolmogorov *et al.* [1937]) and it is well known (see for example Bramson [1983]) that if $M(t)$ is the position of the right-most particle of a BBM at time t,

$$\lim_{t\to\infty} \frac{M(t)}{t} = \sqrt{2}. \tag{6.1}$$

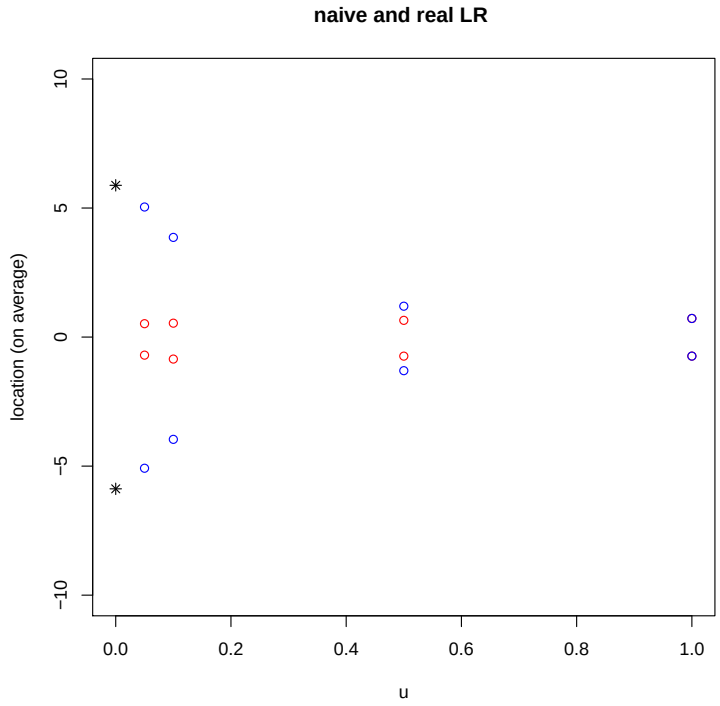

Figure 6.6: BBM approximation for $n = 300$. Here we have taken $n = 300$ and ran 200 simulations for each value of $u \in \{0.05, 0.1, 0.5, 1\}$. The plot shows the average position of the true and naive left/right-most paths at time $1/u$. We have added the locations of the left/right-most particles predicted by the BBM approximation for this value of n, they are indicated by the black stars.

We would like to see if for small values of u the process is indeed similar to a BBM, or at least, if the position of the right-most particle bears some resemblance to what we would expect in the case of BBM. In order to do so we need to calculate the position of the right-most particle of a BBM that has branching rate $\sqrt{n}$. This can be easily obtained from (6.1), using the self similarity of Brownian motion. If $\mathcal{Z}_t$ is a standard BBM, it follows that $\alpha^{-1/2}\mathcal{Z}_{\alpha t}$ is a BBM with branching rate α. Therefore, for every $n < \infty$, the location of the right-most particle $M_n(t)$ of a BBM with branching rate $\sqrt{n}$ satisfies:

$$\lim_{t\to\infty} \frac{M_n(t)}{t} = \sqrt{2\sqrt{n}}. \tag{6.2}$$

We can further assess whether Ξ_t^n indeed behaves in a similar way to $\alpha^{-1/2}\mathcal{Z}_{\alpha t}$ with $\alpha = u\sqrt{n}$ for

small values of u. In order to do so, we first must fix (a large) n, and then look at the expected location of the right-most particle at time $1/u$ for values of u close to 0.

Figure 6.5 provides a good initial test. For $n = 1000$, the right-most particle of the corresponding BBM should be around $\sqrt{2\sqrt{1000}} \sim 7.9$, and we see in the plot that for $u = 1/8$, we are already half way there. Though there are very few points in the plot, they already suggest that the average position of the right-most path creeps up as u gets smaller.

We decided to further test our hypothesis. Unfortunately, running the simulation for small values of u soon becomes too expensive in terms of running time, so we scaled down the parameter n to 300, and ran 200 simulations of the process. The results are shown in Figure 6.6. In this plot we have also added the approximate location of the left/right-most particle in a BBM with branching rate $\sqrt{300}$. The plot suggests that for small values of u, the predictions of the BBM approximation become more and more accurate.

6.3 The two dimensional case

The reason we care about the rate of convergence in two dimensions is that if we have any hope of saying something meaningful about the genealogy of a given sample, or in general about the evolution of a population under our model, we need the BBM we have obtained as its limit to be a good approximation of the pre-limiting models. Even for n on the order of the size of the population of humans or ants, for example, $\log(n)$ is a rather small number, and thus it is not a good enough bound on the rate of convergence in most practical applications. Therefore, we would like to see if the convergence actually happens faster. Since we don't have an exact expression for $\lim_{n\to\infty} \log(n)\mathbf{p_n}$ (where $\mathbf{p_n}$ is the probability that two branches that arise from a selective event separate), we can't compare our simulations to a particular BBM with the right branching rate. Nevertheless, we know that the sequence $(\Xi_t^n)_{n\in\mathbb{N}}$ converges, and so we will attempt to check whether the number and location of the ancestral lineages seems to become stationary at a relatively early stage. To do so, we run simulations of Ξ^n up to time 1 for

