Non-asymptotic mixing of the MALA algorithm

August 20, 2010

N. Bou-Rabee[1], **M. Hairer**[1,2], **E. Vanden-Eijnden**[1]

[1] Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012-1185
Email: nawaf@cims.nyu.edu, mhairer@cims.nyu.edu, eve2@cims.nyu.edu
[2] Mathematics Institute, The University of Warwick, Coventry CV4 7AL, U.K.
Email: M.Hairer@Warwick.ac.uk

Abstract

The Metropolis-Adjusted Langevin Algorithm (MALA), originally introduced to sample exactly the invariant measure of certain stochastic differential equations (SDE) on infinitely long time intervals, can also be used to approximate pathwise the solution of these SDEs on finite time intervals. However, when applied to an SDE with a nonglobally Lipschitz drift coefficient, the algorithm may not have a spectral gap even when the SDE does. This paper reconciles MALA's lack of a spectral gap with its ergodicity to the invariant measure of the SDE and finite time accuracy. In particular, the paper shows that its convergence to equilibrium happens at exponential rate up to terms exponentially small in time-stepsize. This quantification relies on MALA's ability to exactly preserve the SDE's invariant measure and accurately represent the SDE's transition probability on finite time intervals.

Keywords: Stochastic Differential Equations, Metropolis-Hastings algorithm, Weak Accuracy, Spectral Gap, Geometric Ergodicity

Subject classification: 60J05 (65C30, 65C05)

1 Introduction

The Metropolis-Adjusted Langevin Algorithm (MALA), originally proposed by Roberts and Tweedie [RT96b, RT96a], is a technique to sample exactly complex, high-dimensional probability distributions. MALA fits the general framework of the Metropolis-Hastings method [MRTT53, Has70] and can be viewed as a special case of smart and hybrid Monte-Carlo algorithms [RDF78, DKPR87]. The main idea of MALA is to obtain the proposal moves from the forward Euler discretization of an SDE whose invariant measure is the target distribution one seeks to sample. Besides being ergodic with respect to this invariant measure by construction, it was shown recently that MALA also captures the dynamical behavior of the solutions to the SDE [BV10]. Therefore MALA has the nice feature that it can be used to estimate finite time dynamical properties along infinitely long trajectories of ergodic SDEs.

Still, one issue with MALA is its theoretical rate of convergence, see for example [RT96a, CWG+08]. When applied to measures with tails that are lighter than Gaussian, it is known that MALA does not exhibit a geometric rate of convergence to equilibrium even though the exact solution to the SDE does. The main reason is that the proposal moves generated by forward Euler are not globally stable. Indeed for any time-stepsize one can find an energy value above which the drift in forward Euler gives proposed moves that increase the energy, in contrast to the exact drift in the SDE which always centers the solution towards lower energy values. Since higher energy values have a lower equilibrium probability weight, these proposed moves are typically rejected. While these rejections ensure that MALA is ergodic, at high energy values they prevent MALA from having a spectral gap.

The question we investigate in this paper is how severe this problem is in practical applications. Above we have argued that the main cause of the lack of geometric convergence is the behavior of the chain at high energy values. Since the chain is unlikely to reach such high energy values over finite time horizons, one does not expect their influence to be significant. In practice, it is the behavior of MALA on finite but very long times that is of interest, since this behavior is what one would experience when running the algorithm on a computer. The goal of this article is to quantify the non-asymptotic behavior of MALA.

The main result of this paper states that the convergence of MALA to its equilibrium distribution happens at exponential rate up to terms exponentially small in time-stepsize. This can be formulated in the following way, and will later be reformulated rigorously as Theorem 3.1:

Claim. *Let P_h^n denote the n-step transition probability of MALA and μ its equilibrium measure. Set $P = P_h^{\lfloor 1/h \rfloor}$. Under natural assumptions on the target distribution $\mu(d\boldsymbol{x}) = Z^{-1}\exp(-U(\boldsymbol{x}))\,d\boldsymbol{x}$ (see Assumption 2.1), for h small enough and for all $\boldsymbol{x} \in \mathbb{R}^n$ satisfying $U(\boldsymbol{x}) < E_0$ there exist positive constants $\varrho \in (0,1)$, $C_1(E_0)$ and C_2 independent of h such that the bound*

$$\|P^k(\boldsymbol{x},\cdot) - \mu\|_{\mathrm{TV}} \le C_1(E_0)(\varrho^k + e^{-C_2/h^{1/4}})\,, \tag{1.1}$$

holds for all $k \in \mathbb{N}$.

Observe from (1.1) that the distance of MALA to equilibrium is bounded by the sum of two terms. The first term converges to 0 exponentially fast and essentially gives the speed of convergence to equilibrium for the exact solution to the underlying SDE. The second term on the other hand remains bounded away from 0 as $k \to \infty$. This term arises from the lack of a spectral gap in MALA, but its important feature is that it is exponentially small in h. Therefore, its importance will be negligible in applications for most practical purposes.

The crux of the proof is the demonstration that MALA inherits some of the convergence properties of the solution to the underlying SDE up to exponentially small terms. This proof relies on finite time accuracy of MALA, ergodicity of MALA with respect to the exact equilibrium measure of the SDE, and an application of Harris' theorem. In fact, if MALA did not exactly preserve the equilibrium measure of the SDE, the second term in (1.1) would not be exponentially small in the time-stepsize. For example, if MALA was replaced simply by the uncorrected Euler approximations to the SDE, then one would expect the size of the error term to be $\mathcal{O}(h)$.

The estimate (1.1) does not imply that MALA does not converge to the equilibrium of the SDE. In fact, it is known [RT96a] that the TV distance between MALA and the equilibrium measure vanishes in the limit as $k \to \infty$. However, this asymptotic property provides no insight on the nonasymptotic behavior of MALA which is the main focus of this paper. In fact, even though the upper bound in (1.1) does not converge to zero in the limit $k \to \infty$, it is the sharpest known bound on finite time intervals.

The power $1/4$ in the exponentially small term in (1.1) is due to the second-order weak accuracy of the proposal moves generated by the forward Euler scheme, and the conditions we impose on the potential energy. In particular, it can be traced back to the appearance of the factor $U^4(\boldsymbol{x})$ appearing in the statement of Lemma 5.3. Under the assumptions made in this paper, this power is sharp.

At the technical level, the main novelty of the proof of our result is twofold. First, we prove finite-time accuracy of MALA in the total variation norm in our setting. While accuracy in total variation of the forward Euler algorithm is known [BT95], it is essential for our analysis to cover situations where the drift of the underlying SDE is not globally Lipschitz continuous. Furthermore, we need to keep track of the dependency of the error estimates with respect to the initial condition. The main idea for this result is to first obtain an error estimate in some weaker Wasserstein distance, and then to strengthen this into a total variation estimate by making use of the regularising properties of the one-step transition probabilities of the forward Euler algorithm. Second, we show that on a very large set, MALA admits a Lyapunov function of the type $\Phi(\boldsymbol{x}) = \exp(\theta U(\boldsymbol{x}))$ for suitable $\theta > 0$. Since U is allowed to grow much faster than quadratically at infinity, this Lyapunov function fails to be integrable with respect to *any* Gaussian measure, including of course the transition probabilities of forward Euler. While this leads to technical complications, having such a fast-growing Lyapunov function is a crucial ingredient of our proof, as this is the key to obtaining bounds that are exponentially small in h.

The remainder of this paper is organized as follows. In Section 2, we will state the main assumptions required for the proof of our main result. Along the way, we recall that MALA is ergodic. In Section 3, the proof of the main result is provided. This proof relies crucially on comparison with a 'patched' MALA algorithm, where the chain is reflected at the boundaries of a large level set. The accuracy of this patched algorithm is investigated in Section 4. Finally, Section 5 shows that Φ is a Lyapunov function for the MALA algorithm (at least on a large domain), which provides the strong *a priori* bounds required for our analysis.

Acknowledgements

We wish to acknowledge stimulating discussions with Jianfeng Lu, Gareth Roberts, Andrew Stuart, and Jonathan Weare. The research of NBR was supported in part by NSF Fellowship DMS-0803095. The research of MH was supported by an EPSRC Advanced Research Fellowship EP/D071593/1 and a Royal Society Wolfson Research Merit Award. The research of EVE was supported in part by NSF grants DMS-0718172 and DMS-0708140, and ONR grant N00014-04-1-6046.

2 A short overview of the MALA algorithm

2.1 Overdamped Langevin equations

In this paper we focus on overdamped Langevin dynamics on an energy landscape defined by a potential energy function $U \in \mathcal{C}^4(\mathbb{R}^n, \mathbb{R})$:

$$d\boldsymbol{Y} = -\nabla U(\boldsymbol{Y})dt + \sqrt{2\beta^{-1}}d\boldsymbol{W}, \quad \boldsymbol{Y}(0) = \boldsymbol{x} \in \mathbb{R}^n \ . \tag{2.1}$$

Here $\nabla U : \mathbb{R}^n \to \mathbb{R}^n$ denotes the gradient of the function U, $\boldsymbol{W}$ 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 stated in Assumption 2.1 below, the solution to (2.1) is geometrically ergodic with an invariant probability measure μ that possesses the following density $\pi(\boldsymbol{x})$ with respect to Lebesgue measure [Has80, RT96a]:

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

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

Before stating assumptions on the potential energy, let us fix some notation. For a function $G \in \mathcal{C}^r(\mathbb{R}^n, \mathbb{R})$ and an integer $r > 1$, let ∇G and $D^r G$ be the gradient and the rth derivative of G, respectively. Let $|\cdot|$ denote the Euclidean vector norm and $\|\cdot\|$ the Frobenius norm. Let $\mathcal{L}$ denote the generator of (2.1) defined for any $G \in C^2(\mathbb{R}^n, \mathbb{R})$ as

$$\mathcal{L}G(\boldsymbol{x}) \stackrel{\text{def}}{=} -\nabla U(\boldsymbol{x}) \cdot \nabla G(\boldsymbol{x}) + \beta^{-1} \Delta G(\boldsymbol{x}) \ . \tag{2.3}$$

For any $t \geq 0$, let Q_t denote the transition probabilities of $\boldsymbol{Y}$. We will generally make an abuse of notation and use the same symbol for a Markov transition kernel and the associated Markov operator. That is, for any measurable bounded function $\varphi : \mathbb{R}^n \to \mathbb{R}$, we define $Q_t \varphi : \mathbb{R}^n \to \mathbb{R}$ as

$$(Q_t \varphi)(\boldsymbol{x}) \stackrel{\text{def}}{=} \int_{\mathbb{R}^n} Q_t(\boldsymbol{x}, d\boldsymbol{y}) \varphi(\boldsymbol{y}) \ .$$

Throughout this article, we will make the following assumptions on the potential energy. Not all of these assumptions will be required for every statement, but we find it notationally convenient to have a single set of assumptions to refer to.

Assumption 2.1. *The potential energy* $U \in \mathcal{C}^4(\mathbb{R}^n, \mathbb{R})$ *satisfies the following.*

A) One has $U(\boldsymbol{x}) \geq 1$ *and, for any* $C > 0$ *there exists an* $E > 0$ *such that*

$$U(\boldsymbol{x}) \geq C(1 + |\boldsymbol{x}|^2)\,,$$

for all $U(\boldsymbol{x}) > E$.

B) There exist constants $c \in (0, \beta)$, $d > 0$ *and* $E > 0$ *such that*

$$\Delta U(\boldsymbol{x}) \leq c|\nabla U(\boldsymbol{x})|^2 - dU(\boldsymbol{x})\,, \tag{2.4}$$

for all $\boldsymbol{x} \in \mathbb{R}^n$ *satisfying* $U(\boldsymbol{x}) > E$.

C) The Hessian of U *is bounded from below in the sense that there exists* $C \geq 0$ *such that*

$$D^2U(\boldsymbol{x})(\boldsymbol{\eta}, \boldsymbol{\eta}) \geq -C|\boldsymbol{\eta}|^2\,,$$

uniformly for all $\boldsymbol{x}, \boldsymbol{\eta} \in \mathbb{R}^n$.

D) There exists a constant $C > 0$ *such that the first four derivatives of the potential energy* $U \in C^4(\mathbb{R}^n, \mathbb{R})$ *are bounded by the potential energy itself, that is*

$$\|D^4U(\boldsymbol{x})\| \vee \|D^3U(\boldsymbol{x})\| \vee \|D^2U(\boldsymbol{x})\| \vee |\nabla U(\boldsymbol{x})| \leq CU(\boldsymbol{x})\,,$$

for all $\boldsymbol{x} \in \mathbb{R}^n$. *Recall, the function* $\vee$ *returns the argument with the maximum value.*

Remark 2.2. It follows immediately from Assumption 2.1 (A) above that exists a constant $E_c > 0$ such that

$$\mu(\{U(\boldsymbol{x}) \geq E\}) \leq e^{-\frac{\beta E}{2}}\,, \tag{2.5}$$

for all $E > E_c$. Indeed, it suffices to note that

$$\begin{aligned}\mu(\{U(\boldsymbol{x}) \geq E\}) &= \frac{1}{Z}\int_{U(\boldsymbol{x})\geq E} e^{-\beta U(\boldsymbol{x})}d\boldsymbol{x} \leq \frac{e^{-3\beta E/4}}{Z}\int_{U(\boldsymbol{x})\geq E} e^{-\beta U(\boldsymbol{x})/4}d\boldsymbol{x}\\ &\leq Ce^{-3\beta E/4} < e^{-\beta E/2}\,,\end{aligned}$$

where the second to last inequality follows from point (A) above, and the last inequality holds for E sufficiently large.

Remark 2.3. The only place where we actually use the fact that $U(\boldsymbol{x})$ grows like $|\boldsymbol{x}|^2$ is in the proof of Lemma 5.3 below. On the other hand, the statement of that approximation result would certainly be true also for potentials that grow slower at ∞. However, such potentials would not be of interest for the present work. Indeed, if the potential grows slower than $|\boldsymbol{x}|^2$ and no slower than $|\boldsymbol{x}|$, then MALA can be shown to be exponentially ergodic, so that the results in this article would be superfluous. If the potential grows slower than $|\boldsymbol{x}|$, then MALA will not be exponentially ergodic because the true solution of the SDE will not be either.

Remark 2.4. Assumption 2.1 (C) is equivalent to the existence of $C > 0$ such that ∇U satisfies the one-sided Lipschitz property

$$\langle -\nabla U(\boldsymbol{x}) + \nabla U(\boldsymbol{y}), \boldsymbol{x} - \boldsymbol{y}\rangle \le C|\boldsymbol{x} - \boldsymbol{y}|^2, \ \ \forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n \;.$$

All of these conditions are satisfied, for example, if U is smooth and $U(\boldsymbol{x}) \approx |\boldsymbol{x}|^\alpha$ with $\alpha > 2$ for large values of $\boldsymbol{x}$. However, they also allow for potentials that have very asymmetric growth at infinity, and they even allow for the potential to grow at exponential speed. As a consequence of Assumption 2.1 (B), one has the following drift condition on the transition probability of the solution.

Lemma 2.5. *Let* $\Theta \colon \mathbb{R}_+ \to \mathbb{R}$ *be a* $\mathcal{C}^2$ *function such that there exist* $u_0 > 0$ *and* $\alpha > 0$ *such that* $\Theta(u) > 0$, $\Theta'(u) > 0$, $u\Theta'(u) > \alpha\Theta(u)$, *and* $\Theta''(u) \le (\beta - c)\Theta'(u)$ *for* $u > u_0$. *(Here, the constant c is the one appearing in (2.4) above.) Then, there exist positive constants* K_Θ *and* γ_Θ *such that*

$$\mathcal{L}(\Theta \circ U) \le K_\Theta - \gamma_\Theta\,(\Theta \circ U)\;.$$

In particular,

$$(Q_t(\Theta \circ U))(\boldsymbol{x}) \le e^{-\gamma_\Theta t}\Theta(U(\boldsymbol{x})) + \frac{K_\Theta}{\gamma_\Theta}(1 - e^{-\gamma_\Theta t}) \tag{2.6}$$

holds for every $t \ge 0$ *and for every* $\boldsymbol{x} \in \mathbb{R}^n$.

