Convergence and Existence for Polymorphic Adaptive Dynamics Jump and Degenerate Diffusion Models

Nicolas Champagnat*

March 17, 2003

Abstract

The canonical equation of adaptive dynamics has recently been proposed as a model for Darwinian evolution by mutation-selection in a monomorphic population (where all individual, except mutants, are identical). Here we derive rigorously a canonical equation in the case of n-morphic evolution (when n types of individuals co-exist). Starting with a jump process describing the evolution of the population (the so-called trait substitution sequence), and taking the proper limit, we obtain a precise mathematical justification of this canonical equation, both for symmetrical and asymmetrical mutations. We then propose a diffusion model for adaptive dynamics in the case of symmetrical mutations, allowing for evolution in any direction of the trait space. We prove a weak existence result for this process which has discontinuous and degenerate parameters at evolutionary singularities.

Keywords: Convergence of jump processes, Weak existence of degenerate diffusions, Martingale problem, Polymorphic asymmetrical adaptive dynamics models, Canonical equation of adaptive dynamics.

1 Introduction

Adaptive dynamics (Hofbauer and Sigmund [8], Marrow *et al.* [12] and Metz *et al.* [13]) is a recent branch of evolutionary ecology that proposes new models of Darwinian evolution involving three mechanisms: reproduction, that transmits the trait through generations, mutations, that generate variability in the trait values, and selection between traits, that results from ecological interactions between individuals and their environment.

More precisely, in a population subject to *clonal reproduction* (*e.g.* reproduction in asexual), each individual is characterized by a quantitative *adaptive trait*: a real vector representing individual features affecting reproduction and survival, and subject to mutation, such as individual size, age at maturity, size of prefered preys, pathogen virulence... Each clone has the same trait as its progenitor, unless affected by a *mutation*. When a mutant appears, the background population is called *resident*, and the mutant initiates its own population with only one individual. *Invasion* occurs if the mutant population significantly increases in frequency with respect to the resident one. Finally, a population in which all individuals bear the same trait value is called *monomorphic*. A non-monomorphic population is called *polymorphic*. Among these *n-morphic* populations, in which the individual trait assumes exactly n different values. At any time, the *population state* is described by the set of trait

*Université Paris 10, MODAL'X, Bât. G, 200 av. de la République, 92001 Nanterre Cedex, France, e-mail: champagn@clipper.ens.fr

values represented in the population. Adaptive dynamics modelling seeks to describe how this population state changes through time.

An important advance of adaptive dynamics theory is the so-called "canonical equation" (Dieckmann and Law [4]), derived from a monomorphic trait substitution sequence model (Metz *et al.* [14]) under the assumption of small mutations (*i.e.* the trait of a mutant is only slightly different from the trait of its progenitor). This assumption allows useful simplifications and leads to a classification of "evolutionary singularities" (Metz *et al.* [14]) — particular trait values where evolution may come to a halt, or start diversifying the population. Problem with the canonical equation approach to adaptive dynamics is that it is still lacking a rigorous mathematical basis, and has been restricted so far to evolution in monomorphic populations, even on heuristic grounds.

In this paper, after giving a precise description of the biological process (section 2), we adapt the trait substitution sequence model of [14] to n-morphic populations. This model is in fact a Markov jump process in the trait space, and we give a precise pathwise representation for it. In section 3, we prove that this jump process converges to a n-morphic canonical equation when mutation jumps converge to 0. This result gives a comprehensive mathematical justification for the canonical equations of [4] and its generalizations in Champagnat *et al.* [2]. Finally, in section 4, we propose a new model of adaptive dynamics: a diffusion process having as infinitesimal generator an order-one approximation of the generator of the jump process of section 2, when mutation jumps converge to 0. We obtain a second-order operator with degenerate and non-continuous coefficients, to which the classical theory of diffusion processes does not apply. In the case of symmetrical mutations (*i.e.* the law of the mutant trait is symmetrical in the trait space, which is a common biological assumption [2]), we prove the weak existence of solutions to the corresponding SDE in Theorem 2.

Let us discuss the biological scope of this study. The n-morphic trait substitution sequence of section 2 is a slight generalization of models of [14], very similar to the monomorphic trait substitution sequence of [4] for n species. However, this new model and the n-morphic canonical equation that we obtain in section 3 extend dynamical models of evolution beyond "branching points" — special evolutionary singularities where a monomorphic population evolves into polymorphism. Until now, one was able to describe the evolution of a monomorphic population (thanks to the canonical equation of [4]), and to determine the trait values at which this population could become polymorphic (thanks to the branching conditions at evolutionary singularities of [14]). Our work shows that, between two branching events, the population evolves according to the n-morphic canonical equation, where n is the number of traits co-existing.

The diffusion model of section 4 answers another biological problem. The main drawback of the trait substitution models of [14] and [4] is that they are grounded on the biological assumption of deterministic evolution (see assumptions (BH2) and (BH3) of section 2) which is valid in infinite population. This leads to an unrealistic phenomenon: in these models, evolution may only be strictly "directional": it is possible only in particular directions of the trait space, depending on the sign of the fitness function (defined in section 2). On the contrary, in finite populations, any mutant trait (even deleterious, *i.e.* competitively inferior) may invade the resident population by chance, so that evolution is possible in any direction. The diffusion process of section 4, even if it is derived from the infinite population model of trait substitution sequence, accounts for such a phenomenon.

Let us also mention another important biological application of this diffusion process. Evolution, when it reaches a steady state, can proceed by two different strategies (see Rand and Wilson [16]): radiation (appearance, coexistence and divergence of different trait values in the population), or punctualism (rapid evolution from one evolutionary steady state to another). Our diffusion model of section 4 could help study the process of punctualism, by means of a large deviation principle. This topic is work in progress.

Notations $\mathcal{B}_b(E)$ *(resp.*$\mathcal{C}^k(E), \mathcal{C}_b^k(E), \mathcal{C}_c^k(E)$*) denotes the set of functions from* E *to* $\mathbb{R}$ *that are Borel and bounded (resp. of class* $\mathcal{C}^k$*, resp. bounded, of class* $\mathcal{C}^k$ *with bounded* i^{th} *order derivatives, for* $1 \leq i \leq k$*, resp. of class* $\mathcal{C}^k$ *with compact support).*

2 n-morphic trait substitution sequence

We consider an asexual population. We restrict ourselves to the case of a single species in order to keep notations simple (all results of this paper can be extended to the multispecific case with no difficulty). Each individual is characterized by a trait value x in a given trait space $\mathcal{X}$, convex open subset of $\mathbb{R}^d$ for some fixed $d \geq 1$. $\mathcal{X}$ may be bounded or not. This paper studies the evolution of n-morphic populations, for a fixed integer n. This means that there are exactly n trait values represented in the background resident population at any time a mutant arises.

2.1 Biological premises

Our model is based on the following four biological assumptions:

(BH1) The time scales of mutation and selection are separated, so that selection is given the time to eliminate unfit mutants or promote competitively superior ones before a new mutant appears.

(BH2) In a polymorphic population, in the absence of new mutations, the population size of each trait converges to a unique, positive equilibrium depending only on the set of traits initially represented.

(BH3) There is some quantity $f(x; x_1, \ldots, x_n)$, called "fitness" (see Metz *et al.* [13]), that measures the selective advantage (or disadvantage) of a mutant individual with trait x in the equilibrium population corresponding to the resident trait values $x_1, \ldots, x_n$: if $f(x; x_1, \ldots, x_n) > 0$, invasion is possible, and if $f(x; x_1, \ldots, x_n) \leq 0$, invasion is impossible.

(BH4) When a mutant trait invades a resident n-morphic population, the ecology of the system drives the mutant progenitor's population to extinction, and the new n-morphic community is ecologically stable.

(BH5) In a resident n-morphic population with trait values $x_1, \ldots, x_n$, for any $i \in \{1, \ldots, n\}$,

$$f(x_i; x_1, \ldots, x_n) = 0. \tag{1}$$

(BH1) means that mutations are rare. This is a reasonable assumption since we are only concerned with mutations of the genotype having effect on the phenotype, *i.e.* nonsynonymous mutations in coding parts of the genome, which are very rare. (BH2) makes the ecology of the system as simple as possible (for more general hypotheses, see [13] and [14]). Much theoretical work has been devoted to identify fitness functions under specific biological assumptions (see [14] or Diekmann [3]). $f(x; x_1, \ldots, x_n)$ is generally the growth rate of a mutant individual with trait x in the equilibrium population made of trait values $x_1, \ldots, x_n$. (BH4) is a generalization of the principle of mutual exclusion of [4]. Together with (BH1), (BH4) ensures that the population can be considered n-morphic at any evolutionary time (see [14] and Geritz *et al.* [7] for a more detailled discussion). (BH5) is a very frequently made assumption (see [13], [14] or [4]): actually, since the population $x_1, \ldots, x_n$ is in equilibrium, the growth rate of an individual with trait x_i in this population is necessarily 0.

In the equilibrium population (given by (BH2)) in which individuals have trait values $x_1, \dots, x_n$, the following notations are defined

$X = (x_1, \dots, x_n)$ is the state of the population (we will often use the term "population X" for the equilibrium population above)

$b(x; X)$ is the rate of birth for a mutant individual with trait x in a resident population X

$d(x; X)$ is the rate of death of a mutant individual with trait x in a resident population X

$n_i(X)$ is the size (number of individuals) of the sub-population of trait x_i

$\mu(x)$ is the probability that a mutation occurs in a birth from an individual with trait x

$p(x, dh)$ is the law of $h = x' - x$, where x' is the mutant trait borned from an individual with trait x. It is a probability measure on $\mathbb{R}^d$, and since x' must be in the trait space $\mathcal{X}$, the support of $p(x, \cdot)$ is a subset of

$$\mathcal{X} - x = \{y - x; y \in \mathcal{X}\} \tag{2}$$

Until section 4, no symmetry assumption will be made for $p(x, \cdot)$.

As suggested by Metz *et al.* [13] and Dieckmann and Law [4], the fitness $f(x; X)$ of a mutant trait x in the resident population X, is given by

$$f(x; X) = b(x; X) - d(x; X). \tag{3}$$

This is the growth rate of a mutant individual x in the population X.

2.2 Description of the process

Assuming (BH1), (BH2), (BH3) and (BH4), we can now give a precise description of our n-morphic trait substitution sequence model for the evolution of the population. It is inspired from [14] and [4], and biological justifications of the use of the parameters are the same as therein.

1. We start with a n-morphic population with traits $x_1, \dots, x_n$ at ecological equilibrium (see assumptions (BH1) and (BH2)).

2. Consider n independent exponential random variables with respective parameters $\mu(x_i)b(x_i; X)n_i(X)$ for $1 \leq i \leq n$. Suppose that the smallest value is the i^{th} one. After waiting this amount of time, choose around x_i a mutant trait $x'_i = x_i + h$, where h follows the law $p(x_i, dh)$. This step is the same as waiting an exponential time with rate $\sum_{i=1}^n \mu(x_i)b(x_i; X)n_i(X)$, and then choosing the trait of the mutant's progenitor in the following way: for $i \in \{1, \dots, n\}$, this trait is x_i with probability $\frac{\mu(x_i)b(x_i;X)n_i(X)}{\sum_{i=1}^n \mu(x_i)b(x_i;X)n_i(X)}$.

3. Because of (BH3), the mutant population invades the resident one with probability $\frac{[f(x'_i;X)]_+}{b(x'_i;X)}$, where $[a]_+ = a \vee 0$. If invasion does not occur, the process returns to the first step.

4. If invasion occurs, assumption (BH4) entails that the resident subpopulation with trait x_i disappears, and the process goes back to the first step with the new n-morphic composition $x_1, \dots, x_{i-1}, x'_i, x_{i+1}, \dots, x_n$.

In biological models, the trait space $\mathcal{X}$ is always taken open. However, for the theorems of sections 3 and 4, we will have to consider processes in a Polish space (complete metric separable). So let us extend the process above to the state space $\overline{\mathcal{X}^n}$ in a way suggested by ecological considerations: trait values at the boundary $\partial\mathcal{X}^n = \overline{\mathcal{X}^n} \setminus \mathcal{X}^n$ of the trait space are often considered as non-viable, or, at least, completely deleterious, so that evolution cannot reach them (if it does, it is equivalent to the extinction of the population). So we extend our process by assuming it stays constant (no evolution) if its initial state is in the boundary of $\mathcal{X}^n$.

