PERSISTENCE OF SMALL NOISE AND RANDOM INITIAL CONDITIONS

BY J. BAKER, P. CHIGANSKY, K. HAMZA AND F. C. KLEBANER

Abstract

The effect of small noise in a smooth dynamical system is negligible on any finite time interval; in this paper we study situations where the effect persists on intervals increasing to ∞. Such an asymptotic regime occurs when the system starts from an initial condition that is sufficiently close to an unstable fixed point. In this case, under appropriate scaling, the trajectory converges to a solution of the unperturbed system started from a certain *random* initial condition. In this paper we consider the case of one-dimensional diffusions on the positive half-line; this case often arises as a scaling limit in population dynamics.

1. Introduction

In this paper we study a new approximation for a stochastic process that is generated by a nonlinear dynamical system started in the vicinity of an unstable fixed point. The processes we consider can be represented as deterministic dynamics perturbed by small noise. The well-known results of Kurtz [9] in the context of Markov jump processes or of Freidlin and Wentzell [4] in the context of diffusions state that the effect of some small noise is negligible on any finite time interval. This is known as the fluid limit approximation, which implies that, under appropriate conditions, the small noise limit of the stochastic process solves the approximating ordinary differential equation.

We are interested in situations for which the stochastic process starts near an unstable fixed point, such as 0, in which case the usual fluid approximation results in 0, and thus is uninformative. To give a more adequate approximation, we let intervals increase to ∞. This yields a fluid approximation with a *random* initial condition. The randomness comes via some random variable W, itself being the limit of the scaled linearized stochastic system, now as time goes to ∞. The new initial condition is given by a suitable transformation of W, being derived from the nonlinear deterministic dynamics in the fluid approximation. Such an approximation with a random initial condition is given in [1] for some nonlinear birth–death processes and in [3] for a discrete-time model of the polymerase chain reaction.

In this paper we consider the case of one-dimensional diffusions. Such processes arise, for example, as approximations to the Wright–Fisher model from evolutionary biology. Recently, heuristics for random initial conditions were given in [10], arguing for gamma-type distributions, that is, a sum of independent exponentials. Our theory yields a random Poisson sum of similar exponentials.

The relevant approximation philosophy can be traced back to the work of Kendall [6] and Whittle [13] in the context of a Markovian SIR epidemic process. It was observed that such processes should behave much like a linear branching process near the unstable fixed point and elsewhere should look more like a solution of the deterministic equations. However, it emerges that rigorous implementation of the Kendall–Whittle heuristics is a nontrivial matter: it requires quite different and model-specific techniques.

The main difficulty while working with dynamical systems on increasing time intervals is that Grönwall's inequality, which is the most common tool in this kind of analysis, ceases to be effective. A more efficient approximation can be constructed by means of a two-stage procedure. The whole time interval is split into two parts, on each of which the solution of the perturbed system is approximated in different ways: first it is coupled to linearized stochastic dynamics and then to its extrapolation along the deterministic flow that is generated by the unperturbed system.

The key ingredient of the method is a coupling that is tailored to the particular type of dynamics on a case-by-case basis. For the density-dependent birth–death processes as in [1], this can be done via an auxiliary collection of Poisson processes, properly synchronized with the jumps of the original system. For one-dimensional diffusions, this approach is no longer feasible; instead, we realize the coupling by using a Yamada–Watanabe type of approximation by a linear process, driven by the same Brownian motion.

2. The main result

Consider the stochastic differential equation (SDE) on $\mathbb{R}_+$

$$\mathrm{d}X_t^\varepsilon = f(X_t^\varepsilon)\,\mathrm{d}t + \sqrt{\varepsilon\sigma(X_t^\varepsilon)}\,\mathrm{d}B_t, \qquad t \geq 0, \tag{2.1}$$

where B_t is standard Brownian motion, $\varepsilon > 0$ is a small parameter, and the functions f and σ satisfy the following assumptions.

(A$_1$) The functions $f\colon \mathbb{R}_+ \mapsto \mathbb{R}$ and $\sigma\colon \mathbb{R}_+ \mapsto \mathbb{R}_+$ are twice continuously differentiable with $f(0) = \sigma(0) = 0$, $f'(0) > 0$, $\sigma'(0) > 0$, and bounded second derivatives.

(A$_2$) $\sigma(\cdot)$ and its derivative are bounded.

(A$_3$) $f(\cdot)$ satisfies the drift condition $(y - x)(f(y) - f(x)) \leq f'(0)(y - x)^2$ for $x, y \in \mathbb{R}_+$.

Assumption (A$_1$) makes 0 an unstable fixed point of both (2.1) and the ordinary differential equation (ODE) obtained by removing the stochastic perturbation in (2.1),

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = f(x_t), \qquad t \geq 0. \tag{2.2}$$

Smoothness of the coefficients and the drift condition (A$_3$) are sufficient for the existence of a unique strong solution of (2.1) for any initial point $X_0^\varepsilon \in \mathbb{R}_+$ (see, e.g. [7]). Similarly, the deterministic equation (2.2) admits a unique continuous solution subject to any $x_0 \in \mathbb{R}_+$. Henceforth, $\phi_t(x)$ denotes the flow generated by (2.2).

The theory of small random perturbations (see, e.g. [4]) asserts that the effect of noise on any *fixed* time interval $[0, T]$ is negligible as $\varepsilon \to 0$.

Theorem 2.1. *Let X_t^ε satisfy (2.1) with (A$_1$)–(A$_3$) and $X_0^\varepsilon \xrightarrow{\mathbb{P}} x_0 \in \mathbb{R}_+$ as $\varepsilon \to 0$. Then, for any T,*

$$\sup_{t \leq T} |X_t^\varepsilon - x_t| \xrightarrow{\mathbb{P}} 0 \quad as \; \varepsilon \to 0,$$

where x_t is the solution of (2.2) subject to the initial condition x_0.

Since 0 is a fixed point of the limit dynamics in (2.2), this theorem implies that the solution of (2.1), started from a small positive initial condition $X_0^\varepsilon = \varepsilon > 0$, converges to 0 on any fixed bounded interval, i.e. for every finite $T \geq 0$,

$$\sup_{t \leq T} |X_t^\varepsilon| \overset{\mathbb{P}}{\to} 0 \quad \text{as } \varepsilon \to 0.$$

On the other hand, since the fixed point is unstable and the initial condition is nonzero, with positive probability, the trajectory X_t^ε may be pushed out of the vicinity of the origin and, after a sufficiently large period of time, may reach a significant magnitude. Such behaviour is missing from Theorem 2.1, but is captured in the following result.