Proof. Using the specific form of $\mathcal{L}$, it follows that for $U(\boldsymbol{x}) > u_0$, we have

$$\begin{aligned}
\mathcal{L}(\Theta \circ U) &= (\Theta' \circ U)\,\mathcal{L}U + \frac{1}{\beta}(\Theta'' \circ U)|\nabla U|^2 \\
&= \Big(\frac{(\Theta'' \circ U)}{\beta} - (\Theta' \circ U)\Big)\,|\nabla U|^2 + \frac{(\Theta' \circ U)}{\beta}\Delta U \\
&\le \frac{1}{\beta}((\Theta'' \circ U) - (\beta - c)(\Theta' \circ U))\,|\nabla U|^2 - \frac{d}{\beta}(\Theta' \circ U)U \\
&\le -\frac{d\alpha}{\beta}(\Theta \circ U)\;.
\end{aligned}$$

The result then follows at once from the fact that the condition $u\Theta'(u) > \alpha\Theta(u)$ implies that $\Theta(u) \to \infty$ as $u \to \infty$. □

Remark 2.6. The condition of Lemma 2.5 holds for example for $\Theta(u) = \exp(\theta u)$, provided that $\theta < \beta - c$. It also holds for $\Theta(u) = u^\ell$ for every $\ell > 0$ and for $\Theta(u) = u^\ell \exp(\theta u)$ with the same constraints on ℓ and θ. This will be useful in the sequel. Throughout this article, we will write $\Phi(\boldsymbol{x}) = \exp(\theta U(\boldsymbol{x}))$ for some unspecified $\theta < \beta - c$, so that

$$\mathcal{L}\Phi \le K - \gamma\Phi\;. \tag{2.7}$$

When the precise value of θ matters, we will denote the corresponding function by Φ_θ.

As a consequence of the ellipticity of the SDE (2.1), one has the following minorization condition on the solution's transition probability.

Lemma 2.7. *For every $t > 0$ and $E > 0$, there exists $\epsilon > 0$ such that*

$$\|Q_t(\boldsymbol{x}, \cdot) - Q_t(\boldsymbol{y}, \cdot)\|_{\text{TV}} \leq 2(1-\epsilon)\,, \tag{2.8}$$

for all $\boldsymbol{x}, \boldsymbol{y}$ satisfying $U(\boldsymbol{x}) \vee U(\boldsymbol{y}) < E$.

Remark 2.8. Here and in the sequel, the total variation distance between two probability measures is defined as

$$\|\mu - \nu\|_{\text{TV}} = 2 \sup_A |\mu(A) - \nu(A)| \,,$$

where the supremum runs over all measurable sets. In particular, the total variation distance between two probability measures is two if and only if they are mutually singular.

Proof of Lemma 2.7. It follows from the ellipticity of the equations that there exists a function $q(t, \boldsymbol{x}, \boldsymbol{y})$ smooth in all of its arguments (for $t > 0$) such that the transition probabilities are given by $Q_t(\boldsymbol{x}, d\boldsymbol{y}) = q(t, \boldsymbol{x}, \boldsymbol{y})\, d\boldsymbol{y}$. Furthermore, q is strictly positive (see, e.g., Lemma 2.2 of [Tal02]). Hence, by the compactness of the set $\{\boldsymbol{x} : U(\boldsymbol{x}) < E\}$, one can find a probability measure η and a constant $\epsilon > 0$ such that,

$$Q_t(\boldsymbol{x}, \cdot) > \epsilon \eta(\cdot)$$

for any $\boldsymbol{x}$ satisfying $U(\boldsymbol{x}) < E$. This condition implies the following transition probability $\tilde{Q}_t$ is well-defined:

$$\tilde{Q}_t(\boldsymbol{x}, \cdot) = \frac{1}{1-\epsilon} Q_t(\boldsymbol{x}, \cdot) - \frac{\epsilon}{1-\epsilon} \eta(\cdot)$$

for any $\boldsymbol{x}$ satisfying $\Phi(\boldsymbol{x}) < E$. Therefore,

$$\|Q_t(\boldsymbol{x}, \cdot) - Q_t(\boldsymbol{y}, \cdot)\|_{\text{TV}} = (1-\epsilon) \|\tilde{Q}_t(\boldsymbol{x}, \cdot) - \tilde{Q}_t(\boldsymbol{y}, \cdot)\|_{\text{TV}}$$

for all $\boldsymbol{x}, \boldsymbol{y}$ satisfying $U(\boldsymbol{x}) \vee U(\boldsymbol{y}) < E$. Since the TV norm is bounded by 2, one obtains the desired result. □

Harris' theorem can now be invoked to conclude the transition probability of the true solution converges at a geometric rate to its equilibrium measure. For the reader's convenience, we state the precise version used in this article. For a proof, see the monograph [MT09], or [HM08] for a shorter and somewhat more constructive version. Harris' theorem essentially states that if a Markov chain $\mathcal{P}$ on an arbitrary (Polish) state space $\mathcal{X}$ admits a Lyapunov function such that its sublevel sets are 'small', then it is exponentially ergodic. More precisely, Harris' theorem applies to any Markov chain that satisfies the following assumptions:

Assumption 2.9 (Drift Condition)**.** *There exists a function* $\Phi : \mathcal{X} \to \mathbb{R}^+$ *and constants* $\gamma \in (0, 1)$ *and* $K \geq 0$*, such that the Markov chain* $\mathcal{P}$ *satisfies*

$$(\mathcal{P}\Phi)(\boldsymbol{x}) \leq \gamma\Phi(\boldsymbol{x}) + K \,, \tag{2.9}$$

for all $\boldsymbol{x} \in \mathcal{X}$*.*

Assumption 2.10 (Associated 'Minorization' Condition)**.** *There exists a constant* $\alpha \in (0, 1)$ *so that the Markov chain* $\mathcal{P}$ *satisfies*

$$\|\mathcal{P}(\boldsymbol{x}, \cdot) - \mathcal{P}(\boldsymbol{y}, \cdot)\|_{\mathrm{TV}} \leq 2(1 - \alpha) \,, \tag{2.10}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ *with* $\Phi(\boldsymbol{x}) + \Phi(\boldsymbol{y}) \leq 4K/(1 - \gamma)$*, where* K *and* γ *are the constants from Assumption 2.9.*

Note that in this statement, we have normalised the total variation distance between two probability measures in such a way that it is equal to 2 if and only if the measures are mutually singular. One then has:

Theorem 2.11 (Harris' theorem)**.** *Suppose a Markov chain* $\mathcal{P}(\boldsymbol{x}, dy)$ *on* $\mathbb{R}^n$ *satisfies Assumptions 2.9 and 2.10. Then there exists a unique invariant measure* μ *for* $\mathcal{P}$ *and there are constants* $C > 0$ *and* $\varrho < 1$*, both depending only on the constants* γ*,* K *and* α *appearing in the assumptions, such that*

$$\|\mathcal{P}^n(\boldsymbol{x}, \cdot\,) - \mu\|_{\mathrm{TV}} \leq C\varrho^n\Phi(\boldsymbol{x}) \,,$$

for any $\boldsymbol{x} \in \mathbb{R}^n$*.*

With this tool at hand, we obtain the following exponential ergodicity result for the solutions to (2.1):

Theorem 2.12. *Let* U *be a potential function satisfying Assumption 2.1. Then, for every* $\theta \in (0, \beta - c)$ *there exist positive constants* $\delta \in (0, 1)$ *and* C *such that*

$$\|Q_t^k(\boldsymbol{x}, \cdot) - \mu\|_{\mathrm{TV}} \leq C\delta^k \exp(\theta U(\boldsymbol{x})) \tag{2.11}$$

for all $t > 0$ *and all* $\boldsymbol{x} \in \mathbb{R}^n$*.*

Proof. According to Remark 2.6, for every $\theta \in (0, \beta - c)$, $\exp(\theta U)$ is a Lyapunov function for the Markov chain Q_t. Moreover, by Lemma 2.7 it satisfies a minorization condition on every sublevel set of U. Hence, Harris' theorem implies that (2.11) holds. □

Next we recall some integration strategies for (2.1) and summarize their properties. In particular, we discuss to what extent these strategies preserve the geometric rate of convergence of the true solution.

2.2 Forward Euler

Let the time-stepsize h be given, set $t_k = hk$ for $k \in \mathbb{N}$, and consider the following forward Euler discretization of (2.1):

$$\tilde{\boldsymbol{X}}_{k+1} = \tilde{\boldsymbol{X}}_k - h\nabla U(\tilde{\boldsymbol{X}}_k) + \sqrt{2\beta^{-1}}(\boldsymbol{W}(t_{k+1}) - \boldsymbol{W}(t_k)) \ , \quad \tilde{\boldsymbol{X}}_0 = \boldsymbol{x} \in \mathbb{R}^n \ . \tag{2.12}$$

Here $\tilde{\boldsymbol{X}}_k$ should be viewed as an approximation to $\boldsymbol{Y}_k \stackrel{\text{def}}{=} \boldsymbol{Y}(t_k)$. The iteration rule (2.12) defines a Markov chain that possesses a transition probability with the following smooth, strictly positive transition density:

$$q_h(\boldsymbol{x}, \boldsymbol{y}) = (4\pi\beta^{-1}h)^{-n/2} \exp\left(-\frac{|\boldsymbol{y} - \boldsymbol{x} + h\nabla U(\boldsymbol{x})|^2}{4\beta^{-1}h} \right) . \tag{2.13}$$

Hence, the chain is irreducible with respect to Lebesgue measure.

If ∇U is globally Lipschitz and h is small enough, forward Euler (2.12) can be shown to be exponentially ergodic with respect to a probability distribution that is a first-order approximant to the equilibrium distribution of the SDE (2.1). This property is typically established using a Talay-Tubaro expansion of the global weak error of forward Euler [TT90].

When ∇U is nonglobally Lipschitz, forward Euler is a transient Markov chain for any $h > 0$. In fact, all moments of forward Euler are unbounded on long time-intervals for any initial condition $\boldsymbol{x} \in \mathbb{R}^n$. To be precise for any integer $\ell \geq 1$ and for any $h > 0$

$$\mathbb{E}^{\boldsymbol{x}} |\tilde{\boldsymbol{X}}_k|^{\ell} \to \infty \quad \text{as} \quad k \to \infty \ , \tag{2.14}$$

where $\mathbb{E}^{\boldsymbol{x}}$ denotes the expectation conditional on $\tilde{\boldsymbol{X}}_0 = \boldsymbol{x}$, see e.g. [MSH02, Tal02]. This instability implies that an equilibrium trajectory of forward Euler does not sample any probability distribution. As is well known in the literature, a Metropolis-Hastings method can stochastically stabilize forward Euler.

2.3 MALA Algorithm

A Metropolis-Hastings method is a Monte-Carlo method for producing samples from a known probability distribution [MRTT53, Has70]. The method generates a Markov chain from a given proposal Markov chain as follows. A proposal move is computed according to the proposal chain and accepted with a probability that ensures the Metropolized chain is ergodic with respect to the given probability distribution. Here we shall focus on the Metropolized forward Euler integrator defined in terms of the equilibrium density π (2.2) and the transition density q_h (2.13).

Given a time-stepsize h and input state $\boldsymbol{X}_k$ the algorithm calculates a proposal move using the forward Euler updating scheme in (2.12):

$$\boldsymbol{X}^{\star}_{k+1} = \boldsymbol{X}_k - h\nabla U(\boldsymbol{X}_k) + \sqrt{2\beta^{-1}}(\boldsymbol{W}(t_{k+1}) - \boldsymbol{W}(t_k)) \ , \tag{2.15}$$

and accepts this proposal with a probability

$$\alpha_h(\boldsymbol{x}, \boldsymbol{y}) = 1 \wedge \frac{q_h(\boldsymbol{y}, \boldsymbol{x})\pi(\boldsymbol{y})}{q_h(\boldsymbol{x}, \boldsymbol{y})\pi(\boldsymbol{x})} \ . \tag{2.16}$$

In other words, if $\zeta_k \sim U(0,1)$ is an i.i.d. sequence of uniformly distributed random variables, the update is defined as:

$$\boldsymbol{X}_{k+1} = \begin{cases} \boldsymbol{X}^\star_{k+1} & \text{if } \zeta_k < \alpha_h(\boldsymbol{X}_k, \boldsymbol{X}^\star_{k+1}) \\ \boldsymbol{X}_k & \text{otherwise} \end{cases} \tag{2.17}$$

for $k \in \mathbb{N}$. To be consistent with the literature, we will refer to the Metropolized forward Euler integrator as the Metropolis-adjusted Langevin algorithm (MALA) [RT96b]. We emphasize that MALA is a special case of the smart and hybrid Monte-Carlo algorithms which are older and more general sampling methods, see [RDF78, DKPR87]. By construction, MALA preserves the invariant measure μ of (2.1). This implies for any $g : \mathbb{R}^n \to \mathbb{R}$,

$$\mathbb{E}^\mu(g(\boldsymbol{X}_k)) = \int_{\mathbb{R}^n} g(\boldsymbol{x})\mu(d\boldsymbol{x}), \;\; \forall \;\; k \in \mathbb{N} \,. \tag{2.18}$$

Here $\mathbb{E}^\mu$ denotes expectation conditioned on the initial distribution of the integrator being the equilibrium distribution of the SDE (2.1):

$$\mathbb{E}^\mu(g(\boldsymbol{X}_k)) = \int_{\mathbb{R}^n} \mathbb{E}^{\boldsymbol{x}}(g(\boldsymbol{X}_k))\, \mu(d\boldsymbol{x}), \quad \boldsymbol{X}_0 = \boldsymbol{x} \in \mathbb{R}^n \,.$$

Moreover, it is quite standard to show that MALA gives rise to an ergodic Markov chain. Indeed, denoting by P_h the transition probabilities defined by (2.17), one has

Theorem 2.13 (Roberts and Tweedie, [RT96a]). *Let U be a potential satisfying Assumption 2.1. For any $h > 0$ the k-step transition probability of MALA converges to μ in the total variation metric on probability measures, that is*

$$\lim_{k\to\infty} \|P_h^k(\boldsymbol{x}, \cdot) - \mu\|_{\mathrm{TV}} = 0 \,,$$

for all $\boldsymbol{x} \in \mathbb{R}^n$.

If ∇U is globally Lipschitz and h is small enough, MALA is geometrically ergodic (see Theorem 4.1 of [RT96a]). However, if ∇U is nonglobally Lipschitz, MALA is not geometrically ergodic even though the solution to the SDE is (see Theorem 4.2 of [RT96a]). Specifically, one can prove the following.

Theorem 2.14 (Roberts and Tweedie, [RT96a]). *Let U be a potential satisfying Assumption 2.1. If*

$$\liminf_{|\boldsymbol{x}|\to\infty} \frac{|\nabla U(\boldsymbol{x})|}{|\boldsymbol{x}|} > \frac{2\beta}{h} \tag{2.19}$$

then MALA operated at time-stepsize h is not geometrically ergodic.

If (2.19) holds, the tail of the equilibrium density is no heavier than Gaussian. For example, if $U(x) = x^4/4$ then

$$\liminf_{|x|\to\infty} \frac{|\nabla U(x)|}{|x|} = \infty \ .$$

In this case the theorem states MALA is not geometrically ergodic, in contrast to the true solution of the SDE. The main purpose of this article is to argue that, up to errors that are *exponentially* small in the time-step size h, the convergence of the transition probabilities of MALA towards equilibrium still takes place at an exponential rate. The next section gives a precise statement of this result, as well as an overview of its proof.

