Available online at www.sciencedirect.com
SCIENCE DIRECT®

Physica A 335 (2004) 385–412

PHYSICA A

www.elsevier.com/locate/physa

White noise limits for discrete dynamical systems driven by fast deterministic dynamics

Dror Givon, Raz Kupferman*

Institute of Mathematics, The Hebrew University, Jerusalem 91904, Israel

Received 15 September 2003

Abstract

We study a class of singularly perturbed dynamical systems that have fast and slow components, $\varepsilon \ll 1$ being the fast to slow timescale ratio. The fast components are governed by a strongly mixing discrete map, which is iterated at time intervals ε. The slow components are governed by a first-order finite-difference equation that uses a time step ε. As ε tends to zero, the fast components may be eliminated, giving rise to SDEs for the slow components. The emerging stochastic calculus is, in the general case, of neither Itô nor Stratonovich type, but depends on the correlation time of the mixing process.

PACS: 05.40.−a

Keywords: Scale separation; SDEs; Mixing

1. Introduction

Many problems in science exhibit the interaction of dynamics characterized by very different time scales. A class of such systems, which has been the focus of extensive study for a century, is characterized by an explicit distinction between "fast" and "slow" variables—systems with *scale separation*. Often, the problem may be formulated by a set of ordinary differential equations (ODEs):

$$\frac{dx}{dt} = f(x, y) ,$$

* Corresponding author.

E-mail addresses: givon@math.huji.ac.il (D. Givon), raz@math.huji.ac.il (R. Kupferman).

URLs: http://www.math.huji.ac.il/givon, http://www.math.huji.ac.il/razk

0378-4371/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2003.12.019

$$\frac{dy}{dt} = \frac{g(x,y)}{\varepsilon}, \tag{1.1}$$

where $\varepsilon \ll 1$; $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ are the "slow" and "fast" variables, respectively. The parameter ε is the fast to slow timescale ratio.

Averaging methods are concerned with the derivation of "effective dynamics", which approximate the evolution of the slowly varying components $x(t)$ without having to solve for the fast components $y(t)$ [1,2]. Thus, one looks for a new system of ODEs—the "reduced", or "effective" system

$$\frac{dX}{dt} = F(X)$$

so that $X(t) \in \mathbb{R}^n$ converges to $x(t)$ as $\varepsilon \to 0$. There exist many variants to this basic idea: the fast dynamics may take place on an infinite-dimensional space, the equations may be non-autonomous, there can be more than one fast time scale, and some of the components may satisfy stochastic dynamics; see Ref. [3] for a recent literature review.

A situation of interest is where the fast dynamics is ergodic, that is, for almost every $x \in \mathbb{R}^n$ the dynamics defined by the ODE

$$\frac{dy}{dt} = g(x,y) \tag{1.2}$$

(with x viewed as a parameter) is ergodic. Then, Anosov's theorem (e.g. Ref. [2]) proves the uniform convergence of $x(t)$, as $\varepsilon \to 0$, to the solution $X(t)$ of the reduced equation

$$\frac{dX}{dt} = \int f(X(t),z)\, \mu_{X(t)}(dz),$$

where μ_x is the invariant measure (on $\mathbb{R}^m$) associated with the fast dynamics (1.2).

Consider now a different class of systems with scale separation:

$$\frac{dx}{dt} = f_1(x,y) + \frac{f_0(x,y)}{\sqrt{\varepsilon}},$$

$$\frac{dy}{dt} = \frac{g(x,y)}{\varepsilon}, \tag{1.3}$$

where the fast dynamics is ergodic, and $f_0(x,y)$ averages to zero under μ_x for all x. In this class of problems both the x and y equations contain fast dynamics but the dynamics in y is an order of magnitude faster. Although $f_0(x,y)$ averages to zero under the y-dynamics, the fluctuations may influence the x-dynamics to leading order, i.e., they may not vanish as $\varepsilon \to 0$. (Systems of form (1.3) can also be identified, upon a rescaling of time, with systems of form (1.1) on time intervals of the order of $1/\sqrt{\varepsilon}$.)

In Ref. [3], the authors and Stuart considered such a system where x satisfies a scalar equation and $y \in \mathbb{R}^3$ satisfies the Lorenz equations. Numerical experiments suggest that $f_0(x,y)/\sqrt{\varepsilon}$ might converge, as $\varepsilon \to 0$, to white noise. Whether this is indeed the case remains an open question. The situation is considerably simpler if the fast dynamics is stochastic, in which case one may apply classical perturbation analyses by Kurtz [4] and by Papanicolaou and co-workers [5,6]; see also more recent analyses by

Majda et al. [7]. Papanicolaou's analysis in Ref. [6] may also apply for deterministic systems with random initial data, however, the assumptions for which the general theorem applies are usually hard to verify for given systems.

Discrete variants of systems of form (1.3) were constructed and studied by Beck [8,9]. In this paper we are concerned with such discrete systems, in which the small parameter ε is also the discretization step. We describe Beck's work using notations and terminology consistent with system (1.3). The fast variable, y, assumes values in some bounded interval $Y \subset \mathbb{R}$. The function $y(t)$ is constant on time intervals $[t_n, t_{n+1})$, where $t_n = n\varepsilon$. Every ε time units a new value of y is assigned by transforming its previous value under a discrete map T; let $y_n = y(t_n)$, then $y_{n+1} = Ty_n$. The initial value y_0 is randomly drawn from a probability measure, $\mathbb{P}$, invariant under T; expected values with respect to $\mathbb{P}$ are denoted by $\mathbb{E}$. The map T is assumed to have sufficiently strong mixing properties in a sense to be specified below. The dynamics of the slow variables x are defined, in analogy with (1.3), so that $(x_n, y_n) = (x(t_n), y(t_n))$ satisfy the discrete system

$$\begin{aligned} \frac{x_{n+1} - x_n}{\varepsilon} &= f_1(x_n, y_n) + \frac{f_0(x_n, y_n)}{\sqrt{\varepsilon}} , \\ y_{n+1} &= Ty_n . \end{aligned} \tag{1.4}$$

We are concerned with the case where $f_0(x, y)$ averages to zero under iterations of the discrete map T acting on the y variables, i.e.,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f_0(x, T^k y) = \int f_0(x, y)\, \mathbb{P}(\mathrm{d}y) = 0 .$$

A generic example of such a dynamical system $(Y, \mathscr{B}, \mathbb{P}, T)$ is $Y = [0, 1]$, $\mathscr{B}$ the σ-field of Borel sets on $[0, 1]$, and $T : [0, 1] \mapsto [0, 1]$ defined by $Ty = 2y \bmod 1$; it is easy to see that $\mathbb{P} =$ the Lebesgue measure on $[0, 1]$ is invariant under T. The function $f_0(x, y) = \cos(2\pi y)$ has the required property of averaging to zero under $\mathbb{P}$.

While the slow dynamics is defined in discrete time, it is useful to retain the notion of a process in continuous time. In this work we take the function $x(t)$ to be piecewise-linear, with straight lines connecting the points (t_n, x_n). This choice allows us to carry our analysis within the space of continuous functions, which simplifies some of the proofs.

Consider first the case $f_0(x, y) = f_0(y)$ and $f_1(x, y) = 0$, with $f_0(y)$ continuous. Denoting the slow variables by $x_n = x_n^{\mathrm{B}}$, (1.4) can be summed up explicitly, to find

$$x_n^{\mathrm{B}} = x_0 + \varepsilon \sum_{k=0}^{n-1} \frac{f_0(y_k)}{\sqrt{\varepsilon}} \tag{1.5}$$

or in continuous time,

$$x^{\mathrm{B}}(t) = x_0 + \varepsilon \sum_{k=0}^{\lfloor t/\varepsilon \rfloor - 1} \frac{f_0(y_k)}{\sqrt{\varepsilon}} + (t - \varepsilon \lfloor t/\varepsilon \rfloor) \frac{f_0(y_{\lfloor t/\varepsilon \rfloor})}{\sqrt{\varepsilon}} . \tag{1.6}$$

As $\varepsilon \to 0$ the last term on the right-hand side vanishes uniformly (recall that y_n is bounded and $f_0(y)$ continuous), hence the entire focus is on the second term, which is the sum of $\mathcal{O}(1/\varepsilon)$ terms of order $\mathcal{O}(\sqrt{\varepsilon})$. Each of the summands is a random variable whose expected value is zero, which suggests that this sum may converge by a central limit theorem (CLT) to a Gaussian variable. However, one cannot use the CLT in its standard form because the $f_0(y_k)$ are not independent; the y_k are all iterates of the deterministic map T: $y_k = T^k y_0$. Yet, if the map T has sufficiently strong mixing property (which amounts to asymptotic independence) and the function f_0 satisfies certain boundedness conditions, then it may be shown that the second term in (1.6) converges weakly to a Gaussian random variable. Moreover, $x^{\mathrm{B}}(t)$ converges weakly in $C[0,\tau]$ (the space of continuous functions on $[0,\tau]$ endowed with the sup-norm topology; τ is arbitrary) to $\sigma B(t)$, where $B(t)$ is standard Brownian motion and

$$\sigma^2 = \mathbb{E}[f_0(y)^2] + 2\sum_{n=1}^{\infty} \mathbb{E}[f_0(y) f_0(T^n y)] .$$

The proof of this statement is the subject of a chapter in Billingsley [10]. A precise formulation will be given in the next section.

Thus, (1.5) defines a family of random processes $x^{\mathrm{B}}(t)$, whose weak limit, as $\varepsilon \to 0$, $X^{\mathrm{B}}(t)$, satisfies the stochastic differential equation (SDE):

$$\mathrm{d}X^{\mathrm{B}}(t) = \sigma\, \mathrm{d}B(t) .$$

The summand $f_0(y_k)/\sqrt{\varepsilon}$ plays here the role of white noise—the derivative (in the sense of distributions) of Brownian motion.

This observation lead Beck to speculate that random processes defined by dynamical systems of form (1.4) could, under certain conditions, weakly converge to SDEs of more general type. Specifically, he considered the case where $f_0(x,y) = f_0(y)$ and $f_1(x,y) = -\gamma x$, $\gamma > 0$. Denoting here the slow variables by $x_n = x_n^{\mathrm{OU}}$, (1.4) reduces to the linear (in x) equation

$$\frac{x_{n+1}^{\mathrm{OU}} - x_n^{\mathrm{OU}}}{\varepsilon} = -\gamma x_n^{\mathrm{OU}} + \frac{f_0(y_n)}{\sqrt{\varepsilon}} . \tag{1.7}$$

In view of the identification of $f_0(y_n)/\sqrt{\varepsilon}$ with an approximation to white noise, it is natural to expect that the function $x^{\mathrm{OU}}(t)$ defined through the discrete equation (1.7) weakly converges to the solution $X^{\mathrm{OU}}(t)$ of the SDE:

$$\mathrm{d}X^{\mathrm{OU}}(t) = -\gamma X^{\mathrm{OU}}(t)\,\mathrm{d}t + \sigma\,\mathrm{d}B(t) ,$$

i.e., $x^{\mathrm{OU}}(t)$ converges to an Ornstein–Uhlenbeck process [11]. This assertion was indeed proved by Beck in Ref. [9] and supported by extensive numerical simulations in Refs. [8,9].

The question is whether Beck's theorem may be extended to more general dynamics of form (1.4), and notably to nonlinear equations and multiplicative noise. In Section 3, we analyze the linear system (1.7) considered by Beck and provide a new convergence proof, based on a continuity argument. Specifically, we identify the relation between $x^{\mathrm{B}}(t)$ and $x^{\mathrm{OU}}(t)$ as a continuous mapping in $C[0,\tau]$. Since weak convergence is preserved under continuous mappings, the convergence of $x^{\mathrm{B}}(t)$, as $\varepsilon \to 0$, implies the convergence of $x^{\mathrm{OU}}(t)$. All that remains is to show that the limiting process is indeed an OU process.

