

Pathwise Accuracy and Ergodicity of Metropolized Integrators for SDEs

NAWAF BOU-RABEE
Courant Institute

ERIC VANDEN-EIJNDEN
Courant Institute

Abstract

Metropolized integrators for ergodic stochastic differential equations (SDEs) are proposed that (1) are ergodic with respect to the (known) equilibrium distribution of the SDEs and (2) approximate pathwise the solutions of the SDEs on finite-time intervals. Both these properties are demonstrated in the paper, and precise strong error estimates are obtained. It is also shown that the Metropolized integrator retains these properties even in situations where the drift in the SDE is nonglobally Lipschitz, and vanilla explicit integrators for SDEs typically become unstable and fail to be ergodic. © 2009 Wiley Periodicals, Inc.

1 Introduction

Purpose

This paper considers SDEs whose solution is ergodic with respect to a known probability distribution. The paper analyzes integrators for such SDEs that

- (1) are ergodic with respect to the exact equilibrium distribution of the SDE on infinite time intervals, and
- (2) strongly converge to the solutions of the SDE on finite time intervals.

Pure sampling methods can accomplish (1), but they typically do not approximate the solution to the SDE. Integrators for SDEs certainly satisfy (2), but they often are divergent on infinite-time intervals or ergodic with respect to a different equilibrium distribution [16, 20, 31, 33]. This paper shows that a Metropolized integrator can simultaneously accomplish these goals.

Motivation

This paper is motivated by SDEs that arise in molecular dynamics (MD). Although the numerical analysis of general SDEs is well studied (see, e.g., [20, 31]), the relevance of this analysis to SDEs that arise in MD is limited. Indeed, such SDEs are characterized by complex drift vector fields that have limited regularity, and one is typically interested in their solutions over long time intervals. Let us elaborate briefly on the consequences of these two observations.

The drift vector field of SDEs that arise in MD involve the gradient of an empirically derived potential energy function. Due to singularities in the potential energy, this drift vector field possesses limited regularity; e.g., it is nonglobally Lipschitz. The drift vector field is also computationally costly to evaluate because the potential force involves intricate short- and long-range interactions between atoms. As a result, implicit methods are typically too costly in this application area.

Moreover, the lack of regularity in the drift of SDEs that arise in MD causes explicit discretizations to be stochastically unstable in general. This implies that, in principle, such schemes cannot be used to sample the equilibrium distribution of the SDE. Some approaches to stabilize explicit discretizations of SDEs include

- adaptive time stepping [13],
- multiple time stepping [35],
- method of rejecting exploding trajectories [21], and
- using bounded random numbers instead of Gaussian ones.

These methods are often ergodic but typically introduce discretization errors in the sampling (i.e., the equilibrium distribution of the numerical scheme is not exactly the same as the distribution of the SDE). A standard way to remove these errors is to resort to a Monte Carlo method designed for sampling, e.g., hybrid Monte Carlo algorithms [6, 11]. This procedure, however, has the effect that the dynamics of the method becomes unrelated to that of the SDE. Indeed, in this context, the main question often becomes how to make the sampling as efficient as possible by accelerating the rate of convergence of the method.

Here we show that a Metropolis-adjusted explicit integrator can both be ergodic with respect to the known equilibrium distribution of the SDE and capture the dynamical behavior of the solutions of this SDE. To establish this second property, one needs to estimate the effect on the dynamics of the rejections in the Metropolis-Hasting method. The analysis of this effect in the context of a Metropolized integrator for SDEs with the special structure relevant to MD is one of the main objectives of this paper. For a recent application of the techniques presented in this paper to constant-temperature molecular dynamics, the reader is referred to [3].

Organization of the Paper

In Section 2, the main results of the paper are presented without proofs and put into context. In Section 3, numerical validation of the main results is provided. The proofs of the main results are organized according to the type of dynamic and discretization considered. In Section 4, the proofs pertaining to a forward Euler-Maruyama discretization of overdamped Langevin dynamics and its Metropolis adjustment can be found. In Section 5, the proofs pertaining to a Störmer-Verlet-based discretization of inertial Langevin dynamics and its Metropolis adjustment can be found. A conclusion is provided in Section 6 followed by an appendix.

2 Main Results

2.1 Overdamped Langevin

In the first part of the paper, we shall focus on overdamped Langevin dynamics on an energy landscape defined by a potential energy function $U \in C^\infty(\mathbb{R}^n, \mathbb{R})$:

$$(2.1) \quad dY = -\nabla U(Y)dt + \sqrt{2\beta^{-1}} dW.$$

Here $\nabla U(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ denotes the gradient of the function U , W is a standard n -dimensional Wiener process, or Brownian motion, and $\beta > 0$ is a parameter referred to as the inverse temperature. Under certain regularity conditions on the potential energy, the solution to (2.1) is geometrically ergodic with an invariant probability measure μ that possesses the following density $\pi(\mathbf{x})$ with respect to Lebesgue measure [8, 25, 26]:

$$(2.2) \quad \pi(\mathbf{x}) = Z^{-1} \exp(-\beta U(\mathbf{x}))$$

where $Z = \int_{\mathbb{R}^n} \exp(-\beta U(\mathbf{x}))d\mathbf{x}$.

Next we recall various integration strategies for (2.1) and summarize their properties.

Forward Euler-Maruyama

Let N and h be given, set $T = Nh$ and $t_k = hk$ for $k = 0, \dots, N$, and consider the following explicit Euler-Maruyama discretization of (2.1):

$$(2.3) \quad \tilde{X}_{k+1} = \tilde{X}_k - h\nabla U(\tilde{X}_k) + \sqrt{2\beta^{-1}}(W(t_{k+1}) - W(t_k))$$

where \tilde{X}_k is the numerical approximation of $Y(t_k)$. The iteration rule (2.3) defines a Markov chain that possesses a transition kernel with the following smooth, strictly positive probability transition density:

$$(2.4) \quad q_h(\mathbf{x}, \mathbf{y}) = (4\pi\beta^{-1}h)^{-n/2} \exp\left(-\frac{|\mathbf{y} - \mathbf{x} + h\nabla U(\mathbf{x})|^2}{4\beta^{-1}h}\right).$$

Hence the chain is irreducible with respect to Lebesgue measure. To be consistent with the literature, the discrete-time Markov chain generated by Euler-Maruyama is called the unadjusted Langevin algorithm (ULA) [26].

If ∇U is globally Lipschitz and h is small enough, ULA (2.3) can often be shown to be

- first-order strongly convergent to the solution orbits of (2.1) on finite-time intervals, and
- geometrically ergodic on infinite-time intervals with respect to an invariant measure that is a first-order approximant to the invariant measure μ of (2.1).

The second property is typically established using a Talay-Tubaro expansion of the global weak error of ULA [33].

On the other hand, when ∇U is nonglobally Lipschitz, ULA is typically a transient Markov chain for any $h > 0$ [19, 22]. In fact, all moments of Euler-Maruyama are unbounded on long-time intervals for any initial condition $\mathbf{x} \in \mathbb{R}^n$, i.e., for any integer $\ell \geq 1$ and for any $h > 0$

$$(2.5) \quad \mathbb{E}^{\mathbf{x}} \{ |\tilde{X}_k|^{2\ell} \} \rightarrow \infty \quad \text{as } k \rightarrow \infty$$

where $\mathbb{E}^{\mathbf{x}}$ denotes the expectation conditional on $\tilde{X}_0 = \mathbf{x}$. See, e.g., [16, 32]. As is well-known in the literature, a Metropolis-Hastings method can stochastically stabilize ULA.

Metropolized Forward Euler-Maruyama

A Metropolis-Hastings method is a quite general Monte Carlo method for sampling from a known probability distribution [9, 18]. The method generates a Markov chain from a given proposal distribution as follows: The algorithm computes a proposed move according to the proposal distribution and accepts this move with a probability that ensures the Metropolized chain is ergodic with respect to the given probability distribution. Here we shall focus on the Metropolized Euler-Maruyama integrator defined in terms of the equilibrium density π (2.2) and the transition density q_h (2.4).

Let $\zeta_k \sim U(0, 1)$ for $k = 0, \dots, N-1$. Given h and X_k the algorithm calculates a proposed move using the Euler-Maruyama updating scheme in (2.3),

$$(2.6) \quad X_{k+1}^* = X_k - h\nabla U(X_k) + \sqrt{2\beta^{-1}}(\mathbf{W}(t_{k+1}) - \mathbf{W}(t_k)),$$

and accepts this proposal with probability

$$(2.7) \quad \alpha_h(\mathbf{x}, \mathbf{y}) = 1 \wedge \frac{q_h(\mathbf{y}, \mathbf{x})\pi(\mathbf{y})}{q_h(\mathbf{x}, \mathbf{y})\pi(\mathbf{x})}.$$

That is, the update is defined as

$$(2.8) \quad X_{k+1} = \begin{cases} X_{k+1}^* & \text{if } \zeta_k < \alpha_h(X_k, X_{k+1}^*), \\ X_k & \text{otherwise,} \end{cases}$$

for $k = 0, \dots, N-1$. To be consistent with the literature, we will refer to the Metropolized Euler-Maruyama integrator as the Metropolis-adjusted Langevin algorithm (MALA) [26]. By construction, MALA preserves the invariant measure μ of (2.1). If ∇U is globally Lipschitz and h small enough, one can often show that MALA is geometrically ergodic [26]. See, e.g., [26, theorem 4.1]. However, if ∇U is nonglobally Lipschitz, MALA is often no longer geometrically ergodic. Still, for any $g : \mathbb{R}^n \rightarrow \mathbb{R}$ with $\mu(g) < \infty$,

$$(2.9) \quad \mathbb{E}_\mu \mathbb{E}^{\mathbf{x}} \{ g(X_k) \} = \int_{\mathbb{R}^n} g d\mu \quad \forall k \in \mathbb{N}$$

where $\mathbb{E}_\mu \mathbb{E}^{\mathbf{x}}$ denotes expectation conditioned on the initial distribution being the equilibrium distribution of (2.1), i.e.,

$$\mathbb{E}_\mu \mathbb{E}^{\mathbf{x}} \{g(\mathbf{X}_k)\} = \int_{\mathbb{R}^n} \mathbb{E}^{\mathbf{x}} \{g(\mathbf{X}_k)\} \mu(d\mathbf{x}).$$

The identity (2.9) is obviously a significant improvement to (2.5).

When ∇U is nonglobally Lipschitz, MALA is not geometrically ergodic because ULA is transient (see theorem 4.2 of [26]). To correct this problem, Roberts and Tweedie introduced a modification of MALA that truncates the drift in regions where the underlying Euler method can be explosive. Specifically, the proposed move (2.6) is modified to a MALA algorithm with bounded drift:

$$(2.10) \quad \mathbf{Z}_{k+1}^* = \mathbf{Z}_k - h \frac{\nabla U(\mathbf{Z}_k)}{1 \vee h |\nabla U(\mathbf{Z}_k)|} + \sqrt{2\beta^{-1}(\mathbf{W}(t_{k+1}) - \mathbf{W}(t_k))}.$$

When $|\nabla U(\mathbf{Z}_k)| < \frac{1}{h}$, (2.10) is the Euler-Maruyama update, and otherwise, the proposed drift in (2.10) preserves the direction of the drift in ULA but normalizes the amplitude in all degrees of freedom. The Metropolis-Hastings method with this modified proposed move will be referred to as the Metropolis-adjusted Langevin truncated algorithm (MALTA) [26]. In comparison to a random-walk-based Metropolis algorithm, one can often show that MALTA is geometrically ergodic [1, 12, 26]. While this ad hoc correction to MALA makes MALTA geometrically ergodic, the relation of MALTA to the original diffusion (2.1) remains to be established.

Main Result I: Pathwise Convergence of MALA and MALTA

In the paper we shall consider the situation where ∇U is nonglobally Lipschitz. Roughly speaking, MALTA corrupts Euler-Maruyama orbit by the random rejections in (2.8) with the modified proposed (2.10) and the ad hoc truncation of the drift. Nevertheless, we show in the paper that MALTA approximates pathwise the solution of (2.1). The precise statement is as follows:

THEOREM 2.1 (MALTA Pathwise Accuracy) *Let the assumptions in Section 4.1 hold. Then for every $E_0 > 0$ and $T > 0$, there exists $h_c(E_0) > 0$ and $C(T, E_0) > 0$ such that for all $h < h_c$, for all \mathbf{x} , $U(\mathbf{x}) \leq E_0$, and for all $t \in [0, T]$,*

$$(\mathbb{E}^{\mathbf{x}} \{|\mathbf{Z}_{\lfloor t/h \rfloor} - \mathbf{Y}(t)|^2\})^{1/2} \leq C(T, E_0)h^{3/4}.$$

This assumption remains valid for SDEs of the type (2.1) in which the drift is nonglobally Lipschitz. A key ingredient in the proof is geometric ergodicity of MALTA. This property implies bounds on moments of MALTA at finite times. The precise bounds can be found in Lemma 4.11.

Since MALA is not geometrically ergodic when ∇U is nonglobally Lipschitz, we are unable to obtain such bounds on moments of MALA at finite times. However, such bounds can be derived if MALA’s initial condition is restricted to the

equilibrium distribution of (2.1). This is because MALA by design preserves this distribution. Hence for MALA we are able to prove the following theorem. We stress that MALA's pathwise accuracy does not require this restriction on initial conditions.

THEOREM 2.2 (MALA Pathwise Accuracy) *Let the assumptions in Section 4.1 hold. For all $T > 0$ there exists $h_c > 0$ and $C(T) > 0$ such that for all positive $h < h_c$ and for all $t \in [0, T]$,*

$$(\mathbb{E}_\mu \mathbb{E}^{\mathbf{x}} \{|X_{\lfloor t/h \rfloor} - Y(t)|^2\})^{1/2} \leq C(T)h^{3/4}.$$

The proofs of Theorems 2.1 and 2.2 as well as the precise statements of the assumptions on which they rely are presented in Section 4.1. The proofs build on results of global accuracy for numerical methods applied to SDEs with uniformly Lipschitz drift [20] and with one-sided Lipschitz drift [10]. Section 3 will illustrate these results on a simple but representative example with a nonglobally Lipschitz drift.

We remark that if ∇U is globally Lipschitz the rate of convergence of MALA is still $h^{3/4}$. Under this strong assumption on the drift in (2.1), geometric ergodicity of MALA can be attained without truncating the drift, and hence one can extend Theorem 2.2 to any initial condition.

2.2 Inertial Langevin

Next we shall consider Langevin equations defined in terms of a Hamiltonian function $H \in C^\infty(\mathbb{R}^{2n}, \mathbb{R})$:

$$(2.11) \quad d\mathbf{Y} = \mathbb{J} \nabla H(\mathbf{Y}) dt - \gamma \mathbf{C} \nabla H(\mathbf{Y}) dt + \sqrt{2\gamma\beta^{-1}} \mathbf{C} d\mathbf{W}$$

where the following matrices have been introduced:

$$\mathbb{J} = \begin{bmatrix} 0 & \mathbf{I} \\ -\mathbf{I} & 0 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{I} \end{bmatrix}.$$

Here \mathbf{W} is a standard $2n$ -dimensional Wiener process, or Brownian motion, $\beta > 0$ is a parameter referred to as the inverse temperature, and $\gamma > 0$ is referred to as the friction coefficient. Write $\mathbf{Y}(t) = (\mathbf{Q}(t), \mathbf{P}(t))$ where $\mathbf{Q}(t)$ and $\mathbf{P}(t)$ represent the instantaneous configuration and momentum of the system, respectively. We shall assume

$$H(\mathbf{q}, \mathbf{p}) = \frac{1}{2} \mathbf{p}^\top \mathbf{M}^{-1} \mathbf{p} + U(\mathbf{q}),$$

where \mathbf{M} is a symmetric, positive definite mass matrix and U is a potential energy function.

There are two main differences between the overdamped and inertial Langevin dynamics. First, the solution of (2.1) is a reversible stochastic process, whereas the solution of (2.11) is not. Second, the diffusion in (2.11) is only applied to momentum degrees of freedom, whereas the diffusion in (2.1) is applied to all

degrees of freedom. The consequences of these differences to the discretization of (2.11) and its Metropolis adjustment will be the main focus of Section 5.

Despite the degenerate diffusion in (2.11), under certain regularity conditions on U , the solution to this SDE is geometrically ergodic with respect to an invariant probability measure μ with the following density [32]:

$$(2.12) \quad \pi(\mathbf{q}, \mathbf{p}) = Z^{-1} \exp(-\beta H(\mathbf{q}, \mathbf{p})),$$

where $Z = \int_{\mathbb{R}^{2n}} \exp(-\beta H(\mathbf{q}, \mathbf{p})) d\mathbf{q} d\mathbf{p}$.

Geometric Langevin Algorithm

Let N and h be given, and set $T = Nh$ and $t_k = hk$ for $k = 0, \dots, N$. We shall consider an integrator for (2.11) based on splitting the Langevin equations into Hamilton’s equations for the Hamiltonian H :

$$(2.13) \quad \begin{cases} d\mathbf{Q} = \mathbf{M}^{-1} \mathbf{P} dt \\ d\mathbf{P} = -\nabla U(\mathbf{Q}) dt \end{cases}$$

and Ornstein-Uhlenbeck equations

$$(2.14) \quad \begin{cases} d\mathbf{Q} = 0 \\ d\mathbf{P} = -\gamma \mathbf{M}^{-1} \mathbf{P} dt + \sqrt{2\beta^{-1}\gamma} d\mathbf{W}. \end{cases}$$

The solution of Hamilton’s equations will be approximated by a symplectic integrator, to be specific, the discrete Hamiltonian map of a self-adjoint variational integrator [15], while the exact flow will be used for the Ornstein-Uhlenbeck equations. These flows will be composed in a Strang-type splitting to obtain a pathwise approximant to the solution of inertial Langevin equations, which we will refer to as the geometric Langevin algorithm (GLA).

