Small jumps asymptotic of the moving optimum Poissonian SDE

Elma Nassar* and Etienne Pardoux**

February 12, 2018

* Aix Marseille Univ, CNRS, Centrale Marseille, I2M, Marseille, France
elma_nassar@hotmail.com
** Aix Marseille Univ, CNRS, Centrale Marseille, I2M, Marseille, France
etienne.pardoux@univ-amu.fr, corresponding author.

Abstract

We consider a Poissonian SDE for the lack of fitness of a population subject to a continuous change of its environment, and an accumulation of advantageous mutations. We neglect the time of fixation of new mutations, so that the population is monomorphic at all times. We consider the asymptotic of small and frequent mutations. In that limit, we establish a law of large numbers and a central limit theorem. For small enough mutations, the original process is Harris recurrent and ergodic. We show in which sense the limits as $t \to \infty$ of the law of large and number and central limit theorem give a good approximation of the invariant probability measure of the original process.

Keys words: Poissonian SDE, Law of large numbers, Central limit theorem, Approximation of invariant measure, canonical equation of adaptive dynamics, moving optimum model.

Introduction

The present work is motivated by the moving optimum model in theoretical biology, which aims at evaluating the possibility for mutations to rescue a given population undergoing a linear change in its environment which deteriorates its survival conditions. We refer the reader to [8] and [11] for the presentation of this model. The authors have set in [13] a rigorous mathematical study of the moving optimum model by introducing a stochastic differential equation driven by a Poisson point process describing the evolution of a quantitative one–dimensional phenotypic trait in accordance with the biological description of this evolutionary rescue model. They studied the large time behavior of its solution, which is Harris recurrent when the speed of the environment v is smaller than the mean effect of the beneficial mutations m per unit time, transient if $v > m$. In the case of equality between the two parameters, the solution of the stochastic differential equation can either be transient or Harris recurrent depending upon additional technical conditions.

One is mainly interested in the positive recurrent case. However, the limitation of the ability to draw biological conclusions from these result is due to the difficulty to compute explicitly any quantity related to the invariant probability measure. This led us to study the small jumps limit, which is obtained by multiplying the jumps' sizes by ε, dividing the rates by ε^2, and then letting $\varepsilon \to 0$. In this paper, we study the limit as $\varepsilon \to 0$ in the such rescaled multidimensional version of the SDE from [13]. More precisely, we consider the SDE in $\mathbb{R}^d$

$$X_t^\varepsilon = X_0^\varepsilon - vt + \int_0^t \int_{\mathbb{R}^d} \int_0^1 \varepsilon\alpha\Gamma(X_{s-}^\varepsilon, \varepsilon\alpha, \xi) N_\varepsilon(ds, d\alpha, d\xi),$$

where N_ε is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}^d \times [0,1]$, with the intensity

$$\left(\frac{1}{\varepsilon^2} ds\right) \times \nu(d\alpha) \times d\xi.$$

Our first main result is

Theorem 1. *If $X_0^\varepsilon \to x_0$ in probability as $\varepsilon \to 0$, then $X_t^\varepsilon \underset{\varepsilon\to 0}{\longrightarrow} \bar{x}_t$ in probability, locally uniformly in t, where, with a given $d \times d$ matrix L,*

$$\frac{d\bar{x}_t}{dt} = -v + L\bar{x}_t, \quad \bar{x}_0 = x_0.$$

The above limiting ODE can be interpreted as the canonical equation of adaptive dynamics in the context of a changing environment, see [3] and [2]. This is discussed in greater detail in [9]. The next step is to establish a Central Limit Theorem. Indeed, we define $U_t^\varepsilon = \varepsilon^{-1/2}(X_t^\varepsilon - \bar{x}_t)$, and show

Theorem 2. *Assume that $X_0^\varepsilon = \bar{x}_0$. Then $U^\varepsilon \Rightarrow U$, where U is an Ornstein-Uhlenbeck process:*

$$\begin{aligned} dU_t &= LU_t dt + \Lambda^{\frac{1}{2}}(\bar{x}_t) dB_t, \\ U_0 &= 0, \end{aligned}$$

with B being a d-dimensional standard Brownian motion. In other words,

$$U_t = \int_0^t e^{L(t-s)} \Lambda^{\frac{1}{2}}(\bar{x}_s) dB_s.$$

Moreover it is not hard to show that as $t \to \infty$, both $\bar{x}_t \to \bar{x}_\infty$ and the law of U_t converges to a Gaussian law $\mathcal{N}(0, \bar{S}^2)$.

Both $\bar{x}_\infty$ and $\bar{S}^2$ can be easily computed with high accuracy, given the parameters of the model. On the other hand, we show that there exists $\varepsilon_0 > 0$ such that, for any $0 < \varepsilon \leq \varepsilon_0$, the process X_t^ε is Harris recurrent and possesses a unique invariant probability measure μ^ε. Since our motivation for studying the small jumps limit is to get informations about μ^ε, it is desirable to show that the pair $(\bar{x}_\infty, \bar{S}^2)$ gives a precise approximation of the invariant measure μ^ε, for small ε. This is a delicate question, since it amounts in a sense to invert the two limits $\varepsilon \to 0$ and $t \to \infty$.

We first show that the collection of probability measures $\{\mu^\varepsilon,\ \varepsilon \leq \varepsilon_0\}$ is tight. It is then not too difficult to deduce that $\mu^\varepsilon \Rightarrow \delta_{\bar{x}_\infty}$, as $\varepsilon \to 0$. We want to prove more, namely that μ^ε is close to ν^ε, which is the law of $\bar{x}_\infty + \sqrt{\varepsilon}\xi$, if $\xi \simeq \mathcal{N}(0, \bar{S}^2)$. This is done by analysing the probability measure

$$\mu_t^\varepsilon(A) = \frac{1}{t} \int_0^t \mathbf{1}_A(X_s^\varepsilon) ds$$

for large t and small ε.

We believe that those results are original, and have an interest, not only for the specific model which we study, but could also be useful in different frameworks, where a process converges in law to a limiting process, and one wants to compare large time behaviors. In our case, numerical simulations tend to indicate that the approximation is valid even for not very small values of ε, see [9].

The proofs of the first two theorems follow the following scheme : prove tightness, and the identify the unique possible limit point. Harris recurrence for small enough ε is proved using a criterion due to Meyn and Tweedie [12].

The paper is organized as follows: After giving some useful notations, section 1 gives a detailed presentation of the model, from the biological literature. Furthermore we present the stochastic differential equation that describes the evolution of the vector phenotypic lag between the population and its environment, explaining the fixation mechanism of mutations.

In section 2, we prove Theorem 1 and Theorem 2.

Section 3 is dedicated to the study of the large time behavior of X_t^ε which will turn out to be positive Harris recurrent for ε sufficiently small, admitting thus a unique invariant probability measure. Then we proceed to prove that the sequence of invariant measures is tight and converges in law. We finally give a precise statement which describes in which sense the invariant measure is well approximated by a combination of the limit as $t \to \infty$ of the LLN and the CLT limits. This involves a sort of interchange of limits as $\varepsilon \to 0$ and as $t \to \infty$, which seems to be new.

Notation

We will deal with processes X_t with "càdlàg" paths, i.e. paths which are right–continuous and have a left limit everywhere. This left limit is denoted $X_{t-} := \lim_{s\uparrow t, s<t} X_s$, and for any t such that $X_t \neq X_{t-}$, we denote by $\Delta X_t = X_t - X_{t-}$ the jump of X at time t. Note that a càdlàg path has at most countably many jumps on $\mathbb{R}_+$.

We remind that the quadratic variation of a scalar discontinuous bounded variation locally square integrable martingale M_t is the sum of the squares of its jumps and is denoted by :

$$[M]_t = \sum_{s\leq t} |\Delta M_s|^2.$$

Its predictable quadratic variation $\langle M\rangle_t$ is the unique increasing predictable process such that $[M]_t - \langle M\rangle_t$, and hence also $M_t^2 - \langle M\rangle_t$ is a local martingale.

In the d-dimensional case, we define the quadratic variation of a discontinuous bounded variation locally square integrable martingale M_t as :

$$[[M]]_t = \sum_{s\leq t} \Delta M_s \otimes \Delta M_s.$$

Its predictable quadratic variation $\langle\langle M\rangle\rangle_t$ is the unique S^d-valued predictable increasing process such that both $[[M]]_t - \langle\langle M\rangle\rangle_t$ and $M_t \otimes M_t - \langle\langle M\rangle\rangle_t$ are S^d-valued local martingales. Here S^d denotes the set of symmetric positive semi–definite $d \times d$ matrices. Note that $[[M]]_t$ (resp. $\langle\langle M\rangle\rangle_t$) is the matrix whose i, j element is $[M^i, M^j]_t$ (resp. $\langle M^i, M^j\rangle_t$).

We shall use the notation

$$[M]_t = \sum_{0<s\leq t} \|\Delta M_s\|^2 = \mathrm{Tr}[[M]]_t, \quad \text{and } \langle M\rangle_t = \mathrm{Tr}\langle\langle M\rangle\rangle_t,$$

so that $\|M_t\|^2 - \langle M\rangle_t$ is an $\mathbb{R}$-valued local martingale, and the notations in the scalar and vector case are coherent. See [17] for more details.

Suppose now that X_t is a d–dimensional process which takes the form

$$X_t = X_0 + \int_0^t A_s ds + M_t,$$

where A_t is d–dimensional and M_t is a d-dimensional martingale with paths of bounded variation (hence discontinuous). Then, if $\|\cdot\|$ and $(\cdot,\cdot)$ denote the Euclidean norm and scalar product in $\mathbb{R}^d$, we have the following

$$\|X_t\|^2 = \|X_0\|^2 + 2\int_0^t (X_s, A_s)ds + 2\int_0^t (X_{s-}, dM_s) + [M]_t. \tag{1}$$

This is a change of variable formula which is the content (in the case where X_t is scalar) of Theorem 31 in [17]. Its proof is quite elementary and does not require any stochastic calculus.

Finally in this paper we will repeatedly write integrals of vector–valued (or matrix–valued) functions of their argument, as we already did in the definition of X_t above, where the vector $\int_0^t A_s ds$ was such that its i–th coordinates equals $\int_0^t A_s^i ds$ if A_s^i stands for the i–th coordinate of A_s. Below in (7), $\int_{(x,\alpha)\leq 0} \alpha\otimes\alpha\ \nu(d\alpha)$ (where $\{\alpha,\ (x,\alpha)\leq 0\}$ is a subset of $\mathbb{R}^d$) stands for the $d\times d$ matrix whose (i,j) entry is $\int_{(x,\alpha)\leq 0} \alpha_i\alpha_j\ \nu(d\alpha)$, and $\int_{(x,\alpha)\leq 0} \alpha\ \nu(d\alpha)$ stands for the d–dimensional vector whose i–th coordinate is $\int_{(x,\alpha)\leq 0} \alpha_i\ \nu(d\alpha)$.

