

# IRREDUCIBILITY AND GEOMETRIC ERGODICITY OF HAMILTONIAN MONTE CARLO

BY ALAIN DURMUS<sup>1</sup>, ÉRIC MOULINES<sup>2</sup> AND EERO SAKSMAN<sup>3</sup>

<sup>1</sup>CNRS, ENS, Université Paris-Saclay, [alain.durmus@ens-paris-saclay.fr](mailto:alain.durmus@ens-paris-saclay.fr)

<sup>2</sup>Faculty of Computer Sciences, HSE University, [eric.moulines@hse.ru](mailto:eric.moulines@hse.ru)

<sup>3</sup>Department of Mathematics and Statistics, University of Helsinki, [eero.saksman@helsinki.fi](mailto:eero.saksman@helsinki.fi)

Hamiltonian Monte Carlo (HMC) is currently one of the most popular Markov Chain Monte Carlo algorithms to sample smooth distributions over continuous state space. This paper discusses the irreducibility and geometric ergodicity of the HMC algorithm. We consider cases where the number of steps of the Störmer–Verlet integrator is either fixed or random. Under mild conditions on the potential  $U$  associated with target distribution  $\pi$ , we first show that the Markov kernel associated to the HMC algorithm is irreducible and positive recurrent. Under more stringent conditions, we then establish that the Markov kernel is Harris recurrent. We provide verifiable conditions on  $U$  under which the HMC sampler is geometrically ergodic. Finally, we illustrate our results on several examples.

**1. Introduction.** We consider the Hamiltonian Monte Carlo (HMC), a Metropolis–Hastings algorithm to sample from a target probability density  $\pi$  on  $\mathbb{R}^d$ . This method was first proposed by [9] in computational physics. It was later introduced to the statistics community in [22] and quickly gained popularity; see, for example, [15], Chapter 9, [13, 23].

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  $U : \mathbb{R}^d \rightarrow \mathbb{R}$  is a continuously differentiable function. Hamiltonian dynamics describes the evolution of a physical system which consists in the *position*  $q \in \mathbb{R}^d$  and the *momentum*  $p \in \mathbb{R}^d$ . The total energy of the system is given by the Hamiltonian function  $H$  defined for  $(q, p) \in \mathbb{R}^d \times \mathbb{R}^d$  by

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

Note that other choices of kinetic energy have been proposed recently; see, for example, [16] and [17]. The system  $(q(t), p(t))_{t \geq 0}$  then evolves according to Hamilton’s equations on  $\mathbb{R}^d \times \mathbb{R}^d$ ,

$$(1) \quad \frac{d}{dt} \begin{bmatrix} q(t) \\ p(t) \end{bmatrix} = \begin{bmatrix} p(t) \\ -\nabla U(q(t)) \end{bmatrix}.$$

The Hamiltonian flow associated with (1) preserves the extended target distribution (see [23] and [5]) with density  $\tilde{\pi}$  given for any  $(q, p) \in \mathbb{R}^{2d}$  by

$$(2) \quad \tilde{\pi}(q, p) = Z^{-1} \exp(-H(q, p)), \quad Z = \int_{\mathbb{R}^{2d}} \exp(-H(q, p)) dq dp.$$

---

Received May 2019; revised December 2019.

*MSC2020 subject classifications.* 60J05, 60J22, 65C05.

*Key words and phrases.* Hamiltonian Monte Carlo, Markov Chain Monte Carlo, irreducibility, geometric ergodicity.

The distribution  $\tilde{\pi}$  is the independent product of  $\pi$  and the  $d$ -dimensional Gaussian distribution with zero mean and identity covariance matrix. Therefore, sampling from  $\tilde{\pi}$  allows to get samples from  $\pi$  by marginalization. Since the Hamiltonian flow leaves  $\tilde{\pi}$  invariant, it has been suggested to sample  $\tilde{\pi}$  by sampling independently the momentum variable from a standard Gaussian distribution and then integrating the Hamiltonian flow during either a fixed or a random duration leading to an *idealized* version of HMC. The irreducibility and geometric ergodicity of this algorithm has been studied in [4]; see Section 3.2 for a discussion.

In most cases, it is not possible to compute explicitly the solutions of (1); discretization must be used instead. In this paper, we consider the Störmer–Verlet integrator which proceeds as follows. Let  $h \in \mathbb{R}_+^*$  be a time step and  $T \in \mathbb{N}^*$  be a number of iterations. The sequence  $(q_\ell, p_\ell)_{\ell \in \{0, \dots, T\}}$ , starting from  $(q_0, p_0) \in \mathbb{R}^d \times \mathbb{R}^d$  is defined by the recursion

$$\begin{cases} p_{\ell+1/2} = p_\ell - (h/2)\nabla U(q_\ell) \\ q_{\ell+1} = q_\ell + hp_{\ell+1/2} \\ p_{\ell+1} = p_{\ell+1/2} - (h/2)\nabla U(q_{\ell+1}). \end{cases}$$

This sequence defines a discrete dynamical system given for  $\ell \in \{0, \dots, T - 1\}$  by

$$(q_{\ell+1}, p_{\ell+1}) = \Psi_{h/2}^{(1)} \circ \Psi_h^{(2)} \circ \Psi_{h/2}^{(1)}(q_\ell, p_\ell) = \Phi_h^{(1)}(q_\ell, p_\ell),$$

where for each  $t \in \mathbb{R}_+$ ,  $\Psi_t^{(1)}, \Psi_t^{(2)} : \mathbb{R}^{2d} \rightarrow \mathbb{R}^{2d}$  are given for all  $(q, p) \in \mathbb{R}^{2d}$  by  $\Psi_t^{(1)}(q, p) = (q, p - t\nabla U(q))$  and  $\Psi_t^{(2)}(q, p) = (q + tp, p)$ . Define the sequence of iterates  $\{\Phi_h^{\circ(\ell)} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{R}^d : \ell \in \mathbb{N}^*\}$  for  $\ell \geq 1$  by induction

$$(3) \quad \Phi_h^{\circ(\ell+1)} = \Phi_h^{\circ(\ell)} \circ \Phi_h^{\circ(1)}.$$