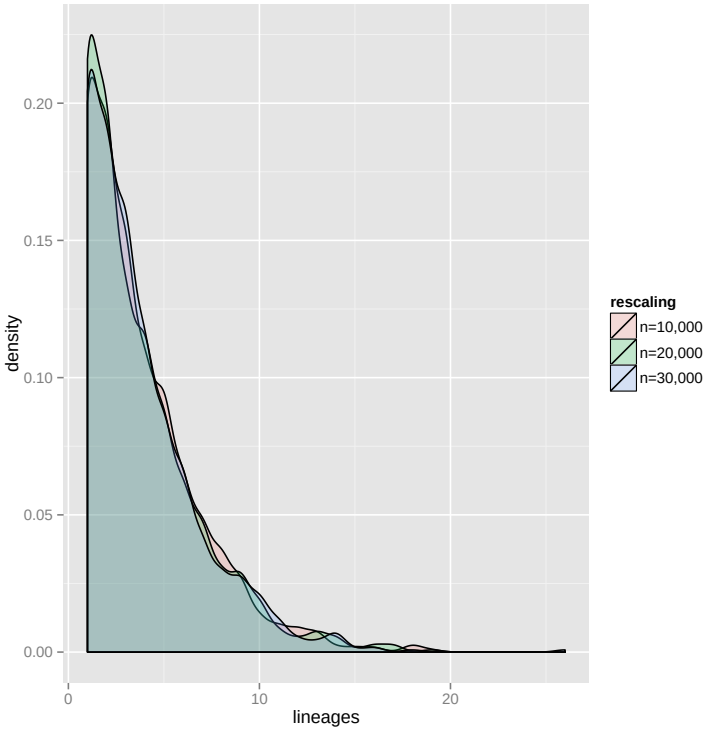

Figure 6.7: Count distribution. This plots compares the empirical distribution of the number of potential ancestral lineages at time 1 for $n = 10000, 20000, 30000$.

$n = 10000, 20000, 30000$, and compare some summary statistics in the hope they reveal that the sequence is not varying too much any more at this point. In particular we will compare the lineage counts, and the maximal distance achieved by the process.

In Figure 6.7 we can see how the empirical distributions of the number of potential ancestors at time 1 compare for the different values of n. It's remarkable how close all three distributions are, and it does seem as though for this particular statistic a reasonable stationarity has been achieved. Once again, we don't claim that this is a rigorous test, but it certainly looks like a reasonable indicator for our purposes.

Figure 6.8 plots the empirical distribution of the distance of the particle furthest away from the origin at time 1. Just as in the previous example, all three distributions are very similar, sharing the same shape and location of the mode. We notice that in both cases $n = 20000$ seems to be the more anomalous of the distributions, particularly in the second figure, where it seems

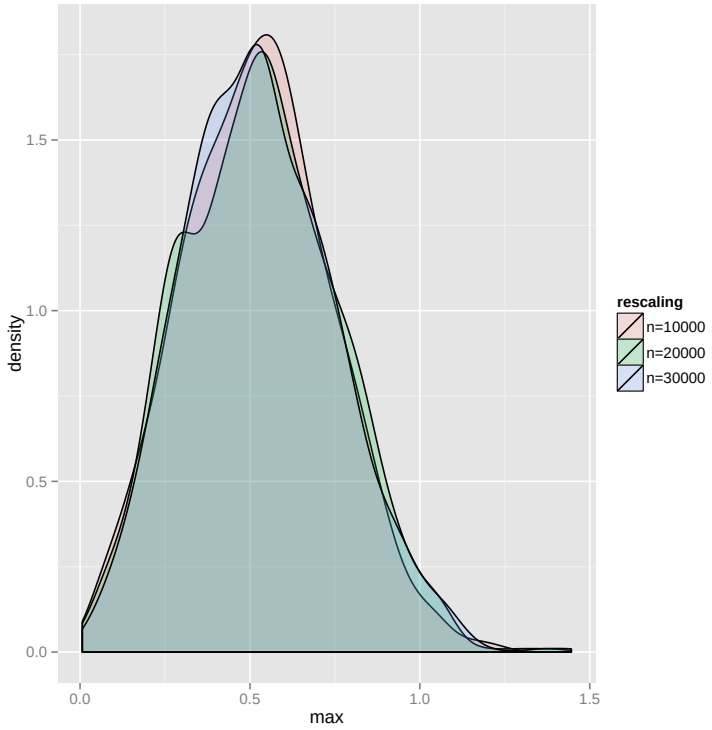

Figure 6.8: Maximal displacement distribution. This plots compares the empirical distribution of the location of the particle farthest away from the origin at time 1 for $n = 10000, 20000, 30000$.

to be almost bimodal. But given that $n = 30000$ doesn't suffer from this irregularity suggests that it is due to sampling error.

In any case, both of these plots give us reason for optimism. Even if they of are no guarantee of convergence, they do suggest that relevant statistics of the process might be close to stationarity for relatively small values of n (notice that $\log(30000) \sim 10$). In this case it would be possible to realise meaningful inference about the process Ξ^n with a BBM approximation, provided we have a good estimate of $\lim_{n\to\infty} \log(n)\mathbf{p_n}$. It is also gives a way of estimating the probability of success itself, which should be a quantity of interest for any given population undergoing selection.

CHAPTER 7

Conclusions

In this thesis we have focused on the (backwards in time) dual process of the spatial Λ Fleming-Viot process with selection. In particular we were interested in obtaining diffusive rescaling limits for the systems of branching and coalescing random walks that correspond to the potential ancestral lineages of a sample of the population.

Our first two chapters gave a concise introduction to the field of population genetics and in particular to spatial models, providing the background and the motivation necessary for constructing the spatial Λ Fleming-Viot process. We later introduced natural selection, and considered the dual process that arises when we add selection the SΛFV process.

We were able to identify the limit in two and higher dimensions as a branching Brownian

motion, although in the two dimensional case we had to take selection to be much stronger in order to arrive at a non trivial limit. The two dimensional result is of independent mathematical interest since it provides a way of making sense of a process where the branching rate goes to infinity but the population remains finite (since we can find an 'effective' bounded branching rate).

