Adaptation of a population to a changing environment under the light of quasi-stationarity

Aurélien Velleret *

September 14, 2021

Abstract

We analyze the long-term stability of a stochastic model designed to illustrate the adaptation of a population to variation in its environment. A piecewise-deterministic process modeling adaptation is coupled to a Feller logistic diffusion modeling population size. As the individual features in the population become further away from the optimal ones, the growth rate declines, making population extinction more likely. Assuming that the environment changes deterministically and steadily in a constant direction, we obtain the existence and uniqueness of the quasi-stationary distribution, the associated survival capacity and the Q-process. Our approach also provides several exponential convergence results (in total variation for the measures). From this synthetic information, we can characterize the efficiency of internal adaptation (i.e. population turnover from mutant invasions). When the latter is lacking, there is still stability, but because of the high level of population extinction. Therefore, such a characterization must be based on specific features of this quasi-ergodic regime.

Keywords: mobile optimum, quasi-stationary distribution, evolution, ecology, jump processes, Markov process in continuous time and continuous space

1 Introduction

1.1 Eco-evolutionary motivations

Our objective is to study the relative contribution of mutations with various strong effects to the adaptation of a population. Our goal is therefore to analyze a model as simple as possible in which these mutations are filtered according to the advantage they provide. This advantage can be immediately significant (better growth rate of the mutant subpopulation) or play a role in the future adaptation (the population is doomed without mutants). The stochastic model considered takes into account these two aspects. It extends the one introduced by [KH09] and described more formally in [NP17] and [KNP18].

Similarly, we assume that the population is described by a certain value $\hat{x} \in \mathbb{R}^d$, hereafter referred to as its trait. For the sake of a simple theoretical model, spatial dispersion as well as phenotypic heterogeneity (at least for the individual features of interest) are neglected. We therefore that the population is monomorphic at all times and that $\hat{x}$ represents the phenotype of the individuals in the population. Nonetheless, we allow for variations of this trait $\hat{x}$ due to stochastic events, namely...
when a subpopulation issued from a mutant with trait $\hat{x} + w$ manages to persist and invade the "resident" population. In the model, such events are assumed to occur instantaneously.

The main novelty of our approach is that we couple this "adaptive" process with a Feller diffusion process $N$ with a logistic drift. This diffusion describes the dynamics of the population size in a limit where it is large. We mean here that individual birth and death events have a negligible impact, but that the accumulation of these events has a visible and stochastic effect. In particular, the introduction of the "size" in the model allows us to easily translate the notion of maladaptation, in the form of a poor growth rate.

For the long-time dynamics, we are mainly interested in considering only surviving populations, that is conditioning the process upon the fact that the population size has not decreased to 0. The implication of taking size into account is twofold. On the one hand, extinction occurs much more rapidly when adaptation is poor. Indeed, the population size is then very rapidly declining. So a natural selection effect can be observed at the population level. On the other hand, the better the adaptation, the larger the population size can be and the more frequent the birth of new mutants in the population. In our simple model, a mutant trait that is better suited for the survival of the population as a whole is also characterized by a greater probability that the resident population gets invaded, once a single mutant is introduced. Compared to the case of a fixed size as in [NP17] and [KNP18], this second implication means a stabilizing effect for the phenotype when the population size is large enough; but also a destabilizing effect when the population size decreases. This is in contrast to natural selection at the individual level (which is the main effect detailed in [KH09]). Indeed, when adaptation is already nearly optimal, very few among the mutants that appear in the population can successfully maintain themselves and eventually invade the resident trait.

Let us assume here that mutations can allow the individuals to survive in these new environments. In this context, how resilient is the population to environmental changes? Is there a clear threshold to the rate of change that such a population can handle? How can we describe the interplay between the above properties?

To begin to answer these questions, and like [KH09], we assume for simplicity that the environmental change is given by a constant speed translation of the profile of fitness, with $v$ this speed and $e_1$ the direction of this change. In practice, this means that the growth rate of the population at time $t$ is expressed as a function of $x := \hat{x} - v t e_1$, for a monomorphic population with trait $\hat{x}$ at time $t$. Naturally, the phenotypic lag $x$ becomes the main quantity of interest for varying $t$. Likewise, we can express as a function of $x$ and $w$ the probability that a mutant individual, with mutation $w$, will lead to the invasion of a resident population with trait $\hat{x}$ at time $t$. This probability should indeed be related to the difference between the growth rate at $x$ and at $x + w$, although we will not require any precise relationship in the results of this paper. Furthermore, we assume that the distribution of the additive effect for the new mutations is constant over time and independent of the trait $\hat{x}$ of the population before the mutation (thus independent of $x$ in the moving frame of reference).

In this context, we can exploit the notion of quasi-stationary distribution (QSD, cf Remark 2.2.3) to characterize what would be an equilibrium for these dynamics prior to extinction. The main contribution of the current paper is to ensure that this notion is unambiguously defined here. To the best of our knowledge, this is the first time that the existence and uniqueness of the QSD is proved for a piecewise deterministic process coupled to a diffusion.

By our proof, we also provide a justification of the notion of typical relaxation time and extinction time. The quasi-stationary description is well suited provided the latter is much longer than the former. As can be verified by simulations, typical convergence to the QSD is exponential in such cases. However, the marginal starting from certain initial conditions may take long before it approximates the QSD, mainly in cases where extinction is initially very likely.

In the following subsections of the introduction, we present the stochastic process under consideration then some elementary notations. The main results are described in Section 2 starting
with our hypothesis in Subsection 2.1 and the theorem statement in Section 2.2. In Subsection 2.3
we discuss its interpretation in terms of ecology and evolution. Its connection with related adap-
tation models is given in Subsection 2.4 and with the classical techniques of quasi-stationarity in
Subsection 2.5. The rest of the paper is devoted to proofs. We prove the existence and uniqueness
of the process in Section 3 and introduce in the next Section 4 the main theorems on which our
main Theorem 2.1 is based. Two alternative hypotheses are considered, with some variations in the
proofs. We choose to group the theorems in the three following sections according to the property
of the process they imply for the various sets of assumptions. The definition of a specific sigma-field
and its property are reported in the Appendix, as well as some illustrations of the asymptotic profiles
given by simulations.

1.2 The stochastic model
Following [KH09] as explained in the introduction for the definition of the adaptive component, the
system that describes the combined evolution of the population size and of its phenotypic lag is then
given by:

\[
\begin{align*}
X_t &= x - vt e_1 + \int_{[0,t] \times \mathbb{R}^d \times (\mathbb{R}^+)^2} \psi \varphi_0 (X_s- , N_s, w, u_f , u_g) M(ds, dw, du_f , du_g) \\
N_t &= n + \int_0^t \left( r(X_s) N_s - \gamma_0 \times (N_s)^2 \right) ds + \sigma \int_0^t \sqrt{N_s} dB_s,
\end{align*}
\]

where \( N_t \) describes the size of the population and \( X_t \) the phenotypic lag of this population.

Here, \( v > 0 \) is the speed of environmental change (in direction \( e_1 \)), \( B_t \) is a standard \( \mathcal{F}_t \) Brownian
motion and \( M \) is a Poisson Random Measure (PRaMe) on \( \mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^+ \), also adapted to \( \mathcal{F}_t \), with
intensity:

\[
\pi(ds, dw, du_f , du_g) = ds \nu(dw) du_f du_g,
\]

where \( \nu(dw) \) is a measure describing the distribution of new mutations, and:

\[
\varphi_0(x, n, w, u_f , u_g) = 1_{\{u_f \leq f_0(n)\}} \times 1_{\{u_g \leq g(x,w)\}}.
\]

The independence between \( M \) and \( B \) is automatically deduced from the following proposition.

**Proposition 1.2.1.** A Brownian Motion and a PRaMe that are adapted to the same filtration and
such that their increase after time \( t \) is independent from \( \mathcal{F}_t \) are necessarily independent.

**Proof of Proposition 1.2.1.** Thanks to Theorem 2.1.8 of [DiT13], if \( X_1, X_2 \) are additive func-
tionals and semi-martingales with respect to a common filtration, both starting from zero, and such
that their quadratic covariations \( [X_1, X_2] \) is a.s. zero, then the random vector \( (X_1(t) - X_1(s),
X_2(t) - X_2(s)) \) is independent of \( \mathcal{F}_s \), for every \( 0 \leq s \leq t \). Moreover, the vector \( (X_1, X_2) \) of additive processes
is independent.

Note \( B \) the Brownian Motion and \( M \) the PRaMe on \( \mathbb{R}^+ \times \mathcal{X} \) For any test function \( F : \mathcal{X} \mapsto \mathbb{R},
\) define \( Z(t) := \int_{[0,t] \times \mathcal{X}} F(x) M(ds, dx) \). Both \( Z \) and \( B \) are additive functionals and semi-martingales
with respect to the filtration \( \mathcal{F}_t \), both starting from zero. \( Z \) being a jump process and \( B \) continuous,
their quadratic covariation equals a.s. 0. Since it applies to any \( F \), exploiting Theorem 2.18 of
[DiT13] implies that \( B \) and \( M \) are independent. \( \square \)

In the model of the moving optimum originally considered in [KH09], \( X = 0 \) corresponds to the
optimal state in terms of some reproductive value function \( R(x) \), for \( x \in \mathbb{R} \). This function \( R \) is also
assumed to be symmetrical and decreasing with $|X|$. Here we consider a possibly multidimensional state space for $X$ and will usually not require any assumption on the related function $g$.

$X$ is described as the phenotypic lag because $X_t + v t e_1$ is the character of the individuals at time $t$ in the population while in this original model, the mobile optimum is located at trait $v t e_1$. These assumptions on the fitness landscape are natural, and we abide by them in our simulations. Nonetheless, they are mainly assumed for simplicity and we have chosen here to be as general as possible in the definition of $r$. $X_t$ is thus a lag as compared to the trait $v t e_1$ that is merely a reference value.

$g(X_t, w)$ is the mutation kernel that describes the rate of fixation at which a mutant subpopulation of trait $X_t + v t e_1 + w$ invades a resident population of trait $X_t + v t e_1$. Although the rate at which the mutations occur in one individual can reasonably assumed to be symmetrical in $w$, it is clearly not the case for $g$. In a large population, the filtering of considering only fixing mutations highly restricts the occurrence of strongly deleterious mutations, strongly favors strongly advantageous mutations. For mutations with little effects, there is only a slight bias. To cover both of these situations, we consider in our analysis both the case where any mutation effect is permitted and the case where only advantageous ones are. Although the latter case will raise more difficulty in terms of accessibility of the domain, the core of the argument is quite the same and the simulations seem to provide similar results in both cases.

The term $f(N_t)$ is introduced to model the fact that for a constant mutation rate by individual, the mutation rate for the population is all the larger than the population size is large. $f(N_t) := N_t$ is the first reasonable choice, but we may also be interested in introducing an effect of the population size in the fixation rate.

$N$ follows the equation of a Feller logistic diffusion where the growth rate $r$ at time $t$ only depends on $X_t$, while the strength of competition $c$ and the coefficient of diffusion $\sigma$ are kept constant. Such a process is the most classical ones for the dynamics of a large population size in a continuous space setting and such that explosion is prevented. It is described in [La05] (with fixed growth rate), notably as a limit of some individual-based model. $\sigma$ is related to the proximity between to uniformly sampled individuals in terms of their filiation links: $1/\sigma^2$ scales as the population size and is sometimes describes as the "effective population size".

From a biological perspective, $X$ has no reason to explode. Under our assumption [H11] below, such explosion is clearly prevented. Yet, we won’t focus on conditions ensuring non-explosion for $X$. Indeed, it would mean (by assumption [H8] below) that the growth rate becomes extremely negative. It appears very natural to consider that it would lead to the extinction of the population. So, we define the extinction time as:

$$\tau_0 := \inf\{t \geq 0 ; N_t = 0\} \land \sup\{k \geq 1\} T^k_X,$$

where $T^k_X := \inf\{t \geq 0 ; \|X_t\| \geq k\}$.

Because it simplifies many of our calculations, in the following, we will consider $Y_t := \frac{2}{\sigma} \sqrt{N_t}$ rather than $N_t$.

**Fact 1.2.2.** With the previous notations, $(X, Y)$ satisfies the following SDE:

$$\begin{cases}
X_t = x - v t e_1 + \int_{[0,t] \times \mathbb{R}^d \times \mathbb{R}_+^2} w \varphi(X_s, Y_s, w, u_f, u_g) \ M(ds, dw, du_f, du_g), \\
Y_t = y + \int_0^t \psi(X_s, Y_s) \ ds + B_t,
\end{cases}$$

where we define:

$$\psi(x, y) = -\frac{1}{2y} + \frac{r(x)}{2} - \gamma y^3,$$

with $\gamma := \frac{\gamma_0 \sigma^2}{8}$,

$$\varphi(x, y, w, u_f, u_g) := \varphi_0 \left(x, \sigma^2 y^2/4, w, u_f, u_g\right).$$

Thus with $f(y) := f_0[\sigma^2 y^2/4]$, $\varphi(x, y, w, u_f, u_g) = 1_{\{u_f \leq f(y)\}} \times 1_{\{u_g \leq g(x, w)\}}$. 






An elementary application of the Ito formula proves this fact.

The aim of the following theorems is to describe the law of the marginal of the process \((X, Y)\) at large time \(t\) conditionally upon the fact that the extinction has not occurred, in short the MCNE at time \(t\). Considering the conditioning at the current time leads to considering properties of quasi-stationarity; while a conditioning at a much more future time leads to a Markov process usually referred to as the Q-process, in some sense the process conditioned on never going extinct. The two aspects are clearly complementary and our approach will treat both in the same framework, in the spirit initiated by [CV16].

1.3 Elementary notations

In the following, the notation \(k \geq 1\) is to be understood as \(k \in \mathbb{N}\) while \(t \geq 0\) –resp. \(c > 0\)– should be understood as \(t \in \mathbb{R}_+ := [0, \infty)\) –resp. \(c \in \mathbb{R}_+^* := (0, \infty)\). In this context (with \(m \leq n\)), we denote classical sets of integers by: \(\mathbb{Z}_+ := \{0, 1, 2, \ldots\}\), \(\mathbb{N} := \{1, 2, 3, \ldots\}\), \([m, n] := \{m, m+1, \ldots, n-1, n\}\), where the notation := makes explicit that we define some notation by this equality. For maxima and minima, we usually denote: \(s \vee t := \max\{s, t\}\), \(s \wedge t := \min\{s, t\}\). Accordingly, for a function \(\varphi\), \(\varphi^\wedge\) –resp. \(\varphi^\vee\)– will be the notation for a lower-bound –resp. for an upper-bound– of \(\varphi\). \(C^0(X, Y)\) denotes the set of continuous functions from any \(X\) to any \(Y\). \(\mathcal{B}(X)\) is the set of bounded functions from any \(X\) to \(\mathbb{R}\). \(\mathcal{M}(X)\) and \(\mathcal{M}_1(X)\) denote the sets of resp. positive measures and probability measures on any state space \(X\). Numerical indices are rather indicated in superscript, while specifying notations are often in subscript. By notation, \(\{y \in c : A(y)\ \& \ B(y)\}\) denotes the set of values \(y\) of \(c\) such that both \(A(y)\) and \(B(y)\) hold true. Likewise, for two probabilistic conditions \(A\) and \(B\) on \(\omega \in \Omega\), and a r.v. \(X\), we may use \(E(X ; A, B)\) instead of \(E(X1_{\Gamma})\) where \(\Gamma := \{\omega \in \Omega : A(\omega) \\& \ B(\omega)\}\).

2 Exponential convergence to the QSD

2.1 Hypothesis

We will consider two different sets of assumptions, including or rejecting the possibility for deleterious mutations to invade the population.

First, the following set \((H)\) of assumptions can always be assumed throughout the paper, although some assumptions may be mentioned as not involved.

\([H1]\) \(f \in C^0(\mathbb{R}_+^*, \mathbb{R}_+)\) is positive.

\([H2]\) \(g \in C^0(\mathbb{R}^d \times \mathbb{R}_+^*, \mathbb{R}_+)\) and is bounded on any \(K \times \mathbb{R}^d\), where \(K\) is a compact set of \(\mathbb{R}^d\).

\([H3]\) \(r\) is locally Lipschitz-continuous on \(\mathbb{R}^d\) and \(r(x)\) tends to \(-\infty\) as \(\|x\|\) tends to \(\infty\).

\([H4]\) \(\nu(\mathbb{R}^d) < \infty\). Moreover, there exist \(\theta, \nu_\Lambda > 0\) and \(\eta \in (0, \theta)\) such that:

\[\nu(dw) \geq \nu_\Lambda 1_{B(\theta+\eta)\setminus B(\theta-\eta)} dw,\]

where \(B(R)\), for \(R > 0\), denotes the open ball of radius \(R\) centered at the origin.

\([H5]\) provided \(d \geq 2\), \(\nu(dw) \ll dw\) and the density \(g(x, w) \nu(w)\) (for a jump from \(x\) to \(x + w\)), of the jump size law w.r.t. Lebesgue’s measure, satisfies:

\[
\forall x_\nu > 0, \quad \sup \left\{ \frac{g(x, w) \nu(w)}{\int_{\mathbb{R}^d} g(x, w') \nu(w') dw'} ; \ |x| \leq x_\nu, w \in \mathbb{R}^d \right\} < \infty.
\]

5
When we allow deleterious mutations to invade the population, we actually mean that the rate of invasion is always positive, leading to the following assumption:

\[(D) \quad g \text{ is positive.}\]

Otherwise, we consider the case where deleterious mutations are forbidden, in the sense the rate is zero for mutations that would induce an increase of \(\|X\|\). The invasion rate of advantageous mutations is however still assumed to be positive. This is stated in the following assumption \((A)\) as the alternative to \((D)\).

\[(A) \quad \text{For any } x, w \in \mathbb{R}^d, \quad \|x + w\| < \|x\| \text{ implies } g(x, w) > 0 \quad \|x + w\| \geq \|x\| \text{ implies } g(x, w) = 0.\]

Remarks 2.1.1. * For \(d = 1\), no condition on the density of \(g \times \nu\) as in \([H5]\) is required.

* It is quite natural to assume that \(f(0) = 0\) and \(f(y)\) tends to \(\infty\) as \(y\) tends to \(\infty\), but we will not need those assumptions.

* 1 is the natural bound with the above-mentioned biological interpretation of \([H2]\). Yet an extension can be introduced where \(g\) is not exactly the fixation probability, cf. Corollary 2.2.6.

* Under \([H2]\) and \([H4]\) (since \(v(\mathbb{R}^d) < \infty\)), over any finite time-interval, only a finite number of mutations can occur. We also need lower-bounds on the probability of specific events which roughly prescribe the dynamics of \(X\). This is where the lower-bound on the density of \(v\) is exploited as well as the positivity of \(g\), deduced from either Assumption \((D)\) or \((A)\).

* The fact that \(r(x)\) tends to \(-\infty\) as \(x\) tends to \(\infty\) makes it easier to prove that the process remains mostly confined, say in the time-interval \([0, t]\) under the condition that \(t < \tau_\vartheta\). We would be able to state an explicit value \(r_\lambda\) depending on the other parameters so that the proof holds while assuming that the limsup of \(r(x)\) is upper-bounded by \(-r_\lambda\) when \(\|x\|\) tends to infinity (instead of being necessarily equal to \(-\infty\)).

2.2 Statement of the main theorems

First, we need to ensure that the model specified by equation \((S)\) properly defines a unique solution, which is stated in the next Proposition.

Proposition 2.2.1. Assume that Assumption \((H)\) holds. Then, for any initial condition \((x, y) \in \mathbb{R}^d \times \mathbb{R}_+\), there is a unique strong solution \((X_t, Y_t)_{t \geq 0}\) in the Skorokhod space satisfying \((S)\) for any \(t < \tau_\vartheta\), and \(X_t = Y_t = 0\) for \(t \geq \tau_\vartheta\), where \(\tau_\vartheta := \sup_{\{n \geq 1\}} T^Y \land \sup_{\{n \geq 1\}} T^X\), \(T^Y := \inf\{t \geq 0, Y_t \leq 1/n\}\), \(T^X := \inf\{t \geq 0, \|X_t\| \geq n\}\).

Remarks 2.2.2. This proposition makes it possible to express \(\tau_\vartheta\) as \(\inf\{t \geq 0, Y_t = 0\}\).

We exploit the notion of uniform exponential quasi-stationary convergence as previously introduced in \([Ve21b]\).

Definition 1. For any linear and bounded semi-group \((P_t)_{t \geq 0}\) acting on a Polish state space \(Z\), we say that \(P\) displays a uniform exponential quasi-stationary convergence with characteristics \((\alpha, h, \lambda) \in \mathcal{M}_1(Z) \times B(Z) \times \mathbb{R}\) if \((\alpha \| h) = 1\) and there exists \(C, \gamma > 0\) such that for any \(t > 0\) and for any measure \(\mu \in \mathcal{M}(Z)\) with \(\|\mu\|_{TV} \leq 1\):

\[\|e^{\lambda t} \mu P_t(ds) - (\mu \| h)\alpha(ds)\|_{TV} \leq Ce^{-\gamma t}.\]  

(2.1)

Remarks 2.2.3. * As shown in Fact 2.2.2 of \([Ve21b]\) it implies that for any \(t > 0\), \(\alpha P_t(ds) = e^{-\lambda t} \alpha(ds)\). Any measure satisfying this property is called a quasi-stationary distribution.
It is elementary that $h_1 : x \mapsto e^{\lambda t} P_t (\delta_z 1)$ converge in the uniform norm to $h$. We call $h$ the survival capacity because $e^{\lambda t} P_t (\delta_z 1) = \mathbb{P}_x (t < \tau_0) / \mathbb{P}_x (t < \tau_0)$ enables to compare the likelihood of survival with respect to the initial conditions.

Since $h_{t+h'} = e^{\lambda t} P_t h'$, one can then easily deduce that $e^{\lambda t} P_t h = h$. It is also obvious that $h$ is necessarily non-negative.

* By “characteristics”, we express that they are uniquely defined.

Our main theorem is stated as follows, with $Z := \mathbb{R}^d \times \mathbb{R}^+_*$:

**Theorem 2.1.** Assume that Assumption (H) holds. Suppose that either (D) or (A) holds. If $d \geq 2$, assume finally that [H5] holds. Then, the semi-group $P$ associated to the process $Z := (X, Y)$ and extinction at time $\tau_0$ displays a uniform exponential quasi-stationary convergence with some characteristics $(\alpha, h, \lambda) \in M_1 (Z) \times B(Z) \times \mathbb{R}_+$. Moreover, $h$ is positive.

**Remarks 2.2.4.** We refer to Corollary 2.2.1 in [Ve21b], for the implied convergence result of the renormalized semi-group to $\alpha$. The fact that $h$ is positive implies that there is no other QSD in $M_1 (Z)$.

In [Ve21b] is also provided an analysis of the so-called Q-process, whose properties are as follow:

**Theorem 2.2.** Under the same assumptions as in Theorem 2.1, with $(\alpha, h, \lambda)$ the characteristics of exponential convergence of $P$, the following properties hold:

(i) **Existence of the Q-process:**
There exists a family $(Q_{x,y})_{x,y} \in Z$ of probability measures on $\Omega$ defined by:

$$\lim_{t \to \infty} \mathbb{P}_{x,y} (\Lambda_s \mid t < \tau_0) = Q_{x,y} (\Lambda_s),$$

for all $\mathcal{F}_s$-measurable set $\Lambda_s$. The process $(\Omega; (\mathcal{F}_t)_{t \geq 0}; (X_t, Y_t)_{t \geq 0}; (Q_{x,y})_{x,y} \in Z)$ is an $Z$-valued homogeneous strong Markov process.

