Muller's Ratchet with Compensatory Mutations

Diplomarbeit

von

Paul R. Staab

betreut von

Professor Dr. Peter Pfaffelhuber

Version mit kleineren Korrekturen

Fakultät für Mathematik und Physik
Albert-Ludwigs-Universität Freiburg im Breisgau
August 2011

Contents

1. Introduction

> "I will not attempt here to predict from theory the quantitative effect of the ratchet mechanism. Involving natural selection, mutation, and genetic drift at many linked loci, the problem poses enormous difficulties for the application of population genetics theory. But the possible significance of the phenomenon makes it important that some theoretical treatment should be attempted. The ratchet mechanism has been unjustly ignored by theoretical population genetics." Felsenstein [1974]

While the problems Joseph Felsenstein saw arising from a mathematical model in population genetics back in 1974 have not been solved to the present day, *Muller's ratchet* has in the meanwhile gotten the attention of biomathematicians and has become a well-known model in population genetics.

The importance he claimed it has is given by the biological question to which Muller's ratchet is a possible answer:

> Why has sex evolved?

From an evolutionary point of view, the step from asexual to sexual reproduction is a rather harsh one, as the population suddenly needs to spend a significant part of its resources on individuals that only contribute a minor part to reproduction: males. Population geneticists have therefore searched for an advantage that justifies this effort and have found a biological phenomenon called (sexual) recombination: During meiosis, a process that is part of the production of gametes, either the maternal or the paternal version of every chromosome is transferred to the gamete. As this happens independently for each chromosome, (nearly) all gametes contain a mixture of the genomes of their producer's parents.[1]

It has already been suggested by Fisher [1930] and Muller [1932] that the possibility of recombinations may be an advantage of sexual reproduction because beneficial mutations – small spontaneous changes in the genome that increase the "fitness" of an individual – can be combined in a later descendant, even if they occur in different individuals. While the importance of this effect was discussed heavily[2], Muller [1964] gave another argument looking at mutations that decrease an individuals fitness: Assume

[1] To be precise, recombination refers to the "shuffling" of genetic material in general. The described phenomenon is just one form of recombination, the so called interchromosomal recombination. While there are other important mechanisms leading to recombinations (e.g. crossing-overs), it suffices for our purpose to imagine it in the described way.

[2] See Felsenstein [1974] for a summary.

that such deleterious mutations constantly occur inside a population. As genomes are typically large, there is only a very small chance that two different mutations will occur in the very same place in the genome. Hence, the reversion of one mutation by another one is so unlikely that we can ignore this effect. Without recombination, an individual passes its whole genome – including all mutations – to all its offspring. Thus, the total number of mutations in the population can decrease only if individuals with many mutations have no descendants. However, once all mutants have by chance acquired one mutation, there will never be an individual free of mutations again. As this will happen again and again, the population will gather more and more deleterious mutations and will finally face extinction. Because the accumulation of mutations happens in an irreversible way, like the "clicks" of a ratchet moving forward notch by notch, this argument became known as Muller's ratchet.

This argument was qualitatively accepted in the following years and has by now become a textbook model for the advantage of sexual reproduction in the biological literature. Even though Joseph Felsenstein proved to be right in the difficulties he had seen in determining the quantitative effect of Muller's ratchet. Up to now the literature concerned with Muller's ratchet relies on computer simulations and approximating models to determine how fast the fitness of a population decreases (e.g. Haigh [1978], Lynch and Gabriel [1990], Stephan et al. [1993], Gessler [1995], Gordo and Charlesworth [2000], Etheridge et al. [2008]). Though this is the key question in the analysis of Muller's Ratchet, an exact determination of this "click rate" still seems to be out of reach.

Along with other extensions of the classical model, like a diploid version of Muller's ratchet (Pamilo et al. [1987], Charlesworth and Charlesworth [1997]) and a model with variable population size (Gabriel et al. [1993]), the behavior of Muller's ratchet with so called *compensatory mutations* has been discussed in the past. In contrast to the normal model, we relax the above mentioned assumption that deleterious mutations cannot be compensated by a second mutation. Hence, there is a small chance for the occurrence of a back-mutation. We assume that the probability for such a compensatory mutation is γ for every malicious mutation in the genome. However, deleterious mutations typically arise with a rate of λ per genome, whereas the rate of compensatory mutations only scales with the number of acquired mutations. Haigh [1978] concluded that compensatory mutations occur far too rarely to counter the ratchet effect. However, as the rate of compensatory mutations increases with the number of acquired deleterious mutations – while the rate of occurrence of the latter is constant – it seems reasonable that the ratchet will eventually reach an equilibrium state if compensatory mutations are present.

It has been discussed if this equilibrium can be reached by natural populations, or if it lies above the maximal number of deleterious mutations that a population can bear. On the one hand, Antezana and Hudson [1997] have argued that the former could be important for small viruses. On the other hand, Loewe [2006] concluded that the latter is likely for human mitochondria and Chao [1990] and Smith [1978] claimed that compensatory mutations occur far too rarely to have a significant effect in a population undergoing Muller's ratchet. However, new insights in molecular biology in the past decades have shown that (back-)mutation is not the only mechanism in living cells that can compensate the effects of unfavorable mutations. Maisnier-Patin and Andersson [2004] suggested at least five different ways how this could theoretically happen. Poon and Chao [2005] found experimental evidence of such compensations in the DNA bacteriophage φX174. Maier et al. [2008] and Depperschmidt et al. [2011] studied the plasmid genomes of mosses and concluded that deleterious mutations are compensated by RNA editing, a couple of mechanisms by which specific bases in the genome are altered after the sequence has been transcribed from DNA to RNA.

Summarizing all these mechanisms, it is reasonable to assume that the rate of compensatory mutations can be high enough to affect a natural population. Therefore we extend the classical model of Muller's ratchet by compensatory mutations and extend most of the known results about the ratchet to that case. In particular, we show that the ratchet with compensatory mutations has exactly one equilibrium state in which the malicious effects of mutations on one side cancels out with the beneficial effects from purifying selection and compensatory mutations on the other (Corollaries 4.20 & 4.26). Furthermore, the population always reaches this equilibrium as the time goes towards infinity (also Corollary 4.26). To achieve these results, we start with some necessary topological preliminaries (Chapter 2). Thereafter, we state the corresponding results for the classical ratchet, where they can be obtained in a time-discrete model, mostly due to the work of Maia et al. [2003] (Chapter 3). In Chapter 4, we introduce compensatory mutations and start with a brief look at an extended version of Haigh's time-discrete model (Chapter 4.1). Afterwards we move on to time-continuous models. Here we will use the diffusion approximation of Muller's ratchet, which is based on a stochastic differential equation (SDE). We translate the SDE into a martingale problem and show that it uniquely defines a stochastic process (Chapter 4.2). Afterwards we increase the population size towards infinity, were the SDE turns into an ordinary differential equation (ODE). We solve the ODE and derive the mentioned results (Chapter 4.3). Finally, we present computer simulations to see how large the "finite population effect" is for various combinations of parameters (Chapter 5).

2. Preliminaries

The mathematical modeling of Muller's ratchet is typically realised as a Wright-Fisher model with countable infinite different (geno-)types. Hence, it is a Fleming-Viot process that takes values in the space of probability measure $\mathcal{P}(\mathbb{N})$ on the natural numbers $\mathbb{N} = \{0, 1, 2, \ldots\}$. We will often refer to this space as the infinite dimensional simplex $\mathbb{S}$, defined as

$$\mathbb{S} := \left\{ (x_0, x_1, \ldots) \in [0,1]^{\mathbb{N}} : \sum_{i \in \mathbb{N}} x_i = 1 \right\}.$$

The identification of both spaces is made via the one-to-one relationship

$$\sum_{i \in \mathbb{N}} p_i \delta_i \mapsto (p_i)_{i \in \mathbb{N}},$$

where δ_i denotes the Dirac measure on i and p_i is the corresponding point weight. We start by deriving some necessary topological preliminaries about the simplex $\mathbb{S}$.

Notation. Throughout this manuscript we will denote vectors $\underline{x} \in \mathbb{S}$ with an underline and will refer to its i-th coordinate x_i with a subscript i.

At first we show that $\mathbb{S}$ is complete and separable. Therefore, we equip $\mathbb{N}$ with the discrete metric

$$d_{\mathbb{N}}(n, m) = \begin{cases} 0 & \text{if } n = m, \\ 1 & \text{otherwise.} \end{cases}$$

Thus $\mathbb{N}$ becomes complete and separable. It follows from standard theory that $\mathbb{S}$ with Prohorov metric d_P is complete and separable as well. Moreover, we can explicitly calculate d_P in this case.

Lemma 2.1. *The Prohorov metric on $\mathbb{S}$ is equal to the total variation metric. That is, for $\underline{x}, \underline{y} \in \mathbb{S}$,*

$$d_P\left(\underline{x}, \underline{y}\right) = \sum_{i \in \mathbb{N}} \frac{|x_i - y_i|}{2}.$$

Proof. Let $\underline{x}, \underline{y} \in \mathbb{S}$. For a given set $F \subset \mathbb{N}$ denote its open ε-hull with

$$F^{\varepsilon} = \{ n \in \mathbb{N} : d_{\mathbb{N}}(n, m) < \varepsilon \quad \text{for a } m \in F \}.$$

As $F^\varepsilon = F$ for $\varepsilon \leq 1$, we have

$$\begin{aligned} d_P\left(\underline{x}, \underline{y}\right) &:= \inf\{\varepsilon > 0 : \sum_{i\in\mathbb{N}} x_i \delta_i(F) \leq \sum_{i\in\mathbb{N}} y_i \delta_i(F^\varepsilon) + \varepsilon \quad \text{for all } F \subset \mathbb{N}\} \\ &= \inf\{\varepsilon > 0 : \sum_{i\in\mathbb{N}} (x_i - y_i)\delta_i(F) \leq \varepsilon \quad \text{for all } F \subset \mathbb{N}\}. \end{aligned}$$

The next step is to verify that

$$d_P\left(\underline{x}, \underline{y}\right) = \sum_{i\in\mathbb{N}} (x_i - y_i)\mathbb{1}_{\{x_i > y_i\}}.$$

For "$\leq$", observe that

$$\sum_{i\in\mathbb{N}} (x_i - y_i)\delta_i(F) \leq \sum_{i\in\mathbb{N}} (x_i - y_i)\mathbb{1}_{\{i\in F,\, x_i > y_i\}} \leq \sum_{i\in\mathbb{N}} (x_i - y_i)\mathbb{1}_{\{x_i > y_i\}}$$

for all $F \subset \mathbb{N}$. For "$\geq$", assume that there exists $\tilde{\varepsilon}$ with

$$\sum_{i\in\mathbb{N}} (x_i - y_i)\delta_i(F) \leq \tilde{\varepsilon} < \sum_{i\in\mathbb{N}} (x_i - y_i)\mathbb{1}_{\{x_i > y_i\}}$$

for all $F \subset \mathbb{N}$. This is a contradiction for $F = \{i \in \mathbb{N} : x_i > y_i\}$. Now because

$$\sum_{i\in\mathbb{N}} (x_i - y_i)\mathbb{1}_{\{x_i > y_i\}} = \sum_{i\in\mathbb{N}} (x_i - y_i)(1 - \mathbb{1}_{\{y_i > x_i\}}) = \sum_{i\in\mathbb{N}} (y_i - x_i)\mathbb{1}_{\{y_i > x_i\}},$$

we get

$$\sum_{i\in\mathbb{N}} |x_i - y_i| = \sum_{i\in\mathbb{N}} (x_i - y_i)\mathbb{1}_{\{x_i > y_i\}} + (y_i - x_i)\mathbb{1}_{\{y_i > x_i\}} = 2\sum_{i\in\mathbb{N}} (x_i - y_i)\mathbb{1}_{\{x_i > y_i\}}$$

□

We will later see that the distributions which have finite exponential moments are of crucial importance for Muller's ratchet, as they will turn out to be its 'natural' state space. Therefore we define the subspace $\mathbb{S}^\circ$ of such distributions as well as a metric d_P° that makes $\mathbb{S}^\circ$ complete and separable.

Definition 2.2. For $\xi \in \mathbb{R}$, we define the ξ-exponential moment h_ξ of a discrete probability distribution $\underline{x} \in \mathbb{S}$ as

$$h_\xi(\underline{x}) := \sum_{k=0}^{\infty} x_k e^{\xi k}$$

and define the subspace $\mathbb{S}^\circ$ as

$$\mathbb{S}^\circ := \{\underline{x} \in \mathbb{S} : h_\xi(\underline{x}) < \infty \text{ for all } \xi \in \mathbb{R}\}.$$

For $x, y \in \mathbb{S}$ we define

$$d_P^\circ(\underline{x}, \underline{y}) := d_P(\underline{x}, \underline{y}) + \int_0^\infty e^{-t}(1 \wedge \sup_{0 \le \xi \le t} |h_\xi(\underline{x}) - h_\xi(\underline{y})|)\, dt. \tag{2.1}$$

The idea to use d_P° is taken from Ethier and Shiga [2000]. The next Lemma shows that $(\mathbb{S}^\circ, d_P^\circ)$ is indeed Polish.

Lemma 2.3.

1. *d_P° is a metric on $\mathbb{S}^\circ$.*
2. *Let $\underline{x} \in \mathbb{S}$ and $\underline{x}^1, \underline{x}^2, \ldots \in \mathbb{S}^\circ$. Then $d_P^\circ(\underline{x}, \underline{x}^n) \to 0$ iff $d_P(\underline{x}, \underline{x}^n) \to 0$ and $\sup_n h_\xi(\underline{x}^n) < \infty$ for all $\xi \ge 0$.*
3. *The metric space $(\mathbb{S}^\circ, d_P^\circ)$ is complete and separable.*

Proof. 1. This is obvious.
2. For "$\Rightarrow$", $d_P(\underline{x}, \underline{x}^n) \to 0$ follows directly as both summands in (2.1) are non-negative. Assume that $\sup_n h_\xi(\underline{x}^n) = \infty$ for some $\xi \in \mathbb{R}$. Then there exists a subsequence $\underline{x}^{n_k}$ with $|h_\xi(\underline{x}) - h_\xi(\underline{x}^{n_k})| \ge 1$ as $h_\xi(\underline{x})$ is finite. Hence

$$d_P^\circ(\underline{x}, \underline{x}^{n_k}) \ge \int_\xi^\infty e^{-t} dt > 0,$$

which is a contradiction to $d_P^\circ(\underline{x}, \underline{x}^n) \to 0$.
For "$\Leftarrow$", notice that $h_\xi(\underline{x})$ is the composition of a power series with e^ξ and therefore is continuous in ξ for $\underline{x} \in \mathbb{S}^\circ$. Hence it suffices to show that $h_\xi(\underline{x}^n) \to h_\xi(\underline{x})$ as $n \to \infty$ for all $\xi \ge 0$. So let $\varepsilon > 0$. We first show that there is an N such that $\sup_n \sum_{k=N}^\infty x_k^n e^{\xi k} < \varepsilon$. Otherwise for all N there would be an n with $\sum_{k=N}^\infty x_k^n e^{\xi k} \ge \varepsilon$. As

$$h_{\xi+1}(\underline{x}^n) \ge \sum_{k=N}^\infty x_k^n e^{(\xi+1)k} \ge e^N \sum_{k=N}^\infty x_k^n e^{\xi k} \ge e^N \xrightarrow{N \to \infty} \infty$$

that would be a contradiction to $sup_n h_{\xi+1}(\underline{x}^n) < \infty$. Hence we have as well

$$\sup_n \sum_{k=N}^\infty x_k e^{\xi k} < \varepsilon.$$

As $d_P(\underline{x}, \underline{x}^n) \to 0$, we have

$$\lim_{n\to\infty} |h_\xi(\underline{x}^n) - h_\xi(\underline{x})| \le \lim_{n\to\infty} \sum_{k=0}^{N-1} |x_k^n - x_k| e^{\xi k} + \sum_{k=N}^\infty x_k^n e^{\xi k} + \sum_{k=N}^\infty x_k e^{\xi k} < 2\varepsilon.$$

3. For separability, note that every $\underline{x} \in \mathbb{S}^\circ$ can be approximated by probability vectors with entries in $\mathbb{Q}$ and only finitely many non-zero entries. For completeness, take a Cauchy-sequence $\underline{x}^1, \underline{x}^2, ...$ with respect to d_P°. Since the sequence is also Cauchy with respect to the complete metric d_P, there is $\underline{x} \in \mathbb{S}$ with $d_P(\underline{x}^n, \underline{x}) \xrightarrow{n\to\infty} 0$. Moreover, $(h_\xi(\underline{x}^n))_{n=1,2,...}$ is Cauchy in $\mathbb{R}_+$ by construction, so $\sup_n h_\xi(\underline{x}) < \infty$ and so $d_P^\circ(\underline{x}^n, \underline{x}) \to 0$ as $n \to \infty$ by 1. □

We now turn to a certain class of functions on $\mathbb{S}$, which we will later need as a domain for the generator when we define Muller's ratchet via a martingale problem.

Definition 2.4.

1. For a complete and separable metric space (E, r), we denote the class of measurable functions from E to $\mathbb{R}$ with $\mathcal{M}(E)$ and the class of bounded functions in $\mathcal{M}(E)$ with $\mathcal{B}(E)$. Furthermore, we say a function $f \in \mathcal{M}(E)$ is *exponentially bounded* if there exist constants $C, \xi > 0$ such that

$$|f(x)| < Ce^{\xi x} \quad \text{for all } x \in E. \tag{2.2}$$

   We denote the class of such functions with $\mathcal{B}^e(E)$.

2. A class of functions $\mathcal{F} \subseteq \mathcal{M}(E)$ is said to be an *algebra*, if $1 \in \mathcal{F}$ and for all $\alpha, \beta \in E$ and $f, g \in \mathcal{F}$ also $\alpha f + \beta g \in \mathcal{F}$ and $fg \in \mathcal{F}$. We denote the algebra generated by a set $\mathcal{F} \subseteq \mathcal{M}(E)$ with $\mathcal{A}(\mathcal{F})$.

3. For $\varphi_1, \ldots, \varphi_n \in \mathcal{M}(\mathbb{N})$ we define $f_{\varphi_1,\ldots,\varphi_n} : \mathbb{R}^\mathbb{N} \to \mathbb{R}$ by

$$f_{\varphi_1,\ldots,\varphi_n}(\underline{x}) := \langle \underline{x}, \varphi_1 \rangle \cdots \langle \underline{x}, \varphi_n \rangle \quad \text{with} \quad \langle \underline{x}, \varphi_i \rangle := \sum_{k\in\mathbb{N}} x_k \varphi_i(k).$$

   Talking about $f_{\varphi_1,\ldots,\varphi_n}$ as functions on $\mathbb{S}$ or $\mathbb{S}^\circ$, we canonically refer to the restrictions of $f_{\varphi_1,\ldots,\varphi_n}$ to those spaces.

4. We define the algebra $\overline{\mathcal{F}}$ on $\mathbb{R}^\mathbb{N}$ by

$$\overline{\mathcal{F}} := \mathcal{A}(\{f_{\varphi_1,\ldots,\varphi_n} : n \in \mathbb{N}, \varphi_i \in \mathcal{B}^e(\mathbb{N}) \text{ for } i = 1, \ldots, n\}).$$

While it is quite obvious that every $f \in \overline{\mathcal{F}}$ is continuous on $(\mathbb{S}, d_P)$, it is a bit harder to see that this remains true for the restrictions to $(\mathbb{S}^\circ, d_P^\circ)$. As this will be important later, we state the following Lemma.

Lemma 2.5. *Every $f \in \overline{\mathcal{F}}$ is continuous on $\mathbb{S}^\circ$ with respect to the topology generated by the metric d_P°.*

Proof. It suffices to show the assertion for the function $\langle \cdot, \varphi \rangle$ with $0 \leq \varphi \in \mathcal{B}^e(\mathbb{N})$. Let C and ξ be the ones from (2.2) and $\underline{x}, \underline{x}^1, \underline{x}^2, ... \in \mathbb{S}^\circ$ with $r^\circ_{TV}(\underline{x}^n, \underline{x}) \xrightarrow{n\to\infty} 0$. Then, by the definition of d°_P, $\sum_{k=0}^\infty x_k^n C e^{\xi k} \xrightarrow{n\to\infty} \sum_{k=0}^\infty x_k C e^{\xi k}$. Thus,

$$\sum_{k=0}^\infty x_k^n \varphi(k) = \int_0^\infty \sum_{k=0}^\infty \mathbb{1}_{\{\varphi(k)>t\}} x_k^n dt \xrightarrow{n\to\infty} \int_0^\infty \sum_{k=0}^\infty \mathbb{1}_{\{\varphi(k)>t\}} x_k dt = \sum_{k=0}^\infty x_k \varphi(k)$$

by Fubini and dominated convergence, [Kallenberg, 1997, Theorem 1.21], since d_P metrizes weak convergence on $\mathcal{P}(\mathbb{N})$. $\square$

Finally, we will need another criteria under which continuous functions on $(\mathbb{S}, d_P)$ are also continuous on $(\mathbb{S}^\circ, d^\circ_p)$.

Lemma 2.6. *Let $t \mapsto \underline{x}(t)$ take values in $\mathbb{S}^\circ$, be continuous with respect to the topology generated by d_P and $\sup_{0\leq t\leq T} \sum_{k=0}^\infty x_k(t) e^{\xi k} < \infty$ for all $\xi \geq 0$ and $T > 0$. Then, $t \mapsto \underline{x}(t)$ is continuous with respect to the topology generated by d°_P.*

Proof. It suffices to show that

$$\lim_{s\to t} \sum_{k=0}^\infty |x_k(s) - x_k(t)| e^{\xi k} = 0$$

for all $t \geq 0$ and $\xi \geq 0$. Let $T > 0$ and choose a $K \in \mathbb{R}_+$ such that

$$\sup_{0\leq t\leq T} \sum_{k=0}^\infty x_k(t) e^{2\xi k} \leq K.$$

Then, necessarily, $x_k(t) e^{\xi k} \leq K e^{-\xi k}$ for all $k \in \mathbb{N}$ and $0 \leq t \leq T$. For $\varepsilon > 0$ choose $m \in \mathbb{N}$ such that $\sum_{k=m+1}^\infty K e^{-\xi k} \leq \varepsilon/2$. Then,

$$\lim_{s\to t} \sum_{k=0}^\infty |x_k(s) - x_k(t)| e^{\xi k} \leq \lim_{s\to t} \sum_{k=0}^m |x_k(s) - x_k(t)| e^{\xi k} + \sum_{k=m+1}^\infty (x_k(s) + x_k(t)) e^{\xi k} \leq \varepsilon.$$

Letting $\varepsilon \to 0$ gives the assertion. $\square$

3. Muller's Ratchet

Before we introduce compensatory mutations into the classical model of Muller's ratchet, we give a short review of some of the literature about Muller's ratchet. Thereby, we will focus on the time-discrete stochastic model that was introduced by Haigh [1978] and its deterministic large population limit.

As dealing with the discrete systems described in this chapter is delicate, we will later work primarily with time-continuous models. However, the results we will derive for the ratchet with compensatory mutations are similar to the ones we present here. Hence this chapter can be treated as a gentle introduction to the next one.

We start this chapter by defining the mathematical model of Muller's ratchet. We discuss its behaviour and turn our attention towards the so-called equilibrium points afterwards. These are equilibrium states in which the opposing evolutionary forces of mutation and selection cancel out each other. As the main result in this chapter, we state all of these points (Theorem 3.3, p. 15). While it is not hard to show that these points indeed fulfill the equilibrium condition, we need the mentioned large population limit approximation to show that no other equilibria exist. In contrast to the normal ratchet, this approximation can be explicitly given (Theorem 3.5, p. 17). Afterwards it is easy to prove the uniqueness of the equilibrium points (Corollary 3.6, p. 19).

3.1. Haigh's Model

The mathematical model of Muller's ratchet was introduced by Haigh [1978]. It describes a population that evolves according to Wright-Fisher dynamics and is affected by the evolutionary forces of *mutation*, *selection* and *genetic drift*.

For mathematical simplicity, we assume that every generation consists of exactly N individuals. When going from one, say the t-th, generation to the next, each of the individuals of the $(t+1)$-th generation chooses a parent independently of all others. The random fluctuations that arise from this sampling are called *genetic drift*.

The probability that a parent is picked decreases with each mutation in its genome. We refer to this probability as the "fitness" of an individual and assume that it is proportional to $(1-\alpha)^k$ for an individual with k mutations, where α is a coefficient that determines the strength of *selection*.

In addition to the mutations an offspring inherits from its parent, new *mutations* may appear throughout its life. We assume that this happens at a constant rate. Therefore we take the number of new mutations an individual acquires in a generation to be Poisson distributed with parameter λ. As written before, we ignore that a mutation can theoretically be compensated by another mutation and refer to the number of mutations of an individual as its “type”.

To model these forces, we take $\underline{X}(t) = (X_0(t), X_1(t), \ldots)$ to be a time discrete, $\mathbb{S}$-valued stochastic process that states the empirical distribution of the types for every generation $t \in \mathbb{N}$. Thus, $X_k(t)$ is the percentage of individuals that carry exactly k mutations in generation t. Now given $\underline{X}(t)$, we calculate $\underline{X}(t+1)$ in three steps:

(i) First, we let each of the N offspring choose a parent independently of each other and according to the fitness of the parents. Therefore, we model the types of the N descendants by i.i.d. random variables $H_1(t), \ldots, H_N(t)$ with

$$\mathbf{P}[H_1(t) = k \mid \underline{X}(t)] = \frac{(1-\alpha)^k X_k(t)}{W(t)}$$

where

$$W(t) = \sum_{i \in \mathbb{N}} X_k(t)\,(1-\alpha)^i$$

is the mean fitness of the population in generation t.

(ii) Second, we add a $\mathrm{Poi}\,(\lambda)$-distributed number J_i of new mutations to the type of offspring i, where $J_1(t), \ldots, J_N(t)$ are again independent of each other and furthermore do not depend on $H_1(t), \ldots, H_N(t)$.

(iii) Finally, we calculate the frequency of each type in the offspring generation:

$$X_k(t+1) = \frac{1}{N}\#\{i : H_i(t) + J_i(t) = k\}.$$

It is obvious that a process that follows these dynamics is a Markov chain. As (i) to (iii) determine the transition probabilities, it defines a process unique in law. Thus, we can use the three steps to define Muller's Ratchet.

Definition 3.1 (Muller's ratchet)**.** Let $N \in \mathbb{N}$, $\lambda > 0$, $\alpha \in [0,1)$ and $\underline{x} \in \mathbb{S}$. We call a $\mathbb{S}$-valued Markov chain $\big(\underline{X}(t)\big)_{t\in\mathbb{N}} = \big(X_0(t), X_1(t), \ldots\big)_{t\in\mathbb{N}}$ *Muller's ratchet* starting in $\underline{x}$ if $\underline{X}(0) = \underline{x}$ and $\underline{X}$ evolves according to the dynamics described by steps (i) to (iii). We continue to denote *Muller's ratchet* with $\underline{X}$ throughout this chapter.

Of course, the ratchet with population size N takes (almost surely) only distributions as values, for which all probability weights are multiples of $\frac{1}{N}$. We denote the set of such distributions with

$$\mathbb{S}_N := \{\underline{x} \in \mathbb{S} : x_i = \frac{a_i}{N} \text{ with } a_i \in \mathbb{N} \text{ for all } i \in \mathbb{N}\} \tag{3.1}$$

Many of the unsolved problems of the ratchet are due to the implicit definition of the ratchet, i.e. the definition via dynamics. No one has so far managed to calculate the transition probabilities of $\underline{X}$ explicitly, or to give another more "closed" definition of Muller's ratchet. One of the main challenges here is that the different components $X_k(t+1)$ are not independent of each other by step (iii) of the definition. Dealing with the implicit definition is a delicate thing. However we can easily calculate the conditioned expectation and variance of Muller's ratchet. Therefore observe that given $\underline{X}(t)$,

$$NX_k(t+1) = \#\{i : H_i(t) + J_i(t) = k\}$$

is binomially distributed with parameters N and $p_k(t+1)$, which is defined as

$$\begin{aligned} p_k(t+1) &:= \mathbf{P}[H_1(t) + J_1(t) = k \,|\, \underline{X}(t)] \\ &= \sum_{i=0}^{k} \mathbf{P}[H_1 = i \,|\, \underline{X}(t)]\, \mathbf{P}[J_1 = k - i] \\ &= \sum_{i=0}^{k} \frac{(1-\alpha)^i \cdot X_i(t)}{W(t)} e^{-\lambda} \frac{\lambda^{(k-i)}}{(k-i)!}. \end{aligned} \tag{3.2}$$

Hence,

$$\mathbf{E}[X_k(t+1) \mid \underline{X}(t)] = \frac{1}{N} N p_k(t+1) = p_k(t+1) \tag{3.3}$$

and

$$\begin{aligned} \mathbf{Var}[X_k(t+1) \mid \underline{X}(t)] &= \frac{1}{N^2} N p_k(t+1)(1 - p_k(t+1)) \\ &= \frac{1}{N} p_k(t+1)(1 - p_k(t+1)). \end{aligned} \tag{3.4}$$

As mentioned before, most of the literature about Muller's ratchet aims to predict its click rate. To formulate this mathematically, we define a process $K^* = (K^*(t))_{t \geq 0}$ that states the number of mutations of the fittest class

$$K^*(t) := \inf\{k : X_k(t) > 0\}.$$

Now, given a specific path $t \mapsto \underline{X}(t)(\omega)$, we say that the ratchet has *clicked* between times $\tau < \tau'$ if

$$K^*\big(\tau'\big)(\omega) > K^*(\tau)(\omega)\,.$$

It is important that Muller's ratchet does not change its behavior after a click, when we measure the type frequencies relative to the fittest class. Therefore, notice that new mutations are completely independent of the current mutational load and fitness is measured relative to the best class, i.e.

$$\begin{aligned}
&\mathbf{P}[H_1(t) = K^*(t) + k \,|\, \underline{X}(t) = (0, \ldots, 0, x_{K^*}, x_{K^*+1}, \ldots)] \\
&= \frac{(1-\alpha)^{K^*(t)+k}\, x_{K^*(t)+k}}{\sum_{i=0}^{\infty}(1-\alpha)^{K^*(t)+i} x_{K^*(t)+i}} \\
&= \mathbf{P}[H_1(t) = k \,|\, \underline{X}(t) = (x_{K^*}, x_{K^*+1}, \ldots)]\,.
\end{aligned} \tag{3.5}$$

Hence, we expect that the vector $\underline{X}(t)$ will retain its basic shape, but the masses move one step in direction of higher X_k after each click (Figure 3.1). This phenomenon has been named "traveling wave" in the literature (e.g. Higgs and Woodcock [1995], Rouzine et al. [2003]).

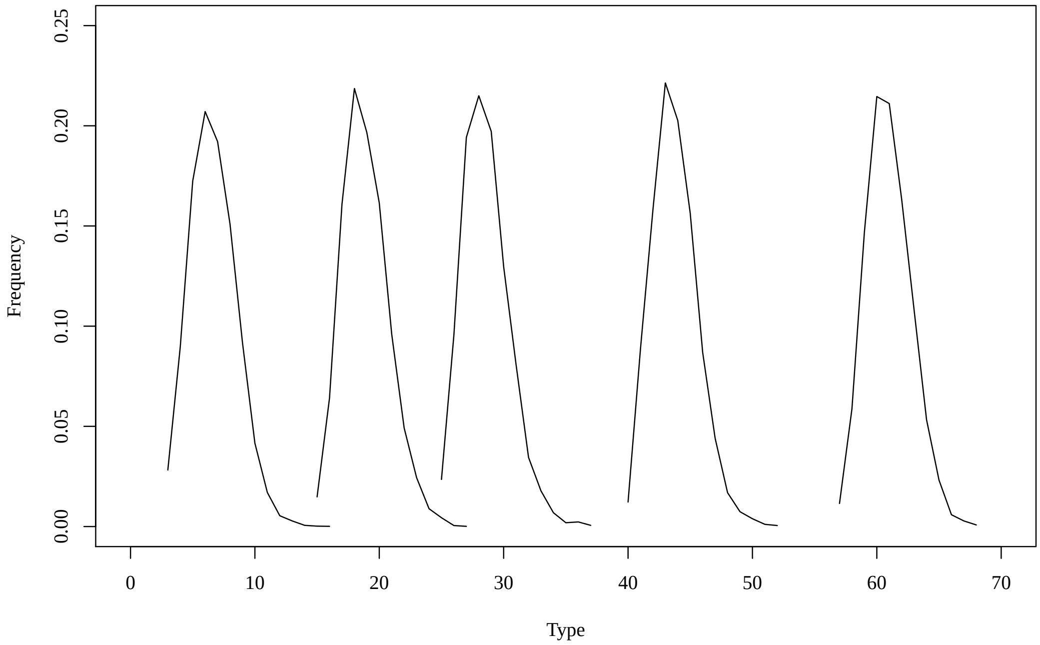

Figure 3.1.: The empirical distribution of types are drawn after $2N$, $10N$, $20N$, $30N$ and $40N$ generations respectively (from left to right). Here $N = 10000$, $\lambda = 0.1$ and $\alpha = 0.028$.