Theorem 2.2. *Let X_t^ε satisfy (2.1) with (A₁)–(A₃) subject to $X_0^\varepsilon = \varepsilon \in (0, 1)$; define $T_\varepsilon :=$ $\log(1/\varepsilon)/f'(0)$. Then, for any $T > 0$,*

$$\sup_{t \in [0,T]} |X_{T_\varepsilon + t}^\varepsilon - x_t| \overset{\mathbb{P}}{\to} 0 \quad as \ \varepsilon \to 0, \tag{2.3}$$

where x_t is the solution of (2.2) subject to the initial condition $x_0 = H(W)$. The function $H(x)$ and the random variable W are respectively the limit of the scaled flow

$$H(x) = \lim_{t \to \infty} \phi_t(x e^{-f'(0)t}), \qquad x \geq 0, \tag{2.4}$$

and the almost-sure martingale limit

$$W := \lim_{t \to \infty} e^{-f'(0)t} Y_t$$

of the solution to the linearized system

$$Y_t = 1 + \int_0^t f'(0) Y_s \, ds + \int_0^t \sqrt{\sigma'(0) Y_s} \, dB_s. \tag{2.5}$$

Remark 2.1. (a) Both the random variable W and the function $H(\cdot)$, arising in the limit, admit explicit characterizations. As recalled in Section 4.2 below, W has the compound Poisson distribution with rate $2a$ and exponentially distributed jumps with mean $1/(2a)$. The function $H(\cdot)$ admits the explicit expression (4.2), elaborated in Proposition 4.1 below.

(b) Approximation (2.3) reveals that the trajectory of (2.1), when started near the unstable fixed point and observed at a suitably increasing sequence of times, converges to the solution of the same deterministic system (2.2) as in Theorem 2.1, but now it has a random initial condition. Note that $\mathbb{P}(x_0 = 0) = \mathbb{P}(W = 0) = e^{-2a} > 0$ and, hence, the limiting trajectory can be 0 with probability greater than 0. This corresponds to the event on which the process X_t^ε is absorbed at 0 in some finite time. On the event $\{W > 0\}$, the trajectories converge to a nontrivial curve whose initial point is random. This type of randomness has been observed in biological models of sweeps (see, e.g. [10]).

Example 2.1. (*Wright–Fisher diffusion with selection.*) The SDE

$$dX_t^\varepsilon = a X_t^\varepsilon (1 - X_t^\varepsilon) \, dt + \sqrt{\varepsilon} \sqrt{X_t^\varepsilon (1 - X_t^\varepsilon)} \, dB_t$$

fits the above framework with $f(x) = ax(1-x)$ and $\sigma(x) = x(1-x)$. This equation describes the evolution of an allele frequency in a population and is known as the Wright–Fisher diffusion with selection coefficient a. It has two absorbing states, 0 and 1, so that, for $X_0^\varepsilon \in [0, 1]$, the trajectory is confined to the unit interval, $X_t^\varepsilon \in [0, 1]$ for all $t \geq 0$. In particular, all the

assumptions of Theorem 2.2 are satisfied: $f(\cdot)$ and $\sigma(\cdot)$ can be defined arbitrarily outside the state space $[0, 1]$ and, hence, their derivatives can be chosen to be bounded, etc., so that assumptions (A_1)–(A_3) are met.

Using the expression for the deterministic flow

$$\phi_t(x) = \frac{x\mathrm{e}^{at}}{1 - x + x\mathrm{e}^{at}},$$

it follows that

$$H(x) = \lim_{t \to \infty} \phi_t(x\mathrm{e}^{-at}) = \frac{x}{1 + x}.$$

Hence, the random initial condition is given by

$$x_0 = \frac{W}{W + 1}.$$

Example 2.2. (*Balancing selection model.*) The SDE

$$\mathrm{d}X_t^\varepsilon = aX_t^\varepsilon(1 - X_t^\varepsilon)(1 - 2X_t^\varepsilon)\,\mathrm{d}t + \sqrt{\varepsilon}\sqrt{X_t^\varepsilon(1 - X_t^\varepsilon)}\,\mathrm{d}B_t$$

fits the above framework with $f(x) = ax(1 - x)(1 - 2x)$ and $\sigma(x) = x(1 - x)$. The fluid limit is given by the ODE

$$\frac{\mathrm{d}x_t}{\mathrm{d}t} = ax_t(1 - x_t)(1 - 2x_t), \qquad t \geq 0,$$

which generates the flow

$$\phi_t(x) = \frac{1}{2}\left(1 - \frac{1 - 2x}{\sqrt{4x(1 - x)(\mathrm{e}^{at} - 1) + 1}}\right), \qquad x \in \left(0, \tfrac{1}{2}\right).$$

It now follows that

$$H(x) = \lim_{t \to \infty} \phi_t(x\mathrm{e}^{-at}) = \frac{1}{2}\left(1 - \frac{1}{\sqrt{4x + 1}}\right).$$

Hence, the random initial condition in Theorem 2.2 for this model is given by

$$x_0 = \frac{1}{2}\left(1 - \frac{1}{\sqrt{4W + 1}}\right).$$

3. On some recent related work

In this section we describe some recent research on approximations to population processes that start near an unstable fixed point of the approximating system.

3.1. Random time shift

Recent work in [2] contains an important step in the rigorous realization of the ideas of Kendall [6] and Whittle [13]; its setting is of nonlinear birth–death processes. The essence of this work is that the stochastic system is similar to the approximating deterministic system except for a *random time shift*. To explain the setting, consider the 'bare bones' evolution model of Klebaner *et al.* [8], in which a mutation appears in an established population, after which the two subpopulations compete. This is treated as a pure-jump Markov process $Z^K(t)$ on $\mathbb{Z}_+^2$, in

which process the first component counts wild-type individuals, initially around their carrying capacity, and the second counts mutant individuals. The transition rates are as follows:

$$
\begin{aligned}
z &\to z + (1, 0) &&\text{at rate} && a_1 z_1, \\
z &\to z + (-1, 0) &&\text{at rate} && z_1((z_1/K) + \gamma(z_2/K)), \\
z &\to z + (0, 1) &&\text{at rate} && a_2 z_2, \\
z &\to z + (0, -1) &&\text{at rate} && z_2(\gamma(z_1/K) + (z_2/K));
\end{aligned}
$$

here K is a large parameter that controls the carrying capacities of the populations, and a_1, a_2, and γ are positive constants.

Initially, the first component $Z_1^K(0)$ has a value near its carrying capacity $a_1 K$ and no mutants are present in the population, so $Z_2^K(0-) = 0$. At time 0 a number of mutant individuals are introduced, so $Z_2^K(0+)$ is a (small) positive integer, where 'small' means much less than K. The mutants and wild-type individuals differ only through their birth rates a_1 and a_2. Each species has *per capita* death rate given by the density of its own population, together with an additional component of γ times the density of individuals of the other species. If $\gamma > 1$, members of the other species result in a higher mortality rate than if they were of the same species; if $\gamma < 1$, they result in a lower mortality rate, favouring the possibility of coexistence. If $a_2 < \gamma a_1$, the mutants have negligible chance of survival, but, if $a_2 > \gamma a_1$, there is a positive probability that the mutant strain may become established. In this case, if $a_1 > \gamma a_2$ also, the two populations eventually come to coexist.