Let us define, for $1 \leq i \leq n$,

$$g_i(x;X) = \mu(x_i)b(x_i;X)n_i(X)\frac{f(x;X)}{b(x;X)}. \tag{4}$$

The generator of the Markov process described above writes

$$L\varphi(X) = \begin{cases} \sum_{i=1}^{n} \int_{\mathbb{R}^d} (\varphi(X+(h)_i) - \varphi(X))[g_i(x_i+h;X)]_+ p(x_i,dh) & \text{if } X \in \mathcal{X}^n \\ 0 & \text{if } X \in \partial\mathcal{X}^n \end{cases} \tag{5}$$

for all $\varphi \in \mathcal{B}_b(\overline{\mathcal{X}^n})$, with the notation $(h)_i = (0,\dots,0,h,0,\dots,0)$ where h is at the i^{th} coordinate. The paths of the process X are in the Skorohod space $\mathbb{D}(\mathbb{R}_+,\overline{\mathcal{X}^n})$ of right continuous and left limited functions from $\mathbb{R}_+$ to $\overline{\mathcal{X}^n}$.

This formula gets simpler if we extend $p(x,dh)$ to $\overline{\mathcal{X}}$ and g_i to $\overline{\mathcal{X}^{n+1}}$ for $1 \leq i \leq n$ as follows: for $x \in \partial\mathcal{X}$, set $p(x,\cdot) = \delta_x$ (the Dirac mass at x), and for $(x,X) \in \partial\mathcal{X}^{n+1}$, set $g_i(x;X) = 0$. Then, for $\varphi \in \mathcal{B}_b(\overline{\mathcal{X}^n})$,

$$L\varphi(X) = \sum_{i=1}^{n} \int_{\mathbb{R}^d} (\varphi(X+(h)_i) - \varphi(X))[g_i(x_i+h;X)]_+ p(x_i,dh). \tag{6}$$

If, for $1 \leq i \leq n$, g_i is bounded, this generator defines a unique semigroup, so that the process is unique in law (see Ethier and Kurtz [5]). We will now prove its existence by giving an explicit pathwise representation that will be useful later on, under the following assumptions:

(Ha) $p(x,\cdot)$ has finite and bounded on $\mathcal{X}$ second order moments (*i.e.* $\int \|h\|^2 p(x,dh)$ is a bounded function of $x \in \mathcal{X}$), and is absolutely continuous with respect to some Radon measure ν on $\mathbb{R}^d$ for any $x \in \mathcal{X}$ (ν is often in applications the Lebesgue measure on $\mathbb{R}^d$). Then

$$\forall x \in \mathcal{X}, \quad p(x,dh) = p(x,h)\nu(dh). \tag{7}$$

Moreover, there is some function $p : \mathbb{R}^d \to \mathbb{R}$ such that

$$\forall x \in \mathcal{X}, \quad p(x,h) \leq p(h). \tag{8}$$

(Hb) For all $i \in \{1,\dots,n\}$, g_i is bounded by some constant κ on $\overline{\mathcal{X}^{n+1}}$.

In these conditions, following on Fournier and Méléard [6], it is possible to give an explicit pathwise construction of a process $(X(t), t \geq 0)$ generated by (6):

Definition 1 *Let us $(\Omega,\mathcal{F}_t,\mathbf{P})$ be a sufficiently rich filtered probability space. On $(\Omega,\mathcal{F}_t,\mathbf{P})$, let us define n independent Poisson point measures $N_i(dh,d\theta,ds)$ on $\mathbb{R}^d \times [0,1] \times \mathbb{R}_+$ with intensity*

$$q_i(dh,d\theta,ds) = p(h)\nu(dh)\kappa d\theta ds \quad \textit{for } 0 \leq i \leq n. \tag{9}$$

Let us also consider also a random variable X_0 on $\overline{\mathcal{X}^n}$, independent of the Poisson point measures N_i.

Then, assuming (Ha) and (Hb), we can define for any $\omega \in \Omega$ and $t \geq 0$

$$X(t,\omega) = X_0(\omega) + \sum_{i=1}^{n} \int_0^t \int_0^1 \int_{\mathbb{R}^d} (h)_i \mathbf{1}_{\left\{\theta \leq \frac{[g_i(x_i(s-,\omega)+h;X(s-,\omega))]_+}{\kappa} \frac{p(x_i(s-,\omega),h)}{p(h)}\right\}} N_i(dh,d\theta,ds)(\omega), \tag{10}$$

where $X(t,\omega) = (x_1(t,\omega),\ldots,x_n(t,\omega))$. Note that, since $p(x,\cdot)$ and g_i have been extended by δ_x on $\partial\mathcal{X}$ and by 0 $\partial\mathcal{X}^{n+1}$ respectively, this process is actually constant when $X_0 \in \partial\mathcal{X}^n$.

Note also that, in the expression above, g_i may be calculated at points out of its domain, when h is such that $x_i(s-,\omega) + h \notin \overline{\mathcal{X}}$. But in this case, $p(x_i(s-,\omega),h) = 0$, so let us admit that $[g_i(x_i(s-,\omega)+h;X(s-,\omega))]_+ p(x_i(s-,\omega),h) = 0$.

Note that the Poisson point measure N_i is well defined since ν is a Radon measure, and that the process $(X(t), t \geq 0)$ is well defined since

$$\begin{aligned}\int_0^t \int_0^1 \int_{\mathbb{R}^d} \|h\|^2 \mathbf{1}_{\left\{\theta \leq \frac{[g_i(x_i(s-)+h;X(s-))]_+}{\kappa} \frac{p(x_i(s-),h)}{p(h)}\right\}} q_i(dh,d\theta,ds)\\ \leq \int_0^t \int_{\mathbb{R}^d} \|h\|^2 [g_i(x_i(s-)+h;X(s-))]_+ p(x_i(s-),h)\nu(dh)ds < +\infty,\end{aligned} \tag{11}$$

because of (Ha) and (Hb).

We can show:

Proposition 1 *Assume (Ha) and (Hb). Then, L, defined in (6), is the infinitesimal generator of the Markov process $(X(t), t \geq 0)$ of Definition 1.*

Proof

For any $\varphi \in \mathcal{B}(\overline{\mathcal{X}^n})$, Itô's formula for jump processes writes

$$\begin{aligned}\varphi(X(t)) = \varphi(X_0) + \sum_{i=1}^{n} \int_0^t \int_0^1 \int_{\mathbb{R}^d} [\varphi(X(s-)+(h)_i) - \varphi(X(s-))] \times \\ \mathbf{1}_{\left\{\theta \leq \frac{[g_i(x_i(s-)+h;X(s-))]_+}{\kappa} \frac{p(x_i(s-),h)}{p(h)}\right\}} N_i(dh,d\theta,ds).\end{aligned} \tag{12}$$

Taking expectation, we get

$$\begin{aligned}\mathbf{E}[\varphi(X(t)) - \varphi(X_0)] &= \sum_{i=1}^{n} \mathbf{E}\left[\int_0^t \int_0^1 \int_{\mathbb{R}^d} [\varphi(X(s-)+(h)_i) - \varphi(X(s-))] \right.\\ &\qquad \left. \mathbf{1}_{\left\{\theta \leq \frac{[g_i(x_i(s-)+h;X(s-))]_+}{\kappa} \frac{p(x_i(s-),h)}{p(h)}\right\}} q_i(dh,d\theta,ds)\right] \\ &= \sum_{i=1}^{n} \mathbf{E}\left[\int_0^t \int_{\mathbb{R}^d} [\varphi(X(s-)+(h)_i) - \varphi(X(s-))] \right.\\ &\qquad \left. [g_i(x_i(s-)+h;X(s-))]_+ p(x_i(s-),dh)ds\right]\end{aligned} \tag{13}$$

and, taking $X_0 = X \in \overline{\mathcal{X}^n}$ constant, it is straightforward to show that $(\mathbf{E}[\varphi(X(t))] - \varphi(X))/t \to L\varphi(X)$ when $t \to 0$. □

3 Polymorphic canonical equation of adaptive dynamics

As explained in the introduction, we now intend to apply to the process X the biological limit of small mutations, more precisely, of small covariance matrix for $p(x,\cdot)$.

3.1 Rescaled process

Let us assume that $p(x,\cdot)$ has finite second order moments, and let us denote by $K(x)$ its covariance matrix:

$$\begin{gathered} K(x) = [k_{ij}(x)]_{1\leq i,j\leq d}\,, \\ \text{where } k_{ij}(x) = \int_{\mathbb{R}^d} h_i h_j p(x,dh) - \int_{\mathbb{R}^d} h_i p(x,dh) \int_{\mathbb{R}^d} h_j p(x,dh) \\ \text{for } i \text{ and } j \text{ in } \{1,\ldots,d\},\ \text{with } h=(h_1,\ldots,h_d). \end{gathered} \tag{14}$$

If $p(x,\cdot)$ was Gaussian with mean 0 and covariance matrix σI for all $x\in\mathcal{X}$, the limit of small mutation would simply correspond to $\sigma\to 0$. The analogous in our case is to introduce a parameter $\varepsilon>0$ and to replace $p(x,dh)$ by $p(x,dh/\varepsilon)$ (*i.e.* the image of the measure $p(x,\cdot)$ by the function $h\mapsto\varepsilon h$). Then, $\int_{\mathbb{R}^d} h_i h_j p(x,dh/\varepsilon) = \varepsilon^2 k_{ij}(x)$ goes to zero as $\varepsilon\to 0$.

So, assuming (Ha) and (Hb), let us consider, on the probability space $(\Omega,\mathcal{F}_t,\mathbf{P})$ of Definition 1, a family of random variables $\{X_0^\varepsilon\}_{0<\varepsilon\leq 1}$, and a family of Markov jump processes $\{(X_t^\varepsilon, t\geq 0)\}_{0<\varepsilon\leq 1}$, with paths in $\mathbb{D}(\mathbb{R}_+,\overline{\mathcal{X}^n})$, defined by

$$X_t^\varepsilon = X_0^\varepsilon + \sum_{i=1}^n \int_0^t \int_0^1 \int_{\mathbb{R}^d} (\varepsilon h)_i \mathbf{1}_{\left\{\theta\leq \frac{[g_i(x_i^\varepsilon(s-)+\varepsilon h; X_{s-}^\varepsilon)]_+}{\kappa}\frac{p(x_i^\varepsilon(s-),h)}{p(h)}\right\}} N_i\left(dh, d\theta, \frac{ds}{\varepsilon^2}\right), \tag{15}$$

where we used the notation $X_t^\varepsilon = (x_1^\varepsilon(t),\ldots,x_n^\varepsilon(t))$. Since the jumps of X^ε get smaller as $\varepsilon\to 0$, we need to accelerate jumps (otherwise, the process would be constant in the limit). That is why we wrote $N_i(dh,d\theta,ds/\varepsilon^2)$ instead of $N_i(dh,d\theta,ds)$: this is the proper rescaling of time that gives the canonical equation of adaptive dynamics of [4] in the monomorphic case (see [2]).

Since $\mathcal{X}$ is a convex set, as long as $\varepsilon\leq 1$, the jumps governed by the measure $p(x,dh/\varepsilon)$ cannot make the process go out of the state space $\mathcal{X}^n$ (that is why the processes X^ε is only defined for $0<\varepsilon\leq 1$).