Using the same approach, we then extend in Section 4 the treatment to non-autonomous equations of the form

$$\frac{x_{n+1} - x_n}{\varepsilon} = f_1(x_n, t_n) + \beta(t_n) \frac{f_0(y_n)}{\sqrt{\varepsilon}} . \tag{1.8}$$

For T and $f_0(y)$ satisfying the same assumptions as before and $f_1(x,t)$ continuous in time and Lipschitz continuous in x, we prove that $x(t)$ converges weakly in $C[0,\tau]$ to the solution $X(t)$ of the nonlinear SDE

$$\mathrm{d}X(t) = f_1(X,t)\,\mathrm{d}t + \sigma\beta(t)\,\mathrm{d}B(t) . \tag{1.9}$$

These results are supported by numerical experiments. In particular, we compare the approach to the limit to that obtained with the white noise approximated by i.i.d. random variables, say, $\sigma\xi_n/\sqrt{\varepsilon}$, where $\xi_n \sim \mathcal{N}(0,1)$, which yields an Euler-type approximation to SDE (1.9).

The interesting situation is when the (approximate) white noise term is multiplied by a function of x_n or y_n. This is where the delicacies of stochastic calculus emerge. A general analysis of such situations is beyond the scope of this paper. In Section 5, we examine three examples in which the limiting SDEs can be derived. Those turn out to be, in general, neither Itô nor Stratonovich SDEs; the relevant stochastic calculus depends on the rate at which the correlations of $f_0(y_n)$ decay. This situation is in contrast to the Stratonovich SDEs obtained in the $\varepsilon \to 0$ limit if the x-dynamics were governed by differential equations, rather than difference equations. The emergence of intermediate stochastic calculus can be understood in light of recent results by Pavliotis et al. [12,13], who study SDE limits in scale separated systems where the fast dynamics generate colored noise, and the slow dynamics satisfy a second-order equation with small inertia (see also the related work by Graham and Schenzle [14]). Then, intermediate stochastic calculus arises in the presence of a "competition" between the inertial time and the noise correlation time. In our case, it is the discretization which introduces an intrinsic time scale which is comparable with the noise correlation time.

2. Background: limit theorems for sums of weakly dependent variables

In this section, we describe the setting under which sums of form (1.6) weakly converge to Brownian motion as $\varepsilon \to 0$. Our presentation follows closely Beck [9], with notations adapted such to fit into the framework presented in Section 1.

Let then

$$x^{\mathrm{B}}(t) = \varepsilon \sum_{k=0}^{\lfloor t/\varepsilon \rfloor - 1} \frac{f_0(y_k)}{\sqrt{\varepsilon}} + (t - \varepsilon\lfloor t/\varepsilon \rfloor) \frac{f_0(y_{\lfloor t/\varepsilon \rfloor})}{\sqrt{\varepsilon}} , \tag{2.1}$$

where at this point $y_1, y_2, \ldots$ is an arbitrary sequence of random variables on a probability space $(Y, \mathscr{B}, \mathbb{P})$ and f_0 is a real-valued function on Y (the y_n will be related back to a dynamical system further below). The simplest case is when the variables y_n are i.i.d. If the function $f_0(y)$ has mean zero and finite variance σ^2, then Donsker's theorem states that $x^{\mathrm{B}}(t)$ weakly converges, as $\varepsilon \to 0$, to Brownian motion [10],

$$x^{\mathrm{B}} \Rightarrow \sigma B \quad \text{in } C[0,\tau] .$$

The next level of complexity is where the variables $y_0, y_1, \dots$ are not independent, but y_n and y_{n+m} become asymptotically independent as m increases. To this end, the notion of ϕ-mixing sequences is introduced.

Let $y_0, y_1, \dots$ be a stationary sequence of random variables on $(Y, \mathscr{B}, \mathbb{P})$, and define

$$\mathscr{B}^n = \sigma(y_0, y_1, \dots, y_n) ,$$

$$\mathscr{B}_n = \sigma(y_{n+1}, y_{n+2}, \dots) ,$$

to be the σ-algebras generated by the sequence up to the nth element, and from the $(n+1)$th element on, respectively.

Definition 2.1. The stationary sequence y_n is called ϕ-mixing if for every $k \geqslant 0$ and $m \geqslant 1$, and for every $E_1 \in \mathscr{B}^k$, and $E_2 \in \mathscr{B}_{k+m}$

$$|\mathbb{P}(E_1 \cap E_2) - \mathbb{P}(E_1)\mathbb{P}(E_2)| \leqslant \phi(m)\mathbb{P}(E_1) ,$$

where $\phi(m)$ is a non-negative function such that $\phi(m) \to 0$ as $m \to \infty$.

Thus, events related to the sequence y_n up to its kth element and events related to the sequence from its $(k+m)$th element on are only weakly dependent for large m. The function ϕ is a bound on the degree of dependence between such events. It is easy to see that if y_n is a ϕ-mixing sequence and $f_0(y)$ is a measurable real-valued function, then the sequence $\eta_n = f_0(y_n)$ is also ϕ-mixing, with a different function ϕ.

Donsker's theorem may then be generalized for ϕ-mixing sequences [10]. Suppose that $f_0(y_n)$ is a ϕ-mixing sequence with $\phi(n)$ satisfying $\sum_{n=1}^{\infty} \sqrt{\phi(n)} < \infty$. If $f_0(y)$ has mean zero and finite variance, then the series

$$\sigma^2 = \mathbb{E}[f_0^2(y_0)] + 2\sum_{k=1}^{\infty} \mathbb{E}[f_0(y_0) f_0(y_k)]$$

converges absolutely, and $x^{\mathrm{B}}(t)$, given by (2.1), weakly converges to Brownian motion, $x^{\mathrm{B}} \Rightarrow \sigma B$ in $C[0,\tau]$.

We are concerned in this paper with the case where the sequence $y_0, y_1, \dots$ is generated by the iterates of a deterministic map T on the initial element y_0. Thus, y_0 is random but $y_1 = Ty_0$, $y_2 = T^2 y_0$, etc., are deterministically determined by y_0. Clearly, the sequence y_n is not ϕ-mixing, yet the notion of ϕ-mixing sequences may still be exploited. This requires the introduction of *generating partitions*.

Consider the dynamical system $(Y, \mathscr{B}, \mathbb{P}, T)$ and let $\mathscr{A}$ be a finite alphabet. A *partition* ξ of $(Y, \mathscr{B})$ is a measurable map $Y \mapsto \mathscr{A}$ (to each point y in Y corresponds a letter a in the alphabet $\mathscr{A}$). Let Σ be the space of sequences $\{a_0, a_1, \dots\}$, where $a_n \in \mathscr{A}$ (the space of infinite words over the alphabet $\mathscr{A}$). The map T together with the partition ξ define a map from Y to Σ:

$$y \mapsto \xi_T(y) = \{\xi(y), \xi(Ty), \xi(T^2 y), \dots\} .$$

Note that ξ_T maps T into a left-shift operator S in the sense that $\xi_T(Ty) = S(\xi_T(y))$ for all $y \in Y$. The partition ξ is called *generating* if ξ_T admits a retraction; namely, if there exists a map $\pi : \Sigma \mapsto Y$ (a map that converts infinite words over $\mathscr{A}$ back into element of Y) such that (i) π is measurable, (ii) π acts as a left inverse, $\pi(\xi_T(y)) = y$, and (iii) π maps the shift operator into T in the sense that $\pi(S\varsigma) = T\pi(\varsigma)$ for all $\varsigma \in \Sigma$. Thus, a partition is called generating if the elements in Y can be identified, in a reversible way, with infinite words over some alphabet, and the dynamical system maps into a symbolic dynamical system.

The generic example [10,9] is $Y = [0,1]$, $\mathscr{B}$ the σ-algebra of Borel sets on $[0,1]$, $Ty = 2y \bmod 1$, and $\mathbb{P}$ the Lebesgue measure on $[0,1]$, which is invariant under T. A generating partition is $\xi(y) = \lfloor 2y \rfloor$, i.e., the alphabet is $\mathscr{A} = \{0,1\}$ and $\xi(y) = 0$ if $y < 1/2$ and $\xi(y) = 1$ if $y \geqslant 1/2$. Then, ξ_T maps every $y \in [0,1]$ to its standard binary representation $a_0a_1a_2\ldots$, whereas π maps every sequence $a_0a_1a_2\ldots$ of $\{0,1\}$ into the number $y = a_0 2^{-1} + a_1 2^{-2} + \cdots$.

To any dynamical system $(Y, \mathscr{B}, T, \mathbb{P})$ with generating partition ξ we associate a sequence of random variables ξ_n on Y, $\xi_n(y) = \xi(T^n y)$. Since T is measure preserving, the sequence is stationary. We may now define the notion of ϕ-mixing for dynamical systems:

Definition 2.2. A dynamical system $(Y, \mathscr{B}, T, \mathbb{P})$ is said to be ϕ-mixing if it admits a generating partition ξ, such that the sequence of random variables $\xi_n = \xi \circ T^n$ is ϕ-mixing.

Note that in the above example $\phi(n) \equiv 0$, because the digits ξ_n of the binary representation are strictly independent. It is worth emphasizing again that the iterates $T^n y$ are not independent; y determines all the $T^n y$ deterministically. It is only the corresponding symbols, $\xi(T^n y)$, which are independent.

We now turn our attention back to sum (2.1). The summands are proportional to $f_0(y_k) = f_0(T^k y_0)$. Let ξ be a generating partition on $(Y, \mathscr{B}, \mathbb{P}, T)$; it maps f_0 into a function F_0 on Σ:

$$F_0(\varsigma) = f_0(\pi(\varsigma)) .$$

Thus, $\eta_k(y) = f_0(T^k y)$ can be written in the equivalent form $\eta_k(y) = F_0(S^k \varsigma)$, where $\varsigma = \xi_T(y)$. Note that η_k is a function of the sequence of random variables $(\xi_k, \xi_{k+1}, \ldots)$. Suppose that $F_0(\varsigma)$ depends only on a finite truncation of the infinite sequence ς. Then, if $(\xi_0, \xi_1, \ldots)$ is ϕ-mixing (i.e., the dynamical system is ϕ-mixing), then so it the sequence $(\eta_0, \eta_1, \ldots)$, and provided that the corresponding function $\phi(n)$ decays sufficiently fast, then $x^{\mathrm{B}}(t)$ weakly converges to Brownian motion. This result may be extended to cases where $F_0(\varsigma)$ depends on the entire sequence, provided that it can be approximated well enough by a function that depends only on a finite truncation of ς.

Theorem 2.1. *Let $(Y, \mathscr{B}, \mathbb{P}, T)$ be a ϕ-mixing dynamical system with generating partition ξ, such that $\sum_{n=0}^{\infty} \sqrt{\phi(n)} < \infty$. Let $\xi_n(y) = \xi(T^n y)$ and $\mathscr{B}^k = \sigma(\xi_0, \xi_1, \ldots, \xi_k)$.*

If $f_0(y)$ is a real valued function on Y with the property that

$$\sum_{k=1}^{\infty} (\mathbb{E}|f_0 - \mathbb{E}[f_0|\mathscr{B}^k]|^2)^{1/2} < \infty ,$$

where $\mathbb{E}[f_0|\mathscr{B}^k]$ is the conditional expectation of $f_0(y)$ given the first k digits of $\xi_T(y)$, then the series

$$\sigma^2 = \mathbb{E}[f_0^2(y_0)] + 2\sum_{k=1}^{\infty} \mathbb{E}[f_0(y_0)f_0(y_k)] \tag{2.2}$$

converges absolutely and $x^{\mathrm{B}}(t)$, given by (2.1), weakly converges, as $\varepsilon \to 0$, to $\sigma B(t)$.

