

A PERTURBATION THEORY FOR ERGODIC MARKOV CHAINS AND APPLICATION TO NUMERICAL APPROXIMATIONS*

T. SHARDLOW[†] AND A. M. STUART[‡]

Abstract. Perturbations to Markov chains and Markov processes are considered. The unperturbed problem is assumed to be geometrically ergodic in the sense usually established through the use of Foster–Lyapunov drift conditions. The perturbations are assumed to be uniform, in a weak sense, on bounded time intervals. The long-time behavior of the perturbed chain is studied. Applications are given to numerical approximations of a randomly impulsive ODE, an Itô stochastic differential equation (SDE), and a parabolic stochastic partial differential equation (SPDE) subject to space-time Brownian noise. Existing perturbation theories for geometrically ergodic Markov chains are not readily applicable to these situations since they require very stringent hypotheses on the perturbations.

Key words. Markov chains, ergodic theory, numerical approximations, random impulses, stochastic differential equations, stochastic partial differential equations

AMS subject classifications. 60J10, 60J27, 65U, 60H10, 60H15, 34A37

PII. S0036142998337235

1. Introduction. It is frequently of interest to understand how ergodic properties of Markov chains persist under various kinds of perturbations. Here perturbations to (discrete time) Markov chains and (continuous time) Markov processes evolving in a Banach space are considered. In both cases the perturbation is assumed to be a discrete time Markov chain, and our primary motivation is to understand the *numerical* approximation of Markov chains and processes. The unperturbed Markov chain is assumed to be *geometrically ergodic*, implying exponential convergence of expectations of functions from a certain class; the general framework of geometric ergodicity within which we operate is taken from the work of Meyn and Tweedie [23, 24] based on Foster–Lyapunov drift conditions. The perturbed Markov chains are assumed to be close to the unperturbed problem in a *weak sense*: the error in expectations of functions is small, uniformly on compact time intervals disjoint from the origin, for functions in the same class. Perturbation theories for geometrically ergodic Markov chains do exist already, but it turns out that the class of perturbations considered there is typically too restrictive to admit application to the numerical methods considered here—at least for the finite time approximation results that we are currently able to obtain for these numerical methods. At the end of section 3 we will relate our perturbation theory to an existing perturbation theory due to Kartashov [16, 17, 18]. Properties of ergodic Markov chains under perturbation have been studied in many other contexts; for example, in SDEs the idea of approximating white noise by a

*Received by the editors April 13, 1998; accepted for publication (in revised form) September 16, 1998; published electronically March 23, 2000.

<http://www.siam.org/journals/sinum/37-4/33723.html>

[†]Scientific Computing and Computational Mathematics Program, Durand 257, Stanford University, Stanford, CA 94305-4040 (shardlow@ima.umn.edu). Supported by the National Science Foundation under grant DMS-95-04879 and continued working at IMA, University of Minnesota, 514 Vincent Hall, Minneapolis, MN 55455-0436. Current address: OCIAM, Mathematical Institute, 24–29 St. Giles, Oxford, OX1 3LB, UK.

[‡]Scientific Computing and Computational Mathematics Program, Durand 257, Stanford University, Stanford, CA 94305-4040 (stuart@sccm.stanford.edu). Current address: Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK.

broad-band Gaussian noise process is of interest, and this is studied in [3] and further in [21].

In section 2 our notation and framework is established. A general theory is developed in section 3, and then, in section 4, applications are described for three problems where, in all cases, the perturbation arises from numerical approximation. The first concerns an ODE subject to random impulses, the second to an Itô stochastic differential equation (SDE), and the third to a parabolic stochastic partial differential equation (SPDE) subject to space-time to Brownian noise.

The numerical simulation of ergodic stochastic processes in the context of finite dimensional SDEs has been studied by Talay [30] in the case where the generator of the process is uniformly parabolic; see also [31, 13, 20]. However, in many applications this uniform parabolicity does not hold; our theory encompasses problems for which the generator is not uniformly parabolic, albeit at a reduced rate of convergence when compared with the estimates in [30]. For stochastic parabolic PDEs, much is known about approximation properties on finite time intervals [6, 11, 14, 27], but the theory presented here enables us to prove long-time weak convergence properties in the geometrically ergodic case; such results have not been obtained before, to the best of our knowledge.

2. Preliminaries. In the following $\mathbb{N} = \{1, 2, 3, \dots\}$ and $\mathbb{Z}^+ = \{0, 1, 2, \dots\}$. Let \mathbb{S} be a Banach space and $\mathbb{B}(\mathbb{S})$ be the corresponding Borel σ -algebra. We consider Markov chains

$$(2.1) \quad \{u_n, n \in \mathbb{Z}^+\}$$

of random variables from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\mathbb{S}, \mathbb{B}(\mathbb{S}))$. For an initial distribution μ , let the probability triple $(\Omega, \mathcal{F}, \mathbb{P}^\mu)$ generate the chain (2.1). We use the notation δ^x to denote a point mass at $x \in \mathbb{S}$. Expectation with respect to \mathbb{P}^μ will be denoted by \mathbb{E}^μ . Finally, we define measures μ_n on $(\mathbb{S}, \mathbb{B}(\mathbb{S}))$, parameterized by $\mu_0 = \mu$, according to

$$\mu_n(B) = \mathbb{P}^\mu\{\omega \in \Omega : u_n \in B\}, \quad B \in \mathbb{B}(\mathbb{S}).$$

We approximate the Markov chain (2.1) by a Markov chain

$$(2.2) \quad \{u_n^\varepsilon, n \in \mathbb{Z}^+\}$$

of random variables from a probability space $(\Omega_\varepsilon, \mathcal{F}_\varepsilon, \mathbb{P}_\varepsilon)$ to $(\mathbb{S}, \mathbb{B}(\mathbb{S}))$. For an initial distribution μ , the probability triple $(\Omega_\varepsilon, \mathcal{F}_\varepsilon, \mathbb{P}_\varepsilon^\mu)$ is assumed to generate the chain (2.2). Expectation with respect to $\mathbb{P}_\varepsilon^\mu$ will be denoted by \mathbb{E}^μ ; this should cause no confusion as it will always be clear from the context which underlying probability space is giving rise to the expectation. In many applications the underlying probability spaces for (2.1) and (2.2) will be the same, but this is not necessarily the case, for example, when weak approximations of stochastic differential equations are studied. We define measures μ_n^ε on $(\mathbb{S}, \mathbb{B}(\mathbb{S}))$, parameterized by $\mu_0^\varepsilon = \mu^\varepsilon$, according to

$$\mu_n^\varepsilon(B) = \mathbb{P}_\varepsilon^\mu\{\omega \in \Omega_\varepsilon : u_n^\varepsilon \in B\}, \quad B \in \mathbb{B}(\mathbb{S}).$$

We will consider also the approximation of time-continuous Markov processes by time-discrete Markov chains of the form (2.2). Specifically we consider a stochastic process $\{u(t), t \geq 0\}$ of random variables from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\mathbb{S}, \mathbb{B}(\mathbb{S}))$. For an initial distribution μ , the probability triple $(\Omega, \mathcal{F}, \mathbb{P}^\mu)$ generates this process. For each fixed

$\omega \in \Omega$, we thus have a path $u(\cdot) := u(\cdot; \omega)$. We define measures $\mu(t)$ on $(\mathbb{S}, \mathbb{B}(\mathbb{S}))$, parameterized by $\mu(0) = \mu$, according to

$$\mu(t)(B) = \mathbb{P}^\mu\{\omega \in \Omega : u(t) \in B\}, \quad B \in \mathbb{B}(\mathbb{S}).$$

If time is discretized with $t_n = n\varepsilon$, we then approximate the sampled chain $\{u(n\varepsilon), n \in \mathbb{Z}^+\}$ by a time-discrete Markov chain $\{u_n^\varepsilon, n \in \mathbb{Z}^+\}$ assumed to be of the form (2.2).

The notation $\lfloor x \rfloor$ will be used to denote the largest integer no larger than x . Throughout the paper C denotes a constant whose actual value may change between instances.

3. Basic theory. The following two assumptions will be used in proving the basic results of this paper concerning approximation of (2.1) by (2.2).

Assumption AI. The Markov chain (2.1) is geometrically ergodic. Specifically fix $G : \mathbb{S} \rightarrow [1, \infty)$ and define $\mathcal{G} = \{\text{measurable } G : \mathbb{S} \rightarrow \mathbb{R}, |G| \leq \bar{G}\}$. Assume that $\bar{G} \in \mathcal{L}^1(\mathbb{S}, \mathbb{B}(\mathbb{S}), \mu_n) \forall n \geq 0$. For some $R_1 > 0$, $r > 1$ and some set $\mathcal{G}_0 \subseteq \mathcal{G}$ containing \bar{G} , further assume that

$$\sup_{G \in \mathcal{G}_0} |\mathbb{E}^{\delta_x} G(u_n) - \pi(G)| \leq R_1 r^{-n} \bar{G}(x) \quad \forall n \geq 0$$

with π the unique invariant measure on $(\mathbb{S}, \mathbb{B}(\mathbb{S}))$ generated by (2.1).