3 Main Results

We now state and prove the main result of the paper. Throughout this section, P_h will denote the one-step transition probabilities of the MALA algorithm as defined in Section 2.3 above. We will also use throughout this section the shorthand notation $P = P_h^{\lfloor 1/h \rfloor}$ for the evolution of MALA over one unit of 'physical time'.

Theorem 3.1. *Let U be a potential function satisfying Assumption 2.1, and let P be as above. Then, there exists $\bar{\delta} \in (0,1)$ and, for every $E_0 > 0$, there exist positive constants C_1, C_2, and $h_c(E_0)$ such that MALA's distance to stationarity satisfies*

$$\|P^k(\boldsymbol{x}, \cdot) - \mu\|_{\mathrm{TV}} \leq C_1 \Phi(\boldsymbol{x}) (\bar{\delta}^k + e^{-C_2/h^{1/4}}) \ ,$$

for all $k \in \mathbb{N}$, all stepsizes $h < h_c$, and all $\boldsymbol{x}$ satisfying $U(\boldsymbol{x}) < E_0$.

To quantify MALA's distance to stationarity, $\|P^k(\boldsymbol{x}, \cdot) - \mu\|_{\mathrm{TV}}$, we adopt a patching argument. The point of the patching argument is to use compactness to boost a local property of MALA to a global property. The main ingredient of this argument is a version of MALA with reflection on the boundaries of certain compact sets.

To introduce this patched version of MALA, set $R_h = \{\boldsymbol{x} : U(\boldsymbol{x}) < E_h\}$, where $E_h = E_\star h^{-1/4}$ for a constant $E_\star$ yet to be determined. The 'patched MALA' algorithm is then defined as a Metropolized version of forward Euler with a reflecting boundary condition at the boundary of R_h. This boundary condition is enforced by setting the target distribution in MALA to be the equilibrium distribution μ conditional on being in R_h. This distribution possesses the following density with respect to Lebesgue measure:

$$\bar{\pi}(\boldsymbol{x}) = Z_h^{-1} e^{-\beta U(\boldsymbol{x})} \mathbf{1}_{R_h}(\boldsymbol{x}) \ , \tag{3.1}$$

where $Z_h = \int_{R_h} \exp(-\beta U(\boldsymbol{x})) d\boldsymbol{x}$ and $\mathbf{1}_{R_h}$ is the indicator function for the set $R_h \in \mathbb{R}^n$.

To be more precise, given a time-stepsize h and input state $\bar{\boldsymbol{X}}_k \in R_h$, the algorithm calculates a proposal move using the forward Euler updating scheme in (2.12):

$$\bar{\boldsymbol{X}}^{\star}_{k+1} = \bar{\boldsymbol{X}}_k - h\nabla U(\bar{\boldsymbol{X}}_k) + \sqrt{2\beta^{-1}}(\boldsymbol{W}(t_{k+1}) - \boldsymbol{W}(t_k)) \, , \tag{3.2}$$

and accepts this proposal with a probability

$$\bar{\alpha}_h(\boldsymbol{x}, \boldsymbol{y}) = \begin{cases} 1 \wedge \frac{q_h(\boldsymbol{y},\boldsymbol{x})\pi(\boldsymbol{y})}{q_h(\boldsymbol{x},\boldsymbol{y})\pi(\boldsymbol{x})} & \text{if } \boldsymbol{x} \in R_h \\ 0 & \text{otherwise .} \end{cases} \tag{3.3}$$

In other words, if $\zeta_k \sim U(0,1)$ is an i.i.d. sequence of uniformly distributed random variables, the update is defined as:

$$\bar{\boldsymbol{X}}_{k+1} = \begin{cases} \bar{\boldsymbol{X}}^{\star}_{k+1} & \text{if } \zeta_k < \bar{\alpha}_h(\bar{\boldsymbol{X}}_k, \bar{\boldsymbol{X}}^{\star}_{k+1}) \\ \bar{\boldsymbol{X}}_k & \text{otherwise} \end{cases} \tag{3.4}$$

for $k \in \mathbb{N}$. We stress that patched MALA always remains in R_h since it rejects all moves to R_h^c. Let $\bar{P}_h$ denote the transition probability of patched MALA. Let $\bar{\mu}$ denote the invariant measure of $\bar{P}_h$ with density $\bar{\pi}$. The invariant measures of $\bar{P}_h$ and P_h are related by:

$$\bar{\mu}(A) = \frac{\mu(A \cap R_h)}{\mu(R_h)} \tag{3.5}$$

for all measureable sets A. Set $\bar{P} = \bar{P}_h^{\lfloor 1/h \rfloor}$. With this notation we are ready to prove Theorem 3.1.

Proof of Main Result. This proof relies on Lemmas 3.2 and 3.3 provided below. Using the triangle inequality, we bound the distance of P^k to stationarity by

$$\begin{aligned} \|P^k(\boldsymbol{x},\cdot) - \mu\|_{\text{TV}} &\le \|P^k(\boldsymbol{x},\cdot) - \bar{P}^k(\boldsymbol{x},\cdot)\|_{\text{TV}} + \|\bar{P}^k(\boldsymbol{x},\cdot) - \bar{\mu}\|_{\text{TV}} + \|\bar{\mu} - \mu\|_{\text{TV}} \\ &\overset{\text{def}}{=} I_1 + I_2 + I_3 \, . \end{aligned} \tag{3.6}$$

We now bound all three terms separately.

Lemma 3.2 bounds I_1 in (3.6) using a coupling between MALA and patched MALA, and the coupling characterization of the total variation distance. The lemma states for every $E_0 > 0$ there exist positive constants $\tilde{C}_1$ and h_c such that

$$I_1 = \|P^k(\boldsymbol{x},\cdot) - \bar{P}^k(\boldsymbol{x},\cdot)\|_{\text{TV}} \le \tilde{C}_1 \Phi(\boldsymbol{x}) e^{-\beta E_h} k \, , \tag{3.7}$$

for all $h < h_c$ and every $\boldsymbol{x}$ satisfying $U(\boldsymbol{x}) < E_0$.

Lemma 3.3 bounds I_2 in (3.6) by using Harris' theorem, Theorem 2.11. This lemma relies on a drift and minorization condition for patched MALA. The lemma states that patched MALA is exponentially ergodic, that is, for every $\bar{\delta} \in (\delta, 1)$ and $E_0 > 0$, there exist positive constants C_3 and h_c such that

$$I_2 = \|\bar{P}^k(\boldsymbol{x},\cdot) - \mu\|_{\text{TV}} \le C_3 \Phi(\boldsymbol{x}) \bar{\delta}^k \, , \tag{3.8}$$

for all $h < h_c$ and for all $\boldsymbol{x}$ satisfying $U(\boldsymbol{x}) < E_0$.

To bound I_3, we use the characterisation of $\bar{\mu}$ in (3.5) and the definition of the total variation distance, to get

$$\|\bar{\mu} - \mu\|_{\mathrm{TV}} = 2\mu(R_h^c) \leq 2e^{-\beta E_h/2} \ , \tag{3.9}$$

where we used Remark 2.2 to obtain the inequality.

Combining the bounds (3.7), (3.8) and (3.9) yields

$$\|P^k(\boldsymbol{x}, \cdot) - \mu\|_{\mathrm{TV}} \leq \tilde{C}_1 \Phi(\boldsymbol{x}) e^{-\beta E_h} k + C_3 \Phi(\boldsymbol{x}) \bar{\delta}^k + 2e^{-\beta E_h/2} \ . \tag{3.10}$$

Since the total variation distance between a Markov chain and its invariant measure is nonincreasing in the TV norm, the linear dependence on k can be eliminated as follows. Set $k = \lceil h^{-1/4} \rceil$ in (3.10) to obtain:

$$\tilde{C}_1 \Phi(\boldsymbol{x}) e^{-\beta E_h} \lceil h^{-1/4} \rceil + C_3 \Phi(\boldsymbol{x}) e^{\ln(\bar{\delta}) \lceil h^{-1/4} \rceil} + 2e^{-\beta E_h/2} \ .$$

Since $E_h \propto h^{-1/4}$, there exist positive constants C_1 and C_2 such that

$$\|P^k(\boldsymbol{x}, \cdot) - \mu\|_{\mathrm{TV}} \leq C_1 \Phi(\boldsymbol{x}) (e^{-C_2/h^{1/4}} + \bar{\delta}^k) \ .$$

for all $k \in \mathbb{N}$ and every $\boldsymbol{x}$ satisfying $U(\boldsymbol{x}) < E_0$. This observation concludes the proof. □

The next lemma bounds I_1 in (3.6) using the drift condition obtained in Lemma 3.5.

Lemma 3.2. *Provided that $E_\star$ is sufficiently small there exist positive constants C_1, C_2 and h_c such that*

$$\sup_{t \in [0,T]} \|P_h^{\lfloor t/h \rfloor}(\boldsymbol{x}, \cdot) - \bar{P}_h^{\lfloor t/h \rfloor}(\boldsymbol{x}, \cdot)\|_{\mathrm{TV}} \leq C_1 \Phi(\boldsymbol{x}) e^{-C_2/h^{1/4}} (1 + T)$$

for all $\boldsymbol{x} \in R_h$, every $h < h_c$, and every $T > 0$.

Proof. The measures $P_h(\boldsymbol{x}, \cdot)$ and $\bar{P}_h(\boldsymbol{x}, \cdot)$ are *not* the same, even for a point $\boldsymbol{x} \in R_h$, since their invariant distributions are different. In particular, patched MALA rejects all proposed moves to R_h^c. However, if the input state and proposed move are in R_h, the acceptance probabilities of the two chains are the same. Hence, if we initiate the two chains in R_h, and drive them by the same realization of noise, we obtain a coupling between the two chains such that they are identical up until the first time MALA hits R_h^c. Based on this observation, we obtain a bound on the total variation difference between the transition probabilities of the two chains in the following way.

Let $\{\boldsymbol{X}_k\}$ and $\{\bar{\boldsymbol{X}}_k\}$ be instances of the Markov chains with respective transition probabilities P_h and $\bar{P}_h$, driven by the same realization of the noise $\boldsymbol{W}$, the same realisation of the acceptance variables ζ_k, and with identical initial conditions $\boldsymbol{X}_0 = \bar{\boldsymbol{X}}_0 = \boldsymbol{x} \in R_h$. As argued above, we then have $\boldsymbol{X}_k = \bar{\boldsymbol{X}}_k$ for $k \leq n$ provided that the first time MALA hits R_h^c is greater than n. Let τ_h denote the first time that $\boldsymbol{X}_k$ hits R_h^c. The coupling characterization of the total variation distance implies that,

$$\|P_h^n(\boldsymbol{x},\cdot) - \bar{P}_h^n(\boldsymbol{x},\cdot)\|_{\mathrm{TV}} \le 2\mathbb{P}^{\boldsymbol{x}}(\boldsymbol{X}_n \neq \bar{\boldsymbol{X}}_n) \le 2\mathbb{P}^{\boldsymbol{x}}(\tau_h \le n) \ .$$

At this stage, one of our main ingredients is the fact that the function $\Phi(\boldsymbol{x}) = \exp(\theta U(\boldsymbol{x}))$ is a Lyapunov function for the MALA algorithm, see Proposition 5.2 below. The probability of MALA first hitting R_h^c before time n can therefore be expressed as

$$\mathbb{P}^{\boldsymbol{x}}(\tau_h \le n) = \sum_{k=1}^{n} \mathbb{P}^x(\Phi(\boldsymbol{X}_k) > e^{\theta E_h}) \le e^{-\theta E_h} \sum_{k=1}^{n} \mathbb{E}^x \Phi(\boldsymbol{X}_k)$$

where we made use of Chebychev's inequality. We now note that we can apply Proposition 5.2 since $E_h < h^{-1/2}$ for h sufficiently small. Since $E_h = E_\star h^{-1/4}$, we can make $E_\star$ sufficiently small so that there exists some $\bar{\gamma} > 0$ such that

$$\mathbb{E}^x \Phi(\boldsymbol{X}_1) \le e^{-\bar{\gamma} h} \Phi(\boldsymbol{x}) + Kh \ .$$

Combining this with the previous bound, we obtain

$$\mathbb{P}^{\boldsymbol{x}}(\tau_h \le n) \le e^{-\theta E_h} \sum_{k=1}^{n} (e^{-\bar{\gamma} k h} \Phi(\boldsymbol{x}) + \frac{Kh}{1 - e^{-\bar{\gamma} h}}) \ .$$

Summing over k and using the fact that $E_h \propto h^{-1/4}$ yields the existence of positive constants C_1 and C_2 such that

$$\mathbb{P}^{\boldsymbol{x}}(\tau_h \le n) \le C_1 \Phi(\boldsymbol{x}) e^{-C_2/h^{1/4}} (1 + T) \ , \tag{3.11}$$

which is indeed the desired result. □

The following lemma proves a geometric rate of convergence for the Markov chain $\bar{P}$. Recall $R_h = \{\boldsymbol{x} : U(\boldsymbol{x}) < E_h\}$. The key tool used is Harris' theorem, Theorem 2.11.

Lemma 3.3. *For every $\bar{\delta} \in (\delta, 1)$, there exist positive constants C and h_c such that*

$$\|\bar{P}^k(\boldsymbol{x},\cdot) - \mu\|_{\mathrm{TV}} \le C\Phi(\boldsymbol{x})\bar{\delta}^k$$

for all $\boldsymbol{x} \in R_h$ and $h < h_c$. In particular, $\bar{\delta}$ is independent of time-stepsize.

Proof. To prove this result, we use once again Harris' theorem. The verification of its conditions for the Markov chain $\bar{P}$ is precisely the content of Lemmas 3.4 and 3.5 below. □

In the next lemma, a minorization condition for patched MALA is derived using finite time accuracy of patched MALA in the TV norm (see Lemma 4.1).

Lemma 3.4. *Let U be a potential function satisfying Assumption 2.1. Let ϵ be the constant appearing in the minorization condition of the true solution (see Lemma 2.7), and let $\bar{P}$ be as above. For every $E > 0$ and $\bar{\epsilon} \in (0, \epsilon)$, there exists a positive constant h_c such that*

$$\|\bar{P}(\boldsymbol{x}, \cdot) - \bar{P}(\boldsymbol{y}, \cdot)\|_{\mathrm{TV}} \le 2(1 - \bar{\epsilon}) , \tag{3.12}$$

for all $\boldsymbol{x}, \boldsymbol{y}$ satisfying $U(\boldsymbol{x}) \vee U(\boldsymbol{y}) \le E$ and $h \le h_c$.

Proof. According to Lemma 2.7, the bound (3.12) holds when $\bar{P}$ is replaced by Q_1, the transition probability for the true solution $\boldsymbol{Y}$ at time one. Combining this with Lemma 4.1 below, we thus obtain

$$\begin{aligned}\|\bar{P}(\boldsymbol{x}, \cdot) - \bar{P}(\boldsymbol{y}, \cdot)\|_{\mathrm{TV}} &\le 2(1 - \epsilon) + 2 \sup_{\Phi(\boldsymbol{x}) \le E} \|\bar{P}(\boldsymbol{x}, \cdot) - Q_1(\boldsymbol{x}, \cdot)\|_{\mathrm{TV}} \\ &\le 2(1 - \epsilon) + C(E)\sqrt{h} \ .\end{aligned}$$

Choosing h sufficiently small so that $C(E)\sqrt{h} < 2(\epsilon - \bar{\epsilon})$, the claim follows. □

In the next lemma, we derive a drift condition for patched MALA using its single-step accuracy in representing the Lyapunov function Φ. Deriving this drift condition requires a generalization of Theorem 7.2 in [MSH02] to Lyapunov functions that are neither globally Lipschitz nor essentially quadratic.

