Ergodicity For SDEs and Approximations: Locally Lipschitz Vector Fields And Degenerate Noise *

J.C. Mattingly[1], A.M. Stuart[2] and D.J. Higham[3].

Abstract

The ergodic properties of SDEs, and various time discretizations for SDEs, are studied. The ergodicity of SDEs is established by using techniques from the theory of Markov chains on general state spaces. Application of these Markov chain results leads to straightforward proofs of ergodicity for a variety of SDEs, in particular for problems with degenerate noise and for problems with locally Lipschitz vector fields. The key points which need to be verified are the existence of a Lyapunov function inducing returns to a compact set, a uniformly reachable point from within that set, and some smoothness of the probability densities; the last two points imply a minorization condition. Together the minorization condition and Lyapunov structure give geometric ergodicity. Applications include the Langevin equation, the Lorenz equation with degenerate noise and gradient systems. The ergodic theorems proved are strong, yielding exponential convergence of expectations for classes of measurable functions restricted only by the condition that they grow no faster than the Lyapunov function.

The same Markov chain theory is then used to study time-discrete approximations of these SDEs. It is shown that the minorization condition is robust under approximation. For globally Lipschitz vector fields this is also true of the Lyapunov condition. However in the locally Lipschitz case the Lyapunov condition fails for explicit methods such as Euler-Maruyama; it is, in general, only inherited by specially constructed implicit discretizations. Examples of such discretization based on backward Euler methods are given, and approximation of the Langevin equation studied in some detail.

Key Words: Geometric Ergodicity, Stochastic Differential Equations, Langevin equation, Monotone, Dissipative and Gradient Systems, Additive Noise, Hypoelliptic and Degenerate Diffusions, Time-Discretization.

[1]Mathematics Department, Building 380, Stanford University, Stanford CA 94305, USA. Supported by the National Science Foundation under grant DMS-9971087

[2]Mathematics Institute, Warwick University, Coventry, CV4 7AL, England. Supported by the Engineering and Physical Sciences Research Council of the UK under grant GR/N00340.

[3]Mathematics Department, University of Strathclyde, Glasgow G1 1XH, Scotland. Supported by the Engineering and Physical Sciences Research Council of the UK under grant GR/M42206.

1 Introduction

The primary objective of this paper is to study ergodicity of dynamical systems subject to noise, especially stochastic differential equations (SDEs) with additive noise and their time discretizations. In particular, we are interested in problems where the noise is degenerate and the vector field governing the deterministic flow is not necessarily globally Lipschitz. Such situations arise frequently in applications. A secondary objective is to develop a framework for establishing ergodicity which is both simple and, at the same time, sufficiently adaptable to allow application to both SDEs and their time-discretizations, especially those appropriate for computer simulation. We show that explicit methods such as Euler-Maruyama can fail to be ergodic, even when the underlying SDE is ergodic; we introduce variants of the backward Euler method which overcome this difficulty.

To fulfill the secondary objective, section 2 contains the statement of a theory of geometric ergodicity for Markov chains. (We use the term geometric ergodicity to mean the existence of an invariant measure π to which there is exponentially fast convergence.) No essentially new ideas are presented in section 2, but the treatment is self-contained and applicable in a straightforward way to both continuous and discrete time; in addition it is unencumbered by machinery required for situations more general than those of interest to us. Our treatment is influenced by the work of Meyn and Tweedie ([16, 17]), Durrett [3] and Has'minskii ([9]). Indeed many of the ergodicity results in this paper could be proved by combining various results in [16, 17]. However, by stating and proving a theorem tailored to our needs, we believe that the subsequent material is made more accessible; we prove the theorem with a straightforward coupling argument given in the Appendix. Our approach, when proving ergodicity for SDEs, is to use knowledge of the deterministic flow explicitly. Restricting to globally Lipschitz drift terms and a non-degenerate diffusion matrix, it is possible to prove ergodicity with little knowledge of the flow. Yet many interesting equations which do not meet these stringent conditions are nonetheless tractable, given a little knowledge of the underlying noise free dynamics; this is the case for non-linear stochastic equations that have deterministic counterparts for which the geometry of the phase space is well understood.

Sections 3, 4 and 5 are devoted to a variety of applications. In all cases, the noise free equations are dissipative in the general sense of [8]. The Lyapunov functions used to prove this dissipativity are natural candidates for establishing the supermartingale structure (outside a compact set) which underlies the theory of geometric ergodicity in section 2; see [5] for a treatise on the use of Lyapunov functions to study the ergodicity of countable state-space Markov chains. Our results are complementary to the work of Kleimann [1, 11] where invariant control sets are used to partition the state space, and Markov chain properties studied on these distinct control sets. We essentially work in settings where there is exactly one invariant control set.

Section 3 is concerned with the Langevin equation, describing the motion of a particle subject to a central force and interacting with a heat bath [6]. The noise is degenerate because it acts directly only on the momentum co-ordinates and not positions. We generalize previous results in [31] where semigroup techniques were employed. Section 4 is concerned with monotone and dissipative problems where the underlying deterministic flow has an equilibrium point with non-trivial stable manifold. The basic idea for monotone problems comes from [4] where it is used to study Galerkin approximations of the Navier-Stokes equation at arbitrary Reynolds number and subject to degenerate noise. In the Navier-Stokes equations, the underlying deterministic flow is monotone; here the approach is generalized considerably to allow study of a variety of dissipative problems, including the Lorenz equations subject to degenerate noise. Ideas similar to those in section 4 are employed in section 5 to study gradient systems with, possibly degenerate, noise. Gradient systems with non-degenerate noise are thoroughly investigated in [20] where, confusingly in the context of this paper, such problems are referred to as Langevin diffusions; here we reserve the terminology Langevin diffusions for the particle-in-a-heat-bath models of section 3, noting that in the absence of inertia these models reduce to the gradient problems of section 5.

Our presentation rests on two fundamental assumptions. The first is the existence of a Lyapunov function. This implies that outside some compact region C in the center of the phase space the dynamics move inward on average. Loosely, this allows us to restrict our attention to this central compact region. Secondly, there exists a neighborhood $\mathcal{N}$ of some distinguished point in C which is uniformly reachable from inside C and the probability densities are smooth in C; this leads to a minorization condition. In sections 3,4 and 5

we basically verify these two assumptions for a variety of SDEs so that the induced measures converge exponentially to the stationary distribution. Polynomial rates of convergence are obtained under a range of conditions in [32].

The remaining sections study the effect of time-discretization on the problems in sections 3, 4 and 5. Since the noise may be degenerate, and the vector fields not globally Lipschitz, existing theories establishing ergodicity of numerical methods [28, 29, 7] do not apply. We show that the geometric ergodicity theory of section 2 can be used to prove ergodicity of a variety of approximation methods applied to these problems. The key points to understand are how time-discretization affects a minorization condition and how it effects a Lyapunov structure. The former is robust to a wide range of approximations (see section 6), being a property on a compact set. The Lyapunov condition, however, is more sensitive: by constructing examples, we show that explicit methods such as Euler-Maruyama are transient, and hence not ergodic, for any choice of time-step; related examples may be found in [20, 30]. In section 7 we study globally Lipschitz diffusions where the Lyapunov structure *is* inherited by a wide range of approximations, including Euler-Maruyama. In section 8 we study locally Lipschitz diffusions, showing that certain implicit numerical methods can be constructed to inherit a Lyapunov structure; this work builds on related studies of deterministic problems (see [27, Chapters 4 and 5]). Some illustrative numerical experiments are described in Section 9. Detailed conclusions about numerical approximation are summarized at the start of Section 6.

2 Geometric Ergodicity

In this section we state Theorem 2.5, guaranteeing geometric ergodicity, which is sufficiently general to enable application to both a variety of SDEs (possibly with degenerate noise) and various time-discrete approximations. The proof is inspired and guided by those in [3, 16, 17] but is self-contained and tailored to our specific needs; it is given in the Appendix.

Consider a Markov process $x(t)$ ($t \in \mathbb{R}^+$) or a Markov chain $x(t)$ ($t \in \mathbb{Z}^+$) on a state space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Here $\mathcal{B}(\mathbb{R}^d)$ denotes the Börel σ-algebra on $\mathbb{R}^d$. To help combine our treatment of continuous and discrete time, we set $\mathbb{T} = \mathbb{R}^+$ (resp. $\mathbb{Z}^+$) for the Markov process (resp. chain) case. Throughout the remainder of the paper $\mathcal{B}_\delta(x)$ denotes the open ball of radius δ centered at x. We denote the transition kernel of the Markov process or chain by

$$P_t(x, A) \stackrel{\text{def}}{=} \mathbb{P}(x(t) \in A | x(0) = x), \quad t \in \mathbb{T}, x \in \mathbb{R}^d, A \in \mathcal{B}(\mathbb{R}^d) \ .$$

Assumption 2.1 *The Markov chain or process $\{x(t)\}$ with transition kernel $P_t(x, A)$ satisfies, for some fixed compact set $C \in \mathcal{B}(\mathbb{R}^d)$, the following:*

i) for some $y^* \in int(C)$ there is, for any $\delta > 0$, a $t_1 = t_1(\delta) \in \mathbb{T}$ such that

$$P_{t_1}\Big(x, \mathcal{B}_\delta(y^*)\Big) > 0 \quad \forall x \in C;$$

ii) for $t \in \mathbb{T}$ the transition kernel possesses a density $p_t(x, y)$, precisely

$$P_t(x, A) = \int_A p_t(x, y) dy \quad \forall x \in C, \ A \in \mathcal{B}(\mathbb{R}^d) \cap \mathcal{B}(C),$$

and $p_t(x, y)$ is jointly continuous in $(x, y) \in C \times C$.

Consider the Markov chain formed by sampling at the rate $T \in \mathbb{T}$, with the kernel $P(x, A) \stackrel{\text{def}}{=} P_T(x, A)$. Let $\{x_n\}_{n \in \mathbb{Z}^+}$ be the Markov chain generated by this kernel. We use a Lyapunov function to control the return times to C. In the following $\mathcal{F}_n$ denotes the σ-algebra of events up to and including the n^{th} iteration.

Assumption 2.2 *There is a function $V : \mathbb{R}^d \to [1, \infty)$, with $\lim_{x \to \infty} V(x) = \infty$, and real numbers $\alpha \in (0, 1)$, and $\beta \in [0, \infty)$ such that*

$$\mathbb{E}[V(x_{n+1}) | \mathcal{F}_n] \leq \alpha V(x_n) + \beta.$$

The basic conclusion of the next lemma, whose proof is given in the Appendix, is known as the *minorization condition.*

Lemma 2.3 *Let Assumption 2.1 hold. There is a choice of $T \in \mathbb{T}$, an $\eta > 0$, and a probability measure ν, with $\nu(C^c) = 0$ and $\nu(C) = 1$, such that*

$$P(x, A) \geq \eta\nu(A) \quad \forall A \in \mathcal{B}(\mathbb{R}^d), x \in C.$$

Throughout this paper we will study the following SDE and its approximations:

$$dx = Y(x)dt + \Sigma dW, \quad x(0) = y, \tag{2.1}$$

where $x \in \mathbb{R}^d$, $Y : \mathbb{R}^d \to \mathbb{R}^d$ and W is a standard m-dimensional Brownian motion for some $m \leq d$. The fixed matrix Σ is in $\mathbb{R}^{d\times m}$ and is assumed to have linearly independent columns. We use $\mathbb{E}^y$ to denote expectation under (2.1), with the given initial data. To establish geometric ergodicity for this SDE we may use the following assumption which implies Assumption 2.2.

Assumption 2.4 *There is a function $V : \mathbb{R}^d \to [1, \infty)$, with $\lim_{x\to\infty} V(x) = \infty$, and real numbers $a \in (0, \infty)$, $d \in (0, \infty)$ such that*

$$\mathcal{A}\{V(x)\} \leq -a\{V(x)\} + d, \tag{2.2}$$

where $\mathcal{A}$ is the generator for (2.1) given by

$$\mathcal{A}g = \sum_{i=1}^{d} Y_i \frac{\partial g}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} \left[\Sigma\Sigma^T\right]_{ij} \frac{\partial^2 g}{\partial x_i \partial x_j}. \tag{2.3}$$

This is just the infinitesimal version of Assumption 2.2. To see this note that, by the Itó formula,

$$dV = \mathcal{A}\{V\}dt + \text{Martingale}$$

so that, if $\mathcal{F}_s$ is the $\sigma-$algebra of all events up to time s, it follows that

$$\mathbb{E}^y\{V(x(t))|\mathcal{F}_s\} \leq e^{-a(t-s)}V(x(s)) + \frac{d}{a}[1 - e^{-a(t-s)}] \ . \tag{2.4}$$

If $x_n = x(nT)$, so that $\{x_n\}_{n=0}^{\infty}$ is a Markov chain, then (2.4) shows that Assumption 2.2 holds for this Markov chain: with $\alpha = e^{-aT}$ and $\beta = d/a$.

In what follows, we will use the shorthand notation $|f| \leq V$ to mean $|f(x)| \leq V(x)$ for all x.

Theorem 2.5 *Let $x(t)$ denote the Markov chain or process with transition kernel $P_t(x, A)$. Let $\{x_n\}_{n\in\mathbb{Z}^+}$ denote the embedded Markov chain with transition kernel $P(x, A) = P_T(x, A)$. There is a $T > 0$ for which the following holds. Let the Markov chain $\{x_n\}_{n\in\mathbb{Z}^+}$ satisfy the minorization condition (or Assumption 2.1) and Assumption 2.2 (or Assumption 2.4 when $x_n = x(nT)$ for (2.1)) with C given by*

$$C = \left\{x : V(x) \leq \frac{2\beta}{\gamma - \alpha}\right\} \tag{2.5}$$

for some $\gamma \in (\alpha^{\frac{1}{2}}, 1)$. Then there exists a unique invariant measure π. Furthermore there is $r(\gamma) \in (0, 1)$ and $\kappa(\gamma) \in (0, \infty)$ such that for all measurable $f : |f| \leq V$

$$|\mathbb{E}^{x_0} f(x_n) - \pi(f)| \leq \kappa r^n V(x_0).$$

3 The Langevin Equation

In this section we prove geometric ergodicity of the Langevin equation. The ergodicity of this equation is established in [31] by semigroup methods, but under somewhat restrictive hypotheses on the drift, requiring its first and second derivatives to be globally bounded, and on the class of functions whose expectations converge to values under the limit measure; furthermore, no rate of convergence is given. By employing the Markov chain techniques of section 2, we obtain geometric convergence, for a large class of test function, under considerably weaker conditions on the drift. This result is stated in Theorem 3.2. A concurrent recent study of a specific instance of this problem is taken up in [30].

Let W be a standard d-dimensional Brownian Motion, $F : \mathbb{R}^d \to \mathbb{R}$, $\sigma \in \mathbb{R}^{d\times d}$ and $\rho_i \in \mathbb{R}^d$ be the ith column of σ; we assume that the ρ_i are linearly independent so that σ is invertible. Consider the Langevin SDE for $q, p \in \mathbb{R}^d$ the position and momenta of a particle of unit mass, namely

$$dq = pdt, \tag{3.1}$$
$$dp = -\gamma p dt - \nabla F(q)dt + \sigma dW. \tag{3.2}$$

In the case $d = \sigma = 1$, for example, there is a known invariant measure with density

$$\rho(p, q) = \exp\{-\gamma[\frac{p^2}{2} + F(q)]\}.$$

We apply the theory of section 2 to prove ergodicity of (3.1)–(3.2) under the following condition:

Condition 3.1 *The function $F \in C^\infty(\mathbb{R}^d, \mathbb{R})$ and satisfies*

- *$F(q) \geq 0$ for all $q \in \mathbb{R}^d$.*
- *There exists an $\alpha > 0$ and $\beta \in (0, 1)$ such that*

$$\frac{1}{2}\langle \nabla F(q), q\rangle \geq \beta F(q) + \gamma^2 \frac{\beta(2-\beta)}{8(1-\beta)}||q||^2 - \alpha$$

A polynomial F growing at infinity like $\|q\|^{2l}$, with l a positive integer, will satisfy the assumptions; a simple example for expository purposes is

$$F(q) = \frac{1}{4}(\|q\|^2 - 1)^2 \ . \tag{3.3}$$

Under Condition 3.1 it is possible, using the Lyapunov function V below, to prove global in time existence and uniqueness of solutions to (3.1)–(3.2) – see Chapter III, Theorem 4.1 in [9]. It is expedient to write (3.1)–(3.2) in the abstract form (2.1) where now

$$x = \begin{pmatrix} q \\ p \end{pmatrix} \in \mathbb{R}^{2d},\ W = \begin{pmatrix} W_1 \\ \vdots \\ W_d \end{pmatrix} \in \mathbb{R}^d,\ Y(x) = \begin{pmatrix} p \\ -\gamma p - \nabla F(q) \end{pmatrix},\ \Sigma = \begin{pmatrix} O \\ \sigma \end{pmatrix}. \tag{3.4}$$

Here each W_i is an independent standard one-dimensional Brownian motion and $O \in \mathbb{R}^{d\times d}$ is the zero matrix. Note that we may write

$$\Sigma dW = \sum_{i=1}^{d} X_i dW_i,\ X_i = \begin{pmatrix} 0 \\ \rho_i \end{pmatrix},\ 0 \in \mathbb{R}^d \text{ and } \rho_i \in \mathbb{R}^d \ . \tag{3.5}$$

For (3.1)–(3.2), it is useful to define the Lyapunov function

$$V(x) \stackrel{\text{def}}{=} \frac{1}{2}||p||^2 + F(q) + \frac{\gamma}{2}\langle p, q\rangle + \frac{\gamma^2}{4}||q||^2 + 1 \tag{3.6}$$

with which we define

$$\mathcal{G}_l = \{\text{measurable } g : \mathbb{R}^{2d} \to \mathbb{R} \text{ with } |g| \leq V^l\}. \tag{3.7}$$

Theorem 3.2 *Let Condition 3.1 hold. Then the SDE* (3.1)–(3.2) *with* $x(t) = (q(t)^T, p(t)^T)^T$ *has a unique invariant measure* π *on* $\mathbb{R}^{2d}$. *Fix any* $l \geq 1$. *If* $x(0) = y$ *then there exists* $C = C(l) > 0$, $\lambda = \lambda(l) > 0$ *such that, for all* $g \in \mathcal{G}_l$,

$$\left|\mathbb{E}^y g(x(t)) - \pi(g)\right| \leq CV(y)^l e^{-\lambda t} \quad \text{for all } t \geq 0. \tag{3.8}$$

Proof The result follows from an application of Theorem 2.5. First note that

$$V(x) \geq 1 + \frac{1}{8}||p||^2 + \frac{\gamma^2}{12}||q||^2, \tag{3.9}$$

using Condition 3.1(i). Thus $V(x)^l \to \infty$ as $||x|| \to \infty$. Lemma 3.3 shows that if $\mathcal{A}$ is the generator of the process governed by (3.1)–(3.2), that is,

$$\mathcal{A}g = \sum_{i=1}^{2d} Y_i \frac{\partial g}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{2d} \left[\Sigma\Sigma^T\right]_{ij} \frac{\partial^2 g}{\partial x_i \partial x_j} \tag{3.10}$$

then

$$\mathcal{A}\{V(x)^l\} \leq -a_l\{V(x)^l\} + d_l$$

for some $a_l, d_l > 0$. Thus, by the discussion at the end of section 2, Assumption 2.2 holds for the time T sampled chain $x_n = x(nT)$.

