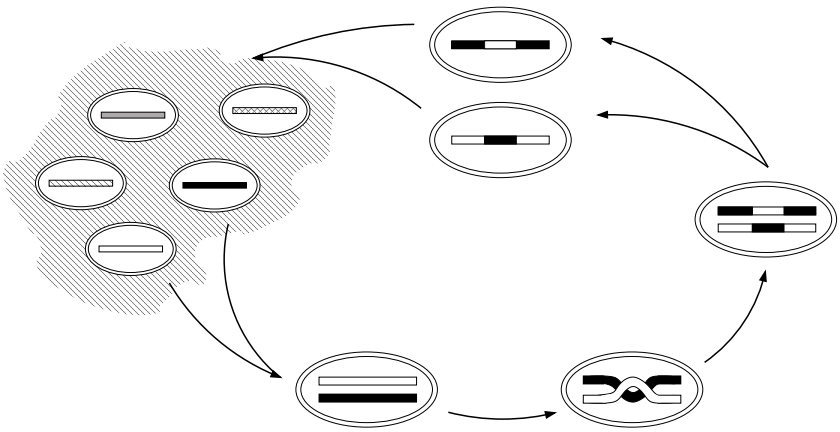
Dynamic and probabilistic aspects of recombination

Dissertation

zur Erlangung des akademischen Grades
Doktor der Mathematik (Dr. math.)

vorgelegt an der Fakultät für Mathematik der Universität Bielefeld



eingereicht von

M. Sc. Frederic L. P. Alberti

im September 2020

Acknowledgement

I am deeply indebted to Professors Ellen and Michael Baake for the outstanding supervision of my doctorate; it is impossible to overstate how essential their insightful feedback and unwavering patience have been to my work. I am particular grateful to Professor Ellen Baake for sparking my interest in mathematical population genetics and for guiding me in my research. I would also like to thank Professor Michael Baake, not only for also providing valuable impulses to my research as well and for proofreading parts of my work, but also for helping me to further expand my mathematical horizon, in particular by introducing me to the fascinating world of aperiodic order. Special thanks belong to Carolin Hermann, whose masters thesis provided the groundwork for parts of this thesis.

Furthermore, I would like to thank Dr. Sebastian Hummel for proofreading parts of this thesis. Besides, I want to thank him, along with Nicole Althermeler, Fernando Cordero, Luigi Esercito and Enrico di Gaspero for their company and many interesting discussions during coffee breaks. Similarly, I have to thank Dr. Dirk Frettlöh, Philiip Gohlke, Dr. Neil Mañibo and Dr. Dan Rust for their company in the pub. My studies have been supported by the DFG SPP Priority Programme 1590 'Probabilistic structures in evolution' as well as the CRC 1283 'Taming uncertainty and profiting from randomness and low regularity in analysis, stochastics and their applications'.

I am deeply grateful to my mother for her support and encouragement throughout all of my studies.

The selection-recombination equation is a large, nonlinear system of differential equations that are impossible to solve explicitly.

Ethan Akin (*The Geometry of Population Genetics*, 1979)

Contents

1 Introduction

Since the pioneering work in evolution and genetics by Darwin [Dar59] and Mendel [Men66], the advent of large scale sequencing technology and increasing accessibility of computing power have brought about a proliferation of quantitative methods in these fields [Wak04]. The models of *mathematical population genetics* provide a rigorous framework to describe how the genetic composition of idealised populations is shaped over time by various evolutionary forces, such as natural selection, mutation, random genetic drift, recombination and migration. There is an enormous variety of models, varying not only in the specific subset of evolutionary forces considered, but also in whether the population is finite or infinite, or whether evolution is assumed to act over the course of discrete generations or in continuous time. The role of these models is twofold; on the one hand, they provide a theoretical framework to generate testable hypotheses and thus help to guide empirical research. On the other hand, they often exhibit interesting mathematical features, which are worthy of investigation in their own right [Cas88].

Of all evolutionary forces commonly considered, the treatment of recombination is particularly challenging, due to its highly nonlinear nature. Since its first introduction by Jennings [Jen17] and Robbins [Rob18] more than a hundred years ago, the *deterministic recombination equation* defied all attempts at its solution for a century. It was ultimately solved by Baake et al. in 2016 [BBS16; BB16] by relating the complicated dynamics forward in time to a much simpler stochastic process that describes the ancestral lineages of single individuals backward in time. This theme, the interplay between the dynamics of the solution of an ODE model for recombination forward in time and probabilistic aspects of the corresponding ancestral process backward in time, will be developed further in this thesis, culminating in an explicit solution of the *selection*-recombination equation for single-crossover with single-site selection.

1.1 Background

Recombination describes the reshuffling of genetic information that occurs during the reproductive cycle of sexually reproducing organisms; it is ubiquitous among virtually all organisms on earth [ABHJ, Ch. 2]. Biologically, this is realised via one or more so-called *crossover events* between two parental sequences, a mechanism that was discovered by Morgan [Mor11] in 1911. It was first formalised in 1917 by Jennings [Jen17] via a discrete dynamical system that models the evolution of a diploid population at two diallelic loci. The word *locus* refers to the position of a particular gene within a chromosome and *diallelic* means that this gene

occurs in two different variants (or *alleles*). Shortly after, in 1918, this work was continued by Robbins [Rob18], still restricted to the case of two loci. In 1944, Geiringer [Gei44] considered a more general equation, for the first time with more than two loci and an arbitrary number of alleles at each of them. She investigated the asymptotic behaviour and, in the case of three loci, stated the general form of the solution in terms of linear combinations of certain ansatz functions. This was subsequently generalised by Bennett [Ben54] to an arbitrary number of loci. Later, it was seen that this method leads to a linearisation of the original model, an approach which was then developed further within the sophisticated framework of *genetic algebras* [MR83; Lyu92] and became known by the name of *Haldane linearisation*. Somewhat surprisingly, these efforts yielded relatively few concrete results.

While most research in mathematical population genetics had initially focussed on deterministic models, stochastic models became more and more fashionable, starting with the seminal works of Fisher [Fis30], Wright [Wri31], Malécot [Mal48], Feller [Fel51], and Moran [Mor58]. The major benefit of stochastic models lies in their ability to take into account the random fluctuations in type frequency that occur due to random reproduction. These fluctuations are particularly important if one considers either very small populations, or the large time scales on which evolution usually takes place. Of great theoretical importance are graphical constructions, which were introduced by Harris [Har78] in the general context of additive, set-valued Markov processes. In our setting, these constructions enable us, thanks to the time-reversibility of the Poisson process, to trace back the ancestry of a finite sample of individuals. The resulting *ancestral processes*, running *backward* in time, are again Markovian and contain information about the evolution in the original *forward* direction of time. In the case of recombination, this leads to the *ancestral recombination graph*, see [Hud83; GM96; GM97; LPS; JFS15; BS16]. Formally, the forward and backward processes are related via the notion(s) of *duality* [Lig10; JK14]. For an introduction to this approach, see [Wak09] and the seminal work by Kingman [Kin82a; Kin82b]. Over time, various evolutionary forces have been incorporated, and the corresponding ancestral processes have been studied extensively. For a survey of the state of the art, see the monographs [Ewe04; Dur08; Eth11].

In light of these new developments, Baake et al. have recently reconsidered the deterministic recombination equation in a very general setting, even allowing for an arbitrary number of parents. To accommodate the needs of quantitative genetics [Bür00, Ch. IV], even uncountable sets of alleles are admitted. The deterministic recombination equation can be constructed from a sequence of (stochastic) Moran models with increasing population sizes via a dynamical law of large numbers [Kur70; EK86] and the corresponding ancestral processes remain random in the infinite-population limit. The limiting process is a simple *partitioning process* on the set of partitions of the genetic sequence, and the solution of the *linear* Kolmogorov forward equation yields a solution of the original nonlinear equation. In addition, this sheds new light on the idea of Haldane linearisation, outside the framework of genetic algebras [BB16]. Previously, these authors obtained an explicit, albeit somewhat involved, recursive solution via lattice theoretic techniques [BBS16]. For a recent review, see [BB].

While the solution of the recombination equation was a significant achievement in its own right, the *selection-recombination equation*, which describes recombination along with natural selection, has remained unsolved. Indeed, explicit solutions have been deemed out of reach (as illustrated by the opening quote of Akin) and research has instead focussed on either its asymptotic behaviour or on special cases in which recombination and selection act on different time scales [NHB99]. Only in the special case of two neutral loci linked to one locus under selection, an approximate solution was given in [SSL06]. While sufficiently precise, it does not convey any hope for generalisation.

Last but not least, an interesting perspective on population genetics can be found in the monograph by Akin [Aki79]. There, deterministic models of population genetics are reformulated in the language of differential geometry and a number of qualitative results are proved — it is shown, for instance, that the entropy of the type distribution is a Lyapunov function for the recombination equation. Indeed, it is even a potential for a generalised gradient system, as was later shown by Hofbauer [Hof17]. This follows by reinterpreting the recombination equation as the law of mass action of a strongly reversible chemical reaction network, which was done by Müller and Hofbauer [HM15].

1.2 Outline

This thesis is structured as follows. First, in Chapter 2, we recall the pure recombination equation, both in discrete and continuous time, along with its basic properties and establish the notation that will be used in subsequent chapters. Due to its importance for the rest of this work, we briefly review the dual partitioning process.

In Chapter 3, we reinterpret the recombination equation for an arbitrary number of parents as a strongly reversible chemical reaction network, generalising the result by Müller and Hofbauer, at least in the case of finite sets of alleles. In particular, we again obtain a representation as a generalised gradient system. Later, we also see how the monotonicity of the partitioning process implies the gradients structure of the evolution of its law, regardless of the underlying type space. Last not least, we consider the finite-dimensional (but nonlinear) system of equations that was derived in [BBS16] for the coefficients in an ansatz for the solution of the recombination equation. We will see that it can be understood as the law of mass action for a network of chemical reactions among the partitions of the genetic sequence.

The heart of this thesis is Chapter 4, where the *explicit* solution of the selection-recombination equation is presented, in the case of one selected site, located at an arbitrary position within the sequence, and linked to an arbitrary number of neutral sites. It is stated in the form of iterated integrals, and we will show that this structure is intimately connected to the model's genealogical structure which is subsequently distilled into three distinct (but related) dual stochastic processes, each yielding different insight. Via their Markov semigroups (which we derive in closed form!), the solution of the original differential equation mentioned above can be stated explicitly. As an important technical tool, we introduce a non-commutative

generalisation of the product of probability measures. This allows us to greatly streamline the necessary computations and also illuminates the *algebraic* structure of the model that underlies our solution.

Finally, in Chapter 3, we apply the methods of [BB16] to a dynamical system that describes the joint action of recombination and *migration*, in *discrete* time. This sheds additional light on results obtained previously by Bürger [Bür00] via the classical theory of dynamical systems. Again, the dual process will be a partitioning process, this time labelled by the origins of the parents. Finally, its *quasi-stationary* distribution [CMS13] is investigated, based on ideas from [Mar17].

The first chapter is based on a single-author paper by the author of this thesis [Alb], the second chapter is joint work with Ellen Baake [AB], and the third chapter is joint work with Ellen Baake, Ian Letter and Servet Martínez [ABLM]; in both cases, all authors contributed equally. All manuscripts have been submitted for publication.

2 The recombination equation in discrete and continuous time

Let us set the stage by recalling the recombination equation in discrete and continuous time. We want to model the evolution of an infinitely large population of haploid individuals under the influence of genetic recombination. For our purposes, a *(genetic) type* will be a finite sequence of letters, indexed by the set

$$S = \{1, \dots, n\}$$

of sequence sites; they can either be thought of as the set of nucleotide positions in a DNA sequence or genetic loci on a chromosome. The letter at each site $i \in S$ is chosen from its own alphabet X_i. If S is interpreted as the set of nucleotide positions, the X_i are the set $\{A, C, G, T\}$ of nucleotides, but they can be more general. For the purposes of *quantitative genetics* [Bür00, Ch. IV], for instance, it is useful to allow general locally compact Hausdorff spaces as alphabets. In any case, a genetic type will be thought of as an element of the *type space*

$$X := \prod_{i \in S} X_i, \tag{2.1}$$

which is endowed with the product topology. If X_i is finite, we endow it with the discrete topology.

In addition to sequences defined over the entire set of sequence sites, we will also be interested in sequences that are only defined over a subset of S; keep in mind that even with modern technology, it is usually infeasible to observe evolution along an entire genome at once. We thus define, for every $U \subseteq S$, the *marginal type space*

$$X_U := \prod_{i \in U} X_i$$

with respect to U. When U is empty, $X_\varnothing$ is the empty Cartesian product, and thus the set with a single element, namely the empty sequence $\boldsymbol{e}$. We drop the subscript when $U = S$, that is, we simply write X as in (2.1) instead of X_S. We denote by $\mathcal{P}(X_U)$ the set of all probability measures on X_U.

Remark 2.1. In the general case of arbitrary Hausdorff spaces, we always understand the term 'probability measure' to mean '*Borel* probability measure'. ♢

More generally, we will at some point also need to consider *signed* measures. For any compact topological space M, the set of finite signed (Borel-)measures on M is denoted by $\mathcal{M}(M)$.

Clearly, $\mathcal{M}(M)$ is a real vector space; equipped with the total variation norm, it is indeed a Banach space, although we will not make use of this fact. Consider a finite collection $(M_i)_{i\in I}$ of such spaces and assume that the M_i are all finite. Then, the set of finite signed measures on the cartesian product $\bigtimes_{i\in I} M_i$ can be identified with the tensor product $\bigotimes_{i\in I} \mathcal{M}(M_i)$; the product $\bigotimes_{i\in I} \nu_i$ of $\nu_i \in \mathcal{M}(M_i)$ can be read either as a measure product or an elementary tensor, according to personal preference. More explicitly, we identify for each $(m_i)_{i\in I} \in \prod_{i\in I} M_i$ the point (or Dirac) measure $\delta_{(m_i)_{i\in I}}$ on $(m_i)_{i\in I}$ with the (tensor or measure) product $\bigotimes_{i\in I} \delta_{m_i}$. Finally, (again for finite M), the set $\mathcal{P}(M)$ of probability measures on M can be identified with the $(|M|-1)$-dimensional standard simplex, i.e.

$$\mathcal{P}(M) = \left\{ \sum_{m\in M} \nu(m)\delta_m : \sum_{m\in M} \nu(m) = 1, \nu(m) \in [0,1] \text{ for all } m \in M \right\}.$$

For $V \subseteq S$, we denote by π_V the canonical projection to X_V, which maps any sequence $x_U = (x_i)_{i\in U}$ defined over any $U \supseteq V$ to the subsequence $x_U^V := (x_i)_{i\in V}$, also called the *marginal type* of x_U with respect to V. To keep the notation simple, we use the same symbol for all projections to a given X_V, irrespective of their domains.

We abbreviate the *push-forward* $\pi_V.\nu_U$ of $\nu_U \in \mathcal{P}(X_U)$ (or, more generally, of $\nu_U \in \mathcal{M}(X_U)$) under π_V by ν_U^V. More explicitly,

$$\nu_U^V(E) := \nu_U(\pi_V^{-1}(E)) = \nu_U(E \times X_{U\setminus V}) \tag{2.2}$$

for all measurable $E \subseteq X_V$. In words, $\nu_U^V(E)$ is the probability that the letters at the sites in V of a random sample from ν_U match those of some type in E.

The following result is elementary, but useful.

Lemma 2.1. *Let $U, V \subseteq S$, $U \cap V = \varnothing$, and let $\nu_U \in \mathcal{M}(X_U)$ and $\nu_V \in \mathcal{M}(X_V)$. Then, for any $W \subseteq U \cup V$, we have*

$$(\nu_U \otimes \nu_V)^W = \nu_U^{U\cap W} \otimes \nu_V^{V\cap W}.$$

Proof. Note that $X_W = X_{U\cap W} \times X_{V\cap W}$. Let us fix $E_{U\cap W} \subseteq X_{U\cap W}$ and $E_{V\cap W} \subseteq X_{V\cap W}$. Then, for any $W \subseteq U \cup V$,

$$\begin{aligned}
(\nu_U \otimes \nu_V)^W (E_{U\cap W} \times E_{V\cap W}) &= (\nu_U \otimes \nu_V)(E_{U\cap W} \times E_{V\cap W} \times X_{(U\cup V)\setminus W}) \\
&= (\nu_U \otimes \nu_V)\big((E_{U\cap W} \times X_{U\setminus W}) \times (E_{V\cap W} \times X_{V\setminus W})\big) \\
&= \nu_U(E_{U\cap W} \times X_{U\setminus W}) \cdot \nu_V(E_{V\cap W} \times X_{V\setminus W}) \\
&= \nu_U^{U\cap W}(E_{U\cap W}) \cdot \nu_V^{V\cap W}(E_{V\cap W}).
\end{aligned}$$

□

Remark 2.2. It is important to note that Lemma 2.1 remains true if $U\cap W = \varnothing$ or $V\cap W = \varnothing$. Assume, for instance, that $U \cap W = \varnothing$. As the empty Cartesian product $X_\varnothing$ is the singleton $\{\boldsymbol{e}\}$, where $\boldsymbol{e}$ is the empty sequence, $\nu_U^{U\cap W}$ is then the unique measure on $\{\boldsymbol{e}\}$ with the same

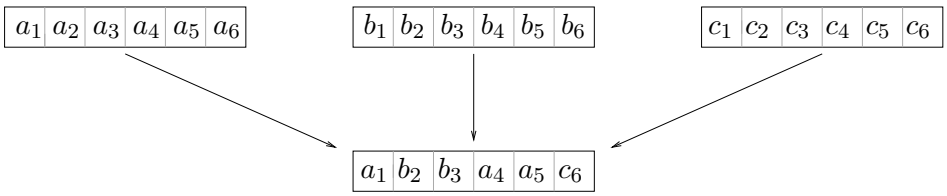


Figure 2.1. A mating event involving 3 parents. One parent contributes the letters at sites $1, 4$ and 5 of its genetic sequence, another the letters at sites 2 and 3 and the third one only the letter at site 6. Thus, the offspring is recombined according to $\mathcal{A} = \{\{1, 4, 5\}, \{2, 3\}, \{6\}\}$.

total mass as ν_U and can be treated as the scalar $\nu_U(X_U)$, in the sense that

$$\nu_U^{U\cap W} \otimes \nu_V^{V\cap W} = \nu_U(X_U)\nu_V^{V\cap W};$$

when ν_U is a probability measure, this simplifies to

$$\nu_U^{U\cap W} \otimes \nu_V^{V\cap W} = \nu_V^{V\cap W};$$

This convention will simplify several calculations later on. ♢

Whenever a group of individuals mate, they produce an offspring whose type sequence is pieced together from fragments of those of its ancestors[1] This group may consist of a single individual[2], two, or arbitrarily many. Obviously, only the first two cases are of biological relevance. However, the case of an arbitrary number of parents is an interesting generalisation from a mathematical perspective, and requires little additional effort.

A central role in the description of recombination in this general setting is played by the partitions of S and its subsets. Recall that a *partition* of an arbitrary set M is a set of pairwise disjoint, non-empty subsets, called *blocks*, of M whose union is M; We denote the set of partitions of M by $\boldsymbol{P}(M)$. Be careful not to confuse $\boldsymbol{P}(M)$ with $\mathcal{P}(M)$, which denotes the set of probability measures on M. Partitions can be used to describe the process of recombination as follows. To any offspring, we can associate an $\mathcal{A} \in \boldsymbol{P}(S)$ to describe how its genetic sequence has been pieced together from its parents. More precisely, the letters at sites $i, j \in S$ are inherited from the same parent if and only if i, j are in the same block of $\mathcal{A}$; we say that the individual is *recombined according to* $\mathcal{A}$; compare Fig. 2.1. If we are only interested in the offspring's type along a subset $U \subseteq S$, we can take $\mathcal{A}$ to be a partition of U. Assuming that the current (marginal) type distribution (along U) in the population is given

[1] While recombination, strictly speaking, does *not* occur during reproduction itself, this is not relevant in the simplified setting of our model; simply put, as we are working at the level of gametes, the word 'reproduction' refers, in this context, to the formation of new germ cells prior to mating.

[2] Obviously, in sexual populations, every individual has two parents. However, crossover events are fairly rare and thus, the entire sequence is often inherited from a single parent. From a modelling perspective, we can think of such an individual as the offspring of a single parent.

by $\nu_U \in \mathcal{P}(X_U)$, the type of the offspring is given by

$$\mathcal{R}^U_{\mathcal{A}}(\nu_U) := \bigotimes_{A\in\mathcal{A}} \nu_U^A; \tag{2.3}$$

this formalises the idea that the offspring 'chooses' its parents independently from the current generation. The (nonlinear) operators

$$\mathcal{R}^U_{\mathcal{A}} : \mathcal{P}(X_U) \to \mathcal{P}(X_U), \quad \nu_U \mapsto \bigotimes_{A\in\mathcal{A}} \nu_U^A$$

for $\mathcal{A} \in \boldsymbol{P}(U)$ are called *recombinators*, and have been introduced in [BB03]. Whenever $U = S$, we suppress the subindex and write $\mathcal{R}$ instead of $\mathcal{R}^S$.

The following representation of the recombinator will play an important role in the next chapter. For this, assume we are given $\mathcal{A} \in \boldsymbol{P}(S)$ and a collection $(x_A)_{A\in\mathcal{A}}$ of marginal types $x_A \in X_A$. Then, we denote by

$$\bigsqcup_{A\in\mathcal{A}} x_A \in X$$

the type (defined over the entire sequence) obtained by glueing the x_A together; its letter at site $i \in S$ is the letter at site i of x_{A^*} where A^* is the unique block of $\mathcal{A}$ that contains i.

Lemma 2.2. *Let X be a finite type space. Then, for $\nu \in \mathcal{P}(X)$ and $\mathcal{A} = \{A_1, \ldots, A_k\} \in \boldsymbol{P}(S)$, we have*

$$\mathcal{R}_{\mathcal{A}}(\nu) = \sum_{x^{(1)},\ldots,x^{(k)}\in X} \nu\big(x^{(1)}\big) \cdot \ldots \cdot \nu\big(x^{(k)}\big) \delta_{\bigsqcup_{i=1}^k \pi_{A_i}(x^{(i)})}.$$

Proof. Let us write $\widetilde{\mathcal{R}}$ for the map on $\mathcal{P}(X)$ defined by the right-hand side. Then, for all $y \in X$,

$$\widetilde{\mathcal{R}}(\nu)(y) = \sum_{\substack{x^{(1)},\ldots,x^{(k)}\in X \\ \pi_{A_i}(x^{(i)})=y^{A_i}\forall i}} \nu\big(x^{(1)}\big) \cdot \ldots \cdot \nu\big(x^{(k)}\big) = \prod_{i=1}^k \nu\big(\pi_{A_i}^{-1}(y)\big),$$

which implies the identity claimed; recall that y^{A_i} denotes the subsequence of y over A_i. □

Remark 2.3. Whenever we enumerate the blocks of a partition of S, i.e. write

$$\mathcal{A} = \{A_1, \ldots, A_{|\mathcal{A}|}\},$$

we order the blocks such that A_1 is the block that contains 1 and, for all $2 \leqslant k \leqslant |\mathcal{A}|$, A_k is the block that contains the smallest element not contained in $\bigcup_{j=1}^{k-1} A_j$. ♢

Notation 2.3. As expressions of the form

$$\delta_{\bigsqcup_{i=1}^k \pi_{A_i}(x^{(i)})}$$

are difficult to read, we simplify the notation by formally identifying each element m of some finite set M with the associated point measure δ_m. Under this convention, the statement from Lemma 2.2 reads

$$\mathcal{R}_{\mathcal{A}}(\nu) = \sum_{x^{(1)},\dots,x^{(k)} \in X} \nu(x^{(1)}) \cdot \ldots \cdot \nu(x^{(k)}) \bigsqcup_{i=1}^{k} \pi_{A_i}(x^{(i)}).$$

♢

Remark 2.4. Sampling parents independently from the same type distribution as described by Eq. (2.3) is only valid in the setting of an infinite population. If the population were finite, the parents would need to be sampled *without* replacement. The corresponding *sampling functions* can, however, be expressed in terms of recombinators (which describe sampling *with* replacement); this is an application of the inclusion-exclusion principle in the form of the Möbius inversion formula [BEP16; Aig79]. ♢

Remark 2.5. Let us also mention at this point that, in Section 4.4, we will introduce a generalisation of $\otimes$ to products of measures that are defined on X_U and X_V with $U \cap V \neq \varnothing$. In particular, they may be marginals with respect to overlapping subsets of S. This will prove to be a handy alternative to the use of recombinators in the description of recombination. ♢

Now, we can fomulate the basic recombination model, first in discrete time. We denote the type distribution in generation $t \in \mathbb{N}_0$ by $\mu_t \in \mathcal{P}(X)$, and assume that generations do not overlap; that is, between generations t and $t+1$, the entire population is replaced by new offspring. We write $r_{\mathcal{A}}$ for the proportion of the offspring that are recombined according to $\mathcal{A}$. As the entire population is replaced, we demand that these proportions add up to one; thus, the type distribution is replaced by the convex combination of the recombined type distributions from Eq. (2.3),

$$\mu_{t+1} = \sum_{\mathcal{A} \in \boldsymbol{P}(S)} r_{\mathcal{A}} \mathcal{R}_{\mathcal{A}}(\mu_t), \tag{2.4}$$

where we refer to the collection $r = (r_{\mathcal{A}})_{\mathcal{A} \in \boldsymbol{P}(S)}$ as the *recombination distribution*.

In continuous time, the role of the recombination distribution r is taken by the collection

$$\varrho = (\varrho_{\mathcal{A}})_{\mathcal{A} \in \boldsymbol{P}(S)}$$

of non-negative recombination *rates*. The intuition is that, for each $\mathcal{A} \in \boldsymbol{P}(S)$ and during each infinitesimally short time interval $[t, t + \mathrm{d}t]$, a proportion of size $\varrho_{\mathcal{A}}\,\mathrm{d}t$ of the population is replaced by offspring that are recombined according to $\mathcal{A}$ (as explained above). Denoting the type distribution at time $t \in \mathbb{R}_{\geqslant 0}$ by ω_t, this translates into the ordinary differential equation

$$\dot{\omega}_t = \sum_{\mathcal{A} \in \boldsymbol{P}(S)} \varrho_{\mathcal{A}} (\mathcal{R}_{\mathcal{A}} - \mathrm{id})(\omega_t). \tag{2.5}$$

We will refer to Eqs. (2.4) and (2.5) as the *recombination equations* in discrete and continuous

time; note that we use different letters, namely ω (in the continuous setting) and μ (in the discrete setting) to avoid confusion.

Remark 2.6. Let us mention that Eqs. (2.4) and (2.5) may alternatively be obtained as large population limits from the (discrete-time) Wright–Fisher model and (continuous-time) Moran model with recombination via a dynamical law of large numbers. In the continuous-time case, this follows from the standard theory of *density dependent families*; see [EK86, Thm. 11.2.1]. We are going to elaborate on this in Chapter 4 in the context of the selection-recombination equation. In discrete time, we refer the interested reader to [BW14, Prop. 1] for the worked (elementary) argument in the special case of single-crossover recombination. ♢

Keep in mind that the choice of the set S of sequence sites was rather arbitrary; in the motivation of the recombination equation we may just as well replace S by $U \subseteq S$; indeed, S itself should be thought of as a mere subset of the entire genome. Then, we would expect to obtain equations of a similar form for $\mu^U = (\mu_t^U)_{t\in\mathbb{N}_0}$ and $\omega^U = (\omega_t^U)_{t\geqslant 0}$. This is indeed the case, and is known as *marginalisation consistency*. To state this formally, we need the concept of an induced partition.

Given $\mathcal{A} \in \boldsymbol{P}(S)$ and $U \subseteq S$, we denote by

$$\mathcal{A}|_U := \{A \cap U : \varnothing \neq A \cap U,\, A \in \mathcal{A}\}$$

the partition of U induced by $\mathcal{A}$.

Now, we can state the marginalisation consistency as follows, in both continuous and discrete time. But first, note that an inductive application of Lemma 2.1 yields

$$\mathcal{R}_{\mathcal{A}}(\nu)^U = \mathcal{R}^U_{\mathcal{A}|_U}(\nu^U) \tag{2.6}$$

for all $\nu \in \mathcal{P}(X)$, $\mathcal{A} \in \boldsymbol{P}(S)$ and $U \subseteq S$; see also [BB16, Lem. 1].

Theorem 2.4 ([BB16, Prop. 3])**.** *Let $U \subseteq S$ and let ω be a solution of the recombination equation in continuous time* (2.5)*. Then, ω^U satisfies the* marginalised recombination equation *in continuous time, i.e.*

$$\dot{\omega}_t^U = \sum_{\mathcal{A}\in\boldsymbol{P}(U)} \varrho^U_{\mathcal{A}} (\mathcal{R}^U_{\mathcal{A}} - \mathrm{id}\,)(\omega_t^U),$$

where the marginal recombination rates $\varrho^U_{\mathcal{A}}$ *are given by*

$$\varrho^U_{\mathcal{A}} = \sum_{\substack{\mathcal{B}\in\boldsymbol{P}(S)\\ \mathcal{B}|_U=\mathcal{A}}} \varrho_{\mathcal{B}}.$$

Theorem 2.5 ([BB16, Lem. 3])**.** *Let $U \subseteq S$ and let μ be a solution of the recombination*

equation in discrete time (2.4). *Then,* μ^U *satisfies the marginalised equation*

$$\mu^U_{t+1} = \sum_{\mathcal{A}\in\boldsymbol{P}(U)} r^U_{\mathcal{A}} \mathcal{R}^U_{\mathcal{A}}(\mu^U_t),$$

where the marginal recombination distribution r^U *is given by*

$$r^U_{\mathcal{A}} = \sum_{\substack{\mathcal{B}\in\boldsymbol{P}(S)\\ \mathcal{B}|_U=\mathcal{A}}} r_{\mathcal{B}}.$$

Both theorems are an immediate consequence of Eq. (2.6).

There are a few additional notions around partitions, which will be essential in what follows. First, we can compare two partitions $\mathcal{A}$ and $\mathcal{B}$. If every block of $\mathcal{A}$ is contained within some block of $\mathcal{B}$, we say that $\mathcal{A}$ is *finer* than $\mathcal{B}$ and write $\mathcal{A} \preccurlyeq \mathcal{B}$. This defines a partial order on each $\boldsymbol{P}(U)$, with unique minimal element

$$\underline{0}_U := \{\{i\} : i \in U\}$$

and unique maximal element

$$\underline{1}_U := \{U\}.$$

When $U = S$, we drop the subscript and simply write $\underline{1}$ and $\underline{0}$ rather than $\underline{1}_S$ and $\underline{0}_S$. For two partitions $\mathcal{A}$ and $\mathcal{B}$, we use

$$\mathcal{A} \wedge \mathcal{B} := \{A \cap B : A \in \mathcal{A},\, B \in \mathcal{B} \text{ and } A \cap B \neq \varnothing\}$$

to denote their *coarsest common refinement*; it is the coarsest partition that is finer than both $\mathcal{A}$ and $\mathcal{B}$, and unique as such.

Remark 2.7. For all partitions $\mathcal{A}$ and $\mathcal{B}$, there exists also a unique finest partition among all partitions coarser than both $\mathcal{A}$ and $\mathcal{B}$, denoted by $\mathcal{A} \vee \mathcal{B}$. Therefore, the partitions (of any finite set) form a *complete lattice* [Aig79]. ♢

In our discussion of marginalisation consistency, we have introduced the concept of an induced partition. Conversely, assume we are given a partition $\mathcal{A}$ of U and a collection $(\mathcal{B}_A)_{A\in\mathcal{A}}$, where $\mathcal{B}_A$ is a partition of A for each $A \in \mathcal{A}$. Then,

$$\bigcup_{A\in\mathcal{A}} \mathcal{B}_A$$

is a partition of U.

Remark 2.8. It is not difficult to see that for any two partitions $\mathcal{A}, \mathcal{B} \in \boldsymbol{P}(U)$, $\mathcal{B} \preccurlyeq \mathcal{A}$ if and

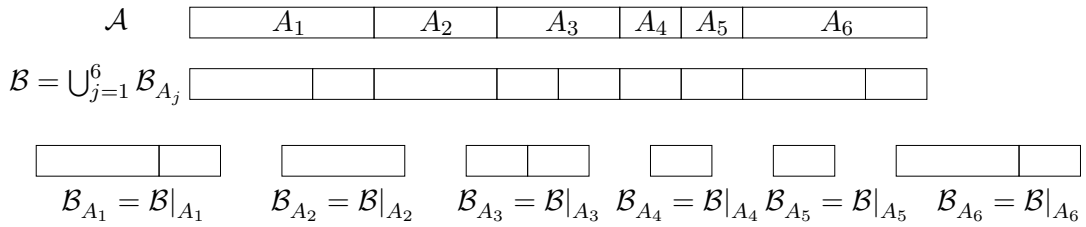


Figure 2.2. At the top, a partition of S. In the middle, a partition finer than $\mathcal{A}$, which gives rise to partitions of the blocks of $\mathcal{A}$ (bottom). Conversely, one can start with the collection of partitions at the bottom and join them to obtain a partition that is finer than $\mathcal{A}$.

only if

$$\mathcal{B} = \bigcup_{A \in \mathcal{A}} \mathcal{B}|_A .$$

For a fixed $\mathcal{A} \in \boldsymbol{P}(U)$, this implies the following bijection between all $\mathcal{B} \in \boldsymbol{P}(U)$ with $\mathcal{B} \preccurlyeq \mathcal{A}$ and all collections $(\mathcal{B}_A)_{A \in \mathcal{A}}$ of partitions of the individual blocks of $\mathcal{A}$. Namely, given $\mathcal{B} \preccurlyeq \mathcal{A}$, we obtain the collection $(\mathcal{B}|_A)_{A \in \mathcal{A}}$ of induced partitions. Conversely, given $(\mathcal{B}_A)_{A \in \mathcal{A}}$, we set $\mathcal{B} := \bigcup_{A \in \mathcal{A}} \mathcal{B}_A$. Note that $\mathcal{B} \preccurlyeq \mathcal{A}$ and $\mathcal{B}|_A = \mathcal{B}_A$ for all $A \in \mathcal{A}$. See Fig. 2.2 for an illustration. ♢

On the level of recombinators, we make the following observation.

Lemma 2.6. *Let $\mathcal{A} \in \boldsymbol{P}(S)$ and let $\mathcal{B}_A \in \boldsymbol{P}(A)$ for all $A \in \mathcal{A}$. Then,*

$$\mathcal{R}_{\bigcup_{A \in \mathcal{A}} \mathcal{B}_A}(\nu) = \bigotimes_{A \in \mathcal{A}} \mathcal{R}^A_{\mathcal{B}_A}(\nu^A)$$

for all $\nu \in \mathcal{P}(X)$. □

We close this preliminary chapter by recalling two fundamental results. First, a reduction of the (potentially) infinite-dimensional measure-valued ODE (2.5) to a finite-dimensional one. For the sake of simplicity, we only consider the case $U = S$.

Theorem 2.7 ([BBS16, Thm. 1])**.** *Every solution ω of (2.5) has the form*

$$\omega_t = \sum_{\mathcal{A} \in \boldsymbol{P}(S)} a_t(\mathcal{A}) \mathcal{R}_{\mathcal{A}}(\omega_0), \tag{2.7}$$

where the coefficients $a_t(\mathcal{A})$ satisfy the coupled nonlinear differential equations

$$\dot{a}_t(\mathcal{A}) = -\sum_{\mathcal{B}} \varrho(\mathcal{B}) \cdot a_t(\mathcal{A}) + \sum_{\dot{\mathcal{B}} \succcurlyeq \mathcal{A}} \left(\prod_{i=1}^{|\mathcal{B}|} \sum_{\substack{\mathcal{C} \in \boldsymbol{P}(S) \\ \mathcal{C}|_{B_i} = \mathcal{A}|_{B_i}}} a_t(\mathcal{C}) \right) \varrho(\mathcal{B}),$$

with initial value $a_0(\underline{1}) = 1$ and $a_0(\mathcal{A}) = 0$, otherwise. The sums run over all partitions of U, where the underdot marks the summation variable. □

This result is immediate after inserting the ansatz (2.7) into (2.5) and using Lemma 2.6 together with Eq. (2.6). In vector notation (compare Notation 2.3), writing $a_t := \sum_{\mathcal{A}\in\boldsymbol{P}(S)} a_t(\mathcal{A})\mathcal{A}$, this system can be written as

$$\dot{a}_t = -\sum_{\mathcal{A}}\sum_{\mathcal{B}} \varrho(\mathcal{B})\cdot a_t(\mathcal{A})\mathcal{A} + \sum_{\mathcal{A}}\sum_{\mathcal{B}\dot{\succcurlyeq}\mathcal{A}} \left(\prod_{i=1}^{|\mathcal{B}|} \sum_{\substack{\mathcal{C}\in\boldsymbol{P}(S)\\ \mathcal{C}|_{B_i}=\mathcal{A}|_{B_i}}} a_t(\mathcal{C})\right)\varrho(\mathcal{B})\mathcal{A} \tag{2.8}$$

The second result we want to mention is the connection to the so-called *partitioning process* [BB16; BEP16]. The partitioning process $\Sigma = (\Sigma_t)_{t\geqslant 0}$ or $\Sigma = (\Sigma_t)_{t\in\mathbb{N}_0}$ is a Markov chain with values in $\boldsymbol{P}(S)$, that, just like the recombination equation, comes in both a continuous and discrete-time version; it describes how the genome of an individual is pieced together from those of its ancestors.

Since similar concepts will be discussed later in more detail, we content ourselves at this point with a rough sketch of the general idea in the case of discrete time. It is best understood if Σ starts at $\underline{1}$. The single block of $\underline{1}$ represents the genome of an individual (Bob, say), sampled from the population at present. Recall that for each $\mathcal{A} \in \boldsymbol{P}(S)$, Bob is with probability $r_{\mathcal{A}}$ recombined according to $\mathcal{A}$. Accordingly, in the first time step from 0 to 1, this single block is replaced by the blocks of $\mathcal{A}$ which represent the different parts of Bob's genome that are contributed by the different parents. The genomes of these ancestors are in turn pieced together from the genomes of their ancestors. This means that during the next time step, from 1 to 2, every block A of $\mathcal{A}$ is again replaced by a random partition of it, according to the corresponding marginal recombination distribution r^A. This is then iterated until we arrive at Σ_t which describes how Bob's genome is pieced together from its ancestors that lived t generations before the present. Finally, their types are sampled, indepedently of each other, from the initial type distribution μ_0.

Formally, the last paragraph boils down to the following stochastic representation for the solution of the recombination equation. In continuous time, we have

$$\mathbb{E}[\mathcal{R}_{\Sigma_t}(\omega_0) \mid \Sigma_0 = \underline{1}] = \omega_t, \quad t \in \mathbb{R}_{\geqslant 0}, \tag{2.9}$$

while in discrete time, we have

$$\mathbb{E}[\mathcal{R}_{\Sigma_t}(\mu_0) \mid \Sigma_0 = \underline{1}] = \mu_t, \quad t \in \mathbb{N}_{\geqslant 0}.$$

More generally, admitting arbitrary $\mathcal{A} \in \boldsymbol{P}(S)$ as initial values for the partitioning process, the following holds, respectively in continuous and discrete time; see also [BB, Remark 3.5].

$$\mathbb{E}[\mathcal{R}_{\Sigma_t}(\omega_0) \mid \Sigma_0 = \mathcal{A}] = \mathcal{R}_{\mathcal{A}}(\omega_t), \quad t \in \mathbb{R}_{\geqslant 0},$$

$$\mathbb{E}[\mathcal{R}_{\Sigma_t}(\mu_0) \mid \Sigma_0 = \mathcal{A}] = \mathcal{R}_{\mathcal{A}}(\mu_t), \quad t \in \mathbb{N}_{\geqslant 0}.$$

This is an example of a so-called duality relation [JK14] between Markov processes, if we interpret ω as a Markov process with deterministic transitions. The transitions of Σ are in continuous time described by the rate matrix $\mathcal{Q}$, where

$$Q(\mathcal{A},\mathcal{B}) := \begin{cases} 0, & \text{if } \mathcal{B} \not\preccurlyeq \mathcal{A}, \\ \varrho^A_{\mathcal{B}_A}, & \text{if } \mathcal{B} = (\mathcal{A}\setminus\{A\})\cup\mathcal{B}_A \\ -\sum_{\mathcal{C}\neq\mathcal{A}} Q(\mathcal{A},\mathcal{C}), & \text{if } \mathcal{A}=\mathcal{B} \end{cases} \tag{2.10}$$

and in discrete time by the Markov matrix T with entries

$$T_{\mathcal{A}\mathcal{B}} = \begin{cases} 0, & \text{if } \mathcal{B} \not\preccurlyeq \mathcal{A}, \\ \prod_{A\in\mathcal{A}} r^A_{\mathcal{B}|_A}, & \text{if } \mathcal{B} \preccurlyeq \mathcal{A}. \end{cases} \tag{2.11}$$

Let us finish this Chapter by remarking that Σ is a variant of the *ancestral recombination graph* [Hud83; GM96; GM97; BS16].

3 Genetic recombination as a gradient flow

In this chapter, we consider Eq. (2.5), the general recombination equation in continuous time. In the biologically relevant case of only two parents, that is, $\varrho_{\mathcal{A}} = 0$ for $|\mathcal{A}| \geqslant 3$ and finite sets of alleles, this equation was reinterpreted by Hofbauer and Müller [HM15] as the law of mass action for a network of chemical reactions between gametes. This reaction network is strongly reversible and general theory [Mie11; Yon12] on chemical reaction networks therefore implies that it can be represented as a generalised gradient system, with respect to entropy [Hof17]. This strengthens an earlier result by Akin [Aki79, Thm III.2.5] that entropy is a strong Lyapunov function for recombination; for a somewhat weaker statement, see [Lyu92, Thm. 6.3.5].

The first goal in this chapter is to generalise this to the setting of arbitrary partitions, in particular, to allow for an arbitrary number of parents. In the setting of finite sets of alleles, we shall generalise the results of [HM15] directly to the multi-parent case; secondly, we will show that the law of associated partitioning process evolves like a gradient system, a property swhich ultimately comes down to the monotonicity of its sample paths. Last not least, we reconsider the nonlinear system Eq. (2.8) and reinterpret it as the law of mass action for a chemical reaction network among the partition of the set of sequence sites.

The rest of this chapter is organised as follows. First, we recall the necessary basic notions from chemical reaction network theory. Then, we discuss the results on the gradient structure of the reaction network equivalent to the recombination equation, along with the necessary background in differential geometry. Section 3.3 explains the gradient structure of a particular class of Markov chains, which contains in particular the partitioning process.

3.1 Chemical reaction networks

Let us recapitulate some basic notions in chemical reaction network theory, taylored to our purposes. A general introduction to the field can be found in [Fei19].

Let $\mathcal{S}$ (not to be confused with S, the set of sequence sites) be a finite set, the elements of which will be thought of as the *reacting species* in a *chemical reaction network* (CRN), that is, a finite collection of *chemical reactions*, which are represented by symbolic expressions of the form

$$r_1 + \ldots + r_{m_1} \xrightarrow{\kappa} s_1 + \ldots + s_{m_2}. \tag{3.1}$$

Here, the r_i and s_i are reacting species (not necessarily distinct) and $\kappa > 0$ is the *reaction*

constant. The left and right-hand sides in Eq. (3.1) are called the complexes of *substrates* and *products*. In our setting, we will always have $m_1 = m_2 = m$, as we will see later.

Remark 3.1. Recall from Notation 2.3 that we formally identified the elements of a finite set with the corresponding point or Dirac measures. In this sense, the addition in Eq. (3.1) can be understood as addition of vectors in the space of signed measures on $\mathcal{S}$. Furthermore, we understand all vectors as column-vectors, as this will be the more natural choice in the context of this chapter. In particular, this entails that the standard scalar product of any two vectors v and w can be written as $v^\intercal w$ (where $\intercal$ denotes transposition) whereas $vw^\intercal$ denotes the matrix that maps any other vector u to $\langle w, u\rangle v$. ♢

Of particular interest are *strongly reversible* CRNs. They are usually defined as CRNs in which the forward reaction constant agrees with the backward reaction constant for every reaction. In the present setting, where we think of reactions as unidirectional, it is more convenient to phrase this slightly differently.