(ii) **Weighted exponential ergodicity of the Q-process:**

The measure $\beta (dx, dy) := h(x, y) \alpha (dx, dy)$ is the unique invariant probability measure under $\mathcal{Q}$. Moreover, for any $\mu \in M_1 (Z)$ satisfying $\mu (1/h) < \infty$ and $t \geq 0$:

$$\| \mathcal{Q} \mu (X_t, Y_t) \in (dx, dy) - \beta (dx, dy) \|_{TV} \leq C \| \mu - \mu (1/h) \beta \|_{1/h} e^{-\gamma t},$$

where $\mathcal{Q} \mu (dw) := \int_Z \mu (dx, dy) Q_{x,y} (dw)$, $\| \mu (1/h) \beta \|_{TV} = \| \mu (dx, dy) / h(x, y) \|_{TV}$.

**Remarks 2.2.5.**  
* For the total variation norm, considering $(X, Y)$ or $(X, N)$ is equivalent.

* The constant $\mu (1/h)$ in (2.3) is optimal up to a factor 2, in the sense that for any $u > 0$:

$$\| \mu - u \alpha \|_{1/h} \geq \| \mu - \mu (1/h) \beta \|_{1/h} / 2$$

(cf Remark 2.2.5 of [Ve21b]).

* Since $v$ tends to $-\infty$ as $\| x \|_2$ tends to infinity, it is natural to assume that mutations leading $X$ to be large have a very small probability of fixation. Notably, it means that we highly expect the upper-bound of $g$ in [H2], uniform over $w$.

* Under hypothesis (A), one may expect the real probability of fixation $g(x, w)$ to be at most of order $O(||w||)$ for small values of $w$ (and locally in $x$). In such a case, we can allow $\nu$ to satisfy a smaller integrability condition than [H4] while forbidding observable accumulation of mutations.

**Corollary 2.2.6.** Assume that Assumption (H) and (A) hold. Suppose that $\int_{\mathbb{R}} (|w| \wedge 1) \nu (dw) < \infty$ while $\bar{g} : (x, w) \mapsto g(x, w)/(|w| \wedge 1)$ is bounded on any $K \times \mathbb{R}^d$ for $K$ a compact set of $\mathbb{R}^d$. If $d \geq 2$, assume additionally that [H5] holds. Then, the conclusions of Theorem 2.1 and Theorem 2.2 hold true.
Proof of Corollary 2.2.6. \((X, Y)\) is solution of \((S)\) iff it is solution of:

\[
\begin{align*}
X_t &= x - vt e_1 + \int_{[0,t] \times \mathbb{R}^d \times \mathbb{R}^+} w \tilde{\varphi} (X_{s^-}, Y_s, w, u_f, \tilde{u}_g) \tilde{M}(ds, dw, du_f, du_g), \\
Y_t &= y + \int_0^t \psi (X_s, Y_s) ds + B_t,
\end{align*}
\]

where \(\tilde{M}\) is a PRAme of intensity \(ds \tilde{\nu}(dw) du_f \tilde{d}u_g\),

\[
\tilde{\nu}(dw) := \nu(dw)/(|w| \wedge 1),
\]

\[
\tilde{\varphi}(x, y, w, u_f, \tilde{u}_g) = \varphi(x, y, w, u_f, \tilde{u}_g \times (|w| \wedge 1)),
\]

where \(\tilde{\varphi}\) is defined as \(\varphi\) with \(g\) replaced by \(\tilde{g}\).

Thanks to the condition on \(\nu\), \([H4]\) holds with \(\tilde{\nu}\) instead of \(\nu\). Thanks to the condition on \(g\), \([H2]\) still holds with \(\tilde{g}\) instead of \(g\). Conditions \(A\) and \([H5]\) are equivalent for the systems \((g, \nu)\) and \((\tilde{g}, \tilde{\nu})\). Consequently, if we prove Theorem 2.1 and Theorem 2.2 with \([H2]\) and \([H4]\), the results follow under the assumptions of Corollary 2.2.6 \(\square\)

2.3 Eco-evolutionary implications of these results

One of the major motivations for the present analysis is to make a distinction, as rigorous as possible, between an environmental change to which the population can spontaneously adapt to and a change that imposes too much a pressure. We recall that in [NP17], the authors obtain a clear and explicit distinction between an environmental change to which the population can spontaneously adapt to and a change that imposes too much a pressure. We note that in [NP17], the authors obtain a clear and explicit distinction between an environmental change to which the population can spontaneously adapt to and a change that imposes too much a pressure. Thus, it might seem that imposes too much a pressure. We recall that in [NP17], the authors obtain a clear and explicit distinction between an environmental change to which the population can spontaneously adapt to and a change that imposes too much a pressure. Thus, it might seem that imposes too much a pressure.

One of the major motivations for the present analysis is to make a distinction, as rigorous as possible, between an environmental change to which the population can spontaneously adapt to and a change that imposes too much a pressure. We recall that in [NP17], the authors obtain a clear and explicit distinction between an environmental change to which the population can spontaneously adapt to and a change that imposes too much a pressure. Thus, it might seem that imposes too much a pressure. We recall that in [NP17], the authors obtain a clear and explicit distinction between an environmental change to which the population can spontaneously adapt to and a change that imposes too much a pressure. Thus, it might seem that imposes too much a pressure.

In fact, the reason why this threshold is so distinct in [NP17] is that their model is based on the following underlying assumption: The poorer the current adaptation is, the more efficiently mutations are able to fix, provided that they are then beneficial. In our case, a population that is too poorly adapted is almost doomed to a rapid extinction, because the population size cannot be maintained at large values. Instead, long-term survival is triggered by dynamics that maintain the population adapted. Looking back at the history of surviving populations, it means that we are likely to observe that the process has mostly remained confined outside of deadly areas.

In order to establish this distinction between environmental changes that are sustainable and those that endanger the population, we need a criterion that quantifies the stability of such core regions. Our results provide two exponential rates whose comparison is enlightening: if the extinction rate is of the same order as the convergence rate or larger, it means that the dynamics is strongly dependent upon the initial condition. If the convergence is much faster, the dynamics shall rapidly become similar regardless the initial condition. This is at least the case for initial conditions that are not too risky (i.e. where \(h\) is not too small). This criterion takes into account the intrinsic sustainability of the mechanisms involved in the adaptation to the current environmental change, but does not involve the specific initial state of adaptation.

Looking at the simulation results, the convergence in total variation indeed appears to happen at some exponential rate, provided that extinction does not abruptly wipe out a large part of the distribution at a given time. To obtain a generic estimate of the exponential rate at which the effect of the initial condition is lost, the decay in total variation appears however computationally expensive and not very meaningful. Although they are not as clearly justified, it seems more practical to exploit the decay in time of the correlations of \(X\) and/or \(N\) starting from the QSD profile. It does not seem very difficult to compare the extinction rate from this estimate. This is especially true in the case where \(X\) is of dimension one, as one can directly estimate the dynamics of the density
and thus the extinction rate. Furthermore, it is quite reassuring to see that the inclusion or not deleterious mutations (for which the invasion probability is expected be positive but very small) is not crucial in the present proof. We do not see much difference by looking at the simulations.

Much more can be said by looking at the simulation estimates of the quasi-stationary (QSD), the quasi-ergodic distribution (QED), and the survival capacity. It is planned to detail these simulation results in a later article, but let us already give some insights into the comparison between the QSDs and the QEDs provided in Appendix B. We see that although the QSDs look very different at the three different values of mutation rates, the QEDs are in fact very similar. When extinction plays a notable role, a tail appears on the QSD from the area of concentration of the QED to an area where the population size is close to zero. The shape of the tail and the fact that it does not appear on the QED nor for larger mutation rate suggests that it corresponds in some sense to a path towards a rapid extinction. These regions are clearly more unstable than the core areas where the Q-process remains confined. This is probably due to this decline in population size when the level of maladaptation becomes more pronounced. This confinement by the conditioning upon survival only weakens in the recent past. Conditions most likely leading to extinction are allowed provided the delay is sufficiently large before extinction actually occurs.

2.4 Quasi-ergodicity of related models

The current paper completes the illustrations given in Subsection 4.2 of Ve21a and Sections 4-5 of Ve21b. If the model of the current paper was in fact the original motivation for the techniques presented in these two papers, we can focus more closely on each of the difficulties thanks to these various illustrations. In any of them, the adaptation of the population to its environment is described by some process $X$ solution to some SDE of the form:

$$X_t = x - \int_0^t V_s \, ds + \int_0^t \Sigma_s \cdot dB_s + \int_{[0,t] \times \mathbb{R}^d \times \mathbb{R}_+} w \mathbf{1}_{\{u \leq U_s(w)\}} M(ds, dw, du),$$

where $B$ is a $\mathcal{F}_t$-adapted Brownian Motion and $M$ a $\mathcal{F}_t$-adapted PRaMe. $V_s$ and $\Sigma_s$ a priori depend on $X_s$, $U_s$ on $X_{s-}$ and possibly on a coupled process $N_t$ describing the population size. Like the product $f(Y_t) g(X_{t-}, w)$ in equation (S), one specifies in $U_t(w)$ the rate at which a mutations of effect $w$ invades the population at time $s$. $V_s$ both relates to the speed of the environmental change and to the mean effects of the mutations invading the population at time $s$ in a limit of very frequent mutations of very small effects. $\Sigma_s$ then relates to the undirected fluctuations both of the environment and in the effects of this large number of small fixating mutations.

We can relate the current coupling of $X$ and $N$ to an approximation given by the autonomous dynamics of a process $Y$ similar to $X$. For the approximation to be as valid as possible, the law of $Y$ should be biased by some extinction rate (depending at time $t$ on the value $Y_t$) and its jump rate should be adjusted. By these means, we would take an implicit account of what would be the fluctuations of $N$ if $X$ would be around the value of $Y_t$. This approximation is particularly reasonable when the characteristic fluctuations of $N$ around its quasi-equilibrium are much quicker than the effect of the growth rate changing over time with the adaptation. Its validity is less clear when the extinction has a strong effect on the establishment of the quasi-equilibrium.

The exponential quasi-stationary convergence is treated in Subsection 4.2 of Ve21a for a coupling $(X, N)$ that behaves as an elliptic diffusion, while Sections 4 and 5 of Ve21b deal with some cases of a biased autonomous process $Y$ that behaves as a piecewise-deterministic process. For such a process with jumps, it is manageable yet technical to deal with restrictions on the allowed directions or sizes of jumps, while imposing $V_t$ to stay at zero actually makes the proof harder than choosing $V_t := v \times t$. While the proofs of (A1) and (A3) highly depends on such local properties of the dynamics, the ones of (A2) for these semi-groups rely on a common intuition. Although we allow $X$ to live in an unbounded domain, the maladaptation of the process when it is far from the optimal position...
constraints $X$ to stay confined conditionally upon survival. This effect of the maladaptation has been modeled either directly on the growth rate of the coupled process $N_t$ or with some averaged description in terms of extinction rate. Such confinement property for the coupled process is in fact the main novelty of [Ve21a] and notably illustrated in Subsection 4.2.4. For simplicity, we have dealt there with a locally elliptic process, for which the Harnack inequality is known to greatly simplify the proof, as observed previously for instance in [CV21]. The proof of this confinement is actually simpler with $Y$ behaving as an autonomous process under the pressure of a death rate going to infinity outside of compact sets. The proof in this case is naturally adapted from the proof of (A2) given in Subsection 4.1.2 of [Ve21a].

Assume for now that the fluctuations of $N$ are much quicker than the change of the growth rate in the domain where the population is well-adapted. Then we conjecture that considering the autonomous process $Y$ (including the bias by the extinction rate) instead of the coupled process $(X, N)$ would produce very similar results: the extinction rates and the rates of stabilization to equilibrium should be close between these models, while the QSD profile of $X$ should be similar to the one of $Y$.

The drop in the quality of the approximation when extinction has a crucial contribution must have a quite limited effect for our concern, which is to compare the extinction rate to the rate of stabilization to equilibrium, see Subsection 2.3. Indeed, as long as the extinction rate is not way larger than the rate of stabilization to equilibrium, such domains of maladaptation are strongly avoided when looking in the past of surviving populations. On the other hand, the population is almost doomed when it enters these domains, so that we should be able to neglect the contribution to the extinction rate of the dynamics of the process there.

2.5 The mathematical perspective on quasi-stationarity

The subject of quasi-stationarity is now quite vast and a considerable literature is dedicated to it, as suggested by the bibliography collected by Pollett [Po15]. Some insights into the subject can be found in general surveys like [CMS13], [DP13] or more specifically for population dynamics in [MV12]. However, it appears that that much remains to be done for the study of strong Markov processes both on a continuous space and in a continuous time, without any property of reversibility. For general recent results, besides [Ve21a] and [Ve21b] that we exploit, we refer to [CV18b], [BCGM19], [CG20], [FRS20] or [GNW20]. The difficulty is increased when the process is discontinuous (because of the jumps in $X$) and multidimensional, since the property of reversibility becomes all the more stringent and new difficulties arise (cf e.g. Appendix A of [CCM17]).

Thus, ensuring the existence and uniqueness of the QSD is already some breakthrough, and we are even able to ensure an exponential rate of convergence in total variation to the QSD and similar results on the Q-process. This model is in fact a very interesting illustration of the new technique which we exploit. Notably, we see how conveniently our conditions are suited for exploiting the Girsanov transform as a way to disentangle couplings (here between $X$ and $N$, that are respectively the evolutionary component and the demographic one).

Our approach relies on the general result presented in [Ve21b], which, as a continuation of [Ve21a], has been originally motivated by this problem. In [Ve21a], the generalization of Harris recurrence property at the core of the results of [CV16] is extended to deal with exponential convergence which are not uniform with respect to the initial condition. The fine control over the MCNE has opened the way for the approach developed in [Ve21b] to deal with continuous-time and continuous-space strong Markov processes with discontinuous trajectories.

After their seminal article [CV16], these same authors have obtained quite a number of extensions, for instance with multidimensional diffusions [CCV18], inhomogeneous in time processes [CV18a], and various examples of processes in a countable space notably with the use of Lyapunov functions,
Theorem 1.1 in [BCGM19]). Yet, in the approach of [CV18b] for continuous-time and continuous
space Markov process, the rather abstract assumption (F3) appears tightly bound to the Harnack
inequality. The similar Assumption (A4) in [BCGM19] is also left without further guidance, while
the assumption of a strong Feller property in [FRS20] and [GNW20] appears too restrictive. For
discontinuous processes, these two properties generally do not hold true, which is what motivated
us to look for an alternative statement in [Ve21a] or [CV18b]. Exploiting the result of [CV18b],
it may be possible to ensure the properties of exponential quasi-ergodicity for such a discontinuous process as the one of this article, keeping a certain dependence on the initial condition. At least, the conditions they provide as well as the ones from [BCGM19] are necessarily implied by our convergence result (cf Theorem 2.3 of [CV20]
and Theorem 3) appears tightly bound to the Harnack inequality. Yet, in the approach of [CV18b] for continuous-time and continuous
space Markov process, the rather abstract assumption (F3) appears tightly bound to the Harnack
inequality. The similar Assumption (A4) in [BCGM19] is also left without further guidance, while
the assumption of a strong Feller property in [FRS20] and [GNW20] appears too restrictive. For
discontinuous processes, these two properties generally do not hold true, which is what motivated
us to look for an alternative statement in [Ve21a] or [CV18b]. Exploiting the result of [CV18b],
it may be possible to ensure the properties of exponential quasi-ergodicity for such a discontinuous process as the one of this article, keeping a certain dependence on the initial condition. At least, the conditions they provide as well as the ones from [BCGM19] are necessarily implied by our convergence result (cf Theorem 2.3 of [CV20]
and Theorem 3) appears tightly bound to the Harnack inequality. Yet, in the approach of [CV18b] for continuous-time and continuous
space Markov process, the rather abstract assumption (F3) appears tightly bound to the Harnack
inequality. The similar Assumption (A4) in [BCGM19] is also left without further guidance, while
the assumption of a strong Feller property in [FRS20] and [GNW20] appears too restrictive. For
discontinuous processes, these two properties generally do not hold true, which is what motivated
us to look for an alternative statement in [Ve21a] or [CV18b].

3 Proof of Proposition 2.2.1

Uniqueness:
Step 1: A priori upper-bound on the jump rate.
Assume that we have a solution \((X_t, Y_t)_{t \leq T}\) to (S) until some (stopping) time \(T\) (i.e. for any
\(t < T\)) satisfying \(T \leq t \vee T^\nu \vee T^R\) for some \(t \vee > 0, m, n \geq 1\) (see Equation \([3.1] \)). We know from
[H3] that the growth rate of the population remains necessarily upper-bounded by some \(\nu\) until
\(T\). Thus, we deduce a stochastic upper-bound \((Y_t, Y_t)_{t \geq 0}\) on \(Y\):

\[
Y_t = y + \int_0^t \psi(Y_s) ds + B_t \quad \text{where} \quad \psi(y) = -\frac{1}{2} y + \frac{\nu}{2} y - \gamma y^3, \tag{3.1}
\]

which is thus independent of \(M\). Since \(\psi(y) \leq \nu y/2\), it is classical that \(Y\) and a fortiori \(Y\)
cannot explode before \(T\), see for instance Lemma 3.3 in [BM13] or [La05] where such a process is
described in detail.

Under \([H2]\), the jump rate of \(X\) is uniformly bounded until \(T\) by:

\[
\nu(\mathbb{R}^d) \times \{g(x', w) : x' \in \bar{B}(0, n), w \in \mathbb{R}^d\} \times \sup\{f(y') : y' \leq \sup_{s \leq t} Y_s\} \times \infty a.s.
\]

Step 2: Identification until \(T\).
In any case, this means that the behavior of \(X\) until \(T\) is determined by the value of \(M\) on a
(random) domain associated to an a.s. finite intensity. Thus, we need a priori to consider only a
finite number \(K\) of "potential" jump, that we can describe as the points \((T_j^i, W^i, U^i_j, U^j_0)_{i \leq K}\) in the
increasing order of the times \(T_j^i\).

From the a priori estimates, we know that for any \(t < T_j^i \wedge T\) : \(X_t = x - v_t\). By the improper
notation \(t < T_j^i \wedge T\), we mean \(t < T_j\) if \(K > 1\) (since \(T_j < T\) by construction) and \(t < T\) if \(K = 0\), i.e.
when there is no potential jump before \(T\). We then consider the solution \(\hat{Y}\) of:

\[
\hat{Y}_t = y + \int_0^t \psi(x - v_s, \hat{Y}_s) ds + B_t.
\]

It is not difficult to adjust the proof of [YW71] to this time-inhomogeneous setting, with \([H3]\),
so as to prove the existence and uniqueness of such a solution until any stopping time \(T \leq \tau_0\), where
\(\tau_0 := \inf\{t \geq 0, Y_t = 0\}\). Besides, \(\hat{Y}\) is independent of \(M\) and must coincide with \(Y\) until \(T\) if \(K = 0\), i.e.
when there is no potential jump before \(T\), i.e. \(K = 0\), we have identified \((X_t, Y_t)\) for \(t \leq T\) as \(X_t = x - v_t, Y_t = \hat{Y}_t\). Else, at time \(T_j^i\), we check
whether \( U_1 \leq f(T_1) \) and \( U_1 \leq g(x - v T_1) \), \( W^1 \). If it holds, necessarily \( X(T_1) = x - v T_1 + W^1 \), else \( X(T_1) = x - v T_1 \). Doing the same inductively for the following time-intervals \([T^k_j, T^{k+1}_j] \), we identify the solution \((X, Y)\) until \( T \).

**Step 3: Uniqueness of the global solution.**

Now, consider two solutions \((X, Y)\) and \((X', Y')\) of \((S)\) defined up to respectively \( \tau_0 \) and \( \tau'_0 \) as in Proposition \[2.2.1\] with in addition \( X_t = Y_t = 0 \) for \( t \geq \tau_0 \), and \( X'_t = Y'_t = 0 \) for \( t \geq \tau'_0 \).

On the event \( \{\sup_m T^m = \tau_0 \land \tau'_0 \} \), we deduce by continuity of \( Y' \) that \( T^m = T'^m \) so that \( \tau_0 = \tau'_0 \). On the event \( \{\sup_n T^n = \tau_0 \leq \tau'_0 < \infty \} \), for any \( n \) and \( t_\nu > 0 \) there exists \( m \geq 1 \) and \( n' \geq n \) such that \( T^n = t_\nu < T'^m = T'_m \) and \( \|X(T^n) \land t_\nu\| \lor \|X'(T'^m) \land t_\nu\| < n' < \infty \). Thanks to Step 2, \((X, Y)\) and \((X', Y')\) must coincide until \( T = (t_\nu + 1) \land T'^m \land T'_m \land T'^{n'} \land T_X \land t_\nu < T \) (with the fact that \( X \) and \( X' \) are right-continuous). This proves that \( T^n \land t_\nu = T'^m \land t_\nu \), and with \( t_\nu, n \to \infty \) that \( \tau_0 = \tau'_0 \).

By symmetry between the two solutions, we have a.s. \( \tau_0 = \tau'_0 \), \( \forall t < \tau_0 \), \( X_t = X'_t \) and \( \forall t \geq \tau_0 \), \( X_t = X'_t = 0 \). It concludes the proof of the uniqueness. \(\Box\)

**Existence.** We see that the identification obtained for the uniqueness clearly defines the solution \((X, Y)\) until some \( T = T(t_\nu, n) \) such that either \( T = t_\nu \) or \( Y_T = 0 \) or \( \|X_T\| \geq n \). Thanks to the uniqueness property and the a priori estimates, this solution coincide with the ones for larger values of \( t_\nu \) and \( n \). Thus, it indeed produces a solution up to time \( \tau_0 \).

### 4 Main properties leading to the proof of Theorem 2.1

#### 4.1 General criteria for the proof of exponential quasi-stationary convergence

The proof of Theorem 2.1 relies on the set of Assumptions (AF) presented in [Ve21b], and that we recall next. (AF) is stated in the general context of a càdlàg process \( Z \) on a Polish state \( Z \), with extinction at time still denoted \( \tau_0 \). The notations are changed from [Ve21b] to avoid confusion with the current ones, \( Z \) corresponding now to the couple \((X, Y)\). We introduce the following notations for the exit and first entry times of any set \( D \):

\[
T_D := \inf \{ t \geq 0 \, ; \, Z_t \notin D \}, \quad \tau_D := \inf \{ t \geq 0 \, ; \, Z_t \in D \}.
\]

(4.1)

The assumptions involved in (AF) are the following ones.

(A0\(_S\)) There exists a sequence \((D_\ell)_{\ell \geq 1}\) of closed subsets of \( Z \) such that for any \( \ell \geq 1 \), \( D_\ell \subset \text{int}(D_{\ell+1}) \) (with \( \text{int}(D) \) the interior of \( D \)).

(A1) There exists a probability measure \( \zeta \in \mathcal{M}_1(Z) \) such that, for any \( \ell \geq 1 \), there exists \( L > \ell \) and \( c, t > 0 \) such that:

\[
\forall z \in D_\ell, \quad \mathbb{P}_z [Z_t \in dx \, ; \, t < \tau_0 \land T_{D_\ell}] \geq c \zeta (dz).
\]

(A2) \( \sup_{\{z \in Z\}} \mathbb{E}_z (\exp [\rho (\tau_0 \land \tau_E)]) < \infty \).

(A3\(_F\)) for any \( \epsilon \in (0, 1) \), there exist \( t_\tilde{\alpha}, c > 0 \) such that for any \( z \in E \) there exists a stopping time \( U_A \) such that:

\[
\{\tau_0 \land t_\tilde{\alpha} \leq U_A \} = \{U_A = \infty \} \quad \text{and} \quad \mathbb{P}_z (U_A = \infty, t_\tilde{\alpha} < \tau_0) \leq \epsilon \exp (-\rho t_\tilde{\alpha}),
\]

(4.2)
while for some stopping time $V$:

$$
\mathbb{P}_x(Z(U_A) \in dz' \; ; \; U_A < \tau_0) \leq c \mathbb{P}_x(Z(V) \in dz' \; ; \; V < \tau_0). \quad (4.3)
$$

We further require that there exists a stopping time $U_A^\infty$ extending $U_A$ in the following sense:

* $U_A^\infty := U_A$ on the event $\{\tau_0 \wedge U_A < \tau^1_E\}$, where $\tau^1_E := \inf\{s \geq t_\infty \colon Z_s \in E\}$.
* On the event $\{\tau^1_E \leq \tau_0 \wedge U_A\}$ and conditionally on $\mathcal{F}_{\tau^1_E}$, the law of $U_A^\infty - \tau^1_E$ coincides with the one of $\tilde{U}_A^\infty$ for a realization $\tilde{Z}$ of the Markov process $(Z_t, t \geq 0)$ with initial condition $\tilde{Z}_0 := Z(\tau^1_E)$ and independent of $Z$ conditionally on $Z(\tau^1_E)$.