Assumption AII. The Markov chains (2.1) and (2.2) satisfy, for some $\zeta > 1$ and $R_2 > 0$,

$$\sup_{G \in \mathcal{G}_0} |\mathbb{E}^{\delta_x} G(u_n) - \mathbb{E}^{\delta_x} G(u_n^\varepsilon)| \leq R_2 \bar{G}(x) \zeta^n \varepsilon \quad \forall n \geq 0.$$

The first assumption characterizes the dependence of the rate of convergence of expectations under the chain to their limiting value in terms of the deterministic initial value x . Typically such results are proved by deriving a Foster–Lyapunov type drift condition involving \bar{G} , together with construction of an irreducibility measure; see [23, Chapter 16] and [24]. Indeed for the randomly impulsive ODEs considered in this paper, the condition in Assumption AI is typically satisfied for any $G \in \mathcal{G}$. The second assumption states that expectations under (2.1) or (2.2) remain close, uniformly over a bounded number of transitions, and the dependence on deterministic initial data is again specified. Note, however, that the class of functions \mathcal{G}_0 for which Assumption AII, holds is typically smaller than \mathcal{G} itself, at least for the currently available estimates we use for numerical approximations of Markov chains and processes.

We will prove the following theorems using these assumptions. For the second, we require two definitions: a function $\bar{G} : \mathbb{S} \rightarrow [1, \infty)$ is *norm-like* if

$$\lim_{n \rightarrow \infty} \left(\inf_{x \in \mathbb{S} \setminus C_n} \bar{G}(x) \right) = \infty$$

for a sequence $C_n \subset \mathbb{S}$ of compact sets with $C_n \uparrow \mathbb{S}$. A Markov chain is *weak Feller* if the transition kernel maps the space of bounded continuous functions on \mathbb{S} into itself.

THEOREM 3.1. *Under Assumptions AI and AII, there are $K > 0$, $\varepsilon_c > 0$ such that, for all $0 < \varepsilon < \varepsilon_0$, there is $N = N(\varepsilon)$:*

$$(3.1) \quad \sup_{G \in \mathcal{G}_0} |\mathbb{E}^{\delta_x} G(u_n^\varepsilon) - \pi(G)| \leq 2K \max\{\pi(\bar{G}), \bar{G}(x)\} \varepsilon^\gamma \quad \forall n \geq N,$$

where $\gamma = \log r / \log(\zeta^2 r) \in (0, 1)$. If the Markov chain (2.2) has an invariant probability measure π^ε , then

$$\sup_{G \in \mathcal{G}_0} |\pi^\varepsilon(G) - \pi(G)| \leq 6K\pi(\bar{G})\varepsilon^\gamma.$$

COROLLARY 3.2. *Let \mathbb{S} be finite dimensional and the functional \bar{G} be norm-like, and let Assumptions AI and AII hold. If the perturbed chain (2.2) is weak Feller, then it has at least one invariant probability measure π^ε .*

The theorem confines stabilization of expected values under (2.2) to a small neighborhood of their limit under (2.1). The corollary shows that, under some regularity conditions and in finite dimensions, the perturbed chain has an invariant measure, which, in general, is not unique (see the Appendix for an example). The corollary applies only to finite dimensional vector spaces; those interested in existence of invariant measures on infinite dimensional spaces should look in [5] and also at the papers [1] and [28].

Note that if the perturbed chain is ergodic, then Theorem 3.1 implies that time averages of G under the perturbed chain will converge to a limit $\pi^\varepsilon(G)$ which is $\mathcal{O}(\varepsilon^\gamma)$ close to the time average $\pi(G)$ under the original chain. It remains an open and interesting question to study when, indeed, ergodicity is inherited by the perturbed chain under the type of assumptions made in this paper.

Proof of Theorem 3.1. Choose $N \in \mathbb{R}$ such that $\varepsilon\zeta^{2N} = r^{-N}$ so that $r^{-N} = \varepsilon^\gamma$. By choosing ε_c sufficiently small we can ensure that $N \geq 2$ for all $\varepsilon \leq \varepsilon_c$. With $N \geq 2$ we have, by Assumptions AI and AII,

$$\begin{aligned} \sup_{G \in \mathcal{G}_0} |\mathbb{E}^{\delta_x} G(u_n) - \pi(G)| &\leq rR_1\varepsilon^\gamma\bar{G}(x) \quad \forall n \geq N - 1, \\ \sup_{G \in \mathcal{G}_0} |\mathbb{E}^{\delta_x} G(u_n^\varepsilon) - \mathbb{E}^{\delta_x} G(u_n)| &\leq R_2\varepsilon^\gamma\bar{G}(x) \quad \forall n \leq 2N. \end{aligned}$$

Thus,

$$(3.2) \quad \sup_{G \in \mathcal{G}_0} |\mathbb{E}^{\delta_x} G(u_n^\varepsilon) - \pi(G)| \leq (rR_1 + R_2)\varepsilon^\gamma\bar{G}(x) \quad \forall n : N - 1 \leq n \leq 2N.$$

Let $K = rR_1 + R_2$ and assume, for induction, that the following holds for some $M \geq 1$:

$$(3.3) \quad \begin{aligned} \sup_{G \in \mathcal{G}_0} |\mathbb{E}^{\delta_x} G(u_n^\varepsilon) - \pi(G)| &\leq \pi(\bar{G}) \sum_{j=1}^{M-1} (K\varepsilon^\gamma)^j + \bar{G}(x)(K\varepsilon^\gamma)^M \\ \forall n : NM \leq n < N(M + 1). \end{aligned}$$

Note that (3.3) holds for $M = 1$ by (3.2).

Let p be any integer $N(M + 1) \leq p < N(M + 2)$. Let n be the closest integer to $(p - N)$ subject to the constraint that n lies in $[NM, N(M + 1))$. Then, since $N \geq 2$, $p = n + L$ where $N - 1 \leq L \leq 2N$. Now, by conditional expectation and the Markov property, for any $G \in \mathcal{G}_0$,

$$\begin{aligned} |\mathbb{E}^{\delta_x} G(u_p^\varepsilon) - \pi(G)| &= |\mathbb{E}^{\delta_x} G(u_{n+L}^\varepsilon) - \pi(G)| \\ &= |\mathbb{E}^{\mu_L^\varepsilon} G(u_n^\varepsilon) - \pi(G)| \\ &= \left| \int_{\mathbb{S}} [\mathbb{E}^{\delta_x} G(u_n^\varepsilon) - \pi(G)] \mu_L^\varepsilon(dx) \right| \\ &\leq \int_{\mathbb{S}} |\mathbb{E}^{\delta_x} G(u_n^\varepsilon) - \pi(G)| \mu_L^\varepsilon(dx). \end{aligned}$$

Thus, by the inductive hypothesis (3.3),

$$\begin{aligned}
 |\mathbb{E}^{\delta_x} G(u_p^\varepsilon) - \pi(G)| &\leq \int_{\mathbb{S}} \pi(\bar{G}) \sum_{j=1}^{M-1} (K\varepsilon^\gamma)^j + \bar{G}(x)(K\varepsilon^\gamma)^M \mu_L^\varepsilon(dx) \\
 (3.4) \qquad \qquad \qquad &= \pi(\bar{G}) \sum_{j=1}^{M-1} (K\varepsilon^\gamma)^j + (K\varepsilon^\gamma)^M \mu_L^\varepsilon(\bar{G}).
 \end{aligned}$$

But $\mu_L^\varepsilon(\bar{G}) = \mathbb{E}^{\delta_x} \bar{G}(u_L^\varepsilon)$ and so, by (3.2), since $\bar{G} \in \mathcal{G}_0$,

$$\begin{aligned}
 (3.5) \qquad \qquad \mu_L^\varepsilon(\bar{G}) &= \pi(\bar{G}) + [\mathbb{E}^{\delta_x} \bar{G}(u_L^\varepsilon) - \pi(\bar{G})] \\
 &\leq \pi(\bar{G}) + K\varepsilon^\gamma \bar{G}(x)
 \end{aligned}$$

and combining (3.4) and (3.5) gives

$$\begin{aligned}
 |\mathbb{E}^{\delta_x} G(u_p^\varepsilon) - \pi(G)| &\leq \pi(\bar{G}) \sum_{j=1}^{M-1} (K\varepsilon^\gamma)^j + (K\varepsilon^\gamma)^M \pi(\bar{G}) + (K\varepsilon^\gamma)^{M+1} \bar{G}(x) \\
 &= \pi(\bar{G}) \sum_{j=1}^M (K\varepsilon^\gamma)^j + (K\varepsilon^\gamma)^{M+1} \bar{G}(x).
 \end{aligned}$$

Since $G \in \mathcal{G}_0$ is arbitrary, (3.3) is established for all $M \geq 1$ by induction. Now reduce ε_c so that $K\varepsilon^\gamma \leq \frac{1}{2}$ for $0 < \varepsilon < \varepsilon_c$. Then (3.3) implies, for $0 < \varepsilon < \varepsilon_c$,

$$\begin{aligned}
 \sup_{G \in \mathcal{G}_0} |\mathbb{E}^{\delta_x} G(u_n^\varepsilon) - \pi(G)| &\leq \max\{\pi(\bar{G}), \bar{G}(x)\} \sum_{j=1}^M (K\varepsilon^\gamma)^j \\
 &\leq 2 \max\{\pi(\bar{G}), \bar{G}(x)\} K\varepsilon^\gamma,
 \end{aligned}$$

giving (3.1). Average (3.1) over data distributed according to π^ε to obtain

$$\sup_{G \in \mathcal{G}_0} |\pi^\varepsilon(G) - \pi(G)| \leq 2K \{\pi(\bar{G}) + \pi^\varepsilon(\bar{G})\} \varepsilon^\gamma.$$

Thus, possibly by further reduction of ε_c ,

$$\pi^\varepsilon(\bar{G}) \leq 2\pi(\bar{G}),$$

and the required result follows. \square

Proof of Corollary 3.2. Under the stated hypotheses, (3.1) holds and so Proposition 12.1.4 in [23] applies, giving the existence of an invariant probability measure. \square

We consider also the approximation of the time-continuous Markov process $\{u(t)\}_{t \geq 0}$ by (2.2). For this we will use the following modifications of AI–AII.