1 The model

The model from Matuszewski *et al.* [11] is set up as follows: a population of constant size N is subject to Gaussian stabilizing selection with a moving optimum that increases linearly with speed vector $v \in \mathbb{R}^d$. That is, at time t, the phenotypic lag between an individual with trait value $z \in \mathbb{R}^d$ and the optimum equals $x = z - vt \in \mathbb{R}^d$, and the corresponding fitness is

$$\mathcal{W}(x) = \exp\left(-x'\Sigma^{-1}x\right), \tag{2}$$

where Σ describes the shape of the fitness landscape. For the adaptive-walk approximation, the population is assumed to be monomorphic at all times (i.e., its state is completely characterized by x). Mutations arise at rate $\Theta/2 = N\mu$ (where μ is the *per-capita* mutation rate and $\Theta = 2N\mu$ is a standard population-genetic parameter), and their phenotypic effects α are drawn from a distribution $p(\alpha)$. We neglect the possibility of fixation for deleterious mutations. Yet even beneficial mutations have a significant probability of being lost due to the effects of genetic drift while they are rare. A mutation with effect α that arises in a population with phenotypic lag x has a probability of fixation

$$g(x,\alpha) = \begin{cases} 1 - \exp(-2s(x,\alpha)) & \text{if } s(x,\alpha) > 0, \\ 0 \text{ otherwise} \end{cases} \tag{3}$$

where

$$s(x,\alpha) = -(2x+\alpha)'\Sigma^{-1}\alpha \tag{4}$$

is the selection coefficient. Once a mutation gets fixed, it is assumed to do so instantaneously (which is of course a simplification which is not realistic), and the phenotypic lag x of the population is updated accordingly.

We make the following assumptions (see [11]):

1. v is a horizontal vector, $v = (v_1, 0, \dots, 0)'$ with $v_1 > 0$.
2. Σ is isotropic, i.e. $\Sigma = \sigma^2 \mathbf{I}_{\mathbb{R}^d}$.

It is always possible to reduce the situation to our assumptions, via a change of variables.

The evolution of the phenotypic lag of the population can be described by the following equation:

$$X_t = x_0 - vt + \int_{[0,t]\times\mathbb{R}^d\times[0,1]} \alpha\Gamma(X_{s^-},\alpha,\xi)N(ds,d\alpha,d\xi). \tag{5}$$

Here, N is a Poisson point process over $\mathbb{R}_+ \times \mathbb{R}^d \times [0,1]$ with intensity $ds\ \nu(d\alpha)\ d\xi$ where $\nu(d\alpha)$ is the measure of new mutations and

$$\Gamma(x,\alpha,\xi) = \mathbf{1}_{\{\xi \le g(x,\alpha)\}},$$

where the fixation probability $g(x,\alpha)$ of a mutation of size α that hits the population when the lag is x, as defined by (3) and (4), can be expressed as

$$g(x,\alpha) = \left(1 - \exp\left(2\sigma^{-2}\left(2x+\alpha,\alpha\right)\right)\right) \times \mathbf{1}_{\{(2x+\alpha,\alpha)\le 0\}}.$$

Following the model by [11], we consider that

$$\nu(d\alpha) = \frac{\Theta}{2} p(\alpha) d\alpha, \tag{6}$$

where p is the density of a centered multidimensional Gaussian distribution $\mathcal{N}(0,M)$, M being a positive definite symmetric matrix. Under the above assumptions about the speed vector v and the fitness matrix Σ, M is generally not an isotropic matrix. $\Theta/2$ is the rate at which new mutations are "proposed".

The points of this Poisson Point Process (T_i, A_i, Ξ_i) are such that the (T_i, A_i) form a Poisson Point Process over $\mathbb{R}_+ \times \mathbb{R}^d$ of the mutations that hit the population with intensity $ds\nu(d\alpha)$, and the Ξ_i are i.i.d. $\mathcal{U}[0,1]$, globally independent of the Poisson Point Process of the (T_i, A_i). T_i's are the times when mutations are proposed and A_i's are the effect sizes of those mutations. The Ξ_i are auxiliary variables determining fixation: a mutation gets instantaneously fixed if $\Xi_i \leq g(X_{T_i}, A_i)$, and is lost otherwise.

Note that the limit of the probability of fixation as $\|x\| \to \infty$, while $\frac{x}{\|x\|}$ remains a constant unit vector, is $\mathbf{1}_{\{(x,\alpha)\leq 0\}}$. This means that when the process is sufficiently far away from 0, the fixation mechanism tends to accept all mutations inside the half space $\left(\frac{x}{\|x\|}, \alpha\right) \leq 0$.

Define the covariance matrix of fixed mutations :

$$\bar{V}(x) = \int_{(x,\alpha)\leq 0} \alpha \otimes \alpha \; \nu(d\alpha). \tag{7}$$

Proposition 1. *Under the definition of ν given by* (6)*, $\bar{V}(x)$ is independent of the direction of x and*

$$\bar{V} = \frac{\Theta}{4} M.$$

Proof. The additional factor 1/2 comes from the fact that we integrate over a half space. What is not obvious a priori is that $\bar{V}$ does not depend upon x.

We assume without loss of generality that we have chosen as orthonormal basis of $\mathbb{R}^d$ a basis made of eigenvectors of M, the covariance matrix of p. In other words,

$$M = P'DP,$$

where D is a diagonal matrix and P is the matrix representing the change of basis such that $P' = P^{-1} = P^*$. First, we will show that

$$\frac{1}{(2\pi)^{\frac{d}{2}}(\det D)^{\frac{1}{2}}} \int_{(x,\alpha)\leq 0} \alpha \otimes \alpha \; e^{-\frac{1}{2}\alpha' D^{-1}\alpha} d\alpha = \frac{D}{2}. \tag{8}$$

This is equivalent to showing that for $X_1, \dots, X_d$ being mutually independent zero mean Gaussian random variables, and for any vector $a = (a_1, \dots, a_d) \in \mathbb{R}^d$, with the notation $\mathbb{E}[Z; A] = \mathbb{E}(Z\mathbf{1}_A)$,

$$\mathbb{E}\left[X_j^2; (X, a) \leq 0\right] = \frac{1}{2}\mathbb{E}\left[X_j^2\right], \quad \text{for any } j \in \{1, \dots, d\}$$

and

$$\mathbb{E}\left[X_j X_\ell; (X, a) \leq 0\right] = 0, \quad \text{for any } j \neq \ell \in \{1, \dots, d\}.$$

The first of these two identities follows from the fact that if X and Y are two mutually independent zero mean Gaussian random variables, then

$$\mathbb{E}[X^2; Y < X] = \frac{1}{2}\mathbb{E}[X^2].$$

Indeed, if F_Y denotes the distribution function of the zero mean Gaussian r.v. Y, and σ^2 is the variance of X, since $F_Y(x) + F_Y(-x) = 1$ for all $x \in \mathbb{R}$,

$$\begin{aligned}
\mathbb{E}[X^2; Y \leq X] &= \mathbb{E}[X^2 F_Y(X)] \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty x^2 [F_Y(x) + F_Y(-x)] e^{-x^2/2\sigma^2} dx \\
&= \frac{1}{2}\mathbb{E}[X^2].
\end{aligned}$$

We now establish the second formula. All we have to compute is the following quantity, where X, Y, Z are mutually independent zero mean Gaussian random variables, and a, b are arbitary real numbers,

$$\begin{aligned}
\mathbb{E}[XY; Z \leq aX + bY] &= \mathbb{E}[XY F_Z(aX + bY)] \\
&= \frac{1}{\sigma\tau 2\pi} \int_{\mathbb{R}}\int_{\mathbb{R}} xy F_Z(ax + by) e^{-\frac{x^2}{2\sigma^2} - \frac{y^2}{2\tau^2}} dxdy \\
&= \frac{1}{\sigma\tau 2\pi} \int_0^\infty\int_0^\infty xy[F_Z(ax+by) + F_Z(-ax-by) \\
&\qquad - F_Z(-ax+by) - F_Z(ax-by)]dxdy \\
&= 0,
\end{aligned}$$

since clearly $F_Z(ax + by + F_Z(-ax - by) = F_Z(-ax + by) + F_Z(ax - by) = 1$.

Now that (8) is established, the change of variables $\alpha = P'\tilde{\alpha}$ in the integral formula for $\bar{V}$ yields

$$\bar{V} = \frac{\Theta}{2}\left(P'\frac{D}{2}P\right) = \frac{\Theta}{4}M.$$

□

2 Small Jumps Limit

We now introduce the rescaling

$$\tilde{\alpha} = \varepsilon\alpha \quad \text{and} \quad \tilde{s} = \frac{s}{\varepsilon^2} \text{ with } \varepsilon > 0$$

of the jumps and the time, respectively. In other words, we rewrite our process (5) as

$$X_t^\varepsilon = X_0^\varepsilon - vt + \int_0^t \int_{\mathbb{R}^d} \int_0^1 \varepsilon\alpha\Gamma(X_{s-}^\varepsilon, \varepsilon\alpha, \xi) N_\varepsilon(ds, d\alpha, d\xi),$$

where the intensity measure of the Poisson Point Process N_ε is

$$\left(\frac{1}{\varepsilon^2} ds\right) \times \nu(d\alpha) \times d\xi.$$

The above SDE can be rewritten as

$$X_t^\varepsilon = X_0^\varepsilon - vt + \int_0^t \frac{1}{\varepsilon^2} m_\varepsilon(X_s^\varepsilon) ds + \mathcal{M}_t^\varepsilon, \tag{9}$$

with

$$\begin{aligned} m_\varepsilon(x) &= \int_{\mathbb{R}^d} \varepsilon\alpha g(x, \varepsilon\alpha) \nu(d\alpha), \\ \text{where} \quad g(x, \varepsilon\alpha) &\leq 2\sigma^{-2} \left\|(2x + \varepsilon\alpha, \varepsilon\alpha)\right\| \mathbf{1}_{\{(2x+\varepsilon\alpha, \varepsilon\alpha) \leq 0\}} \\ &\leq 4\sigma^{-2}\varepsilon \left\|(x, \alpha)\right\| \mathbf{1}_{\{(x,\alpha)\leq 0\}}, \end{aligned}$$

and, if $\overline{N}_\varepsilon$ denotes the compensated Poisson measure N_ε, i.e.

$$\overline{N}_\varepsilon(ds, d\alpha, d\xi) = N_\varepsilon(ds, d\alpha, d\xi) - \frac{1}{\varepsilon^2} ds\nu(d\alpha) d\xi,$$

$$\mathcal{M}_t^\varepsilon = \int_0^t \int_{\mathbb{R}^d} \int_0^1 \varepsilon\alpha\Gamma(X_{s-}^\varepsilon, \varepsilon\alpha, \xi) \overline{N}_\varepsilon(ds, d\alpha, d\xi)$$

is a martingale. We will sometimes consider the above SDE with a random initial condition X_0^ε. We insist upon the fact that in such a case the Poisson Point Process N_ε will be assumed to be independent of the initial condition X_0^ε.