It is often useful to look at the vector of type frequencies $\underline{Y}$ relative to the current best class, hence

$$\underline{Y}(t) := \big(X_{K^*(t)+k}(t)\big)_{k\in\mathbb{N}},$$

as it forms a recurrent Markov chain on $\mathbb{S}_N$. Therefore notice, that mutation and selection are opposing each other in a way that the former always increase the number of deleterious mutations while the latter tries to minimize it. We will see that both forces drive the population towards a balance state. We call such states equilibrium points. As we expect that $\underline{X}$ stays in such a point once it has reached it, we define them as follows.

Definition 3.2. For a Markov chain $(\underline{X}(t))_{t\in\mathbb{N}}$ on $\mathbb{S}$, we call every $\underline{\pi} \in \mathbb{S}$ with

$$\mathbf{E}[\underline{X}(t+1) \,|\, \underline{X}(t) = \underline{\pi}] = \underline{\pi}$$

an *equilibrium point* (of $\underline{X}$).

However, the random fluctuations of genetic drift bars the population from staying in an equilibrium point forever. In fact, Haigh [1978] suggested that mutation and selection are strong enough to dominate genetic drift when the ratchet is away from an equilibrium. Therefore, the population drives towards an equilibrium point with strong force. When it approaches this point, mutation and selection cancel each other out more and more, and genetic drift becomes more important. Eventually, the random fluctuations will lead to the extinction of the current best class, and the ratchet clicks. Our main concern in this chapter is to prove that the Poisson distribution with parameter θ,

$$\underline{\pi} = \underline{\pi}^0 := \left(e^{-\theta}\frac{\theta^k}{k!}\right)_{k\in\mathbb{N}} \in \mathbb{S} \qquad \text{with } \theta := \frac{\lambda}{\alpha}, \tag{3.6}$$

and its right shifts $\underline{\pi}^j$, so vectors with

$$\pi_i^j = \begin{cases} 0, & \text{if } i < j, \\ \pi_{i-j}, & \text{otherwise}, \end{cases} \tag{3.7}$$

for $j \in \mathbb{N}$ are the equilibrium points of $\underline{X}$. We state this as a theorem.

Theorem 3.3. *The equilibrium points of $\underline{X}$ are exactly the distributions $\underline{\pi}^j$ with $j \in \mathbb{N}$.*

Proof. "Uniqueness" will be a consequence of Corollary 3.6 later. For "existence", let $j \in \mathbb{N}$. By Equation (3.3) we have,

$$\mathbf{E}\big[X_k(t+1) \mid \underline{X}(t) = \underline{\pi}^j\big] = 0 \qquad \text{for } k < j,$$

and for $k \in \mathbb{N}$,

$$\begin{aligned}
\mathbf{E}\big[X_{j+k}(t+1) \mid \underline{X}(t) = \underline{\pi}^j\big] &= \sum_{i=0}^{j+k} \frac{(1-\alpha)^i \cdot \pi_i^j}{\sum_{\ell=0}^{\infty}(1-\alpha)^\ell \pi_\ell^j} e^{-\lambda} \frac{\lambda^{(j+k-i)}}{(j+k-i)!} \\
&= \sum_{i=0}^{k} \frac{(1-\alpha)^i \cdot \pi_i}{\sum_{\ell=0}^{\infty}(1-\alpha)^\ell \pi_\ell} e^{-\lambda} \frac{\lambda^{(k-i)}}{(k-i)!} \\
&= \frac{e^{-\lambda}}{\sum_{\ell=0}^{\infty} \frac{(1-\alpha)^\ell \theta^\ell}{\ell!}} \sum_{i=0}^{k} \frac{(1-\alpha)^i \theta^i}{i!} \frac{\lambda^{(k-i)}}{(k-i)!} \\
&= \frac{e^{-\lambda}}{e^{(1-\alpha)\theta}} \frac{\lambda^k}{k!} \sum_{i=0}^{k} \left(\frac{(1-\alpha)\theta}{\lambda}\right)^i \frac{k \cdot \ldots \cdot (k-i+1)}{i!} \\
&= e^{-(\lambda\alpha^{-1}(1-\alpha)+\lambda)} \frac{\lambda^k}{k!} \left(\frac{(1-\alpha)}{\alpha} + 1\right)^k \\
&= e^{-\theta} \frac{\theta^k}{k!}. \qquad \square
\end{aligned}$$

To prove the "uniqueness" of the $\underline{\pi}^j$, we examine the convergence behavior of an approximation of Muller's ratchet, the ratchet without the effects of genetic drift. By the above argumentation, this approximation never clicks, and therefore will give us an impression of the ratchet's behavior in the time between two clicks.

3.2. The Approximation With Infinite Population Size

One of the first ideas in the examination of Muller's ratchet was to simplify the model by ignoring the effects of genetic drift. Therefore notice that according to the law of large numbers, $X_k(t+1)$ becomes almost surely deterministic if we increase the population size N towards infinity. We take this as motivation for the following definition.

Definition 3.4. Let $\lambda > 0$, $\alpha \in [0, 1)$ and $\underline{x} \in \mathbb{S}$. We call the time discrete, deterministic process $\underline{x}(t) = (x_i(t))_{i\in\mathbb{N}}$ with $\underline{x}(0) = \underline{x}$ and

$$x_k(t+1) := p_k(t) = \sum_{i=0}^{k} \frac{(1-\alpha)^i \cdot x_i(t)}{W(t)} e^{-\lambda} \frac{\lambda^{(k-i)}}{(k-i)!}$$

the *large population limit* of Muller's ratchet.

As $\underline{x}$ and $\underline{X}$ equal each other in (conditioned) expectation, they have exactly the same equilibrium points. Hence, we already know from Theorem 3.3 that all right shifts of $\underline{\pi}$ are equilibrium points of $\underline{x}$ as well. In return, we use the more simple deterministic process to prove the absence of other equilibrium points. As the almost sure convergence implies convergence in distribution, we can furthermore assume that the claimed approximating behavior holds.

The large population limit of Muller's ratchet is in particular useful because the recursion given by Definition 3.4 has been explicitly solved by Maia et al. [2003].

Theorem 3.5 (Maia, Botelho, and Fontanari [2003]). *For all $t \in \mathbb{N}$, $\underline{x}(t)$ is given by*

$$x_k(t) = \frac{\exp(-\theta_t)}{\sum_{j=0}^{\infty} x_j(0)\,(1-\alpha)^{jt}} \sum_{j=0}^{k} \frac{\theta_t^{k-j}}{(k-j)!} x_j(0)\,(1-\alpha)^{jt}, \tag{3.8}$$

where $\theta_t = \frac{\lambda}{\alpha}\left[1-(1-\alpha)^t\right]$.

Proof. The proof is based on the fact that we can uniquely determine a discrete probability distribution $\underline{x} \in \mathbb{S}$ by its *generating function*

$$G[z, \underline{x}] := \sum_{i=0}^{\infty} x_i z^i, \qquad z \in [0,1]$$

(e.g. see Klenke [2006, Chapter 3]). Hence, we translate $\underline{x}(t)$ into terms of generating functions, solve the recursion there and translate the solution back to a probability distribution afterwards.

First, notice that

$$G[1-\alpha, \underline{x}(t)] = \sum_{i=0}^{\infty} x_i(t)\,(1-\alpha)^i = W(t).$$

Using the equation

$$\sum_{i=j}^{\infty} e^{-\lambda} \frac{\lambda^{i-j}}{(i-j)!} z^i = e^{-\lambda} z^j \sum_{i=0}^{\infty} \frac{(\lambda z)^i}{i!} = z^j e^{\lambda(z-1)} \tag{3.9}$$

we obtain the recursion

$$\begin{aligned} G[z, \underline{x}(t+1)] &= \sum_{i=0}^{\infty} x_i(t+1)\, z^i \\ &= \sum_{i=0}^{\infty} \sum_{j=0}^{i} \frac{(1-\alpha)^j x_j(t)}{W(t)} e^{-\lambda} \frac{\lambda^{i-j}}{(i-j)!} z^i \\ &= \sum_{j=0}^{\infty} \sum_{i=j}^{\infty} \frac{(1-\alpha)^j x_j(t)}{W(t)} e^{-\lambda} \frac{\lambda^{i-j}}{(i-j)!} z^i \\ &= \sum_{j=0}^{\infty} \frac{(1-\alpha)^j x_j(t)}{W(t)} \sum_{i=j}^{\infty} \frac{\lambda^{i-j}}{(i-j)!} z^i \\ &= \sum_{j=0}^{\infty} \frac{(1-\alpha)^j x_j(t)}{W(t)} z^j e^{\lambda(z-1)} \\ &= e^{\lambda(z-1)} \frac{G[z(1-\alpha), \underline{x}(t)]}{G[(1-\alpha), \underline{x}(t)]} \end{aligned} \tag{3.10}$$

where we can reorder the summands because they are all non-negative. We use (3.10) to prove by induction that

$$G[z, \underline{x}(t)] = e^{\theta_t(z-1)} \frac{G[z(1-s)^t, \underline{x}(0)]}{G[(1-\alpha)^t, \underline{x}(0)]}. \tag{3.11}$$

While (3.11) is obvious for $t = 0$, we have

$$\begin{aligned} G[z, \underline{x}(t+1)] &\overset{(3.10)}{=} e^{\lambda(z-1)} \frac{G[z(1-\alpha), \underline{x}(t)]}{G[1-\alpha, \underline{x}(t)]} \\ &\overset{(3.11)}{=} e^{\lambda(z-1)} \frac{e^{\theta_t(z(1-\alpha)-1)} \cdot G[z(1-\alpha)^{t+1}, \underline{x}(0)]}{e^{\theta_t((1-\alpha)-1)} \cdot G[(1-\alpha)^{t+1}, \underline{x}(0)]} \\ &= e^{\lambda(z-1)+\theta_t(z(1-\alpha)-1)+\theta_t\alpha} \frac{G[z(1-\alpha)^{t+1}, \underline{x}(0)]}{G[(1-\alpha)^{t+1}, \underline{x}(0)]} \\ &= e^{\theta_{t+1}(z-1)} \frac{G[z(1-\alpha)^{t+1}, \underline{x}(0)]}{G[(1-\alpha)^{t+1}, \underline{x}(0)]} \end{aligned}$$

because

$$\begin{aligned} \lambda(z-1) + \theta_t(z(1-\alpha)-1) + \theta_t\alpha &= \frac{\lambda}{\alpha}\left(z - z(1-\alpha)^{t+1} - 1 + (1-\alpha)^t - \alpha(1-\alpha)^t\right) \\ &= \frac{\lambda}{\alpha}\left(z - z(1-\alpha)^{t+1} - 1 + (1-\alpha)^{t+1}\right) \\ &= (z-1)\,\frac{\lambda}{\alpha}(1-(1-\alpha)^{t+1}). \end{aligned}$$

Finally, we again write (3.11) as a generating function to identify its coefficients.

$$\begin{aligned}
G[z, \underline{x}(t)] &= e^{\theta_t(z-1)} \frac{G[z(1-\alpha)^t, \underline{x}(0)]}{G[(1-\alpha)^t, \underline{x}(0)]} \\
&= \frac{1}{G[(1-\alpha)^t, \underline{x}(0)]} \sum_{j=0}^{\infty} x_j(0)\,(1-\alpha)^{jt} z^j e^{\theta_t(z-1)} \\
&\overset{(3.9)}{=} \frac{1}{G[(1-\alpha)^t, \underline{x}(0)]} \sum_{j=0}^{\infty} x_j(0)\,(1-\alpha)^{jt} \sum_{k=j}^{\infty} \frac{\theta_t^{k-j}}{(k-j)!} z^k e^{-\theta_t} \\
&= \frac{1}{G[(1-\alpha)^t, \underline{x}(0)]} \sum_{j=0}^{\infty} \sum_{k=j}^{\infty} x_j(0)\,(1-\alpha)^{jt} \frac{\theta_t^{k-j}}{(k-j)!} z^k e^{-\theta_t} \\
&= \frac{1}{G[(1-\alpha)^t, \underline{x}(0)]} \sum_{k=0}^{\infty} \sum_{j=0}^{k} x_j(0)\,(1-\alpha)^{jt} \frac{\theta_t^{k-j}}{(k-j)!} z^k e^{-\theta_t} \\
&= \sum_{k=0}^{\infty} \left(\frac{e^{-\theta_t}}{G[(1-\alpha)^t, \underline{x}(0)]} \sum_{j=0}^{k} x_j(0)\,(1-\alpha)^{jt} \frac{\theta_t^{k-j}}{(k-j)!} \right) z^k \\
&= \sum_{k=0}^{\infty} \left(\frac{e^{-\theta_t}}{\sum_{j=0}^{\infty} x_j(0)\,(1-\alpha)^{jt}} \sum_{j=0}^{k} x_j(0)\,(1-\alpha)^{jt} \frac{\theta_t^{k-j}}{(k-j)!} \right) z^k \qquad \square
\end{aligned}$$

Note that by (3.8), $x_k(t) = 0$ if and only if $x_j(0) = 0$ for all $j \leq k$. Hence – as argued above – the large population limit of Muller's ratchet never clicks.

Using Theorem 3.5, we can both prove the uniqueness in Theorem 3.3 and show that the equilibrium points attract Muller's ratchet with infinite population size. We define $\underline{y}$ for $\underline{x}$ as $\underline{Y}$ for $\underline{X}$, i.e. as the type frequencies relative to the fittest class.

Corollary 3.6. *For any starting distribution $y(0)$ and $k \in \mathbb{N}$, we have $y_k(t) \to \pi_k$ for $t \to \infty$. In particular, $\underline{\pi}$ is the only equilibrium point for $\underline{y}$.*

Proof. Recall from Theorem (3.5) that

$$y_k(t) = \frac{\exp(-\theta_t)}{\sum_{j=0}^{\infty} y_j(0)\,(1-\alpha)^{jt}} \sum_{j=0}^{k} \frac{\theta_t^{k-j}}{(k-j)!} y_j(0)(1-\alpha)^{jt}. \tag{3.12}$$

Obviously $\theta_t = \theta\left[1-(1-\alpha)^t\right] \to \theta$. As the following calculation shows, the sums in (3.12) are dominated by their $y_0(0)$ term:

$$0 \leq \lim_{t\to\infty} \sum_{j=1}^{\infty} y_j(0)(1-\alpha)^{jt} \leq \lim_{t\to\infty} \sum_{j=1}^{\infty} \left((1-\alpha)^t\right)^j = \lim_{t\to\infty} \frac{1}{1-(1-\alpha)^t} - 1 = 0$$

and

$$0 \leq \lim_{t\to\infty} \sum_{j=1}^{k} \frac{\theta_t^{k-j}}{(k-j)!} y_j(0)(1-\alpha)^{jt} \leq e^{\theta} \lim_{t\to\infty} \sum_{j=1}^{\infty} \left((1-\alpha)^t\right)^j = 0$$

using the limit of the geometric series for $(1-\alpha)^t$ and the fact that $\frac{\theta_t^{k-j}}{(k-j)!} \leq e^{\theta_t} \leq e^{\theta}$. Hence we have

$$\begin{aligned}
\lim_{t\to\infty} y_k(t) &= \lim_{t\to\infty} \frac{\exp(-\theta_t)}{\sum_{j=0}^{\infty} y_j(0)\,(1-\alpha)^{jt}} \sum_{j=0}^{k} \frac{\theta_t^{k-j}}{(k-j)!} y_j(0)(1-\alpha)^{jt} \\
&= \frac{\exp(-\theta)}{y_0(0)} \frac{\theta^k}{k!} y_0(0) \\
&= \frac{\theta^k}{k!} \exp(-\theta)
\end{aligned}$$

where the second equality is valid because $y_0(0) > 0$ by definition. □

Taking (3.5) into account, we conclude that the ratchet with infinite population size with best class k will always tend towards $\underline{\pi}^k$. If the population size N is big enough, the ratchet with finite population size should behave similarly. However, after a click – which should occur only rarely for such N – the ratchet will tend to $\underline{\pi}^{k+1}$ instead.

The question which values of N are actually "big enough" for different combinations of λ and α is unanswered yet. However, as the large population limit does not click, N surely is too small if Muller's ratchet clicks frequently. Different authors made predictions for when this should be the case. For example, Haigh [1978] suggested that the ratchet clicks frequently if the expected size of the best class $n_0 = Ne^{-\theta}$ is less than one individual, while Etheridge et al. [2008] concluded that the same should occur if $\frac{N\lambda}{N\alpha \cdot \log(N\lambda)}$ is less then 0.5. Anyway, the magnitude of the "finite population effect" seems to be closely related to the click rate of the ratchet.

4. Compensatory Mutations

As motivated in the introduction, we introduce *compensatory mutations* into the classical model of Muller's ratchet in this chapter. Therefore we assume that every acquired mutation has a probability $\gamma \in [0, 1]$ to become compensated. At first, we take a brief look at an extended version of Haigh's discrete model. As results are difficult to derive here, we will soon go on to two time-continuous approximations that have proved to be useful for the classical model. The first one is the popular diffusion approximation of Muller's ratchet, a Fleming-Viot process described in Etheridge et al. [2008]. Similar to the approximation for the discrete model, the second one is the (weak) large population limit of the diffusion approximation, which is almost surely deterministic. This model is defined as the solution of an ordinary differential equation (ODE), while the diffusion limit relies on a stochastic differential equation (SDE). Hence, we refer to the latter model as ODE approximation of Muller's ratchet. We will introduce both models directly including compensatory mutations. However, we do explicitly allow $\gamma = 0$, i.e. the absence of such mutations. In this case, in particular the results about the ODE approximation are well-known (see Etheridge et al. [2008]).

Our agenda for this chapter is similar to the previous one. First we show that the SDE has a unique solution (Theorem 4.7, p. 35). For that, it will be crucial to reformulate the SDE as a martingale problem (Theorem 4.6, p. 30). Afterwards, we show that the ODE approximation is indeed a (weak) limit of solutions of the SDE (Theorem 4.13, p. 40) and that the ODE is well-posed (Theorem 4.16, p. 42). Then, we will introduce cumulants and derive the existence of an equilibrium point (Corollary 4.20, p. 45). Analogous to the Maia-Botelho-Fontanari-Theorem we explicitly solve the ODE (Theorem 4.21, p. 46) using a stochastic particle system afterwards. Finally we prove that the ODE approximation always convergences towards the equilibrium point (Corollary 4.26, p. 51).[1]

4.1. The Discrete Model

It is not difficult to introduce compensatory mutations into Haigh's model of Muller's ratchet. For instance, we can add an additional step between steps (i) and (ii) in Definition 3.1 by removing a $\mathrm{Bin}(H_i, \gamma)$-distributed number K_i of mutations from

[1] The results in this chapter are joint work of Peter Pfaffelhuber, Anton Wakolbinger and the author. A shorter version of this manuscript, which is prepared for publication (Pfaffelhuber et al. [2011]), is attached in the Appendix. Note that we were able to relax the conditions on the exponential moments of the ratchet there.

every individual. Hence the conditioned expectation p_k in (3.2) changes to

$$\begin{aligned}
p_k(t+1) &:= \mathbf{P}[H_1 + J_1 - K_1 = k \,|\, \underline{X}(t)] \\
&= \sum_{i=0}^{k} \mathbf{P}[J_1 = k-i] \sum_{j=i}^{\infty} \mathbf{P}[H_1 = j \,|\, \underline{X}(t)] \cdot \mathbf{P}[K_1 = j-i \,|\, \underline{X}(t)\,, H_1 = j] \\
&= \sum_{i=0}^{k} e^{-\lambda} \frac{\lambda^{(k-i)}}{(k-i)!} \sum_{j=i}^{\infty} \frac{(1-s)^j \cdot X_j(t)}{W(t)} \binom{j}{j-i} \gamma^{j-i}(1-\gamma)^i.
\end{aligned}$$

With compensatory mutations present, an individual can have less mutations than its parent. Hence, clicks can be reverted and the ratchet mechanism breaks (Figure 4.1). In fact, as the expected number of compensatory mutations of an individual with a parent of type H_i is $\gamma \cdot H_i$, the compensatory effect will increase with an increasing mutational load of the population. Because the intensity of deleterious mutations is constant, it is reasonable to assume that the accumulation of mutations slows down and will eventually stop.

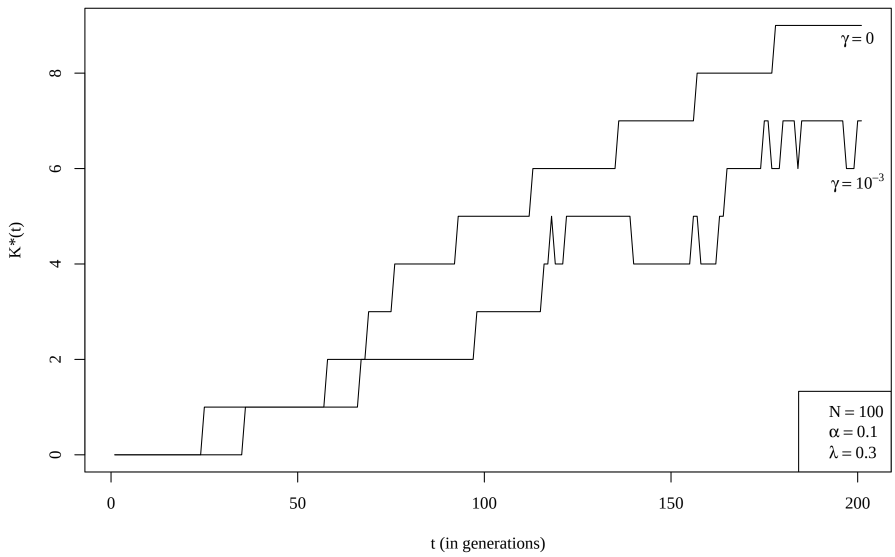

Figure 4.1.: Evolution of the number of mutations K^* of the fittest class in time, once non-decreasing without compensatory mutations ($\gamma = 0$) and once also decreasing with compensatory mutations present ($\gamma = 10^{-3}$).

Without the ratchet effect present, we can always reach every state in $\mathbb{S}_N$ (as defined in (3.1)) with positive probability. Hence, the ratchet with compensatory mutations is more similar to the vector of type frequencies relative to the current fittest class $\underline{Y}$ in the previous chapter than to the classical ratchet. Furthermore we can hope that if there is an equilibrium point, then it is unique. However, the more complicated situation prevents us from applying any of the approaches we used in the previous chapter. In particular, we were not able to derive a result similar to the Maia-Botelho-Fontanari-Theorem for $\gamma > 0$ in the discrete system. Therefore, we go on to the before mentioned continuous models.

4.2. The Diffusion Approximation

In this chapter, we are using diffusion theory to approximate Muller's ratchet (with compensatory mutations) by a time continuous stochastic process. Similar as it has been done for the original ratchet (e.g. in Stephan et al. [1993], Etheridge et al. [2008]), we translate the dynamics in the definition of Muller's ratchet with compensatory mutations into a stochastic differential equation (SDE) on $\mathbb{S}$ defined by

$$dX_k(t) = b_k(\underline{X}(t))dt + \sum_{l\neq k} \sqrt{\frac{1}{N} X_k(t)\, X_l(t)}\, dW_{kl} \tag{$*$}$$

for $k \in \mathbb{N}$, where $X_{-1} := 0$,

$$b_k(\underline{x}) := \alpha \sum_l (l-k)x_k x_l + \lambda\,(x_{k-1} - x_k) + \gamma\,((k+1)x_{k+1} - kx_k) \tag{4.1}$$

and W_{kl} are independent, 1-dimensional Brownian motions for $k > l$ and $W_{kl} := -W_{lk}$ for $k < l$. For $\alpha = \lambda = \gamma = 0$, this is the classical diffusion approximation of the Wright-Fisher dynamics. Again, we add selection, mutation and compensatory mutations to this model via the three corresponding terms in (4.1). As $(1-\alpha)^k \sim 1-\alpha k$ for small α, we have the occurrence of $-\alpha k$ in the selection term. The mutation term reflects the flow of individuals of type k to type $k+1$ and from $k-1$ to k, both with constant rate λ. The compensatory mutation term is similar, but the rates depend on the current number of mutations. We briefly recall the necessary notations for stochastic differential equations.

Definition 4.1 (SDE)**.** We say that a tuple $(\Omega, \mathscr{F}, \mathbf{P}, (W_{kl})_{k<l}, \underline{X})$, consisting of a filtered probability space $(\Omega, \mathscr{F}, \mathbf{P})$, an array $(W_{kl})_{k<l}$ of $\mathscr{F}$-Brownian motions satisfying the conditions mentioned above and an adapted process $\underline{X}$, is a *(weak) solution* of the SDE ($*$) with initial distribution μ, if $X(0)_*\mathbf{P} = \mu$ and it satisfies the SDE or, more explicitly,

$$X_k(t) = X_k(0) + \int_0^t b_k(\underline{X}(s))\, ds + \sum_{l\neq k} \int_0^t \sqrt{\frac{1}{N} X_k(s)\, X_l(s)}\, dW_{kl}(s)$$

for all $k \in \mathbb{N}$. We say that *(weak) existence* holds for an initial distribution if there is a weak solution with this initial distribution. Analogous, *uniqueness* holds for an initial distribution if all weak solutions for it are equal in distribution. We say the SDE is *well-posed* if existence and uniqueness holds for every initial distribution. For Muller's ratchet, we always assume that $\mathscr{F}$ is the natural filtration of $\underline{X}$.

It is a well-known result by Stroock and Varadhan that stochastic differential equations can equivalently be formulated as martingale problems under sufficient conditions (e.g.

[Kallenberg, 1997, Theorem 18.7]). The diffusion approximation of Muller's ratchet was defined as a SDE, because this is quite typical for Wright-Fisher models. However, a "typical" Wright-Fisher model has only a finite number of different types, whereas the type space of Muller's ratchet is $\mathbb{N}$. Therefore it is realised as a Fleming-Viot process, which are typically defined as martingale problems (e.g. Ethier and Kurtz [1993]). Hence, we want to formulate Muller's ratchet as a martingale problem in order to apply results from the theory of Fleming-Viot processes. This requires further investigation, as the conditions of the Stroock-Varadhan-Theorem are not given here, mainly because $\mathbb{S}$ is an infinite dimensional space.

4.2.1. Characterisation As A Martingale Problem

As we have done for stochastic different equations, we start by recalling the basic notations in the theory of martingale problems.

Definition 4.2 (Martingale problem)**.** Let (E, r) be a complete and separable metric space, $\mathbf{P}_0 \in \mathcal{P}(E)$, $\mathcal{F} \subseteq \mathcal{M}(E)$ and G a linear operator on $\mathcal{M}(E)$. A distribution $\mathbf{P}$ of an E-valued stochastic process $\mathcal{X} = (\mathcal{X}_t)_{t\geq 0}$ is called a solution of the $(E, \mathbf{P}_0, G, \mathcal{F})$-martingale problem if $\mathcal{X}_0$ has distribution $\mathbf{P}_0$, almost all paths of $\mathcal{X}$ are cadlag and $N^f = (N_t^f)_{t\geq 0}$ defined by

$$N_t^f = f(\mathcal{X}_t) - f(\mathcal{X}_0) - \int_0^t Gf(\mathcal{X}_s)ds \tag{4.2}$$

is a $\mathbf{P}$-martingale with respect to the canonical filtration for all $f \in \mathcal{F}$. We say that a process $\mathcal{X}$ solves the $(E, \mathbf{P}, G, \mathcal{F})$-martingale problem if its distribution does so and refer to G as its (infinitesimal) generator. Moreover, we call a martingale problem well-posed if there is a unique solution $\mathbf{P}$.

Hence, we need a domain $\mathcal{F}$ and a generator G to define a martingale problem on $\mathbb{S}$. We define the ones that correspond to SDE $(*)$ in the next definition. In fact, we state two different domains $\mathcal{F}$ and $\overline{\mathcal{F}}$ for technical reasons. We will show in Theorem 4.6 that the corresponding martingale problems are equivalent.

Definition 4.3 (Domains and generator for Muller's Ratchet)**.** Recall Definition 2.4. In addition to $\overline{\mathcal{F}}$, we define the algebra

$$\mathcal{F} := \mathcal{A}\left(\{f_{\varphi_1,\dots,\varphi_n} : n \in \mathbb{N}, \varphi_i \in \mathcal{B}(\mathbb{N}) \text{ with } |\mathrm{supp}(\varphi_i)| < \infty \text{ for } i = 1, \dots, n\}\right)$$

and the generator G^α on $\overline{\mathcal{F}}$ with domain $\overline{\mathcal{F}}$ by

$$
\begin{aligned}
G^\alpha f(\underline{x}) &= G^\alpha_{\text{sel}} f(\underline{x}) + G_{\text{mut}} f(\underline{x}) + G_{\text{cm}} f(\underline{x}) + G^N_{\text{res}} f(\underline{x}) \\
G^\alpha_{\text{sel}} f(\underline{x}) &= \alpha \sum_{k=0}^\infty \sum_{\ell=0}^\infty (\ell - k) x_\ell x_k \frac{\partial}{\partial x_k} f(\underline{x}), \\
G_{\text{mut}} f(\underline{x}) &= \lambda \sum_{k=0}^\infty (x_{k-1} - x_k) \frac{\partial}{\partial x_k} f(\underline{x}), \\
G_{\text{cm}} f(\underline{x}) &= \gamma \sum_{k=0}^\infty ((k+1) x_{k+1} - k x_k) \frac{\partial}{\partial x_k} f(\underline{x}) \qquad \text{and} \\
G^N_{\text{res}} f(\underline{x}) &= \tfrac{1}{2N} \sum_{k,\ell=0}^\infty x_k (\delta_{kl} - x_\ell) \frac{\partial^2}{\partial x_k \partial x_\ell} f(\underline{x})
\end{aligned}
$$

with $\alpha, \lambda, \gamma \in [0, \infty)$, $N \in (0, \infty)$.

We will now prove a lemma that is the technical key for the diffusion approximation. Basically, it says that if Muller's ratchet starts with a distribution in $\mathbb{S}^\circ$, then it will always take values in this space.

Lemma 4.4 (Bounds on exponential moments)**.** *Let $\underline{x} \in \mathbb{S}^\circ$ and $\underline{X} = (\underline{X}(t))_{t \geq 0}$ be a solution of the $(\mathbb{S}, \delta_{\underline{x}}, G^\alpha, \mathcal{F})$-martingale problem and recall that $h_\xi(\underline{x}) := \sum_{k \in \mathbb{N}} x_k e^{\xi k}$. Then, for all $T > 0$, there is $C > 0$, depending on T and all model parameters with*

$$
\mathbf{E}\big[\sup_{0 \leq t \leq T} h_\xi(\underline{X}(t))^2\big] \leq C \cdot h_{2\xi+2}(\underline{x}). \tag{4.3}
$$

In particular, we have $\sup_{0 \leq t \leq T} h_\xi(\underline{X}(t)) < \infty$ almost surely.