Meanwhile, for the one dimensional case we identified the Brownian net to be the limit of the dual process. In this way we showed that the results of Sun and Swart [2008] extend to continuous space-time random walks. Furthermore, if we are concerned with the range of a sub population undergoing selection in one dimension, the leftmost and rightmost paths in the Brownian net provide a natural way of estimating it.

From a more biological perspective our results are relevant to populations evolving in spatial continua. As outlined in the introduction, we aimed to answer the question 'When can we hope to detect a signal of natural selection in data?'. What we found is that the answer was dependent on the dimension of the space in which the population evolves.

In contrast Maruyama [1970] showed that under certain conditions the structure of a population has little effect on the probability that an advantageous allele fixes in the population. In his article some rather strong assumptions are made: the population is living on islands, each island has constant total population size and its contribution to the next generation is in proportion to that size. Much subsequent work retained Maruyama's assumptions, and so it is often assumed that spatial structure has no influence on the accumulation of favourable genes. Nevertheless, Barton [1993] showed that introducing local extinctions and colonisations into a subdivided population could significantly change the fixation probability. This work was extended in, for example, Cherry [2003] and Whitlock [2002]. By taking a fixed parameter u in our model, we are considering a population with small neighbourhood size, and, since a positive fraction of the local population is replaced by new offspring, our reproduction events are somewhat akin to the extinction and colonisation of Barton [1993].

Our main result shows that our ability to detect selection is critically dependent on spatial

dimension. In higher dimensions spatial structure has a weak effect on the trace left by selection. But in two dimensions selection has to be stronger, and even stronger in one dimension, if we are to detect it in the current population. The explanation is that in low dimensions individuals with an advantageous gene have no choice but to compete with close relatives that carry the same gene before they can escape a particular region, whereas in higher dimensions they are more likely to neighbour individuals that do not share the selective advantage.

7.1 Directions for future work

Chapter 6 provides us with two natural courses to continue our investigations, one corresponding to the one dimensional case and one to the two dimensional case.

By fixing $u = 1$ in the one dimensional case we were able to draw from the Brownian net and web literature and extend their results to our settings. Nevertheless, it is natural to wonder what can be said if we are to take $u < 1$, and allow lineages to jump over each other. Our simulations showed that the relationship between the behaviour of left/rightmost paths and the parameter u is anything but linear, and it also showcased that the behaviour of the process for small values of u seems qualitatively different from the case $u = 1$. Since for very small values of u we expect the behaviour of the rightmost ancestral lineages to be approximated by the location of the rightmost particle of a branching Brownian motion, it would be interesting to see if it is possible to characterise the behaviour of the location of left/rightmost paths for fixed values of $u \in (0, 1)$ as a weighted average of the location of a Brownian motion with drift and the location of the right most particle of a BBM. It would also be of mathematical interest to consider the limiting behaviour as $u \to \infty$, which should depend on the ratio of u and n.

For the two dimensional case, our numerical results suggest that the rate of convergence of the dual process to branching Brownian motion should be much faster than logarithmic, and thus we are interested in whether we can improve the bounds we obtained in our proof. Better analytic bounds would provide a good basis for using the BBM approximation to do inference

on genetic data.

APPENDIX A

Spaces of sets of paths

In this appendix we construct a natural state space on which to consider convergence of branching/coalescing systems of càdlàg paths necessary for the proofs in Chapter §5. We also establish some useful relationships to the (continuous path) equivalents introduced by Fontes *et al.* [2004]. The results presented here are valid for any dimension $d \in \mathbb{N}$.

A.1 Càdlàg paths

We definte the tanh function on $[-\infty, \infty]^d$ by pointwise operations, that is $\tanh((x_i)_{i=1}^d) = (\tanh(x_i))_{i=1}^d$. Note that $[-\infty, \infty]^d$ is a complete separable metric space equipped with the

metric

$$(x, y) \mapsto ||\tanh(x) - \tanh(y)||.$$

Let

$$\begin{aligned} C[-\infty] &= \{f : [-\infty, \infty] \to [-\infty, \infty]^d;\ f \text{ is càdlàg}\} \\ C[s] &= \{f : [s, \infty] \to [-\infty, \infty]^d;\ f \text{ is càdlàg}\} \quad \text{for } s \in (-\infty, \infty) \\ C[\infty] &= \{f : \{\infty\} \to [-\infty, \infty]^d\}. \end{aligned}$$

For each $s \in [-\infty, \infty]$ and $f \in C[s]$ we define a function $\bar{f}$ as follows. Let

$$\kappa(t) = \kappa_t = \tanh^{-1}(t)$$

and note that κ is an order preserving homeomorphism between $[-1, 1]$ and $[-\infty, \infty]$. We use the symbol κ in place of $\tanh^{-1}$ to denote a change of time rather than rescaling of space. Then if $f \in C[-\infty]$ we define

$$\bar{f}(t) = \frac{\tanh(f(\kappa_t))}{1 + |\kappa_t|} \tag{A.1}$$

for $t \in [-1, 1]$. Similarly, if $s \in (-\infty, \infty)$ and $f \in C[s]$ then for $t \in [\kappa^{-1}(\sigma_f), 1]$ we define $\bar{f}$ as above and for $t \in [-1, \kappa^{-1}(\sigma_f)]$ we define $\bar{f}(t) = \bar{f}(\kappa^{-1}(\sigma_f))$. Finally, for $f \in C[\infty]$ we define $\bar{f}(\infty) = 0$.

