LONG-TERM BEHAVIOUR OF LARGE MECHANICAL SYSTEMS WITH RANDOM INITIAL DATA

R. KUPFERMAN

Institute of Mathematics, The Hebrew University, Jerusalem 91904 Israel

A. M. STUART and J. R. TERRY

Mathematics Institute, University of Warwick, Coventry CV4 7AL, England

P. F. TUPPER

*Department of Mathematics and Statistics, McGill University,
Montréal, Québec, H3A 2K6, Canada*

We study the long-time behaviour of large systems of ordinary differential equations with random data. Our main focus is a Hamiltonian system which describes a distinguished particle attached to a large collection of heat bath particles by springs. In the limit where the size of the heat bath tends to infinity, the trajectory of the distinguished particle can be weakly approximated, on finite time intervals, by a Langevin stochastic differential equation. We examine the long-term behaviour of these trajectories, both analytically and numerically. We find ergodic behaviour manifest in both the long-time empirical measures and in the resulting auto-correlation functions.

Keywords: Hamiltonian systems; stochastic differential equations; weak convergence; ergodicity.

AMS Subject Classification:

1. Introduction

In many applications of molecular dynamics the desired information is low dimensional, even when the governing model equations contain a large number of degrees of freedom. An important example is the extraction of conformational dynamics for biomolecules. Conformational dynamics may be stochastic in nature, even when the underlying model is deterministic [1]. Since numerical methods are typically needed to study the equations of motion [2], it is clearly important to be able to evaluate them by their ability to extract the correct macroscopic information. This applies both to straightforward time-stepping methods, and to more sophisticated methods such as those proposed in [3, 4]. In this context the development of simple model problems is important: the equations of motion for a bio-molecule exhibiting

conformational dynamics are extremely complex and not always appropriate for thorough investigation of algorithmic capability.

In [5] a variant of the Ford–Kac model of a heat bath [6] was used to study standard time-stepping algorithms by their ability to correctly predict macroscopic quantities. It is of interest to use similar models to study the more sophisticated algorithms in [3, 4]. However, the set-up of the model problem in [5] is not appropriate for the study of large-time dynamics because of a periodicity inherent in the construction. The primary purpose of this paper is to propose and study a related class of model problems which are suitable for the study of long-time dynamics. These model problems have the form of deterministic differential equations with emergent stochastic dynamics. Some of the models proposed here have already been used in [7] to evaluate the algorithm in [3].

In Sec. 2 we introduce a Hamiltonian system which will be the primary focus of our study; it has the form of a distinguished particle attached by springs to a large number of heat bath particles. By eliminating the heat bath variables, an integro-differential equation (IDE) is found for the distinguished particle. Section 3 contains some basic analysis of Monte–Carlo approximations of integrals and stochastic processes, enabling an understanding of the memory kernel and forcing which appear in the IDE. In Sec. 4 we show that the distinguished particle in the Hamiltonian system can be approximated by the solution of a Langevin SDE. All the approximation results in Secs. 3 and 4 are weak convergence results on finite time intervals. Section 5 contains some analysis of the large-time behaviour of the Monte Carlo approximation of an Ornstein–Uhlenbeck (OU) process introduced in Sec. 3. This analysis forms the motivation for the numerical experiments in Sec. 6 which elucidate the long-time behaviour of large Hamiltonian systems of ODEs introduced in Sec. 2. The results show a close relationship between these ODEs and their approximating SDEs, when comparisons are made for empirical measures and autocovariance functions, and when the limiting SDE is ergodic.

The derivation of SDEs from ODEs has a long history [8, 9]. Because of the explicit nature of our problem this abstract machinery, which has been developed considerably over the last few decades [10], is not required for the analysis undertaken here; we make use of straightforward tools from weak convergence of probability measures [11]. Our main contribution is to construct an explicit family of model problems which are useful in the study of algorithms for macroscopic properties of molecular systems. Through analysis and numerical experiments we establish the properties of these model problems.

2. The Hamiltonian System

The model problem under investigation is a Hamiltonian system defined by the Hamiltonian,

$$H(Q_n, P_n, q, p) = \frac{1}{2}P_n^2 + V(Q_n) + \sum_{j=1}^{n} \frac{p_j^2}{2m_j} + \sum_{j=1}^{n} \frac{k_j}{2}(q_j - Q_n)^2, \qquad (2.1)$$

where Q_n, P_n are the position and momentum of a distinguished particle of unit mass in a potential field $V(\cdot)$. The vectors $q = (q_1, q_2, \ldots, q_n)$ and $p = (p_1, p_2, \ldots, p_n)$ are the coordinates and momenta of n particles that are referred to as "heat bath" particles. The jth heat bath particle has mass m_j and interacts with the distinguished particle via a linear spring with stiffness constant k_j. If the distinguished particle was held fixed it would be the anchor point of n independent oscillators with frequencies $\omega_j = (k_j/m_j)^{1/2}$. This model is a variant of the well-known Ford, Kac, Mazur model [12, 6] (see also [13, 14]). The subscript n in Q_n, P_n labels the size of the heat bath (it should not be confused with the subscript n in (q_n, p_n) which refers to the nth heat bath particle); this subscript is introduced to have a convenient notation when comparing systems of variable size.

Hamilton's equations of motion are

$$\dot{Q}_n = P_n\,,$$

$$\dot{P}_n = -V'(Q_n) + \sum_{j=1}^{n} k_j(q_j - Q_n)\,,$$

$$\dot{q}_j = p_j/m_j\,,$$

$$\dot{p}_j = -k_j(q_j - Q_n)\,,$$

(2.2)

supplemented with initial conditions $Q_n(0) = Q_0$, $P_n(0) = P_0$, $q_j(0) = q_j^0$, and $p_j(0) = p_j^0$. It is further assumed that the initial data for the heat bath particles are randomly drawn from a Gibbs distribution with inverse temperature β. The Gibbs measure is conditioned by the (non-random) initial data Q_0 and P_0. For fixed Q_n, P_n the Hamiltonian (2.1) is quadratic in q, p, and hence the corresponding measure is Gaussian. It is easily verified that

$$q_j^0 = Q_0 + (1/\beta k_j)^{1/2}\xi_j\,,$$

$$p_j^0 = (m_j/\beta)^{1/2}\eta_j\,,$$

where $\xi_j, \eta_j \sim \mathcal{N}(0, 1)$ are mutually independent sequences of i.i.d. random variables.

System (2.2) is a model problem for the situation when a particle interacts with a system of many degrees of freedom (a heat bath) having a broad and dense spectrum. In this context it is natural to choose the parameters k_j and m_j such that, as n increases, the set of ω_j covers an increasingly large range of frequencies in an increasingly dense manner. A simple choice that satisfies this requirement is to take the frequencies ω_j random and uniformly distributed in $[0, n^a]$,

$$\omega_j = n^a \nu_j\,, \qquad \nu_j \text{ i.i.d.}\,, \qquad \nu_1 \sim \mathcal{U}[0, 1]\,,$$

for some $a \in (0, 1)$. Another alternative is to take the ω_j non-random and equally distributed in $[0, n^a]$. Both choices are considered below.

Initially we choose the spring constants k_j as follows:

$$k_j = f^2(\omega_j)\,\Delta\omega\,, \qquad f^2(\omega) = \frac{2\alpha}{\pi\tilde{\beta}}\frac{1}{\alpha^2 + \omega^2}\,, \tag{2.3}$$

with $\alpha, \tilde{\beta} > 0$ and $\Delta\omega = n^a/n$ being the mean spectral density. The reason for this choice and notation will become apparent in the next section, where generalizations of 2.3 are also considered.

The probability space is induced by the three mutually independent sequences of i.i.d. random variables ν_j, ξ_j and η_j. In several instances below, we derive properties of the system that hold almost surely with respect to certain variables but are of statistical nature with respect to the remaining variables. As a result, we need distinct notations for integration over the ν, ξ and η components of the probability space. For example, we will denote by $\mathbb{E}_\nu$ the expected values with respect to the ν variables only, by $\mathbb{E}_{\xi\eta}$ the expected values with respect to ξ and η variables, and by $\mathbb{E}_{\nu\xi\eta}$ the expected values with respect to the entire probability space; when no confusion should arise we will use the shorter notation $\mathbb{E}$. Similarly, Var_ν denotes the variance with respect to the ν variables, $\mathbb{P}_\nu$ denotes probability with respect to the ν variables and so on.

The variables q_j, p_j in (2.2) can be integrated explicitly, giving rise to an inhomogeneous integro-differential equation for the distinguished particle trajectory $Q_n(t)$:

$$\ddot{Q}_n(t) + \int_0^t K_n(t-s)\dot{Q}_n(s)\, ds + V'(Q_n(t)) = Z_n(t)\,, \tag{2.4}$$

where

$$K_n(t) = \sum_{j=1}^n f^2(\omega_j) \cos(\omega_j t)\, \Delta\omega\,, \tag{2.5}$$

$$Z_n(t) = \beta^{-1/2} \sum_{j=1}^n f(\omega_j)\, [\xi_j\, \cos(\omega_j t) + \eta_j\, \sin(\omega_j t)]\, (\Delta\omega)^{1/2}\,. \tag{2.6}$$

Equation (2.4) is a projection of the $(2n+2)$-dimensional system (2.2) onto the two-dimensional subspace (Q_n, P_n). It describes the rate of change of (Q_n, P_n) as function of their present and past values. The history dependence is encapsulated by the memory kernel $K_n(t)$. The function $Z_n(t)$ is a forcing that depends, for fixed ν_j, on the initial data ξ_j, η_j. It is a stationary zero-mean Gaussian process; its autocovariance function satisfies the so-called fluctuation–dissipation relation

$$\mathbb{E}_{\xi\eta}[Z_n(t)Z_n(s)] = \beta^{-1}\, K_n(t-s)\,;$$

this holds for every choice of frequencies. Equation (2.4) is an instance of the Mori–Zwanzig projection formalism [15–18], and is also known as a generalized Langevin equation.

3. Weak Convergence of the Forcing

In this section we study the $n \to \infty$ limit of the memory kernel $K_n(t)$ and the forcing $Z_n(t)$; see Theorem 3.1 and Corollary 3.2. As part of this analysis we also investigate

the convergence of the kernel $K_n(t)$. Before undertaking a rigorous analysis we note that $K_n(t)$, given by (2.5) with $f(\omega)$ given by (2.3), can be viewed as a Monte–Carlo approximation of the integral,

$$\frac{2\alpha}{\pi\tilde{\beta}} \int_0^{n^a} \frac{\cos(\omega t)}{\alpha^2 + \omega^2}\, d\omega\,,$$

and this integral tends to the Fourier cosine representation of $\tilde{\beta}^{-1}e^{-\alpha t}$ as $n \to \infty$. Similarly, $Z_n(t)$, given by 2.6, can be viewed as a Monte–Carlo approximation of the stochastic integral

$$\left(\frac{2\alpha}{\pi\beta\tilde{\beta}}\right)^{1/2} \left[\int_0^{n^a} \frac{\cos(\omega t)}{(\alpha^2 + \omega^2)^{1/2}}\, dB_1(\omega) + \int_0^{n^a} \frac{\sin(\omega t)}{(\alpha^2 + \omega^2)^{1/2}}\, dB_2(\omega) \right],$$

where $B_1(\omega)$ and $B_2(\omega)$ are independent Brownian motions. (Note that the Itô and Stratonovich interpretations of this integral are the same.) In view of the form of the spectral density it follows that, as $n \to \infty$, this stochastic integral tends to the stationary Ornstein–Uhlenbeck process $U(t)$ with covariance $(\beta\tilde{\beta})^{-1}e^{-\alpha t}$ (e.g. Grimmett and Stirzaker [19], p. 407). Recall that $U(t)$ solves the stochastic differential equation

$$
\begin{aligned}
dU(t) &= -\alpha U(t)\, dt + (2\alpha/\beta\tilde{\beta})^{1/2} dB(t)\,, \\
U(0) &\sim (\beta\tilde{\beta})^{-1/2}\mathcal{N}(0,1)\,,
\end{aligned}
\tag{3.1}
$$

where $B(t)$ is standard Brownian motion independent of $U(0)$. These heuristic arguments are made rigorous in the remaining part of this section. Throughout this section it will be assumed that $\omega_j = n^a\nu_j$, $a \in (0,1)$, and $\Delta\omega = n^a/n$, with $\nu_j \sim \mathcal{U}[0,1]$ and $\xi_j, \eta_j \sim \mathcal{N}(0,1)$ being three mutually independent sequences of i.i.d. random variables. We also make occasional references to the straightforward alternative of choosing the frequencies nonrandom and equally distributed in $[0, n^a]$.