Returning to the example with $Y = [0,1]$ and $Ty = 2y \bmod 1$, we take $f_0(y) = \cos(2\pi y)$. The function f_0 is differentiable, hence it satisfies the assumptions of the theorem and it is a matter of a simple calculation to show that $\sigma^2 = 1/2$, i.e., $x^{\mathrm{B}} \Rightarrow \frac{1}{\sqrt{2}} B$.

3. From deterministic dynamics to SDEs: a first example

Consider system (1.7) considered in Refs. [8,9]:

$$\begin{aligned} y_{n+1} &= Ty_n \\ \frac{x_{n+1}^{\mathrm{OU}} - x_n^{\mathrm{OU}}}{\varepsilon} &= -\gamma x_n^{\mathrm{OU}} + \frac{f_0(y_n)}{\sqrt{\varepsilon}} , \end{aligned} \tag{3.1}$$

where T and $f_0(y)$ are assumed to satisfy the assumptions on Theorem 2.1. Recall that a continuous piecewise-linear function, $x^{\mathrm{OU}}(t)$, is defined by

$$x^{\mathrm{OU}}(t) = x_{\lfloor t/\varepsilon \rfloor}^{\mathrm{OU}} + (t/\varepsilon - \lfloor t/\varepsilon \rfloor)(x_{\lfloor t/\varepsilon \rfloor + 1}^{\mathrm{OU}} - x_{\lfloor t/\varepsilon \rfloor}^{\mathrm{OU}}) .$$

In this section, we provide a new proof that, as $\varepsilon \to 0$, $x^{\mathrm{OU}}(t)$ weakly converges in $C[0,\tau]$ to the OU processes $X^{\mathrm{OU}}(t)$ defined by the SDE,

$$\mathrm{d}X^{\mathrm{OU}}(t) = -\gamma X^{\mathrm{OU}}(t)\,\mathrm{d}t + \sigma\,\mathrm{d}B(t) , \tag{3.2}$$

where σ is given by (2.2).

Our proof is based on the well-known fact that weak convergence is preserved under a continuous mapping [10, p. 31]. That is, if $\mathscr{F}$ is a continuous mapping in $C[0,\tau]$, and $x(t)$ weakly converges in $C[0,\tau]$ as $\varepsilon \to 0$, then $\mathscr{F}[x(t)]$ is also weakly convergent. We will construct such an $\mathscr{F}$ so that $x^{\mathrm{OU}} = \mathscr{F}[x^{\mathrm{B}}]$, where $x^{\mathrm{B}}(t)$ is given by (2.1). Thus, the weak convergence of $x^{\mathrm{B}}(t)$ implies the weak convergence of $x^{\mathrm{OU}}(t)$. It only remains then to show that the limit is indeed $X^{\mathrm{OU}}(t)$.

Theorem 3.1. *Let $(Y, \mathscr{B}, \mathbb{P}, T)$ and f_0 satisfy the conditions of Theorem* 2.1, *and let $x^{\mathrm{OU}}(t)$ and $X^{\mathrm{OU}}(t)$ be given by* (3.1) *and* (3.2), *respectively. Then, as $\varepsilon \to 0$,*

$$x^{\mathrm{OU}} \Rightarrow X^{\mathrm{OU}} \quad \text{in } C[0,\tau] .$$

Proof. We construct a continuous map $\mathscr{F}$ on $C[0,\tau]$, such that $x^{\mathrm{OU}} = \mathscr{F}[x^{\mathrm{B}}]$. Because $C[0,\tau]$ is a complete metric space, it is sufficient to construct an $\mathscr{F}$ that is uniformly continuous on a dense subset; it then has a unique extension, which is uniformly continuous, to the whole space [15, p. 118].

Define $\mathscr{E} \subset C[0,\tau]$ to be the set of all piecewise linear functions whose points of non-differentiability are rational. Obviously $\mathscr{E}$ is dense in $C[0,\tau]$. By restricting the values of ε to rational numbers we guarantee that $x^{\mathrm{B}}(t) \in \mathscr{E}$ and $x^{\mathrm{OU}}(t) \in \mathscr{E}$.

Let $h(t)$ be an arbitrary function in $\mathscr{E}$, and denote by M the smallest common denominator (SCD) of its points of non-differentiability. Given $h(t)$, we define a corresponding discrete function,

$$h_n = h\left(\frac{n}{M}\right), \quad n = 0, 1, \ldots, \lfloor M\tau \rfloor$$

(note that the points $t_n = n/M$ contain all the points of non-differentiability of h), and use it to construct another discrete function, x_n^h, satisfying

$$\frac{x_{n+1}^h - x_n^h}{1/M} = -\gamma x_n^h + \frac{h_{n+1} - h_n}{1/M}\ , \tag{3.3}$$

with $x_0^h = x_0$. The continuous process $x^h(t)$ is defined by

$$x^h(t) = x_{\lfloor tM \rfloor}^h + (tM - \lfloor tM \rfloor)(x_{\lfloor tM \rfloor + 1}^h - x_{\lfloor tM \rfloor}^h)\ . \tag{3.4}$$

This defines a mapping on $\mathscr{E}$, $h \mapsto x^h$, which we denote by $\mathscr{F}$. It is easily verified that $\mathscr{F}[x^{\mathrm{B}}] = x^{\mathrm{OU}}$.

In Lemma 3.1 below we show that $\mathscr{F}$ is a uniformly continuous mapping in $\mathscr{E}$. Hence, it can be uniquely extended as a uniformly continuous mapping in $C[0,\tau]$; since no confusion should arise, we denote its extension by $\mathscr{F}$ as well. It then follows that the weak convergence $x^{\mathrm{B}}(t) \Rightarrow \sigma B(t)$, by Theorem 2.1, implies that $x^{\mathrm{OU}}(t)$ weakly converges to the limiting process, $\mathscr{F}[\sigma B]$.

It remains to identify the process $\mathscr{F}[\sigma B]$. Since the limit $\mathscr{F}[\sigma B]$ does not depend on the sequence in $\mathscr{E}$ that approximates B, we can use

$$\mathscr{F}[\sigma B] = \lim_{n \to \infty} \mathscr{F}[\sigma x^{\mathrm{w}}]\ ,$$

where $x^{\mathrm{w}}(t)$ is any differentiable approximation to standard Brownian motion. Then, it is well-known that $\mathscr{F}[\sigma x^{\mathrm{w}}]$ converges in probability in $C[0,\tau]$ to the solution of (Stratonovich) SDE (3.2) [16] (since the coefficient of $\mathrm{d}B$ is constant, there is no difference here between Itô and Stratonovich interpretations). □

Lemma 3.1. *$\mathscr{F}$ is uniformly continuous on the subset $\mathscr{E}$.*

Proof. Let $\eta > 0$, and define $\delta = \eta/2\mathrm{e}^{|\gamma|\tau}$. Let $g, h \in \mathscr{E}$ such that $\|g - h\| \leqslant \delta$. Without loss of generality, we assume that the functions g, h have the same points of non-differentiability, with M their SCD. As above we define $x^g = \mathscr{F}[g]$ and $x^h = \mathscr{F}[h]$.

Then, for $m = 0, 1, \ldots, \lfloor M\tau \rfloor$,

$$
\begin{aligned}
|x_m^h - x_m^g| &= \left| -\frac{\gamma}{M} \sum_{i=1}^{m-1} (x_i^h - x_i^g) + \sum_{i=1}^{m-1} (h_{i+1} - h_i) - \sum_{i=1}^{m-1} (g_{i+1} - g_i) \right| \\
&= \left| -\frac{\gamma}{M} \sum_{i=1}^{m-1} (x_i^h - x_i^g) + (h_m - h_0) - (g_m - g_0) \right| \\
&\leqslant \frac{\gamma}{M} \sum_{i=1}^{m-1} |x_i^h - x_i^g| + |h_m - g_m| + |h_0 - g_0| \\
&\leqslant \frac{\gamma}{M} \sum_{i=1}^{m-1} |x_i^h - x_i^g| + 2\delta \; .
\end{aligned}
$$

A discrete version of Gronwall's inequality yields

$$
|x_m^h - x_m^g| \leqslant 2\delta \left(1 + \frac{\gamma}{M} \right)^{m-1} \leqslant 2\delta \mathrm{e}^{|\gamma|\tau} = \eta \; .
$$

Thus, for all $\eta > 0$ there exists a $\delta > 0$, such that $\|g - h\| \leqslant \delta$ implies that $|\mathscr{F}[g] - \mathscr{F}[h]| \leqslant \eta$ on all points $t_m = m/M$. Since the maximum distance between functions in $\mathscr{E}$ (that have the same points of non-differentiability) is attained at a point of non-differentiability, this implies the uniform continuity of the map $\mathscr{F}$. □

3.1. Numerical experiment

We now illustrate the behavior of the deterministic process (3.1) with a computer experiment. We took $y_0 \sim \mathscr{U}[0,1]$ and $Ty = 2y \bmod 1$. In some of the experiments we used $f_0(y) = \cos(2\pi y)$ and in other $f_0(y) = \sqrt{2}(y - 1/2)$; in both cases $\sigma^2 = 1/2$. Note that because of the finite representation of real numbers on a computer, T cannot be implemented using real number multiplication; it is necessary to actually generate a sequence of pseudo-random bits.

In Fig. 1 we plot the empirical distribution of $x(t)$ for $\gamma = 1$ and three different values of ε for a sampling time of 10^5 units. The dashed line corresponds to $f_0(y) = \cos(2\pi y)$ and the dash–dotted line corresponds to $f_0(y) = \sqrt{2}(y - 1/2)$. Note that in the first case the distribution is asymmetric, whereas it is symmetric in the second case. The empirical distribution is compared to the empirical distribution generated by an Euler scheme (dotted line), i.e., by (3.1) with $f_0(y_n)$ replaced by independent normal variables $\mathscr{N}(0, \sigma^2)$. The solid line is the empirical distribution of the OU process, which is a normal distribution with variance $1/2\sigma^2$. As expected, all the curves approach the asymptotic limit for small ε.

In Fig. 2 we plot the corresponding auto-covariance functions. Here again, all the curves approach the limiting exponential curve for small ε. Note that $x(t)$ generated using $f_0(y) = \cos(2\pi y)$ has the same empirical auto-covariance as the Euler approximation, which may be shown by a straightforward calculation (the $f_0(y_n)$ are uncorrelated in both cases).

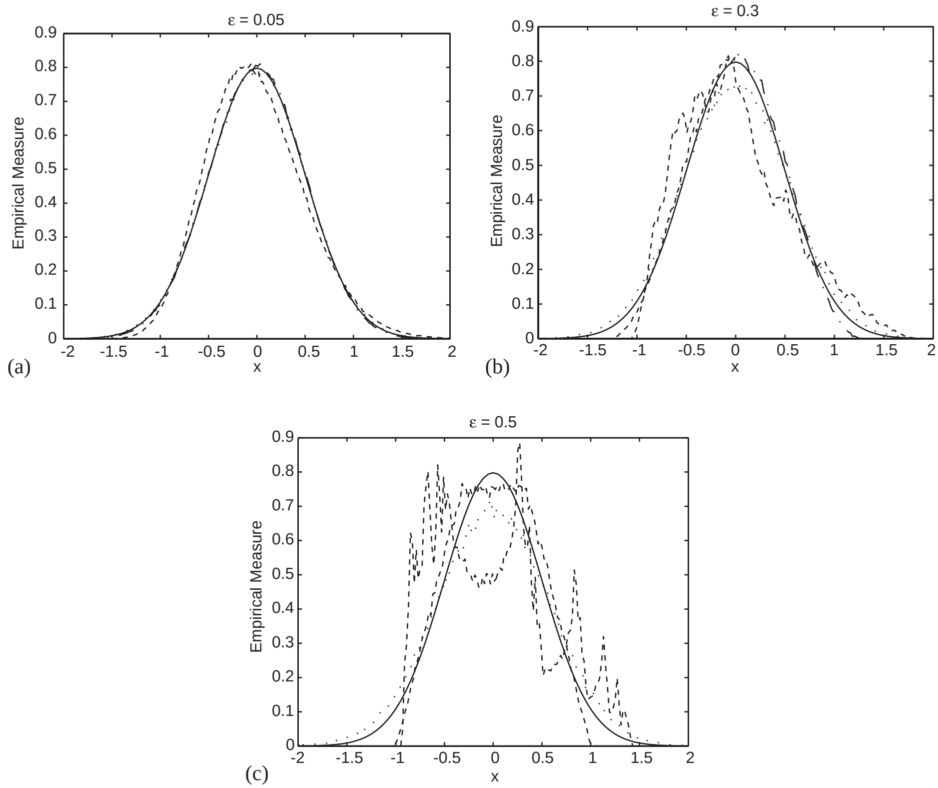

Fig. 1. Dashed lines: the empirical measure of $x(t)$ solving (3.1) for $\gamma = 1$, $f_0(y) = \cos(2\pi y)$, and (a) $\varepsilon = 0.05$, (b) $\varepsilon = 0.3$, (c) $\varepsilon = 0.5$. Dash–dotted lines: same for $f_0(y) = \sqrt{2}(y - 1/2)$. Dotted-lines same for $x(t)$ generated by an Euler approximation of the OU equation. Solid line: the empirical measure of the limiting OU process.