Set, for all  $\ell \geq 1$ ,

$$(4) \quad \tilde{\Phi}_h^{\circ(\ell)} = \text{proj} \circ \Phi_h^{\circ(\ell)},$$

where  $\text{proj} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  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  $\ell \in \{1, \dots, T\}$ ,  $(q_\ell, p_\ell) = \Phi_h^{\circ(\ell)}(q_0, p_0)$  and  $q_\ell = \tilde{\Phi}_h^{\circ(\ell)}(q_0, p_0)$ . We now have all the background required to describe the HMC algorithm. 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, which are constructed to leave the extended distribution  $\tilde{\pi}$  invariant; see [23], [12] and [5], Theorem 5.7. In the first step, we draw  $G_{k+1}$  from the  $d$ -dimensional normal distribution with zero mean and identity covariance matrix, independent of  $\{(Q_j, P_j)\}_{j=0}^k$ . In the second step, we set the initial conditions  $(Q_k, G_{k+1})$  and compute the position and the momentum after  $T$  leapfrog steps. This move is accepted with probability  $\alpha_H\{(Q_k, G_{k+1}), \Phi_h^{\circ(T)}(Q_k, G_{k+1})\}$  where for all  $(q, p) \in \mathbb{R}^d \times \mathbb{R}^d$ ,  $(\tilde{q}, \tilde{p}) \in \mathbb{R}^d \times \mathbb{R}^d$ ,

$$(5) \quad \alpha_H\{(q, p), (\tilde{q}, \tilde{p})\} = \min[1, \exp(H(q, p) - H(\tilde{q}, \tilde{p}))].$$

It may be shown that  $\tilde{\pi}$  is invariant (see (2)) with respect to the Markov kernel defined by the HMC algorithm on the extended state space  $\mathbb{R}^d \times \mathbb{R}^d$ ; see [12]. Hence,  $\pi$  is a stationary distribution for the Markov chain  $(Q_k)_{k \geq 0}$ , which is the process which we are interested in. The number of steps  $T$  is either a deterministic quantity or a random variable independent of the current state. If the number of steps  $T = 1$ , then the algorithm reduces to the Metropolis Adjusted Langevin Algorithm (MALA) [25].

Recently, the theory on HMC have been addressed by many authors; see [1, 2, 6, 28, 29] and in depth discussions of the HMC methodology can be found in [1, 5, 23]. This

paper addresses two important issues in the analysis of HMC algorithm: irreducibility and geometric ergodicity.

Irreducibility plays an essential role in the theory of Markov chains. In particular, it implies that the invariant distribution, when it exists, is unique. The classical approach to derive irreducibility of Hastings–Metropolis algorithms on  $\mathbb{R}^d$  outlined, for example, in [20] [26], is to use that the proposal distribution admits a (sufficiently regular) transition density with respect to the Lebesgue measure. For HMC, this condition does not necessarily hold. HMC has been shown to be irreducible in [7] in the case where the state space is compact and the potential is twice continuously differentiable. In [2], under appropriate conditions, irreducibility is shown for a version of HMC where the number of leapfrog steps  $T$  is random, independent of the proposal, and such that  $T = 1$  with positive probability. Under such assumption, irreducibility of HMC boils down to irreducibility of MALA which has been established in [25]. In this paper, we establish the irreducibility of the HMC algorithm under a general tail condition of the target density which significantly relaxes the assumptions of [7] and [2]. This result follows from a general irreducibility result for iterative Markov models which we believe to be of independent interest; see the [Appendix](#). Our main tool to establish irreducibility is the degree theory for continuous maps [24].

In a second part, we establish the geometric ergodicity of the HMC sampler under the assumptions that the potential  $U$  is homogeneous outside a ball or is a perturbation of a homogeneous potential. Our assumptions imply that the proposal kernel of HMC satisfies an “inwards acceptance” property [25]. Our results complement the recent paper [2] which provides a variety of conditions under which the HMC algorithm is not geometrically ergodic.

In [4], a variant of HMC, referred to as the Randomized Hamiltonian Monte Carlo (RHMC), is analyzed. This method is associated with a continuous-time Markov process for which  $\tilde{\pi}$  given by (2) is invariant [4], Proposition 3.1. However, sampling such a process requires the exact Hamiltonian flow which allows to bypass the acceptance-rejection step and makes the analysis easier. Bypassing the discretization step nevertheless reduces the applicability of the results, since direct integration of the Hamiltonian flow is most of the time not an option. We discuss a simple example showing that the conditions in [4] upon which RHMC is geometrically ergodic are not sufficient in the case of HMC.

The paper is organized as follows. In Section 2, conditions upon which the HMC kernel, associated with  $(Q_k)_{k \in \mathbb{N}}$ , is irreducible, recurrent and Harris-recurrent are given. In Section 3, conditions under which the HMC kernel is  $V$ -uniformly geometrically ergodic are developed and discussed. The proofs of the main results of Section 2 are gathered in Section 4. Note that these proofs rely on technical results established in the Supplementary Material [11], Section S1. Some general irreducibility results which are of independent interest, are stated in the [Appendix](#). Section S2 of the Supplementary Material contains the proof for the statements of Section 3. Finally, our results are illustrated through several examples in [11], Section S4.