To verify Assumption 2.1(ii), we define

$$\mathcal{L} = Lie\{Y, X_1, \ldots, X_d\},$$

namely the Lie algebra generated by $\{Y, X_1, \ldots, X_d\}$. Let $\mathcal{L}_0$ be the ideal in $\mathcal{L}$ generated by $\{X_1, \ldots, X_d\}$. By results in [2, 18, 13], it suffices to show that $\mathcal{L}_0$ spans $\mathbb{R}^{2d}$ to verify Assumption 2.1(ii). Note that

$$X_i = \begin{pmatrix} 0 \\ \rho_i \end{pmatrix} \quad \& \quad [X_i, Y] = DYX_i = \begin{pmatrix} 0 & I \\ -d^2F(q) & -\gamma I \end{pmatrix}\begin{pmatrix} 0 \\ \rho_i \end{pmatrix} = \begin{pmatrix} \rho_i \\ -\gamma\rho_i \end{pmatrix}$$

Thus, since σ has linearly independent columns $\{\rho_i\}_{i=1}^d$,

$$\{X_1, \ldots, X_d, [X_1, Y], \ldots, [X_d, Y]\}$$

span $\mathbb{R}^{2d}$ as required.

Lemma 3.4, found at the end of this section, proves that, in any positive time, any open set may be reached with positive probability. Thus Assumption 2.1(i) holds with any choice of C and y^*. Hence Theorem 2.5 shows that for some $r \in (0,1)$ and $\kappa > 0$

$$\left|\mathbb{E}^y g(x_n) - \pi(g)\right| \leq \kappa r^n V^l(y),$$

where $x_n = x(nT)$, any $T > 0$. To complete the proof we use an argument from [17]. Let $t_n = nT + \delta$, $\delta \in [0, T)$. Then, by conditioning on $\mathcal{F}_\delta$,

$$\left|\mathbb{E}^y g(x(t_n)) - \pi(g)\right| = \left|\mathbb{E}^y g\big(x(nT+\delta)\big) - \pi(g)\right| \leq \kappa r^n \mathbb{E}^y V(x(\delta))^l.$$

Applying (2.4) gives

$$\left|\mathbb{E}^y g\big(x(t_n)\big) - \pi(g)\right| \leq \kappa r^n \left[e^{-a_l\delta}V(y)^l + \frac{d_l}{a_l}\right];$$

defining λ by $e^{-\lambda} = r^{1/T}$ we obtain the required result by re-defining $\kappa \to \kappa(1 + d_l/a_l)e^{\lambda T}$:

$$\left|\mathbb{E}^y g(x(t_n)) - \pi(g)\right| \leq \kappa e^{-\lambda t_n}\left[1 + V(y)^l\right]$$

where $e^{-\lambda} = r^{1/T}$. The proof is complete. □

The previous theorem rested on two lemmas which we now establish.

Lemma 3.3 *Let Condition 3.1 hold. For every $l \geq 1$, there exists $a_l \in (0,\infty)$ and $d_l \in (0,\infty)$ such that, for equation* (3.1)–(3.2) *with $\mathcal{A}$ given by* (3.10)*,*

$$\mathcal{A}\{V(x)^l\} \leq -a_l\{V(x)^l\} + d_l \ .$$

Proof We do the case $l = 1$ first. Let

$$\begin{aligned}
Y_i(x) &= p_i, && i = 1, \cdots, d \\
Y_i(x) &= -\gamma p_i - \frac{\partial F}{\partial q_i}(q), && i = d+1, ..., 2d \\
\frac{\partial V}{\partial x_i} &= \frac{\partial F}{\partial q_i}(q) + \frac{\gamma}{2}p_i + \frac{\gamma^2}{2}q_i, && i = 1, \ldots, d \\
\frac{\partial V}{\partial x_i} &= p_i + \frac{\gamma}{2}q_i && i = d+1, \ldots, d
\end{aligned}$$

The following inequality, proved in Lemma 2.2 of [23] as a consequence of Condition 3.1, will be useful to us:

$$-\frac{1}{2}\|p\|^2 - \frac{1}{2}\langle \nabla F(q), q\rangle \leq \alpha - \beta[V(x) - 1]. \tag{3.11}$$

Using (3.11) to bound the inner-product we obtain

$$\begin{aligned}
\sum_{i=1}^{2d} Y_i \frac{\partial V}{\partial x_i} =& \langle p, \nabla F(q)\rangle + \frac{\gamma}{2}\|p\|^2 + \frac{\gamma^2}{2}\langle p, q\rangle \\
&- \gamma||p||^2 - \langle p, \nabla F(q)\rangle - \frac{\gamma^2}{2}\langle p, q\rangle - \frac{\gamma}{2}\langle q, \nabla F(q)\rangle \\
=& -\frac{\gamma}{2}||p||^2 - \frac{\gamma}{2}\langle q, \nabla F(q)\rangle \\
\leq& \ \gamma\big[\alpha - \beta(V-1)\big] \\
=& \ \gamma[\alpha + \beta] - \gamma\beta V \ .
\end{aligned}$$

Also,

$$\Sigma\Sigma^T = \begin{pmatrix} 0 & 0 \\ 0 & \sigma\sigma^T \end{pmatrix}$$

and thus

$$\sum_{i,j=1}^{2d} [\Sigma\Sigma^T]_{ij} \frac{\partial^2 V}{\partial x_i \partial x_j} = \sum_{i,j=1}^{d} [\sigma\sigma^T]_{ij} \frac{\partial^2 V}{\partial x_{d+i}\partial x_{d+j}} = \sum_{i=1}^{d} [\sigma\sigma^T]_{ii} \frac{\partial^2 V}{\partial p_i^2} \ .$$

But

$$\frac{\partial^2 V}{\partial p_i^2} = 1 \quad \& \quad \frac{1}{2}\sum_{i=1}^{d}[\sigma\sigma^T]_{ii} = \frac{1}{2}\sum_{i,j=1}^{d}\sigma_{ij}^2 = \frac{1}{2}\|\sigma\|_F^2 \stackrel{\text{def}}{=} \mathcal{E},$$

where $\|\cdot\|_F$ is the Frobenius norm on matrices. Combining, we have

$$\mathcal{A}V(x) \leq \gamma[\alpha + \beta + \frac{\mathcal{E}}{\gamma} - \beta V],$$

as required. Now we calculate $\mathcal{A}\{V(x)^l\}$. To this end, note that

$$\frac{\partial}{\partial x_i}\left\{V(x)^l\right\} = l\{V(x)\}^{l-1}\frac{\partial V}{\partial x_i}$$
$$\frac{\partial^2}{\partial x_i \partial x_j}\left\{V(x)^l\right\} = \frac{\partial}{\partial x_j}\left\{l\{V(x)\}^{l-1}\frac{\partial V}{\partial x_i}\right\},$$
$$\text{and } \mathcal{A}\{V(x)^l\} = l\{V(x)\}^{l-1}\mathcal{A}V + \frac{1}{2}\sum_{i,j=1}^{d}[\sigma\sigma^T]_{ij}l(l-1)V(x)^{l-2}\frac{\partial V}{\partial p_i}\frac{\partial V}{\partial p_j}.$$

But

$$\frac{\partial V}{\partial p_i} = p_i + \frac{\gamma}{2}q_i$$

and hence, by using (3.9), we obtain

$$\frac{1}{2}l(l-1)\sum_{i,j=1}^{d}[\sigma\sigma^T]_{ij}\frac{\partial V}{\partial p_i}\frac{\partial V}{\partial p_j} \le \chi V(x)$$

for some $\chi > 0$. Thus

$$\mathcal{A}V(x)^l \le lV(x)^{l-1}\mathcal{A}V(x) + \chi V(x)^{l-1}\ .$$

By the calculation for $l = 1$,

$$\begin{aligned}\mathcal{A}V(x)^l &\le lV(x)^{l-1}[d - aV(x)] + \chi V(x)^{l-1}\\ &= -alV(x)^l + (ld + \chi)V(x)^{l-1}.\end{aligned}$$

By choosing $a_l < al$ and d_l sufficiently large we obtain

$$\mathcal{A}V(x)^l \le -a_l V(x)^l + d_l$$

as required. □

Lemma 3.4 *Let Condition 3.1 hold. For all $x \in \mathbb{R}^{2d}, t > 0$ and open $\mathcal{O} \subset \mathbb{R}^{2d}$, the transition kernel for (3.1)–(3.2) satisfies $P_t(x, \mathcal{O}) > 0$.*

Proof It suffices to consider the probability of hitting an open ball of radius δ, $\mathcal{B}_\delta$, centered at y^+. Consider the associated control problem, derived from (3.1)–(3.2),

$$\frac{dX}{dt} = Y(X) + \Sigma\frac{dU}{dt}\ . \tag{3.12}$$

For any $t > 0$, any $y \in \mathbb{R}^{2d}$, and any $y^+ \in \mathbb{R}^{2d}$, we can find smooth $U \in C^1([0,t], \mathbb{R}^d)$ such that (3.12) is satisfied and $x(0) = y$, $x(t) = y^+$. To see this set $x = (Q^T, \frac{dQ}{dt}^T)^T$ and note that

$$\frac{d^2Q}{dt^2} + \gamma\frac{dQ}{dt} + \nabla F(Q) = \sigma\frac{dU}{dt}\ .$$

Choose Q to be a C^∞ path such that, for the given $t > 0$,

$$\begin{pmatrix} Q(0) \\ \frac{dQ}{dt}(0)\end{pmatrix} = y, \quad \begin{pmatrix} Q(t) \\ \frac{dQ}{dt}(t)\end{pmatrix} = y^+.$$

This can be achieved, for example, by polynomial interpolation between the end points, using a cubic in time with vector coefficients in $\mathbb{R}^d$. Since σ is invertible, $\frac{dU}{dt}$ is defined by substitution and will be as smooth as ∇F – hence C^∞. Also $U(0)$ can be taken as 0.

Now

$$x(t) = y + \int_0^t Y(x(s))ds + \Sigma W(t)$$
$$X(t) = y + \int_0^t Y(X(s))ds + \Sigma U(t) \ .$$

Note that the event

$$\sup_{0\le s\le t} ||W(t) - U(t)|| \le \epsilon$$

occurs with positive probability for any $\epsilon > 0$, since the Weiner measure of any such tube is positive (Theorem 4.20 of [26]). Assuming this event occurs, note that

$$||x(t) - X(t)|| \le \int_0^t ||Y(x(s)) - Y(X(s))||ds + ||\Sigma||\epsilon \ .$$

Since F is locally Lipschitz so is Y and thus it follows that

$$\sup_{0\le t\le T} ||x(t) - X(t)|| \to 0 \text{ as } \epsilon \to 0 \ .$$

By choice of ϵ, we can hence ensure $||x(t) - X(t)|| \le \delta$ and the result follows. □

4 Monotone and Dissipative Problems

We now consider the SDE (2.1) where again $x \in \mathbb{R}^d, W \in \mathbb{R}^m, Y : \mathbb{R}^d \to \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d\times m}, m \le d$. The columns of Σ, $\{\rho_j\}_{j=1}^m$ are assumed to be linearly independent. We prove geometric ergodicity in Theorem 4.4 under the following dissipativity condition concerning the deterministic flow.

Condition 4.1 *The function $Y \in C^\infty(\mathbb{R}^d, \mathbb{R}^d)$ and $\exists \alpha, \beta > 0$:*

$$\langle Y(x), x\rangle \le \alpha - \beta\|x\|^2.$$

This condition means that, sufficiently far from the origin, the Markov process defined by (2.1) moves inward on average. For the deterministic counterpart ($\Sigma \equiv 0$), it implies dissipativity in the sense of [8]. If $m = d$ the work of [9] implies ergodicity under this condition. Below we add a further condition ((4.3)) which will enable us to establish ergodicity even when $m < d$. Calculations analogous to those in Lemma 3.3 enable proof of the following:

Lemma 4.2 *Let Condition 4.1 hold. For every $l \ge 1$ there exists $a_l \in (0, \infty)$ and $d_l \in (0, \infty)$ such that, for equation* (2.1) *with $\mathcal{A}$ given by* (2.3)*,*

$$\mathcal{A}\{\|x\|^{2l}\} \le -a_l\{\|x\|^{2l}\} + d_l.$$

From (2.4) it follows that

$$\mathbb{E}\{\|x(t)\|^{2l}|\mathcal{F}_s\} \le e^{-a_l(t-s)}\|x(s)\|^{2l} + \frac{d_l}{a_l}[1 - e^{-a_l(t-s)}]$$

so that

$$\mathbb{E}\{[1 + \|x(t)\|^{2l}]|\mathcal{F}_s\} \le e^{-a_l(t-s)}[1 + \|x(s)\|^{2l}] + \frac{d_l + a_l}{a_l}[1 - e^{-a_l(t-s)}]. \tag{4.1}$$

Theorem 3.6 in [14, Chapter 2] establishes global existence and uniqueness for (2.1), under Condition 4.1. By the discussion at the end of section 2 we deduce that Assumption 2.2 holds for the time T sampled SDE.

We now make assumptions that, when combined with the Lyapunov structure (4.1), will induce ergodicity. The assumptions are stated entirely in terms of the dynamics of the deterministic counterpart of (2.1) ($\Sigma \equiv 0$) and the vectors ρ_i forming the columns of Σ. Under Condition 4.1 equation (2.1) without noise must have at least one equilibrium point. Without loss of generality, we place this at the origin. We let $\phi(\cdot, t)$ denote the deterministic flow for (2.1) with $\Sigma = 0$ and denote by $\mathcal{S}$ the stable manifold of 0. The next condition encodes the basic idea that, if by a combination of alternating pure noise and pure deterministic flow we can reach $\mathcal{S}$, it will be possible to satisfy Assumption 2.1(i).