4. Extension to more general SDEs

In this section we extend the treatment to systems of the form,

$$y_{n+1} = T y_n ,$$

$$\frac{x_{n+1} - x_n}{\varepsilon} = f_1(x_n, t_n) + \beta(t_n)\,\frac{f_0(y_n)}{\sqrt{\varepsilon}} , \tag{4.1}$$

where $t_n = n\varepsilon$ and T and $f_0(y)$ are assumed to satisfy the assumptions of Theorem 2.1. The continuous piecewise-linear function, $x(t)$, is defined by

$$x(t) = x_{\lfloor t/\varepsilon \rfloor} + (t/\varepsilon - \lfloor t/\varepsilon \rfloor)(x_{\lfloor t/\varepsilon \rfloor + 1} - x_{\lfloor t/\varepsilon \rfloor}) .$$

In this section we prove that $x(t)$ weakly converges in $C[0,\tau]$, as $\varepsilon \to 0$, to the process $X(t)$ defined by the SDE,

$$\mathrm{d}X(t) = f_1(X, t)\,\mathrm{d}t + \sigma\beta(t)\,\mathrm{d}B(t) , \tag{4.2}$$

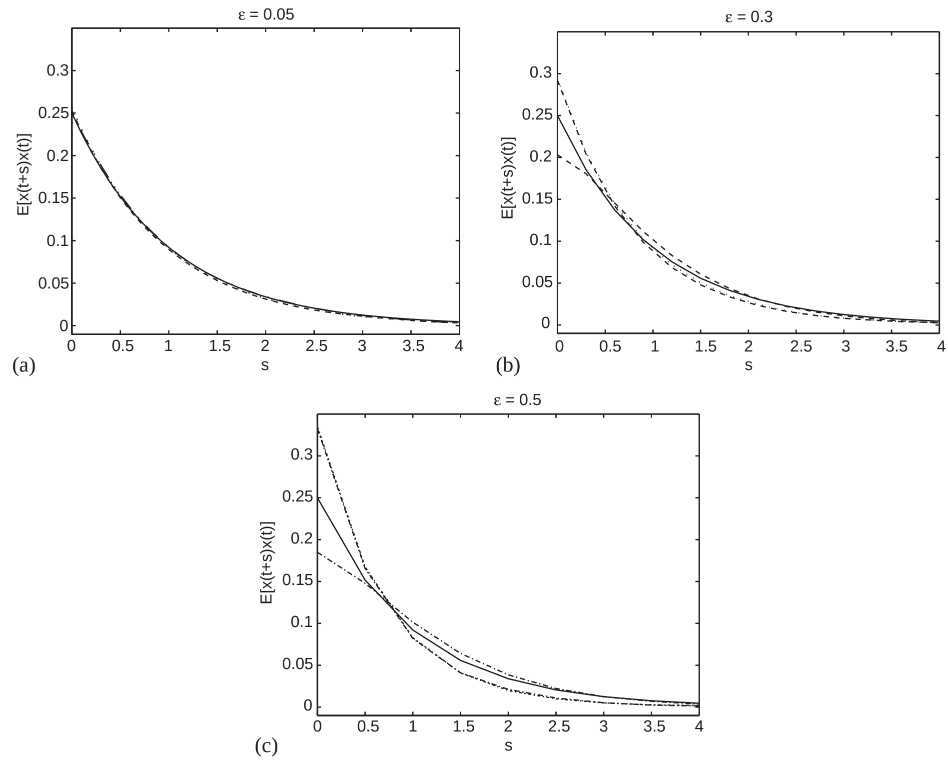

Fig. 2. Dashed lines: the empirical auto-covariance of $x(t)$ solving (3.1) for $\gamma = 1$, $f_0(y) = \cos(2\pi y)$, and (a) $\varepsilon = 0.05$, (b) $\varepsilon = 0.3$, (c) $\varepsilon = 0.5$. Dash–dotted lines: same for $f_0(y) = \sqrt{2}(y - 1/2)$. Dotted lines same for $x(t)$ generated by an Euler approximation of the OU equation. Solid line: the empirical auto-covariance of the limiting OU process.

where σ is still given by (2.2). Our proof uses the same technique that was used in Section 3—the fact that weak convergence is preserved under a continuous mapping.

Theorem 4.1. *Let* $(Y, \mathscr{B}, \mathbb{P}, T)$ *and* f_0 *satisfy the conditions of Theorem* 2.1, *and let* $x(t)$ *and* $X(t)$ *be given by* (4.1) *and* (4.2), *respectively. Assume that* $f_1(z,t)$ *is Lipschitz in z with constant L, and that* $\beta(t)$ *is of bounded variation. Then, as* $\varepsilon \to 0$,

$$x \Rightarrow X \quad \text{in } C[0,\tau] .$$

Proof. Following the proof of Theorem 3.1, we construct a uniformly continuous map $\mathscr{F}$ on $\mathscr{E}$, such that $x = \mathscr{F}[x^{\mathrm{B}}]$. Let $h(t)$ be an arbitrary function in $\mathscr{E}$, and denote by M the SCD of its points of non-differentiability. Given $h(t)$, we define a corresponding discrete function,

$$h_n = h\left(\frac{n}{M}\right), \quad n = 0, 1, \ldots, \lfloor M\tau \rfloor$$

and use it to construct another discrete function, x_n^h, satisfying

$$\frac{x_{n+1}^h - x_n^h}{1/M} = f_1(x_n^h, n/M) + \beta(n/M)\frac{h_{n+1} - h_n}{1/M} \tag{4.3}$$

with $x_0^h = x_0$. The continuous process $x^h(t)$ is defined as in (3.4). This defines a mapping $\mathscr{F}$ on $\mathscr{E}$, $x^h = \mathscr{F}[h]$. Again, it is easily verified that $x = \mathscr{F}[x^{\mathrm{B}}]$. In Lemma 4.1 below we show that $\mathscr{F}$ is a uniformly continuous mapping on $\mathscr{E}$, so that $\mathscr{F}[x^{\mathrm{B}}] \Rightarrow \mathscr{F}[B]$. The identification of the process $\mathscr{F}[B]$ is the same as in the proof of Theorem 3.1. □

Lemma 4.1. *$\mathscr{F}$ is uniformly continuous on the subset $\mathscr{E}$.*

Proof. Let $\eta > 0$. Define

$$V = \sup_{0=t_0<\cdots<t_n=\tau} \sum_{i=0}^{n-1} |\beta(t_{i+1}) - \beta(t_i)| \tag{4.4}$$

to be the variation of $\beta(t)$ in $[0, \tau]$, and

$$\delta = \frac{\eta}{(V + 2\|\beta\|)\mathrm{e}^{L\tau}} . \tag{4.5}$$

Let $g, h \in \mathscr{E}$ such that $\|g - h\| \leqslant \delta$. Without loss of generality we assume that the functions g, h have the same points of non-differentiability, with M their SCD. If $x^g = \mathscr{F}[g]$ and $x^h = \mathscr{F}[h]$, then for all $m = 0, 1, \ldots, \lfloor M\tau \rfloor$,

$$\begin{aligned}
|x_m^h - x_m^g| &= \Bigg| \frac{1}{M} \sum_{i=1}^{m-1} \left[f_1\left(x_i^h, \frac{i}{M}\right) - f_1\left(x_i^g, \frac{i}{M}\right) \right] \\
&\quad + \sum_{i=1}^{m-1} \beta\left(\frac{i}{M}\right)(h_{i+1} - g_{i+1} - h_i + g_i) \Bigg| \\
&\leqslant \frac{L}{M} \sum_{i=1}^{m-1} |x_i^h - x_i^g| + \Bigg| \sum_{i=1}^{m-1} \left[\beta\left(\frac{i+1}{M}\right) - \beta\left(\frac{i}{M}\right) \right](h_i - g_i) \\
&\quad - \beta(0)(h_0 - g_0) + \beta\left(\frac{m-1}{M}\right)(h_{m-1} - g_{m-1}) \Bigg| ,
\end{aligned}$$

where we used the Lipschitz property and summation by parts. Substituting (4.4) and (4.5),

$$|x_m^h - x_m^g| \leqslant \frac{1}{M} \sum_{i=1}^{m-1} L|x_i^h - x_i^g| + (V + 2\|\beta\|)\delta ,$$

and by the discrete Gronwall inequality,

$$|x_m^h - x_m^g| \leqslant (V + 2\|\beta\|)\delta \left(1 + \frac{L}{M}\right)^{m-1} \leqslant (V + 2\|\beta\|)\delta \mathrm{e}^{|L|} = \eta ,$$

which implies the uniform continuity of the map $\mathscr{F}$. □

4.1. Numerical experiment

Consider the following deterministic process:

$$y_{n+1} = Ty_n \, ,$$
$$\frac{x_{n+1} - x_n}{\varepsilon} = -V'(x_n) + \frac{f_0(y_n)}{\sqrt{\varepsilon}} \tag{4.6}$$

with $V(x) = \frac{x^4}{4} - \frac{x^2}{2}$, $y_0 \sim \mathcal{U}[0,1]$, $Ty = 2y \bmod 1$, and $f_0(y) = \cos(2\pi y)$.

Theorem 4.1 asserts that $x(t)$ weakly converges to the solution of the SDE,

$$\mathrm{d}X(t) = -V'(X(t))\,\mathrm{d}t + \sigma\,\mathrm{d}B(t) \, , \tag{4.7}$$

$\sigma = 1/\sqrt{2}$, which describes a noise-driven over-damped particle in a potential well $V(x)$. The process $X(t)$ is ergodic with distribution $Z^{-1}\exp[-4V(X)]$, and Z is a normalizing constant.