We start with two useful lemmas:

Lemma 3.1. *Suppose that h is a bounded, positive, real-valued function with bound c_1, such that $h(\omega) \leq c_2\, \omega^{-1}$ for some $c_2 > 0$. Let*

$$S_n = \sum_{j=1}^n h(\omega_j)\, \Delta\omega\,. \tag{3.2}$$

Then,

$$\mathbb{P}\left(\lim_{n\to\infty} |S_n - \mathbb{E}_\nu S_n| = 0 \right) = 1\,. \tag{3.3}$$

Proof. It is sufficient to show that there exists an integer b such that the sequence

$$\sigma_n = \mathbb{E}_\nu |S_n - \mathbb{E}_\nu S_n|^{2b}$$

is summable: then, by the Borel–Cantelli lemma (Billingsley [20], p. 59),

$$\mathbb{P}\left(|S_n - \mathbb{E}_\nu S_n|^{2b} > \varepsilon \text{ i.o.} \right) = 0$$

6   *R. Kupferman et al.*

for any $\varepsilon > 0$ (i.o. = infinitely often), which in turn implies (3.3).

We rewrite σ_n as follows:

$$\sigma_n = \mathbb{E}_\nu \left| \frac{n^a}{n} \sum_{j=1}^n [h(\omega_j) - \mathbb{E}_\nu h(\omega_j)] \right|^{2b} = n^{2b(a-1)} \sum_{j_1=1}^n \cdots \sum_{j_{2b}=1}^n V_{j_1,j_2,\ldots,j_{2b}}, \quad (3.4)$$

where

$$V_{j_1,j_2,\ldots,j_{2b}} = \mathbb{E}_\nu \left\{ [h(\omega_{j_1}) - \mu_n][h(\omega_{j_2}) - \mu_n] \ldots [h(\omega_{j_{2b}}) - \mu_n] \right\} \quad (3.5)$$

are the centred joint moments of degree $2b$ of $h(\omega_j)$, $j = 1, 2, \ldots, n$, and $\mu_n = \mathbb{E}_\nu h(\omega_1)$. Note that our assumptions on h imply

$$\mu_n \le n^{-a}(c_1 + c_2 \log n^a). \quad (3.6)$$

Since $h(\omega_j)$ and $h(\omega_i)$ are independent for $i \ne j$, then many of the joint moments vanish; specifically, every $V_{j_1,j_2,\ldots,j_{2b}}$ that contains an index which appears only once vanishes. To estimate σ_n we regroup the $2b$-tuple sum (3.4) by the number k of distinct indices in (3.5); k assumes values from 1 to b because each index must occur at least twice, otherwise (3.5) is zero. The number of terms corresponding to a given k can be bounded by $c_3 n^k$, where $c_3 > 0$ is a constant that depends on b, but not on n nor k (there are n^k ways to "decode" a k-letter pattern with an n-letter alphabet). Each of the $V_{j_1,j_2,\ldots,j_{2b}}$ which corresponds to a given k is of the form,

$$V_{j_1,j_2,\ldots,j_{2b}} = \left(\frac{1}{n^a} \int_0^{n^a} [h(\omega) - \mu_n]^{m_1} \, d\omega \right) \cdots \left(\frac{1}{n^a} \int_0^{n^a} [h(\omega) - \mu_n]^{m_k} \, d\omega \right),$$

where $m_1, m_2, \ldots, m_k \ge 2$ and $m_1 + m_2 + \cdots + m_k = 2b$. Now,

$$\frac{1}{n^a} \int_0^{n^a} [h(\omega) - \mu_n]^m \, d\omega = \sum_{\ell=0}^m \binom{m}{\ell} (-\mu_n)^{m-\ell} \frac{1}{n^a} \int_0^{n^a} h^\ell(\omega) \, d\omega \,.$$

The terms $\ell = 0, 1$ sum up to $(-m+1)(-\mu_n)^m$, which by (3.6), tends to zero faster than n^{-a} because $m \ge 2$. For $\ell = 2, 3, \ldots, m$ the integral over h^ℓ converges, by our assumptions on h. Thus, there exists a constant $M > 1$ such that for all n sufficiently large $| \int_0^{n^a} [h(\omega) - \mu_n]^m \, d\omega | \le M$, and

$$|V_{j_1,j_2,\ldots,j_{2b}}| \le \left(\frac{M}{n^a} \right)^k. \quad (3.7)$$

Combining (3.4) and (3.7) we get, for n sufficiently large,

$$\sigma_n \le n^{2b(a-1)} \sum_{k=1}^b c_3 n^k \left(\frac{M}{n^a} \right)^k \le b \, c_3 M^b n^{2b(a-1)} n^{b(1-a)} \,,$$

which is summable if we take $b > 1/(1 - a)$. This completes the proof. $\qquad \square$

Lemma 3.2. *Let g be a real-valued function on $[0, \infty)$. Assume that g is bounded, and that there exist constants $c > 0$ and $\lambda > 1/2$ such that $g(\omega) \leq c/\omega^\lambda$. Define*

$$
\begin{aligned}
R_n(t) &= \sum_{j=1}^{n} g^2(\omega_j) \cos(\omega_j t) \, \Delta\omega \, , \\
R(t) &= \int_0^\infty g^2(\omega) \cos(\omega t) \, d\omega \, .
\end{aligned}
\tag{3.8}
$$

Then, for any $T \in (0, \infty)$,

1. *ν-almost surely, $R_n(t)$ converges pointwise to $R(t)$, $t \in [0, T]$;*
2. *$R_n \to R$ in $L^2(\Omega, L^2[0, T])$.*

Proof. Let $t \in [0, T]$. We start by showing that $R_n(t)$ converges pointwise to $R(t)$ in the mean square. We write

$$
\mathbb{E}_\nu |R_n(t) - R(t)|^2 = \mathbb{E}_\nu |R_n(t) - \mathbb{E}_\nu R_n(t)|^2 + |R(t) - \mathbb{E}_\nu R_n(t)|^2.
\tag{3.9}
$$

The first term on the right-hand side is the variance of $R_n(t)$ which can be estimated by

$$
\mathrm{Var}_\nu R_n(t) = n \Delta\omega^2 \, \mathrm{Var}_\nu [g^2(\omega_j) \cos(\omega_j t)] \leq \frac{n^{2a}}{n} \frac{1}{n^a} \int_0^{n^a} g^4(\omega) \, d\omega \leq C_1 \, \Delta\omega \, ,
\tag{3.10}
$$

where the boundedness of the integral $C_1 = \int_0^\infty g^4(\omega) \, d\omega < \infty$ follows from our assumptions on g. To estimate the second term we note that the expected value of $R_n(t)$ is

$$
\mathbb{E}_\nu R_n(t) = \int_0^{n^a} g^2(\omega) \cos(\omega t) \, d\omega \, ,
$$

thus

$$
|R(t) - \mathbb{E}_\nu R_n(t)| \leq \int_{n^a}^\infty g^2(\omega) \, d\omega \leq \int_{n^a}^\infty \left(\frac{c}{\omega^\lambda} \right)^2 \, d\omega = \frac{c^2}{(2\lambda - 1)} \frac{1}{n^{a(2\lambda - 1)}} \, .
\tag{3.11}
$$

Substituting (3.10) and (3.11) into (3.9) gives

$$
\mathbb{E}_\nu |R_n(t) - R(t)|^2 \leq \frac{C_1}{n^{1-a}} + \frac{C_2}{n^{2a(2\lambda - 1)}} \, ,
\tag{3.12}
$$

where $C_2 = c^4/(2\lambda - 1)^2$.

Equation (3.12) implies that $R_n \to R$ in $L^2(\Omega, L^2[0, T])$ as:

$$
\|R_n - R\|_{L^2(\Omega, L^2[0,T])}^2 = \int_0^T \mathbb{E}_\nu |R_n(t) - R(t)|^2 \, dt \leq \frac{TC_1}{n^{1-a}} + \frac{TC_2}{n^{2a(2\lambda - 1)}} \, .
\tag{3.13}
$$

Finally, $\mathbb{E}_\nu R_n(t)$ converges pointwise to $R(t)$ by 3.11, whereas Lemma 3.1 with $h(\omega) = g^2(\omega) \cos(\omega t)$ implies that $R_n(t) - \mathbb{E}_\nu R_n(t)$ decays to zero ν-almost surely. We conclude that ν-almost surely $R_n(t)$ converges pointwise to $R(t)$. $\qquad\square$

Comments. 1. The rate of convergence of R_n is expected to be optimally fast if a is chosen to balance the two terms in 3.13, which suggests the choice of $a = [1 + 2(2\lambda - 1)]^{-1}$.

2. Convergence in $L^2(\Omega, L^2[0,T])$ implies convergence in $L^2(\Omega, L^1[0,T])$, which in turn implies convergence in probability in $L^1[0,T]$. This weaker form of convergence is sufficient for establishing the weak convergence of the trajectories Q_n; see Sec. 4.

3. For ω_j deterministic and equally spaced, $\omega_j = j\,\Delta\omega$, $R_n(t)$ can be viewed as an approximation to $R(t)$ by quadrature. It is straightforward to show that in this case R_n converges to R both pointwise and in $L^1[0,T]$.

Verifying that $f(\omega)$ given by 2.3 satisfies the assumptions on g in Lemma 3.2 with $\lambda = 1$, we conclude:

Corollary 3.1. *Let $K_n(t)$ be given by (2.5), then*

1. *For any t on a finite, but arbitrary, time interval $[0,T]$, and ν-almost surely, $K_n(t)$ converges pointwise to*

$$K(t) = \frac{2\alpha}{\pi\tilde{\beta}} \int_0^\infty \frac{\cos(\omega t)}{\alpha^2 + \omega^2}\, d\omega = \tilde{\beta}^{-1} e^{-\alpha t}\,.$$

2. *$K_n \to K$ in $L^2(\Omega, L^2[0,T])$.*

The rate of convergence is expected to be optimal if we take $a = 1/3$.

The next theorem establishes the weak convergence of random series of the form (2.6). The convergence of $Z_n(t)$ to the stationary Ornstein–Uhlenbeck process $U(t)$ follows as an immediate consequence. Weak convergence, here, is in the probability space induced by the random variables ξ_j and η_j, and holds for almost every choice of frequencies (ν-almost surely). Thus "$\Rightarrow$" denotes weak convergence with respect to the sequences ξ_j, η_j (for a general reference on weak convergence see Billingsley [11]).

Theorem 3.1. *Let $g(t)$ be a real-valued function satisfying the assumptions of Lemma 3.2, with $R_n(t)$ and $R(t)$ given by (3.8). Define*

$$Y_n(t) = \sum_{j=1}^n g(\omega_j)[\xi_j\,\cos(\omega_j t) + \eta_j\,\sin(\omega_j t)](\Delta\omega)^{1/2}\,. \tag{3.14}$$

Then, ν-almost surely, $Y_n \Rightarrow Y$ in $C[0,T]$ where Y is the stationary Gaussian process with mean zero and autocovariance function $R(t)$.