Remark 1 *This is the only place where the assumption that $\mathcal{X}$ is convex is used. In fact, it would be sufficient to assume that $\exists\varepsilon_0>0$ such that $\forall\varepsilon<\varepsilon_0$, $\forall x\in\mathcal{X}$, $\varepsilon Supp(p(x,\cdot))\subset\mathcal{X}$, where $\varepsilon Supp(p(x,\cdot)) = \{\varepsilon y; y\in Supp(p(x,\cdot))\}$, and where $Supp(p(x,\cdot))$ is the support of the measure $p(x,\cdot)$. However, all the biological models involve convex set of traits.*

One can prove, exactly as in Proposition 1, the

Proposition 2 *Assuming (Ha) and (Hb), the infinitesimal generator of X^ε writes for $\varphi\in\mathcal{B}_b(\overline{\mathcal{X}^n})$*

$$L^\varepsilon\varphi(X) = \frac{1}{\varepsilon^2}\sum_{i=1}^n \int_{\mathbb{R}^d} (\varphi(X+(\varepsilon h)_i) - \varphi(X))[g_i(x_i+\varepsilon h; X)]_+ p(x_i,dh). \tag{16}$$

3.2 Convergence result

We first need to define metrics on sets of probability measures:

Definition 2 *For any $k \geq 1$, ρ_k is the Kantorovich metric (see Rachev [15]) defined on the set of probability measures on some measurable subset S of $\mathbb{R}^d$ with finite k^{th} order moments, given by*

$$\rho_k(P_1, P_2) = \inf \int_{S^2} c_k(x, y) R(dx, dy), \tag{17}$$

where the infimum is taken over the set of measures $R(dx, dy)$ on S^2 with marginales $P_1(dx)$ and $P_2(dy)$, and where

$$c_k(x, y) = \|x - y\| \sup\{\|x\|^{k-1}, \|y\|^{k-1}\}. \tag{18}$$

We will use the following property of the metric ρ_k, as a consequence of the dual form of the Kantorovich-Rubinstein metrics (Rachev [15])

Proposition 3 *For any probability measures P_1 and P_2 on $S \subset \mathbb{R}^d$ having finite k^{th} order moments,*

$$\sup \left\{ \int_S \psi d(P_1 - P_2);\ \textit{for } \psi \textit{ continuous and bounded on } S \right.$$
$$\left. \textit{such that } |\psi(x) - \psi(y)| \leq c_k(x, y) \right\} \leq \rho_k(P_1, P_2). \tag{19}$$

We will use more precisely the following consequence:

Corollary 1 *Consider a family $\{q(x, \cdot), x \in D\}$ of probability measures on $S \subset \mathbb{R}^d$ having finite k^{th} order moments, where $D \subset \mathbb{R}^m$. If $x \mapsto q(x, \cdot)$ is continuous (resp. Lipschitz) on D for the metric ρ_k, then, for any continuous function $\psi : S \to \mathbb{R}_+$ such that $|\psi(h) - \psi(h')| \leq K c_k(h, h')$ for some constant K, the function $x \to \int_S \psi(h) q(x, dh)$ is continuous (resp. Lipschitz) on D.*

Proof

Take q and ψ as in the statement. For all $N > 0$ and $a \in \mathbb{R}$, set $\xi_N(a) = a\mathbf{1}_{-N \leq a \leq N} + N\mathbf{1}_{a > N} - N\mathbf{1}_{a < -N}$ and $\psi_N = \xi_N \circ \psi$. ψ_N is bounded and continuous, and satisfies $|\psi_N(h) - \psi_N(h')| \leq K c_k(h, h')$ for all h and h' in S, so we can apply Proposition 3: for all x and x' in D, and $N > 0$,

$$\left| \int_S \psi_N(h)(q(x, dh) - q(x', dh)) \right| \leq K \rho_k(q(x, \cdot), q(x', \cdot)). \tag{20}$$

Now, by the dominated convergence Theorem (since $q(x, \cdot)$ and $q(x', \cdot)$ have finite k^{th} order moments), letting $N \to +\infty$, we get

$$\left| \int_S \psi(h)(q(x, dh) - q(x', dh)) \right| \leq K \rho_k(q(x, \cdot), q(x', \cdot)), \tag{21}$$

and this gives the result both when $x \mapsto q(x, \cdot)$ is continuous or Lipschitz. □

Before stating our convergence result, let us list all the assumptions involved in this theorem. We will, according to cases, assume (Hd) or (He).

(Hc) For all $i \in \{1, \dots, n\}$, $g_i(x; X)$ is continuous and $\mathcal{C}^1$ with respect to the first vector x on $\overline{\mathcal{X}^{n+1}}$, and $\nabla_1 g_i$, the gradient vector of g_i with respect to the first variable, is bounded and Lipschitz on $\overline{\mathcal{X}^{n+1}}$.

(Hd) The probability measure $p(x, \cdot)$ has finite and bounded third order moments on $\overline{\mathcal{X}}$, and $x \mapsto p(x, dh)$ is Lipschitz for the metric ρ_2 on $\overline{\mathcal{X}}$.

(He) The probability measures $p(x, \cdot)$ has finite and bounded third order moments, and its covariance matrix $K(x)$ has Lipschitz entries on $\mathcal{X}$.

Corollary 1 shows that (Hd) implies (He).

Now, let us state our first result:

Theorem 1 *Assume (Ha), (Hb), (Hc) and (Hd). Suppose also that the family of initial states* $\{X_0^\varepsilon\}_{0<\varepsilon\leq 1}$ *is bounded in* $\mathbb{L}^1(\overline{\mathcal{X}^n})$ *and converges in law to a random variable* X_0 *as* $\varepsilon \to 0$.

Then X^ε *converges when* $\varepsilon \to 0$, *for the Skorohod topology of* $\mathbb{D}\left(\mathbb{R}_+, \overline{\mathcal{X}^n}\right)$, *to the process* $(X(t) = (x_1(t), \dots, x_n(t)), t \geq 0)$ *with initial state* X_0 *and which sample paths are the unique solution to the (deterministic) system of differential equation*

$$\frac{dx_i}{dt} = \int_{\mathbb{R}^d} h[h \cdot \nabla_1 g_i(x_i; X)]_+ p(x_i, dh) \qquad \text{for } 1 \leq i \leq n. \tag{22}$$

In the case where $p(x, \cdot)$ *is a symmetrical measure on* $\mathbb{R}^d$ *for all* $x \in \mathcal{X}$, *this convergence holds under the less restrictive assumptions (Ha), (Hb), (Hc) and (He), and equation (22) gets the simpler form*

$$\frac{dx_i}{dt} = \frac{1}{2} K(x_i) \nabla_1 g_i(x_i; X) \qquad \text{for } 1 \leq i \leq n. \tag{23}$$

In the case of monomorphic evolution ($n = 1$), we recover the classical form of the canonical equation introduced in [4].

3.3 Proof of Theorem 1

First, the Cauchy-Lipschitz Theorem ensures global existence and unicity for the solutions to (22) in $\overline{\mathcal{X}^n}$: $X \mapsto \int_{\mathbb{R}^d} h[h \cdot \nabla_1 g_i(x_i; X)]_+ p(x_i, dh)$ is Lipschitz on $\overline{\mathcal{X}^n}$. This is actually a consequence of (Hc) and (Hd): $\nabla_1 g_i(x_i; X)$ is bounded by some constant K and K-Lipschitz, and $p(x, \cdot)$ has second order moments bounded by M_2, so

$$\begin{aligned} &\left| \int_{\mathbb{R}^d} h[h \cdot \nabla_1 g_i(x_i; X)]_+ p(x_i, dh) - \int_{\mathbb{R}^d} h[h \cdot \nabla_1 g_i(x_i'; X')]_+ p(x_i', dh) \right| \\ &\qquad \leq \int_{\mathbb{R}^d} \|h\| \times |[h \cdot \nabla_1 g_i(x_i; X)]_+ - [h \cdot \nabla_1 g_i(x_i'; X')]_+| p(x_i, dh) \\ &\qquad\qquad + \left| \int_{\mathbb{R}^d} h[h \cdot \nabla_1 g_i(x_i'; X')]_+ (p(x_i, dh) - p(x_i', dh)) \right|. \end{aligned} \tag{24}$$

Using the fact that $|[a]_+ - [b]_+| \leq |a - b|$, the first term is smaller than

$$\int_{\mathbb{R}^d} \|h\| |h \cdot (\nabla_1 g_i(x_i; X) - \nabla_1 g_i(x_i'; X'))| p(x_i, dh) \leq K(2\|X - X'\|) M_2, \tag{25}$$

where M_2 is a bound for the second order moments of $p(x, \cdot)$ for all $x \in \overline{\mathcal{X}}$, given by (Hd). The second term is bounded by some constant times $\|x_i - x_i'\|$ thanks to Corollary 1 for $k = 2$:

if we denote by ξ the vector $\nabla_1 g_i(x_i; X)$, and by ψ the function $h \mapsto h[h \cdot \xi]_+$, then

$$\begin{aligned}\|\psi(h) - \psi(h')\| &\leq \|(h-h')[h\cdot\xi]_+\| + \|h'([h\cdot\xi]_+ - [h'\cdot\xi]_+)\| \\ &\leq \|\xi\|\|h-h'\|\|h\| + \|h'\| \times |[h\cdot\xi]_+ - [h'\cdot\xi]_+| \\ &\leq \|\xi\| c_2(h,h') + \|h'\| \times |h\cdot\xi - h'\cdot\xi| \leq 2\|\xi\| c_2(h,h').\end{aligned} \tag{26}$$

Hence, X is well defined (note that, since $\nabla_1 g_i(x_i; X) = 0$ for $X \in \partial\mathcal{X}^n$, if for some $t \geq 0$ $X(t) \in \partial\mathcal{X}^n$, X is constant after t).

In the case of symmetrical mutations, a simple computation shows that

$$\begin{aligned}\int_{\mathbb{R}^d} h[h\cdot\nabla_1 g_i(x_i;X)]_+ p(x_i, dh) &= \frac{1}{2}\int_{\mathbb{R}^d} h(h\cdot\nabla_1 g_i(x_i;X)) p(x_i, dh) \\ &= \frac{1}{2} K(x_i)\nabla_1 g_i(x_i;X),\end{aligned} \tag{27}$$

so that equation (22) writes (23). To prove that this function is Lipschitz, one only needs assumptions (Hc) and (He) instead of (Hc) and (Hd).

The remaining of the proof will be divided in three steps: first, we show that the laws of X^ε form a tight family, second, that a uniform convergence result holds for the generators, and finally, that any accumulation point of the laws of X^ε when $\varepsilon \to 0$ is solution to the martingale problem associated with the process X with initial state X_0. As a solution to a (deterministic) SDE, the law of X is characterized by this martingale problem, so X^ε converges in law (for the Skorohod topology in $\mathbb{D}(\mathbb{R}_+, \overline{\mathcal{X}^n})$) to X as $\varepsilon \to 0$ (see *e.g.* Jacod and Shiryaev [9]).

This method requires to calculate the infinitesimal generator L^0 of the process X: for $\varphi \in \mathcal{C}^1(\overline{\mathcal{X}^n})$ and for a non-random $X_0 \in \overline{\mathcal{X}^n}$,

$$L^0\varphi(X_0) = \lim_{t\to 0}\frac{\varphi(X_t) - \varphi(X_0)}{t} = A(X_0)\cdot\nabla\varphi(X_0), \tag{28}$$

where $A(X_0) = (a_1(X_0), \dots, a_n(X_0)) \in (\mathbb{R}^d)^n$ is defined by

$$a_i(X) = \int_{\mathbb{R}^d} h[h\cdot\nabla_1 g_i(x_i;X)]_+ p(x_i, dh). \tag{29}$$

As shown above, A is Lipschitz on $\overline{\mathcal{X}^n}$.

Tightness of $\{\mathbf{P}^\varepsilon_{X_0^\varepsilon}\}_{\varepsilon>0}$ Let us denote by $\mathbf{P}^\varepsilon_{X_0^\varepsilon}$ the law on $\mathbb{D}(\mathbb{R}_+, \overline{\mathcal{X}^n})$ of the jump process X^ε with initial random state X_0^ε.