Definition 3.1. A CRN is called *strongly reversible* if it can be partitioned into pairs, each consisting of a reaction,

$$r_1 + \ldots + r_m \xrightarrow{\kappa} s_1 + \ldots + s_m,$$

together with its backward reaction,

$$s_1 + \ldots + s_m \xrightarrow{\kappa} r_1 + \ldots + r_m.$$

♢

Given a CRN, it is natural to inquire about the dynamics of the probability vector

$$c_t = \sum_{s\in\mathcal{S}} c_t(s)\delta_s = \sum_{s\in\mathcal{S}} c_t(s)s$$

of normalised concentrations of reacting species. In our case, as the left and right-hand sides in Eq. (3.1) contain the same number of reacting species, the total mass is preserved and may therefore be normalised to one.

The *law of mass action* translates the collection of formal expressions (3.1) into a system of coupled differential equations for $c = (c_t)_{t\geqslant 0}$. It assumes that each reaction occurs with a rate that is proportional to the concentration of each of the substrates, and hence to their product; the proportionality factor is the reaction constant κ in Eq. (3.1). As each reaction decreases the concentration of substrates and increases the concentration of products, we obtain the following system of ordinary differential equations,

$$\dot{c}_t = \sum \kappa c_t(r_1)\cdot\ldots\cdot c_t(r_m)(s_1 + \ldots + s_m - r_1 - \ldots - r_m), \tag{3.2}$$

where, again, the reacting species $s_1, \ldots, s_m$ and $r_1, \ldots, r_m$ are identified with the corresponding point measures $\delta_{s_1}, \ldots, \delta_{s_m}$ and $\delta_{r_1}, \ldots, \delta_{r_m}$, in accordance with Notation 2.3. The sum

is taken over all reactions that make up the CRN. We refer the interested reader to [EK86, Ex. 11.1.C] for a probabilistic variation on this theme.

We now return to recombination. In [HM15], genetic recombination is treated as a CRN with the types as reacting species, in the special case of two parents. For example, recombination according to $\mathcal{A} = \{\{1,2\},\{3\}\}$ translates to reactions of the form

$$(x_1, x_2, x_3) + (y_1, y_2, y_3) \xrightarrow{\kappa} (x_1, x_2, y_3) + (y_1, y_2, x_3).$$

This describes the process of recombination at the molecular level; first, the parental sequences (x_1, x_2, x_3) and (y_1, y_2, y_3) are split in two, according to $\mathcal{A}$. Then, two new sequences are obtained by joining the leading part of one sequence with the trailing part of the other, and vice versa. For each (ordered) pair of types and each partition $\mathcal{A}$, the reaction constant is $\kappa = \frac{\varrho(\mathcal{A})}{2}$; this is a special case of Theorem 3.2, which is stated below. In the general case of more than two parents, the basic idea is still the same; for any partition $\mathcal{C}$, take $|\mathcal{C}|$ types, split each of them according to $\mathcal{C}$, rearrange the parts and join them back together. Note that this last step is somewhat ambiguous; already in the three-parent case, this can be done in at least two different ways; either,

$$(x_1, x_2, x_3) + (y_1, y_2, y_3) + (z_1, z_2, z_3) \longrightarrow (x_1, y_2, z_3) + (y_1, z_2, x_3) + (z_1, x_2, y_3), \tag{3.3}$$

or

$$(x_1, x_2, x_3) + (y_1, y_2, y_3) + (z_1, z_2, z_3) \longrightarrow (x_1, z_2, y_3) + (z_1, y_2, x_3) + (y_1, x_2, z_3). \tag{3.4}$$

One way of resolving this ambiguity is to order the substrates and define the reaction accordingly. Thus, there may be many different reactions that share the same complex of substrates. More precisely, for every partition $\mathcal{C} \in \boldsymbol{P}(S)$ and each ordered tuple $(x^{(1)}, \dots, x^{(|\mathcal{C}|)}) \in X^{|\mathcal{C}|}$, we define a chemical reaction via the following graphical construction, illustrated in Figure 3.1. First, just as in the two-parent case, the $|\mathcal{C}|$ type sequences are broken up into the subsequences $\pi_{C_j}(x^{(i)})$ over the blocks of $\mathcal{C}$. Then, these fragments are arranged on a two-dimensional, $|\mathcal{C}|$-periodic grid (or discrete torus), where $\pi_{C_j}(x^{(i)})$ is placed in the i-th column and j-th row. Finally, the products are formed by joining the fragments along each diagonal line, running from north-west to south-east through the grid. Alternatively, one may think about moving the i-th row $i-1$ places to the left, and then joining the fragments in each column. More formally, every choice of $\mathcal{C}$ and $(x^{(1)}, \dots, x^{|\mathcal{C}|})$ defines a reaction

$$\sum_{j=1}^{|\mathcal{C}|} x^{(j)} \xrightarrow{\frac{\varrho(\mathcal{C})}{|\mathcal{C}|}} \sum_{j=1}^{|\mathcal{C}|} \bigsqcup_{i=1}^{|\mathcal{C}|} \pi_{C_i}(x^{(i+j-1)}), \tag{3.5}$$

where the indices are to be read modulo $|\mathcal{C}|$.

Notice that the right-hand side depends on the order of the substrates, while the left-hand

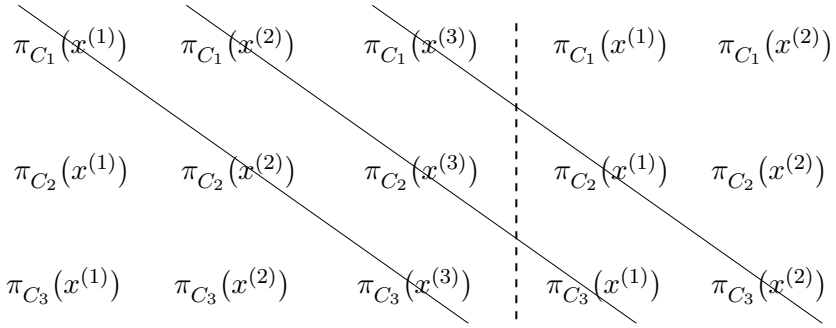


Figure 3.1. An illustration of the reaction scheme for $|\mathcal{C}| = 3$. The types $x^{(1)}$, $x^{(2)}$ and $x^{(3)}$ are each split according to $\mathcal{C}$ and then joined back together as indicated by the diagonal lines. For the sake of clarity, the first two columns of the diagram are repeated after the vertical line.

side does not. For instance, in our earlier example with three sites and parents (that is, $\mathcal{C}$ is $\underline{0}$, the trivial partition into singletons), the choice $x^{(1)} = x = (x_1, x_2, x_3), x^{(2)} = y = (y_1, y_2, y_3), x^{(3)} = z = (z_1, z_2, z_3)$ leads to Eq. (3.3), while exchanging the roles of the second and third type leads to Eq. (3.4).

Theorem 3.2. *For finite X, the general recombination equation (Eq.* (2.5)*) can be written as the law of mass action for the* CRN *comprised of all reactions* (3.5)*, one for every choice of $\mathcal{C}$ and $(x^{(1)}, \dots, x^{(|\mathcal{C}|)})$. More concisely,* (2.5) *is equivalent to*

$$\dot{\omega}_t = \sum_{\mathcal{C} \in \boldsymbol{P}(S)} \sum_{x^{(1)}, \dots, x^{(|\mathcal{C}|)} \in X} \frac{\varrho(\mathcal{C})}{|\mathcal{C}|} \omega_t(x^{(1)}) \cdot \ldots \cdot \omega_t(x^{(|\mathcal{C}|)}) \left(\sum_{j=1}^{|\mathcal{C}|} \left(\bigsqcup_{i=1}^{|\mathcal{C}|} \pi_{C_i}(x^{(i+j-1)}) - x^{(j)} \right) \right).$$

Proof. We show that

$$(\mathcal{R}_{\mathcal{C}}(\nu) - \nu) = \frac{1}{|\mathcal{C}|} \sum_{x^{(1)}, \dots, x^{(|\mathcal{C}|)}} \nu(x^{(1)}) \cdot \ldots \cdot \nu(x^{(|\mathcal{C}|)}) \sum_{j=1}^{|\mathcal{C}|} \left(\bigsqcup_{i=1}^{|\mathcal{C}|} \pi_{C_i}(x^{(i+j-1)}) - x^{(j)} \right)$$

holds for all $\mathcal{C} \in \boldsymbol{P}(S)$ and all $nu \in \mathcal{P}(X)$. Recall that, by Lemma 2.2, we have

$$\begin{aligned}
\mathcal{R}_{\mathcal{C}}(\nu) &= \sum_{x^{(1)}, \dots, x^{(|\mathcal{C}|)} \in X} \nu(x^{(1)}) \cdot \ldots \cdot \nu(x^{(|\mathcal{C}|)}) \bigsqcup_{i=1}^{|\mathcal{C}|} \pi_{C_i}(x^{(i)}) \\
&= \frac{1}{|\mathcal{C}|} \sum_{j=1}^{|\mathcal{C}|} \sum_{x^{(1)}, \dots, x^{(|\mathcal{C}|)} \in X} \nu(x^{(1)}) \cdot \ldots \cdot \nu(x^{(|\mathcal{C}|)}) \bigsqcup_{i=1}^{|\mathcal{C}|} \pi_{C_i}(x^{(i+j-1)}) \\
&= \frac{1}{|\mathcal{C}|} \sum_{x^{(1)}, \dots, x^{(|\mathcal{C}|)} \in X} \nu(x^{(1)}) \cdot \ldots \cdot \nu(x^{(|\mathcal{C}|)}) \sum_{j=1}^{|\mathcal{C}|} \bigsqcup_{i=1}^{|\mathcal{C}|} \pi_{C_i}(x^{(i+j-1)}).
\end{aligned}$$

Here, the second equality is obtained by replacing the product $\nu(x^{(1)}) \cdot \ldots \cdot \nu(x^{(|\mathcal{C}|)})$ with its cyclic permutation,

$$\nu(x^{(1-j+1)}) \cdot \ldots \cdot \nu(x^{(|\mathcal{C}|-j+1)}),$$

and subsequently renaming the indices; recall that indices are to be read modulo $|\mathcal{C}|$. Similarly, keeping in mind that $\sum_{x\in X} \nu(x) = 1$ because ν is a probability measure, we obtain

$$\begin{aligned}\nu = \sum_{x\in X} \nu(x)x =& \frac{1}{|\mathcal{C}|} \sum_{j=1}^{|\mathcal{C}|} \sum_{x\in X} \Big(\sum_{y\in X} \nu(y) \Big)^{j-1} \nu(x) \Big(\sum_{y\in X} \nu(y) \Big)^{|\mathcal{C}|-j} x \\ =& \frac{1}{|\mathcal{C}|} \sum_{x^{(1)},\ldots,x^{(|\mathcal{C}|)}} \nu(x^{(1)}) \cdot \ldots \cdot \nu(x^{(|\mathcal{C}|)}) (x^{(1)} + \ldots + x^{(|\mathcal{C}|)}),\end{aligned}$$

which completes the argument. □

We have thus seen that in the case of finite type spaces, genetic recombination can be reinterpreted as a CRN, also in the case of an arbitrary number of parents. Next, we show that this network is strongly reversible.

Theorem 3.3. *The* CRN *from Theorem* 3.2 *is strongly reversible in the sense of Definition* 3.1.

Proof. Let $\mathcal{C} \in \boldsymbol{P}(S)$ be fixed. Define $\varphi : X^{|\mathcal{C}|} \to X^{|\mathcal{C}|}$ via

$$\varphi(x^{(1)}, \ldots, x^{(|\mathcal{C}|)}) := \Big(\bigsqcup_{i=1}^{|\mathcal{C}|} \pi_{C_i}(x^{(i+|\mathcal{C}|-1)}), \ldots, \bigsqcup_{i=1}^{|\mathcal{C}|} \pi_{C_i}(x^{(i)}) \Big).$$

Note that $\varphi(x^{(1)}, \ldots, x^{(|\mathcal{C}|)})$ contains the products in the reaction defined by $\mathcal{C}$ together with $(x^{(1)}, \ldots, x^{(|\mathcal{C}|)})$, in reverse order (compare Eq. 3.5). A short reflection on Fig. 3.1 reveals that φ is an involution and thus partitions $X^{|\mathcal{C}|}$ into orbits that contain one or two elements each. Consider first an orbit with two elements $(x^{(1)}, \ldots, x^{(|\mathcal{C}|)})$ and $(y^{(1)}, \ldots, y^{(|\mathcal{C}|)})$. Then, the associated reactions form a forward-backward reaction pair,

$$\sum_{j=1}^{|\mathcal{C}|} x^{(j)} \xrightarrow{\frac{\varrho(\mathcal{C})}{|\mathcal{C}|}} \sum_{j=1}^{|\mathcal{C}|} y^{(j)} \quad \text{and} \quad \sum_{j=1}^{|\mathcal{C}|} y^{(j)} \xrightarrow{\frac{\varrho(\mathcal{C})}{|\mathcal{C}|}} \sum_{j=1}^{|\mathcal{C}|} x^{(j)}.$$

When, on the other hand, an equivalence class consists of a single element, the reaction defined by it is void in the sense that its contribution to the right-hand side of the system (3.2) of differential equations vanishes, since its product and substrate complex agree. □

Next, we consider the connection to gradient systems.

3.2 The gradient system

In this section, we need a few basic notions from differential (particularly Riemannian) geometry, which we recall here for the convenience of the reader. For further background we refer the reader to [Wal04], in particular to Chapter 5.

For a real-valued differentiable function V defined on (some subset of) $\mathbb{R}^d$, a function $\boldsymbol{C}$ with the same domain and values in the positive semi-definite symmetric matrices, a *generalised gradient system* (with respect to V) is an ordinary differential equation of the form

$$\dot{x} = C(x)\nabla V(x), \tag{3.6}$$

where we suppressed the time argument. Here,

$$\nabla := \sum_{i=1}^{d} \hat{e}_i \frac{\partial}{\partial x_i}$$

is the nabla symbol and $\hat{e}_1, \ldots, \hat{e}_d$ denote the standard basis vectors of $\mathbb{R}^d$. The real-valued function V can be thought of as a potential. Here, x does not denote a type, but rather a point in $\mathbb{R}^d$.

Given $x \in \mathbb{R}^d$, a vector v in $T_x(\mathbb{R}^d)$, the tangent space of $\mathbb{R}^d$ at x, and a continuously differentiable curve γ in $\mathbb{R}^d$ with $\gamma(0) = x$ and $\gamma'(0) = v$, recall that the *directional derivative* of V in direction v is given by

$$\mathrm{d}V(x)(v) := \frac{\mathrm{d}}{\mathrm{d}t} V(\gamma(t))|_{t=0}.$$

The one-form $\mathrm{d}V$ is called the *exterior derivative* of V; note that it can be defined analogously for any real-valued function on a smooth manifold, and, in particular, does not depend on the Euclidean structure of $\mathbb{R}^d$. One has, by an application of the chain rule,

$$\mathrm{d}V(x)(v) = \sum_{j=1}^{d} \gamma'(0)_j \frac{\partial}{\partial x_j} V(x) = \langle \gamma'(0), \nabla V(x) \rangle, \tag{3.7}$$

where $\langle \cdot, \cdot \rangle$ denotes the standard scalar product on $\mathbb{R}^d$. Replacing the standard scalar product by a general Riemannian metric $\langle\!\langle \cdot, \cdot \rangle\!\rangle_x$ on $\mathbb{R}^d$ (that is, a positive definite, symmetric bilinear form on the tangent space, which varies smoothly, depending on the base point), Eq. (3.7) can be used to define the *gradient* of V with respect to this metric [Wal04, Ex. 108], denoted by $\mathrm{grad}_{\langle\!\langle \cdot, \cdot \rangle\!\rangle}(V)$; it is the unique vectorfield that satisfies

$$\mathrm{d}V(x)(v) = \langle\!\langle v, \mathrm{grad}_{\langle\!\langle \cdot, \cdot \rangle\!\rangle}(V)(x) \rangle\!\rangle_x$$

for all x and v. Geometrically, this means that, unless x is an equilibrium, $\mathrm{grad}_{\langle\!\langle \cdot, \cdot \rangle\!\rangle}(V)(x)$ points in the direction of steepest ascent of V at the point x, with respect to the chosen

metric. In particular, if $C(x)$ in Eq. (3.6) is invertible and we consider the metric,

$$\langle\langle u, w\rangle\rangle_x := \langle u, C(x)^{-1} w\rangle,$$

we see that

$$\mathrm{grad}_{\langle\langle\cdot,\cdot\rangle\rangle}(V)(x) = C(x)\mathrm{grad}_{\langle\cdot,\cdot\rangle}(V)(x) = C(x)\nabla V(x).$$

Thus, Eq. (3.6) can be thought of as a gradient system in the classical sense, if we replace the Euclidean metric on $\mathbb{R}^d$ by a Riemannian one, at least when $C(x)$ is invertible.

The interpretation is somewhat more delicate when $C(x)$ fails to be invertible. Intuitively, one might think of the kernel of $C(x)$ as a set of forbidden directions, and try to restrict attention to submanifolds which partition the space and are in each point x tangent to the image of C. However, this interpretation is only valid when the image of C is *integrable* in the sense that whenever Y and Z are two vectorfields such that $Y(x) \in \mathrm{Im}\ C(x)$ and $Z(x) \in \mathrm{Im}\ C(x)$ for all x, then also $[Y, Z](x) \in \mathrm{Im}\ C(x)$ for all x, where $[Y, Z]$ denotes the Lie bracket of Y and Z; this is the content of frobenius' Theorem [Wal04, Thm. 1.9.2]. The situation when Im C is not integrable can be understood via the theory of sub-Riemannian manifolds. Roughly speaking, this theory is concerned with Riemannian metrics which may take the value $+\infty$; see [BR96] for an overview.

Remark 3.2. To demonstrate the non-triviality of the condition of integrability, consider the following two vector fields on $\mathbb{R}^3$.

$$X_1 := \frac{\partial}{\partial x_1} \quad \text{and} \quad X_2 := x_1 \frac{\partial}{\partial x_3} + \frac{\partial}{\partial x_2}.$$

Then,

$$[X_1, X_2] = \frac{\partial}{\partial x_3},$$

which is nowhere in the span of X_1 and X_2; thus, proving integrability in our case (and for the gradient systems arising in chemical reaction network theory in general) might be an interesting problem in its own right. ♢

We remark that, under the assumption that (3.6) has a unique equilibrium, the potential V is always a strong (global) Lyapunov function (by which we mean that V is strictly increasing along non-constant solutions). This is because

$$\langle \nabla V(x), \dot{x}\rangle = \langle \nabla V(x), C(x)\nabla V(x)\rangle \geqslant 0,$$

by the positive semi-definiteness of $C(x)$. Equality holds if and only if $\nabla V(x)$ is in the kernel of $C(x)$ (implying that $\dot{x} = 0$), hence, if and only if the system is in equilibrium.

We have seen in the previous section that the general recombination equation, interpreted as a chemical reaction network, is strongly reversible. Thus, it is a gradient system in the sense of Eq. (3.6), by standard theory; compare [Yon12; Mie11], where this is proved in much

greater generality. For the sake of completeness, we include the simple proof of this fact, in the special case needed for our purposes.

Theorem 3.4. *The law of mass action for any strongly reversible* CRN *can be written as a generalised gradient system,*

$$\dot{c}_t = C(c)\nabla F(c),$$

where

$$F(c) := -\sum_{s\in\mathcal{S}} \Big(c(s)\log\big(c(s)\big) - c(s) \Big)$$

is called the negative free energy *and C is a continuous function on $\mathcal{P}(S)$, which is smooth on its interior and takes values in the positive semi-definite matrices.*

Proof. Due to strong reversibility (see Definition 3.1), the law of mass action takes the form

$$\dot{c}_t = \sum \Big(\prod_{i=1}^{m} c(r_i) - \prod_{i=1}^{m} c(s_i) \Big) \sum_{i=1}^{m} (s_i - r_i),$$

where the outer sum is taken over all forward-backward reaction pairs in the network. Define for $r, t \geqslant 0$,

$$L(r,t) := \frac{r-t}{\log\,(r) - \log\,(t)}.$$

It is a straightforward exercise to verify that L defines a continuous, non-negative function on $\mathbb{R}^2_{\geqslant 0}$, which is smooth on $\mathbb{R}^2_{>0}$. Note that

$$\nabla F(c) = -\sum_{s\in\mathcal{S}} \log\big(c(s)\big)s.$$

Thus, defining (for each forward-backward reaction pair)

$$M(c) := L\Big(\prod_{i=1}^{m} c(r_i), \prod_{i=1}^{m} c(s_i) \Big) \Big(\sum_{i=1}^{m} (s_i - r_i) \Big) \Big(\sum_{i=1}^{m} (s_i - r_i) \Big)^{\mathsf{T}},$$

we see by the multiplication rule for the logarithm that

$$\Big(\prod_{i=1}^{m} c(r_i) - \prod_{i=1}^{m} c(s_i) \Big) \sum_{i=1}^{m} (s_i - r_i) = M(c)\nabla F(c).$$

Here, we also used that $s^{\mathsf{T}}\nabla F(c) = -\log\big(c(s)\big)$ for all $s \in \mathcal{S}$. Since a non-negative linear combination of positive semi-definite, symmetric matrices is symmetric and positive semi-definite, the claim follows. □

Remark 3.3. Since the total mass, $\sum_{s\in\mathcal{S}} c_t(s)$, is preserved in our case, we may replace the

negative free energy F in Theorem 3.4 by the *entropy*,

$$H(c) := -\sum_{s\in\mathcal{S}} c(s)\log\big(c(s)\big).$$

For the solution of the recombination equation (Eq. (2.5)) this has the following consequence. It is a well-known fact that, when considering the set of probability measures on a product space which all have the same marginals, the product measure of these marginals is a maximiser for the entropy. As the one-dimensional marginals are preserved under recombination (in absence of mutation or selection), the fact that Eq. (2.5) can be written as a generalised gradient system with respect to H reflects on the fact that the solution approaches linkage equilibrium; compare [Bür09, Theorem 3.1]. ♢

3.2.1 Explicit examples

Combining Theorems 3.4,3.2 and 3.3, for finite X, there exists a Function C, defined on $\mathcal{P}(X)$ with values in the symmetric positive semi-definite matrices such that

$$\dot{\omega}_t = C(\omega_t)\nabla F(\omega_t)$$

is equivalent to the recombination equation (2.5). Our goal is now to write down the function $\nu \mapsto C(\nu)$ for $\nu \in \mathcal{P}(X)$ explicitly for concrete examples. The most simple one is the classical case with two parents and two diallelic loci (compare [HM15, Ex. 1]). Then, we have the reaction

$$(0,0)+(1,1) \overset{\varrho}{\leftrightarrow} (1,0)+(0,1).$$

Identifying $(0,0)$ with the first, $(0,1)$ with the second, $(1,0)$ with the third and $(1,1)$ with the fourth basis vector in $\mathbb{R}^4$, the matrix $C(\nu)$, as constructed in the proof of Theorem 3.4 can be written as

$$\varrho L\big(\nu(0,0)\nu(1,1), \nu(1,0)\nu(0,1)\big) \begin{pmatrix} 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix},$$

where L is as in the proof of Theorem 3.4.

Next, we treat the slightly more complicated example of three diallelic loci (but still 2 parents); compare [HM15, Ex. 2]. Again, we denote the two alleles by 0 and 1. We denote the type (i_1, i_2, i_3) by $g_{4i_i+2i_2+i_3}$; in other words, the index of a type is just the type itself, read as a binary integer. For example, we refer to $(0,0,0)$ by g_0 and to $(1,0,1)$ by g_5, and identify g_i with the canonical $i+1$-th basis vector of $\mathbb{R}^8$.

Now, by the proof of Theorem 3.4, we associate to each reaction pair of the form

$$g_{i_1} + g_{i_2} \overset{\kappa}{\leftrightarrow} g_{j_1} + g_{j_2}, \tag{3.8}$$

an 8×8 matrix $M(\nu)$ with entries

$$M_{ij}(\nu) := \begin{cases} \kappa L(\nu(g_{i_1})\nu(g_{i_2}), \nu(g_{j_1})\nu(g_{j_2})), & \text{if } g_{i-1} \text{ and } g_{j-1} \text{ are on the same side of (3.8)}, \\ -\kappa L(\nu(g_{i_1})\nu(g_{i_2}), \nu(g_{j_1})\nu(g_{j_2})), & \text{if } g_{i-1} \text{ and } g_{j-1} \text{ are on different sides of (3.8)}, \\ 0, & \text{otherwise} \end{cases}$$

and $C(\nu)$ is then given by summing these matrices over all forward-backward reaction pairs in the network. To keep things tidy, instead of summing over all forward-backward reaction pairs, we write down the sums over each different linkage class seperately; this allows to take advantage of the following symmetry implied by our choice of indices. Namely, as 1s are only exchanged between gametes but their relative positions in the sequence remains unchanged, the sum of indices is the same for each complex that are in the same linkage class, of which there are seven; six consisting of only one forward-backward reaction pair each, and one consisting of six such pairs. Assume that M belongs to a reaction within a complex where the indices sum to ℓ. Then, it is easy to see that we have $M_{i,j} = M_{\ell-i+2,j} = M_{i,\ell-j+2} = M_{\ell-i+2,\ell-j+2}$. This means that, for ℓ odd, M is of the form

$$\left(\begin{array}{c|c|c} A & \phi A & 0 \\ \hline \ominus A & \ominus \phi A & 0 \\ \hline 0 & 0 & 0 \end{array}\right) \text{ if } \ell \leqslant 7 \text{ and } \left(\begin{array}{c|c|c} 0 & 0 & 0 \\ \hline 0 & \ominus \phi A & \ominus A \\ \hline 0 & \phi A & A \end{array}\right) \text{ for } \ell > 7,$$

where ϕ denotes the reversal of columns and $\ominus$ denotes the reversal of rows within a matrix and A is a $\frac{\ell+1}{2} \times \frac{\ell+1}{2}$ matrix if $\ell \leqslant 7$ and a $\frac{14-\ell+1}{2}$ matrix if $\ell > 7$. For ℓ even , M is of the form

$$\left(\begin{array}{c|c|c|c} A & 0 & \phi A & 0 \\ \hline 0 & 0 & 0 & 0 \\ \hline \ominus A & 0 & \ominus \phi A & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right) \text{ if } \ell \leqslant 7 \text{ and } \left(\begin{array}{c|c|c|c} 0 & 0 & 0 & 0 \\ \hline 0 & \ominus \phi A & 0 & \ominus A \\ \hline 0 & 0 & 0 & 0 \\ \hline 0 & \phi A & 0 & A \end{array}\right) \text{ for } \ell > 7,$$

where A is now an $\frac{\ell}{2} \times \frac{\ell}{2}$ matrix if $\ell \leqslant 7$ and $\frac{14-\ell}{2}$ if $\ell > 7$; Here, the extra 0 between the reflected copies of A comes from the fact that reactions of the form

$$g_i + g_i \overset{\kappa}{\leftrightarrow} g_i + g_i$$

do not contribute to the system. Let us now write these matrices A for the different linkage classes. We abbreviate the function $\nu \mapsto L(\nu(g_{i_1})\nu(g_{i_2}), \nu(g_{j_1}))\nu(g_{j_2}))$ by $L_{i_1 i_2, j_1 j_2}$. For all

$1 \leqslant i \leqslant 3$, ϱ_i denotes the recombination rate for the partition $\{\{i\}, \{1,2,3\} \setminus \{i\}\}$. For the first six linkage classes in [HM15, Ex. 2], each consisting of one reaction, we have in place of A

$$\begin{pmatrix} (\varrho_1+\varrho_2)L_{06,24} & 0 & -(\varrho_1+\varrho_2)L_{06,24} \\ 0 & 0 & 0 \\ -(\varrho_1+\varrho_2)L06,24 & 0 & (\varrho_1+\varrho_2)L_{06,24} \end{pmatrix}, \begin{pmatrix} (\varrho_1+\varrho_2)L_{17,35} & 0 & -(\varrho_1+\varrho_2)L_{17,35} \\ 0 & 0 & 0 \\ -(\varrho_1+\varrho_2)L_{17,35} & 0 & (\varrho_1+\varrho_2)L_{17,35} \end{pmatrix},$$
$$\begin{pmatrix} (\varrho_1+\varrho_3)L_{05,14} & -(\varrho_1+\varrho_3)L_{05,14} & 0 \\ -(\varrho_1+\varrho_3)L_{05,14} & (\varrho_1+\varrho_3)L_{05,14} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 \\ 0 & (\varrho_1+\varrho_3)L_{27,36} & -(\varrho_1+\varrho_3)L_{27,36} \\ 0 & -(\varrho_1+\varrho_3)L_{27,36} & (\varrho_1+\varrho_3)L_{27,36} \end{pmatrix},$$

representing the reactions $g_0 + g_6 \overset{\varrho_1+\varrho_2}{\longleftrightarrow} g_4 + g_2$, $g_1 + g_7 \overset{\varrho_1+\varrho_2}{\longleftrightarrow} g_5 + g_3$, $g_0 + g_5 \overset{\varrho_1+\varrho_3}{\longleftrightarrow} g_4 + g_1$, $g_2 + g_7 \overset{\varrho_1+\varrho_3}{\longleftrightarrow} g_6 + g_3$ and

$$\begin{pmatrix} (\varrho_2+\varrho_3)L_{03,12} & -(\varrho_2+\varrho_3)L_{03,12} \\ -(\varrho_2+\varrho_3)L_{03,12} & -(\varrho_2+\varrho_3)L_{03,12} \end{pmatrix}, \begin{pmatrix} (\varrho_2+\varrho_3)L_{47,56} & -(\varrho_2+\varrho_3)L_{47,56} \\ -(\varrho_2+\varrho_3)L_{47,56} & (\varrho_2+\varrho_3)L_{47,56} \end{pmatrix},$$

representing the reactions $g_0 + g_3 \overset{\varrho_2+\varrho_3}{\longleftrightarrow} g_2 + g_1$ and $g_4 + g_7 \overset{\varrho_2+\varrho_3}{\longleftrightarrow} g_5 + g_6$. Finally, the last linkage class, comprised of the six reactions $g_2 + g_5 \overset{\varrho_1}{\longleftrightarrow} g_6 + g_1$, $g_6 + g_1 \overset{\varrho_2}{\longleftrightarrow} g_4 + g_3$, $g_4 + g_3 \overset{\varrho_1}{\longleftrightarrow} g_0 + g_7$, $g_0 + g_7 \overset{\varrho_2}{\longleftrightarrow} g_2 + g_5$, $g_2 + g_5 \overset{\varrho_3}{\longleftrightarrow} g_4 + g_3$, $g_6 + g_1 \overset{\varrho_3}{\longleftrightarrow} g_0 + g_7$, is represented by

$$\begin{pmatrix} \varrho_1 L_{07,34}+\varrho_2 L_{07,25}+\varrho_3 L_{16,07} & -\varrho_3 L_{16,07} & -\varrho_2 L_{07,25} & -\varrho_1 L_{07,34} \\ -\varrho_3 L_{16,07} & \varrho_1 L_{16,25}+\varrho_2 L_{16,34}+\varrho_3 L_{16,07} & -\varrho_1 L_{16,25} & -\varrho_2 L_{16,34} \\ -\varrho_2 L_{25,07} & -\varrho_1 L_{25,16} & \varrho_1 L_{25,16}+\varrho_2 L_{25,07}+\varrho_3 L_{25,34} & -\varrho_3 L_{25,34} \\ -\varrho_1 L_{34,07} & -\varrho_2 L_{34,16} & -\varrho_3 L_{34,25} & \varrho_1 L_{34,07}+\varrho_2 L_{34,16}+\varrho_3 L_{34,25} \end{pmatrix}.$$

3.3 Markov chains with strictly monotone orbits and the partitioning process

We have seen how the result of Müller and Hofbauer [HM15] generalises in the setting of an arbitrary number of parents, at least for finite type spaces. For more general, potentially uncountable type spaces, this approach fails because it is not clear how to even make sense of the notion of the concentration of individual types, unless ω_t is pure point. Now, we show how the evolution of the law of the partitioning process, related to ω via Eq. (2.9) can be written as a gradient system. Ultimately, this is due to the monotonicity of its sample paths; recall from (2.10) that the transition rate from $\mathcal{A}$ to $\mathcal{B}$ vanishes whenever $\mathcal{B} \not\preccurlyeq \mathcal{A}$. In particular, the number of blocks increases strictly in each transition.

Definition 3.5. Let $\mathcal{X} = (\mathcal{X}_t)_{t\geqslant 0}$ be a continuous-time Markov chain on a finite state-space E, with rate matrix $\big(Q(i,j)\big)_{i,j\in E}$; it is called a *Markov chain with strictly monotone orbits* (MCsmo) (with respect to a real-valued function W on E) if $Q(i,j) > 0$ implies that $W(j) > W(i)$. ♢

Recall that the distribution of a finite-state Markov chain $\mathcal{X}$ can be interpreted as a probability

vector,

$$p_t^{\mathcal{X}} := \sum_{i \in E} p_t^{\mathcal{X}}(i) i,$$

which evolves in time according to the differential equation,

$$\dot{p}_t^{\mathcal{X}} = \sum_{i \in E} \sum_{j \in E} p_t^{\mathcal{X}}(i) Q(i,j)(j-i). \tag{3.9}$$

If $\mathcal{X}$ has strictly monotone orbits in the sense of Definition 3.5, Eq. (3.9) can be written as a generalised gradient system, as defined in Section 3.2.

Theorem 3.6. *Let $\mathcal{X}$ be a* MCsmo *with respect to W and define $\Psi : \mathcal{P}(E) \to \mathbb{R}$,*

$$\Psi(p) := \sum_{i \in E} p(i) W(i).$$

Then, Eq. (3.9)*, which describes the time evolution of $p^{\mathcal{X}} := \left(p_t^{\mathcal{X}}\right)_{t \geqslant 0}$, can be written as*

$$\dot{p}_t^{\mathcal{X}} = K(p_t^{\mathcal{X}}) \nabla \Psi(p_t^{\mathcal{X}}),$$

where K takes values in the symmetric, positive semi-definite matrices, is continuous on $\mathcal{P}(E)$ and smooth on its interior.

Proof. Define

$$K(p) := \sum_{\substack{i,j \in E \\ Q(i,j) > 0}} \frac{p(i) Q(i,j)}{W(j) - W(i)} (j-i)(j-i)^{\mathsf{T}}.$$

Since Ψ is linear with (constant) gradient

$$\nabla \Psi = \sum_{i \in E} W(i) i,$$

we have $(j-i)^{\mathsf{T}} \nabla \Psi = W(j) - W(i)$ and thus,

$$K(p) \nabla \Psi = \sum_{\substack{i,j \in E \\ Q(i,j) > 0}} \frac{p(i) Q(i,j)}{W(j) - W(i)} (W(j) - W(i))(j-i) = \sum_{i,j \in E} p(i) Q(i,j)(j-i).$$

Inserting $p_t^{\mathcal{X}}$ for p, this is exactly the right-hand side of Eq. (3.9). □

As mentioned before, the partitioning process is a process of successive refinement; in every non-silent transition, the number of blocks increases at least by one. Thus, it is a MCsmo with respect to the number of blocks.

Corollary 3.7. *The law p^{Σ} of the partitioning process with generator $\mathcal{Q}$ given in Eq.* (2.10)

satisfies a generalised gradient system with respect to N given by

$$N(p) = \sum_{\mathcal{A}\in\boldsymbol{P}(S)} p(\mathcal{A})|\mathcal{A}|.$$

Let us give an explicit example.

Example 3.1. Let us consider a Markov chain with 4 states A, B, C, D and jump rates $q(A, B) = q(A, C) = 1$ and $q(B, D) = q(C, D) = 2$. All other transition rates are 0. This is a Markov chain with strictly monotone orbits in the sense of Definition 3.5, with respect to W given by $W(A) = 1, W(B) = W(C) = 2, W(D) = 3$. Upon identifying A, B, C, D with the standard basis of $\mathbb{R}^4$, the linear differential equation describing the dynamics of its distribution reads

$$\dot{p}_t = \begin{pmatrix} -2 & 0 & 0 & 0 \\ 1 & -2 & 0 & 0 \\ 1 & 0 & -2 & 0 \\ 0 & 2 & 2 & 0 \end{pmatrix} p_t, \tag{3.10}$$

and can be rewritten as

$$\dot{p}_t = \begin{pmatrix} 2p_t(A) & -p_t(A) & -p_t(A) & 0 \\ -p_t(A) & p_t(A) + 2p_t(B) & 0 & -2p_t(B) \\ -p_t(A) & 0 & 2p_t(C) + p_t(A) & -2p_t(C) \\ 0 & -2p_t(B) & -2p_t(C) & 2p_t(C) + 2p_t(B) \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 2 \\ 3 \end{pmatrix}. \tag{3.11}$$

Here, the vector $(1, 2, 2, 3)^\mathsf{T}$ is the gradient (with respect to the euclidean metric) of

$$\Psi(p) = p(A) + 2p(B) + 2p(C) + 3p(D).$$

Also, the matrix is symmetric and it is positive semi-definite, as it can be written as a sum of positive semi-definite matrices (as long as $p(A), p(B), p(C) \geqslant 0$),

$$p(A) \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} -1 & 1 & 0 & 0 \end{pmatrix} + p(A) \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} -1 & 0 & 1 & 0 \end{pmatrix} + 2p(B) \begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & -1 & 0 & 1 \end{pmatrix} + 2p(C) \begin{pmatrix} 0 \\ 0 \\ -1 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & -1 & 1 \end{pmatrix},$$

evaluated at $p = p_t$. Thus, Eq. (3.11) is a generalised gradient system in the sense of Eq. (3.6). Note that the coefficient matrix in Eq. (3.10) is not diagonalisable; this is because the eigenvalue -2 has algebraic multiplicity 3, but the associated eigenspace is merely two-dimensional and spanned by $(0, 1, -1, 0)^\mathsf{T}$ and $(0, 1, 0, -1)^\mathsf{T}$. Its general solution will therefore contain terms of the form te^{-2t}. This seems to be in contradiction with the fact that generalised gradient system can not have resonant solutions of the form $t^k e^{\lambda t}$ for $k \geqslant 1$. One has to keep in mind, however, that the gradient representation only holds on the non-negative cone (which is invariant for the system). Note also that the problematic generalised eigenspace only has a trivial intersection with $\mathbb{R}^4_{\geqslant 0}$. ♢

Let us conclude this section with one additional example.

Example 3.2. Let us now consider the actual partitioning process, for three loci. We have the five partitions $\mathcal{A}_1 = \{\{1,2,3\}\}, \mathcal{A}_2 = \{\{1\},\{2,3\}\}, \mathcal{A}_3 = \{\{1,3\},\{2\}\}, \mathcal{A}_4 = \{\{1,2\},\{3\}\}$ and $\mathcal{A}_5 = \{\{1\},\{2\},\{3\}\}$. Identifying $\mathcal{A}_i$ with the i-th basis vector in $\mathbb{R}^5$, the generator $\mathcal{Q}$ of the partitioning process (compare Eq. (2.10)) reads

$$\begin{pmatrix} -\varrho_1-\varrho_2-\varrho_3 & \varrho_1 & \varrho_2 & \varrho_3 & 0 \\ 0 & -\varrho_2-\varrho_3 & 0 & 0 & \varrho_2+\varrho_3 \\ 0 & 0 & -\varrho_1-\varrho_3 & 0 & \varrho_1+\varrho_3 \\ 0 & 0 & 0 & -\varrho_1-\varrho_2 & \varrho_1+\varrho_2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

where $\varrho_1, \varrho_2, \varrho_3$ are as in Subsection 3.2.1, and he gradient system then for the distribution p_t^Σ then reads

$$\dot{p}_t^\Sigma = \begin{pmatrix} D_1 & -p_t(\mathcal{A}_1)\varrho_1 & -p_t(\mathcal{A}_1)\varrho_2 & -p_t(\mathcal{A}_1)\varrho_3 & 0 \\ -p_t(\mathcal{A}_1)\varrho_1 & D_2 & 0 & 0 & -p_t(\mathcal{A}_2)(\varrho_2+\varrho_3) \\ -p_t(\mathcal{A}_1)\varrho_2 & 0 & D_3 & 0 & -p_t(\mathcal{A}_3)(\varrho_1+\varrho_3) \\ -p_t(\mathcal{A}_1)\varrho_3 & 0 & 0 & D_4 & -p_t(\mathcal{A}_4)(\varrho_1+\varrho_2) \\ 0 & -p_t(\mathcal{A}_2)(\varrho_2+\varrho_3) & -p_t(\mathcal{A}_3)(\varrho_1+\varrho_3) & -p_t(\mathcal{A}_4)(\varrho_1+\varrho_2) & D_5 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 2 \\ 2 \\ 3 \end{pmatrix},$$

where $D_1, \dots, D_5$ are chosen such that the rows sum to 0 and $(1,2,2,3)^\mathsf{T}$ is the gradient ∇N of the mean number of blocks N, defined in Corollary 3.7. Again, the maximum of the potential, the partition $\{\{1\},\{2\},\{3\}\}$ characterises linkage equilibrium ('all sites come from independent ancestors'). ♢

3.4 Nonlinear partitioning as a chemical reaction network

We have seen in the previous chapter that the evolution of the law of the partitioning process can be rewritten as a linear generalised gradient system. We now consider the nonlinear system from Theorem 2.7. We will see that it, too, can be interpreted as the law of mass action for a network of chemical reactions between the partitions of S. Its construction is very similar to the network from Section 3.1.

To motivate this result, imagine that at time $t = 0$, we paint every gamete in a different color. As described in Theorem 3.5 and Fig. 3.1, for every $\mathcal{C} \in \boldsymbol{P}(S)$, every randomly chosen $|\mathcal{C}|$-tuple of gametes undergoes a chemical reaction as in Eq. (3.5) at rate $\frac{\varrho(\mathcal{C})}{|\mathcal{C}|}$. But now, instead of investigating the effect on the type distribution, we ask how the initially assigned colors are mixed in the process. To this end, we attach to each individual a partition of its sites by grouping together all sites with the same color.

Now, consider the j-th gamete that results from such a reaction (compare Eq. (3.5)); for two sites k and ℓ in this individual to have the same color, they must come from the same individual on the left-hand side (this is due to the fact that the tuple was chosen randomly and, as there are infinitely many colors in the population, the probability that the same color occurs in more than one individual in the chosen sample is negligible). More formally, there must be an i between 1 and $|\mathcal{C}|$ such that k and ℓ are both in C_i. If that is true, both sites come from the $i + j - 1$-th individual, and thus must share the same block of $\mathcal{A}_{i+j-1}$. Put

more concisely, this means that k and ℓ belong to the same block of the induced partition $\mathcal{A}_{i+j-1}|_{C_i}$ for some $i \in \{1, \dots, |\mathcal{C}|\}$. Equivalently, this means that the partition that describes the coloring of the j-th product gamete is given precisely by

$$\bigcup_{i=1}^{|\mathcal{C}|} \mathcal{A}_{i+j-1}|_{C_i}.$$

For an illustration, see Fig. 3.2. Thus, the reaction network from Section 3.1 translates to the system consisting of the reactions

$$\sum_{j=1}^{|\mathcal{C}|} \mathcal{A}_j \xrightarrow{\frac{\varrho(\mathcal{C})}{|\mathcal{C}|}} \sum_{j=1}^{|\mathcal{C}|} \bigcup_{i=1}^{|\mathcal{C}|} \mathcal{A}_{i+j-1}|_{C_i}, \tag{3.12}$$

one for each $\mathcal{C}$ and every $|\mathcal{C}|$-tuple of partitions of S; as always, indices are to be read mod $|\mathcal{C}|$. These reactions are of the same form as the ones between gametes in Eq. (3.5), after replacing the type fragments $\pi_{C_i}(x^{(i+j-1)})$ with the induced partitions $\mathcal{A}_{i+j-1}|_{C_i}$.

We finish by showing that the law of mass action of this chemical reaction network is precisely the nonlinear system from Theorem 2.7.