This type of splitting of inertial Langevin equations is quite natural and has recently been used in simulations of molecular dynamics (see [5, 37]), dissipative particle dynamics (see [28, 29]), and inertial particles (see [23]). This paper is geared towards applications in molecular dynamics where Langevin integrators (including the ones cited above) have been based on generalizations of the widely used Störmer-Verlet integrator. The Störmer-Verlet integrator is attractive for molecular dynamics because it is an explicit, symmetric, second-order accurate, variational integrator for Hamilton’s equations. In molecular dynamics it was popularized by Loup Verlet in 1967. Other popular generalizations of the Störmer-Verlet integrator to Langevin equations include Brünger, Brooks, and Karplus (BBK) [4], van Gunsteren and Berendsen (vGB) [36], and the Langevin impulse (LI) methods [30]. The LI method is also based on a splitting of Langevin equations, but it is different from the splitting considered here. The long-time statistical properties of GLA were recently analyzed in the context of uniformly Lipschitz potential forces in [2].

An example of a GLA that will be analyzed in Section 5 is the Störmer-Verlet-based scheme:

$$(2.15) \quad \begin{cases} \tilde{\mathbf{Q}}_{k+1} = \tilde{\mathbf{Q}}_k + h\mathbf{M}^{-1}e^{-\gamma\mathbf{M}^{-1}h/2}\tilde{\mathbf{P}}_k - \frac{h^2}{2}\mathbf{M}^{-1}\nabla U(\tilde{\mathbf{Q}}_k) \\ \quad + h\sqrt{2\beta^{-1}\gamma}\int_{t_k}^{t_k+h/2}\mathbf{M}^{-1}e^{-\gamma\mathbf{M}^{-1}(t_k+h/2-s)}d\mathbf{W}(s), \\ \tilde{\mathbf{P}}_{k+1} = e^{-\gamma\mathbf{M}^{-1}h}\tilde{\mathbf{P}}_k - \frac{h}{2}e^{-\gamma\mathbf{M}^{-1}h/2}(\nabla U(\tilde{\mathbf{Q}}_k) + \nabla U(\tilde{\mathbf{Q}}_{k+1})) \\ \quad + \sqrt{2\beta^{-1}\gamma}\int_{t_k}^{t_k+h}e^{-\gamma\mathbf{M}^{-1}(t_k+h-s)}d\mathbf{W}(s). \end{cases}$$

It is straightforward to show (2.15) possesses a smooth probability transition function with respect to Lebesgue measure even though the noise is only applied to momenta in (2.14). For globally Lipschitz potential forces, it is possible to show that (2.15) is a first-order, strongly accurate integrator for (2.11) and is often geometrically ergodic with respect to an invariant measure that is a first-order accurate approximation to the exact invariant measure of (2.11) [2]. However, if the potential force is nonglobally Lipschitz, then this discretization is plagued with the same transient behavior as forward Euler-Maruyama applied to the reversible Langevin diffusion process. As before, a Metropolis-Hastings method is proposed to correct the discretization error in the invariant measure and stochastically stabilize this discretization.

MAGLA

Let $\zeta_k \sim U(0, 1)$ for $k = 0, \dots, N-1$. To Metropolize GLA (5.5), we adopt the method of Metropolizing an inertial Langevin integrator introduced in [27]. In that paper, the authors present and test this Metropolization technique using a potpourri of integrators including the Ricci-Ciccotti algorithm as a proposed move [24]. The method presented in Scemama et al. involves the involution $\varphi : \mathbb{R}^{2n} \rightarrow \mathbb{R}^{2n}$:

$$\varphi(\mathbf{q}, \mathbf{p}) = (\mathbf{q}, -\mathbf{p}).$$

This involution is introduced since the stochastic process defined by composing the solution of (2.11) with φ is reversible with respect to μ .

The Metropolis-adjusted geometric Langevin algorithm (MAGLA) will be based on the probability transition density of GLA, which we will refer to as q_h . An explicit formula for this transition density is derived in the body of the paper (see (5.6)). Given h and $(\mathbf{Q}_k, \mathbf{P}_k)$, MAGLA computes a proposed move $(\mathbf{Q}_{k+1}^*, \mathbf{P}_{k+1}^*)$ according to a step of GLA composed with a momentum flip. For example, for the Störmer-Verlet-based GLA, this proposed move is given explicitly by

$$(2.16) \quad \begin{cases} \mathbf{Q}_{k+1}^* = \mathbf{Q}_k + h\mathbf{M}^{-1}e^{-\gamma\mathbf{M}^{-1}h/2}\mathbf{P}_k - \frac{h^2}{2}\mathbf{M}^{-1}\nabla U(\mathbf{Q}_k) \\ \quad + h\sqrt{2\beta^{-1}\gamma}\int_{t_k}^{t_k+h/2}\mathbf{M}^{-1}e^{-\gamma\mathbf{M}^{-1}(t_k+h/2-s)}d\mathbf{W}(s), \\ \mathbf{P}_{k+1}^* = e^{-\gamma\mathbf{M}^{-1}h}\mathbf{P}_k - \frac{h}{2}e^{-\gamma\mathbf{M}^{-1}h/2}(\nabla U(\mathbf{Q}_k) + \nabla U(\mathbf{Q}_{k+1}^*)) \\ \quad + \sqrt{2\beta^{-1}\gamma}\int_{t_k}^{t_k+h}e^{-\gamma\mathbf{M}^{-1}(t_k+h-s)}d\mathbf{W}(s). \end{cases}$$

The algorithm accepts this proposed move with probability

$$(2.17) \quad \alpha_h((\mathbf{q}_0, \mathbf{p}_0), (\mathbf{q}_1, \mathbf{p}_1)) = 1 \wedge \frac{q_h((\mathbf{q}_1, \mathbf{p}_1), (\mathbf{q}_0, -\mathbf{p}_0))\pi(\mathbf{q}_1, \mathbf{p}_1)}{q_h((\mathbf{q}_0, \mathbf{p}_0), (\mathbf{q}_1, -\mathbf{p}_1))\pi(\mathbf{q}_0, \mathbf{p}_0)}.$$

In other words, the MAGLA update is defined as

$$(2.18) \quad \mathbf{X}_{k+1} := (\mathbf{Q}_{k+1}, \mathbf{P}_{k+1}) = \begin{cases} (\mathbf{Q}_{k+1}^*, \mathbf{P}_{k+1}^*) & \text{if } \zeta_k < \alpha_h((\mathbf{Q}_k, \mathbf{P}_k), (\mathbf{Q}_{k+1}^*, \mathbf{P}_{k+1}^*)), \\ \varphi(\mathbf{Q}_k, \mathbf{P}_k) & \text{otherwise,} \end{cases}$$

for $k = 0, \dots, N - 1$. Observe that when a proposed move is rejected, the momentum is “flipped,” and MAGLA involves a modified detailed balance condition (2.17) as opposed to the usual detailed balance condition used with MALA (2.7). The reason MAGLA uses a modified detailed balance is related to the non-reversibility of the solution to (2.11). In particular, the transition density of the solution to (2.11) does not satisfy detailed balance but does satisfy a modified detailed balance condition. It is straightforward to show that MAGLA preserves μ and hence is not transient. In fact, it is often easy to classify MAGLA as an ergodic Markov chain even if the potential force is nonglobally Lipschitz. In the next paragraph, we quantify the pathwise convergence of MAGLA.

Main Result II: Pathwise Convergence of MAGLA

It is straightforward to show that GLA provides a first-order globally accurate integrator for (2.11) provided the potential force is globally Lipschitz [2]. As discussed MAGLA involves a momentum flip whenever a proposed move is rejected. This momentum flip was introduced since the stochastic process defined by composing the solution of (2.11) with a momentum flip satisfies detailed balance with respect to π . Despite this momentum flip, MAGLA provides a pathwise approximant to (2.11) as summarized by the following theorem. As in Theorem 2.2, we will restrict the initial conditions of MAGLA to be distributed according to the equilibrium distribution of (2.11). As before, this assumption implies bounds on relevant moments of MAGLA that follow from the fact that MAGLA preserves the measure μ .

THEOREM 2.3 (Störmer-Verlet-Based MAGLA Pathwise Accuracy from Equilibrium) *Let the assumptions in Section 4.1 hold on the potential energy. Then for every $T > 0$, there exists $h_c > 0$ and $C(T) > 0$ such that for all $h < h_c$, for all $\mathbf{x} \in \mathbb{R}^{2n}$, and for all $t \in [0, T]$,*

$$(\mathbb{E}_\mu \mathbb{E}^{\mathbf{x}} \{|X_{\lfloor t/h \rfloor} - Y(t)|^2\})^{1/2} \leq C(T)h.$$

The assumptions in Section 4.1 on the potential energy hold even if potential force is nonglobally Lipschitz.

Unlike the rejections in (2.8), a rejection in (2.18) leads to a momentum flip and hence a loss of local accuracy, even though MAGLA still approximates pathwise

the solution to (2.11) because the probability of these rejections as a function of time step size is small. These issues, including a proof of Theorem 2.3 and the lemmas on which it is based, can be found in Section 5.

3 Numerical Illustration

3.1 Overdamped Langevin

Here we illustrate the results above on the simple example of a particle diffusing in a quartic potential $U(x) = x^4/4$ with inverse temperature $\beta > 0$. The overdamped dynamics of this system is given by

$$(3.1) \quad dY = -Y^3 dt + \sqrt{2\beta^{-1}} dW, \quad Y(0) = x.$$

This SDE admits a unique invariant measure with the following density:

$$\pi(x) = Z^{-1} \exp(-\beta x^4/4).$$

Let T and h be given. Set $N = \lfloor T/h \rfloor$ and $t_k = hk$ for $k = 0, \dots, N$. Consider the following forward Euler approximation to (3.1):

$$(3.2) \quad \tilde{X}_{k+1} = \tilde{X}_k - h\tilde{X}_k^3 + \sqrt{2\beta^{-1}}(W(t_{k+1}) - W(t_k)), \quad \tilde{X}_0 = x.$$

The drift in (3.2) is destabilizing in the region

$$B_h = \{x : |1 - hx^2| > 1\}.$$

This property of Euler-Maruyama's discrete drift is the essential reason that Euler-Maruyama defines a transient Markov chain. Indeed, using property (2.5) is proved in [16, lemma 63]. Despite this shortcoming, Euler-Maruyama can be used to generate a proposed move in a Metropolis-Hastings method designed to sample from π .

Figure 3.1 illustrates the difference between a given realization of the exact solution to (3.1), the Euler-Maruyama integrator (3.2), and the Metropolized integrator when initiated at the edge of B_h and driven by the same realization of the Wiener process W . We emphasize that in this figure the time step is held fixed and chosen to be large. The figure shows the Euler-Maruyama orbit explodes, the solution orbit rapidly contracts to the origin, and the Metropolis-Hastings method initially stagnates, but eventually tracks the solution apparently better over time. We remark that the ability of the Metropolized orbit to track the solution better over time is not a generic property of Metropolized integrators. In this case it is a consequence of (3.1) satisfying a "one force, one solution" principle (see [7]). That is, for every realization of the Wiener process, the solution possesses a random attractor that the Metropolized orbit is drawn to.

Figure 3.2 shows a longer-time realization of the Metropolized integrator and the solution. The solution at this temperature frequently visits higher energy values. The proposed moves to higher energy values are less likely to be accepted by the Metropolized integrator. At the same time, ergodicity implies the Metropolized

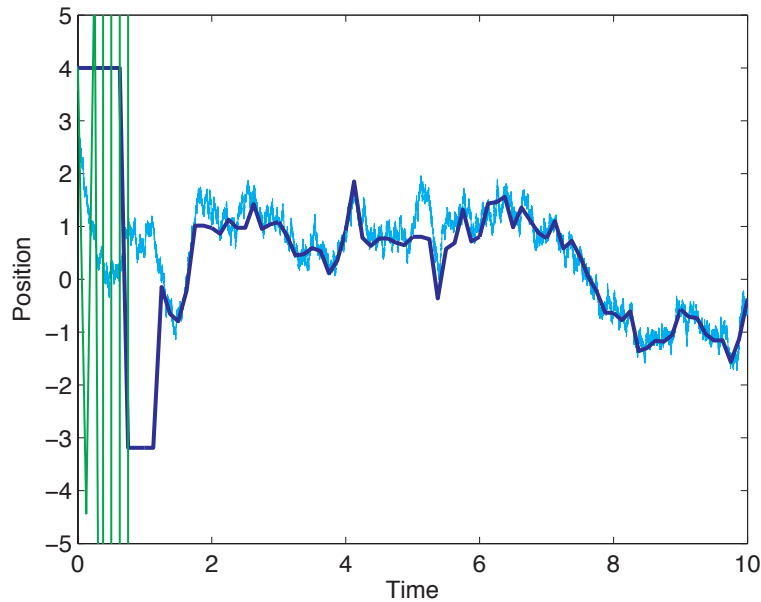


FIGURE 3.1. EXPLODING, STAGNATING, AND CONTRACTING ORBITS. For $h = 0.3125$, $\beta = 1.0$, and $x = 4$, this figure illustrates a realization of the solution to (3.1) (cyan), the Euler-Maruyama approximation (3.2) (green), and MALA (blue) initiated at the edge of B_h and driven by the same realization of the Wiener process W . Observe that Euler-Maruyama is explosive while the Metropolized orbit is not. Also observe that the position of the exact solution contracts rapidly while the Metropolized orbit stagnates before contracting due to proposed moves being rejected.

integrator must visit these higher energy values, and as often as the exact solution along this long time interval, since both chains preserve π . Consequently, the Metropolized integrator intermittently stagnates as shown in the figure, and of course, loses pathwise accuracy.

This observation does not contradict Theorems 2.1 and 2.2. Keep in mind that the time step size for this particular orbit is held fixed and chosen to be large. Such stagnations at high energy become less likely to occur along orbits on finite-time intervals as the time step becomes smaller.

Figures 3.3 and 3.4 confirm this pathwise convergence. In particular, the figures show an $\mathcal{O}(h^{3/4})$ rate of pathwise convergence of the Metropolized integrator for this example using MALTA and an out-of-equilibrium initial condition, and MALA with an initial condition restricted to the equilibrium distribution of (3.1).

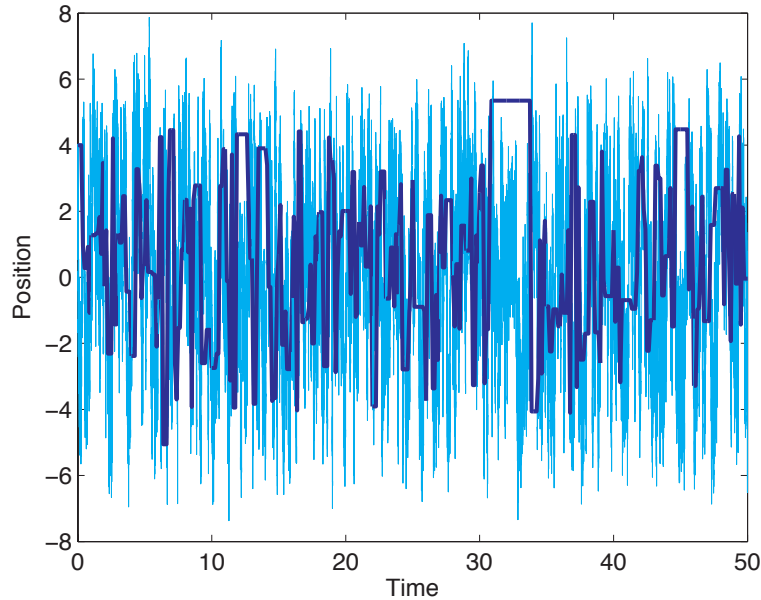


FIGURE 3.2. LONG-TIME BEHAVIOR OF MALA. For $h = 0.3125$, $\beta = 0.01$, and $x = 4.0$, a long-time realization of the solution to (3.1) (cyan) and MALA (blue). Observe that the solution frequently visits higher energy values that are not reached by the Metropolized integrator. But ergodicity requires that the Metropolized integrator visit these higher energy levels and for about the same ratio of time spent by the solution, since both chains sample from π . Hence the Metropolized integrator intermittently stagnates at higher energy values as shown in the figure.