We will use the Rebolledo criterion for tightness (see Joffe and Metivier [10]) of the laws of families of semimartingales: X^ε writes

$$X_t^\varepsilon(\omega) = X_0^\varepsilon(\omega) + \sum_{i=1}^n \int_0^t \int_0^1 \int_{\mathbb{R}^d} H_i(h,\theta,s,\omega) N_i\left(dh, d\theta, \frac{ds}{\varepsilon^2}\right)(\omega) \tag{30}$$

where $H_i(h,\theta,s,\omega)$ is the previsible process

$$H_i(h,\theta,s,\omega) = (\varepsilon h)_i \mathbf{1}_{\left\{\theta \leq \frac{[g_i(x_i^\varepsilon(s-,\omega)+\varepsilon h; X^\varepsilon(s-,\omega))]_+}{\kappa}\frac{p(x_i^\varepsilon(s-,\omega),h)}{p(h)}\right\}}. \tag{31}$$

Since, by (Hd) (or (He) in the symmetrical case), $p(x,\cdot)$ has bounded second order moments, for any $i \in \{1,\dots,n\}$, H_i is square integrable with respect to the intensity $q_i(dh, d\theta, ds)$ of the Poisson point measure N_i (defined in Definition 1). Then one can write $X^\varepsilon = M^\varepsilon + V^\varepsilon$, where M^ε and V^ε are respectively a square-integrable martingale and a process with finite variation, given by

$$\begin{aligned} M_t^\varepsilon &= \sum_{i=1}^n \int_0^t \int_0^1 \int_{\mathbb{R}^d} H_i(h,\theta,s)\tilde{N}_i\left(dh, d\theta, \frac{ds}{\varepsilon^2}\right) \\ \text{and } V_t^\varepsilon &= X_0^\varepsilon + \sum_{i=1}^n \int_0^t \int_0^1 \int_{\mathbb{R}^d} H_i(h,\theta,s) q_i\left(dh, d\theta, \frac{ds}{\varepsilon^2}\right) \end{aligned} \tag{32}$$

where $\tilde{N}_i$ is the compensated Poisson measure $N_i - q_i$. Moreover, since the H_i charge independent subspaces of $\mathbb{R}^{nd}$, the Meyer process of the martingale M^ε is given by

$$\langle M^\varepsilon \rangle_t = \sum_{i=1}^n \int_0^t \int_0^1 \int_{\mathbb{R}^d} H_i^2(h,\theta,s) q_i\left(dh, d\theta, \frac{ds}{\varepsilon^2}\right) \tag{33}$$

where the square of H_i is taken coordinate by coordinate (*i.e.* for $h = (h_i)_{1\leq i\leq nd} \in \mathbb{R}^{nd}$, $h^2 = (h_i^2)_{1\leq i\leq nd}$). We want to show that $\langle M^\varepsilon\rangle$ and V^ε satisfy Aldous criterion (see [10]).

Using equation (31) for H_i and equation (9) for q_i, we obtain for any $T>0$, $\xi>0$, $\eta>0$ and $\theta>0$, and for any family $\{\tau_\varepsilon\}$ of stopping times,

$$\begin{aligned} &\|\langle M^\varepsilon\rangle_{\tau_\varepsilon+\theta} - \langle M^\varepsilon\rangle_{\tau_\varepsilon}\| \\ &\leq \sum_{i=1}^n \int_{\tau_\varepsilon}^{\tau_\varepsilon+\theta} \int_{\mathbb{R}^d} \|h^2\| |g_i(x_i(s-)+\varepsilon h; X_{s-}^\varepsilon)| p(x_i^\varepsilon(s-), h)\nu(dh)ds \\ &\leq n\kappa M_2 \theta, \end{aligned} \tag{34}$$

where κ is defined in (Hb). For the second inequality, we used that $\|h^2\| = \sqrt{\sum h_i^4} \leq \sqrt{(\sum h_i^2)^2} = \|h\|^2$. This gives one part of the Aldous criterion for $\langle M^\varepsilon\rangle$. For the other part, we have to prove the tightness of $\{\langle M^\varepsilon\rangle_t\}_{\varepsilon>0}$ for t in a dense subset of $\mathbb{R}_+$. But this is obvious since (34) for $\tau_\varepsilon = 0$ gives that $\|\langle M^\varepsilon\rangle_t\| \leq n\kappa M_2 t$.

Using the fact that $|g_i(x+\varepsilon h; X)| \leq \varepsilon K\|h\|$ (remember that $\nabla_1 g_i$ is bounded by K, and that (BH5) implies that $g_i(x;X) = 0$ as soon as x is one coordinate of X), a very similar calculation shows that $\|V_{\tau_\varepsilon+\theta}^\varepsilon - V_{\tau_\varepsilon}^\varepsilon\| \leq nKM_2\theta$. In the same way, using the fact that X_0^ε is $\mathbb{L}^1$, it is easy to see that $\|V_t^\varepsilon\| \leq \|X_0^\varepsilon\| + nKM_2 t$, so the tightness of V_t^ε is a consequence of the tightness of X_0^ε, which holds since X_0^ε converges in law to the random variable X_0. Then, Aldous criterion is also true for V^ε.

So, by Rebolledo's criterion, the laws $\mathbf{P}_{X_0^\varepsilon}^\varepsilon$ of X^ε form a tight family.

Convergence of the generators Let us now prove that

$$\forall \varphi \in \mathcal{C}_b^2\left(\overline{\mathcal{X}^n}\right),\ L^\varepsilon\varphi \to L^0\varphi \text{ uniformly on } \overline{\mathcal{X}^n}, \tag{35}$$

where L^ε is defined in (16) and L^0 is defined in (28).

For $X \in \partial\mathcal{X}^n$, there is no problem since everything is 0. So let us fix some $X \in \mathcal{X}^n$. We can write

$$L^0\varphi(X) = \sum_{i=1}^n \int_{\mathbb{R}^d} (h\cdot\nabla_i\varphi(X))[h\cdot\nabla_1 g_i(x_i;X)]_+ p(x_i, dh), \tag{36}$$

where $\nabla_i\varphi(X)$ is the gradient vector of $\varphi(X)$ considered as a fonction of the i^{th} coordinate $x_i \in \mathcal{X}$ of $X \in \mathcal{X}^n$. Note that this equation can be obtained by expanding to the first order

φ and g_i for $1 \leq i \leq n$ with respect to ε in the expression (16) for L^ε. To make this precise, let us fix a function φ in $\mathcal{C}_b^2\left(\overline{\mathcal{X}^n}\right)$ and let us write

$$\begin{aligned}
&|L^\varepsilon\varphi(X) - L^0\varphi(X)| \\
&\leq \sum_{i=1}^n \int_{\mathbb{R}^d} \left|\frac{\varphi(X+(\varepsilon h)_i)-\varphi(X)}{\varepsilon}\right| \times \left|\left[\frac{g_i(x_i+\varepsilon h; X)}{\varepsilon}\right]_+ - [h\cdot\nabla_1 g_i(x_i;X)]_+\right| p(x_i,dh) \\
&\qquad + \sum_{i=1}^n \int_{\mathbb{R}^d} [h\cdot\nabla_1 g_i(x_i;X)]_+ \times \left|\frac{\varphi(X+(\varepsilon h)_i)-\varphi(X)}{\varepsilon} - h\cdot\nabla_i\varphi(X)\right| p(x_i,dh). \quad (37)
\end{aligned}$$

Let us call B_i and C_i the quantities inside the integral in the i^{th} term of the first and the second sum, respectively. Now, φ is $\mathcal{C}^1$, $g_i(x_j;X)=0$ for all i and j in $\{1,\ldots,n\}$ by (BH5), and, by (Hc), for $1\leq i\leq n$, $g_i(x;X)$ is $\mathcal{C}^1$ with respect to the first variable x. So, we can find θ_1, θ_2 and θ_3 in $[0,1]$ depending on X, h and i such that

$$\begin{aligned}
B_i &= |h\cdot\nabla_i\varphi(X+(\theta_1\varepsilon h)_i)| \times |[h\cdot\nabla_1 g_i(x_i+\theta_2\varepsilon h;X)]_+ - [h\cdot\nabla_1 g_i(x_i;X)]_+| \\
\text{and}\quad C_i &= [h\cdot\nabla_1 g_i(x_i;X)]_+ \times |h\cdot\nabla_i\varphi(X+(\theta_3\varepsilon h)) - h\cdot\nabla_i\varphi(X)|.
\end{aligned} \quad (38)$$

Now, since φ is $\mathcal{C}^2$ with bounded first and second order derivatives, and, because of (Hc), we can choose a number K such that for all $i\in\{1,\ldots,n\}$, $\nabla_i\varphi$ and $\nabla_1 g_i$ are both K-Lipschitz and bounded by K on $\mathcal{X}^n$ and $\mathcal{X}^{n+1}$ respectively. Then

$$\begin{aligned}
B_i &\leq K\|h\| \times |h\cdot\nabla_1 g_i(x_i+\theta_2\varepsilon h;X) - h\cdot\nabla_1 g_i(x_i;X)| \leq \varepsilon K^2\|h\|^3 \\
\text{and}\quad C_i &\leq K\|h\| \times \|h\|K\|\theta_3\varepsilon h\| \leq \varepsilon K^2\|h\|^3.
\end{aligned} \quad (39)$$

It remains to put this bound in equation (37) to obtain:

$$|L^\varepsilon\varphi(X) - L^0\varphi(X)| \leq 2\varepsilon K^2 \sum_{i=1}^n \int_{\mathbb{R}^d} \|h\|^3 p(x_i,h)\nu(dh), \quad (40)$$

and the integrals of the right hand side are, by (Hd) (or (He) in the symmetrical case), bounded on $\mathcal{X}$, which ends the proof of (35).

Martingale problem for P Finally, let us use (35) to show that any accumulation point $\mathbf{P}$ of the family of laws $\{\mathbf{P}^\varepsilon_{X_0^\varepsilon}\}$ on $\mathbb{D}(\mathbb{R}_+,\overline{\mathcal{X}^n})$ solves the martingale problem for L^0 with initial condition X_0. Let us endow the space $\mathbb{D}(\mathbb{R}_+,\overline{\mathcal{X}^n})$ with the canonical filtration $\mathcal{G}_t$, and for any $\varphi\in\mathcal{C}^2(\overline{\mathcal{X}^n})$, let us define on this space the process

$$M_t^\varphi(w) = \varphi(w_t) - \varphi(w_0) - \int_0^t L\varphi(w_s)ds. \quad (41)$$

We have to show that, under $\mathbf{P}$, M^φ is a local $\mathcal{G}_t$-martingale. We already know that under $\mathbf{P}^\varepsilon_{X_0^\varepsilon}$, $M_t^{\varepsilon,\varphi}(w) = \varphi(w_t)-\varphi(w_0)-\int_0^t L^\varepsilon\varphi(w_s)ds$ is a local $\mathcal{G}_t$-martingale with initial state X_0^ε. Since by (Hb), g_i is bounded for all $i\in\{1,\ldots,n\}$, this is a square-integrable martingale as soon as $\varphi\in\mathcal{C}_c^2(\overline{\mathcal{X}^n})$.

Consider an extracted sequence $\{\mathbf{P}^{\varepsilon_k}_{X_0^{\varepsilon_k}}\}$ of $\{\mathbf{P}^\varepsilon_{X_0^\varepsilon}\}$ converging to $\mathbf{P}$ on $\mathbb{D}(\mathbb{R}_+,\overline{\mathcal{X}^n})$. Let us consider a fixed $\varphi\in\mathcal{C}_c^2(\overline{\mathcal{X}^n})$. Then $L^0\varphi$ is continuous on $\overline{\mathcal{X}^n}$. Now, fix $s>0$ and $t>s$, and consider p real numbers $0\leq t_1<\ldots<t_p\leq s$ for some $p\geq 1$, and a continuous bounded

function $q : (\mathbb{R}^{nd})^p \to \mathbb{R}$. We can write

$$
\begin{aligned}
&\left|\mathbf{E}_{\mathbf{P}}\left\{q(w_{t_1},\,\ldots,w_{t_p})\left[\varphi(w_t)-\varphi(w_s)-\int_s^t L^0\varphi(w_u)du\right]\right\}\right| \\
&\qquad \le \left|\mathbf{E}_{\mathbf{P}^{\varepsilon_k}_{X_0^{\varepsilon_k}}}\left\{q(w_{t_1},\ldots,w_{t_p})\left[\varphi(w_t)-\varphi(w_s)-\int_s^t L^{\varepsilon_k}\varphi(w_u)du\right]\right\}\right| \\
&\qquad + \left|\mathbf{E}_{\mathbf{P}^{\varepsilon_k}_{X_0^{\varepsilon_k}}}\left\{q(w_{t_1},\ldots,w_{t_p})\int_s^t [L^{\varepsilon_k}\varphi(w_u)-L^0\varphi(w_u)]du\right\}\right| \qquad (42)\\
&\qquad + \left|\mathbf{E}_{\mathbf{P}}\left\{q(w_{t_1},\ldots,w_{t_p})\left[\varphi(w_t)-\varphi(w_s)-\int_s^t L^0\varphi(w_u)du\right]\right\}\right. \\
&\qquad\qquad \left. -\mathbf{E}_{\mathbf{P}^{\varepsilon_k}_{X_0^{\varepsilon_k}}}\left\{q(w_{t_1},\ldots,w_{t_p})\left[\varphi(w_t)-\varphi(w_s)-\int_s^t L^0\varphi(w_u)du\right]\right\}\right|.
\end{aligned}
$$

The first term of the right-hand side is 0 since $M^{\varepsilon_k,\varphi}$ is a $\mathbf{P}^{\varepsilon_k}_{X_0^{\varepsilon_k}}$-martingale. Because of the uniform convergence of generators (35), the second term converges to 0 when k goes to infinity. Finally, the third term also goes to 0 when k goes to infinity since $\mathbf{P}^{\varepsilon_k}_{X_0^{\varepsilon_k}} \Rightarrow \mathbf{P}$ and since $w \mapsto q(w_{t_1},\ldots,w_{t_p})\left[\varphi(w_t)-\varphi(w_s)-\int_s^t L^0\varphi(w_u)du\right]$ is bounded and continuous for the weak topology on $\mathbb{D}([0,t],\overline{\mathcal{X}^n})$. Since the left-hand side does not depend on k, it is 0.