Lemma 3.5. *Let U be a potential function satisfying Assumption 2.1 and let γ be the constant appearing in the drift condition (2.7). For every $\bar{\gamma} \in (0, \gamma/2)$, there exist positive constants $E_\star$ and h_c such that*

$$\mathbb{E}^{\boldsymbol{x}}\big(\Phi(\bar{\boldsymbol{X}}_{\lfloor 1/h \rfloor})\big) \le e^{-\bar{\gamma}} \Phi(\boldsymbol{x}) + K , \tag{3.13}$$

for all $\boldsymbol{x} \in R_h$ and all $h < h_c$.

Proof. We will actually show that

$$\mathbb{E}^{\boldsymbol{x}}\big(\Phi(\bar{\boldsymbol{X}}_1)\big) \le (1 - \bar{\gamma} h)\Phi(\boldsymbol{x}) + Kh ,$$

from which the required bound follows by induction, noting that $U(\bar{\boldsymbol{X}}_k) \le E_h$ for every $k > 0$ by construction.

We decompose the expression that we want to bound as

$$\mathbb{E}^{\boldsymbol{x}}\big(\Phi(\bar{\boldsymbol{X}}_1)\big) = \mathbb{E}^{\boldsymbol{x}}\big(\Phi(\bar{\boldsymbol{X}}_1) , \ \bar{\boldsymbol{X}}_1^\star \in R_h\big) + \Phi(\boldsymbol{x})\mathbb{P}^{\boldsymbol{x}}\big(\bar{\boldsymbol{X}}_1^\star \in R_h^c\big) \ .$$

Since

$$\mathbb{E}^{\boldsymbol{x}}\big(\Phi(\bar{\boldsymbol{X}}_1) \mid \bar{\boldsymbol{X}}_1^\star \in R_h\big) = \mathbb{E}^{\boldsymbol{x}}(\Phi(\boldsymbol{X}_1) \mid \boldsymbol{X}_1^\star \in R_h) \le \mathbb{E}^{\boldsymbol{x}}(\Phi(\boldsymbol{X}_1)) ,$$

it follows that

$$\mathbb{E}^{\boldsymbol{x}}\big(\Phi(\bar{\boldsymbol{X}}_1)\big) \le \mathbb{E}^{\boldsymbol{x}}(\Phi(\boldsymbol{X}_1))\mathbb{P}^{\boldsymbol{x}}\big(\bar{\boldsymbol{X}}_1^\star \in R_h\big) + \Phi(\boldsymbol{x})\mathbb{P}^{\boldsymbol{x}}\big(\bar{\boldsymbol{X}}_1^\star \in R_h^c\big) \ .$$

Since $E_h < h^{-1/2}$ for h sufficiently small, we can apply Proposition 5.2 to the first term in this expression, thus obtaining

$$\mathbb{E}^{\boldsymbol{x}}\big(\Phi(\bar{\boldsymbol{X}}_1)\big) \le \Big(1 + \mathbb{P}^{\boldsymbol{x}}\big(\bar{\boldsymbol{X}}_1^\star \in R_h\big)(e^{-\gamma h} - 1 + CE_\star^4 h)\Big)\Phi(\boldsymbol{x}) + Kh\;.$$

By making $E_\star$ sufficiently small, the requested bound now follows, provided that we can find a lower bound on $\mathbb{P}^{\boldsymbol{x}}\big(\bar{\boldsymbol{X}}_1^\star \in R_h\big)$ that is arbitrarily close to $\frac{1}{2}$ for small values of h.

Recall that we have the identity

$$\mathbb{P}^{\boldsymbol{x}}\big(\bar{\boldsymbol{X}}_1^\star \in R_h\big) = (4\pi\beta^{-1}h)^{-n/2}\int_{R_h}\exp\left(-\frac{|\boldsymbol{y}-\boldsymbol{x}+h\nabla U(\boldsymbol{x})|^2}{4\beta^{-1}h}\right)d\boldsymbol{y}\;.$$

Using $(a+b)^2 \le 2a^2 + 2b^2$ and Assumption 2.1 (D), it follows that we can bound this by

$$\begin{aligned}\mathbb{P}^{\boldsymbol{x}}\big(\bar{\boldsymbol{X}}_1^\star \in R_h\big) &\ge (4\pi\beta^{-1}h)^{-n/2}\exp\left(-\beta\frac{hE_h^2}{2}\right)\int_{R_h}\exp\left(-\beta\frac{|\boldsymbol{y}-\boldsymbol{x}|^2}{2h}\right)d\boldsymbol{y}\\ &= 2^{-n/2}\exp\left(-\beta\frac{h^{1/2}E_\star^2}{2}\right)\mathbb{P}(\boldsymbol{\xi}+\boldsymbol{x}\in R_h)\;,\end{aligned} \tag{3.14}$$

where $\boldsymbol{\xi}$ denotes a Gaussian random variable with distribution $\mathcal{N}(0,\beta^{-1}h)$. In order to bound this term, denote by $\boldsymbol{n}(\boldsymbol{x})$ the unit vector opposite the direction of the gradient of U at $\boldsymbol{x}$, i.e. $\boldsymbol{n}(\boldsymbol{x}) = -\nabla U(\boldsymbol{x})/|\nabla U(\boldsymbol{x})|$. We claim that for every $\delta > 0$, there exists $C > 0$ and $E_0 > 0$ such that for every unit vector $\boldsymbol{m}$ with $\langle \boldsymbol{m}, \boldsymbol{n}(\boldsymbol{x})\rangle \ge \delta$, we have $U(\boldsymbol{x}+\kappa\boldsymbol{m}) \le U(\boldsymbol{x})$, provided that $\kappa \le CU(\boldsymbol{x})^{-1/2}$ and $U(\boldsymbol{x}) \ge E_0$.

Indeed, consider the function $f(\kappa) = U(\boldsymbol{x}+\kappa\boldsymbol{m}) - U(\boldsymbol{x})$. Then f is a smooth function such that $f(0) = 0$ and $f'(0) \le -\delta|\nabla U(\boldsymbol{x})| \le -C_1\delta\sqrt{U(\boldsymbol{x})}$ for some C_1 by Assumption 2.1. Furthermore, one has $f''(\kappa) \le C_2 U(\boldsymbol{x})$ for some C_2, as long as $f(\kappa) \le 0$. Combining these, we see that $f'(\kappa) < 0$ (and therefore $f(\kappa) < 0$) for every $\kappa < \delta C_1/(C_2\sqrt{U(\boldsymbol{x})})$, as claimed.

For every $\boldsymbol{x} \in R_h$, we now define a set $A(\boldsymbol{x}) \subset S^{n-1}$ by $A(\boldsymbol{x}) = \{\boldsymbol{m} : \boldsymbol{x}+\kappa\boldsymbol{m} \in R_h\,\forall\kappa \le E_h^{-1}\}$. As a consequence of our previous claim, for any $\alpha < \frac{1}{2}$ there exists h_c such that if $h < h_c$, one has $\inf_{\boldsymbol{x}\in R_h}|A(\boldsymbol{x})|/|S^{n-1}| \ge \alpha$, where $|\cdot|$ denotes the surface measure on the sphere. Denoting by $\boldsymbol{B}(\boldsymbol{x},r)$ the ball of radius r centered at $\boldsymbol{x}$, we conclude that

$$\mathbb{P}(\boldsymbol{\xi}+\boldsymbol{x}\in R_h) \ge \mathbb{P}(\boldsymbol{\xi}+\boldsymbol{x}\in R_h\cap\boldsymbol{B}(\boldsymbol{x},E_h^{-1})) \ge \alpha\mathbb{P}(|\boldsymbol{\xi}| \le h^{1/4}E_\star^{-1})\;,$$

where we used Assumption 2.1 (E) to obtain the last inequality. By making h sufficiently small, this expression can be made arbitrarily close to α, and the prefactor in (3.14) can be made arbitrarily close to 1, thus yielding the required bound. □

4 Accuracy of the Patched MALA Algorithm

When all of the derivatives of U are bounded, accuracy in the total variation distance for forward Euler has been derived using a Talay-Tubaro expansion and Malliavin integration by parts [BT95]; see also [TT90]. In this section we treat the situation where the derivatives of U are unbounded. The order of accuracy obtained below is not sharp, but the proof is constructive and is sufficient for MALA to inherit a minorization condition from the true solution. To sharpen the estimate, retrace the steps of the proof in [BT95] and replace boundedness of the coefficients by some coercivity.

Lemma 4.1. *Let U be a potential satisfying Assumption 2.1. Let $\bar{P}_h$ and Q_h denote the transition probability of patched MALA and the true solution, respectively. Then, for every $T > 0$, there exists $C(T) > 0$ such that for all $h < 1$, the bound*

$$\|\bar{P}_h^{\lfloor t/h \rfloor}(\boldsymbol{x}, \cdot) - Q_h^{\lfloor t/h \rfloor}(\boldsymbol{x}, \cdot)\|_{\text{TV}} \leq C(T)\sqrt{h} U^3(\boldsymbol{x}) , \tag{4.1}$$

is valid for all $\boldsymbol{x} \in \mathbb{R}^n$ and all $t \in [0, T]$.

Proof. This estimate is a consequence of Lemmas 4.2 and 4.6 below. Let $\tilde{P}_h$ denote the transition probability of forward Euler (2.12). The triangle inequality implies that,

$$\begin{aligned} &\|\bar{P}_h^{\lfloor t/h \rfloor}(\boldsymbol{x}, \cdot) - Q_h^{\lfloor t/h \rfloor}(\boldsymbol{x}, \cdot)\|_{\text{TV}} \leq \\ &\qquad \|\bar{P}_h^{\lfloor t/h \rfloor}(\boldsymbol{x}, \cdot) - \tilde{P}_h^{\lfloor t/h \rfloor}(\boldsymbol{x}, \cdot)\|_{\text{TV}} + \|\tilde{P}_h^{\lfloor t/h \rfloor}(\boldsymbol{x}, \cdot) - Q_h^{\lfloor t/h \rfloor}(\boldsymbol{x}, \cdot)\|_{\text{TV}} \ . \end{aligned}$$

According to Lemma 4.6, the first term is bounded by $C(T)\sqrt{h}U^3(\boldsymbol{x})$. According to Lemma 4.2, the second term is bounded by $C(T)\sqrt{h}U^2(\boldsymbol{x})$. Hence, the desired error estimate is obtained. □

Lemma 4.2. *Let U be a potential satisfying Assumption 2.1. Let $\tilde{P}_h$ and Q_h denote the transition probability of forward Euler and the true solution, respectively. Then, for every $T > 0$, there exists $C(T) > 0$ such that for all $h < 1$*

$$\|\tilde{P}_h^{\lfloor t/h \rfloor}(\boldsymbol{x}, \cdot) - Q_h^{\lfloor t/h \rfloor}(\boldsymbol{x}, \cdot)\|_{\text{TV}} \leq C(T)\sqrt{h} U^2(\boldsymbol{x}) ,$$

for all $\boldsymbol{x} \in \mathbb{R}^n$ and all $t \in [0, T]$.

Proof. We bound the TV distance between forward Euler and the true solution using Lemmas 4.3, 4.4, and 4.5 as follows. Using the triangle inequality, we split the quantity that we wish to bound as

$$\begin{aligned} \|\tilde{P}_h^{\lfloor t/h \rfloor}(\boldsymbol{x}, \cdot) - Q_h^{\lfloor t/h \rfloor}(\boldsymbol{x}, \cdot)\|_{\text{TV}} &\leq \|\tilde{P}_h^{\lfloor t/h \rfloor}(\boldsymbol{x}, \cdot) - (\tilde{P}_h \circ Q_h^{\lfloor t/h \rfloor - 1})(\boldsymbol{x}, \cdot)\|_{\text{TV}} \\ &\quad + \|(\tilde{P}_h \circ Q_h^{\lfloor t/h \rfloor - 1})(\boldsymbol{x}, \cdot) - Q_h^{\lfloor t/h \rfloor}(\boldsymbol{x}, \cdot)\|_{\text{TV}} \\ &\stackrel{\text{def}}{=} I_1 + I_2 \ . \end{aligned} \tag{4.2}$$

We can rewrite the first term of (4.2) as

$$I_1 = \mathbb{E}^{\boldsymbol{x}} \|\tilde{P}_h(\tilde{\boldsymbol{X}}_{\lfloor t/h \rfloor -1}, \cdot) - \tilde{P}_h(\boldsymbol{Y}_{\lfloor t/h \rfloor -1}, \cdot)\|_{\mathrm{TV}} ,$$

which, using Lemma 4.4, is bounded by

$$\begin{aligned} I_1 \le & \frac{1}{\sqrt{2\beta^{-1}h}} \mathbb{E}^{\boldsymbol{x}} (|\tilde{\boldsymbol{X}}_{\lfloor t/h \rfloor -1} - \boldsymbol{Y}_{\lfloor t/h \rfloor -1}| \wedge 1) \\ & + \sqrt{\frac{h}{2\beta^{-1}}} \mathbb{E}^{\boldsymbol{x}} (|\nabla U(\tilde{\boldsymbol{X}}_{\lfloor t/h \rfloor -1}) - \nabla U(\boldsymbol{Y}_{\lfloor t/h \rfloor -1})| \wedge 1) . \end{aligned}$$

Strong accuracy of forward Euler in a bounded metric (see Lemma 4.3) then yields

$$I_1 \le C\sqrt{h} U^2(\boldsymbol{x}) .$$

The second term of (4.2) is bounded by

$$I_2 \le \mathbb{E}^{\boldsymbol{x}} (\|\tilde{P}_h(\boldsymbol{Y}_{\lfloor t/h \rfloor -1}, \cdot) - Q_h(\boldsymbol{Y}_{\lfloor t/h \rfloor -1}, \cdot)\|_{\mathrm{TV}}) .$$

From Lemma 4.5 and (2.7), it follows that I_2 is bounded by $ChU^2(\boldsymbol{x})$, and the claim follows. □

Even though forward Euler is numerically unstable for drifts that are not globally Lipschitz, one can prove the following 'strong accuracy' for forward Euler in a bounded metric. As the proof shows, boundedness of the metric plays the role of stability of the numerical scheme.

Lemma 4.3. *Let U be a potential satisfying Assumption 2.1. Let $\tilde{\boldsymbol{X}}$ and $\boldsymbol{Y}$ denote forward Euler and the true solution, respectively. Then, for every $T > 0$ there exists $C(T) > 0$ such that*

$$\mathbb{E}^{\boldsymbol{x}} (|\tilde{\boldsymbol{X}}_{\lfloor t/h \rfloor} - \boldsymbol{Y}(\lfloor t/h \rfloor h)| \wedge 1) \le C(T) h U^2(\boldsymbol{x}) ,$$

holds for all $\boldsymbol{x} \in \mathbb{R}^n$, all $h \le 1$, and all $t \in [0, T]$.