3.2 Inertial Langevin

Consider a simple particle with Hamiltonian given by

$$H(q, p) = \frac{p^2}{2} + U(q)$$

where $U(q) = q^4/4$. The inertial Langevin dynamics of this system is given by

$$(3.3) \quad \begin{cases} dQ = P dt, \\ dP = -Q^3 dt - \gamma P dt + \sqrt{2\gamma\beta^{-1}} dW, \end{cases}$$

with initial condition $(Q(0), P(0)) = (q_0, p_0)$. This SDE admits a unique invariant measure with the following density:

$$\pi(q, p) = Z^{-1} \exp(-\beta H(q, p)).$$

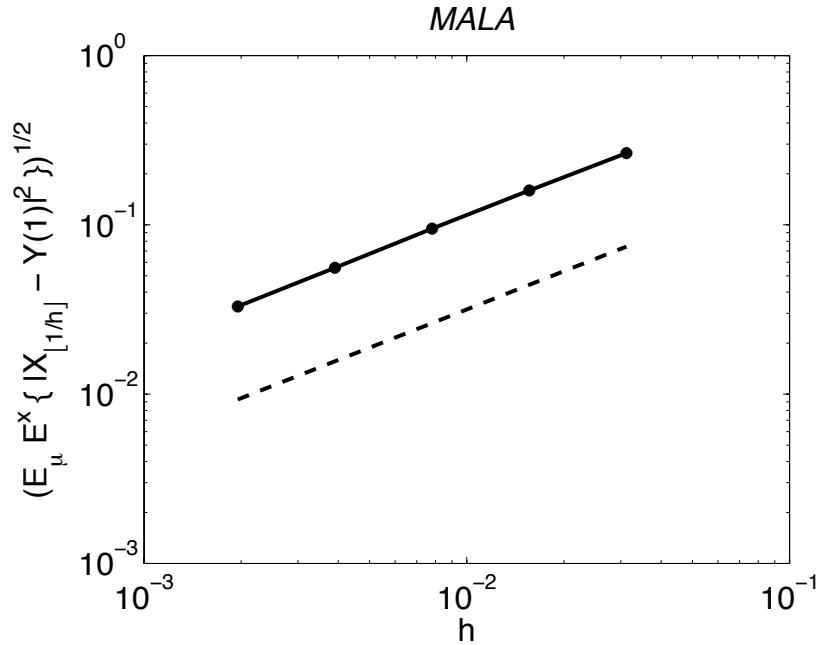


FIGURE 3.3. PATHWISE ACCURACY OF MALA. This log-log plot illustrates that despite random rejections of integration steps and the non-globally Lipschitz nature of the drift, MALA remains mean-squared and hence strongly convergent to solutions of the SDE (3.1). The dashed line represents a reference slope of $h^{3/4}$. The solid line represents an empirical estimate of the mean-squared order of accuracy of MALA for an inverse temperature $\beta = 1$, time span of $T = 1$, and an ensemble of initial conditions drawn from the equilibrium distribution of the SDE using inverse transform sampling. The estimated rate of convergence agrees with that reported in Theorem 2.2. A total of 10^6 realizations were used to obtain this estimate.

Let T and h be given. Set $N = \lfloor T/h \rfloor$ and $t_k = hk$ for $k = 0, \dots, N$. Consider the following Störmer-Verlet-based GLA discretization of (3.3):

$$(3.4) \quad \begin{cases} \tilde{Q}_{k+1} = \tilde{Q}_k + h e^{-\gamma h/2} \tilde{P}_k - \frac{h^2}{2} \tilde{X}_k^3 \\ \quad + \sqrt{2\beta^{-1}\gamma} \int_{t_k}^{t_k+h/2} e^{-\gamma(t_k+h/2-s)} dW(s), \\ \tilde{P}_{k+1} = e^{-\gamma h} \tilde{P}_k - \frac{h}{2} e^{-\gamma h/2} (\nabla U(\tilde{Q}_k) + \nabla U(\tilde{Q}_{k+1})) \\ \quad + \sqrt{2\beta^{-1}\gamma} \int_{t_k}^{t_k+h} e^{-\gamma(t_k+h-s)} dW(s) \end{cases}$$

with initial condition $(\tilde{Q}_0, \tilde{P}_0) = (q_0, p_0)$. The inertial Langevin integrator (3.4) is plagued with the same transient behavior as the Euler-Maruyama integrator applied to the reversible cubic oscillator dynamics. This occurs at points in phase

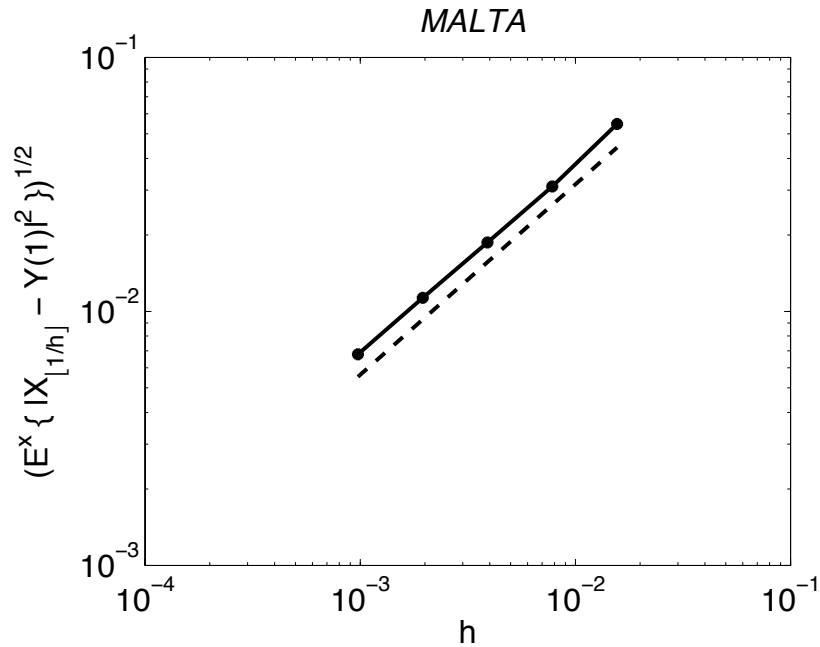


FIGURE 3.4. PATHWISE ACCURACY OF MALTA. This log-log plot illustrates that despite random rejections of integration steps and the ad hoc truncation of the drift, MALTA remains mean-squared, and hence strongly convergent to solutions of the SDE (3.1). The dashed line represents a reference slope of $h^{3/4}$. The solid line represents an empirical estimate of the mean-squared order of accuracy of MALTA for an inverse temperature $\beta = 1$, time span of $T = 1$, and initial condition $x = 0.1$. The estimated rate of convergence agrees with that reported in Theorem 2.1. A total of 10^6 realizations were used to obtain this estimate.

space where the underlying explicit Störmer-Verlet integrator in (3.4) becomes linearly unstable.

A Metropolis-Hastings method can stochastically stabilize (3.4). At the same time, as stated in Theorem 2.3, the method is also pathwise accurate with respect to the solution of (3.3) when initiated from equilibrium. This accuracy is achieved despite the loss of local accuracy in momentum that occurs at every rejection.

Figure 3.5 is consistent with the strong $\mathcal{O}(h)$ rate of convergence of MAGLA for this example.

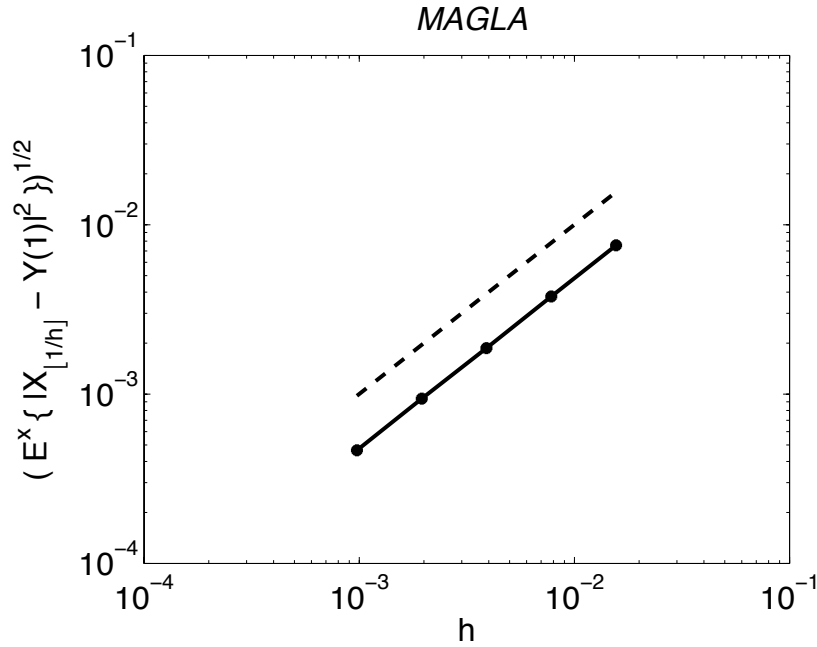


FIGURE 3.5. PATHWISE ACCURACY OF MAGLA. This plot illustrates that despite momentum flips that occur at every rejection of a proposed move and the nonglobally Lipschitz nature of the drift, MAGLA remains mean-squared and hence strongly convergent to solutions of the SDE (3.3). In this log-log plot, the dashed line represents a reference slope of h . The solid line shows a sample average of the mean-squared order of accuracy of MAGLA, for $T = 1$, $\beta = \gamma = 1$, and $(q_0, p_0) = (0.1, 0.0)$. The estimated rate of convergence is consistent with that reported in Theorem 2.3. A total of 10^5 realizations were used to obtain this plot.

4 Overdamped Langevin

4.1 Preliminaries

Structural Assumptions

For a function $G \in C^\infty(\mathbb{R}^n, \mathbb{R})$ and an integer $r > 1$, let ∇G and $D^r G$ be the gradient and the r^{th} derivative of G , respectively. Let $|\cdot|$ denote the Euclidean vector norm and $\|\cdot\|$ the Frobenius norm. Let L denote the generator of (2.1) defined for any $G \in C^2(\mathbb{R}^n, \mathbb{R})$ as

$$(4.1) \quad LG(\mathbf{x}) = -\nabla U(\mathbf{x}) \cdot \nabla G(\mathbf{x}) + \beta^{-1} \text{Tr}[D^2 G(\mathbf{x})].$$

Assumptions 4.1. The following structural assumptions will be invoked in this paper regarding (2.1):

- (A) There exists a real constant $K > 0$ such that

$$U(\mathbf{x}) \geq K|\mathbf{x}| \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

- (B) For every integer $\ell \geq 1$ there exist real constants $\delta_\ell > 0$ and $M_\ell > 0$ such that the operator (4.1) applied to the ℓ^{th} power of the potential energy satisfies

$$L\{U(\mathbf{x})^\ell\} \leq -\delta_\ell U(\mathbf{x})^\ell + M_\ell \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

- (C) There exists a real constant $K > 0$ such that

$$|\nabla U(\mathbf{x}) - \nabla U(\mathbf{y})| \leq K(U(\mathbf{x}) + U(\mathbf{y}))|\mathbf{x} - \mathbf{y}| \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

- (D) There exists a real constant $K > 0$ such that

$$\langle -\nabla U(\mathbf{x}) + \nabla U(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \leq K|\mathbf{x} - \mathbf{y}|^2 \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

- (E) There exists a real constant $K > 0$ such that

$$\|D^3 U(\mathbf{x})\| \vee \|D^2 U(\mathbf{x})\| \vee |\nabla U(\mathbf{x})| \leq K(1 + U(\mathbf{x})) \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

We stress that these structural assumptions hold even if the drift in (2.1) is non-globally Lipschitz. For example, the assumptions hold for potentials $U(x) = x^{2p}$ and $U(x) = x^{2p} - x^2$ for any positive integer $p > 0$ and, in particular, the cubic oscillator example treated in Section 3. These structural assumptions imply that the Markov process defined by the solution to (2.1) possesses a unique invariant probability measure explicitly given by

$$(4.2) \quad \mu(d\mathbf{x}) := \pi(\mathbf{x})d\mathbf{x}$$

where $\pi(\mathbf{x}) = Z^{-1} \exp(-\beta U(\mathbf{x}))$ with $Z = \int_{\mathbb{R}^n} \exp(-\beta U(\mathbf{x}))d\mathbf{x}$. These structural assumptions also imply the following estimates on higher moments of solutions to (2.1), which we state as a lemma.

LEMMA 4.2 (Estimates on Higher Moments of Solution) *Let the assumptions in Section 4.1 hold. For every integer $\ell \geq 1$ we have the following estimates on higher moments of the solution to (2.1):*

- (i) Given $M_\ell > 0$ and $\delta_\ell > 0$ from Assumption 4.1(B),

$$\mathbb{E}^{\mathbf{x}}\{U(\mathbf{Y}(t))^\ell\} \leq U(\mathbf{x})^\ell + \frac{M_\ell}{\delta_\ell}, \quad \forall \mathbf{x} \in \mathbb{R}^n, \forall t > 0.$$

- (ii) There exists $K_\ell > 0$ such that

$$\mathbb{E}^{\mathbf{x}}\{|\mathbf{Y}(t)|^\ell\} \leq K_\ell(1 + U(\mathbf{x})^\ell), \quad \forall \mathbf{x} \in \mathbb{R}^n, \forall t > 0.$$

- (iii) There exists $K_\ell > 0$ such that

$$\mathbb{E}^{\mathbf{x}}\{|\mathbf{Y}(t) - \mathbf{x}|^{2\ell}\} \leq K_\ell(1 + t^\ell U(\mathbf{x})^{2\ell})t^\ell, \quad \forall \mathbf{x} \in \mathbb{R}^n, \forall t > 0.$$

The proof of this lemma is presented in Section A.

The structural assumptions on (2.1) also imply a Lipschitz condition on its solution, which we state as another lemma.

LEMMA 4.3 (Regularity of Solutions) *Let the assumptions in Section 4.1 hold. For $s \leq t$, let $Y_{t,s}(\mathbf{x})$ denote the evolution operator of the solution to (2.1): with $Y_{s,s}(\mathbf{x}) = \mathbf{x}$ and for $r \leq s \leq t$ recall the Chapman-Kolmogorov identity $Y_{t,s} \circ Y_{s,r}(\mathbf{x}) = Y_{t,r}(\mathbf{x})$. Set*

$$\Delta = Y_{s+h,s}(\mathbf{x}) - Y_{s+h,s}(\mathbf{y}) - (\mathbf{x} - \mathbf{y}).$$

For all $K > 0$ there exists $h_c > 0$ such that for all positive $h < h_c$, for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, and for all $s \geq 0$,

$$\begin{aligned} \mathbb{E}\{|Y_{s+h,s}(\mathbf{x}) - Y_{s+h,s}(\mathbf{y})|^2\} &\leq |\mathbf{x} - \mathbf{y}|^2(1 + Kh), \\ \mathbb{E}\{|\Delta|^2\} &\leq Kh^2(1 + U(\mathbf{x})^2 + U(\mathbf{y})^2)|\mathbf{x} - \mathbf{y}|. \end{aligned}$$

The proof of this lemma is also given below in Section A.

Unadjusted Langevin Algorithm

Recall that the transition kernel for ULA (2.3) possesses a density $q_h(\mathbf{x}, \mathbf{y})$ with respect to the Lebesgue measure given by

$$(4.3) \quad q_h(\mathbf{x}, \mathbf{y}) = (4\pi\beta^{-1}h)^{-n/2} \exp\left(-\frac{|\mathbf{y} - \mathbf{x} + h\nabla U(\mathbf{x})|^2}{4\beta^{-1}h}\right).$$

It is derived by a simple change of variables relating the probability density of the Euler-Maruyama update to the probability density of a multivariate Gaussian with zero mean and variance $2\beta^{-1}h$. ULA is irreducible with respect to the Lebesgue measure because its transition density is smooth and strictly positive. However, as mentioned in the introduction, in Langevin diffusions with nonglobally Lipschitz drift, ULA is transient, i.e., stochastically unstable. Nevertheless, under the assumptions on the drift, it is straightforward to obtain the following single-step accuracy for the forward Euler-Maruyama scheme.

LEMMA 4.4 (Local Pathwise Accuracy of Euler-Maruyama) *Let the assumptions in Section 4.1 hold. For all $K > 0$ there exists an h_c such that for all positive $h < h_c$ and for all $\mathbf{x} \in \mathbb{R}^n$,*

(i) *the local mean-squared error of (2.3) satisfies*

$$\mathbb{E}^{\mathbf{x}}\{|\tilde{X}_1 - Y(h)|^2\} \leq K(1 + U(\mathbf{x})^4)h^3;$$

(ii) *the local mean deviation of (2.3) satisfies*

$$|\mathbb{E}^{\mathbf{x}}\{\tilde{X}_1 - Y(h)\}| \leq K(1 + U(\mathbf{x})^2)h^2.$$

The above lemma is proven in Section A.

4.2 MALA and Its Properties

Ergodicity

MALA randomly rejects integration steps produced by Euler-Maruyama with a probability designed to ensure that the composite chain preserves the equilibrium measure of (2.1). Recall that this probability is given by

$$(4.4) \quad \alpha_h(\mathbf{x}, \mathbf{y}) = 1 \wedge \frac{q_h(\mathbf{y}, \mathbf{x})\pi(\mathbf{y})}{q_h(\mathbf{x}, \mathbf{y})\pi(\mathbf{x})}.$$

The off-diagonal transition probability kernel of the composite chain has density

$$p_h(\mathbf{x}, \mathbf{y}) = q_h(\mathbf{x}, \mathbf{y})\alpha_h(\mathbf{x}, \mathbf{y}).$$

The probability of remaining at the same point, or *stagnation probability*, is given by

$$r_h(\mathbf{x}) = 1 - \int_{\mathbb{R}^n} p_h(\mathbf{x}, \mathbf{z})d\mathbf{z}.$$

In sum, the Metropolis-Hastings transition kernel is

$$(4.5) \quad P_h(\mathbf{x}, d\mathbf{y}) = p_h(\mathbf{x}, \mathbf{y})d\mathbf{y} + r_h(\mathbf{x})\delta_{\mathbf{x}}(d\mathbf{y}).$$

where $\delta(\cdot)$ denotes the Dirac delta measure in \mathbb{R}^n . It is straightforward to show that this P_h is a well-defined probability transition kernel. Let $\mathcal{B}(\mathbb{R}^n)$ denote the smallest σ -algebra containing all the open subsets of \mathbb{R}^n . For all $\mathbf{x} \in \mathbb{R}^n$ and $A \in \mathcal{B}(\mathbb{R}^n)$, the k -step transition probability kernel is iteratively defined as

$$P_h^k(\mathbf{x}, A) = \int_{\mathbb{R}^n} P_h(\mathbf{x}, d\mathbf{y})P_h^{k-1}(\mathbf{y}, A), \quad k \geq 1,$$

with $P^0(\mathbf{x}, A)$ equal to the characteristic function of the subset A of \mathbb{R}^n , i.e., $P^0(\mathbf{x}, A) = \chi_A(\mathbf{x})$.