Assumption ACI. The Markov process $\{u(t)\}_{t \geq 0}$ is geometrically ergodic. Specifically fix $\bar{G} : \mathbb{S} \rightarrow [1, \infty)$ and define $\mathcal{G} = \{\text{measurable } G : \mathbb{S} \rightarrow \mathbb{R}, |G| \leq \bar{G}\}$. Assume that $\bar{G} \in \mathcal{L}^1(\mathbb{S}, \mathbb{B}(\mathbb{S}), \mu(t)) \forall t \geq 0$. For some $R_1 > 0, \omega > 0$ and some set $\mathcal{G}_0 \subseteq \mathcal{G}$ containing \bar{G} , further assume that

$$\sup_{G \in \mathcal{G}_0} |\mathbb{E}^{\delta_x} G(u(t)) - \pi(G)| \leq R_1 e^{-\omega t} \bar{G}(x) \quad \forall t \geq 0$$

with π the unique invariant measure on $(\mathbb{S}, \mathbb{B}(\mathbb{S}))$.

Assumption ACII. The Markov chain $\{u_n^\varepsilon\}_{n \geq 0}$ approximates the time-continuous Markov chain $\{u(t)\}_{t \geq 0}$ in the following way: for some $\kappa > 0, R_2 > 0$ and $s > 0$,

$$\sup_{G \in \mathcal{G}_0} |\mathbb{E}^{\delta_x} G(u(n\varepsilon)) - \mathbb{E}^{\delta_x} G(u_n^\varepsilon)| \leq R_2 \bar{G}(x) e^{\kappa n \varepsilon} \varepsilon^s \quad \forall n : n\varepsilon \geq 1.$$

Note that for SDEs the condition in ACI is typically satisfied for any $G \in \mathcal{G}$. However, for the stochastic PDEs considered here the condition is currently verified only for a proper subset of \mathcal{G} in which a Lipschitz condition is satisfied. Hence the formulation of ACI in terms of \mathcal{G}_0 . In any case, for our current methods of analysis, we can only verify ACII on a proper subset of \mathcal{G} for the applications to SDEs and SPDEs considered here.

The following theorem may be proved similarly to Theorem 3.1—the necessary modifications to the proofs are outlined below.

THEOREM 3.3. *Under Assumptions ACI and ACII there is $K > 0, \varepsilon_c > 0$ such that for all $0 < \varepsilon < \varepsilon_c$, there is $T = T(\varepsilon)$:*

$$(3.6) \quad \sup_{G \in \mathcal{G}_0} |\mathbb{E}^{\delta_x} G(u_n^\varepsilon) - \pi(G)| \leq 2K \max\{\pi(\bar{G}), \bar{G}(x)\} \varepsilon^\gamma \quad \forall n : n\varepsilon \geq T,$$

where $\gamma = \frac{\omega}{\omega + 2\kappa} \in (0, 1)$. If the Markov chain (2.2) has an invariant probability measure π^ε , then

$$\sup_{G \in \mathcal{G}_0} |\pi^\varepsilon(G) - \pi(G)| \leq 6K \pi(\bar{G}) \varepsilon^\gamma.$$

COROLLARY 3.4. *Let \mathbb{S} be finite dimensional and the functional \bar{G} be norm-like, and let Assumptions ACI and ACII hold. If the perturbed chain (2.2) is weak Feller, then it has at least one invariant probability measure π^ε .*

Proofs of Theorem 3.3 and Corollary 3.4. To prove Theorem 3.3, choose $\varepsilon_c < 1$ sufficiently small so that there is $T > 1$ solving

$$e^{-\omega T} = e^{2\kappa T} \varepsilon^s, \quad 0 < \varepsilon < \varepsilon_c.$$

Then $e^{-\omega T} = \varepsilon^{\gamma s}$. By combining ACI and ACII, we deduce that

$$(3.7) \quad \sup_{G \in \mathcal{G}_0} |\mathbb{E}^{\delta_x} G(u_n^\varepsilon) - \pi(G)| \leq K \varepsilon^{s\gamma} \bar{G}(x), \quad T - \varepsilon_c \leq n\varepsilon \leq 2T,$$

where $K = R_1 \varepsilon^{\varepsilon_c \omega} + R_2$. Now assume for induction that

$$(3.8) \quad \sup_{G \in \mathcal{G}_0} |\mathbb{E}^{\delta_x} G(u_n^\varepsilon) - \pi(G)| \leq \pi(\bar{G}) \sum_{j=1}^{M-1} (K \varepsilon^{s\gamma})^j + \bar{G}(x) (K \varepsilon^{s\gamma})^M$$

$$\forall n : MT \leq n\varepsilon \leq (M + 1)T.$$

The induction now proceeds similarly to the proof of Theorem 3.1: choose p such that $(M + 1)T \leq p\varepsilon \leq (M + 2)T$ and define n to be the unique integer that is closest to $(p\varepsilon - T)/\varepsilon$ and is such that $MT \leq n\varepsilon \leq (M + 1)T$. Then define $N := p - n$. It follows that $T - \varepsilon \leq N\varepsilon \leq T + \varepsilon$. In particular, $T - \varepsilon_c \leq N\varepsilon \leq 2T$ and (3.7) holds. The induction step can again be made by using (3.7); the details are omitted. Averaging as in Corollary 3.2 gives the Corollary 3.4. \square

REMARK. Let Q (resp., Q^ε) denote the transition kernel for the chain (2.1) (resp., (2.2)) and Π denote the lift of π to a transition kernel. The invariant measure is a solution of the eigenvalue problem

$$\pi Q = \pi, \quad \pi(\mathbb{S}) = 1,$$

and in essence our interest is focused on perturbation theory for this eigenvalue problem. Such problems have, of course, been studied before. If, for given $v : \mathbb{S} \rightarrow [1, \infty)$, we define a norm on measures on \mathbb{S} by

$$\|\mu\|_v = \int_{\mathbb{S}} v(x)|\mu|(dx),$$

then the induced operator norm for transition kernels P is

$$\|P\|_v = \sup_{x \in \mathbb{S}} \frac{\int_{\mathbb{S}} v(y)|P(x, dy)|}{v(x)}.$$

This norm is discussed in some detail in [23, Chapter 16] and also in [16, 17]. (Note that if $v \equiv 1$, then $\|\cdot\|_v$ reduces to the total variation norm). It is straightforward to show that

$$\|\mu\|_v \leq \varepsilon \iff \left| \int_{\mathbb{S}} h(x)\mu(dx) \right| \leq \varepsilon \quad \forall h : |h| \leq v$$

and that

$$\|P\|_v \leq \varepsilon \iff \left| \int_{\mathbb{S}} h(y)P(x, dy) \right| \leq \varepsilon v(x) \quad \forall h : |h| \leq v.$$

In the case where $\mathcal{G}_0 \equiv \mathcal{G}$, Assumption AI states that

$$\|Q^n - \Pi\|_{\mathcal{G}} \leq \frac{R_1}{r^n} \quad \forall n \geq 0$$

and AII states that

$$\|Q^n - (Q^\varepsilon)^n\|_{\mathcal{G}} \leq R_2 \zeta^n \varepsilon \quad \forall n \geq 0.$$

Under these two conditions, Kartashov shows in [17] that Q^ε has an invariant probability measure $\mathcal{O}(\varepsilon)$ close to π , a stronger result than we are able to prove. (This is essentially a differentiability result for the invariant measures; related issues are discussed in [10].) However, while AI can often be extended to the whole of \mathcal{G} , our current analysis of numerical methods does not enable us to extend AII in this way for the problems considered here. Hence, we are unable to apply the Kartashov theory. This motivates the development of the present theory. (See [19] for a discussion of similar issues.)

4. Applications. We present applications of the foregoing theory to randomly impulsed ODEs, to Itô SDEs, and to stochastic parabolic PDEs forced by space-time Brownian noise.

