On the convergence of Hamiltonian Monte Carlo

Alain Durmus *, Éric Moulines ** and Eero Saksman †

LTCI, Telecom ParisTech & CNRS,
46 rue Barrault, 75634 Paris Cedex 13, France.
e-mail: *alain.durmus@telecom-paristech.fr

Centre de Mathématiques Appliquées, UMR 7641,
Ecole Polytechnique,
route de Saclay, 91128 Palaiseau cedex, France.
e-mail: **eric.moulines@polytechnique.edu

University of Helsinki, Department of Mathematics and Statistics,
P.O. Box 68 , FIN-00014 University of Helsinki, Finland
 e-mail: †eero.saksman@helsinki.fi

Abstract: This paper discusses the stability properties of the Hamiltonian Monte Carlo (HMC) algorithm used to sample from a positive target density \( \pi \) on \( \mathbb{R}^d \), with either a fixed or a random number of integration steps. Under mild conditions on the potential \( U \) associated with \( \pi \), we show that the Markov kernel associated to the HMC algorithm is irreducible and recurrent. Under some additional conditions, the Markov kernel may be shown to be Harris recurrent. Besides, verifiable conditions on \( U \) are derived which imply geometric convergence.

AMS 2000 subject classifications: primary 60J05, 60J22, 65C05.
Keywords and phrases: Hamiltonian Monte Carlo, Markov Chain Monte Carlo, irreducibility, geometric ergodicity.

1. Introduction

We consider in this paper the Hamiltonian Monte Carlo (HMC), or Hybrid Monte Carlo algorithm, a Metropolis-Hastings algorithm to sample from a target probability density \( \pi \) on \( \mathbb{R}^d \). This method has been first proposed in the physics literature in [5]. It has been popularized in the computational statistics community by the landmark paper by [11] to infer Bayesian neural networks. Since then, it has been consistently reported in numerous works that HMC achieves better performance than more classical MCMC methods; see among many others [17, 7, 2, 16, 12] and the references therein.

In sharp contrast, theoretical properties of the HMC algorithm are still not completely understood despite several attempts over these past few years; see [9]. The analysis of HMC raises several challenges. First, the proposal of the HMC algorithm may not have a density with respect to the Lebesgue measure. Hence, contrary to most Metropolis-Hasting algorithm, the kernel of the HMC does not define a T-chain. Establishing the irreducibility of the HMC kernel is therefore much more challenging than for classical Metropolis-Hastings type
algorithm. Finally, verifiable conditions - weak enough to hold for most models of interests- under which the HMC kernel is geometric ergodicity are still missing.

The main aim of this paper is to contribute to fill the gap between theory and practice. First, we establish the irreducibility of the HMC algorithm under a general tail condition of the target density. This result follows from a general irreducibility result for iterative Markov models (derived under conditions which are weaker than the ones reported in the literature) which we believe to be of independent interest; see Section 5. Our main tool to establish irreducibility is the degree theory for continuous maps.

Second, we establish the geometric ergodicity of the HMC sampler under the assumptions that \( \log \pi \) is homogeneous outside a ball (or is a perturbation of an homogeneous function) and that the level sets are convex.

The paper is organized as follows. In Section 2, the HMC algorithm is presented and the main notations of the paper are introduced. In Section 3, conditions upon which the HMC kernel is irreducible, recurrent and Harris-recurrent are presented. Comparison with earlier results will be mainly given in this Section. In Section 4, conditions under which the HMC kernel if \( V \)-uniformly geometrically ergodic are developed and discussed. Some general irreducibility results which are of independent interest, are stated in Section 5. The proofs are gathered in Section 6.

**Notations**

Denote by \( \mathbb{R}_+ \) and \( \mathbb{R}_+^* \), the set of non-negative and positive real numbers respectively. Denote by \( \| \cdot \| \) the Euclidean norm on \( \mathbb{R}^d \). Denote by \( B(\mathbb{R}^d) \) the Borel \( \sigma \)-field of \( \mathbb{R}^d \), \( \mathcal{F}(\mathbb{R}^d) \) the set of all Borel measurable functions on \( \mathbb{R}^d \) and for \( f \in \mathcal{F}(\mathbb{R}^d) \), \( \| f \|_\infty = \sup_{x \in \mathbb{R}^d} |f(x)| \). Denote by \( \text{Leb} \) the Lebesgue-measure on \( \mathbb{R}^d \). For \( \mu \) a probability measure on \( (\mathbb{R}^d, B(\mathbb{R}^d)) \) and \( f \in \mathcal{F}(\mathbb{R}^d) \) a \( \mu \)-integrable function, denote by \( \mu(f) \) the integral of \( f \) w.r.t. \( \mu \). For \( f \in \mathcal{F}(\mathbb{R}^d) \), set \( \| f \|_\infty = \sup_{x \in \mathbb{R}^d} |f(x)| \). Let \( V: \mathbb{R}^d \to [1, \infty) \) be a measurable function. For \( f \in \mathcal{F}(\mathbb{R}^d) \), the \( V \)-norm of \( f \) is given by \( \| f \|_V = \| f/V \|_\infty \). For two probability measures \( \mu \) and \( \nu \) on \( (\mathbb{R}^d, B(\mathbb{R}^d)) \), the \( V \)-total variation distance of \( \mu \) and \( \nu \) is defined as

\[
\| \mu - \nu \|_V = \sup_{f \in \mathcal{F}(\mathbb{R}^d), \| f \|_V \leq 1} \left| \int_{\mathbb{R}^d} f(x)d\mu(x) - \int_{\mathbb{R}^d} f(x)d\nu(x) \right|
\]

If \( V \equiv 1 \), then \( \| \cdot \|_V \) is the total variation denoted by \( \| \cdot \|_{\text{TV}} \). For all \( x \in \mathbb{R}^d \) and \( M > 0 \), we denote by \( B(x, M) \), the ball centered at \( x \) of radius \( M \). Let \( M \) be a \( d \times m \)-matrix, then denote by \( M^T \) and \( \det(M) \) (in the case \( m = d \)) the transpose and the determinant of \( M \) respectively. Let \( k \geq 1 \). Denote by \( (\mathbb{R}^d)^{\otimes k} \) the \( k \)th tensor power of \( \mathbb{R}^d \), for all \( x, y \in \mathbb{R}^d \), \( x \otimes y \in (\mathbb{R}^d)^{\otimes 2} \) the tensor product of \( x \) and \( y \), and \( x^{\otimes k} \in (\mathbb{R}^d)^{\otimes k} \) the \( k \)th tensor power of \( x \). For all \( x_1, \ldots, x_k \in \mathbb{R}^d \), set \( |x_1 \otimes \cdots \otimes x_k| = \sup_{i \in \{1, \ldots, k\}} |x_i| \). We let \( \mathcal{L}((\mathbb{R}^d)^{\otimes k}, \mathbb{R}^\ell) \) stand for the set of linear maps from \( (\mathbb{R}^n)^{\otimes k} \) to \( \mathbb{R}^\ell \) and for \( L \in \mathcal{L}((\mathbb{R}^d)^{\otimes k}, \mathbb{R}^\ell) \), we denote by \( \| L \| \) the operator norm of \( L \). Let \( f: \mathbb{R}^d \to \mathbb{R}^\ell \) be a Lipschitz function, namely there
exists $C \geq 0$ such that for all $x, y \in \mathbb{R}^d$, $\|f(x) - f(y)\| \leq C \|x - y\|$. Then we denote $\|f\|_{\text{Lip}} = \inf \{\|f(x) - f(y)\| / \|x - y\| \mid x, y \in \mathbb{R}^d, x \neq y\}$. Let $k \geq 0$ and $U$ be an open subset of $\mathbb{R}^d$. Denote by $C^k(U, \mathbb{R}^\ell)$ the set of all $k$ times continuously differentiable functions from $U$ to $\mathbb{R}^\ell$. Let $\Phi \in C^k(U, \mathbb{R}^\ell)$. Write $J\Phi$ for the Jacobian matrix of $\Phi \in C^1(\mathbb{R}^d, \mathbb{R}^\ell)$, and $D^k\Phi : U \to L((\mathbb{R}^d) \otimes k, \mathbb{R}^\ell)$ for the $k$th differential of $\Phi \in C^k(\mathbb{R}^d, \mathbb{R}^\ell)$. For smooth enough function $f : \mathbb{R}^d \to \mathbb{R}$, denote by $\nabla f$ and $\nabla^2 f$ the gradient and the Hessian of $f$ respectively. Let $A \subset \mathbb{R}^d$. We write $\overline{A}$, $\mathring{A}$ and $\partial A$ for the closure, the interior and the boundary of $A$, respectively.

2. Description of the Hamiltonian Monte Carlo algorithm

Consider a target probability density $\pi$ on $\mathbb{R}^d$ with respect to the Lebesgue measure, defined for all $q \in \mathbb{R}^d$ by

$$\pi(q) = e^{-U(q)} \int_{\mathbb{R}^d} e^{-U(\tilde{q})} d\tilde{q},$$

where $U : \mathbb{R}^d \to \mathbb{R}$ is a continuously differentiable function. Note that this representation implies that the density is nonzero everywhere (this can be relaxed; see [12, Section 5.5.1]).

The key idea behind HMC is to exploit the measure-preserving properties of Hamiltonian flow. For simplicity, we restrict our analysis to the phase space $\mathbb{R}^{2d}$, equipped with the $2d \times 2d$ canonical structure matrix

$$J = \begin{bmatrix} 0_d & I_d \\ -I_d & 0_d \end{bmatrix},$$

where $0_d$ is the $d \times d$ zero matrix and $I_d$ is the $d \times d$ identity matrix. The state of the system consists of the position $q \in \mathbb{R}^d$ and the momentum $p \in \mathbb{R}^d$. The position corresponds to the variables of interest and the momentum, one for each position variable, is an auxiliary variable. The Hamiltonian function $H$ is defined for $(q, p) \in \mathbb{R}^d \times \mathbb{R}^d$ by

$$H(q, p) = U(q) + \|p\|^2 / 2,$$

where $U(q)$ plays the role of the potential and $\|p\|^2 / 2$ stands for the kinetic energy.

The Hamiltonian dynamics $(q(t), p(t))_{t \geq 0}$ is defined as the solution of the homogeneous system of ordinary differential equation on $\mathbb{R}^{2d}$,

$$\begin{cases} \dot{q}(t) &= \frac{\partial H}{\partial p}(q(t), p(t)) = p(t) \\ \dot{p}(t) &= -\frac{\partial H}{\partial q}(q(t), p(t)) = -\nabla U(q(t)) \end{cases}. $$

(3)

Under weak regularity conditions, the Hamiltonian dynamics (3) has a unique solution on $\mathbb{R}_+$ for a given initial condition $(q(0), p(0))$. These solutions have
The volume preservation states that if we apply the mapping \( \varphi \) from the state \((q(t), p(t))\) at time \( t \in \mathbb{R}_+^* \) to the state \((q(t+s), p(t+s))\) = \((\varphi_s^1(q(t), p(t)), \varphi_s^2(q(t), p(t)))\) at time \( t+s \) is one-to-one; its inverse can be obtained by negating \( p \), applying \( \varphi_s \) and negating \( p \) again: for any \( t \in \mathbb{R}_+^* \), \((q(t), p(t)) = (\varphi_s^1(q(t+s), -p(t+s))), -\varphi_s^2(q(t+s), p(t+s))\). This is the reversibility property. The conservation of the Hamiltonian property simply states that Hamiltonian dynamics keeps the Hamiltonian invariant: for any \( s, t \in \mathbb{R}_+^* \), \( H(q(t), p(t)) = H(q(t+s), p(t+s)) \). The volume preservation states that if we apply the mapping \( \varphi_s \) (for any \( s \in \mathbb{R}_+^* \)) to the points in some region \( R \) of the phase space with volume \( V \), the image of \( R \) under \( \varphi_s \) will also have volume \( V \). Volume preservation is a consequence of Hamiltonian dynamics being symplectic: for \( s \in \mathbb{R}_+^* \), \( B_s^T J B_s = J \) where \( B_s \) stands for the Jacobian matrix of the mapping \( \varphi_s \) and \( J \) is the structure matrix (2). Consider the extended target distribution with density

\[
\tilde{\pi}(q, p) \propto \exp(-H(q, p)),
\]

where \( H(q, p) \) is the Hamiltonian defined in (2). Note that reversibility and volume preservation properties of the Hamiltonian flow \((\varphi_s)_{s \geq 0}\) imply that \( \tilde{\pi} \) is invariant for the Markov semi-group associated with this flow.