Fig. 3 shows the empirical distribution for $x(t)$ solving (4.6) for three values of ε and a sampling time of 10^5 units. As before, we compare the empirical distribution to

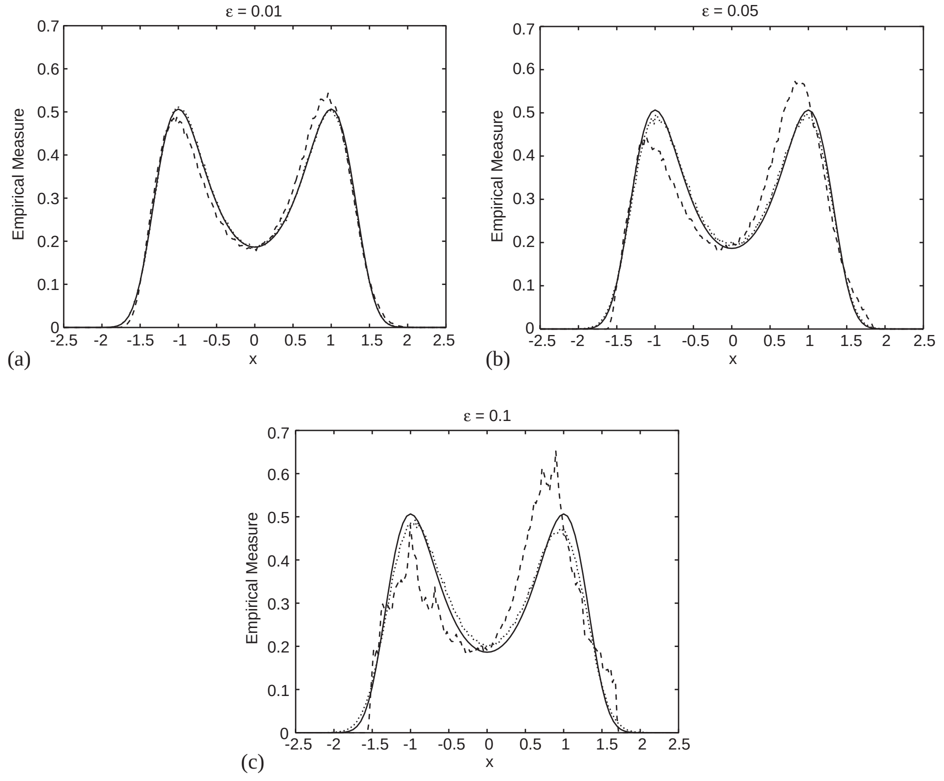

Fig. 3. Dashed lines: the empirical distribution of $x(t)$ solving (4.6) for (a) $\varepsilon = 0.01$, (b) $\varepsilon = 0.05$, and (c) $\varepsilon = 0.1$. Dotted lines: empirical distribution of an Euler approximation to the SDE (4.7) with time step ε. Solid lines: empirical distribution of the limiting process (4.7).

that obtained from an Euler approximation to (4.7) with time step ε. The two curves are compared with the empirical distribution of the limiting process $X(t)$. The results clearly indicate the convergence of the empirical distribution to the predicted limit. Note once again the asymmetry of the distribution for $x(t)$ solving (4.6), which is still noticeable even for ε as small of 0.01.

5. Examples with multiplicative noise

In the last two sections, we studied dynamical systems that converge to SDEs in which the noise term is additive—the prefactor of the (approximate) Brownian increment is at most a function of time and does not depend on the random process. In this section, we study three examples in which the noise is multiplicative, in which case it is not clear a priori how to interpret stochastic integration. The two most common interpretations of stochastic integration are Itô's, which assumes that Brownian increments are independent of the present or past state of the process, and the Stratonovich interpretation, which applies when Brownian motion is approximated by a sequence of differentiable functions.

5.1. First example

Consider the following discrete system:

$$\begin{aligned} y_{n+1} &= T y_n , \\ \frac{x_{n+1}-x_n}{\varepsilon} &= \frac{f_0(y_n)}{\sqrt{\varepsilon}} , \\ \frac{z_{n+1}-z_n}{\varepsilon} &= x_n \frac{f_0(y_n)}{\sqrt{\varepsilon}} \end{aligned} \tag{5.1}$$

with $x_0 = 0$ and $z_0 = 0$; T and f_0 satisfy the assumptions of Theorem 2.1.

The discrete function x_n is identical to x_n^{B} defined by (1.5); its piecewise-linear interpolant, $x(t) = x^{\mathrm{B}}(t)$ weakly converges to $\sigma B(t)$, with σ given by (2.2). Since $f_0(y_n)/\sqrt{\varepsilon}$ are the increments of x_n^{B}, we expect $z(t)$, the piecewise-linear interpolant of z_n, to weakly converge to the solution $Z(t)$ of the SDE

$$\mathrm{d}Z(t) = \sigma^2 B(t)\,\mathrm{d}B(t), \qquad Z(0) = 0 .$$

The question is in which sense has this equation to be interpreted. When interpreted in the sense of Itô its solution is $Z_I(t) = (\sigma^2/2)[B^2(t) - t]$, whereas the Stratonovich calculus yields $Z_S(t) = (\sigma^2/2)B^2(t)$. We will show that $z(t)$ converges to a stochastic integral $\sigma^2 \int B\,\mathrm{d}B$, which is neither in the sense of Itô nor Stratonovich. Specifically, $z(t) \Rightarrow Z(t)$, where

$$Z(t) = \frac{\sigma^2}{2}\left[B^2(t) - \frac{\hat{\sigma}^2}{\sigma^2}\,t\right] \tag{5.2}$$

and $\hat{\sigma}^2 = \mathbb{E} f_0^2(y)$.

To obtain this result we start by rewriting z_n in the following form:

$$\begin{aligned} z_n &= \sum_{k=0}^{n-1} x_k(x_{k+1} - x_k) \\ &= \frac{1}{2}\sum_{k=0}^{n-1} [x_{k+1}^2 - x_k^2 - (x_{k+1} - x_k)^2] \\ &= \frac{1}{2} x_n^2 - \frac{1}{2}\sum_{k=0}^{n-1} \varepsilon f_0^2(y_k) \\ &\equiv z_n^{(1)} - z_n^{(2)} . \end{aligned}$$

Note that this equation holds also for the piecewise-linear interpolants, $z(t) = z^{(1)}(t) - z^{(2)}(t)$.

We now show that $z^{(1)}(t)$ weakly converges to $(\sigma^2/2)B^2(t)$, and that $z^{(2)}(t)$ converges almost-surely (a.s.) to the (non-random) function $(\hat{\sigma}^2/2)t$, whence the validity of (5.2). Note that, in general, $z^{(1)} \Rightarrow Z^{(1)}$ and $z^{(2)} \Rightarrow Z^{(2)}$ does not imply the weak convergence of the sums $z^{(1)} + z^{(2)} \Rightarrow Z^{(1)} + Z^{(2)}$. Indeed, for the sum to converge we need the pair $(z^{(1)}, z^{(2)})$ to weakly converge in the product space $C[0,\tau] \times C[0,\tau]$. This is guaranteed, however, if one of the summands converges in a stronger sense, say, in probability [10].

We start by arguing that $z^{(1)}(t)$ weakly converges to $(\sigma^2/2)B^2(t)$. This may seem obvious since $x^{\mathrm{B}} \Rightarrow \sigma B$, and the squaring of a function is a continuous operation in $C[0,\tau]$. More generally, if $h : \mathbb{R} \mapsto \mathbb{R}$ is continuous, then it is continuous as an operator in $C[0,\tau]$, hence $h(x^{\mathrm{B}}) \Rightarrow h(\sigma B)$ in $C[0,\tau]$. This argument require some care, since $z^{(1)}(t) \neq \frac{1}{2}[x^{\mathrm{B}}(t)]^2$; these two functions only agree on the points $t_n = n\,\varepsilon$, and the square of the linear interpolant of x_n^{B} does not coincide with the interpolant of $(x_n^{\mathrm{B}})^2$. In Appendix A we show that if $h : \mathbb{R} \mapsto \mathbb{R}$ is differentiable and $h'(z)$ can be bounded by exponential growth, then the piecewise-linear interpolant of the discrete function $h(x_n^{\mathrm{B}})$ weakly converges to $h(\sigma B)$.

It remains to show that $z^{(2)}(t)$, converges a.s. in $C[0,\tau]$ to the non-random function $(\hat{\sigma}^2/2)t$. Since the maximum of a piecewise-linear function is attained at a point of non-differentiability, it is sufficient to show that a.s.

$$\lim_{\varepsilon\to 0} \max_{0 \leqslant n \leqslant \lfloor \tau/\varepsilon \rfloor} \left| z_n^{(2)} - \frac{\hat{\sigma}^2}{2}\, t_n \right| = 0 ,$$

where $t_n = n\,\varepsilon$.

Let $\delta > 0$ be given. Since the sequence $f_0(y_n)$ is ergodic, then it follows from Birkhoff's ergodic theorem that there exists an ε_δ such that for all $\varepsilon < \varepsilon_\delta$ and $\delta/\varepsilon < n \leqslant \lfloor \tau/\varepsilon \rfloor$,

$$\left| \frac{1}{n}\sum_{k=0}^{n-1} f_0^2(y_k) - \hat{\sigma}^2 \right| \leqslant \delta$$

or equivalently,

$$\left|z_n^{(2)} - \tfrac{\hat\sigma^2}{2} t_n\right| \leqslant \frac{n\varepsilon}{2}\,\delta \leqslant \frac{\tau}{2}\,\delta\ .$$

For $0 \leqslant n \leqslant \delta/\varepsilon$ we have

$$|z_n^{(2)}| \leqslant \frac{M}{2}\,\delta, \qquad \frac{\hat\sigma^2}{2}\,t_n \leqslant \frac{\hat\sigma^2}{2}\,\delta\ ,$$

where $M = \max_{y\in Y} f_0^2(y)$. Thus, for every $\delta > 0$ there exists an ε_δ such that for all $\varepsilon < \varepsilon_\delta$,

$$\max_{0\leqslant n\leqslant\lfloor\tau/\varepsilon\rfloor} \left|z_n^{(2)} - \frac{\hat\sigma^2}{2}\,t_n\right| \leqslant \frac{1}{2}\max(\tau, M, \hat\sigma^2)\,\delta$$

which concludes our proof.

5.2. Second example

We now consider a generalization of the previous example. Let $h : \mathbb{R} \to \mathbb{R}$ be a twice continuously differentiable function, and consider the following discrete processes:

$$\begin{aligned} &y_{n+1} = Ty_n\ , \\ &\frac{x_{n+1} - x_n}{\varepsilon} = \frac{f_0(y_n)}{\sqrt{\varepsilon}}\ , \\ &\frac{z_{n+1} - z_n}{\varepsilon} = h'(x_n)\frac{f_0(y_n)}{\sqrt{\varepsilon}}\ , \end{aligned} \tag{5.3}$$

where $x_0 = 0$ and $z_0 = 0$; T and f_0 satisfy the assumptions of Theorem 2.1. By the same argument as above, we expect z_n to weakly converge to $Z(t)$ satisfying the SDE:

$$\mathrm{d}Z(t) = h'(\sigma B)\,\sigma\,\mathrm{d}B, \qquad Z(0) = 0\ . \tag{5.4}$$

For Itô and Stratonovich SDEs the solution are

$$\begin{aligned} &Z_I(t) = h(\sigma B(t)) - h(0) - \frac{\sigma^2}{2}\int_0^t h''(\sigma B(s))\,\mathrm{d}s\ , \\ &Z_S(t) = h(\sigma B(t)) - h(0)\ . \end{aligned}$$

Here again, the relevant stochastic calculus is of neither type; we will show that the piecewise-linear interpolant $z(t)$ weakly converges to $Z(t)$ given by

$$Z(t) = h(\sigma B(t)) - h(0) - \frac{\hat\sigma^2}{2}\int_0^t h''(\sigma B(s))\,\mathrm{d}s\ , \tag{5.5}$$

where $\hat\sigma$ is the same as in Section 5.1.

Consider the discrete function $h(x_n)$. By Taylor's expansion, and using (5.3), we get

$$\frac{h(x_{n+1})-h(x_n)}{\varepsilon}=h'(x_n)\frac{f_0(y_n)}{\sqrt{\varepsilon}}+\tfrac{1}{2}h''(x_n)f_0^2(y_n)+\mathcal{O}(\varepsilon^{1/2})\,. \tag{5.6}$$

A comparison of (5.6) with the equation for z_n shows that

$$z_n=h(x_n)-h(x_0)-\frac{1}{2}\sum_{k=0}^{n-1}\varepsilon h''(x_k)f_0^2(y_k)+\mathcal{O}(\varepsilon^{3/2})\,,$$

which we rewrite as follows:

$$z_n=z_n^{(1)}-z_n^{(2)}-z_n^{(3)}+\mathcal{O}(\varepsilon^{3/2})\,,$$

where

$$z_n^{(1)}=h(x_n)-h(x_0)\,,$$

$$z_n^{(2)}=\frac{\hat\sigma^2}{2}\sum_{k=0}^{n-1}\varepsilon\, h''(x_k)\,,$$

$$z_n^{(3)}=\frac{1}{2}\sum_{k=0}^{n-1}\varepsilon\, h''(x_k)[f_0^2(y_k)-\hat\sigma^2]\,.$$

This decomposition is also valid for the piecewise-linear interpolants.