Proof. We define

$$
h_{\xi,m}(\underline{x}) := \sum_{k=0}^m x_k e^{\xi k} + e^{\xi m} \left(1 - \sum_{k=0}^m x_k\right) = e^{\xi m} + \sum_{k=0}^m x_l (e^{\xi k} - e^{\xi m}) \in \mathcal{F}
$$

for $m \in \mathbb{N}$. Note that for $\underline{x} \in \mathbb{S}$,

$$
h_{\xi,m}(\underline{x}) = \sum_{k=0}^\infty x_k e^{\xi (k \wedge m)}
$$

We compute

$$
\begin{aligned}
G_{\mathrm{mut}} h_{\xi,m}(\underline{x}) &= \lambda \sum_{k=0}^{m} (x_{k-1} - x_k)(e^{\xi k} - e^{\xi m}) \\
&= \lambda \sum_{k=0}^{m-1} x_k \Big(e^{\xi(k+1)} - e^{\xi k} \Big) \geq 0, \\
G_{\mathrm{cm}} h_{\xi,m}(\underline{x}) &= \lambda \sum_{k=0}^{m} ((k+1)x_{k+1} - kx_k)(e^{\xi k} - e^{\xi m}) \\
&= \lambda \sum_{k=0}^{m} kx_k \Big(e^{\xi(k-1)} - e^{\xi k} \Big) \leq 0, \qquad (4.4) \\
G_{\mathrm{sel}}^{\alpha} h_{\xi,m}(\underline{x}) &= \alpha \sum_{k=0}^{m} \sum_{\ell=0}^{\infty} (\ell - k) x_\ell x_k (e^{\xi k} - e^{\xi m}) \\
&= \alpha \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} (\ell - k) x_\ell x_k (e^{\xi(k\wedge m)} - e^{\xi m}) \\
&= \alpha \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} (\ell - k) x_\ell x_k e^{\xi(k\wedge m)} \leq 0,
\end{aligned}
$$

where the latter equality only holds for $x \in \mathbb{S}$. For the final inequality, assume that Z is an $\mathbb{N}$-valued random variable with distribution $\underline{x}$. Then we have

$$
G_{\mathrm{sel}}^{\alpha} h_{\xi,m}(\underline{x}) = -\alpha\, \mathbf{COV}[Z, e^{\xi(Z\wedge m)}] \leq 0,
$$

since for non-decreasing functions f, g and i.i.d. random variables Z and Z'

$$
\mathbf{E}\big[\big(f(Z) - f(Z')\big)\big(g(Z) - g(Z')\big)\big] \geq 0
$$

as both factors have the same sign. It follows that

$$
2\,\mathbf{COV}[f(Z), g(Z)] = 2\,\mathbf{E}[f(Z)g(Z)] - 2\,\mathbf{E}[f(Z)]\,\mathbf{E}[g(Z)] \geq 0.
$$

Second, we prove that a similar bound as (4.3) holds at fixed time points. We write

$$
\begin{aligned}
\frac{d}{dt}\mathbf{E}[h_{\xi,m}(\underline{X}(t))] = \mathbf{E}\big[G^{\alpha} h_{\xi,m}(\underline{X}(t))\,\big] &\leq \mathbf{E}\big[G_{\mathrm{mut}} h_{\xi,m}(\underline{X}(t))\,\big] \\
&\leq \lambda \mathbf{E}\Big[e^{\xi} \sum_{k=0}^{\infty} X_k(t)\, e^{\xi(k\wedge m)} - \sum_{k=0}^{\infty} X_k(t) e^{\xi(k\wedge m)} \Big] \\
&= \lambda(e^{\xi} - 1)\mathbf{E}[h_{\xi,m}(\underline{X}(t))].
\end{aligned}
$$

So, by Gronwall's inequality,

$$\mathbf{E}[h_{\xi,m}(\underline{X}(t))] \leq h_{\xi,m}(\underline{x}) \cdot \exp\big(\lambda t(e^{\xi} - 1)\big)$$

which also implies

$$\mathbf{E}[h_{\xi}(\underline{X}(t))] \leq h_{\xi}(\underline{x}) \cdot \exp\big(\lambda t(e^{\xi} - 1)\big) \tag{4.5}$$

by monotone convergence.
Third, (4.4) implies that

$$\Big(h_{\xi,m}(\underline{X}(t)) - \int_0^t G_{\mathrm{sel}}^{\alpha} h_{\xi,m}(\underline{X}(s)) + G_{\mathrm{cm}} h_{\xi,m}(\underline{X}(s))\, ds\Big)_{t\geq 0}$$

is a submartingale. Now, by Doob's submartingale inequality and the fact that $(h_{\xi}(\underline{x}))^2 \leq h_{2\xi}(\underline{x})$ for all $\underline{x} \in \mathbb{S}$ and $\xi \geq 0$,

$$\begin{aligned}
&\mathbf{E}[\sup_{0\leq t\leq T} (h_{\xi,m}(\underline{X}(t)))^2] \\
&\quad \leq \mathbf{E}\Big[\sup_{0\leq t\leq T}\Big(h_{\xi,m}(\underline{X}(t)) - \int_0^t G_{\mathrm{sel}}^{\alpha} h_{\xi,m}(\underline{X}(s)) + G_{\mathrm{mut}} h_{\xi,m}(\underline{X}(s))\, ds\Big)^2\Big] \\
&\quad \leq 4 \cdot \mathbf{E}\Big[\Big(h_{\xi,m}(\underline{X}(T)) - \int_0^T G^{\alpha} h_{\xi,m}(\underline{X}(s)) + h_{\xi,m}(\underline{X}(s)) ds\Big)^2\Big] \\
&\quad \leq 8 \cdot \Big(\mathbf{E}\big[(h_{\xi,m}(\underline{X}(T)))^2\big] \\
&\qquad\qquad + \int_0^T \int_0^T \mathbf{E}[(G_{\mathrm{sel}}^{\alpha} h_{\xi,m}(\underline{X}(s)) + G_{\mathrm{cm}} h_{\xi,m}(\underline{X}(s))) \\
&\qquad\qquad\qquad \cdot (G_{\mathrm{sel}}^{\alpha} h_{\xi,m}(\underline{X}(r)) + G_{\mathrm{cm}} h_{\xi,m}(\underline{X}(r))) dr ds\Big) \\
&\quad \leq 8 \cdot \Big(\mathbf{E}\big[h_{2\xi}(\underline{X}(T))\big] + (\alpha+\gamma)^2 \int_0^T \int_0^T \mathbf{E}[h_{\xi+1}(\underline{X}(s)) h_{\xi+1}(\underline{X}(r))] dr ds\Big) \\
&\quad \leq 8 \cdot \mathbf{E}\big[h_{2\xi}(\underline{X}(T))\big] + 16 \cdot (\alpha+\gamma)^2 T \int_0^T \mathbf{E}\big[h_{2\xi+2}(\underline{X}(s))\big] ds \\
&\quad \leq C \cdot h_{2\xi+2}(\underline{x})
\end{aligned}$$

for $C = (8 + 16\alpha^2 T^2) \exp\big(\lambda T(e^{2\xi+2} - 1)\big)$ by (4.5). □

We need another small result to prove Theorem 4.6.

Lemma 4.5. *Let $f_{\varphi_1,\dots,\varphi_n}$ be as in Definition 2.4. For $k, l \in \mathbb{N}$, its first and second order partial derivatives are*

$$\frac{\partial}{\partial x_k} f_{\varphi_1,\dots,\varphi_n}(\underline{x}) = \sum_{i=1}^{n} \varphi_i(k) f_{\varphi_1,\dots,\widehat{\varphi_i},\dots,\varphi_n}(\underline{x})$$

and

$$\frac{\partial^2}{\partial x_k \partial x_l} f_{\varphi_1,\ldots,\varphi_n}(\underline{x}) = \sum_{\substack{i,j=1,\\ i\neq j}}^{n} \varphi_i(k)\varphi_j(l) f_{\varphi_1,\ldots,\widehat{\varphi_i},\ldots,\widehat{\varphi_j},\ldots,\varphi_n}(\underline{x})$$

where $\widehat{\varphi}_i := 1$ *for all* i.

Proof. We use an induction over n. While $n = 1$ is trivial, we have

$$\begin{aligned}
\frac{\partial}{\partial x_k} f_{\varphi_1,\ldots,\varphi_n}(\underline{x}) &= \frac{\partial}{\partial x_k}\left(f_{\varphi_1,\ldots,\varphi_{n-1}}(\underline{x}) \sum_{i=0}^{\infty} x_i \varphi_n(i)\right) \\
&= \sum_{i=1}^{n-1} \varphi_i(k) f_{\varphi_1,\ldots,\widehat{\varphi_i},\ldots,\varphi_{n-1}}(\underline{x}) \sum_{i=0}^{\infty} x_i \varphi_n(i) \\
&\quad + f_{\varphi_1,\ldots,\varphi_{n-1}}(\underline{x}) \varphi_n(k) \\
&= \sum_{i=1}^{n} \varphi_i(k) f_{\varphi_1,\ldots,\widehat{\varphi_i},\ldots,\varphi_n}(\underline{x}).
\end{aligned}$$

For the second order derivatives, we calculate

$$\begin{aligned}
\frac{\partial^2}{\partial x_k \partial x_l} f(\underline{x}) &= \frac{\partial}{\partial x_l} \sum_{i=1}^{n} \varphi_i(k) f_{\varphi_1,\ldots,\widehat{\varphi_i},\ldots,\varphi_n}(\underline{x}) \\
&= \sum_{\substack{i,j=1,\\ i\neq j}}^{n} \varphi_i(k)\varphi_j(l) f_{\varphi_1,\ldots,\widehat{\varphi_i},\ldots,\widehat{\varphi_j},\ldots,\varphi_n}(\underline{x})
\end{aligned}$$ □

We define the following notation to abbreviate calculations with stochastic integrals.

Notation. For suitable stochastic processes H and X, we denote the Itô integral with

$$\int H_s dX_s = H \cdot X.$$

To evaluate the integral on time t, we write

$$\int_0^t H_s dX_s = H \cdot X_t.$$

Analogously, we denote the Lebesgue-measure with ν and the integral of $f : \mathbb{R} \to \mathbb{R}$ from 0 to t as

$$\int_0^t f(s) ds = f \cdot \nu(t)$$

After these necessary preparations, we are now ready to prove the maybe most technical theorem of this manuscript.

Theorem 4.6. *Let $\underline{X}$ be a process with values in $\mathbb{S}$ and $\underline{X}(0) = \underline{x} \in \mathbb{S}^\circ$. Then, the following are equivalent:*

1. *The process $\underline{X}$ solves the $(\mathbb{S}, \delta_{\underline{x}}, G^\alpha, \mathcal{F})$-martingale problem.*
2. *The process $\underline{X}$ solves the $(\mathbb{S}^\circ, \delta_{\underline{x}}, G^\alpha, \overline{\mathcal{F}})$-martingale problem.*
3. *The process $\underline{X}$ is a weak solution of the SDE $(*)$ with initial distribution $\delta_{\underline{x}}$.*

Proof. 2. $\Rightarrow$ 1.: Obviously $\mathcal{F} \subseteq \overline{\mathcal{F}}$. The assertion follows directly.
1. $\Rightarrow$ 2.: Let $\underline{X}$ solve the $(\mathbb{S}, \delta_{\underline{x}}, G^\alpha, \mathcal{F})$-martingale problem. By Lemma 4.4, $\underline{X}$ takes values in $\mathbb{S}^\circ$ a.s. and hence is a solution of the $(\mathbb{S}^\circ, \delta_{\underline{x}}, G^\alpha, \overline{\mathcal{F}})$-martingale problem if N^f (as in (4.2) with $G = G^\alpha$) is a martingale for every $f \in \overline{\mathcal{F}}$. However, it suffices to show the martingale property for the generator of $\overline{\mathcal{F}}$, as it is closed under multiplication, $N^f(t)$ is linear in f and linear combinations of martingales are martingales.
So for a fixed f, there exist $\varphi_1, \dots, \varphi_n \in \mathcal{B}^e(\mathbb{S})$ such that $f = f_{\varphi_1,\dots,\varphi_n}$. We approximate f with functions $f^j := f_{\varphi_1^j,\dots,\varphi_n^j} \in \mathcal{F}$ where $\varphi_i^j(l) := \varphi_i(l)\mathbb{1}_{\{l\le j\}}$ with $j, l \in \mathbb{N}$. Hence, $f^j \xrightarrow{j\to\infty} f$ pointwise and similar to Lemma 4.5,

$$\frac{\partial}{\partial x_k} f^j \xrightarrow{j\to\infty} \frac{\partial}{\partial x_k} f \quad \text{and} \quad \frac{\partial^2}{\partial x_k \partial x_l} f^j \xrightarrow{j\to\infty} \frac{\partial^2}{\partial x_k \partial x_l} f,$$

both again pointwise.
We now want to prove that also $G^\alpha_{\text{sel}} f^j(\underline{x}) \to G^\alpha_{\text{sel}} f(\underline{x})$. Therefore, observe that

$$\sup_j |\langle \varphi_i^j, \underline{x}\rangle| \le \sum_{l=0}^\infty \sup_j |\varphi_i^j(l)| x_l \le \sum_{l=0}^\infty |\varphi_i(l)| x_l \le C \sum_{l=0}^\infty e^{\xi l} x_l < \infty$$

for some $C, \xi > 0$ and all $1 \le i \le n$, as φ_i is exponentially bounded and $\underline{x}$ has all exponential moments. Hence,

$$\sup_j |\frac{\partial}{\partial x_k} f^j(\underline{x})| \le \sup_j |\sum_{i=1}^n \varphi_i^j(k) f_{\varphi_1^j,\dots,\widehat{\varphi_i^j},\dots,\varphi_n^j}(\underline{x})| \tag{4.6}$$

$$\le \sum_{i=1}^n |\varphi_i(k)| C_i \le \sum_{i=1}^n \widetilde{C}_i e^{\xi_i k} \le C e^{\xi k} \tag{4.7}$$

again for suitable $C, C_i, \widetilde{C}_i, \xi_i, \xi > 0$. Now

$$\sum_{k=0}^\infty \sup_j |\alpha \sum_{l=0}^\infty (l-k) x_l x_k \frac{\partial}{\partial x_k} f^j(\underline{x})| \le \sum_{k=0}^\infty x_k\, C e^{\xi k} \left| \left(\sum_{l=0}^\infty l x_l \right) - k \right|$$

$$\le \sum_{k=0}^\infty x_k\, \widetilde{C} e^{\widetilde{\xi} k} < \infty,$$

and the convergence of $G^{\alpha}_{\text{sel}} f^j(\underline{x})$ follows by dominated convergence. Analogously,

$$G_{\text{mut}} f^j(\underline{x}) \xrightarrow{j\to\infty} G_{\text{mut}} f(\underline{x}) \quad \text{and} \quad G_{\text{cm}} f^j(\underline{x}) \xrightarrow{j\to\infty} G_{\text{cm}} f(\underline{x})\,.$$

Finally observe that analogous to (4.6), also

$$\sup_j \Big|\frac{\partial^2}{\partial x_k \partial x_l} f_{\varphi_1^j,\ldots,\varphi_n^j}(\underline{x})\Big| \le C e^{\xi(k+l)}$$

and therefore

$$\begin{aligned} \sum_{k,l=0}^{\infty} \sup_j \Big|x_k(\delta_{kl} - x_l)\frac{\partial^2}{\partial x_k \partial x_l} f^j\Big| &\le C \sum_{k=0}^{\infty} x_k e^{\xi k} \sum_{l=0}^{\infty} |(\delta_{kl} - x_l)e^{\xi l}| \\ &\le \widetilde{C} \sum_{k=0}^{\infty} x_k e^{2\xi k} < \infty. \end{aligned} \tag{4.8}$$

This gives $G^{\alpha} f^j(\underline{x}) \to G^{\alpha} f(\underline{x})$. Hence, N^f is a martingale:

$$\begin{aligned} \mathbf{E}\Big[N^f(t) \mid (X_r)_{r\le s}\Big] &= \mathbf{E}\Big[\lim_{j\to\infty} N^{f^j}(t) \mid (X_r)_{r\le s}\Big] \\ &= \lim_{j\to\infty} \mathbf{E}\Big[N^{f^j}(t) \mid (X_r)_{r\le s}\Big] \\ &= \lim_{j\to\infty} N^{f^j}(s) \\ &= N^f(s) \end{aligned}$$

where we again use dominated convergence.
3. $\Rightarrow$ 1.: Assume that $\underline{X}$ is a solution of the SDE $(*)$. We calculate the covariation process $([X_i, X_j]_t)_{t\ge 0}$ of X_i and X_j as

$$\begin{aligned} [X_i, X_j] &= \left[\sum_{k\neq i} \sqrt{N^{-1} X_i X_k} \cdot W_{ki}, \sum_{l\neq j} \sqrt{N^{-1} X_j X_l} \cdot W_{lj}\right] \\ &= \sum_{\substack{k\neq i\\ l\neq j}} N^{-1} \sqrt{X_i X_k X_j X_l} \cdot [W_{ki}, W_{lj}] \\ &= \begin{cases} N^{-1}(-1) X_i X_j \cdot \nu & \text{for } i \neq j \\ N^{-1} X_i (1 - X_i) \cdot \nu & \text{for } i = j \end{cases} \\ &= N^{-1} X_i (\delta_{ij} - X_j) \cdot \nu \end{aligned}$$

Now as every $f \in \mathcal{F}$ only depends on finitely many coordinates, we can use Itô's formula, which yields

$$\begin{aligned}
df(\underline{X}(t)) &= \sum_{i=0}^{\infty} f_i(\underline{X}(t))dX_i(t) + \sum_{i,j=0}^{\infty} \frac{1}{2} f_{ij}(\underline{X}(t))d[X_i, X_j]_t \\
&= \sum_{i=0}^{\infty} f_i(\underline{X}(t)) \left(b_i(\underline{X}(t))dt + \sum_{j \neq i} a_{ij}(\underline{X}(t))dW_{ij} \right) \\
&\quad + \sum_{i,j=0}^{\infty} \frac{1}{2} f_{ij}(\underline{X}(t))N^{-1}X_i(1 - X_i)dt \\
&= \sum_{i=0}^{\infty} f_i(\underline{X}(t)) \sum_{j \neq i} a_{ij}(\underline{X}(t))dW_{ij} + G^{\alpha} f(\underline{X}(t))dt.
\end{aligned}$$

with $a_{ij}(\underline{x}) := \sqrt{\frac{1}{N}x_i x_j}$, $f_i := \frac{\partial}{\partial x_i} f$ and $f_{ij} := \frac{\partial^2}{\partial x_i \partial x_j} f$. Hence,

$$\begin{aligned}
dN_t^f &= df(\underline{X}(t)) - G^{\alpha} f(\underline{X}(t)) \\
&= \sum_{i=0}^{\infty} f_i(\underline{X}(t)) \sum_{j \neq i} a_{ij}(\underline{X}(t))\, dW_{ij}.
\end{aligned} \tag{4.9}$$

Clearly every $a_{ij}(\underline{X}(t))\, dW_{ij}$ is a martingale because $a_{ij}(\underline{X}(t))$ is bounded almost surely. The same is true for $\sum_{j \neq i} a_{ij}(\underline{X}(t))dW_{ij}$, as we can again use dominated convergence because

$$\begin{aligned}
&\mathbf{E}\left[\left(\sum_{j \neq i} \int_0^t \sqrt{\frac{1}{N} X_i(t)\, X_j(t)}\, dW_{ij} \right)^2 \right] \\
\leq &\mathbf{E}\left[\sum_{\substack{j \neq i, \\ k \neq i}} \left| \int_0^t \sqrt{\frac{1}{N} X_i(t)\, X_j(t)} dW_{ij} \right| \left| \int_0^t \sqrt{\frac{1}{N} X_i(t)\, X_k(t)} dW_{ik} \right| \right] \\
\leq &\sum_{\substack{j \neq i, \\ k \neq i}} \mathbf{E}\left[\left(\int_0^t \sqrt{\frac{1}{N} X_i(t)\, X_j(t)} dW_{ij} \right)^2 \right] \mathbf{E}\left[\left(\int_0^t \sqrt{\frac{1}{N} X_i(t)\, X_k(t)} dW_{ik} \right)^2 \right] \\
= &\left(\sum_{j \neq i} \mathbf{E}\left[\int_0^t \frac{1}{N} X_i(s)\, X_j(s)\, ds \right] \right)^2 \\
= &\left(\frac{1}{N} \mathbf{E}\left[\int_0^t X_i(s)\, (1 - X_i(s))\, ds \right] \right)^2 < \infty
\end{aligned}$$

using the Cauchy-Schwarz inequality in the 2nd step and Itô's isometry in the 3rd one. This finishes the proof of 2. because the first sum in (4.9) is effectively finite.
2. $\Rightarrow$ 3.: Let $\underline{X}$ solve the $(\mathbb{S}^\circ, \delta_{\underline{x}}, G^\alpha, \overline{\mathcal{F}})$-martingale problem. We have to construct Brownian motions $(W_{kl})_{k\leq l}$ on $\mathbb{S}$ such that $(*)$ holds. To do so, we use Doob's integral representation theorem (e.g. Kallenberg [1997, Theorem 16.12]):

Theorem. *Let M be a continuous local $\mathscr{F}$-martingale in $\mathbb{R}^d$ with $M_0 = 0$ such that the covariation of two coordinate processes M_i and M_j is*

$$[M_i, M_j]_t = \int_0^t \sum_{k=1}^{d} \sum_{l>k}^{d} V^i_{k,l}(s) V^j_{k,l}(s)\, ds \quad a.s.$$

for some $\mathscr{F}$-progressive processes $V^i_{k,l}$, $1 \leq i \leq d, 1 \leq k < l \leq d$. Then there exist some independent standard Brownian motions $(W_{kl})_{1\leq k<l\leq d}$ with respect to a standard extension of $\mathscr{F}$ such that $M_i = \sum_{k=1}^d \sum_{l>k}^d V^i_{k,l} \cdot W_{k,l}$.

First notice that we can reformulate the SDE $(*)$ to

$$dX_i(t) = b_i(\underline{X}(t))dt + \sum_{k=0}^{\infty} \sum_{l>k} (\delta_{ik} - \delta_{il}) \sqrt{\frac{1}{N} X_k(t)\, X_l(t)}\, dW_{kl}. \tag{4.10}$$

Analogously to the proof of Theorem 18.7 in Kallenberg [1997] we can calculate that for

$$M_i(t) = X_i(t) + X_i(0) - \int_0^t b_i(X_i(s))\, ds, \tag{4.11}$$

with b_i as in $(*)$, the quadratic covariation of M_i and M_j is

$$[M_i, M_j] = \frac{1}{N} X_i(s)\, (\delta_{ij} - X_j(s)) \cdot \nu.$$

As Doob's theorem requires $\mathbb{R}^d$ as state space, we summarize all coordinates (X_i) with $i \geq R$ for fixed $R \in \mathbb{N}$ into a single process $\widetilde{X}_R := 1 - \sum_{i=0}^{R-1} X_i$ and lift the Brownian motions to $\mathbb{S}$ with a projective limit argument later. Furthermore, we define

$$\widetilde{M}_R(t) := \widetilde{X}_R(t) - \widetilde{X}_R(0) - \int_0^t b_i\left(\widetilde{X}_R(s)\right) ds = 1 - \sum_{i=0}^{R-1} M_i(t)$$

which clearly is a martingale as well. With the first line according to (4.10), we define

$$V^i_{k,l} := \begin{cases} (\delta_{ik} - \delta_{il}) \sqrt{\frac{1}{N} X_k X_l} & \text{for } 1 \leq i < R, 0 \leq k < l < R, \\ (\delta_{ik} - \delta_{iR}) \sqrt{\frac{1}{N} X_k \sum_{l\geq R} X_l} & \text{for } 0 \leq i < R, 0 \leq k < R, l = R, \\ 0 & \text{for } i = R, 0 \leq k < l < R, \\ -\sqrt{\frac{1}{N} X_k \sum_{l \geq R} X_l} & \text{for } i = l = R, 0 \leq k < R. \end{cases}$$

Now, for $i \leq j < R$, we have

$$\begin{aligned}\sum_{k=0}^{R}\sum_{l>k}^{R} V^i_{k,l}V^j_{k,l} &= \sum_{k=0}^{R-1}\sum_{l>k}^{R-1} V^i_{k,l}V^j_{k,l} + \sum_{k=0}^{R-1} V^i_{k,R}V^j_{k,R} \\ &= \delta_{ij}\left(\frac{1}{N}X_i\sum_{k=0}^{R-1}(1-\delta_{ik})X_k\right) + (1-\delta_{ij})(-\frac{1}{N}X_iX_j) + \delta_{ij}\left(\frac{1}{N}X_i\sum_{k\geq R}X_k\right) \\ &= \frac{1}{N}X_i\left(\delta_{ij}-X_j\right)\end{aligned}$$

and therefore

$$\sum_{k=0}^{R}\sum_{l>k}^{R} V^i_{k,l}V^j_{k,l}\cdot\nu = [M_i, M_j].$$

For $i < j = R$ we get

$$\begin{aligned}\sum_{k=0}^{R}\sum_{l>k}^{R} V^i_{k,l}V^R_{k,l}\cdot\nu &= -\sum_{k=0}^{R-1}\delta_{ik}\frac{1}{N}X_k\sum_{l\geq k}X_l\cdot\nu \\ &= -\frac{1}{N}X_i\left(1-\sum_{l=0}^{R-1}X_l\right)\cdot\nu \\ &= -\sum_{l=0}^{R-1}[M_i, M_l] = [M_i, 1-\sum_{l=0}^{R-1}M_l] = [M_i, M_R]\end{aligned}$$

as $[M_i, 1] = 0$. Finally, for $i = j = R$,

$$\begin{aligned}\sum_{k=0}^{R}\sum_{l>k}^{R} V^R_{k,l}V^R_{k,l}\cdot\nu &= \sum_{k=0}^{R-1}\frac{1}{N}X_k\left(1-\sum_{l=0}^{R-1}X_l\right)\cdot\nu \\ &= \sum_{k=0}^{R-1}\sum_{l=0}^{R-1}\frac{1}{N}X_k\left(\delta_{kl}-X_l\right)\cdot\nu \\ &= \sum_{k,l=0}^{R-1}[M_i, M_l] = [1-\sum_{k=0}^{R-1}M_k, 1-\sum_{l=0}^{R-1}M_l] = [M_R, M_R].\end{aligned}$$

Hence, we can apply the representation theorem, and get independent standard Brownian motions $(W_{kl})_{0\leq k<l\leq R}$ such that

$$M_i = \sum_{k=0}^{R}\sum_{l>k}^{R} V^i_{k,l}\cdot W_{k,l}$$

for $i \leq R$. Now, assume we have constructed two arrays of standard Brownian motions $(W_{kl})_{0\leq k<l\leq R}$ and $(\widetilde{W}_{kl})_{0\leq k<l\leq \widetilde{R}}$ by the above construction for R and $\widetilde{R}$ with $R < \widetilde{R}$. As the Brownian motions are independent within an array, we get

$$[W_{k,l}, W_{m,n}]_t = \delta_{km}\delta_{ln} t = [\widetilde{W}_{k,l}, \widetilde{W}_{m,n}]_t$$

for $k, l, m, n < R$. As a consequence of Lévy's characterization of the Brownian motion, $(W_{kl})_{0\leq k<l<R}$ and $(\widetilde{W}_{kl})_{0\leq k<l<R}$ are equal in distribution. Hence, their projective limit for $R \to \infty$ exists in form of independent standard Brownian motions $(W_{kl})_{0\leq k<l<\infty}$ with

$$M_i = \sum_{k=0}^{\infty}\sum_{l>k}^{\infty} V_{k,l}^i \cdot W_{k,l}.$$

Substituting this into (4.11) finishes the proof. □

4.2.2. Existence And Uniqueness Of A Solution

Now with Theorem 4.6 at hand, we can prove that SDE $(*)$ is well-posed by verifying the same for the martingale problem. Here, a corresponding result is known for bounded generators (Dawson [1993]). However, for Muller's ratchet $G_{\mathrm{sel}}^{\alpha}$ is not bounded for $\alpha > 0$. We therefore use a Girsanov theorem to reduce the general system to this case. First, we state the well-posedness of the SDE as the main result we want to prove in this section.

Theorem 4.7 (Well-posedness of the SDE)**.** *Let $\mathbf{P}_0 \in \mathcal{P}(\mathbb{S}^\circ)$ and $\underline{X}(0) \sim \mathbf{P}_0$ with $\mathbf{E}_{\mathbf{P}_0}[h_\xi(\underline{X}(0))] < \infty$ for all $\xi > 0$. Then, for $N \in (0,\infty)$ and $\alpha, \lambda, \gamma \in [0,\infty)$, the system $(*)$ starting in $\underline{X}(0)$ has a unique weak solution $\underline{X} = (\underline{X}(t))_{t\geq 0}$ in the space $\mathcal{C}_{\mathbb{S}^\circ}([0,\infty))$ of continuous functions on $\mathbb{S}^\circ$. We refer to the process $\underline{X}$ as* diffusion approximation of Muller's ratchet *(with compensatory mutations) with selection coefficient α, mutation rate λ, compensatory mutation rate γ and population size N.*

In terms of martingale problems, this theorem translates to the following proposition.

Proposition 4.8 (Well-posedness of the martingale problem)**.** *Let $\underline{X}(0)$ and $\mathbf{P}_0$ be as in Theorem 4.7, G^α as in Definition 4.3, $\alpha, \lambda, \gamma \in [0,\infty)$, $N \in (0,\infty)$ and $\overline{\mathcal{F}}$ be as in Definition 2.4. Then, the $(\mathbb{S}^\circ, \mathbf{P}_0, G_{\mathcal{X}}^\alpha, \overline{\mathcal{F}})$-martingale problem is well-posed and has continuous paths on $\mathbb{S}^\circ$. Its solution is equal in distribution to the diffusion approximation of Muller's ratchet with compensatory mutations.*

For the proof of Proposition 4.8, we need two well-known results from the theory of semimartingales. Briefly recall that a sequence of stopping times $(\tau_n)_{n\in\mathbb{N}}$ is said to be a *localizing sequence* if $\tau_n \nearrow \infty$ almost surely. An adapted process M is said to be a *local martingale* if there exists a localizing sequence such that the stopped process M^{τ_n} is a martingale for every n. Obviously, every martingale is a local martingale, while the reversal is not true in general, but under the following condition.

Lemma 4.9. *Let $\mathcal{N} = (N_t)_{t\geq 0}$ be a local martingale. If $\mathbf{E}[\sup_{0\leq t\leq T} N_t] < \infty$ for all $T > 0$, then $\mathcal{N}$ is a martingale.*

Proof. See e.g. Protter [2005], Theorem I.51. □

The second result is a Girsanov type theorem. It is the key step in the proof of Proposition 4.8, as it will allow us the reduce the martingale problem to the case without selection, where the generator is bounded.

Theorem 4.10 (Girsanov Theorem for continuous semimartingales)**.** *If $\mathcal{L} = (L_t)_{t\geq 0}$ is a continuous $\mathbf{P}$-martingale for some probability measure $\mathbf{P}$, then $\mathcal{Z} = (Z_t)_{t\geq 0}$, given by $Z_t = e^{L_t - \frac{1}{2}[\mathcal{L}]_t}$, is a continuous local martingale.*
If $\mathcal{Z}$ is a martingale as well, $\mathcal{N} = (N_t)_{t\geq 0}$ is a $\mathbf{P}$-local martingale and $\mathbf{Q}$ is defined via

$$\frac{d\mathbf{Q}}{d\mathbf{P}}\Big|_{\mathcal{F}_t} = Z_t,$$

then $\mathcal{N} - [\mathcal{L}, \mathcal{N}]$ is a $\mathbf{Q}$-local martingale.

Proof. See e.g. Kallenberg [1997], Theorem 16.19 and Lemma 16.21. □

Combining Lemma 4.9 and the Girsanov theorem, we can show now that a solution of the martingale problem for selection coefficient α is also a solution for the one with selection coefficient $\alpha' > \alpha$ on a modified probability space.

Proposition 4.11 (Change of measure)**.** *For $\underline{x} \in \mathbb{S}$, we denote with*

$$\kappa_1(\underline{x}) := \sum_{k=0}^{\infty} k x_k \quad \textit{and} \quad \kappa_2(\underline{x}) := \sum_{k=0}^{\infty} (k - \kappa_1(\underline{x}))^2 x_k \tag{4.12}$$

the expectation and variance of $\underline{x}$. Now let $\underline{x} \in \mathbb{S}^\circ$ and $\underline{X} = (\underline{X}(t))_{t\geq 0}$ be a solution of the $(\mathbb{S}^\circ, \delta_{\underline{x}}, G^\alpha, \overline{\mathcal{F}})$-martingale problem and denote its distribution by $\mathbf{P}^\alpha$. Then, the process $\mathcal{Z}^{\alpha,\alpha'} = (Z_t^{\alpha,\alpha'})_{t\geq 0}$, given by

$$\begin{aligned} Z_t^{\alpha,\alpha'} = \exp\Big(N(\alpha - \alpha')\Big(&\kappa_1(\underline{X}(t)) - \kappa_1(\underline{X}(0)) \\ &- \int_0^t \lambda - \gamma\kappa_1(\underline{X}(s)) - \frac{\alpha + \alpha'}{2}\kappa_2(\underline{X}(s))\, ds\Big)\Big) \end{aligned} \tag{4.13}$$

is a $\mathbf{P}^\alpha$-local martingale. If $\alpha' > \alpha$, it is even a $\mathbf{P}^\alpha$-martingale and the probability measure $\mathbf{P}^{\alpha'}$, defined by

$$\frac{d\mathbf{P}^{\alpha'}}{d\mathbf{P}^\alpha}\Big|_{\mathcal{F}_t} = Z_t^{\alpha,\alpha'},$$

solves the $(\mathbb{S}^\circ, \delta_{\underline{x}}, G^\alpha, \overline{\mathcal{F}}))$-martingale problem.

Remark. The quantities $\kappa_1(\underline{x})$ and $\kappa_2(\underline{x})$ are called the first and second cumulant of $\underline{x}$. Cumulants have proven to be useful for Muller's ratchet. We will take a closer look at them in Section 4.3.2.

