CONVERGENCE OF NUMERICAL TIME-AVERAGING AND STATIONARY MEASURES VIA POISSON EQUATIONS*

JONATHAN C. MATTINGLY†, ANDREW M. STUART‡, AND M. V. TRETYAKOV§

Abstract. Numerical approximation of the long time behavior of a stochastic differential equation (SDE) is considered. Error estimates for time-averaging estimators are obtained and then used to show that the stationary behavior of the numerical method converges to that of the SDE. The error analysis is based on using an associated Poisson equation for the underlying SDE. The main advantages of this approach are its simplicity and universality. It works equally well for a range of explicit and implicit schemes, including those with simple simulation of random variables, and for hypoelliptic SDEs. To simplify the exposition, we consider only the case where the state space of the SDE is a torus, and we study only smooth test functions. However, we anticipate that the approach can be applied more widely. An analogy between our approach and Stein's method is indicated. Some practical implications of the results are discussed.

Key words. stochastic differential equations, ergodic limits, convergence of weak schemes, time averaging, Poisson equation

AMS subject classifications. Primary, 65C30; Secondary, 60H35, 37H10, 60H10

DOI. 10.1137/090770527

1. Introduction. In many applications one is interested in estimating the invariant measure of a stochastic differential equation (SDE) by running a numerical scheme which approximates the time dynamics of the SDE. Two common approaches are to use the numerical trajectories to construct an empirical measure by the time-averaging trajectories or by averaging many different realizations to obtain a finite ensemble average (see, for example, [30]). In either case one produces an approximation to the numerical method's invariant measure and is immediately presented with a number of questions, including the following: (i) Does the numerical scheme have a stationary measure to which it converges quickly as time goes to infinity? (ii) How close is the numerical method's stationary measure to the stationary measure of the underlying SDE? (iii) How close is the time-averaging estimator to the stationary measure of the underlying SDE? This article mainly focuses on questions (ii) and (iii).

The first question is addressed in many papers, including [39, 43, 27, 23, 20]. There are also a number of works where the second question has been considered. In [43] it was shown that several numerical methods for SDEs, including the forward Euler–Maruyama scheme, have unique stationary measures which converge at the expected rate to the unique stationary measure of the SDE. Moreover, in [45] an expansion of the numerical integration error in powers of time step was obtained which allows one

†Department of Mathematics, Center of Nonlinear and Complex Systems, Center for Theoretical and Mathematical Science, and Department of Statistical Science, Duke University, Durham, NC 27708-0320 (jonm@math.duke.edu). This author's work was supported by NSF grants DMS-0616710 and DMS-0449910 and a Sloan Fellowship.

‡Mathematics Institute, University of Warwick, Coventry CV4 7AL, United Kingdom (A.M. Stuart@warwick.ac.uk). This author's work was supported by the ERC and EPSRC.

§Department of Mathematics, University of Leicester, Leicester LE1 7RH, United Kingdom (M.Tretyakov@le.ac.uk). This author's work was partially supported by the EPSRC research grant EP/D049792/1.

to use the Richardson–Romberg extrapolation to improve the accuracy. These works discuss the convergence in the case of relatively smooth test functions. In [2, 3], the authors use techniques from Malliavin calculus to establish smoothing properties of the discretization scheme. This allows one to prove the convergence of the averages of test functions which are only bounded. In turn, this shows that the distance between the numerical method's stationary measure is close to the stationary measure of the SDE in the total-variation norm. In [41], a general method of deducing closeness of the stationary measure from an error estimate on a finite time interval is proposed, though in general it does not provide optimal estimates of the convergence rate. In the earlier works a global Lipschitz assumption was imposed on the coefficients of the SDE which was lifted in [40, 27, 44, 7]. The papers [43, 45] deal with elliptic SDEs, while the hypoelliptic case is treated in [2, 3, 44, 27, 26, 7].

In this paper we obtain error estimates for time-averaging estimators. The error estimates are given in terms of a term of the same order of magnitude as the weak finite time error of the numerical integrator used and the length of the simulation. Control of the time-averages is then used to prove closeness of the numerical method's stationary measure to the stationary measure of the SDE. The convergence rate obtained is the same as the weak convergence rate obtained on a finite time interval. This highlights the general principle that, for time-dependent problems, finite time approximation properties can be transferred to infinite time approximation properties in the presence of suitable stability. Here the underlying stability is the geometric ergodicity of the SDE.

The main tool in our analysis is an associated Poisson equation for the underlying SDE. In this error analysis, the Markov chain generated by the numerical method need not be uniquely ergodic. It is shown that any stationary measure of the numerical method will be close to the unique stationary measure of the underlying SDE. We see the main advantages of the current analysis as being its simplicity and universality. In particular, we obtain the error estimates in the case of numerical schemes with any reasonable simple random variables (including the discrete random variables used in weak Euler schemes [29]). Furthermore, our approach works equally well for a range of explicit and implicit schemes, and, perhaps more importantly, it works for hypoelliptic SDEs. It is comparatively short, and the analysis leverages classical results on PDEs. It should be possible to extend this approach to methods with adaptive step-sizes (see [22, 26]). It should also be adaptable to the stochastic PDE (SPDE) setting using the Poisson equation in infinite dimensions (see related applications of Poisson equations in [8, 10]).

Our goal here has not been to give the most general results. We have picked the simplifying setting of a compact phase space, namely, the d-dimensional torus, and relatively smooth test functions. The extension to the whole space requires control of the time spent outside of the center of the phase space. Under additional assumptions in the spirit of [27], we believe that versions of the present results can be proven in $\mathbf{R}^d$.

In section 2, we introduce the basic setting for the underlying SDE and discuss the hypoellipticity assumptions. In section 3, we describe the class of numerical approximations we consider. We give examples of explicit and implicit methods which fit our framework. In section 4.1, we discuss the auxiliary Poisson equation which will be used to prove the main results, and we give properties of its solution. In section 4.2, we warm up by showing how to prove a law of large numbers, for the SDE, using an auxiliary Poisson equation. Section 5.1 contains the main results of the article which give a number of senses in which the numerical time-averaging estimators are close

to the corresponding stationary time-average of the SDE. In section 5.2 we extend the results of section 5.1 to numerical methods of higher orders, and in section 5.3 the Richardson–Romberg (Talay–Tubaro) error expansion is considered. Section 6 uses the results of section 5 to give a quantitative estimate on the distance of the numerical stationary measures from the underlying stationary measure for the SDE. In section 6.2 we make some general comments about analogies with Stein's method for proving the convergence of distributions to a limiting law. Some practical implications of the results of section 5 are discussed in section 7, where we classify the errors of the numerical time-averaging estimators and, in particular, pay attention to the statistical error.

2. SDE setting. Let $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbf{P})$, $t \geq 0$, be a filtered probability space and $W(t) = (W_1(t), \ldots, W_m(t))^\top$ be an m-dimensional $\{\mathcal{F}_t\}_{t \geq 0}$-adapted standard Wiener process. Consider the Markov process $X(t)$ on the torus[1] $\mathbf{T}^d$ whose time evolution is governed by the Itô SDE

$$(2.1) \qquad dX(t) = f(X(t))\, dt + g(X(t))\, dW(t),$$

where $f = (f_1, \ldots, f_d)^\top \colon \mathbf{T}^d \to \mathbf{R}^d$ and $g \colon \mathbf{T}^d \to \mathbf{R}^{d \times m}$. We assume that f and g are Lipschitz continuous (and hence uniformly bounded) functions on $\mathbf{T}^d$. Under these assumptions, (2.1) has a unique pathwise global solution. The generator of the Markov process $X(t)$ is

$$(2.2) \qquad \mathcal{L} \overset{\text{def}}{=} f.\nabla + \frac{1}{2} a : \nabla \nabla$$

$$= \sum_{i=1}^d f_i \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^d a_{ij} \frac{\partial^2}{\partial x_i \partial x_j},$$

where $a(x) = g(x) g^\top(x)$. For a twice differentiable function $\phi \colon \mathbf{T}^d \to \mathbf{R}$ and $x \in \mathbf{T}^d$ we have

$$(\mathcal{L}\phi)(x) = f(x).\nabla\phi(x) + \frac{1}{2} a(x) : \nabla\nabla\phi(x)$$

$$= \sum_{i=1}^d f_i(x) \frac{\partial\phi}{\partial x_i}(x) + \frac{1}{2} \sum_{i,j=1}^d a_{ij}(x) \frac{\partial^2\phi}{\partial x_i \partial x_j}(x).$$

The operator $\mathcal{L}$ generates a strong Markov semigroup P_t for $t \geq 0$, which maps smooth bounded functions to smooth bounded functions. It can also be defined by $(P_t\phi)(x) = \mathbf{E}_x \phi(X_t)$, where $X_0 = x$. By duality, P_t also induces a semigroup which acts on probability measures. In the name of notational economy, we will also denote this semigroup by P_t.

The kth column of g, which we denote by $g^{(k)}$, can be viewed as a function from $\mathbf{T}^d \to \mathbf{R}^d$. Hence $\{f, g^{(1)}, \ldots, g^{(m)}\}$ is a collection of vector fields on $\mathbf{T}^d$ and (2.1) can be rewritten as

$$dX(t) = f(X(t))\, dt + \sum_{k=1}^m g^{(k)}(X(t))\, dW_k(t).$$

---

[1] In physical applications, a process on the torus can be of interest when periodic boundary conditions are imposed on the problem. A typical example is a noisy gradient system which may be used to sample from the Gibbs distribution with a periodic potential [38]. We also note that our results can potentially be applied to approximating Lyapunov exponents on compact smooth manifolds [14].

This form of (2.1) makes it clear that there is a deterministic drift in the direction of f and independent random kicks in each of m directions $\{g^{(1)}, \ldots, g^{(m)}\}$. We will require that the randomness injected into the dynamics in the $g^{(k)}$ directions affects all directions sufficiently to produce a "smoothing" effect at the level of probability densities. Uniform ellipticity is the simplest assumption which ensures "sufficient spreading" of the randomness. However, it is far from necessary, and many important examples are not uniformly elliptic. For our purposes, it is sufficient that the system is hypoelliptic.

The conditions given in our standing Assumption 1, given below, are enough to ensure hypoellipticity. To state them, we need to recall the definition of the Lie-bracket (or commutator) of two vector fields. In our setting, given two vector fields $h, \tilde{h}$ on $\mathbf{T}^d$ with $h = (h_1, \ldots, h_d)^\top$ and $\tilde{h} = (\tilde{h}_1, \ldots, \tilde{h}_d)^\top$, one defines the Lie-bracket as $[h, \tilde{h}](x) \stackrel{\text{def}}{=} (h.\nabla\tilde{h})(x) - (\tilde{h}.\nabla h)(x)$. Hence the jth component of $[h, \tilde{h}](x)$ is given by $\sum_{i=1}^{d} \left(h_i(x)\frac{\partial \tilde{h}_j}{\partial x_i}(x) - \tilde{h}_i(x)\frac{\partial h_j}{\partial x_i}(x) \right)$. The vector $[h, \tilde{h}](x)$ may be thought of as the new direction generated by infinitesimally following h, then $\tilde{h}$, then $-h$, and finally $-\tilde{h}$. Our most general assumption will be that the collection of all the brackets generated by the randomness spans the tangent space at all points. To track how the noise spreads, we introduce the following increasing set of vector fields:

$$\Lambda_0 = \mathrm{span}\{f, g^{(1)}, \ldots, g^{(m)}\}, \quad \Lambda_{n+1} = \mathrm{span}\{h, [\bar{h}, h] : h \in \Lambda_n, \ \bar{h} \in \Lambda_0\}.$$

The following is our basic assumption governing the stationary measure of the system and its smoothness.

Assumption 1. We assume that one of the following two assumptions hold:
(i) *Elliptic setting.* The matrix-valued function $a(x) = g(x)g^\top(x)$ is uniformly positive definite: there exists $\alpha > 0$ so that, for all $z \in \mathbf{R}^d$ and $x \in \mathbf{T}^d$, $\alpha|z|^2 \le a(x)z.z$.
(ii) *Hypoelliptic setting.* The functions f and g are $C^\infty(\mathbf{T}^d, \mathbf{R}^d)$ and such that, for some n and all $x \in \mathbf{T}^d$, $\Lambda_n = \Lambda_n(x) = \mathbf{R}^d$ and (2.1) possesses a unique stationary measure.