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

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

If  $V \equiv 1$ , then  $\|\mu - \nu\|_V$  is the total variation distance denoted by  $\|\mu - \nu\|_{TV}$ . 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  $x \in \mathbb{R}^d, y \in \mathbb{R}^\ell, x \otimes y \in (\mathbb{R}^d)^{\otimes 2}$  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  $x_1, \dots, x_k \in \mathbb{R}^d$ , set  $\|x_1 \otimes \dots \otimes x_k\| = \sup_{i \in \{1, \dots, k\}} \|x_i\|$ . We let  $\mathcal{L}((\mathbb{R}^d)^{\otimes k}, \mathbb{R}^\ell)$  stand for the set of linear maps from  $(\mathbb{R}^d)^{\otimes k}$  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  $f : \mathbb{R}^d \rightarrow \mathbb{R}^\ell$  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  $\|f\|_{Lip} = \inf\{\|f(x) - f(y)\|/\|x - y\|, x, y \in \mathbb{R}^d, x \neq y\}$ . 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  $D^k \Phi : U \rightarrow \mathcal{L}((\mathbb{R}^d)^{\otimes k}, \mathbb{R}^\ell)$  for the  $k$ th differential of  $\Phi \in C^k(\mathbb{R}^d, \mathbb{R}^\ell)$ . For smooth enough functions  $f : \mathbb{R}^d \rightarrow \mathbb{R}$ , denote by  $\nabla f$  and  $\nabla^2 f$  the gradient and the Hessian of  $f$ , respectively. Let  $A \subset \mathbb{R}^d$ . We write  $\bar{A}, A^\circ$  and  $\partial A$  for the closure, the interior and the boundary of  $A$ , respectively. For any  $n_1, n_2 \in \mathbb{N}, n_1 > n_2$ , we take the convention that  $\sum_{k=n_2}^{n_1} = 0$ .

**2. Ergodicity of the HMC algorithm.** For  $h > 0$  and  $T \in \mathbb{N}^*$ , consider the Markov kernel  $P_{h,T}$  associated with the Markov chain of the HMC algorithm  $(Q_k)_{k \in \mathbb{N}}$ , given for all  $q \in \mathbb{R}^d$  and  $A \in \mathcal{B}(\mathbb{R}^d)$  by

$$(6) \quad P_{h,T}(q, A) = \int_{\mathbb{R}^d} \mathbb{1}_A(\tilde{\Phi}_h^{\circ(T)}(q, \tilde{p})) \alpha_H\{(q, \tilde{p}), \Phi_h^{\circ(T)}(q, \tilde{p})\} \frac{e^{-\|\tilde{p}\|^2/2}}{(2\pi)^{d/2}} d\tilde{p} + \delta_q(A) \int_{\mathbb{R}^d} [1 - \alpha_H\{(q, \tilde{p}), \Phi_h^{\circ(T)}(q, \tilde{p})\}] \frac{e^{-\|\tilde{p}\|^2/2}}{(2\pi)^{d/2}} d\tilde{p},$$

where  $\tilde{\Phi}_h^{\circ(T)}, \Phi_h^{\circ(T)}$  and  $\alpha_H$  are defined by (3)–(4) and (5), respectively. In this section, we establish conditions upon which the Markov kernel  $P_{h,T}$  is irreducible or (Harris) recurrent. For  $\beta \in [0, 1]$ , we consider the following assumption on the potential  $U$ .

H 1 ( $\beta$ ).  $U$  is continuously differentiable and:

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

$$\|\nabla U(q) - \nabla U(x)\| \leq L_1 \|q - x\|.$$

(ii) there exists  $M_1 \geq 0$  such that for all  $q \in \mathbb{R}^d, \|\nabla U(q)\| \leq M_1 \{1 + \|q\|^\beta\}$ .

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  $n$  be an integer and  $\mu$  be a non-trivial measure on  $\mathcal{B}(\mathbb{R}^d)$ . A set  $C \in \mathcal{B}(\mathbb{R}^d)$  is called a  $(n, \mu)$ -small set for  $P$  if for all  $x \in C$  and  $A \in \mathcal{B}(\mathbb{R}^d), P^n(x, A) \geq \mu(A)$ . A set  $A \in \mathcal{B}(\mathbb{R}^d)$  is said to be accessible for  $P$  if for all  $x \in \mathbb{R}^d, \sum_{i=1}^\infty P^i(x, A) > 0$ . A nontrivial  $\sigma$ -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 [21], our notion of irreducibility is referred to as  $\phi$ -irreducibility, where  $\phi$  is an irreducibility measure; here irreducibility therefore means  $\phi$ -irreducibility).  $P$  is said to be a T-kernel if there exists a kernel  $T$  on  $\mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d)$  and a sequence of nonnegative numbers  $(a_i)_{i \in \mathbb{N}^*}$  satisfying  $\sum_{i=1}^{+\infty} a_i = 1$ , such that (i) for any  $x \in \mathbb{R}^d, T(x, \mathbb{R}^d) > 0$ ; (ii) for any  $A \in \mathcal{B}(\mathbb{R}^d), x \mapsto T(x, A)$  is lower semicontinuous; (iii) for any  $x \in \mathbb{R}^d, A \in \mathcal{B}(\mathbb{R}^d), \sum_{i=1}^{+\infty} a_i P^i(x, A) \geq T(x, A)$ .  $T$  is referred to as a continuous component of  $P$ .

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  $N_A = \sum_{i=0}^{+\infty} \mathbb{1}_A(X_i)$  is the number of visits to  $A$ . The set  $A$  is Harris recurrent if for any  $x \in A$ ,  $\mathbb{P}_x(N_A = +\infty) = 1$ . The Markov kernel  $P$  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  $A$ ,  $\mathbb{P}_x(N_A = +\infty) = 1$ .