Theorem 3.8. *The nonlinear system of ordinary differential equations that describes the dynamics of the coefficients in* (2.7) *can be written as the law of mass action for the* CRN *comprised of all reactions* (3.12)*. More concisely, the system from Theorem 2.7 is equivalent to*

$$\dot{a}_t = \sum_{\mathcal{C}} \sum_{\mathcal{A}_1, \dots, \mathcal{A}_{|\mathcal{C}|}} \frac{\varrho(\mathcal{C})}{|\mathcal{C}|} a_t(\mathcal{A}_1) \cdot \ldots \cdot a_t(\mathcal{A}_{|\mathcal{C}|}) \left(\sum_{j=1}^{|\mathcal{C}|} \left(\bigcup_{i=1}^{|\mathcal{C}|} \mathcal{A}_{i+j-1}|_{C_i} - \mathcal{A}_j \right) \right),$$

where the summation is over $\boldsymbol{P}(S)$*.*

Proof. We will use the following identity (the proof of which will conclude the proof of the theorem),

$$\prod_{i=1}^{|\mathcal{B}|} \sum_{\substack{\mathcal{C} \in \boldsymbol{P}(S) \\ \mathcal{C}|_{B_i} = \mathcal{A}|_{B_i}}} a(\mathcal{C}) = \frac{1}{|\mathcal{B}|} \sum_{j=1}^{|\mathcal{B}|} \sum_{\mathcal{A}_1, \dots, \mathcal{A}_{|\mathcal{B}|}} \delta\Big(\mathcal{A}, \bigcup_{i=1}^{|\mathcal{B}|} \mathcal{A}_{i+j-1}|_{B_i}\Big) \cdot a(\mathcal{A}_1) \cdot \ldots \cdot a(\mathcal{A}_{|\mathcal{B}|}), \tag{3.13}$$

valid for all $\mathcal{B} \succcurlyeq \mathcal{A}$ and all $a \in \mathbb{R}^{\boldsymbol{P}(S)}$, where $\mathcal{B} = \{B_1, \dots, B_{|\mathcal{B}|}\}$. Inserting (3.13), we see that the second sum on the right-hand side of Eq. (2.8),

$$\sum_{\mathcal{A}} \sum_{\mathcal{B} \succcurlyeq \mathcal{A}} \left(\prod_{i=1}^{|\mathcal{B}|} \sum_{\substack{\mathcal{C} \in \boldsymbol{P}(S) \\ \mathcal{C}|_{B_i} = \mathcal{A}|_{B_i}}} a_t(\mathcal{C}) \right) \varrho(\mathcal{B}) \mathcal{A},$$

can be written as

$$\sum_{\mathcal{A}} \sum_{\mathcal{B} \succcurlyeq \mathcal{A}} \left(\frac{\varrho(\mathcal{B})}{|\mathcal{B}|} \sum_{j=1}^{|\mathcal{B}|} \sum_{\mathcal{A}_1,\dots,\mathcal{A}_{|\mathcal{B}|}} \delta\Big(\mathcal{A}, \bigcup_{i=1}^{|\mathcal{B}|} \mathcal{A}_{i+j-1}|_{B_i}\Big) a_t(\mathcal{A}_1) \cdot \ldots \cdot a_t(\mathcal{A}_{|\mathcal{B}|}) \mathcal{A} \right). \tag{3.14}$$

Notice that the second argument of the Kronecker function is always finer than $\mathcal{B}$. Thus, the whole summand vanishes whenever $\mathcal{B} \succcurlyeq \mathcal{A}$ does not hold. We may therefore ignore the restriction $\mathcal{B} \succcurlyeq \mathcal{A}$ in the inner sum, which allows us then to change the order of summation. After using the Kronecker function to perform the summation with respect to $\mathcal{A}$, what remains is

$$\sum_{\mathcal{B}} \frac{\varrho(\mathcal{B})}{|\mathcal{B}|} \sum_{\mathcal{A}_1,\dots,\mathcal{A}_{|\mathcal{B}|}} a(\mathcal{A}_1) \cdot \ldots \cdot a(\mathcal{A}_{|\mathcal{B}|}) \sum_{j=1}^{|\mathcal{B}|} \bigcup_{i=1}^{|\mathcal{B}|} \mathcal{A}_{i+j-1}|_{B_i}.$$

Up to renaming $\mathcal{B}$ with $\mathcal{C}$, this is exactly the first part of the law of mass action for the CRN described above. Using the same argument as in the proof of Theorem 3.2, the first sum in Eq. (2.8),

$$-\sum_{\mathcal{B}} \varrho(\mathcal{B}) \sum_{\mathcal{A}} a(\mathcal{A}) \mathcal{A},$$

can be rewritten as

$$-\sum_{\mathcal{B}} \frac{\varrho(\mathcal{B})}{|\mathcal{B}|} \sum_{\mathcal{A}_1,\dots,\mathcal{A}_{|\mathcal{B}|}} a_t(\mathcal{A}_1) \cdot \ldots \cdot a_t(\mathcal{A}_{|\mathcal{B}|}) (\mathcal{A}_1 + \ldots + \mathcal{A}_{|\mathcal{B}|}).$$

Up to renaming $\mathcal{B}$ with $\mathcal{C}$, this completes the proof, provided Eq. (3.13) is correct. To show this, we start by expanding the right hand side,

$$\begin{aligned} \prod_{i=1}^{|\mathcal{B}|} \sum_{\substack{\mathcal{C} \in \boldsymbol{P}(S) \\ \mathcal{C}|_{B_i} = \mathcal{A}|_{B_i}}} a(\mathcal{C}) &= \sum_{(\mathcal{A}_1,\dots,\mathcal{A}_{|\mathcal{B}|}) \in \mathcal{G}(\mathcal{A})} a(\mathcal{A}_1) \cdot \ldots \cdot a(\mathcal{A}_{|\mathcal{B}|}) \\ &= \sum_{\mathcal{A}_1,\dots,\mathcal{A}_{|\mathcal{B}|}} \delta\Big(\mathcal{A}, \bigcup_{i=1}^{|\mathcal{B}|} \mathcal{A}_i|_{B_i}\Big) a(\mathcal{A}_1) \cdot \ldots \cdot a(\mathcal{A}_{|\mathcal{B}|}) \end{aligned}$$

where $\mathcal{G}(\mathcal{A})$ is the set of all $|\mathcal{B}|$-tupels $(\mathcal{A}_1, \dots, \mathcal{A}_{|\mathcal{B}|})$ of partitions with $\mathcal{A}_i|_{B_i} = \mathcal{A}|_{B_i}$. Since $\mathcal{A} \preccurlyeq \mathcal{B}$ implies that

$$\mathcal{A} = \mathcal{A}|_{B_1} \cup \ldots \cup \mathcal{A}|_{B_{|\mathcal{B}|}},$$

$(\mathcal{A}_1, \dots, \mathcal{A}_{|\mathcal{B}|}) \in \mathcal{G}(\mathcal{A})$ if and only if

$$\mathcal{A} = \bigcup_{i=1}^{|\mathcal{B}|} \mathcal{A}_i|_{B_i}.$$

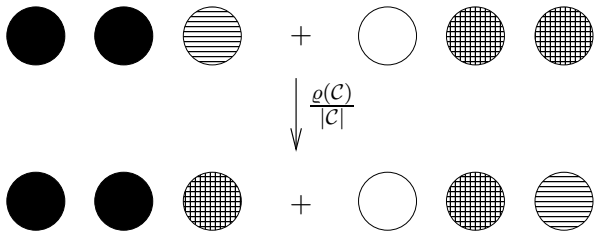


Figure 3.2. A reaction between two gametes with 3 loci, corresponding to the partition $\mathcal{C} = \{\{1,2\},\{3\}\}$. This means that the leading two sites of the left gamete on the top is combined with the trailing third site of the gamete on the right, and the leading two sites of the gamete to the right are combined with the trailing third of the left one. Here, the coloring of the sites is represented by different patterns. The partitions associated with the gametes are as follows. For the substrate complex (top), we have $\mathcal{A}_1 = \{\{1,2\},\{3\}\}$ and $\mathcal{A}_2 = \{\{1\},\{2,3\}\}$, and the product complex (bottom) consists of $\mathcal{A}_1|_{C_1} \cup \mathcal{A}_2|_{C_2} = \{\{1,2\}\} \cup \{\{3\}\} = \{\{1,2\},\{3\}\}$ and $\mathcal{A}_1|_{C_2} \cup \mathcal{A}_2|_{C_1} = \{\{3\}\} \cup \{\{1\},\{2\}\} = \{\{1\},\{2\},\{3\}\}$.

Now, as in the proof of Theorem 3.2, we replace the product (for $1 \leqslant j \leqslant |\mathcal{B}|$)

$$a(\mathcal{A}_1) \cdot \ldots \cdot a(\mathcal{A}_{|\mathcal{B}|})$$

by

$$a(\mathcal{A}_{1-j+1}) \cdot \ldots \cdot a(\mathcal{A}_{|\mathcal{B}|-j+1})$$

and subsequently rename the summation indices. Thus,

$$\begin{aligned} &\sum_{\mathcal{A}_1,\ldots,\mathcal{A}_{|\mathcal{B}|}} \delta\Big(\mathcal{A}, \bigcup_{i=1}^{|\mathcal{B}|} \mathcal{A}_i|_{B_i}\Big) a(\mathcal{A}_1) \cdot \ldots \cdot a(\mathcal{A}_{|\mathcal{B}|}) \\ &= \frac{1}{|\mathcal{B}|} \sum_{j=1}^{|\mathcal{B}|} \sum_{\mathcal{A}_1,\ldots,\mathcal{A}_{|\mathcal{B}|}} \delta\Big(\mathcal{A}, \bigcup_{i=1}^{|\mathcal{B}|} \mathcal{A}_{i+j-1}|_{B_i}\Big) a(\mathcal{A}_1) \cdot \ldots \cdot a(\mathcal{A}_{|\mathcal{B}|}), \end{aligned}$$

which finishes the proof of Eq. (3.13) and hence, of the theorem. □

Despite their similar appearance, there is one crucial difference between the CRN from Section 3.1, and the one above. Because the products are pieced together from partitions of subsets induced by the substrates, the total number of blocks on the right-hand side is in general strictly larger than on the left-hand side. This implies that this network is *not* reversible, and the question whether it can be interpreted as a gradient system remains open. The loss of reversibility appears to be the coarse-graining of the information in our system that we performed by transitioning from the (potentially infinite) set of types to the finite set of partitions. This is vaguely reminiscent of the common phenomenon in statistical mechanics where the projection of the underlying (high-dimensional) microscopic model to a smaller set of macroscopic degrees of freedom leads to a loss of reversibility.

4 Ancestral lines under selection and recombination

After investigating aspects of the dynamics of pure recombination in the previous chapter, we now want to attack the *selection-recombination equation*, which describes evolution under the joint action of recombination *and selection*. Here, selection means that fit individuals flourish at the expense of less fit ones. This equation first appeared in the literature in a paper by Kimura [Kim65] in 1956 and has since been studied intensely; see [Bür00, Ch. II] for a comprehensive review. The selection-recombination dynamics is more complex than that of pure recombination; in particular, it displays Hopf bifurcations and stable limit cycles in certain parameter regimes [Aki74]. Much research has been devoted to the case where recombination is much faster than selection, so that time-scale separation applies and the dynamics is confined to a specific manifold [NHB99].

In this chapter, our goal is to understand the selection-recombination equation with one selected site and single crossovers, to provide a systematic and transparent approach that also generalises to an *arbitrary* number of sites, and to establish an *exact* solution via a recursion. We do this by extending the approach used in [BBS16; BB16] for the pure recombination equation; namely, we trace back the (potential) ancestral lines of individuals in the current population, this time by a variant of the *ancestral selection-recombination graph* [DK99; LK12; BP18] for an arbitrary number of sites. This gives rise to a Markov process on the set of *weighted* partitions of the set of sequence sites; this process is dual to the selection-recombination equation. The corresponding Markov semigroup is available in closed form, and the resulting stochastic representation yields deep insight into the genealogical content of the solution of the differential equation. Moreover, it gives access to the long-term behaviour.

This chapter is organised as follows. Sections 4.1 and 4.2 introduce the selection-recombination equation, both in its own right and in terms of a dynamical law of large numbers of the corresponding Moran model, which describes a *finite* population under selection and recombination and, via its graphical construction, provides the foundation for the genealogical arguments to follow later. In Section 4.3, we revisit *marginalisation consistency*, which, in the presence of selection, is more subtle and only true for certain subsets, but all the more interesting. A recursive integral representation of the solution is given in Section 4.4. The core of this chapter consists of Sections 4.5 and 4.6, where we construct the backward process and provide the genealogical argument behind our recursion, together with Section 4.7, where the dual process is formulated and the formal duality result is proved. Finally, the explicit solution is presented in Section 5.6, and its long-term behaviour is investigated.

4.1 The selection-recombination equation

The selection-recombination equation is based on the pure recombination equation (compare Eq. (2.5)) in continuous time, together with an additional selection term, which we will explain later.

We restrict our attention to the special case of *single-crossover*. This means that in (2.5), we assume that $\varrho_{\mathcal{A}} = 0$ whenever $\mathcal{A}$ is *not* an interval partition into two parts, i.e. whenever $\mathcal{A}$ is *not* of the form $\{[1 : i], [i+1 : n]\}$ for some $i \in S$, where $[a : b]$ denotes the discrete interval $\{a, a+1, \ldots, b-1, b\}$, which is empty if $b < a$. Moreover, we assume that $X_i = \{0, 1\}$, that is, there are two alleles at each locus. Thus, the selection-recombination equation has the form

$$\dot{\omega}_t = \Psi_{\mathrm{rec}}(\omega_t) + \Psi_{\mathrm{sel}}(\omega_t), \tag{4.1}$$

where

$$\Psi_{\mathrm{rec}}(\omega_t) := \sum_{\mathcal{A} \in \boldsymbol{P}(S)} \varrho_{\mathcal{A}} (\mathcal{R}_{\mathcal{A}} - \mathrm{id})(\omega_t)$$

is the right-hand side of Eq. (2.5).

Remark 4.1. We remark that our assumption of single-crossover recombination is in line with the biological reality. In each generation, crossover events are typically rare and approximately independent; thus, the probability of multiple simultaneous crossovers is negligible. ♢

To explain the selection term $\Psi_{\mathrm{sel}}(\omega_t)$, we start by fixing a site $1 \leqslant i_* \leqslant n$, which we will refer to as the *selected site*, and set $S^* := S \setminus \{i_*\}$ (note that $\mathrm{card}(S^*) = n - 1$). An individual of type $x \in X$ is deemed to be *fit* or *of beneficial type* if $x_{i_*} = 0$ and *unfit* or *of deleterious type* otherwise, regardless of the letters at all other sites.

Selection then works as follows. Unfit individuals produce offspring at rate 1, while fit individuals reproduce at a higher rate $1 + s$, $s > 0$. Put differently, every individual, regardless of its type, has the neutral reproduction rate 1, while the fit individuals have an additional (selective) rate s. Upon reproduction, every offspring individual replaces another randomly chosen individual from the population so that the total population size remains constant. The net effect of the difference in reproduction rate between the fit and unfit type is that, in every infinitesimal time step, a certain (infinitesimal) fraction of the population is replaced by offspring of the subpopulation of fit individuals.

Let us be more precise. We write

$$f(\nu) := \nu\big(\pi_{i_*}^{-1}(0)\big) = \nu^{\{i_*\}}(0) \tag{4.2}$$

for the proportion of fit individuals in a population with type distribution ν, and the *selection operator* $F : \mathcal{P}(X) \to \mathcal{P}(X)$ via

$$F(\nu)(x) = (1 - x_{i_*})\nu(x). \tag{4.3}$$

Taking advantage of the interpretation of $\mathcal{M}(X)$ as a tensor product (compare the discussion after Remark 2.1), the selection operator can also be written as

$$F = P_{i_*} \otimes \mathrm{id}_{S^*}, \tag{4.4}$$

where $P_{i_*} := \delta_0 \delta_0^{\mathsf{T}}$ (recall that measures are interpreted as *column* vectors). Here, the subscripts indicate the site(s) at which the matrices act.

In words, F is the orthogonal projection to the subspace spanned by all elements of the form

$$\delta_0 \otimes v \text{ with } v \in \bigotimes_{i \in S^*} \mathcal{M}(X_i) = \mathcal{M}(X_{S^*}).$$

Furthermore, we define $b(\nu)$ and $d(\nu)$ via

$$f(\nu) b(\nu) = F\nu \tag{4.5}$$

and

$$\bigl(1 - f(\nu)\bigr) d(\nu) = (1 - F)\nu, \tag{4.6}$$

respectively (thus averting the danger of division by zero); here and in what follows, we write $F\nu$ instead of $F(\nu)$ when there is no risk of confusion. The measure $b(\nu)$ (the measure $d(\nu)$) is the type distribution in the beneficial (deleterious) subpopulation.

Finally, we assume that, in each infinitesimal time interval $[t, t+\mathrm{d}t]$, the number of individuals replaced by offspring of the subpopulation of fit individuals is proportional to $f(\omega_t)$. Thus the selection term is given by

$$\Psi_{\mathrm{sel}}(\omega_t) = s f(\omega_t) \bigl(b(\omega_t) - \omega_t\bigr), \tag{4.7}$$

where $b(\omega_t)$ is as in (4.5), and we refer to the proportionality factor s as the *selection intensity*.

Remark 4.2. The additive structure (4.1) of the selection-recombination is a consequence of the independence of recombination and selection and reflects the assumption that both selection and recombination are rare, so that one can neglect the possibility that recombination happens during selective reproduction; see Remark 4.3 below, and [Hof85] for the worked argument in the analogous case of the selection-mutation equation. ♢

To better exploit the additional structure we gain by restricting to single crossover, we introduce the following partial order on S.

Definition 4.1. For two sites $i, j \in S$, we say that i precedes j, or $i \preccurlyeq j$, if either $i_* \leqslant i \leqslant j$ or $j \leqslant i \leqslant i_*$. We write $i \prec j$ if $i \preccurlyeq j$ and $i \neq j$. We furthermore define the *i-tail* as the set

$$D_i := \{j \in S \mid i \preccurlyeq j\}$$

of all sites that succeed i, including i itself. We define the *i-head* C_i to be the complement

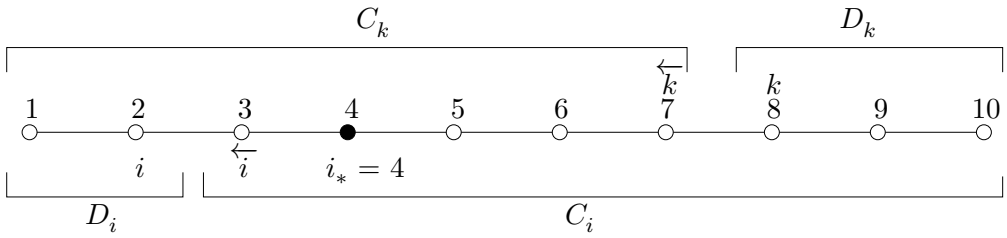


Figure 4.1. A sequence of length 10 with selected site, and two examples of predecessor, head, and tail; see text for more.

of the i-tail, $C_i := S \setminus D_i = \overline{D_i}$ (throughout, the overbar will denote the complement with respect to S); see Figure 4.1. Note that $D_{i_*} = S$ and $C_{i_*} = \varnothing$. Finally, if $i \neq i_*$, we denote by $\overleftarrow{i}$ the *predecessor* of i; that is, the maximal $j \in S$ with $j \prec i$ (note that $\overleftarrow{i} = i_*$ is possible). ♢

Remark 4.1. 1. Let us stress the fact that, in this chapter, $\preccurlyeq$ now defines a partial order on S rather than refinement of partitions.

2. The definition may appear awkward in that $i_* \in D_{i_*}$ but $i_* \in C_i$ for $i \in S^*$. However, it will become clear in Section 4.7 why this is exactly the way it must be.

3. In the limiting case $s = 0$, we may single out any site as the selected one; say $i_* = n$, so that $D_i = [1 : i]$ and $C_i = [i + 1 : n]$. ♢

For $i \in S^*$, we now define a slightly different version of the recombinator $R_i : \mathcal{P}(X) \to \mathcal{P}(X)$ by

$$R_i(\nu) := \mathcal{R}_{\{C_i, D_i\}}(\nu) = \nu^{C_i} \otimes \nu^{D_i}, \tag{4.8}$$

with the notation of (2.2) and (2.3); we will also write $R_i\nu$ instead of $R_i(\nu)$. Note the use of 'R' rather than '$\mathcal{R}$'. With this, Eq. (4.1) now reads

$$\boxed{\dot{\omega}_t = \sum_{i \in S^*} \varrho_i (R_i - \mathrm{id})\omega_t + s\big(F - f(\omega_t)\big)\omega_t = \Psi_{\mathrm{rec}}(\omega_t) + \Psi_{\mathrm{sel}}(\omega_t)} \tag{4.9}$$

with *recombination rates* $\varrho_i \geqslant 0$ for $i \in S^*$; for consistency, we set $\varrho_{i_*} := 0$. This means that, for $i < i_*$ ($i > i_*$), a single-crossover event takes place between sites i and $i + 1$ (sites $i - 1$ and i) with rate ϱ_i; in any case, we say that recombination happens at site i. This way, we address the *links* between neighbouring sites, as in [BB03], but in a way that depends on the location of the selected site.

4.2 The Moran model with selection and recombination

In order to gain a better understanding of Eq. (4.9) and to prepare for the genealogical arguments to follow in Section 4.5, we briefly recall the Moran model with selection and recombination. This is a *stochastic* model that describes selection and recombination in a *finite* population, from which (4.9) is recovered via a dynamical law of large numbers. We will use the representation as an *interacting particle system* (IPS). The Moran IPS works with N individuals, labelled $1 \leqslant \alpha \leqslant N$, each equipped with a (random) type $\Xi_t(\alpha) \in X$ (of (2.1)) at time t, which behaves as follows.

- Every individual β reproduces asexually at a fixed rate according to its fitness. That is, unfit individuals reproduce at rate 1 whereas fit individuals reproduce at rate $1+s$, where $s > 0$ is again the selection intensity. Upon reproduction, the single offspring inherits the parent's type and replaces a uniformly chosen individual α in the population (possibly its own parent). We will realise the different reproduction rates of the two types by distinguishing between *neutral reproduction events*, which happen at rate 1 to all individuals regardless of their type, and *selective reproduction events*, which are additionally performed by fit individuals at rate s. This distinction is a crucial ingredient in the ancestral selection graph [KN97].
- At rate ϱ_i, $i \in S^*$, individual β reproduces sexually, choosing a partner γ uniformly at random, possibly β itself. (Biologically, this means that we include the possibility of *selfing*.) The offspring is of type $(\Xi_{C_i}(\beta), \Xi_{D_i}(\gamma))$ and replaces another uniformly chosen individual α, possibly one of its own parents.

Formally, we can thus define the Moran IPS as a continuous-time Markov chain with states $\xi = (\xi(\alpha))_{1\leqslant\alpha\leqslant N} \in X^N$ and the following transitions when $\Xi_t = (\Xi_t(\alpha))_{1\leqslant\alpha\leqslant N} = \xi$:

$$\xi \to \xi^{\alpha\leftarrow\beta}_{\text{neut}} \text{ at rate } \frac{1}{N} \text{ for all } 1 \leqslant \alpha, \beta \leqslant N, \tag{4.10}$$

$$\xi \to \xi^{\alpha\leftarrow\beta}_{\text{sel}} \text{ at rate } \frac{s}{N} \text{ for all } 1 \leqslant \alpha, \beta \leqslant N, \quad \text{and} \tag{4.11}$$

$$\xi \to \xi^{\alpha\leftarrow(\beta,\gamma,i)}_{\text{rec}} \text{ at rate } \frac{\varrho_i}{N^2} \text{ for all } 1 \leqslant \alpha, \beta, \gamma \leqslant N \text{ and } i \in S^* \tag{4.12}$$

where, for $1 \leqslant \varepsilon \leqslant N$, the new state vectors explicitly read

$$\xi^{\alpha\leftarrow\beta}_{\text{neut}}(\varepsilon) = \begin{cases} \xi(\beta), & \varepsilon = \alpha, \\ \xi(\varepsilon), & \text{otherwise,} \end{cases} \qquad \xi^{\alpha\leftarrow\beta}_{\text{sel}}(\varepsilon) = \begin{cases} \xi(\beta), & \varepsilon = \alpha \text{ and } \xi_{i_*}(\beta) = 0, \\ \xi(\varepsilon), & \text{otherwise,} \end{cases} \tag{4.13}$$

and

$$\xi^{\alpha\leftarrow(\beta,\gamma,i)}_{\text{rec}}(\varepsilon) := \begin{cases} (\xi_{C_i}(\beta), \xi_{D_i}(\gamma)), & \varepsilon = \alpha, \\ \xi(\varepsilon), & \text{otherwise.} \end{cases}$$

Remark 4.3. The reader may wonder at this point why we include both sexual and asexual reproduction in our model. However, the 'asexual' reproduction events are actually sexual ones in which no recombination has occurred; that is, $C = \varnothing$ and $D = S$, so the offspring is a full copy of the first parent, and the second parent is irrelevant. Selective reproduction never occurs together with recombination due to the independence built into the SRE. ♢

For our purpose, it is particularly profitable to take advantage of the *graphical representation* of the Moran IPS, see Figure 4.2. Here, every individual is represented by a horizontal line, lines are labelled $1 \leqslant \alpha \leqslant N$ from bottom to top, and reproduction events are represented by arrows between the lines with the parent at the tail, the offspring at the tip, and the offspring replacing the individual at the target line (arrows pointing to their own tails have no effect and are omitted). In line with (4.13) and for reasons to become clear when taking the ancestral perspective in Section 5, we distinguish two types of arrows: *neutral* arrows (with normal arrowheads), which appear between every ordered pair of lines at rate $1/N$ regardless of the types of the lines; and *selective* arrows (with star-shaped arrowheads), which are laid down at rate s/N between every ordered pair of lines, again regardless of the types. Similarly, a recombination event in which the individual at line α is replaced by the joint offspring of lines β and γ is encoded as a square (on the α-th line) in which the recombination site i is inscribed and which is accompanied by two arms connecting to the parents and labelled C or D, indicating which of the parents contributes the i-head and i-tail, respectively. These graphical elements appear at rate ϱ_i/N^2 for every ordered triple of lines and every $i \in S^*$. If both arms connect to the same parent, the recombination event turns into a neutral reproduction event.

Remark 4.4. In view of this graphical construction, another perspective on the transition rates in the Moran IPS is natural. We can say that, with rates ϱ_i, each individual is replaced by the joint offspring of two uniformly chosen parents with the crossover point at site i. Likewise, at rate 1, each individual is replaced by the offspring of a *single* uniformly chosen parent individual; and with rate $s|\{1 \leqslant \alpha \leqslant N : \Xi_{i_*,t}(\alpha) = 0\}|$, it is replaced by the offspring of a parent individual chosen uniformly from the subset of fit individuals. This point of view will be particularly useful when looking back in time in Section 4.5. ♢

The fact that we use different kinds of arrows for the two types of reproduction events (rather than simply letting fit individuals shoot reproduction arrows at a faster rate) reflects the distinction between neutral and selective reproduction. The advantage of this strategy is that it allows for an *untyped* construction of the Moran IPS; that is, we first lay down the graphical elements between the lines regardless of the types and only then assign an initial type configuration. This type configuration is finally propagated forward in time under the rule that only individuals of beneficial type use the selective arrows to place their offspring, while neutral arrows and the arms of recombination events are used by all individuals, regardless of type.

Consider now the process $Z^{(N)} := (Z_t^{(N)})_{t\geqslant 0}$, where $Z_t^{(N)}$ is the empirical measure

$$Z_t^{(N)} := \frac{1}{N}\sum_{\alpha=1}^{N} \delta_{\Xi_t(\alpha)};$$

Proposition 3.1 in [Cor17b] in combination with Theorem 2.1 from [Ess16] (see also [BEP16]) shows that, as $N \to \infty$ without rescaling of parameters or time, the processes $Z^{(N)}$ converge almost surely locally uniformly to the solution $\omega = (\omega_t)_{t\geqslant 0}$ of the deterministic SRE (4.9) for every finite time horizon, whenever $Z_0^{(N)}$ converges to ω_0. This is because the Moran models, indexed with population size, form a density-dependent family, for which a dynamical law of large numbers applies; see [EK86, Thm. 3.2, Ch. 11].

For completeness, we will quote this theorem and see how it applies to our situation. In the terminology of [EK86], a *density-dependent family* is a sequence $(X^{(N)})_{N\geqslant 1}$ of Markov chains in continuous time, each defined on its own state space $E^{(N)}$ which is assumed to be of the form

$$E^{(N)} = E \cap \frac{1}{N}\mathbb{Z}^d,$$

where E is some subset of $\mathbb{R}^d$. It is further assumed that there exists a collection $(\beta_\ell)_{\ell\in\mathbb{Z}^d}$ of non-negative functions on E which describe the transition rates $q^{(N)}(e_1, e_2)$ of these Markov chains via

$$q^{(N)}(e_1, e_2) = N\beta_{N(e_2-e_1)}(e_1).$$

For such a family of Markov chains, the following theorem holds.

Theorem 4.2 ([EK86, Thm. 3.2, Ch. 11])**.** *Let $(X^{(N)})_{N\geqslant 1}$ be a density-dependent family as above. Suppose that, for each compact $K \subseteq E$,*

$$\sum_{\ell\in\mathbb{Z}^d} |\ell| \sup_{e_1\in K} \beta_\ell(e_1) < \infty \tag{4.14}$$

and that $G := \sum_{\ell\in\mathbb{Z}^d} \ell\beta_\ell$ is Lipschitz continuous. If $X_0^{(N)} \to x_0 \in E$ as $N \to \infty$, then for every $t \geqslant 0$

$$\lim_{N\to\infty} \sup_{0\leqslant s\leqslant t} |X_s^{(N)} - X_s| = 0 \text{ almost surely,}$$

where X is the unique solution of

$$\dot{X}_t = G(X_t) \tag{4.15}$$

with initial condition $X(0) = x_0$. □

Remark 4.5. Clearly, Eq. (4.14) ensures that G is a well-defined continuous function on E. Implicitly, we assume that a global solution to (4.15) exists; its uniqueness is guaranteed by the assumed Lipschitz continuity of G via the Picard–Lindelöf Theorem. ♢

For us, the role of E will be played by the simplex $\mathcal{P}(X)$ of probability measures, which can be thought of as a subset of $\mathbb{R}^{2^n}$. Clearly, as $Z^{(N)}$ records the relative frequencies of types

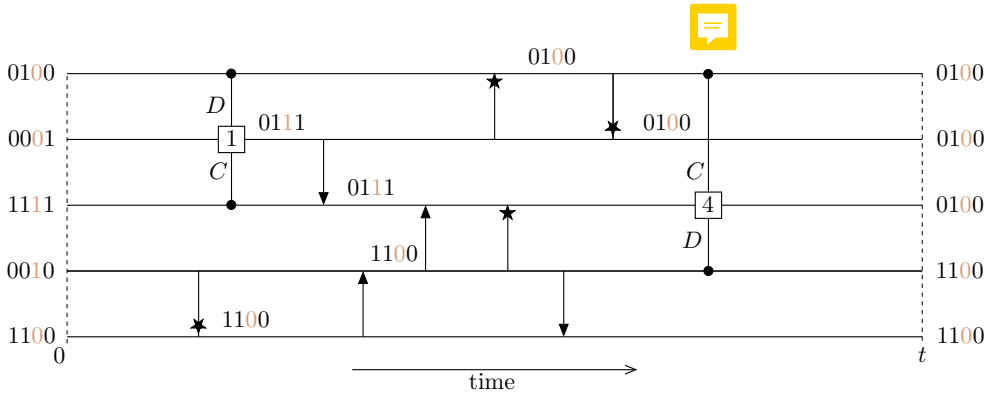


Figure 4.2. Graphical representation of the Moran IPS. Time runs from left to right. Arrows corresponding to neutral reproduction events are depicted with normal arrowheads, selective arrows with star-shaped arrowheads; recombination events are symbolised by squares containing the recombination point, and arms connecting to the parents that contribute the head (C) and tail (D) segments. The selected site is marked in light brown.

in a population of total size N, this process takes values in $\mathcal{P}(X) \cap \frac{1}{N}\mathbb{Z}^{2^n}$. Its transition rates can be described as follows (see also Remark 4.4). A transition from state $\zeta \in \mathcal{P}(X)$ to $\zeta + \frac{1}{N}(\delta_y - \delta_x)$ occurs if an individual of type x is replaced by an individual of type y. This happens either due to neutral reproduction (4.10) with rate $N\zeta(y)\zeta(x)$, due to selective reproduction (4.11) with rate $s(1-y_{i_*})N\zeta(y)\zeta(x)$ and due to recombination (4.12) at site i with rate $\varrho_i R_i(\zeta)(y)N\zeta(x)$. Thus, the total transition rate is given by

$$q^{(N)}(\zeta, \zeta + \frac{1}{N}(\delta_y - \delta_x)) = N\big[\zeta(y)\zeta(x) + s(1-y_{i_*})\zeta(y)\zeta(x) + \sum_{i \in S^*} \varrho R_i(\zeta)(y)\zeta(x)\big] = N\beta_{\delta_y - \delta_x}(\zeta)$$

where

$$\beta_{\delta_y - \delta_x}(\zeta) = \zeta(y)\zeta(x) + s(1-y_{i_*})\zeta(y)\zeta(x) + \sum_{i \in S^*} \varrho_i R_i(\zeta)(y)\zeta(x).$$

Since individuals are replaced one by one, all possible transitions are of this form and $q^{(N)}(\zeta, \zeta') = 0$ for all ζ' which are not of the form $\zeta + \delta_y - \delta_x$ for some $x, y \in X$. Due to the finiteness of X, the summability condition (4.14) is trivially fulfilled and we have

$$\sum_{x,y \in X} \beta_{\delta_y - \delta_x}(\zeta)(\delta_y - \delta_x) = (\zeta - \zeta) + s(F\zeta - f(\zeta)\zeta) + \sum_{i \in S^*} \varrho_i(R_i\zeta - \zeta) = \Psi_{\mathrm{sel}}(\zeta) + \Psi_{\mathrm{rec}}(\zeta).$$

That the right-hand side is Lipschitz continuous follows from the Lipschitz continuity of the recombinators (compare [BBS16, Prop. 1]) and thus, of Ψ_{rec}. The Lipschitz continuity of Ψ_{sel} is immediate. Thus, Theorem 4.2 yields

$$\lim_{N \to \infty} \sup_{0 \leqslant s \leqslant t} |Z_s^{(N)} - \omega_s| = 0,$$

under the assumption that $Z_0^{(N)} \to \omega_0$ as $N \to \infty$, where $\omega = (\omega_t)_{t \geqslant 0}$ solves Eq. (4.9).

4.3 Marginalisation consistency

Let us now turn to the dynamics of the *marginal type distributions* under selection and recombination. As the results of this section will not play a pivotal role for the core of the chapter, the impatient reader may skip this section at first reading. However, knowledge of marginalisation consistency will help to understand the graphical constructions in Sec. 4.5, and is also of independent interest. Furthermore, the current section will enable the reader to appreciate some of the difficulties and pitfalls inherent in the selective case.

For $A \subseteq S$, we define the *marginal recombinators* $R_i^A : \mathcal{P}(X_A) \to \mathcal{P}(X_A)$ by

$$R_i^A \nu := \nu^{A \cap C_i} \otimes \nu^{A \cap D_i} \tag{4.16}$$

for $i \in A \setminus i_*$, where C_i and D_i denote the head and tail for i as before, and we use the shorthand $A \setminus j$ for $A \setminus \{j\}$.

Remark 4.6. Note that $\pi_A.R_i\omega = R_i^A \omega^A$ for all $A \subseteq S$ and $\omega \in \mathcal{P}(X)$, and $R_i^A = \mathrm{id}$ if A is contained either in C_i or D_i, that is, if $\{C_i, D_i\}|_A = \{A\}$ (compare [BBS16, Lemma 1]). ♢

Consider now the marginal $\omega^A = (\omega_t^A)_{t \geqslant 0}$ of the solution ω of the recombination equation. In the *neutral case* ($s = 0$), it is well known (compare Theorem 2.4) that the marginal satisfies the *marginalised recombination equation*

$$\dot{\omega}_t^A = \pi_A.\Psi_{\mathrm{rec}}(\omega_t) = \sum_{i \in A \setminus i_*} \varrho_i^A \big(R_i^A \omega_t^A - \omega_t^A\big) =: \Psi_{\mathrm{rec}}^A(\omega_t^A) \tag{4.17}$$

with initial condition $\omega_0^A = \pi_A.\omega_0$ and *marginal recombination rates*

$$\varrho_i^A := \sum_{\substack{j \in S^* \\ \{C_j, D_j\}|_A = \{C_i, D_i\}|_A}} \varrho_j \quad \text{for all } i \in A \setminus i_*; \tag{4.18}$$

see Figure 4.3 for an illustration. In particular,

$$\dot{\omega}_t^{\{i\}} = 0 \quad \text{for } i \in S^* \tag{4.19}$$

since $R_i^{\{i\}} = \mathrm{id}$. Eq. (4.17) follows from Remark 4.6, the linearity of $\pi_A.$ and the definition of the marginal rates in Eq. (4.18).

Unfortunately, this property does not generalise to the selective case. The reason is that Ψ_{sel} also depends on the proportion $f(\omega_t) = \omega_t^{\{i_*\}}(0)$ of fit individuals and that we lose this information by projecting onto a factor with respect to a subset of S not containing i_*. When A *does* contain i_*, however, we clearly have

$$f(\nu) = f^A(\nu^A) \quad \text{for any } \nu \in \mathcal{P}(X), \tag{4.20}$$

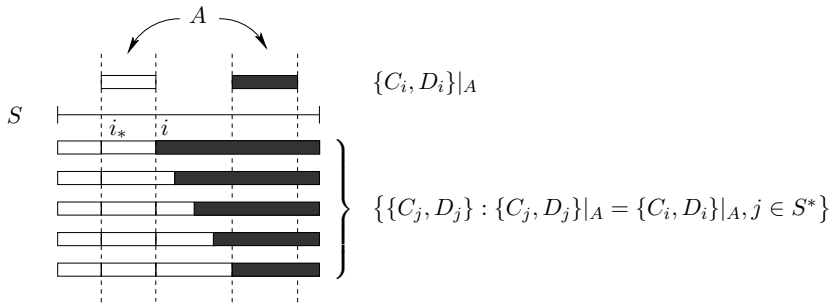


Figure 4.3. The partitions in Eq. (4.18) that define the marginal recombination rates.

where f^A is defined analogously to (4.2), but with S replaced by A. Moreover, the selection operator defined in (4.3) acts consistently on subsystems that contain the selected site, that is,

$$\pi_A.F\nu = F^A\nu^A \quad \text{for } A \ni i_*,$$

where the marginalised selection operator is given by

$$F^A(\nu^A)(x_A) = \begin{cases} \nu^A(x_A), & \text{if } x_{i_*} = 0, \\ 0, & \text{otherwise.} \end{cases}$$

In view of these considerations, we can define $\Psi^A_{\text{sel}} : \mathcal{P}(X_A) \to \mathcal{P}(X_A)$ via

$$\Psi^A_{\text{sel}}(\nu^A) := s\big(F^A - f^A(\nu^A)\big)\nu^A$$

such that $\pi_A.\Psi_{\text{sel}}(\nu) = \Psi^A_{\text{sel}}(\nu^A)$ for $A \ni i_*$ and all $\nu \in \mathcal{P}(X)$. Combining this with (4.17), we obtain the following result.

Theorem 4.3 (marginalisation consistency of the SRE)**.** *Let ω be a solution of the* SRE (4.9)*. Let $A \subseteq S$ contain i_*. Then, the marginal $\omega^A := (\omega^A_t)_{t\geqslant 0}$ solves the* marginal SRE*,*

$$\dot\omega^A_t = s\Big(F^A(\omega^A_t) - f^A(\omega^A_t)\omega^A_t\Big) + \sum_{i\in A\setminus i_*} \varrho^A_i\big(R^A_i\omega^A_t - \omega^A_t\big),$$

with the marginal recombination rates *ϱ^A_i given in* (4.18)*. In particular, ω^A is independent of all ϱ_i with i such that $\{C_i, D_i\}|_A = \{A\}$; or equivalently, with i such that $i \succ j$ for all $j \in A$ comparable to i.* □

Remark 4.7. The problem of marginalisation (in)consistency was already observed by Ewens and Thomson [ET77] in 1977 for the discrete-time SRE; see also the review in [Bür00, pp. 69–72]. For Theorem 4.3 to hold, the assumption that A contains the selected site is crucial: It is otherwise impossible to find a closed expression for the projection of the selective part in (4.9) in terms of the marginal measure, because we lose the information about the proportion of fit

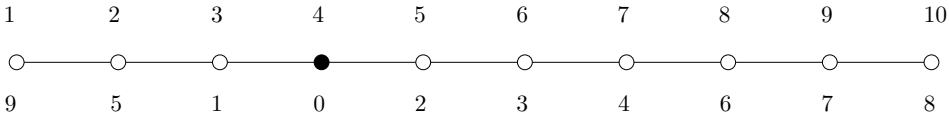


Figure 4.4. A nondecreasing permutation of sites. The original labels of the sites, $1 \leqslant i \leqslant n$, are at the top;below each site with label i, we have noted the corresponding k for which $i_k = i$.

individuals in the case that $i_* \notin A$. It is indeed a common pitfall to assume that Theorem 4.3 holds for arbitrary A. This is also implicit in [BB03]; see the corresponding erratum. ♢

4.4 Recursive solution of the selection-recombination equation

The first main result in this chapter will be a recursive solution of the SRE. The recursion will start at i_* and work along the site indices in agreement with the partial order introduced in Definition 4.1. If the original indices are used, the recursion must be formulated individually for every choice of i_*; in particular, it looks quite different depending on whether i_* is at one of the ends or in the interior of the sequence. To establish the recursion in a unified framework, we introduce a relabelling; let us fix a nondecreasing (in the sense of the partial order from Definition 4.1) permutation $(i_k)_{0\leqslant k\leqslant n-1}$ of S (compare Fig. 4.4) and denote the corresponding heads and tails by upper indices, that is, $C^{(k)} := C_{i_k}$ and $D^{(k)} := D_{i_k}$ (compare Figure 4.1). Note that $i_0 = i_*$, $D^{(0)} = S$ and $C^{(0)} = \varnothing$ and also that this choice of permutation implies that for all $\ell \geqslant k$, one has either $D^{(\ell)} \subseteq D^{(k)}$ (if $\ell \succcurlyeq k$) or $D^{(\ell)} \subseteq C^{(k)}$ (if ℓ and k are incomparable). Furthermore, we define $\varrho^{(k)} := \varrho_{i_k}$ and $R^{(k)} = R_{i_k}$ for $k > 0$.

We now proceed as follows. First, we recapitulate the solution of the pure selection equation, that is, we solve (4.9) in the special case that all recombination rates vanish. Then, in accordance with the labelling given by $(i_k)_{1\leqslant k\leqslant n-1}$, we will successively add sites at which we allow recombination. This can be formalised as follows.