In either case in Assumption 1, we have that (2.1) has a unique stationary measure, henceforth denoted by μ, which has a density with respect to Lebesgue measure on $\mathbf{T}^d$.

As was already mentioned, the goal of this paper is to give a simple, yet robust, proof that time-averages obtained using a class of numerical methods for simulating (2.1) is close to the corresponding ergodic limits of (2.1). To this end, we now describe the class of numerical methods we will consider.

3. Numerical approximations. We will begin by considering a class of simple numerical approximation of (2.1) given by the generalized Euler–Maruyama method on the torus $\mathbf{T}^d$,

(3.1)
$$\begin{cases} X_{n+1} = X_n + F(X_n, \Delta)\Delta + G(X_n, \Delta)\eta_{n+1}\sqrt{\Delta}, \\ X_0 = x_0, \end{cases}$$

where Δ is the time increment, $F \colon \mathbf{T}^d \times (0,1) \to \mathbf{R}^d$, $G \colon \mathbf{T}^d \times (0,1) \to \mathbf{R}^{d \times m}$, and $\eta_n = (\eta_{n,1}, \ldots, \eta_{n,m})^\top$ is an $\mathbf{R}^m$-valued random variable and $\{\eta_{n,i} : n \in \mathbf{N}, i \in \{1, \ldots, m\}\}$ is a collection of independent and identically distributed (i.i.d.) real-valued random variables satisfying

$$\mathbf{E}\eta_{n,i} = \mathbf{E}\eta_{n,i}^3 = 0, \qquad \mathbf{E}\eta_{n,i}^2 = 1, \qquad \mathbf{E}\eta_{n,i}^{2r} < \infty$$

for a sufficiently large $r \geq 2$.[2] More general methods will be considered in section 5.2.

For example, $\eta_{1,1} \sim \mathcal{N}(0,1)$ satisfies these assumptions as well as $\eta_{1,1}$ with the law

$$(3.2) \qquad\qquad \mathbf{P}(\eta_{1,1} = \pm 1) = 1/2.$$

We also note that the above two choices of $\eta_{1,1}$ guarantee existence of finite moments of $\eta_{1,1}$ of any order. Further, we assume that η_n, $n \in \mathbf{N}$, is defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and is $\{\mathcal{F}_{t_n}\}$-adapted for $t_n = n\Delta$.

Since we want the dynamics of (3.1) to be "close" to those of (2.1), we make the following assumption on the "local error" of (3.1).

Assumption 2. For some $C > 0$, all $x \in \mathbf{T}^d$, and all Δ sufficiently small,

$$(3.3) \qquad\qquad |F(x, \Delta) - f(x)| + |G(x, \Delta) - g(x)| \leq C\Delta.$$

Under this assumption we expect that $X_n \approx X(t_n)$, where as before $t_n = n\Delta$. In fact, it follows from the general theory [29] that such numerical schemes have first-order weak convergence on finite time intervals.

In analogy to (2.2), we define an operator associated to (3.1) by

$$\mathcal{L}^{\Delta}(\,\cdot\,) = F(\,\cdot\,, \Delta).\nabla + \frac{1}{2} A(\,\cdot\,, \Delta){:}\nabla\nabla,$$

where $A(x, \Delta) = G(x, \Delta) G^{\top}(x, \Delta)$. While not the generator of the Markov process (3.1), it is the generator's leading order part since

$$(3.4) \qquad \mathbf{E}\Big(\phi(X_1) - \phi(x_0)\Big) = (\mathcal{L}^{\Delta}\phi)(x_0)\Delta + O(\Delta^2) \quad \text{as} \quad \Delta \to 0.$$

Since (by Assumption 2) we have that

$$(3.5) \qquad\qquad |\mathcal{L}^{\Delta}\phi - \mathcal{L}\phi|_{\infty} \leq C\Delta |D^2\phi|_{\infty},$$

where $D^2\phi$ is the second derivative, we deduce that

$$(3.6) \qquad \mathbf{E}\Big(\phi(X_1) - \phi(x_0)\Big) = (\mathcal{L}\phi)(x_0)\Delta + O(\Delta^2) \quad \text{as} \quad \Delta \to 0.$$

It is reasonable to expect that the distribution of the dynamics of the SDE and its approximation will be close to one another since the leading part of the generator of the approximation (3.1) is close to that of the original process (2.1).

We close this section by giving two approximation methods which satisfy Assumption 2.

Example 3.1 (explicit Euler–Maruyama). We define

$$(3.7) \qquad\qquad X_{n+1} = X_n + f(X_n)\Delta + g(X_n)\eta_{n+1}\sqrt{\Delta},$$

and hence $F(x, \Delta) = f(x)$ and $G(x, \Delta) = g(x)$.

Example 3.2 (implicit split-step). We define

$$X_{n+1}^* = X_n + f(X_{n+1}^*)\Delta,$$
$$X_{n+1} = X_{n+1}^* + g(X_{n+1}^*)\eta_{n+1}\sqrt{\Delta},$$

---

[2]We do not specify r in the statements of the forthcoming theorems since common numerical schemes use random variables with bounded moments up to any order. At the same time, in each proof in section 5.1 it is not difficult to recover the number of bounded moments required.

and hence $F(x, \Delta) = f(y)$ and $G(x, \Delta) = g(y)$, where y is an element of $\{y \in \mathbf{T}^d : y = x + f(y)\Delta\}$, which is closest to x.

Remark 3.3. From (3.6) and the backward Kolmogorov equation for (2.1), Assumption 2 is equivalent to

$$(3.8) \qquad \mathbf{E}\phi(X_1) - \mathbf{E}\phi(X(\Delta)) = O(\Delta^2)$$

for all sufficiently smooth ϕ, provided that $X_0 = X(0)$. In the language of consistency and stability of a numerical method used in the introduction, (3.6) and (3.8) provide an appealing way to characterize the degree of local consistency of a numerical method. This will be the starting point for our discussion of higher order methods in section 5.2.

4. Poisson equation.

4.1. Background. It is a "meta-theorem" in averaging, homogenization, and ergodic theory that if one can solve the relevant Poisson equation, then one can prove results about the desired limit of a time-average. A central theme of this paper is to show how an appropriate Poisson equation can be used to analyze the long time average of a given function $\phi \colon \mathbf{T}^d \to \mathbf{R}$, evaluated along a numerical approximation of (2.1), and obtain information about its closeness to the corresponding ergodic limits.

In this section, for motivation, we illustrate key ideas related to the Poisson equation, purely in the continuous time setting. To this end, recalling that μ is the unique stationary measure of (2.1) and $\mathcal{L}$ its generator, given $\phi \colon \mathbf{T}^d \to \mathbf{R}$, we define the stationary average of ϕ by

$$(4.1) \qquad \bar{\phi} = \int_{\mathbf{T}^d} \phi(z)\,\mu(dz)$$

and let ψ solve the Poisson equation

$$(4.2) \qquad \mathcal{L}\psi = \phi - \bar{\phi}.$$

Under Assumption 1, (4.2) possesses a unique solution which is at least as smooth as ϕ. For $k \in \mathbf{N}$, we denote by $W^{k,\infty}$ the space of functions from $\mathbf{T}^d$ to $\mathbf{R}$ such that the function and all of its partial derivatives up to order k are essentially bounded. We then have the following result whose proof we sketch.

THEOREM 4.1. *Under Assumption* 1(i), *given any* $\phi \in W^{k,\infty}$, *with* $k \in \mathbf{N} \cup \{0\}$ *there exists a unique solution* $\psi \in W^{k+2,\infty}$ *to* (4.2). *Under Assumption* 1(ii) *there exists a* $\delta > 0$ *so that given any* $\phi \in W^{k,\infty}$, *with* $k \in \mathbf{N} \cup \{n \geq 2\}$, *there exists a unique solution* $\psi \in W^{k+\delta,\infty}$ *to* (4.2).

Proof. Let $\triangle$ denote the Laplacian on the d-dimensional torus. In either the hypoelliptic or elliptic case, one knows that for some positive α and C

$$\|(-\triangle)^\alpha \phi\|_{L^2} \leq C(\|\mathcal{L}\phi\|_{L^2} + \|\phi\|_{L^2})$$

for all smooth ϕ. (See [18] and recall that our space is compact.) This implies that $\mathcal{L}$ has compact resolvent and discrete spectrum (consisting of isolated points). Thus $\mathcal{L}$ has a spectral gap and is invertible on the complement of the span of the eigenfunctions with zero eigenvalue. The space is nothing other than the functions which have zero mean with respect to an invariant measure for the corresponding semigroup P_t (i.e., μ such that $P_t\mu = \mu$ or, in terms of the generator, $\mathcal{L}^*\mu = 0$). Since the system has a unique invariant measure due to Assumption 1, we see that $\phi - \bar{\phi}$ has zero projection

onto the space spanned by eigenfunctions with eigenvalue zero. Hence we know there exists a function u which solves (4.2) weakly, by the Fredholm alternative.

This leaves the regularity. In the uniformly elliptic case, see [17, 21, 42]. In the hypoelliptic case, the results follow from Theorem 7.3.4 of [42] or [18]. This states that there is $\delta > 0$ such that, if $\phi \in W^{k,p}$, then $u \in W^{k+\delta,p}$ for $p \in [2, \infty)$ and $k \in \mathbf{N}$. Since our $\phi \in W^{k,p}$ for all $p \geq 2$, we know that $u \in \bigcap_{p \geq 1} W^{k+\delta,p}$. Since our space has finite measure, we know that $W^{k+\delta,\infty} = \bigcap_{p \geq 2} W^{k+\delta,p}$. □

Remark 4.2. One of the principal technical issues that must be addressed in order to extend the results in this paper from the $\mathbf{T}^d$ case to a general ergodic SDE on $\mathbf{R}^d$ is proof of the existence of nice, well-controlled solutions to the above Poisson equation. Many of the needed results can be found in [35, 36, 37].

4.2. A strong law of large numbers: An illustrative example. We now show how to use the Poisson equation to prove a law of large numbers for (2.1). This idea appears frequently in the literature [6, 13, 34, 31, 38]. Nonetheless, since this technique is central to our investigation of discrete time approximations to diffusions, we first highlight the main ideas in continuous time. However, before we begin, it is worth noting that, at least formally, the solution to the Poisson equation can be written as

$$(4.3) \qquad \psi(x) \stackrel{\text{def}}{=} - \int_0^\infty P_s(\phi - \bar{\phi})(x)ds.$$

This can be interpreted as the total fluctuation over time of $P_t\phi$ from $\bar{\phi}$; and hence, it is not surprising that ψ can be used to control convergence to equilibrium.

As stated in Theorem 4.1, when Assumption 1 holds, given any $\phi\colon \mathbf{T}^d \to \mathbf{R}$ with $\phi \in W^{2,\infty}$, there exists a unique $\psi\colon \mathbf{T}^d \to \mathbf{R}$ which solves (4.2). Furthermore, $\psi \in W^{2,\infty}$, and hence Itô's formula then tells us that

$$(4.4) \quad \psi(X(t)) - \psi(x_0) = \int_0^t \left(\phi(X(s)) - \bar{\phi}\right) ds + \int_0^t (\nabla\psi)(X(s)) \cdot g(X(s))dW(s).$$

Rearranging this, we obtain that

$$(4.5) \qquad \frac{1}{t}\int_0^t \phi(X(s))\, ds - \bar{\phi} = \frac{\psi(X(t)) - \psi(x_0)}{t} - \frac{1}{t}M(t),$$

where $M(t) = \int_0^t(\nabla\psi)(X(s)) \cdot g(X(s))\, dW(s)$. Since ψ is bounded, the first term on the right-hand side goes to zero as $t \to \infty$. To see that the last term also goes to zero as $t \to \infty$, observe that

$$\frac{1}{t^2}\mathbf{E}\big(M(t)^2\big) = \frac{1}{t^2}\mathbf{E}\langle M\rangle(t) \leq \frac{K}{t}$$

for some $K > 0$ independent of time since $\nabla\psi$ and g are both bounded on $\mathbf{T}^d$. More precisely, we have shown that for any initial x_0

$$(4.6) \qquad \mathbf{E}\left(\frac{1}{T}\int_0^T \phi(X(s))ds - \bar{\phi}\right)^2 \leq \frac{K}{T},$$

which is a quantitative version of what is often called the mean ergodic theorem.