Condition 4.3 *For some fixed $R, T_1 > 0$ the following holds: given any $\delta > 0$ and any $x \in \mathcal{B}_R(0)$ there exists an integer N, and a sequence of non-negative τ_i with $\sum_{i=1}^N \tau_i < T_1$, and $\{a_{i,j}\}_{i,j=1}^{N,m}$ with $a_{i,j} \in \mathbb{R}$, so that $\phi(z_N(x), t) \in \mathcal{B}_\delta(0)$ for all $t \in [0, T_1]$. Here $z_N(x)$ is defined by*

$$z_0 = x$$
$$z_{n+1} = \phi(z_n, \tau_{n+1}) + \sum_{j=1}^m a_{n+1,j}\rho_j, \quad n = 0, \ldots, N-1.$$

Define

$$\mathcal{G}_l = \{\text{measurable } g : \mathbb{R}^d \to \mathbb{R} \text{ with } |g(x)| \le 1 + \|x\|^{2l}\}.$$

Let $T = T_1$ and define, for $\gamma_l \in (\alpha_l^{\frac{1}{2}}, 1)$,

$$\beta_l = \frac{d_l + a_l}{a_l}, \quad \alpha_l = e^{-a_l T}, \quad r_l = \frac{2\beta_l}{\gamma_l - \alpha_l}.$$

Theorem 4.4 *Let Conditions 4.1, 4.3 hold with R chosen so that $\{x : 1 + \|x\|^{2l} \le r_l\} \subseteq \mathcal{B}_R(0)$, some $l > 0$. If the transition kernel for (2.1) has density $p_t(x, y)$ which is jointly continuous in (x, y) for every fixed $t > 0$ then (2.1) has a unique invariant measure π and, if $x(0) = y$ then there exists $\kappa = \kappa(l) > 0$ and $\lambda = \lambda(l) > 0$ such that, for all $g \in \mathcal{G}_l$,*

$$\left|\mathbb{E}^y g(x(t)) - \pi(g)\right| \le \kappa[1 + \|y\|^{2l}]e^{-\lambda t} \quad \text{for all } t \ge 0. \tag{4.2}$$

Proof We use Theorem 2.5. Lemma 4.2 (equation (4.1)) implies that Assumption 2.2 holds with $V(x) = 1 + \|x\|^{2l}$. We have assumed Assumption 2.1(ii), so it remains to show part (i) of that assumption. Define $T_0 = \sum_{i=1}^N \tau_i$ noting that $T_0 < T_1$. The set C is $\{x : V(x) \le r_l\} \subseteq \mathcal{B}_R(0)$. Now for any $\delta_1 < T_1 - T_0$ define $U(t)$ by, for $l = 1, \ldots, N$,

$$U(t) = \begin{cases} 0 & t \in I_l^- \stackrel{\text{def}}{=} [t_{l-1}, t_{l-1} + \tau_l) \\ \dfrac{N}{\delta_1}\displaystyle\sum_{j=1}^m a_{l,j}\rho_j & t \in I_l^+ \stackrel{\text{def}}{=} [t_{l-1} + \tau_l, t_l) \\ 0 & t \in [t_N, T_1] \end{cases}$$

where

$$t_l = \frac{l\delta_1}{N} + \sum_{j=1}^l \tau_j, \quad |I_l^-| = \tau_l, \quad |I_l^+| = \frac{\delta_1}{N}.$$

Notice that by construction $t_N \le T_0 + \delta_1$ and hence $t_N < T_1$. If

$$X(t) = x + \int_0^t Y(X(s))ds + U(t)$$

then by choosing δ_1 sufficiently small, so that the effect of U dominates the drift Y for $t \in I_l^+$, we have that $X(t) \in \mathcal{B}_{3\delta/2}(0)$ for $t \in [t_N, t_N + T_1]$ and any initial $x \in C$. Since $t_N < T_1 \le T_1 + t_N$ we have that $X(T_1) \in \mathcal{B}_{3\delta/2}(0)$ for any initial $x \in C$.

By continuity, there is some tube about $U(t)$ so that the system forced by a Brownian motion in that tube will have $X(T_1) \in \mathcal{B}_{2\delta}(0)$. Since the Weiner measure of any such tube is positive (Theorem 4.20 of [26]) Assumption 2.1(i) is proven.

Thus we have proved geometric ergodicity for the Markov chain found by sampling the SDE at rate $T = T_1$. To obtain convergence for the continuous time Markov process from that of the embedded chain we proceed as in Theorem 3.2 for the Langevin equation. □

Remark It is worth noting that a similar theorem could be proved whenever, in the absence of noise, one has some globally attracting compact structure. Here we consider the simplest case when the structure is a point. Similar ideas can be used when, for example, there is an attracting periodic orbit in the deterministic flow, such as in the Van Der Pol oscillator. □

Example If $\phi(\cdot, t)$ is exponentially monotone so that, for some $c > 0$

$$\langle Y(a), a\rangle \leq -c\|a\|^2 \quad \forall a \in \mathbb{R}^d,$$

then Condition 4.1 holds with $\alpha = 0$ and $\beta = c$. Also

$$\|\phi(x,t)\| \leq e^{-ct}\|x\|$$

and so Condition 4.3 holds for any $0 < \delta < R$ with $N = 1$, $\tau_1 = \frac{1}{c}\ln(R/\delta)$ and $a_{1,j} \equiv 0$. Any $T_1 > \tau_1$ can be used. This is independent of the form of the noise which can hence be degenerate, provided the underlying smooth density assumption can be satisfied. As a specific instance of this example consider the problem

$$\begin{aligned} dy &= [-y + yz]dt + dw. \\ dz &= [-z - y^2]dt. \end{aligned}$$

Here $\rho_1 = (1,0)^T$, $[[Y,\rho_1],\rho_1] = (0,-2)^T$ and so smoothness follows from [13], recalling the definition and significance of $\mathcal{L}_0$ from section 3. Geometric ergodicity follows from Theorem 4.4. This approach is used to establish the ergodicity of arbitrary Galerkin approximations of the Navier Stokes equations (at any Reynolds number) in [4]. □

Example Consider a problem in the form

$$\begin{aligned} dv &= a(v,z)dt + \sigma dw. \\ dz &= [-bz + c(v,z)]dt, \end{aligned}$$

where $b > 0$. We assume that, for each t, $v \in \mathbb{R}^d, w \in \mathbb{R}^d$ and $z \in \mathbb{R}$, whilst $\sigma \in \mathbb{R}^{d\times d}$ is invertible; we also assume that $(v,z) = (0,0)$ is an equilibrium point of the deterministic flow $\phi(\cdot,t)$ ($\sigma \equiv 0$) and that $a(0,z) \equiv 0$ and $c(0,z) \equiv 0$. Clearly $v \equiv 0$ is part of the stable manifold of $(0,0)$. To establish Condition 4.3 the idea is to choose noise to move onto the stable manifold and then flow to the origin without noise. As σ is invertible then Condition 4.3 can be realized with $N = 2, \tau_1 = 0$ and $\tau_2 = \frac{1}{b}\ln(R/\delta)$; there exists $p \in \mathbb{R}^d$ such that $\sigma p = -v(0)$ and then $(a_{1,1}, a_{1,2})^T = p$ whilst $(a_{2,1}, a_{2,2})^T = 0$. Any $T_1 > \tau_2$ can be used.

This approach applies to the Lorenz equations

$$\begin{aligned} dx &= [\sigma(y-x)]dt + dW_1 \\ dy &= [rx - y - xz]dt + dW_2 \\ dz &= [xy - bz]dt \end{aligned} \tag{4.3}$$

with $v = (x,y)$. Condition 4.1 holds here for a range of parameter values (including those where chaos is observed, see [25]) and (recalling the definition and significance of $\mathcal{L}_0$ from section 3), for $\rho_1 = (1,0,0)^T$ and $\rho_2 = (0,1,0)$ we have $[[Y,\rho_2],\rho_1] = (0,0,1)^T$ so that the density is smooth [13]. Hence the equations are geometrically ergodic by Theorem 4.4. □

5 Gradient Systems

In this section we study equation (2.1) in the case where $Y(x) = -\nabla F(x)$ and is hence a gradient. Specifically we consider the problem

$$dx = -\nabla F(x)dt + \Sigma dW, \quad x(0) = x_0, \tag{5.1}$$

where $x \in \mathbb{R}^d, W \in \mathbb{R}^m, F : \mathbb{R}^d \to \mathbb{R}, \Sigma \in \mathbb{R}^{d\times m}$ and $m \le d$. The columns of Σ are assumed to be linearly independent. We also define $B = \Sigma\Sigma^T \in \mathbb{R}^{d\times d}$. This problem is studied in [20] by use of the theory of geometrically ergodic Markov chains, as we do in this paper. In that work $m = d$ (non-degenerate noise) and the Lyapunov function used is $V(x) = \exp\{\zeta F(x)\}$ for some $\zeta \in (0, 1)$. Here we allow degenerate noise and use $V(x) = 1 + F(x)^l$ leading to weaker dependence of the time to equilibrium on initial data than in [20], but also to correspondingly smaller classes of allowable test function; however in some cases the overall bounds may lead to improved estimates of the necessary time to approximate a stationary distribution. We make the following conditions concerning F :

Condition 5.1 *The function F satisfies:*

- $F \in C^\infty(\mathbb{R}^d, \mathbb{R}), F \ge 0, F(a) \to \infty$ *as* $|a| \to \infty$.
- *For all $l > 0$ there are $\alpha_l' > 0, \beta_l' > 0$ with*

$$|\nabla F(a)|^2 + \alpha_l' \ge \frac{1}{2}B : \partial^2 F(a) + \frac{(l-1)}{2F(a)}(\nabla F(a)\nabla F(a)^T) : B + \beta_l' F(a).$$

In the preceding, the colon denotes the inner-product on matrices which induces the Frobenius norm. The conditions are satisfied if, for example, $F(x)$ is smooth, bounded below and, as $\|x\| \to \infty$, grows as follows, for some integer $p \ge 1$:

$$\begin{aligned} F(x) &\sim \|x\|^{2p} \\ \nabla F(x) &\sim \|x\|^{2p-1} \qquad \text{as } \|x\| \to \infty. \\ \partial^2 F(x) &\sim \|x\|^{2p-2} \end{aligned}$$

The next lemma is the key result that follows from these conditions.

Lemma 5.2 *Let Condition 5.1 hold. Then, for equation* (5.1) *with $\mathcal{A}$ given by* (2.3)*,*

$$\mathcal{A}\{F(x)^l\} \le l\{\alpha_l - \beta_l F(x)^l\}$$

for all $l > 0$. Thus, for any $l > 0$,

$$\mathbb{E}\{F(x(t))^l | \mathcal{F}_s\} \le \frac{\alpha_l}{\beta_l}[1 - e^{-l\beta_l(t-s)}] + e^{-l\beta_l(t-s)}F(x(0))^l.$$

Proof Straightforward calculation shows that

$$\begin{aligned} \mathcal{A}\{F(x)^l\} &= \textstyle\sum_{i=1}^d lF(x)^{l-1}\frac{\partial F}{\partial x_i}[-\frac{\partial F}{\partial x_i}] \\ &+ \textstyle\frac{1}{2}\sum_{i,j=1}^d \{lF(x)^{l-1}\frac{\partial^2 F}{\partial x_i \partial x_j}B_{ij} \quad + \quad l(l-1)F^{l-2}\frac{\partial F}{\partial x_i}\frac{\partial F}{\partial x_j}B_{ij}\}. \end{aligned}$$

The first result follows, by use of Condition 5.1(ii) and the fact that

$$\alpha' x^{l-1} - \beta' x^l \le \alpha - \beta x^l \quad \forall x \ge 0$$

for suitably chosen α, β. The second follows from the discussion at the end of section 2. □

As in the previous section, we now make Condition 4.3 which, when combined with the Lyapunov structure of Lemma 5.2, will induce ergodicity. Also we define

$$\mathcal{G}_l = \{\text{measurable } g : \mathbb{R}^{2d} \to \mathbb{R} \text{ with } |g(x)| \le 1 + F(x)^l\}.$$

The following theorem may be proved in exactly the same way that Theorem 4.4 is proved, with r_l as defined there:

Theorem 5.3 *Let Conditions 5.1, 4.3 hold with R chosen so that $\{x : 1 + F(x)^l \le r_l\} \subseteq \mathcal{B}_R(0)$ for some $l > 0$. If the transition kernel for (5.1) has density $p_t(x, y)$ which is jointly continuous in (x, y) for every fixed $t > 0$ then (5.1) has a unique invariant measure π and, if $x(0) = y$ then there exists $\kappa = \kappa(l) > 0$ and $\lambda = \lambda(l) > 0$ such that, for all $g \in \mathcal{G}_l$,*

$$\left|\mathbb{E}^y g(x(t)) - \pi(g)\right| \le \kappa[1 + F(y)^l]e^{-\lambda t} \quad \text{for all } t \ge 0. \tag{5.2}$$

Example Let $d = 2, x = (u, v)$ and

$$F(x) = \frac{1}{4}(1 - u^2)^2 + \frac{v^2}{2}.$$

Clearly Condition 5.1(i) is satisfied and a little calculation reveals that Condition 5.1(ii) can be satisfied. Thus, to apply Theorem 5.3, it remains to check the smoothness condition on the transition kernel together with Condition 4.3. It is useful here to recall the definition of the ideal $\mathcal{L}_0(x)$ from section 3 and that existence and smoothness of the transition kernel density $p_t(x, y)$ follows if $\mathcal{L}_0$ has full rank at all points. In this example this condition requires that $\text{rank}(\mathcal{L}_0(x)) = 2$ for all x. If $m = 1$ and $\Sigma = (a, b)^T$ then three cases arise:

- $a = 0, b = 1$. In this case, the smoothness fails because $\mathcal{L}_0$ is spanned by $(0, 1)^T$ as is apparent from the fact that no noise enters the u equation. Furthermore there is no point reachable from the whole space $\mathbb{R}^2$ with positive probability: $u(0) > 1$ (resp. < -1) implies $u(t) > 1$ (resp. < -1) with probability one. Hence the problem cannot be ergodic on $\mathbb{R}^2$.
- $a = 1, b = 0$. In this case, the smoothness fails because $\mathcal{L}_0$ is spanned by $(1, 0)^T$ as is apparent from the fact that no noise enters the v equation. However, Condition 4.3 can be satisfied here, using $(u, v) = (1, 0)$ in place of the origin, taking $N = 1$ and then $\tau_1 = 0$. The problem is ergodic, but the invariant measure is singular in v and so Theorem 5.3 does not apply.
- $a \ne 0$ and $b \ne 0$. The same argument as in the previous case shows that Condition 4.3 holds. Smoothness of $p_t(x, y)$ is also satisfied since $\{\Sigma, [Y, \Sigma], [[Y, \Sigma], \Sigma]\}$ spans $\mathbb{R}^2$ at each $x = (u, v)$. Hence the problem is geometrically ergodic in this case. □

6 Time Discretization

6.1 Introduction

Our primary objective in this, and subsequent sections, is to study the ergodic properties of discretizations of the SDE (2.1). Recall that the case of degenerate noise, $m < d$, is allowed. We will make a variety of assumptions about Y and Σ that imply the geometric ergodicity of the SDE. Our aim is then to study whether discretizations provided by numerical methods have an analogous property. We study an abstract family of approximations and focus on three specific numerical methods. With the notation

$$\Delta W_n \stackrel{\text{def}}{=} W(t_{n+1}) - W(t_n),\ t_n = n\Delta t,\ X_n \approx x(t_n) \tag{6.1}$$

we see that the ΔW_n form an i.i.d. family distributed as $\sqrt{\Delta t}\mathcal{N}(0, I)$, with I the $m \times m$ identity.

The first scheme, known as the forward Euler or *Euler–Maruyama scheme*, is as follows [12]:

$$X_{n+1} = X_n + \Delta t Y(X_n) + \Sigma \Delta W_n. \tag{6.2}$$

The second, which we call the *stochastic Backward Euler method*, is

$$X_{n+1} = X_n + \Delta t Y(X_{n+1}) + \Sigma \Delta W_n. \tag{6.3}$$

The third, which we call the *split-step stochastic Backward Euler method*, is

$$\begin{aligned} X_\star &= X_n + \Delta t Y(X_\star), \\ X_{n+1} &= X_\star + \Sigma \Delta W_n. \end{aligned} \tag{6.4}$$

Note that (6.2) is an explicit method, whereas (6.3) and (6.4) are implicit, requiring the solution of a (generally) nonlinear equation at each step. All the methods are examples of the general family

$$X_{n+1} = H(X_n, \Delta W_n), \quad X_0 = y; \tag{6.5}$$

solvability of the implicit equations must be established for the implicit methods to be written in this form.

We briefly summarize our results for the numerical approximation of (2.1).

- The two keys points in our approach to proving ergodicity are the existence of a *minorization condition* on a compact set C, together with a *Lyapunov function* inducing repeated returns into C. The minorization condition tends to persist for all reasonable approximations, relying on properties on a compact set (established later in this section). The Lyapunov condition, since it is a property on non-compact sets, is more sensitive to the choice of discretization and is inherited only by specially constructed methods.

- If the vector field Y is not globally Lipschitz then the Euler–Maruyama scheme does not preserve ergodicity, in general. We give an example of an ergodic SDE whose Euler–Maruyama approximation tends to infinity for any $\Delta t > 0$, with positive probability (established later in this section).

- If the vector field Y is globally Lipschitz and the Lyapunov function is essentially quadratic (a term defined below) then any reasonable method, including (6.2), (6.3) and (6.4), will inherit ergodicity of (2.1) for time-steps below a sufficiently small level that is independent of initial data. The key point is that all reasonable methods inherit the Lyapunov structure under these conditions on the vector field and the Lyapunov function for the SDE (section 7).

- Under a variety of natural structural assumptions, for which Y is not necessarily globally Lipschitz, one or other of the stochastic backward Euler methods may be proved ergodic, for sufficiently small time-step independent of initial data. The key point is to find methods which replicate the Lyapunov structure (section 8).

- In many cases where the numerical method is ergodic, the invariant measure for the method converges to that for (2.1), in a metric closely related to that induced by a (Lyapunov-function) weighted total variation norm, as the time-step converges to zero (sections 7 and 8).

Related issues have arisen in [20] where it was shown that the Euler-Maruyama scheme can be transient when applied to geometrically ergodic SDEs as we do here. However, rather than studying the use of implicit schemes to overcome this, the authors studied "Metropolis-adjusted" algorithms based on the Euler-Maruyama method with possible step rejection.