Define the density process $X^K(t) = Z^K(t)/K$, and let

$$
f(x) = \begin{pmatrix} a_1 x_1 - x_1^2 - \gamma x_1 x_2 \\ a_2 x_2 - x_2^2 - \gamma x_1 x_2 \end{pmatrix}.
$$

If $X^K(0) \to x(0)$ as $K \to \infty$ then, by Theorem 3.1 of [9], for $T > 0$ and as $K \to \infty$,

$$
\sup_{0 \le t \le T} \| X^K(t) - x(t) \| \xrightarrow{\mathbb{P}} 0,
$$

where $x(t)$ solves ODE (2.2) subject to the initial condition $x(0) = (a_1, 0)$. This is an unstable fixed point for the deterministic dynamics above. Hence, the limit trajectory is constant and equal to $x(0)$ on any finite interval $[0, T]$.

The heuristic reasoning for the approximation in [2] is as follows. Initially, X^K is near $(a_1, 0)$ and the mutant component $Z_2^K(t)$ is approximated by a linear birth–death process $Y(t)$ with *per capita* birth rate a_2, death rate γa_1, and corresponding survival rate $\beta = a_2 - \gamma a_1$; also, $Y(t)$ starts from $Y(0) := Z_2^K(0)$. Since, for a birth–death process Y, the process $Y(t)e^{-\beta t}$ is a martingale with a nondegenerate limit W, say, we then argue that

$$
X_2^K(t) = \frac{Z_2^K(t)}{K} \approx \frac{Y(t)}{K} \approx \frac{1}{K} e^{\beta t} W = \frac{1}{K} e^{\beta(t + \beta^{-1} \log W)}. \tag{3.1}
$$

On the other hand, the dynamics at (2.2) after linearization are governed by $f(\cdot)$ near the fixed point $(a_1, 0)$, leading to $\dot{x}(t) \approx Bx(t)$, where the Jacobian matrix

$$
B = \begin{pmatrix} -a_1 & -\gamma a_1 \\ 0 & a_2 - \gamma a_1 \end{pmatrix}.
$$

Hence, $x(t) \approx e^{Bt} x(0)$ and

$$x_2(t) \approx \frac{1}{K} e^{\beta t} Z_2^K(0) = \frac{1}{K} e^{\beta(t+\beta^{-1} \log Z_2^K(0))} = \frac{1}{K} e^{\beta(t+\beta^{-1} \log \mathbb{E}W)}, \qquad (3.2)$$

where we have used the martingale property $Y(0) = \mathbb{E}Y(t)e^{-\beta t} = \mathbb{E}W$. Comparing the two approximations in (3.1) and (3.2), we conclude that the stochastic and deterministic paths differ by the random time shift $\beta^{-1}(\log W - \log \mathbb{E}W)$. These heuristics are made precise in [2, Theorem 1.2], where, for a certain class of stochastic systems, including the 'bare bones' model above as a special case, there is derived a nonasymptotic approximate solution in terms of trajectories of the corresponding ODE shifted by the random quantity $\log W - \log \mathbb{E}W$.

3.2. Random initial condition

Fluid approximation on increasing time intervals also leads to differential equations with *random initial conditions*. This phenomenon was recently studied in [1] in the context of density-dependent populations. Let Z_t^K be a $\mathbb{Z}_+$-valued continuous-time birth–death process with respective *per capita* birth and death rates

$$\lambda - (\lambda - \mu) g_1 \left(\frac{z}{K} \right) \quad \text{and} \quad \mu + (\lambda - \mu) g_2 \left(\frac{z}{K} \right),$$

where $\lambda > \mu \geq 0$ are fixed constants, and $g(\cdot) = g_1(\cdot) + g_2(\cdot)$ is an increasing function with $g(0) = 0$ and $g(x_\infty) = 1$ for some $x_\infty \in (0, \infty)$. As before, K is a parameter, thought of as being large, that is representative of the carrying capacity of the population.

The process Z_t^K admits the decomposition

$$Z_t^K = Z_0^K + (\lambda - \mu) \int_0^t Z_s^K \left(1 - g\left(\frac{Z_s^K}{K} \right) \right) \mathrm{d}s + M_t, \qquad t \geq 0, \qquad (3.3)$$

where M_t is a martingale with predictable quadratic variation

$$\langle M \rangle_t = \int_0^t Z_s^K \left(\lambda + \mu - (\lambda - \mu) \left[g_2\left(\frac{Z_s^K}{K} \right) - g_1\left(\frac{Z_s^K}{K} \right) \right] \right) \mathrm{d}s.$$

On dividing both asides of (3.3) by K, the density process $X_t^K := Z_t^K / K$ satisfies

$$X_t^K = X_0^K + (\lambda - \mu) \int_0^t X_s^K (1 - g(X_s^K)) \, \mathrm{d}s + \frac{\tilde{M}_t}{\sqrt{K}}, \qquad t \geq 0,$$

where $\langle \tilde{M}_t \rangle := \langle M_t \sqrt{K} \rangle$ depends only on X_t^K. Then, under appropriate technical conditions, Theorem 3.1 of [9] implies that the density process X_t^K converges as $K \to \infty$ to the solution of the ODE

$$\frac{\mathrm{d}}{\mathrm{d}t} x_t = (\lambda - \mu) x_t (1 - g(x_t)), \qquad t \geq 0, \qquad (3.4)$$

subject to $x_0 = \lim_{K \to \infty} X_0^K$. Thus, if the initial population size Z_0^K is constant with respect to K and, hence, the initial limit density $x_0 = 0$, then we obtain the trivial limit $x_t = 0$ for $t \in [0, T]$.

The main result in [1] asserts that, for any $T > 0$ and $T_K := (\log K)/(\lambda - \mu)$, as $K \to \infty$,

$$\sup_{t \leq T} |X_{t+T_K}^K - x_t| \overset{\mathbb{P}}{\to} 0,$$

where x_t solves (3.4) subject to the random initial condition $x_0 := G^{-1}(\log W)$ with

$$G(x) := \int_0^x \frac{g(u)}{u[1-g(u)]} \, du + \log x, \qquad x \in \mathbb{R}_+, \tag{3.5}$$

and the random variable W is the almost-sure limit $\lim_{t\to\infty} W_t$ of the martingale

$$W_t := e^{-(\lambda-\mu)t} Y_t,$$

where Y_t is the continuous-time Galton–Watson branching process with constant per capita rates λ and μ, suitably defined on the same probability space.