Definition 4.4. For $\varrho^{(1)}, \dots, \varrho^{(n-1)}$ as above and every $k \in [0 : n-1]$, we set

$$\Psi_{\mathrm{rec}}^{(k)} := \sum_{\ell=1}^{k} \varrho^{(\ell)}\big(R^{(\ell)} - \mathrm{id}\,\big), \quad \Psi^{(k)} := \Psi_{\mathrm{sel}} + \Psi_{\mathrm{rec}}^{(k)}$$

(with the usual convention that the empty sum is 0). We then define the SRE *truncated at* k as the differential equation

$$\dot{\omega}_t^{(k)} = \Psi^{(k)}(\omega_t^{(k)}).$$

Furthermore, we understand $(\omega^{(k)})_{0\leqslant k\leqslant n-1}$ as the family of the corresponding solutions, all with the same initial condition ω_0. In particular, $\omega^{(0)}$ is the solution of the pure selection

equation

$$\dot{\omega}_t^{(0)} = \Psi_{\mathrm{sel}}(\omega_t^{(0)}) = s f(\omega_t^{(0)})\big(b(\omega_t^{(0)}) - \omega_t^{(0)}\big). \tag{4.21}$$

We also define $\psi^{(k)} = (\psi_t^{(k)})_{t\geqslant 0}$ as the flow semigroup associated to the differential equation defined via $\Psi^{(k)}$. In line with (4.9), we have $\omega = \omega^{(n-1)}$ (which is to say $\omega_t = \omega_t^{(n-1)}$ for all $t \geqslant 0$) and $\Psi = \Psi^{(n-1)}$, and we likewise set $\psi = \psi^{(n-1)}$. We will also write φ instead of $\psi^{(0)}$ for the (pure) selection semigroup. ♢

Proposition 4.5. *The solution of the pure selection equation* (4.21) *with initial condition* $\omega_0 \in \mathcal{P}(X)$ *is given by*

$$\omega_t^{(0)} = \varphi_t(\omega_0) = \frac{\mathrm{e}^{st} F(\omega_0) + (1-F)(\omega_0)}{\mathrm{e}^{st} f(\omega_0) + 1 - f(\omega_0)}, \quad t \geqslant 0, \tag{4.22}$$

with f *and* F *as given in* (4.2) *and* (4.3)*. In particular,*

$$f(\omega_t^{(0)}) = \frac{\mathrm{e}^{st} f(\omega_0)}{\mathrm{e}^{st} f(\omega_0) + 1 - f(\omega_0)} \tag{4.23}$$

is increasing over time and $\omega_t^{(0)} = \varphi_t(\omega_0)$ *is a convex combination of the initial type distributions of the fit (that is, beneficial) and unfit (that is, deleterious) subpopulations introduced in Eqs.* (4.5) *and* (4.6)*, namely,*

$$\omega_t^{(0)} = f(\omega_t^{(0)}) b(\omega_0) + \Big(1 - f(\omega_t^{(0)})\Big) d(\omega_0).$$

This in particular implies

$$b\Big(\varphi_t(\omega_0)\Big) = b(\omega_0) \text{ and } d\Big(\varphi_t(\omega_0)\Big) = d(\omega_0). \tag{4.24}$$

Proof. By straightforward verification. To see Eq. (4.24), recall that the fitness operator F is a projection and $b(\omega)$ is in the image of F, while $d(\omega)$ is in the image of $1 - F$ for any $\omega \in \mathcal{P}(X)$. □

Remark 4.8. Eq. (4.23) generalises the well-known solution of the selection equation for a single site, which is simply a logistic equation; compare [Dur08, p. 198]. Eq. (4.24) reflects the plausible fact that, while the proportion of fit individuals increases at the cost of the unfit ones (as quantified in Eq. (4.22)), the type composition *within* the set of fit types remains unchanged, and likewise for the set of unfit types. ♢

The main result in this section is the following recursion formula for the family of solutions of the (truncated) SREs.

Theorem 4.6. *The family of solutions* $(\omega^{(k)})_{1\leqslant k\leqslant n-1}$ *of Definition* 4.4 *satisfies the recursion*

$$\omega_t^{(k)} = \mathrm{e}^{-\varrho^{(k)} t} \omega_t^{(k-1)} + \pi_{C^{(k)}}.\omega_t^{(k-1)} \otimes \pi_{D^{(k)}}. \int_0^t \varrho^{(k)} \mathrm{e}^{-\varrho^{(k)}\tau} \omega_\tau^{(k-1)} \,\mathrm{d}\tau$$

for $1 \leqslant k \leqslant n-1$ *and* $t \geqslant 0$*, where* $\omega^{(0)}$ *is the solution of the pure selection equation given in Proposition* 4.5*.*

We will first give an analytic proof. Then, in the next section, we will give a genealogical proof of the recursion by means of the ancestral selection-recombination graph (ASRG), which will provide additional insight.

To deal with the nonlinearity of recombination and to exploit the underlying *linear* structure (see [BB16]) more efficiently, we now introduce a variant of the product of two measures that are defined on X_A and X_B, where A and B need not be disjoint. Namely, given a subset U of S, sets $I, J \subseteq U$, and *signed* measures ν_I, ν_J on X_I and X_J, respectively, we define

$$\nu_I \boxtimes \nu_J := (\pi_{I\setminus J}.\nu_I) \otimes \nu_J,$$

which is a signed measure on $X_{I\cup J}$ (recall that $\pi_{\varnothing}.\nu = \nu(X_I)$ for all signed measures ν on X_I, $I \subseteq S$ in line with Remark 2.2). Note that we use ν_I here to mean any signed measure on X_I, whereas we abbreviate by ν^I the specific signed measure on X_I that is obtained from ν on X via $\nu^I = \pi_I.\nu$.

Proposition 4.7. *Let* $U \subseteq S$*. For* $I, J, K \subseteq U$ *and signed measures* ν_I, ν_J, ν_K *on* X_I*,* X_J*, and* X_K*, respectively, the operation* $\boxtimes$ *has the following properties.*

(i) $(\nu_I \boxtimes \nu_J) \boxtimes \nu_K = \nu_I \boxtimes (\nu_J \boxtimes \nu_K)$ (*associativity*).

(ii) *If* $I \cap J = \varnothing$*, we have* $\nu_I \boxtimes \nu_J = \nu_I \otimes \nu_J = \nu_J \boxtimes \nu_I$ (*reduction to tensor product and commutativity*).

(iii) *If* $I \subseteq J$*, then* $\nu_I \boxtimes \nu_J = \nu_I(X_I)\nu_J$ (*cancellation property*).

Proof. For associativity, note that

$$\begin{aligned}(\nu_I \boxtimes \nu_J) \boxtimes \nu_K &= \big((\pi_{I\setminus J}.\nu_I) \otimes \nu_J\big) \boxtimes \nu_K = \big(\pi_{(I\cup J)\setminus K}.(\pi_{I\setminus J}.(\nu_I) \otimes \nu_J)\big) \otimes \nu_K \\ &= \pi_{I\setminus(J\cup K)}.\nu_I \otimes \pi_{J\setminus K}.\nu_J \otimes \nu_K = \pi_{I\setminus(J\cup K)} \otimes (\nu_J \boxtimes \nu_K) = \nu_I \boxtimes (\nu_J \boxtimes \nu_K),\end{aligned}$$

where we have used in the third step that $((I \cup J) \setminus K) \cap (I \setminus J) = I \setminus (J \cup K)$.

When $I \cap J = \varnothing$, one has

$$\nu_I \boxtimes \nu_J = \pi_{I\setminus J}.\nu_I \otimes \nu_J = \pi_I.\nu_I \otimes \nu_J = \nu_I \otimes \nu_J = \nu_J \otimes \nu_I,$$

which implies the claimed reduction to $\otimes$ and thus commutativity. Finally, for $I \subseteq J$,

$$\nu_I \boxtimes \nu_J = (\pi_{I\setminus J}.\nu_I) \otimes \nu_J = (\pi_{\varnothing}.\nu_I) \otimes \nu_J = \nu_I(X_I)\nu_J$$

establishes the cancellation property. □

Under the conditions of Proposition 4.7, we now denote by $\nu_J \boxplus \nu_K$ the formal sum of ν_J and ν_K (and use $\boxminus$ for the corresponding formal difference). Note that the formal sum turns into a proper sum (and hence $\boxplus$ reduces to $+$) when $I = J$. Furthermore, we define

$$\nu_I \boxtimes (\nu_J \boxplus \nu_K) := (\nu_I \boxtimes \nu_J) \boxplus (\nu_I \boxtimes \nu_K). \tag{4.25}$$

Clearly, the right-hand side reduces to a proper sum when $I \cup J = I \cup K$.

Generalising the formal sum above, we define $\mathfrak{A}(X_U)$ to be the real vector space of formal sums

$$\nu := \lambda_1 \nu_{U_1} \boxplus \ldots \boxplus \lambda_q \nu_{U_q},$$

where $q \in \mathbb{N}$, $\lambda_1, \ldots, \lambda_q \in \mathbb{R}$, $U_1, \ldots, U_q \subseteq U \subseteq S$, and $\nu_{U_1}, \ldots, \nu_{U_q}$ are signed measures on $X_{U_1}, \ldots, X_{U_q}$, respectively. We also write $\nu(X_U) := \sum_{i=1}^{q} \lambda_i \nu_{U_i}(X_{U_i})$.

Remark 4.9. If one extends the definition of $\boxtimes$ canonically to all of $\mathfrak{A}(X_U)$ (recalling that the projections are linear), $(\mathfrak{A}(X_U), \boxtimes)$ becomes an associative, unital algebra with neutral element $\mathbf{1}$, the measure with weight 1 on $X_\varnothing$. Note that, when multiplying $\nu \in \mathfrak{A}(X_I)$ and $\mu \in \mathfrak{A}(X_J)$ for disjoint I and J, the multiplication introduced above agrees with the measure product. ♢

Now, we can rewrite $\Psi_{\mathrm{rec}}^{(k)}$ of Definition 4.4 as

$$\boxed{\Psi_{\mathrm{rec}}\big(\omega_t^{(k)}\big) = \omega_t^{(k)} \boxtimes \Big(\bigboxplus_{\ell=1}^{k} \varrho^{(\ell)}\big(\pi_{D_\ell}.\omega_t^{(k)} - \mathbf{1}\big)\Big);} \tag{4.26}$$

note that the right-hand side indeed reduces to a proper (rather than a formal) sum of measures via (4.25), because $\omega_t^{(k)}$ lives on X_S and $D_\ell \subseteq S$ for $1 \leqslant \ell \leqslant k$, so that each term is a measure on X_S.

We shall see later that, when combined with selection, this representation has an advantage over the use of recombinators because it nicely brings out the recursive structure; this will streamline calculations and connect to the graphical construction in a natural way. The fact that the head alone determines the fitness of an individual manifests itself in the right-multiplicativity of Ψ_{sel} and its associated flow φ (compare Definition 4.4) as follows.

Lemma 4.8. *For all $\mu \in \mathcal{P}(X)$ and all $\nu \in \mathfrak{A}(X_{S^*})$,*

$$F(\mu \boxtimes \nu) = F(\mu) \boxtimes \nu.$$

If, in addition, $\nu(X_{S^*}) = 1$, one has

$$\Psi_{\mathrm{sel}}(\mu \boxtimes \nu) = \Psi_{\mathrm{sel}}(\mu) \boxtimes \nu$$

and therefore

$$\varphi_t(\mu \boxtimes \nu) = \varphi_t(\mu) \boxtimes \nu$$

for every $t \geqslant 0$.

Proof. To keep the notation simple, we assume $U_1, U_2 \subseteq S^*$ and $\nu = \nu_{U_1} \boxplus \nu_{U_2}$ with signed measures ν_{U_1} and ν_{U_2} on X_{U_1} and X_{U_2}, respectively. By the tensor product representation of F from (4.4), we have

$$\begin{aligned} F(\mu \boxtimes \nu_{U_1} + \mu \boxtimes \nu_{U_2}) &= F(\mu \boxtimes \nu_{U_1}) + F(\mu \boxtimes \nu_{U_2}) = F(\pi_{\overline{U_1}}.\mu \otimes \nu_{U_1}) + F(\pi_{\overline{U_2}}.\mu \otimes \nu_{U_2}) \\ &= (P_{i_*} \otimes \mathrm{id}_{\overline{U_1}\setminus i_*})(\pi_{\overline{U_1}}.\mu) \otimes \mathrm{id}_{U_1}(\nu_{U_1}) + (P_{i_*} \otimes \mathrm{id}_{\overline{U_2}\setminus i_*})(\pi_{\overline{U_2}}.\mu) \otimes \mathrm{id}_{U_2}(\nu_{U_2}) \\ &= \pi_{\overline{U_1}}.(P_{i_*} \otimes \mathrm{id}_{S^*})(\mu) \otimes \mathrm{id}_{U_1}(\nu_{U_1}) + \pi_{\overline{U_2}}.\,(P_{i_*} \otimes \mathrm{id}_{S^*})\,(\mu) \otimes \mathrm{id}_{U_2}(\nu_{U_2}) \\ &= F(\mu) \boxtimes \nu_{U_1} + F(\mu) \boxtimes \nu_{U_2}, \end{aligned}$$

which gives the first claim. Taking the first claim together with the fact that $f(\mu \boxtimes \nu) = f(\mu)$ if $\nu(X_{S^*}) = 1$, we get the second and the third claim. □

Now, the proof of Theorem 4.6 is straightforward.

Proof of Theorem 4.6. Let $\Psi^{(k)}$ be as in Definition 4.4. With the shorthand

$$\nu_t^{(k-1)} := \pi_{D^{(k)}}.\int_0^t \varrho^{(k)} \mathrm{e}^{-\varrho^{(k)}\tau} \omega_\tau^{(k-1)}\,\mathrm{d}\tau,$$

one has $\nu_t^{(k-1)}(X_{D^{(k)}}) = 1 - \mathrm{e}^{-\varrho^{(k)}t}$, and the right-hand side of the recursion formula from Theorem 4.6 can be expressed as

$$\mu_t^{(k)} := \omega_t^{(k-1)} \boxtimes (\mathrm{e}^{-\varrho^{(k)}t}\mathbf{1} \boxplus \nu_t^{(k-1)}). \tag{4.27}$$

First, we show that

$$\mu_t^{(k)} \boxtimes \pi_{D^{(\ell)}}.\mu_t^{(k)} = \big(\omega_t^{(k-1)} \boxtimes \pi_{D^{(\ell)}}.\omega_t^{(k-1)}\big) \boxtimes (\mathrm{e}^{-\varrho^{(k)}t}\mathbf{1} \boxplus \nu_t^{(k-1)}) \tag{4.28}$$

for all $1 \leqslant \ell \leqslant k$. To see this, write the left-hand side as $\omega_t^{(k-1)} \boxtimes A \boxtimes B$, where

$$A := \mathrm{e}^{-\varrho^{(k)}t}\mathbf{1} \boxplus \nu_t^{(k-1)}$$

and

$$B := \pi_{D^{(\ell)}}.\big(\omega_t^{(k-1)} \boxtimes (\mathrm{e}^{-\varrho^{(k)}t}\mathbf{1} \boxplus \nu_t^{(k-1)})\big) = \pi_{D^{(\ell)}}.\mu_t^{(k)}.$$

Recall that, by our monotonicity assumption on the permutation of sites, we have either $D^{(k)} \subseteq D^{(\ell)}$ or $D^{(k)} \cap D^{(\ell)} = \varnothing$. In the first case, (4.28) follows by cancelling A using Proposition 4.7 (note that $A(X_{D^{(k)}}) = 1$). In the second case, B is just $\pi_{D^{(\ell)}}.\omega_t^{(k-1)}$, and so $A \boxtimes B = B \boxtimes A$, again by Proposition 4.7. Now we compute, using (4.26) and (4.27) in the first step, (4.28) and Lemma 4.8 in the second, Definition 4.4 in the third, and Proposition 4.7

in the last:

$$\begin{aligned}\Psi^{(k)}(\mu_t^{(k)}) &= \Psi_{\mathrm{sel}}(\omega_t^{(k-1)} \boxtimes (\mathrm{e}^{-\varrho^{(k)}t}\mathbf{1} \boxplus \nu_t^{(k-1)})) + \sum_{\ell=1}^{k} \varrho^{(\ell)} \mu_t^{(k)} \boxtimes (\pi_{D^{(\ell)}}.\mu_t^{(k)} \boxminus \mathbf{1}) \\ &= \bigl(\Psi_{\mathrm{sel}}(\omega_t^{(k-1)}) + \sum_{\ell=1}^{k} \varrho^{(\ell)} \omega_t^{(k-1)} \boxtimes (\pi_{D^{(\ell)}}.\omega_t^{(k-1)} \boxminus \mathbf{1})\bigr) \boxtimes (\mathrm{e}^{-\varrho^{(k)}t}\mathbf{1} \boxplus \nu_t^{(k-1)}) \\ &= \bigl(\Psi^{(k-1)}(\omega_t^{(k-1)}) + \varrho^{(k)} \omega_t^{(k-1)} \boxtimes (\pi_{D^{(k)}}.\omega_t^{(k-1)} \boxminus \mathbf{1})\bigr) \boxtimes (\mathrm{e}^{-\varrho^{(k)}t}\mathbf{1} \boxplus \nu_t^{(k-1)}) \\ &= \dot{\omega}_t^{(k-1)} \boxtimes (\mathrm{e}^{-\varrho^{(k)}t}\mathbf{1} \boxplus \nu_t^{(k-1)}) + \omega_t^{(k-1)} \boxtimes (\varrho^{(k)}\mathrm{e}^{-\varrho^{(k)}t} \pi_{D^{(k)}}.\omega_t^{(k-1)} \boxminus \varrho^{(k)}\mathrm{e}^{-\varrho^{(k)}t}\mathbf{1}).\end{aligned}$$

Identifying $\varrho^{(k)}\mathrm{e}^{-\varrho^{(k)}t}\pi_{D^{(k)}}.\omega_t^{(k-1)}$ with $\dot{\nu}_t^{(k-1)}$, we see that the last line is just the time derivative of $\mu_t^{(k)}$ of (4.27). □

Remark 4.10. We could have proved Theorem 4.6 also without the help of formal sums and the new operations $\boxplus, \boxminus, \boxtimes$. However, we decided on the current presentation in order to familiarise the reader with this — admittedly somewhat abstract — formalism, as it is the key to stating the duality result in Section 4.7 in closed form. It will also allow us later to state the solution itself in closed form; see Corollary 4.26. ♢

Remark 4.11. Note that the only property of the selection operator that entered the proof of Theorem 4.6 is the second property in Lemma 4.8, namely, $\Psi_{\mathrm{sel}}(\omega \boxtimes \nu) = \Psi_{\mathrm{sel}}(\omega) \boxtimes \nu$ for all $\nu \in \mathfrak{A}(X_{S^*})$ with $\nu(X_{S^*}) = 1$. Therefore, the result remains true if Ψ_{sel} is replaced by a more general operator with this property. In particular, Theorem 4.6 remains true when frequency-dependent selection and/or mutation at the selected site is included. ♢

Remark 4.12. Applying Theorem 4.3 to $A = \{i_*\}$ shows that the marginal type frequency at the selected site is unaffected by recombination. More generally, consider the set

$$L^{(k)} := \{i_0 = i_*, i_1, \ldots, i_k\}$$

and note that $L^{(k)} \setminus i_*$ is exactly the set of recombination sites that are considered up to and including the k-th iteration. Obviously, marginalisation consistency holds for $L^{(k)}$ for all $0 \leqslant k \leqslant n-1$. Since $\varrho_i^{L^{(k)}} = \varrho_i$ for $i \in L^{(k)} \setminus i_*$, Remark 4.6 and Eq. (4.18) together with Definition 4.4 give

$$\pi_{L^{(k)}}.\dot{\omega}_t = \pi_{L^{(k)}}. \sum_{i \in L^{(k)} \setminus i_*} \varrho_i (R_i \omega_t - \omega_t) = \pi_{L^{(k)}}.\Psi_{\mathrm{rec}}^{(k)}(\omega_t) = \pi_{L^{(k)}}.\dot{\omega}_t^{(k)},$$

and so $\pi_{L^{(k)}}.\omega_t^{(k)} = \pi_{L^{(k)}}.\omega_t$. This implies that if one is only interested in the marginal with respect to $L^{(k)}$, then one may stop the iteration after the k-th step. ♢

An important application of Theorem 4.6 is the following recursion for the first-order correlation functions $\omega_t^{(k)} - R^{(k)}\omega_t^{(k)}$ between the type frequencies at the sites contained in $C^{(k)}$ and

those contained in $D^{(k)}$, for solutions of the truncated equations. These objects, which are referred to as *linkage disequilibria* in the biological literature, are also of independent interest; compare [Dur08, Ch. 3.3].

Lemma 4.9 (correlation functions)**.** *The family of solutions* $(\boldsymbol{\omega}^{(k)})_{0\leqslant k\leqslant n-1}$ *of Definition* 4.4 *satisfies, for* $1 \leqslant k \leqslant n-1$*,*

$$(\mathrm{id} - R^{(k)})\omega_t^{(k)} = \mathrm{e}^{-\varrho^{(k)}t}(\mathrm{id} - R^{(k)})\omega_t^{(k-1)}.$$

Proof. By direct verification via Theorem 4.6, using $R^{(k)}\omega_t^{(k)} = \omega_t^{(k)} \boxtimes \pi_{D^{(k)}}.\omega_t^{(k)}$. □

4.5 Looking back in time: the ancestral selection-recombination graph

Our next goal is to reveal the genealogical content of the recursive solution formula of Theorem 4.6. We will accomplish this by a change of perspective: Instead of focusing on the evolution of the type distribution (in the entire population) forward in time as described by the SRE (4.9), we will analyse instead the type distribution at time t by tracing back the genealogy of a given individual.

The crucial tool for this purpose is the *ancestral selection-recombination graph* (ASRG) of [DK99; LK12; BP18]. As the name suggests, it is a combination of the *ancestral selection graph* (ASG) of [KN97] and the *ancestral recombination graph* (ARG) of [Hud83; GM96; GM97]. We will introduce the ASRG here as taylored to meet the selection-recombination differential equation. The purpose of the graph is to trace back all lines that may carry information about the type (and the ancestry) of an individual at present, so that a Markov structure is obtained. This is similar to [Cor17a; BCH18] for the selection part and to [BBS16; BB16] for the recombination part, where the ancestral graphs consist of all *potentially ancestral lines* of an individual at present. At this point, we will understand the notion of *potentially ancestral* in a broad sense, including lines that are potentially ancestral to some line in the graph, but not necessarily to the individual at present. It will indeed turn out that some of these lines are not potentially ancestral to the present individual itself (that is, in this stricter sense, the notion of potential ancestry is not transitive); such lines will be pruned away later on. We will first consider the case of a finite population of size N, before taking the limit $N \to \infty$. Recalling the definition of the Moran IPS in Section 4.2, we can sample from the type distribution at present time t via the following procedure (compare Fig. 4.5).

(1) Select an arbitrary label α from $\{1, \dots, N\}$ for the individual to be considered.

(2) Construct the untyped version of the Moran IPS.

(3) Start the graph by tracing back the single line emerging from the individual at present time t. Proceed as follows in an iterative way in the backward direction of time until

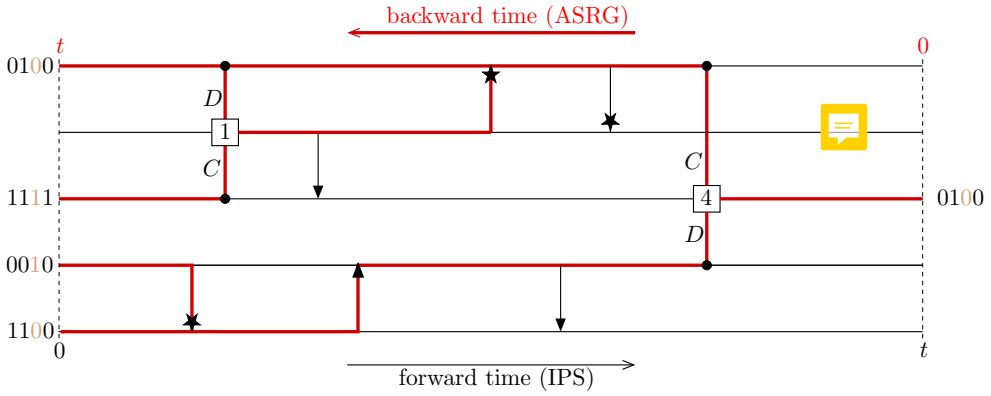


Figure 4.5. Sampling from the type distribution at present using the graphical representation of the Moran IPS. The ASRG is marked in red and the selected site in light brown. Notice the two different time axes for the IPS and the ASRG, respectively; while the types are propagated through the IPS from left to right, the genealogy is constructed in the opposite direction, starting with a present-day individual on the right.

the initial time is reached; note that forward time 0 (forward time t) corresponds to backward time t (backward time 0).

(3a) If a line currently in the graph is hit by the tip of a neutral arrow, it is relocated to the line at the tail.

(3b) If a line in the graph is hit by a selective arrow, we trace back both its potential ancestors, namely the incoming branch (at the tail of the arrow) and the continuing branch (at the tip). That is, we add the incoming line to the graph, which results in a *branching event*.

(3c) If a line is hit by a recombination square at site i, we have a *splitting event* and trace back the lines that contribute the head (C_i) and the tail (D_i), respectively, while the line hit by the square is discontinued.

(4) Assign types to all lines in the graph at time 0 by sampling without replacement from the initial counting measure NZ_0 (compare Section 4.2). Then, propagate the types forward along the lines obtained in step (3), according to the same rules as in the Moran IPS. That is, selective branchings are resolved by applying the *pecking order* derived from the Moran IPS and illustrated in Fig. 4.6, namely: the incoming branch is parental to the descendant line if it has a 0 at the selected site; otherwise, the continuing branch is parental. Splitting events are resolved by piecing together heads and tails. This way, a type is associated with every line element of the graph.

The graph resulting from steps (1)–(3), along with the graphical elements indicating reproduction and recombination, is called the *untyped* ASRG, whereas the outcome of step (4) is the *typed* ASRG. While steps (3a) and (3c) are obvious, let us comment on the crucial

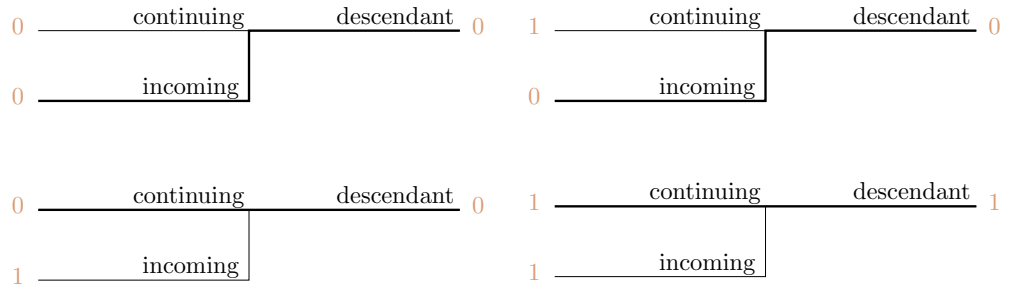


Figure 4.6. The pecking order between incoming line and continuing line, and the resulting type of the descendant. In each case, the ancestral line is bold. To keep the picture simple, we have only indicated the letter at the selected site. Likewise, the picture applies to the case $n = i_* = 1$.

branching step (3b). It builds on the special role of the selective arrows in the Moran IPS and reflects the fact that whether the incoming or the continuing branch is the true parent depends on the type of the incoming branch, which is not known in the untyped situation; in this sense, every branching event encodes a case distinction. Let us also mention that, in all events (3a)–(3c), it may happen that a line coalesces with a line that is already in the graph. Likewise, it is possible that, in a splitting event, the same parent contributes both the head and the tail; the event then turns into a relocation.

Steps (1)–(4) yield the type of the present individual considered, but also serve to elucidate the true ancestry of each site in this individual. In step (4), the paths along which the individuals contributing to the type of the present-day individual are propagated are called *(true) ancestral lines*, as opposed to the *potentially ancestral lines* in the untyped ASRG. More precisely, for $i \in S$, the path along which the type of the ancestor of site i is propagated is called the *ancestral line of site* i. It is obtained explicitly by adding step

(5) For each $i \in S$, trace back the ancestry of site i by starting from the individual at present, following back the true ancestral line (as determined in step (4)) in every branching event. This is the bold line in Fig. 4.6, and the one following either the C or D branch at every splitting event, depending on whether $i \in C$ or $i \in D$. That is, we remove from the ASRG those lines that do not contribute genetic material to site i in the present individual.

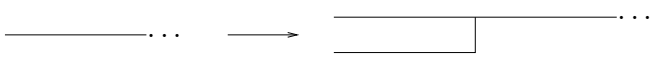
Clearly, in step ((2)), we need not construct the full graphical representation of the interacting particle system. Instead, it suffices to consider those events that occur on the lines in the ASRG of the sampled individual, that is, the lines (to be) traced back in step (3). We therefore obtain the same ASRG (in distribution) if steps (2) and (3) are replaced by the following single one.

(2'&3') Starting from the single line at forward time t, move backward in time and independently at rates $1, s$, and ϱ_i, let each line in the graph be hit by neutral arrows, selective arrows,

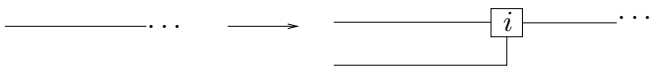
and recombination events at site i, $i \in S^*$, with the (potential) parent individual(s) chosen uniformly without replacement from the entire population in all cases; update the graph accordingly.

Note that we make use of the homogeneity of the Poisson process here, which entails that the graphical elements are laid down according to the same law in either direction of time. As we let N tend to infinity, another simplification results. Namely, the probability of choosing, for any kind of event, parent(s) already contained in the genealogy is of order $1/N$; the same is true for the probability to choose the same parent twice in a recombination event. In the limit $N \to \infty$, therefore, the probability that a coalescence happens when a neutral arrow is met will vanish. Likewise, selective reproduction (recombination) events will always result in branching (splitting) into two lines, with the incoming branch (both arms) outside the current set of lines. Furthermore, we disregard the position of the lines within the IPS; this is allowed because the types associated with each line form a *permutation-invariant* or *exchangeable* family of random variables. In particular, relocations may be safely ignored. The resulting random graph is called the ASRG in the law of large numbers regime. Since we will only be concerned with this limit in the remainder of the paper, we will often omit this specification.

Definition 4.10. For any given $t > 0$, the *ancestral selection-recombination graph* (ASRG) *in the law of large numbers* (LLN) *regime* is a random graph-valued function in backward time starting from a single node at time 0 and growing from right to left until time t, where branching events



occur at rate s on every line, and splitting events



occur at rate ϱ_i, $i \in S^*$, per line; all events are mutually independent. The right-most node is called the *root* of the ASRG and the leftmost nodes are called the *leaves*. ♢

Note that the graph, grown until any finite time t, is almost surely finite. Note also that we dispense with the star-shaped arrowheads used in the interacting particle system for the selective branchings; rather, we use the convention that the incoming branch be placed below the continuing branch. This is again allowed due to exchangeability. For the same reason, we dispense with the labelling of the recombination arms and instead adopt the convention that the sites in the head always come from the individual on the upper line, which we place on the same level as the descendant line. The sites in the tail are provided by the line attached

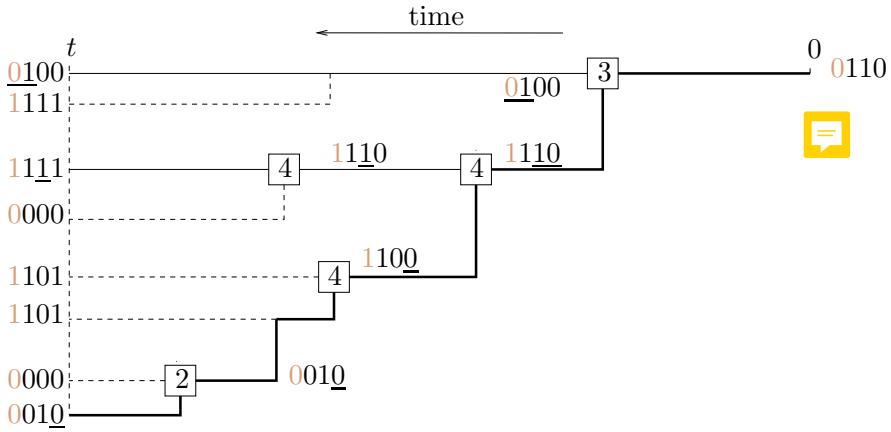


Figure 4.7. Tracing back the ancestry of an individual with 4 sites $i_* = i_0 = 1$, $i_1 = 2$, $i_2 = 3$ and $i_3 = 4$ under selection and recombination; the selected site $i_* = 1$ is light brown. The bold line is ancestral to site 4, the thin solid lines are ancestral to sites 1, 2, or 3, and the dashed lines are not ancestral to any site. Each branch is decorated with its type, and the sites to which it is ancestral are underlined.

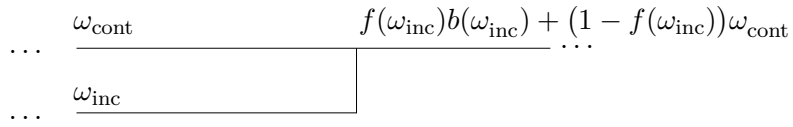
from below. For an example realisation of the ASRG and the construction of the type of an individual at present along with the ancestral line of one specific site, see Fig. 4.7.

For our purposes, the important point about the ASRG is that it implies the following procedure for sampling from ω_t. First, construct a realisation of the ASRG, run for time t. Then, assign types to its leaves, sampled independently from ω_0. These are then propagated through the graph in the same way as described above.

Remark 4.13. In order to connect the graphical constructions in this section to the viewpoint from the previous section, let us describe the type propagation in slightly more formal terms. Given a realisation of the ASRG of length t, we assign a type distribution to each node as follows. First, each leaf is assigned the initial type distribution ω_0. Each internal node v arises either from a branching or a splitting event. In the case of a branching, let ω_{inc} and ω_{cont} be the type distributions associated to the nodes that connect to v via the incoming and continuing branch. Then, we associate to v the distribution

$$\omega_v := f(\omega_{\text{inc}})b(\omega_{\text{inc}}) + (1 - f(\omega_{\text{inc}}))\omega_{\text{cont}},$$

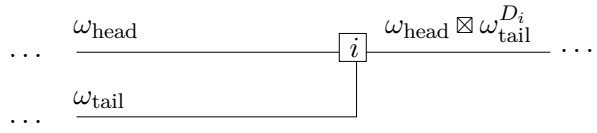
that is,



Likewise, if v is due to splitting (at site i, say), we associate with it the distribution

$$\omega_{\text{head}} \boxtimes \omega_{\text{tail}}^{D_i},$$

where ω_{head} and ω_{tail} are the distributions associated to the nodes that connect to v via the ancestral lines of the head and tail, respectively,



Finally, the distribution for the root individual is just the same as that of the unique internal node connected to it. ♢

Example 4.1. In the case of pure selection ($k = 0$), our ASRG reduces to an ordered version of the ASG in the deterministic limit; this is equivalent to a special case of the *pruned lookdown ASG* in the LLN regime, as introduced in [Cor17a; BCH18] in the context of a probabilistic representation of the solution of the deterministic selection-mutation equation. Since the contribution of coalescence events vanishes in this regime, the number of lines in the graph, that is, the number of potential ancestors of an individual sampled at time t, becomes a simple Yule process $K = (K_t)_{t\geqslant 0}$ with branching rate s. This is a continuous-time branching process where, at any time t, every individual branches into two at rate s independently of all others. In the case considered here, the process starts with $K_0 = 1$. Clearly, the pecking order implies that the individual at present will be drawn from the unfit subpopulation $d(\omega_0)$ if all K_t potential ancestors are of deleterious type; this happens with probability $(1 - f(\omega_0))^{K_t}$. Likewise, the individual will be sampled from the fit subpopulation $b(\omega_0)$ if at least one potential ancestor is of beneficial type (with probability $1 - (1 - f(\omega_0))^{K_t}$). Thus, we obtain the type distribution by averaging over all realisations of the Yule process at time t:

$$\begin{aligned}\omega_t^{(0)} &= \varphi_t(\omega_0)\\ &= \mathbb{E}\big[(1 - f(\omega_0))^{K_t} \mid K_0 = 1\big] d(\omega_0) + \big(1 - \mathbb{E}\big[(1 - f(\omega_0))^{K_t} \mid K_0 = 1\big]\big) b(\omega_0).\end{aligned} \tag{4.29}$$

This is a *stochastic representation* of the solution of the selection equation.

It is well known that K_t, given $K_0 = 1$, follows $\mathrm{Geom}(\mathrm{e}^{-st})$ (compare [Fel68, Ch. II.4] or [SO94, Ex. 2.19]), where $\mathrm{Geom}(\sigma)$ denotes the distribution of the number of independent Bernoulli trials with success probability σ up to and including the first success. The probability generating function is given by

$$g(z) = \mathbb{E}\big[z^{K_t} \mid K_0 = 1\big] = \frac{\mathrm{e}^{-st} z}{1 - (1 - \mathrm{e}^{-st}) z}. \tag{4.30}$$

Consequently,

$$\mathbb{E}\big[(1 - f(\omega_0))^{K_t} \mid K_0 = 1\big] = \frac{\mathrm{e}^{-st}(1 - f(\omega_0))}{\mathrm{e}^{-st}(1 - f(\omega_0)) + f(\omega_0)} = 1 - f(\omega_t^{(0)}) \tag{4.31}$$

with $f(\omega_t^{(0)})$ of Proposition 4.5. Inserting this into (4.29), we obtain $\omega_t^{(0)}$ of Proposition 4.5.

Anticipating the results in Section 6, this can be viewed as a special case of the general duality relation with respect to the duality function

$$h(m,\nu) = \big(1-f(\nu)\big)^m d(\nu) + \big(1-\big(1-f(\nu)\big)^m\big) b(\nu) \tag{4.32}$$

(compare Definition 5.9 and Proposition 4.18), which is the distribution of an individual's type at present, given it has m potential ancestors, which are sampled from the type distribution $\nu \in \mathcal{P}(X)$. Note that the right-hand side of Eq. (4.32) is a convex combination of the probability measures $d(\nu)$ and $b(\nu)$; this is ultimately due to the fact that while the proportions of fit and unfit types in the population change over time due to selection, the type compositions within the fit and unfit subpopulations remain constant. ♢

Example 4.2. Likewise, in the case of pure recombination, the ASRG reduces to the stochastic partitioning process $\Sigma = (\Sigma_t)_{t\geqslant 0}$ explained at the end of Chapter 2. ♢

We can now gear up for the genealogical proof of the recursion formula in Theorem 4.6. (Recall that the start of the recursion, the solution $\omega^{(0)}$ of the pure selection equation, was already considered in Example 4.1). To this end, we reuse the nondecreasing permutation $(i_k)_{0\leqslant k\leqslant n-1}$ of sites defined in Section 4.4 and, in perfect analogy with the family $(\omega^{(k)})_{0\leqslant k\leqslant n-1}$, define for $0 \leqslant k \leqslant n-1$ the ASRG *truncated at* k to be an ASRG with $\varrho^{(\ell)} = 0$ for all $\ell > k$. We denote the ASRG truncated at k by $\mathrm{ASRG}^{(k)}$, or by $\mathrm{ASRG}_t^{(k)}$ if we also want to indicate its duration. Clearly, the $\mathrm{ASRG}^{(k)}$ is the ASRG that corresponds to $\omega^{(k)}$. In particular, $\mathrm{ASRG}^{(0)}$ is just the ASG (without recombination), and the type at the root of an $\mathrm{ASRG}_t^{(k)}$ follows $\omega_t^{(k)}$. The key ingredient to the genealogical proof of the recursion is the following proposition, which links the type of the root of an $\mathrm{ASRG}^{(k)}$ to the type at the root of an $\mathrm{ASRG}^{(k-1)}$, or two independent copies thereof.

Proposition 4.11. *For $1 \leqslant k \leqslant n-1$ and any given $t > 0$, let B be a Bernoulli variable with success probability $1-\mathrm{e}^{-\varrho^{(k)}t}$. Conditional on $\{B = 1\}$, let T be an $\mathrm{Exp}(\varrho^{(k)})$ random variable conditioned on being $\leqslant t$, where $\mathrm{Exp}(\sigma)$ denotes the exponential distribution with parameter σ. Furthermore, denote by $\mathfrak{X} \in X$ the type at the root of an $\mathrm{ASRG}_t^{(k-1)}$, and by $\widetilde{\mathfrak{X}}$ the type at the root of an $\mathrm{ASRG}_T^{(k-1)}$, independent of the $\mathrm{ASRG}_t^{(k-1)}$ that delivers $\mathfrak{X}$. The type $\mathfrak{Z}$ at the root of an $\mathrm{ASRG}_t^{(k)}$ is then, in distribution, given by*

$$\boxed{\mathfrak{Z} = (1-B)\mathfrak{X} + B\big(\pi_{C^{(k)}}(\mathfrak{X}), \pi_{D^{(k)}}(\widetilde{\mathfrak{X}})\big).}$$

Before we prove this proposition, let us give some intuition for it. We work with the untyped $\mathrm{ASRG}_t^{(k)}$, obtained via steps (1) and (2'&3'), and consider the line ancestral to $D^{(k)}$. It is clear that this is a single line because, due to the partial order, none of the splitting events in the $\mathrm{ASRG}^{(k)}$ partitions $D^{(k)}$. Note that, at this point, the location of the true ancestral line is not yet known, since this is only decided in step (4), when propagating the types forward after sampling the initial types, as in Figure 4.7.

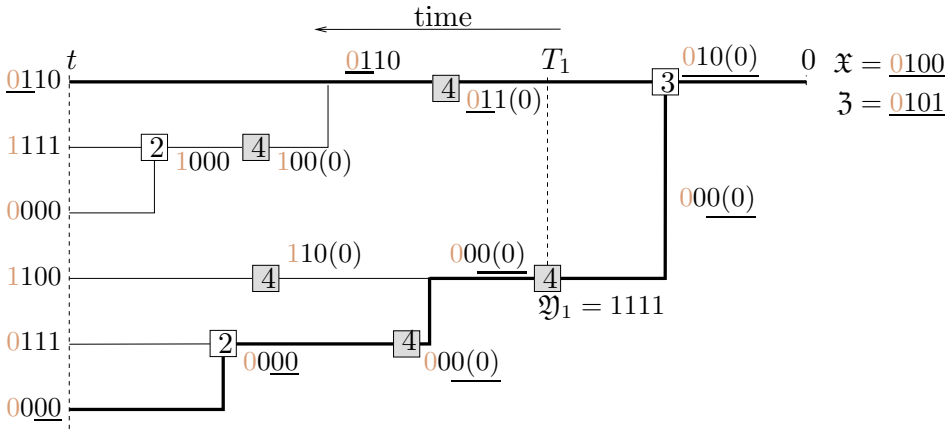


Figure 4.8. Determining the type at the root of a cASRG$^{(4)}$. The graph is a cASRG$^{(4)}$, the selected site is light brown, ancestral lines in the ASRG$^{(3)}$ are printed in bold, and ancestral letters are underlined. The shaded recombination squares indicate splitting events at site 4, where a new copy of an ASRG$^{(4)}$ is attached for the remaining time. Parentheses mark the 4th site in the ASRG$^{(3)}$ that is replaced by the tail of the new copy. Thus, $\mathfrak{X}$ is obtained by ignoring the shaded squares as well as the parentheses, and $\mathfrak{Z}$ is then obtained by replacing the 0 in brackets in the type of the lower branch of the rightmost recombination event by the 1 from $\mathfrak{Y}_1$.

We now distinguish two cases. With probability $\mathrm{e}^{-\varrho^{(k)}t}$, no splitting at site i_k has happened along this line, so the tail is 'glued' to the head. Thus, $\mathfrak{Z}$ may be constructed as in the absence of recombination events at site i_k, that is, via an $\mathrm{ASRG}_t^{(k-1)}$; this gives the first term on the right-hand side. With probability $1-\mathrm{e}^{-\varrho^{(k)}t}$, a splitting at site i_k has happened along the ancestral line of $D^{(k)}$. We then consider the time of the last, that is, of the *leftmost* splitting event at site i_k on the line in question and identify this time with $t-T$ (since such splitting events occur at rate $\varrho^{(k)}$ and due to the homogeneity of the Poisson process, T is indeed distributed as stated). The ancestry of the sites in $C^{(k)}$ is then unaffected by the split and thus follows an $\mathrm{ASRG}_t^{(k-1)}$; this is in line with the marginalisation consistency of Theorem 4.3. But the sites contained in $D^{(k)}$ now come from a different individual. Since $t-T$ is the time of the *leftmost* splitting event, we know that no further splits at site i_k have occured at any point further back in the past. This means that, at this point, the tail of the individual at the root of an independent $\mathrm{ASRG}_T^{(k-1)}$ enters the ancestral line. The combination of head and tail as described gives the second term on the right-hand side.

In order to turn these heuristics into a proof, we have to make the construction of the ancestral line of $D^{(k)}$ explicit. To this end, we mimick the recursion forward in time by coupling the $\mathrm{ASRG}_t^{(k)}$ in a suitable manner to an $\mathrm{ASRG}_t^{(k-1)}$. To keep things as transparent as possible, and to reduce the number of lines to be visualised, we introduce the following simplified construction; see Fig. 4.8.