Our basic tool for proving ergodicity is Theorem 2.5. Its proof relies on two key facts concerning a Markov chain $\{x_n\}_{n\in\mathbb{Z}^+}$ with transition kernel $P(x, A)$. Assumption 2.1 implies (see Lemma 2.3) the *minorization condition*: $\exists \eta > 0$ and a probability measure ν with $\nu(C) = 1 - \nu(C^c) = 1$ satisfying

$$P(x, A) \geq \eta\nu(A) \quad \forall A \in \mathcal{B}(\mathbb{R}^d), x \in C. \tag{6.6}$$

Assumption 2.2 is the *Lyapunov condition*. Thus understanding the effect of numerical approximation of an SDE on ergodicity boils down, in this context, to understanding how the minorization and Lyapunov conditions are affected by approximation. We will see that the former is rather insensitive, since it is a property on a compact set C, whilst the latter can be destroyed unless special discretizations are employed, since it is a property on the whole space.

6.2 The Minorization Condition

Let $\{x(t)\}_{t\in\mathbb{R}^+}$ be a Markov process generated by (2.1) and let $\{X_n\}_{n\in\mathbb{Z}^+}$ be a strong approximation generated by the numerical method (6.5), constructed so that $X_n \approx x(n\Delta t)$. Define

$$\begin{aligned} P_t(x, A) &\stackrel{\text{def}}{=} \mathbb{P}(x(t) \in A | x(0) = x) \\ P_{n,\Delta t}(x, A) &\stackrel{\text{def}}{=} \mathbb{P}(X_n \in A | X_0 = x). \end{aligned}$$

The following condition holds for a wide variety of numerical methods, including those of interest to us, when applied to a wide variety of SDEs. In particular the convergence criterion is a consequence of standard strong convergence results, which are uniform across compact sets of initial data.

Condition 6.1 *Fix $n\Delta t = t$ and then the following hold for all Δt sufficiently small. For any open set $\mathcal{O}$ and compact C*

$$\sup_{x\in C} |P_{n,\Delta t}(x,\mathcal{O}) - P_t(x,\mathcal{O})| \to 0$$

as $\Delta t \to 0$. Furthermore, for $n \geq n_0$, $P_{n,\Delta t}$ has a density $p_{n,\Delta t}$, so that

$$P_{n,\Delta t}(x,A) = \int_A p_{n,\Delta t}(x,y)dy,$$

and $p_{n,\Delta t}(x,y)$ is differentiable in (x,y) with derivative bounded independently of Δt sufficiently small, for $n\Delta t$ fixed.

Theorem 6.2 *Let Assumption 2.1 hold for $x(t)$ solving (2.1) and assume, in addition, that the density $p_t(x,y)$ is jointly continuous in $(t,x,y) \in \mathbb{T}\times C\times C$. Assume also that Condition 6.1 holds. Then there is a choice of $M \in \mathbb{Z}^+$ such that the minorization condition holds for the chain $\{X_{nM}\}_{n\in\mathbb{Z}^+}$ generated by the numerical method (6.5).*

Proof Assumption 2.1(i), together with continuity of the density in t, implies that

$$P_t(y^*, \mathcal{B}_{\delta'}(y^*)) \geq \gamma > 0 \quad \forall t \in [t_2-\Delta, t_2+\Delta].$$

Hence, for all $t \in [t_2-\Delta, t_2+\Delta]$, there is a $z^* \in \mathcal{B}_{\delta'}(y^*)$ such that

$$p_t(y^*, z^*) \geq 2\epsilon > 0.$$

Thus, for the same interval of t, Assumption 2.1(ii) implies that there exists $z^*, \epsilon_1, \epsilon_2 > 0$ such that,

$$p_t(y,z) \geq \epsilon > 0 \text{ for all } y \in \mathcal{B}_{\epsilon_1}(y^*) \text{ and } z \in \mathcal{B}_{\epsilon_2}(z^*). \tag{6.7}$$

By reduction of ϵ_2 if necessary, we may ensure that $\mathcal{B}_{\epsilon_2}(z^*) \subset C$. Let $n \geq n_0$ and assume for contradiction that for $n\Delta t = t$ there exists $y \in \mathcal{B}_{\epsilon_1}(y^*)$ such that

$$p_{n,\Delta t}(y,z) \leq \frac{1}{2}\epsilon \text{ for all } z \in \mathcal{B}_{\epsilon_2}(z^*).$$

This implies that

$$\inf_{y\in\mathcal{B}_{\epsilon_1}(y^*)} P_{n,\Delta t}(y, \mathcal{B}_{\epsilon_2}(z^*)) \leq \frac{1}{2}\epsilon_2\epsilon$$

whereas (6.7) gives

$$\inf_{y\in\mathcal{B}_{\epsilon_1}(y^*)} P_{n\Delta t}(y, \mathcal{B}_{\epsilon_2}(z^*)) \geq \epsilon_2\epsilon.$$

Reduction of Δt, and use of Condition 6.1, gives a contradiction. Thus, provided

$$n\Delta t \in [t_2-\Delta, t_2+\Delta], \Delta t \leq \Delta t_c, n \geq n_0 \tag{6.8}$$

we deduce that there exists $\bar{z} \in \mathcal{B}_{\epsilon_2}(z^*)$:

$$p_{n,\Delta t}(y^*, \bar{z}) \geq \frac{1}{2}\epsilon$$

and then, by continuity, that

$$p_{n,\Delta t}(y,z) \geq \frac{1}{4}\epsilon \text{ for all } y \in \mathcal{B}_{\delta_1}(y^*) \text{ and } z \in \mathcal{B}_{\delta_2}(\bar{z}).$$

Note that δ_1, δ_2 and $\bar{z}$ may depend upon Δt but that we can assume $\mathcal{B}_{\delta_2}(\bar{z}) \subset C$ without loss of generality, and that $\delta_1, \delta_2 > 0$ uniformly for Δt sufficiently small, because of the derivative conditions on the density for the method.

Thus, assuming (6.8),

$$\begin{aligned} P_{n,\Delta t}(y, A) &= \int_A p_{n,\Delta t}(y, z)dz \\ &\geq \int_{A \cap \mathcal{B}_{\delta_2}(\bar{z})} p_{n,\Delta t}(y, z)dz \\ &\geq \frac{1}{4}\epsilon\lambda(A \cap \mathcal{B}_{\delta_2}(\bar{z})) \end{aligned}$$

for all $y \in \mathcal{B}_{\delta_1}(y^*)$. (Here $\lambda(\cdot)$ is Lebesgue measure on $\mathbb{R}^d$).

By Assumption 2.1(i), we know that t_1 may be chosen so that

$$P_{t_1}(x, \mathcal{B}_{\delta_1}(y^*)) > 0 \quad \forall x \in C.$$

The continuity of $p_\cdot(\cdot, y)$ can be transfered to $P_\cdot(\cdot, A)$ by dominated convergence. Hence we have, since C is compact,

$$\inf_{x \in C} P_t(x, \mathcal{B}_{\delta_1}(y^*)) \geq \gamma > 0$$

for all $t \in [t_1 - \Delta, t_1 + \Delta]$, possibly by reduction of Δ. By approximation, reducing Δt_c if necessary and using Condition 6.1, we deduce that

$$\inf_{x \in C} P_{n,\Delta t}(x, \mathcal{B}_{\delta_1}(y^*)) \geq \frac{1}{2}\gamma > 0$$

for all $n, \Delta t$ satisfying

$$n\Delta t \in [t_1 - \Delta, t_1 + \Delta], \quad \Delta t \leq \Delta t_c, \quad n \geq n_0.$$

By reducing Δ further so that it is less than Δt_c we can find, for all $\Delta t \leq \Delta t_c$, integers n_i such that

$$n_i \Delta t_i \in [t_i - \Delta, t_i + \Delta], \quad i = 1, 2.$$

Now set $M = n_1 + n_2$ and note that, for all $x \in C$,

$$\begin{aligned} P_{M,\Delta t}(x, A) &\geq \int_{\mathcal{B}_{\delta_1}(y^*)} p_{n_1,\Delta t}(x, y) P_{n_2,\Delta t}(y, A)dy \\ &\geq \frac{1}{4}\epsilon\lambda(A \cap \mathcal{B}_{\delta_2}(\bar{z})) \int_{\mathcal{B}_{\delta_1}(y^*)} p_{n_1,\Delta t}(x, y)dy \\ &= \frac{1}{4}\epsilon\lambda(A \cap \mathcal{B}_{\epsilon_2}(\bar{z})) P_{n_1,\Delta t}(x, \mathcal{B}_{\delta_1}(y^*)) \\ &\geq \frac{1}{8}\epsilon\gamma\lambda(\mathcal{B}_{\delta_2}(\bar{z}))\nu(A), \end{aligned}$$

where $\nu(\cdot)$ is Lebesgue measure restricted to $\mathcal{B}_{\delta_2}(\bar{z})$ and normalized to be a probability measure. Thus we have

$$P_{M,\Delta t}(x, A) \geq \eta\nu(A) \quad \forall A \in \mathcal{B}(\mathbb{R}^d)$$

and $x \in C$ where $\eta = \frac{1}{8}\epsilon\gamma\lambda(\mathcal{B}_{\delta_2}(\bar{z}))$. Since $\mathcal{B}_{\delta_2}(\bar{z}) \subset C$, we have $\nu(C^c) = 0$ and $\nu(C) = \nu(\mathcal{B}_{\delta_2}(\bar{z})) = 1$, as required. □

Note that, alternatively, the minorization condition can often be established directly, by mimicking the techniques used for the SDE; we do this in sections 7 and 8.

6.3 The Lyapunov Condition

Although the minorization condition is robust to discretization, the Lyapunov condition is not. Consider the SDE (2.1) with $d = 1$, $Y(x) = -x^3$ and $\Sigma = 1$ so that

$$dx = -x^3 dt + dW. \tag{6.9}$$

From section 4 we know that this SDE is ergodic and, for example, $V(x) = 1 + x^2$ is a Lyapunov function since

$$\mathcal{A}x^2 = -2x^4 + 1 \leq -4x^2 + 3.$$

When the Euler–Maruyama method (6.2) is applied to (6.9) this Lyapunov structure is lost, as Lemma 6.3 shows: it follows from Lemma 6.3(i) that the numerical solution is not ergodic in the sense we have used it so far—namely exponential convergence of induced measures to a unique limit—and from Lemma 6.3(ii) that it is not ergodic in a second commonly used sense—namely almost sure convergence of time-averages to a limit independent of the sample path. Hence the lemma shows that, in the case of non-globally Lipschitz vector fields, numerical methods do not automatically preserve ergodicity, even for small stepsizes. The example motivates the work in section 8 where positive results about ergodicity are proved for certain implicit methods. Note, however, that if the Lyapunov function V is quadratic, and the vector field Y globally Lipschitz, then the Lyapunov condition is preserved for all reasonable approximations, not just specially constructed ones – see section 7.

Lemma 6.3 *Consider the SDE (6.9), noting that it is geometrically ergodic (see section 4). When the Euler–Maruyama method (6.2) is applied to the SDE, the following results hold.*

(i) *If* $\mathbb{E}[X_0^2] \geq \frac{2}{\Delta t}$ *then* $\mathbb{E}[X_n^2] \to \infty$ *as* $n \to \infty$.

(ii) *For any* $X_0 \in \mathbb{R}$ *and any* $\Delta t > 0$

$$\mathbb{P}\left(|X_n| \geq \frac{2^n}{\sqrt{\Delta t}}, \quad \forall\, n \geq 1\right) > 0.$$

Proof **(i)** We have

$$X_{n+1} = X_n(1 - \Delta t X_n^2) + \Delta W_n. \tag{6.10}$$

Squaring and taking expected values gives

$$\mathbb{E}[X_{n+1}^2] = \mathbb{E}[X_n^2(1 - 2\Delta t X_n^2 + \Delta t^2 X_n^4)] + \Delta t. \tag{6.11}$$

Since $1 - 2z + z^2 \geq -1 + \frac{1}{2}z^2$ for all $z \in \mathbb{R}$, we may weaken (6.11) to

$$\mathbb{E}[X_{n+1}^2] \geq \mathbb{E}[-X_n^2 + \tfrac{1}{2}\Delta t^2 X_n^6] + \Delta t = -\mathbb{E}[X_n^2] + \tfrac{1}{2}\Delta t^2 \mathbb{E}[X_n^6] + \Delta t. \tag{6.12}$$

We have $\left(\mathbb{E}[X_n^2]\right)^3 \leq \mathbb{E}[X_n^6]$ from Jensen's inequality, and hence,

$$\mathbb{E}[X_{n+1}^2] \geq \mathbb{E}[X_n^2]\left(\tfrac{1}{2}\Delta t^2 \mathbb{E}[X_n^2]^2 - 1\right) + \Delta t. \tag{6.13}$$

Now if $\mathbb{E}[X_0^2] \geq \frac{2}{\Delta t}$ then we see from (6.13) that $\mathbb{E}[X_1^2] \geq \mathbb{E}[X_0^2] + \Delta t$ and iterating this argument we find that

$$\mathbb{E}[X_n^2] \geq \mathbb{E}[X_0^2] + n\Delta t.$$

Hence $\mathbb{E}[X_n^2] \to \infty$ as $n \to \infty$. This proves **(i)**.

(ii) We deal first with the case where $|X_0|^2 < \frac{4}{\Delta t}$. Assume that the following events arise:

$$|\Delta W_0| \geq \frac{4}{\sqrt{\Delta t}} + \Delta t\left(\frac{2}{\sqrt{\Delta t}}\right)^3, \tag{6.14}$$

$$|\Delta W_n| \leq \frac{2^n}{\sqrt{\Delta t}}, \quad \text{for } n \geq 1. \tag{6.15}$$

Since $|X_0|^2 < \frac{4}{\Delta t}$, it follows from (6.14) that

$$|\Delta W_0| \geq \frac{2}{\sqrt{\Delta t}} + |X_0| + \Delta t |X_0|^3.$$

Hence, using (6.10),

$$|X_1| \geq |\Delta W_0| - |X_0| - \Delta t |X_0|^3 \geq \frac{2}{\sqrt{\Delta t}}. \tag{6.16}$$

Now, consider the induction hypothesis

$$|X_k| \geq \frac{2^k}{\sqrt{\Delta t}}, \quad 1 \leq k \leq n, \tag{6.17}$$

which, from (6.16), holds for $n = 1$. Using (6.17) gives

$$[1 - \Delta t X_n^2] \leq 1 - 2^{2n} \leq 1 - 4 = -3,$$

and hence, from (6.10), (6.15),

$$|X_{n+1}| \geq \frac{2^n}{\sqrt{\Delta t}} 3 - |\Delta W_n| \geq \frac{2^{n+1}}{\sqrt{\Delta t}}.$$

So, by induction, (6.17) holds for all n.

It remains to show that the events (6.14)–(6.15) occur with positive probability. (Recall that the ΔW_n are independent, $\mathcal{N}(0, \Delta t)$ random variables.) Clearly (6.14) occurs with positive probability. Now, for some constants D, E and $\widehat{n}$ we have, for $n \geq \widehat{n}$,

$$\begin{aligned}
\mathbb{P}\left(|\Delta W_n| \leq \frac{2^n}{\sqrt{\Delta t}}\right) &= 1 - \frac{2}{\sqrt{2\pi\Delta t}} \int_{\frac{2^n}{\sqrt{\Delta t}}}^{\infty} \exp(-x^2/(2\Delta t))\, dx \\
&= 1 - D \int_{\frac{2^n}{\sqrt{2}\Delta t}}^{\infty} \exp(-y^2)\, dy \\
&\geq 1 - D \int_{\frac{2^n}{\sqrt{2}\Delta t}}^{\infty} \exp(-y)\, dy \\
&\geq 1 - D \exp(-2^n E).
\end{aligned}$$

By increasing $\widehat{n}$ if necessary we have

$$\log\left(1 - D\exp(-2^n E)\right) \geq -2D\exp(-2^n E) \geq -Fr^n, \quad n \geq \widehat{n},$$

for constants F and r with $0 < r < 1$. It follows that

$$\log\left(\prod_{n \geq \widehat{n}} \mathbb{P}\left(|\Delta W_n| \leq \frac{2^n}{\sqrt{\Delta t}}\right)\right) \geq -\sum_{n \geq \widehat{n}} Fr^n = -G,$$

for some finite constant $G > 0$. Hence,

$$\prod_{n \geq \widehat{n}} \mathbb{P}\left(|\Delta W_n| \leq \frac{2^n}{\sqrt{\Delta t}}\right) \geq \exp(-G) > 0.$$

Since each of the finite number of independent events $|\Delta W_n| \leq \frac{2^{n+1}}{\sqrt{\Delta t}}$ for $n < \widehat{n}$ has positive probability, the result follows.

In the case where $|X_0|^2 \geq \frac{4}{\Delta t}$ a similar approach can be used, based on the events

$$|\Delta W_n| \leq \frac{2^n}{\sqrt{\Delta t}}, \quad \text{for } n \geq 0.$$

□

A rather general, but less detailed, analysis of similar issues may be found in section 3 of [20]. (We retain our explicit calculations with a concrete example as they are instructive for intuition.) Furthermore a similar result to our Lemma 6.3 (i) is also contained in the recent paper [30].

7 Globally Lipschitz Vector Fields

We assume that, by appropriate choice of t_1, t_2 etc. the transition kernel $P_t(x, A)$ for the SDE (2.1) satisfies Assumptions 2.1 and 2.2. Theorem 2.5 then implies that the SDE is geometrically ergodic. We would like to establish conditions under which the same can be said of the three numerical methods (6.2), (6.3) and (6.4). We do this by appealing to Theorem 2.5. However, we start simply by considering the effect of approximation on Lyapunov conditions.