$\rho$ as stated in Assumptions (A2) and (A3$F$) is required to be strictly larger than the following "survival estimate":

$$
\rho_S := \sup\{\gamma \geq 0 \; ; \; \sup_{L \geq 1, t \geq 0} e^{\gamma t} \mathbb{P}_x(t < \tau_0 \wedge T_{D_L}) = 0\} \lor 0.
$$

We are now in position to state (AF):

"(A1) holds for some $\zeta \in \mathcal{M}_1(\mathcal{Z})$ and a sequence $(D_\ell)_\ell$ satisfying (A0$S$). Moreover, there exist $\rho > \rho_S$ and a closed set $E$ such that $E \subset D_\ell$ for some $\ell \geq 1$ and such that (A2) and (A3$F$) hold."

As stated next by gathering the results of Theorems 2.2, 2.3 and Corollary 2.2.3 of [Ve21Ib], (AF) implies the convergence results that we aim, noting that the sequence $(D_\ell)_\ell$ will cover the whole space. Some additional properties of approximations are also obtained, where the process is localized to large $D_L$ by extinction.

**Theorem 4.1.** Provided that (AF) holds, the semi-group $P_t$ associated to the process $Z$ with extinction at time $\tau_0$ displays a uniform exponential quasi-stationary convergence with some characteristics $(\alpha, h, \lambda) \in \mathcal{M}_1(\mathcal{Z}) \times B(\mathcal{Z}) \times \mathbb{R}$.

Moreover, consider for any $L \geq 1$ the semi-group $P_L$ for which the definition of $\tau_0$ is replaced by $\tau^L_0 := \tau_0 \wedge T_{D_L}$. Then, for any $L \geq 1$ sufficiently large, $P_L$ displays a uniform exponential quasi-stationary convergence with some characteristics $(\alpha^L, h^L, \lambda_L) \in \mathcal{M}_1(D_L) \times B(D_L) \times \mathbb{R}_+$. The associated versions of (2.1) hold true with constants that can be chosen uniformly in $L$. As $L$ tends to infinity, $\lambda_L$ converges to $\lambda$ and $\alpha^L, h^L$ converge to $\alpha, h$ in total variation and pointwise respectively.

If in addition, $\bigcup_{\ell \geq 1} D_\ell = \mathcal{Z}$, then $h$ is positive and the results of Theorem 2.2 on the $Q$-process hold also true.

**Remarks 4.1.1.** Under (AF), the $Q$-process can generally be defined on $\mathcal{H} := \{z \in \mathcal{Z} \colon h(z) > 0\}$ and the fact that $h$ is positive is not required or may be proven as a second step. The proof of Theorem 4.1 however provides lower-bound of $h$ on any $D_\ell$, so that $\mathcal{Z} = \bigcup_{\ell \geq 1} D_\ell$ is a practical assumption for the proof that $h$ is positive.

**Remarks 4.1.2.** The assumption (A3$F$) appears certainly technical and its usage is the main focus of [Ve21Ib]. It is referred to as the “Absorption with failure” property and makes it possible to upper-bound the asymptotic survival probability from initial condition $z$ as compared to the one from initial condition $\zeta$. To this purpose, a coupling is introduced where (4.3) makes it possible to “absorb” most trajectories. Since failures where $U_A = \infty$ while $t_\infty < \tau_0$ are allowed, this step is to be iterated and the probability of such failure is to be controlled through (4.2).

For the proof of Theorem 2.1 the sequence $(D_\ell)_{\ell \geq 1}$ is defined as follows:

$$
D_\ell := B(0, \ell) \times [1/\ell, \ell], \quad (4.4)
$$

where $B(0, \ell)$ denotes the closed ball of radius $\ell$ for the Euclidian norm.
Forbidding deleterious mutations in the case of unidimensional \( X \) will make our proof a bit more complicated. This case is thus treated later on. The expression "with deleterious mutation" will be used a bit abusively to discuss the model under \((D)\). On the other hand, the expression "with only advantageous mutation" will refer to the case where \((A)\) holds.

These criteria are proved to hold true under the assumptions of Theorem 2.1 in the following Theorems 4.2-6. We see in Subsection 4.2.1 how these theorems together with Theorem 4.1 imply Theorem 2.1. In the next subsections, we then prove Theorems 4.2-6. By mentioning first the mixing estimate, we wish to highlight the constraint on the reachable domain under hypothesis \((A)\). The order of the proofs is different and done for the clarity of their presentation. The mixing estimates are handled similarly under the different sets of assumptions and directly exploited in the proofs of the absorption estimates. The escape estimates are very close to the ones of previously considered models, so more easily dealt with.

4.2 The whole space is accessible: with deleterious mutations or \( d \geq 2 \)

4.2.1 Mixing property and accessibility

With deleterious mutations, the whole space becomes accessible. It is in fact also the case with only advantageous mutations, provided \( d \geq 2 \):

**Theorem 4.2.** Suppose Assumption \((H)\). For \( d = 1 \), assume \((D)\). For \( d \geq 2 \), assume either \((D)\) or \((A)\). Then, for any \( l_1, l_M \geq 1 \), there exists \( L > l_1 \lor l_M \) and \( c, t > 0 \) such that:

\[
\forall (x_I, y_I) \in D_{l_1}, \quad P_{(x_I, y_I)}[(X, Y)t \in (dx, dy) ; t < \tau_0 \lor T_{D_L}] \geq c 1_{D_{l_M}}(x, y)dx dy.
\]  

(4.5)

**Remarks 4.2.1.**

- **(4.1)** is exploited when defining \( T_{D_L} := \inf \{ t \geq 0 ; (X, Y)t \notin D_L \} \).
- **Theorem 4.2** implies in particular that the density w.r.t. Lebesgue’s measure of any QSD is uniformly lower-bounded on any \( D_{l_1} \).
- In the case where \((D)\) holds, \( L := l_1 \lor l_M + \theta \) can be chosen. The choice of \( t \) cannot generally be made arbitrary, at least for \( d = 1 \), since the lower-bound of the density of jump sizes is only valid for jumps of size close to \( \theta \). Under \((A)\) with \( d \geq 2 \), the constraint that jumps must be advantageous makes the convenient choice of \( L \) less clear.

4.2.2 Escape from the Transitory domain

**Theorem 4.3.** Assume Assumptions \((H)\). Then, for any \( \rho > 0 \), there exists \( l_E \geq 1 \) such that \((A2)\) holds with \( E := D_{l_E} \).

**Remarks 4.2.2.** Heuristically, it means that the killing rate can be made arbitrarily large by adding killing when hitting some compact \( D_{l} \) that sufficiently covers \( Z = \mathbb{R} \times \mathbb{R}^*_+ \).

4.2.3 Absorption with failures

We need some reference set on which our reference measure has positive density. With the constants \( \theta \) and \( \eta \) involved in \([H4]\) let:

\[
\Delta := B(-\theta e_1, \eta) \times [1/2, 2].
\]  

(4.6)

This choice (rather arbitrary), is made in such a way that the uniform distribution on \( \Delta \) can be taken as the lower-bound in the conclusions of Theorems 4.5 and 4.2.

Including deleterious mutations or with \( d \geq 2 \), we will exploit the following theorem for sets \( E \) of the form \( E := D_{l_E} \), where \( l_E \) is determined thanks to Theorem 4.3. But the theorem holds generally for any closed subsets \( E \) of \( \mathbb{R}^d \times \mathbb{R}^*_+ \), for which there exists \( l \geq 1 \) such that \( E \subset D_l \), property that we briefly denote as \( E \in D \).
Theorem 4.4. Suppose Assumption (H). For \( d = 1 \), assume (D). For \( d \geq 2 \), assume either (D) or (A). Then, for any \( \rho > 0 \), \( \epsilon \in (0, 1) \) and \( E \in \mathcal{D} \), there exist \( t_{\zeta}, c > 0 \) which satisfy the following property for any \((x, y) \in E \) and \((x_{\zeta}, y_{\zeta}) \in \Delta \). There exists a stopping time \( U_A \) such that:
\[
\{\tau_0 \wedge t_{\zeta} \leq U_A\} = \{U_A = \infty\} \quad \text{and} \quad \mathbb{P}_{(x, y)}(U_A = \infty, t_{\zeta} < \tau_0) \leq \epsilon \exp(-\rho t_{\zeta}),
\]
and an additional stopping time \( V \) such that:
\[
\mathbb{P}_{(x, y)}[(X(U_A), Y(U_A)) \in (dx', dy') \ ; \ U_A < \tau_0] \\
\leq c \mathbb{P}_{(x_{\zeta}, y_{\zeta})}[(X(V), Y(V)) \in (dx', dy') \ ; \ V < \tau_0].
\]
Moreover, there exists a stopping time \( U_A^\infty \) satisfying the following properties:

- \( U_A^\infty := U_A \) on the event \( \{\tau_0 \wedge U_A < \tau_E^1\} \), where \( \tau_E^1 := \inf\{s \geq t_{\zeta} : (X_s, Y_s) \in E\} \).
- On the event \( \{\tau_E^1 < \tau_0\} \cap \{U_A = \infty\} \), and conditionally on \( \mathcal{F}_{\tau_E^1} \), the law of \( U_A^\infty - \tau_E^1 \) coincides with the one of \( \tilde{U}_E^\infty \) for the solution \((\tilde{X}, \tilde{Y})\) of:
\[
\begin{cases}
\tilde{X}_r = X(\tau_E^1) - v r e_1 + \int_{[0,r] \times \mathbb{R}^d \times \mathbb{R}^d} w \varphi \left( \tilde{X}_{s-}, \tilde{X}_s, w, u_f, u_g \right) \tilde{M}(ds, dw, du_f, du_g) \\
\tilde{Y}_r = Y(\tau_E^1) + \int_0^r \psi \left( \tilde{X}_s, \tilde{Y}_s \right) ds + \tilde{B}_r,
\end{cases}
\]
where \( r \geq 0 \), \( \tilde{M} \) and \( \tilde{B} \) are independent copies of respectively \( M \) and \( B \).

4.2.4 Proof of Theorem 2.1 as a consequence of Theorems 4.2-3

- First, it is clear that the sequence \((\mathcal{D}_\ell)_{\ell} \) satisfies both \((A0_2)\) and \( \bigcup_{\ell \geq 1} \mathcal{D}_\ell = \mathcal{Z} \).
- \((A1)\) holds true thanks to Theorem 4.2 where \( \zeta \) is the uniform distribution over \( \Delta \) - cf (4.6).
- Theorem 4.3 implies \((A2)\) for any \( \rho \), and we also require that \( \rho \) is chosen such that:
\[
\rho > \rho_S := \sup \{\gamma \geq 0 \ ; \ \sup_{L \geq 1} \inf_{t > 0} e^{\gamma t} \mathbb{P}_\zeta(t < \tau_0 \wedge T_D^\ell) = 0\} \lor 0.
\]

Thanks to Lemma 3.0.2 in [Ve21a] and \((A1)\), we know that \( \rho_S \) is upper-bounded by some value \( \tilde{\rho}_S \). In order to satisfy \( \rho > \rho_S \), we set \( \rho := 2\tilde{\rho}_S \). Thanks to Theorem 4.3, we deduce \( E = \mathcal{D}_E^\ell \) such that assumption \((A2)\) holds for this value of \( \rho \).

- Finally, Theorem 4.4 implies that assumption \((A3_F)\) holds true, for \( E \) and \( \rho \). In the adaptation of (4.7) where \((x_{\zeta}, y_{\zeta})\) is replaced by \( \zeta \), \( V \) is specified by the initial condition \((x_{\zeta}, y_{\zeta}) \in \Delta \) chosen uniformly according to \( \zeta \).

This concludes the proof of the assumption \((AF)\) with \( \bigcup_{\ell \geq 1} \mathcal{D}_\ell = \mathcal{Z} \). Exploiting Theorem 4.1, it implies Theorems 2.1 and 2.2 in the case where, besides Assumption (H), either \((D)\) holds or \( d \geq 2 \) and \((A)\) holds.

4.3 No deleterious mutations in the uni-dimensional case

4.3.1 Mixing property and accessibility

When only advantageous mutations are allowed and \( d = 1 \), as soon as the size of jumps is bounded, the process can’t access some portion of space (there is a limit in the \( X \) direction). We could prove that the limit is related to the quantity:
\[
L_A := \sup \{M; \nu[2M, +\infty) > 0\} \in (\theta/2, \infty).
\]
The accessible domains with maximal extension would then be rather of the form:
\[
[-\ell, L_A - 1/\ell] \times \mathbb{R}^{d-1}.
\]
For some \( \ell \geq 1 \), to simplify the proof, the limit \( L_A \) will however not appear in the next statements. We just wanted to point out this potential constraint on the visited domain. In fact, the \( X \) component is assumed to be negative in the following definition of the accessibility domains:

\[
\Delta_E := \{ [-L, 0] \times [1/\ell, \ell]; \ L, \ell \geq 1 \}.
\]

**Theorem 4.5.** Assume \( d = 1 \), Assumption (H) and (A). Then, for any \( \ell I \geq 1 \) and \( E \in \Delta_E \), there exists \( L > \ell I \) and \( c, t > 0 \) such that:

\[
\forall (x_I, y_I) \in D_{\ell I}, \quad \mathbb{P}(x_I, y_I) [(X_t, Y_t) \in (dx, dy); t < \tau_0 \land T_{D_L}] \geq c 1_E(x, y) dx \, dy.
\]

**Remarks 4.3.1.** Theorem 4.5 implies that the density w.r.t. Lebesgue’s measure of any QSD is uniformly lower-bounded on any \( E \) of the form given by (4.9).

### 4.3.2 Escape from the Transitory domain

**Theorem 4.6.** Assume \( d = 1 \), Assumptions (H) and (A). Then, for any \( \rho > 0 \), there exists \( E \in \Delta_E \) such that (A2) holds.

**Remarks 4.3.2.** Heuristically, it means that the asymptotic killing rate can be made arbitrarily large by adding killing when hitting some compact \( E \) that sufficiently covers \( \mathbb{R}^- \times \mathbb{R}^+ \).

### 4.3.3 Absorption with failures

**Theorem 4.7.** Suppose Assumption (H) and (A). Then, for any \( \rho > 0 \), \( \epsilon \in (0, 1) \) and \( E \in \Delta_E \), there exist \( t \gamma, c > 0 \) which satisfy the same property as in Theorem 4.4.

**Remarks 4.3.3.** The definition of \( \Delta \) is chosen to apply for both theorems.

### 4.3.4 Proof of Theorem 2.1 as a consequence of Theorems 4.5-6

The argument being very similar to the one for the case \( d \geq 2 \) or with (D), we go briefly through it.

- (A1) holds thanks to Theorem 4.5 with again the choice of \( \zeta \) uniform on \( \Delta \).
- Thanks to Theorem 4.6 and similarly as in the proof exploiting Theorem 1.3 in Subsection 4.2.4, we deduce that there exists \( E \in \Delta_E \) such that (A2) holds with some value \( \rho > \rho_S \).
- Finally, (A3F) holds for these choices of \( \rho \) and \( E \), thanks to Theorem 4.4.

This concludes the proof of the assumption (AF) with \( \bigcup_{\ell \geq 1} D_{\ell} = \mathcal{Z} \). Exploiting Theorem 4.1, it implies Theorems 2.1 and 2.2 in the case where \( d = 1 \), Assumptions (H) and (A) hold. □

### 4.4 Structure of the proof

To allow for fruitful comparison, the proofs are gathered according to the properties resp. (A1), (A2) and (A3F) they ensure. We first prove Theorems 4.3 and 4.6 in Section 5 since they are the simplest and the closest to the proofs in [Ve21b] and the remaining theorems are more closely related. We then prove Theorems 4.2 and 4.5 in Section 6, and finally Theorems 4.4 and 4.7 in Section 7.

### 5 Escape from the transitory domain

The most straightforward way to prove exponential integrability of first hitting times is certainly via Lyapunov methods. Yet, we highly doubt that this can be achieved as easily as we present next given the interplay between the different domains on which the escape is to be justified.
5.1 With deleterious mutations or \( d \geq 2 \)

Theorem 4.3 is a direct consequence of Proposition 4.2.2 in [Ve21a] and the proof is thus omitted. The process mainly considered in [Ve21a] is similar to this one in that there is also a coupling between a population size process \( N \) and an adaptation process \( X \). Both population size processes are defined in the same way in their relation to the process \( X \) as

\[
N_t = n + \int_0^t (r(X_s)N_s - \gamma_0 \times (N_s)^2) \, ds + \sigma \int_0^t \sqrt{N_s} \, dB_s.
\]

Contrary to the current model, \( X \) is not evolving in [Ve21a] as piecewise deterministic with jumps, but as a diffusion process. If it changes significantly the proof of the other assumptions (A1) and (A3), this proof of (A2) actually does not depend at all on the dynamics of \( X \), as expressed in Proposition 4.2.2 of [Ve21a]. The proof developed in the next subsection is an extension of this one and illustrates the technique.

Let us give a few hints of how it works. The proof relies on uniform couplings which ensure that with a probability close to 1, the population size experience drastic decrease sufficiently quickly, be it when it starts at a very large value, when the adaptation is very poor (large \( \|X\| \)) or when the population is close to extinction. In addition, we simply need to prove that the probability of large increase is also very exceptional.

In practice, we distinguish between 3 different sets of initial conditions depending on which of the above situation is to be considered, like the sets \( T_\infty, T_X^\infty \) and \( T_0 \) from Figure 1. The above-mentioned estimations provide relations between the 3 exponential moments of return starting from the different sets of initial conditions. The decrease estimate proves that, prescribing a fixed time interval, the process exits during this time interval the set of conditions he starts in with a probability sufficiently close to 1. The increase estimate makes it possible to control the probability of trajectories rapidly navigating between the different sets of conditions.

5.2 Without deleterious mutations, \( d = 1 \)

In this section, we prove Theorem 4.6, i.e.:

Suppose that \( d = 1 \), assumption \((H)\) and \((A)\) hold. Then:

\[
\forall \rho > 0, \quad \exists E \in \Delta_E, \quad \sup_{(x,y) \in \mathbb{R} \times \mathbb{R}_+} \mathbb{E}_{(x,y)} (\exp [\rho (\tau_E \wedge \tau_0)]) < \infty.
\]

5.2.1 Decomposition of the transitory domain

The proof is very similar to the one of Subsection 4.2.4 of [Ve21b] except that, due to Theorem 4.7, the domain \( E \) cannot be chosen as large. We thus need to consider another subdomain of \( T \), that will be treated specifically thanks to \((A)\).

The complementary \( T \) of \( E \) is then made up of 4 subdomains: "\( y = \infty \)", "\( y = 0 \)" , "\( x > 0 \)" , and "\( \|x\| = \infty \)" , according to the figure 1. Thus, we define:

- \( T_\infty^Y := ((-\infty, -L) \cup (0, \infty)) \times (y_\infty, \infty) \cup [-\ell, 0] \times [\ell, \infty) \) ("\( y = \infty \)"),
- \( T_0 := (-L, L) \times [0, 1/\ell] \) ("\( y = 0 \)"),
- \( T_+ := (0, L) \times (1/\ell, y_\infty] \) ("\( x > 0 \)"),
- \( T_\infty^X := \mathbb{R} \setminus (-L, L) \times (1/\ell, y_\infty] \) ("\( \|x\| = \infty \)").
Until the process reaches $E$ or extinction, it is likely to escape any region either from below or from the side into $T_+$, the reversed transitions being unlikely. As long as $X_t > 0$, $\|X_t\|$ must decrease (see Fact 5.2.5 in Subsection 5.2.4). Once the process has escaped $\{x \geq L_A\}$, there is no way (by allowed jumps and $v$) that it reaches it afterwards.

With some threshold $t_\vee$ (meant to ensure finiteness but whose effect shall vanish as it tends to $\infty$), let us first introduce the exponential moments of each area (remember that $\tau_E$ is the hitting time of $E$):

- $\mathcal{E}_Y^\infty := \sup_{(x,y) \in \mathcal{T}_\infty} E_{(x,y)}[\exp(\rho V_E)]$,
- $\mathcal{E}_X^\infty := \sup_{(x,y) \in \mathcal{T}_\infty} E_{(x,y)}[\exp(\rho V_E)]$,
- $\mathcal{E}_0 := \sup_{(x,y) \in \mathcal{T}_0} E_{(x,y)}[\exp(\rho V_E)]$,
- $\mathcal{E}_X := \sup_{(x,y) \in \mathcal{T}_+} E_{(x,y)}[\exp(\rho V_E)]$,

where $V_E := \tau_E \land \tau_{\partial} \land t_\vee$. Implicitly, $\mathcal{E}_Y^\infty$, $\mathcal{E}_X^\infty$, $\mathcal{E}_X$ and $\mathcal{E}_0$ are functions of $\rho$, $L$, $\ell$, $y_\infty$ that need to be specified.

### 5.2.2 A set of inequalities

Like in Subsection 4.2.4 in [Ve21a], we first state some inequalities between these quantities, summarized in Propositions 5.2.1, 5.2.2, 5.2.3 and 5.2.4 that follow. Thanks to these inequalities, we prove in Subsection 5.2.3 that those quantities are bounded. This will end the proof of Theorem 4.6.

**Proposition 5.2.1.** Suppose Assumption (H). Then, given any $\rho > 0$, there exist $y_\infty > 0$ and $C_y^\infty \geq 1$ such that for any $\ell > y_\infty$ and $L > 0$:

$$
\mathcal{E}_Y^\infty \leq C_y^\infty \left(1 + \mathcal{E}_X^\infty + \mathcal{E}_X\right).
$$

(5.1)
Proposition 5.2.2. Suppose Assumption (H) and [H3]. Then, given any \( \rho > 0 \), there exists \( C_X^\infty \geq 1 \) which satisfies the following property for any \( \epsilon^X, \gamma_\infty > 0 \). There exists \( L > 0 \) and \( \ell^X > \gamma_\infty \) such that for any \( \ell \geq \ell^X \):

\[
\mathcal{E}_\infty^X \leq C_\infty^X (1 + \mathcal{E}_0 + \mathcal{E}_X) + \epsilon^X \mathcal{E}_\infty^Y.
\]  

(5.2)

Proposition 5.2.3. Suppose Assumption (H) and (A). Then, given any \( \rho, L > 0 \), there exists \( C_X \geq 1 \) which satisfies the following property for any \( \epsilon^+, \gamma_\infty > 0 \). For any \( \ell \) sufficiently large (\( \ell \geq \ell^+ > \gamma_\infty \)):

\[
\mathcal{E}_X \leq C_X (1 + \mathcal{E}_0) + \epsilon^+ \mathcal{E}_\infty^Y.
\]  

(5.3)

Proposition 5.2.4. Suppose Assumption (H). Then, given any \( \rho, \epsilon^0, \gamma_\infty > 0 \), there exists \( C_0 \geq 1 \) which satisfies the following property for any \( L \) and \( \ell \) sufficiently large (\( \ell \geq \ell^0 > \gamma_\infty \)):

\[
\mathcal{E}_0 \leq C_0 + \epsilon^0 \left( \mathcal{E}_\infty^Y + \mathcal{E}_\infty^X + \mathcal{E}_X \right).
\]  

(5.4)

The proofs of Proposition 5.2.1, 5.2.2 and 5.2.4 can be taken mutatis mutandis from the ones of Propositions respectively 4.2.1, 4.2.2 and 4.2.3 of Ve21a. The idea behind them is exactly the same as in the explanation provided for Subsection 5.1, made more precise with the statement of the propositions. The justification relies on the estimate of what has happened in a time-interval \([0, t_0]\) for some arbitrary \( t_0 > 0 \): by a proper definition of the domain boundaries, observing a transition to some domain with lower-population sizes (or to extinction) is justified to be a very likely event, with a probability larger than \( 1 - \exp(-\rho t_0) \). Transitions to domain with a larger population size shall be handled as very exceptional, where an additional threshold is involved to specify transitions that are considered to happen during the time-interval \((0, t_0]\). The only difference in the proofs is that transitions into \( T_+ \) are distinguished in the current paper, which makes appear the term \( \mathcal{E}_X \) with factors resp. \( C_\infty^X, C_\infty^X \) and \( \epsilon^0 \). The reason for \( \epsilon^0 \) to be as small as needed is that spontaneous extinction during some finite time-interval \([0, t_0]\) can be made as likely as needed, while potential transitions (including those towards \( T_+ \)) are only considered at the end of the time-interval \([0, t_0]\).