3.3. An application to a polymerase chain reaction

So far, theoretical considerations have been our main motivation, but fluid approximations on growing intervals can also be of practical interest. One example is an application to a polymerase chain reaction (PCR) suggested in [3]. The model contains the Michaelis–Menten constant K which is large in terms of the number of molecules. A PCR typically starts from a very low concentration of initial DNA strands which are not observable at first, but become so after a number of iterations. This seems to be consistent with the theory, proposed in [3], which predicts that DNA concentration rises to a measurable level after the number of repetitions is of order $\log K$. Once the reaction becomes observable, the analytic approximation features random initial conditions, and these can have practical implications. Since this model is in discrete time and uses discrete-time techniques, we omit further details here.

3.4. Contribution of this paper

As mentioned in the introduction, the key element of our program is the realization of a coupling to the linear stochastic process. While we use the broad term 'coupling', it can mean different things in different situations. For the multi-dimensional birth–death processes of [2], the coupling is done by applying an abstract general result of Thorisson [12]. The nonlinear and linear processes are coupled trajectory-wise of full probability on the limit set. This technique relies on our ability to evaluate the total variation distance between the nonlinear and linear processes. For the one-dimensional birth–death processes of [1], the coupling is done by constructing two linear processes with constant rates such that the nonlinear process is sandwiched between them. This is done in such a way that in the limit both linear processes, and hence also the nonlinear process, converge to the same limit under appropriate scaling.

The results of the present paper are closer in spirit to those of [1]. Here coupling uses the same driving Brownian motion for both the original nonlinear process and its approximation with linear drift, and the latter is a Feller branching diffusion (2.5). To show that these processes are close, we use a smooth approximation to the absolute value function, akin to the Yamada–Watanabe approach to the analysis of one-dimensional diffusions with nonsmooth coefficients.

Our other contribution is of a more conceptual flavour: it is the identification of the nonlinearity $H(\cdot)$ as the limit of the scaled flow (2.4), generated by the differential equation in the fluid approximation

$$H(x) = \lim_{t\to\infty} \phi_t(x e^{-f'(0)t}).$$

In the one-dimensional case this limit admits the closed-form expression $H(x) = G^{-1}(\log x)$, where $G(\cdot)$ is defined in (4.2) below (cf. (3.5)). We conjecture that this new characterization remains valid in higher-dimensional models, such as that considered in [2], and as such it may prove to be useful in further exploration of the subject.

4. Proof of Theorem 2.2

Without loss of generality, we fix the normalization $\sigma'(0) = 1$ and set $a := f'(0) > 0$. The main step in the proof is to establish the convergence in (2.3) at $t = 0$, namely,

$$X_{T_\varepsilon}^\varepsilon \xrightarrow{\mathbb{P}} H(W) \quad \text{as } \varepsilon \to 0. \tag{4.1}$$

The rest of the proof then follows by a change of time scale. Indeed, on writing $\tilde{X}_t^\varepsilon = X_{T_\varepsilon + t}^\varepsilon$ and $\tilde{B}_t = B_{T_\varepsilon + t} - B_{T_\varepsilon}$, from (2.1) we obtain

$$\tilde{X}_t^\varepsilon = \tilde{X}_0^\varepsilon + \int_0^t f(\tilde{X}_s^\varepsilon) \, \mathrm{d}s + \int_0^t \sqrt{\varepsilon \sigma(\tilde{X}_s^\varepsilon)} \, \mathrm{d}\tilde{B}_s,$$

and the result follows from (4.1) by Theorem 2.1.

The proof of (4.1) consists of a number of steps given as propositions in the rest of the paper. First we establish the existence of a nontrivial limit function $H(\cdot)$ that appears in the random initial condition. Next we consider an auxiliary Feller branching diffusion and its martingale limit W. Then we show the convergence of processes on finite intervals under appropriate rescaling. Finally, all these ingredients are assembled together to construct the main approximation, yielding the statement of Theorem 2.2.

4.1. The function $H(\cdot)$

Proposition 4.1. *The limit in (2.4) exists, uniformly on compact sets. It is given by $H(x) = G^{-1}((\log x)/a)$, where*

$$G(x) := \int_0^x \left(\frac{1}{f(u)} - \frac{1}{au} \right) \mathrm{d}u + \frac{\log x}{a}. \tag{4.2}$$

Proof. The function f is continuously differentiable by (A$_1$), so the flow $\phi_t(x)$ is differentiable in both variables and the derivative $\phi_t'(x) := \partial_x \phi_t(x)$, $x > 0$, satisfies both $\phi_0'(x) = 1$ and

$$\frac{\mathrm{d}}{\mathrm{d}t} \phi_t'(x) = f'(\phi_t(x)) \phi_t'(x), \qquad t \geq 0.$$

Let

$$x^* = \begin{cases} \infty & \text{if } f(x) > 0 \text{ for all } x > 0, \\ \inf\{x > 0 \colon f(x) = 0\} & \text{otherwise.} \end{cases}$$

Since the interval $(0, x^*)$ is invariant under the flow, $f(\phi_t(x)) > 0$, and

$$\phi_t'(x) = \exp\left(\int_0^t f'(\phi_s(x)) \, \mathrm{d}s \right) = \exp\left(\int_0^t \frac{f'(\phi_s(x))}{f(\phi_s(x))} \, \mathrm{d}\phi_s(x) \right) = \frac{f(\phi_t(x))}{f(x)}.$$

Define $h_t = \phi_t(x\mathrm{e}^{-at})$. Then h_t satisfies

$$
\begin{aligned}
h_t' = \frac{\mathrm{d}}{\mathrm{d}t} \phi_t(x\mathrm{e}^{-at}) &= f(h_t) - ax\mathrm{e}^{-at} \phi_t'(x\mathrm{e}^{-at}) \\
&= f(h_t) - ax\mathrm{e}^{-at} \frac{f(\phi_t(x\mathrm{e}^{-at}))}{f(x\mathrm{e}^{-at})} \\
&= f(h_t) \left[1 - \frac{ax\mathrm{e}^{-at}}{f(x\mathrm{e}^{-at})} \right].
\end{aligned}
$$

Let $t_0 \geq 0$ be any point such that $xe^{-at_0} \in (0, x^*)$; then, by rearranging and integrating,

$$\int_{t_0}^{t} \frac{dh_s}{f(h_s)} = \int_{t_0}^{t} \left(1 - a\frac{xe^{-as}}{f(xe^{-as})} \right) ds = \int_{xe^{-at}}^{xe^{-at_0}} \left(\frac{1}{au} - \frac{1}{f(u)} \right) du, \qquad t \geq t_0.$$

