The emergence of a birth-dependent mutation rate: causes and consequences

Florian Patout, R Forien, M Alfaro, Julien Papaïx, L Roques

The emergence of a birth-dependent mutation rate: causes and consequences.

F. Patout [a], R. Forien [a], M. Alfaro [a, b], J. Papaïx [a] and L. Roques [a]

[a] BioSP, INRAE, 84914, Avignon France

[b] Université de Rouen Normandie, CNRS,

Laboratoire de Mathématiques Raphaël Salem, Saint-Etienne-du-Rouvray, France

Abstract

In unicellular organisms such as bacteria and in most viruses, mutations mainly occur during reproduction. Thus, genotypes with a high birth rate should have a higher mutation rate. However, standard models of asexual adaptation such as the 'replicator-mutator equation' often neglect this effect. In this study, we investigate the emergence of a positive dependence between the birth rate and the mutation rate in models of asexual adaptation and the consequences of this dependence. We show that it emerges naturally at the population scale, based on a large population limit of a stochastic time-continuous individual-based model with elementary assumptions. We derive a reaction-diffusion framework that describes the evolutionary trajectories and steady states in the presence of this dependence. When this model is coupled with a phenotype to fitness landscape with two optima, one for birth, the other one for survival, a new trade-off arises in the population. Compared to the standard approach with a constant mutation rate, the symmetry between birth and survival is broken. Our analytical results and numerical simulations show that the trajectories of mean phenotype, mean fitness and the stationary phenotype distribution are in sharp contrast with those displayed for the standard model. Here, we obtain trajectories of adaptation where the mean phenotype of the population is initially attracted by the birth optimum, but eventually converges to the survival optimum, following a hook-shaped curve which illustrates the antagonistic effects of mutation on adaptation.

1 Introduction

The effect of the mutation rate on the dynamics of adaptation is well-documented, both experimentally (e.g, Giraud et al., 2001; Anderson et al., 2004) and theoretically. Regarding theoretical work, since the first studies on the accumulation of mutation load (Haldane, 1937; Kimura and Maruyama, 1966), several modelling approaches have investigated the effect of the mutation rate on various aspects of the adaptation of asexuals. This includes lethal mutagenesis theory (Bull et al., 2007;

Bull and Wilke, 2008), where too high mutation rates lead to extinction, evolutionary rescue (Anciaux et al., 2019) or the invasion of a sink (Lavigne et al., 2020). The evolution of the mutation rate per se is also the subject of several models (André and Godelle, 2006; Lynch, 2010).

Though it is now established that mutation rates are not always uniform among individuals of the same species, most models that describe the dynamics of adaptation of phenotypically structured populations assume a constant mutation rate across phenotypes (e.g., Desai and Fisher, 2011; Gerrish et al., 2007; Sniegowski and Gerrish, 2010; Alfaro and Carles, 2014; Gandon and Mirrahimi, 2017; Gil et al., 2019). Variations in the individual mutation rate per generation can be caused by genotypic variability (Sharp and Agrawal, 2012), environmental factors (Hoffmann and Hercus, 2000) or more generally 'G x E' interactions. In a spirit of parsimony, the above-mentioned modelling approaches deliberately ignore these processes but do take into account a certain variability in the reproductive success. As some cancer studies emphasize, with the observation of dose-dependent mutation rates (Liu et al., 2015), the mutation rate at the population scale can be correlated with the reproductive success, through the individual birth rate. The main goal of the current study is to determine in which context the variability of the birth rate must result in a variability in the rate at which new mutations arise in the population. We also study the consequences of such birth rate - mutation rate dependence on the evolutionary trajectory of the population.

In multicellular sexual organisms, this dependence does not emerge naturally. Germline mutations, which can be transmitted to the offspring, appear during cellular division, and therefore, are not *a priori* correlated to birth events. This means that, for a given germline mutation rate, such organisms who live a long time would accumulate as much mutations *per units of time* as one living for a shorter period of time, but with more birth events. The picture is very different for unicellular organisms such as bacteria. In that case, mutations occur during reproduction, by means of binary fission (Van Harten, 1998; Trun and Trempy, 2009), meaning that individuals with a high birth rate should have a higher mutation rate (they produce more mutant offspring per unit of time). This is also true for viruses, as mutations mostly arise during replication (Sanjuán and Domingo-Calap, 2016). The probability of mutation during the replication is even greater in RNA viruses as their polymerase lacks the proofreading activity found in the polymerase of DNA viruses (Lauring et al., 2013).

The consequences of the dependence between birth rate and mutation rate are not easy to anticipate as birth rate is also involved in trade-offs with other life-history traits. Such trade-offs play a crucial role in shaping evolution (Stearns, 1989). They create evolutionary compromises, for instance between dispersal and reproduction (Nathan, 2001; Smith et al., 2014; Helms and Kaspari, 2015; Xiao et al., 2015) or between the traits related to survival and those related to birth (Taylor, 1991). In this last case, we expect that the consequences of the trade-off on the dynamics of adaptation strongly depend on the existence of a positive correlation between the birth rate and the mutation rate. High mutation rates tend to promote adaptation when the population is far from equilibrium (Sniegowski et al., 2000) but eventually have a detrimental effect due to a higher mutation load (Anciaux et al.,

2019) when it approaches a mutation-selection equilibrium. This ambivalent effect of mutation may therefore lead to complex trajectories of adaptation when the birth and mutation rates are correlated.

In the classical models describing the dynamics of adaptation of a phenotypically structured population, the breeding values at a set of n traits are described by a vector $\mathbf{x} \in \Omega \subset \mathbb{R}^n$, that we call here 'phenotype' for simplicity. The effect of mutations on the phenotype distribution is described through a linear operator $\mathcal{M}$ which does not depend on the parent phenotype $\mathbf{x}$. The operator $\mathcal{M}$ can be described with a convolution product involving a mutation kernel (Champagnat et al., 2006; Gil et al., 2017) or with a Laplace operator (Kimura, 1964; Lande, 1975; Alfaro and Carles, 2014; Hamel et al., 2020), corresponding to a diffusion approximation of the mutation effects. Under the diffusion approximation, $\mathcal{M}(\cdot) = D\,\Delta(\cdot)$ with $D > 0$ a constant coefficient which is proportional to the mutation rate and to the mutational variance at each trait. The Malthusian fitness $m(\mathbf{x})$, *i.e.* the Malthusian growth rate of individuals with phenotype $\mathbf{x}$, is defined as the difference between the birth rate $b(\mathbf{x})$ and death rate $d(\mathbf{x})$ of this class of individuals:

$$m(\mathbf{x}) = b(\mathbf{x}) - d(\mathbf{x}). \tag{1}$$

The following generic equation then describes the combined effects of mutation and selection on the dynamics of the phenotype density $f(t, \mathbf{x})$ under a diffusive approximation of the mutation effects:

$$\partial_t f(t, \mathbf{x}) = D\,\Delta(f)(t, \mathbf{x}) + f(t, \mathbf{x}) m(\mathbf{x}), \tag{2}$$

in the absence of density-dependent competition, or

$$\partial_t f(t, \mathbf{x}) = D\,\Delta(f)(t, \mathbf{x}) + f(t, \mathbf{x}) \left(m(\mathbf{x}) - \int_\Omega f(t, \mathbf{y})\, d\mathbf{y} \right), \tag{3}$$

if density-dependent competition is taken into account. In both cases, the equation satisfied by the frequency $q(t, \mathbf{x}) = f(t, \mathbf{x})/N(t)$ (with $N(t) = \int_\Omega f(t, \mathbf{x}) d\mathbf{x}$ the total population size) is

$$\partial_t q(t, \mathbf{x}) = D\,\Delta(q)(t, \mathbf{x}) + q(t, \mathbf{x})(m(\mathbf{x}) - \overline{m}(t)), \qquad (\mathcal{Q}_{stand})$$

with $\overline{m}(t)$ the mean fitness in the population:

$$\overline{m}(t) = \int_\Omega m(\mathbf{x})\, q(t, \mathbf{x})\, d\mathbf{x}. \tag{4}$$

These models allowed a broad range of results in various biological contexts: concentration around specific traits (Diekmann et al., 2005; Lorz et al., 2011; Martin and Roques, 2016); explicit solutions (Alfaro and Carles, 2014; Biktashev, 2014; Alfaro and Carles, 2017); moving and/or fluctuating optimum (Figueroa Iglesias and Mirrahimi, 2019; Roques et al., 2020); anisotropic mutation effects (Hamel et al., 2020). Then can aslo be extended in order to take migration events into account (Bouin and Mirrahimi, 2013; Débarre et al., 2013; Lavigne et al., 2020).

With these models, the dynamics of adaptation and the equilibria only depend on the birth and death rates through their difference $m(\mathbf{x}) = b(\mathbf{x}) - d(\mathbf{x})$. Thus,

these models do not discriminate between phenotypes for which both birth and death rates are high compared to those for which they are both low, given that the difference is constant. However, as explained above for unicellular organisms, the mutation rate may be positively correlated with the birth rate which could generate an imbalance in favour of one of the two strategies: having a high birth rate vs. having a high survival rate. To acknowledge the role of phenotype-dependent birth rate and the resulting asymmetric effects of fertility and survival in a deterministic setting, a new paradigm is necessary.

In this work, we consider the case of mutations that occur during the reproduction of asexual organisms. We assume that the probability of mutation per birth event U *does not* depend on the phenotype of the parent. On the other hand, following classical adaptive landscape approaches (Tenaillon, 2014), the birth and death rates do depend on the phenotype. Using these basic assumptions, we consider in Section 2 standard stochastic individual-based models of adaptation with mutation and selection. We present how the standard model ($\mathcal{Q}_{stand}$) appears naturally as a large population limit of both a discrete time model and a continuous time model when the variance of mutation effects is small and when selection is weak (*i.e.* when the variations of the birth and death rates across the phenotype space are very small). In this work, however, we are interested in a particular setting where this assumption is not satisfied. In this case, using results from Fournier and Méléard (2004) for the continuous-time model, we argue that, when the mutation variance is small, a more accurate approximation of the mutation operator is given by $\mathcal{M}(q) = D\Delta(b(\cdot)\, q)$, leading to a new equation of the form

$$\partial_t q(t,\mathbf{x}) = D\Delta(b\, q)(t,\mathbf{x}) + q(t,\mathbf{x})(m(\mathbf{x}) - \overline{m}(t)). \qquad (\mathcal{Q}_b)$$

Here, the mutation operator $D\Delta(b(\mathbf{x})\, q)(t,\mathbf{x})$ depends on the phenotype $\mathbf{x}$ through the birth rate $b(\mathbf{x})$, translating the fact that new mutants appear at a higher rate when the birth rate increases.

In Section 3, we use $(\mathcal{Q}_b)$ to study the evolution of the phenotype distribution when the population is subjected to a trade-off between a birth optimum and a survival optimum, and we highlight the main differences with the standard approach ($\mathcal{Q}_{stand}$). More specifically, we study the evolution of the phenotype distribution in the presence of a fitness optimum where b and d are both large (the birth or reproduction optimum), and a survival optimum, where b and d are both small, but such that the difference $b - d$ is symmetrical.

Based on analytical results and numerical simulations, we compare the trajectories of adaptation and the equilibrium phenotype distributions between these two approaches and we check their consistency with the underlying individual-based models. We discuss these results in Section 4. Mathematical proofs are postponed to Appendix A.

2 Emergence of a birth-dependent mutation rate in an individual-based setting

In this section, we present how the standard equation ($\mathcal{Q}_{stand}$) and the new model ($\mathcal{Q}_b$) with birth-dependent mutation rate are obtained from large population limits of stochastic individual-based models. We first state a convergence result due to Fournier and Méléard (2004) which provides the convergence of the phenotype distribution of the population to the solution of an integro-differential equation, when the size of the population tends to infinity. We then show that, when the variance of the mutation effects is small, this equation yields the new model equation ($\mathcal{Q}_b$). This shows how a dependence between the birth rate and the rate at which new mutant appear in the population arises, even though the probability of mutation per birth event U does not depend on the phenotype of the parent. We then treat the case of weak selection, and show how the model ($\mathcal{Q}_{stand}$) is obtained as a large population limit of the phenotype distribution with a specific time scaling, using results in Champagnat et al. (2008). We also state an analogous result for the discrete time model, where in the same regime of weak selection and small mutation effects, we show the convergence to the solution of ($\mathcal{Q}_{stand}$) as the population size tends to infinity, on the same timescale as the other model.

In this individual-based setting, we consider a finite population where each individual carries a phenotype in a bounded open set $\Omega \subset \mathbb{R}^d$. If the individuals at time t have traits $\{\mathbf{x}_1, \dots, \mathbf{x}_{N_t}\}$, we record the state of the population through the empirical measure

$$\nu_t^K = \frac{1}{K} \sum_{i=1}^{N_t} \delta_{\mathbf{x}_i}, \qquad t \geq 0,$$

where $\delta_{\mathbf{x}}$ is the Dirac measure at the point $\mathbf{x} \in \Omega$ and N_t is the current number of individuals in the population. Note that the number of individuals N_t in this stochastic individual-based setting does not correspond to the quantity $N(t)$ defined in the Introduction. In fact, these two quantities will be related via the scaling parameter $K > 0$: $N(t) = \lim_{K\to\infty} N_t/K$ (or $N(t) = \lim_{K\to\infty} N_{t/\varepsilon_K}/K$ if time is rescaled, see the definition of ε_K below).

Let $M_F(\Omega)$ denote the space of finite measures on Ω, endowed with the topology of weak convergence. For any $\nu \in M_F(\Omega)$ and any measurable and bounded function $\phi : \Omega \to \mathbb{R}$, we shall write

$$\langle \nu, \phi \rangle = \int_\Omega \phi(\mathbf{x}) \nu(d\mathbf{x}).$$

2.1 Derivation of the model ($\mathcal{Q}_b$) with birth-dependent mutation rate

We first consider a continuous time stochastic individual-based model where individuals die and reproduce at random times depending on their phenotype and the current population size. We let $b : \Omega \to \mathbb{R}_+$ and $d : \Omega \to \mathbb{R}_+$ be two bounded and

measurable functions, and we assume that an individual with phenotype $\mathbf{x} \in \Omega$ reproduces at rate $b(\mathbf{x})$ and dies at rate $d(\mathbf{x}) + c_K N_t$ for some $c_K > 0$. This parameter c_K measures the intensity of competition between the individuals in the population, and prevents the population size from growing indefinitely. Each newborn individual either carries the phenotype of its parent, with probability $1 - U$, or, with probability U, carries a phenotype $\mathbf{y}$ chosen at random from some distribution $\rho(\mathbf{x}, \mathbf{y})d\mathbf{y}$, where $\mathbf{x}$ is the phenotype of its parent.