Define  $\vartheta_1 : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ , for any  $s \in \mathbb{R}_+$  by

$$(7) \quad \vartheta_1(s) = 1 + s/2 + s^2/4.$$

We consider below values of the stepsize  $h$  and the number of iterations satisfying

$$(8) \quad [\{1 + hL_1^{1/2}\vartheta_1(hL_1^{1/2})\}^T - 1] < 1.$$

For all  $h > 0$  and  $T \in \mathbb{N}^*$ , we have

$$\{1 + hL_1^{1/2}\vartheta_1(hL_1^{1/2})\}^T - 1 \leq e^{hL_1^{1/2}T\vartheta_1(hL_1^{1/2}T)} - 1$$

using that  $\vartheta_1$  is nondecreasing. Then, setting  $\bar{S} = cL_1^{-1/2}$  where  $c = 0.521$  is the unique positive root of the equation  $c\vartheta_1(c) = \log(2)$ , all  $T \in \mathbb{N}^*$  and  $h \in (0, \bar{S}/T)$  satisfy (8). Note that conversely, if  $h > 0$  and  $T \in \mathbb{N}^*$  satisfies (8), necessarily  $h \in (0, L_1^{-1/2})$  because for any  $s > 0$ ,  $\vartheta_1(s) \geq 1$ . In addition, since  $e^{\log(2)s} \leq (1 + s)$  for all  $s \in (0, 1)$ ,  $T$  and  $h$  satisfy  $hT \leq \bar{S} = L_1^{-1/2}$ .

**THEOREM 1.** *Assume H 1( $\beta$ ) for some  $\beta \in [0, 1]$  and that  $U$  is twice continuously differentiable. Then, for all  $T \in \mathbb{N}^*$ , and  $h > 0$  satisfying (8) and  $q \in \mathbb{R}^d$ , there exists a  $C^1(\mathbb{R}^d, \mathbb{R}^d)$ -diffeomorphism  $\tilde{q} \mapsto \tilde{\Psi}_h^{(T)}(q, \tilde{q})$  such that for any  $p \in \mathbb{R}^d$ ,*

$$(9) \quad \text{if } q_T = \tilde{\Phi}_h^{\circ(T)}(q, p) \quad \text{then } p = \tilde{\Psi}_h^{(T)}(q, q_T).$$

Moreover,

(i) *The Markov kernel  $P_{h,T}$ , is a T-kernel; more precisely, for any  $B \in \mathcal{B}(\mathbb{R}^d)$ ,*

$$(10) \quad P_{h,T}(q, B) = T_{h,T}(q, B) + \delta_q(B)(2\pi)^{-d/2} \int_{\mathbb{R}^d} [1 - \alpha_H\{(q, \tilde{p}), \Phi_h^{\circ(T)}(q, \tilde{p})\}] e^{-\|\tilde{p}\|^2/2} d\tilde{p},$$

where the kernel  $T_{h,T}$  is a continuous component of  $P_{h,T}$  and is given by

$$(11) \quad T_{h,T}(q, B) = (2\pi)^{-d/2} \int_B \bar{\alpha}_H(q, \tilde{q}) e^{-\|\tilde{\Psi}_h^{(T)}(q, \tilde{q})\|^2/2} D_{\tilde{\Psi}_h^{(T)}(q, \cdot)}(\tilde{q}) d\tilde{q},$$

setting for  $q, \tilde{q} \in \mathbb{R}^d$ ,  $\bar{\alpha}_H(q, \tilde{q}) = \alpha_H\{(q, \tilde{\Psi}_h^{(T)}(q, \tilde{q})), \Phi_h^{\circ(T)}(q, \tilde{\Psi}_h^{(T)}(q, \tilde{q}))\}$  and  $D_{\tilde{\Psi}_h^{(T)}(q, \cdot)}(\tilde{q}) = |\det(\mathbf{J}_{\tilde{\Psi}_h^{(T)}(q, \cdot)}(\tilde{q}))|$ .

(ii) *The Markov kernel  $P_{h,T}$  is irreducible and the Lebesgue measure is an irreducibility measure. Moreover,  $P_{h,T}$  is aperiodic, Harris recurrent and all the compact sets are 1-small. Therefore, for all  $q \in \mathbb{R}^d$ ,*

$$(12) \quad \lim_{n \rightarrow +\infty} \|\delta_q P_{h,T}^n - \pi\|_{TV} = 0.$$

**PROOF.** The proof is postponed to Section 4.1.  $\square$

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  $h$  and the number of iterations  $T$ . The result is less quantitative and the proof is more involved: we use degree theory for continuous mapping (the main notions required in the proof are recalled in the [Appendix](#)).

**THEOREM 2.** *Let  $h > 0$  and  $T \in \mathbb{N}^*$  and assume either:*

- (a) **H 1** ( $\beta$ ) for some  $\beta \in [0, 1)$ ,
- (b) **H 1** (1) and that  $T \in \mathbb{N}^*$  and  $h > 0$  satisfy (8).

*Then:*

- (i) *the HMC kernel  $P_{h,T}$  defined by (6) is irreducible, aperiodic, the Lebesgue measure is an irreducibility measure and any compact set of  $\mathbb{R}^d$  is small.*
- (ii)  *$P_{h,T}$  is recurrent and for  $\pi$ -almost every  $q \in \mathbb{R}^d$ ,  $\lim_{n \rightarrow +\infty} \|\delta_q P_{h,T}^n - \pi\|_{TV} = 0$ .*

**PROOF.** The proof is postponed to Section 4.2.  $\square$

To the best of the author’s knowledge, the first results regarding the irreducibility of the HMC algorithm are established in [7] under the assumption that  $U$  and  $\|\nabla U\|$  are bounded above (in [7] the state space of is a  $d$ -dimensional torus). Irreducibility has also been tackled in [2]: 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, [2] shows that the kernel associated with the HMC algorithm is irreducible. Under this condition, the proof 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 [25]. Finally, [4], Proposition 3.7, shows that RHMC is irreducible under the condition that  $U$  is at least quadratic. Note that Theorem 2 establishes irreducibility of HMC of subquadratic potential. However, leap-frog integrator is not numerically stable for lighter than Gaussian target density, therefore, other kind of integrators should be used instead; see, for example, [14], Chapter VI.