The function $z^{(1)}(t)$ weakly converges to $h(\sigma B(t))-h(0)$ by the same argument as in the previous section. The function $z^{(2)}(t)$ weakly converges to the integral,

$$\frac{\hat\sigma^2}{2}\int_0^t h''(\sigma B(s))\,\mathrm{d}s$$

as can easily be shown by a continuity argument. In general the limit of a sum of two weakly convergent sequences does not equal the sum of their limits. But because $x\to(x,x)$ is a continuous operation from $C[0,\tau]$ to $C[0,\tau]\times C[0,\tau]$, $z^{(1)}(t)$ and $z^{(2)}(t)$ are both continuous functions of $x(t)$, and addition is a continuous operation from $C[0,\tau]\times C[0,\tau]$ to $C[0,\tau]$, it follows that

$$z^{(1)}-z^{(2)}\Rightarrow h(\sigma B)-h(0)-\frac{\hat\sigma^2}{2}\int_0^t h''(\sigma B(s))\,\mathrm{d}s\,.$$

To prove the validity of (5.5) it remains to show that $z^{(3)}$ converges to zero *in probability*, i.e., that for all $\delta>0$,

$$\lim_{\varepsilon\to 0}\mathbb{P}\left\{\max_{0\leqslant n\leqslant\lfloor\tau/\varepsilon\rfloor}\left|\frac{1}{2}\sum_{k=0}^{n-1}\varepsilon\, h''(x_k)[f_0^2(y_k)-\hat\sigma^2]\right|\geqslant\delta\right\}=0\,. \tag{5.7}$$

This is proved in Appendix B.

5.3. Third example

Consider the following process:

$$y_{n+1} = Ty_n ,$$

$$\frac{z_{n+1} - z_n}{\varepsilon} = \lambda z_n + z_n \frac{f_0(y_n)}{\sqrt{\varepsilon}} ,$$

where λ is a constant. Here we expect z_n to weakly converge to a solution of an SDE

$$\mathrm{d}Z = \lambda Z\,\mathrm{d}t + \sigma Z\,\mathrm{d}B, \qquad Z(0) = z_0 .$$

Here the Itô and Stratonovich interpretations give

$$Z_I(t) = z_0 \exp[(\lambda - \sigma^2/2)t + \sigma B(t)]$$

$$Z_S(t) = z_0 \exp[\lambda t + \sigma B(t)] .$$

Instead, we obtain that $z(t)$ weakly converges to $Z(t)$ given by

$$Z(t) = z_0 \exp[(\lambda - \hat{\sigma}^2/2)t + \sigma B(t)] \tag{5.8}$$

with $\hat{\sigma}$ the same as in the first two examples.

The equation for z_n being linear, it can be solved explicitly,

$$z_n = z_0 \prod_{k=0}^{n-1} [1 + \lambda\varepsilon + f_0(y_k)\sqrt{\varepsilon}] .$$

Taking the logarithm and expanding in powers of ε:

$$\begin{aligned} \log z_n &= \log z_0 + \sum_{k=0}^{n-1} \log[1 + \lambda\varepsilon + f_0(y_k)\sqrt{\varepsilon}] \\ &= \log z_0 + \sum_{k=1}^{n-1} \left[\lambda\varepsilon + f_0(y_k)\sqrt{\varepsilon} - \frac{1}{2} f_0^2(y_k)\varepsilon + \mathcal{O}(\varepsilon^{3/2}) \right] \\ &= \log z_0 + \lambda t_n + x_n - \frac{1}{2} \sum_{k=1}^{n-1} f_0^2(y_k)\varepsilon + \mathcal{O}(\varepsilon^{1/2}) . \end{aligned}$$

The first two terms on the right-hand side are non-random; the third term converges weakly to $\sigma B(t)$; the fourth term converges in probability to $(\hat{\sigma}^2/2)t$; the fifth term (surely) converges uniformly to zero. Hence, the limits can be added, and the linear interpolant of $\log z_n$ weakly converges to $\log z_0 + \lambda t + \sigma B - (\hat{\sigma}^2/2)t$. By continuity, (5.8) follows.

6. Discussion

(1) This paper studies SDE limits of discrete dynamical systems with scale separation of form (1.4). The discrete system is inspired by the continuous system (1.3), and

has the advantage of allowing explicit examples for which the assumed mixing properties can be verified. As argued below, the discrete setup plays a critical role in the determination of the limiting dynamics.

(2) While we were able to prove a quite general convergence theorem for additive noise, our treatment of multiplicative noise is restricted to three examples. Thus, general systems of the form

$$\frac{x_{n+1}-x_n}{\varepsilon} = f_1(x_n,t_n) + \beta(x_n,t_n)\frac{f_0(y_n)}{\sqrt{\varepsilon}} \;,$$
$$y_{n+1} = Ty_n \tag{6.1}$$

are beyond the scope of this paper. Note that if the "white noise" term, $f_0(y_n)/\sqrt{\varepsilon}$ is replaced by i.i.d. normal variables $\mathcal{N}(0,\sigma^2\varepsilon^{-1})$, then (6.1) is simply the Euler–Maruyama scheme [17] for the Itô SDE,

$$\mathrm{d}X = f_1(X,t)\,\mathrm{d}t + \sigma\beta(X,t)\,\mathrm{d}B(t) \;.$$

Then, the piecewise-linear interpolant, $x(t)$, of x_n converges strongly (in probability) in $C[0,\tau]$ to $X(t)$. If $f_0(y_n)$ are i.i.d., say, ± 1 (i.e., Brownian motion is approximated by random walk), which is a simpler situation than mixing deterministic dynamics, then weak convergence is the only possible mode of convergence. In Ref. [17] it is proven that under mild restrictions on f_0 and β,

$$\lim_{\varepsilon\to 0} \mathbb{E}[g(x(t))] = \mathbb{E}[g(Z(t))]$$

for all $t\in[0,\tau]$ and all sufficiently smooth functions g. Bally and Talay [18] generalize this theorem to functions g that are measurable and bounded.

(3) In the three examples studied in Section 5, the resulting stochastic calculus was found to be intermediate between Itô and Stratonovich. The reason for that is as follows: Itô's calculus assumes that Brownian increments are independent of the current and past state of the system. In our case, where the Brownian increments are approximated by a function of a mixing process, this independence is not satisfied. The factor that determines the limiting stochastic calculus was found out to be the auto-correlation of the sequence $f_0(y_n)$. If the elements of this sequence are uncorrelated (as is the case, for example, if we take $Y=[0,1]$, $Ty=2y \bmod 1$ and $f_0(y)=\cos(2\pi y)$), then $\hat\sigma=\sigma$, and the limit is of Itô type. In all other cases, $\hat\sigma\neq\sigma$, and we obtain a limit which is intermediate between Itô and Stratonovich.

(4) In view of the above, we conjecture that $x(t)$ solving (6.1) weakly converges in $C[0,\tau]$, under quite general conditions, to $X(t)$ solving

$$X(t) = X(0) + \int_0^t f_1(X(s),s)\,\mathrm{d}s + (M)\int_0^t \beta(X(s),s)\,\sigma\,\mathrm{d}B(s) \;,$$

where the stochastic integral $(M)\int$ is defined by the following combination of Itô and Stratonovich integrals:

$$(M)\int_0^t \mathrm{d}B = \alpha(I)\int_0^t \mathrm{d}B + (1-\alpha)(S)\int_0^t \mathrm{d}B \;,$$

and $\alpha = \hat\sigma^2/\sigma^2$. The possible range of α is $(0,\infty)$.

(5) Suppose that rather than constructing discrete equations for the slow dynamics, we kept the y dynamics discrete, with the slow dynamics governed by a differential equation. For example, the approximate Brownian motion could be defined as the solution of the ODE

$$\frac{dx}{dt} = \frac{f_0(y_{\lfloor t/\varepsilon \rfloor})}{\sqrt{\varepsilon}} .$$

It is easy to see that $x(t)$ converges weakly to $\sigma B(t)$. We can then proceed to construct differential systems with multiplicative noise. Consider for example the case

$$\frac{dz}{dt} = h'(x)\,\frac{f_0(y_{\lfloor t/\varepsilon \rfloor})}{\sqrt{\varepsilon}} = h'(x)\,\frac{dx}{dt} .$$

Integration by parts, followed by a continuity argument, shows that $z(t)$ converges weakly to $h(\sigma B)$, i.e. to a Stratonovich interpretation of the stochastic integral. This is not surprising given that Brownian motion is approximated here by a piecewise differentiable function [19].

(6) In general, the Itô interpretation of the stochastic integral is expected to prevail when the correlation time of the driving noise is fast compared to the relaxation time of the rate of change (i.e., velocity) of the process. Conversely, a Stratonovich interpretation is expected to prevail when the driving noise seems smooth on the scale over which the velocity relaxes. Intermediate interpretations are likely to occur when both relaxation times are of the same order. In the present paper, because the slow dynamics are governed by a discrete process, we have a velocity relaxation time of order ε, which is also the timescale of noise correlation.

A detailed study of Itô versus Stratonovich limits when inertia interacts with noise correlation is undertaken by Pavliotis and Stuart [12]. While their study covers a range of different situations, here we only interpret their results within the scope of the present paper. Consider, for example, the second-order ODE with multiplicative noise:

$$\varepsilon^{\gamma}\frac{d^2x}{dt^2} + \frac{dx}{dt} = f(x)\,\frac{f_0(y_{\lfloor t/\varepsilon \rfloor})}{\sqrt{\varepsilon}}$$

with $\gamma > 0$ (in Ref. [12] the setup is stochastic, and the noise term is an Ornstein–Uhlenbeck process with correlation time of order ε). As $\varepsilon \to 0$, one may be tempted to drop the inertial term, and speculate that $x(t)$ weakly converges to the solution $X(t)$ of the SDE

$$dX = \sigma f(X)\, dB . \tag{6.2}$$

The analysis in Ref. [12] shows that when $\gamma < 1$, i.e., the velocity correlation time ε^{γ} is much longer, for small ε, than the noise correlation time, then the limit satisfies (6.2), interpreted in the sense of Itô. When $\gamma > 1$ the noise is smooth on the scale of the velocity correlation time and a Stratonovich correction appears. When $\gamma = 1$, the two times are of the same order, and an intermediate limit holds. Note that we study a first order system, i.e., there is no explicit inertia; the inertial effect is an artifact of the discretization, which uses ε for time step.

Acknowledgements

We are grateful to Noam Berger for elucidating many of the technical aspects of this work, and to Jonathan Mattingly, Carlo Nitsch and Denis Talay for useful discussions. Special thanks to Andrew Stuart for pointing out the relation with Ref. [12] and for his comments on the manuscript. This research was supported in part by the Israel Science Foundation founded by the Israel Academy of Sciences and Humanities, and in part by the Applied Mathematical Sciences subprogram of the Office of Energy Research of the US Department of Energy under Contract DE-AC03-76-SF00098.

Appendix A

Consider $x_n = x_n^{\mathrm{B}}$ given by (1.5), and its piecewise-linear interpolant, $x(t) = x^{\mathrm{B}}(t)$, which weakly converges to $\sigma B(t)$, with σ given by (2.2). Let $h(x) : \mathbb{R} \to \mathbb{R}$ be a continuous function, $z_n = h(x_n)$, and $z(t)$ be its piecewise-linear interpolant. The continuity implies that $h(x(t))$ weakly converges to $h(\sigma B(t))$, however it is not clear whether $z(t)$ weakly converges to $h(\sigma B(t))$ as well.