We can now describe the limiting behaviour of this model when the population size K tends to infinity. The following convergence result can be found for example in (Fournier and Méléard, 2004, Theorem 5.3) and (Champagnat et al., 2008, Theorem 4.2). Let $D([0,T], M_F(\Omega))$ denote the Skorokhod space of càdlàg functions taking values in $M_F(\Omega)$.

Proposition 2.1. *Assume that ν_0^K converges weakly to a deterministic $f_0 \in M_F(\Omega)$ as $K \to +\infty$ and that $c_K = c/K$ for some $c > 0$. Also assume that $\sup_K \mathbb{E}[\langle \nu_0^K, 1\rangle^3] < +\infty$. Then, for any fixed $T > 0$, as $K \to +\infty$,*

$$\left(\nu_t^K, t \in [0,T]\right) \longrightarrow (f_t, t \in [0,T]),$$

in distribution in $D([0,T], M_F(\Omega))$, where $(f_t, t \in [0,T])$ is such that, for any bounded and measurable $\phi : \Omega \to \mathbb{R}$,

$$\langle f_t, \phi\rangle = \langle f_0, \phi\rangle + \int_0^t \langle f_s, b\,\mathcal{M}^*\phi + (b - d - c\langle f_s, 1\rangle)\phi\rangle ds, \tag{5}$$

where

$$\mathcal{M}^*\phi(\mathbf{x}) = U \int_\Omega (\phi(\mathbf{y}) - \phi(\mathbf{x}))\rho(\mathbf{x}, \mathbf{y})d\mathbf{y}.$$

We note that, if f_0 admits a density with respect to the Lebesgue measure, f_t admits a density (denoted by $f(t,\cdot)$) for all $t \geq 0$. In this case, setting $m(\mathbf{x}) = b(\mathbf{x}) - d(\mathbf{x})$ and

$$q(t,\mathbf{x}) = \frac{f(t,\mathbf{x})}{\langle f_t, 1\rangle}, \qquad\qquad \overline{m}(t) = \int_\Omega q(t,\mathbf{x})m(\mathbf{x})d\mathbf{x},$$

we see that the phenotype distribution q solves the following

$$\partial_t q(t,\mathbf{x}) = \mathcal{M}(b\,q)(t,\mathbf{x}) + (m(\mathbf{x}) - \overline{m}(t))\,q(t,\mathbf{x}), \tag{6}$$

where

$$\mathcal{M}(f)(\mathbf{x}) = U\left(\int_\Omega f(\mathbf{y})\rho(\mathbf{y},\mathbf{x})d\mathbf{y} - f(\mathbf{x})\right).$$

In the model (6), due to the coefficient b in $\mathcal{M}(b\,q)$, mutant individuals appear at a higher rate in regions where b is higher. As we shall expose below, this has wide ranging consequences on the qualitative behaviour of the phenotype distribution, which the standard model ($\mathcal{Q}_{stand}$) does not capture. As it turns out, however, very little can be proved about the integro-differential equation (6). If the variance

of mutation effects is sufficently small, however, we can instead study a diffusive approximation of equation (6). Assume that the effects of mutation on phenotype can be described by a mutation kernel J, such that $\rho(\mathbf{y}, \mathbf{x}) = J(\mathbf{x} - \mathbf{y})$. Namely,

$$\mathcal{M}(b\, f)(\mathbf{x}) = U \left(\int_{\mathbf{x}-\Omega} (b\, f)(\mathbf{x} - \mathbf{y})\, J(\mathbf{y})\, d\mathbf{y} - (b\, f)(\mathbf{x}) \right).$$

Formally, we write a Taylor expansion of $(b\, q)(t, \mathbf{x} - \mathbf{y})$ at $\mathbf{x} \in \Omega$:

$$(b\, q)(t, \mathbf{x} - \mathbf{y}) = \sum_{k_1, \ldots, k_n = 0}^{\infty} (-1)^{k_1 + \cdots + k_n} \frac{y_1^{k_1} \cdots y_n^{k_n}}{k_1! \cdots k_n!} \frac{\partial^{k_1 + \cdots + k_n} (b\, q)}{\partial x_1^{k_1} \cdots \partial x_n^{k_n}}(t, \mathbf{x}).$$

We define the central moments of the distribution:

$$\omega_{k_1, \ldots, k_n} = \int_{\mathbb{R}^n} y_1^{k_1} \cdots y_n^{k_n}\, J(y_1, \ldots, y_n)\, dy_1 \ldots dy_n.$$

We make a symmetry assumption on the kernel J which implies that $\omega_{k_1, \ldots, k_n} = 0$ if at least one of the k_i's is odd. Moreover, we assume the same variance λ at each trait: $\omega_{0, \ldots, 0, k_i = 2, 0, \ldots, 0} = \lambda$, and that the moments of order $k_1 + \cdots + k_n \geq 4$ are of order $O(\lambda^2)$. These assumptions are satisfied with the classic isotropic Gaussian distribution of mutation effects on phenotype. For $\lambda \ll 1$, we obtain:

$$U \left(\int_{\mathbf{x}-\Omega} (b\, q)(t, \mathbf{x} - \mathbf{y})\, J(\mathbf{y})\, d\mathbf{y} - (b\, q)(t, \mathbf{x}) \right) \approx \frac{\lambda\, U}{2} \Delta (b\, q)(t, \mathbf{x}) + O(\lambda^2).$$

Thus, when the variance λ of the (symmetric) mutation kernel J is small, we expect that the solution to (6) behaves as the solution to $(\mathcal{Q}_b)$:

$$\partial_t q(t, \mathbf{x}) = D \Delta (b\, q)(t, \mathbf{x}) + (m(\mathbf{x}) - \overline{m}(t))\, q(t, \mathbf{x}),$$

where $D = \lambda\, U / 2$.

Remark 2.2. *We recall that the assumption here is that mutations occur during reproduction. If we had assumed that mutations take place at a constant rate during each individual's lifetime, instead of linking them to reproduction events, we would have obtained a different equation in* (5)*, where the birth rate b does not appear in front of the operator $\mathcal{M}^*$. As a result, in this case, we would have obtained the standard model $(\mathcal{Q}_{stand})$ instead of $(\mathcal{Q}_b)$.*

2.2 Derivations of the standard model $(\mathcal{Q}_{stand})$

The standard model $(\mathcal{Q}_{stand})$ is classically derived by letting the variance of the mutation kernel tend to zero and by rescaling time to compensate for the fact that mutations have very small effects. In order to obtain the convergence of the process $(\nu_t^K, t \geq 0)$ in this regime, one also has to assume that the intensity of selection (measured by $b - d$) is of the same order of magnitude as the variance of the mutation kernel. This corresponds to a weak selection regime, where b and d are almost constant on Ω.

Large population limit of the continuous time model in rescaled timescale We consider the same stochastic individual-based model as above, but we allow b, d and ρ to depend on K. We thus let $b_K(\mathbf{x})$ denote the birth rate of individuals with phenotype $\mathbf{x}$, d_K their death rate, and ρ_K will be the mutation kernel. We then make the following assumption.

Assumption (SE) (frequent mutations with small effects) Let $\varepsilon_K = K^{-\eta}$ for some $0 < \eta < 1$ and assume that ρ_K is a symmetric kernel such that, for all $1 \leq i \leq d$,

$$\int_\Omega (\mathbf{y}_i - \mathbf{x}_i)^2 \rho_K(\mathbf{x}, \mathbf{y}) d\mathbf{y} = \varepsilon_K \, \lambda + o(\varepsilon_K), \qquad \int_\Omega (\mathbf{y}_i - \mathbf{x}_i)^{2+\delta} \rho(\mathbf{x}, \mathbf{y}) d\mathbf{y} = o(\varepsilon_K),$$

for all in $\mathbf{x} \in \Omega$ for some $\lambda > 0$ and $\delta \in]0, 2]$.

This assumption is what justifies the so-called *diffusive approximation*, where the effect of mutations on the phenotype density is modelled by a Laplacian in continuous time.

Assumption (WS) (weak selection) Assume that

$$b_K(\mathbf{x}) = 1 + \varepsilon_K \, b(\mathbf{x}), \qquad d_K(\mathbf{x}) = 1 + \varepsilon_K \, d(\mathbf{x}), \qquad c_K = \frac{\varepsilon_K}{K} c,$$

for some bounded functions $b : \Omega \to \mathbb{R}$, $d : \Omega \to \mathbb{R}$ and some positive c.

The following result then corresponds to Theorem 4.3 in Champagnat et al. (2008). Recall that Ω is assumed to be a bounded open set, and further assume that it has a smooth boundary $\partial\Omega$. Let $C_0^2(\Omega)$ be the set of twice continuously differentiable functions $\phi : \bar{\Omega} \to \mathbb{R}$ such that

$$\nabla\phi(\mathbf{x}) \cdot \vec{\nu}(\mathbf{x}) = 0, \quad \forall \mathbf{x} \in \partial\Omega,$$

where $\vec{\nu}(\mathbf{x})$ is the outward unit normal to $\partial\Omega$.

Proposition 2.3. *Let Assumptions (SE) and (WS) be satisfied. Also assume that ν_0^K converges weakly to a deterministic $f_0 \in M_F(\Omega)$ as $K \to \infty$ and that*

$$\sup_K \mathbb{E}[\langle \nu_0^K, 1 \rangle^3] < +\infty.$$

Then, for any fixed $T > 0$, as $K \to \infty$,

$$\left(\nu^K_{t/\varepsilon_K}, t \in [0, T]\right) \longrightarrow \left(f_t, t \in [0, T]\right),$$

in distribution in $D([0, T], M_K(\Omega))$, where f_t is such that, for any $\phi \in C_0^2(\Omega)$,

$$\langle f_t, \phi \rangle = \langle f_0, \phi \rangle + \int_0^t \langle f_s, D\Delta\phi + (b - d - c\langle f_s, 1 \rangle)\phi \rangle ds,$$

with $D = \frac{U\lambda}{2}$.

For all $t > 0$, f_t admits a density $f(t, \cdot)$ and the phenotype distribution $q(t, \mathbf{x}) = f(t, \mathbf{x})/\langle f_t, 1\rangle$ solves $(\mathcal{Q}_{stand})$:

$$\partial_t q(t, \mathbf{x}) = D\Delta q(t, \mathbf{x}) + (m(\mathbf{x}) - \overline{m}(t))\, q(t, \mathbf{x}).$$

As we can see, we have lost the factor b in the mutation term by taking this limit. This comes from Assumption (WS) which states that $b_K(\mathbf{x}) = 1 + O(\varepsilon_K)$ (the fact that the leading order is 1 simply means that we are measuring time in generations, but it is important that $b_K - d_K = O(\varepsilon_K)$). As a result this equation does not distinguish the birth optimum from the survival optimum.

Large population limit of an individual-based model with non-overlapping generations We now consider a model where generations are non-overlapping, meaning that, between two generations (denoted t and $t + 1$), all the individuals alive at time t first produce a random number of offspring and then die. The population at time $t + 1$ is thus only comprised of the offspring of the individuals alive at time t.

Let $w_K : \Omega \to \mathbb{R}_+$ be a measurable and bounded function and assume that an individual with phenotype $\mathbf{x} \in \Omega$ produces a random number of offspring which follows a Poisson distribution with parameter $w_K(\mathbf{x})$. In order to include competition, we assume that each of these offspring survives with probability $e^{-c_K N_t}$ for some $c_K > 0$, where N_t is the number of individuals in generation t. Each newborn individual either carries the phenotype of its parent, with probability $1 - U$, or, with probability U, carries a phenotype $\mathbf{y}$ chosen at random from some distribution $\rho_K(\mathbf{x}, \mathbf{y})d\mathbf{y}$, where $\mathbf{x}$ is the phenotype of its parent.

We now make several assumptions in order to obtain an approximation of the process as the population size tends to infinity. For the limiting process to be continuous in time, we need to assume that the change in the composition of the population from one generation to the next is very small, and then rescale time by the appropriate factor. This ties our hands somewhat, and we need to assume that w_K is very close to one everywhere in Ω. More precisely, we make the following assumption.

Assumption (WS') Let $\varepsilon_K = K^{-\eta}$ for some $0 < \eta < 1$ and assume that

$$w_K(\mathbf{x}) = \exp\left(\varepsilon_K\, m(\mathbf{x})\right), \qquad\qquad c_K = \frac{\varepsilon_K}{K} c,$$

for some bounded function $m : \Omega \to \mathbb{R}$ and some positive c.

Here, $w_K(\mathbf{x})$ corresponds to the Darwinian fitness (the average number of offspring of an individual with phenotype $\mathbf{x}$), while $m(\mathbf{x})$ corresponds to the Malthusian fitness (*i.e.* the growth rate of the population of individuals with phenotype $\mathbf{x}$). We further assume that ρ_K satisfies Assumption (SE) above.

The large population limit of this process is then given by the following result, which is analogous to similar results in continuous time (for example in Champagnat et al. (2008)). For the sake of completeness, we provide its proof in Appendix A.

Proposition 2.4. *Assume that Assumption (WS') is satisfied, along with (SE). Also assume that, ν_0^K converges weakly to a deterministic $f_0 \in M_F(\Omega)$. Then, for any fixed $T > 0$, as $K \to +\infty$,*

$$\left(\nu^K_{[t/\varepsilon_K]}, t \in [0,T]\right) \longrightarrow (f_t, t \in [0,T]),$$

in distribution in $D([0,T], M_F(\Omega))$, where $(f_t, t \in [0,T])$ is such that, for any $\phi \in C_0^2(\Omega)$,

$$\langle f_t, \phi \rangle = \langle f_0, \phi \rangle + \int_0^t \langle f_s, \mathcal{M}^* \phi + (m - c \langle f_s, 1 \rangle)\phi \rangle \, ds, \tag{7}$$

where

$$\mathcal{M}^* \phi(\mathbf{x}) = \frac{\lambda U}{2} \Delta \phi(\mathbf{x}).$$

We note that, for all $t > 0$, f_t admits a density with respect to the Lebesgue measure. Let $f(t, \cdot) \in L^1(\Omega)$ denote the corresponding density. Then $f(t, \cdot)$ solves the equation

$$\partial_t f(t, \mathbf{x}) = (\mathcal{M} f)(t, \mathbf{x}) + \left(m(\mathbf{x}) - c \int_\Omega f(t, \mathbf{y}) d\mathbf{y} \right) f(t, \mathbf{x}).$$

We also note that if we set

$$q(t, \mathbf{x}) = \frac{f(t, \mathbf{x})}{N(t)}, \qquad \text{with } N(t) = \int_\Omega f(t, \mathbf{x}) d\mathbf{x},$$

then $q(t, \cdot)$ solves $(\mathcal{Q}_{stand})$.