Now $f(\cdot)$ has a bounded second derivative, so, for all sufficiently small $x > 0$,

$$\begin{aligned}
\left| \frac{1}{ax} - \frac{1}{f(x)} \right| &= \left| \frac{f(x) - ax}{axf(x)} \right| \\
&= \frac{1}{axf(x)} \left| \int_0^x \int_0^u f''(v) \, dv \, du \right| \\
&\leq \frac{\|f''\|_\infty}{a} \frac{x}{f(x)} \\
&\leq \frac{\|f''\|_\infty}{a} \frac{x}{ax - x^2\|f''\|_\infty} \\
&\leq C
\end{aligned}$$

for some constant C. In particular, the function $x \mapsto 1/f(x) - 1/(ax)$ is integrable at 0 and we can define $G(x)$ as in (4.2) for $x \in (0, x^*)$. This function is continuous and strictly increasing because $G'(x) = 1/f(x) > 0$ for $x \in (0, x^*)$ and

$$G(h_t) - G(h_{t_0}) = \int_{xe^{-at}}^{xe^{-at_0}} \left(\frac{1}{au} - \frac{1}{f(u)} \right) du. \qquad (4.3)$$

For any fixed $c \in (0, x^*)$, we can write

$$G(x) = \int_0^c \left(\frac{1}{f(u)} - \frac{1}{au} \right) du + \frac{1}{a} \log c + \int_c^x \frac{1}{f(u)} du.$$

Hence, for $x^* \in (0, \infty)$, $\lim_{x \to x^*} G(x) = \infty$ because f has a bounded second derivative. Note that drift condition (A$_3$) implies that $f(x) \leq ax$ for all $x \geq 0$ and, hence, $\lim_{x \to x^*} G(x) = \infty$ also when $x^* = \infty$, i.e. when $f(x) > 0$ for all $x > 0$. Since we also have $\lim_{x \to 0} G(x) = -\infty$, $G(\cdot)$ is a bijection from $(0, x^*)$ onto $\mathbb{R}$ with continuous inverse. Therefore, by (4.3), the limit $H(x) = \lim_{t \to \infty} h_t(x)$ exists and satisfies

$$G(H(x)) = G(h_{t_0}) + \int_0^{xe^{-at_0}} \left(\frac{1}{au} - \frac{1}{f(u)} \right) du,$$

where the convergence is uniform over x on compact sets. This claim follows from the fact that the right-hand side does not depend on the choice of t_0:

$$\begin{aligned}
G(H(x)) &= G(h_{t_0}) + \int_0^{xe^{-at_0}} \left(\frac{1}{au} - \frac{1}{f(u)} \right) du \\
&= \int_0^{h_{t_0}} \left(\frac{1}{f(u)} - \frac{1}{au} \right) du + \frac{1}{a} \log h_{t_0} + \int_0^{xe^{-at_0}} \left(\frac{1}{au} - \frac{1}{f(u)} \right) du \\
&= \int_{xe^{-at_0}}^{h_{t_0}} \frac{1}{f(u)} du + \frac{1}{a} \log x - t_0 \\
&= \frac{1}{a} \log x.
\end{aligned}$$

The last equality holds by the definition of h_{t_0}. $\qquad\square$

Remark 4.1. The function $H(x)$ has two particular properties.

(H$_1$) $H(\cdot)$ is a nontrivial solution of the ODE $H'(x) = f(H(x))/(ax)$, $x > 0$, for which $H(0) = 0$. This property can be seen directly from the explicit formula $H(x) = G^{-1}((\log x)/a)$.

(H$_2$) $H(\cdot)$ solves Schröder's functional equation

$$H(x) = \phi_s \circ H(xe^{-as}) \quad \text{for all } x > 0 \text{ and } s > 0. \tag{4.4}$$

To show (4.4), take the limit $t \to \infty$ in the equation for the semigroup property of the flow $\phi_t(xe^{-at}) = \phi_s \circ \phi_{t-s}(xe^{-as}e^{-a(t-s)})$ and use continuity.

4.2. Feller's branching diffusion

The basic element of the approximation is Feller's branching diffusion

$$Y_t = 1 + \int_0^t aY_s \, \mathrm{d}s + \int_0^t \sqrt{Y_s} \, \mathrm{d}B_s, \qquad t \geq 0, \tag{4.5}$$

driven by the same Brownian motion as in (2.1). The rescaled process $W_t := e^{-at} Y_t$ is a nonnegative martingale with a nondegenerate almost-sure limit

$$W := 1 + \int_0^\infty e^{-as} \sqrt{Y_s} \, \mathrm{d}B_s. \tag{4.6}$$

An explicit expression is available for the Laplace transform of Y_t (see, e.g. [11, Lemma 5, p. 28]):

$$\mathbb{E}e^{-\lambda Y_t} = \exp\left(-\frac{\lambda a e^{at}}{a + \lambda(e^{at} - 1)/2}\right), \qquad \lambda > 0.$$

Thus,

$$\mathbb{E}e^{-\lambda W} = \lim_{t \to \infty} \mathbb{E}e^{-\lambda W_t} = \exp\left(-\frac{2a\lambda}{2a + \lambda}\right) = \mathbb{E}\exp\left(-\lambda \sum_{j=0}^{\Pi} \tau_j\right)$$

with independent random variables $\Pi \sim \mathrm{Poi}(2a)$ and $\tau_j \sim \mathrm{Exp}(2a)$. We show in Section 4.4 that the random variable W emerges in the limit claimed in Theorem 2.2.

4.3. Approximation on bounded intervals

The next lemma shows that, under appropriate scaling, the solution of (2.1) converges to the Feller branching diffusion (4.5) on bounded intervals.

Lemma 4.1. *Let $\overline{X}_t^\varepsilon := \varepsilon^{-1} X_t^\varepsilon$, where X_t^ε is the solution of (2.1) subject to $X_0^\varepsilon = \varepsilon$. Then, for all $t \geq 0$,*

$$\overline{X}_t^\varepsilon \xrightarrow{L^1} Y_t \quad \text{as } \varepsilon \to 0,$$

where Y_t is the solution of (4.5).