However in most cases, it is not possible to compute this flow explicitly and therefore discretisation of (3) must be used instead. Symplectic integrators, like the leap-frog (or Stormer-Verlet) integrator, are discretizations of the Hamiltonian dynamics which preserve reversibility and symplectiness. Given a time step \( h \in \mathbb{R}_+^* \) and a number of iterations \( T \in \mathbb{N}^* \), the leap-frog integrator consists, starting from an initial point \((q_0, p_0) \in \mathbb{R}^d \times \mathbb{R}^d\), in the iterations called leapfrog steps, defined for \( k = 0, \ldots, T \),

\[
\begin{align*}
q_{k+1} &= q_k + h p_{k+1/2} \\
p_{k+1} &= p_{k+1/2} - (\Delta h / 2) \nabla U(q_{k+1}).
\end{align*}
\]

The sequence \((q_k, p_k)_{k \in \{0, \ldots, T\}}\) is an approximation of the solution of (3) at times \( k h : k \in \{0, \ldots, T\} \) started at \((q_0, p_0)\). This sequence defines a discrete dynamical system defined for \( k = 0, \ldots, T - 1 \) by

\[
(q_{k+1}, p_{k+1}) = \Psi_h^{(1)} \circ \Psi_h^{(2)} \circ \Psi_h^{(1)}(q_k, p_k) = \Phi_h^{(1)}(q_k, p_k),
\]

where \( \Psi_h^{(1)}, \Psi_h^{(2)} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{R}^d \) are given for all \((q, p) \in \mathbb{R}^d \times \mathbb{R}^d\) by

\[
\Psi_h^{(1)}(q, p) = (q, p - (h/2) \nabla U(q)), \quad \Psi_h^{(2)}(q, p) = (q + h p, p)
\]

Define the sequence of iterates \( \Phi_h^{(k)} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{R}^d : k \in \mathbb{N}^* \) for \( k \geq 1 \) by induction

\[
\Phi_h^{(k+1)} = \Phi_h^{(k)} \circ \Phi_h^{(1)},
\]
Set for all \( k \geq 1 \),
\[
\tilde{\Phi}_h^{(k)} = \text{proj} \circ \Phi_h^{(k)},
\]  
(8)
where \( \text{proj} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d \) is the projection on the first \( d \) coordinates, for all \((q, p) \in \mathbb{R}^d \times \mathbb{R}^d, \text{proj}(q, p) = q \). Thus, with our notation for all \( k \in \{1, \ldots, T\} \),
\[(q_k, p_k) = \Phi_h^{(k)}(q_0, p_0), \quad q_k = \tilde{\Phi}_h^{(k)}(q_0, p_0).
\]

Note that a simple induction implies that for all \((q_0, p_0) \in \mathbb{R}^d \times \mathbb{R}^d \) and \( k \in \{1, \ldots, T\} \), the \( k \)th iteration of the leap-frog integration takes the form
\[
q_k = q_0 + khp_0 - \frac{kh^2}{2} \nabla U(q_0) - h^2 \Xi_{h,k}(q_0, p_0)
\]
(9)
\[
p_k = p_0 - \frac{h}{2} \left\{ \nabla U(q_0) + \nabla U \circ \tilde{\Phi}_h^{(k)}(q_0, p_0) \right\}
- h \sum_{i=1}^{k-1} \nabla U \circ \tilde{\Phi}_h^{(i)}(q_0, p_0),
\]
(10)
where \( \Xi_{h,k} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d \) is given for all \((q, p) \in \mathbb{R}^d \times \mathbb{R}^d \) by
\[
\Xi_{h,k}(q, p) = \sum_{i=1}^{k-1} (k - i) \nabla U \circ \tilde{\Phi}_h^{(i)}(q, p).
\]
(11)
Because each inner step in the leap-frog step are shear transformations of the phase variable (only the position or the momentum are updated by a quantity that depends only on the variable that do not change), it is clear that this transformation is volume preserving (the Jacobian of each individual transformation is equal to 1). Each inner leap frog step due of its symmetry is also reversible: starting from \((q_{k+1}, -p_{k+1})\), applying the leap-frog step forward and then negating the momentum variable again, we obtain again \((q_k, p_k)\).

It is assumed in this paper that the number of steps \( T \) is set to some fixed value; we will briefly describe how our results can be extended to the case where the number of steps is random but independent of the current state.

We have now all the background required to describe the HMC algorithm. The HMC algorithm is a Metropolis-Hastings algorithm on the phase space aimed at sampling the extended target distribution with density \( \tilde{\pi} \) defined in (4).

Denote by \((Q_k, P_k)\) the value of the position and momentum at the \( k \)th iteration of the algorithm. Each iteration of the algorithm may be decomposed into two steps. The first step modifies only the momentum. The second step may change both the momentum and the position. Both steps are constructed to leave the extended distribution invariant. In the first step, we draw \( G_{k+1} \) from the \( d \)-dimensional normal distribution with zero mean and identity covariance matrix, independent of \( \{(Q_j, P_j)\}_{j=0}^k \). Since \( Q_k \) is not changed and \( G_{k+1} \) is drawn from the stationary distribution of the momentum, this step leaves the joint distribution \( \tilde{\pi} \) invariant.
Consider now the second step. We set the initial conditions \( p_0 = G_{k+1}, q_0 = Q_k \) and compute the position and the momentum after \( T \) leapfrog steps \((q_T, p_T) = \Phi^T_h(q_0, p_0)\). This is our proposal. To correct the target distribution, we simply introduce a Metropolis filter related to \( \tilde{\pi} \). Using the properties of the leap-frog integrator, the acceptance probability is defined for all \((q, p) \in \mathbb{R}^d \times \mathbb{R}^d, (\tilde{q}, \tilde{p}) \in \mathbb{R}^d \times \mathbb{R}^d, \) by

\[
\alpha_H \{(q, p), (\tilde{q}, \tilde{p})\} = \min \left[ 1, \exp \left( H(q, p) - H(\tilde{q}, \tilde{p}) \right) \right].
\] (12)

Finally, with probability \( \alpha_H \{(q_0, p_0), (q_T, p_T)\} \) the move is accepted, \( Q_{k+1} = q_T \) and \( P_{k+1} = -p_T \). Otherwise the move is rejected \((Q_{k+1}, P_{k+1}) = (Q_k, G_{k+1})\).

If the number of steps \( T = 1 \), then the algorithm reduces to the Metropolis Adjusted Langevin Algorithm (MALA).

The sequence \((Q_k)_{k \geq 0}\) is itself a Markov chain with Markov kernel given for all \( q \in \mathbb{R}^d \) and \( A \in \mathcal{B}(\mathbb{R}^d) \) by

\[
P_{h,T}(q, A) = \int_{\mathbb{R}^d} \mathbb{1}_A \left( \Phi^T_h(q, \tilde{p}) \right) \alpha_H \left\{ (q, \tilde{p}), \Phi^T_h(q, \tilde{p}) \right\} \frac{e^{-\|\tilde{p}\|^2/2}}{(2\pi)^{d/2}} \, d\tilde{p} \\
+ \delta_q(A) \int_{\mathbb{R}^d} \left[ 1 - \alpha_H \left\{ (q, \tilde{p}), \Phi^T_h(q, \tilde{p}) \right\} \right] \frac{e^{-\|\tilde{p}\|^2/2}}{(2\pi)^{d/2}} \, d\tilde{p}.
\] (13)

Since \( \tilde{\pi} \) (4) is invariant with respect to the Markov kernel defined by the HMC algorithm on the extended state space \( \mathbb{R}^d \times \mathbb{R}^d \), it naturally implies that \( \pi \) is an invariant probability distribution for \( P_{h,T} \) for all \( h \in \mathbb{R}_+^* \) and \( T \in \mathbb{N}^* \). However, it is important to note that the invariance of \( \pi \) for this kernel is not a sufficient condition for the convergence of algorithm.

### 3. Ergodicity of the HMC algorithm

In this section, we establish conditions upon which the sampler is irreducible and (Harris) recurrent. Not surprisingly these conditions imply regularity conditions and control of the tail. For \( \beta \in [0, 1] \), we consider the following assumption on the potential \( U \).

**H1 (\( \beta \)).** \( U \) is continuously differentiable and

i) there exists \( L_1 \geq 0 \) such that for all \( q, \tilde{q} \in \mathbb{R}^d \),

\[
\| \nabla U(q) - \nabla U(\tilde{q}) \| \leq L_1 \| q - \tilde{q} \|.
\]

ii) there exists \( M_1 \geq 0 \) such that for all \( q \in \mathbb{R}^d \),

\[
\| \nabla U(q) \| \leq M_1 \left\{ 1 + \| q \|^\beta \right\}.
\]

Under the regularity condition **H1**, it is possible to derive useful bounds on the position and the momentum in the intermediate steps of the leap-frog integration.
Lemma 1. Let $h_0 > 0$, $T \in \mathbb{N}^*$ and $\beta \in [0,1]$.

(a) If $H1(\beta)$-i holds, there exists $C \geq 0$ (which depends only on $T, h_0$ and $L_1$) such that for all $h \in (0,h_0)$, $(q_0, p_0) \in \mathbb{R}^d \times \mathbb{R}^d$, $(x_0, v_0) \in \mathbb{R}^d \times \mathbb{R}^d$ and $k \in \{1, \ldots, T\}$

$$\left\| \Phi_h^{s(k)}(q_0, p_0) - \Phi_h^{s(k)}(x_0, v_0) \right\| \leq C(\|q_0 - x_0\| + \|p_0 - v_0\|),$$

where $\Phi_h^{s(k)}$ is defined in (8).

(b) If $H1(\beta)$-ii holds, there exists $C \geq 0$ (which depends only on $T, h_0$ and $M_1$) such that for all $h \in (0,h_0)$, $(q_0, p_0) \in \mathbb{R}^d \times \mathbb{R}^d$ and $k \in \{1, \ldots, T\}$

$$\|q_k - q_0\| \leq C h \left( \|p_0\| + h(1 + \|q_0\|^{\beta}) \right),$$

$$\|p_k - p_0\| \leq C h \left( 1 + \|p_0\|^{\beta} + \|q_0\|^{\beta} \right),$$

where $(q_k, p_k) = \Phi_h^{s,k}(q_0, p_0)$, $\Phi_h^{s,k}$ is defined by (7).

Proof. The proof is postponed to Section 6.1.1. □

We first state our main two results regarding the ergodicity of $P_{h,T}$ for $h \in \mathbb{R}^+_+$ and $T \in \mathbb{N}$. The proof of these results consists in studying the irreducibility properties of the proposal kernel $Q_{h,T}$ associated with $P_{h,T}$ given for all $q \in \mathbb{R}^d$ and $A \in \mathcal{B}(\mathbb{R}^d)$ by

$$Q_{h,T}(q, A) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \mathbb{1}_A \left( \Phi_h^{s(T)}(q, \tilde{p}) \right) e^{-\|\tilde{p}\|^2/2} d\tilde{p},$$

where $\Phi_h^{s(T)}$ is defined in (8). Before going further, we need to briefly recall some definitions pertaining to Markov chains. Let $P$ be a Markov kernel on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Let $m$ be an integer and $\mu$ be a nontrivial measure on $\mathcal{B}(\mathbb{R}^d)$. A set $C \in \mathcal{B}(\mathbb{R}^d)$ is called a $(m, \mu)$-small set for $P$ if for all $x \in C$ and $A \in \mathcal{B}(\mathbb{R}^d)$, $P^m(x, A) \geq \mu(A)$. A set $A \in \mathcal{B}(\mathbb{R}^d)$ is said to be accessible for $P$ if for all $x \in \mathbb{R}^d$, $\sum_{m=1}^{\infty} P^m(x, A) > 0$. A non-trivial $\sigma$-finite measure $\mu$ is an irreducibility measure of $P$ if and only if any set $A \in \mathcal{B}(\mathbb{R}^d)$ satisfying $\mu(A) > 0$ is accessible. The Markov kernel $P$ is said to be irreducible if it admits an accessible small set or equivalently an irreducibility measure (in [10], our notion of irreducibility is referred to as $\phi$-irreducibility, where $\phi$ is an irreducibility measure; here irreducibility therefore means $\phi$-irreducibility).

Let $(X_n)_{n \geq 0}$ be the canonical chain associated with $P$ defined on the canonical space $(\Omega, \mathcal{F}, (\mathbb{P}_x, x \in \mathbb{R}^d))$. A set $A \in \mathcal{B}(\mathbb{R}^d)$ is said to be recurrent if for all $x \in A$, $\mathbb{E}_x[N_A] = \infty$ where $N_A$ is the number of visits to $A$. The set $A$ is Harris recurrent if for any $x \in A$, $\mathbb{P}_x(N_A = \infty) = 1$. A is said to be Harris recurrent if all accessible sets are Harris recurrent. In this case, for all $x \in \mathbb{R}^d$, and all accessible sets $A$, $\mathbb{P}_x(N_A = \infty) = 1$.

Theorem 2. Let $\beta \in [0,1]$. Assume $H1(\beta)$ and that $U$ is twice continuously differentiable. Then for all $T \geq 0$, there exists $\bar{h}_T > 0$ such that for all $h \in (0, \bar{h}_T)$,
the HMC kernel $P_{h,T}$, defined by (13), is irreducible. The Lebesgue measure is an irreducibility measure. Moreover, $P_{h,T}$ is aperiodic and Harris recurrent. Therefore, for all $q \in \mathbb{R}^d$

$$\lim_{n \to +\infty} \| \delta_q P_n^{h,T} - \pi \|_{TV} = 0.$$ 

Proof. The proof is postponed to Section 6.1.2.

In our next result, we relax the second order differentiability condition on $U$, and in the case $\beta < 1$ we even allow for arbitrary large values of the step size $h$. The proof is more involved and requires the use of degree theory for continuous mapping (the main notions of the degree theory required in the proof are briefly recalled in Section 6.3.2). In particular, we use results stated in Section 5 on the class of iterative Markov kernels, to which $Q_{h,T}$ given by (16) belongs.

**Theorem 3.** Assume $H1$ ($\beta$) for some $\beta \in [0, 1].$

(a) Assume $\beta \in [0, 1)$. Then, for all $h > 0$ and $T \in \mathbb{N}^*$, the HMC kernel $P_{h,T}$ defined by (13) is irreducible, the Lebesgue measure is an irreducibility measure and any compact set of $\mathbb{R}^d$ is small.

(b) Assume $\beta = 1$. Then, for all $T \in \mathbb{N}^*$, there exists $\tilde{h}_T > 0$ such that for all $h \in (0, \tilde{h}_T]$, the HMC kernel $P_{h,T}$ defined by (13) is irreducible, the Lebesgue measure is an irreducibility measure and any compact sets of $\mathbb{R}^d$ is a small set.

Moreover, $P_{h,T}$ is aperiodic. Therefore, if either (a) or (b) hold then $P_{h,T}$ is recurrent and for $\pi$-almost every $q \in \mathbb{R}^d$,

$$\lim_{n \to +\infty} \| \delta_q P_n^{h,T} - \pi \|_{TV} = 0.$$ 

Proof. The proof is postponed to Section 6.1.3.