We can view $\frac{1}{T}\int_0^T \phi(X(s))\,ds$ as an estimator for $\bar{\phi}$, and it follows from (4.5) that

$$(4.7) \qquad \mathrm{Bias}\left(\frac{1}{T}\int_0^T \phi(X(s))\,ds\right) \stackrel{\mathrm{def}}{=} \mathbf{E}\left(\frac{1}{T}\int_0^T \phi(X(s))ds - \bar{\phi}\right) = O\left(\frac{1}{T}\right).$$

One can also show that

$$(4.8) \qquad \mathrm{Var}\left(\frac{1}{T}\int_0^T \phi(X(s))\,ds\right) = O\left(\frac{1}{T}\right).$$

The estimate (4.8) easily follows from (4.6) and (4.7) but can also be obtained directly from (4.5) using an additional mixing condition.

To obtain an almost sure (a.s.) result,[3] for any $\varepsilon > 0$ we introduce $A(T;\varepsilon) = \{\frac{1}{T}\sup_{t\le T}|M(t)| > T^{\varepsilon-\frac{1}{2}}\}$. By the Doob inequality for continuous martingales and the fact that $\mathbf{E}|M(T)|^2 \le KT$, we get

$$\mathbf{P}(A(T;\varepsilon)) \le \frac{E|M(T)|^2}{T^{1+2\varepsilon}} \le \frac{K}{T^{2\varepsilon}}.$$

Hence

$$\sum_{n\in\mathbf{N}} \mathbf{P}(A(2^n;\varepsilon)) < \infty,$$

and the Borel–Cantelli lemma implies that there exists an a.s. bounded random variable $C(\omega) > 0$ and n_0 so that for every $n \in \mathbf{N}$ with $n \ge n_0$ one has

$$\frac{1}{2^n}\sup_{t\le 2^n}|M(t)| \le \frac{C(\omega)}{2^{n(1/2-\varepsilon)}}.$$

Hence, with probability one, for every $t \in [2^n, 2^{n+1})$ with $n \ge n_0$ one has

$$\frac{1}{t}|M(t)| \le \frac{1}{2^n}\sup_{2^n\le s\le 2^{n+1}}|M(s)| \le \frac{2}{2^{n+1}}\sup_{s\le 2^{n+1}}|M(s)| \le \frac{2C(\omega)}{2^{(n+1)(1/2-\varepsilon)}} \le \frac{2C(\omega)}{t^{1/2-\varepsilon}}.$$

By combining this estimate with (4.5) and the fact that ψ is bounded, we see that for any $\varepsilon > 0$ and for every $t \ge 2^{n_0}$ one has

$$(4.9) \qquad \left|\frac{1}{T}\int_0^T \phi(X(s))\,ds - \bar{\phi}\right| \le \frac{2|\psi|_\infty}{T} + \frac{C(\omega)}{T^{1/2-\varepsilon}} \qquad \text{a.s.}$$

for some a.s. bounded $C(\omega) > 0$. We note that (4.9) implies that for any initial x_0

$$(4.10) \qquad \lim_{t\to\infty}\frac{1}{t}\int_0^t \phi(X(s))\,ds = \bar{\phi} \qquad \text{a.s.}$$

In other words, the strong law of large numbers holds starting from any initial x_0. Our assumptions are sufficient to ensure that (2.1) has a unique stationary measure μ. Hence, it follows from Birkoff's ergodic theorem that (4.10) holds for μ-a.e. initial x_0. The above argument not only shows that the result holds for *all* x_0 but also gives quantitative estimates on the rate of convergence.

---

[3]This result can also be obtained using Kronecker's lemma (see, e.g., [12]), though the approach we follow gives explicit error estimates which can be useful.

5. Error analysis for the numerical time-average.

5.1. First-order accurate schemes. We now wish to generalize the calculations in the previous section to prove the closeness of the time-averages obtained via the Euler approximation (3.1) to the stationary averages of (2.1). We consider sample path estimates in this section and then convert these to distance estimates in an appropriate metric in section 6.

Let $T = N\Delta$. Introduce the estimator (discrete time-average) $\hat{\phi}_N$ for the stationary average $\bar{\phi}$,

$$(5.1) \qquad \hat{\phi}_N \stackrel{\text{def}}{=} \frac{1}{N} \sum_{n=0}^{N-1} \phi(X_n),$$

where X_n is the SDE approximation defined in (3.1).

We start by proving a mean convergence result, followed by L^2 and a.s. theorems.

THEOREM 5.1. *Let Assumptions 1 and 2 hold. Let H denote $W^{2,\infty}$ in the elliptic setting and $W^{4,\infty}$ in the hypoelliptic setting. Then, for any $\phi \in H$, one has*

$$(5.2) \qquad \left| \mathbf{E}\hat{\phi}_N - \bar{\phi} \right| \leq C\left(\Delta + \frac{1}{T} \right),$$

where $T = N\Delta$ and C is some positive constant independent of Δ and T. Furthermore, the constant C is a linear function of $|\phi|_H$ and otherwise independent of ϕ.[4]

Proof. Under either of the conditions from Assumption 1, we know from Theorem 4.1 that the solution ψ of (4.2) is in $W^{4,\infty}$. In the interest of clarity and definiteness, we will proceed under the first condition in Assumption 1 (the elliptic setting), where $|D^4\psi|_\infty \leq C|D^2\phi|_\infty$, $|D^3\psi|_\infty \leq C|D\phi|_\infty$, $|D^2\psi|_\infty \vee |D\psi|_\infty \vee |\psi|_\infty \leq C|\phi|_\infty$. In the hypoelliptic setting, the kth derivative of ψ is bounded by the same derivative of ϕ, necessitating the change in definition of H between the elliptic and hypoelliptic cases in the statement of the theorem.

For brevity, we write $\phi_n = \phi(X_n)$, $F_n = F(X_n, \Delta)$, $G_n = G(X_n, \Delta)$, $\psi_n = \psi(X_n)$, and $D^k\psi_n = (D^k\psi)(X_n)$, where $(D^k\psi)(z)$ is the kth derivative. We write $(D^k\psi)(z)[h_1, \ldots, h_k]$ for the derivative evaluated in the directions h_j. Defining $\bar{\delta}_n \stackrel{\text{def}}{=} X_{n+1} - X_n = \Delta F_n + \sqrt{\Delta} G_n \eta_{n+1}$, we have

$$(5.3) \qquad \psi_{n+1} = \psi_n + D\psi_n[\bar{\delta}_n] + \frac{1}{2}D^2\psi_n[\bar{\delta}_n, \bar{\delta}_n] + \frac{1}{6}D^3\psi_n[\bar{\delta}_n, \bar{\delta}_n, \bar{\delta}_n] + R_{n+1},$$

where

$$R_{n+1} = \left(\frac{1}{6} \int_0^1 s^3 D^4\psi(sX_n + (1-s)X_{n+1})\, ds \right) [\bar{\delta}_n, \bar{\delta}_n, \bar{\delta}_n, \bar{\delta}_n]$$

is the remainder given by the Taylor theorem. Hence,

$$(5.4) \qquad \psi_{n+1} = \psi_n + \Delta\mathcal{L}^\Delta\psi_n + \Delta^{\frac{1}{2}}D\psi_n[G_n\eta_{n+1}] + \Delta^{\frac{3}{2}}D^2\psi_n[F_n, G_n\eta_{n+1}]$$

$$+ \frac{1}{2}\Delta^2 D^2\psi_n[F_n, F_n] + \frac{1}{6}D^3\psi_n[\bar{\delta}_n, \bar{\delta}_n, \bar{\delta}_n] + r_{n+1} + R_{n+1},$$

---

[4]We indicate the dependence of constants on ϕ since this dependence is used in section 6. Obviously, all the constants appearing in the statements of this and forthcoming theorems also depend on the coefficients of the SDE (2.1) and on the numerical method used.

where

$$r_{n+1} = \frac{\Delta}{2}\Big(D^2\psi_n[G_n\eta_{n+1}, G_n\eta_{n+1}] - A(x,\Delta):\nabla\nabla\psi_n\Big).$$

Summing (5.4) over the first N terms, dividing by $N\Delta$, and using (4.2) produces

$$(5.5)\quad \frac{1}{N\Delta}(\psi_N - \psi_0) = \frac{1}{N}\sum_{n=0}^{N-1}(\phi_n - \bar\phi) + \frac{1}{N}\sum_{n=0}^{N-1}(\mathcal{L}^\Delta - \mathcal{L})\psi_n + \frac{1}{N\Delta}\sum_{i=1}^{3}(M_{i,N} + S_{i,N}),$$

where

$$M_{1,N} = \sum_{n=0}^{N-1} r_{n+1}, \;\; M_{2,N} = \Delta^{\frac{1}{2}}\sum_{n=0}^{N-1} D\psi_n[G_n\eta_{n+1}],$$

$$M_{3,N} = \Delta^{\frac{3}{2}}\sum_{n=0}^{N-1} D^2\psi_n[F_n, G_n\eta_{n+1}],$$

$$S_{1,N} = \frac{\Delta^2}{2}\sum_{n=0}^{N-1} D^2\psi_n[F_n, F_n], \;\; S_{2,N} = \sum_{n=0}^{N-1} R_{n+1}, \;\; S_{3,N} = \frac{1}{6}\sum_{n=0}^{N-1} D^3\psi_n[\bar\delta_n, \bar\delta_n, \bar\delta_n].$$

We will find it convenient to further decompose

$$S_{3,N} = M_{0,N} + S_{0,N},$$

where

$$M_{0,N} = \Delta^{\frac{3}{2}}\sum_{n=0}^{N-1}\Big(D^3\psi_n[G_n\eta_{n+1}, G_n\eta_{n+1}, G_n\eta_{n+1}] + 3\Delta D^3\psi_n[F_n, F_n, G_n\eta_{n+1}]\Big).$$

Notice that $\mathbf{E}[r_{n+1}|\mathcal{F}_{t_n}] = 0$, $\mathbf{E}[\eta_{n+1}|\mathcal{F}_{t_n}] = 0$, and $\mathbf{E}[\eta_{n+1,i}\,\eta_{n+1,j}\,\eta_{n+1,k}|\mathcal{F}_{t_n}] = 0$. Then it is not difficult to see that $M_{i,k}$, $i = 0,\dots,3$, are martingales with respect to $\{\mathcal{F}_{t_k}\}$ and, in particular, $\mathbf{E}M_{i,k} = 0$ for any k.

Since f and g are uniformly bounded, (3.3) implies that F_n and G_n are uniformly bounded in n. Recall that we also know that ψ and its first four derivatives are uniformly bounded. Hence the $S_{i,N}$ are bounded as follows:

(5.6)

$$|S_{1,N}| \le \Delta^2\sum_{n=0}^{N-1} C_1|\phi|_\infty = C_1|\phi|_\infty\Delta T, \;\; \mathbf{E}|S_{2,N}| \le \sum_{n=0}^{N-1}\mathbf{E}|R_{n+1}| \le C_2|D^2\phi|_\infty\Delta T,$$

$$\mathbf{E}|S_{0,N}| \le \Delta^2\sum_{n=0}^{N-1} C_0|D\phi|_\infty = C_0|D\phi|_\infty\Delta T,$$

where the C_i are positive constants which we have labeled for future reference. Similarly, we have (cf. (3.5))

$$\left|\sum_{n=0}^{N-1}(\mathcal{L}^\Delta - \mathcal{L})\psi_n\right| \le \sum_{n=0}^{N-1} C_4|\phi|_\infty\Delta = C_4|\phi|_\infty\Delta N.$$

Notice that this bound and the above bound in $S_{1,N}$ are a.s. bounds with deterministic constants. Finally, observe that $|\psi_N - \psi_0| \le 2|\phi|_\infty$. Applying all of the preceding estimates to (5.5) produces the quoted result. □

Theorem 5.1 implies that (cf. (4.7))

$$\text{(5.7)} \qquad \qquad \text{Bias}(\hat{\phi}_N) = O\left(\Delta + \frac{1}{T}\right).$$

We now consider an L^2 convergence result related to the mean ergodic theorem (4.6) in the continuous case.

THEOREM 5.2. *In the setting of Theorem 5.1, for any $\phi \in H$, one has*

$$\text{(5.8)} \qquad \qquad \mathbf{E}\left(\hat{\phi}_N - \bar{\phi}\right)^2 \le C\left(\Delta^2 + \frac{1}{T}\right),$$

where $T = N\Delta$ and C is some positive constant independent of Δ and T. Furthermore, C depends on ϕ only through $|\phi|_H$ and does so linearly.