Proof of Proposition 4.11. We know from Theorem 4.6 that $\underline{X}$ satisfies the assumptions of Lemma 4.4. The proof is an application of the Theorem (4.10). By assumption and Lemma 2.5 and Lemma 2.6, the processes $\underline{X}$ and $f(\underline{X})$ are continuous for all $f \in \overline{\mathcal{F}}$. Let again $\mathcal{N}^f = (N_t^f)_{t\geq 0}$ be

$$N_t^f := f(\underline{X}(t)) - f(\underline{X}(0)) - \int_0^t G_{\mathcal{X}}^\alpha f(\underline{X}(s))ds.$$

Now for

$$g(\underline{x}) := N(\alpha - \alpha')\kappa_1(\underline{x}) \in \overline{\mathcal{F}}$$

we know that $\mathcal{L} = (L_t)_{t\geq 0}$, defined by

$$L_t := N_t^g = N(\alpha - \alpha')\Big(\kappa_1(\underline{X}(t)) - \kappa_1(\underline{X}(0)) - \int_0^t G^\alpha \kappa_1(\underline{X}(s))ds\Big)$$

is a $\mathbf{P}^\alpha$-(local) martingale. As

$$\begin{aligned}
G_{\text{sel}}^\alpha \kappa_1(\underline{x}) &= \alpha \sum_{k=0}^\infty (\kappa_1(\underline{x}) - k)x_k k = \alpha \left(\kappa_1^2(\underline{x}) - \sum_{k=0}^\infty k^2 x_k\right) = -\alpha\kappa_2(\underline{x}), \\
G_{\text{mut}}\kappa_1(\underline{x}) &= \lambda \sum_{k=0}^\infty (x_{k-1} - x_k)k = \lambda \sum_{k=0}^\infty (k+1)x_k - kx_k = \lambda, \\
G_{\text{cm}}\kappa_1(\underline{x}) &= \gamma \sum_{k=0}^\infty \left((k+1)x_{k+1} - kx_k\right)k = \gamma \sum_{k=0}^\infty (k-1)kx_k - k^2 x_k = -\gamma\kappa_1(\underline{x}),
\end{aligned}$$

$G_{\text{res}}^N \kappa_1(\underline{x}) = 0$ and

$$G_{\text{res}}^N \kappa_1^2(\underline{x}) = \frac{1}{2N}\sum_{l,k=0}^\infty x_k(\delta_{kl} - x_l)2kl = \frac{1}{N}\sum_{k=0}^\infty x_k(k^2 - k\kappa_1(\underline{x})) = \frac{1}{N}\kappa_2(\underline{x}),$$

we have

$$[\mathcal{L}]_t = N^2(\alpha - \alpha')^2 \int_0^t G_{\text{res}}^N \kappa_1^2(\underline{X}(s))\, ds = N(\alpha - \alpha')^2 \int_0^t \kappa_2(\underline{X}(s))ds$$

and

$$\begin{aligned}
&\exp\Big(L_t - \frac{1}{2}[\mathcal{L}]_t\Big) \\
&= \exp\Big(N(\alpha-\alpha')\Big(\kappa_1(\underline{X}(t)) - \kappa_1(\underline{X}(0)) - \int_0^t \lambda - \gamma\kappa_1(\underline{X}(s)) - \frac{\alpha+\alpha'}{2}\kappa_2(\underline{X}(s))ds\Big)\Big) \\
&= Z_t^{\alpha,\alpha'}
\end{aligned}$$

is a local $\mathbf{P}^\alpha$-martingale by Theorem 4.10. Now, if $\alpha < \alpha'$ and $\underline{X}(0)$ has all exponential moments, we have that $\mathbf{E}[\sup_{0\le t\le T} Z_t^{\alpha,\alpha'}] < \infty$ since $e^{\xi\kappa_1(\underline{X}(t))} \le h_\xi(\underline{X}(t))$ by Jensen's inequality for all ξ. Hence, $\mathcal{Z}^{\alpha,\alpha'}$ even is a $\mathbf{P}^\alpha$-martingale by Lemma 4.9.
Now let $f \in \overline{\mathcal{F}}$. By assumption, N^f is a $\mathbf{P}^\alpha$-martingale. We calculate

$$\begin{aligned}
[\mathcal{L}, \mathcal{N}^f]_t &= \int_0^t G_{\mathrm{res}}^N g f(\underline{X}(s)) - g(\underline{X}(s)) G_{\mathrm{res}}^N f(\underline{X}(s))\, ds \\
&= \frac{(\alpha-\alpha')}{2}\int_0^t \sum_{k,l=0}^\infty X_k(s)(\delta_{kl} - X_l(s))\left(l\frac{\partial}{\partial x_k} f + k\frac{\partial}{\partial x_l} f\right) ds \\
&= \frac{(\alpha-\alpha')}{2}\int_0^t \Big(\sum_{k=0}^\infty X_k(s)\left(k - \kappa_1(\underline{X}(s))\right)\frac{\partial}{\partial x_k} f(\underline{X}(s)) \\
&\qquad\qquad + \sum_{l=0}^\infty X_l(s)\left(l - \kappa_1(\underline{X}(s))\right)\frac{\partial}{\partial x_l} f(\underline{X}(s))\Big) ds \\
&= \int_0^t G_{\mathrm{sel}}^{\alpha'} f(\underline{X}(s)) - G_{\mathrm{sel}}^{\alpha} f(\underline{X}(s)) ds.
\end{aligned}$$

Therefore the Girsanov theorem for continuous semimartingales shows that

$$N_t^f - \Big[\mathcal{L}, \mathcal{N}^f\Big]_t = f(\underline{X}(t)) - \int_0^t G^{\alpha'} f(\underline{X}(s)) ds$$

is a $\mathbf{P}^{\alpha'}$-martingale. Hence, $\mathbf{P}^{\alpha'}$ solves the $(\mathbb{S}^\circ, \mathbf{P}_0, G_{\mathcal{X}}^{\alpha'}, \overline{\mathcal{F}})$-martingale problem. □

Using this result, we can now prove the well-posedness of the stochastic differential equation $(*)$.

Proof of Theorem 4.7. The assertions of Theorem 4.7 and Proposition 4.8 are equivalent, as can be seen from Theorem 4.6. So, it suffices to show Proposition 4.8. Moreover, it is enough to consider the case $\mathbf{P}_0 = \delta_{\underline{x}}$ for $\underline{x} \in \mathbb{S}^\circ$.

For $\alpha = 0$, it can be seen from classical theory (e.g. Dawson, 1993, Theorem 5.4.1) that the $(\mathbb{S}, \delta_{\underline{x}}, G^0, \mathcal{F})$-martingale problem has a unique weak solution. Thus, by Theorem 4.6, the $(\mathbb{S}, \delta_{\underline{x}}, G^0, \overline{\mathcal{F}})$-martingale problem is well-posed as well. Denote the unique distribution which solves the martingale problem by $\mathbf{P}^0$. Hence, by Proposition 4.11, there is a change of measure using $\mathcal{Z}^{0,\alpha}$ from $\mathbf{P}^0$ to $\mathbf{P}^\alpha$ and $\mathbf{P}^\alpha$ solves the $(\mathbb{S}^\circ, \mathbf{P}_0, G^\alpha, \overline{\mathcal{F}})$-martingale problem. This establishes the existence.

For uniqueness, recall that the solution of the $(\mathbb{S}, \delta_{\underline{x}}, G^0, \mathcal{F})$-martingale problem is unique. We proceed by contradiction and assume that $\mathbf{P}_1^\alpha$ and $\mathbf{P}_2^\alpha$ are two different solutions of the $(\mathbb{S}^\circ, \delta_{\underline{x}}, G^\alpha, \overline{\mathcal{F}})$-martingale problem. Let $\tau_1, \tau_2, ...$ be a localizing sequence such that $(Z^{\alpha,0}_{t\wedge\tau_n})_{t\geq 0}$, given by (4.13), is both a $\mathbf{P}_1^\alpha$-martingale and a $\mathbf{P}_2^\alpha$-martingale. Since $\mathbf{P}_1^\alpha \neq \mathbf{P}_2^\alpha$, there must be $t \geq 0$ such that the distribution of $\underline{X}(t)$ under $\mathbf{P}_1^\alpha$ and $\mathbf{P}_2^\alpha$ is different (see Theorem 4.4.2 in Ethier and Kurtz [2005]). Hence, there is $f \in \mathcal{F}$ and $n \in \mathbb{N}$ such that $\mathbf{E}_{\mathbf{P}_1^\alpha}[f(\underline{X}(t\wedge\tau_n))] \neq \mathbf{E}_{\mathbf{P}_2^\alpha}[f(\underline{X}(t\wedge\tau_n))]$ as well as

$$\mathbf{E}_{\mathbf{P}_1^\alpha}[Z^{\alpha,0}_{t\wedge\tau_n} f(\underline{X}(t\wedge\tau_n))] \neq \mathbf{E}_{\mathbf{P}_2^\alpha}[Z^{\alpha,0}_{t\wedge\tau_n} f(\underline{X}(t\wedge\tau_n))]. \tag{4.14}$$

However, by the same arguments as in the proof of Proposition 4.11, $Z^{\alpha,0}_{t\wedge\tau_n} \cdot \mathbf{P}_1^\alpha$ as well as $Z^{\alpha,0}_{t\wedge\tau_n} \cdot \mathbf{P}_2^\alpha$ equal $\mathbf{P}^0$ on the σ-algebra $\sigma((\underline{X}(s))_{0\leq s\leq t\wedge\tau_n})$. Hence,

$$\mathbf{E}_{\mathbf{P}_1^\alpha}[Z^{\alpha,0}_{t\wedge\tau_n} f(\underline{X}(t\wedge\tau_n))] = \mathbf{E}_{\mathbf{P}^0}[f(\underline{X}(t\wedge\tau_n))] = \mathbf{E}_{\mathbf{P}_2^\alpha}[Z^{\alpha,0}_{t\wedge\tau_n} f(\underline{X}(t\wedge\tau_n))],$$

in contradiction to (4.14). Uniqueness of the $(\mathbb{S}^\circ, \mathbf{P}_0, G_{\mathcal{X}}, \overline{\mathcal{F}})$-martingale problem follows. □

4.3. The ODE Approximation

Now as we have the well-posedness of the diffusion approximation, we will again consider a deterministic model that approximates the stochastic process if the population size parameter N is large. As mentioned before, the SDE turns into an ODE if we increase N towards infinity (we will specify the type of this convergence later). Hence, we use the deterministic part of the SDE to define the new model.

Definition 4.12. We say a function $\underline{x} : [0, \infty) \mapsto \mathbb{S}$ with $\underline{x}(t) = (x_k(t))_{k\in\mathbb{N}}$ is an *ODE Approximation of Muller's ratchet with compensatory mutations*, if it is a solution of the ordinary differential equation on $\mathbb{S}$ defined by

$$\dot{x}_k = -\alpha\Big(\sum_{\ell=0}^{\infty}(\ell - k)x_\ell\Big)x_k + \lambda(x_{k-1} - x_k) + \gamma((k+1)x_{k+1} - kx_k) \qquad (**)$$

for all $k \in \mathbb{N}$ with $x_{-1} := 0$.

Similar to Chapter 3.2, we will show that the weak limit of the diffusion approximation for $N \to \infty$ solves ODE $(**)$ and that it is its only solution. Then, we will explicitly calculate the solution and derive the long time behavior of the model. This will give us the existence of a unique equilibrium point.

4.3.1. Convergence And Uniqueness

We start with the mentioned convergence that gives us existence of a solution as well as proving the acclaimed approximation behavior.

Theorem 4.13. *Let $\lambda, \alpha \in [0, \infty), \gamma \in (0, \infty)$ and $\underline{x}(0) \in \mathbb{S}^\circ$. If $\underline{X}^{N_i} = \big(\underline{X}^{N_i}(t)\big)_{t\geq 0}$ are solutions of the SDE $(*)$ starting in $\underline{x}(0)$ for $N_1, N_2, \ldots \in \mathbb{N}$ with $N_i \to \infty$, then their weak limit in the space of continuous $\mathbb{S}^\circ$-valued paths $\underline{x} = (\underline{x}(t))_{t\geq 0}$ exists and is a solution of the ODE $(**)$ almost surely.*

Proof. Note that $\underline{X}^{N_i}$ is the solution of the $(\mathbb{S}^\circ, \delta_{\underline{x}}, G^\alpha_{\text{sel}} + G_{\text{mut}} + G_{\text{cm}} + G^{N_i}_{\text{res}}, \mathcal{F})$-martingale problem, while obviously any solution of the $(\mathbb{S}^\circ, \delta_{\underline{x}}, G^\alpha_{\text{sel}} + G_{\text{mut}} + G_{\text{cm}}, \mathcal{F})$-martingale problem is a solution of differential equation $(**)$ almost surely. Since we have for all $f \in \mathcal{F}$ that

$$\sup_{\underline{x}\in\mathbb{S}^\circ} |G^N_{\text{res}} f(\underline{x})| \xrightarrow{N\to\infty} 0$$

as (4.8) also holds for the Uniform norm on $\mathbb{S}^\circ$ by Lemma 4.4, the assertions follows if we can show the compact containment condition (see Ethier and Kurtz [2005], Lemma 4.5.1 and Remark 4.5.2). That means, we have to show that for all $\varepsilon > 0$ and

$T > 0$ there is a compact set $K_{\varepsilon,T} \subseteq \mathbb{S}^\circ$ (with respect to the topology generated by d_P°) such that

$$\sup_N \mathbf{P}(\underline{X}^N(t) \in K_{\varepsilon,T}^c \text{ for a } t \in [0,T]) \leq \varepsilon.$$

Fix $\varepsilon > 0$ and $T > 0$. For $n \in \mathbb{N}$, there is C_n by Lemma 4.4 such that

$$\sup_N \mathbf{P}(\sup_{0 \leq t \leq T} h_n(\underline{X}(t)) > C_n) < \varepsilon 2^{-n}.$$

By Lemma 2.3, a set $K \subseteq \mathbb{S}^\circ$ is relatively compact if for all $\xi > 0$ there is a finite C with $h_\xi(\underline{x}) \leq C$ for all $\underline{x} \in K$. Hence, the closure of

$$K_{\varepsilon,T} := \bigcap_{n=1}^\infty \{\underline{x} \in \mathbb{S}^\circ : h_n(\underline{x}) \leq C_n\}$$

is compact and

$$\sup_N \mathbf{P}(\underline{X}^N(t) \in K_{\varepsilon,T}^c) \leq \sup_N \sum_{n=1}^\infty \mathbf{P}(\sup_{0 \leq t \leq T} h_n(\underline{X}(t)) > C_n) \leq \varepsilon \sum_{n=1}^\infty 2^{-n} = \varepsilon.$$

□

To prove that ODE (∗∗) has no other solutions, we require some preliminary computations which we carry out next. For $\underline{x} \in \mathbb{S}^\circ$, we refer to the function $\xi \mapsto \log h_\xi(\underline{x})$ as the cumulant generating function of $\underline{x}$.

Proposition 4.14 (Dynamics of cumulant generating function)**.** *For any solution $t \mapsto \underline{x}(t)$ of (∗∗) taking values in $\mathbb{S}^\circ$,*

$$\frac{d}{dt} \log h_\xi(\underline{x}(t)) = \alpha \sum_{\ell=0}^\infty \ell x_\ell(t) + \lambda(e^\xi - 1) - (\alpha + \gamma(1 - e^{-\xi})) \frac{d}{d\xi} \log h_\xi(\underline{x}(t)).$$

Proof. Abbreviating $\underline{x} := \underline{x}(t)$, we compute

$$\begin{aligned} h_\xi(\underline{x}) \frac{d}{dt} \log h_\xi(\underline{x}) &= \alpha \sum_{\ell=0}^\infty \sum_{k=0}^\infty (\ell - k) x_\ell x_k e^{\xi k} + \lambda \sum_{k=0}^\infty (x_{k-1} - x_k) e^{\xi k} \\ &\qquad + \gamma \sum_{k=0}^\infty ((k+1) x_{k+1} - k x_k) e^{\xi k} \\ &= \alpha \Big(\Big(\sum_{\ell=0}^\infty \ell x_\ell \Big) h_\xi(\underline{x}) - \frac{d}{d\xi} h_\xi(\underline{x}) \Big) + \lambda(e^\xi - 1) h_\xi(\underline{x}) - \gamma(1 - e^{-\xi}) \frac{d}{d\xi} h_\xi(\underline{x}) \end{aligned}$$

and so

$$\frac{d}{dt} \log h_\xi(\underline{x}) = \alpha \sum_{\ell=0}^\infty \ell x_\ell + \lambda(e^\xi - 1) - (\alpha + \gamma(1 - e^{-\xi})) \frac{d}{d\xi} \log h_\xi(\underline{x}).$$ □

The equation in Proposition 4.14 relates the time-derivative of $\log h_\xi(\underline{x}(t))$ with its ξ-derivative. Such a connection can be very useful, for example when studying so-called duality for Markov processes (see e.g. Ethier and Kurtz [2005], p. 188ff). The next result is a special application of these ideas.

Corollary 4.15 (Duality)**.** *Let $t \mapsto \underline{x}(t)$ and $t \mapsto \underline{y}(t)$ be two solutions of $(**)$ taking values in $\mathbb{S}^\circ$. Moreover $\xi : t \mapsto \xi(t)$ is the solution of $\xi' = -(\alpha + \gamma(1 - e^{-\xi}))$, starting in some $\xi(0)$. Then,*

$$\begin{aligned}\log h_{\xi(0)}(\underline{x}(t)) - \log h_{\xi(0)}(\underline{y}(t)) = \log h_{\xi(t)}(\underline{x}(0)) - \log h_{\xi(t)}(\underline{y}(0))\\ + \int_0^t \sum_{\ell=0}^{\infty} \ell(x_\ell(s) - y_\ell(s)) ds.\end{aligned}$$

Proof. Using Proposition 4.14 and since for any differentiable function $g : \xi \mapsto g(\xi)$,

$$\frac{d}{ds} g(\xi(t-s)) = (\alpha + \gamma(1 - e^{-\xi(t-s)})) \frac{d}{d\xi} g(\xi(t-s)),$$

we obtain

$$\frac{d}{ds}\Big(\log h_{\xi(t-s)}(\underline{x}(s)) - \log h_{\xi(t-s)}(\underline{y}(s)) \Big) = \alpha \sum_{\ell=0}^{\infty} \ell(x_\ell(s) - y_\ell(s)).$$

Now the assertion follows by integrating. □

We use this duality relation to show the uniqueness of the solution of the differential equation.

Theorem 4.16. *The solution of ODE $(**)$ is unique.*

Proof. If $\underline{x}(0) = \underline{y}(0)$, we obtain from Corollary 4.15 that for all $\xi \in \mathbb{R}$ and any $t \geq 0$,

$$\log h_\xi(\underline{x}(t)) - \log h_\xi(\underline{y}(t)) = \int_0^t \sum_{\ell=0}^{\infty} \ell(x_\ell(s) - y_\ell(s)) ds. \tag{4.15}$$

Observing that only the left-hand side depends on ξ, this can only be true if both sides are 0. Formally, taking derivatives with respect to ξ at $\xi = 0$ the last equation gives

$$\sum_{\ell=0}^{\infty} \ell(x_\ell(t) - y_\ell(t)) = 0.$$

Plugging this back into (4.15) gives

$$\log h_\xi(\underline{x}(t)) = \log h_\xi(\underline{y}(t)). \tag{4.16}$$

Since the function $\xi \mapsto \log h_\xi(\underline{x})$ characterizes $\underline{x} \in \mathbb{S}^\circ$ (e.g. see Etheridge et al. [2008]), we obtain that $\underline{x}(t) = \underline{y}(t)$. □

4.3.2. Cumulants

Now we again turn our attention to equilibrium points. First we define them for time-continuous processes as well.

Definition 4.17. We say $\underline{x} \in \mathbb{S}$ is an *equilibrium point* of a time-continuous Markov process $(\underline{Z}(t))_{t\geq 0}$ on $\mathbb{S}$ if

$$\mathbf{E}[\,\underline{Z}(t) \,|\, Z(0) = \underline{x}\,] = \underline{x}$$

for all $t \geq 0$.

In the proof of Theorem 4.6, we have seen that

$$\sum_{l\neq k} \int_0^t \sqrt{\frac{1}{N} X_k(s)\, X_l(s)}\, dW_{kl}(s)$$

is a martingale. Hence, the equilibrium points of the diffusion approximation $\underline{X}$ are again exactly the ones of the ODE approximation $\underline{x}$. Furthermore they are exactly the $\underline{x} \in \mathbb{S}$ with $b_i(x) = 0$, where b_i is as in $(*)$.

Depperschmidt et al. [2011] have identified one of these points by using cumulants, which are the coefficients of the Taylor series of the cumulant generation function.

Definition 4.18 (Cumulants)**.** Let $\underline{x} \in \mathbb{S}^\circ$. The cumulants $(\kappa_k(\underline{x}))_{k=1,2,\dots}$ of $\underline{x}$ are defined by the relation

$$\log \sum_{k=0}^{\infty} x_k e^{\xi k} = \sum_{k=1}^{\infty} \kappa_k(\underline{x}) \frac{\xi^k}{k!}.$$

Note that the cumulants of $\underline{x}$ are closely related to its moments. While κ_1 is equal to the first moment of $\underline{x}$, κ_2 is its variance (in agreement to our notation in Proposition 4.11). A general recursive formula for this relation is also known in the literature (e.g. Ledermann and Lloyd [1981]). We have used before that the cumulant generating function determines an $\underline{x} \in \mathbb{S}^\circ$. The cumulants also do so, as they determine the cumulant generating function. For a more detailed review of cumulants, see Burger [1991].

Mainly because the cumulants of a Poisson distribution, say with parameter θ, have the particularly nice form $\kappa_1 = \kappa_2 = \ldots = \theta$, cumulants have proven to be useful for Muller's ratchet. The system $(**)$ translates nicely into cumulants as well.

Proposition 4.19. *Let $\underline{x} = \underline{x}(t)$ be the solution of the ODE $(**)$. Then*

$$\frac{d}{dt}\kappa_k(\underline{x}) = -\alpha\kappa_{k+1}(\underline{x}) + \lambda + \gamma \sum_{i=1}^{k} (-1)^{i+k-1} \binom{k}{i-1} \kappa_i(\underline{x}) \qquad (4.17)$$

Proof. For this proof, we abbreviate $\kappa_k(\underline{x}(t))$ with κ_k. Loosely following Depperschmidt et al. [2011] we notice that

$$\frac{-\sum_{k=0}^{\infty} kx_k e^{-\xi k}}{\sum_{k=0}^{\infty} x_k e^{-\xi k}} = \frac{d}{d\xi} \log \sum_{k=0}^{\infty} x_k e^{-\xi k} = \frac{d}{d\xi} \sum_{k=1}^{\infty} \kappa_k \frac{(-\xi)^k}{k!} = -\sum_{k=0}^{\infty} \kappa_{k+1} \frac{(-\xi)^k}{k!} \tag{4.18}$$

and calculate

$$\begin{aligned}
\frac{d}{dt} \sum_{k=1}^{\infty} \kappa_k \frac{(-\xi)^k}{k!} &= \frac{d}{dt} \left(\log \sum_{k=0}^{\infty} x_k e^{-\xi k} \right) = \frac{\sum_{k=0}^{\infty} e^{-\xi k} \frac{d}{dt} x_k}{\sum_{k=0}^{\infty} x_k e^{-\xi k}} \\
&= \frac{\sum_{k=0}^{\infty} e^{-\xi k} \left[(\alpha\kappa_1 - \lambda)\, x_k - (\alpha + \gamma)\, kx_k + \gamma(k+1)x_{k+1} + \lambda x_{k-1} \right]}{\sum_{k=0}^{\infty} x_k e^{-\xi k}} \\
&= \alpha\kappa_1 - \lambda + \frac{-(\alpha+\gamma) \sum_{k=0}^{\infty} kx_k e^{-\xi k} + \gamma e^{\xi} \sum_{k=1}^{\infty} kx_k e^{-\xi k}}{\sum_{k=0}^{\infty} x_k e^{-\xi k}} + \lambda e^{-\xi} \\
&= \alpha\kappa_1 - \lambda - (\alpha + \gamma - \gamma e^{\xi}) \frac{\sum_{k=0}^{\infty} kx_k e^{-\xi k}}{\sum_{k=0}^{\infty} x_k e^{-\xi k}} + \lambda e^{-\xi} \\
&\overset{4.18}{=} \alpha\kappa_1 - \lambda - (\alpha + \gamma - \gamma e^{\xi}) \sum_{k=0}^{\infty} \kappa_{k+1} \frac{(-\xi)^k}{k!} + \lambda e^{-\xi} \\
&= -\alpha \sum_{k=1}^{\infty} \kappa_{k+1} \frac{(-\xi)^k}{k!} + \lambda(e^{-\xi} - 1) + \gamma(e^{\xi} - 1) \sum_{k=0}^{\infty} \kappa_{k+1} \frac{(-\xi)^k}{k!}.
\end{aligned} \tag{4.19}$$

To identify the κ_k's, we will now calculate the n-th derivative of the right-hand side of (4.19). To do so, we prove by induction that

$$\frac{d^n}{d^n\xi} \left(\gamma e^{\xi} \sum_{k=0}^{\infty} \kappa_{k+1} \frac{(-\xi)^k}{k!} \right) = \gamma e^{\xi} \sum_{k=0}^{\infty} \sum_{i=0}^{n} (-1)^i \binom{n}{i} \kappa_{k+i+1} \frac{(-\xi)^k}{k!}. \tag{4.20}$$

First

$$\begin{aligned}
\frac{d}{d\xi} \left(\gamma e^{\xi} \sum_{k=0}^{\infty} \kappa_{k+1} \frac{(-\xi)^k}{k!} \right) &= \gamma e^{\xi} \sum_{k=0}^{\infty} \kappa_{k+1} \frac{(-\xi)^k}{k!} - \gamma e^{\xi} \sum_{k=0}^{\infty} \kappa_{k+2} \frac{(-\xi)^k}{k!} \\
&= \gamma e^{\xi} \sum_{k=0}^{\infty} (\kappa_{k+1} - \kappa_{k+2}) \frac{(-\xi)^k}{k!} \\
&= \gamma e^{\xi} \sum_{k=0}^{\infty} \sum_{i=0}^{1} (-1)^i \binom{1}{i} \kappa_{k+i+1} \frac{(-\xi)^k}{k!}
\end{aligned}$$

and

$$\begin{aligned}
\frac{d^{n+1}}{d^{n+1}\xi}\left(\gamma e^{\xi}\sum_{k=0}^{\infty}\kappa_{k+1}\frac{(-\xi)^k}{k!}\right) &= \frac{d}{d\xi}\left(\gamma e^{\xi}\sum_{k=0}^{\infty}\sum_{i=0}^{n}(-1)^i\binom{n}{i}\kappa_{k+i+1}\frac{(-\xi)^k}{k!}\right)\\
&= \gamma e^{\xi}\sum_{k=0}^{\infty}\sum_{i=0}^{n}(-1)^i\binom{n}{i}\kappa_{k+i+1}\frac{(-\xi)^k}{k!} - \gamma e^{\xi}\sum_{k=0}^{\infty}\sum_{i=0}^{n}(-1)^i\binom{n}{i}\kappa_{k+i+2}\frac{(-\xi)^k}{k!}\\
&= \gamma e^{\xi}\sum_{k=0}^{\infty}\sum_{i=0}^{n+1}(-1)^i\left(\binom{n}{i}+\binom{n}{i-1}\right)\kappa_{k+i+1}\frac{(-\xi)^k}{k!}\\
&= \gamma e^{\xi}\sum_{k=0}^{\infty}\sum_{i=0}^{n+1}(-1)^i\binom{n+1}{i}\kappa_{k+i+1}\frac{(-\xi)^k}{k!}.
\end{aligned}$$

As the other summands are easy to derive, (4.20) yields

$$\begin{aligned}
&\frac{d^n}{d^n\xi}\left(-\alpha\sum_{k=1}^{\infty}\kappa_{k+1}\frac{(-\xi)^k}{k!} + \lambda(e^{-\xi}-1) + \gamma(e^{\xi}-1)\sum_{k=0}^{\infty}\kappa_{k+1}\frac{(-\xi)^k}{k!}\right)\\
&= (-1)^n(-\alpha)\sum_{k=0}^{\infty}\kappa_{k+n+1}\frac{(-\xi)^k}{k!} + (-1)^n\lambda e^{-\xi} + \gamma e^{\xi}\sum_{k=0}^{\infty}\sum_{i=0}^{n}(-1)^i\binom{n}{i}\kappa_{k+i+1}\frac{(-\xi)^k}{k!}\\
&\quad + (-1)^n(-\gamma)\sum_{k=0}^{\infty}\kappa_{k+n+1}\frac{(-\xi)^k}{k!}\\
&= (-1)^n\left[-\alpha\sum_{k=0}^{\infty}\kappa_{k+n+1}\frac{(-\xi)^k}{k!} + \lambda e^{-\xi} + \gamma e^{\xi}\sum_{k=0}^{\infty}\sum_{i=1}^{n}(-1)^{i+n-1}\binom{n}{i-1}\kappa_{k+i}\frac{(-\xi)^k}{k!}\right]
\end{aligned}$$

Now deriving the left-hand side of (4.19), we get

$$\frac{d^n}{d^n\xi}\left(\frac{d}{dt}\sum_{k=1}^{\infty}\kappa_k\frac{(-\xi)^k}{k!}\right) = \frac{d^n}{d^n\xi}\left(\sum_{k=1}^{\infty}\frac{d}{dt}\kappa_k\frac{(-\xi)^k}{k!}\right) = (-1)^n\sum_{k=0}^{\infty}\frac{d}{dt}\kappa_{k+n}\frac{(-\xi)^k}{k!}.$$

Evaluation of both functions at $\xi = 0$ finishes the proof. □

Now, observe that for $\kappa_1 = \kappa_2 = \ldots = \theta$ the right-hand side of (4.17) equals

$$-\alpha\theta + \lambda - \gamma\theta = -(\alpha+\gamma)\theta + \lambda$$

what yields the following corollary.

Corollary 4.20. *The Poisson distribution with parameter $\theta = \frac{\lambda}{\alpha+\gamma}$ is an equilibrium point of $\underline{x}$.*

In fact, it is also its only equilibrium point. To prove this, we need to solve the system $(**)$.

4.3.3. The Explicit Solution

Using a stochastic particle system, we now explicitly calculate the solution of ODE $(**)$. Similar to the Maia-Botelho-Fontanari-Theorem, this is the key result for this chapter.

Theorem 4.21. *The solution of the system* $(**)$ *is*

$$x_k(t) = \frac{\sum_{i=0}^{\infty} x_i(0) \sum_{j=0}^{i\wedge k} \binom{i}{j} \Big(\frac{\gamma}{\alpha+\gamma}\big(1 - e^{-(\alpha+\gamma)t}\big)\Big)^{i-j} e^{-j(\alpha+\gamma)t} \frac{1}{(k-j)!} \Big(\frac{\lambda}{\alpha+\gamma}(1 - e^{-(\alpha+\gamma)t})\Big)^{k-j}}{\sum_{i=0}^{\infty} x_i(0) \Big(\frac{\gamma}{\alpha+\gamma} - \frac{\alpha}{\alpha+\gamma} e^{-(\alpha+\gamma)t}\Big)^{i} \exp\Big(\frac{\lambda}{\alpha+\gamma}(1 - e^{-(\alpha+\gamma)t})\Big)}. \tag{4.21}$$

To prove this theorem, we define a Markov jump process that counts the mutations of Muller's ratchet.

Definition 4.22. We define the process $(K_t)_{t\geq 0}$ as the pure Markov jump process that takes values in $\{\dagger, 0, 1, 2, ...\}$, never leaves the state $\dagger$ and jumps

1. from k to $k+1$ at rate λ and
2. from k to $k-1$ at rate $k\gamma$.
3. In addition, the process is killed at k with rate αk, i.e. it jumps to $\dagger$.

This process is closely related to the system $(**)$ by the following proposition.

Proposition 4.23 (Particle representation)**.** *Let* $\underline{x}(0) \in \mathbb{S}^\circ$ *and* $(K_t)_{t\geq 0}$ *be as in Definition 4.22 with initial distribution given by* $\mathbf{P}[K_0 = k] = x_k(0)$*. Then,*

$$x_k(t) := \mathbf{P}[K_t = k | K_t \neq \dagger] \tag{4.22}$$

solves the system in Theorem 4.21.