To the best of the author’s knowledge, the first results regarding the irreducibility of the HMC algorithm are established in [3] under the assumption that $U$ and $\| \nabla U \|$ is bounded above. Note that these assumptions are in general satisfied only for compact state space. Irreducibility has also been tackled in [9]: in this work however, the number of leapfrog steps $T$ is assumed to be random and independent of the current position and momentum. Under this setting and additional conditions which in particular imply that the number of leapfrog steps $T$ is equal to 1 with positive probability, [9] shows that the kernel associated with the HMC algorithm is irreducible. Note as a consequence of the assumptions made, this result is a direct consequence of the irreducibility of the MALA algorithm (a mixture of Markov kernels is irreducible as soon as one component of the mixture is irreducible; the irreducibility of MALA kernel has been established in [14]). In the same setting, [9] also establish geometric ergodicity of the HMC kernel under implicit conditions on the behaviour of the acceptance rate. In contrast, we are able to also deal with the more difficult case of number of leapfrog steps strictly greater than one. Moreover, our conditions for geometric ergodicity are directly verifiable on the potential $U$. 

Note that our results can be easily extended to the case where the number of steps is random. We briefly describe the main arguments to obtain such extension. Let \((a_i)_{i \in \mathbb{N}^*}\) be a sequence of non-negative numbers satisfying \(\sum_{i \in \mathbb{N}^*} a_i = 1\), and \((h_i)_{i \in \mathbb{N}^*}\) be a sequence of positive real numbers. Define the randomized Hamiltonian kernel \(\overline{P}_{h,a}\) on \((\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))\) associated with \((a_i)_{i \in \mathbb{N}^*}\) and \((h_i)_{i \in \mathbb{N}^*}\) by

\[
\overline{P}_{h,a} = \sum_{i \in \mathbb{N}^*} a_i P_{h_i,i} .
\]

**Corollary 4.** Let \(\beta \in [0, 1]\) and assume \(H_1(\beta)\). Let \((a_i)_{i \in \mathbb{N}^*}\) be a non-negative sequence satisfying \(\sum_{i \in \mathbb{N}^*} a_i = 1\) and \((h_i)_{i \in \mathbb{N}^*}\) be a sequence of positive real numbers. Let \(\overline{P}_{h,a}\) be the randomized Hamiltonian kernel associated with \((a_i)_{i \in \mathbb{N}^*}\) and \((h_i)_{i \in \mathbb{N}^*}\).

(a) Under the assumption that \(U\) is twice continuously and there exists \(i \in \mathbb{N}^*, h_i < \tilde{h}_i\) and \(a_i > 0\), where \(\tilde{h}_i\) is defined in Theorem 2, then the conclusions of Theorem 2 hold for \(\overline{P}_{h,a}\).

(b) If \(\beta \in (0, 1)\), then the conclusions of Theorem 3-(a) hold for \(\overline{P}_{h,a}\).

(c) If \(\beta = 1\) and there exists \(i \in \mathbb{N}^*, h_i < \bar{h}_i\) and \(a_i > 0\), where \(\bar{h}_i\) is defined in Theorem 3-(b), then the conclusions of Theorem 3-(b) hold for \(\overline{P}_{h,a}\).

**Proof.** (a) follows from Theorem 2 and Proposition 19. (b) and (c) are straightforward applications of Theorem 3. \(\square\)

### 4. Geometric ergodicity of HMC

In this section, we give conditions on the potential \(U\) which imply that the HMC kernel (8) converges geometrically fast to its invariant distribution. Let \(V : \mathbb{R}^d \to [1, +\infty)\) be a measurable function and \(P\) be a Markov kernel on \((\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))\) with invariant probability measure \(\mu\) on \(\mathcal{B}(\mathbb{R}^d)\). Recall that \(P\) is said to be \(V\)-uniformly geometrically ergodic if \(P\) admits an invariant probability \(\pi\) and there exists \(\rho \in [0, 1)\) and \(\varsigma \geq 0\) such that for all \(q \in \mathbb{R}^d\) and \(k \in \mathbb{N}^*\),

\[
\|P^k(q, \cdot) - \pi\|_V \leq \varsigma \rho^k V(q) .
\]

By [10, Theorem 16.0.1], if \(P\) is aperiodic, irreducible and satisfies a Foster-Lyapunov drift condition, i.e. there exists a small set \(C\) for \(P\), \(\lambda \in [0, 1)\) and \(b < +\infty\) such that for all \(q \in \mathbb{R}^d\),

\[
P V \leq \lambda V + b \mathbb{1}_C ,
\]

then \(P\) is \(V\)-uniformly geometrically ergodic. If a function \(V : \mathbb{R}^d \to [1, \infty)\) satisfies (18); then \(V\) is said to be a Foster-Lyapunov function for \(P\). We first give an elementary condition to establish the \(V\)-uniform geometric ergodicity for a class of generalized Metropolis-Hastings kernel which includes HMC kernels as a special example.
Let $K$ be a proposal kernel on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and $\alpha : \mathbb{R}^d \to [0, 1]$ be an acceptance probability, assumed to be Borel measurable. Consider the Markov kernel $P$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ defined for all $q \in \mathbb{R}^d$ and $A \in \mathcal{B}(\mathbb{R}^d)$ by

$$
P(q, A) = \int_{\mathbb{R}^d} 1_A(\text{proj}(z))\alpha(q, z)K(q, dz) + \delta_q(A)\int_{\mathbb{R}^d} \{1 - \alpha(q, z)\}K(q, dz), \quad (19)$$

where $\text{proj} : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ is the canonical projection onto the first $d$ components. For $h \in \mathbb{R}^*_+$ and $T \in \mathbb{N}^*$, $P_{h,T}$ corresponds to $P$ with $K$ and $\alpha$ given for all $q, p, x \in \mathbb{R}^d$ and $A \in \mathcal{B}(\mathbb{R}^{2d})$ respectively by

$$
K_{h,T}(q, B) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} 1_B(\Phi^{(T)}_h(q, \tilde{p})) e^{-\|\tilde{p}\|^2/2} d\tilde{p} \quad (20)
$$

$$
\tilde{\alpha}_H(q, (x, p)) = \begin{cases} 
\alpha_H \begin{cases} 
\{q, p\}, \Phi^{(T)}_h(q, p) 
\end{cases} & x = \Phi^{(T)}_h(q, p), \\
0 & \text{otherwise}
\end{cases} \quad (21)
$$

where $\Phi^{(T)}_h, \alpha_H : \mathbb{R}^d \to [0, 1]$ and $\Phi^{(T)}_h$ are defined in (8), (12) and (7), respectively. Let $V : \mathbb{R}^d \to [1, +\infty)$ be a measurable function. $V$ naturally extends on $\mathbb{R}^{2d}$ by setting for all $(q, p) \in \mathbb{R}^{2d}$, $V(q, p) = V(q)$. For all $q \in \mathbb{R}^d$, define the two subsets of $\mathbb{R}^{2d}$,

$$
\mathcal{R}(q) = \{z \in \mathbb{R}^{2d}, \alpha(q, z) < 1\}
$$

$$
\mathcal{B}(q) = \{z \in \mathbb{R}^{2d}, V(\text{proj}(z)) \leq V(q)\}. \quad (22)
$$

The set $\mathcal{R}(q)$ is the potential rejection region. Our next result gives a condition on $K$ and $\alpha$ which implies that if $V$ is a Foster-Lyapunov function for $K$ then $P$ satisfies a Foster-Lyapunov drift condition as well. This result is inspired by [14, Theorem 4.1], which is used to show the $V$-uniform geometric ergodicity of the MALA algorithm.

**Proposition 5.** Assume that there exist $\lambda \in [0, 1)$ and $b \in \mathbb{R}_+$ such that

$$
KV \leq \lambda V + b. \quad (23)
$$

and

$$
\lim_{M \to +\infty} \sup_{q \in \mathbb{R}^d : V(q) \geq M} K(q, \mathcal{R}(q) \cap \mathcal{B}(q)) = 0. \quad (24)
$$

Then there exist $\tilde{\lambda} \in [0, 1)$ and $\tilde{b} \in \mathbb{R}_+$ such that $PV \leq \tilde{\lambda} V + \tilde{b}$ where $P$ is given by (19).

**Proof.** By construction, for all $q \in \mathbb{R}^d$, we have (19)

$$
PV(q) - V(q) = \int_{\mathbb{R}^{2d}} \{V(\text{proj}(z)) - V(q)\} \alpha(q, z)K(q, dz) = KV(q) - V(q) + \int_{\mathbb{R}^{2d}} \{V(\text{proj}(z)) - V(q)\} \{\alpha(q, z) - 1\}K(q, dz).
$$
By using (23), this implies for all $q \in \mathbb{R}^d$,

$$PV(q) - V(q) \leq (\lambda - 1)V(q) + b + \int_{\mathbb{R}^{2d}} \{V(proj(z)) - V(q)\} \{\alpha(q, z) - 1\} K(q, dz).$$

(25)

Note that by definition (22) of $\mathcal{R}(q)$ and $\mathcal{B}(q)$

$$\int_{\mathbb{R}^{2d}} \{V(proj(z)) - V(q)\} \{\alpha(q, z) - 1\} K(q, dz)$$

$$\leq \int_{\mathcal{R}(q) \cap \mathcal{B}(q)} \{V(q) - V(proj(z))\} K(q, dz).$$

Therefore by (24), we get

$$\lim_{M \to +\infty} \sup_{\{q \in \mathbb{R}^d : V(q) \geq M\}} \int_{\mathbb{R}^{2d}} \{1 - V(proj(z))/V(q)\} \{\alpha(q, z) - 1\} K(q, dz) = 0.$$ 

The proof then follows from combining this result and (25) since they imply

$$\lim_{M \to +\infty} \sup_{\{q \in \mathbb{R}^d : V(q) \geq M\}} PV(q)/V(q) \leq \lambda.$$ 

We now show that under appropriate conditions, the proposal kernel $K_{h,T}$ and the acceptance $\tilde{\alpha}_H$ given in (20) and (21) satisfy the conditions of Proposition 5 which will imply that the HMC kernel $P_{h,T}$ is $V$-uniformly geometrically ergodic. For $m \in (1, 2]$, consider the following assumption:

$\mathbf{H}_2(m).$ (i) $U \in C^3(\mathbb{R}^d)$ and there exists $A_1 \geq 0$ such that for all $q \in \mathbb{R}^d$ and $k = 2, 3$:

$$\|D^k U(q)\| \leq A_1 \{\|q\| + 1\}^{m-k}.$$ 

(ii) There exist $A_2 \in \mathbb{R}_+^*$ and $R \in \mathbb{R}_+$ such that for all $q \in \mathbb{R}^d$, $\|q\| \geq R$

$$D^2U(q) \{\nabla U(q) \otimes \nabla U(q)\} \geq A_2 \|q\|^{3m-4}.$$ 

(iii) There exist $A_3 \in \mathbb{R}_+^*$ and $A_4 \in \mathbb{R}$ such that for all $q \in \mathbb{R}^d$,

$$\langle \nabla U(q), q \rangle \geq A_3 \|q\|^m - A_4.$$ 

It is easily checked that under $\mathbf{H}_2$, the results of Section 3 can be applied, i.e. $\nabla U$ satisfies $\mathbf{H}_1(m - 1)$; see Lemma 12.

Condition $\mathbf{H}_2(m)$ is satisfied by power functions $q \mapsto c \|q\|^m$. More generally, they are satisfied by $m$-homogeneously quasiconvex functions with convex level sets outside a ball and by lower order perturbations of such functions.

We say that a function $U_0$ is $m$-homogeneous outside a ball of radius $R_1$ if the following conditions are satisfied:
(i) for all \( t \geq 1 \) and \( q \in \mathbb{R}^d, \| q \| \geq R_1, \) \( U(tq) = t^mU_0(q) \).
(ii) for all \( q \in \mathbb{R}^d, \| q \| \geq R_1 \), the level sets \( \{ x : U_0(x) \leq U_0(q) \} \) are convex.

**Proposition 6.** Let \( m \in [1, 2] \) and \( R_1 \in \mathbb{R}_+ \). Assume that the potential \( U \) may be decomposed as

\[
U(q) = U_0(q) + G(q), \quad q \in \mathbb{R}^d, \| q \| \geq R_1,
\]

where the functions \( U_0, G \in C^3(\mathbb{R}^d) \) satisfy the following two conditions:

(i) \( U_0 \) is \( m \)-homogeneously quasiconvex outside a ball of radius \( R_1 \) and \( \lim_{\| q \| \to \infty} U_0(q) = \infty \).

(ii) For \( k = 2, 3 \), \( \lim_{\| q \| \to \infty} \| D^k G(q) \| / \| q \|^{m-k} = 0 \).

Then \( U \) satisfies \( H_2(m) \).

**Proof.** The proof is postponed to Section 6.2.1.

The following proposition establishes that the kernel \( K_{h,T} \) satisfies the Foster-Lyapunov condition (23) under \( H_2 \). For all \( q \in \mathbb{R}_+^d \) define \( V_a : \mathbb{R}^d \to [1, +\infty) \) for all \( q \in \mathbb{R}^d \) by

\[
V_a(q) = \exp \{ a \| q \| \}.
\]

**Proposition 7.** Assume \( H_2(m)-(i)-(iii) \) for some \( m \in (1, 2] \). Let \( T \in \mathbb{N}^+ \). Then the following holds

(a) Assume that \( m \in (1, 2) \). Then for all \( h \in \mathbb{R}_+^* \), there exist \( a \in \mathbb{R}_+^*, \lambda \in [0, 1) \) and \( b \in \mathbb{R}_+ \) such that

\[
K_{h,T}V_a \leq \lambda V_a + b,
\]

where \( K_{h,T} \) is defined by (20).

(b) Assume that \( m = 2 \). Then there exists \( h_0 \in \mathbb{R}_+^* \) such that for all \( h \in (0, h_0] \), there exist \( a \in \mathbb{R}_+^*, \lambda \in [0, 1) \) and \( b \in \mathbb{R}_+ \) which satisfy (27).

**Proof.** The proof is postponed to Section 6.2.2.

To show that the condition (24) of Proposition 5 is satisfied under \( H_2(m) \), we rely on the following important result which implies that if one starts very far away from the origin, the next move will be accepted with an overwhelming probability.

**Proposition 8.** Assume \( H_2(m)-(i)-(ii) \) for some \( m \in (1, 2] \). Let \( T \in \mathbb{N}^+ \), \( h \in \mathbb{R}_+^* \) and \( \gamma \in [0, m - 1) \).