4.1. Randomly impulsed differential equations. Given a vector $\nu \in \mathbb{R}^m$ (the impulse), and initial state u_0 , and a vector field $f : \mathbb{R}^m \rightarrow \mathbb{R}^m$, we consider the equation

$$(4.1) \quad \frac{du}{dt} = f(u) + \sum_{n=1}^{\infty} \theta_n \nu \delta(t - \tau_n), \quad u(0^+) = x + \theta_0 \nu.$$

Here $\delta(\cdot)$ denotes a unit point mass at the origin, the $\{\theta_n\}_{n=0}^{\infty}$ are independent, identically distributed (IID) random variables with probability of $\theta_n = \pm 1$ being

$\frac{1}{2}$, and the waiting times $t_n = \tau_{n+1} - \tau_n$, $n \in \mathbb{Z}^+$, $\tau_0 = 0$, are IID random variables exponentially distributed with parameter λ .¹ Moreover, we assume that the sequences $\{\theta_n\}_{n=0}^\infty$ and $\{t_n\}_{n=0}^\infty$ are independent. Note that $\tau_n \rightarrow \infty$, almost surely (a.s.).

We denote by $S : \mathbb{R}^m \times \mathbb{R}^+ \rightarrow \mathbb{R}^m$ the semigroup assumed to be generated by the problem

$$(4.2) \quad \frac{du}{dt} = f(u), \quad u(0) = U;$$

thus, the solution of (4.2) is $u(t) = S(U, t)$ for $t \geq 0$, and this is defined for all $t \geq 0$ and all $U \in \mathbb{R}^m$. If we define $u_n = u(\tau_n^-)$, $w_n = (\theta_n, t_n)^T$ and use the delta function to derive jump conditions on $u(t)$ across $t = t_n$, we obtain the Markov chain

$$(4.3) \quad u_{n+1} = H(u_n, w_n) := S(u_n + \theta_n \nu, t_n), \quad n \in \mathbb{Z}^+, \quad u_0 = x;$$

see [26]. (Note also that, when viewed in continuous time, (4.1) is an example of a piecewise deterministic Markov process in the general framework of Davis [7].) Let \mathcal{F}_n denote the σ algebra generated by the random variables $\{u_i : i \leq n\}$.

We assume that f satisfies the following conditions with $\|\cdot\|$ denoting the Euclidean norm.

Condition R1. $\exists C > 0 : \|f(u)\| \leq C[1 + \|u\|] \quad \forall u \in \mathbb{R}^m$.

Condition R2. f is globally Lipschitz on \mathbb{R}^m , and the Lipschitz constant K for f satisfies $K < \lambda$.

Condition R3. $\exists \alpha, \beta > 0 : \langle f(u), u \rangle \leq \alpha - \beta \|u\|^2 \quad \forall u \in \mathbb{R}^m$.

Under these conditions, the following lemma holds, as proved in [26].

LEMMA 4.1. *Consider the Markov chain (4.3) with x distributed according to a measure μ on $(\mathbb{R}^m, \mathbb{B}(\mathbb{R}^m))$, satisfying $\mu(\|x\|^2) < \infty$. If, R1–R3 hold, then $\{u_n\}_{n=0}^\infty$ exists \mathbb{P}^μ a.s. and*

$$\|u_{n+1}\|^2 \leq \frac{\alpha}{\beta} + e^{-2\beta t_n} \left(\|u_n\|^2 + 2\theta_n \langle u_n, \nu \rangle + \|\nu\|^2 - \frac{\alpha}{\beta} \right)$$

from which it follows that

$$(4.4) \quad \sup_{n \geq 0} \mathbb{E}^\mu \|u_n\|^2 \leq \max \left\{ \mu(\|x\|^2), \frac{\alpha}{\beta} + \frac{1}{2\beta} \lambda \|\nu\|^2 \right\}$$

and

$$(4.5) \quad \mathbb{E}^\mu \{ \|u_{n+1}\| - \|u_n\| | \mathcal{F}_n \} \leq \left(\sqrt{\frac{\lambda}{\lambda + 2\beta}} - 1 \right) \|u_n\| + \sqrt{\frac{\alpha}{\beta}} + \sqrt{\frac{\lambda}{\lambda + 2\beta}} \|\nu\|.$$

In the one-dimensional case (viz. $m = 1$), it is possible to show by use of (4.5) that the Markov chain (4.3) is geometrically ergodic with respect to the function $\bar{G}(u) := 1 + \|u\|$, provided an irreducibility measure, satisfying certain continuity

¹This example does not rely crucially on the exponential distribution and other densities, provided they are absolutely continuous with respect to Lebesgue measure on \mathbb{R}^+ , could also be considered; see [26]. Neither is the exponential distribution required for the subsequent numerical analysis, although rapid decay of the tails of the distribution is needed; precisely we require that e^{Kt} has finite mean against the distribution of waiting times for impulses where K is the global Lipschitz constant of f . We study only the case of exponential distributions because it allows a simple exposition and arises frequently in applications [9].

properties, exists; see [23] for the general theory and [26] for applications to (4.1). Thus, we will assume the following later in this section.

Condition R4. The Markov chain (4.3) is geometrically ergodic: for $\bar{G}(u) := 1 + \|u\|$, $\exists R > 0$, $r > 1$, such that, for

$$\mathcal{G} = \{G : \mathbb{R}^m \rightarrow [1, \infty), |G| \leq \bar{G}\},$$

we have

$$\sup_{G \in \mathcal{G}} |\mathbb{E}^{\delta_x} G(u_n) - \pi(G)| \leq R\bar{G}(x)r^{-n} \quad \forall n \geq 0,$$

where π is the unique invariant measure on $(\mathbb{R}^m, \mathbb{B}(\mathbb{R}^m))$ generated by (4.3).

We assume that the deterministic part of (4.1) is approximated by a one-step numerical method with map $S_{\Delta t}^1 : \mathbb{R}^m \rightarrow \mathbb{R}^m$ denoting one step of the method with time-step Δt (and $S_{\Delta t}^N$ as its N -fold composition); we assume that this map is defined and continuous on \mathbb{R}^m for all $\Delta t \in [0, \Delta t_c)$, which is true for all Runge–Kutta methods under R2. Given any $t \in \mathbb{R}^+$, we let $N = N(t)$ and $\Delta = \Delta(t)$ be defined by

$$N = \left\lfloor \frac{t}{\Delta t} \right\rfloor, \quad \Delta \in [0, \Delta t) : t = N\Delta t + \Delta.$$

We may thus define, for $w = (\theta, t)^T$,

$$H_{\Delta t}(u, w) := S_{\Delta}^1(S_{\Delta t}^N(u + \theta\nu))$$

and then the Markov chain

$$(4.6) \quad u_{n+1}^{\Delta t} = H_{\Delta t}(u_n^{\Delta t}, w_n), \quad u_0 = x$$

is our numerical approximation to the chain (4.3). (We assume that the same noise drives the original chain and its perturbation.) This too is defined for all $\Delta t \in [0, \Delta t_c)$. Let $\mathcal{F}_n^{\Delta t}$ denote the σ -algebra generated by the random variables $\{u_i^{\Delta t} : i \leq n\}$.

We assume that the one-step map satisfies the approximation property $\exists C > 0$, so that, for all $\Delta t \in [0, \Delta t_c)$,

$$(4.7) \quad \|S_{\Delta t}^1 u - S(\Delta t)u\| \leq C[1 + \|u\|]\Delta t^{p+1} \quad \forall u \in \mathbb{R}^m$$

for some integer $p \geq 1$; this holds for all Runge–Kutta methods under R1, R2, with $p = 1$. Note, however, that this will not, in general, give the optimal rate of convergence. (Other methods with this property are those for which f occurs only once in the elementary differentials defining the error.)

The following lemma may be proved by the techniques of Theorem 7.3.1 in [29] using (4.7).

LEMMA 4.2. *Consider the Markov chain (4.6) with x distributed according to a measure μ on $(\mathbb{R}^m, \mathbb{B}(\mathbb{R}^m))$, satisfying $\mu(\|x\|^2) < \infty$. If R1–R3 hold, then $\{u_n^{\Delta t}\}_{n=0}^{\infty}$ exists \mathbb{P}^{μ} a.s. and*

$$\|u_{n+1}^{\Delta t}\|^2 \leq \gamma + e^{-2\bar{\beta}t_n} (\|u_n^{\Delta t}\|^2 + 2\theta_n \langle u_n^{\Delta t}, \nu \rangle + \|\nu\|^2 - \gamma),$$

where $\gamma = \alpha/\beta + \mathcal{O}(\Delta t)$ and $\bar{\beta} < \beta$ satisfies $\bar{\beta} = \beta + \mathcal{O}(\Delta t)$. From this it follows that, for all Δt sufficiently small,

$$\sup_{n \geq 0} \mathbb{E}^{\mu} \|u_n^{\Delta t}\|^2 \leq \max \left\{ \mu(\|x\|^2), \gamma + \frac{1}{2\bar{\beta}} \lambda \|\nu\|^2 \right\}$$

and

$$(4.8) \mathbb{E}^\mu \{ \|u_{n+1}^{\Delta t} - u_n^{\Delta t}\| | \mathcal{F}_n^{\Delta t} \} \leq \left(\sqrt{\frac{\lambda}{\lambda + 2\beta}} - 1 \right) \|u_n^{\Delta t}\| + \sqrt{\gamma} + \sqrt{\frac{\lambda}{\lambda + 2\beta}} \|\nu\|.$$

The next result addresses the pathwise approximation of (4.3) by (4.6).