Proof. The proof relies on the following theorem ([21], p. 450): Let Y_n be a collection of real-valued almost-surely continuous stochastic processes on $[0,T]$, such that:

1. The finite dimensional distributions of Y_n weakly converge to those of an almost-surely continuous process Y.
2. Tightness: there exist positive constant b, γ, M_1 such that for all n

$$\mathbb{E}|Y_n(t+u) - Y_n(t)|^b \leq M_1|u|^{1+\gamma}\,.$$

Then $Y_n \Rightarrow Y$.

Let Y be the stationary Gaussian process with autocovariance function $R(t)$. By Kolmogorov's continuity condition ([22], p. 53) applied to Gaussian processes, Y has a continuous version if its autocovariance function is Hölder continuous; the Hölder continuity of $R(t)$ follows from the assumed rate of decay of g. To show that the finite-dimensional distributions of Y_n converge weakly to those of Y it is sufficient to show that for any collection of times $0 \leq t_1 < t_2 < \cdots < t_k \leq T$, the joint probability density functions of $(Y_n(t_1), \ldots, Y_n(t_k))$ converge pointwise to the joint probability density function of $(Y(t_1), \ldots, Y(t_k))$. For Gaussian processes this is guaranteed by the pointwise convergence of $R_n(t)$ to $R(t)$, which was established in Lemma 3.2.

It remains to show the tightness property. Let $\theta \in (0, 2\lambda - 1] \cap (0, 2]$, and let b be a sufficiently large integer such that $b\theta = 1 + \gamma$ for some $\gamma > 0$. Then

$$\begin{aligned}
\mathbb{E}_\nu|Y_n(t+u) - Y_n(t)|^{2b} &= (2b-1)!! \left(\mathbb{E}_\nu|Y_n(t+u) - Y_n(t)|^2\right)^b \\
&= 2^b(2b-1)!! \left[R_n(u) - R_n(0)\right]^b \\
&= 2^{2b}(2b-1)!! \left(\sum_{j=1}^n g^2(\omega_j)\sin^2(\tfrac{1}{2}\omega_j u)\,\Delta\omega\right)^b \\
&\leq 2^{2b}(2b-1)!! \left(\sum_{j=1}^n g^2(\omega_j)|\tfrac{1}{2}\omega_j u|^\theta\,\Delta\omega\right)^b \\
&= 2^{b(2-\theta)}(2b-1)!!\,|u|^{b\theta} \left(\sum_{j=1}^n g^2(\omega_j)\omega_j^\theta\,\Delta\omega\right)^b \\
&\leq [2^{b(2-\theta)}(2b-1)!!\,M^b]|u|^{1+\gamma}\,,
\end{aligned}$$

where $(2b-1)!! = 1 \cdot 3 \cdot 5 \cdots (2b-1)$. In the passage from the first to the second line we used the Gaussian property of Y_n and its stationarity. The third line was obtained by substituting the expression (3.8) for R_n. To obtain the fourth line we used the inequality $\sin^2(x) \leq |x|^\theta$, valid for any $\theta \leq 2$. Finally, ν-almost surely there exist a positive constant M such that for any n

$$\sum_{j=1}^n g^2(\omega_j)\omega_j^\theta\,\Delta\omega \leq M\,.$$

This follows from Lemma (3.1), with $h(\omega) = g^2(\omega)\omega^\theta$, which implies that the left hand side converges ν-almost surely to $\int_0^\infty g^2(\omega)\omega^\theta \, d\omega$, which, in turn, is finite by our assumptions on g and θ. This completes the proof. $\square$

Since $f(\omega)$, given by (2.3), satisfies the assumptions on g in Theorem 3.1, and that $\beta^{-1}K(t) = (\beta\tilde{\beta})^{-1}e^{-\alpha t}$ is the autocovariance of the stationary OU process, we conclude:

Corollary 3.2. *Let $Z_n(t)$ be given by (2.6) and let $U(t)$ be the stationary OU process defined by (3.1). Then ν-almost surely $Z_n \Rightarrow U$ in $C[0,T]$.*

Comments. 1. It is straightforward to show weak convergence of Z_n to U in $C[0,T]$ for the case of deterministic, equally spaced frequencies.

2. $Y_n(t)$ can be made to approximate delta-correlated white noise by taking $\tilde{\beta} = \alpha^{-1}$ in (2.3). This corresponds to a memory kernel $K(t) = \alpha \, e^{-\alpha t}$; a limit to a delta function in the sense of distributions is obtained by letting $\alpha \to \infty$.

3. Theorem 3.1 can be easily extended to non-stationary processes. For example,

$$Y_n(t) = \left(\frac{2}{\pi}\right)^{1/2} \sum_{j=1}^{n} \frac{\xi_j}{\omega_j} \sin(\omega_j t) \, (\Delta\omega)^{1/2}$$

can be shown, by similar techniques, to weakly approximate standard Brownian motion in $C[0,T]$.

4. Convergence of Q_n and the Limiting SDE

In this section we show that the $L^1[0,T]$ convergence of K_n to K (in probability) and the weak convergence of Z_n to U in $C[0,T]$ implies the weak convergence in $C^2[0,T]$ of the distinguished particle trajectory Q_n to a limiting process Q. Furthermore we derive an SDE whose trajectories have the same distribution as Q. Recall that here weak convergence is in the probability space induced by the variables ξ_j, η_j.

Theorem 4.1. *Let $Q_n(t)$ be the solution to the randomly-driven IDE (2.4), with $K_n(t)$ and $Z_n(t)$ given by (2.3), (2.5) and (2.6). Assume that $V'(\cdot)$ is globally Lipschitz. Then ν-almost surely Q_n converges weakly in $C^2[0,T]$ to the solution Q of the stochastic IDE*

$$\ddot{Q}(t) + \int_0^t K(t-s)\dot{Q}(s) \, ds + V'(Q(t)) = U(t), \qquad Q(0) = Q_0, \dot{Q}(0) = P_0, \quad (4.1)$$

where $U(t)$ is the OU process (3.1) and $K(t) = \tilde{\beta}^{-1}e^{-\alpha t}$.

Moreover, Q solving (4.1) is equivalent to Q solving the SDE

$$
\begin{aligned}
dQ &= P \, dt, & Q(0) &= Q_0, \\
dP &= [R - V'(Q)] \, dt, & P(0) &= P_0, \qquad\qquad (4.2) \\
dR &= (-\alpha R - \tilde{\beta}^{-1}P) \, dt + (2\alpha/\beta\tilde{\beta})^{1/2} \, dB, & R(0) &\sim (\beta\tilde{\beta})^{-1/2}\mathcal{N}(0,1),
\end{aligned}
$$

where $B(t)$ is standard Brownian motion.

Proof. Corollaries 3.1 and 3.2 imply that ν-almost surely

$$K_n \to K \text{ in probability in } L^1[0, T], \qquad Z_n \Rightarrow U \text{ in } C[0, T].$$

Theorem 4.4 in Billingsley [11] shows that

$$(K_n, Z_n) \Rightarrow (K, U) \text{ in } L^1[0, T] \times C[0, T].$$

Thus the required result follows if we can prove that the mapping $(K, U) \mapsto Q$ defined by

$$\ddot{Q}(t) + \int_0^t K(t - s)\dot{Q}(s)\, ds + V'(Q(t)) = U(t) \tag{4.3}$$

is a continuous mapping from $L^1[0, T] \times C[0, T]$ to $C^2[0, T]$ (weak convergence is preserved under continuous mappings). Integrating the convolution term by parts, introducing a new kernel $\mathcal{K}(t) = \int_0^t K(s)\, ds$, Eq. (4.3) becomes

$$\ddot{Q}(t) + \int_0^t \mathcal{K}(t - s)\ddot{Q}(s)\, ds + V'(Q(t)) = U(t) - P_0\mathcal{K}(t), \tag{4.4}$$

which is a nonlinear equation of Volterra type for $\ddot{Q}(t)$ with continuous kernel $\mathcal{K}(t)$ and continuous forcing $U(t) - P_0\mathcal{K}(t)$. Since $V'(\cdot)$ is globally Lipschitz a straightforward Picard argument gives the existence and uniqueness of $\ddot{Q} \in C[0, T]$ solving (4.4), or equivalently, the existence and uniqueness of $Q \in C^2[0, T]$ solving (4.1). The continuity of $\ddot{Q}$ on $\mathcal{K} \in L^1[0, T]$ and $U - P_0\mathcal{K} \in C[0, T]$ is a standard result (see, for example, [24, 25] and Sec. 12 in [26]).

The equivalence between the stochastic differential system (4.2) and the stochastic integro-differential equation (4.1) follows from a straightforward integration of the first. Here, again, the Itô and Stratonovich interpretations are equivalent. □

Comments. 1. The set-up here is very similar to that used by Nakazawa [14] who also considers approximating an integro-differential stochastic equation via the same Hamiltonian system. The results there are comparable, though the techniques used to obtain them are different.

2. Variants of this model have also been studied in [5, 26–28]. There the forcing functions were approximations to distribution-valued processes, such as white noise, in which case the limiting solution is less regular than in the present case, and Q_n converges in $C^1[0, T]$.

3. In the present model a problem with nonlocal memory can be turned into a Markov process by the introduction of one extra variable $R(t)$. In the context of constructing deterministic model problems with emergent stochastic behaviour this fact is of practical importance: problems whose memory can be described by only a few additional variables constitute an important class where effective dimension reduction can be achieved.

4. For general memory kernels $K(t)$ it is not possible to convert the integro-differential equation into a Markovian system by the introduction of a finite number

of extra variables. Nevertheless, in many problems of interest, the memory kernel can be well approximated by a finite number of decay modes, each corresponding to a characteristic relaxation time of the system. (For an example of such construction in the context of polymeric fluids see [29], p. 262). For a memory kernel of the form

$$K(t) = \sum_{i=1}^{r} a_i e^{-\alpha_i t},$$

the integro-differential equation (4.1) can be converted into a Markovian system by the introduction of r extra variables.

Setting $\tilde{\beta} = (\alpha\gamma)^{-1}$ and letting $\alpha \to \infty$ shows that (Q, P) solving (4.2) converges to (Q, P) solving the standard Langevin equation without memory:

$$dQ = P\,dt\,, \qquad\qquad\qquad Q(0) = Q_0\,,$$
$$dP = -[\gamma P + V'(Q)]\,dt + (2\gamma/\beta)^{1/2}\,dB\,, \qquad P(0) = P_0\,.$$

Recall that in this limit the memory kernel approaches a delta function which is why the memory variable R drops out. Analysis justifying this limiting procedure may be found in [31] for weak convergence and [30] for strong convergence.

5. Long-Term Behaviour

In this section we study analytically the long-term behaviour of the processes (3.14) described in Sec. 3; in the next section the validity of our results are extended to the solution of the integro-differential equation (2.4) by means of numerical experiments.

To illustrate the problem under consideration, consider again the stationary OU process defined as the solution to the SDE

$$dU(t) = -\alpha U(t)\,dt + (2\alpha/\beta)^{1/2}\,dB(t)\,, \qquad U(0) \sim \beta^{-1/2}\,\mathcal{N}(0,1)\,,$$

where $B(t)$ is standard Brownian motion; here and below we take $\tilde{\beta} = 1$. In Sec. 3 it was shown that $U(t)$ was weakly approximated on any *finite* time interval by the following sum

$$U_n(t) = \left(\frac{2\alpha}{\pi\beta}\right)^{1/2} \sum_{j=1}^{n} \frac{\xi_j\,\cos(\omega_j t) + \eta_j\,\sin(\omega_j t)}{(\alpha^2 + \omega_j^2)^{1/2}}\,(\Delta\omega)^{1/2}\,, \qquad (5.1)$$

with ω_j, ξ_j, η_j and $\Delta\omega$ defined as in Sec. 3. It is well known that $U(t)$, $t \in [0, \infty)$ is an ergodic process that has a Gaussian distribution for its invariant measure (see [32], p. 121). Specifically, for any bounded, continuous function h

$$\overline{h(U(t))} \equiv \lim_{T\to\infty} \frac{1}{T} \int_0^T h(U(t))\,dt = \mathbb{E}[h(\zeta)]\,, \qquad \zeta \sim \beta^{-1/2}\,\mathcal{N}(0,1) \qquad (5.2)$$

on almost every trajectory; the overline is a short-hand notation for long time averages.