Proof. From the definition of $(K_t)_{t\geq 0}$, it is clear that for small $\varepsilon > 0$,

$$\begin{aligned} x_k(t+\varepsilon) &= \frac{x_k(t)(1-\alpha k\varepsilon) + \lambda(x_{k-1}(t) - x_k(t))\varepsilon + \gamma((k+1)x_{k+1}(t) - kx_k(t))\varepsilon}{1 - \alpha\sum_j jx_j(t)\varepsilon} + \mathcal{O}(\varepsilon^2) \\ &= x_k(t) + \Big(-\alpha\Big(kx_k(t) - \sum_{j=0}^{\infty} jx_j(t)\Big)x_k(t) + \lambda(x_{k-1}(t) - x_k(t)) \\ &\qquad + \gamma((k+1)x_{k+1}(t) - kx_k(t))\Big)\varepsilon + \mathcal{O}(\varepsilon^2), \end{aligned}$$

which implies the result as $\varepsilon \to 0$. □

Hence we can prove Theorem 4.21 by computing $\mathbf{P}[K_t = k \,|\, K_t \neq \dagger]$. To do so, we need the following notation and a technical lemma.

Definition 4.24. Denote by

$$S_{l,k} := \{\sigma \in S_l : \sigma(1) < \ldots < \sigma(l-k), \sigma(l-k+1) < \ldots < \sigma(l)\}$$

the set of all permutations of $1, \ldots, l$ with two groups of increasing elements of size $l-k$ and k.

Note that, as the elements of one group determine the ones of the other, the size of $S_{l,k}$ is

$$|S_{l,k}| = \binom{l}{k} = \binom{l}{k-l}.$$

The technical lemma basically states that the integrand has equal values on all polytopes gained by connecting l edges of an l-dimensional simplex that share a common vertice.

Lemma 4.25. *Let $f, g : [0, \infty) \mapsto \mathbb{R}$ be bounded functions, $t \in [0, \infty)$ and $l, k \in \mathbb{N}$ with $k \leq l$. We can calculate the following integral as*

$$\begin{aligned} &\int\limits_{0 \leq t_1 \leq \ldots \leq t_l \leq t} \sum_{\sigma \in S_{l,k}} f(t_{\sigma(1)}) \cdots f(t_{\sigma(l-k)}) g(t_{\sigma(l-k+1)}) \cdots g(t_{\sigma(l)})\, d(t_1, \ldots, t_l) \\ = \;& \frac{1}{l!} \binom{l}{l-k} \left(\int_0^t f(t)dt \right)^{l-k} \left(\int_0^t g(t)dt \right)^k \end{aligned}$$

Proof. Once we have shown the above geometrical interpretation of the lemma, i.e. that

$$\begin{aligned} &\int\limits_{0 \leq t_1 \leq \ldots \leq t_l \leq t} \sum_{\sigma \in S_{l,k}} f(t_{\sigma(1)}) \ldots f(t_{\sigma(l-k)}) g(t_{\sigma(l-k+1)}) \ldots g(t_{\sigma(l)})\, d(t_1, \ldots, t_l) \\ = &\int\limits_{0 \leq t_{\tau(1)} \leq \ldots \leq t_{\tau(l)} \leq t} \sum_{\sigma \in S_{l,k}} f(t_{\sigma(1)}) \ldots f(t_{\sigma(l-k)}) g(t_{\sigma(l-k+1)}) \ldots g(t_{\sigma(l)})\, d(t_1, \ldots, t_l) \end{aligned} \tag{4.23}$$

holds for all $\tau \in S_l$, the assertion follows directly as

$$\begin{aligned}
&\int\limits_{0\leq t_1\leq\ldots\leq t_l\leq t} \sum_{\sigma\in S_{l,k}} f(t_{\sigma(1)})\ldots f(t_{\sigma(l-k)})g(t_{\sigma(l-k+1)})\ldots g(t_{\sigma(l)})d(t_1,\ldots,t_l)\\
=&\frac{1}{l!}\int\limits_{[0,t]^l} \sum_{\sigma\in S_{l,k}} f(t_{\sigma(1)})\ldots f(t_{\sigma(l-k)})g(t_{\sigma(l-k+1)})\ldots g(t_{\sigma(l)})d(t_1,\ldots,t_l)\\
=&\frac{1}{l!}\sum_{\sigma\in S_{l,k}}\int\limits_{[0,t]^l} f(t_{\sigma(1)})\ldots f(t_{\sigma(l-k)})g(t_{\sigma(l-k+1)})\ldots g(t_{\sigma(l)})d(t_1,\ldots,t_l)\\
=&\frac{1}{l!}\sum_{\sigma\in S_{l,k}}\int\limits_{[0,t]^l} f(t_1)\ldots f(t_{l-k})g(t_{l-k+1})\ldots g(t_l)d(t_1,\ldots,t_l)\\
=&\frac{1}{l!}\binom{l}{l-k}\left(\int_0^t f(t)dt\right)^{l-k}\left(\int_0^t g(t)dt\right)^k,
\end{aligned}$$

using Fubini's theorem in the last two steps as well as that $|S_l| = l!$ and $|S_{l,k}| = \binom{l}{l-k}$. For (4.23), notice that for $\sigma \in S_{l,k}$ and $\tau \in S_l$, $\sigma \circ \tau \notin S_{l,k}$ in general. However, there exists exactly one permutation $\pi = \pi_{\tau,\sigma} \in S_l$ that just sorts the elements of each $\sigma\circ\tau(\{1,\ldots,l-k\})$ and $\sigma\circ\tau(\{l-k+1,\ldots,l\})$. Hence $\pi\circ\sigma\circ\tau \in S_{l,k}$ and $\pi\circ\sigma\circ\tau((\{1,\ldots,l-k\}) = \sigma\circ\tau((\{1,\ldots,l-k\})$. We additionally define the set

$$S^\tau_{l,k} := \{\pi_{\tau,\sigma}\circ\sigma\circ\tau : \sigma \in S_{l,k}\}$$

that is a subset of $S_{l,k}$ as mentioned above. It is essential for the proof that $S_{l,k} \subset S^\tau_{l,k}$ as well. Thus let $\tilde\sigma \in S^\tau_{l,k}$. Define $\sigma \in S_l$ as

$$i \mapsto \begin{cases} m_i\left(\tilde\sigma\circ\tau^{-1}(\{1,\ldots,l-k\})\right) & \text{if } 1\leq i\leq l-k\\ m_{i-l+k}\left(\tilde\sigma\circ\tau^{-1}(\{l-k+1,\ldots,l\})\right) & \text{otherwise,}\end{cases}$$

where m_i denotes the i-th lowest number of a set. By definition, σ obviously is an element of $S_{l,k}$. Because $\sigma(A) = \tilde\sigma\circ\tau^{-1}(A)$, we have $\sigma\circ\tau(A) = \tilde\sigma(A)$ for both $A = \{1,\ldots,l-k\}$ or $A = \{l-k+1,\ldots,l\}$. Therefore, $\pi\circ\sigma\circ\tau = \tilde\sigma$.

Hence, $S_{l,k} = S^{\tau}_{l,k}$ and therefore

$$\begin{aligned}
&\int\limits_{0\le t_{\tau(1)}\le\dots\le t_{\tau(l)}\le t} \sum_{\sigma\in S_{l,k}} f(t_{\sigma(1)})\dots f(t_{\sigma(l-k)})g(t_{\sigma(l-k+1)})\dots g(t_{\sigma(l)})\,d(t_1,\dots,t_l)\\
&= \int\limits_{0\le t_1\le\dots\le t_l\le t} \sum_{\sigma\in S_{l,k}} f(t_{\sigma\circ\tau^{-1}(1)})\dots f(t_{\sigma\circ\tau^{-1}(l-k)})g(t_{\sigma\circ\tau^{-1}(l-k+1)})\dots g(t_{\sigma\circ\tau^{-1}(l)})\,d(t_1,\dots,t_l)\\
&= \int\limits_{0\le t_1\le\dots\le t_l\le t} \sum_{\sigma\in S_{l,k}} f(t_{\pi\circ\sigma\circ\tau^{-1}(1)})\dots f(t_{\pi\circ\sigma\circ\tau^{-1}(l-k)})\\
&\qquad\qquad\qquad\cdot g(t_{\pi\circ\sigma\circ\tau^{-1}(l-k+1)})\dots g(t_{\pi\circ\sigma\circ\tau^{-1}(l)})\,d(t_1,\dots,t_l)\\
&= \int\limits_{0\le t_1\le\dots\le t_l\le t} \sum_{\sigma\in S^{\tau^{-1}}_{l,k}} f(t_{\sigma(1)})\dots f(t_{\sigma(l-k)})g(t_{\sigma(l-k+1)})\dots g(t_{\sigma(l)})\,d(t_1,\dots,t_l)\\
&= \int\limits_{0\le t_1\le\dots\le t_l\le t} \sum_{\sigma\in S_{l,k}} f(t_{\sigma(1)})\dots f(t_{\sigma(l-k)})g(t_{\sigma(l-k+1)})\dots g(t_{\sigma(l)})\,d(t_1,\dots,t_l) \qquad \square
\end{aligned}$$

We are now ready to prove Theorem 4.21.

Proof of Theorem 4.21. We can directly compute the right hand side of (4.22). In order to do this, note that the process $(K_t)_{t\ge0}$ can be realized as follows:

- Start with K_0 mutations, distributed according to $\underline{x}(0)$.
- New mutations arise at rate λ.
- Every mutation (present from the start or newly arisen) starts an exponential waiting time with parameter $\alpha+\gamma$. If this waiting time expires, then with probability $\frac{\alpha}{\alpha+\gamma}$ the process jumps to †, and with the complementary probability $\frac{\gamma}{\alpha+\gamma}$ the mutation disappears.

With $x_k(t)$ defined by $(**)$, we decompose the probability of the event $\{K_t = k\}$ with respect to the number of mutations present at time 0. If $K_0 = i$, a number $j \le i \wedge k$ of these initial mutations are not compensated by time t and the remaining $i-j$ are compensated. In addition, a number $l \ge k-j$ mutations arise at times $0 \le t_1 \le \dots \le t_l \le t$. From these, $l-k+j$ are compensated and the remaining $k-j$ are not compensated. These arguments lead to the following calculation, where we write $\sim$ for equality up to factors not depending on k. The first $\sim$ comes from the fact that the right hand side is the unconditional probability $\mathbf{P}[K_t = k]$,

$$
\begin{aligned}
x_k(t) \sim &\sum_{i=0}^{\infty} x_i(0) \sum_{j=0}^{i} \binom{i}{j} \Big(\frac{\gamma}{\alpha+\gamma} \big(1 - e^{-(\alpha+\gamma)t}\big)\Big)^{i-j} \cdot e^{-j(\alpha+\gamma)t} \\
&\quad \sum_{l=k-j}^{\infty} \int\limits_{0 \le t_1 \le \dots \le t_l \le t} \lambda^l e^{-\lambda t_1} e^{-\lambda(t_2 - t_1)} \cdots e^{-\lambda(t_l - t_{l-1})} e^{-\lambda(t - t_l)} \\
&\quad \sum_{\sigma \in S_{l,(k-j)}} \prod_{r=1}^{l-k+j} \frac{\gamma}{\alpha+\gamma} \big(1 - e^{-(\alpha+\gamma)(t - t_{\sigma(r)})}\big) \cdot \prod_{s=l-k+j+1}^{l} e^{-(\alpha+\gamma)(t - t_{\sigma(s)})} \, d(t_1, ..., t_l) \\
= &\sum_{i=0}^{\infty} x_i(0) \sum_{j=0}^{i} \binom{i}{j} \Big(\frac{\gamma}{\alpha+\gamma} \big(1 - e^{-(\alpha+\gamma)t}\big)\Big)^{i-j} \cdot e^{-j(\alpha+\gamma)t} \\
&\quad \sum_{l=k-j}^{\infty} \frac{\lambda^l}{l!} e^{-\lambda t} \binom{l}{l-k+j} \Big(\frac{\gamma}{\alpha+\gamma} \int_0^t 1 - e^{-(\alpha+\gamma)(t-r)} dr \Big)^{l-k+j} \Big(\int_0^t e^{-(\alpha+\gamma)(t-s)} ds \Big)^{k-j} \\
\sim &\sum_{i=0}^{\infty} x_i(0) \sum_{j=0}^{i \wedge k} \binom{i}{j} \Big(\frac{\gamma}{\alpha+\gamma} \big(1 - e^{-(\alpha+\gamma)t}\big)\Big)^{i-j} \cdot e^{-j(\alpha+\gamma)t} \frac{\lambda^{k-j}}{(k-j)!} \Big(\int_0^t e^{-(\alpha+\gamma)(t-s)} ds \Big)^{k-j} \\
&\quad \sum_{l=0}^{\infty} \frac{\lambda^l}{l!} \Big(\frac{\gamma}{\alpha+\gamma} \int_0^t 1 - e^{-(\alpha+\gamma)(t-r)} dr \Big)^{l} \\
\sim &\sum_{i=0}^{\infty} x_i(0) \sum_{j=0}^{i \wedge k} \binom{i}{j} \Big(\frac{\gamma}{\alpha+\gamma} \big(1 - e^{-(\alpha+\gamma)t}\big)\Big)^{i-j} \cdot e^{-j(\alpha+\gamma)t} \frac{\lambda^{k-j}}{(k-j)!} \Big(\frac{1}{\alpha+\gamma} (1 - e^{-(\alpha+\gamma)t}) \Big)^{k-j}
\end{aligned}
$$

where the equality is Lemma 4.25. As we have ignored factors not depending on k for simplicity, we must ensure that $\underline{x} \in \mathbb{S}$. Hence, we have to divide through the sum of the right-hand side, which is

$$
\begin{aligned}
&\sum_{i=0}^{\infty} x_i(0) \sum_{j=0}^{i} \binom{i}{j} \Big(\frac{\gamma}{\alpha+\gamma} \big(1 - e^{-(\alpha+\gamma)t}\big)\Big)^{i-j} \cdot e^{-j(\alpha+\gamma)t} \sum_{k=j}^{\infty} \frac{\lambda^{k-j}}{(k-j)!} \Big(\frac{1}{\alpha+\gamma} (1 - e^{-(\alpha+\gamma)t}) \Big)^{k-j} \\
&\qquad = \sum_{i=0}^{\infty} x_i(0) \Big(\frac{\gamma}{\alpha+\gamma} - \frac{\alpha}{\alpha+\gamma} e^{-(\alpha+\gamma)t} \Big)^i \cdot \exp\Big(\frac{\lambda}{\alpha+\gamma} (1 - e^{-(\alpha+\gamma)t}) \Big).
\end{aligned}
$$

Hence,

$$
x_k(t) = \frac{\sum_{i=0}^{\infty} x_i(0) \sum_{j=0}^{i \wedge k} \binom{i}{j} \Big(\frac{\gamma}{\alpha+\gamma} \big(1 - e^{-(\alpha+\gamma)t}\big)\Big)^{i-j} e^{-j(\alpha+\gamma)t} \frac{\lambda^{k-j}}{(k-j)!} \Big(\frac{1}{\alpha+\gamma} (1 - e^{-(\alpha+\gamma)t}) \Big)^{k-j}}{\sum_{i=0}^{\infty} x_i(0) \Big(\frac{\gamma}{\alpha+\gamma} - \frac{\alpha}{\alpha+\gamma} e^{-(\alpha+\gamma)t} \Big)^i \cdot \exp\Big(\frac{\lambda}{\alpha+\gamma} (1 - e^{-(\alpha+\gamma)t}) \Big)}
$$

□

Hence, we have solved the ODE approximation of Muller's ratchet with compensatory mutations. We can now easily derive that the ODE approximation also converges to a Poisson distribution.

Corollary 4.26. *If either $\gamma > 0$ or $x_0(0) > 0$, then*

$$x_k(t) \xrightarrow{t\to\infty} e^{-\theta}\frac{\theta^k}{k!} \quad \textit{for} \quad \theta = \frac{\lambda}{\alpha+\gamma}.$$

In particular, $\left(e^{-\theta}\frac{\theta^k}{k!}\right)_{k\in\mathbb{N}}$ is the only equilibrium point of $\underline{x}$ in this case.

Proof. The proof is analogous to Corollary 3.6 if $\gamma = 0$ and $x_0(0) > 0$. In the case $\gamma > 0$, notice that all terms in the numerator except that for $j = 0$ converge to 0 as $t \to \infty$. Hence, we have

$$\lim_{t\to\infty} x_k(t) = \frac{\sum_{i=0}^{\infty}\left(\frac{\gamma}{\alpha+\gamma}\right)^i \cdot \frac{\lambda^k}{k!} \cdot \frac{1}{(\alpha+\gamma)^k}}{\sum_{i=0}^{\infty}\left(\frac{\gamma}{\alpha+\gamma}\right)^i \cdot \exp\left(\frac{\lambda}{\alpha+\gamma}\right)} = e^{-\theta}\frac{\theta^k}{k!}. \qquad \square$$

Note that for $\gamma = 0$,

$$x_0(t) = \frac{\exp(-\frac{\lambda}{\alpha}(1-e^{-\alpha t}))}{\sum_{j=0}^{\infty} X_j(0)\, e^{-\alpha t j}} \sum_{j=0}^{\infty} \frac{\frac{\lambda}{\alpha}(1-e^{-\alpha t})^{k-j}}{(k-j)!} x_j(0)\, e^{-\alpha t j}$$

is equal to the large population limit of the discrete system (3.8) if we replace $e^{-\alpha t}$ with $(1-\alpha)^t$. There is a good chance that the exploitation of this observation could lead to a solution for a time-discrete deterministic system. However, such an examination would be beyond the scope of the current manuscript.

Furthermore, our proof of Theorem 4.21 relies on a positive selection coefficient. However, it can be shown by deriving (4.21) that the assertion is also true if $\alpha < 0$. In this situation, all mutations are beneficial and γ represents the probability that such a mutation is lost again. The effects of a negative α on the stochastic system also seems to be worth further investigation.

5. Simulations

To study the effect compensatory mutation have on Muller's ratchet, the discrete system described in Section 4.1 was implemented in Java[1] using the Colt libraries for High Performance Scientific Computing[2] developed by the European Organization for Nuclear Research (CERN). The software is named *mrj* (for Muller's ratchet in Java) and is freely available precompiled and in source code at the authors homepage[3]. The source code can also be found in Appendix A.

By default, the software starts with a population free of mutations. To simulate the reproduction according to Muller's ratchet with compensatory mutations, two alternative methods were implemented. As default, the function *reproduction_normal()* is used, which is a straight forward implementation of the discrete dynamics. For each of the N individuals, there is first sampled a parent according to the multinomial distribution with parameters according to the fitness and distribution of the types in the generation before. Then there is a binomial distributed number of mutations removed and a Poisson distributed number added again afterwards. Finally the type frequencies are calculated. The alternative reproduction function *reproduce_alternative()*, which can be activated via the command line option '-a', first calculates the expected type frequencies p_k as in (4.1) for a reasonable number of types k including mutations and compensatory mutations, and does a multinomial sampling of N individuals according to the expected frequencies afterwards.

However, as we could not measure a significant difference in run time of both methods (Figure 5.1), we stuck to the default for our simulations. Note that the run times kept roughly constant even for moderate large values of N.

N	normal	alternative
10^2	4:52	4:55
10^3	4:48	4:55
10^4	4:50	4:49
10^5	28:53	29:48

Figure 5.1.: Run times of a simulation of 10^5 generations using *reproduction_normal()* respectively *reproduce_alternative()* for populations of different size N with $\lambda = 0.1$, $\alpha = 0.03$ and $\gamma = 10^{-4}$.

[1] `http://www.java.com`
[2] `http://acs.lbl.gov/software/colt`
[3] `http://paulstaab.de/science`

The simulations itself were carried out for a wide range of parameters on a distributed grid of twelve desktop computers, better known as the students 'PC-Pool' of the Mathematical Institute in Freiburg. The resulting 1.8 GB of data were analysed using R (Team [2008]).

First, we examined whether compensatory mutations indeed stop the ratchet. As expected, that was sooner or later the case for all simulated combinations of parameters. A typical run consists of a starting phase, where deleterious mutations are accumulated quickly until a plateau is reached, where the ratchet attempts to stay but is disturbed by random fluctuation. This is illustrated in Figure 5.2, where the observed mean number of mutations $\kappa_1(t)$ is drawn for a single run as well as its average.

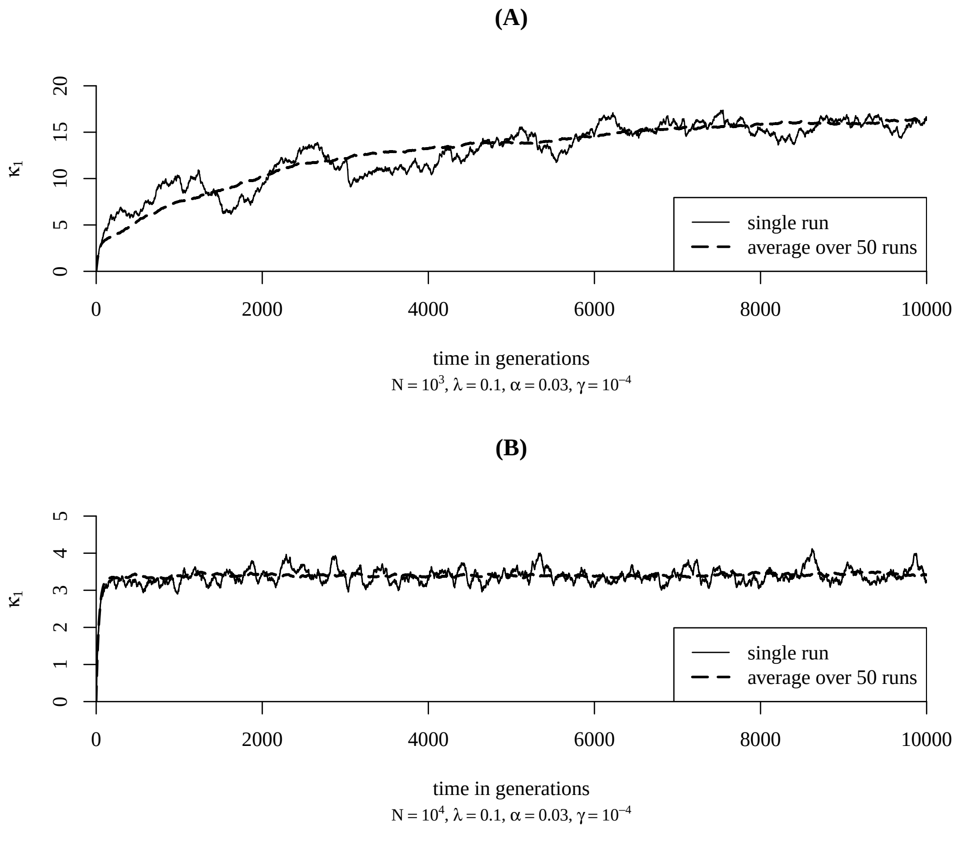

Figure 5.2.: The evolution of the average number of deleterious mutations κ_1 is plotted against time. In addition, the average over 50 different simulations is drawn.

The fluctuations are larger for parameters were the ratchet without compensatory mutation clicks frequently. That is in particular true for small populations as in Figure 5.2.A. Observe that the plateau lays higher in this Figure, whereas it almost exactly takes the expectation value $\theta = \lambda/(\alpha+\gamma)$ of the equilibrium Poisson distribution predicted for $N = \infty$ in Figure 5.2.B.

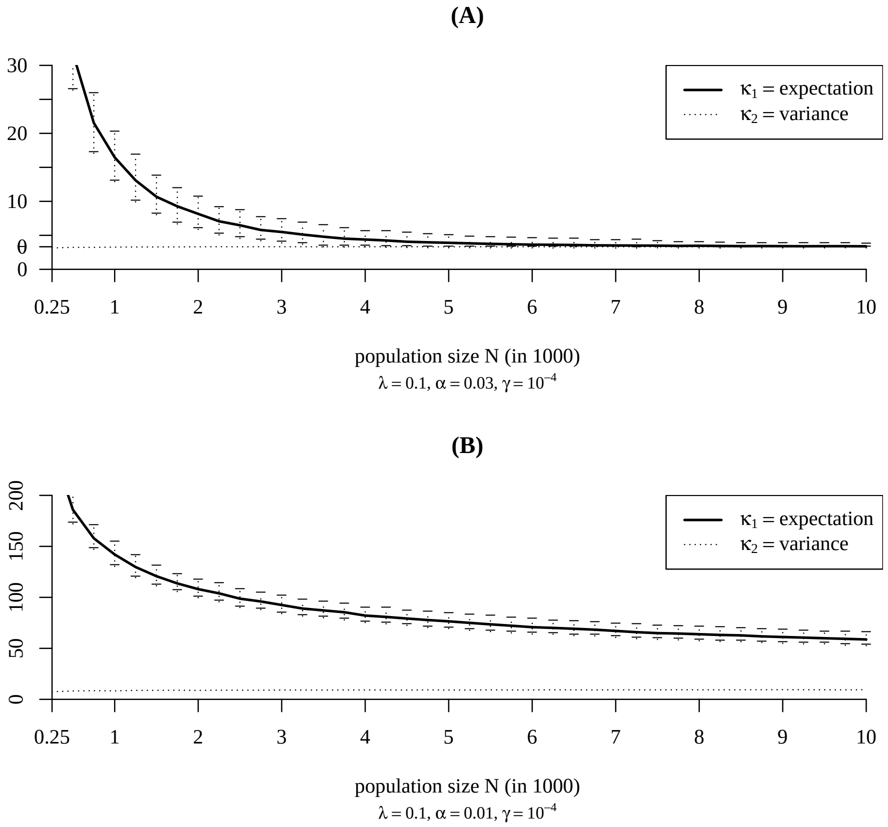

Figure 5.3.: The empirical distribution of κ_1 and κ_2 are evaluated between time $5 \cdot 10^2 N$ and $10^3 N$. The plot for κ_1 includes the resulting 10% and 90% quantiles.

Figure 5.3 demonstrates that the magnitude of the “finite population effect” for the ratchet with compensatory mutations is closely related to the click rate of the classical ratchet. Here, the average value of κ_1 *after the ratchet has reached the described plateau* is draw as function of the population size N. In Figure 5.3.A, the observed value of κ_1 quickly approaches the prediction θ. Hence, we can assume that the population is well approximated by the large population limit even for moderate N. Note that the parameters $\lambda = 0.1$, $\alpha = 0.03$ and $N = 10^4$ would lead to approximately 0.34 clicks per N time units without compensatory mutations. To quantify the magnitude of the fluctuations around the expected value, the 10%- and 90%-quantiles of the empirical distribution of κ_1 are included as well.

In Figure 5.3.B, κ_1 is much greater than the predicted value, even for large populations. As $\alpha = 0.01$ is much smaller here, we would expect approximately 152 clicks per N time units for $\gamma = 0$ (and again for $N = 10^4$). Hence, we can conclude that our approximations hold if the population size N is sufficiently large and λ and α would not lead to frequent clicks of the normal ratchet.

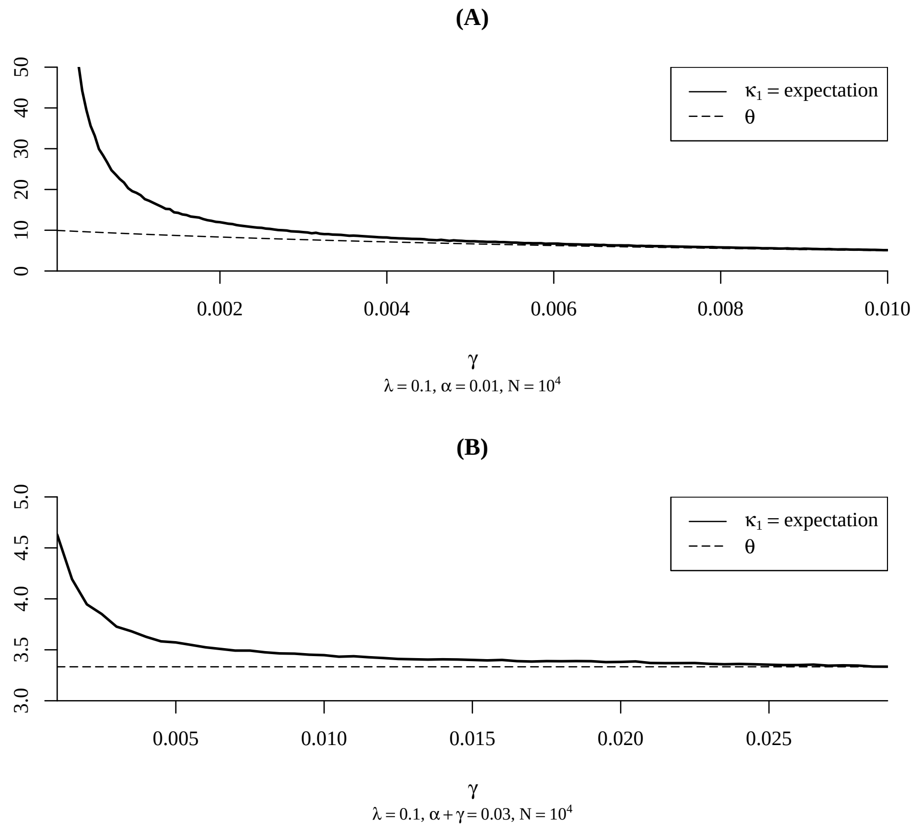

Figure 5.4.: Again average values of κ_1 are drawn, here for different values of γ. In Figure (A) N, α and λ are fixed, whereas α is variable in Figure (B), so that $\alpha + \gamma$ is constant.

Remarkably, the empirical variance κ_2 is always close to the variance of the predicted $\mathrm{Poi}(\theta)$-distribution for all N in both pictures of Figure 5.3. That suggests that the ratchet with compensatory mutations is close to a shifted Poisson distribution most of the time.

As illustrated in Figure 5.4.A, only unrealistic high values of γ assure that frequent clicks become compensated. Note that the parameters λ and α are the same as used in Figure 5.3.B here. Even though the quality of the approximation certainly depends on γ as well, the click rate of the classical ratchet seems to be the more important parameter.

Finally we checked, if the symmetry of α and γ in θ is also reflected for finite population size. However, Figure 5.4.B suggests that the purifying effect of compensatory mutations is more important than the one of selection. This seems reasonable as every mutation can be lost by a compensatory mutation, while only those ones that an individual has additional to the least number of mutations in the population can be lost by selection.

Bibliography

M. Antezana and R. R. Hudson. Era reversibile! point mutations, the ratchet, and the initial success of eukaryotic sex: a simulation study. *Evolutionary Theory*, 11: 209–235, 1997.

Reinhard Burger. Moments, cumulants, and polygenic dynamics. *Journal of Mathematical Biology*, 30(2):199–213, November 1991. ISSN 0303-6812. doi: 10.1007/BF00160336.

Lin Chao. Fitness of RNA virus decreased by muller's ratchet. *Nature*, 348(6300): 454–455, November 1990. doi: 10.1038/348454a0.

Brian Charlesworth and Deborah Charlesworth. Rapid fixation of deleterious alleles can be caused by muller's ratchet. *Genetics Research*, 70(01):63–73, 1997.

Donald Dawson. Measure-valued markov processes. In Paul-Louis Hennequin, editor, *Ecole d'Eté de Probabilités de Saint-Flour XXI - 1991*, volume 1541, pages 1–260. Springer Berlin Heidelberg, 1993. ISBN 978-3-540-56622-9.

A. Depperschmidt, P. Pfaffelhuber, M. Tillich, and S. A. Rensing. Mathematical modelling of plant evolution: Rna editing counteracts muller's ratchet. *Manuscript*, 2011.

A. Etheridge, P. Pfaffelhuber, and A. Wakolbinger. How often does the ratchet click? facts, heuristics, asymptotics. In *Trends in Stochastic Analysis*, volume 353 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, New York, part IV 13 edition, September 2008.

S. N. Ethier and Thomas G. Kurtz. Fleming–Viot processes in population genetics. *SIAM Journal on Control and Optimization*, 31(2):345, 1993. ISSN 03630129. doi: 10.1137/0331019.

Stewart N. Ethier and Thomas G. Kurtz. *Markov processes*. Wiley-Interscience, Hoboken, NJ, 2005. ISBN 978-0-471-76986-6.

Stewart N. Ethier and Tokuzo Shiga. A Fleming-Viot process with unbounded selection. *Journal of Mathematics of Kyoto University*, 40(2):337–361, 2000. ISSN 2154-3321.

Joseph Felsenstein. The evolutionary advantage of recombination. *Genetics*, 78(2): 737–756, 1974.

R. A. Fisher. *The Genetical Theory Of Natural Selection.* At The Clarendon Press, 1930.

W. Gabriel, M. Lynch, and R. Burger. Muller's ratchet and mutational meltdowns. *Evolution*, 47(6):1744–1757, 1993. ISSN 0014-3820. doi: 10.2307/2410218.

Damian D. G. Gessler. The constraints of finite size in asexual populations and the rate of the ratchet. *Genetics Research*, 66(03):241–253, 1995. doi: 10.1017/S0016672300034686.

Isabel Gordo and Brian Charlesworth. The degeneration of asexual haploid populations and the speed of muller's ratchet. *Genetics*, 154(3):1379 –1387, March 2000.