Proof. As in the proof of Theorem 5.1, we provide details in the elliptic case; in the hypoelliptic case the only change is that higher derivatives of ϕ appear in the constants. We begin with (5.5) from the proof of Theorem 5.1 and obtain

$$\text{(5.9)} \qquad \mathbf{E}\left(\frac{1}{N}\sum_{n=0}^{N-1}(\phi_n - \bar{\phi})\right)^2 \le C\mathbf{E}\left\{\frac{(\psi_N - \psi_0)^2}{T^2} + \frac{1}{N^2}\left(\sum_{n=0}^{N-1}(\mathcal{L}^\Delta - \mathcal{L})\psi_n\right)^2 \right.$$
$$\left. + \frac{1}{T^2}\sum_{i=0}^{2}S_{i,N}^2 + \frac{1}{T^2}\sum_{i=0}^{3}M_{i,N}^2\right\},$$

where C is an ever-changing constant. Observe that it follows from (5.6) that $|S_{1,N}|^2 \le C_1^2|\phi|_\infty^2\Delta^2 T^2$. Further, by reasoning similar to that used in getting (5.6), we have

$$\mathbf{E}\,S_{2,N}^2 \le \sum_{n,k=1}^{N}\mathbf{E}|R_n||R_k| \le C|D^2\phi|_\infty^2\Delta^2 T^2,$$

$$\mathbf{E}\,S_{0,N}^2 \le \Delta^4\sum_{k,n=0}^{N-1}C|D\phi|_\infty^2 = C|D\phi|_\infty^2\Delta^2 T^2.$$

Since $M_{1,N}$ is a martingale, we get

$$\mathbf{E}\,M_{1,N}^2 = \sum_{n=0}^{N-1}\mathbf{E}\,r_{n+1}^2 \le \sum_{n=0}^{N-1}C|\phi|_\infty^2\Delta^2 \le C|\phi|_\infty^2\Delta^2 N = C|\phi|_\infty^2\Delta T.$$

Similar reasoning and the bound on $\mathbf{E}\eta_{i,n}^8$ give

$$\mathbf{E}\,M_{2,N}^2 \le \Delta\sum_{n=0}^{N-1}C|\phi|_\infty^2 = C|\phi|_\infty^2 T, \qquad \mathbf{E}\,M_{3,N}^2 \le \Delta^3\sum_{n=0}^{N-1}C|\phi|_\infty^2 = C|\phi|_\infty^2\Delta^2 T,$$

$$\mathbf{E}\,M_{0,N}^2 \le \Delta^3\sum_{n=0}^{N-1}C|D\phi|_\infty^2 = C|D\phi|_\infty^2\Delta^2 T.$$

Using the bounds on $|\sum_{n=0}^{N-1}(\mathcal{L}^\Delta - \mathcal{L})\psi_n|$ and $|\psi_N - \psi_0|$, from the proof of Theorem 5.1 and all the above inequalities, we estimate the right-hand side of (5.9) and arrive at the quoted result. $\quad\square$

We now prove an a.s. version of the preceding two results. Its relation to Theorem 5.2 is the same as the relation of (4.9) to (4.6).

THEOREM 5.3. *In the setting of Theorem 5.1, fixing an $L > 0$ there exists a deterministic constant K (depending linearly on L) so that for all ϕ with $\|\phi\|_H \leq L$, Δ sufficiently small, positive $\varepsilon > 0$, and T sufficiently large one has*

$$(5.10) \qquad \left|\hat{\phi}_N - \bar{\phi}\right| \leq K\Delta + \frac{C(\omega)}{T^{1/2-\varepsilon}} \qquad a.s.,$$

where $T = N\Delta$ and $C(\omega) > 0$ is an a.s. bounded random variable depending on ε and the particular ϕ.

Proof. As in the proof of Theorem 5.1, we provide details in the elliptic case; in the hypoelliptic case the only change is that higher derivatives of ϕ appear in the constants. Starting from (5.5), we have

$$\left|\frac{1}{N}\sum_{n=0}^{N-1}\phi_n - \bar{\phi}\right| \leq \frac{|\psi_N - \psi_0|}{T} + \frac{1}{N}\sum_{n=0}^{N-1}\left|(\mathcal{L}^\Delta - \mathcal{L})\psi_n\right| + \frac{1}{T}\sum_{i=0}^{2}|S_{i,N}| + \frac{1}{T}\sum_{i=0}^{3}|M_{i,N}|.$$

Recall that (see the proof of Theorem 5.1)

$$|\psi_N - \psi_0| \leq 2|\phi|_\infty, \quad \sum_{n=0}^{N-1}\left|(\mathcal{L}^\Delta - \mathcal{L})\psi_n\right| \leq C_4|\phi|_\infty \, \Delta N, \quad |S_{1,N}| \leq C_1|\phi|_\infty \, \Delta T,$$

$$|S_{2,N}| \leq K\Delta^2 \sum_{n=0}^{N-1}|\eta_{n+1}|^4 + K\Delta^4 N, \quad |S_{0,N}| \leq K\Delta^2 \sum_{n=0}^{N-1}|\eta_{n+1}|^2 + K\Delta^3 N,$$

where K is an ever-changing positive deterministic constant independent of Δ and N. Due to the strong law of large numbers, the sum $\frac{1}{N}\sum_{n=0}^{N-1}|\eta_{n+1}|^4$ a.s. converges[5] to $\mathbf{E}|\eta_1|^4$, and thus for almost every sequence $\eta_1, \eta_2, \ldots$ and all sufficiently large N we have $\frac{1}{N}\sum_{n=0}^{N-1}|\eta_{n+1}|^4 < K$ for some deterministic $K > 0$. Hence

$$(5.11) \qquad \left|\frac{1}{N}\sum_{n=0}^{N-1}\phi_n - \bar{\phi}\right| \leq K\left(\Delta + \frac{1}{T}\right) + \Theta_N \qquad a.s.,$$

where $\Theta_N \stackrel{\text{def}}{=} \frac{1}{T}\sum_{i=0}^{3}|M_{i,N}|$.

Now we analyze Θ_N. We have for $r \geq 1$

$$\mathbf{E}\Theta_N^{2r} \leq \frac{K}{T^{2r}}\sum_{i=0}^{3}\mathbf{E}|M_{i,N}|^{2r}$$

(here K depends on r). Recall that $M_{i,k}$ are martingales. If we can prove that $\mathbf{E}|M_{i,N}|^{2r} \leq CN^r$ for all r sufficiently large, then the proof can be completed with the aid of the Borel–Cantelli lemma as in section 4.2.

We will provide the argument for estimating $\mathbf{E}|M_{2,N}|^{2r}$; the other terms are estimated analogously. We rewrite

$$M_{2,N} = \sqrt{\Delta}\sum_{n=0}^{N-1}D\psi_n[G_n\eta_{n+1}] \stackrel{\text{def}}{=} \sqrt{\Delta}\tilde{M}_{2,N}.$$

---

[5]Note that in the case of $\eta_{n,i}$, from (3.2) this sum is equal to m^2.

We have $\tilde{M}_{2,1} = 0$ and for an integer $r > 0$

$$
\begin{aligned}
(5.12) \qquad \mathbf{E}\tilde{M}_{2,k+1}^{2r} &= \mathbf{E}\left(\tilde{M}_{2,k} + D\psi_k[G_k\eta_{k+1}]\right)^{2r} \\
&= \mathbf{E}\left(\tilde{M}_{2,k}^2 + 2\tilde{M}_{2,k}\ D\psi_k[G_k\eta_{k+1}] + (D\psi_k[G_k\eta_{k+1}])^2\right)^r \\
&\leq \mathbf{E}\tilde{M}_{2,k}^{2r} + 2r\mathbf{E}\left(\tilde{M}_{2,k}^{2r-1}\ D\psi_k[G_k\eta_{k+1}]\right) \\
&\quad + K\sum_{l=2}^{2r}\mathbf{E}\left(|\tilde{M}_{2,k}|^{2r-l}|D\psi_k[G_k\eta_{k+1}]|^l\right).
\end{aligned}
$$

Note that the second term is equal to zero. Indeed,

$$
(5.13) \qquad \mathbf{E}\left(\tilde{M}_{2,k}^{2r-1}D\psi_k[G_k\eta_{k+1}]\right) = \mathbf{E}\left(\tilde{M}_{2,k}^{2r-1}\mathbf{E}\left(D\psi_k[G_k\eta_{k+1}]\ |\mathcal{F}_{t_k}\right)\right) = 0.
$$

Using the elementary inequality

$$
ab = \frac{1}{N}\left(abN\right) \leq \frac{1}{N}\left(\frac{a^p}{p} + \frac{b^q N^q}{q}\right), \quad a,b > 0, \quad p,q > 1, \quad \frac{1}{p} + \frac{1}{q} = 1,
$$

we get

(5.14)
$$
\mathbf{E}\left(|\tilde{M}_{2,k}|^{2r-l}|D\psi_k[G_k\eta_{k+1}]|^l\right) \leq \frac{1}{N}\mathbf{E}\left(\frac{2r-l}{2r}|\tilde{M}_{2,k}|^{2r} + \frac{l}{2r}|D\psi_k[G_k\eta_{k+1}]|^{2r}N^{2r/l}\right),
$$
$$
l = 2,\dots,2r.
$$

The relations (5.12)–(5.14) imply

$$
\mathbf{E}\tilde{M}_{2,k+1}^{2r} \leq \mathbf{E}\tilde{M}_{2,k}^{2r}\left(1 + \frac{K}{N}\right) + KN^{r-1},
$$

whence

$$
\mathbf{E}\tilde{M}_{2,N}^{2r} \leq KN^r.
$$

Note that by Jensen's inequality this inequality holds for noninteger $r \geq 1$ as well. Therefore,

$$
\frac{1}{T^{2r}}\mathbf{E}\left|M_{2,N}\right|^{2r} = \frac{\Delta^r}{T^{2r}}\mathbf{E}\left|\tilde{M}_{2,N}\right|^{2r} \leq \frac{K}{T^r}.
$$

Analogously, we obtain

$$
\frac{1}{T^{2r}}\mathbf{E}\left|M_{1,N}\right|^{2r} \leq \Delta^r\frac{K}{T^r}, \quad \frac{1}{T^{2r}}\mathbf{E}\left|M_{3,N}\right|^{2r} \leq \Delta^{2r}\frac{K}{T^r}, \quad \frac{1}{T^{2r}}\mathbf{E}\left|M_{0,N}\right|^{2r} \leq \Delta^{2r}\frac{K}{T^r}.
$$

Thus,

$$
(5.15) \qquad \mathbf{E}\Theta_N^{2r} \leq \frac{K}{T^r}.
$$

The Markov inequality together with (5.15) implies

$$
P\left(\Theta_N > \frac{1}{\Delta^\gamma N^\gamma}\right) \leq \Delta^{2r\gamma}N^{2r\gamma}(\mathbf{E}\Theta_N^{2r}) \leq KT^{r(2\gamma-1)}.
$$

Then, for any $\gamma = 1/2 - \varepsilon$, with $\varepsilon > 0$, there is a sufficiently large $r \geq 1$ such that (recall that Δ is fixed here and $T = \Delta N$)

$$\sum_{N=1}^{\infty} P\left(\Theta_N > \frac{1}{T^{\gamma}}\right) \leq K \sum_{N=1}^{\infty} T^{r(2\gamma - 1)} < \infty.$$

Hence, by the Borel–Cantelli lemma, the random variable $\varsigma \overset{\text{def}}{=} \sup_{T>0} T^{\gamma}\Theta_N$ is a.s. finite, which together with (5.11) implies (5.10). □

We note that Theorems 5.1–5.3 do not require the Markov chain X_n to be ergodic. It is, of course, possible under some additional conditions on the numerical method to prove its ergodicity (see, for example, [43, 44, 27]). If the limit $\lim_{N\to\infty} \hat{\phi}_N$ exists and is independent of X_0 (i.e., if the Markov chain X_n is ergodic), then it follows from Theorem 5.3 that $\lim_{N\to\infty} |\hat{\phi}_N - \bar{\phi}| \leq K\Delta$ a.s., which is consistent with the results of, for example, [43].