(a) If \( m \in (1, 2) \). For all \( h \in \mathbb{R}_+^* \), there exists \( R_H \in \mathbb{R}_+ \) such that for all \( q_0, p_0 \in \mathbb{R}^d \), \( \| q_0 \| \geq R_H \) and \( \| p_0 \| \leq \| q_0 \|^\gamma \), \( H(\Phi_h^{\circ T}(q_0, p_0)) - H(q_0, p_0) \leq 0 \).

(b) If \( m = 2 \). There exist \( R_H \in \mathbb{R}_+ \) and \( h_H \in \mathbb{R}_+^* \) such that for all \( h \in (0, h_H] \), \( q_0, p_0 \in \mathbb{R}^d \), \( \| q_0 \| \geq R_H \) and \( \| p_0 \| \leq \| q_0 \|^\gamma \), \( H(\Phi_h^{\circ T}(q_0, p_0)) - H(q_0, p_0) \leq 0 \).

**Proof.** The proof is postponed to Section 6.2.3.
This result means that far in the tail the HMC proposal are "inward". We illustrate the result of Proposition 8-(a) in Figure 1 for $U$ given by $q \mapsto (\|q\|^2 + \delta)^\kappa$ for $\kappa = 3/4$, $h = 0.9$ and $p_0 \in \mathbb{R}^d$, $\|p_0\| = 1$. Note that this potential satisfies the condition of the proposition. We can observe that choosing the different initial conditions $q_0$ with increasing norm imply that $\tilde{T} = \max\{k \in \mathbb{N}; H(\Phi_h^{c(k)}(q_0, p_0)) - H(q_0, p_0) < 0\}$ increases as well.

We now can establish the geometric ergodicity of the HMC sampler.

**Theorem 9.** Assume $H_2(m)$ for some $m \in (1, 2]$.

(a) If $m \in (1, 2)$, for all $T \in \mathbb{N}^*$ and $h > 0$, the HMC kernel $P_{h,T}$ is geometrically ergodic.

(b) If $m = 2$, then for all $T \in \mathbb{N}^*$ there exists $h_H > 0$ such that for all $h \in (0, h_H)$, $P_{h,T}$ is geometrically ergodic.

**Proof of Theorem 9.** It is enough to consider the case $m \in (1, 2)$ as the proof of the case $m = 2$ (b) follows exactly the same lines taking $h$ small enough.

We check that the conditions of Proposition 5 are satisfied which will conclude the proof. By Proposition 7, (23) holds. Next, note that (24) is a straightforward consequence of

$$
\lim_{M \to +\infty} \sup_{\|q\| \geq M} \int_{\mathcal{R}(q)} K_{h,T}(q, dz) = 0,
$$

where $K_{h,T}$ is defined by (20), $\mathcal{R}(q) = \{z \in \mathbb{R}^d, \alpha_H(q, z) < 1\}$ and $\alpha_H$ is defined by (21). By Proposition 8, there exists $M \geq 0$ such that for all $q \in \mathbb{R}^d$, $\|q\| \geq M$,

$$
\int_{\mathcal{R}(q)} K_{h,T}(q, dz) \leq (2\pi)^{-d/2} \int_{\{\|p\| \geq \|q\|\}} e^{-\|p\|^2/2} dp.
$$
and therefore (28) holds. \hfill \Box

5. Irreducibility for a class of iterative models

In this Section we establish the irreducibility of a Markov kernel associated to a random iterative model. We believe that these results might be of independent interest. Let \( f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d \) be a Borel measurable function and \( \phi : \mathbb{R}^d \to [0, +\infty] \) be a probability density with respect to the Lebesgue measure. Consider Markov kernels \( K \) defined for all \( x \in \mathbb{R}^d \) and \( A \in B(\mathbb{R}^d) \) by

\[
K(x, A) = \int_{\mathbb{R}^d} 1_A(f(x, z))\phi(z)dz.
\]

Note that given an i.i.d. sequence \((Z_k)_{k \geq 1}\) with common density function \( \phi \) and an initial state \( X_0 \) independent of \((Z_k)_{k \geq 1}\), the Markov chain \((X_k)_{k \in \mathbb{N}}\) given for all \( k \geq 0 \) by

\[
X_{k+1} = f(X_k, Z_{k+1}),
\]

has \( K \) as Markov kernel. Define for all \( x \in \mathbb{R}^d \), \( f_x : \mathbb{R}^d \to \mathbb{R}^d \) by \( f_x = f(x, \cdot) \).

Consider the following assumptions.

**G 1** \((R, y_0, M)\). Let \( R, M \in \mathbb{R}^*_+ \) and \( y_0 \in \mathbb{R}^d \).

(i) There exists \( L_f \in \mathbb{R}_+ \) such that for all \( x \in B(0, R) \), \( f_x \) is \( L_f \)-Lipschitz, i.e. for all \( z_1, z_2 \in \mathbb{R}^d \), \( |f_x(z_1) - f_x(z_2)| \leq L_f |z_1 - z_2| \).

(ii) There exists \( M \in \mathbb{R}^*_+ \), such that for all \( x \in B(0, R) \), \( B(y_0, M) \subset f_x(B(0, M)) \).

**G 2.** \( \phi \) is lower semicontinuous and positive on \( \mathbb{R}^d \).

**Theorem 10.** Assume **G 1**\((R, y_0, M)\) and **G 2**. Then \( B(0, R) \) is 1-small for \( K \) for all \( x \in B(0, R) \) and \( A \in B(\mathbb{R}^d) \),

\[
K(x, A) \geq L_f^{-d} \inf_{z \in B(0, M)} \{ \phi(z) \} \text{Leb} \{ A \cap B(y_0, M) \}.
\]

**Proof.** The proof is postponed to Section 6.3.1. \hfill \Box

**Proposition 11.** Let \( g : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d \) and \( R \in \mathbb{R}^*_+ \) satisfy

(i) there exists \( L_{g,R} \in \mathbb{R}_+ \) such that for all \( z_1, z_2, x \in \mathbb{R}^d \), \( |x| \leq R \),

\[
|g(x, z_1) - g(x, z_2)| \leq L_{g,R} |z_1 - z_2|.
\]

(ii) there exist \( C_{R,0}, C_{R,1} \in \mathbb{R}_+ \) such that for all \( x, z \in \mathbb{R}^d \), \( |x| \leq R \)

\[
|g(x, z)| \leq C_{R,0} + C_{R,1} |z|.
\]

Let \( b \in \mathbb{R} \) and define \( f^g : \mathbb{R}^d \times \mathbb{R}^d \) for all \( x, z \in \mathbb{R}^d \) by

\[
f^g(x, z) = bz + g(x, z).
\]

If \( |b| > C_{R,1} \), then \( f^g \) satisfies **G 1**\((R, 0, M)\) for all \( M \in \mathbb{R}^*_+ \) with

\[
\tilde{M} = \{ M + C_{R,0} \}/(|b| - C_{R,1}).
\]

**Proof.** The proof is postponed to Section 6.3.2. \hfill \Box
6. Proofs

In the sequel, $C \geq 0$ is a constant which can change from line to line but does not depend on $h$.

6.1. Proof of Section 3

6.1.1. Proof of Lemma 1

(a) We show by induction that for all $k \in \{1, \cdots, T\}$, there exists $C_k$ (which depends only on $T, h_0$ and $L_1$) such that for all $h \in (0, h_0]$, $(q_0, p_0) \in \mathbb{R}^d \times \mathbb{R}^d$ and $(\tilde{q}_0, \tilde{p}_0) \in \mathbb{R}^d \times \mathbb{R}^d$,

$$|q_k - \tilde{q}_k| \leq C_k(|q_0 - \tilde{q}_0| + |p_0 - \tilde{p}_0|),$$

where $(q_k, p_k) = \Phi^k_{h}(q_0, p_0)$, $(\tilde{q}_k, \tilde{p}_k) = \Phi^k_{h}(\tilde{q}_0, \tilde{p}_0)$. Let $h \in (0, h_0]$, $(q_0, p_0) \in \mathbb{R}^d \times \mathbb{R}^d$ and $(\tilde{q}_0, \tilde{p}_0) \in \mathbb{R}^d \times \mathbb{R}^d$. The case $k = 1$ is immediate by H1-i). Let $k \in \{1, \cdots, T - 1\}$ and assume that the inequality holds for all $i \in \{1, \cdots, k\}$. Then by (9) and H1-i) we get

$$|q_{k+1} - \tilde{q}_{k+1}| \leq |q_0 - \tilde{q}_0| + (k + 1)h |p_0 - \tilde{p}_0|$$

$$+ (1/2)(k + 1)h^2 |\nabla U(q_0) - \nabla U(\tilde{q}_0)| + h^2 \sum_{i=1}^{k} (k + 1 - i) |\nabla U(q_i) - \nabla U(\tilde{q}_i)|$$

$$\leq |q_0 - \tilde{q}_0| + (k + 1)h |p_0 - \tilde{p}_0| + 2^{-1}(k + 1)h^2 L_1 |q_0 - \tilde{q}_0|$$

$$+ h^2 L_1 \sum_{i=1}^{k} (k + 1 - i) |q_i - \tilde{q}_i|.$$ 

An application of the induction hypothesis concludes the proof.

(b) We prove by induction that for all $k \in \{1, \cdots, T\}$ there exists $C_k \geq 0$ (which depends only on $T, h_0$ and $M_1$) such that for all $h \in (0, h_0]$, $(q_0, p_0) \in \mathbb{R}^d \times \mathbb{R}^d$

$$|q_k - q_0| \leq C_k h \left\{ |p_0| + h(1 + |q_0|^\beta) \right\},$$

where $(q_k, p_k) = \Phi^k_{h}(q_0, p_0)$. Let $h \in (0, h_0]$ and $(q_0, p_0) \in \mathbb{R}^d \times \mathbb{R}^d$. The case $k = 1$ is immediate by H1-ii) and (9). Let $k \in \{1, \cdots, T - 1\}$ and assume that the inequalities hold for all $i \in \{1, \cdots, k\}$. Then by (9) and H1-ii), we get

$$|q_{k+1} - q_0| \leq (k + 1)h |p_0| + 2^{-1}(k + 1)h^2 M_1 \left\{ 1 + |q_0|^\beta \right\}$$

$$+ h^2 M_1 \sum_{i=1}^{k} (k + 1 - i) \left\{ 1 + |q_i|^\beta \right\}.$$ 

(31)
By the induction hypothesis and using that \( t \mapsto t^\beta \) is sub-additive, we get for all \( i \in \{1, \cdots, k\} \),
\[
|q_i^1| \leq |q_0| + |q_i - q_0| \leq |q_0|^\beta + \left[C_i h \left( \|p_0\| + h(1 + |q_0|^\beta) \right) \right]^\beta,
\]
for some constants \( C_i \) which only depend on \( T, h_0 \) and \( M_1 \). Plugging this inequality in (31) and using that for all \( x \in \mathbb{R}^d \), \( h \|x\|^\beta \leq h(1 + |x|) \) conclude the proof of (b). As regards to (15), since by definition \( p_{k+1} = p_k - (h/2)\{\nabla U(q_k) + \nabla U(q_{k+1})\} \), using the triangle inequality, H1-ii), (b) to bound \( |q_k| \) and \( |q_{k+1}| \), and the induction hypothesis, we get that there exist some constants \( C_{k+1,1}, C_{k+1,2} \) which only depend on \( T, h_0 \) and \( M_1 \) such that
\[
|p_{k+1} - p_k| \leq |p_k - p_0| + (h/2)\{|\nabla U(q_k)| + |\nabla U(q_{k+1})|\}
\leq C_{k+1,1} h \left\{ 1 + |p_0|^\beta + |q_0|^\beta \right\} + (M_1 h/2) \left\{ 2 + |q_k|^\beta + |q_{k+1}|^\beta \right\}
\leq C_{k+1,1} h \left\{ 1 + |p_0|^\beta + |q_0|^\beta \right\} + (C_{k+2} h/2) \left\{ 1 + |q_0|^\beta + |p_0|^\beta \right\}.
\]
Therefore, the bound on \( |p_{k+1} - p_k| \) is proved, and this concludes the induction and the proof.