Proof. The process $\overline{X}_t^\varepsilon$ satisfies

$$\overline{X}_t^\varepsilon = 1 + \int_0^t \varepsilon^{-1} f(\varepsilon \overline{X}_s^\varepsilon) \, \mathrm{d}s + \int_0^t \sqrt{\varepsilon^{-1} \sigma(\varepsilon \overline{X}_s^\varepsilon)} \, \mathrm{d}B_s. \tag{4.7}$$

First we show that the moments of $\overline{X}_t^\varepsilon$ are bounded, uniformly in ε on any finite time interval. From the drift condition (A$_3$), $f(x) \leq ax$ for all $x \geq 0$. Then the standard localization of the stochastic integral gives

$$\mathbb{E}\overline{X}_t^\varepsilon = 1 + \mathbb{E}\int_0^t \varepsilon^{-1} f(\varepsilon \overline{X}_s^\varepsilon)\, \mathrm{d}s \leq 1 + \int_0^t a\mathbb{E}\overline{X}_s^\varepsilon\, \mathrm{d}s,$$

and, by Grönwall's inequality,

$$\mathbb{E}\overline{X}_t^\varepsilon \leq \mathrm{e}^{at} \quad \text{for all } t > 0. \tag{4.8}$$

Now define $W_t^\varepsilon := \mathrm{e}^{-at}\overline{X}_t^\varepsilon$. Then

$$W_t^\varepsilon = 1 + \int_0^t \mathrm{e}^{-as}(\varepsilon^{-1} f(\varepsilon \overline{X}_s^\varepsilon) - a\overline{X}_s^\varepsilon)\, \mathrm{d}s + \int_0^t \mathrm{e}^{-as}\sqrt{\varepsilon^{-1}\sigma(\varepsilon \overline{X}_s^\varepsilon)}\, \mathrm{d}B_s$$

$$\leq 1 + \int_0^t \mathrm{e}^{-as}\sqrt{\varepsilon^{-1}\sigma(\varepsilon \overline{X}_s^\varepsilon)}\, \mathrm{d}B_s, \tag{4.9}$$

where the inequality holds because $\varepsilon^{-1} f(\varepsilon x) \leq ax$. Since W_t^ε is positive,

$$\mathbb{E}(W_t^\varepsilon)^2 \leq 1 + \|\sigma'\|_\infty \int_0^t \mathrm{e}^{-2as}\mathbb{E}\overline{X}_s^\varepsilon\, \mathrm{d}s \leq 1 + \frac{\|\sigma'\|_\infty}{a},$$

where we have used (4.8). Consequently,

$$\mathbb{E}(\overline{X}_t^\varepsilon)^2 \leq \left(1 + \frac{\|\sigma'\|_\infty}{a}\right)\mathrm{e}^{2at} =: m_2(t). \tag{4.10}$$

For $\varepsilon \in (0, 1)$, let

$$\psi_\varepsilon(u) := \frac{1}{u \log(1/\sqrt{\varepsilon})}\mathbf{1}_{\{\varepsilon \leq u \leq \sqrt{\varepsilon}\}},$$

and define

$$h_\varepsilon(x) := \int_0^{|x|}\int_0^y \psi_\varepsilon(u)\, \mathrm{d}u\, \mathrm{d}y, \qquad x \in \mathbb{R}.$$

Obviously,

$$h_\varepsilon'(x) = \int_0^x \psi_\varepsilon(u)\, \mathrm{d}u = \begin{cases} 0, & 0 < x \leq \varepsilon, \\ \dfrac{\log(x/\varepsilon)}{\log(1/\sqrt{\varepsilon})}, & \varepsilon < x \leq \sqrt{\varepsilon}, \\ 1, & x > \sqrt{\varepsilon}, \end{cases}$$

so $|h_\varepsilon'(x)| \leq 1$. Since $h_\varepsilon(x)$ is symmetric around 0, $h_\varepsilon'(x) \geq \mathbf{1}_{\{x \geq \sqrt{\varepsilon}\}}$ for $x > 0$ and $h_\varepsilon(0) = 0$, so we also have

$$|x| \leq h_\varepsilon(x) + \sqrt{\varepsilon}.$$

The function $h_\varepsilon(x)$ is a smooth approximation of $|x|$, used in the proof of the Yamada–Watanabe theorem and related applications (see, e.g. [5]).

Note from (4.5) and (4.7) that the difference $Z_t^\varepsilon := Y_t - \overline{X}_t^\varepsilon$ satisfies

$$\mathrm{d}Z_t^\varepsilon = aZ_t^\varepsilon\, \mathrm{d}t - (\varepsilon^{-1} f(\varepsilon \overline{X}_t^\varepsilon) - a\overline{X}_t^\varepsilon)\, \mathrm{d}t + \left(\sqrt{Y_t} - \sqrt{\varepsilon^{-1}\sigma(\varepsilon \overline{X}_t^\varepsilon)}\right)\mathrm{d}B_t$$

subject to $Z_0^\varepsilon = 0$. Applying Itô's formula to $h_\varepsilon(Z_t^\varepsilon)$, we conclude that

$$|Z_t^\varepsilon| \le \sqrt{\varepsilon} + h_\varepsilon(Z_t^\varepsilon) = \sqrt{\varepsilon} + \int_0^t I_\varepsilon(s)\,\mathrm{d}s + \frac{1}{2}\int_0^t J_\varepsilon(s)\,\mathrm{d}s + M_\varepsilon(t), \qquad (4.11)$$

where

$$I_\varepsilon(t) := h_\varepsilon'(Z_t^\varepsilon)(aZ_t^\varepsilon - [\varepsilon^{-1}f(\varepsilon\overline{X}_t^\varepsilon) - a\overline{X}_t^\varepsilon]),$$

$$J_\varepsilon(t) := h_\varepsilon''(Z_s^\varepsilon)\left(\sqrt{Y_t} - \sqrt{\varepsilon^{-1}\sigma(\varepsilon\overline{X}_t^\varepsilon)}\right)^2,$$

$$M_\varepsilon(t) := \int_0^t h_\varepsilon'(Z_s^\varepsilon)\left(\sqrt{Y_s} - \sqrt{\varepsilon^{-1}\sigma(\varepsilon\overline{X}_s^\varepsilon)}\right)\mathrm{d}B_s.$$

Since $|h_\varepsilon'(x)| \le 1$ and $|\varepsilon^{-1}f(\varepsilon x) - ax| \le \varepsilon\|f''\|_\infty x^2$, the first term satisfies

$$|I_\varepsilon(s)| \le a|Z_s^\varepsilon| + \varepsilon\|f''\|_\infty(\overline{X}_s^\varepsilon)^2.$$

Furthermore, since $\sqrt{x}$ has global Hölder exponent $\frac{1}{2}$,

$$\left(\sqrt{y} - \sqrt{\varepsilon^{-1}\sigma(\varepsilon x)}\right)^2 \le |y - \varepsilon^{-1}\sigma(\varepsilon x)| \le |y - x| + \varepsilon\|\sigma''\|_\infty x^2, \qquad x, y \in \mathbb{R}_+, \quad (4.12)$$

and, therefore, using the estimate for $h_\varepsilon''(x)$,

$$\begin{aligned}
|J_\varepsilon(t)| &\le |h_\varepsilon''(Z_s^\varepsilon)|\left(\sqrt{Y_t} - \sqrt{\varepsilon^{-1}\sigma(\varepsilon\overline{X}_t^\varepsilon)}\right)^2 \\
&\le \frac{1}{|Z_t^\varepsilon|\log(1/\sqrt{\varepsilon})}\mathbf{1}_{\{\varepsilon \le |Z_t^\varepsilon| \le \sqrt{\varepsilon}\}}(|Z_t^\varepsilon| + \varepsilon\|\sigma''\|_\infty(\overline{X}_t^\varepsilon)^2) \\
&\le \frac{1}{\log(1/\sqrt{\varepsilon})}(1 + \|\sigma''\|_\infty(\overline{X}_t^\varepsilon)^2).
\end{aligned}$$