This result can be extended to more general functions of the stochastic process U, specifically functions that depend on U at more than one point of time. Let $h : \mathbb{R}^r \to \mathbb{R}$ be a bounded continuous function. Let $\{s_p\}_{p=1}^r$ be a sequence in $[0, \infty)$. Then it is known that

$$\overline{h(U(s_1 + t), U(s_2 + t), \ldots, U(s_r + t))}$$

$$\equiv \lim_{T \to \infty} \frac{1}{T} \int_0^T h(U(s_1 + t), U(s_2 + t), \ldots, U(s_r + t)) \, dt = \mathbb{E}[h(\zeta)], \quad (5.3)$$

on almost every trajectory, where ζ is a mean-zero Gaussian vector in $\mathbb{R}^r$ with covariance given by

$$\mathbb{E}[\zeta_p \zeta_q] = \beta^{-1/2} e^{-\alpha |s_q - s_p|} .$$

The latter expression is the autocovariance for the OU process sampled at times s_q and s_p.

We can now ask whether the approximate process $U_n(t)$ satisfies properties analogous to (5.2), (5.3). For both univariate and multivariate h, it will be shown below that:

1. The $T \to \infty$ limit of time averages exists for the approximate process $U_n(t)$, for any n.
2. This limit converges to that of the OU process as $n \to \infty$.

We will establish such ergodic properties for a more general family of stochastic processes. Our methods rely on the explicit trigonometric form of (3.14). In the next section, we extend our observations to the integro-differential equation of Sec. 2 by means of numerical experiments.

It is worth emphasizing that these results do not follow from the analyses of previous sections. Here we consider the behaviour of a single realization over long time intervals whilst previous results concern statistical properties with respect to a large set of realizations and are confined to fixed time intervals. In particular, previous sections do not give convergence uniformly in $[0, \infty)$. Thus, for any given n, the results of previous sections give no guarantee that the sample paths will resemble those of the limiting process for large times.

We investigate the family of random processes (3.14):

$$Y_n(t) = \sum_{j=1}^n g(\omega_j) \left[\xi_j \, \cos(\omega_j t) + \eta_j \, \sin(\omega_j t) \right] (\Delta \omega)^{1/2} ,$$

where g satisfies the assumptions of Lemma 3.2. We have shown that Y_n converges weakly in $C[0, T]$ to the stationary zero-mean Gaussian process Y with autocovariance

$$\mathbb{E}[Y(t)Y(s)] = \int_0^\infty g^2(\omega) \, \cos[\omega(t - s)] \, d\omega .$$

The weak convergence is in the probability space induced by the ξ and η variables, and holds ν-almost surely, i.e. for almost every collection of frequencies.

Let $h : \mathbb{R}^r \to \mathbb{R}$ be a bounded, continuous, real-valued function, and let $s_1, \ldots, s_r$ be a sequence in $[0, \infty)$. The univariate case can be obtained by setting $r = 1$ and $s_1 = 0$. Our first theorem establishes, for every $n \geq 1$, the almost-sure existence of the long-time average of $h(Y_n(s_1 + t), \ldots, Y_n(s_r + t))$. We denote this long time average by $\overline{h(\{Y_n(s_p + t)\}_{p=1}^r)}$. The result relies on the following classical result due to Weyl [33] (see [34], p. 286):

Lemma 5.1. *Let $H : \mathbb{T}^n \mapsto \mathbb{R}$ be a continuous function defined on the n-dimensional torus. Let $\{\omega_j\}_{j=1}^n$ be a set of independent numbers, in the sense that $\sum_{j=1}^n k_j \omega_j = 0$ for integer k_j implies $k_j \equiv 0$. Finally, let $\theta_j(t) = \omega_j t \bmod 2\pi$. Then*

$$\overline{H(\theta_1(t), \ldots, \theta_n(t))} = \frac{1}{(2\pi)^n} \int_0^{2\pi} \cdots \int_0^{2\pi} H(\varphi_1, \ldots, \varphi_n) \, d\varphi_1 \ldots, d\varphi_n \,.$$

Theorem 5.1. *T 5.1 Let $Y_n(t)$ be defined by (3.14) where g satisfies the assumptions of Lemma 3.2. Let h be continuous. Let $\{s_p\}_{p=1}^r$ be a strictly increasing sequence in $[0, \infty)$. Then $\{\nu, \xi, \eta\}$-almost surely:*

$$\overline{h(\{Y_n(s_p + t)\}_{p=1}^r)}$$

$$= \frac{1}{(2\pi)^n} \int_0^{2\pi} \cdots \int_0^{2\pi}$$

$$h\left(\left\{\sum_{j=1}^n g(\omega_j)(\xi_j \, \cos(\omega_j s_p + \varphi_j) + \eta_j \, \sin(\omega_j s_p + \varphi_j))(\Delta\omega)^{1/2}\right\}_{p=1}^r\right)$$

$$d\varphi_1 \ldots d\varphi_n \,. \tag{5.4}$$

Proof. The ω_j are ν-almost surely independent of the properties of Lebesgue measure. Since $\cos(\omega_j(s_p + t))$ and $\sin(\omega_j(s_p + t))$ are functions of $\theta_j(t) = \omega_j t \bmod 2\pi$, (5.4 is an immediate consequence of Lemma 5.1 with

$$H(\varphi_1, \ldots, \varphi_n)$$

$$= h\left(\left\{\sum_{j=1}^n g(\omega_j)(\xi_j \, \cos(\omega_j s_p + \varphi_j) + \eta_j \, \sin(\omega_j s_p + \varphi_j))(\Delta\omega)^{1/2}\right\}_{p=1}^r\right) . \tag{5.5}$$

$\square$

We now rewrite Eq. (5.4) in a slightly different form, appropriate for the subsequent analysis. Let the variables ν_j, ξ_j, and η_j be defined as before, and introduce, in addition, the i.i.d. auxiliary random variables $\varphi_j \sim \mathcal{U}[0, 2\pi]$. Then, define the vector $X_n = (X_{n,1}, \ldots, X_{n,r}) \in \mathbb{R}^r$ by

$$X_{n,p} = \sum_{j=1}^n g(\omega_j)(\xi_j \, \cos(\omega_j s_p + \varphi_j) + \eta_j \, \sin(\omega_j s_p + \varphi_j)) (\Delta\omega)^{1/2} \,. \tag{5.6}$$

Equation (5.4) may be written as follows:

$$\overline{h(\{Y_n(s_p + t)\}_{p=1}^r)} = \mathbb{E}_\varphi h\left(X_n\right),\tag{5.7}$$

where $\mathbb{E}_\varphi$ denotes averaging with respect to the φ variables only. We retain the notations $\mathbb{E}_\nu$, $\mathbb{E}_{\xi\eta}$, etc., for averaging with respect to the ν, ξ and η variables.

The next theorem asserts that $\{\nu, \xi, \eta\}$-almost surely, X_n converges weakly (in the probability space induced by the φ variables) to a Gaussian random vector. Note the difference with Theorem 3.1 where weak convergence occurs in the probability space induced by the variables ξ and η.

Theorem 5.2. *Let X_n be defined by (5.6). Then $\{\nu, \xi, \eta\}$-almost surely X_n converges weakly, as $n \to \infty$, to a Gaussian random vector, $\zeta \in \mathbb{R}^r$, with mean zero and covariance*

$$\mathbb{E}[\zeta_p \zeta_q] = \int_0^\infty g^2(\omega) \cos[\omega(s_p - s_q)]\, d\omega\,.\tag{5.8}$$

Proof. The random vectors (5.6) are of the form:

$$X_n = \sum_{j=1}^n x_j^{(n)}\,,\tag{5.9}$$

where the vectors $x_j^{(n)}$ have components $x_{j,p}^{(n)}$ defined by

$$x_{j,p}^{(n)} = g(\omega_j)\left[\xi_j \cos(\omega_j s_p + \varphi_j) + \eta_j \sin(\omega_j s_p + \varphi_j)\right](\Delta\omega)^{1/2}\,.\tag{5.10}$$

Thus the $x_j^{(n)}$ are are mutually independent random vectors. Contrary to Lemma 3.2, here we consider the ξ and η variables as fixed (in addition to the ν that were the only fixed variables in Lemma 3.2), and weak convergence is sought in the probability space induced by the auxiliary variables φ_j. Thus, in the current setup, (5.9) is a sum of non-Gaussian vectors, and we therefore resort to the central limit theorem.

Specifically, we use the multivariate Lindeberg–Feller theorem (see [35]):

Theorem. (Lindeberg-Feller) *For $n = 1, 2, \ldots$, let $z_j^{(n)}$, $j = 1, 2, \ldots, n$ be independent real-valued random vectors with $\mathbb{E}z_j^{(n)} = 0$, and let $Z_n = \sum_{j=1}^n z_j^{(n)}$. Let Σ be an $r \times r$ matrix. For a vector X, denote its Euclidean norm by $|X|$ and its transpose by X^{T}. Suppose that*

1. $$\lim_{n\to\infty} \mathbb{E}Z_n Z_n^{\mathrm{T}} = \lim_{n\to\infty} \sum_{j=1}^n \mathbb{E}z_j^{(n)} z_j^{(n)\mathrm{T}} = \Sigma\,.$$

2. *For all $\varepsilon > 0$*

$$\lim_{n\to\infty} \sum_{j=1}^n \mathbb{E}(|z_j^{(n)}|^2; |z_j^{(n)}| > \varepsilon) = 0\,,$$

where $\mathbb{E}(x; A)$ denotes the integral of x over the set A.

16   *R. Kupferman et al.*

Then, Z_n converges weakly to a Gaussian random vector with mean zero and co-variance Σ.

It remains to verify that the two conditions of the Lindeberg–Feller theorem are satisfied for $z_j^{(n)} = x_j^{(n)}$, with $\Sigma_{p,q}$ given by the right-hand side of (5.8). These conditions need to be fulfilled $\{\nu, \xi, \eta\}$-almost surely, while integration as referred to in the Lindeberg–Feller theorem is only over the φ variables. This verification is carried out in Lemmas 5.2 and 5.3 below. □

Lemma 5.2. *Let X_n be given by (5.9) with $x_j^{(n)}$ given by (5.10). Then $\{\nu, \xi, \eta\}$-almost surely*

$$\lim_{n \to \infty} \mathbb{E}_\varphi X_{n,p} X_{n,q} = \int_0^\infty g^2(\omega) \cos[\omega(s_p - s_q)] \, d\omega,$$

for $p, q = 1, \ldots, r$.

Proof. Integrating explicitly over the φ_j we get

$$\mathbb{E}_\varphi X_{n,p} X_{n,q} = \sum_{j=1}^n \mathbb{E}_\varphi x_{j,p}^{(n)} x_{j,q}^{(n)}$$

$$= \sum_{j=1}^n g^2(\omega_j) \left(\frac{1}{2} \xi_j^2 \cos[\omega_j(s_p - s_q)] + \frac{1}{2} \eta_j^2 \cos[\omega_j(s_p - s_q)] \right) \Delta\omega.$$

We then make the following three observations:

1. The expected value (with respect to ν, ξ and η) of $\mathbb{E}_\varphi X_{n,p} X_{n,q}$ is

$$\mathbb{E}_{\nu\xi\eta} \left(\mathbb{E}_\varphi X_{n,p} X_{n,q} \right) = \int_0^{n^a} g^2(\omega) \cos[\omega(s_p - s_q)] \, d\omega,$$

which converges, as $n \to \infty$, to $\int_0^\infty g^2(\omega) \cos[\omega(s_p - s_p)] \, d\omega$.