The goal of this section is to prove Theorem 1 below, which says that if $X_0^\varepsilon \to \bar{x}_0$, then X_t^ε converges to the solution $\bar{x}_t$ of a linear ODE, and Theorem 2 below, which says the fluctuations $U_t^\varepsilon := \varepsilon^{-1/2}(X_t^\varepsilon - \bar{x}_t)$ converge to an Ornstein–Uhlenbeck process. We prepare the proofs of those two main results by first establishing two Lemmas and one Proposition. Lemma 1, Lemma 2, 1. and Proposition 2, 1. will be used in the proof of Theorem 1, Lemma 2, 2. Proposition 2, 2. in the proof of Theorem 2. The general idea of the proof of both Theorems is to prove tightness and identify a unique possible limit. In Theorem 1, we do that for the sequence X_t^ε, while in Theorem 2 we apply this strategy not directly to U_t^ε, but rather to the right–hand side of the linear SDE whose solution is U_t^ε.

Lemma 1. *If the collection $\{X_0^\varepsilon,\ 0 < \varepsilon \leq 1\}$ of d–dimensional random vectors is tight, then the following two properties hold:*

1. the collection $\{X^\varepsilon,\ 0<\varepsilon\le 1\}$ *is tight in* $\mathbb{D}(\mathbb{R}_+,\mathbb{R}^d)$*;*

2. we have

$$\langle \mathcal{M}^\varepsilon\rangle_t \xrightarrow[\varepsilon\to 0]{} 0 \textit{ in probability, locally uniformly in } t.$$

PROOF. STEP 1 PROOF OF THE LEMMA UNDER A STRONGER CONDITION We assume in this step that the following holds

$$\sup_{0<\varepsilon\le 1} \mathbb{E}(\|X_0^\varepsilon\|^2)<\infty. \tag{10}$$

It is plain that

$$\frac{1}{\varepsilon^2}\|m_\varepsilon(X_t^\varepsilon)\| \le 4\|X_t^\varepsilon\|\sigma^{-2}\int_{\mathbb{R}^d}\|\alpha\|^2\nu(d\alpha), \tag{11}$$

and

$$\begin{aligned}\langle \mathcal{M}^\varepsilon\rangle_t &= \frac{1}{\varepsilon^2}\int_0^t\int_{\mathbb{R}^d}\|\varepsilon\alpha\|^2 g(X_s^\varepsilon,\varepsilon\alpha)\nu(d\alpha)ds\\ &\le 4\sigma^{-2}\varepsilon\int_0^t\int_{(\alpha,X_s^\varepsilon)\le 0}\|\alpha\|^2\,\|(\alpha,X_s^\varepsilon)\|\,\nu(d\alpha)ds\\ &\le 4\sigma^{-2}\varepsilon\left(\int_{\mathbb{R}^d}\|\alpha\|^3\nu(d\alpha)\right)\int_0^t\|X_s^\varepsilon\|\,ds.\end{aligned} \tag{12}$$

Moreover for all $x\in\mathbb{R}^d$, we have

$$\frac{1}{\varepsilon^2}(m_\varepsilon(x),x)\le 0.$$

On the other hand, from (1),

$$\|X_t^\varepsilon\|^2=\|X_0^\varepsilon\|^2-2\int_0^t(v,X_s^\varepsilon)\,ds+\frac{2}{\varepsilon^2}\int_0^t(m_\varepsilon(X_s^\varepsilon),X_s^\varepsilon)\,ds+2\int_0^t X_{s^-}dM_s^\varepsilon+[\mathcal{M}^\varepsilon]_t$$

Hence, for fixed T, we have for all $0\le t\le T$

$$\begin{aligned}\mathbb{E}\left(\|X_t^\varepsilon\|^2\right) &\le \mathbb{E}\left(\|X_0^\varepsilon\|^2\right)+2v_1\int_0^t\mathbb{E}\|X_s^\varepsilon\|ds+\mathbb{E}\langle\mathcal{M}^\varepsilon\rangle_t\\ &\le \mathbb{E}\left(\|X_0^\varepsilon\|^2\right)+v_1\left(t+\int_0^t\mathbb{E}\left(\|X_s^\varepsilon\|^2\right)ds\right)+2\frac{\varepsilon}{\sigma^2}\int_{\mathbb{R}^d}\|\alpha\|^3\nu(d\alpha)\left(t+\int_0^t\mathbb{E}\left(\|X_s^\varepsilon\|^2\right)ds\right)\\ &\le \mathbb{E}\left(\|X_0^\varepsilon\|^2\right)+\left(v_1+2\frac{\varepsilon}{\sigma^2}\int_{\mathbb{R}^d}\|\alpha\|^3\nu(d\alpha)\right)t\\ &\quad+\left(v_1+2\frac{\varepsilon}{\sigma^2}\int_{\mathbb{R}^d}\|\alpha\|^3\nu(d\alpha)\right)\int_0^t\mathbb{E}\left(\|X_s^\varepsilon\|^2\right)ds\\ &\le\left(\mathbb{E}\left(\|X_0^\varepsilon\|^2\right)+\left(v_1+2\frac{\varepsilon}{\sigma^2}\int_{\mathbb{R}^d}\|\alpha\|^3\nu(d\alpha)\right)T\right)e^{\left[v_1+2\frac{\varepsilon}{\sigma^2}\left(\int_{\mathbb{R}^d}\|\alpha\|^3\nu(d\alpha)\right)\right]T},\end{aligned} \tag{13}$$

since $2\mathbb{E}\|X\| \leq 1 + \mathbb{E}(\|X\|^2)$, and where we have used Gronwall's Lemma (see e.g. Proposition 6.59 in [15]). We deduce from (11), (12) and (13) that the process X^ε is tight in $\mathbb{D}\left(\mathbb{R}_+, \mathbb{R}^d\right)$. Indeed, as noted in Remark 14, part 2 page 119 in [14], if $X_t^\varepsilon = X_0^\varepsilon + \int_0^t F_s^\varepsilon ds + \mathcal{M}_t^\varepsilon$, where $\langle \mathcal{M}^\varepsilon \rangle_t = \int_0^t G_s^\varepsilon ds$, tightness of the collection of real valued r.v.'s $\{\|X_0^\varepsilon\| + \int_0^T (\|F_t^\varepsilon\|^2 + \|G_t^\varepsilon\|^2)dt,\ 0 < \varepsilon \leq 1\}$ for each $T > 0$ implies tightness of X^ε in $D(\mathbb{R}_+; \mathbb{R}^d)$ (this makes use of Aldous' tightness criterion [1]). Note that [14] treats real valued processes, but the proof is identical in the vector valued case.

Finally from (12),

$$\langle \mathcal{M}^\varepsilon \rangle_t \xrightarrow[\varepsilon \to 0]{} 0 \text{ in probability.}$$

This convergence is locally uniformly in t since $t \mapsto \langle \mathcal{M}^\varepsilon \rangle_t$ is increasing.

Step 2 Proof of the Lemma under the original assumption For each $n \geq 1$ let φ_n be a bounded function from $\mathbb{R}^d$ into itself such that $\varphi_n(x) = x$ whenever $\|x\| \leq n$. Let now for each $\varepsilon > 0$ and $n \geq 1$ $\{X_t^{\varepsilon,n},\ t \geq 0\}$ denote the solution of equation (9) starting with the initial condition $X_0^{\varepsilon,n} := \varphi_n(X_0^\varepsilon)$, and

$$\mathcal{M}_t^{\varepsilon,n} = \int_0^t \int_{\mathbb{R}^d} \int_0^1 \varepsilon\alpha\Gamma(X_{s^-}^{\varepsilon,n}, \varepsilon\alpha, \xi)\overline{N}_\varepsilon(ds, d\alpha, d\xi).$$

On the event $\{\|X_0^\varepsilon\| \leq n\}$, $X_t^\varepsilon \equiv X_t^{\varepsilon,n}$ and $\mathcal{M}_t^\varepsilon \equiv \mathcal{M}_t^{\varepsilon,n}$. From the original assumption, for any $\delta > 0$, there exists $n_\delta \geq 1$ such that $\mathbb{P}(\|X_0^\varepsilon\| > n_\delta) \leq \delta/2$ for each $\varepsilon > 0$. Let $T > 0$ be arbitrarily fixed. From step 1 of the proof, there exists a compact subset $K_\delta \subset\subset \mathbb{D}([0,T], \mathbb{R}^d)$ such that $\mathbb{P}(X_\cdot^{\varepsilon,n_\delta} \notin K_\delta) \leq \delta/2$, and $\varepsilon_\delta > 0$ small enough such that $\mathbb{P}(\sup_{0\leq t\leq T}\langle \mathcal{M}^{\varepsilon,n_\delta}\rangle_t \geq \delta) \leq \delta/2$ for all $0 < \varepsilon \leq \varepsilon_\delta$. It follows readily that $\mathbb{P}(X^\varepsilon \notin K_\delta) \leq \delta$ for all $0 < \varepsilon \leq 1$ and $\mathbb{P}(\sup_{0\leq t\leq T}\langle \mathcal{M}^\varepsilon\rangle_t \geq \delta) \leq \delta$ for all $0 < \varepsilon \leq \varepsilon_\delta$. QED ☐

Lemma 2. *For all $x \in \mathbb{R}^d$, we have that*

1. $\varepsilon^{-2} m_\varepsilon(x) \xrightarrow[\varepsilon \to 0]{} Lx$, *where*

$$Lx = 4\sigma^{-2} \int_{(x,\alpha)\leq 0} \alpha \, \|(x,\alpha)\| \, \nu(d\alpha) = -4\sigma^{-2}\bar{V}x = -\Theta\sigma^{-2}Mx. \tag{14}$$

2. $\int \alpha \otimes \alpha \frac{g(x,\varepsilon\alpha)}{\varepsilon} \nu(d\alpha) \xrightarrow[\varepsilon \to 0]{} \Lambda(x)$, *where*

$$\Lambda(x) = 4\sigma^{-2} \int_{(x,\alpha)\leq 0} \| \, (x,\alpha) \, \| \alpha \otimes \alpha\nu(d\alpha), \tag{15}$$

Remark 1. *Note that the above two Lemmas remain true if the measure ν is not Gaussian, but satisfies the following moment condition :*

$$\int_{\mathbb{R}^d} \|\alpha\|^4 \nu(d\alpha) < \infty. \tag{16}$$

In this case the limit $L(x)$ is given by the expression

$$L(x) = 4\sigma^{-2} \int_{(x,\alpha)\leq 0} \alpha \, \|(x,\alpha)\| \, \nu(d\alpha) = -4\sigma^{-2} \bar{V}(x) x,$$

where $\bar{V}$ depends upon the direction of x. The advantage of taking a Gaussian measure (up to a multiplicative constant) of new mutations is the resulting linear behavior of the limit Lx, due to the fact that $\bar{V}(x)$ is a constant matrix. More generally, $\bar{V}(x)$ would depend upon $\frac{x}{\|x\|}$.