Remark A.1 *The function $\bar{f}$ is a time change κ of f coupled with a time dependent compression of space. First, we extend f to $[-\infty, \infty]$ by taking f to be constant on $[-\infty, \sigma_f]$. Then, each time $t \in [-\infty, \infty]$ is mapped to time $\kappa^{-1}(t) \in [-1, 1]$ and, at this time, the space $[-\infty, \infty]^d$ is compressed into $[\frac{-1}{1+|t|}, \frac{1}{1+|t|}]^d$.*

For each $s \in [-\infty, \infty]$ and $f \in C[s]$, $\bar{f}$ is an element of the space

$$G = \{g : [-1, 1] \to [-1, 1]^d;\ g \text{ is càdlàg}\}.$$

We equip G with its usual Skorodhod metric ρ. Recall that for a path f, σ_f is the time at which

the path is first defined. Let

$$d(f_1, f_2) = \rho(\bar{f}_1, \bar{f}_2) \vee |\tanh(\sigma_{f_1}) - \tanh(\sigma_{f_2})| \tag{A.2}$$

where $f_i \in C[t_i]$. It is easily seen that d is a pseudo-metric on $\cup_{s\in[-\infty,\infty]}C[s]$. We define M to be the set of equivalence classes of $\cup_{s\in[-\infty,\infty]}C[s]$ under d and note that d naturally induces a metric on M.

In fact, for $s \in [-\infty, \infty)$, each $f \in C[s]$ has the single element equivalence class $\{f\} \in M$, but all $f \in C[\infty]$ share an equivalence class. For this reason we will typically represent an element of M as a representative from the equivalence class.

We define also the set $\mathcal{K}$ of compact subsets of M, equipped with the Hausdorff metric m and including the empty set $\varnothing$ as an isolated point. We will shortly show (in Lemma A.3) that M is complete and separable; the space $\mathcal{K}$ inherits the properties of completeness and separability from M.

Lemma A.2 *Let $f_n \in M$ for $n \in \mathbb{N}$ and let $f \in M$. Extend f (and similarly f_n) to a function on $[-\infty, \infty]$ by setting $f(t) = f(\sigma_f)$ for $t < \sigma_f$. Then $f_n \to f$ in M if and only if both:*

- *for all $T' \in (0, \infty)$, $f_n \to f$ in $D_{\mathbb{R}}[-T', T']$.*
- *$\sigma(f_n) \to \sigma(f)$ in $[-\infty, \infty]$.*

PROOF: Let $\rho^{T'}$ denote the Skorohod metric on $D_{\mathbb{R}}[-T', T']$. Recall that

$$\rho^{T'}(f_1, f_2) = \inf_{\lambda} \max\left(||I - \lambda||_{T'}, ||f_1 - f_2 \circ \lambda||_{T'}\right)$$

where I is the identity function, $||f||_{T'} = \sup\{|f(t)|\,;\, t \in [-T', T]\}$ and the infimum is taken over all continuous strictly increasing bijections of $[-T', T']$ to itself. We first need to establish the relationship between the Skorohod metric and the map $f \mapsto \bar{f}$.

Note that for each $T \in (0, 1)$ there exist $C_1, C_2 \in (0, \infty)$ (dependent on T) such that

$$C_1|s - t| \leqslant |\kappa(s) - \kappa(t) \leqslant C_2|s - t|.$$

for all $s, t \in [-T, T]$.

Let $T \in (0,1)$, let $f_1, f_2 \in C[-\infty]$ and suppose that $\rho^T(\bar{f}_1, \bar{f}_2) < \epsilon$. Then there is a continuous strictly increasing bijection $\lambda : [-T, T]$ such that $||I - \lambda||_T, ||\bar{f}_1 - \bar{f}_2 \circ \lambda||_T < 2\epsilon$. It is easily seen that $\eta = \kappa \circ \lambda \circ \kappa^{-1}$ is a strictly increasing continuous bijection $\eta : [-\kappa(T), \kappa(T)] \to [-\kappa(T), \kappa(T)]$. Further,

$$\begin{aligned} |t - \eta(t)| &= |\kappa \circ \kappa^{-1}(t) - \kappa \circ \lambda \circ \kappa^{-1}| \\ &\leqslant \frac{C_2}{C_1}|t - \lambda(t)| \\ &\leqslant 2\epsilon \qquad \text{(A.3)} \\ |f_1(t) - f_2 \circ \eta(t)| &= \left|\tanh^{-1}((1+|t|)\bar{f}_1(\kappa_t^{-1})) - \tanh^{-1}((1+|t|)\bar{f}_2(\kappa_{\eta(t)}^{-1}))\right| \\ &= C_2(1+|t|)|\bar{f}_1(\kappa_t^{-1}) - \bar{f}_2(\lambda(\kappa_t^{-1}))| \\ &\leqslant \frac{C_2(1+|T|)}{C_1} 2\epsilon. \qquad \text{(A.4)} \end{aligned}$$

Combining (A.3) and (A.4) we have $\rho^{\kappa(T)}(f_1, f_2) \leqslant 2\epsilon$. Similarly, if instead we assume that $\rho^{\kappa(T)}(f_1, f_2) < \epsilon$, we can establish equivalents of (A.3) and (A.4) and thus show that $\rho^T(\bar{f} - 1, \bar{f}_2) \leqslant 2\epsilon$.