6.1.2. Proof of Theorem 2

To deduce our result, we first show that for all \( T \geq 0 \) there exists \( \tilde{h} > 0 \) such that for all \( h \in (0, \tilde{h}] \), the proposal associated with the HMC algorithm has a positive density with respect to the Lebesgue measure. To show this result by definition, we prove that for any \( T \geq 0 \), there exists \( \tilde{h} > 0 \) such that for all \( h \in (0, \tilde{h}] \), for all \( q \in \mathbb{R}^d \), the function \( p \mapsto \Phi^{(k)}_h(q, p) \) and \( p \mapsto \Xi_{h,k}(q, p) \) are continuously differentiable and for all \( (q, p) \in \mathbb{R}^d \times \mathbb{R}^d \),
\[
\nabla_p \Xi_{h,k}(q, p) = \sum_{i=1}^{k-1} (k-i) \left\{ \nabla^2 U \circ \Phi^{(k)}_i(q, p) \right\} \nabla_p \Phi^{(k)}_i(q, p),
\]
where for all \( q \in \mathbb{R}^d \), \( \nabla_p \Xi_{h,k}(q, p) \) (\( \nabla_q \Phi^{(k)}_i(q, p) \) respectively) is the gradient of the function \( \hat{p} \mapsto \Xi_{h,k}(q, \hat{p}) \) (\( \hat{q} \mapsto \Phi^{(k)}_i(q, \hat{p}) \) respectively) at \( p \in \mathbb{R}^d \). Under H1, sup \( x \in \mathbb{R}^d \nabla^2 U(x) \leq L_1 \) and by Lemma 1-(a), there exists \( C \geq 0 \), such that for all \( q \in \mathbb{R}^d \), sup \( p \in \mathbb{R}^d \left| \nabla_p \Phi^{(k)}_i(q, p) \right| \leq C \). Therefore
\[
\sup_{(q, p) \in \mathbb{R}^d \times \mathbb{R}^d} \nabla_p \Xi_{h,k}(q, p) \leq C.
\]
(32)

For any \( q \in \mathbb{R}^d \), define \( \phi_q : \mathbb{R}^d \to \mathbb{R}^d \) for all \( p \in \mathbb{R}^d \) by
\[
\phi_q(p) = p - (h/T)\Xi_{h,k}(q, p).
\]
By (32), it is a well known fact (see for example [6, Exercise 3.26]) that for all \( h < h_0 \), where
\[
\phi_q \text{ is a diffeomorphism. Therefore by (9), the same conclusion holds for } p \mapsto \overline{\Phi}_T^{(h)}(q, p) \text{ for all } h < h_0 \text{ and } q \in \mathbb{R}^d.
\]
Now let \( T \in \mathbb{N}^*, h \in (0, h_0) \) and \( q \in \mathbb{R}^d \). Denote by \( \overline{\Psi}_q : \mathbb{R}^d \rightarrow \mathbb{R} \) the continuously differentiable inverse of \( p \mapsto \overline{\Phi}_T^{(h)}(q, p) \). By (13), we have for all \( B \in \mathcal{B}(\mathbb{R}^d) \) using the change of variable associated with \( \overline{\Psi}_q \),
\[
P_{h,T}(q, B) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} 1_B(\tilde{p}) \, \alpha_H \left\{ (q, \overline{\Psi}_q(\tilde{p})), \Phi_h^{(T)}(q, \overline{\Psi}_q(\tilde{p})) \right\} e^{-|\overline{\Psi}_q(\tilde{p})|^2/2} d\overline{\Psi}_q(\tilde{p}) d\tilde{p},
\]
where \( D_{\overline{\Psi}_q}(\tilde{p}) = \left| \det(J_{\overline{\Psi}_q}(\tilde{p})) \right| \). Therefore for all \( A \in \mathcal{B}(\mathbb{R}^d) \),
\[
P_{h,T}(q, A) \geq (2\pi)^{-d/2} \text{Leb}(A)
\times \inf_{\tilde{p} \in A} \left\{ \alpha_H \left\{ (q, \overline{\Psi}_q(\tilde{p})), \Phi_h^{(T)}(q, \overline{\Psi}_q(\tilde{p})) \right\} e^{-|\overline{\Psi}_q(\tilde{p})|^2/2} d\overline{\Psi}_q(\tilde{p}) \right\},
\]
with the convention \( 0 \times +\infty = 0 \). Since \( \overline{\Psi}_q \) is a diffeomorphism on \( \mathbb{R}^d \), this result implies that \( P_{h,T} \) is irreducible with respect to the Lebesgue measure and aperiodic. Let \( B \subset \mathbb{R}^d \) be compact. Using (32) there exists \( C \geq 0 \) such that for all \( q, p_1, p_2 \in \mathbb{R}^d \),
\[
|p_1 - p_2| \leq C |\overline{\Phi}_h^{(T)}(q, p_1) - \overline{\Phi}_h^{(T)}(q, p_2)|.
\]
This follows that for all \( p \in \mathbb{R}^d \),
\[
\sup_{q \in B} |\overline{\Psi}_q(p)| \leq C \{ |p| + \sup_{q \in B} |\overline{\Phi}_h^{(T)}(q, 0)| \}.
\]
Using this upper bound and \( J_{\overline{\Psi}_q}(\overline{\Psi}_q(p)) = I_n \) in (35), where \( I_n \) is the identity matrix, we deduce that there exists \( \varepsilon > 0 \) such that for all \( A \in \mathcal{B}(\mathbb{R}^d) \), \( A \subset B \),
\[
\inf_{q \in B} P_{h,T}(q, A) \geq \varepsilon \text{Leb}(A),
\]
and therefore \( B \) is small for \( P_{h,T} \).

A straightforward adaptation of the proof of [18, Corollary 2] shows that \( P_{h,T} \) is Harris recurrent, see Proposition 19 in Appendix A. The desired conclusion then follows from [10, Theorem 13.0.1].

6.1.3. Proof of Theorem 3

Let \( h_0 \in \mathbb{R}_+^* \), \( h \in (0, h_0] \) and \( T \in \mathbb{N}^* \). First note by (13) that for all \( q \in \mathbb{R}^d \), \( K \in \mathbb{R}_+^* \) and \( A \in \mathcal{B}(\mathbb{R}^d) \),
\[
P_{h,T}(q, A) \geq \inf_{\tilde{p} \in A \cap \overline{\Psi}_q(B(0,K))} \left\{ \alpha_H \left\{ (q, \overline{\Psi}_q(\tilde{p})), \Phi_h^{(T)}(q, \overline{\Psi}_q(\tilde{p})) \right\} \right\} Q_{h,T}(q, A \cap B(0,K)),
\]
where $Q_{h,T}$ is defined by (16). Therefore to conclude the proof we only need to show that $Q_{h,T}$ is irreducible with respect to the Lebesgue measure, which can be done using Theorem 10. Indeed $Q_{h,T}$ is of form (29) and it is straightforward to see that it satisfies $G \geq 2$.

We now show that it satisfies $G \geq 1(R,0,M)$ for all $R,M \in \mathbb{R}_+^*$ using Proposition 11. Indeed by (9), for all $q,p \in \mathbb{R}^d$, $\overline{\Phi}_h^{(T)}(q,p) = T h p + g_{T,h}(q,p)$ where $g_{T,h}(q,p) = q - (T h^2/2) \nabla U(q) - h^2 \sum_{i=1}^{T-1} (T-i) \nabla U \circ \sigma_i^h(q,p)$. Therefore, by $H1$-i) and Lemma 1-(a), there exists $\tilde{L} < \infty$ such that

$$\sup_{p_1,p_2,q \in \mathbb{R}^d} \frac{|g_{T,h}(q,p_1) - g_{T,h}(q,p_2)|}{|p_1 - p_2|} \leq \tilde{L} .$$

(37)

In addition, by $H1$-ii) and Lemma 1-(b), there exists $C \geq 0$ such that for all $R \in \mathbb{R}_+^*$ and $q,p \in \mathbb{R}^d$, $|q| \leq R$, we have

$$|g_{T,h}(q,p)| \leq C \left\{ 1 + R + h^2 |p|^2 \right\} .$$

(38)

Therefore, by combining (37) and (38), if $\beta < 1$ or in the case $\beta = 1$, if $h < TC^{-1}$, $g_{T,h}$ satisfies the conditions of Proposition 11 for all $R \in \mathbb{R}_+^*$ and $\overline{\Phi}_h^{(T)}$ satisfies $G \geq 1(R,0,M)$ for all $R,M \in \mathbb{R}_+^*$.

Using this result and Theorem 10, we get that for all $R,M \in \mathbb{R}_+^*$ there exists $\varepsilon > 0$ such that for all $q \in B(0,R)$ and $A \in \mathcal{B}(\mathbb{R}^d)$,

$$Q_{h,T}(q,A) \geq \varepsilon \text{Leb}(A \cap B(0,M)) .$$

Combining this result and (36) concludes the proof of (a) and (b). The last statement then follows from [10, Theorem 14.0.1]

6.2. Proofs of Section 4

6.2.1. Proof of Proposition 6

Note that (i) implies that

$$\inf_{q \in \partial B(0,R_1)} U_0(q) > 0 .$$

(39)

First, $H2$-(i) is straightforward using (ii) and the fact that for all $q \in \mathbb{R}^d$, $||q|| \geq R_1, U_0(q) = (||q||/R_1)^{m} U_0(R_1q/||q||)$ by (i).

In addition, $H2$-(iii) is also easy to check using the Euler’s homogeneous function theorem showing that $\langle \nabla U_0(q), q \rangle = mU_0(q)$ for all $q \in \mathbb{R}^d$, $||q|| \geq R_1$.

We now show that $H2$-(ii) holds. First, since $\lim_{||q|| \to +\infty} U_0(q) = +\infty$ and $U_0$ is continuous for all $K \geq 0$, $L_K = \{ q \in \mathbb{R}^d ; U_0(q) \leq K \}$, is compact. Besides, using (39) and that $U_0$ is continuous, we can define

$$M = \sup_{q \in B(0,R_1)} U_0(q) + 1 \in (1, +\infty) .$$

(40)
and for all $q \notin L_M$,

$$t_q = \sup \{ t \in [0,1] : U_0(tq) = M \} ,$$

which satisfies

$$U_0(t_q q) = M > \sup_{x \in B(0,R_1)} U_0(x) , \ t_q q \in \partial L_M \text{ and } \|t_q q\| > R_1 .$$

(42)

Finally using (i), we get that the set $L_M$ is convex.

To show $\mathbf{H2}$-(ii), we check first that it is sufficient to prove that

$$D^2 U_0(x) \{ \nabla U_0(x) \otimes \nabla U_0(x) \} > 0 \text{ for any } x \in \partial L_M .$$

Indeed note that if this statement holds, since $U \in C^2(\mathbb{R}^d)$ and $\partial L_M$ is compact, we have

$$\varepsilon = \inf_{x \in \partial L_M} D^2 U_0(x) \{ \nabla U_0(x) \otimes \nabla U_0(x) \} > 0 .$$

(43)

Let now $q \notin L_M$ and $t_q$ defined by (41). Since by (i), for all $u \geq 1$ and $z \in \mathbb{R}^d$, $\|z\| \geq R_1$, $U_0(uz) = u^m U_0(z)$, differentiating with respect to $z$, we get

$$\nabla U_0(uz) = u^{m-1} \nabla U_0(z) \text{ and } D^2 U_0(uz) = u^{m-2} D^2 U_0(z).$$

Therefore by (42), we get

$$D^2 U_0(q) \{ \nabla U_0(q) \otimes \nabla U_0(q) \} = t_q^{4-3m} D^2 U_0(t_q q) \{ \nabla U_0(t_q q) \otimes \nabla U_0(t_q q) \} .$$

(44)

Using (42) again and since $\partial L_M$ is compact, we get that there exists $R_2 \geq 0$ such that $t_q \|q\| \in [R_1, R_2]$. Hence by (44), we have

$$D^2 U_0(q) \{ \nabla U_0(q) \otimes \nabla U_0(q) \} \geq \varepsilon \|q\|^{3m-4} \min \left[R_1^{4-3m}, R_2^{4-3m}\right] .$$

Thus $\mathbf{H2}$-(ii) holds for $U_0$. Finally (ii) implies that the function $U = U_0 + G$ satisfies $\mathbf{H2}$-(ii) as well.

Let $x \in \partial L_M$, we now show that $D^2 U_0(x) \{ \nabla U_0(x) \otimes \nabla U_0(x) \} > 0$. By Euler’s homogeneous function theorem and since $M \geq 1$, we have that $\| \nabla U_0(x) \| \geq m > 0$. Denote by $\Pi$ the tangent hyperplane of $\partial L_M$ at $x$, defined by $\Pi = \{ q \in \mathbb{R}^d : \langle \nabla U_0(x), x - q \rangle = 0 \}$. Since $L_M$ is convex, for all $q \in L_M$ and $t \in [0,1]$, $t^{-1}(U_0(tq + (1-t)x) - U_0(x)) \leq 0$. So taking the limit as $t$ goes to 0, we get that $\langle \nabla U_0(x), x - q \rangle \leq 0$. Therefore, $L_M$ is contained in the half-space $\Pi^- = \{ q \in \mathbb{R}^d : \langle \nabla U_0(x), x - q \rangle \leq 0 \}$.

Define the $m$-homogeneous function $\hat{U} : \mathbb{R}^d \to \mathbb{R}_+$ for all $q \in \mathbb{R}^d$ by

$$\hat{U}(q) = M \left| \langle q, \nabla U_0(x) \rangle \right|^m .$$

(45)

Since $U_0(x) = M$, by (40), $\|x\| > R_1$ and therefore there exists $\epsilon_0 \in \mathbb{R}_+^*$ such that

$$B(x, \epsilon_0) \subset \mathbb{R}^d \setminus B(0, R_1) .$$

(46)
We now show that $\tilde{U}(q) \leq U_0(q)$ for all $q \in B(x, \epsilon)$ with
\[ \epsilon = 2^{-1} \min \left[ \epsilon_0, \left\{ \frac{\langle x, \nabla U_0(x) \rangle}{\|\nabla U_0(x)\|^2} \right\} \right]. \]

First consider $q \in \Pi$. We next argue by contradiction that
\[ U_0(q) \geq M = \tilde{U}(q). \tag{47} \]
Indeed assume that $U_0(q) < M$. Then by continuity of $U_0$, we get that $q \in L^\infty_M$. But since $L_M \subset \Pi^-$, we get $q \in (\Pi^-)^c$ which is impossible since $q \in \Pi = \partial\Pi^- = \Pi^- \setminus (\Pi^-)^c$.

Let $q \in B(x, \epsilon)$. Note that $q = x + \|\nabla U_0(x)\|^{-2} \langle q - x, \nabla U_0(x) \rangle \nabla U_0(x) + z$, where $z \in \mathbb{R}^d$ is orthogonal to $\nabla U_0(x)$. Define
\[ u = \frac{\langle x, \nabla U_0(x) \rangle}{\langle q, \nabla U_0(x) \rangle}. \]
Then $uq \in \Pi$ and by (47), $U_0(uq) \geq M$. If $u \geq 1$, using (i) and (45), we get
\[ U_0(q) \geq u^{-m} M = \tilde{U}(q). \tag{48} \]
In turn, if $u < 1$, since $\|q - x\| \leq \epsilon_0$, by (46) and (i), $U_0(q) = u^{-1} U_0(uq)$ and (48) still holds.