Proof. The proof goes by induction over the number of steps, so let us consider one single step first. We then have

$$\tilde{\boldsymbol{X}}_1 - \boldsymbol{Y}_h = \tilde{\boldsymbol{X}}_0 - \boldsymbol{Y}_0 - \int_0^h (\nabla U(\tilde{\boldsymbol{X}}_0) - \nabla U(\boldsymbol{Y}_s)) \, ds ,$$

so that

$$\begin{aligned} |\tilde{\boldsymbol{X}}_1 - \boldsymbol{Y}_h|^2 = & |\tilde{\boldsymbol{X}}_0 - \boldsymbol{Y}_0|^2 - 2h \langle \tilde{\boldsymbol{X}}_0 - \boldsymbol{Y}_0, \nabla U(\tilde{\boldsymbol{X}}_0) - \nabla U(\boldsymbol{Y}_0) \rangle \\ & + 2 \int_0^h \langle \tilde{\boldsymbol{X}}_0 - \boldsymbol{Y}_0, \nabla U(\boldsymbol{Y}_s) - \nabla U(\boldsymbol{Y}_0) \rangle \, ds \\ & + \left| \int_0^h (\nabla U(\tilde{\boldsymbol{X}}_0) - \nabla U(\boldsymbol{Y}_s)) \, ds \right|^2 . \end{aligned}$$

Together with Remark 2.4, this implies that there exists a constant C such that

$$\begin{aligned}|\tilde{\boldsymbol{X}}_1 - \boldsymbol{Y}_h|^2 \leq (1 + Ch)|\tilde{\boldsymbol{X}}_0 - \boldsymbol{Y}_0|^2 + 2h^2|\nabla U(\tilde{\boldsymbol{X}}_0) - \nabla U(\boldsymbol{Y}_0)|^2 \\ + 2\int_0^h \langle \tilde{\boldsymbol{X}}_0 - \boldsymbol{Y}_0, \nabla U(\boldsymbol{Y}_s) - \nabla U(\boldsymbol{Y}_0)\rangle \, ds \\ + 2h\int_0^h |\nabla U(\boldsymbol{Y}_s) - \nabla U(\boldsymbol{Y}_0)|^2 \, ds \, .\end{aligned} \tag{4.3}$$

Note now that if η is any unit vector in $\mathbb{R}^n$, we have the identity

$$\langle \nabla U(\boldsymbol{Y}_s) - \nabla U(\boldsymbol{Y}_0), \eta\rangle = \int_0^s \mathcal{L}\langle \nabla U, \eta\rangle(\boldsymbol{Y}_r)\, dr + \sqrt{\frac{2}{\beta}}\int_0^s D^2U(\boldsymbol{Y}_r)(\eta, d\boldsymbol{W}_r) \, .$$

Since $\|D^2U\| \leq CU$ and $|\mathcal{L}\langle\nabla U, \eta\rangle| \leq CU^2$, it then follows from Remark 2.6 that there exists a constant C such that

$$\begin{aligned}\mathbb{E}|\nabla U(\boldsymbol{Y}_s) - \nabla U(\boldsymbol{Y}_0)|^2 \leq CsU^4(\boldsymbol{Y}_0)\, , \quad \forall s \leq 1 \, , \\ |\mathbb{E}\langle \eta, \nabla U(\boldsymbol{Y}_s) - \nabla U(\boldsymbol{Y}_0)\rangle| \leq CsU^2(\boldsymbol{Y}_0)\, , \quad \forall s \leq 1 \, .\end{aligned}$$

On the other hand, one also has the bound

$$|\nabla U(\tilde{\boldsymbol{X}}_0) - \nabla U(\boldsymbol{Y}_0)|^2 \leq C|\tilde{\boldsymbol{X}}_0 - \boldsymbol{Y}_0|^2 \exp(C|\tilde{\boldsymbol{X}}_0 - \boldsymbol{Y}_0|)U^2(\boldsymbol{Y}_0)\, ,$$

which follows from Assumption 2.1 (D) and Lemma 5.1 below. In the case where $|\tilde{\boldsymbol{X}}_0 - \boldsymbol{Y}_0| \leq 1$, this yields

$$|\nabla U(\tilde{\boldsymbol{X}}_0) - \nabla U(\boldsymbol{Y}_0)|^2 \leq h^{-1}|\tilde{\boldsymbol{X}}_0 - \boldsymbol{Y}_0|^2 + hCU^4(\boldsymbol{Y}_0)\, .$$

Inserting these bounds into (4.3), we see that there is $C > 0$ such that if $|\tilde{\boldsymbol{X}}_0 - \boldsymbol{Y}_0| \leq 1$, then

$$\mathbb{E}|\tilde{\boldsymbol{X}}_1 - \boldsymbol{Y}_h|^2 \leq (1 + Ch)|\tilde{\boldsymbol{X}}_0 - \boldsymbol{Y}_0|^2 + Ch^3U^4(\boldsymbol{Y}_0)\, .$$

Since on the other hand, one obviously has $\mathbb{E}(|\tilde{\boldsymbol{X}}_1 - \boldsymbol{Y}_h|^2 \wedge 1) \leq 1$, we conclude that

$$\mathbb{E}(|\tilde{\boldsymbol{X}}_1 - \boldsymbol{Y}_h|^2 \wedge 1) \leq (1 + Ch)(|\tilde{\boldsymbol{X}}_0 - \boldsymbol{Y}_0|^2 \wedge 1) + Ch^3U^4(\boldsymbol{Y}_0)\, .$$

The requested bound now follows from the *a priori* bounds on the solution $\boldsymbol{Y}_t$ given by Remark 2.6. □

Lemma 4.4. *Let U be a potential satisfying Assumption 2.1. Let $\tilde{P}_h$ denote the transition probability of forward Euler. For every $h < 1$ and for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$,*

$$\|\tilde{P}_h(\boldsymbol{x}, \cdot) - \tilde{P}_h(\boldsymbol{y}, \cdot)\|_{\mathrm{TV}} \leq \frac{1}{\sqrt{2\beta^{-1}h}}|\boldsymbol{x} - \boldsymbol{y}| + \frac{\sqrt{h}}{\sqrt{2\beta^{-1}}}|\nabla U(\boldsymbol{x}) - \nabla U(\boldsymbol{y})|\, .$$

Proof. Recalling Pinsker's inequality:

$$\|\mathcal{N}(0,\sigma) - \mathcal{N}(\boldsymbol{x},\sigma)\|_{\mathrm{TV}} \le \frac{|\boldsymbol{x}|}{\sqrt{\sigma}}\;,$$

we see that the claim follows from the fact that

$$\tilde{P}_h(\boldsymbol{x},\cdot) = \mathcal{N}(\boldsymbol{x} - h\nabla U(\boldsymbol{x}), 2\beta^{-1}h\mathbf{I})\;,$$

where **I** denotes the identity matrix. □

Lemma 4.5. *Let U be a potential satisfying Assumption 2.1. Let $\tilde{P}_h$ and Q_h denote the transition probability of forward Euler and the true solution, respectively. Then, there exists $C > 0$ such that, for every $h < 1$, the bound*

$$\|\tilde{P}_h(\boldsymbol{x},\cdot) - Q_h(\boldsymbol{x},\cdot)\|_{\mathrm{TV}} \le ChU^2(\boldsymbol{x})\;,$$

holds for all $\boldsymbol{x} \in \mathbb{R}^n$.

Proof. We write $E_0 = U(\boldsymbol{x})$ as a shorthand. The bound is trivial if $E_0^2 h \ge 1$, so we can and will assume in the sequel that $E_0^2 h \le 1$. Recall that the transition probabilities Q_h are generated by the solutions at time h to

$$d\boldsymbol{Y} = -\nabla U(\boldsymbol{Y})\,dt + \sqrt{2\beta^{-1}}d\boldsymbol{W}\;, \qquad \boldsymbol{Y}(0) = \boldsymbol{x}\;, \tag{4.4}$$

whereas the transition probabilities $\tilde{P}_h$ of forward Euler can be interpreted as the solution at time h to

$$d\tilde{\boldsymbol{X}} = -\nabla U(\boldsymbol{x})\,dt + \sqrt{2\beta^{-1}}d\boldsymbol{W}\;, \qquad \tilde{\boldsymbol{X}}(0) = \boldsymbol{x}\;. \tag{4.5}$$

Therefore, the required quantity can be bounded from above by the total variation distance between the measures generated by (4.4) and (4.5) on pathspace between times 0 and h. Since only the drift differs in the SDEs (4.4) and (4.5), Girsanov's theorem can be used to quantify the distance between the laws of the solutions at time h to (4.4) and (4.5).

We first replace the potential U by a modified potential $\tilde{U}$ which is bounded, together with all of its derivatives. Indeed, let $\varphi\colon \mathbb{R}_+ \to \mathbb{R}$ be a smooth increasing function such that $\varphi(x) = x$ for $x \le 2$ and $\varphi(x) = 3$ for $x \ge 4$. With this definition at hand, we set

$$\tilde{U}(\boldsymbol{y}) = U(\boldsymbol{x})\varphi(U(\boldsymbol{y})/U(\boldsymbol{x}))\;.$$

It then follows from Assumption 2.1 (D) that there exists a constant C such that

$$|\tilde{U}(\boldsymbol{y})| + \|D\tilde{U}(\boldsymbol{y})\| + \|D^2\tilde{U}(\boldsymbol{y})\| \le CE_0\;, \tag{4.6}$$

uniformly over all $\boldsymbol{y} \in \mathbb{R}^n$.

Before we proceed, we argue that if we define

$$d\tilde{\boldsymbol{Y}} = -\nabla\tilde{U}(\tilde{\boldsymbol{Y}})\,dt + \sqrt{2\beta^{-1}}d\boldsymbol{W}\;, \qquad \tilde{\boldsymbol{Y}}(0) = \boldsymbol{x}\;, \tag{4.7}$$

then, one has $\mathbb{P}(\exists t \leq h \,:\, \boldsymbol{Y}(t) \neq \tilde{\boldsymbol{Y}}(t)) \leq C E_0^2 h$, so that we can replace U by $\tilde{U}$ in (4.4) without any loss of generality. In order to show this, we note that Lemma 2.5 yields the existence of a constant K such that $M(t) = U(\boldsymbol{Y}(t)) - Kt - U(\boldsymbol{x})$ is a supermartingale with quadratic variation process

$$\langle M, M\rangle(t) = 2\beta^{-1} \int_0^t |\nabla U(\boldsymbol{Y}(s))|^2 \, ds \;.$$

Furthermore, for E_0 sufficiently large (independently of h), one has

$$\mathbb{P}(\exists t \leq h \,:\, \boldsymbol{Y}(t) \neq \tilde{\boldsymbol{Y}}(t)) \leq \mathbb{P}(\sup_{t\leq h} M_t \geq \tfrac{1}{2} U(\boldsymbol{x})) \;.$$

It then follows from the exponential martingale inequality [RY99, p. 153] that, for every $\Lambda > 0$, one has the bound

$$\mathbb{P}(\exists t \leq h \,:\, \boldsymbol{Y}(t) \neq \tilde{\boldsymbol{Y}}(t)) \leq \exp(-U^2(\boldsymbol{x})/(8\Lambda)) + \mathbb{P}(\langle M, M\rangle(h) \geq \Lambda) \;.$$

For $\delta > 0$ sufficiently small, the second term in this expression can then be bounded by

$$\begin{aligned}
\mathbb{P}(\langle M, M\rangle(h) \geq \Lambda) &\leq \exp(-\sqrt{\delta\Lambda h^{-1}})\mathbb{E}\exp\sqrt{\delta h^{-1}\langle M, M\rangle(h)} \\
&\leq \exp(-\sqrt{\delta\Lambda h^{-1}})\frac{1}{h}\int_0^h \mathbb{E}\exp\sqrt{2\beta^{-1}\delta|\nabla U(\boldsymbol{Y}(s))|^2}\, ds \\
&\leq \exp(-\sqrt{\delta\Lambda h^{-1}})\frac{1}{h}\int_0^h \mathbb{E}\exp(C\sqrt{\delta}U(\boldsymbol{Y}(s)))\, ds \\
&\leq C\exp(-\sqrt{\delta\Lambda h^{-1}})\exp(C\sqrt{\delta}U(\boldsymbol{x})) \;.
\end{aligned}$$

Here, we have first used Chebychev's inequality, followed by Jensen's inequality, then Assumption 2.1 (D), and finally Lemma 2.5 with δ small enough. Setting $\Lambda = U^2(\boldsymbol{x})h^{-1/3}$, it follows that for h small enough we actually have $\mathbb{P}(\exists t \leq h \,:\, \boldsymbol{Y}(t) \neq \tilde{\boldsymbol{Y}}(t)) \leq 2\exp(-ch^{-1/3})$ for some positive constant c, which is much better than needed.

We now proceed by comparing the true solution and forward Euler for $\tilde{U}$. Denote now by $\mathcal{Q}_h$ the measure on pathspace generated by (4.7), by $\mathcal{P}_h$ the measure on pathspace generated by solutions to (4.5), and by $\mathcal{W}_h$ Wiener measure on $\mathcal{C}([0,h],\mathbb{R}^d)$ with starting point $\boldsymbol{x}$. It then follows from Girsanov's theorem that

$$\begin{aligned}
\frac{d\mathcal{Q}_h}{d\mathcal{W}_h}(\boldsymbol{W}) &= Z_Q^{-1}\exp\Big(-\frac{1}{\sqrt{2\beta^{-1}}}(\tilde{U}(\boldsymbol{W}_h) - \tilde{U}(\boldsymbol{x})) - \beta\int_0^h G(\boldsymbol{W}_t)\, dt\Big) \;, \\
\frac{d\mathcal{P}_h}{d\mathcal{W}_h}(\boldsymbol{W}) &= Z_P^{-1}\exp\Big(-\frac{1}{\sqrt{2\beta^{-1}}}\nabla\tilde{U}(\boldsymbol{x})^T(\boldsymbol{W}_h - \boldsymbol{x}) - h\beta|\nabla\tilde{U}(\boldsymbol{x})|^2\Big) \;,
\end{aligned}$$

for some normalisation factors Z_P and Z_Q, where the function G is given by

$$G(x) = |\nabla\tilde{U}(x)|^2 - \Delta\tilde{U}(x) \;.$$

(See for example [Elw82, Theorem 11A].) In particular, we have

$$\begin{aligned}\frac{d\mathcal{Q}_h}{d\mathcal{P}_h}(\boldsymbol{W}) = Z_h^{-1} \exp\Big(-\frac{1}{\sqrt{2\beta^{-1}}}(\tilde U(\boldsymbol{W}_h) - \tilde U(\boldsymbol{x}) - \nabla \tilde U(\boldsymbol{x})^T(\boldsymbol{W}_h - \boldsymbol{x}))\\ - \int_0^h \beta\Big(G(\boldsymbol{W}_t) - |\nabla \tilde U(\boldsymbol{x})|^2\Big)\,dt\Big) \stackrel{\text{def}}{=} Z_h^{-1}\exp(\mathcal{D}_h(\boldsymbol{W}))\;,\end{aligned}$$

where the normalisation constant Z_h is given by

$$Z_h = \int \exp(\mathcal{D}_h)\,d\mathcal{P}_h\;.$$

By (4.6), there exists a constant $C > 0$ such that the bound

$$|\mathcal{D}_h(\boldsymbol{W})| \le CE_0(|\boldsymbol{W}_h - \boldsymbol{x}|^2 + E_0 h)\;, \tag{4.8}$$

holds for every $\boldsymbol{W}$. As an immediate consequence, for every $c > 0$, there exists a constant $C > 0$ such that

$$\Big|\log \int \exp(c\mathcal{D}_h)\,d\mathcal{P}_h\Big| \le CE_0^2 h$$

for every $h \le 1$. In particular, one has $Z_h = 1 + \mathcal{O}(E_0^2 h)$ and similarly for Z_h^{-1}. Denote now by B_h the set

$$B_h = \{\boldsymbol{W} \,:\, |\mathcal{D}_h(\boldsymbol{W})| \ge 1\}\;.$$

It follows from the bound (4.8) that $\mathcal{P}_h(B_h) \le C\exp(-c/(hE_0))$ for some $c, C > 0$ and for $hE_0^2 \le 1$.