Propositions 2.3 and 2.4 show how the standard model $(\mathcal{Q}_{stand})$ arises as a large population limit of individual-based models in the weak selection regime with small mutation effects. However, as Proposition 2.1 shows, the fact that the birth rate does not appear in the mutation term is a consequence of the weak selection assumption. In the next section, we will focus on a situation corresponding to a strong trade-off between birth and survival. In this case, the weak selection assumption is not satisfied. Thus, the new model $(\mathcal{Q}_b)$ should be more appropriate to study the dynamics of adaptation, at least when generations are overlapping.

In the model with non-overlapping generations, we expect that the model $(\mathcal{Q}_{stand})$ emerges even when the weak selection assumption is not satisfied. From an intuitive perspective, with this model, the expected number of mutants per generation is $U\,N(t)$. Thus, if $N(t)$ is close to the carrying capacity, the overall number of mutants should not depend on the phenotype distribution in the population. However, if one tries to take a large population limit of the discrete time model in the same regime as in Proposition 2.1 (keeping w and ρ fixed and letting the population size tend to infinity), then the phenotype distribution converges to the solution to a deterministic recurrence equation of the form

$$\langle q_{t+1}, \phi \rangle = \left\langle q_t, \tfrac{w}{\overline{w}(t)} \mathcal{M}^* \phi + \tfrac{w}{\overline{w}(t)} \phi \right\rangle,$$

where $\mathcal{M}^*$ is as in (5). We do not study this equation here, but it is interesting to note that the fitness has an effect on the mutations, albeit quite different from that in (5).

In the following section, we use $(\mathcal{Q}_b)$ to study the consequences of a birth-dependent mutation rate on the trade-off between birth and survival, and we compare our results to the standard approach of $(\mathcal{Q}_{stand})$ and to individual-based simulations.

3 Consequences of a birth-dependent mutation rate on the trade-off between birth and death

We focus here on the trajectories of adaptation and the large time dynamics given by the model $(\mathcal{Q}_b)$, with a special attention on the differences with the standard approach $(\mathcal{Q}_{stand})$ which neglects the dependency of mutation rate on birth rate.

In most related studies, the relationships between the phenotype $\mathbf{x}$ and the fitness $m(\mathbf{x})$ is described with the standard Fisher Geometrical Model (FGM) where $m(\mathbf{x}) = r_{max} - \|\mathbf{x}\|^2/2$. This phenotype to fitness landscape model is widely used, see e.g. Tenaillon (2014), Martin and Lenormand (2015). It has shown robust accuracy to predict distributions of pathogens Martin et al. (2007), Martin and Lenormand (2006), and to fit biological data (Perefarres et al., 2014; Schoustra et al., 2016). Here, however, in order to study the trade-off between birth and survival, we shall assume that the death rate d takes the form: $d(\mathbf{x}) = r - s(\mathbf{x})$ for some $r > 0$, such that

$$m(\mathbf{x}) = \underbrace{b(\mathbf{x})}_{\text{birth}} + \underbrace{s(\mathbf{x})}_{\text{survival}} - r, \tag{8}$$

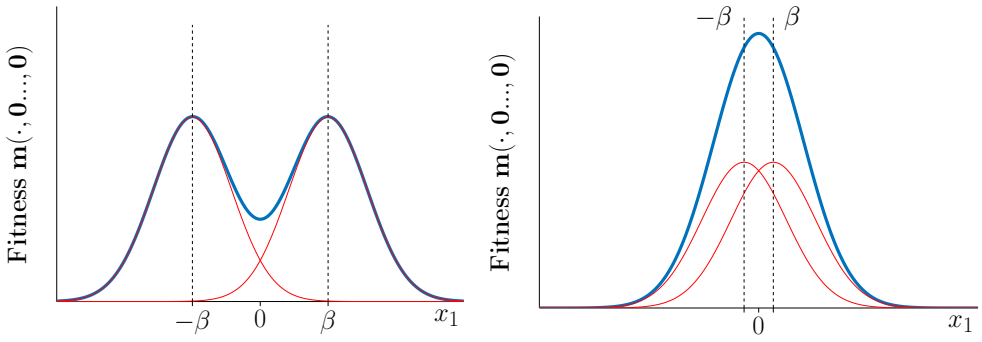
for some function $s : \Omega \to [0, r]$ such that b and s are symmetric about the axis $x_1 = 0$ and we assume that b has a global maximum that is not on this axis. As a result s also has a global maximum, which is the symmetric of that of b. The positive constant r has no impact on the dynamics of the phenotype distribution $q(t, \mathbf{x})$ in model $(\mathcal{Q}_b)$, as it vanishes in the term $m(\mathbf{x}) - \overline{m}(t)$. To keep the model relevant, the constant r must therefore be chosen such that $d(\mathbf{x}) > 0$ for all $x \in \Omega$.

We assume that $b(\mathbf{x})$ reaches its maximum at $\mathcal{O}_b \in \Omega$ and $s(\mathbf{x})$ reaches its maximum at $\mathcal{O}_s \in \Omega$. If one of the optima leads to a higher fitness value, we expect that the corresponding strategy (high birth vs. high survival) will be selected. To avoid such 'trivial' effects, and to analyse the result of the trade-off between birth and survival independently of any fitness bias towards one or the other, we make the following assumptions. The domain Ω is symmetric about the hyperplane $\{x_1 = 0\}$. Next, b and s are positive, continuous over $\overline{\Omega}$ and symmetric in the following sense:

$$b(\mathbf{x}) = s(\iota(\mathbf{x})), \text{ with } \iota(\mathbf{x}) = \iota(x_1, x_2, ..., x_n) = (-x_1, x_2, ..., x_n). \tag{9}$$

The optima are then also symmetric about the axis $x_1 = 0$:

$$\mathcal{O}_b = (\beta, 0, \ldots, 0) \text{ and } \mathcal{O}_s = (-\beta, 0, \ldots, 0),$$



(a) Two optima (b) A single optimum

Figure 1: **Schematic representation of the fitness function $m(\mathbf{x})$ along the phenotype dimension x_1.** In both cases the black dashed lines correspond to the survival optimum $\mathcal{O}_s = (-\beta, \dots, 0)$ (on the left) and the birth optimum $\mathcal{O}_b = (\beta, 0, \dots, 0)$ (on the right). In Figure 1(a) those optima are almost superposed with those of m, which is very different from Figure 1(b). In red we pictured the functions b and $s - r$.

for some $\beta > 0$, so that the birth optimum is situated to the right of $x_1 = 0$ and the survival optimum is situated to the left of $x_1 = 0$. A schematic representation of the birth and survival terms and corresponding fitness function, along the first dimension x_1 is given in Figure 1.

Finally, we assume that the birth rate is larger than the survival rate in the whole half-space around $\mathcal{O}_b$ ($\Omega \cap \{x_1 > 0\}$), and conversely, from (9), the survival rate is higher in the other half-space. In other terms:

$$\begin{aligned} b(x_1, \dots, x_n) &> s(x_1, \dots, x_n), \text{ for all } \mathbf{x} \in \Omega \cap \{x_1 > 0\}, \\ s(x_1, \dots, x_n) &> b(x_1, \dots, x_n), \text{ for all } \mathbf{x} \in \Omega \cap \{x_1 < 0\}. \end{aligned} \tag{10}$$

From the symmetry assumption (9), we know that the hyperplane $\{x_1 = 0\}$ is a critical point for $b+s$ in the direction x_1, that is $\partial_{x_1} b(0, x_2, \dots, x_n) = -\partial_{x_1} s(0, x_2, \dots, x_n)$.

For the well-posedness of the model $(\mathcal{Q}_b)$, and as the integral of $q(t, \mathbf{x})$ over Ω must remain equal to 1 (recall that $q(t, \cdot)$ is a probability distribution), we assume reflective (Neumann) boundary conditions:

$$b(\mathbf{x})(\nabla q(t, \mathbf{x}) \cdot \vec{\nu}(\mathbf{x})) + (\nabla b(\mathbf{x}) \cdot \vec{\nu}(\mathbf{x}))\, q(t, \mathbf{x}) = 0, \mathbf{x} \in \partial\Omega,$$

with $\vec{\nu}(\mathbf{x})$ the outward unit normal to $\partial\Omega$, the boundary of Ω. We also assume a compactly supported initial condition $q_0(\mathbf{x}) = q(0, \mathbf{x})$, with integral 1 over Ω.

3.1 Trajectories of adaptation

The methods developed in Hamel et al. (2020) provide analytic formulas describing the full dynamics of adaptation, and in particular the dynamics of the mean fitness $\overline{m}(t)$, for models of the form $(\mathcal{Q}_{stand})$, i.e., with a constant mutation rate. As far as model $(\mathcal{Q}_b)$ is concerned, due to the birth-dependent term in the mutation operator $D\,\Delta(b\,q)$, the derivation of comparable explicit formulas seems out of reach.

To circumvent this issue, we use numerical simulations to exhibit some qualitative properties of the adaptation dynamics, that we demonstrate next. We focus on the dynamics of the mean phenotype $\overline{\mathbf{x}}(t)$ and of the mean fitness $\overline{m}(t)$, to be compared to the 'standard' case, where the mutation rate does not depend on the phenotype, and to individual-based stochastic simulations with the assumptions of Section 2. In the PDE setting, the mean phenotype $\overline{\mathbf{x}}(t) \in \Omega$ and mean fitness $\overline{m}(t) \in \mathbb{R}_-$ are defined by:

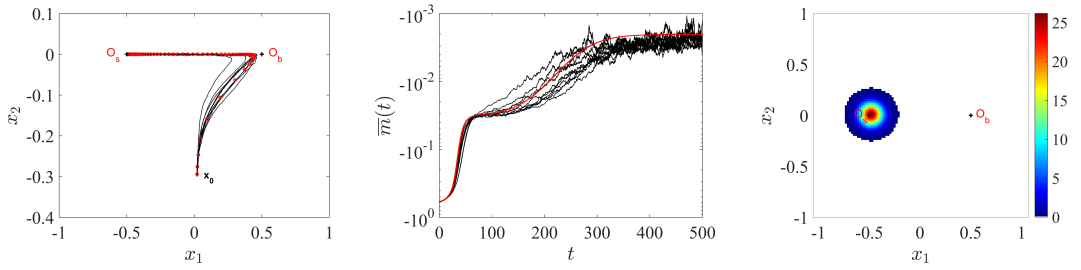
$$\overline{\mathbf{x}}(t) := \int_\Omega \mathbf{x}\, q(t,\mathbf{x})\, d\mathbf{x}, \quad \overline{m}(t) := \int_\Omega m(\mathbf{x})\, q(t,\mathbf{x})\, d\mathbf{x}.$$

Numerical simulations. Our numerical computations are carried out in dimension $n = 2$, starting with an initial phenotype concentrated at some point $\mathbf{x}_0$ in Ω. The trajectories given by the PDE $(\mathcal{Q}_b)$ with a birth-dependent mutation rate are depicted in Figure 2(a), together with 10 replicate simulations of a stochastic individual-based model with overlapping generations (see Section 2). The mean phenotype is first attracted by the birth optimum $\mathcal{O}_b$. In a second time, it converges towards $\mathcal{O}_s$. This pattern leads to a trajectory of mean fitness which exhibits a small 'plateau': the mean fitness seems to stabilize at some value smaller than the ultimate value $\overline{m}_\infty$ during some period of time, before growing again at larger times. The trajectories given by individual-based simulations exhibit the same behaviour.

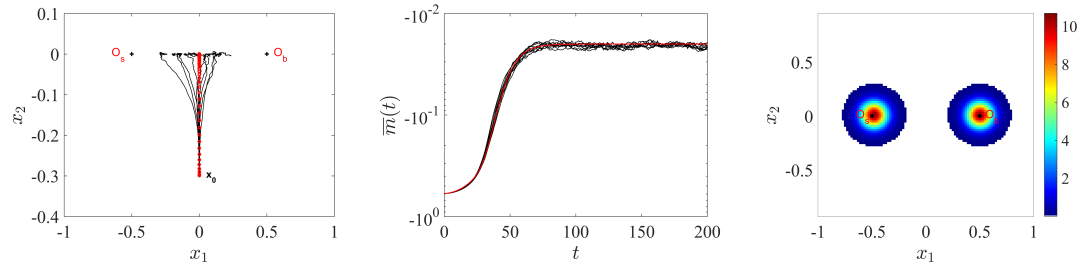
On the other hand, simulation of the standard equation $(\mathcal{Q}_{stand})$ without dependence of the mutation rate with respect to the phenotype (with Neumann boundary conditions), leads to standard saturating trajectories of adaptation, see Figure 2(b) (already observed in Martin and Roques, 2016, with this model).

If the initial population density q_0 is symmetric about the hyperplane $\{x_1 = 0\}$, then so does $q(t,\mathbf{x})$ at all positive times in this case. This is a consequence of the uniqueness of the solution of $(\mathcal{Q}_{stand})$ (which follows from Hamel et al., 2020): we observe that if $q(t,\mathbf{x})$ is a solution of $(\mathcal{Q}_{stand})$ with initial condition q_0, then so does $q(t,\iota(\mathbf{x}))$. By uniqueness, $q(t,\mathbf{x}) = q(t,\iota(\mathbf{x}))$ for all times. This in turns implies that the mean phenotype $\overline{\mathbf{x}}(t)$ remains on the hyperplane $\{x_1 = 0\}$, i.e., at the same distance of the two optima $\mathcal{O}_b$ and $\mathcal{O}_s$. Besides, even if q_0 was not symmetric about $\{x_1 = 0\}$, i.e., if the initial phenotype distribution was biased towards one of the two optima, the trajectory of $\overline{\mathbf{x}}(t)$ would ultimately still converge to the axis $\{x_1 = 0\}$. Again, this is a consequence of the uniqueness of the positive stationary state of $(\mathcal{Q}_{stand})$ (with integral 1), which is itself a consequence of the uniqueness of the principal eigenfunction (up to multiplication) of the operator $\phi \mapsto D\,\Delta\,\phi + m(\mathbf{x})\,\phi$ (this uniqueness result is classical, see e.g. Alfaro and Veruete, 2018). This time, the trajectories given by the model $(\mathcal{Q}_b)$ are in good agreement with those given by an individual-based model with non-overlapping generations (see section 2).