Consider the three times differentiable functions $\phi$ and $\tilde{\phi}$ defined for all $v \in \mathbb{R}$ by
\[ \phi(v) = U_0(x + v\nabla U_0(x)) \quad \text{and} \quad \tilde{\phi}(v) = \tilde{U}(x + v\nabla U_0(x)). \]
First, since for all $q \in B(x, \epsilon)$, $U_0(q) \geq \tilde{U}(q)$, we have
\[ \phi(v) \geq \tilde{\phi}(v), \text{ for all } v \in [-\epsilon/\|\nabla U_0(x)\|, \epsilon/\|\nabla U_0(x)\|]. \tag{49} \]
Moreover, by definition $U_0(x) = \tilde{U}(x)$ and $\nabla \tilde{U}(x)$ is colinear to $\nabla U_0(x)$. Using Euler’s homogeneous function theorem for $U_0$ and $\tilde{U}$, we get that $\nabla \tilde{U}(x) = \nabla U_0(x)$. Therefore $\phi(0) = \tilde{\phi}(0)$, $\phi'(0) = \tilde{\phi}'(0)$. Combining these equalities, (49) and using a Taylor expansion around 0 of order 2 with exact remainder for $\phi$ and $\tilde{\phi}$ shows that necessary
\[ D^2 U_0(x) \{ \nabla U_0(x) \otimes \nabla U_0(x) \} = \phi''(0) \geq \tilde{\phi}''(0) > 0, \]
which concludes the proof.

6.2.2. Proof of Proposition 7

We preface the proof by a useful Lemma.

**Lemma 12.** Assume $H_{2(m)}(i)$ for some $m \in (1, 2]$. Then there exists $C \geq 0$ such that for all $q, x \in \mathbb{R}^d$,
\[ \|\nabla U(q) - \nabla U(x)\| \leq C \min \left( \|q - x\|, \|q - x\|^{m-1} \right). \]
Proof. First by **H2-(i)** we have for all \( q, x \in \mathbb{R}^d \),

\[
\| \nabla U(q) - \nabla U(x) \| = \left\| \int_0^1 \nabla^2 U(x + t(q-x)) \{ q - x \} \, dt \right\| \\
\leq A_1 \| q - x \| \int_0^1 \{ 1 + \| x + t(q-x) \| \}^{-2} \, dt .
\] (50)

Therefore, for all \( q, x \in \mathbb{R}^d \), we get \( \| \nabla U(q) - \nabla U(x) \| \leq A_1 \| q - x \| \). It remains to show that there exists \( C \in \mathbb{R}_+ \) such that for all \( q, x \in \mathbb{R}^d \), we have \( \| \nabla U(q) - \nabla U(x) \| \leq C \| q - x \|^{m-1} \). For all \( q, x \in \mathbb{R}^d \), since \( m \in (1, 2) \), we have

\[
\int_0^1 \{ 1 + \| x + t(q-x) \| \}^{-2} \, dt \leq \int_0^1 \{ 1 + \| x \| - t \| q - x \| \}^{-2} \, dt \\
\leq \int_0^{1 \wedge \| q - x \|^{-1}} \{ 1 + \| x \| - t \| q - x \| \}^{-2} \, dt \\
+ \int_0^1 \{ 1 + t \| q - x \| - \| x \| \}^{-2} \, dt \\
\leq (m-1)^{-1} \| q - x \|^{m-2} .
\]

Plugging this result in (50) concludes the proof.

\( \square \)

Proof of Proposition 7. By Lemma 12 and Lemma 1-(a), for all \( q_0 \in \mathbb{R}^d \), \( p \mapsto \Phi_h^{o(T)}(q_0, p) \) is Lipschitz, with a Lipschitz constant \( L \in \mathbb{R}_+ \) which does not depend \( q_0 \). Therefore by the log-Sobolev inequality [1, Proposition 5.5.1, (5.4.1)] and (16), we get for all \( q_0 \in \mathbb{R}^d \)

\[
K_{h,T} V_a(q_0) \leq \exp \left( (aL)^2 / 2 + a\mathcal{E}(q_0) \right) ,
\]

where

\[
\mathcal{E}(q_0) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \left\| \Phi_h^{o(T)}(q_0, p) \right\| e^{-\| p \|^2/2} \, dp .
\]

Therefore to conclude the proof, it is sufficient to show that

\[
\limsup_{\| q_0 \| \to +\infty} \{ \mathcal{E}(q_0) - \| q_0 \| \} < 0 ,
\] (51)

provided that we take \( a \) small enough. We will soon check that for all \( p_0 \in \mathbb{R}^d \),

\[
\liminf_{\| q_0 \| \to +\infty} \left( \frac{\| q_0 \| - \left\| \Phi_h^{o(T)}(q_0, p_0) \right\|}{1 + \| q_0 \|^{m-1}} \right) > 0 .
\] (52)

Assume that (52) is satisfied. By Lemma 12 and Lemma 1-(b), there exists \( C \), such that for all \( q_0, p_0 \in \mathbb{R}^d \),

\[
\left( \| q_0 \| - \left\| \Phi_h^{o(T)}(q_0, p_0) \right\| \right) / \left( 1 + \| q_0 \|^{m-1} \right) \geq -C(\| p_0 \| + 1) ,
\]
The Fatou’s Lemma then shows that \( \liminf_{\|q_0\| \to +\infty} (\|q_0\| - \delta'(q_0)) / (1 + \|q_0\|^{m-1}) > 0 \), and therefore (51) is true.

It remains to establish (52). Fix \( p_0 \in \mathbb{R}^d \). Denote for all \( k \in \{0, \ldots, T\} \), 
\[ q_k = \Phi_h^{(k)}(q_0, p_0) \]
and consider the following decomposition given by (9):
\[ \|q_T\|^2 = \|q_0\|^2 + A(q_0) - 2h^2B(q_0), \] (53)
where
\[ A(q_0) = 2Th \langle q_0, p_0 \rangle + \left\| Thp_0 - (Th^2/2)\nabla U(q_0) - h^2 \sum_{i=1}^{T-1} (T - i)\nabla U(q_i) \right\|^2 \]
\[ B(q_0) = 2 \left\langle q_0, (T/2)\nabla U(q_0) + \sum_{i=1}^{T-1} (T - i)\nabla U(q_i) \right\rangle. \]

(a) By Lemma 12 and Lemma 1-(b), we get since \( m \in (1, 2) \),
\[ \limsup_{\|q_0\| \to +\infty} \frac{A(q_0)}{\|q_0\|^m} = 0. \] (54)

Note that \( B \) can be written in the form:
\[ B(q_0) = \frac{Th^2}{2} \langle q_0, \nabla U(q_0) \rangle + h^2 \sum_{i=1}^{T-1} (T - i) \langle q_i, \nabla U(q_i) \rangle \]
\[ + h^2 \sum_{i=1}^{T-1} (T - i) \langle q_0 - q_i, \nabla U(q_i) \rangle, \]
Then using Lemma 12, Lemma 1-(b), \( m < 2 \) again and \( \mathbf{H2}(m)-(iii) \) we get
\[ \liminf_{\|q_0\| \to +\infty} \frac{B(q_0)}{\|q_0\|^m} > 0. \] (55)

Combining (54) and (55) in (53) easily yields (52).

(b) Let \( h_0 \in \mathbb{R}^*_+ \) and \( h \in (0, h_0) \). For \( m = 2 \), following the same reasoning, we have that there exist \( C_1, C_2 \in \mathbb{R}^*_+ \) independent of \( h \), such that
\[ \limsup_{\|q_0\| \to +\infty} \frac{A(q_0)}{\|q_0\|^2} \leq C_1 h^4, \text{ and } \liminf_{\|q_0\| \to +\infty} \frac{B(q_0)}{\|q_0\|^2} > C_1 h^2(C_2 - h^2). \]

Then taking \( h \) small enough, (52) follows from plugging this result in (53).

\[ \square \]

6.2.3. Proof of Proposition 8

We preface the proof by several technical preliminary Lemmas.
Lemma 13. Assume $H_2(m)$-(i) for some $m \in (1,2]$. Let $T \in \mathbb{N}^*$ and $\gamma \in (0, m-1)$.

(a) If $m \in (1,2)$ and $h_0 \in \mathbb{R}_+^*$, there exist $\kappa \in \mathbb{R}_+^*$ and $R \in \mathbb{R}_+$ such that for all $h \in (0,h_0]$, $q_0, p_0 \in \mathbb{R}^d$ satisfying $\|p_0\| \leq \|q_0\|^\gamma$ and $\|q_0\| \geq R$, and $i,j,k \in \{1, \ldots, T\}$,

$$\|\Phi_h^{(i)}(q_0, p_0) - \Phi_h^{(j)}(q_0, p_0)\| \leq \kappa h \|\Phi_h^{(k)}(q_0, p_0)\|^{m-1},$$

where $\Phi_h^{(\ell)}$ are defined by (8) for $\ell \in \mathbb{N}^*$.

(b) If $m = 2$, there exist $h_0 \in \mathbb{R}_+^*$, $\kappa \in \mathbb{R}_+^*$ and $R \in \mathbb{R}_+$ such that for all $h \in (0,h_0]$, $q_0, p_0 \in \mathbb{R}^d$ satisfying $\|p_0\| \leq \|q_0\|^\gamma$ and $\|q_0\| \geq R$, and $i,j,k \in \{1, \ldots, T\}$, (56) is satisfied.

Proof. (a) Let $h_0 \in \mathbb{R}_+^*$ and $h \in (0,h_0]$. Denote for all $k \in \{0, \ldots, T\}$ by $(q_k, p_k) = \Phi_h^{(k)}(q_0, p_0)$. By Lemma 12, Lemma 1-(b), there exist $C \geq 0$ and $R_1 \geq 0$ such that for all $q_0, p_0 \in \mathbb{R}^d$ satisfying $\|p_0\| \leq \|q_0\|^\gamma$ and $\|q_0\| \geq R_1$, for all $k \in \{1, \ldots, T\}$, we have

$$\|q_k - q_0\| \leq C h \|q_0\|^{m-1}.$$ (57)

Then there exists $R_2 \geq R_1$ and $\omega > 0$ such that such that for all $q_0, p_0 \in \mathbb{R}^d$ satisfying $\|p_0\| \leq \|q_0\|^\gamma$ and $\|q_0\| \geq R_2$, for all $k \in \{1, \ldots, T\}$,

$$\|q_0\| \leq \omega \|q_k\|.$$ (58)

In addition, using this inequality and (57) again, we get that for all $q_0, p_0 \in \mathbb{R}^d$ satisfying $\|p_0\| \leq \|q_0\|^\gamma$ and $\|q_0\| \geq R_2$, for all $i, j, k \in \{1, \ldots, T\}$,

$$\|q_i - q_j\| \leq 2Ch \|q_0\|^{m-1} \leq 2Ch\omega^{m-1} \|q_k\|^{m-1}.$$ (59)

(b) The proof follows the same lines as (a) upon noting that (58) holds for $h_0$ small enough.

Lemma 14. Assume $H_2(m)$-(i)-(ii) for some $m \in (1,2]$. Then there exist $R_0 \in \mathbb{R}_+$ and $\delta, \eta \in \mathbb{R}_+^*$ such that for all $q, x, z \in \mathbb{R}^d$, with

$$\|q\| \geq R_0, \quad \text{and} \quad \max(\|q - x\|, \|q - z\|) \leq \delta \|q\|,$$

we have

$$D^2U(q) \{\nabla U(x) \otimes \nabla U(z)\} \geq \eta \|q\|^{3m-4}.$$ (60)

Proof. Let $\delta > 0$ and $q, x, z \in \mathbb{R}^d$ satisfy $\max(\|q - x\|, \|q - z\|) \leq \delta \|q\|$. By $H_2$-(i)-(iii) and Lemma 12, it can be easily checked that there exists $C \geq 0$ such that for all $q \in \mathbb{R}^d$, $\|q\| \geq R$,

$$D^2U(q) \{\nabla U(x) \otimes \nabla U(z)\} \geq A_2 \|q\|^{3m-4} - C \left\{1 + \delta^{m-1} \|q\|^{3m-4}\right\}.$$ (61)

Choosing $R_0$ large and $\delta$ small enough concludes the proof.
Lemma 15. Assume that $U$ is twice continuously differentiable. Then for all $q_0, p_0 \in \mathbb{R}^d$ and $h \in \mathbb{R}^*_+$, the following identity holds

$$H \circ \Phi_h^{(1)}(q_0, p_0) - H(q_0, p_0) = h^2 \int_0^1 D^2U(q_t) \{p_0 \otimes \nabla U(q_0)\} \otimes^2 (1/2 - t) \, dt$$

$$+ h^3 \int_0^1 D^2U(q_t) \{p_0 \otimes \nabla U(q_0)\} \{q_1 - q_0\} \otimes^2 (t - 1/4) \, dt$$

$$- \frac{h^4}{4} \int_0^1 D^2U(q_t) \{\nabla U(q_0)\} \{q_1 - q_0\} \otimes^2 \{q_1 - q_0\} \, dt + \frac{h^4}{8} \left\| \int_0^1 \nabla^2 U(q_t) p_0 \, dt \right\|^2$$

$$- \frac{h^5}{8} \left( \int_0^1 \nabla^2 U(q_t) \nabla U(q_0) \, dt, \int_0^1 \nabla^2 U(q_t) p_0 \, dt \right)$$

$$+ \frac{h^6}{32} \left\| \int_0^1 \nabla^2 U(q_t) \nabla U(q_0) \, dt \right\|^2,$$

where $\Phi_h^{(1)}$ is defined in (7) and $q_t = q_0 + t(q_1 - q_0)$ for $t \in [0, 1]$.