RESULT 4.3. Consider the approximation of the Markov chain (4.3) by the Markov chain (4.6), under R1–R3. Then for any measure μ on \mathbb{R}^m with $\mu(\|x\|^2) < \infty$, there is $C > 0$:

$$\mathbb{E}^\mu \|u_n - u_n^{\Delta t}\| \leq C[1 + \mu(\|x\|)] \left(\frac{\lambda}{\lambda - K} \right)^n \Delta t^p \quad \forall n \geq 0.$$

Proof. By use of R2, it follows that

$$(4.9) \quad \|H(u_1, w) - H(u_2, w)\| \leq e^{Kt} \|u_1 - u_2\| \quad \forall u_1, u_2 \in \mathbb{R}^m,$$

where $w = (\theta, t)^T$. Similarly, (4.7) implies that

$$(4.10) \quad \|H(u, w) - H_{\Delta t}(u, w)\| \leq L e^{Kt} [1 + \|u\|] \Delta t^p \quad \forall u \in \mathbb{R}^m.$$

The bound (4.10) is simply the standard convergence result for one-step methods modified by the uniform truncation error bound (4.7), and the global bound on solutions of (4.2) and its numerical approximations, induced by R3.

Now

$$(4.11) \quad \|u_{n+1} - u_{n+1}^{\Delta t}\| \leq \|H(u_n, w_n) - H(u_n^{\Delta t}, w_n)\| + \|H(u_n^{\Delta t}, w_n) - H_{\Delta t}(u_n^{\Delta t}, w_n)\|.$$

Using the properties of the exponential distribution we have, from (4.9),

$$\mathbb{E}^\mu (\|H(u_n, w_n) - H(u_n^{\Delta t}, w_n)\| | \sigma(\mathcal{F}_n, \mathcal{F}_n^{\Delta t})) \leq \left(\frac{\lambda}{\lambda - K} \right) \|u_n - u_n^{\Delta t}\|$$

(where $\sigma(\mathcal{F}_n, \mathcal{F}_n^{\Delta t})$ is the smallest σ -algebra containing both \mathcal{F}_n and $\mathcal{F}_n^{\Delta t}$), and from (4.10),

$$\mathbb{E}^\mu (\|H(u_n^{\Delta t}, w_n) - H_{\Delta t}(u_n^{\Delta t}, w_n)\| | \mathcal{F}_n^{\Delta t}) \leq C_1 [1 + \|u_n^{\Delta t}\|] \Delta t^p.$$

Hence, by (4.11),

$$(4.12) \quad \mathbb{E}^\mu \|u_{n+1} - u_{n+1}^{\Delta t}\| \leq \left(\frac{\lambda}{\lambda - K} \right) \mathbb{E}^\mu \|u_n - u_n^{\Delta t}\| + C_1 \Delta t^p \mathbb{E}^\mu (1 + \|u_n^{\Delta t}\|).$$

By (4.8) there exists $q \in (0, 1)$ and $\xi > 0$ such that

$$\mathbb{E}^\mu \|u_{n+1}^{\Delta t}\| \leq \xi + q \mathbb{E}^\mu \|u_n^{\Delta t}\|$$

so that

$$\mathbb{E}^\mu \|u_n^{\Delta t}\| \leq \left(\frac{\xi}{1 - q} \right) [1 - q^n] + q^n \mathbb{E}^\mu \|u_0^{\Delta t}\|.$$

Thus, (4.12) yields, for $e_n := \mathbb{E}^\mu \|u_n^{\Delta t} - u_n\|$,

$$e_n \leq C_2 \left[\left(\frac{\lambda}{\lambda - K} \right)^n - 1 \right] \Delta t^p + C_3 [1 + \mathbb{E}^\mu \|u_0^{\Delta t}\|] [1 - q^n] \Delta t^p,$$

giving the desired result. \square

In the following, let \mathcal{G} be as defined in R4 and, given a Lipschitz constant $C_{\mathcal{G}_0}$, define the following subset of \mathcal{G} :

$$\mathcal{G}_0 := \{G \in \mathcal{G} : |G(x) - G(y)| \leq C_{\mathcal{G}_0} \|x - y\|\}.$$

Using the pathwise approximation result proved in Result 4.3, together with geometric ergodicity of (4.3), we prove the following.

THEOREM 4.4. *Consider the approximation of the Markov chain (4.3) by the Markov chain (4.6), under R1–R4. Then, there exists $C > 0$, $\delta \in (0, 1)$, and $\Delta t_c > 0$ such that, for any $\Delta t \in [0, \Delta t_c)$, an integer $N = N(\Delta t)$ exists such that, for all $G \in \mathcal{G}_0$,*

$$(4.13) \quad |\mathbb{E}^{\delta_x} G(u_n^{\Delta t}) - \pi(G)| \leq 2C \max\{\pi(\bar{G}), \bar{G}(x)\} \Delta t^{p\delta} \quad \forall n \geq N.$$

Furthermore, the Markov chain (4.6) has an invariant probability measure $\pi^{\Delta t}$ and

$$(4.14) \quad |\pi^{\Delta t}(G) - \pi(G)| \leq 6C\pi(\bar{G})\Delta t^{p\delta}.$$

Proof. By Result 4.3 we have, for all $G \in \mathcal{G}_0$,

$$\begin{aligned} |\mathbb{E}^{\delta_x} G(u_n) - \mathbb{E}^{\delta_x} G(u_n^{\Delta t})| &\leq \mathbb{E}^{\delta_x} |G(u_n) - G(u_n^{\Delta t})| \\ &\leq C_{\mathcal{G}_0} \mathbb{E}^{\delta_x} \|u_n - u_n^{\Delta t}\| \\ &\leq CC_{\mathcal{G}_0} [1 + \|x\|] \left(\frac{\lambda}{\lambda - K} \right)^n \Delta t^p. \end{aligned}$$

Thus AII holds with $\varepsilon = \Delta t^p$. Now AI holds by R4. Thus Theorem 3.1 gives (4.13). Solutions of (4.6) are weak Feller by the continuity properties of $S_{\Delta t}^1$. Therefore, because we are working in finite dimensions, existence of an invariant measure comes from Corollary 3.2, and (4.14) follows from Theorem 3.1. \square

4.2. Itô SDEs. Consider the Itô SDE,

$$(4.15) \quad du = f(u)dt + \sigma(u)dW, \quad u(0) = x,$$

where $u \in \mathbb{R}^m$, $f : \mathbb{R}^m \rightarrow \mathbb{R}^m$, $\sigma : \mathbb{R}^m \mapsto \mathbb{R}^{m \times s}$ and W is an s -dimensional Brownian motion. We make the following assumptions about f and σ , where once again, $\|\cdot\|$ denotes the Euclidean norm and $\|\cdot\|_F$ the Frobenius norm:

Condition I1. $\|f(u)\| \leq C[1 + \|u\|]$ & $\|f(u) - f(v)\| \leq C\|u - v\| \quad \forall u, v \in \mathbb{R}^m$.

Condition I2. $\|\sigma(u) - \sigma(v)\| \leq C\|u - v\| \quad \forall u, v \in \mathbb{R}^m$.

Condition I3. $\exists \alpha, \beta > 0 : \langle f(u), u \rangle \leq \alpha - \beta\|u\|^2 \quad \forall u \in \mathbb{R}^m$.

Condition I4. $\exists \delta \in (0, 2\beta) : \|\sigma(u)\|_F^2 \leq C + \delta\|u\|^2 \quad \forall u \in \mathbb{R}^m$.

Conditions I1 and I2 ensure the existence and uniqueness of solutions \mathbb{P}^{δ_x} a.s. with respect to the product measure induced by the Brownian motion and a point mass at x as a measure on the initial data. (Note that the first point in I1 is implied by the second; the two points are stated separately for clarity.) Furthermore, using I3 and I4, we can prove

$$(4.16) \quad \sup_{t \geq 0} \mathbb{E}^{\delta_x} \|u(t)\|^2 \leq C[1 + \|x\|^2].$$

We define the $m \times m$ matrix A with entries

$$A_{ij}(u) = \sum_{r=1}^s \sigma_{ir}(u)\sigma_{jr}(u)$$

and the second-order differential operator \mathcal{L} (the adjoint of the generator) defined by, for $u = (u_1, \dots, u_m)$,

$$(\mathcal{L}\phi)(u) = \sum_{i=1}^m f_i(u) \frac{\partial \phi}{\partial u_i}(u) + \frac{1}{2} \sum_{i,j=1}^m A_{ij}(u) \frac{\partial^2 \phi}{\partial u_i \partial u_j}(u).$$

It is well known that, if the symmetric nonnegative matrix A is positive definite in a compact set containing $B(0; \sqrt{\alpha/\beta})$ in its interior, then, under I3, the Markov process generated by (4.15) is ergodic and the strong law of large numbers holds [15]. Under this type of strong ellipticity condition on \mathcal{L} it is possible to show that a variety of numerical methods reproduce large-time expectations and time averages accurately [30]. However, many problems of interest do not satisfy the strong ellipticity condition but are still ergodic; we show how the theory of section 3 may be used to verify that numerical methods still reproduce large-time expectations and time averages in such cases.