As in Theorem 2.1, $\xi_n(y) = \xi(T^n y)$ is the nth digit of $\xi_T(y)$, $\mathscr{B}^K = \sigma(\xi_0, \xi_1, \ldots, \xi_K)$ is the σ-algebra generated by the first K digits of $\xi_T(y)$, and $f_0^K = \mathbb{E}[f_0|\mathscr{B}^K]$ is the best approximation of $f_0(y)$ in $L^2(Y, \mathscr{B}, \mathbb{P})$ by a $\mathscr{B}^K$-measurable function. Note that the sequence $f_0^K(y_n)$ is K-dependent, that is, $\{f_0^K(y_i), \ldots, f_0^K(y_l)\}$ and $\{f_0^K(y_{l+n}), \ldots, f_0^K(y_j)\}$ are independent whenever $n > K$, hence it is a ϕ-mixing process.

Proposition A.1. *Assume that there exist $a > 1, K_0 > 0$ such that for all $K > K_0$,*

$$\sup_{y \in Y} |f_0(y) - f_0^K(y)| < a^{-K}. \tag{A.1}$$

If $h'(x)$ is bounded by exponential growth, that is, $|h'(x)| \leqslant c e^{\alpha|x|}$ for some $\alpha, c > 0$, then

$$\lim_{\varepsilon \to 0} \mathbb{P}\left\{ \sup\nolimits_{0 \leqslant t \leqslant \tau} |z(t) - h(x(t))| > \delta \right\} = 0 \tag{A.2}$$

for all $\delta > 0$.

Comments. (1) This proposition determines conditions under which the uniform distance between $z(t)$ and $h(x(t))$ converges to zero in probability. Since $h(x(t))$ weakly converges to $h(\sigma B(t))$, it follows that $z(t)$ weakly converges to $h(\sigma B(t))$.

(2) Note that the assumptions in Theorem 2.1 require a bound on the distance in $L^2(Y, \mathscr{B}, \mathbb{P})$ between $f_0(y)$ and $f_0^K(y)$, while here an $L^\infty(Y, \mathscr{B}, \mathbb{P})$ bound is assumed. Note also that we require exponential convergence while in Theorem 2.1 the requirement is less stringent.

(3) This proposition also holds for $\sup_{y \in Y} |f_0(y) - f_0^K(y)| < K^{-a}$, $a > 1$ and $h'(x)$ bounded by polynomial growth, $|h'(x)| \leqslant c|x|^p$, $p < 2a$.

Proof. Let $0 \leqslant t \leqslant \tau$ and $\varepsilon > 0$. The fact that $h(x(t))$ and $z(t)$ share the same values at integer multiples of ε implies,

$$\begin{aligned}|z(t) - h(x(t))| &\leqslant |z(t) - h(x_{\lfloor t/\varepsilon \rfloor})| + |h(x_{\lfloor t/\varepsilon \rfloor}) - h(x(t))| \\ &= (t/\varepsilon - \lfloor t/\varepsilon \rfloor)|h(x_{\lfloor t/\varepsilon \rfloor+1}) - h(x_{\lfloor t/\varepsilon \rfloor})| + |h(x_{\lfloor t/\varepsilon \rfloor}) - h(x(t))| \\ &= (t/\varepsilon - \lfloor t/\varepsilon \rfloor)\{|h'(x(\theta_1))| + |h'(x(\theta_2))|\}|f_0(y_{\lfloor t/\varepsilon \rfloor})|\sqrt{\varepsilon}\,,\end{aligned}$$

where we used the mean value theorem twice with $\lfloor t/\varepsilon \rfloor \leqslant \theta_1, \theta_2 \leqslant \lfloor t/\varepsilon \rfloor + 1$. Let $\theta \in \{\theta_1, \theta_2\}$ be such that it maximizes $|h'(x(\theta))|$, then

$$|z(t) - h(x(t))| \leqslant 2|h'(x(\theta)) f_0(y_{\lfloor t/\varepsilon \rfloor})|\sqrt{\varepsilon}\,.$$

The boundedness on the growth rate of $h'(x)$ and the fact that $x(t)$ assumes its local extrema on the mesh points implies that (A.2) follows if

$$\lim_{\varepsilon \to 0} \mathbb{P}\left\{\sup_{0 \leqslant t \leqslant \tau} |h'(x_{\lfloor t/\varepsilon \rfloor}) f_0(y_{\lfloor t/\varepsilon \rfloor})\sqrt{\varepsilon}| > \delta\right\} = 0$$

for all $\delta > 0$. Using the boundedness of $f_0(y)$ and the exponential bound, it is sufficient to show that

$$\lim_{\varepsilon \to 0} \mathbb{P}\left\{\max_{0 \leqslant k \leqslant \lfloor \tau/\varepsilon \rfloor} |x_k| > \log \frac{\delta}{\sqrt{\varepsilon}}\right\} = 0$$

for all $\delta > 0$. Define $x_k^K = \sum_{i=0}^{k-1} f_0^K(y_i)\sqrt{\varepsilon}$. (A.1) guarantees that

$$|x_k^K - x_k| \leqslant k\, a^{-K}\sqrt{\varepsilon} \leqslant \frac{\tau}{\sqrt{\varepsilon}}\, a^{-K} \tag{A.3}$$

which implies,

$$\begin{aligned}\mathbb{P}\left\{\max_{0 \leqslant k \leqslant \lfloor \tau/\varepsilon \rfloor} |x_k| > \log \frac{\delta}{\sqrt{\varepsilon}}\right\} &= \mathbb{P}\left\{\max_{0 \leqslant k \leqslant \lfloor \tau/\varepsilon \rfloor} |x_k - x_k^K + x_k^K| > \log \frac{\delta}{\sqrt{\varepsilon}}\right\} \\ &\leqslant \mathbb{P}\left\{\max_{0 \leqslant k \leqslant \lfloor \tau/\varepsilon \rfloor} |x_k^K| + |x_k - x_k^K| > \log \frac{\delta}{\sqrt{\varepsilon}}\right\} \\ &\leqslant \mathbb{P}\left\{\max_{0 \leqslant k \leqslant \lfloor \tau/\varepsilon \rfloor} |x_k^K| > \log \frac{\delta}{\sqrt{\varepsilon}} - \frac{\tau}{\sqrt{\varepsilon}}\, a^{-K}\right\}\,.\end{aligned}$$

Because x_k^K is the sum of variables which are K-dependent, we can split it into a double sum such that each of the inner sums is over independent variables; this exploits the fact that $f_0^K(y_{i+jK})$ are i.i.d. for different values of j. Thus,

$$\begin{aligned}&\mathbb{P}\left\{\max_{0 \leqslant k \leqslant \lfloor \tau/\varepsilon \rfloor} |x_k| > \log \frac{\delta}{\sqrt{\varepsilon}}\right\} \\ &= \mathbb{P}\left\{\max_{0 \leqslant k \leqslant \lfloor \tau/\varepsilon \rfloor} \left|\sum_{i=0}^{K-1} \sum_{j=0}^{(k-K-2)/K} f_0^K(y_{i+jK})\sqrt{\varepsilon}\right| > \log \frac{\delta}{\sqrt{\varepsilon}} - \frac{\tau}{\sqrt{\varepsilon}}\, a^{-K}\right\} \\ &\leqslant \mathbb{P}\left\{\max_{0 \leqslant k \leqslant \lfloor \tau/\varepsilon \rfloor, 0 \leqslant i \leqslant K-1} \left|\sum_{j=0}^{(k-K-2)/K} f_0^K(y_{i+jK})\sqrt{\varepsilon K}\right| > \frac{\log \delta/\sqrt{\varepsilon} - (\tau/\sqrt{\varepsilon})a^{-K}}{\sqrt{K}}\right\}\,.\end{aligned}$$

For sums of i.i.d. variables the reflection principle [10] implies,

$$\mathbb{P}\left\{\max_{0\leqslant k\leqslant\lfloor\tau/\varepsilon\rfloor}|x_k|>\log\frac{\delta}{\sqrt{\varepsilon}}\right\}$$
$$\leqslant 2\mathbb{P}\left\{\max_{0\leqslant i\leqslant K-1}\left|\sum_{j=0}^{\lfloor\tau/\varepsilon\rfloor/K} f_0^K(y_{i+jK})\sqrt{\varepsilon K}\right|>\frac{\log\delta/\sqrt{\varepsilon}-(\tau/\sqrt{\varepsilon})a^{-K}}{\sqrt{K}}\right\}. \tag{A.4}$$

If we choose $K=K(\varepsilon)$ such that

$$\varepsilon K\to 0 \tag{A.5}$$

then $\sum_{j=0}^{\lfloor\tau/\varepsilon\rfloor/K} f_0^K(y_{i+jK})\sqrt{\varepsilon K}$ weakly converges to a Gaussian variable, which implies that for all ε, $\left[\sum_{j=1}^{\lfloor\tau/\varepsilon\rfloor/K} f_0^K(y_{i+jK})\sqrt{\varepsilon K}\right]^2$ are uniformly integrable. If furthermore,

$$\log\frac{\delta}{\sqrt{\varepsilon}}-\frac{\tau}{\sqrt{\varepsilon}}a^{-K}>0\,, \tag{A.6}$$

then Chebyshev's inequality applied to (A.4) implies

$$\mathbb{P}\left\{\max_{1\leqslant k\leqslant\lfloor\tau/\varepsilon\rfloor}|x_k|>\log\frac{\delta}{\sqrt{\varepsilon}}\right\}\leqslant\frac{CK}{[\log\delta/\sqrt{\varepsilon}-(\tau/\sqrt{\varepsilon})a^{-K}]^2} \tag{A.7}$$

for some $C>0$. If we set $K(\varepsilon)=-\log\varepsilon$, then (A.5) and (A.6) are satisfied for sufficiently small ε, and it is easy to see that the right-hand side of (A.7) tends to zero as $\varepsilon\to 0$. □

Appendix B

Proposition B.1. *Assume that there exists an $a>0$ such that for K sufficiently large*

$$\sup_{y\in Y}|f_0(y)-f_0^K(y)|\leqslant a^{-K}\,, \tag{B.1}$$

$$\sup_{y\in Y}|f_0^2(y)-(f_0^K(y))^2|\leqslant a^{-K} \tag{B.2}$$

and that $g(x)$ is Lipschitz with constant L. Then

$$\lim_{\varepsilon\to 0}\mathbb{P}\left\{\max_{0\leqslant k\leqslant\lfloor\tau/\varepsilon\rfloor}\left|\sum_{i=0}^{k} g(x_i)[f_0^2(y_i)-\hat{\sigma}^2]\varepsilon\right|\geqslant\delta\right\}=0 \tag{B.3}$$

for all $\delta>0$.