Remark 5.4. In the series of papers [23, 24, 25, 26, 32, 33] the authors consider the following weighted estimator for the stationary average $\bar{\phi}$:

$$\tilde{\phi}_N = \frac{\sum_{n=1}^{N} w_n \phi(X_n)}{\sum_{n=1}^{N} w_n},$$

where X_n are obtained by an Euler-type scheme with a decreasing time step Δ_n such that $\Delta_n \to 0$ as $n \to \infty$ while positive weights w_n are such that $\sum_{n=1}^{N} w_n \to \infty$ as $N \to \infty$. In particular, they proved that for some Euler-type schemes and $w_n = \Delta_n = \Delta_0 n^{-\alpha}$, $\alpha \in [1/3, 1)$,

$$\text{Bias}\left(\tilde{\phi}_N\right) = \begin{cases} 0, & \alpha \in (1/3, 1), \\ O(1/N^{1/3}), & \alpha = 1/3, \end{cases}$$

$$\text{Var}\left(\tilde{\phi}_N\right) = O\left(\frac{1}{N^{1-\alpha}}\right);$$

i.e., roughly speaking, the error of the estimator $\tilde{\phi}_N$ under the optimal choice of the parameters is

$$\tilde{\phi}_N - \bar{\phi} \sim O\left(\frac{1}{N^{1/3}}\right).$$

We also note that the authors of [23, 24] proved a.s. convergence of weighted empirical measures based on Euler-type schemes with decreasing time step to the invariant measure of the corresponding SDE exploiting the Echeverria–Weiss theorem, which differs from the approach used in this paper.

Let us briefly compare the estimators $\tilde{\phi}_N$ and $\hat{\phi}_N$. For the estimator $\hat{\phi}_N$ from (5.1) with equal weights $w_n = 1$ and with X_n obtained by an Euler-type scheme with constant time step Δ, we can say that the error is (cf. Theorems 5.2 and 5.3)

$$(5.16) \qquad \hat{\phi}_N - \bar{\phi} \sim O\left(\Delta + \frac{1}{N^{1/2}\Delta^{1/2}}\right).$$

If we fix the computational costs (i.e., N and an Euler-type scheme), then the asymptotically optimal choice of Δ for $\hat{\phi}_N$ is $N^{-1/3}$, in which case $O\left(\Delta + 1/N^{1/2}\Delta^{1/2}\right) = O\left(1/N^{1/3}\right)$. It is interesting to note that these optimal orders of errors of both estimators are the same. At the same time, we should emphasize that the term with Δ

in (5.16) (see also $K\Delta$ (5.7) and (5.10)) is related to the numerical integration error while the term with $1/\sqrt{T} = 1/N^{1/2}\Delta^{1/2}$ is related to the statistical error (see also section 7). In the case of large scale simulations (like those in molecular dynamics) the numerical error is usually relatively small thanks to the existing state-of-the-art numerical integrators, and the statistical error prevails. In such common situations one chooses Δ and N to appropriately control the corresponding errors (see further discussion in section 7).

5.2. Higher order schemes. We now consider a more general approximation than (3.1). In addition to the assumptions made in section 2, we assume in this section that the coefficients of the SDE (2.1) and the function ϕ are sufficiently smooth.

Given a function $\bar{\delta} : \mathbf{T}^d \times (0,1) \times \mathbf{R}^m \to \mathbf{T}^d$ and a sequence of $\mathbf{R}^m$-valued i.i.d. random variables $\{\xi_n = (\xi_{n,1}, \ldots, \xi_{n,m}) : n \in \mathbf{N}\}$, we define a general numerical method

$$
(5.17) \qquad \begin{cases} X_{n+1} = X_n + \bar{\delta}(X_n, \Delta, \xi_{n+1}), \\ \quad X_0 = x. \end{cases}
$$

We assume that the ξ_n have sufficiently high finite moments. Clearly our previous class of methods fits into this framework as well as a number of new methods such as implicit Euler. Guided by (3.8), we make the following general assumption about (5.17), after which we will state some easier-to-verify conditions which are equivalent.

Assumption 3. For all $\Delta \in (0,1)$ sufficiently small and all $\phi \in W^{2(p+1),\infty}$, if $X_0 = X(0)$, then

$$
(5.18) \qquad \mathbf{E}\,\phi(X_1) - \mathbf{E}\,\phi(X(\Delta)) = O(\Delta^{p+1}),
$$

where the constant in the error term is uniform over all ϕ with $\|\phi\|_{W^{2(p+1),\infty}} \leq 1$.

To complement the increments of the numerical method $\bar{\delta}$ defined above, we now define the δ increments of the SDE. Namely, for any $x \in \mathbf{T}^d$, we define $\delta : \mathbf{T}^d \times (0,1) \times \Omega \to \mathbf{T}^d$ by

$$
(5.19) \qquad \delta(x,\Delta;\omega) \stackrel{\text{def}}{=} X(\Delta;\omega) - x,
$$

where $X(0;\omega) = x$ and $X(t;\omega)$ solves (2.1). The following proposition, whose proof is given at the end of this section, enables Assumption 3 to be verified.

PROPOSITION 5.5. *Assume that for some $p \in \mathbf{N}$ there exists a positive constant K so that for all $\Delta \in (0,1)$ sufficiently small*

$$
(5.20) \qquad \sup_{\substack{(\alpha_1,\ldots,\alpha_s) \\ \alpha_i \in \{1,\ldots,d\}}} \left| \mathbf{E}\left(\prod_{i=1}^{s} \delta_{\alpha_i} - \prod_{i=1}^{s} \bar{\delta}_{\alpha_i} \right) \right| \leq K\Delta^{p+1}, \; s = 1, \ldots, 2p+1,
$$

$$
(5.21) \qquad \sup_{\substack{(\alpha_1,\ldots,\alpha_{2p+2}) \\ \alpha_i \in \{1,\ldots,d\}}} \mathbf{E} \prod_{i=1}^{2p+2} \left| \bar{\delta}_{\alpha_i} \right| \leq K\Delta^{p+1}.
$$

Then the method satisfies Assumption 3 *with the same p.*

We note that examples of second-order weak schemes ($p = 2$) which satisfy Assumption 3 can be found in many places, including [29, p. 103] and [1]. Higher order methods also exist [29]. We also note that it follows from the general theory [29, p. 100] that the numerical schemes considered in Proposition 5.5 have weak convergence

of order p on finite time intervals. Now we prove a mean convergence result, which is analogous to Theorem 5.1 in the case of Euler-type methods (3.1).

THEOREM 5.6. *Let Assumptions* 1 *and* 3 *hold. Let H denote $W^{2p,\infty}$ in the elliptic setting and $W^{2(p+1),\infty}$ in the hypoelliptic setting. Then for any $\phi \in H$ with $|\phi|_H \leq 1$, consider $\hat\phi_N$ defined by (5.1), where X_n is generated by the numerical method from (5.17) rather than (3.1) as previously. Then*

$$(5.22) \qquad \left|\mathbf{E}\hat\phi_N - \bar\phi\right| \leq C\left(\Delta^p + \frac{1}{T}\right),$$

where $T = N\Delta$ and C is some positive constant independent of Δ and T. Furthermore, the constant C is a linear function of $|\phi|_H$ and otherwise independent of ϕ. In other words,

$$(5.23) \qquad \mathrm{Bias}(\hat\phi_N) = O\left(\Delta^p + \frac{1}{T}\right).$$

Proof of Theorem 5.6. As before, let ψ be the solution to the Poisson equation associated to ϕ given in (4.2). Define $\psi_n = \psi(X_n)$, and let $X(x,t)$ be the solution to (2.1) at time t with initial condition x. From our assumptions, we have that

$$(5.24) \qquad \mathbf{E}\psi_{n+1} = \mathbf{E}\psi(X(X_n, \Delta)) + O(\Delta^{p+1}).$$

Rewriting $\mathbf{E}\psi(X(X_n, \Delta))$ in (5.24) via the Taylor expansion of expectations of SDE solution (see [29, Lemma 2.1.9, p. 99] or [19]), we arrive at

$$(5.25) \qquad \mathbf{E}\psi_{n+1} = \mathbf{E}\psi_n + \sum_{k=1}^{p} \frac{\Delta^k}{k!} \mathbf{E}\left(\mathcal{L}^k\psi\right)(X_n) + O(\Delta^{p+1}).$$

Summing (5.25) over the first N terms and dividing by $N\Delta$, we obtain

$$(5.26) \qquad \frac{\mathbf{E}\psi_N - \psi(x)}{T} = \frac{1}{N}\sum_{n=0}^{N-1} \mathbf{E}\left(\mathcal{L}\psi\right)(X_n) + \sum_{k=2}^{p} \frac{\Delta^{k-1}}{k!}Q_k + O(\Delta^p),$$

where $Q_k = \frac{1}{N}\sum_{n=0}^{N-1} \mathbf{E}\left(\mathcal{L}^k\psi\right)(X_n)$. Using (4.2) and the boundedness of ψ, we have after rearrangement of (5.26)

$$(5.27) \qquad \left|\frac{1}{N}\sum_{n=0}^{N-1}\mathbf{E}\phi(X_n) - \bar\phi\right| \leq \sum_{k=2}^{p}\frac{\Delta^{k-1}}{k!}|Q_k| + K\Delta^p + \frac{K}{T}.$$

Applying arguments to $\mathcal{L}^{k-1}\psi$, $k = 2, \ldots, p$, analogous to those we applied to ψ in (5.26), we have

$$\frac{\mathbf{E}\mathcal{L}^{k-1}\psi(X_N) - \mathcal{L}^{k-1}\psi(x)}{T} = \sum_{i=k}^{p}\frac{\Delta^{i-k}}{(i+1-k)!}Q_i + O(\Delta^{p+1-k}).$$

Therefore (cf. (5.27)),

$$|Q_k| \leq \sum_{i=k+1}^{p}\frac{\Delta^{i-k}}{(i+1-k)!}|Q_i| + K\Delta^{p+1-k} + \frac{K}{T}, \quad k = 2, \ldots, p,$$

and, in particular, $|Q_p| \leq K\Delta + K/T$. Hence $|Q_k| \leq K\Delta^{p+1-k} + K/T$, which together with (5.27) implies (5.22).[6]   ☐

Remark 5.7. If we substitute X_n from the method (5.17) into $\hat{\phi}_N$ from (5.1), it is also possible to prove (analogously to Theorem 5.3) that

$$(5.28) \qquad \left|\hat{\phi}_N - \bar{\phi}\right| \leq K\Delta^p + \frac{C(\omega)}{T^{1/2-\varepsilon}} \quad \text{a.s.}$$

Proof of Proposition 5.5. We note that under the assumptions made the solution ψ of the Poisson equation (4.2) is sufficiently smooth. Expanding $\psi_{n+1} = \psi(X_{n+1})$ in powers of $\bar{\delta}_n \stackrel{\text{def}}{=} X_{n+1} - X_n$ and taking expectation, we get

$$(5.29) \qquad \mathbf{E}\psi_{n+1} = \mathbf{E}\psi_n + \mathbf{E}D\psi_n[\bar{\delta}_n] + \frac{1}{2}\mathbf{E}D^2\psi_n[\bar{\delta}_n, \bar{\delta}_n] + \cdots$$
$$+ \frac{1}{(2p+1)!}\mathbf{E}D^{2p+1}\psi_n[\underbrace{\bar{\delta}_n, \ldots, \bar{\delta}_n}_{2p+1}] + O(\Delta^{p+1}),$$

where the remainder $|O(\Delta^{p+1})| \leq K\Delta^{p+1}$ with K independent of Δ and n. If we replace $\bar{\delta}_n$ by δ_n, defined in (5.19), in (5.29), then, using the conditional version of (5.20), we obtain (with a different remainder $O(\Delta^{p+1})$ than in (5.29))

$$(5.30) \qquad \mathbf{E}\psi_{n+1} = \mathbf{E}\psi_n + \mathbf{E}D\psi_n[\delta_n] + \frac{1}{2}\mathbf{E}D^2\psi_n[\delta_n, \delta_n] + \cdots$$
$$+ \frac{1}{(2p+1)!}\mathbf{E}D^{2p+1}\psi_n[\underbrace{\delta_n, \ldots, \delta_n}_{2p+1}] + O(\Delta^{p+1}).$$

It is not difficult to see that the right-hand side of (5.30) coincides with expectation of the Taylor expansion of $\psi(X(X_n, \Delta))$ around X_n up to a remainder of order Δ^{p+1}. Hence

$$\mathbf{E}\psi_{n+1} = \mathbf{E}\psi(X(X_n, \Delta)) + O(\Delta^{p+1}),$$

and the proof is complete.   ☐

5.3. The Richardson–Romberg (Talay–Tubaro) error expansion. In this section we consider an expansion of the global error $\mathbf{E}\hat{\phi}_N - \bar{\phi}$ in powers of the time step Δ analogous to the Talay–Tubaro result [45]. As in the previous section, we assume here that the coefficients of the SDE (2.1) and the function ϕ are sufficiently smooth. Note that we obtain the expansion both in the elliptic and hypoelliptic settings.