John Haigh. The accumulation of deleterious genes in a population–Muller's ratchet. *Theoretical Population Biology*, 14(2):251–267, October 1978. ISSN 0040-5809. doi: 10.1016/0040-5809(78)90027-8.

Paul G. Higgs and Glenn Woodcock. The accumulation of mutations in asexual populations and the structure of genealogical trees in the presence of selection. *Journal of Mathematical Biology*, 33(7), August 1995. ISSN 0303-6812. doi: 10.1007/BF00184644.

Olav Kallenberg. *Foundations of modern probability.* Springer, New York, 1997. ISBN 9780387227047.

Achim Klenke. *Wahrscheinlichkeitstheorie.* Springer-Verlag GmbH, 1 edition, 2006. ISBN 3540255451.

Walter Ledermann and Emlyn Lloyd. *Probability.* John Wiley & Sons, January 1981. ISBN 0471278211.

L. Loewe. Quantifying the genomic decay paradox due to muller's ratchet in human mitochondrial DNA. *Genetical research*, 87(2):133–160, 2006.

Michael Lynch and Wilfried Gabriel. Mutation load and the survival of small populations. *Evolution*, 44(7):1725–1737, November 1990. ISSN 0014-3820. doi: 10.2307/2409502.

Leonardo P Maia, Daniela F Botelho, and José F Fontanari. Analytical solution of the evolution dynamics on a multiplicative-fitness landscape. *Journal of Mathematical Biology*, 47(5):453–456, November 2003. ISSN 0303-6812. doi: 10.1007/s00285-003-0208-8. PMID: 14605858.

Uwe Maier, Andrew Bozarth, Helena Funk, Stefan Zauner, Stefan Rensing, Christian Schmitz-Linneweber, Thomas Borner, and Michael Tillich. Complex chloroplast RNA metabolism: just debugging the genetic programme? *BMC Biology*, 6(1):36, 2008. ISSN 1741-7007. doi: 10.1186/1741-7007-6-36.

Sophie Maisnier-Patin and Dan I. Andersson. Adaptation to the deleterious effects of antimicrobial drug resistance mutations by compensatory evolution. *Research in Microbiology*, 155(5):360–369, June 2004. ISSN 0923-2508. doi: 16/j.resmic.2004.01.019.

H. J. Muller. Some genetic aspects of sex. *The American Naturalist*, 66(703):118, January 1932. doi: 10.1086/280418.

H.J. Muller. The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 1(1):2–9, May 1964. ISSN 0027-5107. doi: 10.1016/0027-5107(64)90047-8.

Pekka Pamilo, Masatoshi Nei, and Wen-Hsiung Li. Accumulation of mutations in sexual and asexual populations. *Genetics Research*, 49(02):135–146, 1987. doi: 10.1017/S0016672300026938.

Peter Pfaffelhuber, Paul Reinhard Staab, and Anton Wakolbinger. Muller's ratchet with compensatory mutations. *Manuscript*, 2011.

Art Poon and Lin Chao. The rate of compensatory mutation in the DNA bacteriophage φX174. *Genetics*, 170(3):989 –999, July 2005. doi: 10.1534/genetics.104.039438.

Philip E. Protter. *Stochastic Integration and Differential Equations.* Springer, 2nd edition, May 2005. ISBN 3540003134.

Igor M. Rouzine, John Wakeley, and John M. Coffin. The solitary wave of asexual evolution. *Proceedings of the National Academy of Sciences*, 100(2):587 –592, January 2003. doi: 10.1073/pnas.242719299.

J. M Smith. The evolution of sex. *infection*, 54(1), 1978.

Wolfgang Stephan, Lin Chao, and Joanne Guna Smale. The advance of muller's ratchet in a haploid asexual population: Approximate solutions based on diffusion theory. *Genetics Research*, 61(03):225–231, 1993. doi: 10.1017/S0016672300031384.

R. Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing Vienna Austria*, 3(10), 2008.

Erklärung

Hiermit versichere ich, dass ich diese Arbeit selbständig verfasst und nur die angegebenen Quellen und Hilfsmittel benutzt habe.

Datum: Unterschrift:

Danksagung

An dieser Stelle möchte ich allen danken, die mich beim Erstellen dieser Arbeit unterstützt haben. Ihr wart mir eine große Hilfe.

Ein besonderer Dank gebührt Peter Pfaffelhuber. Ohne seine außergewöhnlich engagierte und intensive Betreuung, seinen unerschöpflichen Vorrat an Ideen und seine umfangreichen Vorarbeiten zur Ratsche wäre diese Arbeit zweifellos nicht entstanden.

Weiter danke ich dem Mathematischen Institut hier in Freiburg dafür, dass mir die für die Simulationen nötige Rechenleistung zur Verfügung gestellt wurde, und der Bundesrepublik Deutschland für die finanzielle Unterstützung während meines Studiums.

A. Source Code Of The Simulation Software

Main.java

```
package de.paulstaab.mrj;

public class Main {

 public static void help(){
  System.out.println("MRJ - Muller's ratchet in Java");
  System.out.println("Simulates a population of asexual " +
             "individuals evolving according to Muller's " +
             "ratchet with compensatory mutations.");
  System.out.println("");
  System.out.println("Usage: " +
    "mrj [OPTION] N=1000 t=1000 m=0.4 s=0.1 b=0.0001");
  System.out.println("");
  System.out.println("where");
  System.out.println("N\t is population size");
  System.out.println("t\t is the number of generations to simulate. " +
    "As default t is interpreted as t*N generations (see -e)");
  System.out.println("m\t is the mutation parameter");
  System.out.println("s\t is the selection parameter");
  System.out.println("b\t is the compensatory mutation parameter");
  System.out.println("");
  System.out.println("OPTIONs");
  System.out.println("-a  | --alternative \t " +
    "Uses alternative reproduction function. Much slower.");
  System.out.println("-d  | --dont-rescale \t " +
    "Don't rescale time and parameters by 1/N");
  System.out.println("-h  | --help \t\t " +
    "displays this message.");
  System.out.println("-u  | --human \t\t " +
    "generates an output that is better readable by humans ");
  System.exit(0);
 }


 public static void main(String[] args) {
  try{
   //Get parameters
   int t = 1000;
```

```
int N = 1000;
double m = 0.4;
double s = 0.1;
double b = 0.0001;
int initialMutations = 0;

boolean header = false;
boolean human = false;
boolean alternative = true;
boolean every = false;

for (String arg : args){
 if (arg.startsWith("t")) t=Integer.valueOf(arg.substring(2));
 else if (arg.startsWith("N")) N=Integer.valueOf(arg.substring(2));
 else if (arg.startsWith("m")) m=Double.valueOf(arg.substring(2));
 else if (arg.startsWith("s")) s=Double.valueOf(arg.substring(2));
 else if (arg.startsWith("b")) b=Double.valueOf(arg.substring(2));
 else if (arg.startsWith("i"))
  initialMutations=Integer.valueOf(arg.substring(2));
 else if ( arg.contentEquals(("--header")) ) header = true;
 else if ( arg.contentEquals(("--human")) ) human = true;
 else if ( arg.contentEquals(("--alternative")) ) alternative = true;
 else if ( arg.contentEquals(("--dont-rescale")) ) every = true;
 else if ( arg.contentEquals(("--help")) ) help();
 else if (arg.startsWith("-")) {
  if ( arg.contains("a") ) alternative = true;
  if ( arg.contains("d") ) every = true;
  if ( arg.contains("h") ) help();
  if ( arg.contains("u") ) human = true;
  }
 else {
  throw new IllegalArgumentException("Unknown argument given.");
 }
}

//Initialisation
Population pop = new Population(N, m, s, b, initialMutations,
     alternative, human, every);

if (header) System.out.println( pop.printHeading() );
else{
 System.out.println(pop.toString());
 if (!every) t=t*N;
 int outputEvery = N;
 if (every) outputEvery = 1;
```

```
    //Simulation
    for (int i=1; i<=t; i++){
     pop.reproduce();
     if (i % outputEvery == 0)
      System.out.println( pop.toString() );
    }
   }
  }
  catch (Exception e){
   e.printStackTrace();
  }
 }
}
```

Population.java

```
package de.paulstaab.mrj;

import java.text.NumberFormat;
import java.util.ArrayList;
import java.util.Collections;
import java.util.HashMap;
import java.util.Locale;
import java.util.Map;
import java.util.Map.Entry;

import cern.jet.math.Arithmetic;
import cern.jet.random.Binomial;
import cern.jet.random.Poisson;
import cern.jet.random.engine.MersenneTwister;

public class Population {

 //-----------------------------------------------
 //Variables
 //-----------------------------------------------

 //Used for rounded outputs
 NumberFormat nf = NumberFormat.getInstance(Locale.US);

 //The population
 private HashMap<Integer,Integer> typeList = new HashMap<Integer,Integer>();

 //Model parameters
 private int size = 0;
 private int time = 0;
```

```
private double mutation;
private double backmutation;
private double selection;

//Options
private boolean alternative = false;
private boolean human = false;
private boolean every = false;

//Used for caching values
private double tempFirstMoment;
private double tempMeanFitness;
private int tempFirstMomentCalcTime = -1;
private int tempMeanFitnessCalcTime = -1;

//Statistics
int mutationsOfFittestClass = 0;
int maxMutations = 1;

//Random number generators
MersenneTwister engine =
 new MersenneTwister(new java.util.Date());
Poisson mutationGenerator;



//-----------------------------------------------
//Constructor
//-----------------------------------------------
public Population( int size, double mutation, double selection,
     double backmutation,  int initialMutations,
     boolean alternative, boolean human, boolean every ){

 this.size = size;
 this.typeList.put(initialMutations, size);

 if ( this.checkMutation(mutation) ) {
  this.mutation = mutation;
  this.mutationGenerator = new Poisson(this.mutation, engine);
 }
 else throw new
  IllegalArgumentException("Mutation parameter must be greater than 0");

 if ( this.checkBackMutation(backmutation) )
  this.backmutation = backmutation;
 else throw new
```

```
  IllegalArgumentException("BackMutation parameter must be in [0,1]");

 if ( this.checkSelection(selection) )
  this.selection = selection;
 else throw new
  IllegalArgumentException("Selection parameter must be between 0 and 1");

 if (alternative) this.alternative = true;
 if (human) this.human = true;
 if (every) this.every = true;
}



//-----------------------------------------------
//Parameter checks
//-----------------------------------------------
private boolean checkMutation(double mutation){
 if ( mutation >= 0 ) return true;
 else return false;
}

private boolean checkBackMutation(double backMutation){
 if ( backMutation >= 0 && backMutation <= 1 ) return true;
 else return false;
}

private boolean checkSelection(double selection){
 if ( selection >= 0 && selection <= 1 ) return true;
 else return false;
}



//-----------------------------------------------
//Calcuations
//-----------------------------------------------
private double x(int k){
 double xk = 0;
 if (typeList.containsKey(k)) xk = (double)typeList.get(k) / size;
 return xk;
}

private int multinominialSample(Map<Integer, Double> probWeights)
 throws Exception {
 double random = this.engine.nextDouble();
```

```
 double quantil = 0;
 int last = 0;
 for (Entry<Integer, Double> entry : probWeights.entrySet() ) {
  quantil += entry.getValue();
     if (random < quantil) {
      return( entry.getKey() );
     }
     last = entry.getKey();
 }
 if (quantil < 0.9999 || Double.isNaN(quantil))
  throw new IllegalArgumentException("Error doing multinomial sampling");
 return ( last );  //Error taken in account for speedoptimisation.
      //Should only happen in every 10000 case...
}

private double calcSelection(int i){
 if (x(i) == 0) return 0.0;
 else return x(i) * Math.pow(1-selection, i-mutationsOfFittestClass);
}

private double meanFitness() throws Exception{
 double meanFitness = 0;
 if (this.tempMeanFitnessCalcTime == this.time)
  meanFitness = this.tempMeanFitness;
 else{
  for (int k : this.typeList.keySet()){
   meanFitness += calcSelection(k) ;
  }
  this.tempMeanFitnessCalcTime = this.time;
  this.tempMeanFitness = meanFitness;
 }
 if (meanFitness == 0) throw new Exception("meanFitness = 0!");
 return meanFitness;
}

private double firstMoment()
{
 double firstMoment = 0;
 if (this.tempFirstMomentCalcTime == this.time)
  firstMoment = this.tempFirstMoment;
 else {
  for (int k : this.typeList.keySet()){
   firstMoment += k * x(k);
  }
  this.tempFirstMomentCalcTime = time;
  this.tempFirstMoment = firstMoment;
```

```
 }
 return firstMoment;
}



//----------------------------------------------
//Reproduction functions
//----------------------------------------------
public void reproduce() throws Exception{
 if (alternative) this.reproduce_alternative();
 else this.reproduce_normal();
}

private void reproduce_normal() throws Exception{
 //Compute reproducing probabilites according to selection
 HashMap<Integer,Double> probWeights = new HashMap<Integer,Double>();
 for (int i : this.typeList.keySet()){
  probWeights.put( i , calcSelection(i)/meanFitness() );
 }

 //Draw N siblings
 HashMap<Integer, Integer> typeList= new HashMap<Integer, Integer>();
 int minMutations = Integer.MAX_VALUE;
 for (int i=1; i<=this.size; i++){
  //Draw a sibling,
  int offspring = this.multinominialSample(probWeights);
  //remove mutations
  if (this.backmutation > 0 && offspring > 0)
   offspring -=
    new Binomial(offspring, this.backmutation, this.engine).nextInt();
  //add mutations
  offspring += mutationGenerator.nextInt();

  if (offspring < minMutations) minMutations = offspring;
  //and create a new typeList
  if ( typeList.containsKey(offspring) ){
   int tmp = typeList.get(offspring);
   typeList.remove(offspring);
   typeList.put(offspring, tmp+1);
  }
  else typeList.put(offspring, 1);
 }
 //Save the new population
 this.mutationsOfFittestClass = minMutations;
 this.typeList = typeList;
```

```
 time++;
}

private void reproduce_alternative() throws Exception{
 HashMap<Integer,Double> probWeights = new HashMap<Integer,Double>();
 double sum = 0;

 //calculate expected frequencys for next generation
 int k=-1; while(sum < 0.9999) {
  k++;
  double xk = 0;
  for (int i=Math.max(0,this.mutationsOfFittestClass-15) ; i<=k; i++){
   double xki = 0;
   for (int j=i; j<=this.maxMutations; j++) {
    if (j==0 || (backmutation == 0 && i == j )) xki += calcSelection(j);
    //Can't be calculated by Binomial.class
    else if (backmutation != 0) {
      xki += calcSelection(j)
             * new Binomial(j,backmutation,engine).pdf(j-i);
    }
   }
   //xki = xki * new Poisson(mutation,engine).pdf(k-i);
   xki = xki * Math.pow(mutation, k-i) / Arithmetic.factorial(k-i);
   xk += xki;
   //System.out.println("--- k:"+k+" i:"+i+" xki:"+xki+" xk:"+xk);
  }
  xk = xk * Math.pow(Math.E, -mutation) / meanFitness();
  probWeights.put(k, xk);
  sum += xk;
  //System.out.println("k:" + k + " Xk:" + xk + " sum:" + sum);
 }

 //sample new individuals
 HashMap<Integer, Integer> typeList = new HashMap<Integer, Integer>();
 int minMutations = Integer.MAX_VALUE;
 int maxMutations = 0;
 for (int i=1; i<=this.size; i++){
   //Draw a sibling,
   int offspring = this.multinominialSample(probWeights);

   if (offspring < minMutations) minMutations = offspring;
   if (offspring > maxMutations) maxMutations = offspring;

   //and create a new typeList
   if ( typeList.containsKey(offspring) ){
    int tmp = typeList.get(offspring);
```

```
    typeList.remove(offspring);
    typeList.put(offspring, tmp+1);
   }
   else typeList.put(offspring, 1);
 }
 //Save the new population
 time++;
 this.mutationsOfFittestClass = minMutations;
 this.maxMutations = maxMutations;
 this.typeList = typeList;
}




//-------------------------------------------------
//Output
//-------------------------------------------------
public String toString() {
 if (human) {
  String value = "";

  if (!every) value += "t:" + this.time / this.size + "N ";
  else value += "t:" + this.time + " ";

  value += "N:" + this.size + " ";
  value += "s:" + this.selection + " ";
  value += "m:" + this.mutation + " ";
  value += "b:" + this.backmutation + " ";

  value += "K*:" + Integer.toString(this.mutationsOfFittestClass) + " ";
  value += "M1:" + nf.format(firstMoment() - mutationsOfFittestClass) + " ";

  ArrayList<Integer> sortedKeys =
   new ArrayList<Integer>(this.typeList.keySet());
  Collections.sort(sortedKeys);
  int lastnumber = sortedKeys.get(sortedKeys.size()-1);

  for (int key=0; key <= lastnumber; key++){
   if (sortedKeys.contains(key))
    value +=
     Integer.toString(key)+":" +
     Integer.toString(this.typeList.get(key)) + " ";
   else
    value +=
     Integer.toString(key)+":0 ";
```

```
   }
   return value;
  }

  else {
   String value = "";

   if (!every) value += this.time / this.size + " ";
   else value += this.time + " ";

   value += this.size + " ";
   value += this.selection + " ";
   value += this.mutation + " ";
   value += this.backmutation + " ";
   value += Integer.toString(this.mutationsOfFittestClass) + " ";
   value += nf.format(firstMoment() - mutationsOfFittestClass) + " ";


   ArrayList<Integer> sortedKeys = new ArrayList<Integer>(typeList.keySet());
   Collections.sort(sortedKeys);
   int lastnumber = sortedKeys.get(sortedKeys.size()-1);

   for (int key=this.mutationsOfFittestClass; key <= lastnumber; key++){
    if (sortedKeys.contains(key))
     value += Integer.toString(this.typeList.get(key)) + ";";
    else
     value += "0;";
   }
   return value;
  }
 }

 public String printHeading() {
  return "t N s lambda mu k m1 distribution";
 }
}
```

B. Pfaffelhuber et al. [2011]

Submitted to the Annals of Applied Probability

MULLER'S RATCHET WITH COMPENSATORY MUTATIONS

By P. Pfaffelhuber[°,*], P. R. Staab[°] A. Wakolbinger[‡]

[°]*Albert-Ludwigs-Universität, Freiburg and* [‡]*Goethe-Universität, Frankfurt*

We consider an infinite dimensional system of stochastic differential equations which describes the evolution of type frequencies in a large population. Random reproduction is modeled by a Wright-Fisher noise whose inverse diffusion coefficient N corresponds to the total population size. The type of an individual is the number k of deleterious mutations it carries. We assume that fitness of individuals carrying k mutations is decreased by αk for some $\alpha > 0$. Along the individual lines of descent, (new) mutations accumulate at rate λ per generation, and each of these mutations has a small probability γ per generation to disappear. While the case $\gamma = 0$ is known as (the Fleming-Viot version of) *Muller's ratchet*, the case $\gamma > 0$ is referred to as that of *compensatory mutations* in the biological literature. In the former case ($\gamma = 0$), an ever increasing number of mutations is accumulated over time, while in the latter ($\gamma > 0$) this is prevented by the compensatory mutations which in this sense *halt the ratchet*. We show that the system under consideration ($\gamma \geq 0$) has a unique weak solution. For $N = \infty$, we obtain the solution in a closed form by analyzing a probabilistic particle system that represents this solution. In particular, we show that the unique equilibrium state in the case $\gamma > 0$ and $N = \infty$ is the Poisson distribution with parameter $\lambda/(\gamma + \alpha)$.

1. Introduction and outline. Our objective is the study of a system of a multitype Wright-Fisher SDE (or *Fleming-Viot process)* of the form

$$(*)\qquad \begin{aligned} dX_k = \Big(\alpha\Big(\sum_{\ell=0}^{\infty}(\ell-k)X_\ell\Big)X_k + \lambda(X_{k-1}-X_k) \\ + \gamma((k+1)X_{k+1}-kX_k)\Big)dt + \sum_{\ell\neq k}\sqrt{\tfrac{1}{N}X_kX_\ell}\,dW_{k\ell}, \end{aligned}$$

for $k = 0, 1, \ldots$ with $X_{-1} := 0$ and $\sum_{k=0}^{\infty} X_k = 1$. Here α, λ and γ are (small) non-negative constants, N is a (large) number (or equals infinity in which

---

*Corresponding author; email: p.p@stochastik.uni-freiburg.de

AMS 2000 subject classifications: Primary 92D15, 92D15, 60J70; secondary 60K35, 60H35, 60H10

Keywords and phrases: Muller's ratchet, selection, back mutation, Fleming-Viot process, Girsanov transform

case the last term on the r.h.s of $(*)$ vanishes), and $(W_{k\ell})_{k>\ell}$ is a family of independent Brownian motions with $W_{k\ell} = -W_{\ell k}$.

The interest in this system comes from population genetics (see Section 2 for some background). The equations $(*)$ provide a diffusion approximation of the evolution of the *type frequencies* X_k, $k \in \mathbb{N}_0$ in a population consisting of a large number N of individuals evolving in discrete time. The type k of an individual is given by the number of deleterious mutations it carries. Briefly, the *fitness* (which is proportional to the average number of offspring) of a type-k individual is proportional to $(1-\alpha)^k$, which is approximately $1 - \alpha k$ for small $\alpha > 0$, termed the *selection coefficient.* The parameter λ is the expected number of additional mutations accumulating per individual and generation, and for each of the mutations present, γ is the probability that this mutation disappears in one generation.

In this work we will not be concerned with proving the convergence of the discrete-generation dynamics to the diffusion approximation. Still, in Section 3.3 we will use the discrete generation scheme as just described in order to present a few simulation results which illustrate how certain functionals of the solution of $(*)$ (in particular the mean and the variance of the probability vector $(X_k)_{k=0,1,2,\ldots}$) depend on the model parameters.

Theorem 1 in Section 3.1 states that $(*)$ has a unique weak solution. Note that $(*)$ is an infinite dimensional SDE with an *unbounded non-linear drift coefficient.* Related existence and uniqueness results were obtained by Ethier and Shiga (2000). However, these authors only cover the case of parent-independent mutation and not the situation of $(*)$.

Theorem 2 in Section 3.2 gives the explicit solution of $(*)$ in the (deterministic) case $N = \infty$. This extends results from Haigh (1978) and Etheridge et al. (2008) for the case $\gamma = 0$. In particular, we show that the Poisson weights with parameter $\lambda/(\gamma + \alpha)$ constitute the only equilibrium state of $(*)$ for $N = \infty$. The proofs of Theorems 1 and 2 are given in Sections 4 and 5, respectively. An essential step in the proof of Theorem 2 is Proposition 5.2 which in the case $N = \infty$ provides the solution of $(*)$ in terms of a probabilistic particle system.

2. History and background of the model. For $\gamma = 0$, the system $(*)$ is known as (the Fleming-Viot version of) Muller's ratchet, a population genetic model introduced by Hermann Muller (1964): A clonal population of fixed size reproduces randomly. Each individual carries a number of mutations, all of which are assumed to be deleterious. Fitness decreases linearly with the number of mutations. The offspring of an individual has a small chance to gain a new deleterious mutation. In particular, offspring of an individual has at least as many mutations as the parent and mutation is an irreversible process. Hence, eventually the ratchet will *click* in the sense that the so far fittest type will irreversibly disappear from the population. In this way, the mutation process drives the population to a larger number of deleterious mutations while selection acts in the opposite direction, leading to a form of mutation-selection quasi-balance. Gabriel et al. (1993) consider a related model of a clonally reproducing population in which the evolution of the population size is coupled with the mean fitness of the population, eventually leading to extinction of the population. The prediction of this *mutational meltdown* requires information on the rate at which deleterious mutations accumulate in the population (Loewe, 2006), i.e. on the *rate of Muller's ratchet*.

Various quantitative treatments of Muller's ratchet have been given (Haigh, 1978; Stephan et al., 1993; Gessler, 1995; Higgs and Woodcock, 1995; Gordo and Charlesworth, 2000; Maia et al., 2003; Rouzine et al., 2003; Etheridge et al., 2008; Jain, 2008; Waxman and Loewe, 2010; Audiffren and Pardoux, 2011). The most interesting question concerns the rate of Muller's ratchet. This has so far only been studied by simulations, or approximations which seem ad hoc.

We study an extension of Muller's ratchet where deleterious mutations are allowed to be compensated by (back-)mutations. It is important to note that such *compensatory mutations* are different from *beneficial mutations*, although both increase the fitness of an individual. In our model, compensatory mutations can only remove the effects of previously gained deleterious mutations. In contrast to this, beneficial mutations are usually assumed to have an effect that does not depend on the genetic background.

The possibility of compensatory mutations was discussed already by Haigh (1978) (see also Maynard Smith, 1978). He argued that they rarely occur in realistic parameter ranges, because the deleterious mutation rate is proportional to the full length of the genome of a clonally reproducing individual, while the compensatory mutation rate scales with the length of a single base within the full genome. Therefore, he concluded that compensatory mutations are too rare to halt the accumulation of deleterious mutations in realistic parameter ranges. However, when several deleterious mutations are gained, the total rate of accumulation of deleterious mutations increases and may therefore halt the ratchet. An equilibrium is approached where a certain number of deleterious mutations is fixed. If this number is large enough, these may lead to extinction of the population. While Antezana and Hudson (1997) argue that the effects of compensatory mutations can be an important factor for small viruses, Loewe (2006) concludes that compensatory mutations are still too rare to halt the mutational meltdown of human mitochondria.

Clearly, the relevance of compensatory mutations is greatest for species with a short genome and a high mutation rate. One of the most extreme groups in these respects are RNA viruses (for which the genome length is of the order of 10^3 to 10^4 bases and the per base mutation rate is around 10^{-5} to 10^{-4}). As discussed in Chao (1990), back mutations can hardly stop Muller's ratchet even in this case. We will come back to this numerical example in Section 3.3 below.

The relevance of Muller's ratchet with compensatory mutations is supported by the fact that deleterious mutations might be compensated not only by back mutations at the same genomic position. As discussed by Wagner and Gabriel (1990), restoring the function of a gene which was subject to a mutation is as well possible by mutating a second site within the gene or at another gene. Maisnier-Patin and Andersson (2004) give the following generalizations of (single-base) compensatory mutations: (i) point mutations which restore the RNA secondary structure of a gene or the protein structure, (ii) an up-regulation of gene expression of the mutated gene, (iii) a mutation in another gene restoring the structure of a multi-unit protein complex and (iv) a bypass mechanism where the function of the mutated is taken over by another gene.

Various examples give clear evidence for the existence of compensatory mutations. It has been shown by Poon and Chao (2005) that a deleterious mutation in the DNA bacteriophage phiX174 can be compensated by about nine different intragenic compensatory mutations. This implies that the rate of compensatory mutations can be high enough to halt accumulation of deleterious mutations under realistic scenarios. In fact, compensatory mutations have been observed in various species. Howe and Denver (2008) showed that deletions in protein-coding regions of the mitochondrial genome in *Caenorhabditis briggsae* lead to heteroplasmy, a severe factor in mitochondrial diseases. They also found compensatory mutations leading to a decrease in heteroplasmy. Mutations for antibiotic resistance of bacteria are known to be deleterious in a wild-type population. Fitness can be increased by a compensatory mutation (see e.g. Handel et al., 2006). Plastid genomes of mosses are studied in Depperschmidt et al. (2011). Here, deleterious mutations are compensated by RNA editing, a mechanism by which the base C in DNA is transcribed to U on the level of RNA for specific bases in the genome.

All these examples indicate that the role of compensatory mutations should be taken into account. A relevant question to be addressed in future research is which parameter constellations (of the selection coefficient, the mutation rate, the compensatory mutation rate and the population size) can halt the ratchet before the mutational meltdown leads to extinction of the population.

3. Results. We show that for finite N the system $(*)$ has a unique weak solution (Theorem 1). For the system $(*)$ without noise (i.e. the case $N = \infty$) we provide in Theorem 2 the explicit form of the solution as well as the equilibrium state. For this we use a stochastic particle model (including accumulation and loss of mutations, as well as a state-dependent death rate of the particles) and show in Proposition 5.2 that a solution of $(*)$ with $N = \infty$ is given by the distribution of the particle model conditioned on non-extinction. After stating the theorems, we compare in Section 3.3 the cases of large N with the theoretical considerations for $N = \infty$ using simulations.

3.1. *Existence and uniqueness.* The system $(*)$ of Muller's ratchet with compensatory mutations takes values in the space of probability vectors indexed by $\mathbb{N}_0$, i.e. sequences whose entries are probability weights on $\mathbb{N}_0$. We restrict the state space to the subset of probability vectors with finite exponential moment of a certain order, and show uniqueness in this space. Throughout, we abbreviate $\underline{x} := (x_0, x_1, ...) \in \mathbb{R}_+^{\mathbb{N}_0}$.

Definition 3.1 (Simplex). *The* infinite-dimensional simplex *is given by*

$$\mathbb{S} := \Big\{ \underline{x} \in \mathbb{R}_+^{\mathbb{N}_0} : \sum_{k=0}^{\infty} x_k = 1 \Big\}. \tag{3.1}$$

Moreover, for $\xi > 0$, set

$$h_\xi(\underline{x}) := \sum_{k=0}^{\infty} x_k e^{\xi k} \tag{3.2}$$

and consider elements of $\mathbb{S}$ with ξth exponential moment, forming the space

$$\mathbb{S}_\xi := \big\{ \underline{x} \in \mathbb{S} : h_\xi(\underline{x}) < \infty \big\}. \tag{3.3}$$

Remark 3.2 (Topology on $\mathbb{S}_\xi$). We note that

$$r(\underline{x}, \underline{y}) := \sum_{k=0}^{\infty} e^{\xi k} |x_k - y_k|, \quad \underline{x}, \underline{y} \in \mathbb{S}_\xi, \tag{3.4}$$

defines a complete and separable metric on $\mathbb{S}_\xi$.

THEOREM 1 (Well-posedness of Fleming-Viot system). *Let $\underline{x} \in \mathbb{S}_\xi$ for some $\xi > 0$. Then, for $N \in (0, \infty)$, $\alpha, \lambda, \gamma \in [0, \infty)$, the system $(*)$ starting in $\underline{X}(0) = \underline{x}$ has a unique $\mathbb{S}$-valued weak solution $\mathcal{X} = (\underline{X}(t))_{t \geq 0}$, taking values in the space $\mathcal{C}_{\mathbb{S}_\xi}([0, \infty))$ of continuous functions on $\mathbb{S}_\xi$. The process $\mathcal{X}$ is referred to as* Muller's ratchet with compensatory mutations *with selection coefficient α, mutation rate λ, compensatory mutation rate γ and population size N.*

REMARK 3.3 (Population size N). Resampling models are usually studied either for a finite population of constant size N (e.g. using a Wright-Fisher model), or in the large population limit with a suitable rescaling of time, leading to Fleming-Viot processes. For a bounded fitness function and a compact type space, it is well known that a sequence of (discrete time) Wright-Fisher processes, indexed by N, converges weakly to a Fleming-Viot process (or Wright-Fisher diffusion) if the selection and mutation coefficients are scaled down by N and one unit of time is taken as N generations; see e.g. Ethier and Kurtz (1993). In our situation it may thus be expected (though we do not prove this claim here) that for large N and for αN, λN and γN of order one, the Wright-Fisher process described in Section 3.3, run with a time unit of N generations, is close to the solution of $(*)$, with 1 instead of $\sqrt{1/N}$ as the coefficient of the noise, and αN, λN and γN in place of α, λ and γ. (Note that we do not prove this claim here.) However, this system is $(*)$ with time speeded up by a factor N. In other words, for large N, and αN, λN and γN of order one, the solution of $(*)$ should be close to the corresponding Wright-Fisher model as introduced in Section 3.3, with time unit being one generation. This is the reason why we refer to the model parameter N in $(*)$ as the population size. We use this terminology in interpreting the simulation results for the Wright-Fisher model in Section 3.3.

REMARK 3.4 (Connection to previous work for $\gamma = 0$). For the case $\mu = 0$, variants of Theorem 1 appear in Cuthbertson (2007) and in Audiffren and Pardoux (2011). The latter makes (in the terminology of our Theorem 1) the special choice $\xi = \alpha N$ and refers to Audiffren (2011) for the proof. Cuthbertson (2007) treats also the case of $\alpha < 0$, assuming the existence of all exponential moments of the initial state.

REMARK 3.5 (Strategy of the proof of Theorem 1). For $\alpha = 0$, it follows from classical theory (Dawson, 1993, Theorem 5.4.1) that $(*)$ has a unique weak solution. The same is true if the selection term $\alpha(\sum_{\ell=0}^{\infty}(\ell - k)X_\ell)X_k$ is replaced by a bounded function of $\underline{X}$. This can be shown by a Girsanov change of measure from the case $\alpha = 0$, an idea going back to Dawson (1978) in the measure-valued context. So, the main difficulty in the proof is to deal with the unbounded selection term. This is overcome by showing that the change of measure still works when using $\mathbb{S}_\xi$ as the state space for $\mathcal{X}$.