Further, essentially the same argument as above can be used for general f_1, f_2 in M. To see this, note using our extension of f to $[-\infty, \infty]$, the formula (A.1) holds for all $f \in C[s]$ and all $t \in [-\infty, \infty]$. For the case $\sigma_f = \infty$ we pick any representative element $f \in C[\infty] \in M$ and note that (A.1) holds since $\bar{f}(\infty) = 0$. With (A.1) in hand, the remaining modifications needed for the above argument to go through are trivial. Hence, for each $T \in (0,1)$, $(f_n) \subseteq M$ and $f \in M$,

$$f_n \to f \text{ in } D_{\mathbb{R}}[-\kappa(T), \kappa(T)] \text{ if and only if } \bar{f}_n \to \bar{f} \text{ in } D_{\mathbb{R}}[-T, T]. \qquad \text{(A.5)}$$

We are now ready to prove the lemma. The forwards implication is immediate from (A.5) and (A.2). For the backwards implication, suppose that $f_n \to f$ in $D_{\mathbb{R}}[-\kappa(T), \kappa(T)]$ for each $t \in (0,1)$ and suppose $\sigma(f_n) \to \sigma(f)$. By (A.2) we need only show that $\rho(\bar{f}_n, \bar{f}) \to 0$. Let $\epsilon > 0$ and let $T \in (0,1)$ be such that $\frac{1}{1+|\kappa_T|} < \epsilon$. Then for $T \leqslant |t| \leqslant 1$ we have $|\bar{f}(t)|, |\bar{f}_n(t)| \leqslant \epsilon$, hence

$|\bar{f}_n(t) - \bar{f}(t)| \leqslant 2\epsilon$. By (A.5) we can choose N such that $\rho^T(\bar{f}_n, \bar{f}) \leqslant \epsilon$ for all $n \geqslant N$. We then have

$$\rho(\bar{f}_n, \bar{f}) \leqslant \rho^T(\bar{f}_n, \bar{f}) + 2\epsilon \leqslant 3\epsilon$$

for all $n \geqslant N$, which completes the proof. ■

Lemma A.3 *The metric space (M, d) is complete and separable.*

PROOF: Recall that (G, ρ) and $([-1, 1], |\cdot|)$ are both complete separable metric spaces. It follows immediately from Lemma (A.2) that (M, d) is a metric space. The map $f \mapsto (\bar{f}, \sigma_f)$ is an isometry from M to $G \times [-1, 1]$ (with the product metric $\sup(d(\cdot), |\cdot|)$, hence (M, d) is separable. It remains to show that (M, d) is complete.

Let (f_n) be a Cauchy sequence in M, and extend f to $[-\infty, \infty]$ in the same way as in Lemma A.2. Then $\sigma(f_n)$ is Cauchy and has a limit $\beta \in [-\infty, \infty]$.

By Lemma A.2, for all $T \in (0, 1)$, $(\bar{f}_n)$ is a Cauchy sequence in $D_{\mathbb{R}}[-T, T]$. Since $D_{\mathbb{R}}[-T, T]$ is complete there is some $g^T \in D_{\mathbb{R}}[-T, T]$ such that $\bar{f}_n \to g^T$ in $D_{\mathbb{R}}[-T, T]$. It is easily seen that if $S \in (T, 1)$ then $g^S = g^T$ on $(-S, S)$, hence we can define a function g on $(-1, 1)$ by $g(t) = g^T(t)$ for $|t| \leqslant T$. We extend g to $[-1, 1]$ by defining $g(-1) = g(1) = 0$.

It is immediate from the definition of g that g is càdlàg on $(-1, 1)$. By the same method as in the proof of Lemma A.2, for each $\epsilon > 0$ there exists $T \in (0, 1)$ such that $|g(t)| \leqslant \epsilon$ for $T \leqslant |t| < 1$, hence g is continuous at both 1 and -1. Hence g is càdlàg on $[-1, 1]$.

By construction for all $S \in (0, 1)$ we have $\bar{f}_n \to g$ in $D_{\mathbb{R}}[-S, S]$. Since $|\bar{f}_n(t)| \leqslant \frac{1}{1+\kappa(S)} < 1$ on $[-S, S]$ in fact this convergence takes place in $D_{[-\frac{1}{1+\kappa(S)}, \frac{1}{1+\kappa(S)}]}[-S, S]$, so that $|g(t)| < 1$ for all $t \in [-1, 1]$. Hence we can define

$$f(t) = (1 + \kappa_t) \tanh^{-1}(g(\kappa(t)))$$

for $t \in (-\infty, \infty)$. By (A.5) we then have $f_n \to f$ in $D_{\mathbb{R}}[-T, T]$ for all $T > 0$, hence if we define $\sigma(f) = \beta$, by Lemma A.2 we have $f_n \to f$ in M. ■

A.2 Continuous paths

In the previous section we used the Skorohod metric ρ to build our spaces of càdlàg paths. In fact, in some sense our particular application does not require this level of sophistication; our limit processes will always consists only of continuous paths and for such convergence the supremum metric would be sufficient. However, the supremum metric does not give completeness on spaces of càdlàg paths and as such is unsuitable for probability. In fact, we wish to draw on results from Fontes *et al.* [2004], Ferrari *et al.* [2005] and Sun and Swart [2008] which all use the spaces of continuous paths that were introduced in Fontes *et al.* [2004], based on the supremum metric. There are some (essentially unimportant) issues that we must tidy up in order to bridge the gap between continuous and càdlàg paths.

To this end, let us set up the continuous path spaces from Fontes *et al.* [2004]. We will use the same symbols as in Section A.1 but we add a $\widetilde{\cdot}$ to indicate that the spaces in this section contain only continuous paths. Note that this matches the notation for interpolated paths used in Section 5.2.2.

For $s \in [-\infty, \infty]$, let $\widetilde{C}[s] = \{f \in C[t]\,;\, f \text{ is continuous }\}$. Note that $\bar{f}$ is continuous for all such f, in fact for all $s \in [-\infty, \infty]$ and $f \in \widetilde{C}[s]$, $\bar{f}$ is an element of

$$\{g : [-1,1] \to [-1,1]^d\,;\, g \text{ is continuous}\}.$$

We equip this space with usual supremum metric, $||\cdot,\cdot||_\infty$. We define $\widetilde{M}$ to be the set of equivalence classes of $\cup_{s\in[-\infty,\infty]} C[s]$ under the pseudo-metric

$$\widetilde{d}(f_1, f_2) = ||\bar{f}_1, \bar{f}_2||_\infty \vee |\tanh(\sigma_{\bar{f}_1}) - \tanh(\sigma_{\bar{f}_2})| \tag{A.6}$$

and note that $\widetilde{d}$ induces a metric on $\widetilde{M}$. As for M, we typically identify elements of $\widetilde{M}$ using a representative of the equivalence class (and, as for M, the only such equivalence class containing more than one element is $\widetilde{C}[\infty] \in \widetilde{M}$).