Finally, 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  $(\varpi_i)_{i \in \mathbb{N}^*}$  be a probability distribution on  $\mathbb{N}^*$  and  $(h_i)_{i \in \mathbb{N}^*}$  be a sequence of positive real numbers. Define the randomized Hamiltonian kernel  $\bar{P}_{\mathbf{h}, \boldsymbol{\varpi}}$  on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  associated with  $(\varpi_i)_{i \in \mathbb{N}^*}$  and  $(h_i)_{i \in \mathbb{N}^*}$  by

$$(13) \quad \bar{P}_{\mathbf{h}, \boldsymbol{\varpi}} = \sum_{i \in \mathbb{N}^*} \varpi_i P_{h_i, i}.$$

We denote by  $\text{supp}(\boldsymbol{\varpi}) = \{i \in \mathbb{N}^* : \varpi_i \neq 0\}$  the support of the distribution  $\boldsymbol{\varpi}$ .

**COROLLARY 3.** *Let  $\beta \in [0, 1]$  and assume **H 1**( $\beta$ ). Let  $(\varpi_i)_{i \in \mathbb{N}^*}$  be a probability distribution on  $\mathbb{N}^*$ ,  $(h_i)_{i \in \mathbb{N}^*}$  be a sequence of positive real numbers, and  $\bar{P}_{\mathbf{h}, \boldsymbol{\varpi}}$  be the randomized Hamiltonian kernel associated with  $(\varpi_i)_{i \in \mathbb{N}^*}$  and  $(h_i)_{i \in \mathbb{N}^*}$ .*

(a) *Assume that  $U$  is twice continuously and there exists  $i \in \mathbb{N}^*$  such that  $\{[1 + h_i L_1^{1/2} \vartheta_1(h_i L_1^{1/2})]^i - 1\} < 1$  and  $\varpi_i > 0$  where  $\vartheta_1$  is given by (7). Then the conclusions of Theorem 1(ii) hold for  $\bar{P}_{\mathbf{h}, \boldsymbol{\varpi}}$ .*

(b) *If  $\beta \in [0, 1)$ , then the conclusions of Theorem 2(a) hold for  $\bar{P}_{\mathbf{h}, \boldsymbol{\varpi}}$ .*

(c) *If  $\beta = 1$  and there exists  $i \in \text{supp}(\boldsymbol{\varpi})$  such that  $\{[1 + h_i L_1^{1/2} \vartheta_1(h_i L_1^{1/2})]^i - 1\} < 1$ , then the conclusions of Theorem 2(b) hold for  $\bar{P}_{\mathbf{h}, \boldsymbol{\varpi}}$ .*

**PROOF.** (a) follows from Theorem 1 and Proposition S11. (b) and (c) are straightforward applications of Theorem 2.  $\square$

### 3. Geometric ergodicity of HMC.

3.1. *Main results.* In this section, we give conditions on the potential  $U$  which imply that the HMC kernel (4) converges geometrically fast to its invariant distribution. Let  $V : \mathbb{R}^d \rightarrow [1, +\infty)$  be a measurable function and  $P$  be a Markov kernel on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ . The Markov kernel  $P$  is said to be  $V$ -uniformly geometrically ergodic if  $P$  admits an invariant probability  $\pi$  and there exists  $\rho \in [0, 1)$  and  $\zeta \geq 0$  such that for all  $q \in \mathbb{R}^d$  and  $k \in \mathbb{N}^*$ ,

$$\|P^k(q, \cdot) - \pi\|_V \leq \zeta \rho^k V(q).$$

By [21], Theorem 16.0.1, if  $P$  is aperiodic, irreducible and satisfies a Foster–Lyapunov drift condition, that is, there exists a small set  $C$  for  $P$ ,  $\lambda \in [0, 1)$  and  $b < +\infty$  such that for all  $q \in \mathbb{R}^d$ ,

$$(14) \quad PV \leq \lambda V + b\mathbb{1}_C,$$

then  $P$  is  $V$ -uniformly geometrically ergodic. If a function  $V : \mathbb{R}^d \rightarrow [1, \infty)$  satisfies (14), 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 kernels which includes HMC kernels as a particular example.

Let  $K$  be a proposal kernel on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^{2d}))$  and  $\alpha : \mathbb{R}^{3d} \rightarrow [0, 1]$  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

$$(15) \quad P(q, A) = \int_{\mathbb{R}^{2d}} \mathbb{1}_A(\text{proj}(z))\alpha(q, z)K(q, dz) + \delta_q(A) \int_{\mathbb{R}^{2d}} \{1 - \alpha(q, z)\}K(q, dz),$$

where  $\text{proj} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  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  $B \in \mathcal{B}(\mathbb{R}^{2d})$ , respectively, by

$$(16) \quad K_{h,T}(q, B) = (2\pi)^{-d/2} \int_{\mathbb{R}^d} \mathbb{1}_B(\tilde{\Phi}_h^{\circ(T)}(q, \tilde{p}), \tilde{p})e^{-\|\tilde{p}\|^2/2} d\tilde{p},$$

$$(17) \quad \tilde{\alpha}_H(q, (\tilde{q}, \tilde{p})) = \begin{cases} \alpha_H\{(q, \tilde{p}), \Phi_h^{\circ(T)}(q, \tilde{p})\}, & \text{if } \tilde{q} = \tilde{\Phi}_h^{\circ(T)}(q, \tilde{p}), \\ 0 & \text{otherwise,} \end{cases}$$