Definition 4.12 (collapsed ASRG)**.** Let $1 \leqslant k \leqslant n-1$ be given. A *collapsed* ASRG *truncated at* k, or cASRG$^{(k)}$ for short, is an ASRG$^{(k-1)}$ decorated with i_k-recombination squares that

are laid down according to independent Poisson processes at rate $\varrho^{(k)}$ on every horizontal line segment. ♢

We can then construct a realisation of the $\mathrm{ASRG}_t^{(k)}$ by attaching to every i_k-recombination square of a $\mathrm{cASRG}^{(k)}$ an independent copy of an $\mathrm{ASRG}^{(k)}$ for the remaining time; that is, for any i_k-recombination square at time $\tau \in [0,t]$, we attach an $\mathrm{ASRG}_{t-\tau}^{(k)}$. In this context, therefore, splitting events take the form of attachment events. In the subsequent sampling step, this attachment provides the k-tail while the k-head comes from the original $\mathrm{ASRG}_t^{(k-1)}$. Let us describe now how to utilise the collapsed ASRG to sample a root individual of an $\mathrm{ASRG}_t^{(k)}$, that is, to sample from the distribution $\omega_t^{(k)}$. First, one constructs a realisation of the $\mathrm{cASRG}_t^{(k)}$. Then, types are assigned to the leaves according to ω_0 in an i.i.d. fashion and propagated forward, where selective branchings and splitting (attachment) events are resolved just like in the ASRG. Assume an i_k-square is encountered on a given line at some (forward) time $\tau \in [0,t]$, and the type just before the i_k-square (that is, at time $\tau - 0$) is $\mathfrak{x}$. We then draw a new type $\mathfrak{y}$ from $\omega_\tau^{(k)}$, independently of $\mathfrak{x}$, for the individual contributing the tail. The type on the line then jumps from $\mathfrak{x}$ at time $\tau - 0$ to type

$$\mathfrak{z} = \bigl(\pi_{C^{(k)}}(\mathfrak{x}), \pi_{D^{(k)}}(\mathfrak{y})\bigr)$$

at time τ, see Fig. 4.9. Keeping in mind the original motivation behind Definition 4.12 and thinking of the i_k-squares as splitting events (at site i_k) at which a new realisation of an $\mathrm{ASRG}^{(k)}$ is attached, it is clear that this gives the correct result.

Proof of Proposition 4.11. Let $1 \leqslant k \leqslant n-1$ and $t > 0$ be fixed and let a realisation of the $\mathrm{cASRG}_t^{(k)}$ be given, together with an assignment of types to its leaves. Elements of the proof are illustrated in Fig. 4.8. Note first that

- $\mathfrak{X}$ is, in distribution, equal to the (random) type at the root when ignoring the i_k-squares.

We consider the line ancestral to $D^{(k)}$ in the underlying $\mathrm{ASRG}_t^{(k-1)}$. The location of this line is now well defined, since we sample the types and can perform steps (4) and (5). Note that the line ancestral to $D^{(k)}$ is, at the same time, the line ancestral to $\max(C^{(k)})$, where the maximum is with respect to $\preccurlyeq$ of Definition 4.1; this is because no splits happen at i_k in the $\mathrm{ASRG}_t^{(k-1)}$. We consider the following quantities.

- Let B_1 be the Bernoulli variable that takes the value 0 (the value 1) if there is no (at

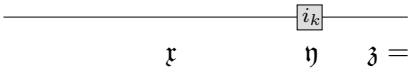


Figure 4.9. Upon encountering an i_k-square, the head of type $\mathfrak{x}$ is combined with the tail of a newly sampled individual (from $\omega_\tau^{(k)}$) to form the type of the descendant.

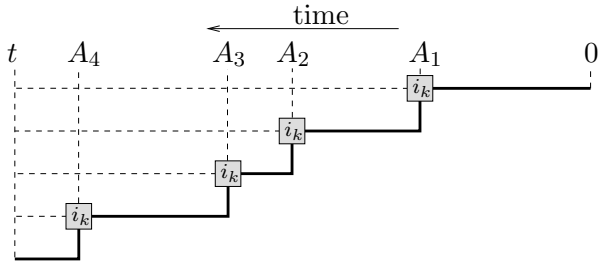


Figure 4.10. The ancestral line of $D^{(k)}$ after expanding all the recombination events arriving at the elements of the set $W \cap [0,t]$ used in the proof of Proposition 4.11. The ancestral lines of the corresponding heads are dashed as they need not be considered any further here. Note that the maximal element A_4 is the leftmost one.

least one) recombination square on the ancestral line of $D^{(k)}$. Clearly, B_1 has success probability $1 - \mathrm{e}^{-\varrho^{(k)}t}$.

- Conditional on $\{B_1 = 1\}$, let T_1 be the waiting time for the first i_k-square, in the backward direction of time, on the line ancestral to $D^{(k)}$ (that is, the *rightmost* i_k-square on this line in our graphical representation). Clearly, T_1 is an $\mathrm{Exp}(\varrho^{(k)})$-random variable conditioned to be $\leqslant t$, and independent of $\mathfrak{X}$.
- Let $\mathfrak{Y}_1 \in X$ be the type at the root of the independent $\mathrm{ASRG}^{(k)}_{t-T_1}$ attached upon encountering the i_k-square at time T_1, that is, an independent sample from $\omega^{(k)}_{t-T_1}$.

We then have (compare Fig. 4.8)

$$\mathfrak{Z} = (1 - B_1)\mathfrak{X} + B_1\big(\pi_{C^{(k)}}(\mathfrak{X}), \pi_{D^{(k)}}(\mathfrak{Y}_1)\big). \tag{4.33}$$

We now iterate Eq. (4.33), see Figure 4.10. In the first step, we draw $\mathfrak{X}$ and B_1 as above. If $B_1 = 1$, we also draw T_1 according to $\mathrm{Exp}(\varrho^{(k)})$, conditioned on being $\leqslant t$. If $B_1 = 0$, we set $\mathfrak{Z} = \mathfrak{X}$. If $B_1 = 1$, by Eq. (4.33) we have to construct $\mathfrak{Y}_1$, which contributes the tail. Since $\mathfrak{Y}_1$ is the type at the root of an $\mathrm{ASRG}^{(k)}_{t-T_1}$, we do this by applying Eq. (4.33) to $\mathfrak{Y}_1$ instead of $\mathfrak{Z}$, that is, we repeat the first step but replace t by $t - T_1$. So we determine whether or not there is a recombination square on the ancestral line of $D^{(k)}$ between 0 and $t - T_1$; if there is one, we determine the waiting time for it, and so forth. More explicitly, let B_2 be the new indicator variable, which is Bernoulli with success probability $1 - \mathrm{e}^{-\varrho^{(k)}(t-T_1)}$. If $B_2 = 0$, let $\mathfrak{X}_1$ be the type at the root of an independent copy of the $\mathrm{ASRG}^{(k-1)}_{t-T_1}$. If $B_2 = 1$, let T_2 be the waiting time for the new event; T_2 follows $\mathrm{Exp}(\varrho^{(k)})$ conditioned to be $\leqslant t - T_1$; and let $\mathfrak{Y}_2$ be the type at the root of an independent $\mathrm{ASRG}^{(k-1)}_{t-T_1-T_2}$. Then, inserting this back into Eq. (4.33), we obtain

$$\mathfrak{Z} = (1 - B_1)\mathfrak{X} + B_1(1 - B_2)\big(\pi_{C^{(k)}}(\mathfrak{X}), \pi_{D^{(k)}}(\mathfrak{X}_1)\big) + B_1B_2\big(\pi_{C^{(k)}}(\mathfrak{X}), \pi_{D^{(k)}}(\mathfrak{Y}_2)\big);$$

note that, if $B_1 = 0$, B_2 has not been declared, but the terms involving it remain well-defined

since B_1 vanishes. Iterating this further gives

$$\mathfrak{Z} = (1 - B_1)\mathfrak{X} + \sum_{i \geqslant 1} B_1 \cdot \ldots \cdot B_i (1 - B_{i+1}) (\pi_{C^{(k)}}(\mathfrak{X}), \pi_{D^{(k)}}(\mathfrak{X}_i)), \tag{4.34}$$

where $\mathfrak{X}_i$ is the type at the root of an independent $\mathrm{ASRG}^{(k-1)}_{t-\sum_{j=1}^{i} T_j}$, and we adhere to the above convention concerning undeclared B_i. Note that, with probability 1, exactly one of the terms on the right-hand side is nonzero; in particular, $B_1 \cdot \ldots \cdot B_i = 0$ whenever $\sum_{j=1}^{i} T_j > t$, so everything is well defined.

Let us now interpret the arrival times T_j of the i_k-squares as arrival times in a Poisson set W with intensity measure $\varrho^{(k)} \mathbb{1}_{t \geqslant 0} \, \mathrm{d}t$ and elements $A_1 < A_2 < \ldots$. When $A_i \leqslant t$, we have $A_i = \sum_{j=1}^{i} T_j$. Furthermore, $B_1 = \mathbb{1}_{\{A_1 \leqslant t\}}$ and, for $i \geqslant 1$,

$$\begin{aligned} B_1 \cdot \ldots \cdot B_i &= \mathbb{1}_{\{A_i \leqslant t\}}, \qquad \text{as well as} \\ B_1 \cdot \ldots \cdot B_i (1 - B_{i+1}) &= \mathbb{1}_{\{A_i \leqslant t < A_{i+1}\}}. \end{aligned} \tag{4.35}$$

We now note that B_1 may also be written as $B_1 = \mathbb{1}_{\{W \cap [0,t] \neq \varnothing\}}$. Together with (4.35), this entails that the nonzero term in (4.34) is the first one if $W \cap [0,t]$ is empty; and if the set is nonempty, then the nonzero term is the one with the index i that satisfies $A_i = \max(W \cap [0,t])$. Conditionally on $B_1 = 1$, we therefore set $T := t - \max(W \cap [0,t])$. The claim then follows by identifying B with B_1, and by noting that, due to the homogeneity of the Poisson process, T has the same distribution as T_1, namely $\mathrm{Exp}(\varrho^{(k)})$ conditioned to be $\leqslant t$. □

Remark 4.14. Remembering the original motivation of the collapsed $\mathrm{ASRG}^{(k)}$, we think of every i_k-square as the anchor point for a new independent copy of the $\mathrm{ASRG}^{(k)}$, which is collapsed to keep things tidy. In the above proof, however, we iteratively expand the i_k-squares *on the ancestral line of* $D^{(k)}$ until there are no more recombination events left on that particular line. Therefore, the Poisson point set W has an interpretation as the collection of all recombination events that occurred on the ancestral line of $D^{(k)}$. The proof has made precise the previously heuristic notion of the last splitting event at site i_k encountered on the ancestral line of $D^{(k)}$ in the backward direction of time; that is, the leftmost event in the graphical representation. For an illustration, see Figure 4.10. ♢

Remark 4.15. When sampling $\mathfrak{Y}_1$ via the newly attached $\mathrm{ASRG}^{(k)}$ in (4.33), the reader may wonder whether one might be able to cut corners and only construct the potential ancestry of the tail — after all, the head of $\mathfrak{Y}_1$ does not enter $\mathfrak{Z}$. However, it cannot be overemphasised that this is not the case! Although $\mathfrak{Y}_1$ only contributes the tail, the branching events in its ancestry can only be resolved if the letter at the selected site is known, whence we need to also trace back the ancestry of the head attached to the new tail. Once more, we are haunted by marginalisation inconsistency due to selection, as discussed in Section 4.3; see in particular Remark 4.7. ♢

We are now set to re-prove Theorem 4.6. Indeed, Proposition 4.11 makes a connection between

the random variable $\mathfrak{Z}$, distributed according to an ASRG$^{(k)}$, and random variables $\mathfrak{X}$ and $\widetilde{\mathfrak{X}}$, distributed according to an ASRG$^{(k-1)}$. This is the crucial observation that we will now exploit.

Genealogical proof of Theorem 4.6. From Proposition 4.11, we can extract the conditional distribution of $\mathfrak{Z}$ given B and T:

$$\mathbb{P}(\mathfrak{Z} = \cdot \mid B, T) = (1-B)\omega_t^{(k-1)} + B\pi_{C^{(k)}}.\omega_t^{(k-1)} \otimes \pi_{D^{(k)}}.\omega_T^{(k-1)}.$$

Theorem 4.6 now follows by integrating out B and T, where we denote the distribution of T by λ.

$$\begin{aligned}\omega_t^{(k)} &= \mathbb{P}(B=0)\omega_t^{(k-1)} + \mathbb{P}(B=1)\pi_{C^{(k)}}.\omega_t^{(k-1)} \otimes \int_0^\infty \pi_{D^{(k)}}.\omega_\tau^{(k-1)}\,\mathrm{d}\lambda(\tau)\\ &= \mathrm{e}^{-\varrho^{(k)}t}\omega_t^{(k-1)} + \pi_{C^{(k)}}.\omega_t^{(k-1)} \otimes \int_0^t \varrho^{(k)}\mathrm{e}^{-\varrho^{(k)}t}\pi_{D^{(k)}}.\omega_\tau^{(k-1)}\,\mathrm{d}\tau,\end{aligned}$$

and we are done. □

4.6 Interlude

Using our insight from the proof of Proposition 4.11, we now informally describe a more efficient version of the ASRG in order to motivate the more elegant dual process and the formal duality results that are detailed and proved in the next section. We start with an untyped ASG, or, equivalently, an ASRG$^{(0)}$, since this marks the beginning of the recursion. Recall that, in the iteration leading from $\omega^{(0)}$ to $\omega^{(1)}$ via the cASRG$^{(1)}$, i_1-recombination squares are laid down at rate $\varrho^{(1)}$ independently on every line of the ASG. But at most one of these squares turns out as relevant; namely the rightmost square on the ancestral line of $D^{(1)}$, if there is such a square. Recall also that the head of the root individual of the ASRG$^{(1)}$, that is its sites in $C^{(1)}$, are delivered by the initial ASG, independently of any recombination squares; while the sites in the tail are delivered by an independent copy of the ASRG$^{(1)}$, attached below the square for the remaining time and processed in the same way as the initial one, in an iterative fashion. This procedure stops when no further recombination square is found on the ancestral line of the tail.

In order to reduce the number of lines and graphical elements in the ASRG$^{(1)}$ to the essential ones, we now start over and decorate the ASG with *at most one recombination event*, which will play the role of the relevant one, see Figure 4.11. Namely, with probability $\mathrm{e}^{-\varrho^{(1)}t}$, we include no event, and both head and tail are delivered by the ASG. With probability $1-\mathrm{e}^{-\varrho^{(1)}t}$, we include one event, which happens at time T_1 distributed according to $\mathrm{Exp}(\varrho^{(1)})$ conditioned to be $\leqslant t$. Since we are in an untyped setting and do not know which of the lines in the ASG will be ancestral to the head, we symbolise the event by an i_1-*bar* that hits all lines at the same time. Below the bar, we attach an independent copy of the ASG starting

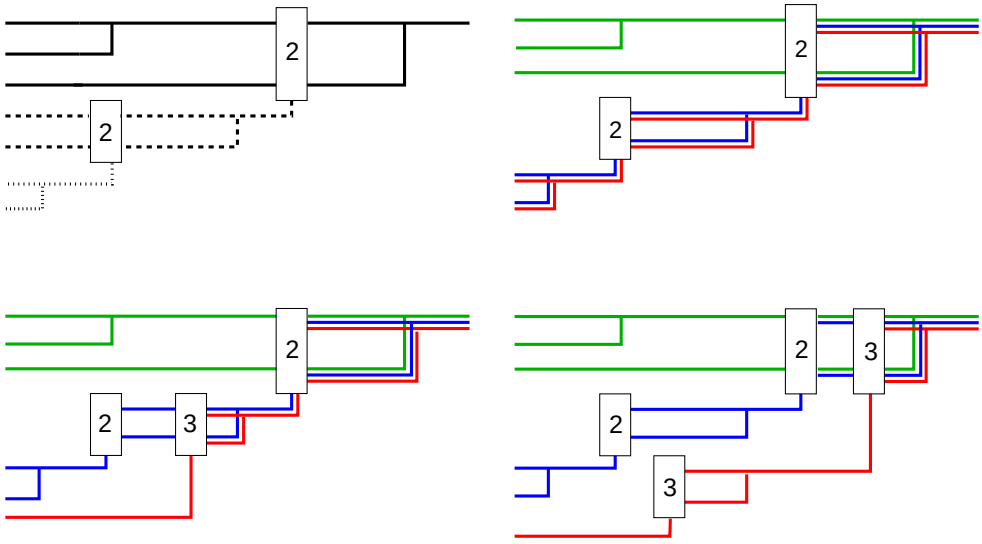


Figure 4.11. Constructing the essential ASRG for three sites with $i_* = 1$. Top left: An ASG decorated with a 2-bar to which decorated ASGs are attached repeatedly; solid, dashed, and dotted lines correspond to steps 1, 2, and 3, respectively. Top right: Labelling and pruning the resulting graph; green, blue and red encode sites 1,2 and 3, respectively. Bottom: Adding 3-bars to the top right graph. If none occurs, the graph remains unchanged. Two different realisations including 3-bars are shown bottom left and bottom right.

with a single line and running for the remaining time. The new ASG is processed in the analogous way, with t replaced by $t - T_1$. This procedure stops when no further i_1-bar is encountered; this is (almost surely) the case after a finite number of steps, see Figure 4.11 (top left). The initial ASG delivers the head, while the last ASG attached delivers the tail. In particular, at every i_1-bar, the tail delivered by the ASG attached below is combined with the head of whichever of the lines running through the bar will turn out to be ancestral to the root of the ASG it belongs to.

We now decorate each line in the graph with the set A of sites in the root individual to which the line is potentially ancestral. This will finally allow us to prune away those lines that are not informative for the type of the root, see Figure 4.11 (top right). We start with the label $A = S$ for the single line at the root. When a branching event occurs to a line labelled A, both branches inherit the label. Upon encountering an i_1-bar, the ASG that continues through the bar to the left is ancestral to $A \cap C^{(1)}$, while the new independent copy attached below is ancestral to $A \cap D^{(1)}$. If $A \cap C^{(1)} = \varnothing$ (this applies in the case of a second and any further i_1-bar), we prune the lines to the left away, because they are neither ancestral to any sites in A at the root, nor do they affect their ancestry. The latter is true because now the same new tail is provided for *all* potential ancestors of the head, at the same moment; in contrast to the original ASRG, where a new tail may compete with others, see Figure 4.8.

We finally work up the recursion by decorating the set of lines potentially ancestral to $D^{(2)}$ with i_2-bars, adding new ASGs, labelling, and pruning in the analogous way, see Figure 4.11

(bottom). That is, with probability $\mathrm{e}^{-\varrho^{(2)}t}$, no i_2-bar appears. With probability $1-\mathrm{e}^{-\varrho^{(2)}t}$, we add an i_2-bar, at a time distributed according to $\mathrm{Exp}(\varrho^{(2)})$ conditioned to be $\leqslant t$. A new ASG labelled $D^{(2)}$ is then attached below, starting with a single line, while the lines that continue through the bar now carry the label $D^{(1)} \cap C^{(2)}$. If a second $\mathrm{Exp}(\varrho^{(2)})$ waiting time still falls within the remaining time, a second i_2-bar occurs, with no lines running through it and a single line labelled $D^{(2)}$ starting a new ASG below; and so on until no further i_2-bar is encountered in the remaining time.

We continue like this until S^* is exhausted. The resulting graph is the *essential ASRG*. Rather than constructing it via recursion over S^* with successive addition of bars, labelling, and pruning, it can also be produced in one go in a Markovian manner, according to the following procedure.

- Start with a single line labelled S.
- Every line independently branches at rate s; both offspring lines inherit the label of the parent.
- Every set of lines that carry the same label, say A, independently receives an i-bar at rate ϱ_i for every $i \in S^*$ with $A \cap D_i \neq \varnothing$, upon which either of the following happens.
  - If $A \cap D_i \neq A$, the lines continue through the bar and change their labels to $A \cap C_i$; a new single line labelled $A \cap D_i$ starts below the bar.
  - If $A \cap D_i = A$, no lines continue through the bar and a new single line labelled A starts below the bar.
- Stop when time horizon t is reached.

Note that the resulting graph may be conceived as a collection of (conditionally) independent ASGs, each with its own label, and joined together by recombination bars. It is now easy to see that all the relevant information can be condensed into a *weighted partitioning process*, namely a Markov process in continuous time that holds, at any time, an interval partition $\mathcal{A}$ of S into the blocks $A \subseteq \mathcal{A}$ of potentially ancestral sites, together with weights v_A giving the number of lines in the respective ASGs. This will be formalised in the next section.

4.7 Duality

For the genealogical proof of the recursive solution in Theorem 4.6, we relied on the graphical construction, which implicitly assumes a duality between the ASRG and the solution of the SRE. Since the ASRG is somewhat unwieldy and difficult to formalise, our goal in this section is to construct a simpler dual processes. Let us begin with our definition of duality for Markov processes, which is a straightforward extension of the standard concept (see [Lig10] or [JK14] for thorough expositions, and [Möh99] for an early application to models of population genetics).

Definition 4.13. Let $\mathcal{X} = (\mathcal{X}_t)_{t\geqslant 0}$ and $\mathcal{Y} = (\mathcal{Y}_t)_{t\geqslant 0}$ be two continuous-time Markov chains with state spaces E and F, respectively. These chains are said to be *dual* with respect to some measurable function $H : E \times F \to \mathbb{R}^d$ if

$$\mathbb{E}[H(\mathcal{X}_t, y) \mid \mathcal{X}_0 = x] = \mathbb{E}[H(x, \mathcal{Y}_t) \mid \mathcal{Y}_0 = y]$$

holds for all $t \geqslant 0, x \in E$, and $y \in F$. Furthermore, H is referred to as a *duality function* for $\mathcal{X}$ and $\mathcal{Y}$. We use the triple $(\mathcal{X}, \mathcal{Y}, H)$ to denote the duality. ♢

Remark 4.16. The slight extension of the standard concept consists in allowing for an $\mathbb{R}^d$-valued duality function instead of the usual real-valued H. This is, of course, equivalent to introducing a family of d real-valued duality functions. It touches on the interesting problem of finding *all* duality functions for a given pair of Markov processes. The corresponding *duality space* has been introduced in [Möh99] and investigated in [Möh13]. ♢

Motivated by our observation at the end of Section 4.6, we now define a suitable dual process for ω, and a corresponding duality function. More precisely, we will find three different processes dual to ω, namely the weighted partitioning process, a Yule process with initiation and resetting, and an initiation process, each linked to ω via a suitable duality function, and each providing different insight.

4.7.1 The weighted partitioning process

For the first dual process, we refer back to the essential ASRG and now show formally that all the information required for reconstructing the genetic type of an individual sampled from the present-day population can be encoded in the form of a *weighted partitioning process* together with the initial condition ω_0. Just as in the neutral case, the partitioning describes how the genotype of a given individual is pieced together from the genetic material of its ancestors. In order to include selection, a positive integer (weight) is assigned to each block, denoting the number of potential ancestors for the sites contained in that block (the number of lines in the ASG labelled with this block). As in the single-site case (compare Figure 4.6), the true ancestor will be of deleterious type if and only if *all* potential ancestors are of deleterious type. We now formally define the weighted partitioning process.

Definition 4.14. The *weighted partitioning process* (WPP) is a continuous-time Markov chain $(\Sigma, V) = (\Sigma_t, V_t)_{t\geqslant 0}$ with (countable) state space

$$F := \bigcup_{k\geqslant 0} (\mathcal{I}_k(S) \times \mathbb{N}_+^k),$$

where $\mathcal{I}_k(S)$ denotes the set of all interval partitions of S into exactly k blocks, and transitions

1. $(\mathcal{A}, v) \longrightarrow (\mathcal{A}, w)$ at rate sv_A if $w_A = v_A + 1$ for some $A \in \mathcal{A}$ and $w_B = v_B$ for all $A \neq B \in \mathcal{A}$.

2. $(\mathcal{A}, v) \longrightarrow (\mathcal{A} \wedge \{C_i, D_i\}, w)$ at rate ϱ_i if, for $i \in S^*$ and the unique $A \in \mathcal{A}$ with $|\{C_i, D_i\}|_A| = 2$, $w_{A\cap C_i} = v_A$, $w_{A\cap D_i} = 1$, and $w_B = v_B$ for all $A \neq B \in \mathcal{A}$.

3. $(\mathcal{A}, v) \longrightarrow (\mathcal{A}, w)$ at rate $\varrho_{\min(A)}^{A\cup\{i_*\}}$ if, for some $A \in \mathcal{A}$, $w_A = 1$ and $w_B = v_B$ for all $B \neq A$ (the minimum is in the sense of $\preccurlyeq$). ♢

Note that transition (3) is silent if $v_A = 1$.

The intuitive explanation for the dynamics of the WPP connects to the essential ASRG at the end of Section 4.6. Clearly, $(\Sigma_t, V_t) = (\mathcal{A}, v)$ represents the set of ASGs present at time t, where each block A of $\mathcal{A}$ corresponds to one ASG with v_A lines. For every $i \in S^*$, every A splits into $A \cap C_i$ and $A \cap D_i$ at rate ϱ_i independently of all other blocks. If this split is nontrivial, then $A \cap C_i$ inherits the weight of A (reflecting the lines that pass through the bar), while the weight of $A \cap D_i$ is set to 1 (reflecting the new ASG attached below the bar and starting with a single line); this gives transition (2). If $A \subseteq C_i$, nothing happens. If $A \subseteq D_i$, the weight is reset to 1 (again reflecting the new ASG attached below the bar); note that this happens whenever the split leaves A intact but separates it from the selected site, which gives rise to the total rate of $\varrho_{\min(A)}^{A\cup\{i_*\}}$ in transition (3). Note also that the marginal Σ is exactly the partitioning process described in Example 4.2. Independently of everything else, every block experiences Yule branching at rate s (transition (1)). Based on the WPP, we now define the corresponding candidate for our duality function.

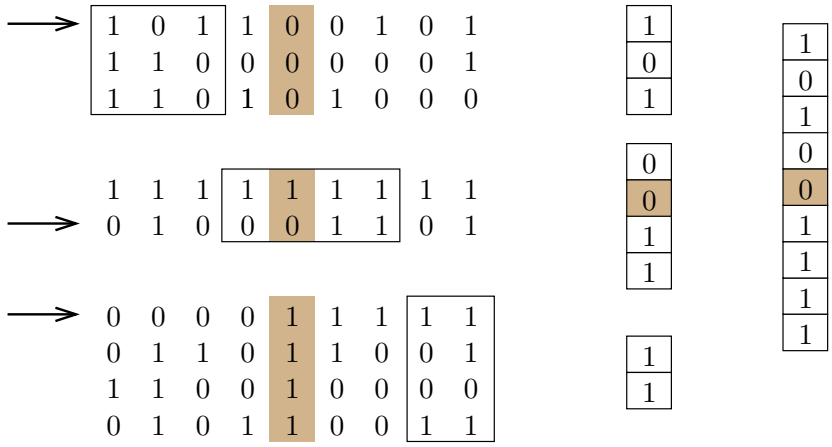


Figure 4.12. Illustration of the duality function H for a WPP in state $(\mathcal{A}, v)$, where $\mathcal{A} = \{\{1,2,3\}, \{4,5,6,7\}, \{8,9\}\}$ and $v_{\{1,2,3\}} = 3$, $v_{\{4,5,6,7\}} = 2$, and $v_{\{8,9\}} = 4$. The selected site is $i_* = 5$ and highlighted in light brown. As prescribed by $(\mathcal{A}, v)$, we sample 3 potential ancestors (displayed horizontally on the left) for the first, 2 for the second, and 4 for the third block of sites, all i.i.d. according to ν. The true ancestor (marked by an arrow) is then sampled uniformly at random from all individuals of beneficial type within the respective samples, except in the case of the third block, since there are no individuals of beneficial type. The resulting marginal types for the individual blocks (middle) are then merged into the sequence on the right. The distribution of this sequence is $H(\mathcal{A}, v; \nu)$.

Definition 4.15. For an interval partition $\mathcal{A}$ of S and associated weights $v := (v_A)_{A\in\mathcal{A}}$ we define

$$H(\mathcal{A}, v; \nu) := \bigotimes_{A\in\mathcal{A}} \pi_A.\Big((1 - f(\nu))^{v_A} d(\nu) + (1 - (1 - f(\nu))^{v_A}) b(\nu)\Big)$$

for all $\nu \in \mathcal{P}(X)$. ♢

The function H has the following meaning, which is illustrated in Figure 4.12. For a given $(\mathcal{A}, v)$ and every $A \in \mathcal{A}$, we sample one sequence according to ν for each of the v_A leaves of the corresponding ASG. The type at the root of this ASG is then distributed according to $b(\nu)$ (according to $d(\nu)$) if at least one of the leaves (none of the leaves) carries a beneficial type, just as in the case of pure selection in Example 4.1. Finally, the sequence at the root of the ASRG is pieced together from the sequences at the roots of the individual ASGs by taking, for every $A \in \mathcal{A}$, the sites in A from the root of the ASG corresponding to A. The resulting sequence is distributed according to $H(\mathcal{A}, v; \nu)$; note that $H(\mathcal{A}, v; \nu)$ may be understood as a probability vector on X, that is, a vector in $\mathbb{R}^{2^n}$. For the time being, let us refrain from proving the resulting duality and proceed to a more convenient representation of the WPP.

4.7.2 The Yule process with initiation and resetting.

Keeping in mind that we are only dealing with single-crossover recombination (and, therefore, only interval partitions), we will take advantage of the following one-to-one correspondence between (weighted) partitions and assignments of non-negative integers to the sites of the sequence (see Figure 4.13). Let a collection $m = (m_k)_{1\leqslant k\leqslant n}$ of non-negative integers with $m_{i_*} > 0$ be given. We then obtain an (interval) partition by the rule that two sites $i \prec j$ belong to the same block if and only if $m_k = 0$ for all $i \prec k \preccurlyeq j$; intuitively, the non-zero integers tell us where to chop up the sequence. We obtain in this way a partition $\mathcal{A}$ in which, for each block $A \in \mathcal{A}$, $m_{\min(A)} > 0$, while $m_i = 0$ for $\min(A) \neq i \in A$ (where the minimum is with respect to $\preccurlyeq$, and is unique since $\mathcal{A}$ is an interval partition). We then assign a weight to block A by setting $v_A := m_{\min(A)}$. Likewise, we may encode a weighted partition as a collection m of integers by assigning the weight of each block to its respective minimal site and 0 to all others. Since i_* is the unique minimal element of S, one always has $m_{i_*} > 0$. Explicitly, $m_{i_*} = v_A$ for the unique A that contains i_* and, for $i \neq i_*$,

$$m_i = \begin{cases} 0, & \text{if } \overleftarrow{i} \text{ and } i \text{ are in the same block,} \\ v_A \text{ for the unique } A \text{ that contains } i, & \text{otherwise,} \end{cases}$$

with $\overleftarrow{i}$ as in Definition 4.1.

The new encoding allows us to rewrite H of Definition 4.15 in a convenient way, where we also take advantage of the formalism introduced in Section 4.4.

Lemma 4.16. *Let H be as in Definition 4.15. For $m \in \mathbb{N}_0^S$ with $m_{i_*} > 0$, let $(\mathcal{A}(m), v(m))$*

v	3	4	1	2
$\mathcal{A}$	$\{\{1,2,3\},\{4,5\},\{6,7,8\},\{9,10\}\}$			
m	0 0 3	0 4	1 0 0	2 0

Figure 4.13. Encoding a weighted partition (top) by a tuple of integers (bottom). The selected site is light brown.

be the weighted partition associated with m, and define

$$\mathcal{H}(m,\nu) := H\big((\mathcal{A}(m), v(m)), \nu\big).$$

Then, one has

$$\mathcal{H}(m,\nu) = \bigboxtimes_{i\in S} h(m_i,\nu)^{D_i} \tag{4.36}$$

where

$$h(m_i,\nu) := \big(1-f(\nu)\big)^{m_i} d(\nu) + \big(1-\big(1-f(\nu)\big)^{m_i}\big) b(\nu) \tag{4.37}$$

for $m_i \neq 0$ and $h(0,\nu) := 1$. The factors are ordered non-decreasingly with respect to $\preccurlyeq$. □

Remark 4.17. When using the product sign $\bigboxtimes$ for products of elements of $\mathfrak{A}(X)$ indexed by S, we *always* understand the factors to be ordered non-decreasingly. ♢

Remark 4.18. At this point, it becomes clear that the special role played by $D_{i_*} = S$ in the definition of the D_i (see Remark 4.1) makes perfect sense. Indeed, the representation (4.36) shows that the contributions to the sequence at the root of the ASRG come from the various ASGs associated to different tails D_i, which are attached to the original one corresponding to $D_{i_*} = S$. This will become even more evident in the context of the initiation process; see the duality function $\mathcal{G}$ in (4.41) and Figure 4.14. ♢

Proof. Recall that, by the minimality of the selected site, we have $C_{i_*} = \varnothing, D_{i_*} = S$ and therefore $\mathcal{H}(m,\nu) = h(m_{i_*},\nu)$ if $m_i = 0$ for all $i \neq i_*$. In all other cases, let i be a maximal site with $m_i \neq 0$. The definitions of $\mathcal{H}$ and H then entail

$$\mathcal{H}(m,\nu) = \mathcal{H}(m',\nu)^{C_i} \otimes h(m_i,\nu)^{D_i} = \mathcal{H}(m',\nu) \boxtimes h(m_i,\nu)^{D_i},$$

where m' is obtained from m by setting m_i to zero. The claim then follows by induction. □

The new encoding also allows us to represent the WPP as a collection of n independent Yule processes with initiation and resetting, to be defined next. In the neutral case, this is similar to the representation of interval partitions in [BB] in terms of the sets of breakpoints.

Definition 4.17. A *Yule process with initiation and resetting* (YPIR) with branching rate $s > 0$, initiation rate $\varrho \geqslant 0$, and resetting rate $r \geqslant 0$ is a continuous-time Markov chain on

$\mathbb{N}_0$ with transitions

$$\begin{array}{lll} \text{(Y)} & m \to m+1 & \text{at rate } sm \text{ for } m > 0, \\ \text{(I)} & 0 \to 1 & \text{at rate } \varrho, \\ \text{(R)} & m \to 1 & \text{at rate } r \text{ for } m > 0. \end{array}$$

Note that transition (R) is silent if $m = 1$. ♢

Given the one-to-one correspondence between $(\mathcal{A}, v)$ and m, it is then easy to see that (Σ, V) is equivalent to a collection M of independent YPIRs, where $M = (M_i)_{i\in S}$. Here, $M_{i_*} = (M_{i_*,t})_{t\geqslant 0}$ is a simple Yule process with branching rate $s > 0$, that is, the degenerate case of a YPIR with initiation and resetting rates $\varrho_{i_*} := r_{i_*} := 0$; for $i \neq i_*$, $M_i = (M_{i,t})_{t\geqslant 0}$ is a YPIR with branching rate s, initiation rate ϱ_i and resetting rate

$$r_i := \sum_{\ell \preccurlyeq i} \varrho_\ell; \tag{4.38}$$

note, in particular, that $r_i \geqslant \varrho_i$. Finally, we also write $M_t := (M_{i,t})_{i\in S}$ for the entire collection at time t. Indeed, the equivalence is clear since the transitions of (Σ, V) and M and the corresponding rates can be matched in a unique way; compare Definitions 4.14 and 4.17. Note that r_i is the total rate at which i is separated from the selected site; it may be understood as the marginal recombination rate $r_i = \varrho_i^{\{i,i_*\}}$, compare (4.18).

Note that our Yule process K (compare Example 4.1) has the law of M_{i_*}. Let us recapitulate from [BCH18] the duality result for the pure selection equation, which is a slight extension of Example 4.1.

Proposition 4.18. *Let K be a Yule process with branching rate s. For $k \geqslant 1$ and $\nu \in \mathcal{P}(X)$, define $h(k, \nu)$ as in Eq.* (4.37)*. Then,*

$$h\big(k, \varphi_t(\nu)\big) = \mathbb{E}\big(h(K_t, \nu) \mid K_0 = k\big),$$

where φ is the selection semigroup.

Proof. Combining Eqs. (4.37), (4.24) and (4.31), one gets

$$\begin{aligned} &h\big(k,\varphi_t(\nu)\big) \\ &\quad = \big(\mathbb{E}\big[(1-f(\nu))^{K_t} \mid K_0 = 1\big]\big)^k d(\varphi_t(\nu)) + \big(1 - \big(\mathbb{E}\big[(1-f(\nu))^{K_t} \mid K_0 = 1\big]\big)^k\big) b(\varphi_t(\nu)) \\ &\quad = \mathbb{E}\big(h(K_t, \nu) \mid K_0 = k\big), \end{aligned}$$

where the last step follows from the fact that a collection of k independent Yule processes, each started with a single line, is equivalent to a Yule process started with k lines. □

Let us still postpone the duality result in the case with recombination to the next section, since the proof is most convenient on the basis of the initiation process.

4.7.3 The initiation process.

Let us first try to gain some intuition by representing the duality function from Lemma 4.16 in terms of (box-)products of elements of the selection semigroup at various times. To this end, recall first from Proposition 4.5 that $\varphi_t(\nu)$ is, for all ν and t, a convex combination of the conditional type distributions $d(\nu)$ and $b(\nu)$, and so is $h(k,\nu)$ for all $k \geqslant 1$, see Eq. (4.37). Since $f(\varphi_t(\nu))$ is strictly increasing in t (compare Proposition 4.5), there exists, for all $k \geqslant 1$ and $s > 0$, a unique $\theta(k) \in \mathbb{R}$ such that $\big(1 - f(\nu)\big)^k = 1 - f\big(\varphi_{\theta(k)}(\nu)\big)$ and thus,

$$h(k,\nu) = \varphi_{\theta(k)}(\nu). \tag{4.39}$$

Note that $\theta(1) = 0$ since $h(1,\nu) = \nu = \varphi_0(\nu)$. Then, setting $\theta(0) := \Delta$ and $\varphi_\Delta(\nu) := 1$ for all ν (in line with $h(0,\cdot) = 1$ in Lemma 4.16), we can write, using the representation from Lemma 4.16,

$$\mathcal{H}(m,\nu) = \bigboxtimes_{i\in S} h(m_i,\nu)^{D_i} = \bigboxtimes_{i\in S} \varphi_{\theta(m_i)}(\nu)^{D_i} =: \mathcal{G}\big(\theta(m),\nu\big), \tag{4.40}$$

where $\theta(m) := (\theta(m_i))_{i\in S}$. More generally, this leads to the ansatz

$$\mathcal{G}(\theta,\nu) := \bigboxtimes_{i\in S} \varphi_{\theta_i}(\nu)^{D_i} \tag{4.41}$$

for a third (putative) duality function. Here, $\theta = (\theta_i)_{i\in S} \in \mathbb{R}_{\geqslant 0}^{i_*} \times (\mathbb{R}_{\geqslant 0} \cup \{\Delta\})^{S^*}$, where the symbol Δ is used to indicate that the factor is absent from the product; thus, we have $\theta_{i_*} \neq \Delta$, in analogy to $m_{i_*} > 0$. Recall that the factors in the product are ordered non-decreasingly w.r.t. $\preccurlyeq$ and note that its value is the same for all such orderings since incomparable factors commute by Proposition 4.7, (ii).

Recall that m in (4.36) corresponds to a partition of S in which each block is weighted by a positive integer, counting the number of lines in the associated ASG (as part of an essential ASRG, see Section 4.6). Similarly, θ in Eq. (4.41) also encodes a partition of S (the role of 0 now being played by Δ), only this time, the blocks are not weighted by the number of lines in the associated ASGs, but by their runtimes (again, seen as part of an essential ASRG). In the sampling step, we average over all realisations of the ASG with the indicated runtime, and thus obtain $\mathcal{G}$ from $\mathcal{H}$ by replacing the factors $h(m_i,\nu)$ in $\mathcal{H}(m,\nu)$ by

$$\varphi_{\theta_i}(\nu) = \mathbb{E}[h(K_{\theta_i},\nu) \mid K_0 = 1];$$

this will later make the connection to the transformation (4.39).

We now give an informal description of the *initiation process* which will take the role of the YPIR. It is a continuous-time Markov process, and its transition rates relate to that of the YPIR as follows. As Δ takes the role of 0, the transition (I) (initiation) in Definition 4.17 corresponds to a transition from Δ to 0. Similarly, as 0 takes the role of 1, a reset (R) (to 1)

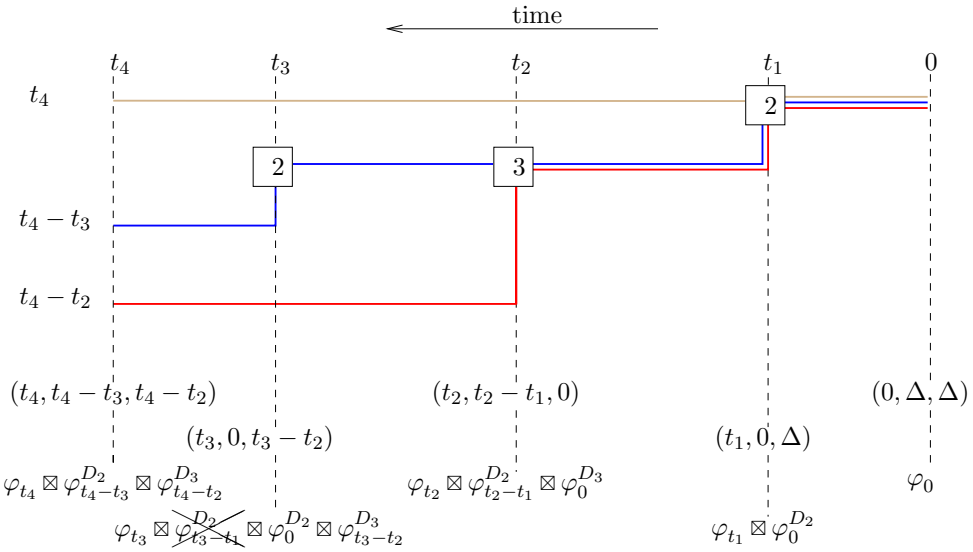


Figure 4.14. A realisation of the essential ASRG, where every ASG is collapsed into a single line. It describes the evolution of a partitioning process whose blocks are weighted by the time that has passed since the corresponding ASG was attached. The colour coding is the same as in Figure 4.11. Light brown, blue and red for site 1,2 and 3; as before, the first site is selected. Below the graph, we indicate the evolution of the associated collection of initiation processes Θ. At the bottom, we see how the function $\mathcal{G}(\Theta_t, \cdot)$, defined in Eq. (4.41), evolves in time. Every factor corresponds to a different line, and attachment of a new line due to an i-recombination event corresponds to multiplication from the right by $\varphi_0^{D_i}$; subsequently, the time index in each factor evolves on its own. Notice the cancellation that occurs at time t_3; it corresponds to the discontinuation of the line at the recombination bar and the reset of the second component of Θ, due to $\{2\} \cap D_2 = \{2\}$.

of the YPIR corresponds to a reset (to 0) of the initiation process. Keeping in mind that (Y) describes the branching of the ASG (and that we now only want to record its runtime), we replace these random jumps by a deterministic and continuous increase. Thus, Θ_t is either Δ, signifying that it has not yet been initiated, or its value is just the time that has passed since the last reset. Finally, when no resetting occurs, we have $\Theta_t = \Theta_0 + t$.

This can be condensed into the following definition; for an illustration, see Fig. 4.14.

Definition 4.19. We define the *initiation process* with initiation rate $\varrho \geqslant 0$ and resetting rate $r \geqslant 0$ as the continuous-time Markov process with values in $\mathbb{R}_{\geqslant 0} \cup \{\Delta\}$ and its generator mapping $u \in C^1(\mathbb{R})$ to $\tilde{u}$, which is defined via

$$\begin{aligned} \tilde{u}(t) &= \dot{u}(t) + r\big(u(0) - u(t)\big) \quad \text{for } t \in \mathbb{R}_{\geqslant 0}, \\ \tilde{u}(\Delta) &= \varrho\big(u(0) - u(\Delta)\big). \end{aligned} \tag{4.42}$$

For later use, we define Θ as the following collection of independent initiation processes, where $\Theta = (\Theta_i)_{i \in S}$. The process $\Theta_i = (\Theta_{i,t})_{t \geqslant 0}$ has initiation rate ϱ_i and resetting rate r_i (compare (4.38)). In particular, since $\varrho_{i_*} = r_{i_*} = 0$, all stochastic contributions in Eq. (4.42) vanish for this choice, and what remains is a purely deterministic drift, that is $\Theta_{i_*,t} = t + \Theta_{i_*,0}$. We denote by $\mathcal{L}_i$ the generator of Θ_i. Furthermore, $\mathcal{L} := \sum_{i \in S} \mathcal{L}_i$, where $\mathcal{L}_i$ acts on the i-th component of the argument. ♢

Note that Θ shares the parameters ϱ_i and r_i with M, but it does not depend on s. Rather, for any given s, Θ and M are related at the level of an expectation, as we now show. First, we prove the duality result for the triple $(\omega, \Theta, \mathcal{G})$. From there, we recover the duality $(\omega, M, \mathcal{H})$ and, equivalently, $(\omega, (\Sigma, V), H)$. The first step is to see that the YPIR and the initiation process are related at the level of expectations.