Remark A.4 *Taking $d = 1$, the spaces $\widetilde{M}$ and $\widetilde{\mathcal{K}}$ are essentially the same spaces of paths on*

which the theory of the Brownian web and net was developed. In previous articles they were written without $\widetilde{\cdot}$s (but in this article we are concerned primarily with càdlàg paths). There are two cosmetic differences:

- *We have identified the values of functions at $\{-\infty, \infty\}$ using equivalence classes of $\widetilde{M}$ instead of using a particular compactification of $[-\infty, \infty]^d$ (see the appendix of Sun and Swart 2008).*
- *Using $\bar{f}$, we have extended $f \in C[\sigma_f]$ across all time and therefore our supremum/Skorohod metrics in (A.2) and (A.6) are over the whole interval of (κ^{-1}-rescaled) time $[-1, 1]$ instead of just $[\kappa^{-1}(\sigma_f), 1]$. This makes no difference for continuous paths but is better suited to càdlàg paths.*

The natural equivalent of Lemma A.2 is already well known (see e.g. the introduction of Sun and Swart 2008).

Lemma A.5 *Let $f_n \in \widetilde{M}$ for $n \in \mathbb{N}$ and let $f \in \widetilde{M}$. Extend f (and similarly f_n) to a function on $(-\infty, \infty)$ by setting $f(t) = f(\sigma_f)$ for $t < \sigma_f$. Then $f_n \to f$ in $\widetilde{M}$ if and only if both:*

- *for all $T \in (0, \infty)$, $f_n \to f$ in $C_{\mathbb{R}}[-T, T]$.*
- *$\sigma(f_n) \to \sigma(f)$.*

It was shown in Fontes *et al.* [2004] that $(\widetilde{M}, \tilde{d})$ is a complete and separable metric space. We define the space $\widetilde{\mathcal{K}}$ of compact subsets of $(\widetilde{M}, \tilde{d})$, equipped with the Hausdorff metric $\tilde{m}$ and with the empty set $\varnothing$ as an isolated point. As with $\mathcal{K}$, the space $\widetilde{\mathcal{K}}$ inherits the properties of completeness and separability.

A.3 Backwards paths

The spaces M and $\mathcal{K}$ are appropriate spaces for considering convergence of sets of (branching/coalescing) forwards paths, but they are not suitable for backwards paths. To remedy this,

if P is a set of backwards paths then we define

$$-P = \{\hat{f}\,;\, f \in P\} \tag{A.7}$$

where $\hat{f}(s) = f(-s)$, and note that $-P \in M$ is a set of forwards paths. With slight abuse of notation, for some mode of convergence (including converge of path valued random variables), if f_n is a sequence of backwards paths and f is a backwards paths, we will say $f_n \to f$ in M if $\hat{f}_n \to \hat{f}$ in M. Similarly, if P_n is a sequence of sets of backwards paths and P is a set of backwards paths we way $P_n \to P$ in $\mathcal{K}$ to mean that $-P_n \to -P$ in $\mathcal{K}$. We apply the same terminology to interpolated paths.

A.4 Connections between càdlàg and continuous paths

We will need a small number of results connecting spaces of continuous paths to spaces of càdlàg paths. As a general principle, a Skorohod space of càdlàg paths continuously embeds the corresponding subspace of continuous paths. This principle holds true in our setting, as is shown in the following lemmas.

Lemma A.6 *The identity map $\iota : \widetilde{M} \to M$ is continuous.*

Proof: Since continuous paths are càdlàg it is immediate that ι maps $\widetilde{M}$ into M. For $T' \in (0, \infty)$, convergence in the supremum metric on $[-T', T']$ implies convergence in the Skorodhod metric, so it follows from Lemmas A.2 and A.5 that ι is a continuous map. ■

Lemma A.7 *The identity map $\iota : \widetilde{\mathcal{K}} \to \mathcal{K}$ is continuous.*

Proof: If P is a compact subset of $\widetilde{M}$ then Lemma A.6 implies that P is a compact subset of M. Hence ι maps $\widetilde{\mathcal{K}}$ into $\mathcal{K}$. If $P_n \to P$ in $\widetilde{\mathcal{K}}$ then also by Lemma A.6 we have $P_n \to P$ in $\mathcal{K}$. ■

The following Lemma will allow us to transfer convergence of sets of interpolated paths in $\widetilde{\mathcal{K}}$ onto convergence of càdlàg paths in $\mathcal{K}$.

Lemma A.8 *Let $f \in \mathcal{P}_n^{\uparrow}(\mathbb{R}^2)$ and recall that $\widetilde{f}$ denotes the interpolation of f. Then $d(f, \widetilde{f}) \leqslant 2\mathcal{R}n^{-1/2}$.*

PROOF: Let us write $g = \widetilde{f}$, for easier notation. From Lemma 5.12 we have $\sup_{t \geqslant \sigma_f} ||f(t) - g(t)|| \leqslant 2\mathcal{R}n^{-1/2}$. Since $\tanh(\cdot)$ is a contraction, from (A.1) we have that $\sup_t ||\bar{f}(t) - \bar{g}(t)|| \leqslant 2\mathcal{R}n^{-1/2}$, which implies the stated result. ■

The following result, which is an application of the Arzela-Ascoli Theorem, can be found within the proof of Lemma 4.6 of Sun and Swart [2008].