where  $\Phi_h^{\circ(T)}$ ,  $\tilde{\Phi}_h^{\circ(T)}$  and  $\alpha_H$  are defined in (3), (4) and (5), respectively. Let  $V : \mathbb{R}^d \rightarrow [1, +\infty)$  be a *norm-like* function, that is, a measurable function such that for all  $M \in \mathbb{R}_+$ , the level sets  $\{q \in \mathbb{R}^d : V(q) \leq M\}$  are compact. Note that if  $V$  is norm-like, for any  $M \in \mathbb{R}_+$ ,  $\{q \in \mathbb{R}^d : V(q) \leq M\}^c$  is non-empty. The 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:

$$(18) \quad \mathcal{R}(q) = \{z \in \mathbb{R}^{2d}, \alpha(q, z) < 1\}, \quad \mathcal{B}(q) = \{z \in \mathbb{R}^{2d}, V(\text{proj}(z)) \leq V(q)\}.$$

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 [25], Theorem 4.1, which is used to show the  $V$ -uniform geometric ergodicity of the MALA algorithm.

PROPOSITION 4. *Let  $V : \mathbb{R}^d \rightarrow [1, +\infty)$  be a norm-like function. Assume moreover that there exist  $\lambda \in [0, 1)$  and  $b \in \mathbb{R}_+$  such that*

$$(19) \quad KV \leq \lambda V + b$$

and

$$(20) \quad \lim_{M \rightarrow +\infty} \sup_{\{q \in \mathbb{R}^d : V(q) \geq M\}} \mathbf{K}(q, \mathcal{R}(q) \cap \mathcal{B}(q)) = 0.$$

Then there exist  $\tilde{\lambda} \in [0, 1)$  and  $\tilde{b} \in \mathbb{R}_+$  such that  $\mathbf{P}V \leq \tilde{\lambda}V + \tilde{b}$  where  $\mathbf{P}$  is given by (15).

PROOF. The proof is postponed to Section S2.1.  $\square$

We show below that under appropriate conditions, the proposal kernel  $\mathbf{K}_{h,T}$  and the acceptance probability  $\tilde{\alpha}_H$  given by (16) and (17) satisfy the conditions of Proposition 4 which imply that the HMC kernel  $\mathbf{P}_{h,T}$  is  $V$ -uniformly geometrically ergodic. For  $m \in (1, 2]$ , consider the following assumption:

H 2 ( $m$ ). There exist  $A_1 \in \mathbb{R}_+^*$  and  $A_2 \in \mathbb{R}$  such that for all  $q \in \mathbb{R}^d$ ,

$$\langle \nabla U(q), q \rangle \geq A_1 \|q\|^m - A_2.$$

For all  $a \in \mathbb{R}_+^*$  and  $q \in \mathbb{R}^d$ , define

$$(21) \quad V_a(q) = \exp(a\|q\|).$$

For  $h \geq 0$ , define

$$(22) \quad \vartheta_2(h) = M_1/L_1^{1/2} + M_1h/2 + L_1^{1/2}M_1h^2/4.$$

PROPOSITION 5.

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

$$(23) \quad \mathbf{K}_{h,T}V_a \leq \lambda V_a + b.$$

(b) Assume H 1(1) and H 2(2). Let  $\tilde{S} > 0$  be such that  $\Theta(\tilde{S}) = A_1$  where the function  $\Theta$  is given by

$$(24) \quad \begin{aligned} \Theta(s) = & 2L_1^{1/2}\vartheta_2(s)\{e^{L_1^{1/2}s\vartheta_1(L_1^{1/2}s)} - 1\} \\ & + 6s^2(M_1^2 + L_1\vartheta_2^2(s)\{e^{L_1^{1/2}s\vartheta_1(L_1^{1/2}s)} - 1\}^2), \end{aligned}$$

with  $\vartheta_1$  and  $\vartheta_2$  defined by (7) and (22), respectively. Let  $\tilde{S} \in (0, \tilde{S})$ . Then, for all  $a \in \mathbb{R}_+^*$ ,  $T \in \mathbb{N}^*$  and  $h \in (0, \tilde{S}/T]$ , there exist  $\lambda \in [0, 1)$  and  $b \in \mathbb{R}_+$  which satisfy (23).

PROOF. The proof is postponed to Section S2.2.  $\square$

We now derive sufficient conditions under which the condition (20) of Proposition 4 is satisfied.

H 3 ( $m$ ). There exist  $F, G : \mathbb{R}^d \rightarrow \mathbb{R}$  such that  $U = F + G$  and satisfying:

(i)  $F \in C^3(\mathbb{R}^d)$  and there exists  $A_3 \in \mathbb{R}_+^*$  such that for all  $q \in \mathbb{R}^d$  and  $k = 2, 3$ ,  $\|D^k F(q)\| \leq A_3\{1 + \|q\|\}^{m-k}$ .

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

$$D^2 F(q)\{\nabla F(q) \otimes \nabla F(q)\} \geq A_4 \|q\|^{3m-4}.$$

(iii)  $G \in C^1(\mathbb{R}^d)$  and there exist  $A_5 \in \mathbb{R}_+^*$  and  $\varrho \in [1, 2(m - 1))$  such that for any  $q, x \in \mathbb{R}^d$ ,

$$\begin{aligned} |G(q)| &\leq A_5(1 + \|q\|)^\varrho, & \|\nabla G(q)\| &\leq A_5(1 + \|q\|)^{\varrho-1}, \\ \|\nabla G(q) - \nabla G(x)\| &\leq A_5 \|q - x\|. \end{aligned}$$

It is easily checked that under H 3, the results of Section 2 can be applied, that is,  $\nabla U$  satisfies H 1( $m - 1$ ); see Lemma S5.

Condition H 2( $m$ ) and H 3( $m$ ) are 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 perturbations of such functions.