THEOREM 4.5 (MALA Ergodicity) *Under the assumptions in Section 4.1 on the potential energy, the k -step transition probability of MALA converges to μ in the following sense:*

$$\sup_{A \in \mathcal{B}(\mathbb{R}^n)} |P_h^k(\mathbf{x}, A) - \mu(A)| \rightarrow 0 \quad \text{as } k \rightarrow \infty, \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

PROOF: The following classification of MALA is quite standard. Since π and q_h are strictly positive and smooth everywhere, MALA is irreducible with respect to Lebesgue measure and aperiodic; see, e.g., lemma 1.2 of [17] and references therein. By the design of the acceptance probability, MALA also preserves the

probability measure μ . To confirm this statement, observe from (4.5) that for any $A \in \mathcal{B}(\mathbb{R}^n)$

$$\begin{aligned} \int_{\mathbb{R}^n} \mu(d\mathbf{x}) P_h(\mathbf{x}, A) &= \int_{\mathbb{R}^n} \left(\int_A \pi(\mathbf{x}) p_h(\mathbf{x}, \mathbf{y}) d\mathbf{y} + \pi(\mathbf{x}) r_h(\mathbf{x}) \chi_A(\mathbf{x}) \right) d\mathbf{x} \\ &= \int_{\mathbb{R}^n} \int_A (\pi(\mathbf{x}) p_h(\mathbf{x}, \mathbf{y}) - \pi(\mathbf{y}) p_h(\mathbf{y}, \mathbf{x})) d\mathbf{y} d\mathbf{x} + \mu(A). \end{aligned}$$

Since $p_h(\mathbf{x}, \mathbf{y})$ satisfies detailed balance, the first term in the above vanishes, and hence,

$$\int_{\mathbb{R}^n} \mu(d\mathbf{x}) P_h(\mathbf{x}, A) = \mu(A).$$

According to corollary 2 of [34], a Metropolis-Hastings algorithm that is irreducible with respect to the same measure it is designed to preserve is positive Harris recurrent. Consequently, MALA is irreducible, aperiodic, and positive Harris recurrent. According to proposition 6.3 of [22], these properties are equivalent to ergodicity of the chain. \square

Remark 4.6. This theorem shows that MALA is ergodic with respect to the equilibrium measure of (2.1). However, the effect of rejections on the pathwise approximation of the solution remains to be quantified.

Pathwise Accuracy

In order to prove equilibrium pathwise accuracy of MALA the following lemmas will be helpful:

LEMMA 4.7 (MALA Stagnation Probability) *Assume 4.1. For all integers $\ell \geq 1$, there exists an $h_c > 0$ and a constant $K_\ell > 0$ such that for all positive $h < h_c$ and for all $\mathbf{x} \in \mathbb{R}^n$,*

$$\mathbb{E}^{\mathbf{x}} \{ (\alpha_h(\mathbf{x}, \mathbf{X}_1^*) - 1)^{2\ell} \} \leq K_\ell (1 + U(\mathbf{x})^{4\ell}) h^{3\ell},$$

where

$$\mathbf{X}_1^* = \mathbf{x} - h \nabla U(\mathbf{x}) + \sqrt{2\beta^{-1}h} \boldsymbol{\eta}.$$

and $\boldsymbol{\eta} \sim \mathcal{N}(0, 1)^n$.

PROOF: Let $R_h(\mathbf{x}) = \{ \mathbf{y} \in \mathbb{R}^n : \alpha_h(\mathbf{x}, \mathbf{y}) < 1 \}$. Then,

$$(4.6) \quad \mathbb{E}^{\mathbf{x}} \{ (\alpha_h(\mathbf{x}, \mathbf{X}_1^*) - 1)^{2\ell} \} = \int_{R_h(\mathbf{x})} \left(\frac{q_h(\mathbf{y}, \mathbf{x}) \pi(\mathbf{y})}{q_h(\mathbf{x}, \mathbf{y}) \pi(\mathbf{x})} - 1 \right)^{2\ell} q_h(\mathbf{x}, \mathbf{y}) d\mathbf{y}.$$

Introduce the function $G : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$,

$$\begin{aligned} G(\mathbf{x}, \mathbf{y}) &= U(\mathbf{y}) - U(\mathbf{x}) - \frac{1}{2} \langle \nabla U(\mathbf{y}) + \nabla U(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle \\ &\quad + \frac{h}{4} (|\nabla U(\mathbf{y})|^2 - |\nabla U(\mathbf{x})|^2). \end{aligned}$$

Using the expressions for q_h and π (cf. (4.3) and (4.2), respectively), one can show

$$\frac{q_h(\mathbf{y}, \mathbf{x})\pi(\mathbf{y})}{q_h(\mathbf{x}, \mathbf{y})\pi(\mathbf{x})} = \exp(-\beta G(\mathbf{x}, \mathbf{y})).$$

Hence (4.6) can be written as

$$(4.7) \quad \mathbb{E}^{\mathbf{x}}\{(\alpha_h(\mathbf{x}, \mathbf{X}_1^*) - 1)^{2\ell}\} = (4\pi\beta^{-1}h)^{-n/2} \int_{R_h(\mathbf{x})} (e^{-\beta G(\mathbf{x}, \mathbf{y})} - 1)^{2\ell} e^{\left(-\frac{|\mathbf{y}-\mathbf{x}+h\nabla U(\mathbf{x})|^2}{4\beta^{-1}h}\right)} d\mathbf{y}.$$

We will evaluate this integral by the Laplace method. Specifically, introduce the map $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^n$

$$\varphi(\boldsymbol{\xi}) = \mathbf{x} + \sqrt{2h\beta^{-1}} \boldsymbol{\xi} + h\nabla U(\mathbf{x}).$$

Set $\tilde{R}_h(\mathbf{x}) = \varphi^{-1}(R_h(\mathbf{x}))$. A change of variables of (4.7) under φ yields

$$(4.8) \quad \mathbb{E}^{\mathbf{x}}\{(\alpha_h(\mathbf{x}, \mathbf{X}_1^*) - 1)^{2\ell}\} = (2\pi)^{-n/2} \int_{\tilde{R}_h(\mathbf{x})} (e^{-\beta G(\mathbf{x}, \mathbf{x} + \sqrt{2h\beta^{-1}} \boldsymbol{\xi} + h\nabla U(\mathbf{x}))} - 1)^{2\ell} e^{-\frac{1}{2}|\boldsymbol{\xi}|^2} d\boldsymbol{\xi}.$$

For h sufficiently small the latter integral can be well approximated by a Taylor expansion as follows: A Taylor expansion of the function G about $h = 0$ yields

$$G(\mathbf{x}, \mathbf{x} + \sqrt{2h\beta^{-1}} \boldsymbol{\xi} + h\nabla U(\mathbf{x})) = \left(\frac{\sqrt{2\beta^{-1}}}{2} \langle \nabla U(\mathbf{x}), D^2 U(\mathbf{x}) \cdot \boldsymbol{\xi} \rangle - \frac{2\beta^{-3/2}}{3} D^3 U(\mathbf{x}) \cdot \boldsymbol{\xi}^3 \right) h^{3/2} + \mathcal{O}(h^2).$$

Substitute this expansion into (4.8) to obtain

$$\begin{aligned} \int_{\tilde{R}_h(\mathbf{x})} (e^{-\beta G(\mathbf{x}, \mathbf{x} + \sqrt{2h\beta^{-1}} \boldsymbol{\xi} + h\nabla U(\mathbf{x}))} - 1)^{2\ell} e^{-\frac{1}{2}|\boldsymbol{\xi}|^2} d\boldsymbol{\xi} = \\ \int_{\tilde{R}_h(\mathbf{x})} \left[h^{3\ell} \left(\frac{\sqrt{2\beta^{-1}}}{2} \langle \nabla U(\mathbf{x}), D^2 U(\mathbf{x}) \cdot \boldsymbol{\xi} \rangle - \frac{2\beta^{-3/2}}{3} D^3 U(\mathbf{x}) \cdot \boldsymbol{\xi}^3 \right)^{2\ell} \right. \\ \left. + \mathcal{O}(h^{3\ell+1}) \right] e^{-\frac{1}{2}|\boldsymbol{\xi}|^2} d\boldsymbol{\xi}. \end{aligned}$$

The Laplace method asserts that the $\mathcal{O}(h^{3\ell+1})$ term can be taken out of the integral, i.e.,

$$\begin{aligned} & \int_{\tilde{R}_h(\mathbf{x})} (e^{-\beta G(\mathbf{x}, \mathbf{x} + \sqrt{2h\beta^{-1}}\boldsymbol{\xi} + h\nabla U(\mathbf{x}))} - 1)^{2\ell} e^{-\frac{\beta}{4}|\boldsymbol{\xi}|^2} d\boldsymbol{\xi} = \\ & \int_{\mathbb{R}^n} h^{3\ell} \left(\frac{\sqrt{2\beta^{-1}}}{2} \langle \nabla U(\mathbf{x}), D^2 U(\mathbf{x}) \cdot \boldsymbol{\xi} \rangle - \frac{2\beta^{-3/2}}{3} D^3 U(\mathbf{x}) \cdot \boldsymbol{\xi}^3 \right)^{2\ell} e^{-\frac{1}{2}|\boldsymbol{\xi}|^2} d\boldsymbol{\xi} \\ & \qquad \qquad \qquad + \mathcal{O}(h^{3\ell+1}). \end{aligned}$$

The proof is completed by invoking Assumption 4.1(E). □

LEMMA 4.8 (Local Accuracy of MALA) *Let the assumptions in Section 4.1 hold. For all $h > 0$ there exists a $C > 0$ such that*

(i) *the local mean-squared error of MALA satisfies*

$$\mathbb{E}^{\mathbf{x}} \{ |X_1 - Y(h)|^2 \} \leq C(1 + U(\mathbf{x})^4)h^{5/2} \quad \forall \mathbf{x} \in \mathbb{R}^n;$$

(ii) *the local mean deviation of MALA satisfies*

$$|\mathbb{E}^{\mathbf{x}} \{ X_1 - Y(h) \}| \leq C(1 + U(\mathbf{x})^3)h^2 \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

PROOF:

(i) By the definition of MALA,

$$\begin{aligned} \mathbb{E}^{\mathbf{x}} \{ |X_1 - Y(h)|^2 \} &= \\ & \mathbb{E}^{\mathbf{x}} \{ |X_1^* - Y(h)|^2 \alpha_h(\mathbf{x}, X_1^*) + |\mathbf{x} - Y(h)|^2 (1 - \alpha_h(\mathbf{x}, X_1^*)) \}. \end{aligned}$$

Hence

$$(4.9) \quad \mathbb{E}^{\mathbf{x}} \{ |X_1 - Y(h)|^2 \} \leq \mathbb{E}^{\mathbf{x}} \{ |X_1^* - Y(h)|^2 \} + \mathbb{E}^{\mathbf{x}} \{ |\mathbf{x} - Y(h)|^2 (1 - \alpha_h(\mathbf{x}, X_1^*)) \}.$$

It is clear from this expression that the local mean-squared error of the Metropolis-Hastings integrator is due to the local mean-squared error of forward Euler-Maruyama and a term due to probable rejections in the Metropolis-Hastings step. By the Cauchy-Schwarz inequality this latter term can be bounded as follows:

$$\begin{aligned} \mathbb{E}^{\mathbf{x}} \{ |\mathbf{x} - Y(h)|^2 (\alpha_h(\mathbf{x}, X_1^*) - 1) \} &\leq \\ & (\mathbb{E}^{\mathbf{x}} \{ |\mathbf{x} - Y(h)|^4 \})^{1/2} \cdot (\mathbb{E}^{\mathbf{x}} \{ (\alpha_h(\mathbf{x}, X_1^*) - 1)^2 \})^{1/2}. \end{aligned}$$

Lemma 4.2 implies there exists a constant $K > 0$ such that

$$(\mathbb{E}^{\mathbf{x}} \{ |\mathbf{x} - Y(h)|^4 \})^{1/2} \leq K(1 + U(\mathbf{x})^2)h,$$

while Lemma 4.7 implies that there exists a constant $K > 0$ such that

$$(\mathbb{E}^{\mathbf{x}} \{ (\alpha_h(\mathbf{x}, X_1^*) - 1)^2 \})^{1/2} \leq K(1 + U(\mathbf{x})^2)h^{3/2}.$$

Thus, there exists a constant $K > 0$ such that

$$\mathbb{E}^{\mathbf{x}} \{ |\mathbf{x} - Y(h)|^2 (1 - \alpha_h(\mathbf{x}, X_1^*)) \} \leq K(1 + U(\mathbf{x})^4)h^{5/2}.$$

Observe that (4.9) is dominated by the error incurred when a proposed move is rejected, and not the $O(h^3)$ error incurred when a proposed move is accepted (cf. Lemma 4.4). Hence, the local mean-squared accuracy of MALA is $O(h^{5/4})$.

(ii) Similar to the proof of part (i),

$$\begin{aligned} \mathbb{E}^{\mathbf{x}}\{X_1 - Y(h)\} &= \\ & \mathbb{E}^{\mathbf{x}}\{(X_1^* - Y(h))\alpha_h(\mathbf{x}, X_1^*) + (\mathbf{x} - Y(h))(1 - \alpha_h(\mathbf{x}, X_1^*))\}. \end{aligned}$$

By the triangle inequality

$$\begin{aligned} |\mathbb{E}^{\mathbf{x}}\{X_1 - Y(h)\}| &\leq |\mathbb{E}^{\mathbf{x}}\{(X_1^* - Y(h))\alpha_h(\mathbf{x}, X_1^*)\}| \\ & \quad + |\mathbb{E}^{\mathbf{x}}\{(\mathbf{x} - Y(h))(1 - \alpha_h(\mathbf{x}, X_1^*))\}| \\ &\leq |\mathbb{E}^{\mathbf{x}}\{(X_1^* - Y(h))\}| \\ & \quad + |\mathbb{E}^{\mathbf{x}}\{(\mathbf{x} - Y(h))(1 - \alpha_h(\mathbf{x}, X_1^*))\}|. \end{aligned}$$

It is clear from this expression that the local mean deviation of the Metropolis-Hastings integrator is due to the local mean deviation of the forward Euler-Maruyama integrator and a term due to possible rejections in the Metropolis-Hastings step. By the Cauchy-Schwarz and Jensen inequalities,

$$\begin{aligned} |\mathbb{E}^{\mathbf{x}}\{X_1 - Y(h)\}|^2 &\leq 2|\mathbb{E}^{\mathbf{x}}\{(X_1^* - Y(h))\}|^2 \\ & \quad + 2|\mathbb{E}^{\mathbf{x}}\{(\mathbf{x} - Y(h))(1 - \alpha_h(\mathbf{x}, X_1^*))\}|^2 \\ &\leq 2|\mathbb{E}^{\mathbf{x}}\{(X_1^* - Y(h))\}|^2 \\ & \quad + 2(\mathbb{E}^{\mathbf{x}}\{|\mathbf{x} - Y(h)|^4\})^{1/2} \cdot (\mathbb{E}^{\mathbf{x}}\{(\alpha_h(\mathbf{x}, X_1^*) - 1)^4\})^{1/2}. \end{aligned}$$

As in part (i) of this proof, the local mean deviation of the forward Euler-Maruyama integrator (see Lemma 4.4) together with Lemma 4.2 and Lemma 4.7 imply that

$$|\mathbb{E}^{\mathbf{x}}\{X_1 - Y(h)\}|^2 \leq K(1 + F(\mathbf{x})^6)h^4. \quad \square$$

LEMMA 4.9 (Uniform-in-Time Bound on Error) *Let the assumptions in Section 4.1 hold. Then for all $h > 0$ and for all $t > 0$, there exists a real constant $C > 0$ such that*

$$\mathbb{E}_\mu \mathbb{E}^{\mathbf{x}}\{|X_{\lfloor t/h \rfloor} - Y(t)|^2\} \leq C.$$

PROOF: This bound is a consequence of ergodicity of the solution to (2.1) and MALA with respect to the probability measure μ . That is, the ergodic property of the integrator and the solution of (2.1) imply that

$$\begin{aligned} &\mathbb{E}_\mu \mathbb{E}^{\mathbf{x}}\{|X_{\lfloor t/h \rfloor} - Y(t)|^2\} \\ &\leq 2\mathbb{E}_\mu \mathbb{E}^{\mathbf{x}}\{|X_{\lfloor t/h \rfloor}\|^2\} + 2\mathbb{E}_\mu \mathbb{E}^{\mathbf{x}}\{|Y(t)|^2\} \\ &= 4 \int_{\mathbb{R}^n} |\mathbf{x}|^2 \mu(d\mathbf{x}). \end{aligned} \quad \square$$

We are now in a position to prove Theorem 2.2, which we restate.