Proof of Lemma 2. Let for all $\varepsilon > 0$ and $x, \alpha \in \mathbb{R}^d$

$$y_\varepsilon = -2\sigma^{-2} \left(2x + \varepsilon\alpha, \alpha\right), \text{ and } y = -4\sigma^{-2} \left(x, \alpha\right).$$

Note that $\varepsilon^{-2} m_\varepsilon(x) = \int_{y_\varepsilon \geq 0} \frac{1-e^{-\varepsilon y_\varepsilon}}{\varepsilon} \alpha \nu(d\alpha)$. We have that

$$\frac{1-e^{-\varepsilon y_\varepsilon}}{\varepsilon} \mathbf{1}_{y_\varepsilon \geq 0} - y \mathbf{1}_{y\geq 0} = \left(\frac{1-e^{-\varepsilon y_\varepsilon}}{\varepsilon} \mathbf{1}_{y_\varepsilon \geq 0} - y_\varepsilon \mathbf{1}_{y_\varepsilon \geq 0} \right) + \left(y_\varepsilon \mathbf{1}_{y_\varepsilon \geq 0} - y \mathbf{1}_{y \geq 0} \right). \tag{17}$$

In addition, since for all $z > 0$, $z - \frac{z^2}{2} \leq 1 - e^{-z} \leq z$,

$$-\varepsilon \frac{y_\varepsilon^2}{2} \mathbf{1}_{y_\varepsilon \geq 0} \leq \frac{1-e^{-\varepsilon y_\varepsilon}}{\varepsilon} \mathbf{1}_{y_\varepsilon \geq 0} - y_\varepsilon \mathbf{1}_{y_\varepsilon \geq 0} \leq 0, \tag{18}$$

From $y_\varepsilon = y - \frac{2\varepsilon}{\sigma^2} \|\alpha\|^2$, $y_\varepsilon \leq y$ and $\mathbf{1}_{y_\varepsilon \geq 0} \leq \mathbf{1}_{y \geq 0}$. This combines with Cauchy–Schwarz entails

$$y_\varepsilon^2 \mathbf{1}_{y_\varepsilon \geq 0} \leq y^2 \mathbf{1}_{y\geq 0} \leq 16\sigma^{-4} \|x\|^2 \|\alpha\|^2.$$

Combining the last inequality with (18), we obtain

$$-8\sigma^{-4} \varepsilon \|x\|^2 \|\alpha\|^2 \leq \frac{1-e^{-\varepsilon y_\varepsilon}}{\varepsilon} \mathbf{1}_{y_\varepsilon \geq 0} - y_\varepsilon \mathbf{1}_{y_\varepsilon \geq 0} \leq 0. \tag{19}$$

Again $\{y_\varepsilon \geq 0\} \subset \{y \geq 0\}$. It follows that

$$\begin{aligned} y_\varepsilon \mathbf{1}_{y_\varepsilon \geq 0} - y \mathbf{1}_{y \geq 0} &= (y_\varepsilon - y) \mathbf{1}_{y \geq 0} - y_\varepsilon \mathbf{1}_{y \geq 0 \setminus y_\varepsilon \geq 0} \\ &= -2\sigma^{-2} \varepsilon \|\alpha\|^2 \mathbf{1}_{y \geq 0} - y_\varepsilon \mathbf{1}_{y \geq 0 \setminus y_\varepsilon \geq 0} \\ &\geq -2\sigma^{-2} \varepsilon \|\alpha\|^2. \end{aligned}$$

We have proved that

$$-2\sigma^{-2}\varepsilon\|\alpha\|^2 \leq y_\varepsilon \mathbf{1}_{y_\varepsilon \geq 0} - y\mathbf{1}_{y\geq 0} \leq 0. \tag{20}$$

We deduce from (17), (19) and (20) that

$$-\varepsilon\left(8\sigma^{-4}\|x\|^2 + 2\sigma^{-2}\right)\|\alpha\|^2 \leq \frac{1-e^{-\varepsilon y_\varepsilon}}{\varepsilon}\mathbf{1}_{y_\varepsilon\geq 0} - y\mathbf{1}_{y\geq 0} \leq 0.$$

Hence,

$$\left\|\alpha\left(\frac{1-e^{-\varepsilon y_\varepsilon}}{\varepsilon}\mathbf{1}_{y_\varepsilon\geq 0} - y\mathbf{1}_{y\geq 0}\right)\right\| \leq \varepsilon\left(8\sigma^{-4}\|x\|^2 + 2\sigma^{-2}\right)\|\alpha\|^3. \tag{21}$$

By integrating (21) with respect to ν, we obtain

$$\left\|\frac{1}{\varepsilon^2}m_\varepsilon(x) - L(x)\right\| \leq \varepsilon\left(8\sigma^{-4}\|x\|^2 + 2\sigma^{-2}\right)\int_{\mathbb{R}^d}\|\alpha\|^3\nu(d\alpha). \tag{22}$$

Hence,

$$\varepsilon^{-2}m_\varepsilon(x) \underset{\varepsilon\to 0}{\longrightarrow} Lx.$$

By a similar argument, we have that

$$\left\|\int \alpha\otimes\alpha\frac{g(x,\varepsilon\alpha)}{\varepsilon}\nu(d\alpha) - \Lambda(x)\right\| \leq \varepsilon\left(8\sigma^{-4}\|x\|^2 + 2\sigma^{-2}\right)\int_{\mathbb{R}^d}\|\alpha\|^4\nu(d\alpha), \tag{23}$$

and we deduce the second result of the Lemma. □

Under the assumption of Lemma 1, we can extract a subsequence which we still denote X^ε by an abuse of notation such that $X^\varepsilon \Rightarrow \bar{X}$ (this notation means here and below that X^ε converges in law towards $\bar{X}$ as $D(\mathbb{R}_+,\mathbb{R}^d)$–valued random variables), and we have the following result:

Proposition 2. *If $X^\varepsilon \Rightarrow \bar{X}$, then*

$$1.\quad \frac{1}{\varepsilon^2}m_\varepsilon(X^\varepsilon_\cdot) \Rightarrow L\bar{X}. \quad \text{in } D(\mathbb{R}_+,\mathbb{R}^d),$$

and

$$2.\quad \frac{1}{\varepsilon}\frac{d}{dt}\langle\langle\mathcal{M}^\varepsilon\rangle\rangle_\cdot \Rightarrow \Lambda(\bar{X}_\cdot) \quad \text{in } D(\mathbb{R}_+,\mathbb{R}^d),$$

Proof of 1. It follows from (22) that for all $\delta, C > 0$, there exists $\varepsilon_{\delta,C}$ such that if $\varepsilon < \varepsilon_{\delta,C}$ and for all $\|x\| \leq C$,

$$\left\|\frac{1}{\varepsilon^2}m_\varepsilon(x) - Lx\right\| \leq \delta,$$

thus, for an arbitrary $T > 0$ and for $\varepsilon < \varepsilon_{\delta,C}$,

$$\mathbb{P}\left(\sup_{t\leq T}\left\|\frac{1}{\varepsilon^2}m_\varepsilon(X_t^\varepsilon) - L(X_t^\varepsilon)\right\| > \delta\right) \leq \mathbb{P}\left(\sup_{t\leq T}\|X_t^\varepsilon\| > C\right).$$

It follows that for all $\delta, C > 0$,

$$\limsup_{\varepsilon\to 0}\mathbb{P}\left(\sup_{t\leq T}\left\|\frac{1}{\varepsilon^2}m_\varepsilon(X_t^\varepsilon) - L(X_t^\varepsilon)\right\| > \delta\right) \leq \sup_{\varepsilon}\mathbb{P}\left(\sup_{t\leq T}\|X_t^\varepsilon\| > C\right).$$

From the tightness of X^ε, for all $\eta > 0$, we can choose $C > 0$ such that

$$\sup_{\varepsilon}\mathbb{P}\left(\sup_{t\leq T}\|X_t^\varepsilon\| > C\right) \leq \eta.$$

Hence for all $\delta, \eta > 0$

$$\limsup_{\varepsilon\to 0}\mathbb{P}\left(\sup_{t\leq T}\left\|\frac{1}{\varepsilon^2}m_\varepsilon(X_t^\varepsilon) - L(X_t^\varepsilon)\right\| > \delta\right) \leq \eta.$$

Moreover $L(X^\varepsilon) \Rightarrow L(\bar{X})$ in $\mathbb{D}(\mathbb{R}_+, \mathbb{R}^d)$ since L is a continuous function. Consequently,

$$\frac{1}{\varepsilon^2}m_\varepsilon(X^\varepsilon) = \frac{1}{\varepsilon^2}m_\varepsilon(X^\varepsilon) - L(X^\varepsilon) + L(X^\varepsilon) \Rightarrow L(\bar{X}).$$

Proof of 2. Note that

$$\frac{d}{dt}\langle\langle\mathcal{M}^\varepsilon\rangle\rangle_t = \int_{\mathbb{R}^d}\alpha\otimes\alpha\; g(X_t^\varepsilon, \varepsilon\,\alpha)\nu(d\alpha).$$

By a similar argument as in the first part of the proof, using this time (23), we establish the second claim, thanks to the fact that

$$\int\|\alpha\|^4\nu(d\alpha) < \infty.$$

□

We can now establish

Theorem 1. *If $X_0^\varepsilon \to x_0$ in probability as $\varepsilon \to 0$, then $X_t^\varepsilon \xrightarrow[\varepsilon\to 0]{} \bar{x}_t$ in probability, locally uniformly in t, where*

$$\frac{d\bar{x}_t}{dt} = -v + L\bar{x}_t, \quad \bar{x}_0 = x_0. \tag{24}$$

PROOF. The assumption and Lemma 1 imply tightness of the collection of processes $\{X^\varepsilon_\cdot,\ 0<\varepsilon\le 1\}$. Hence X^ε converges weakly to a limit along a subsequence. From (9), Proposition 2, 1. and Lemma 1, 2., we deduce that the limit solves the ODE (24). Since the limit $\bar{x}_t$ is uniquely determined and deterministic, the whole collection X^ε_t converges in probability as $\varepsilon\to 0$ towards $\bar{x}_t$, locally uniformly in t. □