We say that a function  $F_1 : \mathbb{R}^d \rightarrow \mathbb{R}$  is  $m$ -homogeneous quasi-convex outside a ball of radius  $R_1$  if the following conditions are satisfied:

- (QC-1) for all  $t \geq 1$  and  $q \in \mathbb{R}^d, \|q\| \geq R_1, F_1(tq) = t^m F_1(q)$ .
- (QC-2) for all  $q \in \mathbb{R}^d, \|q\| \geq R_1$ , the level sets  $\{x : F_1(x) \leq F_1(q)\}$  are convex.

**PROPOSITION 6.** *Let  $m \in [1, 2]$  and  $R_1 \in \mathbb{R}_+$ . Assume that the potential  $U$  may be decomposed as  $U(q) = F_1(q) + F_2(q) + G(q)$ , for any  $q \in \mathbb{R}^d, \|q\| \geq R_1$ , where the functions  $F_1, F_2, G \in C^3(\mathbb{R}^d)$  satisfy the following two conditions:*

- (A)  $F_1$  is  $m$ -homogeneously quasiconvex outside a ball of radius  $R_1$  and  $\lim_{\|q\| \rightarrow +\infty} F_1(q) = \infty$ .
- (B) For  $k = 2, 3, \lim_{\|q\| \rightarrow +\infty} \|D^k F_2(q)\|/\|q\|^{m-k} = 0$ .
- (C)  $G$  satisfies H 3(iii).

Then  $U$  satisfies H 2( $m$ ) and H 3( $m$ ).

**PROOF.** The proof is postponed to Section S2.3.  $\square$

To show that the condition (20) of Proposition 4 is satisfied under H 3( $m$ ), we rely on the following important result which implies that the probability of accepting a move goes to 1 as  $\|q\| \rightarrow \infty$ .

**PROPOSITION 7.** *Assume H 3( $m$ ) for some  $m \in (1, 2]$ . Let  $\gamma \in (0, m - 1)$ .*

- (a) *If  $m \in (1, 2)$ , for all  $T \in \mathbb{N}^*, h \in \mathbb{R}_+^*$ , 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 exists  $\bar{S} > 0$  such that for any  $T \in \mathbb{N}^*$  and  $h \in (0, \bar{S}/T^{3/2}]$ , there exists  $R_H \in \mathbb{R}_+$  satisfying for all  $q_0, p_0 \in \mathbb{R}^d, \|q_0\| \geq R_H$  and  $\|p_0\| \leq \|q_0\|^\gamma, H(\Phi_h^{\circ(T)}(q_0, p_0)) - H(q_0, p_0) \leq 0$ .*

**PROOF.** The proof is postponed to Section S2.4.  $\square$

This result means that far in the tail the HMC proposal are ‘‘inward.’’ We illustrate the result of Proposition 7(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 assumptions of Proposition 7. We can observe that choosing the different initial conditions  $q_0$  with increasing norm imply that  $\tilde{T} = \max\{k \in \mathbb{N}; H(\Phi_h^{\circ(k)}(q_0, p_0)) - H(q_0, p_0) < 0\}$  increases as well.

However, in the case  $m = 2$ , Proposition 7(b) only implies that the HMC proposal is inward only if the step size  $h$  is sufficiently small with respect to the number of leapfrog step  $T$ , that is, is of order  $\mathcal{O}(T^{-3/2})$ . To relax this condition, we strengthen H 3(2) by assuming that  $U$  is a smooth perturbation of a quadratic function.

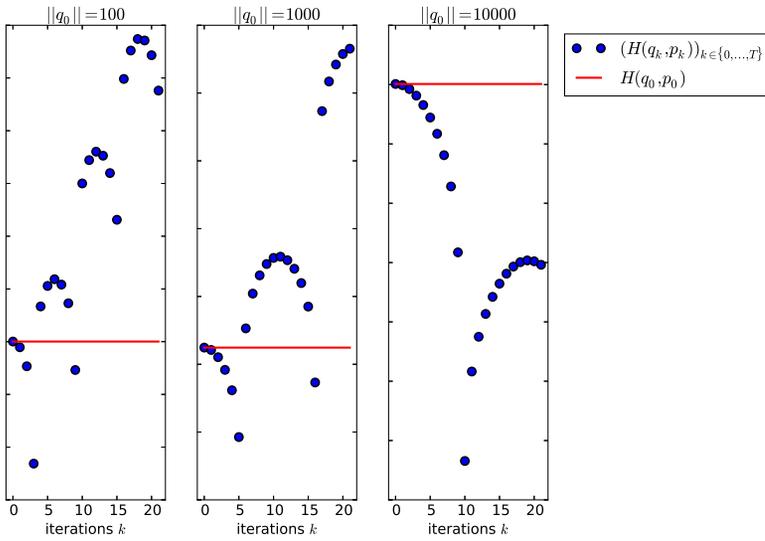


FIG. 1. Behavior of  $(H(\Phi_h^{o(k)}(q_0, p_0)))_{k \in \{0, \dots, T\}}$  for different initial conditions  $q_0$ .

H 4. There exist  $G : \mathbb{R}^d \rightarrow \mathbb{R}$ , continuously differentiable, and a positive definite matrix  $\Pi$  such that for any  $q \in \mathbb{R}^d$ ,  $U(q) = \langle \Pi q, q \rangle / 2 + G(q)$  and there exist  $A_5 \geq 0$  and  $\varrho \in [1, 2)$  such that for any  $q, x \in \mathbb{R}^d$ ,

$$(25) \quad \begin{aligned} |G(q)| &\leq A_5(1 + \|q\|)^\varrho, & \|\nabla G(q)\| &\leq A_5(1 + \|q\|)^{\varrho-1}, \\ \|\nabla G(q) - \nabla G(x)\| &\leq A_5 \|q - x\|. \end{aligned}$$

Note that it is straightforward to check that under H 4, the conditions H 1(1) and H 2(2) hold.