3.2. *The case $N = \infty$.* This case (which is not included in Theorem 1) leads to a deterministic dynamics. For $\gamma = 0$, Haigh (1978) was the first to obtain results on the deterministic evolution of $\mathcal{X}$ in a discrete time setting. These results were later refined by Maia et al. (2003). Here, we work with continuous time, and our next theorem generalizes Proposition 4.1 in Etheridge et al. (2008) to the case $\gamma > 0$. We are dealing with the system

$$\dot{x}_k = \alpha\Big(\sum_{\ell=0}^{\infty}(\ell - k)x_\ell\Big)x_k + \lambda(x_{k-1} - x_k) + \gamma((k+1)x_{k+1} - kx_k) \tag{3.5}$$

for $k = 0, 1, 2, ...$ with $x_{-1} := 0$ and $\sum_{k=0}^{\infty} x_k = 1$.

THEOREM 2. *Let $\alpha, \lambda, \gamma \in [0, \infty)$ and $\underline{x}(0) \in \mathbb{S}_\xi$ for some $\xi > 0$. Then, the system (3.5) has a unique $\mathbb{S}$-valued solution $(\underline{x}(t))_{t\geq 0}$ which takes values in $\mathbb{S}_\xi$. It is given by*

(3.6)
$$x_k(t) = \frac{\sum\limits_{i=0}^{\infty} x_i(0) \sum\limits_{j=0}^{i\wedge k} \binom{i}{j}\big(\frac{\gamma(1-e^{-(\alpha+\gamma)t})}{\alpha+\gamma}\big)^{i-j} e^{-j(\alpha+\gamma)t} \frac{1}{(k-j)!}\big(\frac{\lambda(1-e^{-(\alpha+\gamma)t})}{\alpha+\gamma}\big)^{k-j}}{\sum\limits_{i=0}^{\infty} x_i(0)\big(\frac{\gamma}{\alpha+\gamma} - \frac{\alpha}{\alpha+\gamma}e^{-(\alpha+\gamma)t}\big)^i \cdot \exp\big(\frac{\lambda}{\alpha+\gamma}(1 - e^{-(\alpha+\gamma)t})\big)}.$$

In particular, if either $\gamma > 0$ or $x_0(0) > 0$, then

$$x_k(t) \xrightarrow{t\to\infty} \frac{e^{-\lambda/(\alpha+\gamma)}}{k!} \cdot \Big(\frac{\lambda}{\alpha + \gamma}\Big)^k, \quad k = 0, 1, 2, \ldots \tag{3.7}$$

i.e. the limiting state is the vector of Poisson weights with parameter $\lambda/(\alpha+\gamma)$.

REMARK 3.6 (Equilibria). In the case $\gamma = 0$ it is known already from the work of Haigh (1978) that the vector of Poisson weights with parameter λ/α is an equilibrium state. Moreover, Poisson states which are shifted by $k = 1, 2, ...$ are equilibria as well. This is in contrast to $\gamma > 0$ where only a single equilibrium state exists. Moreover, this equilibrium state depends on the model parameters only through $\lambda/(\alpha + \gamma)$. This is surprising since the dynamics of the process is much different for different combinations of λ, α and γ with the same value of $\lambda/(\alpha + \gamma)$. See Figure 3 for a simulation study of this feature for $N < \infty$.

Remark 3.7 (Connection to the rate of adaptation). Although the strategy of our proof requires that $\alpha \geq 0$ (i.e. the mutations are deleterious), it can be shown by taking the time-derivative of the right hand side of (3.6) that this equation is a solution for $\alpha < 0$ as well. This model is frequently termed *rate of adaptation* and has gained some interest in the case $\gamma = 0$ and $N < \infty$ (Gerrish and Lenski, 1998; Desai and Fisher, 2007; Park and Krug, 2007; Yu et al., 2010).

Taking $\alpha < 0$ in our model, all mutations are beneficial and γ is the rate by which any beneficial mutation is compensated. Interestingly, only in the case $|\alpha| < \gamma$ (i.e. selection is weaker than the compensatory mutation rate) an equilibrium state exists, and is still Poisson with parameter $\lambda/(\gamma - |\alpha|)$. In the case $|\alpha| \geq \gamma$, no equilibrium exists because new beneficial mutations spread through the population quicker than compensatory mutations can halt this process. It will be interesting to investigate the switch between these two scenarios in the case of finite N.

3.3. *Simulations.* We use simulations based on a discrete Wright-Fisher model to study the evolution of the mean fitness, and to investigate the dependence of the mean and the variance of the type frequency distribution on the model parameters. Fixing a population size N, this model is a discrete time Markov chain $(\underline{Y}(t))_{t=0,1,2,\dots}$ taking values in $\{\underline{y} \in \mathbb{S} : N\underline{y} \in \mathbb{N}_0^{\mathbb{N}_0}\}$ and such that

$$\mathbf{P}(\underline{Y}(t+1) = \underline{y}|\underline{Y}(t)) = \binom{N}{Ny_0\ Ny_1\ Ny_2 \cdots} \prod_{j=0}^{\infty} p_j^{Ny_j}$$

where

$$\begin{aligned}
&\text{(i)} \qquad \widetilde{p}_j = \frac{(1-\alpha)^j Y_j(t)}{\sum_{k=0}^{\infty}(1-\alpha)^k Y_k(t)},\\
&\text{(ii)} \qquad \widehat{p}_j = \sum_{m=j}^{\infty} \widetilde{p}_m \binom{m}{j} \gamma^{m-j}(1-\gamma)^j,\\
&\text{(iii)} \qquad p_j = \sum_{l=0}^{j} \widehat{p}_l e^{-\lambda} \frac{\lambda^{j-l}}{(j-l)!}
\end{aligned}$$

for *small* parameters α, λ and γ. The sampling weights $(p_j)_{j=0,1,\dots}$ describe selection, mutation and compensatory mutation. The idea in this scheme (which is standard in population genetics) is that (i) any individual produces a large number of gametes, but an individual with k deleterious mutations only contributes a number proportional to $(1-\alpha)^k$ to the gamete pool, (ii) every deleterious mutation has a small, independent chance γ to be removed while the gamete is built, and (iii) the number of new deleterious mutations is Poisson distributed with parameter λ. After building these gametes, N individuals are randomly chosen from the gamete pool to form the next generation. Since α, γ and λ are assumed to be small, the order in which the three mechanisms (i), (ii), (iii) come into play is negligible. (E.g., if we would assume – in contrast to our simulation scheme above – that compensatory mutations arise before gametes are built proportional to the relative fitness of individuals. Then an individual with a high number of deleterious mutations would produce slightly more gametes than in our simulation scheme.) For our simulations, the working hypothesis is that $(\underline{Y}(Nt))_{t\geq 0}$ behaves similar to $\mathcal{X} = (\underline{X}(Nt))_{t\geq 0}$ where $\mathcal{X}$ is the solution of $(*)$ with parameter N; see Remark 3.3.

We simulated $(\underline{Y}(Nt))_{t\geq 0}$ for various combinations of N, α, λ and γ, starting with $\underline{Y}(0) = \delta_0$, i.e. no deleterious mutations are present at start. Since in reality compensatory mutations are less probable than mutations, we mostly simulate scenarios with $\gamma \ll \lambda$. (For the biological background of this assumption, see Section 2.) Hence, our simulations can be considered as a small perturbation of the case $\gamma = 0$, the case of Muller's ratchet (without compensatory mutations). We compare scenarios where Muller's ratchet clicks rarely with others in which it clicks more frequently. For example, in Figure 1(A) we use $N = 10^3, \lambda = 0.1, \alpha = 0.03$ where the ratchet has about 5.7 clicks in N generations. In Figure 1(B) we use $N = 10^4$ where the ratchet has only about 0.34 clicks in N generations. Both figures show the initial phase of the simulation for a small compensatory mutation rate of $\gamma = 10^{-4}$. Recall that Theorem 2 predicts an equilibrium number of $\lambda/(\alpha + \gamma) \approx 3.3$ deleterious mutations in the case $N = \infty$. This value is reflected in our simulations only in Figure (B) where Muller's ratchet clicks rarely. In Figure (A), not only is the average number of deleterious mutations much larger than the prediction from Theorem 2, but also the fluctuations are much larger than in Figure (B). However, in both parameter constellations we see that the accumulation of deleterious mutations by Muller's ratchet is slowed down (and sooner or later halted) due to the compensatory mutations.

Figure 2 illustrates for a finite N, how far the mean and variance of the number of deleterious mutations deviate from those Poisson distribution, which appears in Theorem 2 for the case $N = \infty$. Again, we see that for fixed α, λ and small compensatory mutation rate γ, the equilibrium for κ_1 is close to $\lambda/(\alpha + \gamma)$ only if the ratchet without compensatory mutations ($\gamma = 0$) does not click too often. If $N = 10^4$ in Figure 2(A), there are approximately 152 clicks in N generations in the absence of compensatory

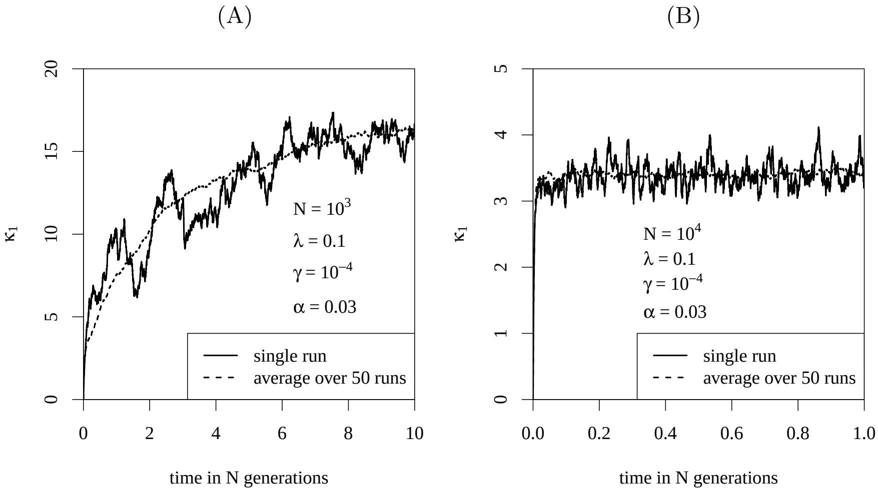

FIG 1. *The evolution of the average number of deleterious mutations κ_1 is plotted. In addition, the single path is compared to the average over 50 different simulations. (A) A parameter combination where Muller's ratchet without compensatory mutations (i.e. $\gamma = 0$) clicks frequently, while it clicks much less frequent in (B).*

mutations while in Figure 2(B) this rate is much lower, approximately 0.34 clicks per N generations (we use the same parameter values α, λ and γ as in Figure 1(B)). These examples show that compensatory mutations halt the ratchet quite efficiently. Note that the parameter values $\lambda = 0.1$ and $\gamma = 10^{-4}$ fit to the evolution of RNA viruses, e.g. for a genome of length 10^3 bases, if the per base mutation rate is 10^{-4} and a population of size 10^4. As our simulations show, the ratchet is halted provided the selection coefficient is large enough. This is in some contrast to Chao (1990) who argues that compensatory mutations are too rare in RNA viruses to halt the ratchet.

Another surprising fact in Figure 2(A) is that the empirical variance of the number of deleterious mutations in the population is always close to the prediction of $\lambda/(\alpha + \gamma)$ from the Poisson state appearing in Theorem 2. This would be compatible with the hypothesis that the type frequencies for the Wright-Fisher model are (in equilibrium) close to a shifted Poisson distribution. The detailed study of the amount of this shift, in particular for a parameter constellation for which $\gamma = 0$ leads to frequent clicks, is a delicate issue. Its answer certainly depends on the rate of clicking of Muller's ratchet without compensatory mutations, a problem which remains unsolved until now.

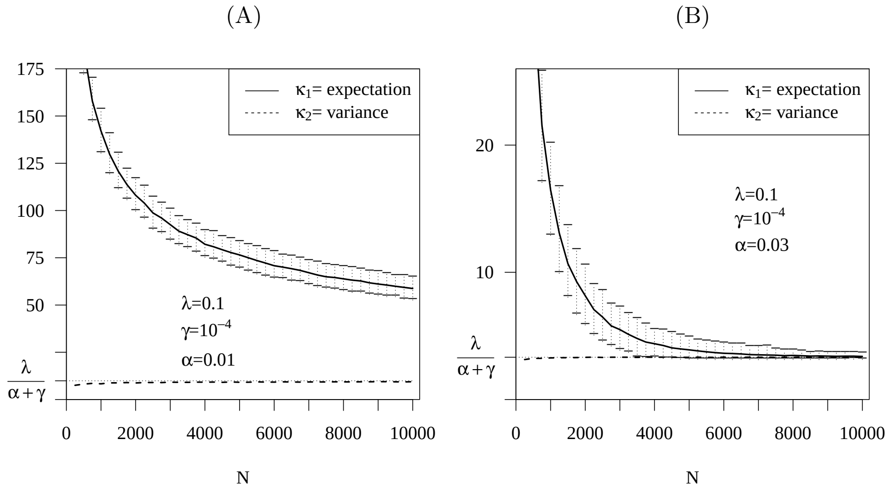

FIG 2. *The empirical distribution of κ_1 and κ_2 are evaluated between generations $5 \cdot 10^2 N$ and $10^3 N$. The plot for κ_1 includes the resulting 10% and 90% quantiles. In absence of compensatory mutations and with $N = 10^4$, the same parameters lead to approximately 152 clicks per N time units for Figure (A), while 0.34 clicks per N time units are obtained for Figure (B).*

Yet another interesting feature seen in Theorem 2 is the symmetric dependence on α and γ of the equilibrium state. We checked in which direction this symmetry is violated for finite N. As seen from Figure 3, compensatory mutations can halt the ratchet more efficiently than selection. The reason is that compensatory mutations reduce the number of mutations no matter how many mutations are fixed in the population, whereas the number of fixed mutations cannot decrease due to selection.

4. Proof of Theorem 1. Our approach is inspired by Ethier and Shiga (2000), who deal with the case of unbounded selection if mutation is parent-independent. In order to prove that $(*)$ has a unique weak solution, we use the equivalent formulation by a martingale problem and show its well-posedness (see Proposition 4.4). We provide bounds on exponential moments for any solution of the martingale problem associated with the generator of $(*)$ in Lemma 4.5. The central step is Proposition 4.8 which provides a Girsanov change of measure by which solutions of the martingale problem for $\alpha = 0$ are transformed to solutions for any $\alpha > 0$. Proposition 4.4 and Theorem 1 then follow because the martingale problem for $\alpha = 0$ is well-posed, and can be transformed to a solution for the martingale problem for

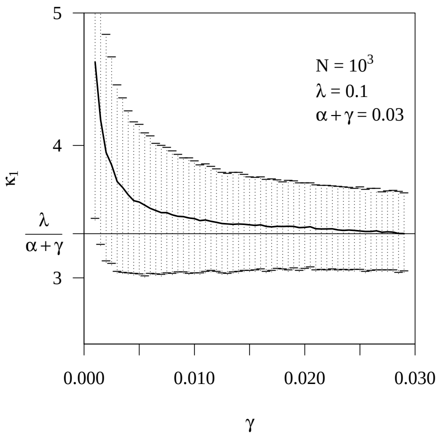

FIG 3. *The parameter $\lambda/(\alpha+\gamma)$ of the Poisson equilibrium state in the case $N=\infty$ is symmetric in α and γ (see Theorem 2). We fix $\alpha+\gamma$ and see that the average number of deleterious mutations is higher for low values of γ. Again, the 10% and 90% quantiles are given.*

$\alpha > 0$ which also solves $(*)$. This shows existence. Uniqueness again follows by using a Girsanov transform.

4.1. *Martingale problem.* We start by defining the generator for the Fleming-Viot system of Muller's ratchet with compensatory mutations. The unboundedness in the selection term of this generator requires particular care in the analysis of the corresponding martingale problem. First we give some notation.

REMARK 4.1 (Notation). For a complete and separable metric space $(\mathbb{E}, r)$, we denote by $\mathcal{P}(\mathbb{E})$ the space of probability measures on (the Borel sets of) $\mathbb{E}$, and by $\mathcal{M}(\mathbb{E})$ (resp. $\mathcal{B}(\mathbb{E})$) the space of real-valued, measurable (and bounded) functions. If $\mathbb{E} \subseteq \mathbb{R}^{\mathbb{N}_0}$, we let $\mathcal{C}^k(\mathbb{E})$ $(\mathcal{C}_b^k(\mathbb{E}))$ be the (bounded), k times partially continuously differentiable functions (with bounded derivatives). Partial derivatives of $f \in \mathcal{C}^2(\mathbb{E})$, $\mathbb{E} \subseteq \mathbb{R}^{\mathbb{N}_0}$ will be denoted by

$$f_k := \frac{\partial f}{\partial x_k}, \qquad f_{k\ell} := \frac{\partial f^2}{\partial x_k \partial x_\ell}, \qquad k, \ell = 0, 1, 2, ... \tag{4.1}$$

DEFINITION 4.2 (Martingale problem). *Let $(\mathbb{E}, r)$ be a complete and separable metric space, $x \in \mathbb{E}$, $\mathcal{F} \subseteq \mathcal{M}(\mathbb{E})$ and G a linear operator on $\mathcal{M}(\mathbb{E})$ with domain $\mathcal{F}$. A (distribution $\mathbf{P}$ of an) $\mathbb{E}$-valued stochastic process*

$\mathcal{X} = (X_t)_{t\geq 0}$ is called a solution of the $(\mathbb{E}, x, G, \mathcal{F})$-martingale problem if $X_0 = x$ and $\mathcal{X}$ has paths in the space $\mathcal{D}_{\mathbb{E}}([0,\infty))$, almost surely, and for all $f \in \mathcal{F}$,

$$\Big(f(X_t) - f(X_0) - \int_0^t Gf(X_s)ds\Big)_{t\geq 0} \tag{4.2}$$

is a **P***-martingale with respect to the canonical filtration. Moreover, the $(\mathbb{E}, x, G, \mathcal{F})$-martingale problem is said to be well-posed if there is a unique solution* **P***.*

For a fixed $\xi > 0$, our state space will be $(\mathbb{S}_\xi, r)$, cf. Definition 3.1 and Remark 3.2. We now specify the generator and its domain.

Definition 4.3 (Generator for Fleming-Viot system).

1. *On $\mathbb{S}$, consider functions of the form*

$$\begin{aligned} f(\underline{x}) &:= f_{\varphi_1,\dots,\varphi_n}(\underline{x}) := \langle \underline{x}, \varphi_1\rangle \cdots \langle \underline{x}, \varphi_n\rangle, \\ \langle \underline{x}, \varphi\rangle &:= \sum_{k=0}^\infty x_k \varphi(k) \end{aligned} \tag{4.3}$$

for $n = 1, 2, \dots$ and $\varphi, \varphi_1, \dots, \varphi_n \in \mathcal{M}(\mathbb{N}_0)$. Let

$$\begin{aligned} \mathcal{F} := &\text{ the algebra generated by} \\ &\{f_{\varphi_1,\dots,\varphi_n} : \varphi_i \in \mathcal{M}(\mathbb{N}_0) \text{ with bounded support}, \\ &\qquad\qquad i = 1, \dots, n,\ n \in \mathbb{N}\}. \end{aligned} \tag{4.4}$$

2. *We define the operator $G^\alpha_{\mathcal{X}}$ as the linear extension of*

$$\begin{aligned} G^\alpha_{\mathcal{X}} f(\underline{x}) &= G^\alpha_{sel} f(\underline{x}) + G_{mut} f(\underline{x}) + G_{cmut} f(\underline{x}) + G^N_{res} f(\underline{x}), \\ G^\alpha_{sel} f(\underline{x}) &= \alpha \sum_{k=0}^\infty \sum_{\ell=0}^\infty (\ell - k) x_\ell x_k f_k(\underline{x}), \\ G_{mut} f(\underline{x}) &= \lambda \sum_{k=0}^\infty (x_{k-1} - x_k) f_k(\underline{x}), \\ G_{cmut} f(\underline{x}) &= \gamma \sum_{k=0}^\infty ((k+1)x_{k+1} - k x_k) f_k(\underline{x}), \\ G^N_{res} f(\underline{x}) &= \tfrac{1}{2N} \sum_{k,\ell=0}^\infty x_k(\delta_{k\ell} - x_\ell) f_{k\ell}(\underline{x}) \end{aligned} \tag{4.5}$$

with $\alpha, \lambda, \gamma \in [0, \infty)$, $N \in (0, \infty)$, for f of the form (4.3) whenever the right hand sides of (4.5) exist (which is certainly the case if $\underline{x}$ has a first moment and the $\varphi_1, ..., \varphi_n$ have bounded support). In particular, for all $f \in \mathcal{F}$ and $\xi > 0$, the function $G^\alpha_{\mathcal{X}} f$ is defined on $\mathbb{S}_\xi$.

3. *For $f = f_{\varphi_1,...,\varphi_n}$, we define $\mathcal{N}_f = (N_f(t))_{t \geq 0}$ by*

$$N_f(t) := f(\underline{X}(t)) - \int_0^t G^\alpha_{\mathcal{X}} f(\underline{X}(s)) ds \tag{4.6}$$

whenever $G^\alpha_{\mathcal{X}} f(\underline{X}(t))$ exists for all $t \geq 0$.

Proposition 4.4 (Martingale problem is well-posed in $\mathbb{S}_\xi$). *Let $\underline{x} \in \mathbb{S}_\xi$ for some $\xi > 0$, $G^\alpha_{\mathcal{X}}$ as in (4.5) and $\alpha, \lambda, \gamma \in [0, \infty), N \in (0, \infty)$ and $\mathcal{F}$ be as in (4.4). Then, the $(\mathbb{S}, \underline{x}, G^\alpha_{\mathcal{X}}, \mathcal{F})$-martingale problem is well-posed and is a process with paths in $\mathcal{C}_{\mathbb{S}_\xi}([0, \infty))$.*

Proposition 4.4 is a crucial step in the proof of Theorem 1. Both proofs are carried out in Section 4.3. Now we start with bounds on exponential moments, which will be crucial in further proofs.

Lemma 4.5 (Bounds on exponential moments). *Let $\underline{x} \in \mathbb{S}_\xi$ for some $\xi > 0$ and $\mathcal{X} = (X(t))_{t \geq 0}$ be a solution of the $(\mathbb{S}, \underline{x}, G^\alpha_{\mathcal{X}}, \mathcal{F})$-martingale problem. Then,*

$$\mathbf{E}[h_\xi(\underline{X}(t))] \leq h_\xi(\underline{x}) \cdot \exp\big(\lambda t (e^\xi - 1)\big) \tag{4.7}$$

and for all $T > 0$ and $\varepsilon > 0$, there is $C > 0$, depending on T, ε, ξ and λ (but not on α, γ, N) with

$$\mathbf{P}\big[\sup_{0 \leq t \leq T} h_\xi(\underline{X}(t)) > C\big] \leq \varepsilon \cdot h_\xi(\underline{x}). \tag{4.8}$$

Proof. Define for $m = 0, 1, 2, ...$ the function $h_{\xi,m} \in \mathcal{F}$ by

$$h_{\xi,m}(\underline{x}) := \sum_{k=0}^m x_k e^{\xi k} + e^{\xi m}\Big(1 - \sum_{k=0}^m x_k\Big) = e^{\xi m} + \sum_{k=0}^m x_k (e^{\xi k} - e^{\xi m})$$

and note that

$$h_{\xi,m}(\underline{x}) = \sum_{k=0}^\infty x_k e^{\xi (k \wedge m)} \qquad \text{for } \underline{x} \in \mathbb{S}.$$

First, we compute
(4.9)

$$\begin{aligned}
G_{\text{mut}} h_{\xi,m}(\underline{x}) &= \lambda \sum_{k=0}^{m} (x_{k-1} - x_k)(e^{\xi k} - e^{\xi m}) = \lambda \sum_{k=0}^{m-1} x_k \Big(e^{\xi(k+1)} - e^{\xi k} \Big) \\
&= \lambda (e^{\xi} - 1) \sum_{k=0}^{m-1} x_k e^{\xi k} \geq 0, \\
G_{\text{cmut}} h_{\xi,m}(\underline{x}) &= \gamma \sum_{k=0}^{m} ((k+1) x_{k+1} - k x_k)(e^{\xi k} - e^{\xi m}) \\
&= \gamma \sum_{k=1}^{m} k x_k \Big(e^{\xi(k-1)} - e^{\xi k} \Big) \leq 0. \\
G^{\alpha}_{\text{sel}} h_{\xi,m}(\underline{x}) &= \alpha \sum_{k=0}^{m} \sum_{\ell=0}^{\infty} (\ell - k) x_\ell x_k (e^{\xi k} - e^{\xi m}) \\
&= \alpha \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} (\ell - k) x_\ell x_k (e^{\xi (k \wedge m)} - e^{\xi m}) \\
&= \alpha \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} (\ell - k) x_\ell x_k e^{\xi (k \wedge m)} \leq 0,
\end{aligned}$$

where the calculation for the term G^{α}_{sel} holds for $\underline{x} \in \mathbb{S}$. (For the last inequality, assume that Z is an $\mathbb{N}_0$-valued random variable with distribution $\underline{x}$. Then, $G_{\text{sel}} h_{\xi,m}(\underline{x}) = -\alpha \, \mathbf{Cov}[Z, e^{\xi(Z \wedge m)}] \leq 0$, since two increasing transformations of a random variable Z have a nonnegative correlation, or in other words, the singleton family $\{Z\}$ is associated.)

In the next step we prove (4.7). We write

$$\begin{aligned}
\frac{d}{dt} \mathbf{E}[h_{\xi,m}(\underline{X}(t))] &= \mathbf{E}\big[G^{\alpha}_{\mathcal{X}} h_{\xi,m}(\underline{X}(t)) \big] \leq \mathbf{E}\big[G_{\text{mut}} h_{\xi,m}(\underline{X}(t)) \big] \\
&= \lambda (e^{\xi} - 1) \cdot \mathbf{E}\Big[\sum_{k=0}^{m-1} X_k(t) e^{\xi k} \Big] \\
&\leq \lambda (e^{\xi} - 1) \mathbf{E}[h_{\xi,m}(\underline{X}(t))].
\end{aligned} \tag{4.10}$$

So, by Gronwall's inequality,

$$\mathbf{E}[h_{\xi,m}(\underline{X}(t))] \leq h_{\xi,m}(\underline{x}) \cdot \exp\big(\lambda t (e^{\xi} - 1) \big)$$

which gives (4.7) by monotone convergence.

Finally, by Doob's submartingale inequality and monotone convergence, using (4.10),

$$\begin{aligned}
\mathbf{P}\Big[\sup_{0\le t\le T} h_\xi(\underline{X}(t)) > C\Big] &= \lim_{m\to\infty}\mathbf{P}\Big[\sup_{0\le t\le T} h_{\xi,m}(\underline{X}(t)) > C\Big] \\
&\le \lim_{m\to\infty}\mathbf{P}\Big[\sup_{0\le t\le T}\Big(h_{\xi,m}(\underline{X}(t)) - \int_0^t G_{\mathrm{cmut}}h_{\xi,m}(\underline{X}(s)) \\
&\qquad\qquad + G_{\mathrm{sel}}h_{\xi,m}(\underline{X}(s))ds\Big) > C\Big] \\
&\le \frac{1}{C}\lim_{m\to\infty}\mathbf{E}\Big[h_{\xi,m}(\underline{X}(T)) - \int_0^T G_{\mathrm{cmut}}h_{\xi,m}(\underline{X}(s)) + G_{\mathrm{sel}}h_{\xi,m}(\underline{X}(s))ds\Big] \\
&\le \frac{1}{C}\lim_{m\to\infty}\Big(h_{\xi,m}(\underline{x}) + \int_0^T \mathbf{E}\big[G_{\mathrm{mut}}h_{\xi,m}(\underline{X}(s))\big]ds\Big) \\
&\le \frac{1}{C}\Big(h_\xi(\underline{x}) + \lambda(e^\xi-1)h_\xi(\underline{x})\int_0^T \exp\big(\lambda s(e^\xi-1)\big)ds\Big)
\end{aligned}$$

and the result follows. □

For the change of measure applied in the next subsection, we will need that the martingale property of N_f extends from $\mathcal{F}$ to a wider class of functions.

Lemma 4.6. *Let $\underline{x}\in\mathbb{S}_\xi$ for some $\xi>0$ and $\mathcal{X}=(\underline{X}_t)_{t\ge0}$ be a solution of the $(\mathbb{S},\underline{x},G^\alpha_{\mathcal{X}},\mathcal{F})$-martingale problem and*

$$f = f_{\varphi_1,\dots,\varphi_n} \text{ be of the form (4.3) with } |\varphi_i(.)|\le Ce^{\zeta\cdot} \text{ for some } C>0 \text{ and } \zeta<\xi, i=1,\dots,n. \tag{4.11}$$

Then, $(N_f(t))_{t\ge0}$ given by (4.6) is a martingale.

Proof. We first observe that $G^\alpha_{\mathcal{X}}f(\underline{X}(t))$ exists for all $t\ge0$, hence N_f is well-defined. For $\varphi\in\mathcal{M}(\mathbb{N}_+)$, let $\varphi^m(k):=\varphi(k\wedge m)$. We note that $\sum_{k=0}^\infty x_k\varphi^m(k) = \varphi(m)+\sum_{k=0}^m x_k(\varphi(k)-\varphi(m))$ for $\underline{x}\in\mathbb{S}$. Hence, for $f_{\varphi_1,\dots,\varphi_n}$ as given in the lemma, the function $f_{\varphi_1^m,\dots,\varphi_n^m}$ coincides on $\mathbb{S}$ with a function in $\mathcal{F}$. Clearly, $(N_{f_{\varphi_1^m,\dots,\varphi_n^m}}(t))_{t\ge0}$ is a martingale by assumption for all $m=0,1,2,\dots$. Using (4.7) and dominated convergence

$$\begin{aligned}
\mathbf{E}[N_{f_{\varphi_1,\dots,\varphi_n}}(t)|(\underline{X}(r))_{r\le s}] &= \lim_{m\to\infty}\mathbf{E}[N_{f_{\varphi_1^m,\dots,\varphi_n^m}}(t)|(\underline{X}(r))_{r\le s}] \\
&= \lim_{m\to\infty} N_{f_{\varphi_1^m,\dots,\varphi_n^m}}(s) = N_{f_{\varphi_1,\dots,\varphi_n}}(s).
\end{aligned}$$

In other words, $(N_f(t))_{t\ge0}$ is a martingale. □

4.2. *Girsanov change of measure.* In Proposition 4.8 we establish a change of measure which shifts the selection coefficient α of Muller's ratchet with compensatory mutations. Two assertions from semimartingale theory which will be required in the proof are recalled in the next remark..

REMARK 4.7. 1. *A condition for a local martingale to be a martingale:* Let $\mathcal{N} = (N_t)_{t\geq 0}$ be a local martingale. If $\mathbf{E}[\sup_{0\leq t\leq T} |N_t|] < \infty$ for all $T > 0$, then $\mathcal{N}$ is a martingale (see e.g. Protter, 2004, Theorem I.51).

2. *Girsanov Theorem for continuous semimartingales:* If $\mathcal{L} = (L_t)_{t\geq 0}$ is a continuous $\mathbf{P}$-martingale for some probability measure $\mathbf{P}$ and assume that $\mathcal{Z} = (Z_t)_{t\geq 0}$, given by $Z_t = e^{L_t - \frac{1}{2}\langle \mathcal{L}\rangle_t}$, is a martingale (where $\langle \mathcal{L}\rangle$ is the predictable quadratic variation of $\mathcal{L}$). If $\mathcal{N} = (N_t)_{t\geq 0}$ is a $\mathbf{P}$-local martingale, and $\mathbf{Q}$ is defined via

$$\frac{d\mathbf{Q}}{d\mathbf{P}}\Big|_{\mathcal{F}_t} = Z_t,$$

then $\mathcal{N} - \langle \mathcal{L}, \mathcal{N}\rangle$ is a $\mathbf{Q}$-local martingale. Here, $\langle \mathcal{L}, \mathcal{N}\rangle$ is the (predictable) covariation process between $\mathcal{L}$ and $\mathcal{N}$. (See e.g. Kallenberg, 2002, 18.19, 18.21.)