Proposition 4.20. *For all $i \in S$, the* YPIR *M_i and the initiation process Θ_i satisfy*

$$\mathbb{E}\big(h(M_{i,t}, \nu) \mid M_{i,0} = m_i\big) = \mathbb{E}\big(\varphi_{\Theta_{i,t}}(\nu) \mid \Theta_{i,0} = \theta(m_i)\big)$$

for all $m_i \in \mathbb{N}_0$ and $t \geqslant 0$.

Proof. It suffices to show that the left- and right-hand side of the statement solve the same initial value problem. By (4.39), the expressions agree at $t = 0$. It remains to be shown that

$$\mathcal{Q}_i h(\cdot, \nu)(m_i) = \mathcal{L}_i \varphi_\cdot(\nu)\big(\theta(m_i)\big),$$

where $\mathcal{Q}_i$ is the generator of M_i, and $\mathcal{L}_i$ that of Θ_i. Comparing Definitions 4.17 and 4.19, it is obvious that the transitions from m to 1 in the YPIR (at rate ϱ_i if $m_i = 0$ and at rate r_i if $m_i > 0$) correspond to transitions to 0 in the initiation process (at rate ϱ_i if $\Theta_i = \Delta$ and at rate r_i if $\Theta_i \in \mathbb{R}_{\geqslant 0}$). The identity (4.39) then implies the equality of the corresponding contributions to the left and right-hand side, i.e.

$$h(1,\nu) - h(m_i,\nu) = \varphi_0(\nu) - \varphi_\Delta(\nu) \quad \text{for } m = 0, \quad \text{and}$$
$$h(1,\nu) - h(m_i,\nu) = \varphi_0(\nu) - \varphi_{\theta(m_i)}(\nu) \quad \text{for } m > 0.$$

Furthermore, it is a direct consequence of Proposition 4.18 together with (4.39) that the time derivative corresponds to branching of the YPIR, that is,

$$\dot{\varphi}_{\theta(m_i)}(\nu) = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\big(h(K_t,\nu) \mid K_0 = m_i)\big)|_{t=0} = sm_i\big(h(m_i+1,\nu) - h(m_i,\nu)\big)$$

by the Kolmogorov backward equation for the Yule process. □

Returning now to $\mathcal{H}$ and $\mathcal{G}$, we obtain immediately, by independence:

Corollary 4.21. *The families M and Θ of independent* YPIRs *and initiation processes satisfy*

$$\mathbb{E}\big(\mathcal{H}(M_t,\nu) \mid M_0 = m\big) = \mathbb{E}\big(\mathcal{G}(\Theta_t,\nu) \mid \Theta_0 = \theta(m)\big)$$

for all $m \in \mathbb{N}_0^n$ and $t \geqslant 0$.

We are now set to state the main result of this section, the duality for $(\omega, \Theta, \mathcal{G})$.

Theorem 4.22. *Let Θ be the family of independent initiation processes introduced in Definition* 4.19*. Then, with $\mathcal{G}$ as in* (4.41)*, we have, for all $\nu \in \mathcal{P}(X)$ and all $\theta \in \mathbb{R}_{\geqslant 0}^{i_*} \times (\mathbb{R}_{\geqslant 0} \cup \{\Delta\})^{S^*}$,*

$$\mathcal{G}\big(\theta, \psi_t(\nu)\big) = \mathbb{E}\big(\mathcal{G}(\theta,\omega_t) \mid \omega_0 = \nu\big) = \mathbb{E}\big(\mathcal{G}(\Theta_t,\nu) \mid \Theta_0 = \theta\big),$$

where $\psi = (\psi_t)_{t\geqslant 0}$ is the flow of the SRE *introduced in Definition* 4.4*.*

Proof. The first equality is clear because ψ is deterministic. For the proof of the second equality (that is, the duality relation), it will be useful to think of the solution of the SRE (4.9) as a deterministic Markov process with generator $\widetilde{\Psi} = \widetilde{\Psi}_{\mathrm{sel}} + \widetilde{\Psi}_{\mathrm{rec}}$ given by

$$\begin{aligned}\widetilde{\Psi} f(\nu) &:= \frac{\mathrm{d}}{\mathrm{d}t} f\big(\psi_t(\nu)\big)|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t} f\big(\nu + t\Psi_{\mathrm{sel}}(\nu) + t\Psi_{\mathrm{rec}}(\nu)\big)_{t=0} \\ &= \frac{\mathrm{d}}{\mathrm{d}t} f\big(\nu + t\Psi_{\mathrm{sel}}(\nu)\big)|_{t=0} + \frac{\mathrm{d}}{\mathrm{d}t} f\big(\nu + t\Psi_{\mathrm{rec}}(\nu)\big)|_{t=0} \\ &=: \widetilde{\Psi}_{\mathrm{sel}} f(\nu) + \widetilde{\Psi}_{\mathrm{rec}} f(\nu)\end{aligned}$$

for all $f \in C^1(\mathcal{P}(X))$.

As in the proof of Proposition 4.20, we are going to show that the left and right-hand side satisfy the same initial value problem. As their values at $t = 0$ obviously agree (see Eq. (4.39)), it suffices to show that

$$\widetilde{\Psi}\mathcal{G}(\theta,\cdot)(\nu) = \mathcal{L}\mathcal{G}(\cdot,\nu)(\theta) \tag{4.43}$$

for all $\nu \in \mathcal{P}(X)$ and all $\theta \in \mathbb{R}_{\geqslant 0}^{i_*} \times (\mathbb{R}_{\geqslant 0} \cup \{\Delta\})^{S^*}$. (Indeed, if (4.43) is satisfied, it trivially

applies to all components of the $\mathbb{R}^{2^n}$-valued function $\mathcal{G}$ and thus establishes duality also in our slightly extended sense; compare Remark 4.16.) First of all, let us note that, since $\widetilde{\Psi}$ is a differential operator, we have

$$\widetilde{\Psi}\big(\mathcal{G}(\theta,\cdot)\big)(\nu) = \sum_{\substack{j\in S\\ \theta_j\neq\Delta}} \Big(\bigboxtimes_{j\not\preccurlyeq \dot{i}\in S\setminus j} \varphi_{\theta_{\dot{i}}}(\nu)^{D_{\dot{i}}}\Big) \boxtimes \big(\widetilde{\Psi}\circ\varphi_{\theta_j}(\nu)\big)^{D_j} \boxtimes \bigboxtimes_{j\prec \dot{i}\in S} \varphi_{\theta_{\dot{i}}}(\nu)^{D_{\dot{i}}} \tag{4.44}$$

by the product rule, where the underdot indicates the summation variable; note that since $\varphi_\Delta(\nu)=1$, factors with $\theta_i=\Delta$ play no role. Hence, in order to evaluate the left-hand side of Eq. (4.43), we only need to compute $\big(\widetilde{\Psi}(\varphi_{\theta_j})(\nu)\big)^{D_j}$ for all $j\in S$ such that $\theta_j\neq\Delta$. Clearly,

$$\big(\widetilde{\Psi}_{\mathrm{sel}}(\varphi_{\theta_j})(\nu)\big)^{D_j} = \big(\dot{\varphi}_{\theta_j}(\nu)\big)^{D_j} \tag{4.45}$$

because φ is the flow of the pure selection equation. For the recombination part, we calculate

$$\begin{aligned}
&\big(\widetilde{\Psi}_{\mathrm{rec}}(\varphi_{\theta_j})(\nu)\big)^{D_j}\\
&\quad= \Big(\frac{\mathrm{d}}{\mathrm{d}h}\varphi_{\theta_j}\big(\nu + h\Psi_{\mathrm{rec}}(\nu)\big)|_{h=0}\Big)^{D_j}\\
&\quad= \Big(\frac{\mathrm{d}}{\mathrm{d}h}\varphi_{\theta_j}\big(\nu\boxtimes\big(\mathbf{1}\boxplus h\sum_{\ell\in S^*}\varrho_\ell(\nu^{D_\ell}\boxminus\mathbf{1})\big)\big)|_{h=0}\Big)^{D_j}\\
&\quad= \Big(\varphi_{\theta_j}(\nu)\boxtimes\frac{\mathrm{d}}{\mathrm{d}h}\big(\mathbf{1}\boxplus h\sum_{\ell\in S^*}\varrho_\ell(\nu^{D_\ell}\boxminus\mathbf{1})\big)|_{h=0}\Big)^{D_j}\\
&\quad= \sum_{\ell\in S^*}\varrho_\ell\big(\varphi_{\theta_j}(\nu)^{D_j}\boxtimes\nu^{D_\ell\cap D_j} - \varphi_{\theta_j}(\nu)^{D_j}\big)\\
&\quad= \sum_{\substack{\ell\in S^*\\ \ell\preccurlyeq j}}\varrho_\ell\big(\varphi_0(\nu)^{D_j} - \varphi_{\theta_j}(\nu)^{D_j}\big) + \sum_{\substack{\ell\in S^*\\ \ell\succ j}}\varrho_\ell\big(\varphi_{\theta_j}(\nu)^{D_j}\boxtimes\varphi_0(\nu)^{D_\ell} - \varphi_{\theta_j}(\nu)^{D_j}\big).
\end{aligned} \tag{4.46}$$

Here, we have used Lemma 4.8 in the third step, and in the last that $\varphi_0(\nu)=\nu$ together with the fact that the sum over sites incomparable to j vanishes because $D_j\cap D_\ell=\varnothing$ if ℓ is incomparable to j. To simplify the first sum, we took advantage of the fact that $\ell\preccurlyeq j$ implies $D_j\subseteq D_\ell$ together with the cancellation rule from Proposition 4.7. Similarly, $\ell\succ j$ implies $D_\ell\subseteq D_j$, which simplifies the second sum. Inserting (4.46) and (4.45) into (4.44) and recalling Eq. (4.38), we have shown so far that

$$\begin{aligned}
&\widetilde{\Psi}\mathcal{G}(\theta,\cdot)(\nu)\\
&\quad= \sum_{\substack{j\in S\\ \theta_j\neq\Delta}}\Big(r_j\big(\mathcal{G}((\theta_{<j},0,\theta_{>j}),\nu)-\mathcal{G}(\theta,\nu)\big) + \frac{\partial}{\partial\theta_j}\mathcal{G}(\theta,\nu) + \sum_{\ell\succ j}\varrho_\ell\big(\mathcal{G}_{j,\ell}(\theta,\nu)-\mathcal{G}(\theta,\nu)\big)\Big),
\end{aligned}$$

where we use the obvious convention that $(\theta_{<j},0,\theta_{>j})$ is obtained from θ by setting θ_j to 0. Furthermore, $\mathcal{G}_{j,\ell}(\theta,\nu)$ (for $t_j\neq\Delta$ and $j\prec\ell$) arises from $\mathcal{G}(\theta,\nu)$ by inserting the factor $\varphi_0(\nu)^{D_\ell}$ at the immediate right of $\varphi_{\theta_j}(\nu)^{D_j}$. That is, if $\mathcal{G}(\theta,\nu)$ is of the form $\mathcal{G}(\theta,\nu)=$

$A \boxtimes \varphi_{\theta_j}(\nu)^{D_j} \boxtimes B$, then

$$\mathcal{G}_{j,\ell}(\theta,\nu) = A \boxtimes \varphi_{\theta_j}(\nu)^{D_j} \boxtimes \varphi_0(\nu)^{D_\ell} \boxtimes B. \tag{4.47}$$

Hence, if we can show that

$$\sum_{\substack{j\in S\\ \theta_j\neq\Delta}} \sum_{\ell\succ j} \varrho_\ell\big(\mathcal{G}_{j,\ell}(\theta,\nu) - \mathcal{G}(\theta,\nu)\big) = \sum_{\substack{\ell\in S^*\\ \theta_\ell=\Delta}} \varrho_\ell\big(\mathcal{G}((\theta_{<\ell},0,\theta_{>\ell}),\nu) - \mathcal{G}(\theta,\nu)\big), \tag{4.48}$$

it follows that $\widetilde{\Psi}\mathcal{G}(\theta,\cdot)(\nu) = \sum_{j\in S} \mathcal{L}_j \mathcal{G}((\theta_{<j},\cdot,\theta_{>j}),\nu)(\theta_j) = \mathcal{L}\mathcal{G}(\cdot,\nu)(\theta)$.

To see Eq. (4.48), notice that, if $j \neq \max\{j' \preccurlyeq \ell : \theta_{j'} \neq \Delta\})$ (in particular, this is the case if $\theta_\ell \neq \Delta$), then $\mathcal{G}_{j,\ell}(\theta,\nu)$ is of the form

$$A \boxtimes \varphi_{\theta_j}(\nu)^{D_j} \boxtimes \varphi_0(\nu)^{D_\ell} \boxtimes \varphi_{\theta_{j'}}(\nu)^{D_{j'}} \boxtimes B' \tag{4.49}$$

for some $j' \preccurlyeq \ell$ due to the site ordering (compare Remark 4.17), where $B = \varphi_{\theta_{j'}} \boxtimes B'$. Since $j' \preccurlyeq \ell$ means $D_\ell \subseteq D_{j'}$, (4.49) is equal to

$$A \boxtimes \varphi_{\theta_j}(\nu)^{D_j} \boxtimes \varphi_{\theta_{j'}}(\nu)^{D_{j'}} \boxtimes B' = \mathcal{G}(\theta,\nu)$$

by the cancellation rule from Proposition 4.7. If $j = \max\{j' \preccurlyeq \ell : \theta_{j'} \neq \Delta\}$, the factors in (4.47) are ordered strictly non-decreasingly w.r.t. $\preccurlyeq$, and no cancellations occur; hence we have $\mathcal{G}_{j,\ell}(\theta,\nu) = \mathcal{G}((\theta_{<\ell},0,\theta_{>\ell}),\nu)$. Thus, we have verified (4.48). □

Remark 4.2. A few comments are in order.

(i) Another approach to recover Theorem 4.22 would be to prove the right multiplicativity for $h(m,\cdot)$ for $m \geqslant 1$ by the same argument as in Lemma 4.8, and to replace φ_t by $h(m,\cdot)$ in the proof of Theorem 4.22.

(ii) Note that nowhere in the proof of Theorem 4.22 have we used the particular form of the selection term; the only property required was the second statement in Lemma 4.8. Therefore, the same procedure can be applied to any single-locus model with linked neutral sites. Examples include the deterministic mutation-selection equation, for which the dual process can then be expressed as a collection of independent *pruned lookdown ASGs* [BCH18; BW18] that are initiated and reset at random.

(iii) It is also instructive to pause and relate the proof of Theorem 4.22 to the genealogical construction detailed above; see Figure 4.14. Recall that the factors $\varphi_{\theta_j}^{D_j}$ in $\mathcal{G}(\theta,\nu)$ correspond to the different independent ASGs that make up the essential ASRG of Section 4.5, and which are ancestral to different sets of sites. At rate ϱ_ℓ, $\ell \in S^*$, each such ASG is hit independently by a recombination bar labelled ℓ, at which a new ASG is started for the tail. This corresponds to right multiplication of $\varphi_{t_j}^{D_j}$ by $\varphi_{t_\ell}^{D_\ell}$. Recall

that in the case of such a multiplication, we had to distinguish the three cases of j being either incomparable to ℓ, $\ell \preccurlyeq j$ and $\ell \succ j$. In the genealogical picture, these cases correspond to the recombination event being either ignored (if ℓ and j are incomparable, which entails that the ASG in question is only ancestral to sites in C_ℓ); a resetting event if $\ell \preccurlyeq j$, which means that the ASG is only ancestral to sites contained in D_ℓ; or an initiation event if $\ell \succ j$, where a new ASG is initiated for the tail. ♢

By Corollary 4.21 and (4.40), Theorem 4.22 also yields the duality of ω and M.

Corollary 4.23. *The family M of* YPIRs *and the solution ω of the* SRE (4.9) *are dual with respect to $\mathcal{H}$ of* (4.32)*, namely*

$$\mathbb{E}[\mathcal{H}(M_t, \nu) \mid M_0 = m] = \mathbb{E}[\mathcal{H}(m, \omega_t) \mid \omega_0 = \nu] = \mathcal{H}(m, \psi_t(\nu)) \tag{4.50}$$

for all $\nu \in \mathcal{P}(X)$ and all initial values $m \in \mathbb{N}_0^S$ with $m_{i_*} > 0$. Here, ψ is the deterministic flow introduced in Definition 4.4. □

The following representations analogous to (4.29) for the solution of the selection-recombination differential equation are now immediate.

Corollary 4.24. *Let $\omega = \psi(\omega_0)$ be the solution of the* SRE (4.9)*. Then, for all $t \geqslant 0$, we have the stochastic representations*

$$\omega_t = \mathbb{E}\Big[\mathcal{H}(M_t, \omega_0) \mid M_{i,0} = \delta(i, i_*) \text{ for } i \in S\Big] = \mathbb{E}\Big[\mathcal{G}(\Theta_t, \omega_0) \mid \Theta_{i_*,0} = 0, \Theta_{i,0} = \Delta \text{ for } i \in S^*\Big]$$

with $\mathcal{H}$ of (4.32) *and $\mathcal{G}$ of* (4.41)*. That is, we average over all realisations of the* WPP *starting from the trivial partition with weight one as represented by the family of* YPIRs*, or the family of initiation processes, started in 0 for $i = i_*$ and started in Δ for $i \in S^*$.* □

4.8 The explicit solution and its long-term behaviour

We have seen in the previous section that the solution of the SRE (forward in time) has a stochastic representation in terms of a collection of independent Yule processes with initiation and resetting. Their semigroups are easily expressed in terms of geometric distributions with random success probability.

Proposition 4.25. *Let $i \in S$ and let M_i be a* YPIR *with branching rate $s > 0$, initiation rate $\varrho_i \geqslant 0$ and resetting rate $r_i \geqslant 0$. If $r_i > 0$, let T_i be a random variable with distribution* $\mathrm{Exp}(r_i)$*; if $r_i = 0$, set $T_i := \infty$ for consistency. The Markov transition semigroup $p_i = (p_{i,t})_{t \geqslant 0}$*

corresponding to M_i is then given by

$$\begin{aligned} p_{i,t}(1,\cdot) &= \mathbb{E}\big[\operatorname{Geom}(\mathrm{e}^{-s(T_i\wedge t)})\big], \\ p_{i,t}(0,n_i) &= \int_0^\infty \varrho_i \mathrm{e}^{-\varrho_i\tau} p_{i,t-\tau}(1,n_i)\,\mathrm{d}\tau + \delta_{0,n_i}\mathrm{e}^{-\varrho_i t}, \\ p_{i,t}(m_i,n_i) &= \int_0^\infty r_i \mathrm{e}^{-r_i\tau} p_{i,t-\tau}(1,n_i)\,\mathrm{d}\tau + \mathrm{e}^{-r_i t}\operatorname{NegBin}(m_i,\mathrm{e}^{-st})(n_i), \quad m_i \geqslant 1, \end{aligned}$$

where $\operatorname{NegBin}(m_i,\sigma)$ *is the negative binomial distribution with parameters* m_i *and* σ*, and we set* $p_{i,t}(1,\cdot) \equiv 0$ *for* $t < 0$*.*

Proof. For the first formula, we argue as in the genealogical proof of Theorem 4.6. After the time of the last resetting event, which is exponentially distributed with parameter r_i, the YPIR experiences no further resetting and hence has the law of a Yule process with branching rate s for the remaining time. The second and third formulae follow from the first by waiting the $\operatorname{Exp}(\varrho_i)$ ($\operatorname{Exp}(r_i)$)-distributed time until the process initiates (resets); recall that $\operatorname{NegBin}(m_i,\sigma)$ is the distribution of the number of independent Bernoulli trials (with the success probability σ) up to and including the m_i-th success. In the degenerate case (for $i = i_*$), $r_i = 0$ and $\varrho_i = 0$, the statement reduces to

$$p_{i,t}(m_i,n_i) = \operatorname{NegBin}(m_i,\mathrm{e}^{-st})(n_i), \quad m \geqslant 1,\ \ p_{i,t}(0,n_i) = \delta(0,n_i)$$

which is just the semigroup of the ordinary Yule process. The consistency in the cases where only one of the parameters ϱ_i or r_i vanishes is seen just as easily. □

Combining Proposition 4.25 with Corollary 4.24 yields a closed expression for the solution ω of the SRE.

Corollary 4.26. *The solution of the* SRE *is given by*

$$\omega_t = p_{i_*,t}h(\cdot,\omega_0)^{D_{i_*}}(1) \boxtimes \bigboxtimes_{i\in S^*} p_{i,t}h(\cdot,\omega_0)^{D_i}(0),$$

where $p_i = (p_{i,t})_{t\geqslant 0}$ *is the semigroup of* M_i *as in Proposition* 4.25*.*

We now turn our attention to the long-term behaviour of the solution. We do so by using its explicit representation in Corollary 4.26. The obvious first step is to consider the asymptotics of the semigroup from Proposition 4.25.

Corollary 4.27. *As in Proposition* 4.25*, let* p_i *be the Markov semigroup of the* YPIR *at site* i *with branching rate* $s > 0$*, initiation rate* ϱ_i*, and resetting rate* $r_i > 0$*. If* $\varrho_i > 0$*, then, for all* $m_i \geqslant 0$*,*

$$\lim_{t\to\infty} p_{i,t}(m_i,\cdot) = \mathbb{E}\big[\operatorname{Geom}(\mathrm{e}^{-sT_i})\big] =: \zeta_i, \tag{4.51}$$

where T_i *follows* $\operatorname{Exp}(r_i)$*. If* $\varrho_i = 0$*, then* $p_{i,t}(0,n_i) = \delta(0,n_i)$*, and Eq.* (4.51) *applies for*

$m_i > 0$. *More explicitly,* ζ_i *in* (4.51) *is given by*

$$\zeta_i(n_i) = \int_0^\infty r_i \mathrm{e}^{-(sn_i+r_i)t}(\mathrm{e}^{st}-1)^{n_i-1}\,\mathrm{d}t. \tag{4.52}$$

Proof. Since a YPIR with $r_i, s > 0$ is irreducible, positive recurrent, and non-explosive (since it is stochastically dominated by a Yule processes with branching rate s, which is non-explosive), there exists a unique asymptotic distribution ζ_i such that $p_{i,t}(m_i, \cdot)$ converges to ζ_i for all initial conditions $m_i > 0$. To see that in fact $\zeta_i = \mathrm{Geom}(\mathrm{e}^{-sT})$, it suffices in the case that $\varrho_i > 0$ to simply let $t \to \infty$ in $p_{i,t}(1, \cdot)$ in Proposition 4.25; note that even when starting in 0, the process will jump to one almost surely. This is not the case if $\varrho_i = 0$; in this case, the process started in 0 will stay there for all times whence the convergence to $\mathrm{Geom}(\mathrm{e}^{-sT})$ then only holds for strictly positive m. □

Remark 4.19. In the degenerate case $\varrho_i = r_i = 0$ (where the YPIR degenerates to an ordinary Yule process), there is no stationary distribution as the Yule process is transitive; the number of lines diverges almost surely. Nonetheless, one may still define (somewhat informally) $\zeta_i(n_i) := 0$ for all $n_i \in \mathbb{N}$ together with $\zeta_i(\infty) = 1$. ♢

If $r_i > 0$, substituting t for $\tau := r_i t$ in Eq. (4.52) gives

$$\zeta_i(n_i) = \int_0^\infty \mathrm{e}^{-(1+\frac{sn_i}{r_i})\tau}(\mathrm{e}^{\frac{s}{r_i}\tau}-1)^{n_i-1}\,\mathrm{d}\tau; \tag{4.53}$$

the long-term behaviour thus depends only on the ratio s/r_i. In particular, (4.53) yields $\zeta_i(1) = r_i/(s + r_i)$. For $r_i \gg s$, therefore, ζ_i is close to a point measure on 1; whereas for $r_i \ll s$, ζ_i puts substantial mass on large values, in line with intuition.

From the representation of the solution in Corollaries 4.25–4.27 together with Eq. (4.37) and $(1-x)^\infty = \delta_{x,0}$ for $x \in [0,1]$, the long-term behaviour of the solution is now immediate.

Corollary 4.28. *Assuming that* $\varrho_i > 0$ *for all* $i \in S^*$*, we have*

$$\omega_\infty := \lim_{t\to\infty} \omega_t = \bigotimes_{i\in S} \pi_i.\Big((1-\gamma_i(1-f(\omega_0)))b(\omega_0) + \gamma_i(1-f(\omega_0))d(\omega_0)\Big)$$

for all initial conditions $\omega_0 \in \mathcal{P}(X)$*. As always,* $f(\omega_0)$ *is the initial frequency of the beneficial type. Furthermore,* $\gamma_{i_*}(x) = \delta_{x,0}$ *(in line with Remark* 4.19 *and* $(1-x)^\infty = \delta_{x,0}$*), and for* $i \in S^*$*,* γ_i *is the probability generating function*

$$\gamma_i(x) := \sum_{n=1}^\infty \zeta_i(n)x^n$$

of ζ_i*.*

Remark 4.20. From Corollary 4.28, it is clear that $\gamma_i(1-f(\omega_0))$ is the probability that site i is drawn from $\pi_i.d(\omega_0)$, or equivalently, that it is associated with $i_* = 1$ at equilibrium;

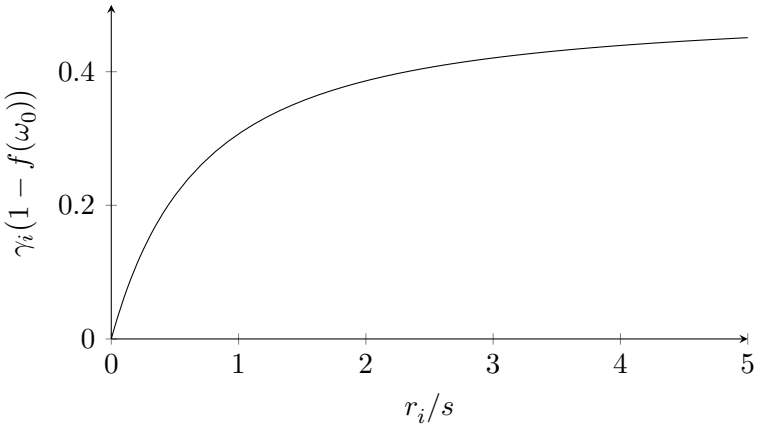


Figure 4.15. Asymptotic probability of site i being drawn from $\pi_i.d(\omega_0)$ as a function of r_i/s. As recombination becomes stronger, the asymptotic probability approaches the initial probability $1 - f(\omega_0) = 1/2$ assumed here.

see Figure 4.15 for an illustration of its parameter dependence. For a site i that is far away from i_* in the sense that its total rate of separation from i_* is large in comparison to the selection strength ($s \ll r_i$), the dynamics is close to that of the pure recombination equation; in particular, the marginals $\pi_i.\omega_t$ are approximately time invariant, in line with the marginalisation consistency (4.19) of the pure recombination equation. Accordingly, the long-term behaviour is governed by $\gamma_i(x) \approx x$. In contrast, in the regime $s \gg r_i$, the behaviour is closer to that of the pure selection equation, in that ζ places much weight on large values, which implies that $\gamma_i(x)$ is very small for small values of x, and the beneficial type prevails. $\diamondsuit$

4.8.1 The evolution of linkage disequilibria during selective sweeps

We close by showing how our results can explain the effect of a selective sweep on the correlation between two neutral sites. A *selective sweep* [SH74] occurs when a new beneficial mutant is introduced into the population and thus also increases the frequency of the letters of the neutral sites of that mutant; these neutral letters thus *hitchhike along* with the beneficial mutation at the selected site. We assume the simplest scenario of two neutral sites L and R that are linked to the single selected site i_*. Following [SSL06], we therefore take $S = \{i_*, L, R\}$, where $i_* \in \{1, 2, 3\}$ is given and $L, R \in S \setminus i_*$ satisfy $L < R$; L and R denote the 'left' and the 'right' neutral site, respectively, see Figure 4.16. We then consider

$$\mathrm{Cor}(\omega_t) := (\pi_{\{L,R\}}.\omega_t)\{(1,1)\} - (\pi_L.\omega_t)\{(1)\}(\pi_R.\omega_t)\{(1)\}; \tag{4.54}$$

due to marginalisation consistency as discussed in Section 4.3, the results are not affected by adding additional neutral sites to S. We will examine how the dynamics of the correlation is affected by the location of the selected site relative to the neutral ones. Indeed, a somewhat

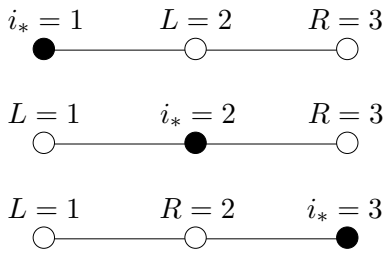


Figure 4.16. The three cases $i_* = 1, i_* = 2$, and $i_* = 3$. The selected site is represented by a bullet, the other two (neutral) sites by circles.

complicated behaviour was observed in Figure 2 of [SSL06] but remained somewhat obscure. A partial explanation was given in [PLS08], which we will complement here.

We are interested in a single, rare beneficial mutation that is introduced into a population that otherwise consists exclusively of unfit individuals. To model this, we pick a single type $x_\mathrm{m} \in \{x \in X : x_{i_*} = 0\}$ and set $\omega_0(\{x_\mathrm{m}\}) := \varepsilon$ (where ε is a small positive number), together with $\omega_0(\{x\}) := 0$ for all $x \in \{x \in X : x \neq x_\mathrm{m}, x_{i_*} = 0\}$. For our numerical solutions, we specifically chose $x_{\mathrm{m},L} = x_{\mathrm{m},R} = 1$ and adjusted the remaining type frequencies such that

- $\mathrm{Cor}(\omega_0) > 0$, and
- for $\varrho_L = \varrho_R = 0$, one has $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Cor}(\omega_t)|_{t=0} > 0$ (in line with hitchhiking of $x_R = 1$ and $x_L = 1$ along with $x_{i_*} = 0$ in the mutant).

For our exact parameter values, see Fig. 4.17.

It is clear that, for $\varrho_L = \varrho_R = 0$, the correlation eventually decays to zero. This is because $\omega_\infty = \delta_{x_\mathrm{m}}$, and the correlation vanishes for any point measure. Let us now investigate how this behaviour changes in the presence of recombination. Here, it is essential to distinguish between recombination events that separate L and R (*separating* recombination) (compare Fig. 4.17 (b)) and those that do not (compare Fig. 4.17 (a)). We denote the set of all sites j such that recombination at site j separates L and R by S_sep, that is,

$$S_\mathrm{sep} = \{j \in S : L \in C_j \text{ and } R \in D_j \text{ or } L \in D_j \text{ and } R \in C_j\}.$$

Likewise, the set of sites j such that recombination at site j separates $\{L, R\}$ from i_* but not from each other is denoted by

$$S_\mathrm{ns} = \{j \in S : \{L, R\} \subseteq C_j \text{ and } i_* \in D_j \text{ or } \{L, R\} \subseteq D_j \text{ and } i_* \in C_j\}.$$

We define $\varrho_\mathrm{sep} := \sum_{j \in S_\mathrm{sep}} \varrho_j$ and $\varrho_\mathrm{ns} := \sum_{j \in S_\mathrm{ns}} \varrho_j$; in other words, ϱ_sep and ϱ_ns are the marginal recombination rates $\varrho^{\{L,R\}}_{\{L\},\{R\}}$ and $\varrho^{\{L,R,i_*\}}_{\{L,R\},\{i_*\}}$. More explicitly, we have for $i_* = 1$

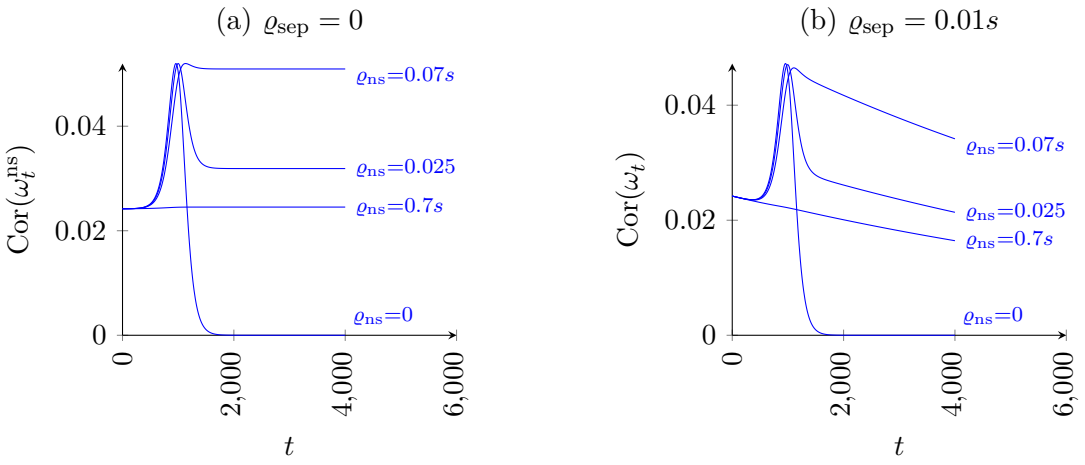


Figure 4.17. Time evolution of the correlation under recombination and selection obtained by evaluating the solution formula from Theorem 4.6. In the left panel, recombination only separates the block $\{L, R\}$ from the selected site, but not L and R from each other. In the right panel, separating recombination is added. The parameters are chosen as follows. $s = 10^{-2}$, Initial type distribution: $\omega_0(\{(0,1,1)\}) = 5 \cdot 10^{-5} = \varepsilon, \omega_0(\{(1,1,1)\}) = 0.38995, \omega_0(\{(1,0,1)\}) = 0.23, \omega_0(\{(1,1,0)\}) = 0.2$ and $\omega_0(\{(1,0,0)\}) = 0.18$.

that $\varrho_{\text{sep}} = \varrho_3$ and $\varrho_{\text{ns}} = \varrho_2$. If the selected site is in the middle, i.e. $i_* = 2$, then we have $\varrho_{\text{sep}} = \varrho_1 + \varrho_3$ and $\varrho_{\text{ns}} = 0$. Finally, if $i_* = 3$, then $\varrho_{\text{sep}} = \varrho_1$ and $\varrho_{\text{ns}} = \varrho_3$.

First, let us consider the effect of recombination separating L and R, as this is somewhat easier to understand.

Theorem 4.29. *Let ω^{ns} be the solution of the selection-recombination equation (4.9) with all ϱ_j with $j \in S_{\text{sep}}$ set to 0. Then, we have*

$$\text{Cor}(\omega_t^{ns}) = \mathrm{e}^{-\varrho_{sep} t} \text{Cor}(\omega_t).$$

Proof. This follows by an iterative application of Lemma 4.9. Note that in all three cases, the labelling in Section 4.4 is such that $S_{\text{sep}} = \{i_1, i_2\}$ (if i_* is in the middle) or $S_{\text{sep}} = \{i_2\}$. □

Next, we examine the effect of recombination at sites in S_{ns}. In the forward-time evolution, they have the effect that the subsequences $(x_2, x_3) = (0,0), (0,1), (1,0)$ partially replace the original tail $(x_2, x_3) = (1,1)$ in the mutant and thwart its establishment as it is 'swept' through the population together with $x_1 = 0$. In the absence of separating recombination, this preserves some of the correlation built up initially. In particular, $\omega_\infty^{\text{ns}}$ is not a point measure. We can compute the limit of the correlation explicitly.

Theorem 4.30. *The limit of* $\mathrm{Cor}(\omega_t^{ns})$ *as* $t \to \infty$ *is given by*

$$\mathrm{Cor}(\omega_\infty^{ns}) = \int_0^\infty \varrho_{ns} \mathrm{e}^{-\tau \varrho_{ns}} (\pi_{\{L,R\}}.\omega_\tau^{(0)})\{(1,1)\}\,\mathrm{d}\tau \\ - \int_0^\infty \varrho_{ns} \mathrm{e}^{-\tau \varrho_{ns}} (\pi_L.\omega_\tau^{(0)})\{(1)\}\,\mathrm{d}\tau \int_0^\infty \varrho_{ns} \mathrm{e}^{-\tau \varrho_{ns}} (\pi_R.\omega_\tau^{(0)})\{(1)\}\,\mathrm{d}\tau.$$

Proof. If i_* is in the middle, then $\varrho_{\mathrm{ns}} = 0$ and the right-hand side vanishes, in line with our earlier observations. Otherwise, we have $S_{\mathrm{ns}} = \{i_1\}$ and the statement follows by letting $t \to \infty$ in Theorem 4.6. □

For an illustration, see Fig. 4.17. In (a), we plotted $\mathrm{Cor}(\omega_t^{\mathrm{ns}})$. Note that this behaviour is only possible if the selected site is not in the middle, and does not seem to have been described previously. In (b), the situation is as in [PLS08].

5 The migration-recombination equation and the labelled partitioning process

Last but not least, we consider the evolution under the joint action of recombination and *migration* of individuals between discrete locations (or *demes*). As before, the model will be deterministic. In contrast to previous chapters, which were set in continuous time, we now focus mainly on discrete time, where generations do not overlap.

This dynamical system is a variant of the migration-selection-recombination equation formulated by Bürger [Bür09] in 2009, who analysed its asymptotic behaviour in the classical dynamical systems setting forward in time. It is our goal to complement this picture by relating this nonlinear system to a linear one by embedding the solution into a space of higher dimension, a technique known as *Haldane linearisation* [MR83; Lyu92] in the context of genetic algebras. This extends the approach taken in [BB16] to the case with migration. The resulting linear system has a natural interpretation as a Markov chain on the set of *labelled* partitions of the set of sequence sites. Intuitively, this Markov chain describes how the genetic material of an individual from the current population is partitioned across an increasing number of ancestors, along with their locations, as the lines of descent are traced back into the past. This backward (or *dual*) process combines a variant of the *ancestral recombination graph* with a variant of the *ancestral migration graph* [Not90; MW06]. It is tractable in the law-of-large-numbers regime considered here; this was previously exploited for the recombination equation (without migration) in [BBS16; Mar17; BB16]; see [BB] for a review. For an application of a similar idea in the context of the ancestral selection graph, see [SW05].

All this leads to a stochastic representation of the solution of the (nonlinear, deterministic) migration-recombination equation in terms of the labelled partitioning process. As a consequence, one obtains an explicit solution of the nonlinear dynamics, simply in terms of powers of the transition matrix of the Markov chain. In particular, the asymptotic behaviour of the recombination-migration equation emerges without any additional effort, via the (unique) absorbing state of the Markov chain. In addition, we will investigate the quasi-limiting behaviour of the labelled partitioning process, based on ideas from [Mar17].

This chapter is organised as follows. In Section 5.1, we set the scene and introduce the model. In Section 5.2, we adapt the notion of *recombinators* to the setting of labelled partitions, and reformulate the model in a compact way. The marginalisation consistency (cf. Theorem 2.5) is established in Section 5.3. The core of the chapter is Section 5.4, where we solve the forward iteration, together with Section 5.5, which establishes the connection to the labelled partitioning process in terms of a duality, together with a genealogical interpretation. Section 5.6 is devoted to the asymptotic properties, namely the limiting and quasi-limiting behaviour,

and Section 5.7 sketches how the approach carries over to continuous time.

5.1 The migration-recombination model

As in previous chapters, we want to model the time evolution of the distribution of the genetic type within a large population. In order to discuss migration, let us fix a geographical structure in the form of a finite set L of discrete *locations* (or *demes*). Then, a type distribution will be given not by one probability measure, but by a *vector* $\mu \in \mathcal{P}(X)^L$ of probability measures $\mu(\alpha)$, one to describe the local type distribution at each location $\alpha \in L$. For any subset $U \subseteq S$, the vector of marginal distributions $\mu^U(\alpha)$ is denoted by μ^U.

We assume that, in each generation, the global type distribution evolves in two stages. First, individuals migrate between locations; then, random mating takes place among individuals at the same location, followed by reproduction involving recombination. Discrete generations will be indexed by $t \in \mathbb{N}_0$, where a population at time t is understood as the population after the t-th round of mating and recombination, but before migration; we will use the corresponding half integers $t + \frac{1}{2}$ to indicate the population after migration, but before mating.

5.1.1 Describing migration

We first consider migration, following the presentation in [Nag92, Ch. 6.2]. The most obvious way to describe migration is via the so-called *forward migration matrix* $\widetilde{M}$. It is a stochastic matrix indexed by L, where the entry $\widetilde{M}(\alpha, \beta)$ is the probability that a randomly chosen individual at location α will migrate to location β in the next generation. However, it is more convenient to work instead with the *backward migration matrix* M. It is also a stochastic matrix, and $M(\alpha, \beta)$ is the probability that a randomly chosen individual that currently lives at location α has migrated from location β. We assume, mainly for the sake of technical convenience (see Remark 5.1 below), that the local population sizes $c(\alpha) \in \mathbb{R}_{>0}$ remain constant over time. This is the case if either

$$c(\alpha) = \sum_{\beta \in L} c(\beta) \widetilde{M}(\beta, \alpha) \tag{5.1}$$

for all $\alpha \in L$, or if population regulation takes place after the migration step. In any case, denoting the location of a randomly sampled individual at time $t + \frac{1}{2}$ by $\ell_{t+\frac{1}{2}}$ and its location in generation t by ℓ_t, we have

$$\begin{aligned} M(\alpha, \beta) = \mathbb{P}\big(\ell_t = \beta \mid \ell_{t+\frac{1}{2}} = \alpha\big) &= \frac{\mathbb{P}\big(\ell_t = \beta, \ell_{t+\frac{1}{2}} = \alpha\big)}{\mathbb{P}\big(\ell_{t+\frac{1}{2}} = \alpha\big)} \\ &= \frac{\mathbb{P}\big(\ell_{t+\frac{1}{2}} = \alpha \mid \ell_t = \beta\big)\mathbb{P}\big(\ell_t = \beta\big)}{\mathbb{P}\big(\ell_{t+\frac{1}{2}} = \alpha\big)} = \frac{c(\beta)}{c(\alpha)} \widetilde{M}(\beta, \alpha). \end{aligned}$$

Note that M is stochastic by definition, i.e. $\sum_{\beta\in L} M(\alpha,\beta) = 1$ and $M(\alpha,\beta) \geqslant 0$ for all $\alpha,\beta \in L$.

Remark 5.1. Dropping the assumption of constant population sizes (but still assuming constant forward migration rates) would result in a time-dependent backward migration matrix. As a consequence, the Markov chain discussed in Section 5.5 would be non-homogeneous. However, the considerations in Section 5.6 about the limiting and quasi-limiting behaviour remain valid if we assume a primitive forward migration matrix; as the population sizes converge to a unique equilibrium, the (time-dependent) backward migration matrix also stabilises asymptotically. ♢

For additional background, see [Nag92, Ch. 6.2]. In what follows, we will work exclusively with the backward migration matrix. Since we are only interested in relative type frequencies, the population sizes $c(\alpha)$ are irrelevant. After migration (but before recombination), the local population at α is therefore given by

$$\mu_{t+\frac{1}{2}}(\alpha) = \sum_{\beta\in L} M(\alpha,\beta)\mu_t(\beta), \tag{5.2}$$

and the metapopulation may be written compactly as

$$\mu_{t+\frac{1}{2}} = M\mu_t; \tag{5.3}$$

again, we interpret μ as a column vector.

5.1.2 Describing recombination

To describe recombination, recall the discrete recombination equation from Chapter 2 (cf. Eq. (2.4)) and the preceding discussion. Two things are different. First, we do not consider only one probability measure, but rather $|L|$ of them, one for each location.

Secondly, random mating occurs after migration, so we have to replace μ_t on the right-hand side of Eq. (2.4) by $\mu_{t+\frac{1}{2}}$.