Initial bias towards the birth optimum, a multidimensional feature. One of the qualitative properties observed in the simulations (Figure 2(a)) is an initial tendency of the trajectory of the mean phenotype $\overline{\mathbf{x}}(t)$ to go towards the birth optimum $\mathcal{O}_b$. We show here that this is a general feature, conditioned by the shape of selection along other dimensions. For simplicity, we denote by $\overline{x}_1(t)$ the mean value of the first trait, that is, the first coordinate of $\overline{\mathbf{x}}(t)$. We consider initial



(a) Model ($\mathcal{Q}_b$)



(b) Model ($\mathcal{Q}_{stand}$)

Figure 2: **Trajectory of adaptation and stationary distribution: model ($\mathcal{Q}_b$) with birth-dependent mutation rate vs standard model ($\mathcal{Q}_{stand}$).** The red circles in the left panels correspond to the position of the mean phenotype $\overline{\mathbf{x}}(t)$ with initial condition concentrated at $\mathbf{x}_0 = (0, -0.3)$, at successive times $t = 0, 1, \ldots, 100$. The central panels describes the dynamics of the mean fitness $\overline{m}(t)$, in a logarithmic scale. The black curves in these panels correspond to 10 replicate simulations of the individual based models of Section 2, with either overlapping generations (upper panels) or non-overlapping generations (lower panels). The right panels correspond to the distribution at $t = 100$. We assumed here that the dimension is $n = 2$, $\mathcal{O}_b = (1/2, 0)$, $\mathcal{O}_s = (-1/2, 0)$ (i.e., $\beta = 1/2$), $b(\mathbf{x}) = b(x_1, x_2) = \exp\left[-(x_1 - \beta)^2/(2\,\sigma_{x_1}^2) - x_2^2/(2\,\sigma_{x_2}^2)\right]$, $s(\mathbf{x}) = \exp\left[-(x_1 + \beta)^2/(2\,\sigma_{x_1}^2) - x_2^2/(2\,\sigma_{x_2}^2)\right]$, $\sigma_{x_1}^2 = \sigma_{x_2}^2 = 1/10$ and $D = 1/4000$.

conditions q_0 that are symmetric about the hyperplane $\{x_1 = 0\}$, and that are localized around a phenotype $\mathbf{x}_0 \in \{x_1 = 0\}$. By localized, we mean that q_0 vanishes outside some compact set that contains $\mathbf{x}_0$. We denote by K_0 the support of q_0, and define $K_0^+ := K_0 \cap \{x_1 > 0\}$ the 'right part of K_0'. We prove the following result (the proof is detailed in Appendix A).

Proposition 3.1. *Let q be the solution of $(\mathcal{Q}_b)$, with an initial condition q_0 which satisfies the above assumptions. Then the following holds.*

- *If $\Delta(x_1 m) \geq 0$ (and $\not\equiv 0$) on K_0^+, then the solution is initially biased towards the birth optimum, that is*

$$\overline{x}_1'(t=0) = 0 \quad \text{and} \quad \overline{x}_1''(t=0) > 0. \tag{11}$$

- *If $\Delta(x_1 m) \leq 0$ (and $\not\equiv 0$) on K_0^+, then the solution is initially biased towards the survival optimum, that is*

$$\overline{x}_1'(t=0) = 0 \quad \text{and} \quad \overline{x}_1''(t=0) < 0. \tag{12}$$

A surprising feature of this proposition is the discussion around the sign of the quantity $\Delta(x_1 m)$. It shows that the local convexity (or concavity) of m around the initial phenotype is important. It stems from the overall shape and symmetry of m. We first illustrate this in dimension 1. In that case, the Laplace operator simply becomes

$$\Delta(xm) = xm''(x) + 2m'(x) =: g(x),$$

and the initial distribution q_0 is located around 0. By the symmetry assumption (9), we know that $m'(0) = 0$ and thus $g(0) = 0$. Therefore, in this one dimensional case, the discussion of Proposition 3.1 about the sign of $\Delta(xm)$ is linked to the sign of $g'(0) = 3m''(0)$, that is the local convexity of m around 0. Equivalently, it is also dictated by a discussion about the shape of m: if m presents a profile with two symmetric optima (camel shape, Figure 1(a)) or a single one located at 0 (dromedary shape, Figure 1(b)), the outcome of the initial bias is different. If m has two local maxima, one around $\mathcal{O}_b$, and the other around $\mathcal{O}_d$, then necessarily m admits a local minimum around 0, see the camel shape in Figure 1. Therefore, as a consequence of Proposition 3.1, there is an initial bias towards the birth optimum. If now m admits a single optimum, as a sum of birth and survival, the critical point 0 is also a global maximum of m. From Proposition 3.1, it means that, reversing it, there is an initial bias towards survival. We refer to Figure 3 for the numerical representations of the different shapes of m, the sign of g and the implications relatively to the trajectories of fitness.

This can be explained as follows. In the case where m has two optima, the population is initially around a minimum of fitness. By symmetry of m there is no fitness benefice of choosing either optimum. However, individuals on the right have a higher mutation rate, which generates variance to fuel and speed-up adaptation, which explains the initial bias towards right. On the other hand, if 0 is the unique optimum of the fitness function, the initial population is already at the optimum. Thus, generating more variance does not speed-up adaptation, but on the contrary generates more mutation load, which explains the initial bias towards left.

In a multidimensional setting, we can follow the same explanations, even if another phenomenon can arise. The reason lies in the following formula:

$$\Delta(x_1 m) = \partial_{x_1 x_1}(x_1 m) + x_1 \sum_{j \geq 2} \partial_{x_j x_j} m. \tag{13}$$

Suppose that, as in Figure 1(a), there is a local minimum around $\mathbf{x}_0$, in the first dimension. Then, the first term of (13) is positive in a neighborhood of $\mathbf{x}_0$ as soon as $x_1 > 0$, as we explained previously. In dimension $n \geq 2$, if the sum of the second derivatives with respect to the other directions is negative, the overall sign of $\Delta(x_1 m)$ may be changed. Such a situation can arise in dimension 2 if $\mathbf{x}_0$ is a saddle point. This phenomenon can be observed on Figure 3. In both Figure 3(a) and Figure 3(b), the fitness function m is camel like along the first dimension, as pictured in Figure 1(a). However, as a consequence of the second dimension, we observe, or not, an initial bias towards the birth optimum. Similarly to the one-dimensional case, one can observe that if the mutational load is too important, here on the second dimension, we do not observe this initial bias. This of course cannot be if $\mathbf{x}_0$ is a local minimum of m in $\mathbb{R}^d$.

Large time behaviour. We now analyze whether the convergence towards the survival optimum at large times observed in Figure 2(a) is a generic behavior. In that respect, we focus on the stationary distribution q_∞ associated with the model $(\mathcal{Q}_b)$. It satisfies equation

$$D\,\Delta(b\,q_\infty)(\mathbf{x}) + m(\mathbf{x})\,q_\infty(\mathbf{x}) = \overline{m}_\infty\,q_\infty(\mathbf{x}), \quad \mathbf{x} \in \Omega, \tag{14}$$

for some $\overline{m}_\infty \in \mathbb{R}$. Setting

$$v(\mathbf{x}) := b(\mathbf{x})\,q_\infty(\mathbf{x}),$$

this reduces to a more standard eigenvalue problem, namely

$$D\,\Delta v(\mathbf{x}) + \frac{m(\mathbf{x})}{b(\mathbf{x})} v(\mathbf{x}) = \overline{m}_\infty \frac{1}{b(\mathbf{x})}\,v(\mathbf{x}), \quad \mathbf{x} \in \Omega, \tag{15}$$

supplemented with Neumann boundaries conditions:

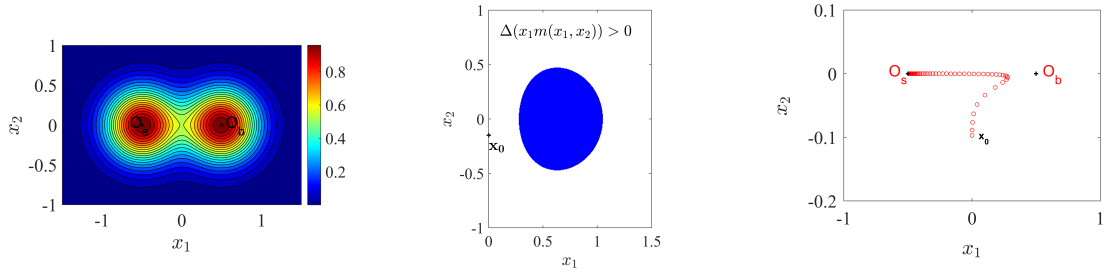
$$\nabla v(\mathbf{x}) \cdot \vec{\nu}(\mathbf{x}) = 0, \quad x \in \partial\Omega. \tag{16}$$

As the factor $1/b(\mathbf{x})$ multiplying $\overline{m}_\infty$ is strictly positive, we can indeed apply the standard spectral theory of Courant and Hilbert (2008) (see also Cantrell and Cosner, 2003). Precisely, there is a unique couple $(v(\mathbf{x}), \overline{m}_\infty)$ satisfying (15)—(16) (with the normalization condition $\int_\Omega v(\mathbf{x})d\mathbf{x} = 1$) such that $v(\mathbf{x}) > 0$ in Ω. The 'principal eigenvalue' $\overline{m}_\infty$ is provided by the variational formula

$$\overline{m}_\infty = \max_{\psi \in W^{1,2}(\Omega)} Q[\psi], \tag{17}$$

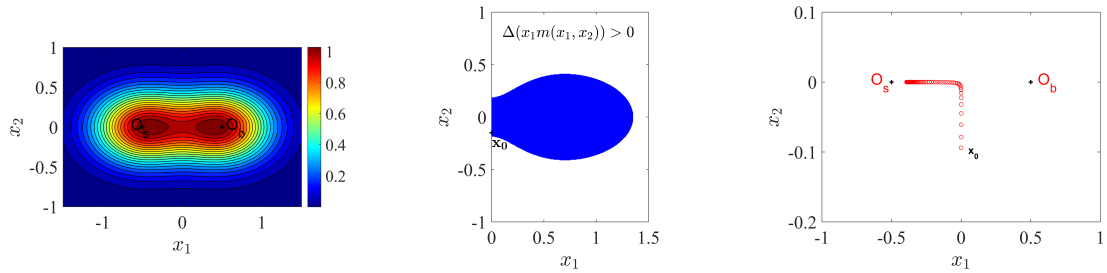
where

$$Q[\psi] := \frac{-D\,\int_\Omega \|\nabla(\psi\,\sqrt{b})\|^2(\mathbf{x})d\mathbf{x} + \int_\Omega m(\mathbf{x})\psi^2(\mathbf{x})d\mathbf{x}}{\int_\Omega \psi^2(\mathbf{x})d\mathbf{x}}.$$



(a) Shape 1: $\Delta(x_1 m(x_1, x_2)) > 0$ on the axis $\{x_1 = 0\}$

fig



(b) Shape 2: $\Delta(x_1 m(x_1, x_2))$ changes sign on the axis $\{x_1 = 0\}$

Figure 3: **Trajectory of adaptation with different shapes of the fitness function.** The left panels depict the fitness function. The central panels describe the sign of $\Delta(x_1 m(x_1, x_2))$ in the region $\{x_1 > 0\}$: this quantity is negative in the blue region and positive otherwise. The right panels depict the corresponding trajectories of the mean phenotype $\overline{\mathbf{x}}(t)$ obtained with the model $(\mathcal{Q}_b)$, at successive times $t = 0, 1, \ldots, 100$. In both cases, the initial condition is concentrated at $\mathbf{x}_0 = (0, -0.1)$, leading to a positive sign of $\Delta(x_1 m(x_1, x_2))$ at $(x_1, x_2) = \mathbf{x}_0$ in the upper panels and a negative sign in the lower panels. The parameter values are the same as in Fig. 2, except for the fitness function of the lower panel, where $\sigma^2_{x_1} = 1/18$ and $\sigma^2_{x_2} = 1/10$.

An immediate consequence of formula (17) is that $\overline{m}_\infty$ is a decreasing function of the mutational parameter D. This means that, as expected, the mutation load increases when the mutational parameter is increased.

We expect the stationary state to 'lean mainly on the left', meaning that the survival optimum is selected at large times, but deriving rigorously the precise shape of q_∞ seems highly involved. However, formula (17) gives us some intuition. First, multiplying (15) by v and integrating, we observe that $Q[v/\sqrt{b}] = Q[\sqrt{b}\,q_\infty] = \overline{m}_\infty$. Thus, formula (17) shows that the shape of q_∞ should be such that $\psi = \sqrt{b}\,q_\infty$ maximizes the Rayleigh quotient Q.

We thus consider each term of Q separately. From Hardy-Littlewood-Pólya rearrangement inequality, the term $\int_\Omega m(\mathbf{x})\psi^2(\mathbf{x})d\mathbf{x}$ is larger when ψ is arranged like m, i.e., ψ takes its largest values where m is large and its smallest values where m is small. Thus, this term tends to promote shapes of ψ which look like m. The other term $-D\int_\Omega \|\nabla(\psi\sqrt{b})\|^2(\mathbf{x})d\mathbf{x}$ tends to promote functions ψ which are proportional to $1/\sqrt{b}$. Finally, the stationary distribution q_∞ should therefore realize a compromise between $1/b$ and $m/\sqrt{b}$. As both functions take their larger values when b is small, we expect q_∞ to be larger close to the survival optimum $\mathcal{O}_s$.

More rigorously, define $\tilde{q}(\mathbf{x}) = q(-x_1, x_2, \dots, x_n) = q(\iota(\mathbf{x}))$. As $\sqrt{b}\,q_\infty$ realises a maximum of Q, we have

$$Q[\sqrt{b}\,q_\infty] \geq Q[\sqrt{s}\,\tilde{q}_\infty].$$

Recalling $s(\mathbf{x}) = b(\iota(\mathbf{x}))$ and using the symmetry of m, this implies that

$$\int_\Omega \|\nabla(\tilde{q}_\infty\sqrt{b\,s})\|^2 = \int_\Omega \|\nabla(q_\infty\sqrt{b\,s})\|^2 \geq \int_\Omega \|\nabla(q_\infty\,b)\|^2. \tag{18}$$

Now, we illustrate that *moralement* this gradient inequality means that the stationary distribution tends to be closer to $\mathcal{O}_s$ than to $\mathcal{O}_b$. In dimension $n = 1$, assume that $b(x) = \exp(-(x-\beta)^2)$ and $s(x) = \exp(-(x+\beta)^2)$. Assume that the domain is large enough so that the integrals over Ω can be accurately approached by integrals over $(-\infty, +\infty)$. Among all functions of the form $h_\gamma(x) = \exp(-(x-\gamma)^2)$, a straightforward computation reveals that

$$\left(\int_{-\infty}^{+\infty} [\partial_x(h_\gamma\sqrt{b\,s})]^2(x)\,dx \geq \int_{-\infty}^{+\infty} [\partial_x(h_\gamma\,b)]^2(x)\,dx\right) \;\Leftrightarrow\; \gamma \leq -\beta/2,$$

which means that the inequality is satisfied by functions h whose maximum is reached at a value $x = \gamma$ closer to $\mathcal{O}_s = -\beta$ than to $\mathcal{O}_b = \beta$.