As a simple example, consider the Langevin equation,

$$(4.17) \quad \frac{d^2x}{dt^2} + \gamma \frac{dx}{dt} + V'(x) = \frac{dW}{dt}, \quad x(0) = x_0, \quad \frac{dx}{dt}(0) = x_1,$$

where W is a one-dimensional Brownian motion. This may be recast in the form (4.15) with $u = (u_1, u_2)^T := (x, \dot{x})^T$, $m = 2$, and $s = 1$:

$$f(u) = \begin{pmatrix} u_2 \\ -V'(u_1) - \gamma u_2 \end{pmatrix}, \quad \sigma(u) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad A = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Clearly A is not positive definite so that the work in [30] cannot be applied to study long-time weak approximation properties of numerical methods. However, the equation may be shown to be ergodic for $\gamma > 0$ (see [32] and [15, section 4.8]) and certain potentials V (see also [1]). If the equation could be further shown to be geometrically ergodic, the theory of the present paper would give insight to the equation's long time numerical approximation. Other work in this directions includes [12] and [22], where the equation is studied when V is quadratic in x so that the equation is linear; however, the results are based on explicit formulae and cannot be generalized to other potentials. The techniques we describe could be.

For (4.17) with V' globally Lipschitz, I1, I2 and I4 hold, although I3 is violated; however I3 simply implies the existence of a positive definite quadratic Lyapunov function and, under appropriate hypothesis on V , a similar construction can be made for (4.17) by use of new coordinates. Returning to the SDE (4.15) under I1–I4, we now assume (see I5 below) that the SDE is geometrically ergodic with respect to a quadratic function $\bar{G}(u) = 1 + \|u\|^2$ on the phase space of the problem. To motivate this choice of Lyapunov function \bar{G} , note that

$$(4.18) \quad \mathcal{L}\bar{G} \leq 2(\alpha + \beta) + \text{tr}(A) - 2\beta\bar{G}.$$

But $\text{tr}(A)$ is equal to the square of the Frobenius norm of σ and therefore, under I4, inequality (4.18) becomes

$$\mathcal{L}\bar{G} \leq 2(\alpha + \beta) + (C - \delta) - (2\beta - \delta)\bar{G},$$

which is a Foster–Lyapunov drift condition on the SDE since $\delta \in (0, 2\beta)$. Such drift conditions have been shown to imply geometric ergodicity, provided all compact sets are petite (see [23, p. 134]) for some skeleton chain (see [24, part II, section 2.3]). For further details, see, for example, [24, Theorem 6.1, part III]. Note, however, that when A is singular (so that the process is not strong Feller), further research is still needed to understand which SDEs are geometrically ergodic—even ergodic—under the drift condition. In any event, we assume the following.

Condition I5. The continuous-time Markov process generated by (4.15) is geometrically ergodic: for $\bar{G}(u) := 1 + \|u\|^2 \exists R > 0, \omega > 0$,

$$\sup_{G \in \mathcal{G}} |\mathbb{E}^{\delta_x} G(u(t)) - \pi(G)| \leq R\bar{G}(x)e^{-\omega t} \quad \forall t \geq 0,$$

where π is the unique invariant measure on $(\mathbb{R}^m, \mathbb{B}(\mathbb{R}^m))$.

Now consider the approximation of (4.15) by the Euler–Maruyama scheme

$$(4.19) \quad U_{j+1} - U_j = f(U_j)\Delta t + \sigma(U_j)\Delta W_j, \quad U_0 = x,$$

where $\Delta W_j = W((j+1)\Delta t) - W(j\Delta t)$. By the techniques described in [20] it may be shown that, for some $K > 0$,

$$(4.20) \quad \mathbb{E}^{\delta_x} (\|U_j - u(j\Delta t)\|^2) \leq C_3 e^{2KJ\Delta t} [1 + \|x\|^2] \Delta t.$$

In the following we use $\mu_n^{\Delta t}$ to denote the measure on $(\mathbb{R}^m, \mathbb{B}(\mathbb{R}^m))$ induced by the Markov chain (4.19).

We now derive a uniform mean-square bound on solutions of (4.19). Rearranging (4.19), squaring, and taking expectations yields

$$\begin{aligned} \mathbb{E}^{\delta_x} \|U_{j+1}\|^2 &\leq \mathbb{E}^{\delta_x} \|U_j\|^2 + 2\Delta t \mathbb{E}^{\delta_x} \langle U_j, f(U_j) \rangle + \mathbb{E}^{\delta_x} \|f(U_j)\|^2 \Delta t^2 + \mathbb{E}^{\delta_x} \text{tr}(A(U_j)) \Delta t \\ &\leq \mathbb{E}^{\delta_x} \|U_j\|^2 + 2\Delta t \mathbb{E}^{\delta_x} (\alpha - \beta \|U_j\|^2) \\ &\quad + C^2 \mathbb{E}^{\delta_x} (1 + \|U_j\|^2) \Delta t^2 + (\delta \mathbb{E}^{\delta_x} \|U_j\|^2 + C) \Delta t \end{aligned}$$

(by using I1–I4 and the independence between ΔW_j and U_j). Iterating this inequality gives a uniform in n bound for Δt sufficiently small since $\delta \in (0, 2\beta)$:

$$(4.21) \quad \sup_{n \geq 0} \mathbb{E}^{\delta_x} \|U_n\|^2 < C[1 + \|x\|^2].$$

We have $\bar{G}(x) = 1 + \|x\|^2$, and, for given $C_{\mathcal{G}_0}$, define

$$\mathcal{G}_0 \{G : |G| \leq \bar{G} \text{ \& \ } |G(x) - G(y)| \leq C_{\mathcal{G}_0} [1 + \|x\| + \|y\|] \|x - y\|\}.$$

Estimate (4.21) implies that $G \in \mathcal{G}_0$ are integrable with respect to $\mu_n^{\Delta t}$.

THEOREM 4.5. *Consider the approximation of the continuous-time Markov process (4.15) by the discrete-time Markov chain (4.19), under I1–I5. Then, there exists $K > 0, \gamma \in (0, 1)$, and $\Delta t_c > 0$ such that for any $\Delta t \in (0, \Delta t_c)$ an integer $N = N(\Delta t)$ exists such that, for all $G \in \mathcal{G}_0$,*

$$(4.22) \quad |\mathbb{E}^{\delta_x} G(U_n) - \pi(G)| \leq 2K \max\{\pi(\bar{G}), \bar{G}(x)\} \Delta t^{\gamma/2} \quad \forall n \geq N.$$

Furthermore, the Markov chain (4.19) has an invariant probability measure $\pi^{\Delta t}$ and

$$(4.23) \quad |\pi^{\Delta t}(G) - \pi(G)| \leq 6K \pi(\bar{G}) \Delta t^{\gamma/2}.$$

Proof. By I5, ACI holds for the given definition of \mathcal{G}_0 ; indeed it holds for all $G \in \mathcal{G}$, which is strictly bigger than \mathcal{G}_0 . Now let $G \in \mathcal{G}_0$. Then

$$\begin{aligned} & |\mathbb{E}^{\delta_x} G(u(n\Delta t)) - \mathbb{E}^{\delta_x} G(U_n)| \\ & \leq \mathbb{E}^{\delta_x} \{C_{\mathcal{G}_0} [1 + \|u(n\Delta t)\| + \|U_n\|] \|u(n\Delta t) - U_n\|\} \\ & \leq C_{\mathcal{G}_0} \{\mathbb{E}^{\delta_x} [1 + \|u(n\Delta t)\| + \|U_n\|]^2 \mathbb{E}^{\delta_x} \|u(n\Delta t) - U_n\|^2\}^{\frac{1}{2}}. \end{aligned}$$

By (4.16), (4.20), and (4.21) we deduce that

$$\sup_{G \in \mathcal{G}_0} |\mathbb{E}^{\delta_x} G(u(n\Delta t)) - \mathbb{E}^{\delta_x} G(U_n)| \leq C \bar{G}(x) \Delta t^{\frac{1}{2}} e^{\alpha n \Delta t} \quad \forall n \Delta t \geq 0.$$

Thus, ACII also holds, yielding (4.22) by Theorem 3.3. Corollary 3.4 applies because we are working in finite dimensions and because the weak Feller property holds for (4.19). This theorem completes the proof. \square

This proof employs I3 and I4 only in establishing the uniform bounds (4.16) and (4.21) for $u(t)$ and U_n ; the proof goes through without these assumptions on the smaller class \mathcal{G}_0 of globally Lipschitz test functions—a case relevant for the Langevin equation mentioned previously.