THEOREM 2.2 (MALA Pathwise Accuracy) *Let the assumptions in Section 4.1 hold. For all $T > 0$ there exist $h_c > 0$ and $C(T) > 0$ such that for all positive $h < h_c$ and for all $t \in [0, T]$,*

$$(\mathbb{E}_\mu \mathbb{E}^{\mathbf{x}} \{|X_{\lfloor t/h \rfloor} - Y(t)|^2\})^{1/2} \leq C(T)h^{3/4}.$$

PROOF: Discretize the interval $[0, T]$ using N equally spaced points in such a fashion that $t_k = kh$ for $k = 0, \dots, N$ with $N = \lfloor T/h \rfloor$. Assume $h > 0$ but otherwise arbitrary for now. The goal in this proof is to obtain a global error estimate for the mean-squared error of MALA by writing the error at the $(k + 1)^{\text{th}}$ step in terms of the error at the k^{th} step. For clarity, we will use a rolling constant K throughout this proof. Set

$$\epsilon_{k+1} = \mathbb{E}_\mu \mathbb{E}^{\mathbf{x}} \{|X_{k+1} - Y_{t_{k+1},0}(\mathbf{x})|^2\}$$

for $k = 0, \dots, N - 1$. Specifically, the proof shows that there exists a constant $K > 0$ such that

$$(4.10) \quad \epsilon_{k+1} \leq (1 + Ah)\epsilon_k + Kh^{5/2}.$$

By unraveling this recursive inequality, the order $h^{3/4}$ mean-squared accuracy becomes apparent. To do this, the expectation of the error at the $(k + 1)^{\text{th}}$ step is conditioned on knowing all events up to the k^{th} step. For this purpose let \mathcal{F}_k denote the sigma-algebra of events up to and including the k^{th} iteration. Also write

$$\begin{aligned} |X_{k+1} - Y_{t_{k+1},0}(\mathbf{x})|^2 &= \\ &|X_{k+1} - Y_{t_{k+1},t_k}(X_k) + Y_{t_{k+1},t_k}(X_k) - Y_{t_{k+1},t_k}(Y_{t_k,0}(\mathbf{x}))|^2. \end{aligned}$$

Expanding this expression gives

$$(4.11) \quad \begin{aligned} &|X_{k+1} - Y_{t_{k+1},0}(\mathbf{x})|^2 \\ &= |X_{k+1} - Y_{t_{k+1},t_k}(X_k)|^2 \\ &\quad + |Y_{t_{k+1},t_k}(X_k) - Y_{t_{k+1},t_k}(Y_{t_k,0}(\mathbf{x}))|^2 \\ &\quad + 2\langle X_{k+1} - Y_{t_{k+1},t_k}(X_k), Y_{t_{k+1},t_k}(X_k) - Y_{t_{k+1},t_k}(Y_{t_k,0}(\mathbf{x})) \rangle. \end{aligned}$$

In what follows the iterated expectation of the three terms on the right-hand side of (4.11) will be bounded separately. These different bounds will then be combined to obtain a bound on the global mean-squared error. For the first term, observe that the local mean-squared accuracy of MALA implies (cf. Lemma 4.8),

$$\mathbb{E}\{|X_{k+1} - Y_{t_{k+1},t_k}(X_k)|^2 \mid \mathcal{F}_k\} \leq K(1 + U(X_k)^4)h^{5/2}.$$

Since MALA preserves μ by design, by the law of total expectation,

$$\mathbb{E}_\mu \mathbb{E}^{\mathbf{x}} \{\mathbb{E}\{|X_{k+1} - Y_{t_{k+1},t_k}(X_k)|^2 \mid \mathcal{F}_k\}\} \leq K(1 + \mu(U^4))h^{5/2}$$

where $\mu(U^4) = \int_{\mathbb{R}^n} U^4 d\mu$. Similarly, for the second term in (4.11), Lemma 4.3 implies that

$$\mathbb{E}_\mu \mathbb{E}^{\mathbf{x}} \{|Y_{t_{k+1}, t_k}(\mathbf{X}_k) - Y_{t_{k+1}, t_k}(Y_{t_k, 0}(\mathbf{x}))|^2\} \leq (1 + Kh)\epsilon_k.$$

This leaves the third term in (4.11).

Set

$$\Delta = Y_{t_{k+1}, t_k}(\mathbf{X}_k) - Y_{t_{k+1}, t_k}(Y_{t_k, 0}(\mathbf{x})) - (\mathbf{X}_k - Y_{t_k, 0}(\mathbf{x})).$$

The third term in (4.11) can be rewritten as

$$\langle \mathbf{X}_{k+1} - Y_{t_{k+1}, t_k}(\mathbf{X}_k), \mathbf{X}_k - Y_{t_k, 0}(\mathbf{x}) \rangle + \langle \mathbf{X}_{k+1} - Y_{t_{k+1}, t_k}(\mathbf{X}_k), \Delta \rangle.$$

The Cauchy-Schwarz inequality implies

$$\begin{aligned} \mathbb{E}_\mu \mathbb{E}^{\mathbf{x}} \{\langle \mathbb{E}\{\mathbf{X}_{k+1} - Y_{t_{k+1}, t_k}(\mathbf{X}_k) \mid \mathcal{F}_k\}, \mathbf{X}_k - Y_{t_k, 0}(\mathbf{x}) \rangle\} &\leq \\ &\mathbb{E}_\mu \mathbb{E}^{\mathbf{x}} \{\|\mathbb{E}\{\mathbf{X}_{k+1} - Y_{t_{k+1}, t_k}(\mathbf{X}_k) \mid \mathcal{F}_k\}\|^2\}^{1/2} \epsilon_k^{1/2}. \end{aligned}$$

It follows from Lemma 4.8 and the invariance of μ under MALA that

$$\mathbb{E}_\mu \mathbb{E}^{\mathbf{x}} \{\|\mathbb{E}\{\mathbf{X}_{k+1} - Y_{t_{k+1}, t_k}(\mathbf{X}_k) \mid \mathcal{F}_k\}\|^2\}^{1/2} \leq K(1 + \mu(U^3))h^2.$$

The Cauchy-Schwarz inequality also implies

$$\begin{aligned} &\mathbb{E}_\mu \mathbb{E}^{\mathbf{x}} \{\mathbb{E}\{\langle \mathbf{X}_{k+1} - Y_{t_{k+1}, t_k}(\mathbf{X}_k), \Delta \rangle \mid \mathcal{F}_k\}\} \\ &\leq \mathbb{E}_\mu \mathbb{E}^{\mathbf{x}} \{\mathbb{E}\{|\mathbf{X}_{k+1} - Y_{t_{k+1}, t_k}(\mathbf{X}_k)|^2 \mid \mathcal{F}_k\}\}^{1/2} \mathbb{E}_\mu \mathbb{E}^{\mathbf{x}} \{\mathbb{E}\{|\Delta|^2 \mid \mathcal{F}_k\}\}^{1/2} \\ &\leq K(1 + \mu(U^2)) \mathbb{E}_\mu \mathbb{E}^{\mathbf{x}} \{\mathbb{E}\{|\Delta|^2 \mid \mathcal{F}_k\}\}^{1/2} h^{5/4}. \end{aligned}$$

However, according to Lemma 4.3,

$$\mathbb{E}_\mu \mathbb{E}^{\mathbf{x}} \{\mathbb{E}\{|\Delta|^2 \mid \mathcal{F}_k\}\}^{1/2} \leq Kh(1 + \mu(U))\epsilon_k^{1/4}.$$

In sum, there exists a real constant $K > 0$ such that the following term bounds from above the third term in (4.11):

$$Kh^{9/4}\epsilon_k^{1/4} + Kh^2\epsilon_k^{1/2}.$$

From Young's inequality it is clear that

$$2h^2K\epsilon_k^{1/2} \leq h\epsilon_k + h^3K^2$$

and

$$2h^{9/4}K\epsilon_k^{1/4} \leq h^{5/2}K^2 + \epsilon_k^{1/2}h^2 \leq h^{5/2}K^2 + \frac{1}{2}(\epsilon_k h + h^3).$$

Combining the bounds on the iterated expectations of the three terms in (4.11) yields (4.10). \square

4.3 MALTA and Its Properties

A necessary condition for geometric ergodicity of a Metropolis-Hastings method is that the rejection or stagnation probability is bounded away from unity (see proposition 5 of [25]). When the drift in (2.3) is nonglobally Lipschitz, ULA is transient and MALA's stagnation probability cannot be globally bounded away from unity. Hence MALA is not geometrically ergodic. MALTA corrects this problem by truncating the drift in regions where the Euler-Maruyama integrator can be explosive. This modification enables a precise control of the moments of the integrator initiated from a nonequilibrium initial condition. With this control we are able to prove pathwise accuracy of MALTA on finite time intervals.

The geometric ergodicity of MALTA for superexponential target densities with nondegenerate level sets was intuitively clear to the investigators who proposed the algorithm in [26]. Subsequently, it was proven in proposition 2.1 of [1] using techniques to prove geometric ergodicity for random walk Metropolis candidate dynamics given in [12]. For the convenience of the reader, the theorem is restated.

THEOREM 4.10 (MALTA Geometric Ergodicity) *Let the assumptions in Section 4.1 hold. For every $h > 0$, there exists a nonnegative function $M : \mathbb{R}^n \rightarrow \mathbb{R}$ and real constant $\rho < 1$ such that the k -step transition probability $P_h^k(\mathbf{x}, \cdot)$ of MALTA converges to μ geometrically fast:*

$$\|P_h^k(\mathbf{x}, \cdot) - \mu\|_M \leq M(\mathbf{x})\rho^k, \quad \forall k \in \mathbb{N}, \forall \mathbf{x} \in \mathbb{R}^n,$$

where we have introduced a TV norm $\|\cdot\|_M$, which is defined for a signed measure ν as

$$\|\nu\|_M = \sup_{f: |f| \leq M} |\nu(f)|.$$

The main implication of geometric ergodicity for our purpose is the following crucial estimate on moments of MALTA. The proof of this lemma is a straightforward but tedious consequence of the results in [1, proposition 2.1 and lemma 6.2].

LEMMA 4.11 (Estimates on Higher Moments of MALTA) *Let the assumptions in Section 4.1 hold. For every $E_0 > 0$ and every $\ell \geq 1$, there exists a $h_c > 0$ and $C(E_0) > 0$ such that for all positive $h < h_c$, for all $\mathbf{x} : U(\mathbf{x}) \leq E_0$, and for all $t > 0$,*

$$\sup_{h < h_c} \mathbb{E}^{\mathbf{x}} \{U(\mathbf{Z}_{\lfloor t/h \rfloor})^\ell\} \leq C(E_0).$$

We are now in measure to prove Theorem 2.1.

THEOREM 2.1 (MALTA Pathwise Accuracy) *Let the assumptions in Section 4.1 hold. Then, for every $E_0 > 0$ and $T > 0$, there exists a $h_c(E_0) > 0$ and a $C(T, E_0) > 0$ such that for all positive $h < h_c$, for all $\mathbf{x} : U(\mathbf{x}) \leq E_0$, and for all $t \in [0, T]$,*

$$(\mathbb{E}^{\mathbf{x}} \{|\mathbf{Z}_{\lfloor t/h \rfloor} - \mathbf{Y}_{t,0}|^2\})^{1/2} \leq C(T, E_0)h^{3/4}.$$

PROOF: The proof of this theorem is similar to the proof of Theorem 2.2 with the following main differences:

- The preservation of the moments of the solution to (2.1) and the Metropolis-Hastings method due to ergodicity are replaced by bounds on moments implied by Lemmas 4.2 and 4.11.
- The $k + 1^{\text{th}}$ step error conditioned on knowing \mathcal{F}_k is split into two parts:

$$\begin{aligned} & \mathbb{E}\{|Z_{k+1} - Y(t_{k+1})|^2 \mid \mathcal{F}_k\} \\ & \leq \mathbb{E}\{|Z_{k+1} - Y(t_{k+1})|^2 \mid \mathcal{F}_k \text{ and } h|\nabla U(\mathbf{Z}_k)| \leq 1\} P(h|\nabla U(\mathbf{Z}_k)| \leq 1) \\ & \quad + \mathbb{E}\{|Z_{k+1} - Y(t_{k+1})|^2 \mid \mathcal{F}_k \text{ and } h|\nabla U(\mathbf{Z}_k)| > 1\} P(h|\nabla U(\mathbf{Z}_k)| > 1). \end{aligned}$$

For the first term on the right-hand side, the results of Theorem 2.2 apply with the provision given in the first difference above. For the second term on the right-hand side, Chebyshev's inequality implies

$$P(h|\nabla U(\mathbf{Z}_k)| > 1) \leq h^2 \mathbb{E}^{\mathbf{x}}\{|\nabla U(\mathbf{Z}_k)|^2\}.$$

This inequality, the assumptions in Section 4.1, Lemma 4.2, and Lemma 4.11 imply that there exists a constant $C(E_0) > 0$ such that

$$\mathbb{E}^{\mathbf{x}}\{|Z_{k+1} - Y(t_{k+1})|^2\} P(h|\nabla U(\mathbf{Z}_k)| > 1) \leq h^2 C(E_0). \quad \square$$

5 Inertial Langevin

Next we shall consider a Metropolized version of a discretization of (2.11) that extends variational integrators to inertial Langevin dynamics. We begin by introducing the so-called geometric Langevin algorithm and discuss its properties. The section then examines the properties of the Metropolis-Hastings adjusted geometric Langevin algorithm, including quantifying its accuracy in approximating inertial Langevin dynamics.

The arguments in this section are very closely related to those in Section 4. However, there are two key differences. First, the solution to (2.11) is no longer a reversible stochastic process. Yet the transition kernel of the solution process composed with a momentum flip does satisfy detailed balance. Second, the diffusion in (2.11) is only applied to momentum and not position. These differences motivate this section on Metropolizing discretizations of nonreversible processes. However, we will omit analysis that is redundant.

5.1 Geometric Langevin Algorithm

Let N and h be given, and set $T = Nh$ and $t_k = hk$ for $k = 0, \dots, N$. We shall consider an integrator for (2.11) based on splitting the Langevin equations into Hamilton's equations for the Hamiltonian H (2.13) and Ornstein-Uhlenbeck equations (2.14). The solution of Hamilton's equations will be approximated by the discrete Hamiltonian map of a variational integrator [15], while the exact flow will be used for the Ornstein-Uhlenbeck equations. These flows will be composed

in a Strang-type fashion to obtain a pathwise approximant to Langevin’s equations, which we will call the geometric Langevin algorithm (GLA).

Variational Integrators

Let $L : \mathbb{R}^{2n} \rightarrow \mathbb{R}$ denote the Lagrangian obtained from the Legendre transform of the Hamiltonian H , given by

$$L(\mathbf{q}, \mathbf{v}) = \frac{1}{2} \mathbf{v}^\top \mathbf{M} \mathbf{v} - U(\mathbf{q}).$$

A variational integrator is defined by a discrete Lagrangian $L_d : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^+ \rightarrow \mathbb{R}$, which is an approximation to the so-called *exact discrete Lagrangian*, which is defined as

$$L_d^E(\mathbf{q}_0, \mathbf{q}_1, h) = \int_0^h L(\mathbf{Q}, \dot{\mathbf{Q}}) dt$$

where $\mathbf{Q}(t)$ solves the Euler-Lagrange equations for the Lagrangian L with endpoint conditions $\mathbf{Q}(0) = \mathbf{q}_0$ and $\mathbf{Q}(h) = \mathbf{q}_1$.

A discrete Lagrangian determines a symplectic integrator as follows: Given $(\mathbf{q}_0, \mathbf{p}_0) \in \mathbb{R}^{2n}$, a variational integrator defines an update $(\mathbf{q}_1, \mathbf{p}_1) \in \mathbb{R}^{2n}$ by the following system of equations:

$$(5.1) \quad \begin{cases} \mathbf{p}_0 = -D_1 L_d(\mathbf{q}_0, \mathbf{q}_1, h), \\ \mathbf{p}_1 = D_2 L_d(\mathbf{q}_0, \mathbf{q}_1, h). \end{cases}$$

Denote this map by $\theta_h : \mathbb{R}^{2n} \rightarrow \mathbb{R}^{2n}$, i.e.,

$$\theta_h : (\mathbf{q}_0, \mathbf{p}_0) \mapsto (\mathbf{q}_1, \mathbf{p}_1),$$

where $(\mathbf{q}_1, \mathbf{p}_1)$ solves (5.1). One can show that θ_h preserves the canonical symplectic form on \mathbb{R}^{2n} and hence is Lebesgue measure preserving [15]. By appropriately constructing L_d , the map θ_h can define an approximation to the flow of Hamilton’s equations for the Hamiltonian H (2.13).

A discrete Lagrangian is self-adjoint if

$$(5.2) \quad L_d(\mathbf{q}_0, \mathbf{q}_1, h) = L_d(\mathbf{q}_1, \mathbf{q}_0, h).$$

Some of the results that follow will be specific to the Störmer-Verlet integrator, which can be derived from the following discrete Lagrangian:

$$(5.3) \quad L_d(\mathbf{q}_0, \mathbf{q}_1, h) = \frac{1}{2h} (\mathbf{q}_1 - \mathbf{q}_0)^\top \mathbf{M} (\mathbf{q}_1 - \mathbf{q}_0) - \frac{h}{2} (U(\mathbf{q}_0) + U(\mathbf{q}_1)).$$

This discrete Lagrangian is clearly self-adjoint.

Ornstein-Uhlenbeck Equations

The following stochastic evolution map $\psi_{t_k+h, t_k} : \mathbb{R}^{2n} \rightarrow \mathbb{R}^{2n}$ defines the stochastic flow of (2.14):

$$(5.4) \quad \psi_{t_k+h, t_k} : (\mathbf{q}, \mathbf{p}) \mapsto \left(\mathbf{q}, e^{-\gamma \mathbf{M}^{-1} h} \mathbf{p} + \sqrt{2\beta^{-1}\gamma} \int_{t_k}^{t_k+h} e^{-\gamma \mathbf{M}^{-1}(t_k+h-s)} d\mathbf{W}(s) \right).$$

For the distribution of the solution, the stochastic flow will be denoted simply by ψ_h . Let o_h denote the transition probability density of ψ_{t_k+h, t_k} . By a change of variables, its transition density is given explicitly by

$$o_h(\mathbf{p}_0, \mathbf{p}_1) = \frac{1}{(2\pi)^{n/2} |\det(\boldsymbol{\Sigma}_h)|} \exp\left(-\frac{1}{2}(\mathbf{p}_1 - e^{-\gamma \mathbf{M}^{-1} h} \mathbf{p}_0)^\top \boldsymbol{\Sigma}_h^{-1} (\mathbf{p}_1 - e^{-\gamma \mathbf{M}^{-1} h} \mathbf{p}_0)\right),$$

where

$$\boldsymbol{\Sigma}_h = \beta^{-1} (\mathbf{Id} - \exp(-2\gamma \mathbf{M}^{-1} h)) \mathbf{M}.$$

Observe that this transition density does not depend on the initial or terminal position, since (5.4) fixes position and the Hamiltonian is separable.