Large mutation effects. This advantage of adaptation towards the survival optimum becomes more obvious when the mutation effects are large. We observed above that $\overline{m}_\infty^D$ (seen here as a function of D) is decreasing. Moreover, from (17), we have, for all $D > 0$,

$$\overline{m}_\infty^D \geq Q[1/\sqrt{b}] = \left(\int_\Omega m b^{-1}\right)\left(\int_\Omega b^{-1}\right)^{-1}$$

Thus $\overline{m}_\infty^D$ admits a limit $\overline{m}_\infty^\infty$ as $D \to \infty$. Moreover, the corresponding stationary states satisfy $\Delta(b\,q_\infty^D)(\mathbf{x}) + q_\infty^D(\mathbf{x})\,(m(\mathbf{x}) - \overline{m}_\infty^D)/D = 0$. Standard elliptic estimates

and Sobolev injections imply that, up to the extraction of some subsequence $D_k \to \infty$, the functions $q_\infty^{D_k}$ converge, as $k \to \infty$, in $C^2(\Omega)$ to a nonnegative solution (with mass 1) of $\Delta(b\, q_\infty^\infty)(\mathbf{x}) = 0$. As such a solution is unique and given by:

$$q_\infty^\infty(\mathbf{x}) = C/b(\mathbf{x}) \text{ with } C = \int_\Omega b^{-1},$$

the whole sequence q_∞^D converges to $C/b(\mathbf{x})$ as $D \to \infty$. Thus, in order to reduce the mutation load, the phenotype distribution tends to get inversely proportional to b in the large mutation regime.

An analytically tractable example. Consider the following form for the birth rate, in dimension $n = 1$:

$$b(x) = \begin{cases} 2 & \text{for } x \in (0, a), \\ 1 & \text{for } x \in (-a, 0), \\ -M & \text{for } x \notin (-a, a). \end{cases} \tag{19}$$

With the assumptions (8) and (19), we get:

$$m(x) = 3 - r, \text{ for } x \in (-a, a),$$

and $m(x) = -(r+2)\, M$ outside $(-a, a)$. Then, we consider the corresponding 1D eigenvalue problem (14) in an interval Ω containing $(-a, a)$. Assuming that the phenotypes are extremely deleterious outside $(-a, a)$ (i.e., $M \gg 1$), we make the approximation $q_\infty(\pm a) = 0$. In this case, we prove (see Appendix A) that

$$\frac{\int_{-a}^0 q_\infty(x)\, dx}{\int_0^a q_\infty(x)\, dx} > \frac{1}{2\sqrt{2-\sqrt{2}} - 2 + \sqrt{2}} > 1.$$

In other word, the stationary distribution has a larger mass to the left of 0 (where s is larger) than to the right (where b is larger).

4 Discussion

We found that a positive dependence between the birth rate and the mutation rate emerges naturally at the population scale, from elementary assumptions at the individual scale. Based on a large population limit of a stochastic individual-based model, we derived a reaction-diffusion framework ($\mathcal{Q}_b$) that describes the evolutionary trajectories and steady states in the presence of this dependence. We compared this approach with stochastic replicate simulations of finite size populations which showed a good agreement with the behaviour of the reaction-diffusion model. These simulations, and our analytical results on ($\mathcal{Q}_b$) demonstrate that taking this dependence into account, or conversely omitting it as in the standard model ($\mathcal{Q}_{stand}$), has far reaching consequences on the description of the evolutionary dynamics. In light of our results, we discuss below the causes and consequences of the positive dependence between the birth rate and the mutation rate.

Birth-dependent mutation rate: causes Even though the probability of mutation per birth event U does not depend on the phenotype of the parent, and therefore on its fitness nor its birth rate, a higher birth rate implies more mutations per unit of time at the population scale. This holds true when mutations mainly occur during reproduction, which is the case for bacteria and viruses. The mathematical derivation of the standard model ($\mathcal{Q}_{stand}$), that does not account for this dependence, generally relies on a weak selection assumption, which de facto implies a very mild variation of the birth rate with the phenotype. In such cases, the mutation rate can safely be assumed to be phenotype-independent at the population scale, even though it is positively correlated with the birth rate. Another exception corresponds to organisms with non-overlapping generations: the simulations in Fig. 2(b) indicate that even with a fitness function that strongly depends on the phenotype, the trajectories of adaptation are adequately described by the model ($\mathcal{Q}_{stand}$). Species with non-overlapping generations include annual plants (but some overlap may exist due to seedbanks), many insect species (e.g. processionary moths Roques, 2015, again some overlap may exist due to prolonged diapause) and fish species (such as some killifishes with annual life cycles Turko and Wright, 2015). In the other situations, our approach reveals that the model ($\mathcal{Q}_b$) will be more relevant.

Birth-dependent mutation rate: consequences When the model ($\mathcal{Q}_b$) is coupled with a phenotype to fitness landscape with two optima, one for birth, the other one for survival, a new trade-off arises in the population. Compared to the standard approach ($\mathcal{Q}_{stand}$), the symmetry between birth and survival is broken. Thus, in a perfectly symmetric situation (symmetric initial condition and fitness function), our analytical results and numerical simulations show new nontrivial strategies for the trajectories of mean phenotype and for the stationary phenotype distribution. These new strategies are in sharp contrast with those displayed for the standard model equation ($\mathcal{Q}_{stand}$) for which the two optima are perfectly equivalent. With the model ($\mathcal{Q}_b$), we obtained trajectories of adaptation where the mean phenotype of the population is initially attracted by the birth optimum, but eventually converges to the survival optimum, following a hook-shaped curve (see Figure 2). It is well-known that increasing the mutation rate has antagonistic effects on adaptation in the FGM (and other models with both deleterious and beneficial mutations) as it generates fitness variance to fuel and speed-up adaptation (Lavigne et al., 2020) but lowers the mean fitness by creating a larger mutation load (e.g. Anciaux et al., 2019). Here, these two effects shape the trajectories of adaptation. When far for equilibrium, the phenotypes with a higher mutation rate tend to be advantaged, leading to trajectories of mean phenotype that are initially biased toward the birth optimum. Then, in a second time, adaptation promotes the survival optimum, as it is associated with a lower mutation load.

Another feature of the model ($\mathcal{Q}_b$) is that the transient trajectory of mean fitness displays plateaus, as observed for instance in Figure 2. This phenomenon of several epochs in adaptation is well documented thanks to the longest ever evolution experiment, the 'Long Term Evolution Experiment' (LTEE). Experimenting on *Eschereschia Coli* bacteria, Wiser et al. (2013) found out that even after more than $70,000$ generations, fitness had not reached its maximum, apparently challenging

the very existence of such a maximum, the essence of Fisher's Geometrical Model. It was then argued that the data could be explained by a two epoch model Good and Desai (2015), with or without saturation. A similar pattern was observed for a RNA virus Novella et al. (1995). Recently, Hamel et al. (2020) showed that the FGM with a single optimum but anistropic mutation effects also leads to plateaus, and they obtained a good fit with the LTEE data. Our study shows that, when coupled with a phenotype to fitness landscape with two optima, the $(\mathcal{Q}_b)$ is also a possible candidate to explain these trajectories of adaptation.

The model $(\mathcal{Q}_b)$ in the mathematical literature Some authors have already considered operator which are closely related to the mutation operator in $(\mathcal{Q}_b)$. For instance Lorz et al. (2011) considered non homogeneous operators of the form $\mathcal{B}(q)(t,\mathbf{x}) = \operatorname{div}(b(\mathbf{x})\nabla q(t,\mathbf{x}))$ within the framework of constrained Hamilton-Jacobi equations. However this operator does not emerge as the limit of a microscopic diffusion process or as an approximation of an integral mutation operator. It is more adapted to the study of heat conduction as it notably tends to homogenize the solution compared to the Fokker-Planck operator $\Delta(b(\mathbf{x})q(t,\mathbf{x}))$, see Figure II.7 in Roques (2013). Finally, the flexible framework of Bürger (2000) allows for heterogeneous mutation rate. Due to the complicated nature of the operator involved, ($\mathcal{B}$ is a kernel operator compact or power compact), the theoretical framework is in turn very intricate. Quantitative results are in consequence either relatively few, and typically consist in existence and uniqueness of solutions, upper or lower bounds on the asymptotic mean fitness (Bürger, 1998, 2000), or concern simpler models (with a discretization of the time or of the phenotypic space), see Hermisson et al. (2002), Redner (2004).

Sexual reproduction How to take into account a phenotype dependent birth rate with a sexual mode of reproduction is an open question to the best of our knowledge. A classical operator to model sexual genetic inheritance in the background adopted in this article is the *infinitesimal operator*, introduced by Fisher, (Fisher, 1918), see Slatkin (1970); Cotto and Ronce (2014) or the review of Turelli (2017). It describes a trait deviation of the offspring around the mean of the phenotype of the parents, drawn from a Gaussian distribution. Mathematically, few studies have tackled the operator, with the notable only exceptions of the derivation from a microscopic point of view of Barton et al. (2017), the small variance and stability of analysis of Calvez et al. (2019); Patout (2020), and finally in Mirrahimi and Raoul (2013); Raoul (2017), with an additional spatial structure, the convergence of the model towards the Kirkpatrick-Barton model when the reproduction rate is large. In all those cases, the reproduction term is assumed to be constant. With the formalism of (2), at the population scale, mating and birth should be positively correlated, which should lead to considering the following variation on the infinitesimal operator, which acts upon the phenotype $x \in \mathbb{R}$ (for simplicity, we take $n = 1$ here for the dimension of

the phenotype):

$$\mathcal{M}(f)(x) :=$$
$$\frac{1}{\sigma\sqrt{\pi}} \iint_{\mathbb{R}^2} \exp\left[-\frac{1}{\sigma^2}\left(x - \frac{x_1+x_2}{2}\right)^2\right] b(x_1) f(x_1) \frac{\omega(x_2) f(x_2)}{\int_{\mathbb{R}} \omega(x_2') f(x_2')\, dx_2'}\, dx_1 dx_2. \quad (20)$$

We try to explain this operator as follows. It describes how an offspring with trait x appears in the population. First, an individual x_1 rings a birth clock, at a rate given by its trait and the distribution of birth events b, as in (1). Next, this individual mates with a second parent x_2, chosen according to the weight ω. Then, the trait of the offspring is drawn from the normal law $\mathcal{N}\left(\frac{x_1+x_2}{2}, \sigma^2\right)$.

As the birth rate of individuals seems a decisive factor in being chosen as a second parent, a reasonable choice would be $\omega = b$ in the formula above. Again, to the best of our knowledge, no mathematical tools have been developed to tackle the issues we raise in this article with this new operator. We can mention the works of Raoul (private communication) about similar operators.

A new trade-off, similar to the one discussed in this article, can also arise with the operator (20). Indeed, coupled with a selection term, as in (1) for instance, a trade-off between birth and survival can appear if b (or ω) and d have different optima. It would be very interesting to follow the trajectories of fitness along time as in Figure 2, to discover if the effects highlighted in this paper for asexual reproduction are still present, and when, with sexual reproduction. Of course, a third factor in the trade-off is also present, through the weight of the choice of the second parent via the function ω. If an external factor favors a second parent around a third optimum, then the effect it has on the population should also be taken into account. The relevance of such a model in an individual based setting, as in Section 2 is also an open question to this day for the operator (20). With the assumption $\omega = b$, the roles of first and second parents are symmetric in the operator (20), and an investigation of the balance between birth and survival could be carried out without additional assumptions.

Appendix A Mathematical proofs

A.1 Proof of Proposition 2.4

For $\phi : \Omega \to \mathbb{R}$ measurable and bounded, let

$$P\phi(\mathbf{x}) = (1 - U_K)\phi(\mathbf{x}_i) + U_K \int_\Omega \phi(\mathbf{y}) \rho_K(\mathbf{x}_i, \mathbf{y}) d\mathbf{y}.$$

Lemma A.1. For $t \in \mathbb{N}$,

$$\langle \nu_t^K, \phi \rangle = \langle \nu_0^K, \phi \rangle + \sum_{s=1}^{t} \langle \nu_s^K, w_K e^{-c_K K \langle \nu_s^K, 1 \rangle} P\phi - \phi \rangle + M_t^K(\phi),$$

where $(M_t^K(\phi), t \geq 0)$ is a local martingale with quadratic variation

$$\sum_{s=1}^{t} \left(\frac{1}{K} \langle \nu_s^K, w_K P\phi^2 \rangle e^{-c_K K \langle \nu_s^K, 1 \rangle} + \langle \nu_s^K, w_K e^{-c_K K \langle \nu_s^K, 1 \rangle} P\phi - \phi \rangle^2 \right).$$

Proof. From the definition of the model, for any $\phi : \Omega \to \mathbb{R}$ measurable and bounded,

$$\begin{aligned}\mathbb{E}\left[\langle \nu_{t+1}^K, \phi\rangle \middle| \nu_t^K\right] &= \frac{1}{K}\sum_{i=1}^{N_t} w_K(\mathbf{x}_i)\, e^{-c_K N_t}\left\{(1-U_K)\phi(\mathbf{x}_i) + U_K \int_\Omega \phi(\mathbf{y})\rho_K(\mathbf{x}_i, d\mathbf{y})\right\} \\ &= \langle \nu_t^K, w_K e^{-c_K K\langle \nu_t^K, 1\rangle} P\phi\rangle.\end{aligned}$$

Hence

$$\mathbb{E}\left[\langle \nu_{t+1}^K, \phi\rangle - \langle \nu_t^K, \phi\rangle \middle| \nu_t^K\right] = \langle \nu_t^K, w_K e^{-c_K K\langle \nu_t^K, 1\rangle} P\phi - \phi\rangle. \tag{21}$$

We now wish to compute

$$\mathbb{E}\left[\left(\langle \nu_{t+1}^K, \phi\rangle - \langle \nu_t^K, \phi\rangle\right)^2 \middle| \nu_t^K\right].$$