THEOREM 5.8. *Let Assumptions* 1 *and* 3 *hold. Let q be a positive integer and H denote $W^{2(p+q+1),\infty}$ in the elliptic setting and $W^{2(p+q+2),\infty}$ in the hypoelliptic setting. For any $\phi \in H$ with $|\phi|_H \leq 1$, consider $\hat{\phi}_N$ defined by* (5.1), *where X_n is generated by the numerical method from* (5.17). *Then*

$$(5.31) \qquad \mathbf{E}\hat{\phi}_N - \bar{\phi} = C_0^p\Delta^p + \cdots + C_q\Delta^{p+q} + O\left(\Delta^{p+q+1} + \frac{1}{T}\right),$$

where $T = N\Delta$;

$$\left|O\left(\Delta^{p+q+1} + \frac{1}{T}\right)\right| \leq K\left(\Delta^{p+q+1} + \frac{1}{T}\right);$$

---

[6]To help with intuitive understanding of the proof, we remark that $\int \mathcal{L}^k\psi(x)\,d\mu(x) = 0$, which is approximated by Q_k with sufficient accuracy.

and the constants $C_0, \ldots, C_q$, K *are independent of* Δ *and* T, *and they are linear functions of* $|\phi|_H$ *and otherwise independent of* ϕ. *In other words,*

$$\text{Bias}(\hat{\phi}_N) = C_0^p \Delta^p + \cdots + C_q \Delta^{p+q} + O\left(\Delta^{p+q+1} + \frac{1}{T}\right).$$

Proof. Here we make use of the Poisson equation again and also exploit an idea used in the proof of the Talay–Tubaro expansion in the finite time case from [29, pp. 106–108]. We prove the theorem in the case of $p = 1$ (i.e., for Euler-type schemes) and $q = 0$ for the clarity of the exposition. It is not difficult to extend the proof to arbitrary $p, q > 0$.

In the case of $p = 1$ for a particular Euler-type scheme, we can write (cf. (5.25))

$$(5.32) \qquad \mathbf{E}\psi_{n+1} = \mathbf{E}\psi_n + \Delta\mathbf{E}\left(\mathcal{L}\psi\right)(X_n) + \Delta^2\mathbf{E}A(X_n) + O(\Delta^3),$$

where $A(x)$ is the coefficient at Δ^2 in the corresponding expansion of $\mathbf{E}\psi_1$ at $X_0 = x$. For instance, in the case of the explicit Euler–Maruyama scheme (3.7),

$$A(x) = \frac{1}{2}\sum_{i,j=1}^{d} f_i f_j \frac{\partial^2 \psi}{\partial x_i \partial x_j} + \frac{1}{2}\sum_{i,j,k=1}^{d} f_i a_{jk}\frac{\partial^3 \psi}{\partial x_i \partial x_j \partial x_k} + \frac{1}{24}\sum_{i,j,k,l=1}^{d} a_{ij} a_{kl}\frac{\partial^4 \psi}{\partial x_i \partial x_j \partial x_k \partial x_l},$$

where all the coefficients and the derivatives of the function ψ are evaluated at x. We do not need the explicit form of $A(x)$ for the proof.

Summing (5.32) over the first N terms, dividing by $N\Delta$, using (4.2), and rearranging the terms, we obtain

$$(5.33) \qquad \mathbf{E}\hat{\phi}_N - \bar{\phi} = \frac{1}{T}(\mathbf{E}\psi_N - \psi_0) - \frac{\Delta}{N}\sum_{n=0}^{N-1}\mathbf{E}A(X_n) + O(\Delta^2)$$

$$= -\frac{\Delta}{N}\sum_{n=0}^{N-1}\mathbf{E}A(X_n) + O\left(\Delta^2 + \frac{1}{T}\right).$$

Due to Theorem 5.1, we have

$$(5.34) \qquad \mathbf{E}\frac{1}{N}\sum_{n=0}^{N-1} A(X_n) - \bar{A} = O\left(\Delta + \frac{1}{T}\right),$$

where the constant $\bar{A}$ is the stationary average of A (see (4.1)). The expansion (5.31) with $p = 1$, $q = 0$, and $C_0 = -\bar{A}$ follows from (5.33) and (5.34). □

6. Error analysis for the numerical stationary measures.

6.1. Distances between true and approximate stationary measures. We now use the results of the previous section to prove that any stationary measure of the numerical method is close to that of the underlining SDE. The existence of stationary measures for the numerical method follows by the Krylov–Bogoliubov construction.[7] We have assumed in (3.4) and (3.6) or Assumption 3 that the finite time dynamics of our method and SDE are close. This can be seen as a form of "consistency." It is

---

[7]The fact that our state space is compact ensures that the time-averaged transition measure forms a tight family of probability measures [17].

reasonable to expect that the long time behavior will be close since our setting of the torus provides the necessary "stability" through ergodicity.

We begin by giving a metric in which we will measure the distance between measures. If μ^Δ is a stationary measure of the numerical method, then for any bounded function $\phi : \mathbf{T}^d \to \mathbf{R}$ and $n \in \mathbf{N}$,

$$(6.1) \qquad \int_{\mathbf{T}^d} \mathbf{E}_z \phi(X_n) \mu^\Delta(dz) = \int_{\mathbf{T}^d} \phi(z) \mu^\Delta(dz),$$

where $\mathbf{E}_z$ denotes the expectation conditional on $X(0) = z$. Fixing an integer $p \geq 1$, we define the metric ρ between two probability measures on $\mathbf{T}^d$ by

$$\rho(\nu_1, \nu_2) = \sup_{\phi \in \mathcal{H}} \left(\int \phi(z) \nu_1(dz) - \int \phi(z) \nu_2(dz) \right),$$

where $\mathcal{H} = \{\phi : \mathbf{T}^d \to \mathbf{R} : |\phi|_H \leq 1\}$ and $H = W^{2p,\infty}$ in the elliptic setting and $H = W^{2(p+1),\infty}$ in the hypoelliptic setting. Observe that since $\phi \in \mathcal{H}$ implies $-\phi \in \mathcal{H}$, one also has the equivalent, and slightly more standard, characterization of ρ given by

$$\rho(\nu_1, \nu_2) = \sup_{\phi \in \mathcal{H}} \left| \int \phi(z) \nu_1(dz) - \int \phi(z) \nu_2(dz) \right|.$$

THEOREM 6.1. *Define the metric ρ as above for some integer $p \geq 1$, and assume that either*

(i) *$p = 1$ and Assumptions 1 and 2 hold; or*

(ii) *Assumptions 1 and 3 hold.*

Then there exists a positive C so that if μ^Δ is any stationary measure of the numerical method (3.1), then

$$\rho(\mu^\Delta, \mu) \leq C\Delta^p.$$

Proof. Since μ^Δ is stationary, we have that $\int \phi(z) \mu^\Delta(dz) = \int \mathbf{E}_z \phi(X_n) \mu^\Delta(dz)$ for any $n \geq 0$, and hence

$$\int \phi(z) \mu^\Delta(dz) = \int \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{E}_z \phi(X_n) \, \mu^\Delta(dz).$$

Then

$$\int \phi(z) \mu^\Delta(dz) - \int \phi(z) \mu(dz) = \int \left[\frac{1}{N} \sum_{n=0}^{N-1} \mathbf{E}_z \phi(X_n) - \bar{\phi} \right] \mu^\Delta(dz)$$

$$\leq \int \left| \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{E}_z \phi(X_n) - \bar{\phi} \right| \mu^\Delta(dz) \leq C \left(\Delta^p + \frac{1}{\Delta N} \right),$$

where in the last estimate we have invoked either Theorem 5.1 or Theorem 5.6 depending on which assumptions hold. Now taking $N \to \infty$ proves the result since the right-hand side is uniform for any $\phi \in \mathcal{H}$. $\square$

We re-emphasize that we have not assumed in this section that the numerical method is uniquely ergodic. Rather, we have shown that any stationary measure of

the numerical system is close to the true stationary measure in the ρ-distance. It is, of course, possible in many settings to show that the numerical method is itself uniquely ergodic (cf. [43, 44, 27]).

Remark 6.2. It follows from Theorem 5.8 that under its assumptions the error $\rho(\mu^\Delta, \mu)$ can be expanded in powers of the time step:

$$\rho(\mu^\Delta, \mu) = C_0^p \Delta^p + \cdots + C_q \Delta^{p+q} + O\left(\Delta^{p+q+1}\right).$$

Remark 6.3. It is also worth contrasting the result of Theorem 6.1 with a short alternative proof. Let P_t denote the Markov semigroup associated with the SDE and P_n^Δ with n steps of a numerical method with step size Δ. Suppose one knows for some metric d on probability measures that $d(P_t\mu_1, P_t\mu_2) \leq Ke^{-\gamma t}d(\mu_1, \mu_2)$ holds for all probability measures μ_i and that for any fixed n, $d(P_n^\Delta\mu^\Delta, P_{n\Delta}\mu^\Delta) \leq C_n\Delta^p$ for some constant C_n, all Δ sufficiently small, and any measure μ^Δ invariant for P^Δ. Then if one fixes an n so that $Ke^{-\gamma n} \leq 1/2$, then

$$d(\mu, \mu^\Delta) = d(P_{n\Delta}\mu, P_n^\Delta\mu^\Delta) \leq d(P_{n\Delta}\mu, P_{n\Delta}\mu^\Delta) + d(P_{n\Delta}\mu^\Delta, P_n^\Delta\mu^\Delta)$$

$$\leq \frac{1}{2}d(\mu, \mu^\Delta) + C_n\Delta^p,$$

and collecting the $d(\mu, \mu^\Delta)$ produces the estimate $d(\mu, \mu^\Delta) \leq 2C_n\Delta^p$ for all Δ sufficiently small. The challenge in implementing such a seemingly simple program is obtaining the two estimates in the same metric d. Typically, the first estimate is available in the total variation distance. On the other hand, the second estimate is usually estimated in distances requiring test functions with a number of derivatives. The recent works [15, 16] allow one to obtain the first estimate in the 1-Wasserstein distance which simplifies matching the two norms. Using this strategy, one can obtain a bound in a stronger norm (such as total variation or 1-Wasserstein metric), but the $d(P_n^\Delta\mu, P_{\Delta n}\mu)$ convergence rate will often be suboptimal in these metrics. On the other hand, one can obtain $d(P_t\mu_1, P_t\mu_2) \leq Ke^{-\gamma t}d(\mu_1, \mu_2)$ with d defined as in the metric ρ above with $H = W^{2(p+1),\infty}$ by using the fact that $\|P_1\phi\|_H \leq C\|\phi\|_\infty$ from some C, and hence $\|P_{t+1}\phi - \bar{\phi}\|_H \leq C\|P_t\phi - \bar{\phi}\|_\infty \leq CKe^{-\gamma t}\|\phi - \bar{\phi}\|_\infty \leq CKe^{-\gamma t}\|\phi - \bar{\phi}\|_H$. This gives a result comparable to the hypoelliptic result in Theorem 6.1 though misses the smoothing in the elliptic result which is embodied in the fact that one can use $H = W^{2p,\infty}$. (Some partial smoothing could have been extracted in the hypoelliptic case in Theorem 6.1 with more care.) See [15] for this program executed in a particular setting.

6.2. Relationship to Stein's method. This section is devoted to outlining the similarity between the current setting and Stein's method.[8] This connection is not fully explored in this work, but we believe that it is insightful to highlight the main idea. Though our goals are different, there are some passing similarities between the details in this paper and [9, 4].