Remark 2. *In case $X^\varepsilon_0 \Rightarrow X_0$, where X_0 is a d–dimensional random vector, the same result still holds, except that the limit is random and the convergence is in law.*

The differential equation (24) represents a deterministic approximation for the stochastic process X^ε in the limit of small jumps. We note that

$$\bar{x}_t \xrightarrow[t\to\infty]{} \bar{x}_\infty = -\frac{M^{-1}v}{\Theta\sigma^{-2}}. \tag{25}$$

To estimate the fluctuations of the process in the small-jumps limit, we now consider the following process

$$U^\varepsilon_t = \frac{X^\varepsilon_t - \bar{x}_t}{\sqrt{\varepsilon}}. \tag{26}$$

Theorem 2. *Assume that $X^\varepsilon_0 = \bar{x}_0$. Then $U^\varepsilon \Rightarrow U$, where U is an Ornstein-Uhlenbeck process:*

$$\begin{aligned} dU_t &= LU_t dt + \Lambda^{\frac{1}{2}}(\bar{x}_t)dB_t,\\ U_0 &= 0, \end{aligned} \tag{27}$$

with B being a d-dimensional standard Brownian motion. In other words,

$$U_t = \int_0^t e^{L(t-s)}\Lambda^{\frac{1}{2}}(\bar{x}_s)dB_s.$$

PROOF. We have that

$$\begin{aligned} U^\varepsilon_t &= \int_0^t \frac{\varepsilon^{-2}m_\varepsilon(X^\varepsilon_s) - L\bar{x}_s}{\sqrt{\varepsilon}}ds + \frac{1}{\sqrt{\varepsilon}}\mathcal{M}^\varepsilon_t,\\ &= L\int_0^t \frac{X^\varepsilon_s - \bar{x}_s}{\sqrt{\varepsilon}}ds + \frac{1}{\sqrt{\varepsilon}}\int_0^t \left(\frac{1}{\varepsilon^2}m_\varepsilon(X^\varepsilon_s) - LX^\varepsilon_s\right)ds + \frac{1}{\sqrt{\varepsilon}}\mathcal{M}^\varepsilon_t\\ &= L\int_0^t U^\varepsilon_s ds + \frac{1}{\sqrt{\varepsilon}}\int_0^t \left(\frac{1}{\varepsilon^2}m_\varepsilon(X^\varepsilon_s) - LX^\varepsilon_s\right)ds + \frac{1}{\sqrt{\varepsilon}}\mathcal{M}^\varepsilon_t. \end{aligned}$$

Thus by the formula for the solution of a linear ODE and an integration by parts,

$$U^\varepsilon_t = \frac{1}{\sqrt{\varepsilon}}\int_0^t e^{L(t-s)}\left(\frac{1}{\varepsilon^2}m_\varepsilon(X^\varepsilon_s) - LX^\varepsilon_s\right)ds + \frac{1}{\sqrt{\varepsilon}}\mathcal{M}^\varepsilon_t + \frac{1}{\sqrt{\varepsilon}}\int_0^t e^{L(t-s)}L\mathcal{M}^\varepsilon_s ds. \tag{28}$$

We deduce from (22) that

$$\frac{1}{\sqrt{\varepsilon}}\int_0^t e^{L(t-s)}\left(\frac{1}{\varepsilon^2}m_\varepsilon(X_s^\varepsilon)-LX_s^\varepsilon\right)ds \underset{\varepsilon\to 0}{\longrightarrow} 0 \tag{29}$$

in probability. Furthermore, $\frac{1}{\sqrt{\varepsilon}}\mathcal{M}^\varepsilon$ are tight martingales by a similar argument as before since

$$\sup_{\{t>0,\ \varepsilon\}}\int_{\mathbb{R}^2}\|\alpha\|^2 g(X_t^\varepsilon,\varepsilon\alpha)\nu(d\alpha)\le\int_{\mathbb{R}^2}\|\alpha\|^2\nu(d\alpha)<\infty.$$

Let us admit for a moment the
Lemma 3. *For any $T>0$, as $\varepsilon\to 0$,*

$$\sup_{t\le T}\left\|\frac{1}{\sqrt{\varepsilon}}\left(\mathcal{M}_t^\varepsilon-\mathcal{M}_{t^-}^\varepsilon\right)\right\|\to 0$$

in probability.

Hence every converging subsequence of $\frac{1}{\sqrt{\varepsilon}}\mathcal{M}^\varepsilon$ converges to a continuous process $\mathcal{M}$ as ε goes to 0, and using Proposition 2, 2., we have (recall (15))

$$\langle\langle\frac{1}{\sqrt{\varepsilon}}\mathcal{M}^\varepsilon\rangle\rangle_t=\int_0^t\alpha\otimes\alpha\frac{g(X_s^\varepsilon,\varepsilon\alpha)}{\varepsilon}\nu(d\alpha)ds\underset{\varepsilon\to 0}{\longrightarrow}\int_0^t\Lambda(\bar{x}_s)ds.$$

Let us now explain why these facts imply that $\mathcal{M}_t$ is a martingale such that

$$\langle\langle\mathcal{M}\rangle\rangle_t=\int_0^t\Lambda(\bar{x}_s)ds.$$

We first note that since $X_0^\varepsilon(\omega)=\bar{x}_0$ for all $\varepsilon>0$ and ω, we deduce from (12) and (13) that for each $t>0$ the collections indexed by ε of random matrices and vectors $\frac{1}{\varepsilon}\langle\mathcal{M}^\varepsilon\rangle_t$, $\frac{1}{\varepsilon}\mathcal{M}_t^\varepsilon\otimes\mathcal{M}_t^\varepsilon$ and $\frac{1}{\sqrt{\varepsilon}}\mathcal{M}_t^\varepsilon$ are uniformly integrable. On the other hand, for any $n\ge 1$, $0<s_1<\cdots<s_n=s<t$, $\Phi\in C_b(\mathbb{R}^{nd})$, we have

$$\mathbb{E}\left\{\Phi\left(\frac{\mathcal{M}_{s_1}^\varepsilon}{\sqrt{\varepsilon}},\dots,\frac{\mathcal{M}_{s_n}^\varepsilon}{\sqrt{\varepsilon}}\right)\left[\frac{\mathcal{M}_t^\varepsilon}{\sqrt{\varepsilon}}-\frac{\mathcal{M}_s^\varepsilon}{\sqrt{\varepsilon}}\right]\right\}=0,$$

$$\mathbb{E}\left\{\Phi\left(\frac{\mathcal{M}_{s_1}^\varepsilon}{\sqrt{\varepsilon}},\dots,\frac{\mathcal{M}_{s_n}^\varepsilon}{\sqrt{\varepsilon}}\right)\left[\frac{1}{\varepsilon}\mathcal{M}_t^\varepsilon\otimes\mathcal{M}_t^\varepsilon-\frac{1}{\varepsilon}\langle\langle\mathcal{M}^\varepsilon\rangle\rangle_t-\frac{1}{\varepsilon}\mathcal{M}_s^\varepsilon\otimes\mathcal{M}_s^\varepsilon+\frac{1}{\varepsilon}\langle\langle\mathcal{M}^\varepsilon\rangle\rangle_s\right]\right\}=0.$$

We may take the limit as $\varepsilon\to 0$ in these two identities, which yields identities which (thanks to the freedom of choice of n, $s_1,\dots,s_n$ and Φ) show that both $\mathcal{M}_t$ and $\mathcal{M}_t\otimes\mathcal{M}_t-\int_0^t\Lambda(\bar{x}_s)ds$ are continuous martingales.

We deduce thanks to a representation theorem of continuous martingales, see e.g. Theorem 4.5.2 from [18], that there exists a d-dimensional Brownian motion B_t such that

$$\mathcal{M}_t = \int_0^t \Lambda^{\frac{1}{2}}(\bar{x}_s) dB_s, \ t \geq 0.$$

This being true for any converging subsequence of $\frac{1}{\sqrt{\varepsilon}}\mathcal{M}^\varepsilon$, the limit is unique (in law). Finally, combining this with (28) and (29), we deduce that

$$U_t^\varepsilon \Rightarrow \int_0^t e^{L(t-s)} \Lambda^{\frac{1}{2}}(\bar{x}_s) dB_s.$$

□

It remains to turn to the

Proof of Lemma 3 It is clearly sufficient to prove the result for any coordinate of the d–dimensional process $\mathcal{M}_t^\varepsilon$, which means that it suffices to prove the result in case $d = 1$, which we assume for the rest of this proof. In other words, p is the density of the Gaussian law $\mathcal{N}(0, \gamma^2)$, with some $\gamma > 0$. Let us define

$$\widetilde{\mathcal{M}}_t^\varepsilon = \int_{[0,t]\times\mathbb{R}\times[0,1]} \varepsilon\alpha N_\varepsilon(ds, d\alpha, d\xi).$$

It is plain that

$$\sup_{t\leq T}\left|\frac{1}{\sqrt{\varepsilon}}\left(\mathcal{M}_t^\varepsilon - \mathcal{M}_{t-}^\varepsilon\right)\right| \leq \sup_{t\leq T}\left|\frac{1}{\sqrt{\varepsilon}}\left(\widetilde{\mathcal{M}}_t^\varepsilon - \widetilde{\mathcal{M}}_{t-}^\varepsilon\right)\right|.$$

Hence all we need to show is that the above right hand side tends to 0 in probability, as $\varepsilon \to 0$. Let $a > 0$ be arbitrary. We need to estimate

$$\mathbb{P}\left(\sup_{t\leq T}|\widetilde{\mathcal{M}}_t^\varepsilon - \widetilde{\mathcal{M}}_{t-}^\varepsilon| > a\sqrt{\varepsilon}\right).$$

It is plain that, with $Z_\varepsilon := N_\varepsilon([0,T]\times\mathbb{R}\times[0,1])$,

$$\begin{aligned}
\mathbb{P}\left(\sup_{t\leq T}|\widetilde{\mathcal{M}}_t^\varepsilon - \widetilde{\mathcal{M}}_{t-}^\varepsilon| > a\sqrt{\varepsilon}\right) &\leq \mathbb{P}\left(Z_\varepsilon > K\varepsilon^{-2}\right) \\
&\quad + \mathbb{P}\left(Z_\varepsilon \leq K\varepsilon^{-2}, \sup_{t\leq T}|\widetilde{\mathcal{M}}_t^\varepsilon - \widetilde{\mathcal{M}}_{t-}^\varepsilon| > a\sqrt{\varepsilon}\right) \\
&= A(K,\varepsilon,T) + B(K,\varepsilon,T).
\end{aligned}$$