PROPOSITION 4.8 (Change of measure). *For $\underline{y} \in \mathbb{S}$, let*

$$\kappa_1(\underline{y}) := \sum_{k=0}^{\infty} k y_k, \qquad \kappa_2(\underline{y}) := \sum_{k=0}^{\infty} (k - \kappa_1(\underline{y}))^2 y_k \tag{4.12}$$

be the expectation and variance of $\underline{y}$, provided they exist. Let $\underline{x} \in \mathbb{S}_\xi$ for some $\xi > 0$ and $\mathcal{X} = (X(t))_{t\geq 0}$ be a solution of the $(\mathbb{S}, \underline{x}, G^\alpha_{\mathcal{X}}, \mathcal{F})$-martingale problem and denote its distribution by $\mathbf{P}^\alpha$. Then, the process $\mathcal{Z}^{\alpha,\alpha'} = (Z^{\alpha,\alpha'}_t)_{t\geq 0}$, given by

$$\begin{aligned} Z^{\alpha,\alpha'}_t = \exp\Big(N(\alpha - \alpha')\Big(&\kappa_1(\underline{X}(t)) - \kappa_1(\underline{x}) \\ &- \int_0^t \lambda - \gamma\kappa_1(\underline{X}(s)) - \frac{\alpha + \alpha'}{2}\kappa_2(\underline{X}(s))ds\Big)\Big) \end{aligned} \tag{4.13}$$

is a $\mathbf{P}^\alpha$-local martingale. If $\alpha' > \alpha$ it is even a $\mathbf{P}^\alpha$-martingale and the probability measure $\mathbf{P}^{\alpha'}$, defined by

$$\frac{d\mathbf{P}^{\alpha'}}{d\mathbf{P}^\alpha}\Big|_{\mathcal{F}_t} = Z^{\alpha,\alpha'}_t$$

solves the $(\mathbb{S}, \underline{x}, G^{\alpha'}_{\mathcal{X}}, \mathcal{F})$-martingale problem.

Proof. The proof is an application of the Girsanov transform for continuous semimartingales; see Remark 4.7.2. By assumption, the process $\mathcal{X}$ is continuous, and so is the processes $(f(\underline{X}(t)))_{t\geq 0}$ for f as in (4.11). Set

$$g(\underline{x}) := N(\alpha - \alpha')\kappa_1(\underline{x})$$

and define $\mathcal{L} = (L_t)_{t\geq 0}$ by

$$\begin{aligned} L_t &= N(\alpha-\alpha')\Big(\kappa_1(\underline{X}(t)) - \kappa_1(\underline{X}(0)) - \int_0^t G^\alpha_{\mathcal{X}}\kappa_1(\underline{X}(s)ds\Big) \\ &= N(\alpha-\alpha')\Big(\kappa_1(\underline{X}(t)) - \kappa_1(\underline{X}(0)) - \int_0^t \lambda - \gamma\kappa_1(\underline{X}(s)) - \alpha\kappa_2(\underline{X}(s))ds\Big). \end{aligned}$$

Then, $\mathcal{L}$ is a $\mathbf{P}^\alpha$-martingale by Lemma 4.6 with quadratic variation

$$\begin{aligned} \langle\mathcal{L}\rangle_t &= N^2(\alpha-\alpha')^2\int_0^t G^N_{\text{res}}\big(\kappa_1(\underline{X}(s))\big)^2 ds \\ &= N(\alpha-\alpha')^2\int_0^t \kappa_2(\underline{X}(s))ds. \end{aligned}$$

For $f\in\mathcal{F}$, let $\mathcal{N}_f = (N_f(t))_{t\geq 0}$ be as in (4.6). Then, for $f = f_\varphi \in \mathcal{F}$,

$$\begin{aligned} \langle\mathcal{L},\mathcal{N}^f\rangle_t &= \int_0^t G^N_{\text{res}}\big(g(\underline{X}(s))f(\underline{X}(s))\big) - g(\underline{X}(s))G^N_{\text{res}}f(\underline{X}(s))ds \\ &= \frac{\alpha-\alpha'}{2}\int_0^t \sum_{k,\ell=0}^\infty X_k(s)(\delta_{k\ell} - X_\ell(s))(\varphi(k)\ell + k\varphi(\ell))ds \\ &= \int_0^t G^{\alpha'}_{\text{sel}}f(\underline{X}(s)) - G^\alpha_{\text{sel}}f(\underline{X}(s))ds. \end{aligned}$$

By an analogous calculation, one checks that the same identity is valid for all $f\in\mathcal{F}$. Since $\mathcal{L}$ is a $\mathbf{P}^\alpha$(-local) martingale, the process $\mathcal{Z}^{\alpha,\alpha'}$ as well is a $\mathbf{P}^\alpha$-local martingale (see Kallenberg, 2002, Lemma 18.21).

If $\alpha<\alpha'$ and $\underline{x}\in\mathbb{S}_\xi$ (and since $e^{\xi\kappa_1(\underline{x})}\leq h_\xi(\underline{x})$ by Jensen's inequality), we have that $\mathbf{E}[\sup_{0\leq t\leq T} Z^{\alpha,\alpha'}_t]<\infty$. Hence, using Remark 4.7.1, we see that $\mathcal{Z}^{\alpha,\alpha'}$ is a $\mathbf{P}^\alpha$-martingale. The above calculations and the Girsanov theorem for continuous semimartingales (recalled in Remark 4.7.2) then show that

$$N_f(t) - \langle\mathcal{L},\mathcal{N}_f\rangle_t = f(\underline{X}(t)) - \int_0^t G^{\alpha'}_{\mathcal{X}}f(\underline{X}(s))ds$$

is a $\mathbf{P}^{\alpha'}$-martingale. Since $f\in\mathcal{F}$ was arbitrary, $\mathbf{P}^{\alpha'}$ solves the $(\mathbb{S},\underline{x},G^{\alpha'}_{\mathcal{X}},\mathcal{F})$-martingale problem. □

4.3. *Proof of Theorem 1.* First we will prove Proposition 4.4 on the well-posedness of the martingale problem for $G^\alpha_{\mathcal{X}}$. The proof of Theorem 1 will then be completed by observing that a process solves the system of SDEs $(*)$ iff it solves the martingale problem for $G^\alpha_{\mathcal{X}}$ (Lemma 4.9).

Proof of Proposition 4.4.
Step 1: Existence of a solution of the martingale problem:
For $\alpha = 0$, it follows from classical theory (e.g. Dawson, 1993, Theorem 5.4.1) that the $(\mathbb{S}, \underline{x}, G^0_{\mathcal{X}}, \mathcal{F})$-martingale problem has a unique solution $\mathbf{P}^0$. By Proposition 4.8, the change of measure using the martingale $\mathcal{Z}^{0,\alpha}$ leads to a distribution $\mathbf{P}^\alpha$ that solves the $(\mathbb{S}, \underline{x}, G^\alpha_{\mathcal{X}}, \mathcal{F})$-martingale problem. This establishes existence.

Step 2: Uniqueness of solutions of the martingale problem:
As in Step 1, let $\mathbf{P}^0$ be the unique solution of the $(\mathbb{S}, \underline{x}, G^0_{\mathcal{X}}, \mathcal{F})$-martingale problem. Assume $\mathbf{P}^\alpha_1$ and $\mathbf{P}^\alpha_2$ are two different solutions of the $(\mathbb{S}, \underline{x}, G^\alpha_{\mathcal{X}}, \mathcal{F})$-martingale problem. Let $\tau_1, \tau_2, ...$ be stopping times with $\tau_n \to \infty$ as $n \to \infty$ such that $(Z^{\alpha,0}_{t\wedge\tau_n})_{t\geq 0}$, given by (4.13), is both a $\mathbf{P}^\alpha_1$-martingale and a $\mathbf{P}^\alpha_2$-martingale. Since $\mathbf{P}^\alpha_1 \neq \mathbf{P}^\alpha_2$, there must be $t \geq 0$ such that the distributions of $\underline{X}(t)$ under $\mathbf{P}^\alpha_1$ and $\mathbf{P}^\alpha_2$ are different (see Theorem 4.4.2 in Ethier and Kurtz, 1986). Hence, there is an $n \in \mathbb{N}$ such that the distributions of $\underline{X}(t\wedge\tau_n)$ under $\mathbf{P}^\alpha_1$ and $\mathbf{P}^\alpha_2$ are different. Since $Z^{\alpha,0}_{t\wedge\tau_n}$ is positive $\mathbf{P}^\alpha_1$-a.s. and $\mathbf{P}^\alpha_2$-a.s., then also the distributions of $\underline{X}(t \wedge \tau_n)$ under $Z^{\alpha,0}_{t\wedge\tau_n} \cdot \mathbf{P}^\alpha_1$ and $Z^{\alpha,0}_{t\wedge\tau_n} \cdot \mathbf{P}^\alpha_2$ are different. However, by the same arguments as in the proof of Proposition 4.8, $Z^{\alpha,0}_{t\wedge\tau_n} \cdot \mathbf{P}^\alpha_1$ as well as $Z^{\alpha,0}_{t\wedge\tau_n} \cdot \mathbf{P}^\alpha_2$ equal $\mathbf{P}^0$ on the σ-algebra $\sigma((\underline{X}(s))_{0\leq s\leq t\wedge\tau_n})$, which contradicts the assumed inequality of $\mathbf{P}^\alpha_1$ and $\mathbf{P}^\alpha_2$. Thus, uniqueness of the $(\mathbb{S}, \underline{x}, G^\alpha_{\mathcal{X}}, \mathcal{F})$-martingale problem follows. □

Lemma 4.9 (Equivalence of SDEs and martingale problem). *For $\underline{x} \in \mathbb{S}_\xi$, a process $\mathcal{X} = (\underline{X}(t))_{t\geq 0}$ is a weak solution of the system of SDEs $(*)$ starting in $\underline{x}$ iff the distribution of $\mathcal{X}$ is a solution to the $(\mathbb{S}, \underline{x}, G^\alpha_{\mathcal{X}}, \mathcal{F})$-martingale problem.*

Proof. 1. Assume that $\mathcal{X} = (\underline{X}(t))_{t\geq 0}$ solves the system of SDEs $(*)$. Then, as a direct consequence of Itô's lemma, the distribution of $\mathcal{X}$ is a solution to the $(\mathcal{S}, \underline{x}, G^\alpha_{\mathcal{X}}, \mathcal{F})$-martingale problem.

2. Conversely, let $\mathcal{X} = (\underline{X}(t))_{t\geq 0}$ solve the $(\mathbb{S}, \underline{x}, G^\alpha_{\mathcal{X}}, \mathcal{F})$-martingale problem. To see that $\mathcal{X}$ is a weak solution of $(*)$, we may appeal to Da Prato and Zabczyk, 1992, Theorem 8.2. Specifically, in their notation, choose H as the Hilbert-space of square-summable elements of $\mathbb{R}^{\mathbb{N}_0}$, $M = (\mathcal{N}_{f^k})_{k=0,1,2,...}$

with $f^k(\underline{x}) := x_k$, let Q be the identity on $\mathbb{R}^{\binom{\mathbb{N}_0}{2}} = \{(w_{k\ell})_{k<\ell} : w_{k\ell} \in \mathbb{R}\}$, and let $\Phi(s) : \mathbb{R}^{\binom{\mathbb{N}_0}{2}} \to \mathbb{R}^{\mathbb{N}_0}$ be given through the matrix $\Phi(s)_{i,k\ell} := (\delta_{ik} - \delta_{i\ell})\sqrt{X_k(s)X_\ell(s)}$. $\square$

5. Proof of Theorem 2. The key element in the proof of Theorem 2 is Proposition 5.2 which represents a solution of (3.5) through a Markov jump process. For uniqueness of the solution we rely on a duality derived in Section 5.2. The proof of Theorem 2 is given in Section 5.3.

5.1. *A particle system.* As a preparation to the proof of Theorem 2, we represent the system of ordinary differential equations by a jump process $(K_t)_{t\geq 0}$. Almost surely, the process will be killed (i.e. hit a cemetery state) in finite time. We show in Proposition 5.2 that a solution of (3.5) is given by the distribution of K_t conditioned on not being killed by time t, $t \geq 0$.

DEFINITION 5.1 (Jump process). *Let $(K_t)_{t\geq 0}$ be a pure Markov jump process which takes values in $\{\dagger, 0, 1, 2, ...\}$ and jumps from k to $k+1$ at rate λ, from k to $k-1$ at rate $k\gamma$, and from k to the cemetery state $\dagger$ with rate αk.*

PROPOSITION 5.2 (Particle representation). *Let $\underline{x}(0) \in \mathbb{S}_\xi$ for some $\xi > 0$ and $(K_t)_{t\geq 0}$ be as in Definition 5.1 with initial distribution given by $\mathbf{P}[K_0 = k] = x_k(0)$. Then,*

$$x_k(t) := \mathbf{P}[K_t = k | K_t \neq \dagger]. \tag{5.1}$$

solves the system (3.5).

PROOF. From the definition of $(K_t)_{t\geq 0}$, it is clear that for small $\varepsilon > 0$,

$$\begin{aligned} x_k(t+\epsilon) &= \frac{x_k(t)(1-\alpha k\varepsilon) + \lambda(x_{k-1}(t) - x_k(t))\varepsilon + \gamma((k+1)x_{k+1}(t) - kx_k(t))\varepsilon}{1 - \alpha\sum_{\ell=0}^\infty \ell x_\ell(t)\varepsilon} \\ &\qquad + \mathcal{O}(\varepsilon^2) \\ &= x_k(t) + \Big(-\alpha\Big(k - \sum_{\ell=0}^\infty \ell x_\ell(t)\Big)x_k(t) + \lambda(x_{k-1}(t) - x_k(t)) \\ &\qquad + \gamma((k+1)x_{k+1}(t) - kx_k(t))\Big)\varepsilon + \mathcal{O}(\varepsilon^2), \end{aligned}$$

which implies the result as $\varepsilon \to 0$. $\square$

5.2. *Dynamics of the cumulant generating function.* The proof of uniqueness of (3.5) requires some preliminary computations which we carry out next. Recall the function h_ξ from (3.2). Note the function $\zeta \mapsto \log h_\zeta(\underline{x})$ is the cumulant generating function of $\underline{x} \in \mathbb{S}$. Cumulants have already be proven to be useful in the study of Muller's ratchet; see Etheridge et al. (2008). Here, we compute the dynamics of the cumulant generating function.

Proposition 5.3 (Dynamics of cumulant generating function). *For any solution $t \mapsto \underline{x}(t)$ of (3.5) taking values in $\mathbb{S}_\xi$ for $\xi > 0$ and $0 < \zeta < \xi$,*

$$\frac{d}{dt}\log h_\zeta(\underline{x}(t)) = \alpha \sum_{\ell=0}^{\infty} \ell x_\ell(t) + \lambda(e^\zeta - 1) - (\alpha + \gamma(1 - e^{-\zeta}))\frac{d}{d\zeta}\log h_\zeta(\underline{x}(t)).$$

Proof. Abbreviating $\underline{x} := \underline{x}(t)$, we compute

$$\begin{aligned} h_\zeta(\underline{x})\frac{d}{dt}\log h_\zeta(\underline{x}) &= \alpha \sum_{\ell=0}^{\infty}\sum_{k=0}^{\infty}(\ell - k)x_\ell x_k e^{\zeta k} + \lambda \sum_{k=0}^{\infty}(x_{k-1} - x_k)e^{\zeta k} \\ &\qquad + \gamma \sum_{k=0}^{\infty}((k+1)x_{k+1} - kx_k)e^{\zeta k} \\ &= \alpha\Big(\Big(\sum_{\ell=0}^{\infty}\ell x_\ell\Big)h_\zeta(\underline{x}) - \frac{d}{d\zeta}h_\zeta(\underline{x})\Big) + \lambda(e^\zeta - 1)h_\zeta(\underline{x}) \\ &\qquad - \gamma(1 - e^{-\zeta})\frac{d}{d\zeta}h_\zeta(\underline{x}) \end{aligned}$$

and so

$$\frac{d}{dt}\log h_\zeta(\underline{x}) = \alpha \sum_{\ell=0}^{\infty}\ell x_\ell + \lambda(e^\zeta - 1) - (\alpha + \gamma(1 - e^{-\zeta}))\frac{d}{d\zeta}\log h_\zeta(\underline{x}).$$

□

The equation in Proposition 5.3 relates the time-derivative of $\log h_\zeta(\underline{x}(t))$ with the ζ-derivative of the same function and leads to a *duality relation* formulated in Corollary 5.4. In Markov process theory, dualities are particularly useful to obtain uniqueness results; cf. Ethier and Kurtz (1986), p. 188ff. Our application in Subsection 5.3 will be in this spirit.

COROLLARY 5.4 (Duality). *Let $t \mapsto \underline{x}(t)$ be a solution of* (3.5) *taking values in $\mathbb{S}_\xi$ for $\xi > 0$. Moreover let $\zeta : t \mapsto \zeta(t)$ be the solution of $\zeta' = -(\alpha + \gamma(1 - e^{-\zeta}))$, starting in some $\zeta(0) < \xi$. Then,*

$$\log h_{\zeta(0)}(\underline{x}(t)) = \log h_{\zeta(t)}(\underline{x}(0)) + \int_0^t \Big(\lambda(e^{\zeta(t-s)} - 1) + \sum_{\ell=0}^{\infty} \ell x_\ell(s)\Big) ds.$$

PROOF. Using Proposition 5.3 and noting, for any differentiable $g : \zeta \mapsto g(\zeta)$, the equality

$$\frac{d}{ds} g(\zeta(t-s)) = (\alpha + \gamma(1 - e^{-\zeta(t-s)}))\frac{d}{d\zeta} g(\zeta(t-s)),$$

we obtain

$$\frac{d}{ds} \log h_{\zeta(t-s)}(\underline{x}(s)) = \lambda(e^{\zeta(t-s)} - 1) + \alpha \sum_{\ell=0}^{\infty} \ell x_\ell(s).$$

Now the assertion follows by integrating. □

5.3. *Proof of Theorem 2.* We proceed in two steps. First, we derive an explicit solution of (3.5) by using Proposition 5.2, i.e. by computing the distribution of the jump process $(K_t)_{t\geq 0}$ conditioned on not being killed by time t. This will result in the right hand side of (3.6). In a second step, we show uniqueness of solutions of (3.5) in $\mathbb{S}_\xi$.

Step 1: Computation of the right hand side of (5.1):
In order to derive an explicit formula for the probability specified in (5.1), we note that the process $(K_t)_{t\geq 0}$ can be realized as the following mutation-couting process:

- Start with K_0 mutations, with the random number K_0 distributed according to $(x_k(0))_{k=0,1,2,\dots}$.
- New mutations arise at rate λ.
- Every mutation (present from the start or newly arisen) starts an exponential waiting time with parameter $\alpha + \gamma$. If this waiting time expires, then with probability $\frac{\alpha}{\alpha+\gamma}$ the process jumps to †, and with the complementary probability $\frac{\gamma}{\alpha+\gamma}$ the mutation disappears.

With $x_k(t)$ defined by (5.1), we decompose the probability of the event $\{K_t = k\}$ with respect to the number of mutations present at time 0. If $K_0 = i$, a number $j \leq i \wedge k$ of these initial mutations are not compensated by time t and the remaining $i - j$ are compensated. In addition, a number

$l \geq k-j$ mutations arise at times $0 \leq t_1 \leq \cdots \leq t_l \leq t$. From these, $l-k+j$ are compensated and the remaining $k-j$ are not compensated. These arguments lead to the following calculation, where we write $\sim$ for equality up to factors not depending on k. The first $\sim$ comes from the fact that the right hand side is the unconditional probability $\mathbf{P}[K_t = k]$,

$$
\begin{aligned}
x_k(t) &\sim \sum_{i=0}^{\infty} x_i(0) \sum_{j=0}^{i\wedge k} \binom{i}{j} \Big(\frac{\gamma}{\alpha+\gamma}\big(1-e^{-(\alpha+\gamma)t}\big)\Big)^{i-j} \cdot e^{-j(\alpha+\gamma)t} \\
&\qquad \sum_{l=k-j}^{\infty} \int_{\substack{\{\mathcal{T}=(t_1,\ldots,t_l:\\ 0\leq t_1\leq\cdots\leq t_l\}}} d(t_1,...,t_l)\lambda^l e^{-\lambda t_1} e^{-\lambda(t_2-t_1)} \cdots e^{-\lambda(t_l-t_{l-1})} e^{-\lambda(t-t_l)} \\
&\qquad \sum_{\substack{\mathcal{S}\subseteq\mathcal{T}\\ |\mathcal{S}|=l-k+j}} \prod_{r\in\mathcal{S}} \frac{\gamma}{\alpha+\gamma}\big(1-e^{-(\alpha+\gamma)(t-r)}\big) \cdot \prod_{s\in\mathcal{T}\setminus\mathcal{S}} e^{-(\alpha+\gamma)(t-s)} \\
&= \sum_{i=0}^{\infty} x_i(0) \sum_{j=0}^{i\wedge k} \binom{i}{j} \Big(\frac{\gamma}{\alpha+\gamma}\big(1-e^{-(\alpha+\gamma)t}\big)\Big)^{i-j} \cdot e^{-j(\alpha+\gamma)t} \\
&\qquad \sum_{l=k-j}^{\infty} \frac{\lambda^l}{l!} e^{-\lambda t} \int_{\substack{\{\mathcal{T}=(t_1,\ldots,t_l:\\ 0\leq t_1,\ldots,t_l\}}} d(t_1,...,t_l) \sum_{\substack{\mathcal{S}\subseteq\mathcal{T}\\ |\mathcal{S}|=l-k+j}} \\
&\qquad\quad \prod_{r\in\mathcal{S}} \frac{\gamma}{\alpha+\gamma}\big(1-e^{-(\alpha+\gamma)(t-r)}\big) \prod_{s\in\mathcal{T}\setminus\mathcal{S}} e^{-(\alpha+\gamma)(t-s)} \\
&\sim \sum_{i=0}^{\infty} x_i(0) \sum_{j=0}^{i\wedge k} \binom{i}{j} \Big(\frac{\gamma}{\alpha+\gamma}\big(1-e^{-(\alpha+\gamma)t}\big)\Big)^{i-j} e^{-j(\alpha+\gamma)t} \\
&\qquad \cdot \frac{\lambda^{k-j}}{(k-j)!} \Big(\int_0^t e^{-(\alpha+\gamma)(t-s)} ds\Big)^{k-j} \\
&\qquad\quad \cdot \sum_{l=0}^{\infty} \frac{\lambda^l}{l!} \Big(\frac{\gamma}{\alpha+\gamma} \int_0^t 1-e^{-(\alpha+\gamma)(t-r)} dr\Big)^l \\
&\sim \sum_{i=0}^{\infty} x_i(0) \sum_{j=0}^{i\wedge k} \binom{i}{j} \Big(\frac{\gamma}{\alpha+\gamma}\big(1-e^{-(\alpha+\gamma)t}\big)\Big)^{i-j} e^{-j(\alpha+\gamma)t} \\
&\qquad\quad \cdot \frac{\lambda^{k-j}}{(k-j)!} \Big(\frac{1}{\alpha+\gamma}(1-e^{-(\alpha+\gamma)t})\Big)^{k-j},
\end{aligned}
$$

where the first '=' comes from the symmetry of the integrand. Summing the right hand side gives

$$\sum_{i=0}^{\infty} x_i(0) \sum_{j=0}^{i} \binom{i}{j} \Big(\frac{\gamma}{\alpha+\gamma}\big(1-e^{-(\alpha+\gamma)t}\big)\Big)^{i-j} e^{-j(\alpha+\gamma)t} \cdot \sum_{k=j}^{\infty} \frac{\lambda^{k-j}}{(k-j)!}\Big(\frac{1}{\alpha+\gamma}(1-e^{-(\alpha+\gamma)t})\Big)^{k-j}$$

$$= \sum_{i=0}^{\infty} x_i(0)\Big(\frac{\gamma}{\alpha+\gamma} - \frac{\alpha}{\alpha+\gamma}e^{-(\alpha+\gamma)t}\Big)^i \cdot \exp\Big(\frac{\lambda}{\alpha+\gamma}(1-e^{-(\alpha+\gamma)t})\Big).$$

Hence,

$$x_k(t) = \frac{\sum\limits_{i=0}^{\infty} x_i(0) \sum\limits_{j=0}^{i\wedge k} \binom{i}{j}\big(\frac{\gamma(1-e^{-(\alpha+\gamma)t})}{\alpha+\gamma}\big)^{i-j} e^{-j(\alpha+\gamma)t}\frac{1}{(k-j)!}\big(\frac{\lambda(1-e^{-(\alpha+\gamma)t})}{\alpha+\gamma}\big)^{k-j}}{\sum\limits_{i=0}^{\infty} x_i(0)\big(\frac{\gamma}{\alpha+\gamma} - \frac{\alpha}{\alpha+\gamma}e^{-(\alpha+\gamma)t}\big)^i \cdot \exp\big(\frac{\lambda}{\alpha+\gamma}(1-e^{-(\alpha+\gamma)t})\big)}$$

which shows (3.6). To see (3.7), it suffices to note that all terms in the numerator except that for $j=0$ converge to 0 as $t\to\infty$. Hence the result is proved.

Step 2: Uniqueness in $\mathbb{S}_\xi$:
Let $(\underline{y}(t))_{t\geq 0}$ be a solution of (3.5) starting in $\underline{y}(0)\in\mathbb{S}_\xi$. From the analogue of Lemma 4.5 in the case $N=\infty$ we have $h_\xi(\underline{y}(t)) \leq h_\xi(\underline{y}(0))\exp(\lambda t(e^\xi - 1)) < \infty$, i.e. $\underline{y}(t)\in\mathbb{S}_\xi$ for all $t\geq 0$.

If $(\underline{x}(t))_{t\geq 0}$ and $(\underline{y}(t))_{t\geq 0}$ are solutions of (3.5) with $\underline{x}(0)=\underline{y}(0)\in\mathbb{S}_\xi$, then we obtain from Corollary 5.4 that for all $0<\zeta<\xi$ and any $t\geq 0$

$$\log h_\zeta(\underline{x}(t)) - \log h_\zeta(\underline{y}(t)) = \int_0^t \sum_{\ell=0}^{\infty} \ell(x_\ell(s)-y_\ell(s))ds. \tag{5.2}$$

Since only the left-hand side depends on ζ, this enforces that both sides vanish. Indeed, taking derivatives with respect to ζ at $\zeta=0$ the previous equality gives

$$\sum_{\ell=0}^{\infty} \ell(x_\ell(t)-y_\ell(t)) = 0, \quad t\geq 0.$$

Plugging this back into (5.2) gives

$$\log h_\zeta(\underline{x}(t)) = \log h_\zeta(\underline{y}(t)), \quad t\geq 0. \tag{5.3}$$

Since the function $\zeta\mapsto \log h_\zeta(\underline{x})$ (for $0<\zeta<\xi$) characterizes $\underline{x}\in\mathbb{S}_\xi$, we obtain that $\underline{x}(t)=\underline{y}(t)$.

Acknowledgments. Part of this work was carried out while PP and AW were visiting the program in Probability and Discrete Mathematics in Mathematical Biology at the University of Singapore in April 2011, whose hospitality is gratefully acknowledged. PP is supported by the BMBF through FRISYS (Kennzeichen 0313921).

References.

Antezana, M. A. and Hudson, R. R. (1997). Era reversibile! Point-mutations, the Ratchet, and the initial success of eukaryotic sex : a simulation study. *Evolutionary Theory*, 11:209–235.

Audiffren, J. (2011). Phd thesis in preparation. *Université de Provence, Marseille.*

Audiffren, J. and Pardoux, E. (2011). Muller's ratchet clicks in finite time. *submitted*, pages 1–29.

Chao, L. (1990). Fitness of RNA virus decreased by Muller's ratchet. *Nature*, 348:454–455.

Cuthbertson, C. (2007). Limits to the rate of adaptation. *DPhil Thesis, University of Oxford*.

Da Prato, G. and Zabczyk, J. (1992). *Stochastic equations in infinite dimensions.* Cambridge University Press.

Dawson, D. (1993). Measure-valued Markov processes. In Hennequin, P., editor, *École d'Été de Probabilités de Saint-Flour XXI–1991*, volume 1541 of *Lecture Notes in Mathematics*, pages 1–260, Berlin. Springer.

Dawson, D. A. (1978). Geostochastic calculus. *Canadian Journal of Statistics*, 6:143–168.

Depperschmidt, A., Pfaffelhuber, P., and Rensing, S. (2011). Mathematical modelling of plant evolution: RNA editing counteracts Mullers ratchet. *Manuscript.*

Desai, M. M. and Fisher, D. S. (2007). Beneficial mutation-selection balance and the effect of linkage on positive selection. *Genetics*, 176:1759–1798.

Etheridge, A., Pfaffelhuber, P., and Wakolbinger, A. (2008). How often does the ratchet click? facts, heuristics, asymptotics. In *Trends in Stochastic Analysis*, volume LMS 353. Cambridge University Press.

Ethier, S. and Kurtz, T. (1986). *Markov Processes. Characterization and Convergence.* John Wiley, New York.

Ethier, S. and Kurtz, T. (1993). Fleming-Viot processes in population genetics. *SIAM J. Contr. Optim.*, 31:345–386.

Ethier, S. N. and Shiga, T. (2000). A FlemingViot process with unbounded selection. *J. Math. Kyoto Univ.*, 40(2):337–361.

Gabriel, W., Lynch, M., and Bürger, R. (1993). Muller's ratchet and mutational meltdowns. *Evolution*, 47(6):1744–1757.

Gerrish, P. and Lenski, R. (1998). The fate of competing beneficial mutations in an asexual population. *Genetica*, 102/103:127–144.

Gessler, D. D. (1995). The constraints of finite size in asexual populations and the rate of the ratchet. *Genet. Res.*, 66(3):241–253.

Gordo, I. and Charlesworth, B. (2000). On the speed of Muller's ratchet. *Genetics*, 156(4):2137–2140.

Haigh, J. (1978). The accumulation of deleterious genes in a population–Muller's Ratchet. *Theor. Popul. Biol.*, 14(2):251–267.

Handel, A., Regoes, R. R., and Antia, R. (2006). The role of compensatory mutations in the emergence of drug resistance. *PLoS Comput. Biol.*, 2:e137.

Higgs, P. G. and Woodcock, G. (1995). The accumulation of mutations in asexual populations and the structure of genealogical trees in the presence of selection. *J. Math. Biol.*, 33:677–702.

Howe, D. K. and Denver, D. R. (2008). Muller's ratchet and compensatory mutation in *Caenorhabditis briggsae* mitochondrial genome evolution. *BMC Evol. Biol.*, 8:62–62.

Jain, K. (2008). Loss of least-loaded class in asexual populations due to drift and epistasis. *Genetics*, 179:2125–2134.

Kallenberg, O. (2002). *Foundations of Modern Probability*. Springer, 2nd edition.

Loewe, L. (2006). Quantifying the genomic decay paradox due to Muller's ratchet in human mitochondrial DNA. *Genet. Res.*, 87:133–159.

Maia, L. P., Botelho, D. F., and Fontatari, J. F. (2003). Analytical solution of the evolution dynamics on a multiplicative fitness landscape. *J. Math. Biol.*, 47:453–456.

Maisnier-Patin, S. and Andersson, D. I. (2004). Adaptation to the deleterious effects of antimicrobial drug resistance mutations by compensatory evolution. *Res. Microbiol.*, 155:360–369.

Maynard Smith, J. (1978). *The Evolution of Sex*. Cambridge University Press.

Muller, H. J. (1964). The relation of recombination and mutational advance. *Mutat. Res.*, 106:2–9.

Park, S.-C. and Krug, J. (2007). Clonal interference in large populations. *PNAS*, 104(46):18135–18140.

Poon, A. and Chao, L. (2005). The rate of compensatory mutation in the DNA bacteriophage phiX174. *Genetics*, 170:989–999.

Protter, P. E. (2004). *Stochastic Integration and Differential Equations*. Springer, 2nd edition.

Rouzine, I., Wakeley, J., and Coffin, J. (2003). The solitary wave of asexual evolution. *Proc. Natl. Acad. Sci.*, 100(2):587–592.

Stephan, W., Chao, L., and Smale, J. (1993). The advance of Muller's ratchet in a haploid asexual population: approximate solutions based on diffusion theory. *Genet. Res.*, 61(3):225–231.

Wagner, G. and Gabriel, W. (1990). What stops Muller's ratchet in the absence of recombination? *Evolution*, 44:715–731.

Waxman, D. and Loewe, L. (2010). A stochastic model for a single click of Mullers ratchet. *J. Theo. Biol*, 264:1120–1132.

Yu, F., Etheridge, A., and Cuthbertson, C. (2010). Asymptotic behavior of the rate of adaptation. *Ann. Appl. Probab.*, 20(3):978–1004.

P. Pfaffelhuber
Fakultät für Mathematik und Physik
Eckerstrasse 1
D-79104 Freiburg
Germany
E-mail: p.p@stochastik.uni-freiburg.de

P. R. Staab
Fakultät für Mathematik und Physik
Eckerstrasse 1
D-79104 Freiburg
Germany

A. Wakolbinger
Fakultät für Informatik und Mathamatik
Robert-Mayer-Str. 10
D-60325 Frankfurt
Germany