Proof. Let $q_0, p_0 \in \mathbb{R}^d$ and $(q_1, p_1) = \Phi_h^{(1)}(q_0, p_0)$. Recall that $H$ is given for all $q, p \in \mathbb{R}^d$ by $H(q, p) = \frac{1}{2} \| p \|^2 + U(q)$. Therefore

$$H(q_1, p_1) - H(q_0, p_0) = (1/2)(\| p_1 \|^2 - \| p_0 \|^2) + U(q_1) - U(q_0).$$

First, Taylor’s formula with exact remainder enables us to write

$$U(q_1) - U(q_0) = \langle \nabla U(q_0), (q_1 - q_0) \rangle + \int_0^1 D^2U(q_t) \{q_1 - q_0\} \otimes^2 (1 - t) \, dt. \quad (59)$$

Since $\nabla U(q_1) = \nabla U(q_0) + \int_0^1 \nabla^2 U(q_t) \{q_1 - q_0\} \, dt$, we get

$$p_1 = p_0 - \frac{h}{2} (\nabla U(q_0) + \nabla U(q_1)) = p_0 - h \nabla U(q_0) - \frac{h}{2} \int_0^1 \nabla^2 U(q_t) \{q_1 - q_0\} \, dt. \quad (60)$$

Using that $q_1 = \Phi_h^{(1)}(p_0, q_0)$, with $\Phi_h^{(1)}$ defined by (8), in (59) and (60), we get

$$U(q_1) - U(q_0) = \langle \nabla U(q_0), hp_0 - (h^2/2) \nabla U(q_0) \rangle$$

$$+ \int_0^1 D^2U(q_t) \{q_1 - q_0\} \otimes^2 (1 - t) dt,$$

and

$$\frac{1}{2}(\| p_1 \|^2 - \| p_0 \|^2) = (h^2/2) \| \nabla U(q_0) \|^2 + \int_0^1 D^2U(q_t) \{q_1 - q_0\} \, dt$$

$$- h \langle p_0, \nabla U(q_0) \rangle - (h/2) \int_0^1 D^2U(q_t) \{p_0 \otimes (q_1 - q_0)\} \, dt$$

$$+ (h^2/2) \int_0^1 D^2U(q_t) \{\nabla U(q_0) \otimes (q_1 - q_0)\} \, dt.$$
Summing these equalities up and observing appropriate cancellations yields

\[ H(q_1, p_1) - H(q_0, p_0) = \int_0^1 D^2U(q_t) \{ q_1 - q_0 \} \otimes^2 (1 - t) dt \]

\[ - (h/2) \int_0^1 D^2U(q_t) \{ p_0 \otimes (q_1 - q_0) \} dt + (h^2/8) \left\| \int_0^1 \nabla^2U(q_t) \{ q_1 - q_0 \} dt \right\|^2 \]

\[ + (h^2/2) \int_0^1 D^2U(q_t) \{ \nabla U(q_0) \otimes (q_1 - q_0) \} dt \]

. (61)

By using \( q_1 = \Phi_h^{\omega_0} (p_0, q_0) \) again in the definition of each \( I_j \) we obtain consecutively

\[ I_1 = h^2 \int_0^1 D^2U(q_t) \{ p_0 \} \otimes^2 (1 - t) dt - h^3 \int_0^1 D^2U(q_t) \{ p_0 \otimes \nabla U(q_0) \} (1 - t) dt \]

\[ + (h^4/4) \int_0^1 D^2U(q_t) \{ \nabla U(q_0) \} \otimes^2 (1 - t) dt \]

\[ I_2 = -(h^2/2) \int_0^1 D^2U(q_t) \{ p_0 \} \otimes^2 dt + (h^3/4) \int_0^1 D^2U(q_t) \{ p_0 \otimes \nabla U(q_0) \} dt \]

\[ I_3 = (h^4/8) \left\| \int_0^1 \nabla^2 U(q_t) p_0 dt \right\|^2 + (h^6/32) \left\| \int_0^1 \nabla^2 U(q_t) \nabla U(q_0) dt \right\|^2 \]

\[ - (h^5/8) \left( \int_0^1 \nabla^2 U(q_t) \nabla U(q_0) dt, \int_0^1 \nabla^2 U(q_t) p_0 dt \right) \]

and

\[ I_4 = (h^3/2) \int_0^1 D^2U(q_t) \{ \nabla U(q_0) \otimes p_0 \} dt \]

\[ - (h^4/4) \int_0^1 D^2U(q_t) \{ \nabla U(q_0) \} \otimes^2 dt \]

Gathering all these equalities in (61) concludes the proof.

Proof of Proposition 8. For all \( h \in \mathbb{R}^*_+, T \in \mathbb{N}^* \) and \( k \in \{0, \ldots, T\} \), denote \( (q_k, p_k) = \Phi_h^{\omega(k)} (q_0, p_0) \). Let \( R_0 \geq 0 \) and \( \delta, \eta > 0 \) be the constants defined in Lemma 14.

(a) Fix \( h \in \mathbb{R}^*_+ \). Since \( m \in (1, 2) \), by Lemma 13-(a) there exists \( R_1 \geq R_0 \) such that for all \( q_0, p_0 \in \mathbb{R}^d \) satisfying \( \|p_0\| \leq \|q_0\|^\gamma \) and \( \|q_0\| \geq R_1 \), for all \( i, j, k \in \{1, \ldots, T\} \),

\[ \|q_i - q_j\| \leq \delta \|q_k\| \]  

(62)

For all \( q_0, p_0 \in \mathbb{R}^d \), consider the following decomposition

\[ H(p_T, q_T) - H(p_0, q_0) = \sum_{k=0}^{T-1} \{ H(p_{k+1}, q_{k+1}) - H(p_k, q_k) \} \]  

(63)
We show that each term in the sum in the right hand side of this equation is nonpositive if \(|q_0|\) is large enough and \(|p_0| \leq |q_0|^{\gamma}\). By Lemma 15, we have

\[
H(q_{k+1}, p_{k+1}) - H(q_k, p_k) = -(h^4/4)A_k + h^2B_k + h^3C_k + (h^4/8)D_k ,
\]

(64)

where

\[
A_k = \int_0^1 D^2U(q_{t,k}) \{\nabla U(q_k)\} \otimes_2 t \ dt
\]

\[
B_k = \int_0^1 D^2U(q_{t,k}) \{p_k\} \otimes_2 (1/2 - t) \ dt
\]

\[
C_k = \int_0^1 D^2U(q_{t,k}) \{p_k \otimes \nabla U(q_k)\} (t - 1/4) \ dt
\]

\[
D_k = \left\| \int_0^1 \nabla^2 U(q_{t,k})p_k \ dt \right\|^2 + (h^2/4) \left\| \int_0^1 \nabla^2 U(q_{t,k}) \nabla U(q_k) \ dt \right\|^2
\]

\[
- h \left\langle \int_0^1 \nabla^2 U(q_{t,k}) \nabla U(q_k) \ dt, \int_0^1 \nabla^2 U(q_{t,k})p_k \ dt \right\>
\]

We will next estimate each of these terms separately.

By (71), Lemma 14 and (58), for all \(q_0, p_0 \in \mathbb{R}^d\), \(|q_0| \geq R_1\) and \(|p_0| \leq |q_0|^\gamma\), we get that there exists \(\eta \in \mathbb{R}^+_+\) such that

\[
\inf_{\|p_0\| \leq |q_0|^\gamma} A_k \geq \eta \|q_k\|^{3m-4} \geq \eta |q_0|^{3m-4}.
\]

(65)

We now bound \(B_k\). First note that since \(q_{t,k} - q_k = -(th^2/2)\nabla U(q_k) + thp_k\) and \(\int_0^1 (1/2 - t) \ dt = 1/2\), we have for all \(q_0, p_0 \in \mathbb{R}^d\),

\[
B_k = \int_0^1 \int_0^t D^3U(q_k + s(q_{k,t} - q_k)) \{p_k^{\otimes 2} \otimes (q_{k,t} - q_k)\} (1/2 - t) ds \ dt
\]

\[
= h \int_0^1 \int_0^t D^3U(q_k + s(q_{k,t} - q_k)) \{p_k\}^{\otimes 3} t(1/2 - t) ds \ dt
\]

\[
- (h^2/2) \int_0^1 \int_0^t D^3U(q_k + s(q_{k,t} - q_k)) \{p_k^{\otimes 2} \otimes \nabla U(q_k)\} t(1/2 - t) ds \ dt.
\]

By (57) and \(\textbf{H2-(i)}\), we get

\[
\|B_k\| \leq C \left\{ 1 + |q_0| \right\}^{m-3} \left\{ h \|p_k\|^3 + h^2 \|p_k\|^2 \|\nabla U(q_k)\| \right\}.
\]

Using in this inequality, \(\textbf{H2-(i)}\), Lemma 12 and Lemma 1-(b), we have for all \(q_0, p_0 \in \mathbb{R}^d\), \(|p_0| \leq |q_0|^\gamma\),

\[
\|B_k\| \leq C \left\{ 1 + |q_0| \right\}^{m-3} \times \left\{ h(1 + |q_0|^\gamma) + h |q_0|^{m-1}\right\}^3 + h^2(1 + |q_0|^\gamma) + h |q_0|^{m-1}\right\}^2 \left\{ 1 + |q_0|^{m-1}\right\},
\]
which implies that there exists \( C \geq 0 \) such that
\[
\limsup_{\|q_0\| \to +\infty} \sup_{\|p_0\| \leq \|q_0\|} \left\{ \left\| B_k \right\| / \|q_0\|^{4m-6} \right\} \leq C h^4 . \tag{66}
\]
Combining \( H2-(i), \) Lemma 12 and Lemma 1-(b) again, we get by crude estimate
\[
\limsup_{\|q_0\| \to +\infty} \sup_{\|p_0\| \leq \|q_0\|} \left\{ \|D_k\| / \|q_0\|^{4m-6} \right\} \leq C h^2 . \tag{67}
\]
Consider now the term \( C_k \) in (64). Using again that \( \int_0^1 (1/2 - t) \, dt = 0 \) and \( q_{t,k} - q_k = -(th^2/2) \nabla U(q_k) + thp_k \), we get
\[
C_k = C_{k,1} + C_{k,2}/4 , \tag{68}
\]
with
\[
C_{k,1} = h \int_0^1 \int_0^t D^3 U(q_k + s(q_{k,t} - q_k)) \left\{ p_k^{\otimes 2} \otimes \nabla U(q_k) \right\} t(t - 1/2)ds \, dt
- (h^2/2) \int_0^1 \int_0^t D^3 U(q_k + s(q_{k,t} - q_k)) \left\{ p_k \otimes (\nabla U(q_k))^{\otimes 2} \right\} t(t - 1/2)ds \, dt
C_{k,2} = \int_0^1 D^2 U(q_{t,k}) \left\{ p_k \otimes \nabla U(q_k) \right\} \, dt .
\]
We bound the two terms in the right hand side of (68) separately. First, using the same reasoning as for \( B_k \), we get that there exists \( C \geq 0 \) such that
\[
\limsup_{\|q_0\| \to +\infty} \sup_{\|p_0\| \leq \|q_0\|} \left\{ \left\| C_{k,1} \right\| / \|q_0\|^{4m-6} \right\} \leq C h^3 . \tag{69}
\]
Regarding \( C_{k,2} \), we have using (10) that for all \( q_0, p_0 \in \mathbb{R}^d \)
\[
C_{k,2} = \int_0^1 D^2 U(q_{t,k}) \left\{ p_0 \otimes \nabla U(q_k) \right\} \, dt
- h \sum_{i=1}^{k-1} \int_0^1 D^2 U(q_{t,k}) \left\{ \nabla U(q_i) \otimes \nabla U(q_k) \right\} \, dt
- (h/2) \int_0^1 D^2 U(q_{t,k}) \left\{ (\nabla U(q_0) + \nabla U(q_k)) \otimes \nabla U(q_k) \right\} \, dt .
\]
By (71) and Lemma 14, the two last terms in the right hand side dominate the first one, therefore we get \( \limsup_{\|q_0\| \to +\infty} \inf_{\|p_0\| \leq \|q_0\|} \gamma \{ C_{k,2} / \|q_0\|^{4m-6} \} \leq 0 \). Combining this result and (69) in (68), we get
\[
\limsup_{\|q_0\| \to +\infty} \sup_{\|p_0\| \leq \|q_0\|} \left\{ \left\| C_k \right\| / \|q_0\|^{4m-6} \right\} \leq C h^3 . \tag{70}
\]
Gathering (65), (66), (67) and (70) in (64), we get using $3m - 4 \geq 4m - 6$ for $m \in (1, 2)$, that for all $k \in \{0, \ldots, T - 1\}$,

$$\limsup_{\|q_0\| \to +\infty} \sup_{\|p_0\| \leq \|q_0\|} \| H(q_{k+1}, p_{k+1}) - H(q_k, p_k) \| / \|q_0\|^{3m-4} < 0,$$

which concludes the proof.

(b) When $m = 2$, by Lemma 13-(b) there exists $R_1 \geq R_0$ and $h_0 > 0$ such that for all $h \in (0, h_0]$, $q_0, p_0 \in \mathbb{R}^d$ satisfying $\|p_0\| \leq \|q_0\|$ and $\|q_0\| \geq R_1$, for all $i, j, k \in \{1, \ldots, T\}$,

$$\|q_i - q_j\| \leq \delta \|q_k\|. \quad (71)$$

Consider now the same decomposition (63) and (64) again. Note that following the same lines as the proof of (a) we get that for all $h \in (0, h_0)$, for all $q_0, p_0 \in \mathbb{R}^d$, $\|q_0\| \geq R_1$ and $\|p_0\| \leq \|q_0\|$ and $\|q_0\| \geq R_1$, for all $i, j, k \in \{1, \ldots, T\}$,

$$\|q_i - q_j\| \leq \delta \|q_k\|. \quad (71)$$

Therefore taking $h_H < \min(h_0, (\eta/C)^{1/2})$ concludes the proof.

\[ \square \]

6.3. Proof of Section 5

6.3.1. Proof of Theorem 10

We first recall a fact from geometric measure theory together with a proof for the reader’s convenience. Let $U \subset \mathbb{R}^d$ be an open set and $\Theta : U \to \mathbb{R}^d$ be a measurable function such that there exist $y_0, \tilde{y}_0 \in \mathbb{R}^d$ and $M, \tilde{M} > 0$ satisfying

$$B(y_0, M) \subset \Theta(B(\tilde{y}_0, \tilde{M})). \quad (72)$$

Define the measure $\lambda_{\Theta}$ on $(\mathbb{R}^d, B(\mathbb{R}^d))$ by setting for any $A \in B(\mathbb{R}^d)$

$$\lambda_{\Theta}(A) \overset{\text{def}}{=} \text{Leb} \left\{ \Theta^{-1}(A) \cap B(\tilde{y}_0, \tilde{M}) \right\}. \quad (73)$$

Note that $\lambda_{\Theta}$ is a finite measure. Therefore by the Lebesgue decomposition theorem (see [15, Section 6.10]) there exist two measures $\lambda_{\Theta}^{(a)}$, $\lambda_{\Theta}^{(s)}$ on $(\mathbb{R}^d, B(\mathbb{R}^d))$, which are absolutely continuous and singular with respect to the Lebesgue measure on $\mathbb{R}^d$ respectively, such that $\lambda_{\Theta} = \lambda_{\Theta}^{(a)} + \lambda_{\Theta}^{(s)}$.