To do this, let $\nu_t^K = \frac{1}{K}\sum_{i=1}^N \delta_{\mathbf{x}_i}$ and let N_i be the number of offspring of individual i and let $(Y_{i,j}, 1 \le j \le N_i)$ denote their types. From the definition of the model, N_i is a Poisson random variable with parameter $w_K(\mathbf{x}_i)e^{-c_K N}$ and the $(Y_{i,j}, j \ge 1)$ are i.i.d. with

$$\mathbb{E}[\phi(Y_{i,j}) \,|\, \mathbf{x}_i] = P\phi(\mathbf{x}_i), \qquad \mathbb{V}[\phi(Y_{i,j}) \,|\, \mathbf{x}_i] = P\phi^2(\mathbf{x}_i) - (P\phi(\mathbf{x}_i))^2.$$

Then we write

$$\begin{aligned}\langle \nu_{t+1}^K, \phi\rangle - \langle \nu_t^K, \phi\rangle &= \frac{1}{K}\sum_{i=1}^N \left(\sum_{j=1}^{N_i} (\phi(Y_{i,j}) - P\phi(\mathbf{x}_i))\right) \\ &\quad + \frac{1}{K}\sum_{i=1}^N \left(N_i - w_K(\mathbf{x}_i)e^{-c_K N}\right) P\phi(\mathbf{x}_i) + \langle \nu_t^N, w_K e^{-c_K N} P\phi - \phi\rangle.\end{aligned}$$

Since the third term is deterministic and the first two terms are uncorrelated,

$$\begin{aligned}\mathbb{E}\left[\left(\langle \nu_{t+1}^K, \phi\rangle - \langle \nu_t^K, \phi\rangle\right)^2 \middle| \nu_t^K\right] &= \frac{1}{K^2}\sum_{i=1}^N w_K(\mathbf{x}_i)e^{-c_K N}\left(P\phi^2(\mathbf{x}_i) - (P\phi(\mathbf{x}_i))^2\right) \\ &\quad + \frac{1}{K^2}\sum_{i=1}^N w_K(\mathbf{x}_i)e^{-c_K N}(P\phi(\mathbf{x}_i))^2 + \langle \nu_t^K, w_K e^{-c_K N} P\phi - \phi\rangle^2.\end{aligned}$$

Rearranging, we arrive at

$$\begin{aligned}&\mathbb{E}\left[\left(\langle \nu_{t+1}^K, \phi\rangle - \langle \nu_t^K, \phi\rangle\right)^2 \middle| \nu_t^K\right] \\ &\qquad = \frac{1}{K}\langle \nu_t^K, w_K P\phi^2\rangle e^{-c_K K\langle \nu_t^K, 1\rangle} + \langle \nu_t^K, w_K e^{-c_K K\langle \nu_t^K, 1\rangle} P\phi - \phi\rangle^2.\end{aligned} \tag{22}$$

This concludes the proof of the lemma. □

Note that, setting

$$\mathcal{M}_K^*\phi(\mathbf{x}) = \frac{U_K}{\varepsilon_K}\int_\Omega (\phi(\mathbf{y}) - \phi(\mathbf{x}))\rho_K(\mathbf{x}, \mathbf{y})d\mathbf{y},$$

equation (21) can also be written

$$\mathbb{E}\left[\langle \nu_{t+1}^K, \phi\rangle - \langle \nu_t^K, \phi\rangle \middle| \nu_t^K\right] \\ = \langle \nu_t^K, w_K\, e^{-c_K K\langle \nu_t^K, 1\rangle}\, \varepsilon_K \mathcal{M}_K^* \phi + (w_K\, e^{-c_K K\langle \nu_t^K, 1\rangle} - 1)\, \phi\rangle.$$

Using Assumption (A1), we then have

$$\mathbb{E}\left[\langle \nu_{t+1}^K, \phi\rangle - \langle \nu_t^K, \phi\rangle \middle| \nu_t^K\right] = \varepsilon_K \langle \nu_t^K, \mathcal{M}_K^*\phi + (m - c\langle \nu_t^K, 1\rangle)\phi\rangle + o(\varepsilon_K \langle \nu_t^K, 1\rangle).$$

We then note that, in the case of Assumption (FE), $\mathcal{M}_K^*\phi = \mathcal{M}^*\phi$, while in the case of Assumption (SE), by a Talyor expansion, for any $\phi \in C_0^2(\Omega)$,

$$\mathcal{M}_K^*\phi(\mathbf{x}) = \frac{\lambda U}{2}\Delta\phi(\mathbf{x}) + o(1),$$

uniformly in $\mathbf{x} \in \Omega$. Finally, note that the first term in (22) is of the order of $1/K$ while the second term is of the order of ε_K^2.

For $N \geq 1$, define a stopping time τ_N^K by

$$\tau_N^K = \inf\{t \geq 0 : \langle \nu_t^K, 1\rangle \geq N\rangle.$$

Lemma A.2. For any fixed $N \geq 1$ and $T > 0$, for any $\phi \in C^2(\Omega)$,

$$\sup_{0 \leq t \leq \lfloor T/\varepsilon_K \rfloor} |M_{t\wedge\tau_N^K}^K(\phi)| \longrightarrow 0,$$

in probability as $K \to \infty$.

Proof. By Doob's martingale inequality,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{0 \leq t \leq \lfloor T/\varepsilon_K \rfloor} |M_{t\wedge\tau_N^K}^K(\phi)|^2\right] \\
&\quad \leq 4\,\mathbb{E}\left[|M_{\lfloor T/\varepsilon_K \rfloor \wedge \tau_N^K}^K(\phi)|^2\right] \\
&\quad \leq 4\,\mathbb{E}\left[\sum_{s=1}^{\lfloor T/\varepsilon_K \rfloor \wedge \tau_N^K} \left\{\frac{1}{K}\langle \nu_s^K, w_K P\phi^2\rangle + \langle \nu_s^K, w_K e^{-c\varepsilon_K\langle \nu_s^K, 1\rangle} P\phi - \phi\rangle^2\right\}\right].
\end{aligned} \tag{23}$$

Clearly, for $0 \leq s \leq \lfloor T/\varepsilon_K \rfloor \wedge \tau_N^K$,

$$|\langle \nu_s^K, w_K P\phi^2\rangle| \leq C\|\phi\|_\infty^2 N. \tag{24}$$

We then note that there exists a function $r_K : \mathbb{R} \to \mathbb{R}$ such that, for all $x \in \mathbb{R}$,

$$e^{\varepsilon_K x} = 1 + \varepsilon_K x\, e^{\varepsilon_K r_K(x)}, \qquad \text{and} \qquad |r_K(x)| \leq |x|.$$

With this notation,

$$w_K(\mathbf{x})e^{-c_K K\langle \nu_s^K, 1\rangle} - 1 = \varepsilon_K m(\mathbf{x}) e^{\varepsilon_K r_K(m(\mathbf{x})) - c\,\varepsilon_K\langle \nu_s^K, 1\rangle} - c\,\varepsilon_K \langle \nu_s^K, 1\rangle e^{\varepsilon_K r_K(-c\langle \nu_s^K, 1\rangle)}.$$

Hence, using the fact that $r_K(\mathbf{x})$ has the same sign as x,

$$|\langle \nu_s^K, (w_K e^{-c\varepsilon_K \langle \nu_s^K, 1\rangle} - 1) P\phi \rangle| \leq C\|\phi\|_\infty \, \varepsilon_K (N + N^2). \tag{25}$$

Finally, $P\phi - \phi = \varepsilon_K \mathcal{M}_K^* \phi$, and, for $\phi \in C_0^2(\Omega)$, under either Assumption (FE) or (SE),

$$\sup_{K>0} \|\mathcal{M}_K^* \phi\|_\infty \leq C_\phi,$$

we have

$$|\langle \nu_s^K, P\phi - \phi \rangle| \leq C_\phi \, \varepsilon_K N. \tag{26}$$

Plugging (24), (25) and (26) in (23), we obtain

$$\mathbb{E}\left[\sup_{0 \leq t \leq \lfloor T/\varepsilon_K \rfloor} |M_{t \wedge \tau_N^K}^K(\phi)|^2 \right] \leq CT \left(\frac{N}{\varepsilon_K K} + \varepsilon_K^2 (N + N^2)^2 \right).$$

Since the right-hand-side tends to zero as $K \to \infty$, this concludes the proof of the lemma. □

Lemma A.3. Fix $T > 0$, and let

$$X_K = \sup_{0 \leq t \leq \lfloor T/\varepsilon_K \rfloor} \langle \nu_t^K, 1 \rangle.$$

Then $(X_K, K > 0)$ is tight in $\mathbb{R}_+$. Moreover, for any $\delta > 0$, N can be chosen such that

$$\limsup_{K \to \infty} \mathbb{P}(\tau_N^K \leq \lfloor T/\varepsilon_K \rfloor) \leq \delta.$$

Proof. Looking at the statement of Lemma A.1, we note that $\mathcal{M}_K^* 1 = 0$ and that

$$\begin{aligned} w_K(\mathbf{x}) e^{-c_K K \langle \nu_s^K, 1\rangle} - 1 &= \left(e^{\varepsilon_K m(\mathbf{x})} - 1\right) e^{-c\varepsilon_K \langle \nu_s^K, 1\rangle} + e^{-c\varepsilon_K \langle \nu_s^K, 1\rangle} - 1 \\ &\leq C\varepsilon_K \langle \nu_s^K, 1 \rangle, \end{aligned}$$

for some constant $C > 0$, using the fact that m is bounded. As a consequence,

$$\begin{aligned} \langle \nu_{t \wedge \tau_N^K}^K, 1 \rangle &\leq \langle \nu_0^K, 1 \rangle + \sum_{s=1}^{t \wedge \tau_N^K} \varepsilon_K C \langle \nu_s^K, 1 \rangle + M_{t \wedge \tau_N^K}^K(1) \\ &\leq \langle \nu_0^K, 1 \rangle + \sum_{s=1}^{t} \varepsilon_K C \langle \nu_{s \wedge \tau_N^K}^K, 1 \rangle + M_{t \wedge \tau_N^K}^K(1). \end{aligned}$$

By Gronwall's inequality, we obtain

$$\langle \nu_{t \wedge \tau_N^K}^K, 1 \rangle \leq \left(\langle \nu_0^K, 1 \rangle + \sup_{0 \leq s \leq t} M_{s \wedge \tau_N^K}^K(1) \right) e^{C\varepsilon_K t}.$$

Hence,

$$\sup_{0\le t\le \lfloor T/\varepsilon_K\rfloor} \langle \nu^K_{t\wedge\tau^K_N}, 1\rangle \le \left(\langle \nu^K_0, 1\rangle + \sup_{0\le t\le \lfloor T/\varepsilon_K\rfloor} M^K_{t\wedge\tau^K_N}(1)\right) e^{CT}.$$

As a result,

$$\begin{aligned}\mathbb{P}(X_K \ge N) &= \mathbb{P}(\tau^K_N \le \lfloor T/\varepsilon_K\rfloor)\\ &\le \mathbb{P}\left(\left[\langle \nu^K_0, 1\rangle + \sup_{0\le t\le \lfloor T/\varepsilon_K\rfloor} M^K_{t\wedge\tau^K_N}(1)\right] e^{CT} \ge N\right)\\ &\le \mathbb{P}\left(\langle \nu^K_0, 1\rangle \ge \frac{1}{2}Ne^{-CT}\right) + \mathbb{P}\left(\sup_{0\le t\le \lfloor T/\varepsilon_K\rfloor} M^K_{t\wedge\tau^K_N}(1) \ge \frac{1}{2}Ne^{-CT}\right).\end{aligned}$$

Since $\langle \nu^K_0, 1\rangle$ is tight, for any $\delta > 0$ we can choose N large enough such that

$$\limsup_{K\to\infty} \mathbb{P}\left(\langle \nu^K_0, 1\rangle \ge \frac{1}{2}Ne^{-CT}\right) \le \delta.$$

In addition, by Lemma A.2, for any $N \ge 1$,

$$\limsup_{K\to\infty} \mathbb{P}\left(\sup_{0\le t\le \lfloor T/\varepsilon_K\rfloor} M^K_{t\wedge\tau^K_N}(1) \ge \frac{1}{2}Ne^{-CT}\right) = 0.$$

Hence we can choose N large enough such that

$$\limsup_{K\to\infty} \mathbb{P}(X_K \ge N) \le \delta,$$

concluding the proof. □

Lemma A.4. For any $T > 0$, the sequence of $M_F(\Omega)$-valued processes

$$\left(\nu^K_{\lfloor t/\varepsilon_K\rfloor}, t \in [0,T]\right), \quad K > 0,$$

is C-tight in $D([0,T], M_F(\Omega))$.

Proof. By (Roelly-Coppoletta, 1986, Theorem 2.1), $\nu^K_{\lfloor \cdot/\varepsilon_K\rfloor}$ is tight if and only if $\langle \nu^K_{\lfloor \cdot/\varepsilon_K\rfloor}, \phi\rangle$ is tight in $D([0,T], \mathbb{R})$ for any ϕ in a dense subset of the space of continuous functions on Ω vanishing at infinity. Hence let $\phi \in C^2(\Omega)$ and let us show that $(\langle \nu^K_{\lfloor t/\varepsilon_K\rfloor}, \phi\rangle, t \in [0,T])$ is tight. By Lemma A.3, it is sufficient to show that $(\langle \nu^K_{\lfloor t/\varepsilon_K\rfloor\wedge\tau^K_N}, \phi\rangle, t \in [0,T])$ is tight for any N large enough. Now, using (25) and (26), for $0 \le s \le \lfloor T/\varepsilon_K\rfloor \wedge \tau^K_N$,

$$|\langle \nu^K_s, w_K e^{-c_K K\langle \nu^K_s, 1\rangle} P\phi - \phi\rangle| \le C_{\phi,N}\, \varepsilon_K,$$

for some constant $C_{\phi,N} > 0$ depending only on ϕ and N. As a result, if $w_\theta(f)$ denotes the modulus of continuity of $f : [0,T] \to \mathbb{R}$, *i.e.*

$$w_\delta(f) = \sup_{\substack{|t-s|\le\theta\\ 0\le s\le t\le T}} |f(t) - f(s)|,$$

we obtain

$$w_\theta\left(\langle \nu^K_{\lfloor \cdot/\varepsilon_K\rfloor\wedge\tau^K_N},\phi\rangle\right) \le C_{\phi,N}\,\theta + 2\sup_{0\le t\le \lfloor T/\varepsilon_K\rfloor}|M^K_{t\wedge\tau^K_N}(\phi)|.$$

Hence, by Lemma A.2, for any $\delta>0$ and $\varepsilon>0$, there exists $\theta>0$ such that

$$\limsup_{K\to\infty}\mathbb{P}\left(w_\theta\left(\langle \nu^K_{\lfloor \cdot/\varepsilon_K\rfloor\wedge\tau^K_N},\phi\rangle\right)>\delta\right)\le\varepsilon.$$

Combined with Lemma A.3, this shows that $(\langle \nu^K_{\lfloor t/\varepsilon_K\rfloor\wedge\tau^K_N},\phi\rangle, t\in[0,T])$ is C-tight for any $\phi\in C^2(\Omega)$ and $N>0$ (see for example (Jacod and Shiryaev, 2003, Proposition VI.3.26)), and the result is proved. □

We can now conclude the proof of the main result.