Since the law of Z_ε is Poisson with mean $(2\varepsilon^2)^{-1}\Theta T$, it follows from Chebychef's inequality that

$$A(K,\varepsilon,T) \leq \frac{\Theta T}{2K}. \tag{30}$$

On the other hand, if $\{Y_k,\ k \geq 1\}$ are i.i.d. r.v.'s with the common distribution $\mathcal{N}(0,1)$, with the notation $a_\varepsilon = \frac{a}{\gamma\sqrt{\varepsilon}}$,

$$\begin{aligned} B(K,\varepsilon,T) &\leq \mathbb{P}\left(\sup_{0\leq k\leq K\varepsilon^{-2}} |Y_k| \geq a_\varepsilon\right) \\ &\leq 2\mathbb{P}\left(\sup_{0\leq k\leq K\varepsilon^{-2}} Y_k \geq a_\varepsilon\right) \\ &= 2 - 2\mathbb{P}\left(\sup_{0\leq k\leq K\varepsilon^{-2}} Y_k \leq a_\varepsilon\right) \\ &= 2 - 2\left(1 - \mathbb{P}(Y_1 > a_\varepsilon)\right)^{K\varepsilon^{-2}}. \end{aligned}$$

Note that, since $\exp(a_\varepsilon Y_1 - a_\varepsilon^2/2)$ has mean 1,

$$\mathbb{P}(Y_1 > a_\varepsilon) = \mathbb{P}(e^{a_\varepsilon Y_1 - a_\varepsilon^2/2} > e^{a_\varepsilon^2/2}) \leq e^{-a_\varepsilon^2/2}.$$

Hence

$$B(K,\varepsilon,T) \leq 2 - 2\left(1 - e^{-a_\varepsilon^2/2}\right)^{K\varepsilon^{-2}}$$

For ε small enough, $e^{-a_\varepsilon^2/2} \leq 1/2$, from which follows that $1 - e^{-a_\varepsilon^2/2} \geq \exp(-e^{-a_\varepsilon^2/2})$, hence

$$B(K,\varepsilon,T) \leq 2\left(1 - \exp\left[-\frac{K}{\varepsilon^2} e^{-\frac{a^2}{2\gamma^2\varepsilon}}\right]\right). \tag{31}$$

For any $\delta > 0$, we first choose $K_\delta = \frac{\Theta T}{\delta}$, so that from (30), $A(K_\delta,\varepsilon,T) \leq \delta/2$ for all $\varepsilon > 0$, and then thanks to (31) ε small enough such that $B(K_\delta,\varepsilon,T) \leq \delta/2$. The result follows. □

Remark 3. *We have exploited the fact that p is the density of a Gaussian distribution. In fact it would suffice that $\mathbb{P}(|Y_1| > a) \leq \frac{C}{a^{4+\delta}}$ for some $C, \delta > 0$. We note that this is exactly what is already necessary for* (16) *to hold.*

The Ornstein–Uhlenbeck process U_t which appears in the last Theorem is a centered Gaussian vector. It is easy to compute its covariance matrix

$$\mathbb{E}(U_t \otimes U_t) = \int_0^t e^{L(t-s)}\Lambda(\bar{x}_s)e^{L(t-s)}ds.$$

as well as the limit of the latter as $t \to \infty$

$$\mathbb{E}(U_t \otimes U_t) \xrightarrow[t\to\infty]{} \bar{S}^2 = \int_0^\infty e^{Lt}\Lambda(\bar{x}_\infty)e^{Lt}dt. \tag{32}$$

It follows readily that the law of U_t converges, as $t \to \infty$, to $\mathcal{N}(0,\bar{S}^2)$. In case $\bar{x}_t = \bar{x}_\infty$ for all $t \geq 0$, then U_t solution of the linear SDE $dU_t = LU_t dt + \Lambda^{\frac{1}{2}}(\bar{x}_\infty)dB_t$ is a time homogeneous Gauss–Markov process, and $\mathcal{N}(0,\bar{S}^2)$ is its unique invariant probability measure, see Theorem 5.6.7 in Karatzas and Shreve [6]. Moreover $(\bar{x}_t, U_t)$ is a time homogeneous Markov process and $\delta_{\bar{x}_\infty} \times \mathcal{N}(0,\bar{S}^2)$ is its unique invariant probability measure.

3 Large time behavior of X_t^ε for small $\varepsilon > 0$

In this section, we shall first prove that provided ε is small enough, X_t^ε is positive Harris recurrent with a unique invariant probability measure, next that the collection of invariant probability measures indexed by ε is tight, and finally that the large time behavior of $\bar{x}_t + \sqrt{\varepsilon}U_t$ is a good approximation of the large time behavior of X_t^ε, thus justifying a sort of interchange of the limits as $\varepsilon \to 0$ and as $t \to \infty$.

3.1 Large time behavior of the process X_t^ε

The aim of this subsection is to prove (in the next statement, $\|\cdot\|$ stands for the total variation norm):

Theorem 3. *There exists $\varepsilon_0 > 0$ such that for any $0 < \varepsilon \le \varepsilon_0$, the process X_t^ε is positive Harris recurrent, with a unique invariant probability measure μ^ε. Moreover X_t^ε is ergodic, in the sense that $\|\mathbb{P}(X_t^\varepsilon \in \cdot | X_0^\varepsilon = x) - \mu^\varepsilon\| \to 0$, as $t \to \infty$, for all $x \in \mathbb{R}^d$.*

Thanks to Theorem 4.2 from [12], the first part Theorem 3 follows from a positive recurrence condition (which is called (CD2) in [12]), and the fact that a certain ball centered at 0 is a petite set. We first establish in the next Lemma and the subsequent Corollary an a priori estimate. The proof of the first part of the Theorem will then consist in two steps : conclude (CD2) from the Corollary, and verify the petite set property. The second part will follow from Theorem 5.1 in [12].

Lemma 4. *There exists three constants $c, C, a > 0$ (whose values depend explicitly upon the parameters of our model) such that for all $0 < \varepsilon \le \frac{1}{4}$, $t > 0$,*

$$\|X_t^\varepsilon\| \le \|X_0^\varepsilon\| + C\, t - c\int_0^t \left(\|X_s^\varepsilon\| \wedge \frac{a}{\varepsilon}\right) ds + \mathcal{N}_t^\varepsilon,$$

where the local martingale $\mathcal{N}_t^\varepsilon$, which is a martingale as soon as $\mathbb{E}[\|X_0^\varepsilon\|] < \infty$, is given as

$$\mathcal{N}_t^\varepsilon = \int_{[0,t]\times\mathbb{R}^d\times[0,1]} (\|X_{s-}^\varepsilon + \varepsilon\alpha\| - \|X_{s-}^\varepsilon\|)\Gamma(X_{s-}^\varepsilon, \varepsilon\alpha, \xi)\bar{M}_\varepsilon(ds, d\alpha, d\xi).$$

Proof. We have

$$\begin{aligned}
\|X_t^\varepsilon\| &= \|X_0^\varepsilon\| - \int_0^t \left(v, \frac{X_s^\varepsilon}{\|X_s^\varepsilon\|}\right) ds + \sum_{s\le t} \left(\|X_{s^-}^\varepsilon + \Delta X_s^\varepsilon\| - \|X_{s^-}^\varepsilon\|\right) \\
&= \|X_0^\varepsilon\| - \int_0^t \left(v, \frac{X_s^\varepsilon}{\|X_s^\varepsilon\|}\right) ds \\
&\quad + \int_{[0,t]\times\mathbb{R}^d\times[0,1]} \left(\|X_{s-}^\varepsilon + \varepsilon\alpha\| - \|X_{s-}^\varepsilon\|\right) \Gamma(X_{s-}^\varepsilon, \varepsilon\alpha, u) N_\varepsilon(ds, d\alpha, du) \\
&= \|X_0^\varepsilon\| + \int_0^t \left[\int_{\mathbb{R}^d} \frac{\|X_s^\varepsilon + \varepsilon\alpha\| - \|X_s^\varepsilon\|}{\varepsilon} \times \frac{g(X_s^\varepsilon, \varepsilon\alpha)}{\varepsilon} \nu(d\alpha) - \left(v, \frac{X_s^\varepsilon}{\|X_s^\varepsilon\|}\right)\right] ds + \mathcal{N}_t^\varepsilon
\end{aligned} \tag{33}$$

Recall that only mutations which improve the fitness may get fixed, since from (3) and (4) $g(x,\alpha) > 0$ requires $s(x,\alpha) > 0$, i.e. $(2x+\alpha, \alpha) < 0$, hence, with again the notation $\Delta X_s^\varepsilon = X_s^\varepsilon - X_{s-}^\varepsilon$,

$$\|X_{s^-}^\varepsilon + \Delta X_s^\varepsilon\| - \|X_{s^-}^\varepsilon\| \le 0,$$

and moreover we shall restrict ourselves below to the jumps such that

$$\|X_{s^-}^\varepsilon + \Delta X_s^\varepsilon\| - \|X_{s^-}^\varepsilon\| \le -\frac{1}{2}\|\Delta X_s^\varepsilon\|.$$

For $x \in \mathbb{R}^d$, let $C(x)$ denote the set of vectors α such that

$$\|x+\alpha\| - \|x\| \le -\frac{1}{2}\|\alpha\|.$$

We define γ, the angle between the two directions $-x$ and α, as follows. If x and α are colinear, then $\gamma = 0$. Otherwise, we consider the plane spanned by x and α. If we restrict ourselves to those α such that $\Gamma(x, \varepsilon\alpha, 0) > 0$, then in the above plane, the angle between the directions $-x$ and α is between $-\pi/2$ and $\pi/2$. We denote by γ that angle, whose sign is unimportant, since only $\cos\gamma$ will enter our computations. From the identity (again we consider only the case where $(x,\alpha) < 0$)

$$\|x+\alpha\| = \sqrt{\|x\|^2 + \|\alpha\|^2 - 2\|x\|\,\|\alpha\|\cos\gamma},$$

we deduce that $\alpha \in C(x)$ iff

$$\|\alpha\| \le \frac{4}{3}(2\cos\gamma - 1)\|x\|.$$

We note in particular that whenever $\alpha \in C(x)$ and $\varepsilon \le 1$, then $\varepsilon\alpha \in C(x)$. Now

$$\begin{aligned}
\int_{\mathbb{R}^d} \frac{\|X_s^\varepsilon + \varepsilon\alpha\| - \|X_s^\varepsilon\|}{\varepsilon} \times \frac{g(X_s^\varepsilon, \varepsilon\alpha)}{\varepsilon} \nu(d\alpha) &\le -\frac{1}{2} \int_{C(X_s^\varepsilon)/\varepsilon} \|\alpha\| \frac{g(X_s^\varepsilon, \varepsilon\alpha)}{\varepsilon} \nu(d\alpha) \\
&\le -(1-e^{-1}) \int_{C(X_{s-}^\varepsilon)/\varepsilon} \|\alpha\| \left(\frac{(2X_s^\varepsilon + \varepsilon\alpha, \alpha)_-}{\sigma^2} \wedge \frac{1}{2\varepsilon}\right) \nu(d\alpha)
\end{aligned}$$

where $(y,z)_-$ stands for the negative part the the scalar product of the two vectors y and z, and we have exploited the elementary inequality

$$1 - e^{-u} \geq (1 - e^{-1})(u \wedge 1), \text{ for all } u \geq 0.$$

We now need to lower bound the factor of $-(1 - e^{-1})$ in the last right hand side. For that sake, we consider the expression

$$\int_{\alpha \in C(x)/\varepsilon} \|\alpha\| \left(\frac{(2x + \varepsilon\alpha, \alpha)_-}{\sigma^2} \wedge \frac{1}{2\varepsilon} \right) \nu(d\alpha).$$