We consider the general family of methods (6.5) for (2.1) and then look at the three Euler methods as special cases. Writing $x_n = x(n\Delta t)$, where $x(t)$ solves (2.1), we consider the following conditions concerning (6.5) and its relation to (2.1).

Condition 7.1 *The function $H \in C^\infty(\mathbb{R}^d \times \mathbb{R}^m, \mathbb{R}^d)$ and satisfies:*

(i) *there exist $c_1 > 0, s > 0$ such that $\mathbb{E}\|X_1 - x_1\|^2 \le c_1[1 + \|y\|^2]\Delta t^{s+2}$ for all $y \in \mathbb{R}^d$;*

(ii) *there exists $c_2 = c_2(r) > 0$ such that $\mathbb{E}\|X_1\|^r \le c_2[1 + \|y\|^r]$, for all $r \ge 1$ and $y \in \mathbb{R}^d$;*

The next result gives conditions under which the numerical method (6.5) inherits a Lyapunov function from the SDE (2.1). We say that V is *essentially quadratic* if there exist $C_i > 0$ so that

$$C_1[1 + ||x||^2] \le V(x) \le C_2[1 + ||x||^2], \quad |\nabla V(x)| \le C_3[1 + \|x\|]. \tag{7.1}$$

Theorem 7.2 *Let Assumption 2.4 hold for (2.1) with $V \to V^l$, $l \ge 1$ and let V be essentially quadratic. If Condition 7.1 holds, then Assumption 2.2 holds for (6.5) with $V \to V^l$.*

Proof We have that

$$\mathbb{E}\{V(X_1)^l\} \le \mathbb{E}\{V(x_1)^l\} + \mathbb{E}|V(X_1)^l - V(x_1)^l|.$$

Assumption 2.4 implies that

$$\mathbb{E}\{V(x(t))^l\} \le e^{-a_l t}V(x(0))^l + \frac{d_l}{a_l}[1 - e^{-a_l t}].$$

Since $V(x)$ is essentially quadratic it follows from (7.1) that there are $c_l > 0$ such that

$$\mathbb{E}\|x(t)\|^{2l} \le c_l^+[1 + \|y\|^{2l}]. \tag{7.2}$$

Thus, by Assumption 2.4 with $V \to V^l$ and since ∇V is linearly bounded and V quadratically bounded,

$$\begin{aligned}
\mathbb{E}\{V(X_1)^l\} &\le e^{-a_l\Delta t}V(y)^l + \frac{d_l}{a_l} \\
&\quad + \mathbb{E}\int_0^1 |\langle \nabla V^l(sX_1 + (1-s)x_1), X_1 - x_1\rangle| ds \\
&\le e^{-a_l\Delta t}V(y)^l + \frac{d_l}{a_l} \\
&\quad + k_1\mathbb{E}\{[1 + \|X_1\|^{2l-1} + \|x_1\|^{2l-1}]\|X_1 - x_1\|\} \\
&\le e^{-a_l\Delta t}V(y)^l + \frac{d_l}{a_l} \\
&\quad + k_2\{\mathbb{E}[1 + \|X_1\|^{4l-2} + \|x_1\|^{4l-2}]\}^{\frac{1}{2}}\{\mathbb{E}\|X_1 - x_1\|^2\}^{\frac{1}{2}}.
\end{aligned}$$

Using, (7.2), Condition 7.1 **(ii)** to bound $\mathbb{E}\|x_1\|^{4l-2}$, $\mathbb{E}\|X_1\|^{4l-2}$ and Condition 7.1 **(i)** to bound $\mathbb{E}\|X_1 - x_1\|^2$, we find from (7.1) that

$$\begin{aligned}
\mathbb{E}\{V(X_1)^l\} &\le e^{-a_l\Delta t}V(y)^l + \frac{d_l}{a_l} \\
&\quad + k_3\{1 + \|y\|^{4l-2}\}^{\frac{1}{2}}\{1 + \|y\|^2\}^{\frac{1}{2}}\Delta t^{1+s/2} \\
&\le e^{-a_l\Delta t}V(y)^l + \frac{d_l}{a_l} + k_4\{1 + \|y\|^{2l}\}\Delta t^{1+s/2} \\
&\le [e^{-a_l\Delta t} + k_5\Delta t^{1+s/2}]V(y)^l + \frac{d_l}{a_l} + k_6\Delta t^{1+s/2}.
\end{aligned}$$

Thus, for $\tilde{a}_l \in (0, a_l)$,

$$\mathbb{E}\{V(X_1)^l\} \le e^{-\tilde{a}_l \Delta t} V(y)^l + \frac{d_l}{\tilde{a}_l} \tag{7.3}$$

by choice of Δt sufficiently small. This is the desired result. □

If Condition 7.1 holds then we may prove the following result, which employs the definition (3.7) and

$$\mathcal{G}_l' = \{g \in \mathcal{G}_l : \quad |g(a) - g(b)| \le k[1 + \|a\|^{2l-1} + \|b\|^{2l-1}]\|a - b\| \quad \forall a, b \in \mathbb{R}^d\}.$$

Theorem 7.3 *Let Assumptions 2.1 and 2.4 hold,, with $V \to V^l$, $l \ge 1$, and let V be essentially quadratic. Thus Theorem 2.5 holds and (2.1) is geometrically ergodic with invariant measure π. If Condition 7.1 holds and if the numerical method (6.5) satisfies the minorization condition when sampled at rate M, then for all Δt sufficiently small, the method has a unique invariant measure $\pi^{\Delta t}$ on $\mathbb{R}^d$. For $l \ge 1$ there exists $\tilde{C} = \tilde{C}(l, \Delta t) > 0$ and $\tilde{\lambda} = \tilde{\lambda}(l, \Delta t) > 0$ such that, for all $g \in \mathcal{G}_l$,*

$$|\mathbb{E}g(X_n) - \pi^{\Delta t}(g)| \le \tilde{C} V(y)^l e^{-\tilde{\lambda} n \Delta t}, \quad \forall n \ge 0.$$

If, in addition,

$$\mathbb{E}\|X_n - x_n\|^2 \le c_3 e^{2c_4 T}[1 + \|y\|^2]\Delta t^s, \text{ for all } 0 \le n\Delta t \le T, \tag{7.4}$$

then there is $K = K(l) > 0$ and $\xi \in (0, 1/2)$ independent of l such that, for all $g \in \mathcal{G}_l'$,

$$|\pi(g) - \pi^{\Delta t}(g)| \le K \Delta t^{s\xi} \pi(V^l). \tag{7.5}$$

Proof Condition 7.1 implies the Lyapunov condition and we have assumed the minorization condition holds for the sampled chain $\{X_{nM}\}$. Thus, by Theorem 2.5, the sampled chain is geometrically ergodic:

$$|\mathbb{E}^y g(X_{lM}) - \pi(g)| \le \kappa \theta^l [1 + V(y)^l].$$

From this we deduce that the unsampled chain is ergodic since, if $n = lM + j$ for integer $j \in [0, M-1]$, conditioning on $\mathcal{F}_j$ gives, for all $g \in \mathcal{G}_l$,

$$|\mathbb{E}^y g(X_{lM+j}) - \pi(g)| \le \kappa \theta^l [1 + \mathbb{E}^y V(X_j)^l].$$

Using (7.3) gives the desired result

$$|\mathbb{E}^y g(X_n) - \pi(g)| \le \kappa_1 \theta_1^n [1 + \mathbb{E}^y V(X_0)^l]$$

for all $g \in \mathcal{G}_l$.

To obtain the second result on convergence of invariant measures we apply Theorem 3.3 in [24]. We need only show

$$|\mathbb{E}^y g(x(n\Delta t) - \mathbb{E}^y g(X_n)| \le C e^{\eta t} V(y)^l \Delta t^s, \quad 0 \le n\Delta t \le t,$$

for all $g \in \mathcal{G}_l'$. Now, for $0 \le n\Delta t \le t$,

$$\begin{aligned} \mathbb{E}|g(x(n\Delta t)) - g(X_n)| &\le C\mathbb{E}\{[1 + \|x(n\Delta t)\|^{2l-1} + \|X_n\|^{2l-1}]\|x(n\Delta t) - X_n\|\} \\ &\le C\mathbb{E}\{1 + \|x(n\Delta t)\|^{4l-2} \\ &\qquad + \|X_n\|^{4l-2}\}^{\frac{1}{2}} \mathbb{E}\{\|x(n\Delta t) - X_n\|^2\}^{\frac{1}{2}}, \end{aligned}$$

so that, by (7.1), (7.2), Condition 7.1 **(ii)** and (7.4),

$$\begin{aligned} |\mathbb{E}g(x(n\Delta t)) - \mathbb{E}g(X_n)| &\le C[1 + \|y\|^{2l-1}] e^{c_4 n \Delta t} [1 + \|y\|] \Delta t^{s/2} \\ &\le C_4 e^{\eta t} V(y)^l \Delta t^{s/2}. \end{aligned}$$

The required result follows. □

Remark If (7.5) holds for all $g \in \mathcal{G}_l$ then it states that π and $\pi^{\Delta t}$ are close in a total variation norm, weighted according to the Lyapunov function. The additional constraints implied by requiring $g \in \mathcal{G}_l'$ lead to a more complex metric. □

The essential point of this theorem is that *approximation properties alone* allow us, in the case of globally Lipschitz Y (which suffices to establish Condition 7.1) and essentially quadratic V, to deduce ergodicity for the numerical method; this is since they imply both the minorization and Lyapunov conditions. Recall, however, that it is sometimes straightforward to deduce the minorization condition directly for the numerical method, without resort to approximation, and that we will use this approach in what follows for the Langevin equation; for other problems, however, it may sometimes be easier to use approximation.

We now give two examples where Condition 7.1 holds, and hence Theorem 7.3 applies, for the three numerical methods defined above. The first example involves the Langevin equation. Note that the hypotheses on F in Corollary 7.4 below are automatically satisfied if F is a positive definite quadratic form. In this case, an appropriate choice for V, which ensures that Assumption 2.4 holds, is (3.6). Using this V, the equation (3.1)–(3.2) is proved to be geometrically ergodic in section 3.

Corollary 7.4 *Consider the Langevin equation* (3.1)–(3.2) *where* $F : \mathbb{R}^m \to \mathbb{R}$ *is essentially quadratic and* $\sigma \in \mathbb{R}^{m\times m}$. *Suppose that the columns of* σ *are linearly independent, and that* F *has the following properties:*

(i) $F \in C^\infty(\mathbb{R}^m, \mathbb{R})$*;*

(ii) ∇F *is globally Lipschitz;*

(iii) $F(q) \geq 0$*;*

(iv) *there exists an* $\alpha > 0$ *and* $\beta \in (0, 1)$ *such that*

$$\frac{1}{2}\langle \nabla F(q), q\rangle \geq \beta F(q) + \gamma^2 \frac{\beta(2-\beta)}{8(1-\beta)}||q||^2 - \alpha.$$

For Δt *sufficiently small the three numerical methods (6.2), (6.3) and (6.4) satisfy Condition 7.1, the minorization condition when sampled at rate* $M = 2$ *and* (7.4) *and hence Theorem 7.3 applies.*

Proof We begin with the Euler–Maruyama scheme (6.2), which gives

$$\begin{aligned} Q_{n+1} &= Q_n + \Delta t P_n, &(7.6)\\ P_{n+1} &= P_n - \Delta t\gamma P_n - \Delta t \nabla F(Q_n) + \sigma \Delta W_n. &(7.7)\end{aligned}$$

Here $Q_n \approx q(n\Delta t)$ and $P_n \approx p(n\Delta t)$. Because σ is invertible it follows that $P_n(y, A)$ has C^∞ density for $n \geq 2$. Explicit construction shows that ΔW_0, ΔW_1 can be chosen to ensure that $(Q_2^T, P_2^T)^T = y^+$ for any starting value y: note that Q_1 is fixed independently of the noise and hence P_1 is forced to ensure Q_2 takes the required value. This value for P_1 determines ΔW_0 uniquely and then ΔW_1 is determined uniquely to ensure the desired value of P_2. Thus Assumption 2.1 holds, and hence the minorization condition by Lemma 2.3.

Considered as an approximation to (2.1), the Euler–Maruyama method may be written as

$$X_{n+1} = X_n + \Delta t Y(X_n) + \Sigma \Delta W_n.$$

Here Y is globally Lipschitz. Also

$$x(\Delta t) = y + \int_0^{\Delta t} Y(x(\tau))d\tau + \Sigma W(\Delta t).$$

Subtracting we find that

$$\|x(\Delta t) - X_1\| \leq \int_0^{\Delta t} L\|x(\tau) - y\|d\tau$$

so that

$$\mathbb{E}\|x(\Delta t) - X_1\|^2 \leq \Delta t L^2 \int_0^{\Delta t} \mathbb{E}\|x(\tau) - y\|^2 d\tau.$$

Further calculation shows that

$$\mathbb{E}\|x(\tau) - y\|^2 \leq C\tau[1 + y^2]$$

and Condition 7.1 **(i)** follows with $s = 1$.

For **(ii)** notice that

$$\begin{aligned} \|X_1\|^p &\leq C[\|y\|^p + \Delta t^p \|Y(y)\|^p + \|\Sigma \Delta W_1\|^p] \\ &\leq C[\|y\|^p + \Delta t^p [\|Y(0)\| + L\|y\|]^p + \|\Sigma \Delta W_1\|^p] \\ &\leq C[1 + \|y\|^p + \|\sigma \Delta W_1\|^p] \end{aligned}$$

and taking expectations gives the desired result. Condition (7.4) is established in [24] with $s = 1$. Theorem 7.3 thus applies with V given by (3.6).

Applying the split-step stochastic backward Euler method (6.4) to (3.1)–(3.2) gives

$$\begin{aligned} Q_{n+1} &= Q_n + \Delta t P_\star & (7.8) \\ P_\star &= P_n - \Delta t \gamma P_\star - \Delta t \nabla F(Q_{n+1}) & (7.9) \\ P_{n+1} &= P_\star + \sigma \Delta W_n. & (7.10) \end{aligned}$$

By the techniques described in subsection 8.1 it is possible to show that, for all Δt sufficiently small, the map $(Q_n, P_n) \to (Q_n, P_\star)$ is uniquely defined, whatever values Q_n, P_n and ΔW_n take. Indeed we may write

$$\begin{aligned} Q_{n+1} &= Q_n + \Delta t f(Q_n, P_n) \\ P_{n+1} &= f(P_n, Q_n) + \sigma \Delta W_n. \end{aligned}$$

Here f is smooth in both arguments and $f(q, \cdot)$ is invertible for all $q \in \mathbb{R}^d$. Thus the method is well-defined. Analysis very similar to that above for the Euler–Maruyama scheme shows that Condition 7.1 holds, together with the minorization condition for the chain sampled at rate $M = 2$. The stochastic backward Euler method (6.3) for (3.1)–(3.2) can be analyzed similarly. □

Our second example where Theorem 7.3 applies involves a dissipativity condition.

Corollary 7.5 *Consider* (2.1) *in the case where $m = d$, Y is globally Lipschitz and the following properties hold*

(i) $Y \in C^\infty(\mathbb{R}^d, \mathbb{R}^d)$,

(ii) $\exists \alpha, \beta > 0$ *such that* $\langle Y(x), x \rangle \leq \alpha - \beta \|x\|^2$ *for all* $x \in \mathbb{R}^d$.

For Δt sufficiently small, the three numerical methods (6.2), (6.3) and (6.4) satisfy Condition 7.1, the minorization condition when sampled at rate $M = 1$ and (7.4) *and hence Theorem 7.3 applies.*

Proof Note that for this SDE, an appropriate choice for the Lyapunov function V is $V(x) = \|x\|^2 + 1$; see section 4, where the equation is proved to be geometrically ergodic. Since the columns of Σ span $\mathbb{R}^d$ in this case it follows every point y^+ is reachable from y in just one step ($N = 1$) by appropriate choice of ΔW_0. Thus minorization is easily verified for all three methods. The remaining arguments follow as in the Langevin case. □

Corollary 7.5 is essentially proved in [28] for the Euler–Maruyama scheme though, in that paper, certain higher order methods are also studied and, furthermore, the rates of convergence of $\pi^{\Delta t}$ to π is optimal. In contrast the use of [24] to prove convergence gives suboptimal rates in this case; it does, however, apply to a different set of test functions.

8 Locally Lipschitz Vector Fields

We now consider the equation (2.1) without the condition that Y is globally Lipschitz. To be concrete we study the Langevin equation (3.1)–(3.2) but similar issues arise for other problems and we briefly outline generalizations at the end of the section.

8.1 The Langevin Equation

For the Langevin problem we impose the structural property that

$$\exists c > 0 : \quad \langle \nabla F(a) - \nabla F(b), a - b \rangle \geq -c\|a - b\|^2. \tag{8.1}$$

This is a one-sided Lipschitz condition on ∇F and it implies that

$$F(a) - F(b) \leq \langle \nabla F(a), a - b \rangle + c\|a - b\|^2. \tag{8.2}$$

In this subsection we replace condition **(ii)** of Corollary 7.4 by the one-sided Lipschitz condition. The function (3.3) is a prototypical example that satisfies conditions **(i)**, **(iii)** and **(iv)** of Corollary 7.4 and (8.1). Analysis similar to that in the previous section shows that the Euler-Maruyama approximation of this problem is not ergodic in general. Here we study the split-step backward Euler method.