Let us recall that Stein's method is a generic tool for finding bounds on a distance between two distributions which then can give quantitative convergence results. Denoting the target distribution as π, the idea is to find an operator $\mathcal{A}$ and a determining class of functions $\mathcal{G}$ so that if, for all $g \in \mathcal{G}$, one has $\int \mathcal{A}g(x)d\widetilde{\pi}(x) = 0$,

---

[8]When this work was nearing completion, a conversation with between JCM and Sourav Chatterjee prompted the authors to write this section reflecting on the relationship between their approach and Stein's method.

then this implies that $\widetilde{\pi} = \pi$.[9] Given any h sufficiently nice, one next solves Stein's equation

$$(6.2) \qquad \mathcal{A}g = h$$

with the tacit assumption that the solution g exists and lies in $\mathcal{G}$. Observe that a basic solvability condition on the above equation requires that $\int h d\pi = 0$ when $\mathcal{A}^*\pi = 0$. In this setting we can consider the equation $\mathcal{A}g = h - \int h d\pi$ for more general, uncentered h. In order to quantify the distance of a given measure $\widetilde{\pi}$ from π, one tries to control $\int (\mathcal{A}g)(z) d\widetilde{\pi}(z)$ by some norm of g which can in turn be controlled by an appropriate norm of h. For definiteness let us assume that

$$\int (\mathcal{A}g)(z) d\widetilde{\pi}(z) \leq \epsilon |g|_{\mathcal{G}} \leq \epsilon C \|h\|$$

if g satisfies (6.2). Then

$$\int h d\widetilde{\pi} - \int h d\pi = \int (\mathcal{A}g)(z) d\widetilde{\pi}(z) \leq \epsilon C \|h\|.$$

By taking the supremum over all h in a given class with $\|h\| \leq 1$, one obtains control over the distance of $\widetilde{\pi}$ from π. In particular, if ϵ is small, then $\widetilde{\pi}$ is close to π.

This is essentially the methodology we have followed. Taking $\mathcal{A}$ to be the generator $\mathcal{L}$ of the Markov process, we are assured that $\mathcal{A}^*\pi = 0$ if π is the Markov process's stationary measure.[10] The basic idea of this note is to show that if $\widetilde{\mathcal{L}}$ is a generator of another Markov process so that $\widetilde{\mathcal{L}} - \mathcal{L}$ is small, then any stationary measure of the second Markov process will be close to π. We will further assume that $\widetilde{\pi}$ is ergodic.

While our paper concerns a mixture of continuous and discrete time, the idea can be more easily demonstrated in a continuous time setting. Let $\mathcal{P}_t$ and $\widetilde{\mathcal{P}}_t$ be strong Markov semigroups on $\mathbf{R}^d$ with generators $\mathcal{L}$ and $\widetilde{\mathcal{L}}$ both defined on some common domain D. Let π and $\widetilde{\pi}$ be stationary measures for $\mathcal{P}_t$ and $\tilde{\mathcal{P}}_t$, respectively. Assume that for some set of bounded functions $\mathcal{G} \subset D$ there exists an $\epsilon > 0$ so that

$$(6.3) \qquad \int (\mathcal{L} - \widetilde{\mathcal{L}}) g d\widetilde{\pi} \leq \epsilon \|g\|_{\mathcal{G}}$$

for any $g \in \mathcal{G}$. Further assume that for some class of bounded, real-valued functions $\mathcal{H}$ on $\mathbf{R}^d$ one can solve $\mathcal{L}g = h - \int h d\pi$ for $h \in \mathcal{H}$ with a solution $g \in \mathcal{G}$ and $\|g\|_{\mathcal{G}} \leq K \|h\|_{\mathcal{H}}$ for a fixed constant K. Then, for all $h \in \mathcal{H}$,

$$\int h d\widetilde{\pi} - \int h d\pi = \int (\mathcal{L}g) d\widetilde{\pi} = \int (\widetilde{\mathcal{L}}g) d\widetilde{\pi} + \int (\mathcal{L} - \widetilde{\mathcal{L}}) g d\widetilde{\pi} \leq \epsilon K \|h\|_{\mathcal{H}},$$

where, in moving from the penultimate expression to the last, we have used $\widetilde{\mathcal{L}}^*\widetilde{\pi} = 0$. Since the right-hand side is uniform in h, we can take the supremum over h. If $\mathcal{H}$ were a rich-enough class to define a metric, we could obtain some estimate of the distance between the two stationary measures in that metric.

---

[9] As a referee correctly observed, the Echeverria–Weiss theorem is useful in identifying when such a condition characterizes an invariant measure and identifying the limiting invariant measure for a sequence measure with $\int \mathcal{A}g(x) d\widetilde{\pi_n}(x) \to 0$ as $n \to \infty$. However, here we are really interested in the next order question—how to use the degree to which the characterizing equation is *not* satisfied to obtain quantitative estimates of the convergence rate.

[10] This is done in some versions of Stein's method. See [4].

If one does not have good control over $\widetilde{\pi}$, then (6.3) can be difficult to obtain. Instead it is often easier to replace (6.3) with

$$(6.4) \qquad \limsup_{T \to \infty} \frac{1}{T} \int_0^T \widetilde{P}_t(\widetilde{\mathcal{L}} - \mathcal{L})g(x)dt \leq \epsilon \|g\|_{\mathcal{G}}$$

for any $g \in \mathcal{G}$ and $\widetilde{\pi}$-a.e. x. As before, we assume that, for some class of functions $\mathcal{H}$ from $\mathbf{R}^d \to \mathbf{R}$, one can solve $\mathcal{L}g = h - \int h d\pi$ for $h \in \mathcal{H}$ with a solution $g \in \mathcal{G}$ such that $\|g\|_{\mathcal{G}} \leq K\|h\|_{\mathcal{H}}$. Observe that, if $\mathcal{G}$ is a class of bounded functions, then the second assumption is always satisfied.

Continuing, since $\widetilde{P}_t \mathcal{L}g(x) = \widetilde{P}_t h(x) - \int h d\pi$,

$$\widetilde{P}_T g(x) - g(x) = \int_0^T \widetilde{P}_t \widetilde{\mathcal{L}}g(x)dt$$

$$= \int_0^T \widetilde{P}_t(\widetilde{\mathcal{L}} - \mathcal{L})g(x)dt + \int_0^T \widetilde{P}_t h(x)dt - T \int_{\mathbf{R}^d} h d\pi.$$

Rearranging, dividing by T, and using $|\widetilde{P}_T g(x)| \leq |g|_\infty$ produces

$$\int_{\mathbf{R}^d} h d\pi - \frac{1}{T} \int_0^T \widetilde{P}_t h(x)dt \leq \frac{2\|g\|_\infty}{T} + \frac{1}{T} \int_0^T \widetilde{P}_t(\widetilde{\mathcal{L}} - \mathcal{L})g(x)dt.$$

By Birkoff's ergodic theorem, we know that for $\widetilde{\pi}$-a.e. x

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T \widetilde{P}_t h(x)dt = \int_{\mathbf{R}^d} h d\widetilde{\pi},$$

so from (6.4) and $\|g\|_{\mathcal{G}} \leq K\|h\|_{\mathcal{H}}$ we obtain

$$\int_{\mathbf{R}^d} h d\pi - \int_{\mathbf{R}^d} h d\widetilde{\pi} \leq \epsilon K \|h\|_{\mathcal{H}}$$

for any $h \in \mathcal{H}$. Hence, in the language of section 6, one has $\rho(\pi, \tilde{\pi}) \leq \epsilon K$.

This argument can be modified in many ways; the restriction that $\mathcal{G}$ and $\mathcal{H}$ are classes of bounded functions can be removed by assuming some control over $\widetilde{P}_t g(x)$ uniform in time. If that control can be maintained using a function which is integrable with respect to $\widetilde{\pi}$, then the requirement that $\widetilde{\pi}$ be ergodic can be removed. Although we do not fully explore these issues here, we believe that the connections made in this subsection are useful.

7. Variance of the empirical time-average. Theorems 5.1 and 5.3 are important for implementing time-averaging in computational practice. There are three types of errors arising in computing stationary averages: (i) the numerical integration error (estimated by $K\Delta$ in (5.7) and (5.10) and by $K\Delta^p$ in (5.23) and (5.28)); (ii) the error due to the distance from the stationary distribution (i.e., the error due to the finite time of integration T estimated by K/T in (4.7), (5.7), and (5.23)); (iii) the statistical error. The first two errors contribute to the bias of the estimator $\hat{\phi}_N$ (see (5.7) and (5.23)). The error of numerical integration is controlled by the time step and the choice of a method. It can be estimated in practice using the Talay–Tubaro expansion (see section 5.3 and [45, 29]) in the usual fashion. The statistical error is contained in the second term of (5.10) and related to the variance of the estimator

$\hat{\phi}_N$ (see the details below and also (4.8)). Both the error due to the finite time of integration T and the statistical error are controlled by the choice of the integration time T (i.e., by the choice of the number of steps N under fixed time step Δ). They correspond to properties of the continuous dynamics of the SDE (2.1), and they are almost independent (assuming a sufficiently small Δ) of a method or time step Δ used.

Let us consider the statistical error. Theorems 5.1 and 5.2 immediately imply that $\text{Var}(\hat{\phi}_N) = O(\Delta^2 + 1/T)$. In order to get a more accurate estimate for the variance of $\hat{\phi}_N$, analogous to the one in the continuous case (4.8), i.e., to have the statistical error of the estimator $\hat{\phi}_N$ controlled by T only, we need to use a mixing-type condition.

The Markov process $X(t)$ defined on the torus $\mathbf{T}^d$ by the SDE (2.1) is uniformly ergodic, it has exponential mixing rates (see, e.g., [11, 28, 5]), and there are some positive C and γ such that for any $t > 0$ and $\theta > 0$ the inequality

$$(7.1) \qquad |\mathbf{E}\phi(X(t))\phi(X(t+\theta)) - \mathbf{E}\phi(X(t))\mathbf{E}\phi(X(t+\theta))| \leq Ce^{-\gamma\theta}$$

holds. Further, as has been mentioned before, it is possible in many settings to show that the Markov chain X_n generated by a numerical method is also geometrically ergodic (cf. [43, 44, 27, 20]). Then, due to the compactness of the phase space, the chain is uniformly ergodic and has an exponential mixing such as in (7.1) [28, Chapter 16]. We note that in the cited papers one of the conditions ensuring the ergodicity of X_n is the requirement that the numerical method uses random variables with densities positive everywhere. We do not address here the question about mixing rate for the Markov chains X_n generated by more general numerical methods treated in this paper, leaving it for further study, but instead we assume that the following relaxed mixing condition is satisfied for X_n, $\phi \in H$, and $l > 0$:

$$(7.2) \qquad |\mathbf{E}\phi(X_k)\phi(X_{k+l}) - \mathbf{E}\phi(X_k)\mathbf{E}\phi(X_{k+l})| \leq \frac{K}{(t_{k+l} - t_k)^2},$$

where $K > 0$ is a constant independent of Δ, k, l. This condition is much weaker than (7.1), but it is sufficient for the proof of the following proposition. At the same time, a faster decorrelation than (7.2) does not improve the estimate (7.3).

PROPOSITION 7.1. *In the setting of Theorem* 5.1 *and under the condition* (7.2), *one has*

$$(7.3) \qquad \text{Var}(\hat{\phi}_N) \leq \frac{K}{T},$$

where $T = N\Delta$ and $K > 0$ is a constant independent of Δ and T.

Proof. It follows from (5.5) that

$$(7.4) \qquad \text{Var}(\hat{\phi}_N) \leq K\frac{\mathbf{E}\,|\psi_N - \mathbf{E}\psi_N|^2}{T^2}$$

$$+ \frac{K}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \mathbf{E}\Big[\big((\mathcal{L}^\Delta - \mathcal{L})\psi_i - \mathbf{E}(\mathcal{L}^\Delta - \mathcal{L})\psi_i\big) \big((\mathcal{L}^\Delta - \mathcal{L})\psi_j - \mathbf{E}(\mathcal{L}^\Delta - \mathcal{L})\psi_j\big) \Big]$$

$$+ \frac{1}{T^2} \sum_{i=0}^{2} \mathbf{E}\,|S_{i,N} - \mathbf{E}S_{i,N}|^2 + \frac{1}{T^2} \sum_{i=0}^{3} \mathbf{E}|M_{i,N}|^2.$$

We have used here that $M_{i,N}$ are martingales. Using the estimates for $\mathbf{E}|M_{i,N}|^2$ from Theorem 5.1, we get that $\sum_{i=0}^{3} \mathbf{E}|M_{i,N}|^2 \leq KT$.