4.3. Parabolic SPDEs. Consider the following SPDE on $L_2(0, 1)$:

$$(4.24) \quad du = [-Au + f(u)]dt + dW(t), \quad u(0) = x,$$

where $A = -\Delta$, the Laplacian with domain $H^2(0, 1) \cap H_0^1(0, 1)$, f is globally Lipschitz from $L_2(0, 1)$ to itself, $W(\cdot)$ is an $L_2(0, 1)$ -valued Wiener process with covariance Q , and dW is interpreted as an Itô integral. If Q has a complete set of eigenvectors e_i , corresponding eigenvalues λ_i , and $\beta_i(\cdot)$ is a set of IID Brownian motions, then the process $W(t)$ may be written as

$$(4.25) \quad W(t) = \sum_{i=1}^{\infty} \lambda_i^{1/2} e_i \beta_i(t).$$

Questions of existence and regularity of this equation are discussed by Da Prato–Zabczyk [4]. The hypotheses gathered here assure existence of a unique mild solution for bounded Q .

It is natural to ask when (4.24) is geometrically ergodic. In [5], ACI is obtained for bounded Lipschitz functions \bar{G} with a given Lipschitz constant and that is subject to the contractivity assumption, $\exists \omega > 0$, such that

$$(4.26) \quad \langle -A(u - v) + f(u) - f(v), u - v \rangle_{L_2(0,1)} \leq -\omega \|u - v\|_{L_2(0,1)}^2 \quad u, v \in L_2(0, 1).$$

This geometric ergodicity can be easily extended to globally Lipschitz test functionals when $\mathbb{E}^{\delta_x} \|u(t)\|_{L_2(0,1)}^2 < \infty$; see [27]. However, the contractivity assumption is highly restrictive. The work of Shardlow [28] proves geometric ergodicity for the space of test functionals dominated by $1 + \|\cdot\|_{L_2(0,1)}$, subject to the covariance operator Q being nonsingular and the condition

$$\exists \alpha, \beta > 0 : \quad \langle -Au + f(u), u \rangle \leq \alpha - \beta \|u\|_{L_2(0,1)}^2.$$

(Some related work may be found in [8].)

Downloaded 02/10/16 to 137.205.50.42. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Motivated by the geometric ergodicity proved in [5] and extended in [27] and [28], we employ the following assumption:

Assumption SI. The continuous-time Markov process generated by (4.24) is geometrically ergodic: given $\bar{G}(u) := 1 + \|u\|_{L_2(0,1)}$ and $C_{\mathcal{G}_0}$, define $\mathcal{G} = \{G : L^2(0,1) \rightarrow \mathbb{R}^+, |G| \leq \bar{G}\}$ and $\mathcal{G}_0 = \{G \in \mathcal{G} : Lip\{G\} \leq C_{\mathcal{G}_0}\}$. Then $\exists R > 0, \omega > 0$:

$$\sup_{G \in \mathcal{G}_0} |\mathbb{E}^{\delta_x} G(u(t)) - \pi(G)| \leq R\bar{G}(x)e^{-\omega t} \quad \forall t \geq 0,$$

where π is the invariant measure on $(L_2(0,1), \mathbb{B}(L_2(0,1)))$ of (4.24).

We consider the numerical approximation of (4.24) by finite differences, in particular, by the θ method in time, the standard three-point approximation in space, and a spectral approximation to W (truncation of (4.25)). For fixed $J \in \mathbb{N}$, this method yields grid functions $U_{nj}, j = 1, \dots, J$ and $n \in \mathbb{N}$, where U_{nj} approximates $u(n\Delta t, j\Delta x)$ ($\Delta x = 1/(J+1)$ is the grid size and Δt is the time step). We prefer to work with continuous interpolants u_n in $L_2(0,1)$ of U_{nj} . To do this, define $A_{\Delta x} : L_2(0,1) \rightarrow L_2(0,1)$ by

$$A_{\Delta x} \sin(k\pi \cdot) = \lambda_k \sin(k\pi \cdot), \quad k = 1, 2, \dots,$$

where $\lambda_{k+nJ} = 4 \sin^2(k\pi\Delta x/2)/\Delta x^2$ ($k = 1, \dots, J, n = 1, \dots$). This operator acts exactly as the standard three-point approximation to the Laplacian on the grid $\{j\Delta x : j = 1, \dots, J\}$. We study the following Markov chain

$$(4.27) \quad u_{n+1} - u_n = [(1-\theta)A_{\Delta x}u_n + \theta A_{\Delta x}u_{n+1} + (1-\theta)f(u_n) + \theta f(u_{n+1})]\Delta t + dW_n,$$

where

$$dW_n = \int_{n\Delta t}^{(n+1)\Delta t} \mathbb{P}_J dW(t);$$

\mathbb{P}_J is the projection of $L_2(0,1)$ onto the first J eigenfunctions of A . This Markov chain, which agrees with the finite-difference scheme outlined above on the grid $\{j\Delta x : j = 1, \dots, J\}$, is discussed in [27].

Other finite-difference approximations to (4.24) are studied in the literature. For example, Gyöngy [14] uses an approximation to the white noise based on integrating over boxes of size $\Delta x \times \Delta t$. He works in a more general setting than we do here (with multiplicative noise and very mild measurability hypotheses on f) and obtains similar results to those below in Lemma 4.6; however, the results in [14] are not directly relevant to the present perturbation theory for invariant measures because the dependence on the L_2 -norm of the initial data and the growth of his estimate in large time are not described. The estimates of [27] given below for the method (4.27) depend on stronger hypothesis on f but are L_2 estimates for L_2 initial data, which, with the geometric ergodicity results available for (4.24), enable us to apply Theorems 3.3.

The assumptions we make about f are as follows:

Assumption SII. There exist constants K_1, K_2 such that

$$\begin{aligned} \|f(x) - f(y)\|_{L_2(0,1)} &\leq K_1 \|x - y\|_{L_2(0,1)}, \quad x, y \in L_2(0,1); \\ \|f(x)\|_{L_2(0,1)} &\leq K_2 + K_1 \|x\|_{L_2(0,1)}, \quad x \in L_2(0,1); \\ \|A^{1/4}f(x)\|_{L_2(0,1)} &\leq K_2 + K_1 \|A^{1/4}x\|_{L_2(0,1)}, \quad x \in L_2(0,1). \end{aligned}$$

(Note that the second point is implied by the first; we list them separately for clarity.) Using, SII, ACII may be established for certain approximations of SPDEs. The following is proved by Shardlow [27].

LEMMA 4.6. *Let hypothesis SII hold and consider approximation of (4.24) by (4.27). Consider initial data $x \in L_2(0, 1)$ for the problem (4.24), and let $(\Delta x, \Delta t) \rightarrow 0$ subject to the stability condition*

$$(4.28) \quad \frac{\Delta t}{\Delta x^2} 4(1 - \theta) \leq 1.$$

For $0 < \delta < 1$, there exists $C, \kappa, \Delta t_c > 0$ such that, for $n\Delta t > 0$ and $0 < \Delta t < \Delta t_c$,

$$(\mathbb{E}^{\delta x} \|u(n\Delta t) - u_n\|_{L_2(0,1)}^2)^{1/2} \leq C\Delta x^{(1-2\delta)/2} e^{\kappa n\Delta t} (1 + \|x\|_{L_2(0,1)}) \times (1 + (n\Delta t)^{(\delta-1)/4}).$$

In particular, for any $G \in \mathcal{G}_0$ the following holds for $n\Delta t > 1$:

$$|\mathbb{E}^{\delta x} G(u(n\Delta t)) - \mathbb{E}^{\delta x} G(u_n)| \leq 2CC_{\mathcal{G}_0} \Delta x^{(1-2\delta)/2} e^{\kappa n\Delta t} (1 + \|x\|_{L_2(0,1)}).$$

The following result is now readily proved by the application of Theorem 3.3. The result is weaker than that proved in sections 4.1–4.2, as we are unable to establish existence of an invariant measure for the numerical approximation. It is likely that an existence result for an invariant measure could be proved by use of a uniform-in-time bound on the numerical solution in a stronger norm ($H_0^1(0, 1)$, for example). However, we do not investigate this issue further.

THEOREM 4.7. *Consider the approximation of the Markov process (4.24) by the Markov chain (4.27), under SI–SII. Then, there exists $C > 0$, $\gamma \in (0, 1)$, and $\Delta t_c > 0$ such that, for any $\Delta t \in [0, \Delta t_c)$, an integer $N = N(\Delta t)$ exists with*

$$(4.29) \quad |\mathbb{E}^{\delta x} G(u_n) - \pi(G)| \leq 2C \max\{\pi(\bar{G}), \bar{G}(x)\} \Delta x^{\gamma/2} \quad \forall n \geq N \quad \forall G \in \mathcal{G}_0.$$

Furthermore, if the Markov chain (4.27) has an invariant probability measure $\pi^{\Delta x}$, then

$$(4.30) \quad |\pi^{\Delta x}(G) - \pi(G)| \leq 6C\pi(\bar{G})\Delta x^{\gamma/2}.$$

5. Conclusions. We have presented an approach to understanding the effects of perturbations on ergodic properties of Markov chains, especially perturbations arising through numerical approximation.