We prove first how to deduce Theorem 4.6, which naturally generalizes the similar argument in Ve21a. Then, we will prove Proposition 5.2.3 that shall provide the main intuition for the proofs of the other propositions. The core idea is that jumps are not allowed here to increase the maladaptation of the process. Thus, the worst-case scenario for the exit time of \( T_+ \) is that the process gets simply drifted by the environmental change at speed \( v \) until \( X \) gets negative.

5.2.3 Proof that Propositions 5.2.1-4 imply Theorem 4.6

We first prove that the inequalities (5.2), (5.3) and (5.4) give a bound on \( \mathcal{E}_\infty^X \wedge \mathcal{E}_\infty^Y \wedge \mathcal{E}_X \wedge \mathcal{E}_0 \). Theorem 4.6 then states that for sufficiently small \( \epsilon^X, \epsilon^+ \) and \( \epsilon^0 \) we have:

\[
\mathcal{E}_\infty^X \leq C_\infty^X (3 + 3 \mathcal{E}_X + 2 \mathcal{E}_0) , \quad \mathcal{E}_\infty^Y \leq C_\infty^Y C_\infty^X (4 + 4 \mathcal{E}_X + 2 \mathcal{E}_0) .
\]

Assuming first that \( \epsilon^X \leq (2 C_\infty^Y C_\infty^X)^{-1} \), we have:

\[
\mathcal{E}_X \leq C_X (2 + 3 \mathcal{E}_0) , \quad \mathcal{E}_\infty^X \leq C_\infty^X C_X (9 + 11 \mathcal{E}_0) , \quad \mathcal{E}_\infty^Y \leq C_\infty^Y C_\infty^X (12 + 14 \mathcal{E}_0) .
\]

Assuming further that \( \epsilon^+ \leq (8 C_\infty^Y C_\infty^X)^{-1} \),

\[
\mathcal{E}_X \leq C_X (2 + 3 \mathcal{E}_0) , \quad \mathcal{E}_\infty^X \leq C_\infty^X C_X (9 + 11 \mathcal{E}_0) , \quad \mathcal{E}_\infty^Y \leq C_\infty^Y C_\infty^X (12 + 14 \mathcal{E}_0) .
\]

Assuming further that \( \epsilon^0 \leq (60 C_\infty^Y C_\infty^X C_\infty^X)^{-1} \) (and exploiting \( 2 \times [14 + 11 + 3] \leq 60 \)):

\[
\mathcal{E}_0 \leq 50 C_0 , \quad \mathcal{E}_X \leq 152 C_X C_0 , \quad \mathcal{E}_\infty^X \leq 559 C_\infty^X C_X C_0 , \quad \mathcal{E}_\infty^Y \leq 712 C_\infty^Y C_\infty^X C_0 ,
\]

In particular \( \sup_{(x,y) \in \mathbb{R} \times \mathbb{R}_+} \mathcal{E}_{(x,y)} \left( \exp [\rho (\tau_E \wedge \tau_0)] \right) = \mathcal{E}_\infty^Y \wedge \mathcal{E}_\infty^X \wedge \mathcal{E}_X \wedge \mathcal{E}_0 < \infty \).
Let us now specify the choice of the various parameters involved. For any given \( \rho \), we obtain from Proposition 5.2.1 the constant \( y_\infty \), and \( C^X_\infty \) which gives us a value \( e^X := (2C^Y_\infty)^{-1} \). We then deduce, thanks to Proposition 5.2.2, some value for \( C^X_\infty, \ell^X \) and \( L \). We can then be fix \( \epsilon^+ := (8C^Y_\infty C^X_\infty)^{-1} \), and deduce, according to Proposition 5.2.3, some value \( C_\epsilon \) and \( \ell^+ > 0 \). Now we fix \( \epsilon^0 := (60C^Y_\infty C^X_\infty)^{-1} \) and choose, according to Proposition 5.2.4, some value \( C_0 \) and \( \ell^0 > 0 \). To make the inequalities (5.2), (5.3) and (5.4) hold, we can just take \( \ell := \ell^X \vee \ell^+ \vee \ell^0 \). With the calculations above, we then conclude Theorem 4.6.

### 5.2.4 Proof of Proposition 5.2.3

Phenotypic lag pushed towards the negatives.

Since the norm of \( X \) decreases at rate at least \( v \) as long as the process stays in \( \tilde{T}_+ := [0, L] \times \mathbb{R}^*_+ \), we know that the process cannot stay in this area during a time-interval larger than \( t_\vee := \frac{L}{v} \). This effect will give us the bound \( C_\epsilon := \exp(\rho L/v) \).

Moreover, we need to ensure that the transitions from \( \mathcal{E}_X \) to \( \mathcal{E}^Y_\infty \) are exceptional enough. This is done exactly as for Proposition 4.2.2 in [Ve21a], by taking \( \ell^+ \) sufficiently larger than \( y_\infty \). The event of having the process reaching \( \ell^+ \) in the time-interval \([0, t_\vee]\) is then exceptional enough.

More precisely, given \( L \) and \( \ell > y_\infty \geq 1 \) and initial condition \((x, y) \in \tilde{T}_+\), let:

\[
C_\epsilon := \exp \left( \frac{\rho L}{v} \right), \quad T := \inf \{ t \geq 0 ; \; X_t \leq 0 \} \wedge V_E \tag{5.5}
\]

**Fact 5.2.5.** Assume that Assumption \((H)\) and \((A)\) hold.

Then, for any initial condition \((x, y) \in \tilde{T}_+, (X, Y)_t \notin \tilde{T}_X^\infty \) a.s. and:

\[
\forall t < T, \quad X_t \leq x - vt \leq L - vt \quad \text{so that } T \leq t_\vee := L/v.
\]

Thanks to Assumption \([H4]\), an immediate induction on the number of jumps previous to \( T \wedge t \) proves that the jumps of \( X \) can only make its value decrease (because it is positive while the absolute value must necessarily decrease). It proves Fact 5.2.5. Thanks to it:

\[
\mathbb{E}_{(x, y)}[\exp(\rho V_E)] = \mathbb{E}_{(x, y)}[\exp(T) ; \; T = V_E] + \mathbb{E}_{(x, y)}[\exp(T) ; \; (X, Y)_t \in \tilde{T}_0] \\
+ C_\epsilon^Y \mathbb{E}_{(x, y)}[\exp(T) ; \; (X, Y)_t \in \tilde{T}^Y] \\
\leq C_\epsilon (1 + E_0) + C_\epsilon \mathbb{E}^Y_{y_\infty} \mathbb{P}_{y_\infty}(T_\uparrow \leq t_\vee)
\]

where \( T_\uparrow := \inf \{ t \geq 0 ; \; Y_t^\uparrow \geq \ell \} \), and \( Y_\uparrow \) is the solution of:

\[
Y_\uparrow := y_\infty + \int_0^t \psi_\vee(Y_\uparrow) \; ds + B_t \quad \text{(again } \psi_\vee(y) := -\frac{1}{2y} + \frac{r_v y}{2} - \gamma y^3). \tag{5.6}
\]

We conclude the proof of Proposition 5.2.3 by noticing that:

\[
\mathbb{P}_{y_\infty}(T_\uparrow \leq t_\vee) \xrightarrow{\ell \to \infty} 0.
\]

### 6 Mixing properties and accessibility

Before we turn to the proofs of Theorems 4.2 and 4.5, we describe the common elementary properties upon which they rely in the three following subsections. The first one gives the trick to disentangle the behavior of the processes \( X \) and \( N \) up to a factor on the densities. Subsection 6.2 deals with the mixing property for the \( Y \) process. These results are exploited in Subsection 6.3 to obtain the elementary mixing properties that allow to deduce \((A2)\). The three next subsections starting from 6.4 deal respectively with the proofs of Theorem 4.2 first under Assumption \((D)\), then under under Assumption \((A)\) and \( d \geq 2 \) and finally with the proof of Theorem 4.5.
General mixing properties

6.1  Construction of the change of probability under \([H4]\)

The idea of this subsection is to prove that we can think of \(Y\) as a Brownian Motion up to some stopping time which will bound \(U_A\). If we get a lower bound for the probability of events in this simpler setup, this will prove that we also get a lower bound in the general setup.

The limits of our control  Let \(t_g, x_\nu > 0, 0 < y_\lambda < \nu, N_J \geq 1\). Our aim is to simplify the law of \((Y_t)_{t \in [0,t_g]}\) as long as \(Y\) stays in \([y_\lambda, y_\nu], \|X\|\) stays in \(\bar{B}(0, x_\nu)\), and at most \(N_J\) jumps have occurred. Thus, let:

\[
T_X := \inf \{ t \geq 0 \mid \|X_t\| \geq x_\nu \}, \quad T_Y := \inf \{ t \geq 0 \mid Y_t \notin [y_\lambda, y_\nu] \}. 
\]

(6.1)

\[
g_\nu := \sup \{ g(x, w) \mid \|x\| \leq x_\nu, w \in \mathbb{R}^d \}, \quad f_\nu := \sup \{ f(y) \mid y \in [y_\lambda, y_\nu] \} 
\]

(6.2)

\[
J := \{(w, u_g, u_f) \in \mathbb{R}^d \times [0, f_\nu] \times [0, g_\nu]\},
\]

so that \(\nu \otimes du_g \otimes du_f(J) = \nu(\mathbb{R}^d) g_\nu f_\nu < \infty\).

Our Girsanov’s transform alters the law of \(Y\) until the stopping time:

\[
T_G := t_g \wedge T_X \wedge T_Y \wedge U_{N_J},
\]

(6.3)

where \(U_{N_J} := \inf \{ t \mid M([0,t] \times J) \geq N_J + 1 \} \).

(6.4)

Note that the \((N_J + 1)\)-th jump of \(X\) will then necessarily occur after \(T_G\).

The change of probability

We define:

\[
L_t := -\int_0^{t \wedge T_G} \psi(X_s, Y_s)dB_s, \quad \text{and} \quad D_t := \exp[L_t - \langle L \rangle_t/2],
\]

(6.5)

the exponential local martingale associated with \(L_t\).

**Theorem 6.1.** Suppose Assumption \((H)\). Then, for any \(t_g, x_\nu > 0\) and \(y_\lambda > \nu > 0\), there exists \(C_G > c_G > 0\) such that a.s. and for any \(t > 0\), \(c_G \leq D_t \leq C_G\). In particular, \(D_t\) is a uniformly integrable martingale and \(\beta_t = B_t - \langle B, L \rangle_t\) is a Brownian Motion under: \(P^G_{(x,y)} := D_\infty \cdot P_{(x,y)}\). Moreover:

\[
\forall \, (x, y) \in \mathbb{R}^d \times \mathbb{R}_+, \quad c_G P^G_{(x,y)} \leq P_{(x,y)} \leq C_G P^G_{(x,y)},
\]

On the event \(\{t \leq T_G\}, Y_t = y + \beta_t\), i.e. \(Y\) has the law of a Brownian Motion under \(P^G_{(x,y)}\) up to time \(T_G\). This means that we can have bounds of the probability for events involving \(Y\) as in our model by considering \(Y\) as a simple Brownian Motion. Meanwhile, the independence between its variations as a Brownian and the Poisson Process still hold due to Proposition 1.2.1

6.1.1  Proof of Theorem 6.1

The proof is achieved by ensuring uniform upper-bounds of \(L_t\) and \(\langle L \rangle_t\), which corresponds to \(L_\infty\) and \(\langle L \rangle_\infty\) for \(t_G\) replaced by \(t \wedge t_G\).
Proof in the case where $r$ is $C^1$

Let
\[ ||r||_{C^1}^G := \sup \{ |r(x)| : x \in \overline{B}(0,x_\vee) \}, \tag{6.6} \]
\[ ||r'||_{C^1}^G := \sup \{ |r'(x)| : x \in \overline{B}(0,x_\vee) \}. \tag{6.7} \]

With $\psi_G^\vee$ an upper-bound of $\psi$ on $\overline{B}(0,x_\vee) \times [y_\wedge, y_\vee]$ (deduced from [H3]) and recalling that $(X,Y)$ belongs to this subset until $T_G$ (see (6.3)):
\[ (L)_\infty = \int_0^{T_G} \psi(X_s,Y_s)^2 ds \leq t_G \times (\psi_G^\vee)^2. \tag{6.8} \]

In the following, we look for bounds on $\int_0^{T_G} \psi(X_s,Y_s)dY_s$, noting that:
\[ L_{T_G} + \int_0^{T_G} \psi(X_s,Y_s)dY_s = \int_0^{T_G} \psi(X_s,Y_s)^2 ds \in [0, t_G \times (\psi_G^\vee)^2]. \]

\[ \int_0^{T_G} \psi(X_s,Y_s)dY_s = \int_0^{T_G} \left( -\frac{1}{2Y_s} + r(X_s) \frac{Y_s}{2} - \gamma(Y_s)^3 \right) dY_s. \tag{6.9} \]

Now, thanks to Itô’s formula:
\[ \ln(Y_{T_G}) = \ln(y) + \int_0^{T_G} \frac{1}{Y_s} dY_s - \frac{1}{2} \int_0^{T_G} \frac{1}{(Y_s)^2} ds \]
\[ \text{thus} \left| \int_0^{T_G} \frac{1}{Y_s} dY_s \right| \leq 2 \left( |\ln(y_\wedge)| \vee |\ln(y_\vee)| \right) + \frac{t_G}{2(y_\wedge)^2} < \infty. \tag{6.10} \]

\[ (Y_{T_G})^4 = y^4 + 4 \int_0^{T_G} (Y_s)^3 dY_s + 6 \int_0^{T_G} (Y_s)^2 ds \]
\[ \text{thus} \left| \int_0^{T_G} (Y_s)^3 dY_s \right| \leq (y_\wedge)^4 / 4 + 3 t_G (y_\wedge)^2 / 2 < \infty. \tag{6.11} \]

\[ r(X_{T_G}^-)(Y_{T_G})^2 = r(x) y^2 + 2 \int_0^{T_G} r(X_s) Y_s dY_s + \int_0^{T_G} r(X_s) ds - v \int_0^{T_G} r'(X_s) (Y_s)^2 ds \]
\[ + \int_{(0,T_G) \times \mathbb{R}^d \times \mathbb{R}_+} \left( r(X_{s-} + w) - r(X_{s-}) \right) \times (Y_s)^2 \]
\[ \times 1_{\{ u_f \leq f(Y_s) \}} 1_{\{ u_g \leq g(X_{s-}, w) \}} M(ds,dw,du_f,du_g). \tag{6.12} \]

Since $\forall s \leq T_G, Y_s \in [y_\wedge, y_\vee]$, we get from [H2] and (6.2):
\[ \forall s \leq T_G, \forall w \in \mathbb{R}^d, \ g(X_{s-}, w) \leq g_\vee, \ f(Y_s) \leq f_\vee, \ \text{and} \ T_G \leq \underline{T}_X. \]

Since moreover $T_G \leq T_X$:
\[ \left| \int_{(0,T_G) \times \mathbb{R}^d \times \mathbb{R}_+} \left( r(X_{s-} + w) - r(X_{s-}) \right) (Y_s)^2 1_{\{ u_f \leq f(Y_s) \}} 1_{\{ u_g \leq g(X_{s-}, w) \}} M(ds,dw,du_f,du_g) \right| \]
\[ \leq 2 N_f \| r \|_{C^1}^G (y_\wedge)^2, \]

22
so that (6.12) leads to:

\[
2 \int_0^{T_G} r(X_s) Y_s \, dY_s \leq (2 (N_J + 1) \|r\|_\infty^G + \|r\|_{Lip}^G \nu_{t_G}) \times (y_\nu)^2 + \|r\|_\infty^G t_G < \infty. \tag{6.13}
\]

Inequalities (6.10), (6.11), (6.13) combined with (6.8) conclude that \(L_\infty\) and \(\langle L \rangle_\infty\) are uniformly bounded. This proves the existence of \(0 < c_G < C_G\) such that a.s. \(c_G \leq D_\infty \leq C_G\).

This statement is a priori adapted for \(t_G\) replaced by \(t \wedge t_G\), yet these bounds are actually the largest for \(t = t_G\). So it entails that \(c_G \leq D_t \leq C_G\) hold uniformly in \(t\). The rest of the proof is only classical application of Girsanov’s transform theory.

**Extension to the case where \(r\) is only Lipschitz-continuous**

(6.10) and (6.11) are still true, so we show that we can find the same bound on \(\int_0^{T_G} r(X_s) Y_s \, dY_s\) where we replace \(\|r\|_\infty^G\) by the Lipschitz-constant \(\|r\|_{Lip}^G\) of \(r\) on \(\mathcal{B}(0, x_\nu)\), by approximating \(r\) by \(C^1\) functions that are \(\|r\|_{Lip}^G\)-Lipschitz continuous.

**Lemma 6.1.1.** Suppose \(r\) is Lipschitz continuous on \(\mathcal{B}(0, x_\nu)\) for some \(x_\nu > 0\). Then there exists \(r_n \in C^1(\mathcal{B}(0, x_\nu), \mathbb{R})\), \(n \geq 1\) such that:

\[\|r_n - r\|_\infty^G \xrightarrow{n \to \infty} 0\quad \text{and} \quad \forall \, n \geq 1, \|r_n\|_\infty^G \leq \|r\|_{Lip}^G.\]

**Proof of Lemma 6.1.1**

We begin by extending \(r\) on \(\mathbb{R}^d\) with \(r_G(x) := r \circ \Pi_G(x)\), where \(\Pi_G\) is the projection on \(\mathcal{B}(0, x_\nu)\) (it is well-known that \(r\) can be extended on \(\mathcal{B}(0, x_\nu)\) with the same Lipschitz constant). Note that this extension \(r_G\) is still \(\|r\|_{Lip}^G\)-Lipschitz continuous. If we define now: \(r_n := r_G * \phi_n \in C^1\), where \((\phi_n)\) is an approximation of the identity of class \(C^1\), then:

\[\forall \, x, y, \|r_n(x) - r_n(y)\| = \left| \int_{\mathbb{R}^d} (r_G(x - z) - r_G(y - z)) \phi_n(z) \, dz \right| \leq \|r\|_{Lip}^G \|x - y\| \int_{\mathbb{R}^d} \phi_n(z) \, dz = \|r\|_{Lip}^G \|x - y\|.
\]

Thus

\[\forall \, n \geq 1, \quad \|r_n\|_\infty^G \leq \|r\|_{Lip}^G, \quad \|r_n - r\|_\infty^G \xrightarrow{n \to \infty} 0. \quad \square\]

**Proof that Lemma 6.1.1 and the case \(r \in C^1\) proves Theorem 6.1**

We just have to prove (6.13) with \(\|r\|_{Lip}^G\) instead of \(\|r\|_{Lip}^G\). If we apply this formula for \(r_n\) and exploit Lemma 6.1.1, we see that there will be some \(C = C(t_G, y_\nu, N_J) > 0\) such that:

\[2 \int_0^{T_G} r_n(X_s) Y_s \, dY_s \leq (2 (N_J + 1) \|r\|_\infty^G + \|r\|_{Lip}^G \nu_{t_G}) (y_\nu)^2 + r_\infty t_G + C \|r - r_n\|_\infty^G.
\]
Thus, it remains to bound:

\[
\left| \int_0^{T_G} (r_n(X_s) - r(X_s)) Y_s \, dY_s \right| \leq t_G y \nu \psi_G^Y \| r - r_n \|_G + |M_n|,
\]

where \( M_n := \int_0^{T_G} (r_n(X_s) - r(X_s)) Y_s \, dB_s \) has mean 0 and variance:

\[
\mathbb{E} ( (M_n)^2 ) = \mathbb{E} \left( \int_0^{T_G} (r_n(X_s) - r(X_s))^2 Y_s^2 \, ds \right) \leq t_G (y \nu)^2 (\| r - r_n \|_G^2) \xrightarrow{n \to \infty} 0.
\]

Thus, we can extract some subsequence \( M_{\phi(n)} \) which converges a.s. towards 0. So that a.s.:\[
\left| \int_0^{T_G} r(X_s) Y_s \, dY_s \right| \leq \liminf_{n \to \infty} \left\{ \left| \int_0^{T_G} r_{\phi(n)}(X_s) Y_s \, dY_s \right| + t_G y \nu \psi_G^Y \| r - r_{\phi(n)} \|_G + |M_{\phi(n)}| \right\} \leq \frac{1}{2} (2 (N_J + 1) \| r \|_G^2 + \| r \|_{L^2_{\nu}} v t_G) (y \nu)^2 + \frac{1}{2} \| r \|_G^2 t_G < \infty.
\]

The proof in the case \( r \in C^1 \) can then be exploited without difficulty. \( \square \)

**6.2 Mixing for \( Y \)**

The proof will rely on Theorem 6.1 and on the following classical property of Brownian Motion:

**Lemma 6.2.1.** Consider any constants \( b_\nu > 0, \epsilon > 0 \) and \( 0 < t_0 \leq t_1 \). Then, there exists \( c_B > 0 \) such that for any \( b \in [0,b_\nu] \) and \( t \in [t_0,t_1] \):

\[
\mathbb{P}_{b_1} \left( B_t \in db \; ; \; \min_{s \leq t_1} B_s \geq -\epsilon, \max_{s \leq t_1} B_s \leq b_\nu + \epsilon \right) \geq c_B \times 1_{[0, b_\nu]}(b) \, db,
\]

where \( B \) under \( \mathbb{P}_{b_1} \) has by definition the law of a Brownian Motion with initial condition \( b_1 \).

Thanks to this lemma and Theorem 6.1 we will be able to control \( Y \) to prove that it indeed diffuses and that it stays in some closed interval \( I_Y \) away from 0. We can then control the behavior of \( X \) independently of the trajectory of \( Y \) by appropriate conditioning of \( M \)–the PRaMe– so as to ensure the jumps we need (conditionally that it remains in \( I_Y \)).

**Proof:** Consider the collection of marginal laws of \( B_t \), with initial condition \( b \in (-\epsilon, b_\nu + \epsilon) \), killed when it reaches \( -\epsilon \) or \( b_\nu + \epsilon \). It is classical that these laws have a density \( u(t;b,b') \), \( t > 0, b' \in [-\epsilon, b_\nu + \epsilon] \), w.r.t. the Lebesgue measure (cf e.g. Section 2.4 in Bass [Ba95] for more details).

\( u \) is a solution to the Cauchy problem with Dirichlet boundary conditions:

\[
\partial_t u(t;b_1,b) = \Delta u(t;b_1,b), \quad \text{for } t > 0, b_1, b \in (-\epsilon, b_\nu + \epsilon),
\]

\[
u t \partial_t u(t;-\epsilon,b) = u(t;b_1, b_\nu + \epsilon) = 0, \quad \text{for } t > 0.
\]

Thanks to the maximum principle (cf e.g. Theorem 4, Subs 2.3.3. in Evans [Ev98]), \( u > 0 \) on \( \mathbb{R}_+ \times [0,b_\nu] \times (-\epsilon, b_\nu + \epsilon) \) and since \( u \) is continuous in its three variables, it is lower-bounded by some \( c_B \) on the compact subset \( [t_0, t_1] \times [0,b_\nu] \times [0,b_\nu] \). \( \square \)
6.3 Mixing for $X$

For clarity, we decompose the "migration" along $X$ into different kinds of elementary steps, as already done in [Ve21]. Let:

$$\mathcal{A} := \tilde{B}(\theta \mathbf{e}_1, \eta/2), \quad \tau_\mathcal{A} := \inf \{ t \geq 0 ; \ X_t \in \mathcal{A} , \ Y_t \in [2,3] \} ,$$

(6.14)

where we assume w.l.o.g. that $\eta \leq \theta/8$ ([2,3] is chosen arbitrarily).