2. The variance (with respect to ν, ξ and η) of $\mathbb{E}_\varphi X_{n,p} X_{n,q}$ tends, as $n \to \infty$, to zero:

$$\mathrm{Var}_{\nu\xi\eta} \left(\mathbb{E}_\varphi X_{n,p} X_{n,q} \right) \qquad = \qquad n \, \mathrm{Var}_{\nu\xi\eta} \mathbb{E}_\varphi x_{1,p}^{(n)} x_{1,q}^{(n)}$$

$$\leq \qquad n \, \mathbb{E}_{\nu\xi\eta} \left(\mathbb{E}_\varphi x_{1,p}^{(n)} x_{1,q}^{(n)} \right)^2$$

$$\leq \qquad n \, \mathbb{E}_\nu \left(g^4(\omega_1) \right) \mathbb{E}_{\xi\eta} \left(\frac{1}{2} \xi_1^2 + \frac{1}{2} \eta_1^2 \right)^2 (\Delta\omega)^2$$

$$= \qquad n \left(\frac{1}{n^a} \int_0^{n^a} g^4(\omega) \, d\omega \right) \frac{1}{4} \left(3 + 2 + 3 \right) (\Delta\omega)^2$$

$$\leq C \, \Delta\omega \to 0,$$

where we have used the mutual independence of the random variables ν, ξ and η, and the boundedness of the integral of $g^4(\omega)$, which follows from our assumptions on g.

3. $\mathbb{E}_\varphi X_{n,p} X_{n,q} - \mathbb{E}_{\nu\xi\eta}(\mathbb{E}_\varphi X_{n,p} X_{n,q})$ tends $\{\nu, \xi, \eta\}$-almost surely to zero. The proof is similar to the proof of Lemma 3.1. It is sufficient to find an integer b such that

$$\sigma_n = \mathbb{E}_{\nu\xi\eta} |\mathbb{E}_\varphi X_{n,p} X_{n,q} - \mathbb{E}_{\nu\xi\eta} \mathbb{E}_\varphi X_{n,p} X_{n,q}|^{2b}$$

is summable, which we write as

$$\sigma_n = \mathbb{E}_{\nu\xi\eta} \left| \frac{n^a}{2n} \sum_{j=1}^{n} \left\{ g^2(\omega_j) \left[\xi^2 \cos[\omega_j(s_p - s_q)] + \eta^2 \cos[\omega_j(s_p - s_q)] \right] - \mu_n \right\} \right|^{2b},$$

where

$$\mu_n = \mathbb{E}_{\nu\xi\eta} \left\{ g^2(\omega_j) \left[\xi^2 \cos[\omega_j(s_p - s_q)] + \eta^2 \cos[\omega_j(s_p - s_q)] \right] \right\}.$$

The rest of the proof follows arguments very similar to those in Lemma 3.1 and so we omit the details.

Thus $\mathbb{E}_\varphi X_{n,p} X_{n,q}$ converges almost surely, and the limit has a vanishing variance, from which we conclude that $\mathbb{E}_\varphi X_{n,p} X_{n,q}$ converges almost surely to its mean value, which in turn converges to $\int_0^\infty g^2(\omega) \cos(\omega(s_p - s_q)) \, d\omega$. $\qquad \square$

Lemma 5.3. *Let X_n be given by (5.9) with $x_j^{(n)}$ given by (5.10). Then $\{\nu, \xi, \eta\}$-almost surely*

$$\lim_{n \to \infty} \sum_{j=1}^{n} \mathbb{E}_\varphi \left(|x_j^{(n)}|^2; |x_j^{(n)}| > \varepsilon \right) = 0$$

for all $\varepsilon > 0$.

Proof. Let g be bounded by the constant M. Then,

$$|x_j^{(n)}|^2 = \sum_{p=1}^{r} g^2(\omega_j)(\xi_j \cos(\omega_j s_p + \varphi_j) + \eta_j \sin(\omega_j s_p + \varphi))^2 \Delta\omega$$

$$\leq r M^2 \Delta\omega (|\xi_j| + |\eta_j|)^2$$

and

$$\mathbb{E}_\varphi(|x_j^{(n)}|^2; |x_j^{(n)}| > \varepsilon) \leq r M^2 \Delta\omega (|\xi_j| + |\eta_j|)^2 \mathbb{P}_\varphi \left(r^{1/2} M(|\xi_j| + |\eta_j|)(\Delta\omega)^{1/2} > \varepsilon \right),$$

and the last expression, which is the probability of an event independent of the variables φ, is simply the indicator function

$$\chi\{r^{1/2} M(|\xi_j| + |\eta_j|)(\Delta\omega)^{1/2} > \varepsilon\}.$$

Thus,

$$\mathbb{E}_\varphi(|x_j^{(n)}|^2; |x_j^{(n)}| > \varepsilon) \leq z_j,$$

where

$$z_j = r M^2 \Delta\omega (|\xi_j| + |\eta_j|)^2 \chi\{r^{1/2} M(|\xi_j| + |\eta_j|)(\Delta\omega)^{1/2} > \varepsilon\}.$$

It is sufficient to show that $\sum_{j=1}^{n} z_j$ tends to zero $\{\xi, \eta\}$-almost surely, and this is guaranteed if it decays sufficiently fast in the mean square. Indeed, using the independence of z_j,

$$
\mathbb{E}_{\xi\eta} \left| \sum_{j=1}^{n} z_j \right|^2 = \mathbb{E}_{\xi\eta} \left| \sum_{j=1}^{n} (z_j - \mathbb{E}_{\xi\eta} z_j) + n \, \mathbb{E}_{\xi\eta} z_j \right|^2
$$

$$
\leq n \, \mathbb{E}_{\xi\eta} |z_1|^2 + n^2 \, (\mathbb{E}_{\xi\eta} z_1)^2
$$

$$
= n \, r^2 M^4 (\Delta\omega)^2 \mathbb{E}_{\xi\eta}([|\xi_1| + |\eta_1|]^4; |\xi_1| + |\eta_1| > \varepsilon/r^{1/2} M(\Delta\omega)^{1/2})
$$

$$
+ n^2 \, r^2 M^4 (\Delta\omega)^2 \{([|\xi_1| + |\eta_1|]^2; |\xi_1| + |\eta_1| > \varepsilon/r^{1/2} M(\Delta\omega)^{1/2})\}^2 \,.
$$

By noting that $|\xi_1| + |\eta_1|$ has an exponentially decaying tail, we can estimate

$$
\mathbb{E}_{\xi\eta} \left([|\xi_1| + |\eta_1|]^{2p}; |\xi_1| + |\eta_1| > C \right) \leq \mathbb{E}_{\xi\eta} \left([\xi_1^2 + \eta_1^2]^p; \xi_1^2 + \eta_1^2 > C^2/2 \right)
$$

$$
= \int_{C/\sqrt{2}}^{\infty} s^{2p+1} e^{-s^2/2} \, ds
$$

$$
= 2^p \Gamma(1 + p, C^2/4) \,,
$$

where $\Gamma(n, x)$ is the incomplete Γ-function. Thus,

$$
\mathbb{E}_{\xi\eta} \left| \sum_{j=1}^{n} z_j \right|^2 \leq 4nr^2 M^4 (\Delta\omega)^2 \left[\Gamma\left(3, \frac{\varepsilon^2}{4rM^2\Delta\omega} \right) + n \, \Gamma^2 \left(2, \frac{\varepsilon^2}{4rM^2\Delta\omega} \right) \right] \,,
$$

which tends to zero exponentially fast as $n \to \infty$. $\qquad\square$

The ergodic property of the long term empirical averages of $h(Y_n(\{s_p\}_{p=1}^r))$, as $n \to \infty$, is a direct consequence of Theorems 5.1 and 5.2:

Corollary 5.1. *Let $Y_n(t)$ and h be defined as in Theorem* 5.1. *Then $\{\nu, \xi, \eta\}$-almost surely:*

$$
\lim_{n \to \infty} \overline{h(\{Y_n(s_p + t)\}_{p=1}^r)} = \mathbb{E}h(\zeta) \,,
$$

where ζ is a Gaussian random vector with mean zero and covariance

$$
\mathbb{E}\zeta_p \zeta_q = \int_0^{\infty} g^2(\omega) \cos[\omega(s_p - s_q)] \, dw \,.
$$

Proof. This follows from (5.7) and from the fact that $\{\nu, \xi, \eta\}$-almost surely the random vectors X_n converge weakly to ζ, hence

$$
\lim_{n \to \infty} \mathbb{E}_{\varphi} h(X_n) = \mathbb{E}h(\zeta) \,. \qquad\square
$$

Comments. 1. Applying our theorems to the particular case of the OU process, we have shown the following: for $U_n(t)$ given by (5.1), for all bounded continuous functions $h : \mathbb{R}^r \to \mathbb{R}$, and all sequences $\{s_p\}_{p=1}^r$, the long-time average

$$
\overline{h(\{U_n(s_p + t)\}_{p=1}^r)} = \lim_{T \to \infty} \frac{1}{T} \int_0^T h(U_n(s_1 + t), \dots, U_n(s_r + t)) \, dt
$$

exists. Furthermore it tends to $\mathbb{E}[h(\zeta)]$ as $n \to \infty$, where ζ is a Gaussian random vector with mean zero and covariance given by

$$
\begin{aligned}
\mathbb{E}\zeta_p\zeta_q &= \int_0^\infty g^2(\omega)\cos[\omega(s_p - s_q)]\,dw \\
&= \frac{2\alpha}{\pi\beta}\int_0^\infty \frac{\cos[\omega(s_p - s_q)]}{\alpha^2 + \omega^2}\,d\omega \\
&= \beta^{-1}e^{-\alpha|s_p - s_q|}\,.
\end{aligned}
\tag{5.11}
$$

That is, the long-time average of bounded continuous functions of the approximate process $U_n(t)$ are, for large n, close to those obtained with $U(t)$, the stationary Gaussian process with covariance $\beta^{-1}\exp(-\alpha|s_p - s_q|)$.

For the case of univariate h and $s_1 = 0$, we obtain that

$$
\lim_{n\to\infty} \overline{h(U_n(t))} = \mathbb{E}h(\xi)\,,
$$

where $\xi \sim \beta^{-1}\mathcal{N}(0,1)$. This tells us that for large n, the empirical distribution function for $U_n(t)$ will approximate that of $U(t)$.

2. Though the above corollary only directly applies to bounded continuous functions h, the result holds true in some other interesting cases.

If we define $h : \mathbb{R}^2 \to \mathbb{R}$ by $h(x_1, x_2) = x_1 x_2$, then the long-term empirical autocovariances of the process $U_n(t)$ can be expressed as

$$
\overline{h(U_n(0), U_n(s))} = \lim_{T\to\infty} \frac{1}{T}\int_0^T h(U_n(t))h(U_n(s+t))\,dt\,.
$$

As h is an unbounded functions, Corollary 5.1 does not apply directly, however the basic result can be extended to cover this case as well. Since h is continuous, Lemma 5.1 shows that the long-time limit exists for each n and is equal to $\mathbb{E}_\varphi X_{n,1}X_{n,2}$, where X_n is defined by (5.6). Lemma 5.2 shows that as $n \to \infty$ this quantity converges to

$$
\int_0^\infty g^2(\omega)\cos(\omega s)\,d\omega = \beta^{-1}e^{-\alpha|s|}\,,
$$

which is the autocovariance of the OU process. So the empirical autocovariance of $U_n(t)$ converges pointwise to that of $U(t)$ as $n \to \infty$.

Another application of our results is to the case of empirical transition probabilities. Let A and B be subintervals of $\mathbb{R}$ such that A has positive length. The empirical transition probability of the process U_n from A to B is given by

$$
\frac{\displaystyle\lim_{T\to\infty} \frac{1}{T}\int_0^T \chi\{U_n(t) \in A, U_n(t+s) \in B\}\,dt}{\lim_{T\to\infty} \frac{1}{T}\int_0^T \chi\{U_n(t) \in A\}\,dt}\,.
$$

This can be shown to converge, as $n \to \infty$, to the transition probability for the OU process to go from A to B in time s. See [23] for details.