For $\|x\| \leq 2$, we lower bound this integral by 0. We now consider the case $\|x\| > 2$. We lower bound the integral by reducing the integration to the set

$$A_\varepsilon(x) = \frac{C(x)}{\varepsilon} \cap \{1 \leq \|\alpha\| \leq 2\}.$$

It is not hard to see that whenever $\alpha \in A_\varepsilon(x)$, $-2 \leq \left(\alpha, \frac{x}{\|x\|}\right) < -1/2$. Indeed, the lower bound is clear and from $\varepsilon\alpha \in C(x)$, we deduce that $2(\varepsilon\alpha, x) + \|\alpha\|^2 \leq \frac{\|\alpha\|^2}{4} - \|x\| \times \|\alpha\|$, whence $(\alpha, \frac{x}{\|x\|}) \leq -\frac{\|\alpha]}{2} \leq -\frac{1}{2}$, since $\|\alpha\| \geq 1$. Moreover, since $\|x\| > 2$, if $\alpha \in A_\varepsilon(x)$,

$$-2\|x\| < (x, \alpha) < -\frac{\|x\|}{2} < -1, \quad \text{because } \varepsilon\|\alpha\|^2 \leq 1,$$

provided $\varepsilon \leq 1/4$. Consequently $(x, \alpha) + \varepsilon\|\alpha\|^2 \leq 0$ and

$$(2x + \varepsilon\alpha, \alpha)_- \geq \|(x, \alpha)\| \geq \frac{\|x\|}{2},$$

so that

$$\begin{aligned} \int_{A_\varepsilon(x)} \|\alpha\| \left(\frac{(2x + \varepsilon\alpha, \alpha)_-}{\sigma^2} \wedge \frac{1}{2\varepsilon} \right) \nu(d\alpha) &\geq \frac{1}{2\sigma^2} \left(\|x\| \wedge \frac{\sigma^2}{\varepsilon} \right) \int_{A_\varepsilon(x)} \|\alpha\| \nu(d\alpha) \\ &\geq \frac{\beta}{2\sigma^2} \left(\|x\| \wedge \frac{\sigma^2}{\varepsilon} \right), \end{aligned}$$

where $\beta = \inf_{\|x\|>2, \varepsilon \leq 1} \int_{A_\varepsilon(x)} \|\alpha\| \nu(d\alpha) > 0$. We have proved that, with $a = \sigma^2$ and $c = (1 - e^{-1})\beta/(2\sigma^2)$,

$$\|X_t^\varepsilon\| \leq \|X_0^\varepsilon\| + v_1 \times t - c \int_0^t \mathbf{1}_{\|X_s^\varepsilon\|>2} \left(\|X_s^\varepsilon\| \wedge \frac{a}{\varepsilon} \right) ds + \mathcal{N}_t^\varepsilon.$$

The result follows with $C = v_1 + 2c$. $\square$

Corollary 1. *There exist* $0 < \varepsilon_0 \leq 1/4$, $b, k > 0$ *and* B *a compact subset of* $\mathbb{R}^d$ *such that for any* $0 < \varepsilon \leq \varepsilon_0$,

$$\|X_t^\varepsilon\| \leq \|X_0^\varepsilon\| - bt + k\int_0^t \mathbf{1}_B(X_s^\varepsilon)ds + \mathcal{N}_t^\varepsilon. \tag{34}$$

Proof. We choose $B = \{x;\ \|x\| \leq 2C/c\}$ and $\varepsilon_0 = \frac{ca}{2C}$. (34) with $b = C$, $k = 2C$ now follows from Lemma 4. □

We finally turn to the

Proof of Theorem 3. Step 1. Verification of the positive recurrence condition The process X_t^ε satisfies (34) which is exactly condition (CD2) from [12]. Indeed, (33) implies that the function $V(x) := \|x\|$ belongs to the domain of the extended generator $\mathcal{A}^\varepsilon$ of the Markov process X_t^ε and that

$$\mathcal{A}^\varepsilon V(x) = \int_{\mathbb{R}^d} \frac{\|x+\varepsilon\alpha\| - \|x\|}{\varepsilon} \times \frac{g(x,\varepsilon\alpha)}{\varepsilon}\nu(d\alpha) - \left(v, \frac{x}{\|x\|}\right).$$

The proof of Corollary 1 implies that

$$\mathcal{A}^\varepsilon V(x) \leq -b + k\mathbf{1}_B(x), \quad \text{again for } V(x) = \|x\|.$$

Step 2. B is a petite set In order to deduce positive Harris recurrence from Theorem 4.2 in [12], it remains to show that $B = \{x;\ \|x\| \leq K\}$, with $K = 2C/c$ is a closed petite set. This will follow if we show that there exist two constants $t_B, C_B > 0$ such that for all $x \in B$,

$$\mathbb{P}(X_{t_B}^\varepsilon \in dy | X_0^\varepsilon = x) \geq C_B \mathbf{1}_{\|y\|\leq 1} dy. \tag{35}$$

We choose $0 < \eta < 1$ and define the event

$$A_{K,\eta}^\varepsilon = \left\{ \begin{array}{c} X_t^\varepsilon \text{ does not jump on the time interval } \left[0, \frac{2K-\eta}{v}\right], \text{ and} \\ X_t^\varepsilon \text{ jumps exactly once on the time interval } \left(\frac{2K-\eta}{v}, \frac{2K}{v}\right]. \end{array} \right\}$$

It is plain that $\mathbb{P}(A_{K,\eta}^\varepsilon) > 0$. Let T^ε denote the first jump time of X_t^ε. We note that, provided that $X_0^\varepsilon = x \in B$, on the event $A_{K,\eta}^\varepsilon$, $X_{T^\varepsilon-}^\varepsilon \in B_{K,\eta}$, where

$$B_{K,\eta} = \{x \in \mathbb{R}^d,\ -3K \leq x_1 \leq -K+\eta,\ -K \leq x_i \leq K, 2 \leq i \leq d\}.$$

. Let Λ_ε denote a random vector which is such that the law of $\varepsilon^{-1}\Lambda_\varepsilon$ has the density p defined by (6). Denote by $f_\varepsilon(y) = \varepsilon p(\varepsilon y)$ the density of the law of Λ_ε. We define

$$\Sigma_\varepsilon = \Lambda_\varepsilon \mathbf{1}_{\{\|X_{T^\varepsilon-}^\varepsilon + \Lambda_\varepsilon\| \leq 1+\eta\}}.$$

On the event $\Sigma_\varepsilon \neq 0$, since $K > 2 > 1 + \eta$, $g(X^\varepsilon_{T-}, \Sigma_\varepsilon) \geq c_{\varepsilon,K,\eta}$, where

$$c_{\varepsilon,K,\eta} = \inf_{x \in B_{K,\eta}, \|x+\alpha\| \leq 1+\eta} g(x, \alpha) > 0.$$

We denote by ξ the random variable with the uniform distribution on the interval $[0, 1]$, which is such that whenever $\xi \leq g(X^\varepsilon_{T-}, \Sigma_\varepsilon)$, the "proposed" jump Σ_ε happens at time T^ε. Recall that $0 \leq \frac{2K}{v} - T^\varepsilon < \frac{\eta}{v}$. On the event $A^\varepsilon_{K,\eta} \cap \{\Sigma_\varepsilon \neq 0\} \cap \{\xi \leq c_{\varepsilon,K,\eta}\}$,

$$X_{\frac{2K}{v}} = X^\varepsilon_{T^\varepsilon -} + \Sigma_\varepsilon + v\left(\frac{2K}{v} - T^\varepsilon\right),$$

and for any $h \in C(\mathbb{R}^d, \mathbb{R}^+)$,

$$\mathbb{E}h\left(X^\varepsilon_{\frac{2K}{v}}\right) \geq \mathbb{P}(A^\varepsilon_{K,\eta} \cap \{\Sigma_\varepsilon \neq 0\}) c_{\varepsilon,K,\eta} a_{\varepsilon,K,\eta} \int_{\|y\| \leq 1} h(y) dy,$$

with

$$a_{\varepsilon,K,\eta} = \inf_{x \in B_{K,\eta}, \|x+y\| \leq 1+\eta} f_\varepsilon(y).$$

(35) follows, with $t_B = \frac{2K}{v}$ and $C_B = \mathbb{P}(A^\varepsilon_{K,\eta} \cap \{\Sigma_\varepsilon \neq 0\}) c_{\varepsilon,K,\eta} a_{\varepsilon,K,\eta}$.

Step 3. Ergodicity We apply Theorem 5.1 from [12]. All we have to show is that their condition $(\mathcal{S})$ holds, namely that all compacts sets are petite for some skeleton chain. In fact in Step 2 of the proof, we have established that B is petite for a skeleton chain. It is easy to verify that the same proof works when replacing B by an arbitrary compact set. □

3.2 Tightness of the invariant probability measure of X^ε_t

It follows from Theorem 3 that for any $\varepsilon \leq \varepsilon_0$, X^ε_t possesses a unique invariant probability measure μ_ε. The aim of this subsection is to prove

Theorem 4. *For any sequence $\varepsilon_n \downarrow 0$, the sequence of invariant measures $\{\mu_{\varepsilon_n},\ n \geq 1\}$ is tight.*

The result will follow from the following statement

Proposition 3. *There exist two constants $a, C' > 0$ such that for any $M > 0$, if $\varepsilon \leq \frac{a}{M} \wedge \frac{1}{4}$,*

$$\int_{\mathbb{R}^d} (\|x\| \wedge M) \mu^\varepsilon(dx) \leq C'.$$

Let us first show how Theorem 4 follows from Proposition 3.