Lemma A.9 *Let $\widetilde{P} \subseteq \widetilde{M}$. Then $\widetilde{P}$ is relatively compact in $\widetilde{M}$ if and only if*

$$m_{\widetilde{P}}(\delta) = \sup \left\{ \left| \frac{\tanh(f(t))}{1+|t|} - \frac{\tanh(f(s))}{1+|s|} \right| ; \, f \in \widetilde{P}, \; s,t \in \mathbb{R}, \; |\tanh(t) - \tanh(s)| \leqslant \delta \right\} \to 0$$

as $n \to \infty$.

References

Arratia, R. A. (1979). *Coalescing Brownian motions on the line*. University of Wisconsin–Madison. 95

Barton, N. (1993). The probability of fixation of a favoured allele in a subdivided population. *Genetical Research*, **62**(02), 149–157. 148

Barton, N., Etheridge, A., Kelleher, J., and Véber, A. (2013a). Inference in two dimensions: allele frequencies versus lengths of shared sequence blocks. *Theoretical population biology*, **87**, 105–119. 47

Barton, N., Etheridge, A., and Véber, A. (2013b). Modelling evolution in a spatial continuum. *Journal of Statistical Mechanics: Theory and Experiment*, **2013**(01), P01002. 35

Barton, N. H., Depaulis, F., and Etheridge, A. M. (2002). Neutral evolution in spatially continuous populations. *Theor. Pop. Biol.*, **61**, 31–48. 34

Barton, N. H., Etheridge, A. M., and Véber, A. (2010a). A new model for evolution in a spatial continuum. *Electron. J. Probab.*, **15**, 162–216. 36, 39, 43, 45, 59, 69, 70

Barton, N. H., Kelleher, J., and Etheridge, A. M. (2010b). A new model for extinction and recolonization in two dimensions: quantifying phylogeography. *Evolution*, **64**(9), 2701–2715. 35, 38

Barton, N. H., Kelleher, J., and Etheridge, A. M. (2013c). Coalescent simulation in continuous space. *Bioinformatics*, **29**(7), 955–956. 47

Berestycki, N. (2009). Recent progress in coaelscent theory. *Ensaios Matemáticos*, **16**. 16, 17, 20, 22

Berestycki, N., Etheridge, A. M., and Hutzenthaler, M. (2009). Survival, extinction and ergodicity in a spatially continuous population model. *Markov Processes and Related Fields*, **15**, 265–288. 36, 37, 47

Berestycki, N., Etheridge, A. M., and Véber, A. (2013). Large-scale behaviour of the spatial Λ-Fleming-Viot process. *Ann. Inst. H. Poincaré (to appear)*, **49**(2), 374–401. 36, 44, 45, 46, 60, 69, 70, 71, 94, 99, 112

Bordenave, C., Gousseau, Y., and Roueff, F. (2006). The dead leaves model: a general tessellation modeling occlusion. *Advances in Applied Probability*, pages 31–46. 125

Bramson, M. (1983). Convergence of solutions of the Kolmogorov equation to traveling waves. *Memoirs of the American Mathematical Society*, **44**(285), 1–190. 141

Cherry, J. L. (2003). Selection in a subdivided population with local extinction and recolonization. *Genetics*, **164**(2), 789–795. 148

Darwin, C. (1859). *On the origin of species by means of natural selection*. John Murray, London. 49

Donnelly, P. J. and Kurtz, T. G. (1996). A countable representation of the Fleming-Viot measure-valued diffusion. *Ann. Probab.*, **24**, 698–742. 24

Donnelly, P. J. and Kurtz, T. G. (1999). Particle representations for measure-valued population models. *Ann. Probab.*, **27**, 166–205. 24, 26

Durrett, R. (1996). *Stochastic calculus. A practical introduction*. CRC Press. 65, 78

Durrett, R. (2008). *Probability models for DNA sequence evolution. 2nd Edition.* Springer-Verlag. 13, 19, 20

Etheridge, A., Veber, A., and Yu, F. (2014). Rescaling limits of the spatial lambda-fleming-viot process with selection. *arXiv preprint arXiv:1406.5884*. 56, 58

Etheridge, A. M. (2000). *An introduction to superprocesses*, volume 20 of *University lecture notes*. Amer. Math. Soc. 22

Etheridge, A. M. (2008). Drift, draft and structure: some mathematical models of evolution. *Banach Center Publ.*, **80**, 121–144. 1, 3, 36, 47

Etheridge, A. M. (2011). Some mathematical models from population genetics. In *École d'été de probabilités de Saint Flour XXXXIX-2009*, volume 2012. Springer-Verlag. 19, 23, 49, 51, 52

Etheridge, A. M. and Kurtz, T. G. (2014). Genealogical constructions of population models. *arXiv preprint arXiv:1402.6724*. 39, 47, 56, 58

Etheridge, A. M. and Véber, A. (2012). The spatial Lambda-Fleming-Viot process on a large torus: genealogies in the presence of recombination. *Ann. Appl. Probab. (to appear)*, **arXiv:1106.4050v2[math.PR]**. 36, 47

Etheridge, A. M., Véber, A., and Yu, F. (2012). Rescaling limits of the spatial Lambda-Fleming-Viot process with selection. *arXiv preprint arXiv:1406.5884*. 59

Ethier, S. N. and Kurtz, T. G. (1986). *Markov processes: characterization and convergence*. Wiley. 116

Evans, S. N. (1997). Coalescing Markov labelled partitions and a continuous sites genetics model with infinitely many types. *Ann. Inst. Henri Poincaré*, **33**, 339–358. 43

Ewens, W. J. (2004). *Mathematical Population Genetics I. Theoretical introduction.* Springer-Verlag. 12