Proof of Proposition 2.4. Consider a converging subsequence, still denoted

$$\left(\nu^K_{\lfloor t/\varepsilon_K\rfloor}, t\in[0,T]\right)$$

and let $(f_t, t\in[0,T])$ be its limit. Since the sequence is C-tight, $t\mapsto f_t$ is continuous and the convergence holds uniformly on $[0,T]$. Moreover, by the Skorokhod embedding theorem, without loss of generality, we can assume that the convergence holds in probability.

The result will be proved if we show that f_t solves (7). Let $\phi\in C^2(\Omega)$. By Lemma A.2,

$$\sup_{0\le t\le\lfloor T/\varepsilon_K\rfloor}\left|\langle\nu^K_{t\wedge\tau^K_N},\phi\rangle-\langle\nu^K_0,\phi\rangle-\sum_{s=1}^{t\wedge\tau^K_N}\langle\nu^K_s, w_K e^{-c_K K\langle\nu^K_s,1\rangle}P\phi-\phi\rangle\right|\longrightarrow 0$$

in probability as $K\to\infty$. In addition,

$$\begin{aligned}
w_K(\mathbf{x})e^{-c_KK\langle\nu^K_s,1\rangle}P\phi(\mathbf{x})-\phi(\mathbf{x})-\varepsilon_K\left(\mathcal{M}^*\phi(\mathbf{x})-(m(\mathbf{x})-c\langle\nu^K_s,1\rangle)\phi(\mathbf{x})\right)&\\
=\left(e^{\varepsilon_K(m(\mathbf{x})-c\langle\nu^K_s,1\rangle)}-1-\varepsilon_K(m(\mathbf{x})-c\langle\nu^K_s,1\rangle)\right)P\phi(\mathbf{x})&\\
+P\phi(\mathbf{x})-\phi(\mathbf{x})-\varepsilon_K\mathcal{M}^*\phi(\mathbf{x})&\\
+\varepsilon_K(m(\mathbf{x})-c\langle\nu^K_s,1\rangle)(P\phi(\mathbf{x})-\phi(\mathbf{x})).&
\end{aligned}$$

Hence, for $0\le s\le\lfloor T/\varepsilon_K\rfloor\wedge\tau^K_N$ and $\phi\in C^2_0(\Omega)$,

$$\begin{aligned}
\left|w_K(\mathbf{x})e^{-c_KK\langle\nu^K_s,1\rangle}P\phi(\mathbf{x})-\phi(\mathbf{x})-\varepsilon_K\left(\mathcal{M}^*\phi(\mathbf{x})-(m(\mathbf{x})-c\langle\nu^K_s,1\rangle)\phi(\mathbf{x})\right)\right|&\\
\le C\varepsilon_K^2+\varepsilon_K\left(\mathcal{M}^*_K\phi(\mathbf{x})-\mathcal{M}^*\phi(\mathbf{x})\right)&\\
\le C'\varepsilon_K^2.&
\end{aligned}$$

As a result,

$$\begin{aligned}
\sup_{0\le t\le\lfloor T/\varepsilon_K\rfloor}\Bigg|\sum_{s=1}^{t\wedge\tau^K_N}\langle\nu^K_s, w_Ke^{-c_KK\langle\nu^K_s,1\rangle}P\phi-\phi\rangle&\\
-\sum_{s=1}^{t\wedge\tau^K_N}\varepsilon_K\langle\nu^K_s,\mathcal{M}^*\phi+(m-c\langle\nu^K_s,1\rangle)\phi\rangle\Bigg|\le C'T\varepsilon_K.&
\end{aligned}$$

Hence, on the event $\{\tau_N^K > \lfloor T/\varepsilon_K \rfloor\}$,

$$\begin{aligned}
&\sup_{0\le t\le T}\left|\langle f_t,\phi\rangle - \langle f_0,\phi\rangle - \int_0^t \langle f_s, \mathcal{M}^*\phi - (m - c\langle f_s,1\rangle)\phi\rangle ds\right| \\
&\qquad \le 2\sup_{0\le t\le T}|\langle \nu^K_{\lfloor t/\varepsilon_K\rfloor},\phi\rangle - \langle f_t,\phi\rangle| + C'T\varepsilon_K + \sup_{0\le t\le \lfloor T/\varepsilon_K\rfloor}\left|M^K_{t\wedge\tau_N^K}(\phi)\right| \\
&\qquad + T\sup_{0\le s\le T}\left|\langle \nu^K_s, \mathcal{M}^*\phi + (m - c\langle \nu^K_s,1\rangle)\phi\rangle - \langle f_s, \mathcal{M}^*\phi + (m - c\langle f_s,1\rangle)\phi\rangle\right|.
\end{aligned}$$

Combined with Lemma A.3 and the (uniform) convergence of $\nu^K_{\lfloor \cdot/\varepsilon_K\rfloor}$ to f, this shows that, for any $\varepsilon > 0$,

$$\mathbb{P}\left(\sup_{0\le t\le T}\left|\langle f_t,\phi\rangle - \langle f_0,\phi\rangle - \int_0^t \langle f_s, \mathcal{M}^*\phi - (m - c\langle f_s,1\rangle)\phi\rangle ds\right| > \varepsilon\right) = 0.$$

It follows that $(f_t, t\in[0,T])$ solves (7), and the result is proved. □

A.2 Proof of Proposition 3.1

Multiplying equation $(\mathcal{Q}_b)$ by x_1, integrating over $\mathbf{x}\in\Omega$ and evaluating at $t=0$, we get

$$\begin{aligned}
\overline{x}_1'(0) &= \partial_t\left(\int_\Omega x_1 q(t,\mathbf{x})d\mathbf{x}\right)(t=0) \\
&= D\int_\Omega x_1\Delta(bq_0)(\mathbf{x})d\mathbf{x} + \int_\Omega x_1 q_0(\mathbf{x})m(\mathbf{x})d\mathbf{x} - \bar{m}(0)\int_\Omega x_1 q_0(\mathbf{x})d\mathbf{x}.
\end{aligned}$$

From Green formula we infer

$$\begin{aligned}
\int_\Omega x_1\Delta(bq_0)(\mathbf{x})d\mathbf{x} &= \int_{\partial\Omega} x_1\,\nabla(bq_0)\cdot\vec{\nu}(\mathbf{x})\,ds - \int_{\partial\Omega}(bq_0)(\mathbf{x})\,\vec{e_1}\cdot\vec{\nu}(\mathbf{x})\,ds, \\
&= 0
\end{aligned}$$

since q_0 is compactly supported in Ω. Moreover, since q_0 and m both satisfy $m(\iota(\mathbf{x})) = m(\mathbf{x})$, $q_0(\iota(\mathbf{x})) = q(\mathbf{x})$,

$$\int_\Omega x_1 q_0(\mathbf{x})m(\mathbf{x})d\mathbf{x} = \int_\Omega x_1 q_0(\mathbf{x})d\mathbf{x} = 0.$$

This shows that $\overline{x}_1'(0) = 0$.

We next turn to the second derivative $\overline{x}_1''(0)$. We differentiate equation $(\mathcal{Q}_b)$ with respect to time, multiply by x_1, integrate over $\mathbf{x}\in\Omega$ and evaluate at $t=0$ to reach

$$\begin{aligned}
\overline{x}_1''(0) &= \partial_{tt}\left(\int_\Omega x_1 q(t,\mathbf{x})d\mathbf{x}\right)(t=0) \\
&= D\int_\Omega x_1\Delta(b\partial_t q(0,\mathbf{x}))d\mathbf{x} + \int_\Omega x_1\partial_t q(0,\mathbf{x})m(\mathbf{x})d\mathbf{x} \\
&\quad - \bar{m}(0)\int_\Omega x_1\partial_t q(0,\mathbf{x})d\mathbf{x} - \bar{m}'(0)\int_\Omega x_1 q_0(\mathbf{x})d\mathbf{x}.
\end{aligned}$$

From the above computation, this reduces to

$$\overline{x}_1''(0) = D\int_\Omega x_1\Delta(b\partial_t q(0,\mathbf{x}))d\mathbf{x} + \int_\Omega x_1\partial_t q(0,\mathbf{x})m(\mathbf{x})d\mathbf{x}.$$

Moreover, since $\partial_t q(0,\mathbf{x})$ is also compactly supported (this follows from equation $(\mathcal{Q}_b)$), Green formula yields

$$\int_\Omega x_1\Delta(b\partial_t q(0,\mathbf{x}))d\mathbf{x} = 0,$$

and we are left with

$$\overline{x}_1''(0) = \int_\Omega x_1\partial_t q(0,\mathbf{x})m(\mathbf{x})d\mathbf{x}. \tag{27}$$

We multiply equation $(\mathcal{Q}_b)$ by $x_1 m(\mathbf{x})$, integrate over $\mathbf{x}\in\Omega$ and evaluate at $t=0$ to obtain

$$\begin{aligned}\int_\Omega x_1\partial_t q(0,\mathbf{x})m(\mathbf{x})d\mathbf{x} &= D\int_\Omega \Delta(bq_0)(\mathbf{x})x_1 m(\mathbf{x})d\mathbf{x} + \int_\Omega x_1 q_0(\mathbf{x})m(\mathbf{x})^2 d\mathbf{x}\\ &\quad -\bar{m}(0)\int_\Omega x_1 q_0(\mathbf{x})m(\mathbf{x})d\mathbf{x}.\end{aligned}$$

By symmetry, the last two terms vanish, and another Green formula leads to

$$\int_\Omega x_1\partial_t q(0,\mathbf{x})m(\mathbf{x})d\mathbf{x} = D\int_\Omega (bq_0)(\mathbf{x})\Delta(x_1 m(\mathbf{x}))d\mathbf{x}. \tag{28}$$

Then, we observe that

$$\int_{\Omega\cap\{x_1<0\}} b(\mathbf{x})\, q_0(\mathbf{x})\Delta(x_1 m(\mathbf{x}))d\mathbf{x} = -\int_{\Omega\cap\{x_1>0\}} s(\mathbf{x})q_0(\mathbf{x})\Delta(x_1 m(\mathbf{x}))d\mathbf{x},$$

as q_0 and m are symmetric about $\{x_1=0\}$, and from (9). Thus,

$$\begin{aligned}\int_\Omega (bq_0)(\mathbf{x})\Delta(x_1 m(\mathbf{x}))d\mathbf{x} &= \int_{\Omega\cap\{x_1<0\}} (bq_0)(\mathbf{x})\Delta(x_1 m(\mathbf{x}))d\mathbf{x}\\ &\quad + \int_{\Omega\cap\{x_1>0\}} (bq_0)(\mathbf{x})\Delta(x_1 m(\mathbf{x}))d\mathbf{x}\\ &= \int_{\Omega\cap\{x_1>0\}\cap K_0} (b-s)(\mathbf{x})\, q_0(\mathbf{x})\,\Delta(x_1 m(\mathbf{x}))d\mathbf{x},\end{aligned}$$

with K_0 the support of q_0 (containing $\mathbf{x}_0$). From this, (27) and (28), we end up with

$$\overline{x}_1''(0) = D\int_{\Omega\cap\{x_1>0\}\cap K_0} (b-s)(\mathbf{x})\, q_0(\mathbf{x})\,\Delta(x_1 m(\mathbf{x}))d\mathbf{x}.$$

We know from (10) that $(b-s)(\mathbf{x})>0$ in $\Omega\cap\{x_1>0\}$. As a result, if $\Delta(x_1 m(\mathbf{x}))$ is nontrivial and nonnegative (nonpositive) on $K_0^+ = K_0\cap\{x_1>0\}$ then $\overline{x}_1''(0)>0$ ($\overline{x}_1''(0)<0$ respectively). This concludes the proof of Proposition 3.1. □

A.3 An explicit solution of the eigenvalue problem

We assume that the dimension is $n = 1$ and

$$b(x) = \begin{cases} 2 & \text{for } x \in (0, a), \\ 1 & \text{for } x \in (-a, 0), \end{cases}$$

and we consider the eigenvalue problem (14) with Dirichlet boundary conditions. As b is discontinuous, the eigenvalue problem must be understood in the weak sense. In particular, we have to solve

$$\begin{cases} D\, q''_{1,\infty}(x) &= (\overline{m}_\infty - r - 3)\, q_{1,\infty}(x),\ x \in (-a, 0), \\ 2D\, q''_{2,\infty}(x) &= (\overline{m}_\infty - r - 3)\, q_{2,\infty}(x),\ x \in (0, a), \end{cases} \tag{29}$$

with the boundary, continuity and flux conditions:

$$\begin{cases} q_{1,\infty}(-a) = q_{2,\infty}(a) = 0, \\ q_{1,\infty}(0) = q_{2,\infty}(0),\ q'_{1,\infty}(0) = 2\, q'_{2,\infty}(0), \end{cases} \tag{30}$$

the positivity conditions $q_{1,\infty}$, $q_{2,\infty} > 0$ and $\overline{m}_\infty - r - 3 < 0$.