We conclude that

$$\begin{aligned}\|\mathcal{Q}_h - \mathcal{P}_h\|_{\text{TV}} = \int |1 - Z_h^{-1}\exp(\mathcal{D}_h)|\,d\mathcal{P}_h \le C\int_{B_h^c} |\mathcal{D}_h|\,d\mathcal{P}_h + \mathcal{O}(E_0^2 h)\\ + \int_{B_h} |1 - Z_h^{-1}\exp(\mathcal{D}_h)|\,d\mathcal{P}_h\\ \le \mathcal{O}(E_0^2 h) + \sqrt{\frac{\mathcal{P}_h(B_h)}{Z_h^2}\Big(\int \exp(2\mathcal{D}_h)\,d\mathcal{P}_h - 1\Big)} = \mathcal{O}(E_0^2 h)\;,\end{aligned}$$

as required. In the last step, we have used the Cauchy-Schwarz inequality. □

Lemma 4.6. *For every $T > 0$, there exists a $C(T) > 0$ such that*

$$\sup_{t\in[0,T]} \|\bar P_h^{\lfloor t/h\rfloor}(\boldsymbol{x},\cdot) - \tilde P_h^{\lfloor t/h\rfloor}(\boldsymbol{x},\cdot)\|_{\text{TV}} \le C(T)\sqrt{h}U^3(\boldsymbol{x})$$

holds for every $h < 1$ and for every $\boldsymbol{x} \in \mathbb{R}^n$.

Proof. Denote by $\tilde{\boldsymbol{X}}_k$ the solution to the forward Euler algorithm after k steps and by $\boldsymbol{X}_k$ the solution to the MALA algorithm. Since both agree until the first time that one step is rejected, it follows from the coupling inequality that we have the bound

$$\|\bar{P}_h^n(\boldsymbol{x},\cdot) - \tilde{P}_h^n(\boldsymbol{x},\cdot)\|_{\mathrm{TV}} \le 2\sum_{k=0}^{n-1} \mathbb{E}^{\boldsymbol{x}}|1-\alpha_h(\tilde{\boldsymbol{X}}_k, \tilde{\boldsymbol{X}}_{k+1})| \;.$$

At this stage, we note that since $\alpha_h \in [0,1]$, it follows from Lemma 5.5 that for every $\alpha > 0$ there exists a $C > 0$ such that the bound

$$\mathbb{E}^{\boldsymbol{x}}|1-\alpha_h(\boldsymbol{x}, \tilde{\boldsymbol{X}}_1)| \le Ch^{3/2}(U(\boldsymbol{x}) \wedge \alpha h^{-1/2})^3 \;,$$

holds for *all* $\boldsymbol{x} \in \mathbb{R}^n$. This is simply because this bound is trivial for $U(\boldsymbol{x}) > 1/\sqrt{h}$.

Making α sufficiently small and combining this with Corollary 5.7, we then obtain

$$\begin{aligned}\mathbb{E}^{\boldsymbol{x}}|1-\alpha_h(\tilde{\boldsymbol{X}}_k, \tilde{\boldsymbol{X}}_{k+1})| &\le Ch^{3/2}\mathbb{E}^{\boldsymbol{x}}(U(\tilde{\boldsymbol{X}}_k) \wedge \alpha h^{-1/2})^3 \\ &\le Ch^{3/2}(U^3(\boldsymbol{x}) + Khk) \;,\end{aligned}$$

for some constant $K > 0$. The claim now follows at once by summing over k. □

5 Local Drift Conditions

This section shows that the single-step accuracy of MALA and forward Euler imply that these algorithms preserve Lyapunov functions of the true solution locally. We refer to this property of a numerical method as a local drift condition. In the lemmas that follow local drift conditions are derived for the MALA and forward Euler algorithms. Deriving such drift conditions requires adapting Theorem 7.2 of [MSH02] to Lyapunov functions that are neither globally Lipschitz nor essentially quadratic. Still, the proofs in this section are strongly inspired by the results in [MSH02].

A key technical issue addressed below is that the natural Lyapunov function of the true solution, namely $\Phi(\boldsymbol{x}) = \exp(\theta U(\boldsymbol{x}))$ grows so fast that it is not in general integrable with respect to a Gaussian measure. In particular, it is not integrable with respect to the transition probabilities of forward Euler. Nevertheless, we will show that the expectation of Φ under one step of MALA is finite and close to the expectation of Φ under the true solution. Integrability of Φ with respect to the transition probability of MALA is a consequence of MALA preserving an equilibrium measure whose tails are lighter than Gaussian.

A first remark which will be useful in this section is that under our assumptions on the potential U, it does not behave 'worse than exponential' in the following sense:

Lemma 5.1. *There exists $C > 0$ such that for every $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, we have*

$$|U(\boldsymbol{x})| \le |U(\boldsymbol{y})| \exp(C|\boldsymbol{x}-\boldsymbol{y}|) \;.$$

Proof. It suffices to differentiate the function $t \mapsto U((1-t)\boldsymbol{x} + t\boldsymbol{y})$, invoke Assumption 2.1 (D), and apply Gronwall's inequality over the interval $t \in [0,1]$. □

Proposition 5.2. *Set $\Phi(\boldsymbol{x}) = \exp(\theta U(\boldsymbol{x}))$. Let $\boldsymbol{X}_1$ denote MALA after one step. Then there exist positive constants C and $\theta \in (0,\beta)$ such that the bound*

$$\mathbb{E}^{\boldsymbol{x}}(\Phi(\boldsymbol{X}_1)) \le (e^{-\gamma h} + CU^4(\boldsymbol{x})h^2)\Phi(\boldsymbol{x}) + \frac{K}{\gamma}(1-e^{-\gamma h})$$

holds for all $\boldsymbol{x} \in \mathbb{R}^n$ satisfying $U(\boldsymbol{x}) < h^{-1/2}$.

Proof. Denoting by $\boldsymbol{Y}(h)$ the true solution after time h, we write

$$\mathbb{E}^{\boldsymbol{x}}(\Phi(\boldsymbol{X}_1)) = \mathbb{E}^{\boldsymbol{x}}(\Phi(\boldsymbol{Y}(h))) + \mathbb{E}^{\boldsymbol{x}}(\Phi(\boldsymbol{X}_1) - \Phi(\boldsymbol{Y}(h))) \ .$$

We know from (2.7) that Φ is a Lyapunov function for the true solution, and hence,

$$\mathbb{E}^{\boldsymbol{x}}(\Phi(\boldsymbol{X}_1)) \le e^{-\gamma h}\Phi(\boldsymbol{x}) + \frac{K}{\gamma}(1-e^{-\gamma h}) + |\mathbb{E}^{\boldsymbol{x}}(\Phi(\boldsymbol{X}_1) - \Phi(\boldsymbol{Y}(h)))| \ .$$

The approximation result between MALA and the true solution given in Lemma 5.3 below then implies the desired result. □

The following lemma states that the single step error of MALA in preserving Φ is $\mathcal{O}(h^2)$ with an error constant that depends on $\Phi(\boldsymbol{y})$ and $U^4(\boldsymbol{y})$ evaluated at the initial condition.

Lemma 5.3. *Set $\Phi(\boldsymbol{x}) = \exp(\theta U(\boldsymbol{x}))$. Let $\boldsymbol{X}_1$ and $\boldsymbol{Y}(h)$ denote MALA and the true solution after one step, respectively. Then there exist positive constants C and $\theta \in (0,\beta)$ such that the bound*

$$|\mathbb{E}^{\boldsymbol{x}}(\Phi(\boldsymbol{X}_1) - \Phi(\boldsymbol{Y}(h)))| \le CU^4(\boldsymbol{x})\Phi(\boldsymbol{x})h^2 \tag{5.1}$$

holds for all $\boldsymbol{x} \in \mathbb{R}^n$ satisfying $U(\boldsymbol{x}) < h^{-1/2}$.

Remark 5.4. Note in particular that the bound (5.1) implies that $\mathbb{E}^{\boldsymbol{x}}\Phi(\boldsymbol{X}_1) < \infty$. This is not obvious *a priori* since $\Phi(\boldsymbol{x})$ grows faster than $\exp|\boldsymbol{x}|^2$ at infinity. As a consequence, this expectation is infinite under the proposal moves.

Proof. Applying Itô's formula twice to the exact solution yields

$$\mathbb{E}^{\boldsymbol{x}}(\Phi(\boldsymbol{Y}(h))) = \Phi(\boldsymbol{x}) + h(\mathcal{L}\Phi)(\boldsymbol{x}) + h^2\int_0^1 (1-t)\,\mathbb{E}^{\boldsymbol{x}}\big(\mathcal{L}^2\Phi(\boldsymbol{Y}(ht))\big)dt \ , \tag{5.2}$$

where $\mathcal{L}$ denotes the generator as in (2.3).

Setting $\boldsymbol{X}(s) = \boldsymbol{x} + s(\boldsymbol{X}_1^\star - \boldsymbol{x})$, we obtain by a simple application of Taylor's formula the following identity for the application of one step of MALA:

$$\begin{aligned}\mathbb{E}^{\boldsymbol{x}}(\Phi(\boldsymbol{X}_1)) &= \Phi(\boldsymbol{x}) \\ &+ \mathbb{E}^{\boldsymbol{x}}(D\Phi(\boldsymbol{x})(\boldsymbol{X}_1^\star - \boldsymbol{x})\alpha_h(\boldsymbol{x}, \boldsymbol{X}_1^\star)) \\ &+ \frac{1}{2}\mathbb{E}^{\boldsymbol{x}}\big(D^2\Phi(\boldsymbol{x})(\boldsymbol{X}_1^\star - \boldsymbol{x}, \boldsymbol{X}_1^\star - \boldsymbol{x})\alpha_h(\boldsymbol{x}, \boldsymbol{X}_1^\star)\big) \\ &+ \frac{1}{6}\mathbb{E}^{\boldsymbol{x}}\big(D^3\Phi(\boldsymbol{x})(\boldsymbol{X}_1^\star - \boldsymbol{x}, \boldsymbol{X}_1^\star - \boldsymbol{x}, \boldsymbol{X}_1^\star - \boldsymbol{x})\alpha_h(\boldsymbol{x}, \boldsymbol{X}_1^\star)\big) \\ &+ \frac{1}{4!}\int_0^1 (1-t)^4\mathbb{E}^{\boldsymbol{x}}\big(D^4\Phi(\boldsymbol{X}(t))(\boldsymbol{X}_1^\star - \boldsymbol{x})^4\alpha_h(\boldsymbol{x}, \boldsymbol{X}_1^\star)\big)\,dt\,.\end{aligned} \tag{5.3}$$

(Here we interpret $D^4\Phi(\boldsymbol{y})(\boldsymbol{x})^4$ as being the quadrilinear form $D^4\Phi(\boldsymbol{y})$ applied to $(\boldsymbol{x}, \boldsymbol{x}, \boldsymbol{x}, \boldsymbol{x})$.) Subtracting (5.2) from (5.3) and using the definition of the forward Euler proposal move to collect this difference in powers of h, we obtain:

$$\mathbb{E}^{\boldsymbol{x}}(\Phi(\boldsymbol{X}_1) - \Phi(\boldsymbol{Y}(h))) = h^{1/2}I_{1/2} + hI_1 + h^{3/2}I_{3/2} + h^2I_2 + R_2\,. \tag{5.4}$$

Here we have introduced:

$$\begin{aligned}I_{1/2} =& \sqrt{2\beta^{-1}}\mathbb{E}^{\boldsymbol{x}}(\alpha_h(\boldsymbol{x}, \boldsymbol{X}_1^\star)D\Phi(\boldsymbol{x})\boldsymbol{\xi}) \\ I_1 =& -\theta\Phi(\boldsymbol{x})|\nabla U(\boldsymbol{x})|^2\mathbb{E}^{\boldsymbol{x}}(\alpha_h(\boldsymbol{x}, \boldsymbol{X}_1^\star) - 1) \\ &+ \beta^{-1}\mathbb{E}^{\boldsymbol{x}}\big(D^2\Phi(\boldsymbol{x})(\boldsymbol{\xi}, \boldsymbol{\xi})(\alpha_h(\boldsymbol{x}, \boldsymbol{X}_1^\star) - 1)\big) \\ I_{3/2} =& -\sqrt{2\beta^{-1}}\mathbb{E}^{\boldsymbol{x}}\big(D^2\Phi(\boldsymbol{x})(\boldsymbol{\xi}, \nabla U(\boldsymbol{x}))\alpha_h(\boldsymbol{x}, \boldsymbol{X}_1^\star)\big) \\ &+ \frac{1}{6}\Big((2\beta^{-1})^{3/2}\mathbb{E}^{\boldsymbol{x}}\big(D^3\Phi(\boldsymbol{x})(\boldsymbol{\xi}, \boldsymbol{\xi}, \boldsymbol{\xi})\alpha_h(\boldsymbol{x}, \boldsymbol{X}_1^\star)\big)\Big) \\ I_2 =& \frac{1}{2}\big(\mathbb{E}^{\boldsymbol{x}}\big(D^2\Phi(\boldsymbol{x})(\nabla U(\boldsymbol{x}), \nabla U(\boldsymbol{x}))\alpha_h(\boldsymbol{x}, \boldsymbol{X}_1^\star)\big)\big) \\ &- \frac{1}{6}\Big(\mathbb{E}^{\boldsymbol{x}}\Big(D^3\Phi(\boldsymbol{x})(\nabla U(\boldsymbol{x}), (-\sqrt{h}\nabla U(\boldsymbol{x}) + \sqrt{2\beta^{-1}}\boldsymbol{\xi})^2)\alpha_h(\boldsymbol{x}, \boldsymbol{X}_1^\star)\Big)\Big) \\ &- \frac{1}{6}\Big(\sqrt{2\beta^{-1}}\mathbb{E}^{\boldsymbol{x}}\Big(D^3\Phi(\boldsymbol{x})(\boldsymbol{\xi}, \nabla U(\boldsymbol{x}), -\sqrt{h}\nabla U(\boldsymbol{x}) + \sqrt{2\beta^{-1}}\boldsymbol{\xi})\alpha_h(\boldsymbol{x}, \boldsymbol{X}_1^\star)\Big)\Big) \\ &- \frac{1}{6}\big(2\beta^{-1}\mathbb{E}^{\boldsymbol{x}}\big(D^3\Phi(\boldsymbol{x})(\boldsymbol{\xi}, \boldsymbol{\xi}, \nabla U(\boldsymbol{x}))\alpha_h(\boldsymbol{x}, \boldsymbol{X}_1^\star)\big)\big) \\ &+ \int_0^1 (1-t)\mathbb{E}^{\boldsymbol{x}}\big(\mathcal{L}^2\Phi(\boldsymbol{Y}(ht))\big)dt \\ R_2 =& \frac{1}{4!}\int_0^1 (1-t)^4\mathbb{E}^{\boldsymbol{x}}\big(D^4\Phi(\boldsymbol{X}(t))(\boldsymbol{X}_1^\star - \boldsymbol{x})^4\alpha_h(\boldsymbol{x}, \boldsymbol{X}_1^\star)\big)\,.\end{aligned}$$

We now bound each of these terms separately. The estimates that follow will often rely on the hypothesis that $U(\boldsymbol{x}) < 1/\sqrt{h}$ together with Assumption 2.1 (D) which implies that the ℓth derivative of Φ satisfies:

$$\|D^\ell\Phi(\boldsymbol{x})\| \leq CU^\ell(\boldsymbol{x})\Phi(\boldsymbol{x})\,, \tag{5.5}$$

for $\ell = 1, 2, 3, 4$.

Since the term $I_{1/2}$ in (5.4) involves an odd function of $\boldsymbol{\xi}$, one can rewrite it as

$$\frac{I_{1/2}}{\sqrt{2\beta^{-1}}} = \mathbb{E}^{\boldsymbol{x}}(\alpha_h(\boldsymbol{x}, \boldsymbol{X}_1^\star) D\Phi(\boldsymbol{x})\boldsymbol{\xi}) = \mathbb{E}^{\boldsymbol{x}}((\alpha_h(\boldsymbol{x}, \boldsymbol{X}_1^\star) - 1) D\Phi(\boldsymbol{x})\boldsymbol{\xi}) \ .$$

Using (5.5), we infer that

$$|I_{1/2}| \leq CU(\boldsymbol{x})\Phi(\boldsymbol{x})\mathbb{E}^{\boldsymbol{x}}\big(|1 - \alpha_h(\boldsymbol{x}, \boldsymbol{X}_1^\star)|^2\big)^{1/2} \leq \tilde{C} U^3(\boldsymbol{x})\Phi(\boldsymbol{x}) h^{3/2} \ ,$$

where we used Lemma 5.5 in the last inequality. One can similarly bound $I_{3/2}$ since it also involves an odd function of $\boldsymbol{\xi}$. The term I_1 is of the form where Lemma 5.5 can be directly applied after using the Cauchy-Schwarz inequality and Assumption 2.1 (D). The terms in I_2 without integrals are bounded in a similar fashion, but without the need to invoke Lemma 5.5.