A classical use of the monotone class Theorem allows to extend this equality to all $\mathcal{G}_s$-measurable bounded function q, so M^φ is a $\mathbf{P}$-martingale.

It remains to extend this result to any function $\varphi \in \mathcal{C}^2(\overline{\mathcal{X}^n})$ by a truncation technique. Fix some $T > 0$. Using expression (30) for X^ε, for any $t \ge 0$, we have

$$
\|X_t^\varepsilon\| \le \|X_0^\varepsilon\| + \sum_{i=1}^n \int_0^t \int_0^1 \int_{\mathbb{R}^d} \|H_i(h,\theta,s)\| N_i\left(dh, d\theta, \frac{ds}{\varepsilon}\right), \tag{43}
$$

so that a calculation similar to (34) gives that $\mathbf{E}\left[\sup_{0\le t\le T}\|X_t^\varepsilon\|\right] \le \mathbf{E}[\|X_0^\varepsilon\|] + nKTM_2$. Since $\{X_0^\varepsilon\}_{\varepsilon>0}$ has been supposed bounded in $\mathbb{L}^1$, $\mathbf{E}\left[\sup_{0\le t\le T}\|X_t^\varepsilon\|\right] = \mathbf{E}_{\mathbf{P}_{X_0^\varepsilon}^\varepsilon}\left[\sup_{0\le t\le T}\|w_t\|\right] \le K_T$ with K_T depending only on T.

The same holds for a process with law $\mathbf{P}$: for any $A > 0$, since $\mathbf{P}^{\varepsilon_k}_{X_0^{\varepsilon_k}} \Rightarrow \mathbf{P}$ on $\mathbb{D}(\mathbb{R}_+, \overline{\mathcal{X}^n})$, we have $\mathbf{E}_{\mathbf{P}}\left[\sup_{0\le t\le T}\|w_t\| \wedge A\right] \le K_T$, and letting A grow to $+\infty$, we get $\mathbf{E}_{\mathbf{P}}\left[\sup_{0\le t\le T}\|w_t\|\right] \le K_T$.

Hence, for $i \ge 1$, the stopping times

$$
T_i = \inf\left\{t \ge 0 : \sup_{0\le t\le T}\|w_t\| \ge i\right\} \tag{44}
$$

satisfy $T_i \to +\infty$, $\mathbf{P}$-a.s as $i \to +\infty$.

Now, fix some $\varphi \in \mathcal{C}^2\left(\overline{\mathcal{X}^n}\right)$. For any $i \ge 1$, $M^\varphi_{T_i\wedge t} = \varphi(w_{T_i\wedge t}) - \varphi(X_0) - \int_0^{T_i\wedge t} L^0\varphi(w_s)ds$ is a $\mathbf{P}$-martingale since $M^\varphi_{T_i\wedge t}$ does not change if we replace φ by a compact-supported function $\tilde\varphi$ equal to φ on the ball $B(0,i)$ of radius i centered at 0. Since $T_i \to +\infty$, $\mathbf{P}$-a.s. as $i \to +\infty$, M^φ is a local martingale.

Moreover, the law of w_0 under $\mathbf{P}$ is the limit in law of the law of w_0 under $\mathbf{P}^{\varepsilon_k}_{X_0^{\varepsilon_k}}$, *i.e.* the law of $X_0^{\varepsilon_k}$. We have supposed that X_0^ε converges in law to X_0, so we have proved that $\mathbf{P}$ solves the martingale problem for L^0 with initial state X_0. □

4 Diffusion model of adaptive dynamics: existence

As explained in the introduction, we now study a diffusion model of adaptive dynamics in the case of symmetrical mutations (we only need this assumption in Theorem 2), the biological interest of which is to allow for evolution in any direction.

4.1 Diffusion generator

The generator of our diffusion model will be obtained by expanding in powers of $\varepsilon > 0$ to the first order the generator L^ε of the trait substitution sequence X^ε defined in (16). When ε is small, which corresponds to the biological assumption of small mutations, an approximation of L^ε will lead to the required second order differential operator.

Proposition 4 *Assume that for all $i \in \{1, \dots, n\}$, $g_i(x; X)$ is $\mathcal{C}^2$ with respect to the first variable x, and that its Hessian matrix $H_1 g_i$ with respect to the first variable is bounded and has Lipschitz entries on $\mathcal{X}^{n+1}$. Assume also that $p(x, \cdot)$ has finite and bounded fifth order moments, and is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$.*

Then, fix $\varphi \in \mathcal{C}_b^3(\overline{\mathcal{X}^n})$. $\forall X \in \mathcal{X}^n$ such that $\nabla_1 g_i(x_i; X) \neq 0$,

$$\begin{aligned} L^\varepsilon \varphi(X) = \sum_{i=1}^n \Bigg(& \int_{\mathbb{R}^d} (h \cdot \nabla_i \varphi(X))[h \cdot \nabla_1 g_i(x_i; X)]_+ p(x_i, dh) \\ & + \frac{\varepsilon}{2} \int_{\mathbb{R}^d} h^* H_i \varphi(X) h [h \cdot \nabla_1 g_i(x_i; X)]_+ p(x_i, dh) \\ & + \frac{\varepsilon}{2} \int_{\mathbb{R}^d} (h \cdot \nabla_i \varphi(X)) \mathbf{1}_{\{h \cdot \nabla_1 g_i(x_i; X) > 0\}} h^* H_1 g_i(x_i; X) h p(x_i, dh) \Bigg) + o(\varepsilon), \end{aligned} \tag{45}$$

and $\forall X \in \mathcal{X}^n$ such that $\nabla_1 g_i(x_i; X) = 0$,

$$L^\varepsilon \varphi(X) = \frac{\varepsilon}{2} \sum_{i=1}^n \int_{\mathbb{R}^d} (h \cdot \nabla_i \varphi(X)) \left[h^* H_1 g_i(x_i; X) h \right]_+ p(x_i, dh) + o(\varepsilon). \tag{46}$$

Proof

Fix some $\varphi \in \mathcal{C}_b^3(\overline{\mathcal{X}^n})$ and some $X \in \mathcal{X}^n$.

Expanding to the second order φ and g_i in equation (16), we can write

$$\begin{aligned} L^\varepsilon \varphi(X) = \sum_{i=1}^n \int_{\mathbb{R}^d} & \Big(h \cdot \nabla_i \varphi(X) + \frac{\varepsilon}{2} h^* H_i \varphi(X) h + o(\|h\|^2 \varepsilon) \Big) \\ & \Big[h \cdot \nabla_1 g_i(x_i; X) + \frac{\varepsilon}{2} h^* H_1 g_i(x_i; X) h + o(\|h\|^2 \varepsilon) \Big]_+ p(x_i, dh), \end{aligned} \tag{47}$$

where $H_i \varphi$ is the Hessian matrix of $\varphi(X)$ considered as a function of the i^{th} coordinate of X.

Using the fact that g_i is $\mathcal{C}^2$ with respect to the first variable, that $H_1 g_i$ is bounded and Lipschitz on $\mathcal{X}^{n+1}$, and that $p(x, \cdot)$ has bounded fifth order moments, one can do the same computation as in the part "convergence of generators" of the proof of Theorem 1 in section 3.3, in order to "get the $o(\varepsilon)$ out of the product":

$$\begin{aligned} L^\varepsilon \varphi(X) = \sum_{i=1}^n \int_{\mathbb{R}^d} & \Big(h \cdot \nabla_i \varphi(X) + \frac{\varepsilon}{2} h^* H_i \varphi(X) h \Big) \\ & \Big[h \cdot \nabla_1 g_i(x_i; X) + \frac{\varepsilon}{2} h^* H_1 g_i(x_i; X) h \Big]_+ p(x_i, dh) + o(\varepsilon). \end{aligned} \tag{48}$$

The computation of Theorem 1 applies without difficulty. Let us omit the details of the calculation.

Now, the result is obvious when $\nabla_1 g_i(x_i;X)=0$.

When $\nabla_1 g_i(x_i;X)\neq 0$, this is more technical. Define

$$\begin{aligned} C=\{h\in\mathbb{R}^d; h\cdot\nabla_1 g_i(x_i;X)+\varepsilon/2h^*H_1g_i(x_i;X)h>0\} \\ \text{and } D=\{h\in\mathbb{R}^d; h\cdot\nabla_1 g_i(x_i;X)>0\}. \end{aligned} \tag{49}$$

Let us rewrite

$$\begin{aligned} L^\varepsilon\varphi(X) &= \sum_{i=1}^n \int_C \Big(h\cdot\nabla_i\varphi(X)+\frac{\varepsilon}{2}h^*H_i\varphi(X)h\Big) \\ &\qquad \Big(h\cdot\nabla_1 g_i(x_i;X)+\frac{\varepsilon}{2}h^*H_1g_i(x_i;X)h\Big)\,p(x_i,dh)+o(\varepsilon) \\ &= \sum_{i=1}^n \Big(\int_C (h\cdot\nabla_i\varphi(X))(h\cdot\nabla_1 g_i(x_i;X))p(x_i,dh) \\ &\quad +\frac{\varepsilon}{2}\int_C (h^*H_i\varphi(X)h)(h\cdot\nabla_1 g_i(x_i;X))p(x_i,dh) \\ &\quad +\frac{\varepsilon}{2}\int_C (h\cdot\nabla_i\varphi(X))(h^*H_1g_i(x_i;X)h)p(x_i,dh)\Big)+o(\varepsilon). \end{aligned} \tag{50}$$

We have to show that each integral over C in the right-hand side differs from the same integral taken over D by a $o(\varepsilon)$ quantity. Let us prove it: first,

$$\begin{aligned} &\left|\int_C (h\cdot\nabla_i\varphi(X))(h\cdot\nabla_1 g_i(x_i;X))p(x_i,dh)-\int_D (h\cdot\nabla_i\varphi(X))(h\cdot\nabla_1 g_i(x_i;X))p(x_i,dh)\right| \\ &\qquad \le \int_{C\cap D^c} |h\cdot\nabla_i\varphi(X)|\times|h\cdot\nabla_1 g_i(x_i;X)|p(x_i,dh) \\ &\qquad\qquad + \int_{C^c\cap D} |h\cdot\nabla_i\varphi(X)|\times|h\cdot\nabla_1 g_i(x_i;X)|p(x_i,dh). \end{aligned} \tag{51}$$

On $C\cap D^c$, $h\cdot\nabla_1 g_i(x_i;X)\le 0$ and $h\cdot\nabla_1 g_i(x_i;X)+\varepsilon/2(h^*H_1g_i(x_i;X)h)>0$, so $|h\cdot\nabla_1 g_i(x_i;X)|\le\varepsilon/2(h^*H_1g_i(x_i;X)h)$. Similarly, the same is true on $C^c\cap D$. Since we assumed that there is some constant K bounding H_1g_i and $\nabla_i\varphi$, the quantity above is smaller than

$$\frac{\varepsilon}{2}K^2\int_{C\cap D^c}\|h\|^3p(x_i,dh)+\frac{\varepsilon}{2}K^2\int_{C^c\cap D}\|h\|^3p(x_i,dh). \tag{52}$$

Now the set $C\cap D^c$ converges to the set $\{h\cdot\nabla_1 g_i(x_i;X)=0 \text{ and } h^*H_1g_i(x_i;X)h>0\}$ as $\varepsilon\to 0$, which has Lebesgue measure 0, and the set $C^c\cap D$ converges to $\emptyset$ as $\varepsilon\to 0$. Since $p(x_i,\cdot)$ has finite third order moments, and is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$, by the dominated convergence Theorem, the quantity in the right-hand side of (51) is $o(\varepsilon)$ as $\varepsilon\to 0$.