Abusing notation and setting $V(p,q) = V(x)$ for $x = (q^T, p^T)^T$ (with V given by (3.6)) we define

$$V_{\Delta t}(p,q) \stackrel{\text{def}}{=} V(p,q) + \frac{\Delta t \gamma}{4}\|p\|^2. \tag{8.3}$$

The following lemma is key to what follows:

Lemma 8.1 *Let* (8.1) *hold and let* $\Delta t \leq \Delta t_c$ *where* $c\Delta t_c^2 = 1 + \gamma \Delta t_c$. *Then the map* $(Q_n, P_n) \to (Q_{n+1}, P_\star)$ *given by (7.8)–(7.9) is uniquely defined for all* Q_n, P_n *and* ΔW_n. *Furthermore, if* $\Delta t \leq \frac{\epsilon\gamma\beta}{8c}$ *for some* $\epsilon \in (0,1)$, *then we have*

$$V_{\Delta t}(P_\star, Q_{n+1}) - V_{\Delta t}(P_n, Q_n) \leq \gamma \Delta t \alpha - \gamma(1-\epsilon)\Delta t \beta V_{\Delta t}(P_\star, Q_{n+1}).$$

□

Proof Solvability is equivalent to finding $P_\star$ such that

$$P_\star - P_n + \gamma \Delta t P_\star + \Delta t \nabla F(Q_n + \Delta t P_\star) = 0,$$

and hence to making

$$\frac{1}{2}\|P_\star - P_n\|^2 + \frac{\gamma \Delta t}{2}\|P_\star\|^2 + F(Q_n + \Delta t P_\star)$$

stationary. Since F is smooth and bounded below at least one such point must exist. For uniqueness, consider two solutions p_1 and p_2 given $(P_n, Q_n) = (p, q)$. Then

$$(1 + \gamma \Delta t) p_i + \Delta t \nabla F(q + \Delta t p_i) = p.$$

Subtracting and using (8.1) gives

$$0 \geq (1 + \gamma \Delta t - c\Delta t^2)\|p_1 - p_2\|^2$$

and uniqueness follows under the required condition on Δt.

For the properties of V define $V_n = V(P_n, Q_n)$ and $V_{n+1} = V(P_\star, Q_{n+1})$; note that

$$\begin{aligned} V_{n+1} - V_n &= \frac{1}{2}\langle P_\star - P_n, P_\star + P_n \rangle + F(Q_{n+1}) - F(Q_n) \\ &\quad + \frac{\gamma}{2}\langle P_\star - P_n, Q_{n+1} \rangle + \frac{\gamma}{2}\langle P_n, Q_{n+1} - Q_n \rangle \\ &\quad + \frac{\gamma^2}{4}\langle Q_{n+1} - Q_n, Q_{n+1} + Q_n \rangle. \end{aligned}$$

From this it may be shown that

$$\begin{aligned} V_{n+1} - V_n &\leq \langle -\gamma\Delta t P_\star - \Delta t \nabla F(Q_{n+1}), P_\star\rangle + F(Q_{n+1}) - F(Q_n) \\ &\quad + \frac{\gamma}{2}\langle -\gamma\Delta t P_\star - \Delta t\nabla F(Q_{n+1}), Q_{n+1}\rangle \\ &\quad + \frac{\gamma}{2}\langle P_n, \Delta t P_\star\rangle + \frac{\gamma^2}{4}\langle \Delta t P_\star, Q_{n+1} + Q_n\rangle. \end{aligned}$$

Thus

$$\begin{aligned} V_{n+1} - V_n &\leq [c\Delta t^2 - \frac{\gamma^2\Delta t^2}{4} - \frac{\gamma\Delta t}{2}]\|P_\star\|^2 + \frac{\Delta t\gamma}{2}\langle P_n - P_\star, P_\star\rangle \\ &\quad - \frac{\gamma\Delta t}{2}\langle \nabla F(Q_{n+1}), Q_{n+1}\rangle. \end{aligned}$$

Using the fact that

$$\langle a - b, b\rangle \leq \frac{1}{2}\|a\|^2 - \frac{1}{2}\|b\|^2$$

and (3.11) we see that

$$\begin{aligned} V_{\Delta t}(P_\star, Q_{n+1}) - V_{\Delta t}(P_n, Q_n) &\leq [c\Delta t^2 - \tfrac{\gamma^2\Delta t^2}{4} - \tfrac{\gamma\Delta t}{2}]\|P_\star\|^2 - \tfrac{\gamma\Delta t}{2}\langle \nabla F(Q_{n+1}), Q_{n+1}\rangle \\ &\leq \left(c\Delta t^2 - \tfrac{\gamma^2\Delta t^2}{4}\right)\|P_\star\|^2 + \gamma\Delta t[\alpha - \beta V(P_\star, Q_{n+1})] \\ &\leq \left(c\Delta t^2 - \tfrac{\gamma^2\Delta t^2}{4} + \tfrac{\gamma^2\Delta t^2\beta}{4}\right)\|P_\star\|^2 + \gamma\Delta t[\alpha - \beta V_{\Delta t}(P_\star, Q_{n+1})]. \end{aligned}$$

Since $\beta \in (0,1)$ it follows that, using $V_{\Delta t}(p,q) \geq \frac{1}{8}\|p\|^2$,

$$\begin{aligned} V_{\Delta t}(P_\star, Q_{n+1}) - V_{\Delta t}(P_n, Q_n) &\leq \left(c\Delta t^2 - \frac{\epsilon\Delta t\beta\gamma}{8}\right)\|P_\star\|^2 + \gamma\Delta t\alpha \\ &\quad - \gamma(1-\epsilon)\Delta t\beta V_{\Delta t}(P_\star, Q_{n+1}) \end{aligned}$$

and the required result follows. □

Corollary 8.2 *Consider the Langevin equation (3.1)–(3.2) under the assumptions of Corollary 7.4 with part **(ii)** of the conditions on F replaced by (8.1). Let*

$$\Delta t \leq \min\{\Delta t_c, \frac{\epsilon\gamma\beta}{8c}, \frac{2}{\gamma}\},$$

where Δt_c is as given in Lemma 8.1. Then the split-step stochastic Backward Euler method is geometrically ergodic and the conclusions of Theorem 2.5 apply.

Proof By (3.6) and (8.3) we see that, for $X_n = (Q_n^T, P_n^T)^T$,

$$\begin{aligned} V_{\Delta t}(X_{n+1}) &= V_{\Delta t}(P_\star, Q_{n+1}) + (1 + \Delta t\gamma/2)\langle P_\star, \sigma\Delta W_n\rangle \\ &\quad + \frac{1}{2}(1 + \frac{\Delta t\gamma}{2})\|\sigma\Delta W_n\|^2 + \frac{\gamma}{2}\langle \sigma\Delta W_n, Q_{n+1}\rangle. \end{aligned}$$

Thus, assuming that $\Delta t\gamma < 2$ to simplify the constants and noting that P_*, Q_{n+1} are independent of ΔW_n,

$$\mathbb{E}\{V_{\Delta t}(X_{n+1})|\mathcal{F}_n\} = \mathbb{E}\{V_{\Delta t}(P_\star, Q_{n+1})|\mathcal{F}_n\} + \mathbb{E}\|\sigma\Delta W_n\|^2.$$

Hence, if $\Delta t\zeta = \frac{1}{2}\mathbb{E}\|\sigma\Delta W_n\|^2$, we have upon application of Lemma 8.1,

$$\mathbb{E}\{V_{\Delta t}(X_{n+1})|\mathcal{F}_n\} = \frac{V_{\Delta t}(X_n) + \Delta t[\gamma\alpha + 2\zeta(1 + \gamma(1-\epsilon)\Delta t\beta)]}{1 + \gamma(1-\epsilon)\Delta t\beta}.$$

Hence Assumption 2.2 holds. Smoothness of the density and reachability of any y^+ from any y in $N = 2$ steps can be established as in section 6, yielding Assumption 2.1. This proves ergodicity for the chain sampled every 2 steps. An argument similar to that used in Theorem 7.3 to similar effect, gives ergodicity for the unsampled chain $\{X_n\}_{n\in\mathbb{Z}^+}$. □

8.2 Dissipative Problems

Throughout this subsection we assume that Condition 4.1 holds. The split-step stochastic backward Euler method applied to (2.1) gives (6.4). Standard calculations (see [27, Chapter 5]) using conditions **(i)** and **(ii)** of Corollary 7.5 show that

$$\|X_\star\|^2 \leq \{1 + \Delta t\beta\}^{-1}\{\|X_n\|^2 + \alpha\Delta t\}$$

and so

$$\mathbb{E}\{\|X_{n+1}\|^2|\mathcal{F}_n\} \leq (1 + \Delta t\beta)^{-l}\{\|X_n\|^2 + \alpha\Delta t + \mathbb{E}\|\Sigma\Delta W_n\|^2\}.$$

Since $\mathbb{E}\|\Sigma\Delta W_n\|^2 = \mathcal{O}(\Delta t)$ we have the required Lyapunov function structure. Thus, provided that the desired minorization condition can be proved, either by approximation or directly, geometric ergodicity follows for this approximation method whenever the underlying SDE (2.1) is geometrically ergodic and satisfies conditions **(i)** and **(ii)** of Corollary 7.5.

A Lyapunov function for the stochastic Backward Euler method (6.3) follows from the preceding analysis. If X_n solves (6.3) then

$$Z_n \stackrel{\text{def}}{=} X_n - \Delta t Y(X_n)$$

solves (6.4). Thus, since $1 + \|Z_n\|^2$ is a Lyapunov function for (6.4), we see that

$$V(x) \stackrel{\text{def}}{=} 1 + \|x - \Delta t Y(x)\|^2$$

is a Lyapunov function for (6.3). That $V(x) \to \infty$ as $\|x\| \to \infty$ follows under Condition 4.1.

8.3 Gradient Systems

For gradient problems it is often the case that the dissipativity structure exploited in the previous subsection also prevails and then the split-step backward Euler method can be shown to be ergodic when applied to geometrically ergodic gradient systems perturbed by noise. However there are examples where this is not the case. Although the ultimate boundedness of $F(X_n)$ implies the ultimate boundedness of $\|X_n\|^2$, *exponential* dissipation in $F(X_n)$ does not imply *exponential* dissipation in $\|X_n\|$, as shown by the example

$$F : \mathbb{R}^2 \to \mathbb{R}^+, \quad F(a_1, a_2) = \left(|a_1| + \log|a_2|\right)^2.$$

It is therefore possibly useful to find numerical methods which preserve the expected Lyapunov structure which we exploited in studying gradient systems in section 5. However, whilst this is possible by use of ideas in [27], we have been unable to find Lyapunov structures which are well-behaved in the limit $\Delta t \to 0$; this is aesthetically unsatisfactory and means that, in principle, the geometric rates of convergence may depend badly on $\Delta t \to 0$ [21]. In practice we do not believe that this occurs and numerical experiments like those in the next section substantiate this claim.

9 Numerical Experiments

We now give some numerical results that are relevant to the foregoing analysis. We use FE, BE and SSBE to denote the Forward Euler method (6.2), the Backward Euler method (6.3) and the split-step Backward Euler method (6.4), respectively. We consider four problems that illustrate results in (sub)sections 7, 8.1, 8.2 and 8.3 respectively.

Lang-Global: the Langevin equation (3.1)–(3.2) with $m = 1$, $F(q) = q^2/2 - (\log(q^2 + 1))/2$, $\gamma = 1$ and $\sigma = 1$, and initial data $p_0 = q_0 = \frac{1}{2}$. Here, the deterministic vector field is globally Lipschitz.

Lang-Local: the Langevin equation (3.1)–(3.2) with $m = 1$, $F(q) = \frac{1}{4}(q^2 - 1)^2$, $\gamma = 1$ and $\sigma = 1$, and initial data $p_0 = q_0 = \frac{1}{2}$. Here, the deterministic vector field is only locally Lipschitz. In this case, the condition (8.1) holds with $c = 1$, and hence Corollary 8.2 applies for SSBE.

Lorenz: the Lorenz equations (4.3) with $\rho = 10$, $r = 28$ and $b = 8/3$ and initial data $x_0 = y_0 = z_0 = 0.5$. This problem is dissipative in the sense of section 4 and hence both BE and SSBE have a Lyapunov function (see subsection 8.2).

Grad-Diss: the gradient system (5.1) with $d = 2$, $F(x_1, x_2) = \frac{1}{2}(\exp(x_1^2) + x_2^2)$, $\sigma = I$, and initial data $x_1(0) = x_2(0) = 0.5$. The problem is also dissipative in the sense of section 4 and our calculations of subsection 8.2 apply for BE and SSBE.

Computations are performed in Matlab [15] using the function `randn` to generate independent $\mathcal{N}(0, 1)$ samples. To apply BE to Lang-Global and Lang-Local, we first eliminate P_{n+1}, leaving a cubic for Q_{n+1}. We take Q_{n+1} to be the real root closest to Q_n, and then substitute this value to give P_{n+1}. Similarly, we apply SSBE to Lang-Global and Lang-Local by solving a cubic polynomial for Q_{n+1}. The same technique is used for the Lorenz equations (4.3); the nonlinearity in the implicit equations for BE and SSBE can be reduced to a cubic polynomial in Y_{n+1} for BE and in $Y_\star$ for SSBE. For Grad-Diss, we implement BE and SSBE by applying a quasi-Newton type nonlinear equation solver.

In all tests, we monitor an approximation to $\mathbb{E}\|X_n\|^2$ that is found by averaging over 1000 paths, using the same paths for each of the three methods. Here, X_n denotes the numerical solution at $t = t_n$ and $\|\cdot\|$ denotes the L_2 norm. We apply the methods over $0 \le t \le 64$ with four different stepsizes.

The Lang-Global results are given in Figure 9.1. Here, we use $\Delta t_i = 2^{-i}$, for $i = 1, 2, 3, 4$. We see that all three methods are well behaved for these stepsizes. The long-time second moments appear to converge to a common limit as $\Delta t \to 0$, with BE and SSBE settling down more quickly than FE.

Figure 9.1: Problem Lang-Global: $\mathbb{E}\|X_n\|^2$ against t_n.

Figure 9.2 relates to Lang-Local, using the same Δt_i values as the previous example. Note that in these (and subsequent) figures the vertical axis for the FE picture uses exponential scaling. We see that the FE solution behaves poorly for $\Delta t = \Delta t_1, \Delta t_2$, suggesting unboundedness of $\mathbb{E}\|X_n\|^2$ as $n \to \infty$. The BE and SSBE solutions behave better, having second moments that are bounded, and appear convergent as $\Delta t \to 0$ in the large-time regime. The FE results for $\Delta t = \Delta t_3, \Delta t_4$ are compatible with those of BE and SSBE.

Figure 9.2: Problem Lang-Local: $\mathbb{E}\|X_n\|^2$ against t_n.

Results for the Lorenz equations are given in Figure 9.3. Here, we use $\Delta t_i = 2^{-i-2}$, for $i = 1, 2, 3, 4$. FE gives unbounded second moments for Δt_1, Δt_2 and Δt_3. For BE and SSBE, this quantity is always bounded and appears convergent to, approximately, the same limit.

Figure 9.4 gives results for Grad-Diss, with $\Delta t_i = 2^{-i}$, for $i = 1, 2, 3, 4$. In this case, FE has unbounded second moments for all stepsizes used. In contrast, BE and SSBE perform well, and convergence in Δt is particularly fast for BE. In further tests with smaller Δt and the same initial data, FE appeared to recover the good behaviour of BE and SSBE, as for the other examples.

The first common theme of all the experiments is that FE blows up unless the time-step is small; we conjecture that, however small the time-step, this method will eventually blow-up, given a long enough time interval. The second common theme is that both BE and SSBE behave well – they produce moments which appear to converge, as $n \to \infty$, to a limiting value which itself converges as $\Delta t \to 0$.

Acknowledgements The authors would like to thank Gérard Ben Arous, Amir Dembo, Persi Diaconis, Weinan E and George Papanicolaou for useful discussions.

Figure 9.3: Problem Lorenz: $\mathbb{E}\|X_n\|^2$ against t_n.

References

[1] L. Arnold and W. Kliemann, *On unique ergodicity for degenerate diffusions*. Stochastics **21**(1987), 41–61.

[2] D. R. Bell. *The Malliavin Calculus*. Longman Scientific & Technical, Harlow, 1987.

[3] R. Durrett, *Stochastic Calculus, A Practical Introduction*. CRC Press, 1996.

[4] W. E and J. C. Mattingly, *Ergodicity for the Navier-Stokes equation with degenerate random forcing: finite dimensional approximation*. Submitted

[5] G. Fayolle, V. A. Malyshev, and M. V. Menshikov. *Topics in the Constructive Theory of Countable Markov Chains*. Cambridge, 1995.

[6] G.W. Ford and M. Kac, *On the quantum Langevin equation*. J. Stat. Phys. **46**(1987), 803–810.

[7] A. Grorud and D. Talay, *Approximation of Lyapunov exponents of nonlinear stochastic differential equations*. SIAM J. Appl. Math. **56**(1996), 627–650.