Strang-Type Splitting

GLA is defined as the following Strang-type splitting pathwise approximant to (2.11):

$$(5.5) \quad \tilde{\mathbf{X}}_{k+1} := (\tilde{\mathbf{Q}}_{k+1}, \tilde{\mathbf{P}}_{k+1}) = \psi_{t_k+h, t_k+h/2} \circ \theta_h \circ \psi_{t_k+h/2, t_k}(\tilde{\mathbf{Q}}_k, \tilde{\mathbf{P}}_k)$$

for $k = 0, \dots, N-1$. For any $A \in \mathcal{B}(\mathbb{R}^{2n})$, the transition probability kernel for GLA is given by

$$Q_h((\mathbf{q}_0, \mathbf{p}_0), A) = \int_{\mathbb{R}^{2n} \times A} o_{h/2}(\mathbf{p}_0, \mathbf{p}_0^*) o_{h/2}(\mathbf{p}_1^*, \mathbf{p}_1) \delta_{\theta_h(\mathbf{q}_0, \mathbf{p}_0^*)}(d\mathbf{q}_1, d\mathbf{p}_1^*) d\mathbf{p}_0^* d\mathbf{p}_1$$

Observe that the zero of the Dirac-delta measure is implicitly defined by

$$\begin{cases} \mathbf{p}_0^* = -D_1 L_d(\mathbf{q}_0, \mathbf{q}_1, h), \\ \mathbf{p}_1^* = D_2 L_d(\mathbf{q}_0, \mathbf{q}_1, h). \end{cases}$$

To make it explicit, we perform a change of variables

$$\begin{aligned} & Q_h((\mathbf{q}_0, \mathbf{p}_0), A) \\ &= \int_{\mathbb{R}^{2n} \times A} o_{h/2}(\mathbf{p}_0, \mathbf{p}_0^*) o_{h/2}(\mathbf{p}_1^*, \mathbf{p}_1) |\det(D_{12} L_d(\mathbf{q}_0, \mathbf{q}_1, h))| \\ & \quad \times \delta_{(-D_1 L_d(\mathbf{q}_0, \mathbf{q}_1, h), D_2 L_d(\mathbf{q}_0, \mathbf{q}_1, h))}(d\mathbf{p}_0^*, d\mathbf{p}_1^*) d\mathbf{q}_1 d\mathbf{p}_1 \end{aligned}$$

From this it follows that the transition density q_h of GLA is given by

$$(5.6) \quad \begin{aligned} q_h((\mathbf{q}_0, \mathbf{p}_0), (\mathbf{q}_1, \mathbf{p}_1)) &= |\det(D_{12}L_d(\mathbf{q}_0, \mathbf{q}_1, h))| \\ &\quad \times o_{h/2}(\mathbf{p}_0, -D_1L_d(\mathbf{q}_0, \mathbf{q}_1, h)) o_{h/2}(D_2L_d(\mathbf{q}_0, \mathbf{q}_1, h), \mathbf{p}_1). \end{aligned}$$

Observe that an explicit characterization of q_h is available even when the variational integrator is implicit.

Störmer-Verlet Based GLA

In this paragraph a local accuracy result is stated for the Störmer-Verlet-based GLA (2.15). Let $\tilde{\mathbf{X}}_1 := (\tilde{\mathbf{Q}}_1, \tilde{\mathbf{P}}_1)$. Recall that given $(\mathbf{q}_0, \mathbf{p}_0) \in \mathbb{R}^{2n}$ at time 0 and $h > 0$, the algorithm computes $\tilde{\mathbf{X}}_1$ by the following explicit update rule:

$$\begin{cases} \tilde{\mathbf{Q}}_1 = \mathbf{q}_0 + h\mathbf{M}^{-1}e^{-\gamma\mathbf{M}^{-1}h/2}\mathbf{p}_0 - \frac{h^2}{2}\mathbf{M}^{-1}\nabla U(\mathbf{q}_0) \\ \quad + h\sqrt{2\beta^{-1}\gamma}\int_0^{h/2}\mathbf{M}^{-1}e^{-\gamma\mathbf{M}^{-1}(h/2-s)}d\mathbf{W}(s), \\ \tilde{\mathbf{P}}_1 = e^{-\gamma\mathbf{M}^{-1}h}\mathbf{p}_0 - \frac{h}{2}e^{-\gamma\mathbf{M}^{-1}h/2}(\nabla U(\mathbf{q}_0) + \nabla U(\tilde{\mathbf{Q}}_1)) \\ \quad + \sqrt{2\beta^{-1}\gamma}\int_0^he^{-\gamma\mathbf{M}^{-1}(h-s)}d\mathbf{W}(s). \end{cases}$$

The proof of the following lemma is similar to Lemma 4.4 and hence is omitted.

LEMMA 5.1 (Local Accuracy of Störmer-Verlet-Based GLA) *Assume Assumptions 4.1(C) and (E) on the potential energy. For all $h > 0$ and $\mathbf{x} = (\mathbf{q}_0, \mathbf{p}_0) \in \mathbb{R}^{2n}$, there exists $C(\mathbf{q}_0, \mathbf{p}_0) > 0$ such that $\mu(C) < \infty$ and*

(i) *the local mean-squared error of (2.15) satisfies*

$$(\mathbb{E}^{\mathbf{x}}\{|\tilde{\mathbf{X}}_1 - \mathbf{Y}(h)|^2\})^{1/2} \leq C(\mathbf{q}_0, \mathbf{p}_0)h^{3/2};$$

(ii) *the local mean deviation of (2.15) satisfies*

$$|\mathbb{E}^{\mathbf{x}}\{\tilde{\mathbf{X}}_1 - \mathbf{Y}(h)\}| \leq C(\mathbf{q}_0, \mathbf{p}_0)h^3.$$

As has been demonstrated, the Markov chain defined by sampling (2.15) every time step possesses a smooth probability transition density with respect to Lebesgue measure. For globally Lipschitz potential forces, this discretization is a first-order, strongly accurate integrator for (2.11). However, if the potential force is nonglobally Lipschitz, the Störmer-Verlet-based GLA is plagued with the same transient behavior as Euler-Maruyama for overdamped Langevin dynamics. As before, a Metropolis-Hastings method is proposed to stochastically stabilize this Markov chain.

5.2 MAGLA and Its Properties

MAGLA is the Metropolis-adjusted GLA. Given h and $(\mathbf{Q}_k, \mathbf{P}_k)$, MAGLA computes a proposed move according to a step of GLA:

$$(\mathbf{Q}_{k+1}^*, \mathbf{P}_{k+1}^*) = \psi_{t_k+h, t_k+h/2} \circ \theta_h \circ \psi_{t_k+h/2, t_k}(\mathbf{Q}_k, \mathbf{P}_k).$$

MAGLA accepts this proposed move with probability

$$\alpha_h((\mathbf{q}_0, \mathbf{p}_0), (\mathbf{q}_1, \mathbf{p}_1)) = 1 \wedge \frac{q_h((\mathbf{q}_1, \mathbf{p}_1), (\mathbf{q}_0, -\mathbf{p}_0))\pi(\mathbf{q}_1, \mathbf{p}_1)}{q_h((\mathbf{q}_0, \mathbf{p}_0), (\mathbf{q}_1, -\mathbf{p}_1))\pi(\mathbf{q}_0, \mathbf{p}_0)}.$$

Altogether, the MAGLA update is given by

$$X_{k+1} := (\mathbf{Q}_{k+1}, \mathbf{P}_{k+1}) = \begin{cases} (\mathbf{Q}_{k+1}^*, \mathbf{P}_{k+1}^*) & \text{if } \zeta_k < \alpha_h((\mathbf{Q}_k, \mathbf{P}_k), (\mathbf{Q}_{k+1}^*, \mathbf{P}_{k+1}^*)) \\ \varphi(\mathbf{Q}_k, \mathbf{P}_k) & \text{otherwise} \end{cases}$$

for $k = 0, \dots, N - 1$. We stress that the momentum flip in the acceptance probability and upon rejection is introduced because inertial Langevin dynamics is irreversible. Yet the exact transition density of the solution to inertial Langevin equations composed with a momentum is reversible; i.e., it does satisfy detailed balance. We will see in the proof of Lemma 5.3 that without this momentum flip, pathwise accuracy cannot be achieved with MAGLA.

MAGLA preserves μ and hence is not transient even when its underlying variational integrator is explicit. In fact, it is often straightforward to classify MAGLA as an ergodic Markov chain even if the potential force is nonglobally Lipschitz. However, with an explicit variational integrator MAGLA is often no longer geometrically ergodic even with momentum flips. To correct this problem, a modification of MAGLA can be implemented where the drift is truncated in regions where an explicit discretization of the drift causes high rejection rates. This modification, which would be analogous to MALTA, is not implemented here. Instead, we will concentrate on proving pathwise accuracy from an initial condition restricted to the equilibrium measure. This restriction enables us to obtain bounds on relevant moments of the Metropolized integrator. To prove pathwise accuracy the following lemmas will be useful.

The following lemma shows for an arbitrary symmetric variational integrator, the acceptance probability of MAGLA is related to the energy change in its variational integrator. This implies that the acceptance probability of MAGLA has mechanical and intrinsic meaning. It also simplifies the subsequent analysis of MAGLA.

LEMMA 5.2 (Acceptance Probability of MAGLA) *Let q_h denote the transition probability density of GLA. Let L_d denote the discrete Lagrangian associated with the variational integrator θ_h . Assume that L_d is self-adjoint (cf. (5.2)). Then the acceptance probability of MAGLA satisfies*

$$\alpha_h((\mathbf{q}_0, \mathbf{p}_0), (\mathbf{q}_1, \mathbf{p}_1)) = 1 \wedge \exp(-\beta\Delta E(\mathbf{q}_0, \mathbf{q}_1)),$$

where we have introduced

$$(5.7) \quad \begin{aligned} \Delta E(\mathbf{q}_0, \mathbf{q}_1) &= \frac{1}{2} D_2 L_d(\mathbf{q}_0, \mathbf{q}_1, h)^\top \mathbf{M}^{-1} D_2 L_d(\mathbf{q}_0, \mathbf{q}_1, h) + U(\mathbf{q}_1) \\ &\quad - \frac{1}{2} D_1 L_d(\mathbf{q}_0, \mathbf{q}_1, h)^\top \mathbf{M}^{-1} D_1 L_d(\mathbf{q}_0, \mathbf{q}_1, h) - U(\mathbf{q}_0). \end{aligned}$$

PROOF: Introduce the abbreviations $D_i L_d = D_i L_d(\mathbf{q}_0, \mathbf{q}_1, h)$ for $i = 1, 2$ and $\mathbf{B}_{h/2} = \exp(-\gamma \mathbf{M}^{-1} h/2)$. Expanding (5.6) yields

$$(5.8) \quad \begin{aligned} & q_h((\mathbf{q}_0, \mathbf{p}_0), (\mathbf{q}_1, \mathbf{p}_1)) \\ &= \frac{|\det(D_{12} L_d(\mathbf{q}_0, \mathbf{q}_1, h))|}{(2\pi)^n |\det(\boldsymbol{\Sigma}_{h/2})|} \\ & \exp\left(-\frac{\beta}{2} ((D_1 L_d + \mathbf{B}_{h/2} \mathbf{p}_0)^\top \mathbf{M}^{-1} (\mathbf{Id} - \mathbf{B}_{h/2}^2)^{-1} (D_1 L_d + \mathbf{B}_{h/2} \mathbf{p}_0) \right. \\ & \quad \left. + (\mathbf{p}_1 - \mathbf{B}_{h/2} D_2 L_d)^\top \mathbf{M}^{-1} (\mathbf{Id} - \mathbf{B}_{h/2}^2)^{-1} (\mathbf{p}_1 - \mathbf{B}_{h/2} D_2 L_d))\right). \end{aligned}$$

The self-adjoint property of the discrete Lagrangian implies that

$$\begin{aligned} D_2 L_d(\mathbf{q}_0, \mathbf{q}_1, h) &= D_1 L_d(\mathbf{q}_1, \mathbf{q}_0, h), \\ D_1 L_d(\mathbf{q}_0, \mathbf{q}_1, h) &= D_2 L_d(\mathbf{q}_1, \mathbf{q}_0, h). \end{aligned}$$

Hence

$$(5.9) \quad \begin{aligned} & q_h((\mathbf{q}_1, \mathbf{p}_1), (\mathbf{q}_0, \mathbf{p}_0)) \\ &= \frac{|\det(D_{12} L_d(\mathbf{q}_0, \mathbf{q}_1, h))|}{(2\pi)^n |\det(\boldsymbol{\Sigma}_{h/2})|} \\ & \exp\left(-\frac{\beta}{2} ((D_2 L_d + \mathbf{B}_{h/2} \mathbf{p}_1)^\top \mathbf{M}^{-1} (\mathbf{Id} - \mathbf{B}_{h/2}^2)^{-1} (D_2 L_d + \mathbf{B}_{h/2} \mathbf{p}_1) \right. \\ & \quad \left. + (\mathbf{p}_0 - \mathbf{B}_{h/2} D_1 L_d)^\top \mathbf{M}^{-1} (\mathbf{Id} - \mathbf{B}_{h/2}^2)^{-1} (\mathbf{p}_0 - \mathbf{B}_{h/2} D_1 L_d))\right). \end{aligned}$$

From (5.8) and (5.9), it follows that

$$\begin{aligned} & \frac{q_h((\mathbf{q}_1, \mathbf{p}_1), (\mathbf{q}_0, -\mathbf{p}_0))}{q_h((\mathbf{q}_0, \mathbf{p}_0), (\mathbf{q}_1, -\mathbf{p}_1))} = \\ & e^{-\frac{\beta}{2} (D_2 L_d^\top \mathbf{M}^{-1} D_2 L_d - D_1 L_d^\top \mathbf{M}^{-1} D_1 L_d - \mathbf{p}_1^\top \mathbf{M}^{-1} \mathbf{p}_1 + \mathbf{p}_0^\top \mathbf{M}^{-1} \mathbf{p}_0)}. \end{aligned}$$

Therefore,

$$\begin{aligned} & \frac{q_h((\mathbf{q}_1, \mathbf{p}_1), (\mathbf{q}_0, -\mathbf{p}_0)) \pi((\mathbf{q}_1, \mathbf{p}_1))}{q_h((\mathbf{q}_0, \mathbf{p}_0), (\mathbf{q}_1, -\mathbf{p}_1)) \pi((\mathbf{q}_0, \mathbf{p}_0))} = \\ & e^{-\beta (\frac{1}{2} D_2 L_d^\top \mathbf{M}^{-1} D_2 L_d + U(\mathbf{q}_1) - \frac{1}{2} D_1 L_d^\top \mathbf{M}^{-1} D_1 L_d - U(\mathbf{q}_0))}. \end{aligned}$$

□

The following lemma is analogous to Lemma 4.7 and, roughly speaking, quantifies how often rejections occur in MAGLA.