Proof. Define

$$u_k=f_0^2(y_k)-\mathbb{E}f_0^2=f_0^2(y_k)-\hat{\sigma}^2\,,$$
$$u_k^K=(f_0^K(y_k))^2-\mathbb{E}(f_0^K)^2\,,$$

where K is selected a posteriori. Using these notations we have,

$$\left|\sum_{i=0}^{k} g(x_i)[f_0^2(y_i) - \hat{\sigma}^2]\varepsilon\right| = \left|\sum_{i=0}^{k} g(x_i)u_i\varepsilon\right|$$
$$\leqslant \left|\sum_{i=0}^{k} [g(x_i) - g(x_i^K)]u_i\varepsilon\right| + \left|\sum_{i=0}^{k} g(x_i^K)(u_i - u_i^K)\varepsilon\right|$$
$$+ \left|\sum_{i=0}^{k} g(x_i^K)u_i^K\varepsilon\right| \quad \text{(B.4)}$$

which implies that

$$\mathbb{P}\left\{\max_{1\leqslant k\leqslant \lfloor\tau/\varepsilon\rfloor} \left|\sum_{i=0}^{k} g(x_i)[f_0^2(y_{i+1}) - \hat{\sigma}^2]\varepsilon\right| \geqslant \delta\right\}$$
$$\leqslant \mathbb{P}\left\{\max_{1\leqslant k\leqslant \lfloor\tau/\varepsilon\rfloor} \left|\sum_{i=0}^{k} [g(x_i) - g(x_i^K)]u_i\varepsilon\right| \geqslant \frac{\delta}{3}\right\}$$
$$+ \mathbb{P}\left\{\max_{1\leqslant k\leqslant \lfloor\tau/\varepsilon\rfloor} \left|\sum_{i=0}^{k} g(x_i^K)(u_i - u_i^K)\varepsilon\right| \geqslant \frac{\delta}{3}\right\}$$
$$+ \mathbb{P}\left\{\max_{1\leqslant k\leqslant \lfloor\tau/\varepsilon\rfloor} \left|\sum_{i=0}^{k} g(x_i^K)u_i^K\varepsilon\right| \geqslant \frac{\delta}{3}\right\} = I_1 + I_2 + I_3 \,. \quad \text{(B.5)}$$

Consider the first two terms. From (B.1),

$$|x_k^K| \leqslant \sum_{i=0}^{k-1} |f_0^K(y_i)|\sqrt{\varepsilon} \leqslant kc\sqrt{\varepsilon} \leqslant \tau c\frac{1}{\sqrt{\varepsilon}} \,,$$
$$|x_k - x_k^K| \leqslant \sum_{i=0}^{k-1} |f_0(y_i) - f_0^K(y_i)|\sqrt{\varepsilon} \leqslant \sqrt{\varepsilon}ka^{-K} \leqslant \frac{1}{\sqrt{\varepsilon}}\tau a^{-K} \,, \quad \text{(B.6)}$$

where $c=\max\{\sup_{y\in Y}|f_0(y)|, \sup_{y\in Y} f_0^2(y), \sup_{y\in Y}|f_0^K(y)|\}$. Using the Lipschitz property of $g(x)$,

$$|g(x_k^K)| \leqslant L|x_k^K| \leqslant L\tau c\frac{1}{\sqrt{\varepsilon}} \,,$$
$$|g(x_k) - g(x_k^K)| \leqslant L|x_k - x_k^K| \leqslant L\tau\frac{1}{\sqrt{\varepsilon}}\, a^{-K} \,. \quad \text{(B.7)}$$

From (B.2) we get, on the other hand,

$$|u_k - u_k^K| \leqslant |f_0^2(y_k) - (f_0^K(y_k))^2| + \mathbb{E}|f_0^2(y_k) - (f_0^K(y_k))^2| \leqslant 2a^{-K} \,. \quad \text{(B.8)}$$

Substituting (B.6)–(B.8) into (B.5) and using the fact that k is at most τ/ε, we get

$$I_1 \leqslant \mathbb{P}\left\{\frac{\tau}{\varepsilon}\,\varepsilon 2cL\tau\,\frac{a^{-K}}{\sqrt{\varepsilon}} \geqslant \frac{\delta}{3}\right\}\ ,$$

$$I_2 \leqslant \mathbb{P}\left\{\frac{\tau}{\varepsilon}\,\varepsilon\,2L\tau c\,\frac{a^{-K}}{\sqrt{\varepsilon}} \geqslant \frac{\delta}{3}\right\}\ .$$

By choosing $K = -\log_a \varepsilon$ we get both I_1 and I_2 to vanish for sufficiently small ε.

Consider now I_3. Note that $\sum_{i=0}^{k} g(x_i^K)u_i^K\varepsilon$ is the sum of K-dependent variables. As in Appendix A, sums of K-dependent sequences can be divided into double sums, where each of the inner sums is over independent variables.

$$\sum_{i=0}^{k} g(x_i^K)u_i^K\varepsilon = \sum_{j=0}^{K-1}\ \sum_{i=0}^{(k-K-1)/K} g(x_{iK+j}^K)u_{iK+j}^K\varepsilon\ .$$

Define

$$g_i = \mathrm{sgn}(g(x_{iK+j}^K))\min(|g(x_{iK+j}^K)|, -\log\varepsilon)\ ,$$

which is the minimum (in absolute value) between $g(x_{iK+j}^K)$ and $(-\log\varepsilon)$. The following calculation shows that for small ε, g_i and $g(x_{iK+j}^K)$ are close in probability. Recall that $g(x)$ is Lipschitz, hence it is bounded by a linear growth rate and therefore,

$$\begin{aligned}\mathbb{P}\{\exists i \text{ s.t. } g_i \neq g(x_{iK+j}^K)\} &= \mathbb{P}\left\{\max_i |g(x_{iK+j}^K)| > -\log\varepsilon\right\}\\ &\leqslant \mathbb{P}\left\{\max_i |x_{iK+j}^K| > \frac{-\log\varepsilon}{L}\right\}\ .\end{aligned}$$

Since the x_i^K weakly converge to Brownian motion, and the maximum is a continuous function on $C[0,\tau]$, it follows that the last expression can be estimated by an analogous expression with x_i^K replaced by Brownian motion,

$$\begin{aligned}\mathbb{P}\{\exists i \text{ s.t. } g_i \neq g(x_{iK+j}^K)\} &\leqslant 2\,\mathbb{P}\left\{\max_{0\leqslant t\leqslant\tau}|B(t)| > \frac{-\log\varepsilon}{L}\right\}\\ &\leqslant 4\,\mathbb{P}\left\{|B(\tau)| > \frac{-\log\varepsilon}{L}\right\}\\ &\leqslant \frac{c\,\tau L^2}{\log^2\varepsilon}\ ,\end{aligned}$$

where we have used the reflection principle and the Chebyshev's inequality. The right-hand side tends to zero for small ε, thus, we can replace everywhere $g(x_{iK+j}^K)$ by g_i as the difference can be made arbitrarily small (in probability).

Define now $W_k = \sum_{i=0}^{k} g_i u_{iK+j}^K\varepsilon$. It is a martingale with respect to the filtration generated by the $\mathscr{B}^{iK+j}$ as by the K-dependence of the y_i^K,

$$\mathbb{E}[g_i u_{iK+j}^K|\mathscr{B}^{(i-1)K+j}] = 0\ .$$

Then W_k satisfy,

$$|W_k - W_{k-1}| = |g_k u^K_{kK+j}\varepsilon| \leqslant (-\log\varepsilon)\, 2c\varepsilon\ .$$

We then use the following inequality:

Lemma B.1 (Azuma and Hoeffding [20,21]). *Let* $0 = X_0, \ldots, X_m$ *be a martingale with*

$$|X_{k+1} - X_k| \leqslant C$$

for all $0 \leqslant k < m$. *Then for all* $\lambda > 0$,

$$\mathbb{P}[X_m > \lambda C\sqrt{m}] < \mathrm{e}^{-\lambda^2/2}\ .$$

Using this inequality and the fact that $(k-K-1)/K < \tau/\varepsilon$,

$$\begin{aligned}
&\mathbb{P}\left\{\left|\sum_{i=0}^{(k-K-1)/K} g_i u^K_{iK+j}\varepsilon\right| > \varepsilon^{1/4}\right\} \\
&\quad \leqslant \mathbb{P}\left\{\left|\sum_{i=0}^{(k-K-1)/K} g_i u^K_{iK+j}\varepsilon\right| > \frac{\varepsilon^{3/4}}{\sqrt{\tau}}\,\frac{(-\log\varepsilon)2c\varepsilon}{(-\log\varepsilon)2c\varepsilon}\sqrt{\frac{k-K-1}{k}}\right\} \\
&\quad \leqslant \exp\left[-\frac{1}{2}\left(\frac{\varepsilon^{3/4}}{\sqrt{\tau}(-\log\varepsilon)2c\varepsilon}\right)\right] \\
&\quad \leqslant \exp\left(-\frac{1}{8c\tau\sqrt{\varepsilon}\log^2\varepsilon}\right)\ .
\end{aligned}$$

Thus, for sufficiently small ε,

$$\begin{aligned}
I_3 &= \mathbb{P}\left\{\max_{1\leqslant k\leqslant\lfloor\tau/\varepsilon\rfloor}\left|\sum_{i=0}^{k} g_i u^K_i\varepsilon\right| > \frac{\delta}{3}\right\} \\
&\leqslant \mathbb{P}\left\{\max_{1\leqslant k\leqslant\lfloor\tau/\varepsilon\rfloor}\left|\sum_{j=0}^{K}\sum_{i=0}^{(k-K-1)/K} g_i u^K_{iK+j}\varepsilon\right| > \varepsilon^{1/4}\right\} \\
&\leqslant K\exp\left(-\frac{1}{8c\tau\sqrt{\varepsilon}\log^2\varepsilon}\right)
\end{aligned}$$

which converges to zero as $\varepsilon \to 0$.

References

[1] J. Sanders, F. Verhulst, Averaging Methods in Nonlinear Dynamical Systems, Springer, New York, 1985.

[2] P. Lochak, C. Meunier, Multiple Phase Averaging for Classical Systems, Springer, New York, 1988.

[3] D. Givon, R. Kupferman, A. Stuart, Extracting macroscopic dynamics: model problems and algorithms, Nonlinearity, 2003, submitted for publication.

[4] T. Kurtz, A limit theorem for perturbed operator semigroups with applications to random evolutions, J. Funct. Anal. 12 (1973) 55–67.
[5] G. Papanicolaou, W. Kohler, Asymptotic theory of mixing stochastic ordinary differential equations, Comm. Pure Appl. Math. XXVII (1974) 641–668.
[6] G. Papanicolaou, Some probabilistic problems and methods in singular perturbations, Rocky Mountain J. Math. 6 (1976) 653–673.
[7] A. Majda, I. Timofeyev, E. Vanden-Eijnden, A mathematical framework for stochastic climate models, Comm. Pure Appl. Math. LIV (2001) 891–947.
[8] C. Beck, G. Roepstorff, From dynamical systems to the Langevin equation, Physica A 145 (1987) 1–14.
[9] C. Beck, Brownian motion from deterministic dynamics, Physica A 169 (1990) 324–336.
[10] P. Billingsley, Convergence of Probability Measures, 2nd Edition, Wiley, New York, 1999.
[11] G. Uhlenbeck, L. Ornstein, On the theory of Brownian motion, Phys. Rev. 36 (1930) 823–841.
[12] G. Pavliotis, A. Stuart, Itô versus Stratonovich white noise limits, preprint 2003.
[13] R. Kupferman, G. Pavliotis, A. Stuart, Ito versus Stratonovich white noise limits for systems with inertia and colored multiplicative noise, 2003, submitted to Phys. Rev. E.
[14] R. Graham, A. Schenzle, Stabilization by multiplicative noise, Phys. Rev. A 26 (1982) 1676–1685.
[15] H. Royden, Real Analysis, The Macmillan Company, New York, 1963.
[16] E. Wong, M. Zakai, Riemann-Stieltjes approximations of stochastic integrals, Z. Wahr. Verb. Geb. 120 (1969) 87–97.
[17] P. Kloeden, E. Platen, Numerical Solution of Stochastic Differential Equations, Springer, Berlin, 1992.
[18] V. Bally, D. Talay, The Euler scheme for stochastic differential equations: error analysis with Malliavin calculus, Math. Comput. Simulation 38 (1995) 35–41.
[19] E. Wong, M. Zakai, On the convergence of ordinary integrals to stochastic integrals, Ann. Math. Statist. 36 (1965) 1560–1564.
[20] W. Hoeffding, Probability inequalities for sums of bounded random variables, J. Amer. Statist. Assoc. 58 (1963) 13–30.
[21] K. Azuma, Weighted sums of certain dependent random variables, Tohoku Math. J. 19 (1967) 357–367.