[8] J.K. Hale. *Asymptotic Behavior of Dissipative Systems*. American Mathematical Society, Providence, RI, 1988.

[9] R. Z. Has'minskii. *Stochastic Stability of Differential Equations*. Sijthoff and Noordhoff, 1980.

[10] Y. Kifer. *Random Perturbations of Dynamical Systems*. Birkhäuser Boston Inc., Boston, MA, 1988.

[11] W. Kliemann, *Recurrence and invariant measures for degenerate diffusions*. The Annals of Probability **15**(1987), 690–707.

[12] P.E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag, New York, 1991.

[13] H. Kunita. Supports of diffusion processes and controllability problems. In *Proceedings of the International Symposium on Stochastic Differential Equations (Res. Inst. Math. Sci., Kyoto Univ., Kyoto, 1976)*, pages 163–185, New York, 1978. Wiley.

[14] X. Mao, *Stochastic Differential Equations and Applications*. Horwood, Chichester, 1997.

[15] The MathWorks, Inc., *MATLAB User's Guide*. Natick, Massachusetts, 1992.

Figure 9.4: Problem Grad-Diss: $\mathbb{E}\|X_n\|^2$ against t_n.

[16] S. Meyn and R.L. Tweedie, *Stochastic Stability of Markov Chains*. Springer-Verlag, New York, 1992.

[17] S.P. Meyn and R.L. Tweedie, *Stability of Markovian processes, I, II and III* Adv. Appl. Prob. **24**(1992), 542–574, **25**(1993), 487–517 and **25**(1933), 518–548.

[18] J. Norris. Simplified Malliavin calculus. In *Séminaire de Probabilités, XX, 1984/85*, pages 101–130. Springer, Berlin, 1986.

[19] S. Orey. *Lecture Notes on Limit Theorems for Markov Chain Transition Probabilities*. Van Nostrand Reinhold Co., London, 1971. Van Nostrand Reinhold Mathematical Studies, No. 34.

[20] G.O. Roberts and R.L. Tweedie *Exponential convergence of Langevin diffusions and their discrete approximations*, Bernoulli **2**(1996), 341–363.

[21] G.O. Roberts and R.L. Tweedie *Bounds on regeneration times and convergence rates for Markov chains*, Stoch. Proc. Applic. **80**(1999), 211–229.

[22] J.S. Rosenthal *Minorization conditions and convergence rates for Markov chain Monte Carlo*, J. Amer. Stat. Ass. **90**(1995), 558–566.

[23] J.M. Sanz-Serna and A.M. Stuart, *Ergodicity of dissipative differential equations subject to random impulses*. J. Diff. Eq. **155**(1999), 262-284.

[24] T. Shardlow and A.M. Stuart, *A perturbation theory for ergodic Markov chains with application to numerical approximation*. SIAM J. Num. Anal.**37**(2000), 1120–1137.

[25] C. Sparrow, *The Lorenz Equations, Bifurcations, Chaos and Strange Attractors*. Springer-Verlag, Berlin, 1982.

[26] D. W. Stroock. *Lectures on Topics in Stochastic Differential Equations*. Tata Institute of Fundamental Research, Bombay, 1982. With notes by Satyajit Karmakar.

[27] A.M. Stuart and A.R. Humphries, *Dynamical Systems and Numerical Analysis*. Cambridge University Press, 1996.

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

[29] D. Talay, *Approximation of upper Lyapunov exponents of bilinear stochastic differential systems.* SIAM J. Num. Anal. **28**(1991), 1141–1164.

[30] D. Talay, *Approximation of the invariant probability measure of stochastic Hamiltonian dissipative systems with non globally Lipschitz co-efficients.* Appears in "Progress in Stochastic Structural Dynamics", Volume 152, 1999, Editors R. Bouc and C. Soize. Publication du L.M.A.-CNRS.

[31] M.M. Tropper, *Ergodic properties and quasideterministic properties of finite-dimensional stochastic systems.* J. Stat. Phys. **17**(1977), 491–509.

[32] A. Y. Veretennikov. *On polynomial mixing bounds for stochastic differential equations.* Stochastic Process. Appl., **70**(1997),115–127.

A Appendix: Proof of Theorem 2.5

Our proof of Theorem 2.5 proceeds in two steps. In STEP 1, relying on Assumption 2.1 (and its consequence the minorization condition), we use a standard construction to find a chain, equivalent in law to $P_t(x, A)$, which makes explicit some uniform behavior.

In STEP 2 we use a Lyapunov function to show that the chain repeatedly returns to a region in which the uniform behavior is valid. Together the two steps give ergodicity.

STEP 1 It is straightforward to see that Assumption 2.1 gives the following lemma.

Lemma A.1 *Let Assumption 2.1 hold. Then there is a $t_2 \in \mathbb{T}$ and a $\delta' > 0$ such that*

- $\mathcal{B}_{\delta'}(y^*) \subseteq C$ *and*
- $P_{t_2}\Big(y^*, \mathcal{B}_{\delta'}(y^*)\Big) > 0.$

We now derive the *minorization condition* on C, the basic conclusion of Lemma 2.3, which is used to characterize and quantify the uniform motion on the set C. Minorization essentially means that the Markov Chain restricted to C satisfies the classical Doeblin condition. The general theory of Markov chains (see [19], [16]) proceeds by use of a deep result which shows that the minorization condition can be satisfied for some sampled version of $\{x(t)\}$, given irreducibility. However, under our assumptions, which are natural for certain dynamical systems perturbed by noise, we can deduce the minorization condition directly. Since this gives rise to more transparent proofs and builds intuition, it is the approach we take here.

Recall that we study the Markov chain $\{x_n\}$ formed by sampling at the rate $T \in \mathbb{T}$, with the kernel $P(x, A) = P_T(x, A)$.

Proof (**Lemma 2.3**) Lemma $A.1$ implies that $P_{t_2}(y^*, \mathcal{B}_{\delta'}(y^*)) > 0$ and it follows from the existence of a density that there is a $z^* \in \mathcal{B}_{\delta'}(y^*) \subseteq C$ such that for some $\epsilon > 0$

$$p_{t_2}(y^*, z^*) \geq 2\epsilon > 0.$$

By Assumption 2.1(ii), there exist $\epsilon_1, \epsilon_2 > 0$ such that

$$p_{t_2}(y, z) \geq \epsilon > 0 \text{ for all } y \in \mathcal{B}_{\epsilon_1}(y^*) \text{ and } z \in \mathcal{B}_{\epsilon_2}(z^*).$$

By reducing ϵ_2 if necessary, we may ensure that $\mathcal{B}_{\epsilon_2}(z^*) \subset C$.

Now

$$P_{t_2}(y, A) = \int_A p_{t_2}(y, z)dz \geq \int_{A \cap \mathcal{B}_{\epsilon_2}(z^*)} p_{t_2}(y, z)dz \geq \epsilon\lambda(A \cap \mathcal{B}_{\epsilon_2}(z^*))$$

for all $y \in \mathcal{B}_{\epsilon_1}(y^*)$. (Here $\lambda(\cdot)$ is Lebesgue measure on $\mathbb{R}^d$).

By Assumption 2.1(i), we know that t_1 may be chosen so that $P_{t_1}(x, \mathcal{B}_{\epsilon_1}(y^*)) > 0$ for any $x \in C$. The continuity of $p_{t_1}(\cdot, y)$ given by Assumption 2.1(ii) can be transfered to $P_{t_1}(\cdot, A)$ by dominated convergence. Hence we have, since C is compact,

$$\inf_{x \in C} P_{t_1}(x, \mathcal{B}_{\epsilon_1}(y^*)) \geq \gamma$$

for some $\gamma > 0$. Now let $T = t_1 + t_2$. Then, for all $x \in C$,

$$\begin{aligned} P(x, A) = P_T(x, A) &\geq \int_{\mathcal{B}_{\epsilon_1}(y^*)} p_{t_1}(x, y)P_{t_2}(y, A)dy \\ &\geq \epsilon\lambda(A \cap \mathcal{B}_{\epsilon_2}(z^*)) \int_{\mathcal{B}_{\epsilon_1}(y^*)} p_{t_1}(x, y)dy \\ &= \epsilon\lambda(A \cap \mathcal{B}_{\epsilon_2}(z^*))P_{t_1}(x, \mathcal{B}_{\epsilon_1}(y^*)) \\ &\geq \epsilon\gamma\lambda(\mathcal{B}_{\epsilon_2}(z^*))\nu(A), \end{aligned}$$

where $\nu(\cdot)$ is Lebesgue measure restricted to $\mathcal{B}_{\epsilon_2}(z^*)$ and normalized to be a probability measure. If $\eta = \epsilon\gamma\lambda(\mathcal{B}_{\epsilon_2}(z^*))$ then we have $P(x, A) \geq \eta\nu(A)$ for all $A \in \mathcal{B}(\mathbb{R}^d)$ and $x \in C$. Since $\mathcal{B}_{\epsilon_2}(z^*) \subset C$, we have $\nu(C^c) = 0$ and $\nu(C) = \nu(\mathcal{B}_{\epsilon_2}(z^*)) = 1$, as required. □

We now use the preceding lemma to build an equivalent Markov chain where the uniform part of the motion is explicit. Recall that we study the Markov chain $\{x_n\}$ formed by sampling at the rate $T \in \mathbb{T}$, with the kernel $P(x, A) = P_T(x, A)$. Because our objective is primarily the study of a noisy dynamical system, we present our proof of ergodicity using random iterated functions [10], although we emphasize that this approach is not necessary; it is possible to work entirely with transition kernels.

We assume that the original chain $\{x_n\}_{x\in\mathbb{Z}^+}$, with kernel $P(x, A)$, is generated by

$$x_{n+1} = h(x_n, w_n), \tag{A.1}$$

with x_0 given. Here the $\omega_n \in \Omega$ are i.i.d. random variables and we have $\mathbb{P}\{x_{n+1} \in A \,|\, x_n\} = P(x_n, A)$. Now define a new transition kernel

$$\tilde{P}(x, A) = \begin{cases} P(x, A) & \forall x \in C^c, \\ \frac{1}{1-\eta}[P(x, A) - \eta\nu(A)] & \forall x \in C. \end{cases} \tag{A.2}$$

(Note that the minorization condition ensures that this kernel is well-defined). We may assume that $\tilde{P}$ is generated by iteration of the random family

$$\tilde{h}(\tilde{x}, \tilde{w}) \text{ with } \tilde{x} \in \mathbb{R}^d, \quad \tilde{\omega} \in \tilde{\Omega}, \tag{A.3}$$

again appealing to the construction in [10]. Then we define the new Markov chain

$$x'_{n+1} = h'(x'_n, \omega'_n). \tag{A.4}$$

Here $\omega'_n \in \Omega'$ are i.i.d. random variables defined below. The function h' is defined by

$$h'(x', \omega') = \mathbf{1}_C(x')[\phi\tilde{h}(x', \tilde{\omega}) + (1 - \phi)\xi] + [1 - \mathbf{1}_C(x')]\tilde{h}(x', \tilde{\omega}) \ . \tag{A.5}$$

The random variable ω'_1 is distibuted as $\omega' = (\tilde{\omega}, \phi, \xi)$ where $\tilde{\omega}, \phi$, and ξ are independent and $\tilde{\omega}$ is distributed as for (A.3), $\mathbb{P}(\phi = 1) = 1 - \eta$, $\mathbb{P}(\phi = 0) = \eta$, and ξ is distributed as ν. Straightforward calculations show that $\mathbb{P}(x'_{n+1} \in A \,|\, x_n) = P(x, A)$ so that (A.1) and (A.4) are equivalent in law.

The advantage of working with (A.4) is that it contains an atom-like structure: if any two independent realizations of the chain lie in C at a time n, and if $\phi_n = 0$ for both realizations, then both chains pick their next value at random according to the law of ξ_1 and hence the laws of the random variables given by sampling either chain at any time after n are the same.

To prove ergodicity we will use a coupling argument and hence, simultaneously, we consider a second copy of this chain whose noise is constructed to be advantageously correlated with that of the x' chain, namely

$$y'_{n+1} = h'(y'_n, \eta'_n), \qquad \eta'_n = (\tilde{W}_n, \phi_n, \xi_n).$$

Here the ϕ_n and ξ_n are the same random variables used to construct ω'_n. The $\tilde{W}_n$ are a new i.i.d sequence distributed in the same way as, but independently from, the $\tilde{\omega}_n$. Notice that the x'_n and y'_n dynamics are independent until $x'_n, y'_n \in C$ and $\phi_n = 0$. Then they both move to ξ_n. This is the key feature of this construction. When $\phi_n = 0$, the entire set C acts as an atom. Movement out of C is uniform irrespective of the point in C. Notice also that, if $P(x, C) = 1$, then the marginals of x and y will converge towards each other exponentially fast because the chance of not coupling is $(1 - \eta)^n$; such issues are discussed for Monte-Carlo Markov-chain techniques in [22].

Step 2 It is hopefully intuitively reasonable that, if the assumption $P(x, C) = 1$ is removed and instead it is simply assumed that the chain spends a lot of time in the set C, then the distributions of the two chains will still converge exponentially. We will make this precise shortly. First we develop the estimates that show that the chain visits C regularly.

We use the Lyapunov function from Assumption 2.2 to control the return times to C. Straightforward calculation shows that this assumption implies the following, at times more computationally useful, condition.

Lemma A.2 *Let Assumption 2.2 hold and let $\gamma \in (\alpha, 1), s \in [1, \infty)$. If*

$$c(s) = \frac{s\beta}{\gamma - \alpha}, \quad C(s) = \{x : V(x) \leq c(s)\}$$

then

$$\mathbb{E}[V(x_{n+1})|\mathcal{F}_n] \leq \gamma V(x_n) + s\beta \mathbf{1}_{C(s)}(x_n). \tag{A.6}$$

In the following we let $c = c(2)$ and $C = C(2)$. Furthermore we use κ to denote a constant independent of initial data for the Markov chain under consideration and independent of any time index $n, k \ldots$ etc. However the actual value of κ may change from occurence to occurence. The following amounts to the Optional Stopping lemma adapted to our setting. Since it is short we include the proof for completeness. It is the key estimate needed to complete the ergodic result.

Lemma A.3 *Let N be any stopping time and fix an $n \geq 0$. Under Assumption 2.2,*

$$\mathbb{E}\left\{V(x_n)\mathbf{1}_{N>n}\right\} \leq \mathbb{E}\left\{V(x_n)\mathbf{1}_{N\geq n}\right\} \leq \kappa\gamma^n \left[V(x_0) + \mathbb{E}\{\sum_{j=1}^{n\wedge N} \gamma^{-j}\mathbf{1}_C(x_{j-1})\}\right] \leq \frac{\kappa[\gamma^n V(x_0) + 1]}{1-\gamma}.$$

At first glance this lemma may seem rather technical. However it gives immediately that the return time to the set C has exponential tails and, with some appeals to the standard theory, the existence of an invariant measure. We defer the proof of the lemma until after these two useful corollaries.