Set $\mu = \sqrt{2\,D}$ and $B := \sqrt{-\overline{m}_\infty + r + 3}/\mu$. We have

$$\begin{cases} q_{1,\infty}(x) = -\dfrac{\sqrt{2}}{B} \cos\left(x\, B\, \sqrt{2}\right) \left(\tan(x\, B\, \sqrt{2}) + \tan(a\, B\, \sqrt{2})\right),\ x \in (-a, 0), \\ q_{2,\infty}(x) = \dfrac{1}{B} \cos\left(x\, B\right) \left(\tan(a\, B) - \tan(x\, B)\right),\ x \in (0, a). \end{cases} \tag{31}$$

The equality $q_{1,\infty}(0) = q_{2,\infty}(0)$ thus implies:

$$\sqrt{2}\, \tan(a\, B\, \sqrt{2}) = -\tan(a\, B). \tag{32}$$

The positivity of $q_{1,\infty}$, $q_{2,\infty}$ implies that $0 < a\, B < \pi/2$. The equation (32) thus admits a unique solution $a\, B \in (\pi/(2\,\sqrt{2}), \pi/2)$ ($a\, B \approx 1.338761890$). Additionally, we have:

$$\frac{\int_{-a}^{0} q_{1,\infty}(x)\, dx}{\int_{0}^{a} q_{2,\infty}(x)\, dx} = -\frac{(1 - \cos(a\, B\, \sqrt{2}))\, \cos(a\, B)}{(1 - \cos(a\, B))\, \cos(a\, B\, \sqrt{2})},$$

and using (32),

$$\begin{aligned} \frac{\int_{-a}^{0} q_{1,\infty}(x)\, dx}{\int_{0}^{a} q_{2,\infty}(x)\, dx} &= \frac{1}{\sqrt{2}} \frac{(1 - \cos(a\, B\, \sqrt{2}))\, \sin(a\, B)}{(1 - \cos(a\, B))\, \sin(a\, B\, \sqrt{2})} \\ &= \frac{1}{\sqrt{2}} \frac{j(a\, B\, \sqrt{2})}{j(a\, B)}. \end{aligned}$$

with $j(x) := (1 - \cos(x))/\sin(x)$. As $j(x\,\sqrt{2})/j(x)$ is increasing on $(\pi/(2\,\sqrt{2}), \pi/2)$, we get:

$$\frac{1}{\sqrt{2}} \frac{j(a\, B\, \sqrt{2})}{j(a\, B)} \geq \frac{1}{\sqrt{2}} \frac{j(\pi/2)}{j(\pi/(2\,\sqrt{2}))} = \frac{1}{\sqrt{2}}\, \frac{1}{j(\pi/(2\,\sqrt{2}))}.$$

As $1/(2\sqrt{2}) < 3/8$ and since j is increasing on $(\pi/(2\sqrt{2}), \pi/2)$,

$$j(\pi/(2\sqrt{2})) < j(3\pi/8) = 1 - \sqrt{2} + \sqrt{4 - 2\sqrt{2}}.$$

Finally,

$$\frac{\int_{-a}^{0} q_{1,\infty}(x)\,dx}{\int_{0}^{a} q_{2,\infty}(x)\,dx} = \frac{1}{\sqrt{2}}\frac{j(a\,B\sqrt{2})}{j(a\,B)} > \frac{1}{\sqrt{2}\,j(3\pi/8)} = \frac{1}{2\sqrt{2-\sqrt{2}} - 2 + \sqrt{2}} > 1.$$

References

Alfaro, M. and R. Carles (2014). Explicit solutions for replicator-mutator equations: Extinction versus acceleration. SIAM Journal on Applied Mathematics 74(6), 1919–1934.

Alfaro, M. and R. Carles (2017). Replicator-mutator equations with quadratic fitness. Proceedings of the American Mathematical Society 145(12), 5315–5327.

Alfaro, M. and M. Veruete (2018). Evolutionary branching via replicator-mutator equations. Journal of Dynamics and Differential Equations, 1–24.

Anciaux, Y., A. Lambert, O. Ronce, L. Roques, and G. Martin (2019). Population persistence under high mutation rate: from evolutionary rescue to lethal mutagenesis. Evolution 73(8), 1517–1532.

Anderson, J. P., R. Daifuku, and L. A. Loeb (2004). Viral error catastrophe by mutagenic nucleosides. Annu. Rev. Microbiol. 58, 183–205.

André, J.-B. and B. Godelle (2006). The evolution of mutation rate in finite asexual populations. Genetics 172(1), 611–626.

Barton, N., A. Etheridge, and A. Véber (2017). The infinitesimal model: Definition, derivation, and implications. Theoretical population biology 118, 50–73.

Biktashev, V. N. (2014). A simple mathematical model of gradual darwinian evolution: emergence of a gaussian trait distribution in adaptation along a fitness gradient. Journal of mathematical biology 68(5), 1225–1248.

Bouin, E. and S. Mirrahimi (2013). A hamilton-jacobi approach for a model of population structured by space and trait. arXiv preprint arXiv:1307.8332.

Bull, J. J., R. Sanjuan, and C. O. Wilke (2007). Theory of lethal mutagenesis for viruses. Journal of Virology 81(6), 2930–2939.

Bull, J. J. and C. O. Wilke (2008). Lethal mutagenesis of bacteria. Genetics 180(2), 1061–1070.

Bürger, R. (1998). Mathematical properties of mutation-selection models. Genetica 102, 279.

Bürger, R. (2000). The mathematical theory of selection, recombination, and mutation, Volume 228. Wiley Chichester.

Calvez, V., J. Garnier, and F. Patout (2019). Asymptotic analysis of a quantitative genetics model with nonlinear integral operator. Journal de l'École polytechnique — Mathématiques 6, 537–579.

Cantrell, R. S. and C. Cosner (2003). Spatial ecology via reaction-diffusion equations. John Wiley & Sons Ltd, Chichester, UK .

Champagnat, N., R. Ferrière, and S. Méléard (2006). Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models. Theoretical population biology 69(3), 297–321.

Champagnat, N., R. Ferrière, and S. Méléard (2008). From individual stochastic processes to macroscopic models in adaptive evolution. Stoch. Models 24(suppl. 1), 2–44.

Cotto, O. and O. Ronce (2014). Maladaptation as a source of senescence in habitats variable in space and time. Evolution 68(9), 2481–2493.

Courant, R. and D. Hilbert (2008). Methods of Mathematical Physics, Vol. I. Interscience, New York.

Débarre, F., O. Ronce, and S. Gandon (2013). Quantifying the effects of migration and mutation on adaptation and demography in spatially heterogeneous environments. Journal of Evolutionary Biology 26(6), 1185–1202.

Desai, M. M. and D. S. Fisher (2011). The balance between mutators and nonmutators in asexual populations. Genetics 188(4), 997–1014.

Diekmann, O., P.-E. Jabin, S. Mischler, and B. Perthame (2005). The dynamics of adaptation: an illuminating example and a Hamilton–Jacobi approach. Theoretical population biology 67(4), 257–271.

Figueroa Iglesias, S. and S. Mirrahimi (2019). Selection and mutation in a shifting and fluctuating environment. HAL Preprint 02320525.

Fisher, R. A. (1918). The correlation between relatives on the supposition of mendelian inheritance. Transactions of the Royal Society of Edinburgh 52, 399–433.

Fournier, N. and S. Méléard (2004). A microscopic probabilistic description of a locally regulated population and macroscopic approximations. The Annals of Applied Probability 14(4), 1880–1919.

Gandon, S. and S. Mirrahimi (2017). A Hamilton–Jacobi method to describe the evolutionary equilibria in heterogeneous environments and with non-vanishing effects of mutations. Comptes Rendus Mathematique 355(2), 155–160.

Gerrish, P. J., A. Colato, A. S. Perelson, and P. D. Sniegowski (2007). Complete genetic linkage can subvert natural selection. Proceedings of the National Academy of Sciences 104(15), 6266–6271.

Gil, M.-E., F. Hamel, G. Martin, and L. Roques (2017). Mathematical properties of a class of integro-differential models from population genetics. SIAM J. Appl. Math. 77(4), 1536–1561.

Gil, M.-E., F. Hamel, G. Martin, and L. Roques (2019). Dynamics of fitness distributions in the presence of a phenotypic optimum: an integro-differential approach. Nonlinearity 32(10), 3485–3522.

Giraud, A., I. Matic, O. Tenaillon, A. Clara, M. Radman, M. Fons, and F. Taddei (2001). Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut. science 291(5513), 2606–2608.

Good, B. H. and M. M. Desai (2015). The impact of macroscopic epistasis on long-term evolutionary dynamics. Genetics 85, 177–190.

Haldane, J. (1937). The effect of variation of fitness. The American Naturalist 71(735), 337–349.

Hamel, F., F. Lavigne, G. Martin, and L. Roques (2020). Dynamics of adaptation in an anisotropic phenotype-fitness landscape. Nonlinear Analysis: Real World Applications 54, 103107.

Helms, J. and M. Kaspari (2015). Reproduction-dispersal tradeoffs in ant queens. Insectes sociaux 62(2), 171–181.

Hermisson, J., O. Redner, H. Wagner, and E. Baake (2002). Mutation–selection balance: ancestry, load, and maximum principle. Theoretical population biology 62(1), 9–46.

Hoffmann, A. A. and M. J. Hercus (2000). Environmental stress as an evolutionary force. Bioscience 50(3), 217–226.

Jacod, J. and A. N. Shiryaev (2003). Limit Theorems for Stochastic Processes (Second ed.), Volume 288 of Grundlehren Der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin.

Kimura, M. (1964). Diffusion models in population genetics. Journal of Applied Probability 1(2), 177–232.

Kimura, M. and T. Maruyama (1966). The mutational load with epistatic gene interactions in fitness. Genetics 54(6), 1337.

Lande, R. (1975). The maintenance of genetic variability by mutation in a polygenic character with linked loci. Genetics Research 26(3), 221–235.

Lauring, A. S., J. Frydman, and R. Andino (2013). The role of mutational robustness in RNA virus evolution. Nature Reviews Microbiology 11(5), 327–336.

Lavigne, F., G. Martin, Y. Anciaux, J. Papaix, and L. Roques (2020). When sinks become sources: adaptive colonization in asexuals. Evolution 74(1), 29–42.

Liu, L. L., F. Li, W. Pao, and F. Michor (2015). Dose-dependent mutation rates determine optimum erlotinib dosing strategies for egfr mutant non-small cell lung cancer patients. PLoS One 10(11), e0141665.

Lorz, A., S. Mirrahimi, and B. Perthame (2011). Dirac mass dynamics in multidimensional nonlocal parabolic equations. Communications in Partial Differential Equations 36(6), 1071–1098.

Lynch, M. (2010). Evolution of the mutation rate. Trends in Genetics 26(8), 345–352.

Martin, G., S. F. Elena, and T. Lenormand (2007). Distributions of epistasis in microbes fit predictions from a fitness landscape model. Nature Genetics 39(4), 555.

Martin, G. and T. Lenormand (2006). The fitness effect of mutations across environments: a survey in light of fitness landscape models. Evolution 60(12), 2413–2427.

Martin, G. and T. Lenormand (2015). The fitness effect of mutations across environments: Fisher's geometrical model with multiple optima. Evolution 69(6), 1433–1447.

Martin, G. and L. Roques (2016). The non-stationary dynamics of fitness distributions: Asexual model with epistasis and standing variation. Genetics 204(4), 1541–1558.

Mirrahimi, S. and G. Raoul (2013). Dynamics of sexual populations structured by a space variable and a phenotypical trait. Theoretical population biology 84, 87–103.

Nathan, R. (2001). The challenges of studying dispersal. Trends in Ecology & Evolution 16(9), 481–483.

Novella, I. S., E. A. Duarte, S. F. Elena, A. Moya, E. Domingo, and J. J. Holland (1995). Exponential increases of RNA virus fitness during large population transmissions. Proceedings of the National Academy of Sciences 92(13), 5841–5844.

Patout, F. (2020). The cauchy problem for the infinitesimal model in the regime of small variance.

Perefarres, F., G. Thebaud, P. Lefeuvre, F. Chiroleu, L. Rimbaud, M. Hoareau, B. Reynaud, and J.-M. Lett (2014, Apr 22). Frequency-dependent assistance as a way out of competitive exclusion between two strains of an emerging virus. Proceedings Of The Royal Society B-Biological Sciences 281(1781), 20133374.

Raoul, G. (2017). Macroscopic limit from a structured population model to the kirkpatrick-barton model. arXiv preprint arXiv:1706.04094.

Redner, O. (2004). Discrete approximation of non-compact operators describing continuum-of-alleles models. Proceedings of the Edinburgh Mathematical Society 47(2), 449–472.

Roelly-Coppoletta, S. (1986). A criterion of convergence of measure-valued processes: Application to measure branching processes. Stochastics: An International Journal of Probability and Stochastic Processes 17(1-2), 43–65.

Roques, A. (2015). Processionary moths and climate change: an update, Volume 427. Springer.

Roques, L. (2013). Modèles de réaction-diffusion pour l'écologie spatiale. Editions Quae.

Roques, L., F. Patout, O. Bonnefon, and G. Martin (2020). Adaptation in general temporally changing environments.

Sanjuán, R. and P. Domingo-Calap (2016). Mechanisms of viral mutation. Cellular and molecular life sciences 73(23), 4433–4448.

Schoustra, S., S. Hwang, J. Krug, and J. A. G. de Visser (2016). Diminishing-returns epistasis among random beneficial mutations in a multicellular fungus. Proceedings of the Royal Society B: Biological Sciences 283(1837), 20161376.

Sharp, N. P. and A. F. Agrawal (2012). Evidence for elevated mutation rates in low-quality genotypes. Proceedings of the National Academy of Sciences 109(16), 6142–6146.

Slatkin, M. (1970). Selection and polygenic characters. Proceedings of the National Academy of Sciences 66(1), 87–93.

Smith, R., C. Tan, J. K. Srimani, A. Pai, K. A. Riccione, H. Song, and L. You (2014). Programmed allee effect in bacteria causes a tradeoff between population spread and survival. Proceedings of the National Academy of Sciences 111(5), 1969–1974.

Sniegowski, P. D. and P. J. Gerrish (2010). Beneficial mutations and the dynamics of adaptation in asexual populations. Philosophical Transactions of the Royal Society B: Biological Sciences 365(1544), 1255–1263.

Sniegowski, P. D., P. J. Gerrish, T. Johnson, and A. Shaver (2000). The evolution of mutation rates: separating causes from consequences. Bioessays 22(12), 1057–1066.

Stearns, S. C. (1989). Trade-offs in life-history evolution. Functional ecology 3(3), 259–268.

Taylor, P. (1991). Optimal life histories with age dependent tradeoff curves. Journal of theoretical biology 148(1), 33–48.

Tenaillon, O. (2014). The utility of Fisher's geometric model in evolutionary genetics. Annual Review of Ecology, Evolution, and Systematics 45, 179–201.

Trun, N. and J. Trempy (2009). Fundamental bacterial genetics. John Wiley & Sons.

Turelli, M. (2017, December). Commentary: Fisher's infinitesimal model: A story for the ages. Theoretical Population Biology 118, 46–49.

Turko, A. and P. Wright (2015). Evolution, ecology and physiology of amphibious killifishes (Cyprinodontiformes). Journal of Fish Biology 87(4), 815–835.

Van Harten, A. M. (1998). Mutation breeding: theory and practical applications. Cambridge University Press.

Wiser, M. J., N. Ribeck, and R. E. Lenski (2013). Long-term dynamics of adaptation in asexual populations. Science, 1364–1367.

Xiao, Z., Z. Zhang, and C. J. Krebs (2015). Seed size and number make contrasting predictions on seed survival and dispersal dynamics: A case study from oil tea camellia oleifera. Forest Ecology and Management 343, 1–8.