3. Although Corollary 5.1 applies to averages in continuous time, it can be extended to discrete time-averages. In the context of the approximation of the OU process, this implies weak convergence of empirical measures of (3.14) to the invariant measure of the limiting SDE. In the next section we study this question in the Hamiltonian context: for the process $Q_n(t)$ in Sec. 4 we study the relationship between its empirical measure and the invariant measure of the SDE it approximates.

4. The question of studying long-term behaviour of processes which are weakly approximated by SDEs over finite time intervals was initiated in [36]. The subject has subsequently been systematized and developed further in [10]. In future work we will study the application of these more general techniques to substantiate the numerical experiments of the next section. Our analysis in this section has relied heavily on the trigonometric form of (5.6).

5. For skew-product maps which, when projected onto the fibre, approximate SDEs on a finite time interval, a numerical study of long-time dynamics of the skew-product, and comparisons with the approximating SDE, is undertaken by Beck *et al.* [37].

6. Numerical Experiments

In this section we describe some numerical experiments which investigate the long-term statistics induced by ordinary differential equations with random initial data. We compute empirical time averages for single trajectories of the process and compare them with the equilibrium ensemble average associated with the approximating stochastic differential system. We do this in the cases where the SDE is known to be ergodic.

Recall that our weak approximation results of Secs. 3 and 4 hold only on finite time intervals; they do not automatically imply anything for long-term behaviour. The analysis in Sec. 5 does concern long-term behaviour, but provides rigorous results only for a limited range of problems. Thus, the numerical experiments considerably extend our understanding.

We will present the results from two sets of experiments. The first is of the forcing process $Z_n(t)$, defined by (2.6), which weakly approximates the OU process. We have established the convergence of long-term averages of this process to those of the OU process in the previous section. However, we will show the results of experiments with this process for illustration. The second set of experiments concerns the fully coupled Hamiltonian system (2.2). Here we do not have any results for long-time statistics, so the experiments can provide useful insight.

6.1. *The Ornstein–Uhlenbeck process*

In Sec. 5 we showed that the empirical measures of finite-dimensional distributions of the approximate OU process (2.6) converge to those of the limiting OU process. Here we illustrate this numerically. We compare the long-term statistics induced by the OU process $U(t)$, defined by (3.1), and the (weakly) approximating process

$Z_n(t)$, defined by (2.6). We use fixed parameters $\alpha = 1$ and $\beta = 2$. Unless explicitly stated otherwise, we take $a = 1/3$, thus maximizing the bound on the rate of convergence according to Theorem 3.1 (see Corollary 3.1).

First we consider the empirical (univariate) distribution of the process Z_n. We choose an n, and then randomly generate appropriately distributed ν_j, ξ_j, η_j for $j = 1, \ldots, n$. Then, rather than generating the sample path induced by these variables, we construct the empirical distribution directly. Equation (5.7) shows that the empirical distribution of Z_n is the same as that of the random variable

$$X_n = \sum_{j=1}^{n} g(\omega_j)(\xi_j \cos \varphi_j + \eta_j \sin \varphi_j)(\Delta\omega)^{1/2} \,.$$

So to sample from the distribution for fixed ν_j, ξ_j, η_j, we simply randomly generate i.i.d. $\varphi_j, j = 1, \ldots, n$ each uniformly distributed on $[0, 2\pi]$ and then compute the resultant X_n. By sampling sufficiently many X_n and storing them in a histogram, we will be able to generate a close approximation to the empirical measure of Z_n. In Fig. 1 we plot the empirical measures for five different realizations of Z_n, for $n = 500$ and for $n = 5000$. For comparison, we also plot the empirical density for the limiting OU process.

We perform similar calculations for the empirical autocovariance functions of Z_n. We can compute explicitly

$$\overline{Z_n(t)Z_n(t+s)} = \frac{1}{2\pi}\Delta\omega \sum_{j=1}^{n}(\eta_j^2 + \xi_j^2)\frac{\cos(\omega_j s)}{1 + \omega_j^2} \,. \tag{6.1}$$

In Fig. 2 we plot the empirical autocovariances for five different realizations of Z_n, for $n = 500$ and $n = 5000$. For comparison, we have plotted the autocovariance for the OU process, $\exp(-s)/2$.

Figures 1 and 2 show the close relationship between the large-time properties of the OU process and its approximation; they also show how this relationship
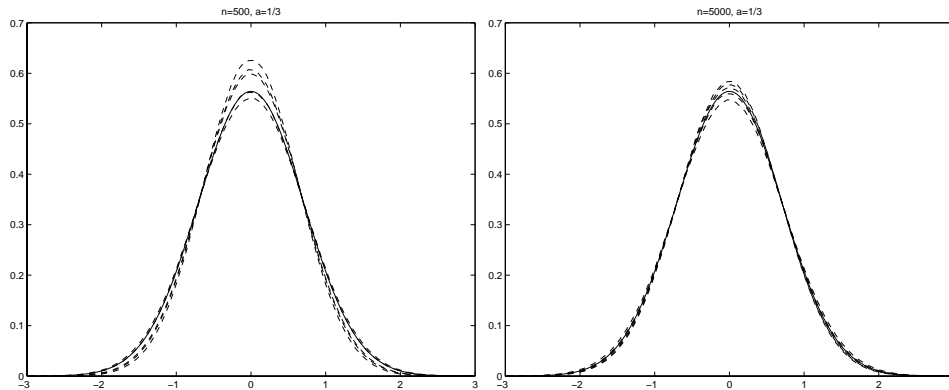

Fig. 1.   Solid lines: empirical distribution of OU process. Dashed lines: empirical distribution obtained from five realizations of the approximate OU process with $n = 500$ and $n = 5000$.
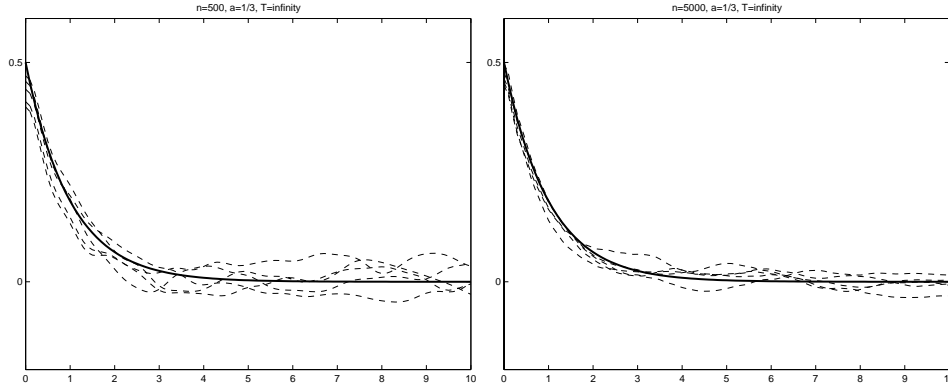
Fig. 2.   Solid lines: autocovariance of OU process. Dashed lines: empirical autocovariances obtained from five realizations of the approximate OU process with $n = 500$ and $n = 5000$.
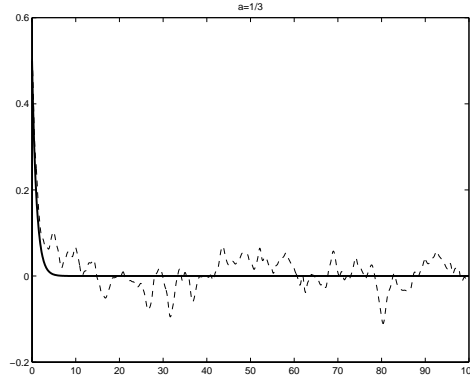

Fig. 3.   Solid line: autocovariance of OU process. Dashed line: empirical autocovariance obtained from approximate OU process with $n = 500$.

improves as n increases. As proven earlier, for any fixed s, the autocovariance function at s converges to that of the OU process as $n \to \infty$. However, if we fix n and consider autocovariance over longer periods of time, we do not see the same decay to zero as we do for the OU process. Figure 3 shows the empirical covariance function for one realization of $Z_n(t)$, with $n = 500$, over the time interval $[0, 100]$. The autocovariance appears to oscillate indefinitely. Indeed, this is to be expected from the expression for the autocovariance (6.1). It is a quasiperiodic function with respect to s, and will thus not decay to zero, but in fact return arbitrarily close to its maximum value infinitely often. This, in turn, is due to the quasiperiodic nature of the approximate process, $Z_n(t)$, which is the sum of finitely many sinusoids.

Finally, Fig. 4 shows the sensitivity of the convergence rate on the parameter a. Both graphs show the empirical autocovariance for increasingly many oscillators. Figure 4(a) is for the optimal value $a = 1/3$ whereas Fig. 4(b) is for $a = 1/2$. Clearly the choice $a = 1/3$ yields closer approximation of the limiting statistics for each n. This is what we would anticipate in view of Corollary 3.1.
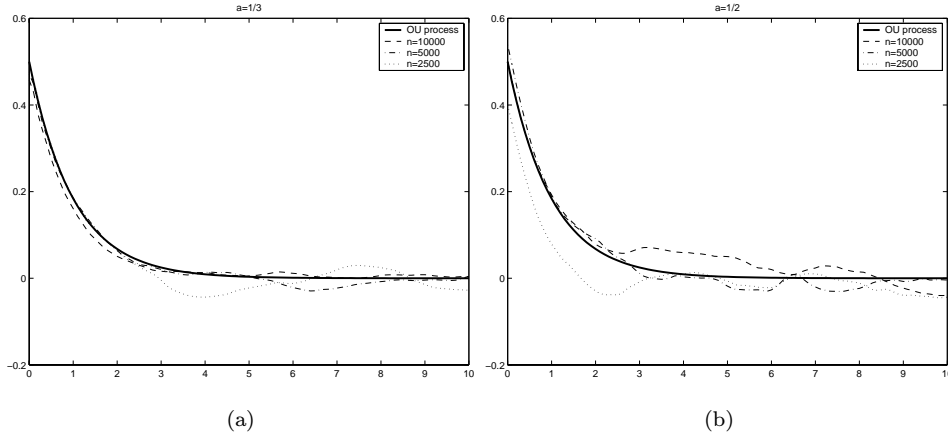
Fig. 4.   Comparison between the equilibrium autocovariance function of $U(t)$ and the empirical functions for a single path of $Z_n(t)$ for $n = 10000$, $n = 5000$, and $n = 2500$. (a) is for $a = 1/3$, which is expected to yield an optimal convergence rate, whereas (b) is for $a = 1/2$.

6.2. *The Hamiltonian system*

We now turn to study the Hamiltonian system (2.2). We use fixed parameters $\alpha = 1$, $\beta = 2$, $\tilde{\beta} = 1$ and $a = 1/3$. Unless stated otherwise we use a sampling/averaging time of $T = 50000$ to calculate empirical measures/autocovariance functions. We test the long-term behaviour of the system for three different potentials $V(\cdot)$,

$$V'(Q) = Q, \tag{6.2}$$

$$V'(Q) = Q^3 - Q, \tag{6.3}$$

$$V'(Q) = Q(1 - Q^2)(1.98 - Q)(2.02 + Q), \tag{6.4}$$

which correspond to single-, double- and triple-well potentials, respectively. For these potentials the Langevin SDE which approximates the motion of the distinguished particle is ergodic [38].

   Figure 5 shows the empirical distribution and autocovariance function for the single-well case, for $n = 500$ and $n = 5000$ oscillators; the sampling time is $T = 50000$. Similarly to the approximate OU process, we see a significant deviation of the empirical distribution from the asymptotic equilibrium distribution when the number of oscillators is too small. The autocovariance function decays initially, but then exhibits a seemingly quasi-periodic behaviour, which persists irrespectively of the averaging time. This is similar to what is proven for the approximate OU process.