The same method proves that

$$\begin{aligned} &\frac{\varepsilon}{2}\int_C (h^*H_i\varphi(X)h)(h\cdot\nabla_1 g_i(x_i;X))p(x_i,dh) \\ &\qquad\qquad = \frac{\varepsilon}{2}\int_D (h^*H_i\varphi(X)h)(h\cdot\nabla_1 g_i(x_i;X))p(x_i,dh)+o(\varepsilon^2) \end{aligned} \tag{53}$$

and that

$$\frac{\varepsilon}{2}\int_C (h\cdot\nabla_i\varphi(X))(h^*H_1g_i(x_i;X)h)p(x_i,dh)$$
$$=\frac{\varepsilon}{2}\int_D (h\cdot\nabla_i\varphi(X))(h^*H_1g_i(x_i;X)h)p(x_i,dh)+o(\varepsilon),\quad (54)$$

as required. □

Negliging the terms of order greater than one in (45) and (46), we obtain the following second order differential operator:

$$\tilde{L}^\varepsilon\varphi(X)=\sum_{i=1}^n\sum_{k=1}^d(b_k^i(X)+\varepsilon\tilde{b}_k^i(X))\frac{\partial_i\varphi}{\partial_i x_k}(X)+\frac{\varepsilon}{2}\sum_{i=1}^n\sum_{1\leq k,l\leq d}a_{kl}^i(X)\frac{\partial_i^2\varphi}{\partial_i x_k\partial_i x_l}(X),\quad (55)$$

where $\partial_i\varphi/\partial_i x_k(X)$ denotes the partial derivative of φ with respect to the k^{th} coordinate of the i^{th} vector $x_i\in\mathcal{X}$ of $X\in\mathcal{X}^n$, and where

$$b_k^i(X)=\int_{\mathbb{R}^d}h_k[\nabla_1 g_i(x_i;X)\cdot h]_+p(x_i,dh),$$
$$\tilde{b}_k^i(X)=\begin{cases}\frac{1}{2}\int_{\{h\cdot\nabla_1 g_i(x_i;X)>0\}}h_k(h^*H_1g_i(x_i;X)h)p(x_i,dh) \\ \quad\text{when } \nabla_1 g_i(x_i;X)\neq 0,\\ \frac{1}{2}\int_{\mathbb{R}^d}h_k[h^*H_1g_i(x_i;X)h]_+p(x_i,dh)\\ \quad\text{when } \nabla_1 g_i(x_i;X)=0,\end{cases}\quad (56)$$
$$\text{and } a_{kl}^i(X)=\int_{\mathbb{R}^d}h_kh_l[h\cdot\nabla_1 g_i(x_i;X)]_+p(x_i,dh).$$

Note that when $p(x,\cdot)$ is symmetrical on $\mathbb{R}^d$ for all $x\in\mathcal{X}$, $\tilde{b}_k^i(X)=0$ for X such that $\nabla_1 g_i(x_i;X)=0$ (make the change of variable $h'=-h$). This is the reason why the symmetry of $p(x,\cdot)$ will be necessary in Theorem 2.

Define $a^i=[a_{kl}^i]_{1\leq k,l\leq d}$ for all $i\in\{1,\dots,n\}$, and a the block diagonal matrix with blocks a^i for $1\leq i\leq n$. Define also $b^i(X)=(b_1^i(X),\dots,b_d^i(X))$ for $1\leq i\leq n$ and $b(X)=(b^1(X),\dots,b^n(X))$, and the same for $\tilde{b}^i(X)$ and $\tilde{b}(X)$, replacing $b_k^i(X)$ by $\tilde{b}_k^i(X)$. We will also set for convenience for all $X\in\mathcal{X}^n$

$$b^\varepsilon(X)=b(X)+\varepsilon\tilde{b}(X).\quad (57)$$

As in section 2.2, we need to extend the generator $\tilde{L}^\varepsilon$ to functions defined on a Polish space. Our method requires to extend it to functions defined on $\mathbb{R}^{nd}$. So let us prolong by 0 the functions b_k^i, $\tilde{b}_k^i$ and a_{kl}^i on $\mathbb{R}^{nd}\setminus\mathcal{X}^n$, and let us rewrite operator $\tilde{L}$ as follows: for any $\varphi\in\mathcal{C}^2(\mathbb{R}^{nd})$ and for any $X\in\mathbb{R}^{nd}$,

$$\tilde{L}^\varepsilon\varphi(X)=b^\varepsilon(X)\cdot\nabla\varphi(X)+\frac{\varepsilon}{2}\sum_{i=1}^n\sum_{1\leq k,l\leq d}a_{kl}^i(X)\frac{\partial_i^2\varphi}{\partial_i x_k\partial_i x_l}(X).\quad (58)$$

4.2 Existence and precise description of X^ε

We now intend to define a Markov process Y^ε on $\mathcal{C}(\mathbb{R}_+,\mathbb{R}^{nd})$ solving the martingale problem for the generator $\tilde{L}^\varepsilon$ defined in (58), *i.e.* a process Y^ε weak solution to the stochastic differential equation

$$dY_t^\varepsilon=b^\varepsilon(Y_t^\varepsilon)dt+\sqrt{\varepsilon}\sigma(Y_t^\varepsilon)dW_t,\quad (59)$$

where $\sigma(X)$ is a given $nd \times k$ matrix such that $\sigma(X)\sigma^*(X) = a(X)$, and where W_t is a standard k-dimensional Brownian motion. As will show Proposition 5, this SDE does not fit the classical hypotheses for weak existence of solution: b^ε is not continuous, and a is not uniformly non-degenerate.

A process solution to (59) has a biological interest by itself, so we will prove its existence without taking into account the hypotheses of Proposition 4 needed to obtain the operator $\tilde{L}^\varepsilon$. We need the assumptions

(H1) $g_i(x;X)$ is $\mathcal{C}^2$ with respect to the first variable x on $\overline{\mathcal{X}^{n+1}}$, $\nabla_1 g_i$ and $H_1 g_i$ are continuous and bounded on $\overline{\mathcal{X}^{n+1}}$, and $\nabla_1 g_i(x_i;X) = 0$ and $H_1 g_i(x_i;X) = 0$ when $X \in \partial\mathcal{X}^n$.

(H2) $p(x,\cdot)$ has finite and bounded third order moments and is continuous on $\overline{\mathcal{X}}$ for the distance ρ_3 of Definition 2, and, for all $x \in \mathcal{X}$, it is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$. Then $p(x,dh) = p(x,h)dh$. Moreover, there is some measurable function $p(h)$ such that for all $x \in \mathcal{X}$ and $h \in \mathbb{R}^d$, $p(x,h) \leq p(h)$, and such that $\int_{\mathbb{R}^d} \|h\|^3 p(h)dh < \infty$.

(H3) Let us define for $1 \leq i \leq n$

$$\begin{aligned} \Gamma_i &= \{X \in \mathcal{X}^n; \nabla_1 g_i(x_i;X) = 0\} \\ \text{and } \Gamma &= \bigcup_{i=1}^n \Gamma_i. \end{aligned} \tag{60}$$

Then, the points of Γ are isolated points of $\mathbb{R}^{nd}$.

Now, let us state some consequences of these assumptions for a, b and $\tilde{b}$:

Proposition 5 *Assume (H1) and (H2). Then, a, b and $\tilde{b}$, defined in (56), are bounded, and a and b are continuous on $\mathbb{R}^{nd}$ (remind that a, b and $\tilde{b}$ has been extended by 0 on $\mathbb{R}^{nd} \setminus \mathcal{X}^n$). Moreover, the matrix $a(X)$ is symmetrical non-negative for $X \in \mathcal{X}^n$, and $\tilde{b}^i$ is continuous on $\mathbb{R}^{nd} \setminus \Gamma_i$.*

Proof

The boundedness properties are obvious. Moreover, for $X \in \mathcal{X}^n$, a is obviously symmetrical, and, given any vector $v = (v_1, \ldots, v_d)$ in $\mathbb{R}^d$ and any $i \in \{1, \ldots, n\}$, an easy calculation shows that

$$v^* a^i(X) v = \int_{\mathbb{R}^d} (h \cdot v)^2 [h \cdot \nabla_1 g_i(x_i;X)]_+ p(x_i,dh) \geq 0. \tag{61}$$

By (H1), $\nabla_1 g_i$ is continuous on $\overline{\mathcal{X}^{n+1}}$, so, for any X and X' in $\overline{\mathcal{X}^n}$, and for $1 \leq k \leq d$ and $1 \leq i \leq n$,

$$\begin{aligned} |b^i_k(X) - b^i_k(X')| \leq & \left| \int_{\mathbb{R}^d} h_k \left([\nabla_1 g_i(x_i;X) \cdot h]_+ - [\nabla_1 g_i(x'_i;X') \cdot h]_+ \right) p(x_i,dh) \right| \\ & + \left| \int_{\mathbb{R}^d} h_k [\nabla_1 g_i(x_i;X) \cdot h]_+ (p(x_i,dh) - p(x'_i,dh)) \right|. \end{aligned} \tag{62}$$

The first term of the right-hand side converges to 0 when $X' \to X$ because $\nabla_1 g_i$ is continuous on $\overline{\mathcal{X}^n}$ and $p(x,\cdot)$ has bounded second order moments, and the second term also converges to 0 by Corollary 1, since $p(x,\cdot)$ is continuous for the metric ρ_3. So b is continuous on $\overline{\mathcal{X}^n}$. A similar computation shows that the same holds for a.

It remains to show the continuity of $\tilde{b}^i$ on $\mathbb{R}^{nd} \setminus \Gamma_i$. (H1) shows that it is continuous on $\mathbb{R}^{nd} \setminus \mathcal{X}^n$. Fix X and X' in $\mathcal{X}^n \setminus \Gamma_i$ and define $S = \{h; h \cdot \nabla_1 g_i(x_i; X) > 0\}$ and $S' = \{h; h \cdot \nabla_1 g_i(x'_i; X') > 0\}$. Then

$$\begin{aligned}
|\tilde{b}^i_k(X) - \tilde{b}^i_k(X')| \leq \frac{1}{2} \Bigg| & \int_{S\cap S'} h_k[h^*(H_1 g_i(x_i; X) - H_1 g_i(x'_i; X'))h] p(x'_i, dh) \\
& + \int_S h_k(h^* H_1 g_i(x_i; X)h)(p(x_i, dh) - p(x'_i, dh)) \\
& - \int_{S\cap S'^c} h_k(h^* H_1 g_i(x_i; X)h) p(x'_i, dh) \\
& - \int_{S^c\cap S'} h_k(h^* H_1 g_i(x'_i; X')h) p(x'_i, dh) \Bigg| \\
\leq \frac{1}{2} \Bigg(& \int_{\mathbb{R}^d} \|h\|^3 \|H_1 g_i(x_i; X) - H_1 g_i(x'_i; X')\| p(x'_i, dh) \\
& + \left| \int_S h_k(h^* H_1 g_i(x_i; X)h)(p(x_i, dh) - p(x'_i, dh)) \right| \\
& + \int_{S\cap S'^c} \|h\|^3 \|H_1 g_i(x_i; X)\| p(x'_i, dh) \\
& + \int_{S^c\cap S'} \|h\|^3 \|H_1 g_i(x'_i; X')\| p(x'_i, dh) \Bigg).
\end{aligned} \tag{63}$$

The first term goes to 0 when $X' \to X$ since $H_1 g_i$ is continuous and $p(x_i, \cdot)$ has finite third order moments; the second term goes to 0 by Corollary 1 as above; using the fact that $H_1 g_i$ is bounded, and hypothesis (H2), the third term is bounded by

$$\int_{S\cap S'^c} K\|h\|^3 p(h) dh \tag{64}$$

and the forth term by

$$\int_{S^c\cap S'} K\|h\|^3 p(h) dh \tag{65}$$

for some constant K. Now, when $X' \to X$, the sets $S \cap S'^c$ and $S^c \cap S'$ converge respectively to $\emptyset$ and $\{h; h \cdot \nabla_1 g_i(x_i; X) = 0\}$, so the dominated convergence Theorem gives the required result, since $\{h : h \cdot \nabla_1 g_i(x_i; X) = 0\}$ has Lebesgue measure 0 (because $X \notin \Gamma$), and since, in (H2), $p(x, \cdot)$ has been supposed absolutely continuous with respect to the Lebesgue measure. □

Note that in general, $\tilde{b}$ is not continuous at points of Γ. For example, if $\mathcal{X} = \mathbb{R}$, $n = 1$ and $p(x, \cdot)$ is symmetrical for all $x \in \mathcal{X}$, then

$$\tilde{b}(x) = \frac{1}{2} \mathrm{sign}[\partial_1 g(x; x)] \partial_1^2 g(x; x) \int_{\mathbb{R}} h^3 p(x, dh), \tag{66}$$

where $\mathrm{sign}(x) = -1$ if $x < 0$, $= 0$ if $x = 0$ and $= 1$ if $x > 0$.