Similarly, by (4.12), the quadratic variation of M_ε is bounded by

$$\begin{aligned}
\langle M_\varepsilon\rangle_t &= \int_0^t (h_\varepsilon'(Z_s^\varepsilon))^2\left(\sqrt{Y_s} - \sqrt{\varepsilon^{-1}\sigma(\varepsilon\overline{X}_s^\varepsilon)}\right)^2\mathrm{d}s \\
&\le \int_0^t |Z_s^\varepsilon|\,\mathrm{d}s + \varepsilon\|\sigma''\|_\infty\int_0^t (\overline{X}_s^\varepsilon)^2\,\mathrm{d}s \\
&\le \int_0^t (Y_s + X_s^\varepsilon)\,\mathrm{d}s + \varepsilon\|\sigma''\|_\infty\int_0^t (\overline{X}_s^\varepsilon)^2\,\mathrm{d}s.
\end{aligned}$$

Since the moments of Y_t and X_t^ε are bounded on $[0, T]$ uniformly in ε, the latter implies that

$$\mathbb{E}\langle M_\varepsilon\rangle_t \le C_T < \infty,$$

where the constant C_T depends only on T. In particular, M_ε is a square-integrable martingale with zero mean. Now taking expectations in (4.11) and using these bounds, we conclude that

$$\mathbb{E}|Z_t^\varepsilon| \le \sqrt{\varepsilon} + a\int_0^t \mathbb{E}|Z_s^\varepsilon|\,\mathrm{d}s + \varepsilon\|f''\|_\infty\int_0^t \mathbb{E}(\overline{X}_s^\varepsilon)^2\,\mathrm{d}s + \frac{\int_0^t (1 + \|\sigma''\|_\infty\mathbb{E}(\overline{X}_s^\varepsilon)^2)\,\mathrm{d}s}{\log(1/\sqrt{\varepsilon})}.$$

Estimate (4.10) and Grönwall's inequality imply that $\mathbb{E}|Z_t^\varepsilon| \to 0$ as $\varepsilon \to 0$. $\qquad\square$

4.4. The approximation

The crux of our proof is an approximation of X_t^ε by means of a deterministic extrapolation of its trajectory onwards from a certain suitably chosen time point. To this end, introduce the deterministic and stochastic flows $\phi_{s,t}(x)$ and $\Phi_{s,t}(x)$ respectively generated by ODE (2.2) and SDE (2.1), namely the solutions of these equations at time t that start at x at time s. Furthermore, for any constant $c \in (\frac{1}{2}, 1)$, let $t_c = c|\log \varepsilon|/a$ and $t_1 = T_\varepsilon = |\log \varepsilon|/a$. With these definitions, $X_{T_\varepsilon}^\varepsilon = \Phi_{t_c,t_1}(X_{t_c}^\varepsilon)$ and

$$X_{T_\varepsilon}^\varepsilon = (\Phi_{t_c,t_1}(X_{t_c}^\varepsilon) - \phi_{t_c,t_1}(X_{t_c}^\varepsilon)) + \phi_{t_c,t_1}(X_{t_c}^\varepsilon). \tag{4.13}$$

To prove the convergence in (4.1), we show that, as $\varepsilon \to 0$, the first term vanishes and the second converges to $H(W)$, where W is the random variable from (4.6).

Lemma 4.2. *As $\varepsilon \to 0$,*

$$\Phi_{t_c,t_1}(X_{t_c}^\varepsilon) - \phi_{t_c,t_1}(X_{t_c}^\varepsilon) \xrightarrow{L^2} 0.$$

Proof. Write $\Phi_t^\varepsilon := \Phi_{t_c,t_c+t}(X_{t_c}^\varepsilon)$ and $\phi_t := \phi_{t_c,t_c+t}(X_{t_c}^\varepsilon)$ for brevity, and define $\delta_t^\varepsilon = \Phi_t^\varepsilon - \phi_t$. Subtracting (2.1) and (2.2), and applying the Itô formula,

$$\mathbb{E}(\delta_t^\varepsilon)^2 = \mathbb{E}\int_0^t 2\delta_s(f(\Phi_s^\varepsilon) - f(\phi_s))\,\mathrm{d}s + \int_0^t \varepsilon \mathbb{E}\sigma(\Phi_s^\varepsilon)\,\mathrm{d}s \le \int_0^t 2a\mathbb{E}(\delta_s)^2\,\mathrm{d}s + \varepsilon t\|\sigma\|_\infty,$$

where we have used assumption (2). By Grönwall's inequality,

$$\mathbb{E}(\Phi_{t_c,t_1}(X_{t_c}^\varepsilon) - \phi_{t_c,t_1}(X_{t_c}^\varepsilon))^2 = \mathbb{E}(\delta_{t_1-t_c}^\varepsilon)^2 \le C_1 \varepsilon t_1 \mathrm{e}^{2a(t_1-t_c)} \le C_2 \varepsilon^{2c-1}|\log \varepsilon| \to 0$$

as $\varepsilon \to 0$, where the convergence holds because $c \in (\frac{1}{2}, 1)$. $\qquad\square$

The next lemma establishes convergence of the second term in (4.13).

Lemma 4.3. *As $\varepsilon \to 0$,*

$$\phi_{t_c,t_1}(X_{t_c}^\varepsilon) \xrightarrow{\mathbb{P}} H(W).$$

Proof. First we show that, as $\varepsilon \to 0$, $W_{t_c}^\varepsilon = \mathrm{e}^{-at_c}\overline{X}_{t_c}^\varepsilon$ converges in L^1 to the limit W from (4.6), associated with the Feller branching diffusion. Indeed, using representations (4.6) and (4.9),

$$W - W_{t_c}^\varepsilon := \int_{t_c}^\infty \mathrm{e}^{-as}\sqrt{Y_s}\,\mathrm{d}B_s - \int_0^{t_c} \mathrm{e}^{-as}(\varepsilon^{-1}f(\varepsilon\overline{X}_s^\varepsilon) - a\overline{X}_s^\varepsilon)\,\mathrm{d}s$$

$$+ \int_0^{t_c} \mathrm{e}^{-as}\left(\sqrt{Y_s} - \sqrt{\varepsilon^{-1}\sigma(\varepsilon\overline{X}_s^\varepsilon)}\right)\mathrm{d}B_s$$

$$=: I_1(\varepsilon) + I_2(\varepsilon) + I_3(\varepsilon).$$

The first term converges to 0 in L^2 because

$$\mathbb{E}I_1(\varepsilon)^2 = \mathbb{E}\left(\int_{t_c}^\infty \mathrm{e}^{-as}\sqrt{Y_s}\,\mathrm{d}B_s\right)^2 = \int_{t_c}^\infty \mathrm{e}^{-2as}\mathbb{E}Y_s\,\mathrm{d}s \to 0,$$

since, for the Feller branching diffusion, $\mathbb{E}Y_s = \mathrm{e}^{as}$. The second term converges to 0 in L^1:

$$\mathbb{E}|I_2(\varepsilon)| \le \int_0^{t_c} \mathrm{e}^{-as}\varepsilon\|f''\|_\infty \mathbb{E}(\overline{X}_s^\varepsilon)^2\,\mathrm{d}s \le C_3\varepsilon \int_0^{t_c} \mathrm{e}^{as}\,\mathrm{d}s \le C_4\varepsilon^{1-c} \to 0.$$