   Similar data are presented in Fig. 6 for the double-well potential, again using sampling time $T = 50000$. It is interesting to note that even for a number of oscillators as small as $n = 500$, there is a very good agreement between the empirical distribution and the equilibrium distribution. On the other hand, there is no apparent improvement as the number of oscillators is increased to $n = 5000$ (for fixed
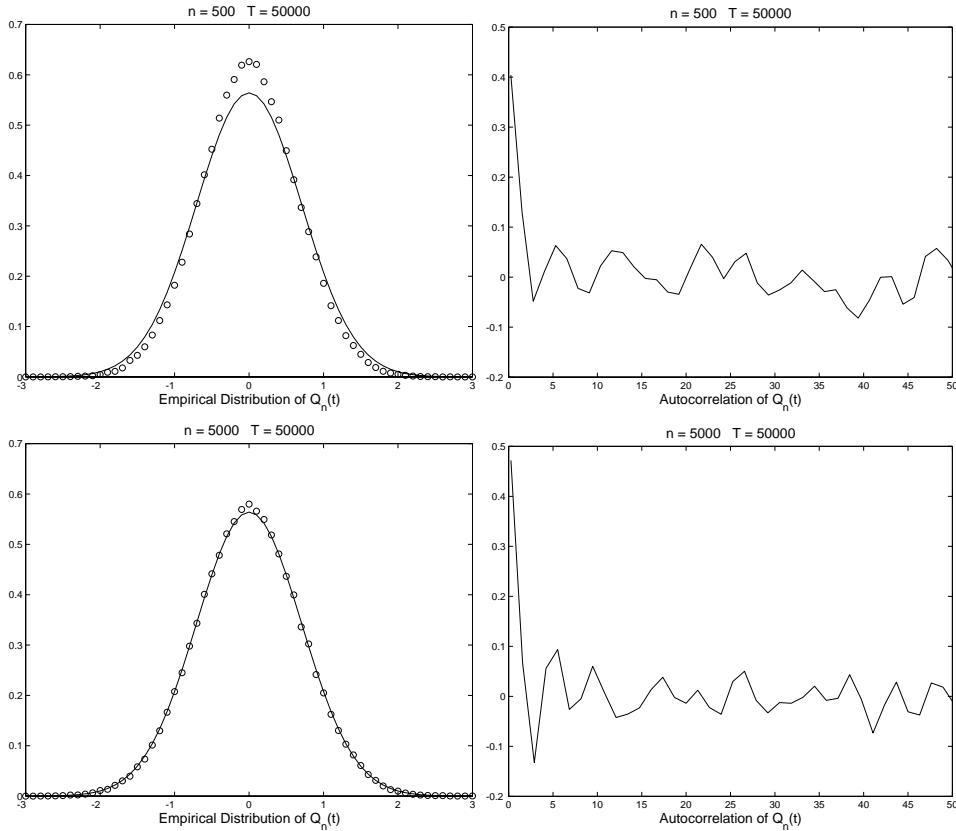
Fig. 5.   Empirical distribution (left, open circles) and empirical autocovariance (right) for the Hamiltonian system 2.2 for a single-well, quadratic potential (6.2). The solid line in the left graphs is the equilibrium distribution. All figures correspond to a sampling time $T = 50000$. The top row corresponds to a calculation with $n = 500$ heat bath particles; the bottom row corresponds to $n = 5000$ heat bath particles.

averaging time). The autocovariance function exhibits much smoother behaviour than in the single-well case, and the persistent quasi-periodic behaviour is weaker, the larger the number of oscillators. This smoother behaviour in the autocovariance is presumably caused by the extra mixing introduced by nonlinearity, and in particular by trajectory separation near the saddle point in the Langevin equation.

Finally we discuss the triple-well case shown in Fig. 7, where a sampling time of $T = 50000$ is again used. For $n = 500$, oscillators the empirical distribution agrees remarkably well with the equilibrium curve. This agreement deteriorates as we increase the number of oscillators. The reason for this surprising fact may be understood by considering the autocovariance function which decays very slowly, reflecting the very long time that the distinguished particle spends in each of the two outermost wells.
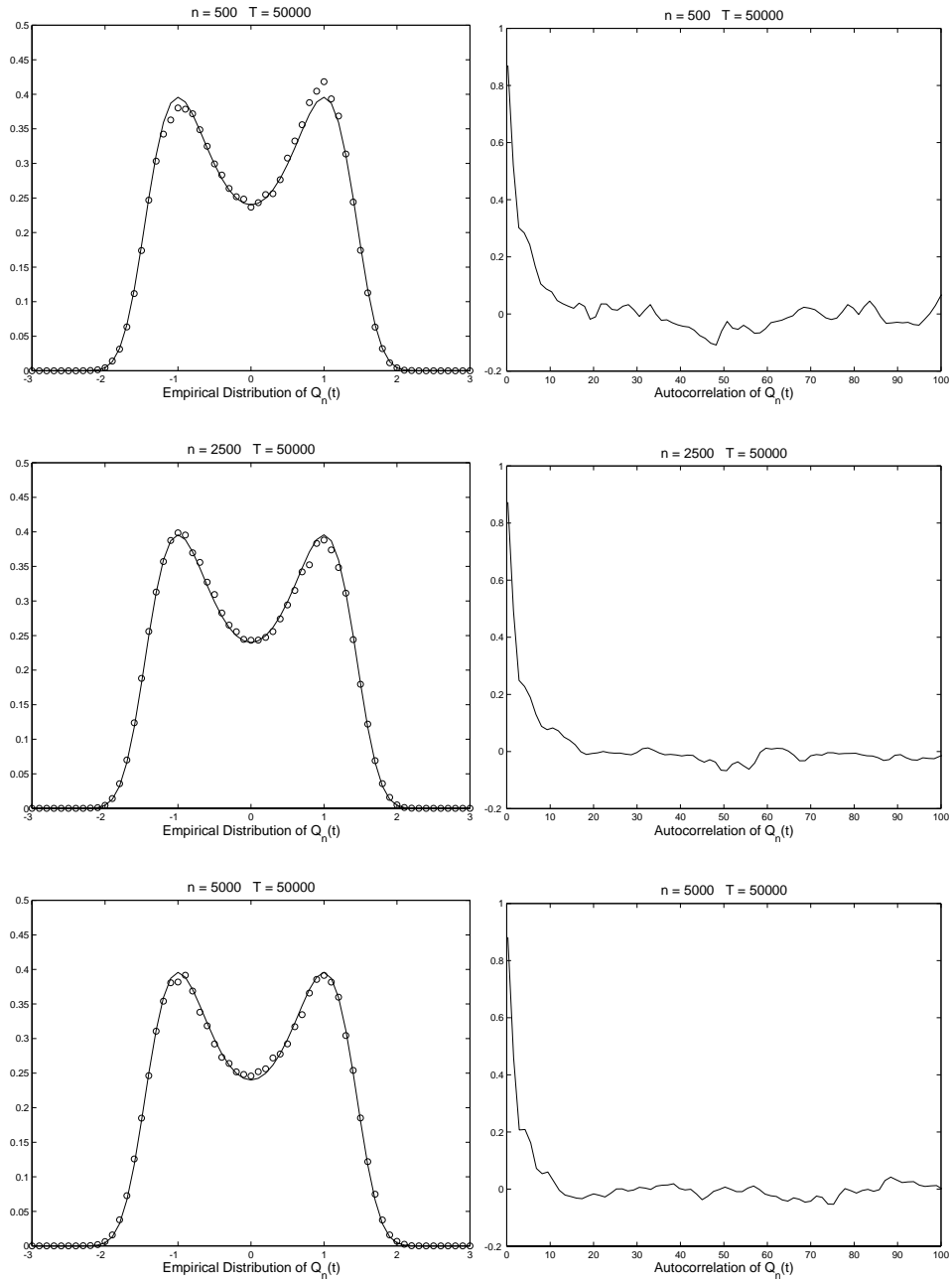
Fig. 6.   Same as Fig. 5 but for $V(\cdot)$ given by the double-well potential (6.3).
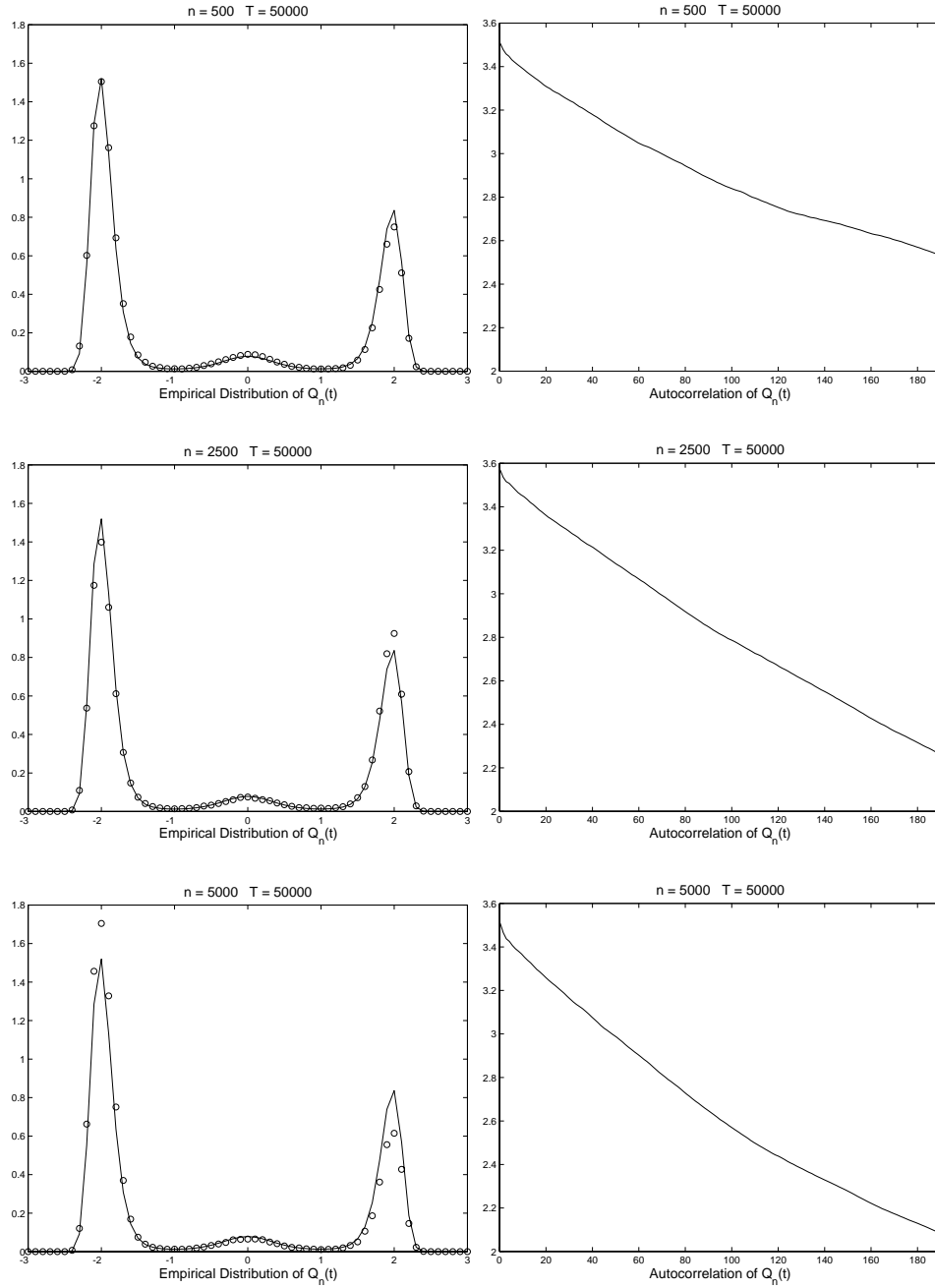
26   *R. Kupferman et al.*

Fig. 7.   Same as Fig. 5 but for $V(\cdot)$ given by the triple-well potential (6.4).