Note now that

$$|\mathcal{L}^2\Phi(\boldsymbol{y})| \leq CU^4(\boldsymbol{y})\Phi(\boldsymbol{y}) \ ,$$

which is a Lyapunov function for the true solution by Remark 2.6, so that the integrand appearing in I_2 is bounded by:

$$|\mathbb{E}^{\boldsymbol{x}}\big(\mathcal{L}^2\Phi(\boldsymbol{Y}(r))\big)| \leq C\mathbb{E}^{\boldsymbol{x}}\big(U^4(\boldsymbol{Y}(r))\Phi(\boldsymbol{Y}(r))\big) \leq CU^4(\boldsymbol{x})\Phi(\boldsymbol{x}) \ . \tag{5.6}$$

Finally, we describe how to bound R_2 in (5.4). It follows from (5.5) that

$$\begin{aligned} R_2 &= \frac{1}{4!}\left|\int_0^1 (1-t)^4 \mathbb{E}^{\boldsymbol{x}}\big(D^4\Phi(\boldsymbol{X}(t))(\boldsymbol{X}_1^\star - \boldsymbol{x})^4 \alpha_h(\boldsymbol{x}, \boldsymbol{X}_1^\star)\big)\right| \\ &\leq C\mathbb{E}^{\boldsymbol{x}}\big(U^4(\boldsymbol{X}(s))\Phi(\boldsymbol{X}(s))|\boldsymbol{X}_1^\star - \boldsymbol{x}|^4 \alpha_h(\boldsymbol{x}, \boldsymbol{X}_1^\star)\big) \ , \end{aligned}$$

so that our claim follows if we can show that the bound

$$\mathbb{E}U^4(\boldsymbol{X}(s))\Phi(\boldsymbol{X}(s))|\boldsymbol{X}^\star(\boldsymbol{\xi}) - \boldsymbol{x}|^4 \alpha_h(\boldsymbol{x}, \boldsymbol{X}^\star(\boldsymbol{\xi})) \leq CU^4(\boldsymbol{x})\Phi(\boldsymbol{x})h^2 \ , \tag{5.7}$$

holds uniformly for $s \in [0, 1]$, where $\boldsymbol{\xi}$ is a normally distributed random variable. Here, we have introduced the shorthand notation

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

Note that for all $\boldsymbol{x}$ satisfying $U(\boldsymbol{x}) < 1/\sqrt{h}$, we have the bound

$$|\boldsymbol{X}^\star(\boldsymbol{\xi}) - \boldsymbol{x}| \leq C\sqrt{h}(1 + |\boldsymbol{\xi}|) \ . \tag{5.8}$$

Hence, to prove (5.7) it suffices to show that

$$\mathbb{E}\big((1 + |\boldsymbol{\xi}|^4)U^4(\boldsymbol{X}(s))\Phi(\boldsymbol{X}(s))\alpha_h(\boldsymbol{x}, \boldsymbol{X}^\star(\boldsymbol{\xi}))\big) \leq CU^4(\boldsymbol{x})\Phi(\boldsymbol{x}) \ .$$

We can then use the Cauchy-Schwarz inequality to get rid of the factor $(1 + |\boldsymbol{\xi}|^4)$, so that it suffices to show that

$$\mathbb{E}(F_\theta(U(\boldsymbol{X}(s)))\alpha_h(\boldsymbol{x}, \boldsymbol{X}^\star(\boldsymbol{\xi}))) \leq CF_\theta(U(\boldsymbol{x})) \ , \tag{5.9}$$

where we defined the shorthand notation

$$F_\theta(u) = u^8 e^{2\theta u} \;.$$

Our next step is to turn the occurrences of $\boldsymbol{X}(s)$ in this expression into $\boldsymbol{X}^\star(\boldsymbol{\xi})$. In order to do this, we use the fact that Assumption 2.1 (C) implies that U is 'almost' convex. Indeed, choose any $x, y \in \mathbb{R}^n$ and set $x_s = (1-s)x + sy$, so that one has the identity

$$U(x_s) = (1-s)U(x) + sU(y) + s(1-s)\int_0^1 \langle \nabla U(x_{st}) - \nabla U(x_{st+1-t}), y-x\rangle\, dt \;.$$

Since $x_{st+1-t} - x_{st} = (1-t)(y-x)$, it then follows from Assumption 2.1 (C) that

$$\begin{aligned} U(x_s) &\le (1-s)U(x) + sU(y) + Cs(1-s)|x-y|^2 \int_0^1 (1-t)\, dt \\ &\le (1-s)U(x) + sU(y) + C|x-y|^2 \;, \end{aligned} \tag{5.10}$$

for some constant C independent of $s \in [0,1]$. Note also that there exists a constant C such that the bound

$$F_\theta(u+v) \le C F_\theta(u) \exp(Cv) \;, \tag{5.11}$$

holds uniformly for all u, v such that $u \ge 1$ and $v \ge 0$.

Since furthermore F_θ is convex, we deduce from (5.11) and (5.10) that

$$F_\theta(U(\boldsymbol{X}(s))) \le C \exp(C|\boldsymbol{X}^\star(\boldsymbol{\xi}) - \boldsymbol{x}|^2)((1-s)F_\theta(U(\boldsymbol{x})) + sF_\theta(U(\boldsymbol{X}^\star(\boldsymbol{\xi})))) \;.$$

To bound the first term in this expression, note that it follows from (5.8) that

$$\mathbb{E} \exp\big(C|\boldsymbol{X}^\star(\boldsymbol{\xi}) - \boldsymbol{x}|^2\big) \le \mathbb{E}\exp\big(Ch(1+|\boldsymbol{\xi}|^2)\big) \le C \;,$$

so that it is bounded by some multiple of $F_\theta(U(\boldsymbol{x}))$.

Combining this bound with (5.9) and the Cauchy-Schwarz inequality, we conclude that it remains to show that

$$\mathbb{E}\big(F_\theta^2(U(\boldsymbol{X}^\star(\boldsymbol{\xi})))\alpha_h(\boldsymbol{x}, \boldsymbol{X}^\star(\boldsymbol{\xi}))\big) \le C F_\theta^2(U(\boldsymbol{x})) \;.$$

Since α_h is bounded, we can reduce ourselves to showing that

$$\mathbb{E}\big(F_\theta^2(U(\boldsymbol{X}^\star(\boldsymbol{\xi})))\alpha_h(\boldsymbol{x}, \boldsymbol{X}^\star(\boldsymbol{\xi}))\,,\; U(\boldsymbol{X}^\star(\boldsymbol{\xi})) \ge U(\boldsymbol{x})\big) \le C F_\theta^2(U(\boldsymbol{x})) \;. \tag{5.12}$$

Note now that one has from the definition (2.16) of α_h the bound

$$\alpha_h(\boldsymbol{x}, \boldsymbol{y}) \le \frac{q_h(\boldsymbol{y}, \boldsymbol{x})}{q_h(\boldsymbol{x}, \boldsymbol{y})} \exp(\beta U(\boldsymbol{x}) - \beta U(\boldsymbol{y})) \;,$$

where q_h denotes the one-step transition probabilities for forward Euler. The left-hand side of (5.12) can therefore be bounded by

$$\int_{U(\boldsymbol{y})\geq U(\boldsymbol{x})} F_\theta^2(U(\boldsymbol{y}))q_h(\boldsymbol{y},\boldsymbol{x})\exp(\beta U(\boldsymbol{x})-\beta U(\boldsymbol{y}))\,d\boldsymbol{y}\;.$$

We break this integral into two regions by setting

$$\mathcal{R}_1=\{\boldsymbol{y}\,:\,U(\boldsymbol{x})\leq U(\boldsymbol{y})\leq \alpha h^{-1/2}\}\;,\quad \mathcal{R}_2=\{\boldsymbol{y}\,:\,U(\boldsymbol{y})\geq \alpha h^{-1/2}\}\;,$$

for some $\alpha > 0$ to be determined.

Observe now that for $\boldsymbol{y}\in\mathcal{R}_1$, one has the bound

$$\begin{aligned}q_h(\boldsymbol{y},\boldsymbol{x}) &= (4\pi\beta^{-1}h)^{-n/2}\exp\Big(-\frac{\beta}{4h}|\boldsymbol{x}-\boldsymbol{y}+h\nabla U(\boldsymbol{y})|^2\Big)\\ &\leq (4\pi\beta^{-1}h)^{-n/2}\exp\Big(-\frac{\beta}{8h}|\boldsymbol{x}-\boldsymbol{y}|^2+\frac{\beta h}{4}|\nabla U(\boldsymbol{y})|^2\Big)\\ &\leq Ch^{-n/2}\exp\Big(-\frac{\beta}{8h}|\boldsymbol{x}-\boldsymbol{y}|^2\Big)\;, \end{aligned}\tag{5.13}$$

where C depends on the choice of α, but not on h. Furthermore, we have the bound

$$\begin{aligned}&F_\theta^2(U(\boldsymbol{y}))\exp(\beta U(\boldsymbol{x})-\beta U(\boldsymbol{y}))\,d\boldsymbol{y}\\ &\qquad\leq F_\theta^2(U(\boldsymbol{x}))\exp(C|\boldsymbol{x}-\boldsymbol{y}|+(\beta-2\theta)(U(\boldsymbol{x})-U(\boldsymbol{y})))\;,\end{aligned}\tag{5.14}$$

where we have used Lemma 5.1 in order to obtain the last inequality. Combining (5.14) and (5.13) and using the fact that $U(\boldsymbol{y})\geq U(\boldsymbol{x})$ on $\mathcal{R}_1$, we obtain indeed the bound

$$\int_{\mathcal{R}_1} F_\theta^2(U(\boldsymbol{y}))q_h(\boldsymbol{y},\boldsymbol{x})\exp(\beta U(\boldsymbol{x})-\beta U(\boldsymbol{y}))\,dy\leq CF_\theta^2(U(\boldsymbol{x}))\;.$$

Finally, in order to bound the integral over $\mathcal{R}_2$, we make use of the fact that $q_h(\boldsymbol{y},\boldsymbol{x})\leq Ch^{-n/2}$, so that , combining this with (5.14), we have the bound

$$\begin{aligned}&\int_{\mathcal{R}_2} F_\theta^2(U(\boldsymbol{y}))q_h(\boldsymbol{y},\boldsymbol{x})\exp(\beta U(\boldsymbol{x})-\beta U(\boldsymbol{y}))\,d\boldsymbol{y}\\ &\quad\leq Ch^{-n/2}F_\theta^2(\boldsymbol{x})\int_{\mathcal{R}_2}\exp(C|\boldsymbol{x}-\boldsymbol{y}|+(\beta-2\theta)(U(\boldsymbol{x})-U(\boldsymbol{y})))\,d\boldsymbol{y}\\ &\quad\leq Ch^{-n/2}F_\theta^2(\boldsymbol{x})\exp(\beta h^{-1/2})\int_{\mathcal{R}_2}\exp(-\delta U(\boldsymbol{y}))\,d\boldsymbol{y}\;,\end{aligned}$$

for some fixed constant $\delta > 0$. Here, we have made use of the fact that $U(\boldsymbol{x})\leq h^{-1/2}$ by assumption, and that U grows faster than quadratically by Assumption 2.1 (A). It follows from (2.5) and the definition of $\mathcal{R}_2$ that

$$\int_{\mathcal{R}_2}\exp(-\delta U(\boldsymbol{y}))\,dy\leq \exp\Big(-\frac{\beta\delta\alpha}{2}h^{-1/2}\Big)\;,$$

so that the requested bound follows, provided that we choose α sufficiently large so that $\alpha > 2\delta^{-1}$. □

The following lemma is useful to bound the average rejection probability of MALA.

Lemma 5.5 (see also [BV10]). *For every $p \in \mathbb{N}$, there exists an $h_c > 0$ and a constant $C > 0$, such that for any $h < h_c$ the bound*

$$\mathbb{E}^{\boldsymbol{x}}\Big(|1-\alpha_h(\boldsymbol{x}, \boldsymbol{X}_1^\star)|^p\Big) \le CU^{2p}(\boldsymbol{x})h^{3p/2}$$

holds for all $\boldsymbol{x} \in \mathbb{R}^n$ satisfying $U(\boldsymbol{x}) < h^{-1/2}$.

Proof. Introduce the function $G : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ given by

$$G(\boldsymbol{x}, \boldsymbol{y}) = U(\boldsymbol{y}) - U(\boldsymbol{x}) - \frac{1}{2}\langle \nabla U(\boldsymbol{y}) + \nabla U(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x}\rangle$$
$$+\frac{h}{4}\Big(|\nabla U(\boldsymbol{y})|^2 - |\nabla U(\boldsymbol{x})|^2\Big) ,$$

and the set

$$R(\boldsymbol{x}) = \{\boldsymbol{y} \in \mathbb{R}^n \mid G(\boldsymbol{x}, \boldsymbol{y}) > 1\} .$$

By (2.16) it follows that, for $\boldsymbol{\xi}$ a normally distributed random variable, one has

$$\mathbb{E}^{\boldsymbol{x}}|1-\alpha_h(\boldsymbol{x}, \boldsymbol{X}_1^\star)|^p = \mathbb{E}|1-(1 \wedge \exp(-\beta G(\boldsymbol{x}, \boldsymbol{X}^\star(\boldsymbol{\xi})))|^p ,$$

where we have used the shorthand notation

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

Since $|1-(1 \wedge e^{-x})| \le |x|$ for every $x \in \mathbb{R}$, it follows that

$$\mathbb{E}^{\boldsymbol{x}}|1-\alpha_h(\boldsymbol{x}, \boldsymbol{X}_1^\star)|^p \le \mathbb{E}|\beta G(\boldsymbol{x}, \boldsymbol{X}^\star(\boldsymbol{\xi}))|^p .$$

Introduce the interpolant

$$\boldsymbol{X}(t) = \boldsymbol{x} - t(h\nabla U(\boldsymbol{x}) - \sqrt{2h\beta^{-1}}\boldsymbol{\xi}) ,$$

so that $\boldsymbol{X}(0) = \boldsymbol{x}$ and $\boldsymbol{X}(1) = \boldsymbol{X}^\star(\boldsymbol{\xi})$. An straightforward but tedious calculation yields the identity

$$G(\boldsymbol{x}, \boldsymbol{X}^\star(\boldsymbol{\xi})) = \frac{h}{2}\int_0^1 D^2U(\boldsymbol{X}(t))(\nabla U(\boldsymbol{X}(t)), \boldsymbol{X}^\star(\boldsymbol{\xi}) - \boldsymbol{x})dt$$
$$+\frac{1}{2}\int_0^1 t(t-1)\, D^3U(\boldsymbol{X}(t))(\boldsymbol{X}^\star(\boldsymbol{\xi}) - \boldsymbol{x})^3\, dt .$$

(Here we interpret $D^3U(x)y^3$ as being the trilinear form $D^3U(x)$ applied to the triple (y, y, y).) Note now that for all $\boldsymbol{x}$ satisfying $U(\boldsymbol{x}) < 1/\sqrt{h}$, we have the bound

$$|\boldsymbol{X}^\star(\boldsymbol{\xi}) - \boldsymbol{x}| \le C\sqrt{h}(1 + |\boldsymbol{\xi}|) .$$

On the other hand, we know from Assumption 2.1 (D) that $\|D^{(k)}U(x)\| \leq CU(x)$ for $k = 1, 2, 3$ and it follows from Lemma 5.1 that

$$U(\boldsymbol{X}(r)) \leq \exp\Big(C\sqrt{h}(1 + |\boldsymbol{\xi}|)\Big)U(\boldsymbol{x}) \,.$$

for all $r \in [0, 1]$. Combining these bounds, we obtain

$$|G(\boldsymbol{x}, \boldsymbol{X}^\star(\boldsymbol{\xi}))| \leq Ch^{\frac{3}{2}}U^2(\boldsymbol{x})(1 + |\boldsymbol{\xi}|)^3 \exp\Big(C\sqrt{h}(1 + |\boldsymbol{\xi}|)\Big) \,,$$

for some constant $C > 0$. Since the expression involving $\boldsymbol{\xi}$ has moments of all orders that are independent of h, the result follows. □

In the following lemmas we prove a local drift condition for forward Euler. As mentioned the 'strong' Lyapunov function $\Phi(\boldsymbol{y})$ is not integrable with respect to the transition probability of forward Euler since its tails are lighter than Gaussian. But $U(\boldsymbol{y})^\ell$ is integrable as a consequence of Assumption 2.1 (D) which ensures that it grows at most exponentially fast. We will show that single-step accuracy of forward Euler implies that it locally inherits this weaker Lyapunov function.