For the last term,

$$\mathbb{E}[(I_3(\varepsilon))^2] = \mathbb{E}\left(\int_0^{t_c} \mathrm{e}^{-as}\left(\sqrt{Y_s} - \sqrt{\varepsilon^{-1}\sigma(\varepsilon\overline{X}_s^\varepsilon)}\right)\mathrm{d}B_s\right)^2$$

$$\leq \int_0^{t_c} \mathrm{e}^{-2as}(\mathbb{E}|Y_s - \overline{X}_s^\varepsilon| + \varepsilon\|\sigma''\|_\infty \mathbb{E}(\overline{X}_s^\varepsilon)^2)\,\mathrm{d}s$$

$$=: \int_0^\infty \mathbf{1}_{\{s\leq t_c\}}g_\varepsilon(s)\,\mathrm{d}s. \tag{4.14}$$

Since $\overline{X}_s^\varepsilon$ converges to Y_s in L^1 by Lemma 4.1, it follows that $g_\varepsilon(s) \to 0$ as $\varepsilon \to 0$ for any fixed $s \geq 0$. Furthermore, exponential bounds on the moments (4.8) and (4.10) imply that

$$\mathbf{1}_{\{s\leq t_c\}}g_\varepsilon(s) \leq C\mathbf{1}_{\{s\leq t_c\}}\mathrm{e}^{-2as}(\mathrm{e}^{as} + \varepsilon\mathrm{e}^{2as})$$
$$= C\mathbf{1}_{\{s\leq t_c\}}\mathrm{e}^{-2as}(\mathrm{e}^{as} + \mathrm{e}^{-(a/c)t_c}\mathrm{e}^{2as})$$
$$\leq C\mathbf{1}_{\{s\leq t_c\}}\mathrm{e}^{-2as}(\mathrm{e}^{as} + \mathrm{e}^{-(a/c)s}\mathrm{e}^{2as})$$
$$\leq C(\mathrm{e}^{-as} + \mathrm{e}^{-(a/c)s}),$$

which is integrable on $\mathbb{R}_+$. Hence, by dominated convergence, the integral at (4.14) converges to 0. Thus, $W_{t_c}^\varepsilon$ converges to W in L^1 as $\varepsilon \to 0$.

Finally, by the definition of t_1 and t_c, as $\varepsilon \to 0$,

$$\phi_{t_c,t_1}(X_{t_c}^\varepsilon) = \phi_{t_c,t_1}(W_{t_c}^\varepsilon\mathrm{e}^{-a(t_1-t_c)}) = \phi_{t_1-t_c}(W_{t_c}^\varepsilon\mathrm{e}^{-a(t_1-t_c)}) \xrightarrow{\mathbb{P}} H(W),$$

where the limit holds by the uniform convergence from Proposition 4.1. □

References

[1] BARBOUR, A. D., CHIGANSKY, P. AND KLEBANER, F. C. (2016). On the emergence of random initial conditions in fluid limits. *J. Appl. Prob.* **53,** 1193–1205.
[2] BARBOUR, A. D., HAMZA, K., KASPI, H. AND KLEBANER, F. C. (2015). Escape from the boundary in Markov population processes. *Adv. Appl. Prob.* **47,** 1190–1211.
[3] CHIGANSKY, P., JAGERS, P. AND KLEBANER, F. C. (2018). What can be observed in real time PCR and when does it show? *J. Math. Biol.* **76,** 679–695.
[4] FREIDLIN, M. I. AND WENTZELL, A. D. (2012). *Random Perturbations of Dynamical Systems* (Fundamental Principles Math. Sci. **260**), 3rd edn. Springer, Heidelberg.
[5] GYÖNGY, I. AND RÁSONYI, M. (2011). A note on Euler approximations for SDEs with Hölder continuous diffusion coefficients. *Stoch. Process. Appl.* **121,** 2189–2200.
[6] KENDALL, D. G. (1956). Deterministic and stochastic epidemics in closed populations. In *Proc. 3rd Berkeley Symp. Math. Statist. Prob., 1954–1955*, Vol. IV, University of California Press, Berkeley, CA, pp. 149–165.
[7] KLEBANER, F. C. (2012). *Introduction to Stochastic Calculus With Applications*, 3rd edn. Imperial College Press, London.
[8] KLEBANER, F. C. *et al.* (2011). Stochasticity in the adaptive dynamics of evolution: the bare bones. *J. Biol. Dynam.* **5,** 147–162.
[9] KURTZ, T. G. (1970). Solutions of ordinary differential equations as limits of pure jump Markov processes. *J. Appl. Prob.* **7,** 49–58.

[10] MARTIN, G. AND LAMBERT, A. (2015). A simple, semi-deterministic approximation to the distribution of selective sweeps in large populations. *Theoret. Pop. Biol.* **101,** 40–46.
[11] PARDOUX, É. (2016). *Probabilistic Models of Population Evolution* (Math. Biosci. Inst. Lecture Ser. **1.6**). Springer, Cham.
[12] THORISSON, H. (2000). *Coupling, Stationarity, and Regeneration*. Springer, New York.
[13] WHITTLE, P. (1955). The outcome of a stochastic epidemic—a note on Bailey's paper. *Biometrika* **42,** 116–122.

J. BAKER, *Monash University*

School of Mathematical Sciences, Monash University, Monash, VIC 3800, Australia.
Email address: jeremy.baker@monash.edu

P. CHIGANSKY, *The Hebrew University of Jerusalem*

Department of Statistics, The Hebrew University, Mount Scopus, Jerusalem 91905, Israel.
Email address: pavel.chigansky@mail.huji.ac.il

K. HAMZA, *Monash University*

School of Mathematical Sciences, Monash University, Monash, VIC 3800, Australia.
Email address: kais.hamza@monash.edu

F. C. KLEBANER, *Monash University*

School of Mathematical Sciences, Monash University, Monash, VIC 3800, Australia.
Email address: fima.klebaner@monash.edu