Under any of the three sets of assumptions considered in the following, the proof is achieved in three steps. The first step is to prove that, with a lower-bounded probability for any initial condition in $\mathcal{D}_t$, $\tau_\mathcal{A}$ is upper-bounded by some constant $\ell_\mathcal{A}$. In the second step, we prove that the process is sufficiently diffuse and that time-shifts are not a problem. In the third step, we specify which sets we can reach from $\mathcal{A}$.

Recall that for any $\ell \geq 1$, $T_{\mathcal{D}_\ell} := \inf \{ t \geq 0 ; \ (X, Y)_t \notin \mathcal{D}_\ell \} < \tau_0$. For $n \geq 3$, let us define $T_{(n)} := T_{\mathcal{D}_{2n}}$. For $n \geq 3$ and $t, c > 0$, let:

$$\mathcal{R}^{(n)}(t, c) := \{ x_F \in \mathbb{R}^d ; \ \forall (x_0, y_0) \in \mathcal{A} \times [1/n, n], \ \mathbb{P}_{(x_0, y_0)}((X,Y)_t \in (dx, dy)) \geq c_\mathcal{D} \delta(t, \mathbf{e}_1, \eta/2)(x) \mathbf{1}_{[1/n, n]}(y) dx dy \} .$$

(6.15)

We will prove the mixing on a global scale by translating local mixing properties into some induction properties of the sets $(\mathcal{R}^{(n)}(t, c))_{t, c > 0}$.

Several local mixing properties require local lower- and upper-bounds on $g$, so that they can only be exploited in specific areas of $\mathbb{R}^d$. In order to provide a general framework for these through Proposition 6.3.2, let us consider the following sequence of sets, indexed by $n \geq 1$:

$$\mathcal{G}_n := \{ x \in \tilde{B}(0,0) ; \ \forall z \in [0, \eta/4], \ \forall \delta \in \tilde{B}(0, \eta/2), \ \forall w \in \tilde{B}(\theta \mathbf{e}_1, \eta), \ g(x - (\theta - z)\mathbf{e}_1 + \delta, w) \geq 1/n, \ \text{and } \forall z \in [-\theta, \eta/4], \ \forall \delta \in \tilde{B}(0, \eta/2), \ \forall w \in \mathbb{R}^d, \ g(x + z\mathbf{e}_1 + \delta, w) \leq n \} .$$

These steps are deduced from the following elementary properties:

**Lemma 6.3.1.** Assume that Assumption (H) hold. Then, for any $n \geq 1$ there exists $c_\mathcal{D} > 0$ such that for any $(x_1, y_1) \in \mathcal{D}_n$ and $u \in [0, u_\mathcal{D}(x_1)]$, where $u_\mathcal{D}(x) := \sup \{ u \geq 0 ; \ (x - vu \mathbf{e}_1) \in \tilde{B}(0,0) \}$:

$$\mathbb{P}_{(x_1, y_1)}((X,Y)_u \in (dx, dy)) \geq c_\mathcal{D} \delta_{(x_1 - vu \mathbf{e}_1)}(dx) \mathbf{1}_{[1/n, n]}(y) dy .$$

In particular, for any $t, c > 0$, $n \geq 3$, the fact that $x$ belongs to $\mathcal{R}^{(n)}(t, c)$ implies the following inclusion:

$$\forall u \in [0, u_\mathcal{D}(x)], \ x - vu \mathbf{e}_1 \in \mathcal{R}^{(n)}(t + u, c \times c_\mathcal{D}) .$$

The proof of Lemma 6.3.1 being easily adapted from the one of the next proposition, it is deferred after the proof of the latter.

**Proposition 6.3.2.** For any $n \geq 3$, there exists $t_p, c_p > 0$ such that for any $x_1 \in \mathcal{G}_n$, for any $x_0 \in \tilde{B}(x_1, \eta/4)$ and $y_0 \in [1/n, n]$:

$$\mathbb{P}_{(x_0, y_0)}((X,Y)_{t_p} \in (dx, dy)) ; t_p < T_{(n)} \geq c_p \delta(t, \mathbf{e}_1, \eta/4)(x) \mathbf{1}_{[1/n, n]}(y) dx dy .$$

A direct application of the Markov property implies the two following results.

**Corollary 6.3.3.** For any $n \geq 3$, there exists $t_p, c_p > 0$ such that for any $t, c > 0$, the following inclusion holds:

$$\{ x \in \mathbb{R}^d ; \ d(x, \mathcal{R}^{(n)}(t, c) \cap \mathcal{G}_n) \leq \eta/4 \} \subset \mathcal{R}^{(n)}(t + t_p, c \times c_p) .$$

**Fact 6.3.4.** For any $t, t', c, c' > 0$ and $n \geq 1$:

$$\mathcal{A} \cap \mathcal{R}^{(n)}(t, c) \neq \emptyset \Rightarrow \mathcal{R}^{(n)}(t', c') \subset \mathcal{R}^{(n)}(t + t', c \times c') .$$

25
Corollary 6.3.3 as a consequence of Proposition 6.3.2  

For $n \geq 3$, let $t_P, c_P > 0$ be prescribed by Proposition 6.3.2. We consider $x_I \in \mathcal{R}(n)(t,c) \cap \mathcal{G}_n$, $x_F$ such that $\|x_F - x_I\| \leq \eta/4$. Combining through the Markov property the fact that $x_I \in \mathcal{R}(n)(t,c)$ and Proposition 6.3.2 we deduce that for any $(x_0, y_0) \in A \times [1/n, n]$:

$$
P_{(x_0, y_0)} [(X,Y)_{t+t_P} \in (dx, dy) \ ; \ t + t_P < T(n)]
\geq c \int_{B(x_I, \eta/2)} \int_{[1/n]} dy_0 \mathbb{P}_{(x_0, y_0)} [(X,Y)_{tP} \in (dx, dy) \ ; \ t_P < T(n)]
\geq c \times \text{Leb}(B(x_I, \eta/4)) \times (n - 1/n) \times c_P \mathbf{1}_{B(x_I, 3\eta/4)}(x) \mathbf{1}_{[1/n, n]}(y) \ dx \ dy
\geq c \times c'_P \mathbf{1}_{B(x_F, \eta/2)}(x) \mathbf{1}_{[1/n, n]}(y) \ dx \ dy,
$$

where $c'_P := \text{Leb}(B(0, \eta/4)) \times (n - 1/n) \times c_P > 0$. This means that $x_F \in \mathcal{R}(n)(t + t_P, c \times c'_P)$. The proof of Corollary 6.3.3 is thus concluded with this choice of $t_P$ and $c'_P$, indeed independent from $x_I$, $x_F$.

\[ \square \]

Proof of Proposition 6.3.2

Step 1: description of the random event. For $n \geq 3$, we set $t_P := \theta/v$, $t_J := \eta/(4v)$, $y_\lambda := 1/(2n)$, $y_\nu := 2n$. Let also:

$$
T^Y := \inf \{ t \geq 0 \ ; \ Y_t \notin [y_\lambda, y_\nu] \},
\quad
f_\lambda := \inf \{ f(y) \ ; \ y \in [y_\lambda, y_\nu] \},
\quad
f_\nu := \sup \{ f(y) \ ; \ y \in [y_\lambda, y_\nu] \}.
$$

(6.16) (6.17)

$f_\nu$ is finite due to [H1]. Thanks to [H1], we know that $f_\lambda$ is positive.

On the event $\{t_P < T^Y\}$, we shall prove that the values of $X$ on $[0, t_P]$ are prescribed as functions of $M$ restricted to the subset:

$$
\mathcal{X}^M := [0, t_P] \times \mathbb{R}^d \times [0, f_\nu] \times [0, n].
$$

(6.18)

Let $x_0 := x_I + \delta_0$ with $x_I \in \mathcal{G}_n$ and $\delta_0 \in B(0, \eta/4)$, and $y_0 \in [1/n, n]$ that we consider as the initial conditions for the process $(X,Y)$.

To ensure one jump of size around $\theta$, at time nearly $t_P$, while “deleting” the contribution of $\delta_0$, let:

$$
\mathcal{J} := [t_P - t_J, t_P] \times B(\theta \mathbf{e}_1 - \delta_0, \eta/2) \times [0, f_\lambda] \times [0, 1/n].
$$

(6.19)

We partition $\mathcal{X}^M = \mathcal{J} \cup \mathcal{N}$, where: $\mathcal{N} := \mathcal{X}^M \setminus \mathcal{J}$. The event mostly under consideration is the following:

$$
\mathcal{W} = \mathcal{W}^{(x_0, y_0)} := \{ t_P < T^Y \} \cap \{ M(\mathcal{J}) = 1 \} \cap \{ M(\mathcal{N}) = 0 \}.
$$

(6.20)

Thanks to Theorem 6.1 (with $x_\nu := n + 2\theta$, $t_G = t_P$, and the same values for $y_\lambda$ and $y_\nu$), there exists $c_G > 0$ such that:

$$
P_{(x_0, y_0)} ((X,Y)_{t_P} \in (dx, dy) \ ; \ \mathcal{W})
\geq c_G \mathbb{P}_{(x_0, y_0)} ((X,Y)_{t_P} \in (dx, dy) \ ; \ \mathcal{W}).
$$

(6.21)

Under the law $\mathbb{P}_{(x_0, y_0)}$, the condition $\{ M(\mathcal{J}) = 1 \}$ is independent of $\{ M(\mathcal{N}) = 0 \}$, of $\{ t_P < T^Y \}$ and of $Y_{t_P}$, cf Proposition 1.2.1. Thus, on the event $\mathcal{W}$, the only “jump” coded in the restriction of $M$ on $\mathcal{J}$ is given as $(T_J, \theta \mathbf{e}_1 - \delta_0 + W, U_f, U_g)$, where $T_J, U_f$ and $U_g$ are chosen uniformly and independently on respectively $[t_P - t_J, t_P]$, $[0, f_\lambda]$ and $[0, 1/n]$, and $\theta \mathbf{e}_1 - \delta_0 + W$ independently according to the restriction of $\nu$ on $B(\theta \mathbf{e}_1 - \delta_0, 3\eta/4)$ (see notably chapter 2.4 in [DV08]). Thanks to $[H4]$, $W$ has a lower-bounded density $d_W$ on $B(0, 3\eta/4)$.

The following fact motivates this description:

26
Fact 6.3.5.  Under $P^{G}_{(x_0,y_0)}$ consider on the event $W$ the r.v. $W = W_J - \theta e_1 + \delta_0$ where $(T_J, W_J, U_f, U_g)$ is the only point encoded by $M$ on $J$. Then, a.s. $X_{t_P} = x_I + W$ and $W$ is included in $\{t_P < T(n)\}$.

Step 2: proof of Fact 6.3.5.

Step 2.1.  We prove that on the event $W$ defined by (6.20):  
$$\forall t < T_J, \quad X_t := x_0 - v t e_1. \quad (6.22)$$
Indeed, $t_P < T^Y$ implies that for any $t \leq T_J$, $Y_t \in [y_\land, y_\lor]$. Thanks to (6.17), any "potential jump" $(T'_J, W', U'_f, U'_g)$ such that $T'_J \leq T_J$ and either $U'_f > f_\lor$ or $U'_g > n$ will be rejected. Thanks to the definition of $T_J$, with (6.18), (6.19) and (6.20), no other jump can occur, thus (6.22) holds.

Note that, in order to prove this rejection very rigorously, we would like to consider the first one of such jumps. This cannot be done however for $(X,Y)$ directly, but is easy to prove for any approximation of $M$ where $u_I$ and $u_g$ are bounded. Since the result does not depend on these bounds and the approximations converge to $(X,Y)$ (and even equal to it before $T_J$ for bounds larger than $(f_\lor, n)$), (6.22) indeed holds.

Step 2.2.  We then prove that the jump at time $T_J$ is surely accepted. Since $x_I \in \Theta_n$, by (6.16) and the definition of $(T_J, W, U_f, U_g)$:  
$$U_f \leq f_\land \leq f(Y_{T_J}), \quad U_g \leq 1/n \leq g(x_0 - v T_J e_1, \theta e_1 - \delta_0 + W)$$  
Thus $X_{T_J} = x_I + \delta_0 - v T_J e_1 + \theta e_1 - \delta_0 + W = x_I + (\theta - v T_J) e_1 + W$.

Step 2.3.  We say that no jump can be accepted after $T_J$, which is proved as in Step 1. This means: $\forall T_J \leq t \leq t_P, \quad X_t = X_{T_J} - v(t - T_J) e_1 = x_I + W$.

This concludes in particular the proof of Fact 6.3.5 with $t = t_P = \theta/v$.  \hfill \Box

Step 3: concluding the proof of Proposition 6.3.2  Note that under $P^G$, $\{M(N) = 0\}$ is also independent of $\{t_P < T^Y\}$ and of $Y_{t_P}$, so that:

$$P^{G}_{(x_0,y_0)}((X,Y)_{t_P} \in (dx, dy) ; W) = P(M(N) = 0) \times P(M(J) = 1) \times P^{G}_{y_0}(Y_{t_P} \in dy ; t_P < T^Y) dW 1_{B(x_I,3\eta/4)}(x) dx. \quad (6.23)$$

From (6.18) and (6.19):

$$P(M(N) = 0) P(M(J) = 1) = (t_J f_\land /n) \times \nu(B(0,3\eta/4)) \times \exp[-t_P f_\lor n \nu(\mathbb{R}^d)] \geq (t_J f_\land d_W/n) \times \nu(B(0,3\eta/4)) \times \exp[-t_P f_\lor \nu(\mathbb{R}^d)] := c_X, \quad (6.24)$$

where the lower-bound $c_X$ is independent of $x_0$ and $y_0$.

Thanks to Lemma 6.2.1 (recall the definitions of $y_\land$ and $y_\lor$ at the beginning of this subsection),

$$P^{G}_{y_0}(Y_{t_P} \in dy ; t_P < T^Y) \geq c_B 1_{[1/n,n]}(y) dy. \quad (6.25)$$

Again, $c_B$ is independent of $x_0$ and $y_0$.

With (6.21), (6.23), (6.24), (6.25), Fact 6.3.5 and setting the value: $c_P := c_G c_X c_B d_W > 0$:

$$\forall x_0 \in B(x_I, \eta/4), \forall y_0 \in [1/n,n], \quad P^{G}_{(x_0,y_0)}((X,Y)_{t_P} \in (dx, dy) ; t_P < T(n)) \geq c_P 1_{B(x_I,3\eta/4)}(x) 1_{[1/n,n]}(y) dx dy.$$  

This ends the proof of Proposition 6.3.2  \hfill \Box
Proof of Lemma 6.3.1 The proof of Lemma 6.3.1 relies on similar principles as the one of Proposition 6.3.2. In this case, $t_P$ is to be replaced by $u \in [0,u_\ell(x_I)]$ and the event under consideration is simply the following:

$$\mathcal{W}' := \{ u < T^Y \} \cap \{ M([0,u] \times \mathbb{R}^d \times [0,f_\ell] \times [0,n]) = 0 \}.$$  

The reasoning given for Step 2.1. can be applied to prove that for any $t \leq u$, $X_t := x_0 - v t \mathbf{e}_1$. We also exploit Theorem 6.1 for the dependence property between $X$ and $Y$ under $\mathbb{P}^{G(x_t,y_t)}$. Lemma 6.2.1 to control the diffusion along the $Y$ coordinate. Note that $c_B$ can be taken independently of $x_I, y_I$ and $t$ (noting that $t$ is uniformly upper-bounded by $2n$). These arguments conclude the proof of the lower-bound on the marginal of $(X,Y)$ on the event \{t < T(n)\}.

Application to the various sets of assumptions

6.4 Proof of Theorem 4.2 under Assumption (D)

We treat in this subsection the mixing of $X$ when both advantageous and deleterious mutations are occurring. More precisely, each step corresponds to each of the following Lemmas:

**Lemma 6.4.1.** Assume that Assumption (H) and (D) hold. Then, for any $m \geq 3$, we can find $n \geq m$ and $c,t > 0$ such that $B(0,m)$ is included in $R^{(n)}(t,c)$.

**Lemma 6.4.2.** Assume that Assumption (H) and (D) hold. Then, there exists $n \geq 3$ which satisfies the following property for any $t_1, t_2 > 0$. There exists $t_R > t_1$ and $c_R > 0$ such that for any $t \in [t_R,t_R + t_2]$ and $(x_0,y_0) \in A \times [2,3]$:

$$\mathbb{P}^{(x_0,y_0)}[(X,Y)_t \in (dx,dy) \cap t < T(n)] \geq c_R \mathbf{1}_A(x) \mathbf{1}_{[2,3]}(y) dx dy.$$  

**Lemma 6.4.3.** Assume that Assumption (H) and (D) hold. Then, for any $\ell_I > 0$, there exists $c_I,t_I > 0$ and $n \geq t_I$ such that:

$$\forall (x,y) \in D_{t_I}, \quad \mathbb{P}^{(x,y)}(\tau_A \leq t_I \cap T(n)) \geq c_I.$$  

In the following Subsections, we prove these three lemmas then how Theorem 4.2 is deduced as a consequence of these.

6.4.1 Step 1: proof of Lemma 6.4.1

Let $x_I = -\theta \mathbf{e}_1$. Since $g$ is positive and continuous under Assumption (D), there exists $n_0$ such that $B(x_I,\eta/2)$ is included in $G_{n_0}$. With $t_0, c_0$ the values associated to $n_0$ through Proposition 6.3.2, we deduce that $x_I \in R^{(n_0)}(t_0,c_0)$.

For $m \geq 3$, let $K := \lfloor 4 \|m + \theta\|/\eta \rfloor + 1$. Similarly, we can choose $n_1$ such that $B(0,m)$ is a subset of $G_{n_1}$. Consider any $x_F \in B(0,m)$ and for $0 \leq k \leq K$, let $x_k := -\theta \mathbf{e}_1 + k/K(x_F + \theta \mathbf{e}_1)$. This choice is made to ensure that $d(x_k,x_{k+1}) \leq \eta/4$ and $\forall k \leq K$, $x_k \in G_{n_1}$. Thanks to Corollary 6.3.3 we deduce by immediate induction over $k \leq K$ that there exist $n_2, t_k, c_k > 0$ independent of $x_F$ such that: $x_k \in R^{(n_2)}(t_k,c_k)$. $t_k$ and $c_k$ are of the form $t_k := t_0 + k t_P$ and $c_k : c_0 \times (c_P)^k$. In particular with $k = K$, and $n := n_2$, Lemma 6.4.1 is proved.  

□
6.4.2 Step 2: proof of Lemma 6.4.2

We keep $x_t := -\theta e_1$ and $x_1 := (\theta + \eta/2) e_1$. Thanks to Lemma 6.4.1, there exists $n, t_1, c_1 > 0$ such that:

$$ \{ x_t + u e_1 ; \ u \in [\eta/6, 5\eta/6] \} \subset \mathcal{R}(n)(t_1, c_1). $$

There exists $t_2, c_2 > 0$ thanks to Lemma 6.3.1 such that: $\forall t \in [t_2, t_2 + 2\eta/(3v)]$, $x_t \in \mathcal{R}(n)(t, c_2)$.

Applying twice Corollary 6.3.3 with the knowledge that $B(x_1, \eta/2)$ is a subset of $G_n$, we deduce that there exists $t_3, c_3 > 0$ such that:

$$ \forall t \in [t_3, t_3 + 2\eta/(3v)], \ A \subset \mathcal{R}(n)(t, c_3). $$

Applying inductively Fact 6.3.4, we deduce the following for any $k \geq 1$:

$$ \forall t \in [kt_3, kt_3 + 2k \eta/(3v)], \ A \subset \mathcal{R}(n)(t, [c_3]^k). $$

Let $t_1, t_2 > 0$ and consider $k \geq 1$ sufficiently large for $kt_3 > t_1$ and $2k \eta/(3v) > t_2$ to hold. Thus, Lemma 6.4.2 is proved with this value of $n$, $t_R := kt_3$ and $c_R := [c_3]^k$. □

6.4.3 Step 3: proof of Lemma 6.4.3

As before, we can find $n \geq \ell_I$ be such that $D_{\ell_I} \subset G_n$. We go backwards in time from $A$ by defining for $t \geq 0$, $c > 0$:

$$ \mathcal{R}'(t, c) := \{ (x, y) \in G_n ; P_{(x,y)} [\tau_A \leq t \wedge T(n)] \geq c \}. $$

It is clear that $A \subset \mathcal{R}'(0, 1)$. Thanks to Proposition 6.3.2 and the Markov property, there exists $t_p, c_p > 0$ such that for any $t, c > 0$:

$$ \{ x \in G_n ; d(x, \mathcal{R}'(t, c)) \leq \eta/4 \} \subset \mathcal{R}'(t + t_p, c \times c_p). $$

Since $D_{\ell_I} \subset G_n$ is bounded, an immediate induction ensures that there exists $t_I, c_I > 0$ such that $D_{\ell_I} \subset \mathcal{R}'(t_I, c_I)$. This concludes the proof of Lemma 6.4.3 □

6.4.4 Theorem 4.2 as a consequence of Lemmas 6.4.1-3

The proof is quite naturally adapted from the one of Lemma 3.2.1 in [Ve21b]. Note that for any $n_1 \leq n_2$, $T(n_1) \leq T(n_2) \leq \tau_0$ holds a.s.

Let $\ell_I, \ell_M \geq 0$. According to Lemma 6.4.3, we can find $c_I, t_I > 0$ and $n_1 \geq \ell_I \wedge \ell_M$ such that for any $(x_I, y_I) \in D_{\ell_I}$:

$$ P_{(x_I,y_I)}(\tau_A \leq t_I \wedge T(n_1)) \geq c_I. \quad (6.26) $$

Let also $n_2 \geq n_1$, $c_R, t_R > 0$ chosen thanks to Lemma 6.4.2 to satisfy that for any $t \in [t_R, t_R + t_I]$ and $(x_0, y_0) \in A \times [2, 3]$:

$$ P_{(x_0,y_0)} [(X, Y)_t \in (dx, dy) ; t < T(n_2)] \geq c_R 1_A(x) 1_{[2,3]}(y) dx,dy. \quad (6.27) $$

Thanks to Lemma 6.4.1 since $D_{\ell_M}$ is a bounded set, we know that there exists $n \geq n_2$, $c_F$ and $t_F > 0$ such that for any $(x_0, y_0) \in A \times [2, 3]$:

$$ P_{(x_0,y_0)} [(X, Y)_{t_k} \in (dx, dy) ; t_k < T(n)] \geq c_F 1_{D_{\ell_M}}(x) 1_{[n/n, n]}(y) dx,dy. \quad (6.28) $$

The fact that $n$ is larger than $n_1$ and $n_2$ implies without difficulty that (6.26) and (6.27) hold with $n_1$ and $n_2$ replaced by $n$, which is how these statements are exploited in the following reasoning.
Let \( t_M := t_I + t_R + t_F \) and \( c_M := c_I \times c_R \times Leb(A) \times c_F \). For any \((x_I, y_I) \in D_{t_I}\), by combining \((6.27), (6.28)\) and the Markov property, we deduce that a.s. on the event \{\( \tau_A \leq t_I \cap T(n) \)\}:

\[
\mathbb{P}_{(X,Y)\mid \tau_A} \left[ (\bar{X}, \bar{Y}) \mid t_M - \tau_A \right] \in (dx, dy) \ ; \ t_M - \tau_A < \bar{T}(n) \right]
\geq c_F \times \mathbb{P}_{(X,Y)\mid \tau_A} \left[ (\bar{X}, \bar{Y}) \mid t_M - t_F - \tau_A \right] \in A \times [2, 3] \ ; \ t_M - t_F - \tau_A < \bar{T}(n) \right]
\times 1_{D_{t_M}}(x) 1_{[1/n, n]}(y) dx dy
\geq c_R \times Leb(A) \times c_F \times 1_{D_{t_M}}(x) 1_{[1/n, n]}(y) dx dy,
\]

where we exploited the knowledge that \( \tau_A \leq t_I \) to deduce that \( t_M - t_F - \tau_A \in [t_R, t_R + t_I] \). By combining this estimate with \((6.20)\) and again the Markov property, we conclude:

\[
\mathbb{P}_{(x_I, y_I)} \left[ (X_{t_M}, Y_{t_M}) \right] \in (dx, dy) \ ; \ t_M < T(n) \]
\geq \mathbb{P}_{(x_I, y_I)}(\tau_A \leq t_I \cap T(n)) \times c_R \times Leb(A) \times c_F \times 1_{D_{t_M}}(x) 1_{[1/n, n]}(y) dx dy
\geq c_M 1_{D_{t_M}}(x) 1_{[1/n, n]}(y) dx dy.
\]

This ends the proof of Theorem \(4.2\) with \( L = 2n \), \( c := c_M \) and \( t := t_M \) under Assumption \((D)\).

\[
6.5 \quad \text{Proof of Theorem } 4.2 \text{ under Assumption } (A) \text{ and } d \geq 2
\]

The proof of Theorem \(4.2\) is handled under Assumption \((A)\) and \( d \geq 2 \) in the same way as in Subsection 6.4.4. Notably, the lemmas that replace Lemmas 6.4.2-3 have identical implications:

**Lemma 6.5.1.** Assume that \( d \geq 2 \) and that Assumption \((H)\) and \((A)\) hold. Then, for any \( m \geq 3 \), we can find \( n \geq m \), \( t, c > 0 \) such that \( B(0, m) \) is included in \( R^{(k)}(t, c) \).