Lemma 5.6. *Let $\tilde{\boldsymbol{X}}_1$ denote forward Euler after one step. Then there exists a constant $C_\ell > 0$ such that for every $E > 0$ and $\ell \in \mathbb{N}$ the bound*

$$\mathbb{E}^{\boldsymbol{x}}(U^\ell(\tilde{\boldsymbol{X}}_1)) \leq (e^{-\gamma_\ell h} + C_\ell U(\boldsymbol{x})^2 h^2)U^\ell(\boldsymbol{x}) + \frac{K_\ell}{\gamma_\ell}(1 - e^{-\gamma_\ell h})$$

holds for all $\boldsymbol{x} \in \mathbb{R}^n$ satisfying $U(\boldsymbol{x}) < h^{-1/2}$.

Proof. Since

$$U^\ell(\boldsymbol{x}) \leq e^{\ell C|\boldsymbol{x}|}U^\ell(\boldsymbol{0})$$

by Lemma 5.1, $U^\ell(\boldsymbol{x})$ is integrable with respect to Gaussian measures for every $\ell \in \mathbb{N}$. Thus, $(\tilde{P}_h U^\ell)(\boldsymbol{x})$ is finite (recall, $\tilde{P}_h$ is the transition probability for forward Euler).

Denoting by $\boldsymbol{Y}(h)$ the true solution after time h, we write

$$\mathbb{E}^{\boldsymbol{x}}(U^\ell(\tilde{\boldsymbol{X}}_1)) = \mathbb{E}^{\boldsymbol{x}}(U^\ell(\boldsymbol{Y}(h))) + \mathbb{E}^{\boldsymbol{x}}(U^\ell(\tilde{\boldsymbol{X}}_1) - U^\ell(\boldsymbol{Y}(h))) \,.$$

Remark 2.6 then implies that there are positive constants γ_ℓ and K_ℓ such that

$$\mathbb{E}^{\boldsymbol{x}}(U^\ell(\tilde{\boldsymbol{X}}_1)) \leq e^{-\gamma_\ell h}U^\ell(\boldsymbol{x}) + \frac{K_\ell}{\gamma_\ell}(1 - e^{-\gamma_\ell h}) + |\mathbb{E}^{\boldsymbol{x}}(U^\ell(\tilde{\boldsymbol{X}}_1) - U^\ell(\boldsymbol{Y}(h)))| \,,$$

and the approximation result between forward Euler and the true solution given in Lemma 5.8 below implies the desired result. □

An immediate corollary of this bound is given by

Corollary 5.7. *For every $\ell > 1$ there exist positive constants α_ℓ and K_ℓ such that the bound*

$$\mathbb{E}^{\boldsymbol{x}}(U(\tilde{\boldsymbol{X}}_1) \wedge \alpha_\ell h^{-1/2})^\ell \le (U(\boldsymbol{x}) \wedge \alpha_\ell h^{-1/2})^\ell + h\, K_\ell\ ,$$

holds for every $\boldsymbol{x} \in \mathbb{R}^n$.

Proof. It follows from Lemma 5.6 that, provided that α_ℓ is small enough, one has

$$\mathbb{E}^{\boldsymbol{x}}(U^\ell(\tilde{\boldsymbol{X}}_1) \wedge \alpha_\ell h^{-1/2}) \le \mathbb{E}^{\boldsymbol{x}}(U^\ell(\tilde{\boldsymbol{X}}_1)) \le U^\ell(\boldsymbol{x}) + h\, K_\ell\ ,$$

for all $\boldsymbol{x}$ such that $U(\boldsymbol{x}) \le \alpha_\ell h^{-1/2}$. On the other hand, one has the obvious bound

$$\mathbb{E}^{\boldsymbol{x}}(U(\tilde{\boldsymbol{X}}_1) \wedge \alpha_\ell h^{-1/2})^\ell \le (\alpha_\ell h^{-1/2})^\ell\ ,$$

which is valid for all $\boldsymbol{x}$. Collecting both bounds concludes the proof. □

Lemma 5.8. *Let $\tilde{\boldsymbol{X}}_1$ and $\boldsymbol{Y}(h)$ denote forward Euler and the true solution after one step, respectively. For every $\ell \in \mathbb{N}$, there exists a constant $C_\ell > 0$ such that the bound*

$$\left|\mathbb{E}^{\boldsymbol{x}}\Big(U^\ell(\tilde{\boldsymbol{X}}_1) - U^\ell(\boldsymbol{Y}(h))\Big)\right| \le C_\ell h^2 U^{\ell+2}(\boldsymbol{x})$$

holds for all $\boldsymbol{x} \in \mathbb{R}^n$ satisfying $U(\boldsymbol{x}) < h^{-1/2}$ and for all $h < 1$.

Proof. Observe that a single step of forward Euler is equivalent in law to the following Langevin diffusion with constant drift:

$$\tilde{\boldsymbol{X}}_1 \sim \boldsymbol{X}(h)$$

where the process $\boldsymbol{X}$ satisfies

$$d\boldsymbol{X} = -\nabla U(\boldsymbol{x})dt + \sqrt{2\beta^{-1}}d\boldsymbol{W}, \quad \boldsymbol{X}(0) = \boldsymbol{x}\ . \tag{5.15}$$

The infinitesimal generator of this process is given by:

$$(\mathcal{L}_h g)(\boldsymbol{y}) = -\nabla U(\boldsymbol{x})^T \nabla g(\boldsymbol{y}) + \beta^{-1}\Delta g(\boldsymbol{y})\ .$$

Since $\mathcal{L}_h U^\ell(\boldsymbol{x}) = \mathcal{L} U^\ell(\boldsymbol{x})$, an exact Itô-Taylor expansion yields,

$$\mathbb{E}^{\boldsymbol{x}}\Big(U^\ell(\tilde{\boldsymbol{X}}_1) - U^\ell(\boldsymbol{Y}(h))\Big) = \mathbb{E}^{\boldsymbol{x}}\Big(\int_0^h \int_0^s \Big(\mathcal{L}_h^2 U^\ell(\boldsymbol{X}(r)) - \mathcal{L}^2 U^\ell(\boldsymbol{Y}(r))\Big) dr ds\Big)\ .$$

The triangle inequality implies

$$\left|\mathbb{E}^{\boldsymbol{x}}\Big(U^\ell(\tilde{\boldsymbol{X}}_1) - U^\ell(\boldsymbol{Y}(h))\Big)\right| \le \int_0^h \int_0^s \left|\mathbb{E}^{\boldsymbol{x}}\Big(\mathcal{L}_h^2 U^\ell(\boldsymbol{X}(r))\Big)\right| dr\, ds + \int_0^h \int_0^s \left|\mathbb{E}^{\boldsymbol{x}}\Big(\mathcal{L}^2 U^\ell(\boldsymbol{Y}(r))\Big)\right| dr\, ds\ . \tag{5.16}$$

Assumption 2.1 (D) implies that there exists a positive constant C such that

$$(\mathcal{L}^2 U^\ell)(\boldsymbol{y}) \le C U^{\ell+2}(\boldsymbol{y}) \;,\; (\mathcal{L}_h^2 U^\ell)(\boldsymbol{y}) \le C U^2(\boldsymbol{x}) U^\ell(\boldsymbol{y}) \;,$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. These inequalities bound the integrands in (5.16). For the second term, we have

$$\left| \mathbb{E}^{\boldsymbol{x}} \left(\mathcal{L}^2 U^\ell(\boldsymbol{Y}(r)) \right) \right| \le C \mathbb{E}^{\boldsymbol{x}} \left(U^{\ell+2}(\boldsymbol{Y}(r)) \right) \le \tilde{C} U^{\ell+2}(\boldsymbol{x}) \;, \tag{5.17}$$

where Remark 2.6 is used in the last step.

To bound the first term, note that

$$\left| \mathbb{E}^{\boldsymbol{x}} \left(\mathcal{L}_h^2 U^\ell(\boldsymbol{X}(r)) \right) \right| \le C U^2(\boldsymbol{x}) \mathbb{E}^{\boldsymbol{x}} \left(U^\ell(\boldsymbol{X}(r)) \right) \tag{5.18}$$

where $r \in [0, h]$. The definition of an Euler step yields

$$\mathbb{E}^{\boldsymbol{x}} \left(U^\ell(\boldsymbol{X}(r)) \right) = (2\pi)^{-n/2} \int_{\mathbb{R}^n} U^\ell(\boldsymbol{x} - r \nabla U(\boldsymbol{x}) + \sqrt{2\beta^{-1} r} \boldsymbol{\xi}) \exp \left(-\frac{|\boldsymbol{\xi}|^2}{2} \right) d\boldsymbol{\xi} \;.$$

Since, by hypothesis,

$$r |\nabla U(\boldsymbol{x})| \le h |\nabla U(\boldsymbol{x})| \le C h |U(\boldsymbol{x})| \le C \sqrt{h} \;,$$

it follows from Lemma 5.1 that

$$U^\ell(\boldsymbol{x} - r \nabla U(\boldsymbol{x}) + \sqrt{2\beta^{-1} r} \boldsymbol{\xi}) \le \exp \left(\ell \sqrt{h} (C + \sqrt{2\beta^{-1}} |\boldsymbol{\xi}|) \right) U^\ell(\boldsymbol{x})$$

for all $r \in [0, h]$. Therefore,

$$\mathbb{E}^{\boldsymbol{x}} (U^\ell(\boldsymbol{X}(r))) \le C U^\ell(\boldsymbol{x}) \;, \tag{5.19}$$

for some $C > 0$ independent of h. Combining (5.19), (5.18) and (5.17) and inserting these bounds into (5.16) yields the required bound. □

6 Conclusion

In this paper we showed that MALA's lack of a spectral gap is not severe. In particular, our main result, Theorem 3.1, states its convergence to equilibrium happens at exponential rate up to terms exponentially small in time-stepsize. This quantification relies on MALA exactly preserving the SDE's invariant measure and accurately representing the SDE's transition probability on finite time intervals. The first property is automatic since the target distribution in the Metropolis-Hastings step is the SDE's equilibrium distribution. Deriving the second property requires a generalization of finite-time estimates for MALA [BV10] and forward Euler [BT95, MSH02]. This derivation involves obtaining new results on the accuracy of MALA and forward Euler with respect to the true solution of the SDE in the context where the drift is not be globally Lipschitz.

A key technical issue addressed in the proof of Theorem 3.1 is that MALA locally inherits a Lyapunov function of the true solution $\Phi(x) = \exp(\theta U(x))$. Since U grows faster than a quadratic function, the function Φ is not integrable with respect to a Gaussian measure including the transition probability of forward Euler. Nevertheless, we prove integrability of Φ with respect to the transition probability of MALA as a consequence of MALA preserving an equilibrium measure whose tails decrease faster than Φ increases.

Finite-time accuracy implied MALA inherits a minorization and local drift condition from the SDE. As a consequence the paper proved that its mixing time is nearby the mixing time of the SDE on compact sets. The patching argument in Theorem 3.1 compares MALA to a version of MALA with reflection on the boundary of these compact sets to boost this local property to a global property plus terms exponentially small in time-stepsize.

Finally, we note that the proof of Lemma 3.2 motivates the following question: is forward Euler a strongly or weakly convergent method on finite time intervals? The answer is no because a necessary condition for a numerical method to converge on finite time intervals is stability which we have shown forward Euler lacks for nonglobally Lipschitz drifts. However, the lemma does motivate using forward Euler as a proposal chain in the Metropolis-Hastings algorithm to sample from the equilibrium measure of the SDE.

References

[BT95] V. BALLY and D. TALAY. The law of the Euler scheme for stochastic differential equations (i):convergence rate of the distribution function. *Probability Theory and Related Fields* **104**, (1995), 43–60.

[BV10] N. BOU-RABEE and E. VANDEN-EIJNDEN. Pathwise accuracy and ergodicity of Metropolized integrators for SDEs. *Comm. Pure and Appl. Math.* **63**, (2010), 655–696.

[CWG+08] F. P. CASEY, J. J. WATERFALL, R. N. GUTENKUNST, C. R. MYERS, and J. P. SETHNA. Variational method for estimating the rate of convergence of Markov-chain Monte-Carlo algorithms. *Phys. Rev. E* **78**, (2008), 046704.

[DKPR87] S. DUANE, A. D. KENNEDY, B. J. PENDLETON, and D. ROWETH. Hybrid Monte Carlo. *Phys. Lett. B* **195**, (1987), 216–222.

[Elw82] K. D. ELWORTHY. *Stochastic differential equations on manifolds*, vol. 70 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1982.

[Has70] W. K. HASTINGS. Monte-Carlo methods using Markov chains and their applications. *Biometrika* **57**, (1970), 97–109.

[Has80] R. Z. HASMINSKII. *Stochastic Stability of Differential Equations*. Sijthoff & Noordhoff, Alphen aan den Rijn, The Netherlands, 1980.

[HM08] M. HAIRER and J. MATTINGLY. Yet another look at Harris' ergodic theorem for Markov chains, 2008. ArXiv:0810.2777 [math.PR].

[MRTT53] N. METROPOLIS, A. W. ROSENBLUTH, A. H. TELLER, and E. TELLER. Equations of state calculations by fast computing machines. *J. Chem. Phys.* **21**, (1953), 1087–1092.

[MSH02] J. C. MATTINGLY, A. M. STUART, and D. J. HIGHAM. Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stoch. Proc. Appl.* **101**, no. 2, (2002), 185–232.

[MT09] S. P. MEYN and R. L. TWEEDIE. *Markov Chains and Stochastic Stability*. Cambridge University Press, New York, NY, 2nd ed., 2009.

[RDF78] P. J. ROSSKY, J. D. DOLL, and H. L. FRIEDMAN. Brownian dynamics as smart Monte Carlo simulation. *J. Chem. Phys.* **69**, (1978), 4628.

[RT96a] G. O. ROBERTS and R. L. TWEEDIE. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2**, (1996), 341–363.

[RT96b] G. O. ROBERTS and R. L. TWEEDIE. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **1**, (1996), 95–110.

[RY99] D. REVUZ and M. YOR. *Continuous martingales and Brownian motion*, vol. 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third ed., 1999.

[Tal02] D. TALAY. Stochastic Hamiltonian systems: Exponential convergence to the invariant measure, and discretization by the implicit Euler scheme. *Markov Processes and Related Fields* **8**, (2002), 1–36.

[TT90] D. TALAY and L. TUBARO. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stoch. Anal. Appl.* **8**, (1990), 94–120.