Thus, the *migration-recombination equation* reads

$$\mu_{t+1}(\alpha) = \sum_{\mathcal{A}\in\boldsymbol{P}(S)} r_{\mathcal{A}} \cdot \bigotimes_{A\in\mathcal{A}} \mu^A_{t+\frac{1}{2}}(\alpha) = \sum_{\mathcal{A}\in\boldsymbol{P}(S)} r_{\mathcal{A}} \cdot \bigotimes_{A\in\mathcal{A}} \sum_{\beta\in L} M(\alpha,\beta)\mu^A_t(\beta), \tag{5.4}$$

where we have used (5.2) and the linearity of marginalisation in the last step.

Remark 5.2. A closely related migration-recombination equation was already considered by Bürger [Bür09]. It differs from our model in two ways. First, Bürger's model describes the evolution of diploid organisms, wheras our model considers the population at the gamete level. The justification is that, in the absence of selection, diploid genotypes are independent combinations of haploid gametes at all stages of the life cycle, that is, one has Hardy–Weinberg equilibrium throughout. Secondly, Bürger's model only allows for mating of at most two parents; while this is sufficient for most applications, the generalisation to recombination patterns involving more than two parents is interesting from a mathematical perspective and, using our methods, does not require additional effort. ♢

5.2 Reformulation of the model

Extending concepts established in [BBS16; Mar17; BB16], we now reformulate the MRE (5.4) in a more compact way. In particular, this involves labelling the blocks of a partition by elements of L to keep track of the location of the ancestors of the various blocks.

Definition 5.1. A labelled partition of $U \subseteq S$ is a collection $\boldsymbol{\mathcal{A}} := \{\boldsymbol{A}_1, \dots, \boldsymbol{A}_m\}$ for some $m \leqslant |U|$, where $\boldsymbol{A}_i = (A_i, \lambda_i)$, $\mathcal{A} = \{A_1, \dots, A_m\}$ is a partition of U, and $\lambda_i \in L$ for $1 \leqslant i \leqslant m$. We call $\mathcal{A}$ the *base* of $\boldsymbol{\mathcal{A}}$, refer to its elements as the blocks of $\boldsymbol{\mathcal{A}}$, and interpret λ_i as the *label* of block A_i. We write $\boldsymbol{LP}(U)$ for the set of all labelled partitions of U. ♢

In order to rewrite Eq. (5.4), we now introduce the *labelled recombinator*. It is the labelled analogue of the recombinator defined in (2.3) for unlabelled partitions. Since we will later also be interested in the evolution of the distribution of subsequences (cf. Section 5.3), we introduce the concept in the required generality right away.

Definition 5.2. Let $U \subseteq S$ and $\boldsymbol{\mathcal{A}} \in \boldsymbol{LP}(U)$. Then, the labelled recombinator (with respect to $\boldsymbol{\mathcal{A}}$), namely $\mathcal{R}^U_{\boldsymbol{\mathcal{A}}} : \mathcal{P}(X_U)^L \to \mathcal{P}(X_U)$, is defined by

$$\mathcal{R}^U_{\boldsymbol{\mathcal{A}}}(\nu) := \bigotimes_{(A,\lambda)\in\boldsymbol{\mathcal{A}}} \nu^A(\lambda);$$

if $U = S$, we will drop the superscript and write $\mathcal{R}_{\boldsymbol{\mathcal{A}}}$ instead of $\mathcal{R}^S_{\boldsymbol{\mathcal{A}}}$. ♢

In words, $\mathcal{R}_{\boldsymbol{\mathcal{A}}}(\nu)$ is the distribution of the type of an offspring individual that is recombined according to $\mathcal{A}$; the parent of the labelled block (A, λ) is sampled from the local population $\nu(\lambda)$. A similar interpretation holds for the *marginal* recombinators; see Theorem 5.5 and

Remark 5.5.

With this, we can now restate Eq. (5.4).

Lemma 5.3. *The* MRE *can be written as*

$$\mu_{t+1} = \sum_{\boldsymbol{\mathcal{A}} \in \boldsymbol{LP}(S)} p_{\boldsymbol{\mathcal{A}}} \mathcal{R}_{\boldsymbol{\mathcal{A}}}(\mu_t) \tag{5.5}$$

with

$$p_{\boldsymbol{\mathcal{A}}} := \left(p_{\boldsymbol{\mathcal{A}}}(\alpha)\right)_{\alpha \in L}$$

and the migration-recombination probabilities

$$p_{\boldsymbol{\mathcal{A}}}(\alpha) := r_{\mathcal{A}} \prod_{(A,\lambda) \in \boldsymbol{\mathcal{A}}} M(\alpha, \lambda).$$

Furthermore, for all $\alpha \in L$,

$$\sum_{\boldsymbol{\mathcal{A}} \in \boldsymbol{LP}(S)} p_{\boldsymbol{\mathcal{A}}}(\alpha) = 1. \tag{5.6}$$

Proof. This follows immediately from Definition 5.2 by expanding the measure product in Eq. (5.4):

$$\begin{aligned}\mu_{t+1}(\alpha) &= \sum_{\mathcal{A} \in \boldsymbol{P}(S)} r_{\mathcal{A}} \bigotimes_{A \in \mathcal{A}} \sum_{\lambda \in L} M(\alpha, \lambda) \mu_t^A(\lambda) = \sum_{\mathcal{A} \in \boldsymbol{P}(S)} \sum_{\boldsymbol{\lambda} \in L^{\mathcal{A}}} r_{\mathcal{A}} \prod_{A \in \mathcal{A}} M(\alpha, \lambda(A)) \bigotimes_{A \in \mathcal{A}} \mu_t^A(\lambda_A) \\ &= \sum_{\boldsymbol{\mathcal{A}} \in \boldsymbol{LP}(S)} p_{\boldsymbol{\mathcal{A}}}(\alpha) \bigotimes_{(A,\lambda) \in \boldsymbol{\mathcal{A}}} \mu_t^A(\lambda) = \sum_{\boldsymbol{\mathcal{A}} \in \boldsymbol{LP}(S)} p_{\boldsymbol{\mathcal{A}}}(\alpha) \mathcal{R}_{\boldsymbol{\mathcal{A}}}(\mu_t),\end{aligned}$$

where, in the third step, we identified the double sum over all partitions of S and all possible vectors of labels of their blocks with the sum over all labelled partitions. The normalisation in (5.6) is a consequence of $\sum_{\mathcal{A} \in \boldsymbol{P}(S)} r_{\mathcal{A}} = 1 = \sum_{\beta \in L} M(\alpha, \beta)$. □

We call the probability distribution $p(\alpha) = \left(p_{\boldsymbol{\mathcal{A}}}(\alpha)\right)_{\boldsymbol{\mathcal{A}} \in \boldsymbol{LP}(S)}$ the *migration-recombination distribution at* α.

Remark 5.3. Lemma 5.3 has a simple stochastic interpretation. To sample the type of an individual in generation $t+1$ (say at location α), we first pick a random labelled partition $\boldsymbol{\mathcal{A}}$ according to $p(\alpha)$ and subsequently sample from $\mathcal{R}_{\boldsymbol{\mathcal{A}}}(\mu_t)$. The intuition behind the formula for $p_{\boldsymbol{\mathcal{A}}}(\alpha)$ in Theorem 5.3 is that the genome is first partitioned across its parents according to $\mathcal{A}$, with probability $r_{\mathcal{A}}$. Subsequently, the labels are reassigned, (conditionally) independently for each block, according to $M(\alpha, \cdot)$ as we trace back the origin of each ancestor. Finally, the offspring type is determined by piecing together (fragments of) independent samples of the ancestral sequences at the appropriate locations, in generation t. This leads to the product measure in Definition 5.2. We will further elaborate on this in Section 5.5. ♢

To continue, we need a few additional concepts around labelled partitions. First, the notion

of an *induced (labelled) partition* is required. For $\varnothing \neq V \subseteq U$ and $\boldsymbol{\mathcal{A}} \in \boldsymbol{LP}(U)$, we denote by $\boldsymbol{\mathcal{A}}|_V$ the labelled partition of V induced by $\boldsymbol{\mathcal{A}}$; it is given by

$$\boldsymbol{\mathcal{A}}|_V := \{(A \cap V, \lambda) : A \cap V \neq \varnothing, (A, \lambda) \in \boldsymbol{\mathcal{A}}\}$$

with base $\mathcal{A}|_V$ as in Chapter 2. Put simply, every block inherits the label of the unique block of the original partition that contains it.

Conversely, given a partition $\mathcal{A}$ of U and a family $(\boldsymbol{\mathcal{B}}_A)_{A\in\mathcal{A}}$ of labelled partitions of its blocks, their union

$$\bigcup_{A\in\mathcal{A}} \boldsymbol{\mathcal{B}}_A$$

is a labelled partition of U; its base is the union

$$\bigcup_{A\in\mathcal{A}} \mathcal{B}_A$$

of the bases $\mathcal{B}_A$.

Finally, given two labelled partitions $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$, we say that $\boldsymbol{\mathcal{B}}$ *is finer than* $\boldsymbol{\mathcal{A}}$ ($\boldsymbol{\mathcal{B}} \preccurlyeq \boldsymbol{\mathcal{A}}$) if $\mathcal{B} \preccurlyeq \mathcal{A}$. The partial order on $\boldsymbol{P}(U)$ thus carries over to a partial order on $\boldsymbol{LP}(U)$. For any $\alpha \in L$, there is a unique maximal element; namely, the labelled partition $\underline{\mathbf{1}}_U^\alpha := \{(U, \alpha)\}$ that consists of a single block with label α. If $U = S$, we drop the subscript.

We have the following analogue to Remark 2.8.

Remark 5.4. It is not difficult to see that $\boldsymbol{\mathcal{B}} \preccurlyeq \boldsymbol{\mathcal{A}}$ if and only if

$$\boldsymbol{\mathcal{B}} = \bigcup_{A\in\mathcal{A}} \boldsymbol{\mathcal{B}}|_A.$$

For a fixed $\mathcal{A} \in \boldsymbol{P}(S)$, this implies the following bijection between the labelled partitions $\boldsymbol{\mathcal{B}}$ with $\mathcal{B} \preccurlyeq \mathcal{A}$ and collections $(\boldsymbol{\mathcal{B}}_A)_{A\in\mathcal{A}}$ of labelled partitions of the individual blocks of $\mathcal{A}$. Given $\boldsymbol{\mathcal{B}}$ with $\mathcal{B} \preccurlyeq \mathcal{A}$, we obtain the collection $(\boldsymbol{\mathcal{B}}|_A)_{A\in\mathcal{A}}$ of labelled partitions induced by $\boldsymbol{\mathcal{B}}$ on the blocks of $\mathcal{A}$. Conversely, given a collection $(\boldsymbol{\mathcal{B}}_A)_{A\in\mathcal{A}}$ of labelled partitions of the blocks of $\mathcal{A}$, we set $\boldsymbol{\mathcal{B}} := \bigcup_{A\in\mathcal{A}} \boldsymbol{\mathcal{B}}_A$; note that $\mathcal{B} \preccurlyeq \mathcal{A}$ and $\boldsymbol{\mathcal{B}}|_A = \boldsymbol{\mathcal{B}}_A$. See also Fig. 5.1. ♢

We will now see that the recombinator for a union of labelled partitions of disjoint subets is the product of the recombinators for the individual labelled partitions..

Lemma 5.4. *Let $\mathcal{A} \in \boldsymbol{P}(S)$ and $\boldsymbol{\mathcal{B}}_A \in \boldsymbol{LP}(A)$ for all $A \in \mathcal{A}$. Then, for all $\nu \in \mathcal{P}(X)^L$,*

$$\mathcal{R}_{\cup_{A\in\mathcal{A}}\boldsymbol{\mathcal{B}}_A}(\nu) = \bigotimes_{A\in\mathcal{A}} \mathcal{R}^A_{\boldsymbol{\mathcal{B}}_A}(\nu^A).$$

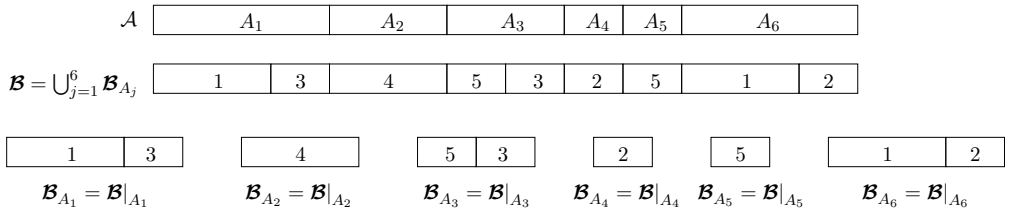


Figure 5.1. At the top, an unlabelled partition of S. In the middle, a labelled refinement of $\mathcal{A}$, which gives rise to labelled partitions of the blocks of $\mathcal{A}$ (bottom). Conversely, one can start with the collection of labelled partitions at the bottom and take their union to obtain a labelled refinement of $\mathcal{A}$.

In particular, for $\boldsymbol{\mathcal{B}} \in \boldsymbol{LP}(S)$ with $\mathcal{B} \preccurlyeq \mathcal{A}$, we have

$$\mathcal{R}_{\boldsymbol{\mathcal{B}}}(\nu) = \bigotimes_{A \in \mathcal{A}} \mathcal{R}^{A}_{\boldsymbol{\mathcal{B}}|_A}(\nu^A).$$

Proof. For the first claim, we write out the labelled recombinators and see that

$$\bigotimes_{A \in \mathcal{A}} \mathcal{R}^{A}_{\boldsymbol{\mathcal{B}}_A}(\nu^A) = \bigotimes_{A \in \mathcal{A}} \bigotimes_{(B,\lambda) \in \boldsymbol{\mathcal{B}}_A} \nu^B(\lambda) = \bigotimes_{(B,\lambda) \in \cup_{A \in \mathcal{A}} \boldsymbol{\mathcal{B}}_A} \nu^B(\lambda) = \mathcal{R}_{\cup_{A \in \mathcal{A}} \boldsymbol{\mathcal{B}}_A}(\nu).$$

For the second claim, see Remark 5.4. □

We now turn to the marginalisation consistency of the MRE, a property that will turn out as the key to its solution.

5.3 Marginalisation consistency

A crucial ingredient is the marginalisation consistency of the model, which was already discussed in Chapter 2 (cf. Theorem 2.5) in absence of migration. Now, we turn to the analogue in the case *with* migration.

Theorem 5.5. *Let $(\mu_t)_{t \in \mathbb{N}_0}$ be a solution of the* MRE (5.5) *and U a non-empty subset of S. Then, $(\mu^U_t)_{t \in \mathbb{N}_0}$ satisfies the* marginal MRE

$$\mu^U_{t+1} = \sum_{\boldsymbol{\mathcal{A}} \in \boldsymbol{LP}(U)} p^U_{\boldsymbol{\mathcal{A}}} \mathcal{R}^U_{\boldsymbol{\mathcal{A}}}(\mu^U_t),$$

where $p^U_{\boldsymbol{\mathcal{A}}}$ is given by

$$p^U_{\boldsymbol{\mathcal{A}}} := \sum_{\substack{\boldsymbol{\mathcal{B}} \in \boldsymbol{LP}(S) \\ \boldsymbol{\mathcal{B}}|_U = \boldsymbol{\mathcal{A}}}} p_{\boldsymbol{\mathcal{B}}} \text{ for } \boldsymbol{\mathcal{A}} \in \boldsymbol{LP}(U).$$

This sum is to be understood component-wise.

Proof. By Lemma 5.3 and the linearity of marginalisation, we have

$$\mu_{t+1}^U = \Big(\sum_{\boldsymbol{\mathcal{B}} \in \boldsymbol{LP}(S)} p_{\boldsymbol{\mathcal{B}}} \mathcal{R}_{\boldsymbol{\mathcal{B}}}(\mu_t) \Big)^U = \sum_{\boldsymbol{\mathcal{B}} \in \boldsymbol{LP}(S)} p_{\boldsymbol{\mathcal{B}}} \big(\mathcal{R}_{\boldsymbol{\mathcal{B}}}(\mu_t)\big)^U .$$

Using Lemma 2.1, we obtain for all $\boldsymbol{\mathcal{B}} \in \boldsymbol{LP}(S)$

$$\big(\mathcal{R}_{\boldsymbol{\mathcal{B}}}(\mu_t)\big)^U = \Big(\bigotimes_{(\tilde{A},\lambda)\in\boldsymbol{\mathcal{B}}} \mu_t^{\tilde{A}}(\lambda) \Big)^U = \bigotimes_{\substack{(B,\lambda)\in\boldsymbol{\mathcal{B}} \\ B\cap U \neq \varnothing}} \mu_t^{B\cap U}(\lambda) = \bigotimes_{(A,\lambda)\in\boldsymbol{\mathcal{B}}|_U} \mu_t^A(\lambda) = \mathcal{R}^U_{\boldsymbol{\mathcal{B}}|_U}(\mu_t^U),$$

where, in the second step, we ignored the factors corresponding to B with $B \cap U = \varnothing$ (cf. Remark 2.2). Thus,

$$\mu_{t+1}^U = \sum_{\boldsymbol{\mathcal{B}} \in \boldsymbol{LP}(S)} p_{\boldsymbol{\mathcal{B}}} \mathcal{R}^U_{\boldsymbol{\mathcal{B}}|_U}(\mu_t^U) = \sum_{\boldsymbol{\mathcal{A}} \in \boldsymbol{LP}(U)} p^U_{\boldsymbol{\mathcal{A}}} \mathcal{R}^U_{\boldsymbol{\mathcal{A}}}(\mu_t^U),$$

which is what we wanted to show. □

The $p^U_{\boldsymbol{\mathcal{A}}}(\alpha)$ are called *marginal migration-recombination probabilities*, and $p^U(\alpha) = \big(p^U_{\boldsymbol{\mathcal{A}}}(\alpha)\big)_{\boldsymbol{\mathcal{A}}\in\boldsymbol{LP}(U)}$ is called the marginal migration-recombination *distribution* (at location α). We will now see that the marginal migration-recombination probabilities have a product structure analogous to that of the migration-recombination probabilities in Lemma 5.3.

Lemma 5.6. *The marginal labelled recombination probabilities $p^U_{\boldsymbol{\mathcal{A}}}(\alpha)$ from Theorem 5.5 can be written as*

$$p^U_{\boldsymbol{\mathcal{A}}}(\alpha) = \Big(\sum_{\substack{\mathcal{B}\in\boldsymbol{P}(S) \\ \mathcal{B}|_U = \mathcal{A}}} r_{\mathcal{A}} \Big) \prod_{(A,\lambda)\in\boldsymbol{\mathcal{A}}} M(\alpha,\lambda).$$

Proof. We write the (given) labelled partition $\boldsymbol{\mathcal{A}}$ as

$$\boldsymbol{\mathcal{A}} = \{(A_1,\lambda_1),\dots,(A_k,\lambda_k)\}.$$

Next, we split the conditional sum over the labelled partitions into the sums over the appropriate partitions and their labels. Thus,

$$p^U_{\boldsymbol{\mathcal{A}}}(\alpha) = \sum_{\substack{\boldsymbol{\mathcal{B}}\in\boldsymbol{LP}(S) \\ \boldsymbol{\mathcal{B}}|_U = \boldsymbol{\mathcal{A}}}} p_{\boldsymbol{\mathcal{B}}}(\alpha) = \sum_{\substack{\mathcal{B}=\{B_1,\dots,B_m\}\in\boldsymbol{P}(S) \\ \{B_1,\dots,B_m\}|_U = \mathcal{A}}} r_{\mathcal{B}} \sum_{\tilde{\lambda}_1,\dots,\tilde{\lambda}_m \in L} \prod_{j=1}^{k} \mathbb{1}_{\tilde{\lambda}_j=\lambda_j} \prod_{j=1}^{m} M(\alpha,\tilde{\lambda}_j), \tag{5.7}$$

where (deviating from the usual convention) the indices are ordered such that $B_j \cap U = A_j$

for all $1 \leqslant j \leqslant k$ and $B_j \cap U = \varnothing$ for $k+1 \leqslant j \leqslant m$. Clearly,

$$\sum_{\tilde{\lambda}_1,\dots,\tilde{\lambda}_m \in L} \prod_{j=1}^{k} \mathbb{1}_{\tilde{\lambda}_j = \lambda_j} \prod_{j=1}^{m} M(\alpha, \tilde{\lambda}_j) \\ = \Big(\sum_{\tilde{\lambda}_1,\dots,\tilde{\lambda}_k} \prod_{j=1}^{k} \mathbb{1}_{\tilde{\lambda}_j = \lambda_j} \prod_{j=1}^{k} M(\alpha, \tilde{\lambda}_j) \Big) \Big(\sum_{\tilde{\lambda}_{k+1},\dots,\tilde{\lambda}_m} \prod_{j=k+1}^{m} M(\alpha, \tilde{\lambda}_j) \Big)$$

with the usual convention that the empty product is 1. Now, we can use the indicator in the first bracket to eliminate the summation, yielding

$$\sum_{\tilde{\lambda}_1,\dots,\tilde{\lambda}_k} \prod_{j=1}^{k} \mathbb{1}_{\tilde{\lambda}_j = \lambda_j} \prod_{j=1}^{k} M(\alpha, \tilde{\lambda}_j) = \prod_{j=1}^{k} M(\alpha, \lambda_j).$$

The second bracket is equal to one, by the stochasticity of M:

$$\sum_{\tilde{\lambda}_{k+1},\dots,\tilde{\lambda}_m} \prod_{j=k+1}^{m} M(\alpha, \tilde{\lambda}_j) = \prod_{j=r+1}^{m} \sum_{\tilde{\lambda} \in L} M(\alpha, \tilde{\lambda}) = 1.$$

Inserting this back into (5.7) finishes the proof. □

Remark 5.5. The same stochastic interpretation as for Eq. (5.5) (see Remark 5.3) holds also for the marginalised system. With probability

$$r^U_{\mathcal{A}} := \sum_{\substack{\mathcal{B} \in \boldsymbol{P}(S) \\ \mathcal{B}|_U = \mathcal{A}}} r_{\mathcal{A}},$$

the subsequence with respect to U of a sampled individual is partitioned across its ancestors according to $\mathcal{A}$. Then, the labels are reassigned independently according to M, reflecting their independent migration. ♢

5.4 Solution of the forward iteration

Next, we use the marginalisation consistency established in the previous section to tame the MRE. As discussed in [BB16] for pure recombination, the main idea is to consider the time evolution of the (column) vector $\mathcal{R}(\mu_t) := \big(\mathcal{R}_{\boldsymbol{\mathcal{A}}}(\mu_t)\big)_{\boldsymbol{\mathcal{A}} \in \boldsymbol{LP}(S)}$, rather than μ_t on its own; note that we recover $\mu_t(\alpha)$ as the $\underline{\mathbf{1}}^{\alpha}$-component of $\mathcal{R}(\mu_t)$.

Theorem 5.7. *The matrix $\boldsymbol{T}$, indexed by $\boldsymbol{LP}(S)$, with entries*

$$\boldsymbol{T}_{\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{B}}} = \begin{cases} 0, & \text{if } \boldsymbol{\mathcal{B}} \not\preccurlyeq \boldsymbol{\mathcal{A}}, \\ \prod_{(A,\lambda) \in \boldsymbol{\mathcal{A}}} p^{A}_{\boldsymbol{\mathcal{B}}|_A}(\lambda), & \text{if } \boldsymbol{\mathcal{B}} \preccurlyeq \boldsymbol{\mathcal{A}}, \end{cases}$$

where the $p^A_{\boldsymbol{\mathcal{B}}|_A}(\lambda)$ *are as in Lemma* 5.6*, is stochastic. Assume that* $(\mu_t)_{t\in\mathbb{N}_0}$ *satisfies the* MRE (5.5)*. Then,* $\mathcal{R}(\mu_t)$ *satisfies the linear recursion*

$$\mathcal{R}(\mu_{t+1}) = \boldsymbol{T}\mathcal{R}(\mu_t).$$

In particular,

$$\mathcal{R}(\mu_t) = \boldsymbol{T}^t\mathcal{R}(\mu_0)$$

(where T^t *denotes the* t*-th power of* $\boldsymbol{T}$*).*

Proof. By Definition 5.2, Theorem 5.5, Remark 5.4, and Lemma 5.4,

$$\begin{aligned}
\mathcal{R}_{\boldsymbol{\mathcal{A}}}(\mu_{t+1}) &= \bigotimes_{(A,\lambda)\in\boldsymbol{\mathcal{A}}} \sum_{\boldsymbol{\mathcal{B}}_A\in\boldsymbol{LP}(A)} p^A_{\boldsymbol{\mathcal{B}}_A}(\lambda)\mathcal{R}^A_{\boldsymbol{\mathcal{B}}_A}(\mu_t^A) \\
&= \sum_{\substack{\boldsymbol{\mathcal{B}}_A\in\boldsymbol{LP}(A)\\ A\in\mathcal{A}}} \Big(\prod_{(A,\lambda)\in\boldsymbol{\mathcal{A}}} p^A_{\boldsymbol{\mathcal{B}}_A}(\lambda)\Big) \bigotimes_{A\in\mathcal{A}} \mathcal{R}^A_{\boldsymbol{\mathcal{B}}_A}(\mu_t^A) \\
&= \sum_{\boldsymbol{\mathcal{B}}\preccurlyeq\boldsymbol{\mathcal{A}}} \Big(\prod_{(A,\lambda)\in\boldsymbol{\mathcal{A}}} p^A_{\boldsymbol{\mathcal{B}}|_A}(\lambda)\Big) \bigotimes_{A\in\mathcal{A}} \mathcal{R}^A_{\boldsymbol{\mathcal{B}}|_A}(\mu_t^A) \\
&= \sum_{\boldsymbol{\mathcal{B}}\preccurlyeq\boldsymbol{\mathcal{A}}} \boldsymbol{T}_{\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{B}}}\mathcal{R}_{\boldsymbol{\mathcal{B}}}(\mu_t),
\end{aligned}$$

where the underdot indicates the summation variable. That $\boldsymbol{T}$ is a stochastic matrix is a straightforward consequence of $p^A(\alpha)$ being a probability distribution on $\boldsymbol{LP}(A)$ for all $A \subseteq S$ and all $\alpha \in L$.

□

We have just witnessed how the solution of a nonlinear system, embedded in a higher dimensional space, turns into the solution of a linear system and may thus be given explicitly, simply via matrix powers. This is an extension of a technique called *Haldane linearisation* [MR83; BB16; BB] to the case with migration. The reason for why this works out so well can be found in the underlying genealogical structure, which is discussed next.

5.5 Stochastic interpretation, genealogical content, and duality

Let us now turn to the probabilistic content of Theorem 5.7. We will see that the appearance of the stochastic matrix T is no coincidence; rather, it has a natural interpretation as the transition matrix of a Markov chain, which describes the random genealogy of a single individual.

Definition 5.8. The *labelled partitioning process* (LPP) is a discrete-time Markov chain $\boldsymbol{\Sigma} =$

$(\boldsymbol{\Sigma}_t)_{t\in\mathbb{N}_0}$ with values in $\boldsymbol{LP}(S)$ and transition matrix $\boldsymbol{T}$, that is,

$$\mathbb{P}(\boldsymbol{\Sigma}_{t+1} = \boldsymbol{\mathcal{B}} \mid \boldsymbol{\Sigma}_t = \boldsymbol{\mathcal{A}}) = \boldsymbol{T}_{\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{B}}}$$

for all $\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \in \boldsymbol{LP}(S)$. ♢

So, $\boldsymbol{\Sigma}_{t+1}$ is constructed from $\boldsymbol{\Sigma}_t$ by independently replacing each labelled block $(A,\lambda) \in \boldsymbol{\Sigma}_t$ by the (labelled) blocks of the labelled partition $\boldsymbol{\mathcal{B}}_A$ with probability $p^A_{\boldsymbol{\mathcal{B}}_A}(\lambda)$.

The genealogical interpretation of $\boldsymbol{\Sigma}$, started in $\underline{\mathbf{1}}^\alpha$, is as follows. Each labelled block (A,λ) of $\boldsymbol{\Sigma}_t$ corresponds to a different ancestor of the individual at present, which was sampled at location α, who lived t generations before the present, at location λ. The elements of A are the sequence sites that are inherited from this ancestor. As we look one generation further into the past, A is replaced by the blocks of a labelled partition $\boldsymbol{\mathcal{B}}_A \in \boldsymbol{LP}(A)$, which describes how the type of that ancestor is, in turn, pieced together from its parents, alive $t+1$ generations before the present. Note that now the labelled partitions of A are relevant rather than those of S. This is because we already know that this ancestor only contributes sites contained in A, whence we only need to trace back the ancestry of these sites. (This reflects the marginalisation consistency of the model, cf. Remark 5.5). Furthermore, the various blocks split independently as the law of large numbers regime assumed here implies that two given individuals never share a common ancestor; thus, their lineages are conditionally independent.

The connection between the solution of the MRE and the genealogical process is formalised in the following theorem, which is a probabilistic restatement of Theorem 5.7 and draws on the notion of duality for Markov processes [Lig10; JK14]; in particular, we think about the solution of the forward-time equation as a Markov chain with deterministic transitions.

Theorem 5.9. *The* LPP *and the solution of the* MRE *are dual with respect to the duality function*

$$(\boldsymbol{\mathcal{A}}, \nu) \mapsto \mathcal{R}_{\boldsymbol{\mathcal{A}}}(\nu).$$

That is, for all $\boldsymbol{\mathcal{A}} \in \boldsymbol{LP}(S)$ *and all* $\mu_0 \in \mathcal{P}(X)^L$*, we have*

$$\mathbb{E}[\mathcal{R}_{\boldsymbol{\Sigma}_t}(\mu_0) \mid \boldsymbol{\Sigma}_0 = \boldsymbol{\mathcal{A}}] = \mathcal{R}_{\boldsymbol{\mathcal{A}}}(\mu_t).$$

In particular, this entails the stochastic representation

$$\mu_t(\alpha) = \mathbb{E}[\mathcal{R}_{\boldsymbol{\Sigma}_t}(\mu_0) \mid \boldsymbol{\Sigma}_0 = \underline{\mathbf{1}}^\alpha]$$

for the solution of the MRE*.*

Proof. We prove the theorem by induction over t. For $t = 0$, there is nothing to show. Assuming now that

$$\mathbb{E}[\mathcal{R}_{\boldsymbol{\Sigma}_t}(\mu_0) \mid \boldsymbol{\Sigma}_0 = \boldsymbol{\mathcal{A}}] = \mathcal{R}_{\boldsymbol{\mathcal{A}}}(\mu_t)$$

for any $t > 0$, we compute, using Theorem 5.7 in the first step, the induction hypothesis in the second, time-homogeneity in the third, and the Markov property in the last:

$$\mathcal{R}_{\boldsymbol{\mathcal{A}}}(\mu_{t+1}) = \sum_{\boldsymbol{\mathcal{B}} \preccurlyeq \boldsymbol{\mathcal{A}}} \boldsymbol{T}_{\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{B}}} \mathcal{R}_{\boldsymbol{\mathcal{B}}}(\mu_t) = \sum_{\boldsymbol{\mathcal{B}} \preccurlyeq \boldsymbol{\mathcal{A}}} \mathbb{P}[\boldsymbol{\Sigma}_1 = \boldsymbol{\mathcal{B}} \mid \boldsymbol{\Sigma}_0 = \boldsymbol{\mathcal{A}}]\, \mathbb{E}[\mathcal{R}_{\boldsymbol{\Sigma}_t}(\mu) \mid \boldsymbol{\Sigma}_0 = \boldsymbol{\mathcal{B}}]$$

$$= \sum_{\boldsymbol{\mathcal{B}} \preccurlyeq \boldsymbol{\mathcal{A}}} \mathbb{P}[\boldsymbol{\Sigma}_1 = \boldsymbol{\mathcal{B}} \mid \boldsymbol{\Sigma}_0 = \boldsymbol{\mathcal{A}}]\, \mathbb{E}[\mathcal{R}_{\boldsymbol{\Sigma}_{t+1}}(\mu) \mid \boldsymbol{\Sigma}_1 = \boldsymbol{\mathcal{B}}] = \sum_{\boldsymbol{\mathcal{B}} \preccurlyeq \boldsymbol{\mathcal{A}}} \mathbb{E}[\mathcal{R}_{\boldsymbol{\Sigma}_{t+1}}(\mu) \mid \boldsymbol{\Sigma}_0 = \boldsymbol{\mathcal{A}}].$$

This proves the statement for $t + 1$. □

Note that the duality function used here is vector-valued. This is a slight extension of the standard notion, since the duality function is usually assumed to take values in $\mathbb{R}$; see also Remark 4.16.

To get used to this probabilistic way of thinking, we now take advantage of the stochastic representation from Theorem 5.9 to construct an explicit solution formula in the case of two sites. When evaluating the expectation, we distinguish two cases. Either, the two sites have not been separated until generation t, which happens with probability $r_{\underline{1}}^t$. In this case, both sites have the same ancestor who comes, with probability $(M^t)_{\alpha\gamma}$, from location γ. If, on the other hand, the sites *have* been separated, we denote by τ the smallest t such that $|\boldsymbol{\Sigma}_t| = 2$. In this case, the letters come from two different parents. Their origins are determined by performing independent random walks on L for the remaining time $t - \tau + 1$. Summing over all possible values for τ and the label of the block upon splitting (which is β with probability $(M^{\tau-1})_{\alpha\beta}$) , we see that

$$\mu_t(\alpha) = r_{\underline{1}}^t (M^t \mu_0)_\alpha + \sum_{\beta \in L} \sum_{\tau=1}^{t} r_{\underline{1}}^{\tau-1} r_{\underline{0}} (M^{\tau-1})_{\alpha\beta} (M^{t-\tau+1}\mu_0)_\beta^{\{1\}} \otimes (M^{t-\tau+1}\mu_0)_\beta^{\{2\}}. \tag{5.8}$$

In the case without migration (i.e, when ignoring the labels), the LPP reduces to the partitioning process mentioned already at the end of Chapter 2. More precisely, the (unlabelled) partitioning process $(\Sigma_t)_{t\geqslant 0}$ is simply the base of the LPP $(\boldsymbol{\Sigma}_t)_{t\geqslant 0}$, and couples to it in a natural way. Likewise, an explicit description of its transition matrix T is obtained from $\boldsymbol{T}$ by marginalising over the labels. Thus, T has the entries (compare Eq. (2.11))

$$T_{\mathcal{A}\mathcal{B}} = \begin{cases} 0, & \text{if } \mathcal{B} \not\preccurlyeq \mathcal{A}, \\ \prod_{A \in \mathcal{A}} r^A_{\mathcal{B}|_A}, & \text{if } \mathcal{B} \preccurlyeq \mathcal{A}, \end{cases} \tag{5.9}$$

and the transition rates for the LPP factorise as

$$\boldsymbol{T}_{\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{B}}} = T_{\mathcal{A}\mathcal{B}} \prod_{(A,\lambda) \in \boldsymbol{\mathcal{A}}} \prod_{(B,\gamma) \in \boldsymbol{\mathcal{B}}|_A} M(\lambda, \gamma),$$

compare Lemma 5.6. Note that $(\Sigma_t)_{t\geqslant 0}$ is a process of progressive refinement, which never

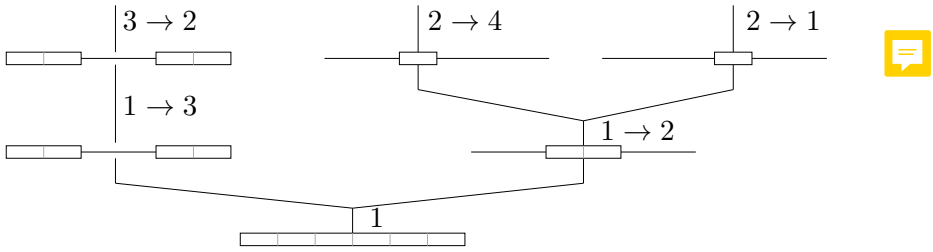


Figure 5.2. An illustration of the LPP starting from $\underline{\mathbf{1}}^1$, the trivial partition consisting of a single block with label 1; the set of locations is $L = \{1, 2, 3, 4\}$. Backward time runs from bottom to top. In each generation, the blocks of the partition are first subject to individual splitting and we trace back the ancestral lines of each fragment; compare Remarks 5.5 and 5.3. The fragments provided by each ancestor are labelled with their locations and we write $\alpha \to \beta$ to indicate migration from α to β. Recall that in the forward-time model, recombination occurs after migration. Thus, when looking backward in time, splitting (or branching) due to recombination occurs before the reassignment of the labels due to migration. In particular, the first event in this example is a splitting of our sequence located in deme 1.

returns to a state $\not\preccurlyeq$ the current state. This is due to the absence of coalescence events in the law of large numbers regime, which means that the ancestral recombination *graph* is actually a *tree*.

Remark 5.6. The LPP can be interpreted as a multitype branching random walk (BRW) on L, with the types given by the subsets of S. The particles move according to the transition kernel M, and, as evident from the product structure of the transitions in Eq. (5.9) and undergo independent branching that is the same at every location; each individual of type A branches with probability $r_{\mathcal{A}}^{A}$ into $|\mathcal{A}|$ individuals of types $A_1, \ldots, A_{|\mathcal{A}|}$, where $\mathcal{A} = \{A_1, \ldots, A_{|\mathcal{A}|}\}$. ♢

5.6 Limiting and quasi-limiting behaviour of the LPP

We assume now that M is primitive (that is, irreducible and aperiodic), which guarantees the existence of and convergence to a unique stable stationary distribution $q = \big(q(\alpha)\big)_{\alpha \in L} \in \mathbb{R}^L$ such that

$$q^{\mathsf{T}} = q^{\mathsf{T}} M, \tag{5.10}$$

where $^{\mathsf{T}}$ denotes the transpose operation. This convergence is uniform in the initial condition.

We also assume that

$$\bigwedge \{\mathcal{A} \in \boldsymbol{P}(S) : r_{\mathcal{A}} > 0\} = \underline{0}. \tag{5.11}$$

That is, the coarsest common refinement of all partitions with positive recombination probability is the trivial partition $\underline{0}$ of S into singletons. This is only a matter of technical convenience; otherwise, we could simply reconsider any set of sites that are not separated by any partition $\mathcal{A}$ with $r_{\mathcal{A}} > 0$ as a single site. Note that Eq. (5.11) implies that $\underline{0}$ is the unique absorbing state of the (unlabelled) partitioning process. We can now explicitly state

the asymptotic behaviour of the MRE.

Theorem 5.10. *Under the above assumptions, one has*

$$\lim_{t\to\infty} \mu_t = \mu_\infty = \big(\mu_\infty(\alpha)\big)_{\alpha\in L},$$

where

$$\mu_\infty(\alpha) = \bigotimes_{i=1}^{n} \mu_\infty^{\{i\}}(\alpha) \tag{5.12}$$

and

$$\mu_\infty^{\{i\}}(\alpha) := \sum_{\beta\in L} q(\beta)\mu_0^{\{i\}}(\beta) \tag{5.13}$$

for $\alpha \in L$. The convergence is geometric, i.e. there is a $\gamma \in (0,1)$ such that

$$\mu_t = \mu_\infty + \mathcal{O}(\gamma^t)$$

as $t \to \infty$, uniformly in μ_0.

This is in line with [Bür09, Theorem 3.1], which states that the solution of (5.4) approaches (at a uniform geometric rate) the submanifold defined by spatial stationarity and linkage equilibrium. Spatial stationarity means that

$$\mu(\alpha) = \sum_{\beta\in L} q(\beta)\mu(\beta)$$

with q of (5.10); and, under the assumption (5.11), linkage equilibrium means that $\mu(\alpha)$ is the product of its one-dimensional marginals, as in Eq. (5.12). However, like the explicit time evolution in Theorem 5.7, the explicit expression in Eq. (5.13) seems to be new.

In view of Theorem 5.9, this result is highly plausible: almost surely (at a uniform geometric rate), the partitioning process will enter its unique absorbing state where all blocks are singletons. Subsequently, the distributions of the independent migration processes (i.e, random walks on L with transition matrix M) associated with each block will converge to the unique stationary distribution q, again at a geometric rate and uniformly in the initial condition. This behaviour is also clear in terms of the BRW picture. At some point, the type of each particle is a singleton, whence the particles stop branching and just keep performing independent random walks; see Remark 5.6.

For the formal proof, note that the uniform convergence of the migration processes follows directly from the primitivity of M via standard theory [KT75, Thm. 2.3, Appendix]. That the partitioning process enters its absorbing state at a uniform geometric rate is the content of the following result.

Lemma 5.11. *Let*

$$\eta := \max_{\mathcal{A}\in \boldsymbol{P}(S)\setminus\{\emptyset\}} T_{\mathcal{A}\mathcal{A}} < 1$$

be the maximal sojourn probability of the (unlabelled) partitioning process and let

$$\tau := \min\{t \in \mathbb{N}_0 : \Sigma_t = \underline{0}\}$$

be its time to absorption. Then, uniformly in the initial distribution,

$$\mathbb{P}(\tau > t) = \mathcal{O}((\eta + \varepsilon)^t)$$

for any $\varepsilon > 0$ as $t \to \infty$.

Proof. Since the state space is finite and the partitioning process never returns to a state $\not\preccurlyeq$ the current state, this Markov chain may jump at most a finite number of times, say m times, before it is absorbed in $\underline{0}$. Thus, for any fixed $\gamma > \eta$,

$$\begin{aligned}\mathbb{P}(\tau > t) &\leqslant \mathbb{P}(\text{the chain has performed at most } m \text{ jumps up to time } t)\\ &\leqslant \sum_{j=0}^{m} \binom{t}{j}(1-\eta)^j \eta^{t-j} \leqslant \sum_{j=0}^{m} \Big(\frac{1-\eta}{\eta}\Big)^j t^m \eta^t = C' t^m \eta^t \leqslant C\eta^t \Big(\frac{\eta+\epsilon}{\eta}\Big)^t = C(\eta+\epsilon)^t,\end{aligned}$$

where $C' = \sum_{j=0}^{m} \left(\frac{1-\eta}{\eta}\right)^j$ and C is chosen sufficiently large. □

Next, we investigate the asymptotic behaviour of the LPP.

Proposition 5.12. *There exists a $\gamma \in (0,1)$ such that*

$$\mathbb{P}(\boldsymbol{\Sigma}_t = \{(\{1\}, \alpha_1), \ldots, (\{n\}, \alpha_n)\}) = \prod_{i=1}^{n} q(\alpha_i) + \mathcal{O}(\gamma^t)$$

as $t \to \infty$, uniformly in $\alpha_1, \ldots, \alpha_n \in L$ and the initial distribution of the LPP*. For any $\boldsymbol{\mathcal{A}} \in \boldsymbol{LP}(S)$ with $\mathcal{A} \neq \underline{0}$,*

$$\mathbb{P}(\boldsymbol{\Sigma}_t = \boldsymbol{\mathcal{A}}) = \mathcal{O}((\eta + \varepsilon)^t)$$

for all $\varepsilon > 0$, again uniformly in the initial distribution.