Let us show the weak existence of solutions to the SDE (59):

Theorem 2 *Assume (H1), (H2) and (H3). Assume also that $p(x, \cdot)$ is a symmetrical measure on $\mathbb{R}^d$ for all $x \in \mathcal{X}$, and define for any $p \in \mathbb{N}^*$ and $X \in \mathbb{R}^{nd}$*

$$a_p(X) = a(X) + \frac{I}{p}, \tag{67}$$

where I is the $nd \times nd$ identity matrix. Let $\sigma_p(X)$ be the only $nd \times nd$ symmetrical positive definite matrix satisfying $\sigma_p \sigma_p^* = a_p$.

Then there is weak existence of a solution to the stochastic differential equation

$$dX_t^{p,\varepsilon} = b^\varepsilon(X_t^{p,\varepsilon})dt + \sqrt{\varepsilon}\sigma_p(X_t^{p,\varepsilon})dW_t. \tag{68}$$

Given a random variable X_0 in $\mathbb{R}^{nd}$, let us denote by $\mathbf{P}_{X_0}^{p,\varepsilon}$ the law of a solution to (68) with initial state X_0.

Then, for any $X_0 \in \mathbb{L}^1$, the family of probability measures $\{\mathbf{P}_{X_0}^{p,\varepsilon}\}$ is tight on $\mathcal{C}([0,T], \mathbb{R}^{nd})$ for any $T > 0$, and any of its accumulation points on $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^{nd})$ is solution to the martingale problem for $\tilde{L}^\varepsilon$. Hence there is weak existence for the SDE (59).

4.3 Proof of Theorem 2

Existence of a weak solution to (68) First, note that a_p has been defined such that $X \mapsto a_p(X)$ is uniformly non degenerate on $\mathbb{R}^{nd}$.

Let us recall how to obtain the unique $nd \times nd$ symmetrical definite positive matrix σ_p such that $\sigma_p \sigma_p^* = a_p$: find an orthonormal basis of $\mathbb{R}^{nd}$ where a_p is diagonal, put the square root of its elements in a new diagonal matrix, and express it back in the first basis. By Proposition 5, the extension of a to $\mathbb{R}^{nd}$ is continuous. So a_p is continuous and uniformly non-degenerate on $\mathbb{R}^{nd}$, and it is classical to show that the same holds for σ_p, as defined above.

Since by Proposition 5, b, $\tilde{b}$ and a_p are bounded on $\mathbb{R}^{nd}$, and since a_p is uniformly non-degenerate, the existence of a weak solution to (68) is a consequence of Problem 3.13 page 305 in Karatzas and Shreve [11], if we can prove the existence of a weak solution to

$$dX_t = \sigma_p(X_t)dW_t, \tag{69}$$

but this is a well-known result since σ_p is bounded and continuous (see Karatzas and Shreve [11] page 323).

Tightness of $\{\mathbf{P}_{X_0}^{p,\varepsilon}\}_{q\geq 1}$ on $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^{nd})$ as $p \to +\infty$. A weak solution $X^{p,\varepsilon}$ of (68) with law $\mathbf{P}_{X_0}^{p,\varepsilon}$, satisfies

$$X_t^{p,\varepsilon} = X_0 + \int_0^t b^\varepsilon(X_s^{p,\varepsilon})ds + \sqrt{\varepsilon}\int_0^t \sigma_p(X_s^{p,\varepsilon})dW_s \tag{70}$$

for some nd-dimensional standard brownian motion W

Since, by Proposition 5, b^ε and σ_p are bounded for $\varepsilon < 1$ and $p \geq 1$, and since X_0 is $\mathbb{L}^1$, the process $X^{p,\varepsilon}$ is $\mathbb{L}^1$, and the Burkholder-Davis-Gundy inequality shows that for any $t \geq 0$

and any $\delta \in]0,1[$

$$
\begin{aligned}
\mathbf{E}\left[\sup_{0\le\theta\le\delta}\|X^{p,\varepsilon}_{t+\theta}-X^{p,\varepsilon}_t\|\right] &\le \mathbf{E}\left[\sup_{0\le\theta\le\delta}\left\|\int_t^{t+\theta} b^\varepsilon(X^{p,\varepsilon}_s)ds\right\|\right] \\
&\quad + K\sum_{i=1}^{nd}\mathbf{E}\left[\sup_{0\le\theta\le\delta}\left|\int_t^{t+\theta}\sum_{j=1}^{nd}(\sigma_p)_{i,j}(X^{p,\varepsilon}_s)dW^j_s\right|\right] \\
&\le \mathbf{E}\left[\int_t^{t+\delta}\|b^\varepsilon(X^{p,\varepsilon}_s)\|ds\right] \qquad (71)\\
&\quad + K\sum_{i=1}^{nd}\mathbf{E}\left[\left(\int_t^{t+\delta}\left(\sum_{j=1}^{nd}(\sigma_p)^2_{i,j}(X^{p,\varepsilon}_s)\right)ds\right)^{1/2}\right] \\
&\le C\delta + Knd\times(nd)^{1/2}C\delta^{1/2}\le K\sqrt{\delta}
\end{aligned}
$$

where the constants K can change from line to line and where C is a constant, given by Proposition 5, bounding b^ε for $\varepsilon < 1$ and σ_q for $q \ge 1$.

By Tchebychef inequality, we obtain for any $t \ge 0$ and $\eta > 0$

$$
\mathbf{P}\left(\sup_{0\le\theta\le\delta}\|X^{p,\varepsilon}_{t+\theta}-X^{p,\varepsilon}_t\|>\eta\right)\le\frac{K\sqrt{\delta}}{\eta}, \qquad (72)
$$

which shows the tightness of the sequence $\{\mathbf{P}^{p,\varepsilon}_{X_0}\}_{p\ge1}$ on $\mathcal{C}([0,T],\mathbb{R}^{nd})$ for any $T \ge 0$ (Billingsley [1]). Let us denote by $\mathbf{P}^\varepsilon_{X_0}$ an accumulation point of this sequence on $\mathcal{C}(\mathbb{R}_+,\mathbb{R}^{nd})$.

Martingale problem for $\mathbf{P}^\varepsilon_{X_0}$ It remains to prove that $\mathbf{P}^\varepsilon_{X_0}$ solves the martingale problem associated with the operator $\tilde{L}^\varepsilon$ defined in (58). So let us denote by $(\mathcal{C}(\mathbb{R}_+,\mathbb{R}^{nd}),\mathcal{G}_t)$ the space of continuous functions from $\mathbb{R}_+$ to $\mathbb{R}^{nd}$ endowed with the canonical filtration, and let us define on this space, for $\varphi \in \mathcal{C}^2(\mathbb{R}^{nd})$, the process

$$
M^\varphi_t(w)=\varphi(w_t)-\varphi(w_0)-\int_0^t\tilde{L}^\varepsilon\varphi(w_s)ds. \qquad (73)
$$

We have to show that, under $\mathbf{P}^\varepsilon_{X_0}$, M^φ is a local $\mathcal{G}_t$-martingale. We already know that under $\mathbf{P}^{p,\varepsilon}_{X_0}$, $M^{p,\varphi}_t(w)=\varphi(w_t)-\varphi(w_0)-\int_0^t\tilde{L}^{p,\varepsilon}\varphi(w_s)ds$ is a local $\mathcal{G}_t$-martingale, where $\tilde{L}^{p,\varepsilon}$ is the generator of $X^{p,\varepsilon}$ defined for $\varphi\in\mathcal{C}^2(\mathbb{R}^{nd})$ and $X\in\mathbb{R}^{nd}$ by

$$
\tilde{L}^{p,\varepsilon}\varphi(X)=b^\varepsilon(X)\nabla\varphi(X)+\frac{\varepsilon}{2}\sum_{i=1}^n\sum_{1\le k,l\le d}(a_p)^i_{kl}(X)\frac{\partial^2_i\varphi}{\partial_ix_k\partial_ix_l}(X). \qquad (74)
$$

Note that this formula can be obtained from the expression (58) for $\tilde{L}^\varepsilon$ by replacing a by $a_p = a + I/p$.

Since, by Proposition 5, a and b^ε are bounded, $M^{p,\varphi}$ is a square integrable martingale under $\mathbf{P}^{p,\varepsilon}_{X_0}$ as soon as $\varphi\in\mathcal{C}^2_c(\mathbb{R}^{nd})$.

The major difference from the proof of Theorem 1 is that, by Proposition 5, $\tilde{b}^i$ may not be continuous at points of Γ_i. We will use an approximation technique, involving some $\mathcal{C}^\infty$ function χ from $\mathbb{R}^{nd}$ to $\mathbb{R}^{nd}$, such that $\chi(X) = X$ for $\|X\| \ge 1$, and such that $\chi(X) = 0$ for $\|X\| \le 1/2$.

Given a function $\varphi \in \mathcal{C}_c^2(\mathbb{R}^{nd})$ and some point $Y \in \mathbb{R}^{nd}$, define for any integer $r \geq 1$ the function

$$\varphi_{Y,r}(X) = \varphi\left(Y + \frac{1}{r}\chi(r(X-Y))\right). \tag{75}$$

Let us list in the following lemma some simple properties of this sequence of function. Its proof is straightforward.

Lemma 1 *If $\|X-Y\| \geq 1/r$, $\varphi_{Y,r}(X) = \varphi(X)$, and if $\|X-Y\| \leq 1/2r$, $\varphi_{Y,r}(X) = \varphi(Y)$. $\varphi_{Y,r}$ uniformly converges to φ on $\mathbb{R}^{nd}$. Moreover, $\varphi_{Y,r} \in \mathcal{C}_c^2(\mathbb{R}^{nd})$, and $\nabla\varphi_{Y,r}$ converges to $\nabla\varphi\mathbf{1}_{\mathbb{R}^{nd}\setminus\{Y\}}$ for the bounded pointwise convergence.*

Now, fix some $\varphi \in \mathcal{C}_c^2(\mathbb{R}^{nd})$, and an extracted sequence $\{\mathbf{P}_{X_0}^{p_k,\varepsilon}\}$ of $\{\mathbf{P}_{X_0}^{p,\varepsilon}\}$ converging to $\mathbf{P}_{X_0}^{\varepsilon}$ for the Skorohod topology on the set $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^{nd})$ of continuous functions from $\mathbb{R}_+$ to $\mathbb{R}^{nd}$. Fix $s > 0$, $t > s$, m real numbers $0 \leq t_1 < \ldots < t_l \leq m$ for some $m \geq 1$, and a continuous bounded function $q : (\mathbb{R}^{nd})^m \to \mathbb{R}$. Remembering that a and a_p only differ on their diagonal elements, we can write

$$\begin{aligned}
&\left|\mathbf{E}_{\mathbf{P}_{X_0}^{\varepsilon}}\left\{q\left(w_{t_1},\ldots,w_{t_m}\right)\left[\varphi(w_t)-\varphi(w_s)-\int_s^t \tilde{L}^{\varepsilon}\varphi(w_u)du\right]\right\}\right| \\
&\quad\leq \left|\mathbf{E}_{\mathbf{P}_{X_0}^{p_k,\varepsilon}}\left\{q\left(w_{t_1},\ldots,w_{t_m}\right)\left[\varphi(w_t)-\varphi(w_s)-\int_s^t \tilde{L}^{p_k,\varepsilon}\varphi(w_u)du\right]\right\}\right| \\
&\quad+\left|\mathbf{E}_{\mathbf{P}_{X_0}^{p_k,\varepsilon}}\left\{q\left(w_{t_1},\ldots,w_{t_m}\right)\left[\varepsilon\sum_{i=1}^n\sum_{j=1}^d\int_s^t \frac{(a_{p_k})^i_{jj}(w_u)-a^i_{jj}(w_u)}{2}\frac{\partial_i^2\varphi}{\partial_i x_j^2}(w_u)du\right]\right\}\right| \\
&\quad+\left|\mathbf{E}_{\mathbf{P}_{X_0}^{\varepsilon}}\left\{q\left(w_{t_1},\ldots,w_{t_m}\right)\left[\varphi(w_t)-\varphi(w_s)-\int_s^t \tilde{L}^{\varepsilon}\varphi(w_u)du\right]\right\}\right. \\
&\qquad\left.-\mathbf{E}_{\mathbf{P}_{X_0}^{p_k,\varepsilon}}\left\{q\left(w_{t_1},\ldots,w_{t_m}\right)\left[\varphi(w_t)-\varphi(w_s)-\int_s^t \tilde{L}^{\varepsilon}\varphi(w_u)du\right]\right\}\right|.
\end{aligned} \tag{76}$$