PROOF OF THEOREM 4. We deduce from the above Proposition that whenever $\varepsilon \le \frac{a}{M} \wedge \frac{1}{4}$,

$$\mu^\varepsilon(\|x\| > M) \le \frac{C'}{M}.$$

Fix $\delta > 0$ arbitrarily small. Let us from now on fix $M \ge C'/\delta$. Let n_0 be such that $\varepsilon_{n_0} \le \frac{a}{M} \wedge \frac{1}{4}$. It follows from the above that for any $n \ge n_0$,

$$\mu_{\varepsilon_n}(\|x\| > M) \le \delta.$$

It is finally easy to find $M' \ge M$ such that

$$\mu_{\varepsilon_n}(\|x\| > M') \le \delta$$

for any $1 \le n \le n_0$, hence the result. $\square$

Now return to the

PROOF OF PROPOSITION 3. It follows readily from Lemma 4 that whenever $X_0^\varepsilon = x_0^\varepsilon$ is deterministic, $M < a/\varepsilon$,

$$\begin{aligned}\mathbb{E}(\|X_t^\varepsilon\|) &\le \|x_0^\varepsilon\| + Ct - c\int_0^t \mathbb{E}\left[\|X_s^\varepsilon\| \wedge (a/\varepsilon)\right] ds \\ &\le \|x_0^\varepsilon\| + Ct - c\int_0^t \mathbb{E}\left[\|X_s^\varepsilon\| \wedge M\right] ds.\end{aligned}$$

Since the function $f(x) = \|x\| \wedge M$ is bounded and continuous, it follows from the last part of Theorem 3 that $\mathbb{E}\left[\|X_t^\varepsilon\| \wedge M\right] \to \int_{\mathbb{R}^d}(\|x\| \wedge M)\mu^\varepsilon(dx)$ as $t \to \infty$. Hence, since the first part of the proof implies that

$$\frac{c}{t}\int_0^t \mathbb{E}\left[\|X_s^\varepsilon\| \wedge M\right] ds \le \frac{\|x_0^\varepsilon\|}{t} + C$$

for all $t > 0$, the result follows by letting $t \to \infty$. $\square$

3.3 Asymptotic analysis of the large time behavior of X_t^ε

We now want to analyze the large time behavior of X_t^ε, for small ε. We first show

Theorem 5. *As $\varepsilon \to 0$, $\mu^\varepsilon \Rightarrow \delta_{\bar{x}_\infty}$.*

Proof We consider the X_t^ε equation started with $X_0^\varepsilon \sim \mu^\varepsilon$. Choose a sequence ε_n which converges to 0 as $n \to \infty$. Since the collection $\{\mu^{\varepsilon_n}\}_{n\geq 1}$ is tight, along a subsequence still denoted the same way, $\mu^{\varepsilon_n} \Rightarrow \mu^0$. It follows from Theorem 1 (in fact its extension explained in Remark 2) that $X_t^{\varepsilon_n} \Rightarrow \overline{X}_t$, where $\overline{X}_t$ solves the ODE

$$\frac{d\overline{X}_t}{dt} = L\overline{X}_t - v, \ \overline{X}_0 \sim \mu^0.$$

But for any $f \in C_b(\mathbb{R}^d)$, $n \geq 1$, $t \to \mathbb{E}f(X_t^{\varepsilon_n})$ is a constant, so this is true in the limit, which implies that $\mu^0 = \delta_{\overline{x}_\infty}$. This shows that the whole collection $\mu^\varepsilon \Rightarrow \delta_{\overline{x}_\infty}$ as $\varepsilon \to 0$. $\square$

We expect from the above results that for small enough ε, the invariant measure μ^ε is close to ν^ε, which is the law of $\bar{x}_\infty + \sqrt{\varepsilon}\xi$, where $\xi \sim \mathcal{N}(0, \bar{S}^2)$ (recall (32)). If we interpret μ^ε as the mass of X_t^ε for large t, this is not really correct, since for some large t, X_t^ε will make a large deviation from $\bar{x}_t$, see [4].

Therefore, we prefer to interpret $\mu^\varepsilon(A)$ as

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t \mathbf{1}_A(X_s^\varepsilon)ds.$$

We now want to give a precise description of μ^ε for small $\varepsilon > 0$. For that sake, let us introduce some notation. For any Borel set A and for all $\varepsilon > 0$, let $A_\varepsilon := \bar{x}_\infty + \sqrt{\varepsilon}A$ and denote by $\lambda_{\bar{S}^2} = \mathcal{N}(0, \bar{S}^2)$ the invariant Gaussian distribution of the Ornstein-Uhlenbeck process U_t.

Theorem 6. *Consider the process X_t^ε, starting at time $t = 0$ from $X_0^\varepsilon = \bar{x}_\infty$. Let A be an arbitrary element of $\mathcal{B}_d$ such that its boundary ∂A has zero d–dimensional Lebesgue measure. For any $\delta > 0$ there exist $t_\delta > 0$ large enough such that for any $t \geq t_\delta$, there exists $\varepsilon_{t,\delta} > 0$ such that for all $\varepsilon \leq \varepsilon_{t,\delta}$, with a probability larger than $1-\delta$, the fraction of the time in the interval $[0,t]$ which X_s^ε spends in the set $A_\varepsilon = \bar{x}_\infty + \sqrt{\varepsilon}A$ belongs to the interval $[\lambda_{\bar{S}^2}(A) - \delta, \lambda_{\bar{S}^2}(A) + \delta]$.*

Proof. As $\varepsilon \to 0$, for any fixed $t > 0$, U_t^ε being defined by (26),

$$\frac{1}{t}\int_0^t \mathbf{1}_{A_\varepsilon}(X_s^\varepsilon)ds = \frac{1}{t}\int_0^t \mathbf{1}_A(U_s^\varepsilon)ds \Rightarrow \frac{1}{t}\int_0^t \mathbf{1}_A(U_s)ds,$$

and

$$\frac{1}{t}\int_0^t \mathbf{1}_A(U_s)ds \to \lambda_{\bar{S}^2}(A) \text{ in probability, as } t \to \infty.$$

Hence for all $\delta > 0$, there exists $t_\delta > 0$ such that for any $t \geq t_\delta$,

$$\mathbb{P}\left(\left|\frac{1}{t}\int_0^t \mathbf{1}_A(U_s)ds - \lambda_{\bar{S}^2}(A)\right| \geq \delta\right) \leq \delta/2.$$

From the Portmanteau theorem,

$$\limsup_{\varepsilon\to 0} \mathbb{P}\left(\left|\frac{1}{t}\int_0^t \mathbf{1}_A(U_s^\varepsilon)ds - \lambda_{\bar{S}^2}(A)\right| \geq \delta\right) \leq \delta/2.$$

Hence there exists $\varepsilon_{t,\delta}$ such that for any $0 < \varepsilon \leq \varepsilon_{t,\delta}$,

$$\mathbb{P}\left(\left|\frac{1}{t}\int_0^t \mathbf{1}_A(U_s^\varepsilon)ds - \lambda_{\bar{S}^2}(A)\right| > \delta\right) \leq \delta.$$

Recall that

$$\mathbf{1}_A(U_s^\varepsilon) = \begin{cases} 1, & \text{if } X_s^\varepsilon \in \bar{x}_\infty + \sqrt{\varepsilon}A, \\ 0, & \text{otherwise,} \end{cases}$$

implying the result. □

Acknowledgements The support of CNRS – EGP/Mission pour l'interdisciplinarité, within the project Évolution et génétique des populations of G. Didier, is gratefully acknowledged.
The authors want to express their gratitude to Micheal Kopp, who introduced them to the model and suggested the study of the small jumps limit, to Nicolas Champagnat, whose clever question during the thesis defense of the first author, led them to the developments of the above section 3, and to two anonymous Referees, whose very detailed and relevant comments helped them to improve significantly the first version of this paper.

References

[1] D. Aldous, Stopping times and tightness, *Ann. Probab.* **6**, 335–340, 1978.

[2] N. Champagnat and S. Méléard, Polymorphic evolution sequence and evolutionary branching. *Probability Theory and Related Fields* **151**, 45–94, 2011.

[3] U. Dieckmann and R. Law, The dynamical theorey of coevolution: a derivation from stochastic ecological processes, *J. Math. Biol.* **34**, 579–612, 1996.

[4] M. Freidlin and A. Wentzell, *Random Perturbations of Dynamical Systems.* Springer, Berlin, Heidelberg, 2012.

[5] J. B. S. Haldane, A mathematical theory of natural and artificial selection. Part V: selection and mutation, *Proc. Camb. Philos. Soc.* **23**, 838âĂŞ 844, 1927.

[6] I. Karatzas and S.E. Shreve, *Brownian motion and stochastic calculus* 2nd ed., Spinger, 1991.

[7] M. Kimura, On the probability of fixation of mutant genes in a population, *Genetics* **47** p. 713, 1962.

[8] M. Kopp and J. Hermisson The genetic basis of phenotypic adaptation II: The distribution of adaptive substitutions in the moving optimum model. *Genetics* **183**, 1453–1476, 2009.

[9] M. Kopp , E. Nassar and E. Pardoux, Phenotypic lag and population extinction in the moving–optimum model : insights from a small–jumps limit, submitted, 2018.

[10] G. Malécot, Les processus stochastiques et la méthode des fonctions génératrices ou caractéristiques. *Publications de l'Institut de Statistiques de l'Université de Paris* **1**, 1–16, 1952.

[11] S. Matuszewski, J. Hermisson, and M. Kopp, Fisher's geometric model with a moving-optimum, *Evolution*, **68**, 2571–2588, 2014.

[12] S. P. Meyn and R. L. Tweedie, Stability of Markovian processes III: Foster- Lyapunov criteria for continuous-time processes, *Advances in Applied Probability*, **25**, no. 3, pp. 518–548, 1993.

[13] E. Nassar and E. Pardoux, On the long time behavior of an SDE driven by a Poisson point process, *Adv. in Appl. Probab.* **49**, 344–367, 2017.

[14] E. Pardoux, *Probabilistic Models of Population Evolution.* Springer, 2016.

[15] E. Pardoux, A. Răşcanu, *Stochastic Differential Equations, Backward SDEs, Partial Differential Equations*, Stochastic Modelling and Applied Probability **69**, 2014.

[16] S. Peischl and M. Kirkpatrick, Establishment of new mutations in changing environments, *Genetics*, 2012.

[17] P. E. Protter, *Stochastic Integration and Differential Equations.* Heidelberg: Springer-Verlag, 2005.

[18] D.W. Stroock and S.R.S. Varadhan, *Multidimensional Diffusion Proceses*, Classics in Mahematics, Springer, 2006.

[19] H. Uecker and J. Hermisson, On the fixation process of a beneficial mutation in a variable environment, *Genetics* **188**, 915–930, 2011.