Proof. Let τ be as in Lemma 5.11. The second statement follows immediately from Lemma 5.11 by noting that

$$\mathbb{P}(\boldsymbol{\Sigma}_t = \boldsymbol{\mathcal{A}}) \leqslant \mathbb{P}(\tau > t).$$

Now, assume that $\boldsymbol{\mathcal{A}}$ is of the form

$$\boldsymbol{\mathcal{A}} = \{(\{1\}, \alpha_1), \ldots, (\{n\}, \alpha_n)\}.$$

Then, for all $\gamma_1 > \eta$,

$$\begin{aligned}\mathbb{P}(\boldsymbol{\Sigma}_t = \boldsymbol{\mathcal{A}}) &= \mathbb{P}\Big(\boldsymbol{\Sigma}_t = \boldsymbol{\mathcal{A}} \mid \tau \leqslant \Big\lfloor \frac{t}{2} \Big\rfloor\Big)\mathbb{P}\Big(\tau \leqslant \Big\lfloor \frac{t}{2} \Big\rfloor\Big) + \mathbb{P}\Big(\boldsymbol{\Sigma}_t = \boldsymbol{\mathcal{A}} \mid \tau > \Big\lfloor \frac{t}{2} \Big\rfloor\Big)\mathbb{P}\Big(\tau > \Big\lfloor \frac{t}{2} \Big\rfloor\Big) \\ &= \mathbb{P}\Big(\boldsymbol{\Sigma}_t = \boldsymbol{\mathcal{A}} \mid \tau \leqslant \Big\lfloor \frac{t}{2} \Big\rfloor\Big) + \mathcal{O}(\gamma_1^t)\end{aligned} \tag{5.14}$$

as $t \to \infty$, where we used Lemma 5.11 in the last step. Furthermore,

$$\begin{aligned}\mathbb{P}\Big(\boldsymbol{\Sigma}_t = \boldsymbol{\mathcal{A}} \mid \tau \leqslant \Big\lfloor \frac{t}{2} \Big\rfloor\Big) &= \mathbb{P}\Big(\Lambda_t^{(i)} = \alpha_i \text{ for all } 1 \leqslant i \leqslant n \mid \tau \leqslant \Big\lfloor \frac{t}{2} \Big\rfloor\Big) \\ &= \prod_{i=1}^{n} \mathbb{P}\Big(\Lambda_t^{(i)} = \alpha_i \mid \tau \leqslant \Big\lfloor \frac{t}{2} \Big\rfloor\Big).\end{aligned} \tag{5.15}$$

Here, the $(\Lambda_t^{(i)})_{t \in \mathbb{N}_{\geqslant \tau}}$ for $i \in L$ are the labels of the (singleton) blocks from time τ onwards; they are given by independent L-valued Markov chains with transition matrix M. By standard theory, we can be sure that, regardless of the initial value, there is a $\gamma_2 \in (0,1)$ so that

$$\mathbb{P}\Big(\Lambda_t^{(i)} = \alpha_i \mid \tau \leqslant \Big\lfloor \frac{t}{2} \Big\rfloor\Big) = q(\alpha_i) + \mathcal{O}(\gamma_2^t),$$

uniformly in α_i. Combining this with Eqs. (5.14) and (5.15) proves the theorem. □

Proof of Theorem 5.10. By Theorem 5.9, Proposition 5.12, and Definition 5.2, we have for some $\gamma \in (0,1)$, independently of μ_0,

$$\begin{aligned}\mu_t(\alpha) &= \mathbb{E}[\mathcal{R}_{\boldsymbol{\Sigma}_t}(\mu_0) \mid \boldsymbol{\Sigma}_0 = \underline{\mathbf{1}}^{\alpha}] \\ &= \sum_{\beta_1,\dots,\beta_n \in L} \Big(\prod_{i=1}^{n} q(\beta_i)\Big)\mathbb{E}[\mathcal{R}_{\boldsymbol{\Sigma}_t}(\mu_0) \mid \boldsymbol{\Sigma}_0 = \underline{\mathbf{1}}^{\alpha}, \boldsymbol{\Sigma}_t = \{(\{1\},\beta_1),\dots,(\{n\},\beta_n)\}] + \mathcal{O}(\gamma^t) \\ &= \sum_{\beta_1,\dots,\beta_n \in L} \bigotimes_{i=1}^{n} q(\beta_i)\mu_0^{\{i\}}(\beta_i) + \mathcal{O}(\gamma^t) \\ &= \bigotimes_{i=1}^{n} \sum_{\beta \in L} q(\beta)\mu_0^{\{i\}}(\beta) + \mathcal{O}(\gamma^t) = \bigotimes_{i=1}^{n} \mu_\infty^{\{i\}}(\alpha) + \mathcal{O}(\gamma^t) = \mu_\infty(\alpha) + \mathcal{O}(\gamma^t)\end{aligned}$$

with μ_∞ as in Theorem 5.10. □

Since the asymptotic behaviour of the LPP is so simple, we now go one step further and inquire about its *quasi-limiting* behaviour; that is, its asymptotic behaviour, conditioned on non-absorption. Recall that the partitioning process is a process of progressive refinement, and never returns to a state finer than the current state. This is very different from the situation considered in [CMS13], where the focus is on irreducible chains.

Unlike the limiting distribution, the quasi-limiting distribution will generally depend on the initial distribution. For convenience of notation, we let the LPP start from a maximal labelled

partition $\underline{\mathbf{1}}^{\alpha}$. However, the following discussion can easily be adapted to the more general setting. In what follows, we will exclude the pathological case of $r_{\underline{0}} = 1$, where the probability of non-absorption is zero, and the conditional distribution we are interested in is not well defined.

We start by recalling the quasi-limiting behaviour of the unlabelled partitioning process $(\Sigma_t)_{t\in\mathbb{N}_0}$, which was already investigated in [Mar17]. We posit throughout that $\Sigma_0 = \underline{1}$. To state the result, we need some additional notation. First, we define the set of states

$$\boldsymbol{P}^{\downarrow}(S) := \{\mathcal{A} \in \boldsymbol{P}(S) : \exists \ell \in \mathbb{N} \text{ s.t. } \left(T^{\ell}\right)_{\underline{1}\mathcal{A}} > 0\}$$

that are *reachable* by $(\Sigma_t)_{t\in\mathbb{N}_0}$ when starting in $\underline{1} = \{S\}$. As before, η denotes the maximal sojourn probability of $(\Sigma_t)_{t\in\mathbb{N}_0}$ (cf. Lemma 5.11). We will also need the set

$$\mathcal{F} := \{\mathcal{A} \in \boldsymbol{P}^{\downarrow}(S) : T_{\mathcal{A}\mathcal{A}} = \eta\}$$

of reachable states with maximal sojourn probability. Note that our assumption $r_{\underline{0}} \neq 1$ guarantees that $\eta > 0$. Finally, we define the *first hitting time* of any given $\mathcal{A} \in \boldsymbol{P}(S)$,

$$\tau_{\mathcal{A}} := \min\{t \in \mathbb{N}_0 : \Sigma_t = \mathcal{A}\},$$

we write $\tau_{\mathcal{F}} := \min_{\mathcal{A}\in\mathcal{F}} \tau_{\mathcal{A}}$ for the first hitting time of $\mathcal{F}$, and, as before, $\tau = \tau_{\underline{0}}$ for the time to absorption. The following result is known; see [Mar17, Thm 5.5].

Theorem 5.13. *For all $\mathcal{A} \in \mathcal{F}$, one has*

$$0 < \mathbb{E}[\eta^{-\tau_{\mathcal{A}}}; \tau_{\mathcal{A}} < \infty] \leqslant \mathbb{E}[\eta^{-\tau_{\mathcal{F}}}; \tau_{\mathcal{F}} < \infty] < \infty.$$

Further, for all $\mathcal{A} \in \boldsymbol{P}(S)$, the limit

$$\mathbb{P}^{\Sigma}_{\mathrm{qlim}}(\mathcal{A}) := \lim_{t\to\infty} \mathbb{P}(\Sigma_t = \mathcal{A} \mid \tau > t)$$

exists and is equal to

$$\frac{\mathbb{E}[\eta^{-\tau_{\mathcal{A}}}; \tau_{\mathcal{A}} < \infty]}{\mathbb{E}[\eta^{-\tau_{\mathcal{F}}}; \tau_{\mathcal{F}} < \infty]} \mathbb{1}_{\mathcal{A}\in\mathcal{F}}.$$

Thus defined, $\mathbb{P}^{\Sigma}_{\mathrm{qlim}}$ is a probability measure on $\boldsymbol{P}(S)$, called the quasi-limiting distribution *of $(\Sigma_t)_{t\geqslant 0}$ (starting from $\underline{1}$).*

Recall that the labels of the different blocks evolve conditionally independently. Thus, we expect the quasi-limiting distribution of the LPP to be similar to the quasi-limiting distribution from Theorem 5.13, garnished with the stationary distribution q of the migration process. More explicitly, we are going to prove the following result.

Theorem 5.14. *For all* $\boldsymbol{\mathcal{A}} \in \boldsymbol{LP}(S)$,

$$\lim_{t\to\infty} \mathbb{P}(\boldsymbol{\Sigma}_t = \boldsymbol{\mathcal{A}} \mid \tau > t) = \Big(\prod_{(A,\lambda)\in\boldsymbol{\mathcal{A}}} q(\lambda) \Big) \mathbb{P}^{\Sigma}_{\mathrm{qlim}}(\mathcal{A}),$$

where q *is the unique stationary distribution* (5.10) *of the migration process.*

Remark 5.7. In Theorem 5.10, we have approximated the solution of the MRE (5.4) by approximating the distribution of the labelled partitioning process by its limiting distribution, given in Proposition 5.12. We can try to improve on this rather coarse estimate by also taking the quasi-limiting distribution into account; at least in principle, the disintegration

$$\mathbb{P}(\boldsymbol{\Sigma}_t = \boldsymbol{\mathcal{A}}) = \mathbb{P}(\boldsymbol{\Sigma}_t = \boldsymbol{\mathcal{A}} \mid \tau \leqslant t)\mathbb{P}(\tau \leqslant t) + \mathbb{P}(\boldsymbol{\Sigma}_t = \boldsymbol{\mathcal{A}} \mid \tau > t)\mathbb{P}(\tau > t)$$

allows us to express the error term in Theorem 5.10 via the quasi-limiting distribution, at least when migration is strong compared to recombination. Acquiring precise asymptotics, however, would require more detailed knowledge about the probability $\mathbb{P}(\tau > t)$ and the rate of convergence of the conditional distribution $\mathbb{P}(\boldsymbol{\Sigma}_t = \boldsymbol{\mathcal{A}} \mid \tau > t)$ to the quasi-limiting distribution. ♢

At the heart of the proof is the observation that any further refinement of any $\mathcal{A} \in \mathcal{F}$ immediately leads to absorption; this was also one of the crucial ingredients in the proof of Theorem 5.13, see [Mar17, Thm 5.5] for the original reference[1].

Lemma 5.15. *For all* $\mathcal{A} \in \mathcal{F}$*, we have*

$$T_{\mathcal{A}\mathcal{A}} + T_{\mathcal{A}\underline{0}} = 1. \tag{5.16}$$

Proof. We show that, for all $\mathcal{A} \in \boldsymbol{P}^{\downarrow}(S)$ with $T_{\mathcal{A}\mathcal{A}} + T_{\mathcal{A}\underline{0}} \neq 1$, one has $\mathcal{A} \notin \mathcal{F}$. Indeed, for any such $\mathcal{A}$, there is a $\mathcal{B} \notin \{\underline{0}, \mathcal{A}\}$ with $T_{\mathcal{A}\mathcal{B}} > 0$. Now, $\mathcal{B} \neq \underline{0}$ means that there is at least one block $B \in \mathcal{B}$ with $|B| > 1$ and the partition

$$\mathcal{B}' := \{B\} \cup \{\{i\} : i \in S \setminus B\} \preccurlyeq \mathcal{A}$$

is reachable by Assumption (5.11) (with S replaced by individual blocks of $\mathcal{A}$). We then have

$$T_{\mathcal{B}'\mathcal{B}'} = r^{B}_{\{B\}} > r^{\tilde{A}}_{\{\tilde{A}\}} \prod_{\substack{A\in\mathcal{A}\\ A\neq\tilde{A},|A|>1}} r^{A}_{\{A\}} = \prod_{A\in\mathcal{A}} r^{A}_{\{A\}} = T_{\mathcal{A}\mathcal{A}},$$

where $\tilde{A}$ is the block in $\mathcal{A}$ that contains B. The inequality holds for the following reason. Either, $|\{A \in \mathcal{A} : |A| > 1\}| > 1$, which implies that the constrained product is not empty and hence smaller than 1; note that $r^{A}_{\{A\}} < 1$ for A with $|A| > 1$. Otherwise, if $\tilde{A}$ is the only

[1] Unfortunately, the proof of this lemma does not seem to have been addressed in the corresponding corrigendum; that is why we decided to give an independent proof here.

block of $\mathcal{A}$ with more than one element, there must be some $\mathcal{C} \in \boldsymbol{P}(S)$ with $r_{\mathcal{C}} > 0$ such that $B \in \mathcal{C}|_{\tilde{A}}$ and thus, $r^{B}_{\{B\}} \geqslant r^{\tilde{A}}_{\{\tilde{A}\}} + r_{\mathcal{C}} > r^{\tilde{A}}_{\{\tilde{A}\}}$. □

Remark 5.8. One might be tempted to assume that the sojourn probability is non-decreasing along every path

$$\underline{1} \succcurlyeq \mathcal{A}_1 \succcurlyeq \mathcal{A}_2 \succcurlyeq \ldots \succcurlyeq \underline{0}$$

from the maximal partition to the absorbing state. To illustrate that this is not true in general, consider the following setup. Let $n = 4$ and assume the recombination distribution given by $r_{\underline{0}} = \frac{1}{2}$, $r_{\{\{1,2\},\{3,4\}\}} = \frac{1}{10}$, $r_{\underline{1}} = \frac{2}{5}$ and $r_{\mathcal{A}} = 0$ otherwise. Then, the sojourn probability of the state $\underline{1}$ is $r_{\underline{1}} = \frac{2}{5}$, while the (finer) state $\{\{1,2\},\{3,4\}\}$ has the smaller sojourn probability

$$r^{\{1,2\}}_{\{1,2\}} r^{\{3,4\}}_{\{3,4\}} = (1 - r_{\underline{0}})^2 = \frac{1}{4}.$$

♢

The idea of the proof of Theorem 5.14 is simple. First, notice that Lemma 5.15 implies that conditional on non-absorption, $(\Sigma_t)_{t\geqslant 0}$ remains constant after $\tau_{\mathcal{F}}$. From then on, the labels keep on evolving according to independent random walks with transition matrix M, and their distributions converge to q. To make this rigorous, we just need to make sure that, conditional on non-absorption at time t, $t - \tau_{\mathcal{F}}$ is large enough.

Lemma 5.16. (a) *There exists $c > 0$ such that $\mathbb{P}(\tau > t) \geqslant c\eta^t$ for all $t \in \mathbb{N}$.*

(b) *Let $\eta' := \max_{\mathcal{A}\in\boldsymbol{P}(S)\setminus(\mathcal{F}\cup\{\underline{0}\})} T_{\mathcal{A}\mathcal{A}}$. Then, for all $\eta'' > \eta'$, there exists $C > 0$ such that $\mathbb{P}(\tau_{\mathcal{F}} \wedge \tau > t) \leqslant C(\eta'')^t$ for all $t \in \mathbb{N}$.*

(c) *There is a $\gamma \in (0,1)$ such that $\lim_{t\to\infty} \mathbb{P}(\tau_{\mathcal{F}} > \gamma t \mid \tau > t) = 0$.*

Proof. First, we show (a). By definition, $\mathcal{F} \subseteq \boldsymbol{P}^{\downarrow}(S)$. Thus, there exists a $t_0 \in \mathbb{N}$ such that $\mathbb{P}(\tau_{\mathcal{F}} = t_0) > 0$. Then, we have for all $t \geqslant t_0$ that

$$\mathbb{P}(\tau > t) \geqslant \mathbb{P}(\tau > t, \tau_{\mathcal{F}} = t_0) = \mathbb{P}(\tau > t \mid \tau_{\mathcal{F}} = t_0)\,\mathbb{P}(\tau_{\mathcal{F}} = t_0) = c'\eta^{t-t_0} = (c'\eta^{-t_0})\eta^t$$

with $c' = \mathbb{P}(\tau_{\mathcal{F}} = t_0)$. Note that we used Lemma 5.15 in the second-last step. Now, simply choose

$$c := \min\Big\{\frac{\mathbb{P}(\tau > t)}{\eta^t} : 0 \leqslant t \leqslant t_0\Big\} \cup \{c'\eta^{-t_0}\}.$$

For the proof of (b), we couple $(\Sigma_t)_{t\in\mathbb{N}_0}$ to another process $(N_t)_{t\in\mathbb{N}_0}$ with values in $\mathbb{N}_0 \cup \{\infty\}$ and $N_0 = 0$. Its dynamics is described as follows. When $\Sigma_{t+1} = \Sigma_t$, then $N_{t+1} := N_t$; when $\Sigma_{t+1} \in \mathcal{F} \cup \{\underline{0}\}$, we set $N_{t+1} := \infty$. In all other cases, we perform a Bernoulli experiment with success probability

$$\frac{1 - \eta'}{1 - T_{\Sigma_t\Sigma_t}}.$$

Upon success, we set $N_{t+1} := N_t + 1$; otherwise, $N_{t+1} := N_t$. Note that the marginal $(N_t)_{t\in\mathbb{N}_0}$ of the coupling $(\Sigma_t, N_t)_{t\in\mathbb{N}_0}$ stochastically dominates yet another stochastic prococess $(K_t)_{t\in\mathbb{N}_0}$, which has independent Bernoulli increments with parameter $1-\eta'$.

As we have argued before, the partitioning process can only jump a finite number of times before hitting either $\underline{0}$ or $\mathcal{F}$. Thus, there is a positive integer m such that, for all $t \in \mathbb{N}$, $\tau \wedge \tau_{\mathcal{F}} > t$ implies $N_t \leqslant m$. Thus,

$$\mathbb{P}(\tau \wedge \tau_{\mathcal{F}} > t) \leqslant \mathbb{P}(N_t \leqslant m) \leqslant \mathbb{P}(K_t \leqslant m) = \sum_{k=0}^{m} \binom{t}{k} (1-\eta')^k (\eta')^{t-k} = P(t)(\eta')^t < C(\eta'')^t,$$

where $P(t)$ is a polynomial with degree $\leqslant m$, and C and η'' are as stated.

Finally, (c) is a straightforward consequence of (a) and (b); after fixing $\eta'' \in (\eta', \eta)$, choose γ such that $(\eta'')^\gamma < \eta$. □

After these preparations, the proof of Theorem 5.14 is not difficult.

Proof of Theorem 5.14. Choose γ as in (c) of Lemma 5.16. We split

$$\mathbb{P}(\boldsymbol{\Sigma}_t = \boldsymbol{\mathcal{A}} \mid \tau > t) = \mathbb{P}(\boldsymbol{\Sigma}_t = \boldsymbol{\mathcal{A}}, \tau_{\mathcal{F}} > \gamma t \mid \tau > t) + \mathbb{P}(\boldsymbol{\Sigma}_t = \boldsymbol{\mathcal{A}}, \tau_{\mathcal{F}} \leqslant \gamma t \mid \tau > t),$$

The first probability tends to zero as $t \to \infty$, due to our choice of γ. The second can be rewritten as

$$\mathbb{P}(\boldsymbol{\Sigma}_t = \boldsymbol{\mathcal{A}} \mid \tau > t, \tau_{\mathcal{A}} \leqslant \gamma t) \cdot \mathbb{P}(\Sigma_t = \mathcal{A}, \tau_{\mathcal{F}} \leqslant \gamma t \mid \tau > t),$$

where we have used that Lemma 5.15 implies $\{\tau > t, \tau_{\mathcal{F}} \leqslant \gamma t, \Sigma_t = \mathcal{A}\} = \{\tau > t, \tau_{\mathcal{A}} \leqslant \gamma t\}$. Here, the second factor converges to $\mathbb{P}^{\Sigma}_{\mathrm{qlim}}(\mathcal{A})$ by the choice of γ and Lemma 5.16 (c).

Now consider the first factor. Together with $\tau > t$ and Lemma 5.15, $\tau_{\mathcal{A}} \leqslant \gamma t$ implies that $\Sigma_s = \mathcal{A}$ for all s between γt and t. During this period, the labels of the blocks of $\mathcal{A}$ evolve independently, and by the uniform convergence to the stationary distribution q, we obtain

$$\lim_{t\to\infty} \mathbb{P}(\boldsymbol{\Sigma}_t = \boldsymbol{\mathcal{A}} \mid \tau > t, \tau_{\mathcal{A}} \leqslant \gamma t) = \prod_{(d,\lambda)\in\boldsymbol{\mathcal{A}}} q(\lambda),$$

which completes the argument. For additional details, see the proof of Proposition 5.12. □

5.7 Recombination and migration in continuous time

Let us close by briefly discussing how our results carry over from the discrete-time to the continuous-time setting. We consider the deterministic migration-recombination equation in continuous time,

$$\dot{\omega}_t(\alpha) = \sum_{\beta\in L} N(\alpha,\beta)\omega_t(\beta) + \sum_{\delta\in\mathbb{S}([n])} (\mathcal{R}_\delta - \mathrm{id})\omega_t(\alpha). \tag{5.17}$$

This is just Eq. (2.5) for each local type distribution, together with an additional migration term. Instead of the stochastic backward migration matrix, we use a Markov generator N on L; its meaning is that between time t and $t + \mathrm{d}t$ and for $\alpha \neq \beta$, an individual at location α is replaced by an individual from location β, with probability $N(\alpha, \beta)\,\mathrm{d}t$; we assume that this happens independently of recombination.

The backward view can be easily adapted as follows. Again, we have an LPP (this time in continuous time) $\boldsymbol{\Sigma}^{\mathrm{c}} = (\boldsymbol{\Sigma}^{\mathrm{c}}_t)_{t\geqslant 0}$. It evolves as follows. At rate $\varrho_{\mathcal{B}}$ for all $\mathcal{B} \in \boldsymbol{P}(S)$, each labelled block A of $\boldsymbol{\Sigma}^{\mathrm{c}}_t$ is split (conditionally) independently into the blocks of the induced partition $\mathcal{B}|_A$; each of these fragments inherits the label from A. In addition and independently, for every $\alpha \in L$, the label of each block with label α is relabelled β at rate $N(\alpha, \beta)$. Somewhat more formally, $\boldsymbol{\Sigma}^{\mathrm{c}}$ is a Markov chain in continuous time with generator $\boldsymbol{\mathcal{Q}}$, defined by its nondiagonal elements

$$\boldsymbol{\mathcal{Q}}_{\boldsymbol{\mathcal{A}}\boldsymbol{\mathcal{B}}} = \begin{cases} \varrho^{A}_{\mathcal{A}|_A}, & \text{if } \boldsymbol{\mathcal{B}} = (\boldsymbol{\mathcal{A}} \setminus \{(A,\lambda)\}) \cup \mathcal{B}|_d \times \{\lambda\} \text{ for some } A \in \mathcal{A}, \\ N(\alpha,\beta), & \text{if } \boldsymbol{\mathcal{B}} = (\boldsymbol{\mathcal{A}} \setminus \{(A,\alpha)\}) \cup \{(A,\beta)\} \text{ for some } A \in \mathcal{A}, \\ 0, & \text{otherwise,} \end{cases}$$

where the marginal recombination rates $\varrho^{A}_{\mathcal{A}|_A}$ are defined in Theorem 2.4.

Note that, in the case without migration and with recombination restricted to single crossovers (that is, partitions of the form $\{\{1,\dots,i\},\{i+1,\dots,n\}\}$ for some $1 \leqslant i \leqslant n$), the continuous-time dynamics has a simple explicit solution, which is due to the fact that crossover events 'rain down' on sequences in an independent Poissonian fashion. See also [LPS] for the (much more involved) extension to the case with (a small amount of) coalescence in the limit of an infinitely large sequence.

But let us return to the full Equation (5.17). As before (compare Theorem 5.9), we have the duality relation

$$\mathcal{R}_{\boldsymbol{\mathcal{B}}}(\omega_t) = \mathbb{E}[\mathcal{R}_{\boldsymbol{\Sigma}^{\mathrm{c}}_t}(\omega_0) \mid \boldsymbol{\Sigma}^{\mathrm{c}}_0 = \boldsymbol{\mathcal{A}}],$$

whence we obtain the solution

$$\omega_t(\alpha) = \sum_{\boldsymbol{\mathcal{A}} \in \boldsymbol{L}\boldsymbol{P}(S)} (\mathrm{e}^{t\boldsymbol{\mathcal{Q}}})_{\underline{\mathbf{1}}^{\alpha}\boldsymbol{\mathcal{A}}} \mathcal{R}_{\boldsymbol{\mathcal{A}}}(\omega_0) \tag{5.18}$$

by solving the associated (linear) Kolmogorov backward equation, in perfect analogy to Theorem 5.7. The duality relation can be proved by a straightforward adaptation of the techniques in [BB16]. Indeed, [BB16] shows that the recombination part of Equation (5.17) is dual to the splitting (or branching) part of $\boldsymbol{\Sigma}^{\mathrm{c}}$. Showing that the migration part is dual to the random walk defined by N is a standard exercise.

Because Equation (5.18) is not very concrete, let us derive a more explicit solution formula for the special case $n = 2$. We give a probabilistic argument, analogous to Equation (5.8). First,

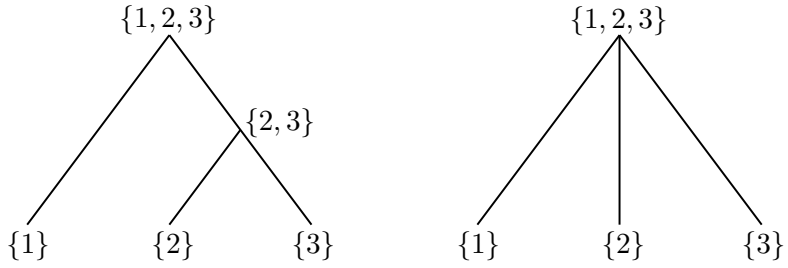


Figure 5.3. The two different tree topologies associated with recombination of three sites; on the left, first one site is separated, before the remaining block of size two is split. On the right, all sites are separated in one recombination event with three parents. It is not difficult to see that the corresponding contributions to the solution would consist of *two* iterated integrals for the left topology, and only one, as in Equation (5.19), for the right topology. Note that by permuting the sites, the left topology actually corresponds to three distinct contributions.

note that with probability $\mathrm{e}^{-\underline{\varrho}_0 t}$, both sites are not separated until time t, that is, $\Sigma_t^{\mathrm{c}} = \underline{1}$; the single block has performed a random walk with transition kernel N for the duration t. Hence, in this case, $\omega_t(\alpha) = (\mathrm{e}^{tN}\omega_0)_\alpha$. On the other hand, if the blocks have been split at time $\tau \in [0, t]$, then both sites have performed independent random walks, starting at time τ and at the location γ where the split took place. In that case, $\omega_t(\alpha) = (\mathrm{e}^{(t-\tau)N}\omega_0)_\gamma^{\{1\}} \otimes (\mathrm{e}^{(t-\tau)N}\omega_0)_\gamma^{\{2\}}$. Integrating over all possible values for τ (keeping in mind that τ is exponentially distributed with mean $\frac{1}{\underline{\varrho}_0}$) and γ (keeping in mind that, at the moment of splitting, the block has label γ with probability $(\mathrm{e}^{\tau N})_{\alpha\gamma}$), we obtain

$$\omega_t(\alpha) = \mathrm{e}^{-\underline{\varrho}_0 t}(\mathrm{e}^{tN}\omega_0)_\alpha + \underline{\varrho}_0 \sum_{\gamma \in L} \int_0^t \mathrm{e}^{-\underline{\varrho}_0 \tau}(\mathrm{e}^{\tau N})_{\alpha\gamma}(\mathrm{e}^{(t-\tau)N}\omega_0)_\gamma^{\{1\}} \otimes (\mathrm{e}^{(t-\tau)N}\omega_0)_\gamma^{\{2\}} \,\mathrm{d}\tau. \quad (5.19)$$

For more than two loci, one can proceed in a similar fashion, disintegrating the solution conditional on the waiting time(s) between splitting events. However, this becomes cumbersome very quickly as the number of contributions, coming from the different realisations of the jump chain of Σ^{c}, grows quickly in the number of sites. In addition, the form of these contributions varies qualitatively, depending on the associated tree topology; see Figure 5.3.

Remark 5.9. It is not difficult to adapt the partitioning process for finite populations as well as in the diffusive limit, as described in [BEP16], to the setting with migration. Put simply, stochastic resampling in the forward process leads to coalescences of blocks in the paritioning process. In the LPP, we have the additional condition that two blocks can only coalesce if they share the same label. However, an exhaustive treatment of the LPP with coalescence is beyond the scope of this work. ♢

6 Summary and Outlook

6.1 Summary

We have seen how probabilistic techniques can yield deep insight into the dynamics of deterministic models of population genetics with recombination. A recurring theme in our analysis was the interplay between the differential (or difference) equation models, forward in time, and their related genealogical processes, backward in time. This generalised the previously observed connection between the deterministic recombination equation and a stochastic partitioning process [BB16].

In Chapter 3, we saw that the pure recombination equation can —for finite sets of alleles— be understood as the law of mass action for a strongly reversible network of chemical reactions; in particular, it can be understood as a generalised gradient system. Regarding the backward-time perspective, we have seen how the monotonicity of the partitioning process implies the gradient-like evolution of its law. Finally, we have identified the nonlinear system of equations in [BBS16] for the coefficients in a suitable ansatz function with the law of mass-action for a network of reactions between partitions of the set of sequence sites; this network, however, turned out to be irreversible.

In Chapter 4, we presented a recursion for the solution of the selection-recombination equation with single-site selection. Starting from the solution of the pure selection equation, this recursion proceeds by successively adding in single crossovers, until we arrived at the full system. The proof was based on a variant of the ASRG [GM96; GM97], without coalescences. We boiled this —rather complex— process down to a weighted partitioning process, a variant of the partitioning process (WPP) with an integer weight signed to each block that represents the number of potential ancestors of the loci in that block. We then further simplified the matter by encoding the WPP as a collection of independent Yule (binary branching) processes with inititation and resetting (YPIR), by exploiting the assumption of single-crossover. The YPIR is a simple Markov chain in continuous time whose transition semigroup is available in closed form, yielding in turn a closed expression for the solution of the selection-recombination equation. For technical reasons, and to further elucidate the underlying structure, we introduced the initiation process which records the amount of time the selection term has acted on each site since the last recombination event. As another important tool , we introduced a non-commutative variant of the measure product, reflecting the different roles played by different parts of the sequence. Last not least, we applied our results to help clarify some issues regarding the time-evolution of linkage disequilibria in the

context of genetic hitchhiking. [SSL06].

Finally, in Chapter 5, we expressed the solution of the migration-recombination equation in terms of a *labelled* partitioning process (LPP), thus generalising the results in [BB16]. We analysed the long-term behaviour of the LPP, which in turn gave us information about the asymptotic behaviour of the deterministic equation, providing a new perspective on and strengthening a result from [Bür09]. We closed by computing the quasi-limiting distribution of this Markov chain.

6.2 Outlook

We have seen (compare Remark 4.2) that the recursive solution formula for the selection-recombination equation in Chapter 4 and the related duality result ultimately hinged on the right-multiplicativity of the selection term with respect to the aforemtioned non-commutative version of the measure product; this is a precise mathematical formulation of the idea that all sites other than the single selected site are 'inert' with respect to selection. We thus expect that our results carry over to the case with mutation and/or frequency dependent selection. It will be interesting to explore the corresponding dual processes, as well as their connections to properties of the forward model.

While deterministic models are an important part of population genetics, the bulk of recent research has been focussing on stochastic models. It would therefore be desirable to incorporate stochastic resampling into our approach. This will be major challenge to overcome as coalescences destroy the conditional independence of the ancestral lines on which our present approach relies. A possible starting point might be the regime where selection and recombination are strong compared to resampling. In this setting, a number of results [JFS15; BS16; JS12] establish the leading terms in an asymptotic expansion of the sampling distribution; it seems worthwhile to investigate if our methods can provide additional insight, in particular for the case with selection which was treated by Jenkins and Song [JS12], based on computations with the generator of the diffusion process forward in time.

Regardless of its connection via duality to the solution of the recombination equation, the partitioning process is an interesting object in its own right; recently, Schertzer et al. [LPS] considered a variant of this process on the positive half line. The authors derived an approximation of the stationary distribution, again in the strong coalescence regime. An intriguing feature is the formation of clusters of ancestral material; however, both the fine-scale structure of these clusters as well as the mechanism behind their formation does not seem to be well understood, and deserve further investigation. Last not least, the amount of trapped material, that is, non-ancestral material enclosed between segments of ancestral material, is of considerable interest in genetics.

Bibliography

[Aig79] M. Aigner, *Combinatorial Theory*, Springer, Berlin (1979).

[Aki74] E. Akin, Cycling in simple genetic systems, *J. Math. Biol.* **13** (1974), 305–324.

[Aki79] E. Akin, *The Geometry of Population Genetics*, Springer, New York (1979).

[Alb] F. Alberti, Genetic recombination as a generalised gradient flow, preprint; arXiv: 2009.09688.

[AB] F. Alberti and E. Baake, Solving the selection-recombination equation: Ancestral lines under selection and recombination, preprint; arXiv: 2003.06831.

[ABLM] F. Alberti, E. Baake, I. Letter and S. Martinez, Solving the migration-recombination equation from a genealogical point of view, preprint; arXiv: 2004.08958.

[ABHJ] B. Alberts, D. Bray, K. Hopkin, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walters, *Essential Cell Biology*, Garland Science, New York (2014).

[BB03] E. Baake and M. Baake, An exactly solved model for recombination, mutation and selection, *Can. J. Math.* **55** (2003), 3–41; arXiv: math/0210422, and erratum *Can. J. Math.* **60** (2008) 264.

[BB16] E. Baake and M. Baake, Haldane linearisation done right: Solving the nonlinear recombination equation the easy way, *Discr. Cont. Dyn. Syst. A* **36** (2016), 6645–6656; arXiv: 1606.05175.

[BB] E. Baake and M. Baake, Ancestral lines under recombination, to appear in: *Probabilistic Structures in Evolution*, E. Baake and A. Wakolbinger (eds.), EMS, Zürich.

[BBS16] E. Baake, M. Baake and M. Salamat, The general recombination equation in continuous time and its solution, *Discr. Cont. Dyn. Syst. A* **36** (2016), 63–95; arXiv: 1409.1378.

[BCH18] E. Baake, F. Cordero and S. Hummel, A probabilistic view on the deterministic mutation-selection equation: dynamics, equilibria, and ancestry via individual lines of descent, *J. Math. Biol.* **77** (2018), 795–820; arXiv: 1710.04573.

[BEP16] E. Baake, M. Esser and S. Probst, Partitioning, duality, and linkage disequilibria in the Moran model with recombination, *J. Math. Biol.* **73** (2016), 161–197; arXiv: 1502.05194.

[BW18] E. Baake and A. Wakolbinger, Lines of descent under selection, *J. Stat. Phys.* **172** (2018), 156–174; arXiv: 1710.08209.

[BW14] E. Baake and U. v. Wangenheim, Single-crossover recombination and ancestral recombination trees, *J. Math. Biol.* **68** (2014), 1371–1402; arXiv: 1206.0950.

[BR96] A. Bellaïche and J. J. Risler, *Sub-Riemannian Geometry*, Birkhäuser, Basel (1996).

[Ben54] J. H. Bennett, On the theory of random mating, *Ann. Human Gen.* **18** (1954), 311–317.

[BS16] A. Bhaskar and Y. S. Song, Closed-form asymptotic sampling distributions under the coalescent with recombination for an arbitrary number of loci, *Adv. Appl. Probab.* **21** (2016), 1–16; arXiv: 1107.4700.

[BP18] S. Bossert and P. Pfaffelhuber, The fixation probability and time for a doubly beneficial mutant, *Stoch. Proc. Appl.* **128** (2018), 4018–4050; arXiv: 1610.06613.

[Bür00] R. Bürger, *The Mathematical Theory of Selection, Recombination, and Mutation*, Wiley, Chichester (2000).

[Bür09] R. Bürger, Multilocus selection in subdivided populations I. Convergence properties for weak or strong migration, *J. Math. Biol.* **58** (2009), 939–978.

[Cas88] H. Caswell, Theory and models in ecology: A different perspective, *Ecol. Modell.* **43** (1988), 33–44.

[CMS13] P. Collet, S. Martínez and J. San Martín, *Quasi-stationary Distributions. Markov Chains, Diffusions and Dynamical Systems*, Springer, Berlin (2013).

[Cor17a] F. Cordero, Common ancestor type distribution: A Moran model and its deterministic limit, *Stoch. Proc. Appl.* **127** (2017), 590–621; arXiv: 1508.06113.

[Cor17b] F. Cordero, The deterministic limit of the Moran model: a uniform central limit theorem, *Markov Proc. Relat. Fields* **23** (2017), 313–324; arXiv: 1508.05231.

[Dar59] C. Darwin, *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, Murray, London (1859).

[DK99] P. Donnelly and T. G. Kurtz, Genealogical processes for Fleming-Viot models with selection and recombination, *Ann. Appl. Probab.* **9** (1999), 1091–1148.

[Dur08] R. Durrett, *Probability Models for DNA Sequence Evolution*, 2nd ed., Springer, New York (2008).

[Ess16] M. Esser, *Recombination Models Forward and Backward in Time*, Dissertation, Universität Bielefeld, 2016; urn: nbn:de:0070-pub-29102790.

[Eth11] A. M. Etheridge, *Some Mathematical Models from Population Genetics: École D'Été de Probabilités de Saint-Flour XXXIX-2009*, Springer, Heidelberg (2011).

[EK86] S. N. Ethier and T. G. Kurtz, *Markov Processes: Characterization and Convergence*, Wiley, New York (1986; reprint 2005).

[Ewe04] W. J. Ewens, *Mathematical Population Genetics I: Theoretical Introduction*, Springer, New York (2004).

[ET77] W. J. Ewens and G. Thomson, Properties of equilibria in multi-locus genetic systems, *Genetics* **87** (1977), 807–819.

[Fei19] M. Feinberg, *Foundations of Chemical Reaction Network Theory*, Springer, Cham (2019).

[Fel68] W. Feller, *An Introduction to Probability Theory and its Applications, Vol. I, 3rd ed.* Wiley, New York (1968).

[Fel51] W. Feller, Diffusion processes in genetics, in: *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (1950), J. Neyman (ed.), Univ. of Calif. Press, Berkeley (1951), 227–246.

[Fis30] R. A. Fisher, *The Genetical Theory of Natural Selection*, Clarendon Press, Oxford (1930).

[Gei44] H. Geiringer, On the probability theory of linkage in Mendelian heredity, *Ann. Math. Stat.* **15** (1944), 25–57.

[GM96] R. C. Griffiths and P. Marjoram, Ancestral inference from samples of DNA sequences with recombination, *J. Comput. Biol.* **3** (1996), 479–502.

[GM97] R. C. Griffiths and P. Marjoram, An ancestral recombination graph, in: *Progress in Population Genetics and Human Evolution*, P. Donnelly and S. Tavaré (eds.), Springer, New York (1997), 257–270.

[Har78] T. E. Harris, Additive set-valued Markov processes and graphical methods, *Ann. Probab.* **6** (1978), 355–378.

[Hof85] J. Hofbauer, The selection-mutation equation, *J. Math. Biol.* **23** (1985), 41–53.

[Hof17] J. Hofbauer, Population dynamics and reaction systems — some crossovers, *Oberwolfach Reports* **28** (2017), 1753–1756.

[HM15] J. Hofbauer and S. Müller, Genetic recombination as a chemical reaction network, *Math. Model. Nat. Phenom.* **10** (2015), 84–99; arXiv: 1503.01155.

[Hud83] R. R. Hudson, Properties of a neutral allele model with intragenic recombination, *Theor. Popul. Biol.* **23** (1983), 183–201.

[JK14] S. Jansen and N. Kurt, On the notion(s) of duality for Markov processes, *Probab. Surveys* **11** (2014), 59–120; arXiv: 1210.7193.

[JFS15] P. A. Jenkins, P. Fearnhead and Y. S. Song, Tractable diffusion and coalescent processes for weakly correlated loci, *Electron. J. Probab.* **20** (2015), 1–26; arXiv: 1405.6863.

[JS12] P. A. Jenkins and Y. S. Song, Padé approximants and exact two-locus sampling distributions, *Ann. Appl. Probab.* **22** (2012), 576–607; arXiv: 1107.3897.

[Jen17] H. S. Jennings, The numerical results of diverse systems of breeding, with respect to two pairs of characters, linked or independent, with special relation to the effects of linkage, *Genetics* **2** (1917), 97–154.

[KT75] S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes*, Academic Press, San Diego, CA (1975).

[Kim65] M. Kimura, A model of a genetic system which leads to closer linkage by natural selection, *Evolution* **10** (1965), 278–287.

[Kin82a] J. F. C. Kingman, Exchangeability and the evolution of large populations, in: *Exchangeability in Probability and Statistics. Proceedings of the International Conference on Exchangeability in Probability and Statistics, Rome, 6th–9th April, 1981, in honour of Professor Bruno de Finetti*, G. Koch and F. Spizzichino (eds.), Elsevier, Amsterdam (1982), 97–112.

[Kin82b] J. F. C. Kingman, The coalescent, *Stoch. Process. Appl.* **13** (1982), 235–248.

[KN97] S. M. Krone and C. Neuhauser, Ancestral processes with selection, *Theor. Popul. Biol.* **51** (1997), 210–237.

[Kur70] T. Kurtz, Solutions of ordinary differential equations as limits of pure jump Markov processes, *J. Appl. Probab.* **7** (1970), 49–58.

[LPS] A. Lambert, V. Miró Pina and E. Schertzer, Chromosome painting: how recombination mixes ancestral colors, preprint; arXiv: 1807.09116.

[LK12] S. Lessard and A. R. Kermany, Fixation probability in a two-locus model by the ancestral recombination-selection graph, *Genetics* **190** (2012), 691–707.

[Lig10] T. M. Liggett, *Continuous Time Markov Processes: an Introduction*, Amer. Math. Soc., Providence, RI (2010).

[Lyu92] Y. I. Lyubich, *Mathematical Structures in Population Genetics*, Springer, Berlin (1992).

[Mal48] G. Malécot, *Les Mathématiques de l'Hérédité*, Masson, Paris (1948).

[Mar17] S. Martínez, A probabilistic analysis of a discrete-time evolution in recombination, *Adv. Appl. Math.* **91** (2017), 115–136; arXiv: `1603.07201`, and erratum *Adv. Appl. Math.* **91** (2017) 115-136.

[MW06] F. A. Matsen and J. Wakeley, Convergence to the island-model coalescent process in populations with restricted migration, *Genetics* **172** (2006), 701–708.

[MR83] D. McHale and G. A. Ringwood, Haldane linearisation of baric algebras, *J. London Math. Soc.* **28** (1983), 17–26.

[Men66] G. Mendel, Versuche über Pflanzen-Hybriden, *Verhandlungen des naturforschenden Vereines in Brünn Bd. IV für das Jahr 1865* (1866), 3–47.

[Mie11] A. Mielke, A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems, *Nonlinearity* **24** (2011), 1329–1346.

[Möh99] M. Möhle, The concept of duality and applications to Markov processes arising in neutral population genetics models, *Bernoulli* **5** (1999), 761–777.

[Möh13] M. Möhle, Duality and cones of Markov processes and their semigroups, *Markov Process. Relat. Fields* **19** (2013), 149–162.

[Mor58] P. A. P. Moran, Random processes in genetics, *Math. Proc. Cam. Phil. Soc.* **54** (1958), 60–71.

[Mor11] T. H. Morgan, Random segregation versus coupling in Mendelian inheritance, *Science* **34** (1911), 384–384.

[Nag92] T. Nagylaki, *Introduction to Theoretical Population Genetics*, Springer, Berlin (1992).

[NHB99] T. Nagylaki, J. Hofbauer and P. Brunovský, Convergence of multilocus systems under weak epistasis or weak selection, *J. Math. Biol.* **38** (1999), 103–133.

[Not90] M. Notohara, The coalescent and the genealogical process in geographically structured populations, *J. Math. Biol.* **29** (1990), 59–75.

[PLS08] P. Pfaffelhuber, A. Lehnert and W. Stephan, Linkage disequilibrium under genetic hitchhiking in finite populations, *Genetics* **179** (2008), 527–537.

[Rob18] R. B. Robbins, Some applications of mathematics to breeding problems III, *Genetics* **3** (1918), 375–389.

[SW05] P. F. Slade and J. Wakeley, The structured ancestral selection graph and the many-demes limit, *Genetics* **169** (2005), 1117–1131.

[SH74] J. M. Smith and J. Haigh, The hitch-hiking effect of a favourable gene, *Gen. Res.* **23** (1974), 23–35.

[SSL06] W. Stephan, Y. S. Song and C. H. Langley, The hitchhiking effect on linkage disequilibrium between linked neutral loci, *Genetics* **172** (2006), 2647–2663.

[SO94] A. Stuart and K. J. Ord, *Kendall's Advanced Theory of Statistics*, Wiley, Chichester (1994).

[Wak04] J. Wakeley, Recent trends in population genetics: More data! More math! Simple models?, *J. Hered.* **95** (2004), 397–405.

[Wak09] J. Wakeley, *Coalescent Theory: An Introduction*, Freeman, New York (2009).

[Wal04] G. Walschap, *Metric Structures in Differential Geometry*, Springer, New York (2004).

[Wri31] S. Wright, Evolution in Mendelian populations, *Genetics* **16** (1931), 97–159.

[Yon12] W. Yong, Conservation-dissipation structure of chemical reaction systems, *Phys. Rev. E* **86** (2012), 1–3.