The approach is quite general, and, because of this, it is likely that the results are not sharp in some of the applications. For example, by use of Malliavin calculus (and the notion of hypo-ellipticity; see, for example, [25]), it may be possible to produce improved rates of convergence for the case of nonuniformly parabolic generators in SDEs, the case which we examined in section 4.2. Some preliminary numerical analysis of methods applied to SDEs with hypo-elliptic generators is presented in [13].

A further interesting question left open here is what additional hypotheses are required to prove ergodicity of the perturbed chains. It would be of interest to answer this question in some generality, but it would also be of interest to address it for some specific applications, such as the Itô SDEs of section 4.2; again the Malliavin calculus would be likely to play an important role in any such studies.

Appendix. We give a simple example of a perturbed Markov chain which cannot be irreducible, whatever the properties of the underlying chain. Consider a Markov chain $\{u_n, n \in \mathbb{Z}^+\}$ on \mathbb{R} , generated by a random map

$$u_{n+1} = H(u_n, w_n),$$

with w_n IID random variables. Fix $\varepsilon > 0$ and partition \mathbb{R} as $\mathbb{R} = \mathbb{W} \cup \mathbb{B}$, where

$$\mathbb{W} = \bigcup_{n \in \mathbb{Z}} [(2n-1)\varepsilon, 2n\varepsilon) \quad \text{and} \quad \mathbb{B} = \bigcup_{n \in \mathbb{Z}} [2n\varepsilon, (2n+1)\varepsilon).$$

We define

$$\mathcal{P}^{\mathbb{B}}u = \begin{cases} u - \varepsilon, & \text{if } u \in \mathbb{W}, \\ u, & \text{otherwise,} \end{cases} \quad \text{and} \quad \mathcal{P}^{\mathbb{W}}u = \begin{cases} u - \varepsilon, & \text{if } u \in \mathbb{B}, \\ u, & \text{otherwise.} \end{cases}$$

Clearly, $\mathcal{P}^{\mathbb{B}}$ maps onto \mathbb{B} and $\mathcal{P}^{\mathbb{W}}$ maps onto \mathbb{W} . Set

$$\mathcal{P}_u = \begin{cases} \mathcal{P}^{\mathbb{B}}, & \text{if } u \in \mathbb{B}, \\ \mathcal{P}^{\mathbb{W}}, & \text{if } u \in \mathbb{W}. \end{cases}$$

Now consider the perturbed Markov chain

$$u_{n+1}^\varepsilon = H^\varepsilon(u_n^\varepsilon, w_n),$$

where

$$H^\varepsilon(u, w) = \mathcal{P}_u H(u, w).$$

If ε is small, this Markov chain will be a small perturbation of the original one. However, if $u_0^\varepsilon \in \mathbb{B}$ (resp., \mathbb{W}) then $u_n^\varepsilon \in \mathbb{B}$ (resp., \mathbb{W}) for all $n \geq 0$. Hence the perturbed problem is not irreducible in any circumstances.

Acknowledgments. We thank Peter Baxendale, Peter Glynn James Norris, and Neil O'Connell for helpful discussions and also the referees for a number of valuable comments.

REFERENCES

- [1] L. ARNOLD AND W. KLIEMANN, *On unique ergodicity for degenerate diffusions*, Stochastics, 21 (1987), pp. 41–61.
- [2] V. BALLY AND D. TALAY, *The Euler scheme for stochastic differential equations: Error analysis with Malliavin calculus*, Math. Comput. Simulation, 38 (1995), pp. 35–41.
- [3] G. BLANKENSHIP AND G. C. PAPANICOLAOU, *Stability and control of stochastic systems with wide-band noise disturbances I*, SIAM J. Appl. Math., 34 (1978), pp. 437–476.
- [4] G. DA PRATO AND J. ZABCZYK, *Stochastic Equations in Infinite Dimensions*, Cambridge University Press, Cambridge, UK, 1992.
- [5] G. DA PRATO AND J. ZABCZYK, *Ergodicity for Infinite Dimensional Systems*, Cambridge University Press, Cambridge, UK, 1996.
- [6] A. M. DAVIE AND J. G. GAINES, *Convergence of implicit schemes for numerical solution of parabolic stochastic partial differential equations*, Math. Comp., to appear.
- [7] M. H. A. DAVIS, *Markov Models and Optimization*, Chapman and Hall, London, 1993.
- [8] W. E. K. KHANIN, A. MAZEL, AND Y. SINAI, *Probability distribution functions for the random forced Burgers equation*, Phys. Rev. Lett., 78 (1997), pp. 1904–1907.
- [9] B. FRISTEDT AND L. GRAY, *A Modern Approach to Probability Theory*, Birkhäuser, Cambridge, MA, 1997.
- [10] P. W. GLYNN AND P. L'ECUYER, *Likelihood ratio gradient estimation for stochastic recursions*, Adv. in Appl. Probab., 27 (1995), pp. 1019–1053.
- [11] W. GRECKSCH AND P. KLOEDEN, *Time-discretized Galerkin approximations of parabolic stochastic PDEs*, Bull. Austral. Math. Soc., 54 (1996), pp. 79–85.
- [12] N. GRONBECH-JENSEN AND S. DONIACH, *Long-time overdamped dynamics of molecular chains*, J. Comput. Chem., 15 (1994), pp. 997–1012.
- [13] A. GRORUD AND D. TALAY, *Approximation of Lyapunov exponents of non-linear stochastic differential equations*, SIAM J. Appl. Math., 56 (1996), pp. 627–650.

- [14] I. GYÖNGY, *Lattice approximation for stochastic quasi-linear parabolic partial differential equations driven by space-time white noise*, *Potential Anal.*, 9 (1998), pp. 1–25, 11 (1999), pp. 1–37.
- [15] R. Z. HASMINSKII, *Stochastic Stability of Differential Equations*, Sijthoff and Noordhoff, Groningen, The Netherlands, 1980.
- [16] N. V. KARTASHOV, *Criteria for uniform ergodicity and strong stability of Markov chains with a common phase space*, *Theory Probab. Math. Statist.*, 30 (1985), pp. 71–89.
- [17] N. V. KARTASHOV, *Inequalities in theorems of ergodicity and stability for Markov chains with a common phase space, Parts I and II*, *Theory Probab. Appl.*, 30 (1986), pp. 247–259 and 505–515.
- [18] N. V. KARTASHOV, *Strong Stable Markov Chains*, VSP, Utrecht, The Netherlands, 1996.
- [19] G. KELLER, *Stochastic stability in some chaotic dynamical systems*, *Monatsh. Math.*, 94 (1982), pp. 313–333.
- [20] P. E. KLOEDEN AND E. PLATEN, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, New York, 1991.
- [21] H. J. KUSHNER, *Approximation and Weak Convergence Methods for Random Processes*, MIT Press, Cambridge, 1984.
- [22] S. P. MEYN AND R. L. TWEEDIE, *Markov Chains and Stochastic Stability*, 2nd ed., Springer-Verlag, London, 1996.
- [23] S. P. MEYN AND R. L. TWEEDIE, *Stability of Markovian Processes, I, II, and III*, *Adv. Appl. Probab.*, 24 (1992), pp. 542–574, 25 (1993), pp. 487–517, and 25 (1993), pp. 518–548.
- [24] B. MISHRA AND T. SCHLICK, *The notion of error in Langevin dynamics*, *J. Chem. Phys.*, 105 (1996), pp. 299–318.
- [25] J. NORRIS, *Malliavin calculus made simple*, *Séminaire de Probabilités XX*, *Lecture Notes in Math.* 1204, Springer-Verlag, New York, pp. 101–130.
- [26] J. M. SANZ-SERNA AND A. M. STUART, *Ergodic properties of dissipative differential equations subject to random impulses*, *J. Differential Equations*, 155 (1999), pp. 262–284.
- [27] T. SHARDLOW, *Numerical Methods for Parabolic Dissipative SPDEs*, *Numer. Funct. Anal. Optim.*, 20 (1999), pp. 121–146.
- [28] T. SHARDLOW, *Geometric ergodicity for stochastic PDEs*, *Stochastic Anal. Appl.*, 17 (1999), pp. 857–869.
- [29] A. M. STUART AND A. R. HUMPHRIES, *Dynamical Systems and Numerical Analysis*, Cambridge University Press, Cambridge, UK, 1996.
- [30] D. TALAY, *Second-order discretization schemes for stochastic differential systems for the computation of the invariant law*, *Stochastics Stochastics Rep.*, 29 (1990), pp. 13–36.
- [31] D. TALAY, *Approximation of upper Lyapunov exponents of bilinear stochastic differential systems*, *SIAM J. Numer. Anal.*, 28 (1991), pp. 1141–1164.
- [32] M. M. TROPPER, *Ergodic properties and quasideterministic properties of finite-dimensional stochastic systems*, *J. Statist. Phys.*, 17 (1977), pp. 491–509.