More insight is gained by examining in Fig. 8 the long-term behaviour of a single path $Q(t)$ of the limiting SDE (4.2). This graph shows that the particle remains in each of the main potential wells during characteristic times of the order of many hundreds of time units. Thus, an averaging time of $T = 50000$, as was used above, may not be sufficient for obtaining equilibrium values. The empirical distribution and autocovariance function for the limiting SDE are shown in Figs. 9 and 10 for averaging times of $T = 10000$ and $T = 50000$. In both cases there is a large discrepancy with the equilibrium curve. Interestingly, then, it appears that the large ODE systems with random data equilibrate considerably faster than the SDE itself when n is not too large. However, we expect that as $n \to \infty$ the equilibration time will approach that of the SDE.
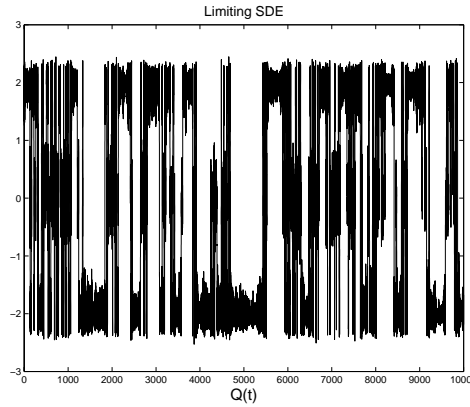

Fig. 8.   A sample path $Q(t)$ solving the stochastic differential system (4.2) for the triple-well potential (6.4).
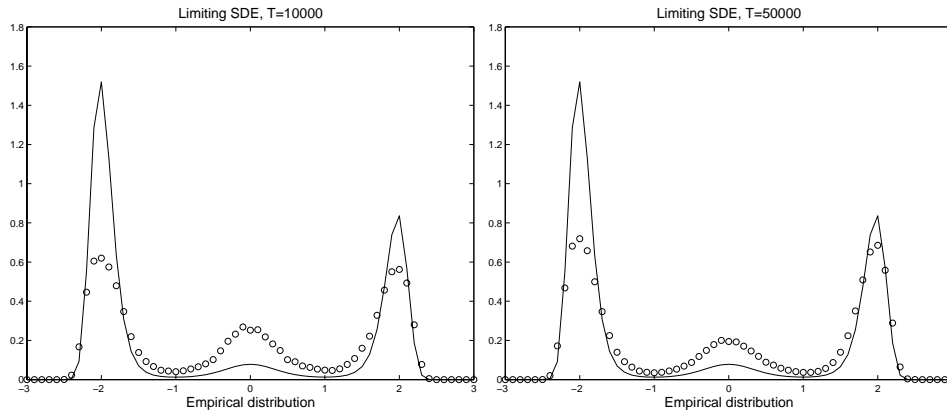

Fig. 9.   Open circles: empirical distribution for a single trajectory $Q(t)$ is the stochastic system (4.2) for a sampling time of $T = 10000$ (left) and $T = 50000$ (right). Solid line: the equilibrium Boltzmann measure.
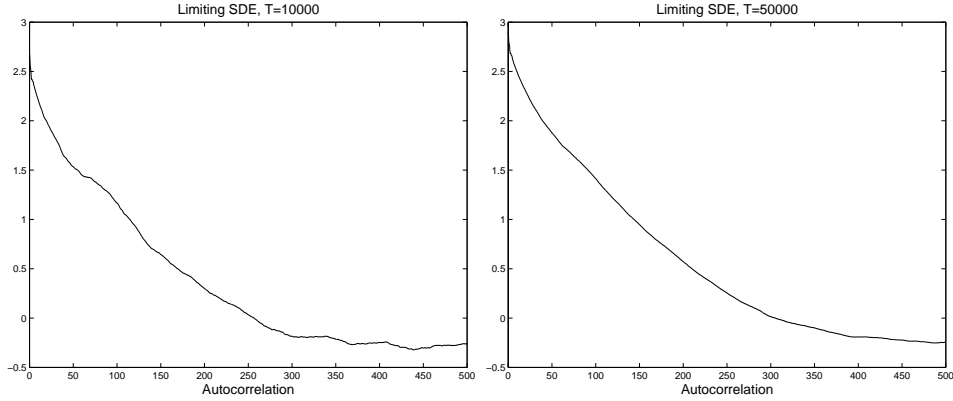
Fig. 10.   Empirical autocovariance function for a single trajectory $Q(t)$ is the stochastic system (4.2) for a sampling time of $T = 10000$ (left) and $T = 50000$ (right).

6.3. *Summary of long-time behaviour*

The primary conclusion of these numerical results is that the long-time behaviour of certain large systems of ordinary differential equations with random data can be understood in terms of ergodic SDEs which approximate projections of the ODEs into low dimensional subspaces. This is manifest in the behaviour of both empirical distributions and autocovariance functions.

However, the observation does need to be qualified somewhat. There is a delicate interplay between the sampling time and the size of n. Our analytical and numerical results show that, for a fixed n, the empirical autocovariance functions for our processes do not decay to zero in the long-time limit, but contain persistent oscillations. For the case of the approximate OU process, this is due to the sample trajectories being quasiperiodic. We conjecture that we observe the same for the Hamiltonian system because it inherits some of the recurrence properties of its quasiperiodic forcing. (The model problem in [39] suffers considerably less from this problem, presumably because of the strong nonlinear effects present in that model.) Furthermore the equilibration times of the large Hamiltonian system can differ substantially from those of the corresponding SDE, even for quite large values of n; this is manifest in the triple-well problem where the correlation time for the SDE is particularly large.

Nonetheless the numerical experiments with the large Hamiltonian system suggest attempting to prove generalizations of theorems analogous to those proven in Sec. 5 for the approximate OU process. Different techniques, however, will be required.

Acknowledgments

References

1. H. Grubmüller and P. Tavan, *Molecular dynamics of conformational substates for a simplified protein model*, J. Chem. Phys. **101** (1994) 5047–5057.
2. P. Deuflhard, J. Hermans, B. Leimkuhler, A. E. Mark, S. Reich and R. D. Skeel, *Computational Molecular Dynamics*: *Challenges*, *Methods*, *Ideas*, *Lecture Notes in Computational Science and Engineering* (Springer, 1999).
3. P. Deuflhard, M. Dellnitz, O. Junge and C. Schütte, *Computation of essential molecular dynamics by subdivision techniques*, in *Computational Molecular Dynamics*: *Challenges*, *Methods*, *Ideas* (Springer, 1999).
4. W. Huisinga, C. Best, R. Roitzsch, C. Schütte and F. Cordes, *From simulation data to conformational ensembles*: *Structure and dynamic based methods*, J. Comp. Chem. **20** (1999) 1760–1774.
5. A. M. Stuart and J. O. Warren, *Analysis and experiments for a computational model of a heat bath*, J. Stat. Phys. **97** (1999) 687–723.
6. G. W. Ford and M. Kac, *On the quantum Langevin equation*, J. Stat. Phys. **46** (1987) 803–810.
7. W. Huisinga, C. Schütte, and A. M. Stuart, *Extracting macroscopic stochastic dynamics*: *Oscillator-driven model problems*, Comm. Pure Appl. Math., in press.
8. G. C. Papanicolaou and S. R. S. Varadhan, *A limit theorem with strong mixing in Banach space and two applications to stochastic differential equations*, Comm. Pure Appl. Math. **XXVI** (1973) 497–524.
9. G. C. Papanicolaou and W. Kohler, *Asymptotic theory of mixing stochastic ordinary differential equations*, Comm. Pure Appl. Math **XXVII** (1974) 641–668.
10. H. Kushner, *Approximation and Weak Convergence Methods for Random Processes*, *with Applications to Stochastic Systems Theory* (MIT Press, 1984).
11. P. Billingsley, *Convergence of Probability Measures* (John Wiley & Sons, 1968).
12. G. W. Ford, M. Kac, and P. Mazur, *Statistical mechanics of assemblies of coupled oscillators*, J. Math. Phys. **6** (1965) 504–515.
13. V. Jakšić and C.-A. Pillet, *Ergodic properties of the non-Markovian Langevin equation*, Lett. Math. Phys. **41** (1997) 49–57.
14. H. Nakazawa, *Quantization of Brownian motion processes in potential fields*, in *Quantum Probability and Applications II*, *Lecture notes in mathematics* (Springer, 1985), pp. 375–387.
15. H. Mori, *Transport, collective motion, and Brownian motion*, Prog. Theor. Phys. **33** (1965) 423–450.
16. R. Zwanzig, *Nonlinear generalized Langevin equations*, J. Stat. Phys. **9** (1973) 215–220.
17. R. Zwanzig, *Problems in nonlinear transport theory*, in *Systems far from Equilibrium*, ed. L. Garrido (Springer, 1980), pp. 198–225.
18. A. Chorin, O.H. Hald and R. Kupferman, *Optimal prediction and the Mori-Zwanzig representation of irreversible processes*, Proc. Nat. Acad. Sci. USA **97** (2000) 2968–2973.
19. G. Grimmett and D. Stirzaker, *Probability and Random Processes* (Oxford Univ. Press, 2001), 3rd edition.
20. P. Billingsley, *Probability and Measure* (John Wiley & Sons, 1995), 3rd edition.

30   *R. Kupferman et al.*

21. I. I Gikhman and A. V. Skorokhod, *Introduction to the Theory of Random Processes* (Dover, 1996).
22. I. Karatzas and S. E. Shreve, *Brownian motion and stochastic calculus* (Springer, 1991).
23. P. F. Tupper, *Constructing stationary gaussian processes from deterministic processes with random initial conditions*, IMA preprint Series 1859, 2002.
24. R.K. Miller, *Nonlinear Volterra Integral Equations* (W. A. Benjamin Inc., 1971).
25. G. Gripenberg, S-O. Londen and O. Staffans, *Volterra Integral and Functional Equations* (Cambridge Univ. Press, 1990).
26. O. H. Hald and R. Kupferman, *Asymptotic and numerical analyses for mechanical models of heat baths*, *J. Stat. Phys.* (2002).
27. B. Cano and A. M. Stuart, *Under-resolved simulations of heat baths*, *J. Comp. Phys.* **169** (2001) 193–214.
28. B. Cano, A. M. Stuart, E. Süli and J. O. Warren, *Stiff oscillatory systems, delta jumps and white noise*, *Found. Comp. Math.* **1** (2001) 69–100.
29. R. B. Bird, R. C. Armstrong and O. Hassager, *Dynamics of Polymeric Liquids* (John Wiley and Sons, 1987), Vol. 1.
30. R.M. Dowell, *Differentiable approximation to Brownian motion on manifolds* Ph.D. Thesis, University of Warwick, 1980.
31. A. J. Majda, I. Timofeyev and E. Vanden Eijnden, *A mathematical framework for stochastic climate models*, *Comm. Pure Appl. Math.* **54** (2001) 891–974.
32. R. Z. Has'minskii, *Stochastic Stability of Differential Equations* (Sijthoff and Noordhoff, 1980).
33. H. Weyl, *Mean motion*, *Amer. J. Math.* **60** (1938) 889.
34. V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer-Verlag, 1989), 2nd edition.
35. A. Dvoretzky, *Asymptotic normality of sums of dependent random vectors*, in *Multivariate analysis IV, Proc. Fourth Int. Symp. Datyon* Ohio, 1975 (North-Holland, 1977), pp. 23–34.
36. G. Blankenship and G. C. Papanicolaou, *Stability and control of stochastic systems with wide-band noise disturbances I*, *SIAM J. Appl. Math.* **34** (1978) 437–476.
37. C. Beck, G. Roepstorff and C. Schroer, *Driven chaotic motion in single- and double-well potentials*, *Physica D* **72** (1994) 211–229.
38. M. M. Tropper, *Ergodic properties and quasideterministic properties of finite-dimensional stochastic systems*, *J. Stat. Phys.* **17** (1977) 491–509.
39. A. Majda and I. Timofeyev, *Remarkable statistic behavior for truncated Burgers-Hopf dynamics*, *Proc. Nat. Acad. Sci. USA* **97** (2000) 12413–12417.