Felsenstein, J. (1975). A pain in the torus: some difficulties with the model of isolation by distance. *Amer. Nat.*, **109**, 359–368. 34

Ferrari, P., Fontes, L., and Wu, X.-Y. (2003). Two-dimensional Poisson Trees converge to the Brownian web. *arXiv preprint math/0304247*. 122, 123, 124

Ferrari, P., Landim, C., and Thorisson, H. (2004). Poisson trees, succession lines and coalescing random walks. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, **40**(2), 141–152. 94

Ferrari, P., Fontes, L., and Wu, X.-Y. (2005). Two-dimensional Poisson trees converge to the Brownian web. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, **41**(5), 851–858. 94, 95, 104, 106, 111, 112, 123, 125, 126, 156

Fleming, W. H. and Viot, M. (1978). Some measure-valued population processes. In *Proc International Conference on Stochastic Analysis, Northwestern University*, pages 97–108. Academic Press. 24

Fontes, L., Isopi, M., Newman, C., Ravishankar, K., *et al.* (2004). The Brownian web: characterization and convergence. *The Annals of Probability*, **32**(4), 2857–2883. 95, 96, 102, 103, 111, 151, 156, 157

Karlin, S. and Taylor, H. M. (1981). *A second course in stochastic processes*. Academic Press. 21

Karr, A. F. (1975). Weak convergence of a sequence of Markov chains. *Probability Theory and Related Fields*, **33**(1), 41–48. 16

Kelleher, J., Etheridge, A. M., and Barton, N. H. (2014). Coalescent simulation in continuous space: Algorithms for large neighbourhood size. *Theoretical Population Biology*. 47, 134, 135

Kimura, M. (1953). Stepping stone model of population. *Ann. Rep. Nat. Inst. Genetics Japan*, **3**, 62–63. 30

Kingman, J. F. C. (1982). The coalescent. *Stoch. Proc. Appl.*, **13**, 235–248. 15

Kolmogorov, A., Petrovskii, I., and Piskunov, N. (1937). A study of the equation of diffusion with increase in the quantity of matter, and its application to a biological problem. *Bjul. Moskovskogo Gos. Univ*, **1**(7), 1–26. 141

Krone, S. M. and Neuhauser, C. (1997). Ancestral processes with selection. *Theor. Pop. Biol.*, **51**, 210–237. 53

Limic, V. and Sturm, A. (2006). The spatial Lambda-coalescent. *Electron. J. Probab.*, **11**(15), 363–393. 42

Malécot, G. (1948). *Les Mathématiques de l'hérédité*. Masson et Cie, Paris. 33

Maruyama, T. (1970). On the fixation probability of mutant genes in a subdivided population. *Genetical research*, **15**(02), 221–225. 148

Matheron, G. (1968). Modele séquentiel de partition aléatoire. *Rapport technique, Centre de Morphologie Mathématique, Fontainebleau*. 125

Micaux, B. (2007). Flots stochastiques d'opérateurs dirigés par des bruits gaussiens et poissonniens. 2007. *These présentée pour obtenir le grade de docteur en sciences de l'université Paris XI, spécialité mathématiques*. 94, 104

Moran, P. A. P. (1958). Random processes in genetics. *Proc. Camb. Phil. Soc.*, **54**, 60–71. 18

Nordborg, M. and Krone, S. M. (2002). Separation of timescales and convergence to the coalescent in structured populations. In M. Slatkin and M. Veuille, editors, *Modern developments in theoretical population genetics: the legacy of Gustave Malécot*. Oxford University Press. 13, 32

Ohta, T. and Kimura, M. (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet. Res. Camb.*, **22**, 201–204. 24

Pitman, J. (1999). Coalescents with multiple collisions. *Ann. Probab.*, **27**, 1870–1902. 26, 27

Rackauskas, A. and Suquet, C. (2013). Functional laws of large numbers in hölder spaces. *ALEA*, **10**(2), 609–624. 117

Sagitov, S. (1999). The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.*, **26**, 1116–1125. 26

Schertzer, E., Sun, R., and Swart, J. M. (2014). *Stochastic flows in the Brownian web and net*. American Mathematical Soc. 96

Sjödin, P., Kaj, I., Krone, S., Lascoux, M., and Nordborg, M. (2005). On the meaning and existence of an effective population size. *Genetics*, **169**, 1061–1070. 13

Sun, R. (2005). Convergence of coalescing nonsimple random walks to the Brownian web. *arXiv preprint math/0501141*. 136

Sun, R. and Swart, J. M. (2008). The Brownian net. *Annals of Probability*, pages 1153–1208. 4, 93, 95, 96, 97, 98, 99, 100, 101, 102, 103, 113, 114, 119, 120, 121, 129, 130, 131, 133, 148, 156, 157, 159

Véber, A. and Wakolbinger, A. (2013). The spatial Lambda-Fleming-Viot process: an event-based construction and a lookdown representation. *Ann. Inst. H. Poincaré Probab. Statist. (to appear)*, **arXiv:1212.5909v2**. 47, 53, 94

Wakeley, J. (2001). The coalescent in an island model of population subdivision with variation among demes. *Theor. Pop. Biol.*, **59**, 133–144. 32

Wakeley, J. and Aliacar, N. (2001). Gene genealogies in a metapopulation. *Genetics*, **159**, 893–905. 32

Whitlock, M. C. (2002). Selection, load and inbreeding depression in a large metapopulation. *Genetics*, **160**(3), 1191–1202. 148

Wilkinson-Herbots, H. M. (2003). Coalescence time and F_{ST} values in subdivided populations with symmetric structure. *Adv. Appl. Prob.*, **35**, 665–690. 31

Wright, S. (1943). Isolation by distance. *Genetics*, **28**, 114–138. 29, 33