Corollary A.4 *Assume the conditions of Lemma A.3 hold. If $\tau_C = \inf\{n > 0 : x_n \in C\}$ then for $n > 0$ and $\gamma \in (\alpha, 1)$ it follows that*

$$\mathbb{P}\{\tau_C > n\} \leq \kappa\gamma^n[V(x_0) + 1]$$

and

$$\mathbb{E}\left(\frac{1}{\gamma}\right)^{\tau_C} \leq \kappa[V(x_0) + 1].$$

Proof (**Corollary A.4**) The definition of τ_C implies that

$$\sum_{j=1}^{n\wedge\tau_C} \gamma^{n-j}\mathbf{1}_C(x_{j-1}) = \gamma^{n-1}\mathbf{1}_C(x_0);$$

also

$$\mathbb{E}\left\{V(x_n)\mathbf{1}_{\tau_C>n}\right\} \geq c\mathbb{E}\left\{\mathbf{1}_{\tau_C>n}\right\} = c\mathbb{P}\{\tau_C > n\}.$$

Using these estimates in Lemma A.3 gives the first result. For the second result notice that

$$\begin{aligned}\mathbb{E}\left(\frac{1}{\gamma}\right)^{\tau_C} &= \sum_{n=1}^{\infty}\left(\frac{1}{\gamma}\right)^n \mathbb{P}(\tau_c = n) \\ &\leq \sum_{n=1}^{\infty}\left(\frac{1}{\gamma}\right)^n \mathbb{P}(\tau_c > n-1).\end{aligned}$$

Since $\gamma \in (\alpha, 1)$ we can employ the first result with $\gamma \to \gamma' \in (\alpha, \gamma)$ to give the desired estimate. □

Corollary A.5 *Under Assumption 2.2, the system possesses an invariant probability measure.*

Proof (**Corollary A.5**) If we take the deterministic stopping time $N = n$ then Lemma A.3 implies that $\sup_{n\geq 0} \mathbb{E}\{V(x_n)\} < \infty$. Fixing an x_0, Chebychev's inequality tells us that the measures defined by

$$\mu_n(A) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{k=0}^{n} \mathbb{P}\big\{x_k \in A\big\}$$

are a tight sequence of measures since the level sets of V bound compact subsets of phase space. Hence once can extract a subsequence which converges to an invariant measure. See [10, 16] for more details. Since the total mass of each μ_n is bounded by one, any limiting measure will be finite and hence can be normalized into a probability measure. □

Proof (**Lemma A.3**) Begin by noticing that the third inequality follows from the second because

$$\sum_{j=1}^{n\wedge N} \gamma^{n-j}\mathbf{1}_C(x_j) \leq \sum_{j=1}^{n} \gamma^{n-j} \leq \frac{1}{1-\gamma}\ .$$

The first inequality holds because $V(x_n)\mathbf{1}_{N>n} \leq V(x_n)\mathbf{1}_{N\geq n}$ for every realization. To see the second claim, define $F(x,n) = \gamma^{-n}V(x)$ and observe that

$$\begin{aligned} F(x_{N\wedge n}, N\wedge n) =& F(x_0,0) + \sum_{j=1}^{N\wedge n} \big[F(x_j,j) - F(x_{j-1},j-1)\big] \\ =& F(x_0,0) + \sum_{j=1}^{n} \mathbf{1}_{N>j-1}\big[F(x_j,j) - F(x_{j-1},j-1)\big]. \end{aligned}$$

Since the event $\{N > j-1\} \in \mathcal{F}_{j-1}$, $F(x_0,0) = V(x_0)$, and

$$\mathbb{E}\big\{F(x_j,j)\big|\mathcal{F}_{j-1}\big\} \leq \gamma^{-j}[\gamma V(x_{j-1}) + 2\beta\mathbf{1}_C(x_{j-1})] = F(x_{j-1},j-1) + 2\gamma^{-j}\beta\mathbf{1}_C(x_{j-1})$$

we have

$$\begin{aligned} \mathbb{E}F(x_{N\wedge n}, N\wedge n) =& V(x_0) + \mathbb{E}\sum_{j=1}^{n} \mathbf{1}_{N>j-1}\mathbb{E}\big\{F(x_j,j) - F(x_{j-1},j-1)\big|\mathcal{F}_{j-1}\big\} \\ \leq& V(x_0) + 2\beta\mathbb{E}\sum_{j=1}^{n} \gamma^{-j}\mathbf{1}_{N>j-1}\mathbf{1}_C(x_{j-1}). \end{aligned}$$

Now observe that $\mathbb{E}F(x_{N\wedge n}, N\wedge n) = \mathbb{E}\big\{\gamma^{-n}V(x_n)\mathbf{1}_{n\leq N}\big\} + \mathbb{E}\big\{\gamma^{-N}V(x_N)\mathbf{1}_{n>N}\big\}$. Since V is positive we can neglect the second of the terms to obtain

$$\mathbb{E}\big\{V(x_n)\mathbf{1}_{n\leq N}\big\} \leq \gamma^n\mathbb{E}F(x_{N\wedge n}, N\wedge n) \leq \gamma^n V(x_0) + 2\beta\mathbb{E}\sum_{j=1}^{n\wedge N} \gamma^{n-j}\mathbf{1}_C(x_{j-1})$$

as required. □

With the estimates of Steps I and II, we are now ready to attack the principle result of this section.

Proof (**Theorem 2.5**) We abuse notation and take $\mathcal{F}_n$ to be the σ-algebra generated by both the $\{x'_n\}_{n\geq 0}$ and $\{y'_n\}_{n\geq 0}$ chains simultaneously. In the following $\mathbb{E}$ with no superscript denotes expectation for the product chain $\{(x'_n, y'_n)\}$ with possibly random data (x'_0, y'_0). Any test function f can be decomposed into two nonnegative functions f^+ and f^- with disjoint support so that $f = f^+ - f^-$. Thus

$$|\mathbb{E}f(x'_n) - \mathbb{E}f(y'_n)| \leq \big|\mathbb{E}f^+(x'_n) - \mathbb{E}f^+(y'_n)\big| + \big|\mathbb{E}f^-(x'_n) - \mathbb{E}f^-(y'_n)\big|\,.$$

We will deal with the two terms on the right hand side simultaneously.

Define the coupling time by

$$\zeta = \inf_{n\geq 0}\{(x'_n, y'_n) \in C \times C, \phi_n = 0\}.$$

Observe that

$$\mathbb{E}f^{\pm}(x'_n) = \mathbb{E}f^{\pm}(x'_n)\mathbf{1}_{n\geq\zeta} + \mathbb{E}f^{\pm}(x'_n)\mathbf{1}_{n<\zeta}$$

and since $\mathbb{E}f^{\pm}(x'_n)\mathbf{1}_{n\geq\zeta} = \mathbb{E}f^{\pm}(y'_n)\mathbf{1}_{n\geq\zeta} \leq \mathbb{E}f^{\pm}(y'_n)$ and $f^{\pm} \leq V$ we obtain

$$\mathbb{E}f^{\pm}(x'_n) \leq \mathbb{E}f^{\pm}(y'_n) + \mathbb{E}V(x'_n)\mathbf{1}_{n<\zeta}.$$

Reversing the roles of x'_n and y'_n produces a second inequality which when combined with the first yields

$$\left|\mathbb{E}f^{\pm}(x'_n) - \mathbb{E}f^{\pm}(y'_n)\right| \leq \max\left\{\mathbb{E}V(x'_n)\mathbf{1}_{n<\zeta}, \mathbb{E}V(y'_n)\mathbf{1}_{n<\zeta}\right\}$$

and hence

$$|\mathbb{E}f(x'_n) - \mathbb{E}f(y'_n)| \leq 2\max\left\{\mathbb{E}V(x'_n)\mathbf{1}_{n<\zeta}, \mathbb{E}V(y'_n)\mathbf{1}_{n<\zeta}\right\}. \tag{A.7}$$

The next lemma, proved at the end of the section, gives the desired control of the right hand side of the above inequality.

Lemma A.6 *In the setting of Theorem 2.5, for any $\gamma \in (\alpha^{\frac{1}{2}}, 1)$ there exists $r \in (0,1)$ so that*

$$\max\left\{\mathbb{E}V(x'_n)\mathbf{1}_{n<\zeta}, \mathbb{E}V(y'_n)\mathbf{1}_{n<\zeta}\right\} \leq \kappa\left[\mathbb{E}(V(x_0) + V(y_0)) + 1\right] r^n.$$

Note that γ enters the result through the definition of C, and hence ζ. Using this estimate, we conclude the proof of Theorem 2.5. To obtain convergence to the invariant measure, we start the y' chain with an invariant distribution π. Then $\mathbb{E}f(y'_n) = \int f(y)d\pi(y) \stackrel{\text{def}}{=} \pi(f)$ for all n. We have from (A.7) and Lemma A.6, starting with product measure $\delta_{x_0} \times \pi$ on the (x', y') chain,

$$|\mathbb{E}^{x_0}f(x_n) - \pi(f)| = |\mathbb{E}f(x'_n) - \mathbb{E}f(y'_n)| \leq 2\kappa\left[V(x_0) + \pi(V) + 1\right] r^n.$$

Since $\pi(V) < \infty$, the result follows.

We conclude this section with the proof of the Lemma A.6 which is the heart of the proof of Theorem 2.5.

Proof (**Lemma A.6**) Instead of determining when both x'_n and y'_n are in C directly, we define a new Lyapunov function to control $V(x'_n)$ and $V(y'_n)$ simultaneously. Set $V'(x,y) = V(x) + V(y)$. If the original chain satisfies Assumption 2.2 then

$$\mathbb{E}\left[V'(x_{n+1}, y_{n+1})\middle|\mathcal{F}_n\right] \leq \alpha V'(x_n, y_n) + 2\beta \tag{A.8}$$

where, recall, $\mathcal{F}_n$ now refers to the $\sigma-$algebra of events up to the nth for the product chain. Hence Lemma A.2 with $s = 1$ and $V \to V'$ implies, for any $\gamma \in (\alpha, 1)$,

$$\mathbb{E}\left[V'(x_{n+1}, y_{n+1})\middle|\mathcal{F}_n\right] \leq \gamma V'(x_n, y_n) + 2\beta\mathbf{1}_{C'}((x_n, y_n))$$

where

$$C' = \left\{(x,y) : V'(x,y) \leq \frac{2\beta}{\gamma - \alpha}\right\}.$$

Clearly if $(x'_n, y'_n) \in C'$ then $(x'_n, y'_n) \in C \times C$. Let

$$\zeta' = \inf_{n\geq 0}\left\{(x'_n, y'_n) \in C', \phi_n = 0\right\},$$

noting that $\zeta \le \zeta'$. Also observe that

$$\begin{aligned}\max\left\{\mathbb{E}V(x'_n)\mathbf{1}_{n<\zeta}, \mathbb{E}V(y'_n)\mathbf{1}_{n<\zeta}\right\} &\le \mathbb{E}V(x'_n)\mathbf{1}_{n<\zeta} + \mathbb{E}V(y'_n)\mathbf{1}_{n<\zeta} \\ &= \mathbb{E}V'(x'_n, y'_n)\mathbf{1}_{n<\zeta} \\ &\le \mathbb{E}V'(x'_n, y'_n)\mathbf{1}_{n<\zeta'}.\end{aligned}$$

Intuitively there are two types of trajectories contributing to $\mathbb{E}V'(x'_n, y'_n)\mathbf{1}_{n<\zeta'}$: those which spend a typical amount of time in C' and those that manage not to visit C' often. The first type contribute little when n is large because it is unlikely that $\phi_j \neq 0$ for all of the visits to C'. This is the same reasoning used when one has the simple Doeblin condition. The other paths contribute little to the expectation when n is large because, by Corollary A.4, it is unlikely that a trajectory stays out of C' for very long. We now make these ideas more precise.

Let τ_k be the time of the kth visit to C'. For notational convenience we define τ_s for any real s by $\tau_s = \tau_{\lceil s \rceil}$ and define $\tau_0 = 0$. Fixing an $a \in (0,1)$, we split $\mathbb{E}V'(x'_n, y'_n)\mathbf{1}_{n<\zeta'}$ into two terms as follows:

$$\begin{aligned}\mathbb{E}V'(x'_n, y'_n)\mathbf{1}_{n<\zeta'} =& \mathbb{E}V'(x'_n, y'_n)\mathbf{1}_{n<\zeta'}\mathbf{1}_{\tau_{an}<n} + \quad \mathbb{E}V'(x'_n, y'_n)\mathbf{1}_{n<\zeta'}\mathbf{1}_{\tau_{an}\ge n} \\ =& \mathbb{E}V'(x'_n, y'_n)\mathbf{1}_{n<\zeta'}\mathbf{1}_{\tau_{an}<n} + \sum_{k=0}^{\lceil an\rceil - 1} \mathbb{E}V'(x'_n, y'_n)\mathbf{1}_{n<\zeta'}\mathbf{1}_{\tau_k<n}\mathbf{1}_{\tau_{k+1}\ge n} \\ =& \qquad (I) \qquad + \qquad (II) \end{aligned} \tag{A.9}$$

Here $n \ge 1$. The first term represents typical behavior, in terms of the number of returns to C', when a is small enough; here we rely on the chance of coupling to dominate. The second term corresponds to unusual behavior of the trajectories and hence will be small. In the following it is convenient to define

$$\overline{V} \stackrel{\text{def}}{=} \sup_{(x',y')\in C'} V'(x', y'), \quad V'_0 \stackrel{\text{def}}{=} V'(x'_0, y'_0).$$

For (I) note that, by Lemma A.3,

$$\begin{aligned}\mathbb{E}V'(x'_n, y'_n)\mathbf{1}_{\tau_{an}<n}\mathbf{1}_{n<\zeta'} &\le \mathbb{E}V'(x'_n, y'_n)\mathbf{1}_{\tau_{an}<n}\mathbf{1}_{\tau_{an}<\zeta'} \\ &= \mathbb{E}\left\{V'(x'_n, y'_n)\mathbf{1}_{\tau_{an}<\zeta'}\middle|\tau_{an}<n\right\}\mathbb{P}\{\tau_{an}<n\} \\ &= \mathbb{E}\left\{\mathbf{1}_{\tau_{an}<\zeta'}\mathbb{E}\left\{V'(x'_n, y'_n)\middle|\tau_{an}<n, \mathcal{F}_{\tau_{an}}\right\}\middle|\tau_{an}<n\right\}\mathbb{P}\{\tau_{an}<n\} \\ &\le \mathbb{E}\left\{\mathbf{1}_{\tau_{an}<\zeta'}\kappa[\bar{V}+1]\middle|\tau_{an}<n\right\}\mathbb{P}\{\tau_{an}<n\} \\ &= \kappa[\bar{V}+1]\mathbb{E}\left\{\mathbf{1}_{\tau_{an}<\zeta'}\mathbf{1}_{\tau_{an}<n}\right\} \le \kappa[\bar{V}+1]\mathbb{E}\left\{\mathbf{1}_{\tau_{an}<\zeta'}\right\} \\ &\le \kappa[\bar{V}+1](1-\eta)^{an}\end{aligned}$$

For (II) let $\gamma \in (\alpha^{\frac{1}{2}}, 1)$ so that $\gamma^2 \in (\alpha, 1)$. For $k = 0$ we have, by Lemma A.3,

$$\mathbb{E}V'(x'_n, y'_n)\mathbf{1}_{n<\zeta'}\mathbf{1}_{\tau_0<n}\mathbf{1}_{\tau_1\ge n} \le \mathbb{E}V'(x'_n, y'_n)\mathbf{1}_{\tau_1\ge n} \le \kappa\gamma^n V'_0. \tag{A.10}$$

For $k \ge 1$, again using Lemma A.3,

$$\begin{aligned}\mathbb{E}V'(x'_n, y'_n)\mathbf{1}_{\tau_{k+1}\ge n}\mathbf{1}_{\tau_k<n}\mathbf{1}_{n<\zeta'} \le& \mathbb{E}V'(x'_n, y'_n)\mathbf{1}_{\tau_{k+1}\ge n}\mathbf{1}_{\tau_k<n}\mathbf{1}_{\tau_k<\zeta'} \\ =& \mathbb{E}\left\{\mathbf{1}_{\tau_k<\zeta'}\mathbb{E}\{V'(x'_n, y'_n)\mathbf{1}_{\tau_{k+1}\ge n}|\mathcal{F}_{\tau_k}, \tau_k<n\}\middle|\tau_k<n\right\}\mathbb{P}\left\{\tau_k<n\right\} \\ \le& \mathbb{E}\left\{\mathbf{1}_{\tau_k<n}\mathbf{1}_{\tau_k<\zeta'}\kappa\gamma^{n-\tau_k}[\overline{V}+1]\right\} \\ \le& \gamma^n\kappa[\overline{V}+1]\{\mathbb{E}\mathbf{1}_{\tau_k<\zeta'}\mathbb{E}\gamma^{-2\tau_k}\}^{\frac{1}{2}} \\ \le& \gamma^n\kappa[\overline{V}+1](1-\eta)^{k/2}\{\mathbb{E}\gamma^{-2\tau_k}\}^{\frac{1}{2}}.\end{aligned}$$

Now

$$\begin{aligned}\mathbb{E}\gamma^{-2\tau_k} =& \mathbb{E}\gamma^{-\sum_{l=1}^k 2(\tau_l-\tau_{l-1})} \\ =& \mathbb{E}\prod_{l=1}^k\left\{\left(\frac{1}{\gamma^2}\right)^{(\tau_l-\tau_{l-1})}\right\} \\ \stackrel{\text{def}}{=}& P_k\end{aligned}$$

Corollary A.4 gives, by conditioning on $\mathcal{F}_{\tau_{k-1}}$ and since $\gamma^2 \in (\alpha, 1)$,

$$P_k \leq \kappa[\overline{V} + 1]P_{k-1}.$$

As $P_1 \leq \kappa[V_0' + 1]$ it follows that

$$\mathbb{E}\gamma^{-2\tau_k} = P_k \leq \kappa^k[\overline{V} + 1]^{k-1}[V_0' + 1].$$

Combining terms produces, for $k \geq 1$ and some $R \geq 1$,

$$\mathbb{E}V'(x_n', y_n')\mathbf{1}_{\tau_{k+1}\geq n}\mathbf{1}_{\tau_k<n}\mathbf{1}_{n<\zeta'} \leq (1-\eta)^{k/2}R^k[V_0' + 1]^{\frac{1}{2}}\gamma^n \leq (1-\eta)^{k/2}R^k\sqrt{2}V_0'\gamma^n \tag{A.11}$$

since $1 + x \leq 2x^2$ for all $x \geq 1$.

With the estimates (A.10), (A.11) in hand we turn to term (II), obtaining

$$\begin{aligned}(II) = &\sum_{k=0}^{\lceil an\rceil-1} \mathbb{E}V'(x_n', y_n')\mathbf{1}_{n<\zeta'}\mathbf{1}_{\tau_k<n}\mathbf{1}_{\tau_{k+1}\geq n}\\ \leq &\sqrt{2}V_0' \sum_{k=0}^{\lceil an\rceil-1} (1-\eta)^{k/2}\gamma^n R^k\\ \leq &\sqrt{2}V_0'\gamma^n R^{an}\sum_{k=0}^{\infty}(1-\eta)^{k/2} \leq \sqrt{2}V_0'\gamma^n R^{an}\frac{1}{\eta'}\end{aligned}$$

where $1 - \eta' = \sqrt{(1-\eta)}$.

Combining our estimates of (I) and (II), we obtain the desired result since $\gamma \in (0, 1)$, and we may choose a sufficiently small so that $\gamma R^a < 1$.