**Lemma 6.5.2.** Assume that \( d \geq 2 \) and that Assumption \((H)\) and \((A)\) hold. Then, there exists \( n \geq 3 \) which satisfies the following property for any \( t_1, t_2 > 0 \). There exists \( t_1 > t_2 \) and \( c_R > 0 \) such that, for any \( t \in [t_1, t_2] \) and \((x_0, y_0) \in A \times [2, 3] \):

\[
\mathbb{P}_{(x_0, y_0)} \left[ (X, Y) \right] \in (dx, dy) \ ; \ t < T(n) \right] \geq c_R 1_{A}(x) 1_{[2, 3]}(y) dx dy.
\]

**Lemma 6.5.3.** Assume that \( d \geq 2 \) and that Assumption \((H)\) and \((A)\) hold. Then, for any \( \ell_1 > 0 \), there exists \( c_1, t_I > 0 \) and \( n \geq \ell_1 \) such that:

\[
\forall (x_0, y_0) \in D_{t_I}, \quad \mathbb{P}_{(x_0, y_0)}(\tau_A \leq \tau_A \cap T(n)) \geq c_A.
\]

Since the implications are the same, the proof of Theorem \(4.2\) under Assumption \((A)\) with \( d \geq 2 \) as a consequence of Lemmas 6.5.1-3 is mutatis mutandis the same as the one given in Subsection 6.4.4. Since deleterious mutations are now forbidden, the proof of Lemma 6.5.1 is much trickier than the one of Lemma 6.4.1. The first step is given by the two following lemmas. To this purpose, given any direction \( u \) on the sphere \( S^d \) of radius 1, we denote its orthogonal component by:

\[
x^{(u)} := x - \langle x, u \rangle u, \quad \text{and specifically for } e_1: \quad x^{(1)} := x - \langle x, e_1 \rangle e_1.
\]

**Lemma 6.5.4.** Assume that \( d \geq 2 \), Assumption \((H)\) and \((A)\) hold. Then, for any \( x \geq 0 \), there exists \( \epsilon \leq \eta / 8 \) which satisfies the following property for any \( n \geq 3 \) \( \cap (2 \theta) \), \( x \in B(0, n) \) and \( u \in S^d \) such that both \( \langle x, u \rangle \geq \theta \) and \( \|x^{(u)}\| \leq x^{(1)} \). There exists \( t_p, c_p > 0 \) such that for any \( t, c > 0 \):

\[
x \in R^{(n)}(t, c) \Rightarrow B(x - \theta u, \epsilon) \subset R^{(n)}(t + t_p, c \times c_p).
\]
Lemma 6.5.5. Assume that $d \geq 2$, Assumption (H) and (A) hold. Then, for any $m \geq 3 \vee (2 \theta)$, there exists $\epsilon \leq \eta/8$ which satisfies the following property for any $x \in B(0, m)$ with $\langle x, \mathbf{e}_1 \rangle \leq 0$. There exists $t_p, c_p > 0$ such that:

$$\forall t, c > 0, \quad x \in \mathcal{R}^{(L)}(t, c) \Rightarrow B(x, \epsilon) \subset \mathcal{R}^{(L)}(t + t_p, c \times c_p).$$

Lemma 6.5.5 is actually directly implied from Lemma 6.3.1 (first applied for a time-interval $[0, \theta/v]$) then Lemma 6.5.4 with $u := \mathbf{e}_1$, combined with the Markov property. Subsection 6.5.1 is dedicated to the proof of Lemma 6.5.4

6.5.1 Step 1: proof of Lemma 6.5.4

Fix $x_0 > 0$. Consider $\epsilon > 0$ that is to be fixed later, but assume already that $\epsilon \leq \theta/8$. We recall that $\eta \leq \theta/8$ if assumed w.l.o.g. Let $n \geq 3 \vee (2 \theta)$, $x_0 \in B(0, n)$ and $u \in \mathcal{S}_d$ such that both $\langle x_0, u \rangle \geq \theta$ and $\|x_0^{(1,u)}\| \leq x_\epsilon$ hold.

Compared to Proposition 6.3.2, the first main difference is that the jump is now almost instantaneous. The second is that, in order that $g_\lambda > 0$, we have way less choice in the value of $w$ when $\|x_0^{(1,u)}\|$ is large. In particular, the variability of any particular jump will not be sufficient to wipe out the initial diffusion around $x$ deduced from $x \in \mathcal{R}^{(w)}(t, c)$, but will rather make it even more diffuse.

To fix $\epsilon > 0$, let us first compute, for $\delta \in B(0, \eta)$, $w \in B(-\theta u, \epsilon)$:

$$\|x_0 + \delta\| - \|x_0 + \delta + w\|^2 = 2 \langle x_0 + \delta, w \rangle - \|w\|^2$$

$$\geq \left(\frac{7}{4} - \frac{9}{8} \times \left(\frac{1}{4} + \frac{9}{8}\right)\right) \theta^2 - 2 \epsilon x_\epsilon,$$

where we exploited that $\langle u, w \rangle \geq 7 \theta/8$. We note that:

$$c := \frac{7}{4} - \frac{9}{8} \times \left(\frac{1}{4} + \frac{9}{8}\right) = \frac{13}{64} > 0.$$

By taking $\epsilon := \{c\theta^2/(4x_\epsilon)\} \wedge \{\theta/8\}$, we thus ensure that $\|x_0 + \delta\|^2 > \|x_0 + \delta + w\|^2$. Note that $\epsilon$ does not depend on the specific choice of $x_0$.

Let $t_p := \epsilon/(2 \nu)$. The initial condition for $X, Y$ is taken as $x_I \in B(x, \eta/2)$ and $y_I \in [1/n, n]$.

$$g_\lambda := \inf \{g(x, w) : x \in B(x_0, \eta), w \in B(-\theta u, \epsilon)\} > 0,$$

$$X^M := [0, t_p] \times \mathbb{R}^d \times [0, f_\nu] \times [0, n],$$

$$\mathcal{J} := [0, t_p] \times B(-\theta u + (\epsilon/2) \mathbf{e}_1, \epsilon/2) \times [0, f_\lambda] \times [0, g_\lambda].$$

With the same reasoning as in the proof of Proposition 6.3.2 we obtain a change of probability $\mathbb{P}_G^{(x_I, y_I)}$ and an event $W$ on which the r.v. $W$ is uniquely defined from $M$ under $\mathbb{P}_G^{(x_I, y_I)}$ and such that it satisfies a.s.:

$$X_{t_p} = x_I - (\epsilon/2) \mathbf{e}_1 - \theta u + (\epsilon/2) e_1 + W = x_I - \theta u + W,$$

where the density of $W$ is lower-bounded by $dW$ on $B(0, \epsilon/2)$, uniformly over $x_I$ (given $x$), and $y_I$. We thus similarly obtain some constants $c_p, c'_p > 0$ independent of $x_0$ such that for any such $x_0$:

$$\int_{B(x_0, \eta/2)} \int_{[1/n, n]} dxdy \mathbb{P}_{(x_I, y_I)} \left[\left(X, Y\right)_{t_p} \in (dx, dy) : t_p < T(n)\right]$$

$$\geq c_p \int_{B(x_0, \eta/2)} dxdy \mathbb{1}_{B(x_0 - \theta u, \epsilon/2)}(x) \times \mathbb{1}_{[1/n, n]}(y) dx dy$$

$$\geq c'_p \int_{B(x_0 - \theta u, \eta/2 + \epsilon/3)}(x) \times \mathbb{1}_{[1/n, n]}(y) dx dy.$$
We then reason similarly as in the proof of Corollary 6.3.3 as a consequence of Proposition 6.3.2.

Assuming further that \( x_0 \in \mathcal{R}(n)(t,c) \) for some \( t,c > 0 \), we can deduce:

\[
B(x - \theta u, \epsilon/3) \in \mathcal{R}(n)(t + t_P, c \times c_P).
\]

This is exactly the implication of Lemma 6.5.4 stated in terms of \( \epsilon/3 \) instead of \( \epsilon \).

\[ \square \]

6.5.2 Step 2: Lemma 6.5.1 as a consequence of Lemmas 6.5.5 and 6.5.4

**Step 2.1:** \( x_I \in \mathcal{R}(n)(t_0, c_0) \). Let \( x_I := -\theta e_1 \). We check that there exists \( n_1 \geq 1 \) such that \( B(x_I, \eta/2) \) is a subset of \( \mathcal{G}_{n_1} \). Since \( g \) is continuous and thanks to (A), it is sufficient to prove that

\[
\|x_I - ze_1 + \delta\| > \|x_I - ze_1 + \delta + w\|
\]

for any \( z \in [0, \theta] \), \( \delta \in B(0, \eta) \), and \( w \in B(\theta e_1, \eta) \):

\[
\|x_I - ze_1 + \delta\|^2 - \|x_I - ze_1 + \delta + w\|^2 = 2((\theta + z)e_1 - \delta, w) - \|w\|^2 \\
\geq 2[\theta(\theta - \eta) - \eta(\theta + \eta)] - (\theta + \eta)^2 \\
= \theta^2 - 6\theta \eta - 3\eta^2 \geq \frac{13\theta^2}{64} > 0,
\]

since \( \eta \leq \theta/8 \), as assumed above, just after (6.14). Applying twice Proposition 6.3.2 it concludes that there exists \( n_0 \geq 1, t_0, c_0 > 0 \) such that \( x_I \in \mathcal{R}(n)(t_0, c_0) \).

**Step 2.2: under the condition that \( \langle x_F, e_1 \rangle := -\theta \).** The purpose of this step is the following lemma, in which we employ the notation \( \pi_1 : x \mapsto \langle x, e_1 \rangle \).

**Lemma 6.5.6.** For any \( n \geq 1 \) sufficiently large, there exists \( t, c > 0 \) such that \( \pi_1(\theta) \cap B(0, n) \) is a subset of \( \mathcal{R}(n)(t,c) \).

Let \( x_F \in \pi_1(\theta) \cap B(0, n) \), where we assume that \( n \) is larger than \( n_0, 3 \) and 29. First, we define \( u \) as \( e_1 \) if \( x_F^{(1)} = 0 \) and else as \( u := x_F^{(1)}/\|x_F^{(1)}\| \). Note that \( \|x_F^{(1)}\| \leq n \). We consider the value of \( \epsilon \) given by Lemma 6.5.5 for \( x_v := n \) and define:

\[
K := [n\epsilon] + 1, \quad \text{for } 0 \leq k \leq K, \quad x_k := -\theta e_1 + \frac{k\|x_F^{(1)}\|}{K} u.
\]

This choice ensures that for any \( k \in [0, K - 1] \), \( x_{k+1} \in B(x_k, \epsilon) \), while \( x_k \in B(0, n) \), \( \langle x_k, e_1 \rangle \leq 0 \) and \( x_K = x_F \). Thanks to Step 2.1, \( x_0 \in \mathcal{R}(n)(t_0, c_0) \). Thus, by induction over \( k \leq K \) with Lemma 6.5.5 \( x_k \in \mathcal{R}(n)(t_0 + kt_P, c_0 [cP]^k) \). In particular, there exists \( t, c > 0 \) such that \( x \in \mathcal{R}(n)(t,c) \), which concludes Step 2.2.

**Step 2.3: the general case** Assume solely that \( x \in B(0, m) \). We consider the value of \( \epsilon \) given by Lemma 6.5.3 for \( x_v := m \). The choice of \( u \) is as in Step 2.2.

Let:

\[
K := \left\lfloor \frac{m + \theta}{\epsilon} \right\rfloor + 1, \quad \text{so that } \frac{\langle x, e_1 \rangle + \theta}{K} \leq \epsilon, \quad (6.31)
\]

and for \( 0 \leq k \leq K \), \( x_k := (-\theta + (k/K) \times ((x, e_1) + \theta)) e_1 + (K - k) \theta u + x^{(1)}(x) \).

In particular \( \langle x_0, e_1 \rangle = -\theta \), \( x_K = x_F \) while for any \( k \leq K - 1 \), \( x_{k+1} \in B(x_k, \epsilon) \), \( x_k \in B(0, m + K \theta) \) and \( \langle x_k, u \rangle \leq \theta \vee \langle x_k, e_1 \rangle \leq m = x_v \).

Since \( \langle x_0, e_1 \rangle = -\theta \), we can exploit Lemma 6.5.6 to prove that there exists \( n \geq 1 \) and \( t_0, c_0 > 0 \) independent of \( x_F \) such that \( x_0 \in\mathcal{R}(n)(t_0, c_0) \). Thanks to Lemma 6.5.4 and induction on \( k \), we deduce that there exist \( t_P, c_P > 0 \) such that:

\[
x_k \in \mathcal{R}(n)(t_0 + kt_P, c_0 [cP]^k). \quad \text{In particular, there exists } t, c > 0 \text{ such that } x_F \in \mathcal{R}(n)(t,c).
\]

\[ \square \]
6.5.3 Step 3: proof of Lemma [6.5.2]

The proof can be taken mutatis mutandis from the one given in Subsection [6.4.2]. The fact that \( B(x_1, \eta/2) \) is a subset of \( \mathcal{G}_n \) is already proved in Step 2.1 (cf Subsection [6.5.2]), while Lemma [6.5.1] replaces Lemma [6.4.1] with identical implication.

6.5.4 Step 4: proof of Lemma [6.5.3]

**Remarks 6.5.7.** The presented proof efficiently exploits the already known lemmas but is probably very far from optimal in its estimations.

**Step 4.1: study of \( \mathcal{G}_n \).** We look for conditions on \( x \in \mathbb{R}^d \) that ensure that it belongs to \( \mathcal{G}_n \) for some \( n \). Let \( x_0 := x - (\theta - \eta/2)e_1 \). By definition of \( \mathcal{G}_n \), it is necessary that for any \( z \in [0, \eta/4] \), \( \delta \in B(0, \eta/2) \) and \( w \in B(\theta e_1, \eta) \) \( g(x_0 - z e_1 + \delta, w) > 0 \) which under \( (A) \) is equivalent to \( \|x_0 - z e_1 + \delta\| \geq \|x_0 + z e_1 + \delta + w\| \). We first restrict ourselves to the values of \( x \) such that \( \pi_1(x) \leq 0 \), and we compute:

\[
\|x_0 - z e_1 + \delta\|^2 - \|x_0 + z e_1 + \delta + w\|^2 = -2 \langle x_0 + z e_1 + \delta, w \rangle - \|w\|^2 \\
\geq 2 (\pi_1(x_0) - \eta/2) + (\theta - \eta) - 2 (\|x^{(1)}\| + \eta/2) \times \theta - (\theta + \eta)^2 \\
\geq -7 \pi_1(x_0)/32 - \|x^{(1)}\|/4 \times \theta + (7/4) \times (\theta - \eta/2) \times (\theta - \eta) - \eta^2 - (\theta + \eta)^2. \\
\geq -7 \pi_1(x_0)/32 - \|x^{(1)}\|/4 \times \theta + (7 \times 15 \times 7 - 8 \times 7 - 8 \times 8 \times 1) \theta^2/2^9 \\
\geq -7 \pi_1(x_0)/32 - \|x^{(1)}\|/4 \times \theta + 23 \theta^2/2^9.
\]

From these computations, we see that \( g(x_0 - z e_1 + \delta, w) > 0 \) holds true provided \( \pi_1(x) \leq 0 \) and \( \|x_1^{(1)}\| \geq 8 \|x^{(1)}\|/7 \) thus a fortiori if \( |\pi_1(x)| \geq 8 \|x^{(1)}\|/7. \) Since \( g \) is continuous, we deduce that for any \( m \geq 1 \), there exists \( n \geq 1 \) such that \( \mathcal{G}_n \) contains the following set:

\[
\{x \in B(0, m) ; -\pi_1(x) \geq 8 \|x^{(1)}\|/7}\text{.}
\]

**Step 4.2.** Let \( \ell_I \geq 1 \). Thanks to Step 4.1, we can find \( n \geq \ell_I \vee 3 \) such that \( \mathcal{G}_n \) contains the following set:

\[
A_1 := \{x \in B(0, 2\ell_I) ; -\pi_1(x) \geq 8 \|x^{(1)}\|/7\}.
\]

We go backwards in time from \( A \) by defining for \( t \geq 0, c > 0 \):

\[
\mathcal{R}'(t, c) := \{(x_1, y_1) \in \mathcal{G}_n \ ; \ P_{(x, y)}(r_A \leq t \wedge T) \geq c\}.
\]

Similarly as for the proof of Lemma [6.4.3] by exploiting inductively Proposition [6.3.2] we deduce that \( A_1 \) is a subset of \( \mathcal{R}'(t_1, c_1) \) for some \( t_1, c_1 > 0 \).

Consider now any \( x_I \in B(0, \ell_I) \). If \( x_I \notin A_1 \), let \( u_* := 8 \|x^{(1)}\|/(7v) + \pi_1(x)/v \) and \( x_I := x - v u_* e_1 \in A_1 \). If \( x_I \in A_1 \), we simply define \( x_I := x_I \) and \( u_* := 0 \). Since \( \|x^{(1)}\| \leq \ell_I \), this choice necessarily satisfies \( 0 \leq -\pi_1(x_I) = 8 \|x^{(1)}\|/7 \leq 8n/7 \). In any case, \( x_I \in B(0, 2\ell_I) \) thus \( x_I \in A_1 \). Since \( A_1 \subset \mathcal{R}'(t_1, c_1) \) and thanks to Lemma [6.3.1] there exists a value \( c_D > 0 \) uniform over \( x \) such that \( x_I \in \mathcal{R}'(t_1 + u_*c_1 c_D) \). Since \( u_* \) is upper-bounded by \( 2\ell_I \) and the set \( \mathcal{R}'(t, c) \) is increasing with \( t \), it concludes that \( B(0, \ell_I) \) is a subset of \( \mathcal{R}'(t_2, c_2) \) with \( t_2 := t_1 + 2\ell_I \) and \( c_2 := c_1 c_D \). This ends the proof of Lemma [6.5.3].

As mentioned at the beginning of Subsection [6.5], the last step of the proof of Theorem [4.2] can be taken mutatis mutandis from Subsection [6.4.4]. With this, the proof of the theorem is complete.

33
6.6 Proof of Theorem 4.5

We treat in this subsection the mixing for $X$ when simply advantageous mutations are occurring and the phenotype is unidimensional. The proof of Theorem 4.2 is handled under Assumption (A) and $d \geq 2$ in the same way as in Subsection 6.4.4 except that Lemmas 6.4.2, 3 are replaced by the following ones, in respective order. Note that only the first one has a different implication.

Lemma 6.6.1. Assume that $d = 1$, Assumption (H) and (A) hold. Then, for any $m \geq 3$, there exists $n \geq m$, $t, c > 0$ such that $[-m, 0]$ is included in $R^{(n)}(t, c)$.

Lemma 6.6.2. Assume that $d = 1$ and that Assumption (H) and (A) hold. Then, there exists $n \geq 3$ which satisfies the following property for any $t_1, t_2 > 0$. There exists $t_R > t_1$ and $c_R > 0$ such that, for any $t \in [t_R, t_R + t_2]$ and $(x_0, y_0) \in A \times [2, 3]$:

$$\mathbb{P}_{(x_0, y_0)}[(X, Y)_t \in (dx, dy) ; t < T_{(n)}] \geq c_R 1_A(x) 1_{[2,3]}(y) dx dy.$$ 

Lemma 6.6.3. Assume that $d = 1$ and that Assumption (H) and (A) hold. Then, for any $\ell_1 > 0$, there exists $c, t > 0$ and $n \geq \ell_1$ such that:

$$\forall (x_1, y_1) \in D_{\ell_1}, \quad \mathbb{P}_{(x_1, y_1)}(\tau_A \leq t \land T_{(n)}) \geq c.$$

(6.32)

Step 1: proof of Lemmas 6.6.1 and 6.6.2. Consider the calculations given in Step 4.1, Subsection 6.5.4, in this case where there is no contribution from $x^{(1,1)}$, we can conclude that for any $m$, there is $n \geq m$ such that $[-m, 0]$ is included in $G_n$. Adapting the reasoning given resp. in Subsections 6.4.1 and 6.4.2, we can directly conclude the proof of Lemmas 6.6.1 and 6.6.2.

Note that the set first introduced in the proof of Lemma 6.4.2 here takes the form $[-1, \theta + \eta/6, -\theta + 5\eta/6]$. It is included in $[-m, 0]$ for any choice of $m \geq \theta$, so that Lemma 6.6.1 can indeed replace Lemma 6.4.1. □

Step 2: proof of Lemma 6.6.3. Let $(x_1, y_1) \in D_{\ell_1}$.

Case 1: $x_1 \geq -\theta$. Thanks to Lemma 6.3.1 with $u := x_1 + \theta$, there exits $t_+, c_+ > 0$ which satisfies the following property for any $(x_1, y_1) \in D_{\ell_1}$ such that $x_1 \geq -\theta$:

$$\mathbb{P}_{(x_1, y_1)}(\tau_A \leq t_+ \land T_{(n)}) \geq c_+.$$

Case 2: $x_1 < -\theta$. We recall from the proof of Lemmas 6.6.1 that there exists $n \geq 1$ such that $[-\ell_1, 0]$ is included in $G_n$. In this set, the proof of Lemma 6.4.3 given in Subsection 6.4.3 can be directly exploited to prove that there exists $t_-, c_- > 0$ which satisfy the following property for any $(x_1, y_1) \in D_{\ell_1}$ such that $x_1 \leq 0$:

$$\mathbb{P}_{(x_1, y_1)}(\tau_A \leq t_- \land T_{(n)}) \geq c_-.$$

The combination of these two cases with $t := t_+ \lor t_-$ and $c := c_+ \lor c_-$ concludes the proof of Lemma 6.6.3. □

Step 3: concluding the proof of Theorem 4.5. By replacing Lemmas 6.4.1, 6.4.2 and 6.4.3 by Lemmas 6.6.1, 6.6.2 and 6.6.3 in the proof given in Subsection 6.4.4, it is clear that the conclusion of Theorem 4.5 is reached. □
7 Absorption with failures

7.1 Proof of Theorem 4.4 in the case \( d = 1 \)

7.1.1 Definition of the stopping time and its elementary properties

We consider a first process \((X, Y)\) with some initial condition \((x, y) \in E\).