LEMMA 5.3 (Stagnation Probability of Störmer-Verlet-Based MAGLA) *Consider the Metropolized Störmer-Verlet-based GLA (2.15). Let Assumptions 4.1(C) and (E) hold on the potential energy. For any integer $\ell \geq 1$, there exists $h_c > 0$ and $K_\ell > 0$ such that for all positive $h < h_c$ and for all $\mathbf{x} = (\mathbf{q}_0, \mathbf{p}_0) \in \mathbb{R}^n$,*

$$\mathbb{E}^{\mathbf{x}} \{(\alpha_h((\mathbf{q}_0, \mathbf{p}_0), (\mathbf{Q}_1^*, \mathbf{P}_1^*)) - 1)^{2\ell}\} \leq C(\mathbf{q}_0, \mathbf{p}_0)h^{6\ell},$$

where

$$\begin{cases} \mathbf{Q}_1^* = \mathbf{q}_0 + h\mathbf{M}^{-1}e^{-\gamma\mathbf{M}^{-1}h/2}\mathbf{p}_0 - \frac{h^2}{2}\mathbf{M}^{-1}\nabla U(\mathbf{q}_0) \\ \quad + h\sqrt{2\beta^{-1}\gamma}\int_{t_k}^{t_k+h/2}\mathbf{M}^{-1}e^{-\gamma\mathbf{M}^{-1}(t_k+h/2-s)}d\mathbf{W}(s), \\ \mathbf{P}_1^* = e^{-\gamma\mathbf{M}^{-1}h}\mathbf{p}_0 - \frac{h}{2}e^{-\gamma\mathbf{M}^{-1}h/2}(\nabla U(\mathbf{q}_0) + \nabla U(\mathbf{Q}_1^*)) \\ \quad + \sqrt{2\beta^{-1}\gamma}\int_{t_k}^{t_k+h}e^{-\gamma\mathbf{M}^{-1}(t_k+h-s)}d\mathbf{W}(s). \end{cases}$$

PROOF: From Lemma 5.2 it is clear that

$$\begin{aligned} \mathbb{E}^{\mathbf{x}} \{(\alpha_h((\mathbf{q}_0, \mathbf{p}_0), (\mathbf{Q}_1^*, \mathbf{P}_1^*)) - 1)^{2\ell}\} = \\ \int_{\mathbb{R}^{2n}} (\exp(-\beta\Delta E(\mathbf{q}_0, \mathbf{q}_1)) \wedge 1 - 1)^{2\ell} q_h((\mathbf{q}_0, \mathbf{p}_0), (\mathbf{q}_1, \mathbf{p}_1)) d\mathbf{q}_1 d\mathbf{p}_1. \end{aligned}$$

Substitute q_h from (5.6) into the above to obtain

$$\begin{aligned} \mathbb{E}^{\mathbf{x}} \{(\alpha_h((\mathbf{q}_0, \mathbf{p}_0), (\mathbf{Q}_1^*, \mathbf{P}_1^*)) - 1)^{2\ell}\} \\ = \int_{\mathbb{R}^{2n}} |\det(D_{12}L_d(\mathbf{q}_0, \mathbf{q}_1, h))| (\exp(-\beta\Delta E(\mathbf{q}_0, \mathbf{q}_1)) \wedge 1 - 1)^{2\ell} \\ \quad \times o_{h/2}(\mathbf{p}_0, -D_1L_d(\mathbf{q}_0, \mathbf{q}_1, h)) \\ \quad \times o_{h/2}(D_2L_d(\mathbf{q}_0, \mathbf{q}_1, h), \mathbf{p}_1) d\mathbf{q}_1 d\mathbf{p}_1. \end{aligned}$$

Integrate with respect to \mathbf{p}_1 to obtain the following simplified expression:

$$\begin{aligned} \mathbb{E}^{\mathbf{x}} \{(\alpha_h((\mathbf{q}_0, \mathbf{p}_0), (\mathbf{Q}_1^*, \mathbf{P}_1^*)) - 1)^{2\ell}\} \\ = \int_{\mathbb{R}^n} |\det(D_{12}L_d)| \cdot (\exp(-\beta\Delta E(\mathbf{q}_0, \mathbf{q}_1)) \wedge 1 - 1)^{2\ell} \\ \quad \times o_{h/2}(\mathbf{p}_0, -D_1L_d(\mathbf{q}_0, \mathbf{q}_1, h)) d\mathbf{q}_1. \end{aligned}$$

Let $R_h(\mathbf{q}_0) = \{\mathbf{q}_1 \in \mathbb{R}^n : \Delta E(\mathbf{q}_0, \mathbf{q}_1) > 0\}$. Then

$$\begin{aligned}
 & \mathbb{E}^{\mathbf{x}}\{(\alpha_h((\mathbf{q}_0, \mathbf{p}_0), (\mathbf{Q}_1^*, \mathbf{P}_1^*)) - 1)^{2\ell}\} \\
 &= \int_{R_h(\mathbf{q}_0)} |\det(D_{12}L_d)| (e^{-\beta\Delta E(\mathbf{q}_0, \mathbf{q}_1)} - 1)^{2\ell} \\
 & \quad \times o_{h/2}(\mathbf{p}_0, -D_1L_d(\mathbf{q}_0, \mathbf{q}_1, h)) d\mathbf{q}_1.
 \end{aligned}
 \tag{5.10}$$

Let $A_{h/2}$ be the decomposition matrix arising from the Cholesky factorization of $\Sigma_{h/2}$, i.e., $A_{h/2}A_{h/2}^\top = \Sigma_{h/2}$. Introduce the map $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ with $\mathbf{q}_1 = \varphi(\boldsymbol{\eta})$ and defined implicitly by the following relation:

$$-D_1L_d(\mathbf{q}_0, \mathbf{q}_1, h) = \exp(-\gamma\mathbf{M}^{-1}h/2)\mathbf{p}_0 + A_{h/2}\boldsymbol{\eta}.
 \tag{5.11}$$

Set $\tilde{R}_h(\mathbf{q}_0) = \varphi^{-1}(R_h(\mathbf{q}_0))$. A change of variables of (5.10) under the map φ yields

$$\begin{aligned}
 & \mathbb{E}^{\mathbf{x}}\{(\alpha_h((\mathbf{q}_0, \mathbf{p}_0), (\mathbf{Q}_1^*, \mathbf{P}_1^*)) - 1)^{2\ell}\} = \\
 & \quad (2\pi)^{-n/2} \int_{\tilde{R}_h(\mathbf{q}_0)} (e^{-\beta\Delta E(\mathbf{q}_0, \varphi(\boldsymbol{\eta}))} - 1)^{2\ell} e^{-\frac{1}{2}|\boldsymbol{\eta}|} d\boldsymbol{\eta}.
 \end{aligned}
 \tag{5.12}$$

Up to this point the argument has been for a general GLA. Now, the proof is specialized to the Störmer-Verlet-based GLA. For the Störmer-Verlet discrete Lagrangian (5.3),

$$\begin{aligned}
 D_1L_d(\mathbf{q}_0, \mathbf{q}_1, h) &= -\mathbf{M}\frac{\mathbf{q}_1 - \mathbf{q}_0}{h} - \frac{h}{2}\nabla U(\mathbf{q}_0), \\
 D_2L_d(\mathbf{q}_0, \mathbf{q}_1, h) &= \mathbf{M}\frac{\mathbf{q}_1 - \mathbf{q}_0}{h} - \frac{h}{2}\nabla U(\mathbf{q}_1).
 \end{aligned}$$

Substituting these expressions into (5.7) gives

$$\begin{aligned}
 \Delta E(\mathbf{q}_0, \mathbf{q}_1) &= U(\mathbf{q}_1) - U(\mathbf{q}_0) - \frac{1}{2}\langle \nabla U(\mathbf{q}_1) + \nabla U(\mathbf{q}_0), \mathbf{q}_1 - \mathbf{q}_0 \rangle \\
 & \quad + \frac{h^2}{8}(\nabla U(\mathbf{q}_1)^\top \mathbf{M}^{-1} \nabla U(\mathbf{q}_1) - \nabla U(\mathbf{q}_0)^\top \mathbf{M}^{-1} \nabla U(\mathbf{q}_0)).
 \end{aligned}
 \tag{5.13}$$

Substitute the change of variables implicitly defined by (5.11) into (5.13) and then Taylor-expand ΔE about $h = 0$ to obtain

$$\begin{aligned}
 \Delta E\left(\mathbf{q}_0, \mathbf{q}_0 + h\mathbf{M}^{-1}e^{-\gamma\mathbf{M}^{-1}\frac{h}{2}}\mathbf{p}_0 + h\mathbf{M}^{-1}A_{h/2}\boldsymbol{\eta} - \frac{h^2}{2}\mathbf{M}^{-1}\nabla U(\mathbf{q}_0)\right) &= \\
 -\frac{h^3}{12}D^3U(\mathbf{q}_0)(\mathbf{M}^{-1}\mathbf{p}_0)^3 + \frac{h^3}{4}D^2U(\mathbf{q}_0)(\mathbf{M}^{-1}\mathbf{p}_0, \mathbf{M}^{-1}\nabla U(\mathbf{q}_0)) + O(h^{7/2}).
 \end{aligned}$$

As in Lemma 4.7, a standard application of Laplace’s method together with Assumption 4.1(E) gives the desired result. \square

Using Lemma 5.3, we are now in a position to quantify the local accuracy of MAGLA.

LEMMA 5.4 (Local Accuracy of Störmer-Verlet-Based MAGLA) *Assume Assumptions 4.1(C) and (E) for the potential energy. For all $h > 0$ and $\mathbf{x} = (\mathbf{q}_0, \mathbf{p}_0) \in \mathbb{R}^n$, there exists $C(\mathbf{q}_0, \mathbf{p}_0) > 0$ such that $\mu(C) < \infty$ and*

(i) *the local mean-squared error of MAGLA satisfies*

$$(\mathbb{E}^{\mathbf{x}}\{|X_1 - Y(h)|^2\})^{1/2} \leq C(\mathbf{q}_0, \mathbf{p}_0)h^{3/2};$$

(ii) *the local mean deviation of MAGLA satisfies*

$$|\mathbb{E}^{\mathbf{x}}\{X_1 - Y(h)\}| \leq C(\mathbf{q}_0, \mathbf{p}_0)h^3.$$

Remark 5.5. A detailed proof is omitted since similar steps are taken in the proof of Lemma 4.8. We remark that it follows from the definition of MAGLA (2.18) that

$$\begin{aligned} & \mathbb{E}^{\mathbf{x}}\{|X_1 - Y(h)|^2\} \\ & \leq \mathbb{E}^{\mathbf{x}}\{|X_1^* - Y(h)|^2\} \\ & \quad + \mathbb{E}^{\mathbf{x}}\{|\varphi(\mathbf{q}_0, \mathbf{p}_0) - Y(h)|^2(1 - \alpha_h((\mathbf{q}_0, \mathbf{p}_0), (\mathbf{Q}_1^*, \mathbf{P}_1^*)))\}. \end{aligned}$$

This estimate reveals that the local mean-squared error of MAGLA is a sum of the local mean-squared error of GLA (cf. Lemma 5.1) and the error that arises from a potential rejection weighted by the probability of rejection. This error is $O(1)$ because when a rejection happens in (2.18), the momentum is flipped. However, the probability of rejection is within the order of accuracy of the method according to Lemma 5.3.

As discussed, MAGLA involves a momentum flip whenever a proposed move is rejected. Despite this, MAGLA provides a pathwise approximant to (2.11) as summarized by the following theorem:

THEOREM 2.3 (Störmer-Verlet-Based MAGLA Pathwise Accuracy) *Let the assumptions in Section 4.1 hold on the potential energy. Then for every $T > 0$, there exists $h_c > 0$ and $C(T) > 0$ such that for all $h < h_c$, for all $\mathbf{x} \in \mathbb{R}^{2n}$, and for all $t \in [0, T]$,*

$$(\mathbb{E}_\mu \mathbb{E}^{\mathbf{x}}\{|X_{\lfloor t/h \rfloor} - Y(t)|^2\})^{1/2} \leq C(T)h.$$

Remark 5.6. A detailed proof is omitted as it is similar to the proof of Theorem 2.2. Recall that the main tools needed for that proof are local accuracy of MAGLA, which follows from Lemma 5.4, and the invariance of μ under both the transition kernel of the solution to (2.11) and the Markov chain defined by Störmer-Verlet-based MAGLA.

6 Conclusion

This paper examined the pathwise accuracy of Metropolized integrators applied to overdamped and inertial Langevin dynamics. The paper primarily analyzed explicit discretizations of these ergodic SDEs. For explicit discretizations if the drift is nonglobally Lipschitz, discretization errors often lead to stochastically unstable Markov chains. A Metropolis-Hastings method was shown to stochastically stabilize these discretizations. The paper used this stochastic stability to quantify the error induced by the Metropolis-Hastings methods in approximating the solution to the SDE.

Although the paper was restricted to specific discretizations of these SDEs as proposed moves, the strategy to prove the results is rather general. The strategy can be used to prove pathwise accuracy of Metropolized versions of higher-order accurate, adaptive, and multistep discretizations of overdamped or inertial Langevin dynamics (see, e.g., [20, 37]). In fact, this strategy applies to a larger class of SDEs, namely those which possess a transition density that satisfies a detailed balance-type condition.

Our strategy relies on two main ingredients. First, the local error induced by the Monte Carlo method depends upon the time step size in an algebraic fashion. For example, we often found that the local mean-squared error of the Metropolized integrator can be decomposed as

$$\mathbb{E}^{\mathbf{x}} \{|X_1 - Y(h)|^2\} = \underbrace{\mathbb{E}^{\mathbf{x}} \{|X_1^* - Y(h)|^2 \alpha_h(\mathbf{x}, X_1^*)\}}_{\text{accepted proposed move}} + \underbrace{\mathbb{E}^{\mathbf{x}} \{|\varphi(\mathbf{x}) - Y(h)|^2 (1 - \alpha_h(\mathbf{x}, X_1^*))\}}_{\text{rejected proposed move}}$$

where X_1 is a single step of the Metropolis-adjusted discretization, $Y(h)$ is the exact solution of the SDE after a single step, X_1^* is the proposed move step (a single step of the unadjusted discretization), α_h is the probability of accepting a proposed move, and φ is a deterministic map applied if a proposed move is rejected. The first term is bounded by $\mathbb{E}^{\mathbf{x}} \{|X_1^* - Y(h)|^2\}$, i.e., the local order of accuracy of the discretization. The second term is bounded by using the Cauchy-Schwarz inequality, together with bounds on $\mathbb{E}^{\mathbf{x}} \{|\varphi(\mathbf{x}) - Y(h)|^4\}^{1/2}$ and $\mathbb{E}^{\mathbf{x}} \{(1 - \alpha_h(\mathbf{x}, X_1^*))^2\}^{1/2}$. For MALA φ was the identity, so that a rejected proposed was near the solution. To be precise, $\mathbb{E}^{\mathbf{x}} \{|\varphi(\mathbf{x}) - Y(h)|^4\}^{1/2} \leq O(h)$. The likelihood of a rejection occurring turns out to be $\mathbb{E}^{\mathbf{x}} \{(1 - \alpha_h(\mathbf{x}, X_1^*))^2\}^{1/2} \leq O(h^{3/2})$. Since the local mean-squared error of the unadjusted discretization was $O(h^{3/2})$, the local mean-squared error of MALA was dominated by the error incurred by rejected proposed moves, which was $O(h^{5/4})$.

For MAGLA φ was not the identity but involved a momentum flip. This momentum flip was imperative because the exact transition density of inertial Langevin dynamics does not satisfy detailed balance, but a modified detailed balance condition that is derived from composing the solution of inertial Langevin equations with a momentum flip. Consequently, a rejected proposed move for MAGLA loses

local accuracy, i.e., $\mathbb{E}^{\mathbf{x}}\{|\varphi(\mathbf{x}) - Y(h)|^4\}^{1/2} \sim O(1)$. Nevertheless, the probability of rejection turns out to be within the order of accuracy of the SV-based GLA, i.e., $\mathbb{E}^{\mathbf{x}}\{(1 - \alpha_h(\mathbf{x}, \mathbf{X}_1^*))^2\}^{1/2} \leq O(h^3)$. Thus, we found the error incurred by a rejected proposed move to be of the same order as that of an accepted proposed move. Hence the local mean-squared accuracy of MAGLA was $O(h^{3/2})$.

The second ingredient involves bounds on moments of the integrator that are uniform in the time step size. These bounds enabled the local error estimates to be extended to global error estimates and hence pathwise convergence on finite time intervals. These bounds are automatic if initial conditions are restricted to the known equilibrium distribution that the Metropolis-adjusted integrator is designed to sample. To relax this restriction on initial conditions, geometric ergodicity played a key role to obtain such bounds. MALA is often ergodic, but because its proposed chain is often transient, it is not geometrically ergodic. This transience is due to a numerical instability in the forward Euler scheme for large energy values. By suitably truncating the drift at high energy values, one can ensure that the proposed dynamics is not transient. MALA with truncated drift, or MALTA, is known to often be geometrically ergodic on infinite time intervals where MALA is not. The paper used this property to prove MALTA is pathwise convergent on finite time intervals without a restriction on its initial condition. We remark that MALTA can be generalized to explicit higher-order, multistep, and adaptive integrators for overdamped and inertial Langevin equations.

Finally, we emphasize that GLA can be extended to nonflat configuration spaces and multiple time steps. Indeed, its underlying variational integrator can incorporate holonomic constraints (via, e.g., Lagrange multipliers) [38] and multiple time steps to obtain so-called asynchronous variational integrators [14]. A natural choice for the variational integrator would be a SHAKE or RATTLE algorithm [15, 38]. The Ornstein-Uhlenbeck equations (2.14) are defined on a flat vector space since the configuration is held fixed. Moreover, a review of the proof of Lemma 5.2 shows that the probability transition density of GLA can be explicitly characterized even if the configuration space is not flat. By inspection, this density is absolutely continuous with respect to the standard volume measure on that nonflat space. Consequently, GLA on manifolds can be used as proposed dynamics in a Metropolis-Hastings context, and MAGLA can be extended to nonflat configuration spaces to treat, e.g., inertial Langevin equations with holonomic constraints as in [37]. Moreover, MAGLA on manifolds can readily be classified as an ergodic Markov chain on that nonflat phase space.

Appendix: Proof of Lemmas 4.2, 4.3, and 4.4

PROOF OF LEMMA 4.2:

(i) Let $G(\mathbf{x}) := U(\mathbf{x})^\ell$. By the Taylor-Itô formula

$$dG(Y(t)) = L\{G(Y(t))\}dt + \text{martingale}.$$

Assumption 4.1(B) on the generator of (2.1) implies

$$\mathbb{E}^{\mathbf{x}}\{G(\mathbf{Y}(t))\} \leq e^{-\delta_\ell t} G(\mathbf{x}) + \frac{M_\ell}{\delta_\ell} (1 - e^{-\delta_\ell t}).$$

From this expression it is clear that for all $t > 0$ and for all $\mathbf{x} \in \mathbb{R}^n$,

$$\mathbb{E}^{\mathbf{x}}\{G(\mathbf{Y}(t))\} \leq G(\mathbf{x}) + \frac{M_\ell}{\delta_\ell}.$$

(ii) Assumption 4.1(A) implies

$$G(\mathbf{x}) \geq K^\ell |\mathbf{x}|^\ell \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

Hence the upper bound derived in the proof of Lemma 4.2(A) on $\mathbb{E}^{\mathbf{x}}\{G(\mathbf{Y}(t))\}$ implies an upper bound on $\mathbb{E}^{\mathbf{x}}\{|\mathbf{Y}(t)|^\ell\}$.