Proposition 16. Let $U \subset \mathbb{R}^d$ be open and $\Theta : U \to \mathbb{R}^d$ be a Lipschitz function satisfying (72). For any version $\phi_{\Theta}$ of the density of $\lambda_{\Theta}^{(a)}$ with respect to the Lebesgue measure on $\mathbb{R}^d$, it holds

$$\phi_{\Theta}(y) \geq \mathbbm{1}_{B(y_0, M)}(y) \|\Theta\|^{-d}_{\text{Lip}}, \quad \text{Leb-a.e.}$$
Proof. Denote by $L = \|\Theta\|_{\text{Lip}}$. Let $y \in B(y_0, M)$. By (72), we may pick $z \in B(y_0, M)$ such that $\Theta(z) = y$. Let $\delta_0 > 0$ be such that $B(z, \delta_0/L) \subset B(y_0, M)$. Since $\Theta$ is Lipschitz continuous, for all $\delta \in \mathbb{R}_+^*$, $\Theta(B(z, \delta/L) \cap U) \subset B(y, \delta)$. Hence, for all $\delta \in (0, \delta_0]$, we have

$$\lambda_\Theta(B(y, \delta)) \geq L^{-d} \text{Leb}(B(z, \delta)) = L^{-d} \text{Leb}(B(y, \delta)).$$

The claim follows from the differentiation theorem for measures, see [15, Theorem 7.14].

By definition, we have

$$K(x, A) = \int_{\mathbb{R}^d} 1_A(f(x, z)) \phi(z) dz = \int_{\mathbb{R}^d} 1_{f^{-1}(A)}(z) \phi(z) dz$$

$$\geq \inf_{z \in B(0, \tilde{M})} \{ \phi(z) \} \text{Leb}\left\{ f^{-1}(A) \cap B(0, \tilde{M}) \right\} .$$

(74)

By assumption, we get using Proposition 16 that

$$\text{Leb}\left\{ f^{-1}(A) \cap B(0, \tilde{M}) \right\} \geq L_f^{-d} \text{Leb}\left\{ A \cap B(y_0, M) \right\} .$$

(75)

Combining (74) and (75) concludes the proof.

6.3.2. Proof of Proposition 11

We begin the proof by recalling some basic notions of degree theory. Let $D$ be a bounded open set of $\mathbb{R}^d$. Let $f : \overline{D} \to \mathbb{R}^d$ be a continuous function on $\overline{D}$ continuously differentiable on $D$. An element $x \in D$ is said to be a regular point of $f$ if $J_f(x)$ is invertible. An element $y \in f(D)$ is said to be a regular value of $f$ if any $x \in f^{-1}(\{y\})$ is a regular point.

Let $f : \overline{D} \to \mathbb{R}^d$ be a continuous function, $C^\infty$-smooth on $D$. Let $y \in \mathbb{R}^d \setminus f(\partial D)$ be a regular value of $f$. It is shown in [13, Proposition and Definition 1.1] that the set $f^{-1}(\{y\})$ is finite. The degree of $f$ at $y$ is defined by

$$\deg(f, D, y) = \sum_{x \in f^{-1}(\{y\})} \text{sign} \{ \det(J_f(x)) \} .$$

Proposition 17 ([13, Proposition and Definition 2.1]). Let $f : \overline{D} \to \mathbb{R}^d$ be a continuous function and $y \in \mathbb{R}^d \setminus f(\partial D)$.

(a) Then there exists $g \in C(\overline{D}, \mathbb{R}^d) \cap C^\infty(D, \mathbb{R}^d)$ such that $y$ is a regular value of $g$ and $\sup_{x \in \overline{D}} |f(x) - g(x)| < \text{dist}(y, f(\partial D))$.

(b) For all functions $g_1, g_2 : \overline{D} \to \mathbb{R}^d$ satisfying (a),

$$\deg(g_1, D, y) = \deg(g_2, D, y) .$$

Under the assumptions of Proposition 17, the degree of $f$ at $y$ is then defined for any $g : \overline{D} \to \mathbb{R}^d$ satisfying (a) by

$$\deg(f, D, y) = \deg(g, D, y) .$$
Proposition 18 ([13, Proposition 2.4]), Let \( f, g : \overline{\mathbb{D}} \rightarrow \mathbb{R}^d \) be continuous functions. Define \( H : [0, 1] \times \mathbb{R}^d \rightarrow \mathbb{R}^d \) for all \( t \in [0, 1] \) and \( x \in \mathbb{R}^d \) by \( H(t, x) = tf(x) + (1 - t)g(x) \). Let \( y \in \mathbb{R}^d \setminus \mathbb{H}([0, 1] \times \partial \mathbb{D}) \). Then

\[
\deg(f, \mathbb{D}, y) = \deg(g, \mathbb{D}, y).
\]

We have now all the necessary results to prove Proposition 11. Since \( f^g(x, z) = bz + g(x, z) \) and \( g(x, \cdot) \) is Lipschitz with a Lipschitz constant which is uniformly bounded over the ball \( B(0, R) \), \( f^g \) is Lipschitz with bounded Lipschitz constant over this ball. Hence \( \mathbf{G}\{R, 0, M\}(i) \) holds.

For all \( x \in \mathbb{R}^d \), denote by \( f^g_x : z \mapsto f^g(x, z) \) where \( f^g(x, z) = bz + g(x, z) \). Let \( M \in \mathbb{R}_+^* \). We show that for all \( x \in B(0, R) \), \( B(0, M) \subset f_X^g(B(0, \tilde{M})) \), where \( \tilde{M} \) is given by (30), which is precisely \( \mathbf{G}\{R, 0, M\}(ii) \).

Let \( x \in B(0, R) \) and consider the continuous homotopy \( H : [0, 1] \times \mathbb{R}^d \) between the functions \( z \mapsto bz \) and \( f^g_X \) defined for all \( t \in [0, 1] \) and \( z \in \mathbb{R}^d \) by

\[
H(t, z) = tbz + (1 - t)f^g_X(z) = bz + (1 - t)g(x, z).
\]

Then by (ii), since \(|b| \geq C_{R,1} \), for all \( t \in [0, 1] \) and \( z \notin B(0, \tilde{M}) \), where \( \tilde{M} \) is given by (30),

\[
|H(t, z)| \geq |bz| - (1 - t)\{C_{R,0} + C_{R,1}|z|\} \geq M.
\]

In particular, we have \( H([0, 1] \times \partial B(0, \tilde{M})) \subset \mathbb{R}^d \setminus B(0, M) \). Let \( z \in B(0, M) \), then by Proposition 18 we have

\[
\deg(f^g_X, B(0, \tilde{M}), z) = \deg(\text{Id}, B(0, \tilde{M}), z) = 1.
\]

Besides, by [13, Corollary 2.5, Chapter IV], \( \deg(f^g_X, B(0, \tilde{M}), z) \neq 0 \) implies that there exists \( y \in B(0, M) \) such that \( f^g_X(y) = z \). Finally \( \mathbf{G}\{R, 0, M\}(ii) \) follows since this result holds for all \( z \in B(0, M) \).

Appendix A: Harris recurrence for mixture of Metropolis-Hastings type Markov kernels

Let \((X, \mathcal{X})\) be a measurable space and \( \lambda \) be a \( \sigma \)-finite measure on \( \mathcal{X} \). For all \( i \in \mathbb{N}^* \), let \( \alpha_i : X \times X \rightarrow [0, 1] \) be a measurable function and \( k_i : X \times X \rightarrow [0, +\infty] \) be a Markov transition density w.r.t. \( \lambda \). Consider the Markov kernel \( K_i \) on \( X \times \mathcal{X} \) defined by

\[
K_i(x, A) = \int_A \alpha_i(x, y)k_i(x, y)\lambda(dy) + \delta_x(A)r_i(x), \quad x \in X \text{ and } A \in \mathcal{X},
\]

where \( r_i(x) = 1 - \int_X \alpha_i(x, y)k_i(x, y)\lambda(dy) \) for all \( x \in X \). For instance, \( K_i \) may be a Markov kernel defined by a Metropolis-Hastings algorithm, i.e.

\[
\alpha_i(x, y) = \begin{cases} 
1, & \text{if } \pi(x)k_i(x, y) > 0, \\
\min \left\{ 1, \frac{\pi(y)k_i(y, x)}{\pi(x)k_i(x, y)} \right\}, & \text{otherwise,}
\end{cases}
\]
for some probability density \( \pi : X \to (0, +\infty) \) with respect to \( \lambda \). We will use the results below in the case where \( K_i \) is a Markov kernel defined by the HMC algorithm. It is shown in [18, Corollary 2] that if \( K_i \) is irreducible, then \( K_i \) is Harris recurrent. We extend this result to the mixture Markov kernel \( K_a \) defined on \((X, \mathcal{X})\) by

\[
K_a = \sum_{i \in \mathbb{N}^*} a_i K_i \tag{77}
\]

where \((a_i)_{i \in \mathbb{N}^*}\) is a sequence of non-negative numbers satisfying \(\sum_{i \in \mathbb{N}^*} a_i = 1\).

**Proposition 19.** Let \( K_a \) be the Markov kernel given by (77) and associated with the sequence of Markov kernel \((K_i)_{i \in \mathbb{N}^*}\) given by (76). Let \( \pi \) be a probability measure on \((X, \mathcal{X})\). Assume that \( \pi \) and \( \lambda \) are mutually absolutely continuous and for all \( i \in \mathbb{N}^* \), \( \pi \) is invariant for \( K_i \). If \( K_a \) is irreducible and there exists \( i \in \mathbb{N}^* \) such that \( a_i > 0 \) and for all \( x \in X \) \( r_i(x) < 1 \), then \( K \) is Harris recurrent.

**Proof.** A bounded measurable function is said to be harmonic if \( K_a \phi = \phi \). By [10, Theorem 17.1.4, Theorem 17.1.7] a Markov kernel \( K_a \) is Harris recurrent if \( K_a \) is recurrent and any bounded harmonic function \( \phi : \mathbb{R}^d \to \mathbb{R} \) is constant. By [10, Theorem 10.1.1], since \( K_a \) is irreducible and admits \( \pi \) as an invariant probability measure, then \( K_a \) is recurrent. On the other hand, any bounded harmonic function \( \phi \) is \( \pi \)-almost surely equal to \( \pi(\phi) \) by [10, Theorem 17.1.1, Lemma 17.1.1]. Using that \( \pi \) and \( \lambda \) are mutually absolutely continuous, and \( \pi \) is an invariant probability measure for \( K_i \) for all \( i \in \mathbb{N}^* \), we get by (76) that for all \( x \in X \)

\[
K_a \phi(x) = \sum_{i \in \mathbb{N}^*} a_i \{\pi(\phi)(1 - r_i(x)) + \phi(x)r_i(x)\}.
\]

Combining this result with \( K_a \phi = \phi \), we get for all \( x \in X \)

\[
\{\phi(x) - \pi(\phi)\} \sum_{i \in \mathbb{N}^*} a_i \{1 - r_i(x)\} = 0.
\]

The condition that there exists \( i \in \mathbb{N}^* \) such that \( a_i > 0 \) and for all \( x \in X \) \( r_i(x) < 1 \), implies that for all \( x \in X \), \( \phi(x) = \pi(\phi) \).

**References**

[1] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer, Cham, 2014.

[2] S. Byrne and M. Girolami. Geodesic Monte Carlo on Embedded Manifolds. *Scandinavian Journal of Statistics*, 40(4):825–845, December 2013.

[3] E. Cancès, F. Legoll, and G. Stoltz. Theoretical and numerical comparison of some sampling methods for molecular dynamics. *M2AN Math. Model. Numer. Anal.*, 41(2):351–389, 2007.
[4] Bob Carpenter, Andrew Gelman, Matthew Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1), 2017.

[5] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Physics Letters B*, 195(2):216 – 222, 1987.

[6] J. J. Duistermaat and J. A. C. Kolk. *Multidimensional real analysis. I. Differentiation*, volume 86 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2004. Translated from the Dutch by J. P. van Braam Houckgeest.

[7] Mark Girolami and Ben Calderhead. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 73(2):123–214, 2011. With discussion and a reply by the authors.

[8] Benedict Leimkuhler and Sebastian Reich. *Simulating Hamiltonian dynamics*, volume 14 of Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, Cambridge, 2004.

[9] S. Livingstone, M. Betancourt, S. Byrne, and M. Girolami. On the geometric ergodicity of Hamiltonian Monte Carlo. *arXiv preprint arXiv:1601.08057*, 2016.

[10] S. Meyn and R. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, New York, NY, USA, 2nd edition, 2009.

[11] R. M. Neal. Bayesian learning via stochastic dynamics. *Advances in neural information processing systems*, pages 475–475, 1993.

[12] R. M. Neal. Mcmc using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, pages 113–162, 2011.

[13] E. Outerelo and J. M. Ruiz. *Mapping degree theory*, volume 108. American Mathematical Society Providence, RI, 2009.

[14] G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.

[15] W. Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987.

[16] M. R. Schofield, R. J. Barker, A. Gelman, E. R. Cook, and K. R. Briffa. A model-based approach to climate reconstruction using tree-ring data. *J. Amer. Statist. Assoc.*, 111(513):93–106, 2016.

[17] Y. Tang, N. Srivastava, and R. R. Salakhutdinov. Learning generative models with visual attention. In *Advances in Neural Information Processing Systems*, pages 1808–1816, 2014.

[18] L. Tierney. Markov chains for exploring posterior distributions (with discussion). *Ann. Statist.*, 22(4):1701–1762, 1994.