We will prove that considering \( U_A = t_\partial \) is sufficient, except for exceptional behavior of the process. Given \( \epsilon, \rho > 0, \), \( t_\partial \) shall be chosen sufficiently small to ensure that, with probability close to 1 (the thresholds depending on \( \epsilon \) and \( \rho \)), no jump has occurred before time \( t_\partial \), and that the population size has not changed too much. We define:

\[
\delta y := (3\ell_E/\ell_E + 1)^{-1}, \quad y_\wedge := 1/(\ell_E + 1) = \ell_E + 1 > \ell_E + 3\delta y, \quad y_\vee := \ell_E + 1 > \ell_E + 3\delta y,
\]

\[
T_{\delta y} := \inf \{ t \geq 0 ; |Y_t - y_E| \geq 2\delta y \} < \tau_\partial. \quad (7.1)
\]

We recall that we can upper-bound the first jump time of \( X \) by:

\[
T_J := \inf \{ t \geq 0 ; M([0, t] \times J) \geq 1 \}, \quad (7.2)
\]

where \( J \) is defined as in Subsection 6.1.

• On the event \( \{ t_\partial < T_{\delta y} \wedge T_J \wedge \tau_\partial \} \), we set \( U_A := t_\partial \).

• On the event \( \{ T_{\delta y} \wedge T_J \wedge \tau_\partial \leq t_\partial \} \), we set \( U_A := \infty \).

Before we turn to the details of the proof of Theorem 4.4, we first give the main scheme for proving the following lemma, noting that we will not go too deeply in the details of this proof.

**Lemma 7.1.1.** We can define a stopping time \( U_A^\infty \) extending the above definition of \( U_A \) as described in Theorem 4.4.

7.1.2 Step 1: main argument for the proof of Lemma 7.1.1

Recall (with simplified notations) that considering the process \((X, Y)\) with initial condition \((x, y)\), we define for some \( t > 0 : \ U_A := t \) on the event \( \{ t < T_{\delta y} \wedge T_J \} \), \( U_A := \infty \) otherwise,

where \( T_{\delta y} := \inf \{ s \geq 0 ; |X_s - y| \geq 2\delta y \} < \tau_\partial, \) for some \( \delta y > 0, \)

\[
T_J := \inf \{ s \geq 0 ; M([0, s] \times J) \geq 1 \},
\]

\[
J := \mathbb{R}^d \times [0, f_\vee] \times [0, g_\vee]
\]

for some \( f_\vee, g_\vee > 0. \)

Recursively, we also define:

\[
\tau_E^{i+1} := \inf \{ s \geq \tau_E^i + t : X_s \in E \} \wedge \tau_\partial, \text{ and } \tau_E^0 = 0,
\]

and on the event \( \{ \tau_E^i < \tau_\partial \} \), for any \( i \), we set:

\[
T_{\delta y}^i := \inf \{ s \geq \tau_E^i ; |X_s - Y_s(\tau_E^i)| \geq 2\delta y \},
\]

\[
U_j^i := \inf \{ s \geq 0 ; M([\tau_E^i, \tau_E^i + s] \times J) \geq 1 \},
\]

\[
U_A^\infty := \inf \{ \tau_E^i + t ; t \geq 0, \tau_E^i < \infty, \tau_E^i + t < T_{\delta y}^i \wedge U_j^i \},
\]

where in this notation, the infimum equals \( \infty \) if the set is empty, \( T_{\delta y}^i := \infty \) and \( U_j^i = \infty \) on the event \( \{ \tau_\partial \leq \tau_E^i \} \).
We then identify large, and 1 \leq \tau \leq 6.2.1, we deduce a lower-bound on the density of \( Y \) and the reader is spared the details. The main point is that there is a.s. a positive gap between any \( \ell \) as in the proof of Lemma 6.3.1, we have chosen our stopping times to ensure that no jump for \( D_t \) at time \( \tau \leq \infty \) under \( P \), is indeed sufficiently diffused (since we care now for \( \delta y \geq 1 \)).

\[
\tau^i \leq \tau^i_{E(n)} \leq \tau^i_{E(n)} + 1/n + \tau^i_{E(n)} + 1/n.
\]

It is obvious that \( U_A^\infty \) coincide with \( U_A \) on the event \( \{ U_A \wedge \tau^0 \leq \tau^1_E \} \), while the Markov property at time \( \tau^i_E \) and the way \( U_A^\infty \) is defined entails that on the event \( \{ \tau^1_E < U_A \wedge \tau^0 \} \), \( U_A^\infty - \tau^i_E \) has indeed the same law as \( U_A^\infty \) associated the process \( (X, Y) \) solution of the system (4.8) with initial condition \( (X(\tau^1_E), Y(\tau^1_E)) \).

### 7.1.3 Step 2: end of the proof of Theorem 4.4 when \( d = 1 \)

Let \( \ell_E \geq 1, \epsilon, \rho > 0 \) be prescribed. We first require \( t_{\tau} \leq 1 \) to be sufficiently small.

Note that our definitions ensure that for any \( t < t_{\tau} \wedge T_{\delta y} \wedge T_J \), we have a.s.:

\[
(X_t, Y_t) \in [-\ell_E - 1, \ell_E] \times [y, y].
\]

Thanks to Theorem 6.1 with some constant \( C_G \) uniform over any \( (x_E, y_E) \in E \):

\[
P_{(x_E, y_E)}(T_{\delta y} < t_{\tau} \wedge T_J) \leq C_G \mathbb{P}^G_{(x_E, y_E)}(T_{\delta y} < t_{\tau} \wedge T_J) \leq C_G \mathbb{P}^G_0(T_{\delta y} < t_{\tau}) \rightarrow 0 \text{ as } t_{\tau} \rightarrow 0,
\]

where \( T_{\delta y} \) under \( \mathbb{P}^G_0 \) denotes the first time the process \( B \) reaches \( \delta y \), with \( B \) a standard Brownian Motion. Moreover:

\[
P_{(x_E, y_E)}(T_J < t_{\tau} \wedge T_{\delta y}) \leq \mathbb{P}(M([0, t_{\tau}] \times J) \geq 1) \leq \nu(\mathbb{R}) f_{\ell_E} t_{\tau} \rightarrow 0 \text{ as } t_{\tau} \rightarrow 0.
\]

By choosing \( t_{\tau} \) sufficiently small, we can thus ensure the following property for any \( (x_E, y_E) \in E \):

\[
P_{(x_E, y_E)}(U_A = \infty, t_{\tau} \leq \tau_0) \leq \mathbb{P}_{(x_E, y_E)}(T_{\delta y} < t_{\tau} \wedge T_J) + \mathbb{P}_{(x_E, y_E)}(T_J < t_{\tau} \wedge T_{\delta y}) \leq \epsilon e^{-\rho \tau} \leq e \exp(-\rho t_{\tau}). \tag{7.3}
\]

On the event \( \{ t_{\tau} < T_{\delta y} \wedge T_J \} \), we have: \( X_{U_A} = x_E - v t_{\tau} \) and \( Y_{U_A} \in [y_E - \delta y, y_E + \delta y] \). Indeed, as in the proof of Lemma 6.3.1 we have chosen our stopping times to ensure that no jump for \( X \) can occur before time \( T_J \wedge t_{\tau} \wedge T_{\delta y} \). We also rely on the Girsanov transform and Theorem 6.1 to prove that, during the time-interval \([0, t_{\tau}]\), \( Y \) is indeed sufficiently diffused (since we care now for an upper-bound, we can neglect the effect of assuming \( t_{\tau} < T_{\delta y} \)). It makes us conclude that there exists \( D_X > 0 \) such that for any \( x_E \in [-\ell_E, \ell_E] \) and \( y_E \in [1/\ell_E, \ell_E] \):

\[
P_{(x_E, y_E)}([X, Y](U_A) \in (dx, dy) ; U_A < \tau_0] \leq D_X 1_{[y_E-2\delta y, y_E+2\delta y]}(y) \delta_{x_E-v t_{\tau}}(dx) dy \tag{7.4}
\]

With \( \zeta \) the uniform distribution over \( D_1 \), thanks to Theorem 4.2 there exists \( c_M, t_M > 0 \) such that:

\[
P_{\zeta}([X, Y]_{t_M} \in (dx', dy')] \geq c_M 1_{\{(x', y') \in D_{t_M}\}} dx' dy'.
\]

The idea is then to let \( X \) decrease until it reaches \( x_E - v t_{\tau} \) by ensuring that no jump occurs. We then identify \( u \) as the time needed for this to happen. Then, thanks to Theorem 6.1 and Lemma 6.2.1 we deduce a lower-bound on the density of \( Y \) on \([y_E-2\delta y, y_E+2\delta y]\). We have already proved
a stronger result for Lemma 6.3.1 that we let the reader adapt to obtain the following property. For any \( t_\tau > 0 \), there exists \( dx \) which satisfies the following property for any \( x_E \in [-\ell_E, \ell_E] \). There exists a stopping time \( V \) such that:

\[
\mathbb{P}_{\xi} [(X, Y)(V) \in (dx, dy)] \geq d^X \ c_M \ 1_{[y_E - 2 \delta y, y_E + 2 \delta y]}(y) \ \delta_{x_E - v \ t_\tau}(dx) \ dy. \tag{7.5}
\]

The proper definition of \( V \) is given by \( V := t_M + t_\tau + (X_{t_M} - x_E)/v \geq t_M \) on the event \( \{X_{t_M} \in [x_E, x_E + v \ t_\tau]\} \cap \{Y_{t_M} \in [y_E - \delta y/2, y_E + 2 \delta y/2]\} \) (and can be made arbitrary as \( t_M \) otherwise).

Thanks to Lemma 7.1.1, 7.3, 7.4 and 7.5, we conclude the proof of Theorem 4.4 with \( c := D^X/(dx^X \ c_M) \).

\[\square\]

### 7.2 Proof of Theorem 4.7

Except that we exploit Theorem 4.5 instead of 4.2, which constrains the shape of \( E \), the proof is immediately adapted from the previous Subsection 7.1.

\[\square\]

### 7.3 Proof of Theorem 4.4 in the case \( d \geq 2 \)

The difficulty in this case is that, as long as no jump has occurred, \( X_t \) stays confined in the line \( x + \mathbb{R}_+ \mathbf{e}_1 \). The “absorption” thus cannot occur before a jump. Thus, we first wait for a jump to diffuse on \( \mathbb{R}^d \) and then let \( Y \) diffuse independently in the same way as in Subsection 7.1. These two steps are summarized in the following:

**Proposition 7.3.1.** Given any \( \rho > 0 \), \( E \in D \) and \( \epsilon_X \in (0, 1) \), there exists \( t^X \), \( c^X \), \( x^X < 0 \) and \( 0 < y^X \) which satisfies the following property for any \( (x_E, y_E) \in E \). There exists a stopping time \( U^X \) such that:

\[
\{\tau_\theta \wedge t^X \leq U^X \} = \{U^X = \infty\}, \quad \mathbb{P}_{(x_E,y_E)}(U^X = \infty, t^X < \tau_\theta) \leq \epsilon_X \exp(-\rho t^X),
\]

and \( \mathbb{P}_{(x_E,y_E)}(X(U^X) \in dx \setminus Y(U^X) \in [y^X, y^X], U^X < \tau_\theta) \leq c^X \ 1_{B(0,x^X)} \approx (x) \ dx \).

We defer the proof in Subsection 7.3.2.

**Proposition 7.3.2.** Given any \( \rho \), \( x^Y > 0 \), \( 0 < y^X < y^Y \) and \( \epsilon_Y \in (0, 1) \), there exists \( t^Y \), \( c^Y > 0 \) and \( 0 < y^X < y^Y \) which satisfies the following property for any \( (x, y) \in \mathbb{R} \times [y^X, y^Y] \). There exists a stopping time \( T^Y \) such that:

\[
\mathbb{P}_{(x,y)}(T^Y < t^Y \wedge \tau_\theta) \leq \epsilon_Y \exp(-\rho t^Y),
\]

and \( \mathbb{P}_{(x,y)}(X(Y)(t^Y) \in (dx, dy) \setminus (X^Y < t^Y \wedge \tau_\theta) \leq c^Y \delta_{(x-v \ t^Y \mathbf{e}_1)}(dx) 1_{[y^Y, y^Y]}(y) \ dy.\)

The proof of Lemma 7.3.2 is taken mutatis mutandis from the one in Subsection 7.1. It leads to define \( U_A \) as below.

- \( U_A := U^X + t^Y \) on the event \( \{U^X < t^X \wedge \tau_\theta \} \cap \{t^Y < \tilde{\tau}_\theta \wedge \tilde{T}^Y \} \), where \( \tilde{\tau}_\theta \) and \( \tilde{T}^Y \) are defined as respectively \( \tau_\theta \) and \( T^Y \) for the solution \((\hat{X}_t, \hat{Y}_t)\), defined on the event \( \{U^X < t^X \wedge \tau_\theta \} \), of:

\[
\begin{align*}
\hat{X}_t &= X(U^X) - v t \mathbf{e}_1 + \int_{[0, u)} \mathbb{R}^d \times \mathbb{R}^d \ M(ds, dw, du, u_y) \ w \varphi \left( \hat{X}_s, \hat{Y}_s, w, u_f, u_y \right) \ ds + \int_{U^X} dB_c,
\end{align*}
\]

- Else \( U_A := \infty \).

**Fact 7.3.3.** There exists a stopping time \( U_A^\infty \) extending the above definition of \( U_A \) as described in Theorem 4.4 (with \( t = t^X + t^Y \) here).

The proof of Fact 7.3.3 is technical but classical from the way we define \( U^X \) and \( T^Y \) and similar to the proof of Lemma 7.1.1. The reader is spared this proof.
7.3.1 Proof of Theorem 4.4 [4.4] as a consequence of Propositions 7.3.1-2 and Fact 7.3.3

Given \( \rho > 0, \epsilon \in (0, 1) \) and some \( E \in D \), we define \( \epsilon_X := \epsilon/4 \) and deduce from Proposition 7.3.1 the values \( t^X, c^X, x^\Delta, y^\Delta_x, y^\Delta_y \) and the definition for the stopping times \( U^X \) with the associated properties.

With \( \epsilon_Y := \epsilon \exp(-\rho t^X)/2 \), we then deduce from Proposition 7.3.2 the values \( t^Y, c^Y, y^\Delta_x, y^\Delta_y \) and the stopping time \( T^Y \) with the associated properties. Defining, for some \((x, y) \in E, U_A \) as in Fact 7.3.3 and combining these results:

\[
\{ \tau_0 \land (t^X + t^Y) \leq U_A \} = \{ U_A = \infty \},
\]

\[
P_{(x,y)} [(X,Y) (U_A) \in (dx, dy) ; U_A < \tau_0] 
\leq c^X c^Y 1_{B(0, x^\Delta + Y^\Delta)} (x) 1_{[y^\Delta_x, y^\Delta_y]} (y) \; dx \; dy,
\]

\[
P_{(x,y)} (U_A = \infty, t^X + t^Y < \tau_0) \leq \epsilon_X \exp(-\rho t^X) + \epsilon_Y \exp(-\rho t^Y) \leq \epsilon \exp(-\rho [t^X + t^Y]),
\]

where we exploited the definitions of \( \epsilon_X, \epsilon_Y \) and that \( t^Y \leq \ln(2)/\rho \) (i.e. \( 1/2 \leq \exp(-\rho t^Y) \)) in the last inequality.

For the opposite upper-bound, we recall first that \( \zeta \) is chosen to be uniform over the compact space \( \Delta \), that is included in some \( D_x \). Exploiting Theorem 4.5 on this set \( D_x \), we deduce that there exists \( t, c > 0 \) such that:

\[
P_{\zeta} [(X,Y) (t) \in (dx, dy) ; t < \tau_0] \geq c 1_{B(0, x^\Delta + c^Y)} (x) 1_{[y^\Delta_x, y^\Delta_y]} (y) \; dx \; dy.
\]

Combining (7.6)-(7.9) ends the proof of Theorem 4.4 in the case \( d \geq 2 \).

7.3.2 Proof of Proposition 7.3.1

For readability, note that most of the subscripts "\( X \)" (except for \( t^X \)) from Proposition 7.3.1 are removed in this proof.

First, remark that without any jump, \( ||X|| \) tends to infinity, which makes the population almost doomed to extinction. We can thus find some time-limit \( t^\nu \) such that, even with an amplification of order \( \exp(\rho t^\nu) \), the event that the population survived without any mutation occurring in the time-interval \( [0, t^\nu] \) is exceptional enough. With this time-scale, we can find an upper-bound \( y^\nu \) on \( Y \): that the population reaches such size before \( t^\nu \) is an exceptional enough event. For the lower-bound, we exploit the fact that extinction is very strong when the population size is too small. Thus, that the population has survived—at least for a bit—after declining below this lower-bound \( y^\nu \) is also an exceptional enough event.

The last part is to ensure that this first jump is indeed diffuse in \( X \) (which is why we need \( \nu(dw) \) to have a density w.r.t. Lebesgue with the bound of \( [H5] \)).

For \( y^\nu > \ell_E > 1/\ell_E > y^\nu > 0, t^\nu, w^\nu > 0 \) and initial condition \((x, y) \in E \), let:

\[
T^\nu := \inf \{ t \geq 0 ; \; \Delta X_t \neq 0 \},
\]

\[
T^\nu := \inf \{ t \geq 0 ; \; Y_t = y^\nu \}, \quad T^\nu := \inf \{ t \geq 0 ; \; Y_t = y^\nu \} < \tau_0.
\]

On the event \( \{ T^\nu < t^\nu \land T^\nu \land T^\nu \} \cap \{ ||\Delta X_T|| < w^\nu \} \), we define \( U := T^\nu \). Else \( U := \infty \).

To choose \( y^\nu, y^\nu, t^\nu \) and \( w^\nu \), we refer to the following lemmas, which are treated as the four first steps of the proof:

Fact 7.3.4. For any \( \rho, \epsilon_1 > 0 \), there exists \( t^\nu > 0 \) such that:

\[
\forall (x, y) \in E, \quad \mathbb{P}_{(x,y)} (T^\nu < \tau_0) \leq \epsilon_1 \exp(-\rho t^\nu).
\]
Fact 7.3.5. For any \( t_\vee, \epsilon_2 > 0 \), there exists \( y_\vee > 0 \) such that:

\[
\forall (x, y) \in E, \quad P_{(x,y)}(T_{\vee}^y < t_\vee \land \tau_0) \leq \epsilon_2.
\]

Fact 7.3.6. For any \( t_S, \epsilon_3 > 0 \), there exists \( y_\wedge > 0 \) such that:

\[
\forall x \in \mathbb{R}^d, \quad P_{(x,y_\wedge)}(t_S < \tau_0) \leq \epsilon_3.
\]

Fact 7.3.7. For any \( t_\vee, \epsilon_4 > 0 \), there exists \( w_\vee > 0 \) such that:

\[
\forall (x, y) \in E, \quad P_{(x,y)}(\|X_{T_J}\| \geq w_\vee, T_J < t_\vee \land \tau_0) \leq \epsilon_4.
\]

Fact 7.3.8. For any \( t_\vee > 0 \), and any \( y_\vee > \ell_E > 1/\ell_E > y_\wedge > 0 \), there exists \( c, x_\vee > 0 \) such that:

\[
\forall (x, y) \in E, \quad P_{(x,y)}(X(U) \in dx \mid U < \tau_0) \leq c 1_{B(0, x_\vee)}(x) dx.
\]

**Step 1: proof of Fact 7.3.4**

Exploiting assumption \([H3]\), as long as \( \|X\| \) is sufficiently large, we can ensure that the growth rate of \( Y \) is largely negative, leading to a quick extinction. The proof is similar to the one of Lemma 3.2.2 in \[Ve21a\], where more details can be found. We consider the autonomous process \( Y^D \) as an upper-bound of \( Y \) where the growth rate is replaced by \( r_D \). For any \( t_D \) and \( \rho \), there exists \( r_D \) (a priori negative) such that whatever \( Y_D \) the initial condition of \( Y^D \), survival of \( Y^D \) until \( t_D \) (i.e. \( t_D < \tau_0^D \)) happens with a probability smaller than \( \exp(-2 \rho t_D) \).

Thanks to Assumption \([H3]\), we define \( x_\vee \) such that for any \( x, \|x\| \geq x_\vee \) implies \( r(x) \leq r_D \). We then deduce:

\[
\forall (x, y), \quad P_{(x,y)}(\forall t \leq t_D, \|X_t\| \geq x_\vee \mid t_D < \tau_0) \leq \sup_{y_D > 0} P_{y_D}(t_D < \tau_0^D) \leq \exp(-2 \rho t_D).
\]

Let \( t_E := (x_\vee + \ell_E)/v \) and assume \( t_\vee \geq t_E \). A.s. on \( \{t_\vee < T_J \land \tau_0\} \) for any \( (x, y) \in E \):

\[
\forall t_E \leq t \leq t_\vee, \quad \|X(t)\| = \|x - vt e_1\| \geq x_\vee.
\]

Exploiting inductively the Markov property at times \( t_\vee := t_E + k t_D \) for \( k \geq 1 \), we obtain:

\[
\forall (x, y), \quad \exp(\rho t_\vee) P_{(x,y)}(t_\vee < T_J \land \tau_0) \leq \exp(\rho(t_E - k t_D)) \xrightarrow{k \to \infty} 0.
\]

**Step 2: proof of Fact 7.3.5**

This is an immediate consequence of the fact that \( Y \) is upper-bounded by the process \( Y^r \) given in \[3.1\] with initial condition \( \ell_M \). This bound is uniform in the dynamics of \( X_t \) and \( M \) and uniform for any \( (x, y) \in E \). It is classical that a.s. \( \sup_{t \leq t_\vee} Y^r_t < \infty \), which proves the Lemma, see e.g. Lemma 3.3 in \[BM15\].

**Step 3: proof of Fact 7.3.6**

Like in the proof of Proposition 4.2.3 in \[Ve21a\], cf Appendix D, we exploit \( \rho_\gamma \) as the upper-bound of the growth rate of the individuals to relate to the formulas for Continuous State Branching Processes. Referring for instance to \[Pa10\] Subsection 4.2, notably Lemma 5, it is classical that 0 is an absorbing boundary for these processes (we even have explicit formulas for the probability of extinction). This directly entails the result of the present lemma that the probability of extinction tends uniformly to zero as the initial population size tends to zero.

**Step 4: proof of Fact 7.3.7**

On the event \( \{T_J < t_\vee \land \tau_0\} \), for any initial condition \( (x, y) \in E \), there exists a compact \( K \) of \( \mathbb{R}^d \) that contains \( X_t = x - vt \) for any \( t \in [0, T_J] \). Thanks to Assumption \([H2]\), there exists an upper-bound \( g_\vee \) of \( g \) valid on \( K \times \mathbb{R}^d \).

Let \( \epsilon_4 > 0 \) and \( \rho_W := (-1/t_\vee).\log(1 - \epsilon_4) \). We define \( w_\vee \) such that \( \nu(B(0, w_\vee)^c) \leq \rho_W/g_\vee \). Then we can couple the process \( X \) to an exponential r.v. \( T_W \) of mean \( 1/\rho_W \) such that on the
event \( \{T_J < t_\vee \land \tau_0\} \cap \{\|\Delta X_{T_J}\| \geq w_\vee\} \), \( T_J \leq T_W \) holds a.s. We can conclude with the following upper-bound:

\[
\forall (x, y) \in E, \quad P_{(x, y)}(\|\Delta X_{T_J}\| \geq w_\vee, \ T_J < t_\vee \land \tau_0) \leq P(T_W < t_\vee) = 1 - \exp(-\rho_W t_\vee) \leq \epsilon_4.
\]

Note that under Assumption \((A)\), the jump at time \( T_J \) cannot make the process escape \( K \). This provides a deterministic upper-bound \( w_\vee \) such that \( \|\Delta X_{T_J}\| \geq w_\vee \) a.s. on \( \{T_J < t_\vee \land \tau_0\} \).