(iii) From the solution to (2.1) it is apparent that

$$\mathbb{E}^{\mathbf{x}}\{|\mathbf{Y}(t) - \mathbf{x}|^{2\ell}\} = \mathbb{E}^{\mathbf{x}}\left\{\left| -\int_0^t \nabla U(\mathbf{Y}(s)) ds + \sqrt{2\beta^{-1}} \int_0^t d\mathbf{W} \right|^{2\ell}\right\}.$$

The triangle inequality implies that

$$\mathbb{E}^{\mathbf{x}}\{|\mathbf{Y}(t) - \mathbf{x}|^{2\ell}\} \leq \mathbb{E}^{\mathbf{x}}\left\{\left(\left|\int_0^t \nabla U(\mathbf{Y}(s)) ds\right| + \sqrt{2\beta^{-1}} \left|\int_0^t d\mathbf{W}\right|\right)^{2\ell}\right\}.$$

Since the function $g(x) = x^{2\ell}$ is strictly convex,

$$\mathbb{E}^{\mathbf{x}}\{|\mathbf{Y}(t) - \mathbf{x}|^{2\ell}\} \leq 2^{2\ell-1} \mathbb{E}^{\mathbf{x}}\left\{\left|\int_0^t \nabla U(\mathbf{Y}(s)) ds\right|^{2\ell} + (2\beta^{-1})^\ell \left|\int_0^t d\mathbf{W}\right|^{2\ell}\right\}.$$

Recall that the higher-order central moments of a normal random variable $\xi \sim \mathcal{N}(0, t)$ are given by

$$\mathbb{E}\{\xi^{2\ell}\} = \frac{(2\ell)!}{2^\ell \ell!} t^\ell.$$

Since the function $g(x) = x^\ell$ is convex for positive reals,

$$\mathbb{E}^{\mathbf{x}}\left\{\left|\int_0^t d\mathbf{W}\right|^{2\ell}\right\} \leq n^\ell \mathbb{E}^{\mathbf{x}}\{\xi^{2\ell}\} = \frac{(2\ell)!}{\ell!} \left(\frac{nt}{2}\right)^\ell.$$

An application of the Cauchy-Schwarz inequality yields

$$\mathbb{E}^{\mathbf{x}}\left\{\left|\int_0^t \nabla U(\mathbf{Y}(s)) ds\right|^{2\ell}\right\} \leq t^\ell \mathbb{E}^{\mathbf{x}}\left\{\left(\int_0^t |\nabla U(\mathbf{Y}(s))|^2 ds\right)^\ell\right\}.$$

According to Assumption 4.1(E), there exists a constant $K > 0$ such that

$$\int_0^t |\nabla U(\mathbf{Y}(s))|^2 ds \leq K \int_0^t (1 + U(\mathbf{Y}(s))^2) ds.$$

An application of Hölder's inequality and Lemma 4.2(A) imply there exists a possibly different constant $K > 0$ such that

$$\mathbb{E}^{\mathbf{x}} \left\{ \left| \int_0^t \nabla U(\mathbf{Y}(s)) ds \right|^{2\ell} \right\} \leq K t^{2\ell} (1 + U(\mathbf{x})^{2\ell}).$$

Hence,

$$\mathbb{E}^{\mathbf{x}} \{ |\mathbf{Y}(t) - \mathbf{x}|^{2\ell} \} \leq 2^{2\ell-1} \left(t^{2\ell} K (1 + U(\mathbf{x})^{2\ell}) + t^\ell \frac{(2\ell)!}{\ell!} \left(\frac{n}{\beta} \right)^\ell \right).$$

□

PROOF OF LEMMA 4.3:

(i) By the Taylor-Itô formula,

$$\begin{aligned} |\mathbf{Y}_{s+h,s}(\mathbf{x}) - \mathbf{Y}_{s+h,s}(\mathbf{y})|^2 &= |\mathbf{x} - \mathbf{y}|^2 \\ &+ 2 \int_0^h \langle \mathbf{Y}_{s+r,s}(\mathbf{x}) - \mathbf{Y}_{s+r,s}(\mathbf{y}), \nabla U(\mathbf{Y}_{s+r,s}(\mathbf{y})) - \nabla U(\mathbf{Y}_{s+r,s}(\mathbf{x})) \rangle dr. \end{aligned}$$

According to Assumption 4.1(D),

$$\begin{aligned} |\mathbf{Y}_{s+h,s}(\mathbf{x}) - \mathbf{Y}_{s+h,s}(\mathbf{y})|^2 &\leq \\ &|\mathbf{x} - \mathbf{y}|^2 + 2K \int_0^h |\mathbf{Y}_{s+r,s}(\mathbf{x}) - \mathbf{Y}_{s+r,s}(\mathbf{y})|^2 dr. \end{aligned}$$

Gronwall's lemma implies that

$$|\mathbf{Y}_{s+h,s}(\mathbf{x}) - \mathbf{Y}_{s+h,s}(\mathbf{y})|^2 \leq |\mathbf{x} - \mathbf{y}|^2 \exp(2hK).$$

(ii) To prove the second inequality, observe that

$$|\Delta|^2 = \left| \int_0^h (\nabla U(\mathbf{Y}_{s+r,s}(\mathbf{y})) - \nabla U(\mathbf{Y}_{s+r,s}(\mathbf{x}))) dr \right|^2.$$

An application of the Cauchy-Schwarz inequality implies

$$|\Delta|^2 \leq h \int_0^h |\nabla U(\mathbf{Y}_{s+r,s}(\mathbf{y})) - \nabla U(\mathbf{Y}_{s+r,s}(\mathbf{x}))|^2 dr.$$

Assumption 4.1(C) implies

$$|\Delta|^2 \leq h \int_0^h (U(\mathbf{Y}_{s+r,s}(\mathbf{y})) + U(\mathbf{Y}_{s+r,s}(\mathbf{x}))) |\mathbf{Y}_{s+r,s}(\mathbf{y}) - \mathbf{Y}_{s+r,s}(\mathbf{x})|^2 dr.$$

Assumption 4.1(A) and the triangle inequality imply there exists a constant $K > 0$ such that

$$|\Delta|^2 \leq hK \int_0^h (U(\mathbf{Y}_{s+r,s}(\mathbf{y}))^2 + U(\mathbf{Y}_{s+r,s}(\mathbf{x}))^2) |\mathbf{Y}_{s+r,s}(\mathbf{y}) - \mathbf{Y}_{s+r,s}(\mathbf{x})| dr.$$

The Cauchy-Schwarz inequality implies that

$$\mathbb{E}\{|\Delta|^2\} \leq hK \int_0^h (\mathbb{E}\{U(Y_{s+r,s}(y))^4\}^{1/2} + \mathbb{E}\{U(Y_{s+r,s}(x))^4\}^{1/2}) \times \mathbb{E}\{|Y_{s+r,s}(y) - Y_{s+r,s}(x)|^2\}^{1/2} dr.$$

Lemma 4.2 implies with a possibly different constant $K > 0$

$$\mathbb{E}\{|\Delta|^2\} \leq hK(1 + U(y)^2 + U(x)^2) \int_0^h \mathbb{E}\{|Y_{s+r,s}(y) - Y_{s+r,s}(x)|^2\}^{1/2} dr.$$

The first part of this lemma implies, again with a possibly different constant $K > 0$, that

$$\mathbb{E}\{|\Delta|^2\} \leq h^2K(1 + U(y)^2 + U(x)^2)|x - y|. \quad \square$$

PROOF OF LEMMA 4.4:

(i) According to (2.1) and (2.3), the difference between the Euler-Maruyama discretization and the solution to the SDE after a single step takes the form

$$\tilde{X}_1 - Y(h) = - \int_0^h (\nabla U(x) - \nabla U(Y(s))) ds.$$

The Cauchy-Schwarz inequality implies

$$\mathbb{E}^x\{|\tilde{X}_1 - Y(h)|^2\} \leq h \int_0^h \mathbb{E}^x\{|\nabla U(x) - \nabla U(Y(s))|^2\} ds.$$

Assumption 4.1(C) implies that there is a constant $K > 0$ such that

$$\mathbb{E}^x\{|\tilde{X}_1 - Y(h)|^2\} \leq hK \mathbb{E}^x \left\{ \int_0^h (U(x)^2 + U(Y(s))^2) |x - Y(s)|^2 ds \right\}.$$

A second application of Cauchy-Schwarz yields

$$\mathbb{E}^x\{|\tilde{X}_1 - Y(h)|^2\} \leq hK \left(\int_0^h U(x)^2 \mathbb{E}^x\{|x - Y(s)|^2\} ds + \int_0^h \mathbb{E}^x\{U(Y(s))^4\}^{1/2} \mathbb{E}^x\{|x - Y(s)|^4\}^{1/2} ds \right).$$

Lemma 4.2 implies that there exists a constant $K > 0$ such that

$$\mathbb{E}^x\{|\tilde{X}_1 - Y(h)|^2\} \leq Kh^3(1 + U(x)^4).$$

(ii) The Itô-Taylor formula implies that

$$|\mathbb{E}^x\{\tilde{X}_1 - Y(h)\}|^2 = \left| \mathbb{E}^x \left\{ \int_0^h \int_0^s L\{\nabla U(Y(r))\} dr ds \right\} \right|^2.$$

Applying the Cauchy-Schwarz inequality twice gives

$$\begin{aligned} |\mathbb{E}^{\mathbf{x}}\{\tilde{\mathbf{X}}_1 - \mathbf{Y}(h)\}|^2 &\leq h \int_0^h \left| \mathbb{E}^{\mathbf{x}} \left\{ \int_0^s L\{\nabla U(\mathbf{Y}(r))\} dr \right\} \right|^2 ds \\ &\leq h \int_0^h s \int_0^s |\mathbb{E}^{\mathbf{x}}\{L\{\nabla U(\mathbf{Y}(r))\}\}|^2 dr ds. \end{aligned}$$

Jensen's inequality implies that

$$|\mathbb{E}^{\mathbf{x}}\{\tilde{\mathbf{X}}_1 - \mathbf{Y}(h)\}|^2 \leq h \int_0^h s \int_0^s \mathbb{E}^{\mathbf{x}}\{|L\{\nabla U(\mathbf{Y}(r))\}|^2\} dr ds.$$

Assumption 4.1(E) implies the existence of a $K > 0$ such that

$$|\mathbb{E}^{\mathbf{x}}\{\tilde{\mathbf{X}}_1 - \mathbf{Y}(h)\}|^2 \leq h \int_0^h s \int_0^s K(1 + \mathbb{E}^{\mathbf{x}}\{U(\mathbf{Y}(r))^4 g\}) dr ds.$$

An application of Lemma 4.2 gives the required result. □

Acknowledgment. We thank Martin Hairer, Tony Lelièvre, Houman Owhadi, and Christof Schütte for stimulating discussions. N. B.-R. was supported by National Science Foundation Fellowship DMS-0803095. The work of E. V.-E. was supported in part by National Science Foundation grants DMS-0718172 and DMS-0708140 and Office of Naval Research Grant N00014-04-1-6046.

Bibliography

- [1] Atchade, Y. F. An adaptive version for the Metropolis adjusted Langevin algorithm with a truncated drift. *Methodol. Comput. Appl. Probab.* **8** (2006), no. 2, 235–254.
- [2] Bou-Rabee, N.; Owhadi, H. Geometric Langevin algorithm. arXiv:0712:4123v4, 2009.
- [3] Bou-Rabee, N.; Vanden-Eijnden, E. Reconciling Monte Carlo and molecular dynamics. Preprint, 2009.
- [4] Brünger, A.; Brooks, C. L., III; Karplus, M. Stochastic boundary conditions for molecular dynamics simulations of ST2 water. *Chem. Phys. Lett.* **105** (1984), no. 5, 495–500.
- [5] Bussi, G.; Parrinello, M. Accurate sampling using Langevin dynamics. *Phys. Rev. E* **75** (2007), no. 5, 056707, 1–7.
- [6] Duane, S.; Kennedy, A. D.; Pendleton, B. J.; Roweth, D. Hybrid Monte Carlo. *Phys. Lett. B* **195** (1987), no. 2, 216–222.
- [7] E, W.; Khanin, K.; Mazel, A.; Sinai, Y. Invariant measures for Burgers equation with stochastic forcing. *Ann. of Math. (2)* **151** (2000), no. 3, 877–960.
- [8] Has'minskiĭ, R. Z. *Stochastic stability of differential equations*. Monographs and Textbooks on Mechanics of Solids and Fluids: Mechanics and Analysis, 7. Sijthoff & Noordhoff, Alphen aan den Rijn–Germantown, Md., 1980.
- [9] Hastings, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** (1970), no. 1, 97–109.
- [10] Higham, D. J.; Mao, X.; Stuart, A. M. Strong convergence of Euler-type methods for nonlinear stochastic differential equations. *SIAM J. Numer. Anal.* **40** (2002), no. 3, 1041–1063 (electronic).

- [11] Horowitz, A. M. A generalized guided Monte Carlo algorithm. *Phys. Lett. B* **268** (1991), no. 2, 247–252.
- [12] Jarner, S. F.; Hansen, E. Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.* **85** (2000), no. 2, 341–361.
- [13] Lamba, H.; Mattingly, J. C.; Stuart, A. M. An adaptive Euler-Maruyama scheme for SDEs: convergence and stability. *IMA J. Numer. Anal.* **27** (2007), no. 3, 479–506.
- [14] Lew, A.; Marsden, J. E.; Ortiz, M.; West, M. Asynchronous variational integrators. *Arch. Ration. Mech. Anal.* **167** (2003), no. 2, 85–145.
- [15] Marsden, J. E.; West, M. Discrete mechanics and variational integrators. *Acta Numer.* **10** (2001), 357–514.
- [16] Mattingly, J. C.; Stuart, A. M.; Higham, D. J. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stochastic Process. Appl.* **101** (2002), no. 2, 185–232.
- [17] Mengersen K. L.; Tweedie, R. L. Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24** (1996), no. 1, 101–121.
- [18] Metropolis, N.; Rosenbluth, A. W.; Teller, A. H.; Teller, E. Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21** (1953), no. 6, 1087–1092.
- [19] Meyn, S. P.; Tweedie, R. L. *Markov chains and stochastic stability*. 2nd ed. Cambridge University Press, Cambridge, 2009.
- [20] Milstein, G. N.; Tretyakov, M. V. *Stochastic numerics for mathematical physics*. Scientific Computation. Springer, Berlin, 2004.
- [21] Milstein, G. N.; Tretyakov, M. V. Numerical integration of stochastic differential equations with nonglobally Lipschitz coefficients. *SIAM J. Numer. Anal.* **43** (2005), no. 3, 1139–1154 (electronic).
- [22] Nummelin, E. *General irreducible Markov chains and nonnegative operators*. Cambridge Tracts in Mathematics, 83. Cambridge University Press, Cambridge, 1984.
- [23] Pavliotis, G. A.; Stuart, A. M.; Zygalkis, K. C. Calculating effective diffusiveness in the limit of vanishing molecular diffusion. *J. Comput. Phys.* **228** (2009), no. 4, 1030–1055.
- [24] Ricci, A.; Ciccotti, G. Algorithms for Brownian dynamics. *Molecular Phys.* **101** (2003), no. 12, 1927–1931.
- [25] Roberts, G. O.; Tweedie, R. L. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2** (1996), no. 4, 341–363.
- [26] Roberts, G. O.; Tweedie, R. L. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83** (1996), no. 1, 95–110.
- [27] Scemama, A.; Lelièvre, T.; Stoltz, G.; Cancés, E.; Caffarel, M. An efficient sampling algorithm for variational Monte Carlo. *J. Chem. Phys.* **125** (2006), no. 6, 114105.
- [28] Serrano, M.; De Fabritiis, G.; Español, P.; Coveney, P. V. A stochastic Trotter integration scheme for dissipative particle dynamics. *Math. Comput. Simulation* **72** (2006), no. 2-6, 190–194.
- [29] Shardlow, T. Splitting for dissipative particle dynamics. *SIAM J. Sci. Comput.* **24** (2003), no. 4, 1267–1282 (electronic).
- [30] Skeel, R. D.; Izaguirre, J. An impulse integrator for Langevin dynamics. *Molecular Phys.* **100** (2002), no. 24, 3885–3891.
- [31] Talay, D. Simulation and numerical analysis of stochastic differential systems: a review. *Probabilistic methods in applied physics*, chap. 3, 63–106. Lecture Notes in Physics, 451. Springer, Berlin, 1995.
- [32] Talay, D. Stochastic Hamiltonian systems: exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. Inhomogeneous random systems (Cergy-Pontoise, 2001). *Markov Process. Related Fields* **8** (2002), no. 2, 163–198.
- [33] Talay, D.; Tubaro, L. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stochastic Anal. Appl.* **8** (1990), no. 4, 483–509.

- [34] Tierney, L. Markov chains for exploring posterior distributions. *Ann. Statist.* **22** (1994), no. 4, 1701–1728.
- [35] Tuckerman, M. E.; Berne, B. J. Stochastic molecular dynamics in systems with multiple time scales and memory friction. *J. Chem. Phys.* **95** (1991), no. 6, 4389–4396.
- [36] Van Gunsteren, W. F.; Berendsen, H. J. C. Algorithms for Brownian dynamics. *Molecular Phys.* **45** (1982), no. 3, 637–647.
- [37] Vanden-Eijnden, E.; Ciccotti, G. Second-order integrators for Langevin equations with holonomic constraints. *Chem. Phys. Lett.* **429** (2006), no. 1-3, 310–316.
- [38] Wendlandt, J. M.; Marsden, J. E. Mechanical integrators derived from a discrete variational principle. *Phys. D* **106** (1997), no. 3-4, 223–246.

NAWAF BOU-RABEE
Courant Institute
251 Mercer Street
New York, NY 10012
E-mail: nawaf@cims.nyu.edu

ERIC VANDEN-EIJNDEN
Courant Institute
251 Mercer Street
New York, NY 10012
E-mail: eve2@cims.nyu.edu

Received May 2009.