The first term of the right-hand side is 0 since $M^{p_k,\varphi}$ is a $\mathbf{P}_{X_0}^{p_k,\varepsilon}$-martingale. Since $a_p = a + I/p$, the second term is bounded by $\|q\|_\infty\|H\varphi\|_\infty\varepsilon t/p_k$, where $H\varphi$ is the Hessian matrix of φ, so it converges to 0 when $k \to +\infty$. Using equation (58) for $\tilde{L}^{\varepsilon}$, we can bound above the third term by

$$\begin{aligned}
&\left|\mathbf{E}_{\mathbf{P}_{X_0}^{\varepsilon}}\left\{q(w_{t_1},\ldots,w_{t_m})\left[\varphi(w_t)-\varphi(w_s)\right.\right.\right. \\
&\qquad\left.\left.-\int_s^t b(w_u)\cdot\nabla\varphi(w_u)du-\frac{\varepsilon}{2}\sum_{i=1}^n\sum_{1\leq i,j\leq d}\int_s^t a^i_{k,l}(w_u)\frac{\partial_i^2\varphi}{\partial_i x_k\partial_i x_l}(w_u)du\right]\right\} \\
&\quad-\mathbf{E}_{\mathbf{P}_{X_0}^{p_k,\varepsilon}}\left\{q(w_{t_1},\ldots,w_{t_m})\left[\varphi(w_t)-\varphi(w_s)\right.\right. \\
&\qquad\left.\left.\left.-\int_s^t b(w_u)\cdot\nabla\varphi(w_u)du-\frac{\varepsilon}{2}\sum_{i=1}^n\sum_{1\leq i,j\leq d}\int_s^t a^i_{k,l}(w_u)\frac{\partial_i^2\varphi}{\partial_i x_k\partial_i x_l}(w_u)du\right]\right\}\right| \\
&\quad+\varepsilon\sum_{i=1}^n\left|\mathbf{E}_{\mathbf{P}_{X_0}^{\varepsilon}}\left\{q(w_{t_1},\ldots,w_{t_m})\int_s^t \tilde{b}^i(w_u)\cdot\nabla_i\varphi(w_u)du\right\}\right. \\
&\qquad\left.-\mathbf{E}_{\mathbf{P}_{X_0}^{p_k,\varepsilon}}\left\{q(w_{t_1},\ldots,w_{t_m})\int_s^t \tilde{b}^i(w_u)\cdot\nabla_i\varphi(w_u)du\right\}\right|.
\end{aligned} \tag{77}$$

By Proposition 5, a and b are continuous and bounded on $\mathbb{R}^{nd}$, so the first term converges to 0 when $k \to +\infty$, since $\mathbf{P}^{p_k,\varepsilon}_{X_0} \Rightarrow \mathbf{P}^{\varepsilon}_{X_0}$ and since the quantity inside the expectation is continuous and bounded for the weak topology on $\mathcal{C}([0,t],\mathbb{R}^{nd})$.

In order to bound the i^{th} term of the last sum, we will assume that Γ_i is made of a single point $\{Y_i\}$, to keep notations simple. Because of the assumption (H3), all the points of Γ_i are isolated, and one can easily convince himself that the following approximation technique can be extended to the general case. Let us write for any $r \geq 1$

$$\begin{aligned}
&\Bigg|\mathbf{E}_{\mathbf{P}^{\varepsilon}_{X_0}}\left\{q(w_{t_1},\dots,w_{t_m})\int_s^t \tilde{b}^i(w_u)\cdot\nabla_i\varphi(w_u)du\right\} \\
&\qquad\qquad - \mathbf{E}_{\mathbf{P}^{p_k,\varepsilon}_{X_0}}\left\{q(w_{t_1},\dots,w_{t_m})\int_s^t \tilde{b}^i(w_u)\cdot\nabla_i\varphi(w_u)du\right\}\Bigg| \\
&\leq \left|\mathbf{E}_{\mathbf{P}^{\varepsilon}_{X_0}}\left\{q(w_{t_1},\dots,w_{t_m})\int_s^t \tilde{b}^i(w_u)\cdot[\nabla_i\varphi(w_u)-\nabla_i\varphi_{Y_i,r}(w_u)]du\right\}\right| \\
&+ \left|\mathbf{E}_{\mathbf{P}^{p_k,\varepsilon}_{X_0}}\left\{q(w_{t_1},\dots,w_{t_m})\int_s^t \tilde{b}^i(w_u)\cdot[\nabla_i\varphi(w_u)-\nabla_i\varphi_{Y_i,r}(w_u)]du\right\}\right| \\
&+ \Bigg|\mathbf{E}_{\mathbf{P}^{\varepsilon}_{X_0}}\left\{q(w_{t_1},\dots,w_{t_m})\int_s^t \tilde{b}^i(w_u)\cdot\nabla_i\varphi_{Y_i,r}(w_u)du\right\} \\
&\qquad - \mathbf{E}_{\mathbf{P}^{p_k,\varepsilon}_{X_0}}\left\{q(w_{t_1},\dots,w_{t_m})\int_s^t \tilde{b}^i(w_u)\cdot\nabla_i\varphi_{Y_i,r}(w_u)du\right\}\Bigg|.
\end{aligned} \tag{78}$$

By Lemma 1, $\nabla\varphi_{Y_i,r}(X) = 0$ for X such that $\|X - Y_i\| \leq 1/2r$, so $X \mapsto \tilde{b}^i(X)\nabla_i\varphi(X)$ is continuous on $\mathbb{R}^{nd}$ (remind the $\tilde{b}^i$ may not be continuous only in $\Gamma_i = \{Y_i\}$). Since $\mathbf{P}^{p_k,\varepsilon}_{X_0} \Rightarrow \mathbf{P}^{\varepsilon}_{X_0}$, the third term converges to 0 when $k \to +\infty$. By Lemma 1 again, $\tilde{b}^i\nabla_i\varphi_{Y_i,r} \to \tilde{b}^i\nabla_i\varphi\mathbf{1}_{\mathbb{R}^{nd}\setminus\{Y_i\}}$ for the bounded pointwise convergence. But, since we have supposed that $p(x,\cdot)$ is symmetrical for all x, $\tilde{b}^i(Y_i) = 0$, so $\tilde{b}^i\nabla_i\varphi\mathbf{1}_{\mathbb{R}^{nd}\setminus\{Y_i\}} = \tilde{b}^i\nabla_i\varphi$, and by the dominated convergence Theorem, the two first terms of (78) converge to 0 when $r \to +\infty$, for any fixed $k \in \mathbb{N}$.

Finally, choosing first k and then r, we can bound above

$$\left|\mathbf{E}_{\mathbf{P}^{\varepsilon}_{X_0}}\left\{q(w_{t_1},\dots,w_{t_m})\left[\varphi(w_t)-\varphi(w_s)-\int_s^t \tilde{L}^{\varepsilon}\varphi(w_u)du\right]\right\}\right| \tag{79}$$

by an arbitrarily small number, so this quantity is 0.

This equality can be extended to any $\mathcal{G}_s$-measurable bounded function q by the monotone class Theorem, so that M^{φ} is a $\mathbf{P}^{\varepsilon}_{X_0}$-martingale.

Finally, using the same truncation technique as in the proof of Theorem 1, we can extend this result to any function $\varphi \in \mathcal{C}^2(\mathbb{R}^{nd})$. To show that the stopping times T_i defined in equation (44) $\mathbf{P}^{\varepsilon}_{X_0}$-a.s. converge to $+\infty$ when $i \to +\infty$, we can use inequality (71) for $t = 0$ and $\delta = T$: we get

$$\mathbf{E}\left[\sup_{0\leq t\leq T}\|X^{p,\varepsilon}_t - X_0\|\right] \leq C\sqrt{T}, \tag{80}$$

so that, since X_0 is $\mathbb{L}^1$, there is some constant K_T depending on T but not on p such that $\mathbf{E}[\sup_{0\leq t\leq T}\|X^{p,\varepsilon}_t\|] = \mathbf{E}_{\mathbf{P}^{p,\varepsilon}_{X_0}}[\sup_{0\leq t\leq T}\|w_t\|] \leq K_T$. Exactly as in section 3.3, we deduce that M^{φ} is a local martingale. □

References

[1] P. Billingsley, *Convergence of Probability Measures*, John Wiley & Sons (1968).

[2] N. Champagnat, R. Ferrière and G Ben Arous, *The canonical equation of adaptive dynamics: a mathematical view*, Selection **2**, 73-83 (2001).

[3] O. Diekmann, *A beginners guide to adaptive dynamics*, in A. Margheri, C. Rebelo and F. Zondin (ed.), Summer School on Mathematical Biology, CIM, Lisboa, Portugal.

[4] U. Dieckmann and R. Law, *The dynamical theory of coevolution: a derivation from stochastic ecological processes*, Journal of Mathematical Biology **34**, 579-612 (1996).

[5] S. N. Ethier and T. G. Kurtz, *Markov Processes, Characterization And Convergence*, Ed. John Wiley & Sons, New York (1986).

[6] N. Fournier and S. Méléard, *A microscopic probabilistic description of a locally regulated population and macroscopic approximations*, preprint Modal'X 03/4, University Paris-10 (2003).

[7] S. A. H. Geritz, M. Gyllenberg, F. J. A. Jacobs and K. Parvinen, *Invasion dynamics and attractor inheritance*, Journal of Mathematical Biology **44**, 548-560 (2002).

[8] J. Hofbauer and R. Sigmund, *Adaptive dynamics and evolutionary stability.* Applied Mathematics Letters **3**, 75-79 (1990).

[9] J. Jacod and A. N. Shiryaev, *Limit Theorems for Stochastic Processes.* Springer-Verlag, Berlin (1987).

[10] A. Joffe and M. Métivier, *Weak convergence of sequences of semimartingales with applications to multitype branching processes*, Advances in Applied Probability, **18**, 20-65 (1986).

[11] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus.* Springer-Verlag, New York (1988).

[12] P. Marrow, R. Law and C. Cannings, *The coevolution of predator-prey interactions: ESSs and Red Queen dynamics* Proc. R. Soc. Lond. B **250**, 133-141 (1992).

[13] J.A.J Metz, R.M. Nisbet and S.A.H. Geritz, *How should we define 'fitness' for general ecological scenarios?.* Trends in Ecology and Evolution **7**, 198-202 (1992).

[14] J.A.J. Metz, S.A.H. Geritz, G. Meszéna, F.A.J. Jacobs and J.S. van Heerwaarden, *Adaptive Dynamics, a geometrical study of the consequences of nearly faithful reproduction*, in S.J. van Strien and S.M. Verduyn Lunel (ed.), *Stochastic and Spatial Structures of Dynamical Systems*, North Holland, Amsterdam, 183-231 (1996).

[15] S. T. Rachev, *Probability Metrics and the Stability of Stochastic Models*, John Wiley & Sons (1991).

[16] D.A. Rand and H.B. Wilson, *Evolutionary catastrophes, punctuated equilibria and gradualism in ecosystem evolution.* Proc. Roy. Soc. London B **253**, 137-141 (1993).