To estimate the terms with $S_{i,N}$ and $(\mathcal{L}^\Delta - \mathcal{L})\psi$ in (7.4), we exploit the condition (7.2). For example, consider the term with $S_{1,N}$. We have

$$
\mathbf{E}\left|S_{1,N} - \mathbf{E}S_{1,N}\right|^2 = \frac{\Delta^4}{4}\mathbf{E}\left(\sum_{n=0}^{N-1}\left(D^2\psi_n[F_n, F_n] - \mathbf{E}D^2\psi_n[F_n, F_n]\right)\right)^2
$$

$$
= \frac{\Delta^4}{4}\sum_{n=0}^{N-1}\mathbf{E}\left(\left(D^2\psi_n[F_n, F_n] - \mathbf{E}D^2\psi_n[F_n, F_n]\right)\right)^2
$$

$$
+ \frac{\Delta^4}{2}\sum_{i=0}^{N-1}\sum_{j=i+1}^{N-1}\mathbf{E}\Big(\left(D^2\psi_i[F_i, F_i] - \mathbf{E}D^2\psi_i[F_i, F_i]\right)
$$

$$
\times \left(D^2\psi_j[F_j, F_j] - \mathbf{E}D^2\psi_j[F_j, F_j]\right)\Big)
$$

$$
\leq K\Delta^3 T + K\Delta^4\sum_{i=0}^{N-1}\sum_{j=i+1}^{N-1}\frac{1}{(t_j - t_i)^2} \leq K\Delta T.
$$

Analogously analyzing the other terms in (7.4), we arrive at the stated result. □

Although we proved Proposition 7.1 for the method (3.1), it is also valid for a more general class of numerical methods from section 5.2.

In practice one usually estimates the statistical error of $\hat{\phi}_N$ as follows.[11] We run a long trajectory MT split into M blocks of a large length $T = \Delta N$ each. We evaluate the estimators $_m\hat{\phi}_N$, $m = 1, \ldots, M$, for each block. Since T is big and a time decay of correlations is usually fast, $_m\hat{\phi}_N$ can be considered as almost uncorrelated. We compute the sampled variance

$$
\hat{D} = \frac{1}{M-1}\sum_{m=1}^{M}\left(_m\hat{\phi}_N\right)^2 - \left(\frac{1}{M}\sum_{m=1}^{M}{_m\hat{\phi}_N}\right)^2.
$$

For sufficiently large T and M, $\mathbf{E}\,\hat{\phi}_N$ belongs to the confidence interval

$$
\mathbf{E}\,\hat{\phi}_N \in \left(\hat{\phi}_{NM} - c\frac{\sqrt{\hat{D}}}{\sqrt{M}}, \hat{\phi}_{NM} + c\frac{\sqrt{\hat{D}}}{\sqrt{M}}\right),
$$

with probability, for example, 0.95 for $c = 2$ and 0.997 for $c = 3$. Note that $\mathbf{E}\,\hat{\phi}_N$ contains the two errors forming the bias, as explained at the beginning of this section. We also pay attention to the fact that $\hat{D} \sim 1/T$ (cf. (7.3)); i.e., it is inversely proportional to the product ΔN.

Remark 7.2. Instead of the time-averaging considered in this paper, stationary averages can be computed using ensemble averaging, i.e., by the following estimate for the stationary average $\bar{\phi}$:

$$
\bar{\phi} \approx \mathbf{E}\phi(X(t)) \approx \mathbf{E}\phi(\bar{X}(t)) \approx \check{\phi} \stackrel{\text{def}}{=} \frac{1}{L}\sum_{l=1}^{L}\phi\left(\bar{X}^{(l)}(t)\right),
$$

where t is a sufficiently large time, $\bar{X}$ is an approximation of X, and L is the number of independent approximate realizations. The total error $R_{\check{\phi}} \stackrel{\text{def}}{=} \check{\phi} - \bar{\phi}$ consists of

---

[11]Of course, better statistical estimators can be used to quantify the statistical error of the time-averaging, but we restrict ourselves here to a simple one.

three parts: the error ε of the approximation $\bar{\phi}$ by $\mathbf{E}\phi(X(t))$; the error of numerical integration $C\Delta^p$ (here p is the weak order of the method); and the Monte Carlo error which is proportional to $1/\sqrt{L}$. More specifically,

$$\left|\mathrm{Bias}(\check{\phi})\right| = \left|\mathbf{E}\check{\phi} - \bar{\phi}\right| \leq K\Delta^p + \varepsilon, \qquad \mathrm{Var}(\check{\phi}) = O(1/L).$$

Here, in comparison with the time-averaging approach, each error is controlled by its own parameter: sufficiently large t ensures smallness of ε; time step Δ (as well as the choice of a numerical method) controls the numerical integration error; and the statistical error is regulated by choosing an appropriate number of independent trajectories L. For further details see [30].

Acknowledgments. JCM thanks Martin Harier and Sourav Chatterjee for useful discussions and the hospitality of the Warwick Mathematics Institute, where the core of this work was done. The authors thank an anonymous referee and Denis Talay for useful suggestions which improved the manuscript.

REFERENCES

[1] D. F. ANDERSON AND J. C. MATTINGLY, *A weak trapezoidal method for a class of stochastic differential equations*, http://arxiv.org/abs/0906.3475v1, 2009.

[2] V. BALLY AND D. TALAY, *The law of the Euler scheme for stochastic differential equations.* I. *Convergence rate of the distribution function*, Probab. Theory Related Fields, 104 (1996), pp. 43–60.

[3] V. BALLY AND D. TALAY, *The law of the Euler scheme for stochastic differential equations.* II. *Convergence rate of the density*, Monte Carlo Methods Appl., 2 (1996), pp. 93–128.

[4] A. D. BARBOUR, *Stein's method for diffusion approximations*, Probab. Theory Related Fields, 84 (1990), pp. 297–322.

[5] R. BHATTACHARYA, *A central limit theorem for diffusions with periodic coefficients*, Ann. Probab., 13 (1985), pp. 385–396.

[6] R. N. BHATTACHARYA, *On classical limit theorems for diffusions*, SankhyāSer. A, 44 (1982), pp. 47–71.

[7] N. BOU-RABEE AND E. VANDEN-EIJNDEN, *Pathwise accuracy and ergodicity of metropolized integrators for SDEs*, Comm. Pure Appl. Math., 63 (2009), pp. 655–696.

[8] S. CERRAI AND M. FREIDLIN, *Averaging principle for a class of stochastic reaction-diffusion equations*, Probab. Theory Related Fields, 144 (2009), pp. 137–177.

[9] S. CHATTERJEE, *A new approach to strong embeddings*, preprint, http://arxiv.org/abs/0711.0501v1, 2007.

[10] G. DA PRATO AND J. ZABCZYK, *Second Order Partial Differential Equations in Hilbert Spaces*, London Math. Soc. Lecture Note Ser. 293, Cambridge University Press, Cambridge, UK, 2002.

[11] D. DOWN, S. MEYN, AND R. L. TWEEDIE, *Exponential and uniform ergodicity of Markov processes*, Ann. Probab., 23 (1995), pp. 1671–1691.

[12] R. J. ELLIOTT, *A continuous time Kronecker's lemma and martingale convergence*, Stochastic Anal. Appl., 19 (2001), pp. 433–437.

[13] S. ETHIER AND T. KURTZ, *Markov Processes: Characterization and Convergence*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, New York, 1986.

[14] A. GRORUD AND D. TALAY, *Approximation of Lyapunov exponents of nonlinear stochastic differential equations*, SIAM J. Appl. Math., 56 (1996), pp. 627–650.

[15] M. HAIRER AND J. C. MATTINGLY, *Spectral gaps in Wasserstein distances and the $2D$ stochastic Navier-Stokes equations*, Ann. Probab., 36 (2008), pp. 2050–2091.

[16] M. HAIRER, J. C. MATTINGLY, AND M. SCHEUTZOW, *Asymptotic coupling and a weak form of Harris' theorem with applications to stochastic delay equations*, http://arxiv.org/abs/0902.4495, 2009.

[17] R. HASMINSKII, *Stochastic Stability of Differential Equations*, Sijthoff & Noordhoff, Alphen aan den Rijn, The Netherlands, 1980.

[18] L. HÖRMANDER, *The analysis of linear partial differential operators.* III. *Pseudo-differential operators. Reprint of the 1994 edition*, Classics in Mathematics, Springer, Berlin, 2007.

[19] P. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Applications of Mathematics (New York) 23, Springer-Verlag, Berlin, 1992.

[20] S. Klokov and A. Veretennikov, *On mixing and convergence rates for a family of Markov processes approximating SDEs*, Random Oper. Stoch. Equ., 14 (2006), pp. 103–126.

[21] N. Krylov, *Lectures on Elliptic and Parabolic Equations in Hölder Spaces*, Grad. Stud. Math. 12, AMS, Providence, RI, 1996.

[22] H. Lamba, J. C. Mattingly, and A. M. Stuart, *An adaptive Euler-Maruyama scheme for SDEs: Convergence and stability*, IMA J. Numer. Anal., 27 (2007), pp. 479–506.

[23] D. Lamberton and G. Pagès, *Recursive computation of the invariant distribution of a diffusion*, Bernoulli, 8 (2002), pp. 367–405.

[24] D. Lamberton and G. Pagès, *Recursive computation of the invariant distribution of a diffusion: The case of a weakly mean reverting drift*, Stoch. Dyn., 3 (2003), pp. 435–451.

[25] V. Lemaire, *Estimation récursive de la mesure invariante d'un processus de diffusion*, Ph.D. thesis, Université de Marne-la-Vallée, Marne-la-Vallée, France, 2005.

[26] V. Lemaire, *An adaptive scheme for the approximation of dissipative systems*, Stochastic Process. Appl., 117 (2007), pp. 1491–1518.

[27] J. C. Mattingly, A. M. Stuart, and D. J. Higham, *Ergodicity for SDEs and approximations: Locally Lipschitz vector fields and degenerate noise*, Stochastic Process. Appl., 101 (2002), pp. 185–232.

[28] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, Springer-Verlag, New York, 1993.

[29] G. Milstein and M. Tretyakov, *Stochastic Numerics for Mathematical Physics*, Springer, Berlin, Heidelberg, New York, 2004.

[30] G. Milstein and M. Tretyakov, *Computing ergodic limits for Langevin equations*, Phys. D, 229 (2007), pp. 81–95.

[31] G. Pagès, *Sur quelques algorithmes récursifs pour les probabilités numériques*, ESAIM Probab. Stat., 5 (2001), pp. 141–170.

[32] F. Panloup, *Computation of the invariant measure for a Levy driven SDE: Rate of convergence*, Stochastic Process Appl., 118 (2008), pp. 1351–1384.

[33] F. Panloup, *Recursive computation of the invariant measure of a stochastic differential equation driven by a Levy process*, Ann. Appl. Probab., 18 (2008), pp. 379–426.

[34] G. Papanicolaou, *Lecture notes*, unpublished lecture notes, Stanford University, Stanford, CA, 1999.

[35] E. Pardoux and A. Y. Veretennikov, *On the Poisson equation and diffusion approximation. I*, Ann. Probab., 29 (2001), pp. 1061–1085.

[36] E. Pardoux and A. Y. Veretennikov, *On the Poisson equation and diffusion approximation. II*, Ann. Probab., 31 (2003), pp. 1166–1192.

[37] E. Pardoux and A. Y. Veretennikov, *On the Poisson equation and diffusion approximation. III*, Ann. Probab., 33 (2005), pp. 1111–1133.

[38] G. Pavliotis and A. Stuart, *Multiscale Methods: Averaging and Homogenization*, Texts Appl. Math. 53, Springer, New York, 2008.

[39] G. Roberts and R. Tweedie, *Exponential convergence of Langevin distributions and their discrete approximations*, Bernoulli, 2 (1996), pp. 341–363.

[40] G. O. Roberts and R. L. Tweedie, *Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms*, Biometrika, 83 (1996), pp. 95–110.

[41] T. Shardlow and A. M. Stuart, *A perturbation theory for ergodic Markov chains and application to numerical approximations*, SIAM J. Numer. Anal., 37 (2000), pp. 1120–1137.

[42] D. Stroock, *Partial Differential Equations for Probabilists*, Cambridge Stud. Adv. Math. 112, Cambridge University Press, Cambridge, UK, 2008.

[43] D. Talay, *Second order discretization schemes of stochastic differential systems for the computation of the invariant law*, Stochastics Stochastics Rep., 29 (1990), pp. 13–36.

[44] D. Talay, *Stochastic Hamiltonian systems: Exponential convergence to the invariant measure, and discretization by the implicit Euler scheme*, Markov Process. Related Fields, 8 (2002), pp. 163–198.

[45] D. Talay and L. Tubaro, *Expansion of the global error for numerical schemes solving stochastic differential equations*, Stoch. Anal. Appl., 8 (1990), pp. 483–509.