**Step 5: proof of Fact 7.3.8** For \( x_\vee := \ell_E + v t_\vee \), let:

\[
c := \sup \left\{ \frac{g(x, w) \nu(w)}{\int_{\mathbb{R}^d} g(x, w') \nu(w') \, dw'} : \|x\| \leq x_\vee, w \in \mathbb{R}^d \right\} < \infty. \tag{7.12}
\]

We exploit a sigma-field \( F_{T_J}^* \) that includes the whole knowledge of the process until time \( T_J \), except for the size of the jump at this time. It is rigorously defined and studied in Appendix A. Conditionally on \( F_{T_J}^* \) on the event \( \{U < \tau_0\} \in F_{T_J}^* \), the law of \( X(T_J) \) is given by:

\[
g(X[T_J-], x - X[T_J-]) \nu(x - X[T_J-]) \int_{\mathbb{R}^d} g(X[T_J-], w') \nu(w') \, dw'.
\]

Note also that a.s. \( \|X[T_J-]\| \leq \ell_E + v t_\vee = x_\vee \) (since no jump has occurred yet).

Since \( \|\Delta X_{T_J}\| \leq w_\vee \) on the event \( \{U < \tau_0\} \), with \( \bar{x}_\vee := x_\vee + w_\vee \), we get the following upper-bound of the law of \( X(T_J) \):

\[
P_{(x, y)}(X(U) \in dx ; \ U < \tau_0) = P_{(x, y)} \left( \mathbb{E} \left[ X(U) \in dx \mid F_{T_J}^* \right] ; \ U < \tau_0 \right) \leq c 1_{B(0, \bar{x}_\vee)}(x) \, dx.
\]

**Step 6: concluding the proof of Proposition 7.3.1**

Let \( \ell_E, \rho, \epsilon > 0 \). We first deduce \( t_\vee \) thanks to Fact 7.3.4 such that:

\[
\forall (x, y) \in E, \quad P_{(x, y)}(t_\vee < T_J \land \tau_0) \leq \epsilon \exp(-\rho t_\vee)/8. \tag{7.13}
\]

Thanks to Fact 7.3.5 we deduce some \( y_\vee > 0 \) such that:

\[
\forall (x, y) \in E, \quad P_{(x, y)}(T^\vee_Y < t_\vee \land \tau_0) \leq \epsilon \exp(-\rho t_\vee)/8. \tag{7.14}
\]

We could take any value for \( t_S \) (so possibly 1), yet \( t_S = \log(2)/\rho \) seems somewhat more practical. We then deduce \( y_\wedge \) thanks to Fact 7.3.6 such that:

\[
\sup_{x \in \mathbb{R}^d} P_{(x, y_\wedge)}(t_S < \tau_0) \leq \epsilon \exp(-\rho t_\vee)/8.
\]

This implies that for any \( (x, y) \in E \):

\[
P_{(x, y)}(t_\vee + t_S < \tau_0, T^\wedge_Y < t_\vee \land \tau_0 \land T^\vee_Y \land T_J)
\leq E_{(x, y)} \left( \mathbb{P}_{(x, y_\wedge)}(t_S < \tau_0), T^\wedge_Y < t_\vee \land \tau_0 \land T^\vee_Y \land T_J \right)
\leq \epsilon \exp(-\rho t_\vee)/8. \tag{7.15}
\]

\( w_\vee \) is chosen thanks to Fact 7.3.7 such that:

\[
\forall (x, y) \in E, \quad P_{(x, y)}(\|\Delta X_{T_J}\| \geq w_\vee, T_J < t_\vee \land \tau_0) \leq c 1_{B(0, w_\vee)}(x) \, dx.
\]

Thanks to Fact 7.3.8 there exist \( c, x_\vee > 0 \) such that:

\[
\forall (x, y) \in E, \quad P_{(x, y)}(X(U) \in dx ; \ U < \tau_0) \leq c 1_{B(0, x_\vee)}(x) \, dx.
\]
Thanks to the construction of $U$, and noting that $t^X := t_\nu + t_S$, it is clear that $U \geq \tau_\theta \wedge t^X$ is equivalent to $U = \infty$. Combining (7.13), (7.14), (7.15) and (7.16):

$$\mathbb{P}_{(x,y)}(U = \infty, t_\nu + t_S < \tau_\theta) \leq \mathbb{P}_{(x,y)}(t_\nu < T_J \wedge \tau_\theta) + \mathbb{P}_{(x,y)}(T^\nu \leq t_\nu \wedge \tau_\theta)$$

$$+ \mathbb{P}_{(x,y)}(\|\Delta X_{T_J}\| \geq w_\nu, T_J < t_\nu \wedge \tau_\theta) + \mathbb{P}_{(x,y)}(t_\nu + t_S < \tau_\theta, T^\nu < t_\nu \wedge \tau_\theta \wedge T^\nu \wedge T_J)$$

$$\leq \epsilon \exp(-\rho t_\nu)/2 = \epsilon \exp(-\rho t^X).$$

This ends the proof of Proposition 7.3.1.

The proof of Theorem 4.4 in the case $d \geq 2$ is now completed. All the theorems have been proved at this point. There are two appendix, the first one being devoted to the filtration $F^*_{T_J}$ up to the jumping time. We finish in Appendix B with first results of simulations that shall help illustrate the discussion given in Subsection 2.3.

### Appendix A: A specific filtration for jumps

This appendix extends to our case the result already presented in [Ve21b]: there exists a sigma-field $F^*_{T_J}$ which informally “includes the information carried by $M$ and $B$” until the jump time $T_J$ except the realization of the jump itself.

Denote $W$ as the additive effect on $X$ of the first jump of $X$, occurring at time $T_J$. We then define:

$$F^*_{T_J} := \sigma(A_s \cap \{s < T_J\}; s > 0, A_s \in \mathcal{F}_s).$$

**Properties of $F^*_{T_J}$**: If $Z_s$ is $\mathcal{F}_s$-measurable and $s < t \in (0, \infty]$, $Z_s \mathbf{1}_{\{s < T_J \leq t\}}$ is $F^*_{T_J}$-measurable.

**Lemma (A1)**. For any left-continuous and adapted process $Z$, $Z_{T_J}$ is $F^*_{T_J}$-measurable. Reciprocally, $F^*_{T_J}$ is in fact the smallest sigma-algebra generated by these random variables.

In particular, for any stopping time $T$, $\{T_J \leq T\} \in F^*_{T_J}$.

**Lemma (A2)**. For any $h : \mathbb{R} \to \mathbb{R}_+$ measurable, $(x, y) \in (-L, L) \times \mathbb{R}_+$:

$$\mathbb{E}_{(x,y)}\left[h(W) \mid F^*_{T_J}\right] = \frac{\int_{\mathbb{R}} h(w) f(Y_{T_J}) g(X_{T_J-}, w) \nu(dw)}{\int_{\mathbb{R}} f(Y_{T_J}) g(X_{T_J-}, w') \nu(dw')}.$$

**Proof of Lemma (A1)**:

For any left-continuous and adapted process $Z$,

$$Z_{T_J} = \lim_{n \to \infty} \sum_{k=1}^{n} Z_{\frac{k-1}{n}} \mathbf{1}_{\{\frac{k-1}{n} < T_J \leq \frac{k}{n}\}},$$

where by previous property and the fact that $Z$ is adapted: $Z_{\frac{k-1}{n}} \mathbf{1}_{\{\frac{k-1}{n} < T_J \leq \frac{k}{n}\}}$ is $F^*_{T_J}$-measurable for any $k, n$. Reciprocally, for any $s > 0$ and $A_s \in \mathcal{F}_s$:

$$1_{A_s \cup \{s < T_J\}} = \lim_{n \to 1} Z^n_{T_J}, \quad \text{where} \ Z^n_{T} := \{1 \wedge [n(t-s)_+], 1_{A_s}.}$$

Now, for any stopping time $T$, and any $t \geq 0$, $\{t \leq T\} \in \mathcal{F}_t$ and $\{t \leq T\} = \bigcap_{s \leq t} \{s \leq T\}$, thus $\{T_J \leq T\} \cap \{T_J < \infty\} \in F^*_{T_J}$. Similarly:

$$\{T_J = T = \infty\} = \bigcap_{s > 0} \{s < T\} \cap \{s < T_J \leq \infty\} \in F^*_{T_J}.$$  

**Proof of Lemma (A2)**:

Let:

$$Z_t := \frac{\int_{\mathbb{R}} h(w') f(Y_t) g(X_{t-}, w') \nu(dw')}{\int_{\mathbb{R}} f(Y_t) g(X_{t-}, w') \nu(dw')}.$$
which is a left-continuous and adapted process. Thanks to Lemma (A1), $Z_{T_j}$ is $\mathcal{F}_{T_j}^*$-measurable.

We note the two following identities:

$$h(W) = \int_{[0,t] \times \mathbb{R}^d \times \mathbb{R}_+^2} h(w) \mathbf{1}_{\{t=T_j\}} M(dt, dw, du_f, du_g)$$

$$\int_{\mathbb{R}} h(w) f(Y_{T_j}) g(X_{T_j -}, w) \nu(dw)$$

$$= \int_{[0,t] \times \mathbb{R}^d \times \mathbb{R}_+^2} \frac{h(w') f(Y_t) g(X_{T_j -}, w') \nu(dw')}{\int_{\mathbb{R}} f(Y_t) g(X_{T_j -}, w') \nu(dw')} 1_{\{t=T_j\}} M(dt, dw, du_f, du_g),$$

Then, we exploit Palm’s formula to prove that their product with any $Z_s \mathbf{1}_{\{s<T_j\leq r\}}$ has the same average for any $s < r$ and $Z_s \mathcal{F}_s$-measurable:

$$\mathbb{E}_{(x,y)} [h(W) \cdot Z_s \mathbf{1}_{\{s<T_j\}} ; s < T_j \leq r]$$

$$= \mathbb{E}_{(x,y)} \left[ Z_s \int_{[0,t] \times \mathbb{R}^d \times \mathbb{R}_+^2} h(w) \mathbf{1}_{\{t=T_j\}} M(dt, dw, du_f, du_g) ; s < T_j \leq r \right]$$

$$= \mathbb{E}_{(x,y)} \left[ \int_{[0,t] \times \mathbb{R}^d \times \mathbb{R}_+^2} Z_s h(w) \mathbf{1}_{\{s,T_j\}}(t) \mathbf{1}_{\{t=T_j\}} M(dt, dw, du_f, du_g) \right]$$

$$= \mathbb{E}_{(x,y)} \left[ \int_{[0,t] \times \mathbb{R}^d \times \mathbb{R}_+^2} \mathbf{1}_{\{s,T_j\}}(t) Z_s h(w) \mathbf{1}_{\{t=T_j\}} M(dt, \nu(dw) du_f du_g) \right],$$

where, according to Palm’s formula, $\hat{T}_j$ is the first jump of the process $(\hat{X}, \hat{Y})$ encoded by $M+\delta_{(t,w,u)}$ and $B$ (cf e.g. [DV08] Proposition 13.1.VII). Since $(\hat{X}, \hat{Y})$ coincide with $(X, Y)$ at least up to time $t > s$, $Z_s$ was not affected by this change. Moreover:

$$\left\{ t = \hat{T}_j \right\} = \left\{ t \leq T_j \right\} \cap \left\{ u \leq f(Y_t) g(X_{t -}, w) \right\}.$$

Thus:

$$\mathbb{E}_{(x,y)} [h(W) \cdot Z_s \mathbf{1}_{\{s<T_j\}} ; s < T_j \leq r]$$

$$= \mathbb{E}_{(x,y)} \left[ \int_{[0,t] \times \mathbb{R}^d \times \mathbb{R}_+^2} \mathbf{1}_{\{s,T_j\}}(t) Z_s h(w) \mathbf{1}_{\{u_f \leq f(Y_t)\}} \mathbf{1}_{\{u_g \leq g(X_{t -}, w)\}} \mathbf{1}_{\{t \leq T_j\}} M dt \nu(dw) du_f du_g \right],$$

$$= \mathbb{E}_{(x,y)} \left[ Z_s \int_t^r \mathbf{1}_{\{t \leq T_j\}} h(w) f(Y_t) g(X_{t -}, w) \nu(dw) dt \right].$$
On the other hand, and with the same spirit:

\[
\mathbb{E}_{(x,y)} \left[ \int_{[0,T_J]} \frac{h(w') f(Y_{T_J}) g(X_{T_J}, w') \nu(dw')}{{\mathbb{R}} f(Y_t)g(X_t, w') \nu(dw')} \right] = Z_s \; ; \; s < T_J \leq r
\]

\[
= \mathbb{E}_{(x,y)} \left[ Z_s \int_{[0,T_J]} \frac{h(w') f(Y_{T_J}) g(X_{T_J}, w') \nu(dw')}{{\mathbb{R}} f(Y_t)g(X_t, w') \nu(dw')} \right] \times 1_{(t=T_J)} M(dt, dw, du_f, du_g) \; ; \; s < T_J \leq r
\]

\[
= \mathbb{E}_{(x,y)} \left[ \int_{[0,T_J]} \frac{h(w') f(Y_{T_J}) g(X_{T_J}, w') \nu(dw')}{{\mathbb{R}} f(Y_t)g(X_t, w') \nu(dw')} \right] \times 1_{(t=T_J)} M(dt, dw, du_f, du_g)
\]

\[
= \mathbb{E}_{(x,y)} \left[ Z_s \int_{[0,T_J]} h(w') f(Y_{T_J}) g(X_{T_J}, w') \nu(dw') dt \right]
\]

which is indeed the same integral as for \( h(W) \).

Appendix B: Brief overview of characteristic profiles of the quasi-stationary regime obtained by simulations

We provide in this Appendix B some results of a particular choice of three parameters regime whose comparison shall shed light on the discussion given in Subsection [2.3]. We present the profiles of the characteristic distributions and functions of the quasi-stationary regime, namely the QSD, the quasi-ergodic distribution (QED) and the survival capacity (the limiting properties are recalled just beside the figures).

The details of the exploited parameters are as follows. For population size dynamics, the growth rate as a function of \( x \) is here chosen to be of the form \( r(x) = 4 - 30 \times |x| \). A parabolic profile would give very similar results. The competition rate is \( c = 0.1 \), which leads to population sizes at quasi-equilibrium (carrying capacity) close to 40 (in arbitrary units). 2 and 6 are respectively the parameter modified here between the 3 simulation sets: it takes the values \( m \).

The profile of additive effects of mutations is given by \( \nu(dw) = \frac{1}{2w_0} \exp(-|w|/w_0) \). It is therefore symmetrical exponential, with \( w_0 = 0.03 \), so with many small mutations. The effect of population size on the fixation rate is simply proportional \( f_N(n) = m \times n \). The mutation rate \( m \) is the only parameter modified here between the 3 simulation sets: it takes the values \( m = 0.85, m = 0.55 \) and \( m = 0.25 \). The choice of these values is done so that the adaptation is critical at \( m = 0.85 \): for larger \( m \) like \( m = 0.85 \), extinction is kept almost negligible, so that we say that adaptation is spontaneous; whereas for smaller values of \( m \), extinction plays a consequent role and produces differences in shape between the QSD and the QED.

We exploited the following expression for the probability of invasion:

\[
g(x, w) := \frac{N_H(x) \times \Delta r/\sigma}{1 - \exp[-N_H(x) \times \Delta r/\sigma]},
\]

where \( \Delta r := r(x+w) - r(x) \) is the variation of the growth rate between the mutant and the resident, and \( N_H(x) \) is the harmonic mean of a resident population with fixed trait \( x \) (averaged against its
associated QSD). Deleterious mutations are allowed, but their probability of fixation is very reduced if they are strong in relation to population fluctuations. The values of $N_H$ are estimated numerically, with the following profile:

Figure 2: $N_H(x)$ on the left side: the harmonic means of the population size fluctuations of the process $(\tilde{N}_t^x)_{t \geq 0}$ with fixed traits $x$ (given by the associated QSD); the extinction rate of these QSD is on the right.

The formula relies on the Kimura diffusion approximation that has been derived in the case of fixed population size. Assuming rapid size fluctuations, we choose the harmonic mean as the reference by referring to classical approximations obtained in the case of periodically fluctuating population sizes (cf notably [OW97]). More details are given (in French) in my PhD manuscript, and a subsequent paper is planned to discuss these results and the relevance of this estimation. The comparison of such a two-component stochastic model to the individual-based model through the QSD and QED shall be a good test for the relevance of such formula. The kind of dependence in the difference in growth rate seems to play a crucial role for having a QED as much conserved.

These simulations were obtained by calculating the evolution of the densities themselves. This method is related to those of finite volumes, with an explicit numerical scheme and a renormalization of density estimates at each time step. The transitions to $X$ and $N$ are performed successively to reduce the calculation time.
Profiles of the "QSD"

\[ \mathbb{P}(x, n) \left[ (X, N)_t \in (dx, dn) \right| t < \tau_\theta \] 
\[ \xrightarrow{t \to \infty} \alpha(dx, dn) \]

Case of a spontaneous adaptation

Critical regime of adaptation

Profiles of the QED: the invariant measure of the Q-process

\[ \lim_{t \to \infty} \lim_{T \to \infty} \mathbb{P}(x, n) \left[ (X, N)_t \in (dx, dn) \right| T < \tau_\theta \] 
\[ = \beta(dx, dn) \]

Case of a spontaneous adaptation

Critical regime of adaptation

Selection through the extinction of populations

Given the very different profiles obtained for the QSD, it is quite remarkable that the quasi-ergodic measures are similar as much. In particular, we can see that the histories of the surviving populations are still shaped by the maintenance of these populations at large sizes with almost optimal traits, even when such traits are very rare according to the QSD.
Profiles of the survival capacity

\[ h(x, n) := \lim_{t \to \infty} \frac{P(x, n)(t < \tau \alpha)}{P(\alpha)(t < \tau \alpha)} \]

Case of a spontaneous adaptation

Critical regime of adaptation

Selection through the extinction of populations

Profiles 3D of the survival capacity

Spontaneous adaptation  
Critical adaptation  
Adaptation through extinction

References

[BBCM19] Bansaye, V., Cloez, B., Gabriel, P., Marguet, A.; A non-conservative Harris’ ergodic theorem, ArXiv : 1903.03946 (2019, v.1)

[BM15] Bansaye, V., Méléard, S.; Stochastic Models for Structured Populations, Scaling Limits and Long Time Behavior, Stochastics in Biological Systems, V.1, N.4, Springer Switzerland (2015)

[Ba95] Bass, R.; Probabilistic Techniques in Analysis, Probab. and its Appl., Springer (1995)

[BL95] Bürger, R., Lynch, M.; Evolution and Extinction in a Changing Environment: A Quantitative-Genetic Analysis; Evolution, V.49, N.1, pp.151-163 (1995)

[CC+09] Cattiaux, P., et al; Quasi-Stationary Distributions and Diffusion Models in Population Dynamics, The Annals of Probab., V.37, N.5, pp.1926–1969 (2009)

[CCV18] Champagnat, N., Coulibaly-Pasquier, K., Villemonais, D.; Convergence to quasi-stationary distribution for multi-dimensional diffusion processes, Séminaire de Probabilités XLIX, V.2215, pp.165-182, Springer (2018)
[CV16] Champagnat, N., Villemonais, D.; Exponential convergence to quasi-stationary distribution and Q-process, Probab. Theory Relat. Fields, V. 164, pp. 243-283 (2016)

[CV18a] Champagnat, N., Villemonais, D.; Uniform convergence of time-inhomogeneous penalized Markov processes, ESAIM: Probab. and Stat., V.22, pp.129-162 (2018)

[CV18b] Champagnat, N., Villemonais, D.; General criteria for the study of quasi-stationarity, preprint on ArXiv : 1712.08092 (2018, v.2)

[CV20] Champagnat, N., Villemonais, D.; Practical criteria for R-positive recurrence of unbounded semigroups, Electron. Commun. Probab., V.25, N.6, pp.1-11 (2020)

[CV21] Champagnat, N., Villemonais, D.; Lyapunov criteria for uniform convergence of conditional distributions of absorbed Markov processes, Stoch. Proc. and Appl., V.135, pp.51-74 (2021)

[CCM17] Chazottes, R., Collet, P., Méléard, S.; On time scales and quasi-stationary distributions for multitype birth-and-death processes, Ann. Inst. H. Poincaré Probab. Statist., V.55, N.4, pp.2249-2294 (2019)

[CG20] Cloez, B., Gabriel, P.; On an irreducibility type condition for the ergodicity of nonconservative semigroups, Comptes Rendus. Mathématique, V.358, N.6, pp.733-742 (2020)

[CMS13] Collet, P., Martínez, S., San Martín, J.; Quasi-Stationary Distributions, Probab. and Its Appl., Springer, Berlin Heidelberg (2013)

[DV08] Daley, D.J., Vere-Jones, D.; An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure; Probab. and Its Appl., Springer, 2nd ed. (2008)

[DiT13] Di Tella, P., On the Predictable Representation Property of Martingales Associated with Levy Processes; PhD thesis (2013)

[DP13] van Doorn, E.A., Pollett, P.K.; Quasi-stationary distributions for discrete-state models, Eur. J. of Operational Research, V. 230, pp. 1-14 (2013)

[Ev98] C. Evans, L.; Partial Differential Equations; Graduate Studies in Mathematics, V. 19, Am. Math. Society (1998)

[FRS20] Ferré, G., Rousset, M., Stoltz, G.; More on the long time stability of Feynman-Kac semi-groups; Stoch PDE: Anal Comp (2020)

[GNW20] Guillin, A., Nectoux, B., Wu, L.; Quasi-stationary distribution for strongly Feller Markov processes by Lyapunov functions and applications to hypoelliptic Hamiltonian systems, preprint with HAL Id: hal-03068461

[KH09] Kopp, M., Hermisson, J.; The genetic basis of phenotypic adaptation II: The distribution of adaptive substitutions of the moving optimum model; Genetics, V.183, pp. 1453-1476 (2009)

[KNP18] Kopp, M., Nassar, E., Pardoux, E.; Phenotypic lag and population extinction in the moving-optimum model: insights from a small-jumps limit, J. Math. Biol., V.7, N.5, pp.1431-1458. (2018)

[La05] Lambert, A.; The branching process with logistic growth; Ann. Appl. Probab., V.15, N.2, pp.1506-1535 (2005)

[MV12] Méléard, S., Villemonais, D.; Quasi-stationary distributions and population processes, Probab. Surveys, V. 9, pp. 340-410 (2012)
[NP17] Nassar, E., Pardoux, E.; On the long time behavior of the solution of an SDE driven by a Poisson Point Process; Adv. in Appl. Probab., V.49, pp.344-367 (2017)

[NP19] Nassar E., Pardoux, E.; Small jumps asymptotic of the moving optimum Poisonian SDE, Stoch. Proc. and Appl., V.129, N.7, pp.2320-2340. (2019)

[OW97] Otto, S.P., Whitlock, M.C.; The Probability of Fixation in Populations of Changing Size Genetics, V.146, N.2, pp.723-733 (1997)

[Pa16] Pardoux, E; Probabilistic Models of Population Evolution : Scaling Limits, Genealogies and Interactions; Springer (2016)

[Po15] Pollett, P. K.; Quasi-stationary distributions: A bibliography., available at people.smp.uq.edu.au/PhilipPollett/papers/qsds/qsds.html (2015)

[Ve21a] Velleret, A.; Unique quasi-stationary distribution, with a possibly stabilizing extinction; ArXiv :1802.02409 (2021, v.6)

[Ve21b] Velleret, A.; Exponential quasi-ergodicity for processes with discontinuous trajectories, preprint on ArXiv : 1902.01441 (2021, v.4)

[YW71] Yamada, T., Watanabe, S.; On the uniqueness of solutions of stochastic differential equations. J. Math. Kyoto Univ. V.11, pp.155–167 (1971)