The following result shows that it is enough that the decomposition required in H 4 asymptotically holds.

PROPOSITION 8. Assume that there exist  $\Gamma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$  and  $G : \mathbb{R}^d \rightarrow \mathbb{R}$  continuously differentiable such that for any  $q \in \mathbb{R}^d$ ,  $U(q) = \langle \Gamma(q)q, q \rangle / 2 + G(q)$  with  $G$  satisfying (25) and there exist a positive definite matrix  $\Pi \in \mathbb{R}^{d \times d}$ ,  $C_\Gamma \geq 0$  and  $\epsilon_\Gamma > 0$  satisfying for any  $q, x \in \mathbb{R}^d$ ,

$$(26) \quad \|\Gamma(q) - \Pi\| \leq C_\Gamma(1 + \|q\|)^{-\epsilon_\Gamma}, \quad \|D\Gamma(q)\| \leq C_\Gamma(1 + \|q\|)^{-1-\epsilon_\Gamma}$$

$$(27) \quad \|D\Gamma(q) - D\Gamma(x)\| \leq C_\Gamma \|q - x\| / (1 + \|q\| \wedge \|x\|)^2.$$

Then  $U$  satisfies H 4.

PROOF. The proof is postponed to Section S2.5.  $\square$

PROPOSITION 9. Assume H 4 and let  $\gamma \in (0, 1)$ . There exists a constant  $\bar{S} > 0$  such that for all  $T \in \mathbb{N}^*$ ,  $h \in (0, \bar{S}/T]$ , 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  $\|p_0\| \leq \|q_0\|^\gamma$ ,  $H(\Phi_h^{o(T)}(q_0, p_0)) - H(q_0, p_0) \leq 0$ .

PROOF. The proof is postponed to Section S2.6.  $\square$

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

THEOREM 10.

(a) If H 2(m) and H 3(m) hold for some  $m \in (1, 2)$ , then for all  $a \in \mathbb{R}_+^*$ ,  $T \in \mathbb{N}^*$  and  $h > 0$ , the HMC kernel  $P_{h,T}$  is  $V_a$ -uniformly geometrically ergodic, where  $V_a$  is defined by (21).

(b) If H 2(2) and H 3(2) hold, then there exists  $\bar{S} > 0$  such that for all  $a \in \mathbb{R}_+^*$ ,  $T \in \mathbb{N}^*$  and  $h \in (0, \bar{S}/T^{3/2})$ ,  $P_{h,T}$  is  $V_a$ -uniformly geometrically ergodic.

(c) If H 4 holds, then there exists  $\bar{S} > 0$  (depending only on  $\Pi$  and  $A_5$ ) such that for all  $a \in \mathbb{R}_+^*$ ,  $T \in \mathbb{N}^*$  and  $h \in (0, \bar{S}/T)$ ,  $P_{h,T}$  is  $V_a$ -uniformly geometrically ergodic.

PROOF OF THEOREM 10. It is enough to consider (a) as the proof of (b) and (c) follows exactly the same lines taking  $\bar{S}$  small enough. Proposition 5 shows that for all  $T \in \mathbb{N}^*$ ,  $h \in \mathbb{R}_+^*$  and  $a \in \mathbb{R}_+^*$ , there exist  $\lambda \in [0, 1)$  and  $b \in \mathbb{R}_+$  such that the Foster–Lyapunov drift condition  $K_{h,T} V_a \leq \lambda V_a + b$  is satisfied. By Proposition 7, there exists  $R_H \geq 0$  such that for all  $q \in \mathbb{R}^d$ ,  $\|q\| \geq R_H$ ,

$$\int_{\mathcal{R}(q)} K_{h,T}(q, dz) \leq (2\pi)^{-d/2} \int_{\{\|p\| \geq \|q\|^\gamma\}} e^{-\|p\|^2/2} dp,$$

for  $\gamma \in (0, m - 1)$  where  $\mathcal{R}(q) = \{z \in \mathbb{R}^{2d} : \check{\alpha}_H(q, z) < 1\}$  (see (17)), which implies that

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

Since  $V_a$  is norm-like, Proposition 4 implies that for all  $T > 0$  and  $h > 0$ , there exists  $\check{\lambda}$  and  $\check{b}$  (depending upon  $a, h$  and  $T$ ) such that  $P_{h,T} V_a \leq \check{\lambda} V_a + \check{b}$ . For all  $M \geq 0$  the level sets  $\{V_a \leq M\}$  are compact, and hence small by Theorem 2. [8], Corollary 14.1.6, then shows that there exists a small set  $C$ ,  $\check{\lambda} \in [0, 1)$  and  $\check{b} \in [0, 1)$  such that  $P_{h,T} V_a \leq \check{\lambda} V_a + \check{b} \mathbb{1}_C$ . Since  $P_{h,T}$  is aperiodic, the result follows from [8], Theorem 15.2.4.  $\square$

We finally consider the case where the number of leapfrog steps is a random variable independent of the current state.

THEOREM 11.

(a) If H 2(m) and H 3(m) hold for  $m \in (1, 2)$ , then for all probability distributions  $\boldsymbol{\omega} = (\omega_i)_{i \in \mathbb{N}^*}$  on  $\mathbb{N}^*$ , all sequences  $\mathbf{h} = (h_i)_{i \in \mathbb{N}^*}$  of positive numbers, and  $a \in \mathbb{R}_+^*$ , the randomized kernel  $\bar{P}_{\mathbf{h}, \boldsymbol{\omega}}$  (13) is  $V_a$ -uniformly geometrically ergodic, where  $V_a$  is defined by (21).

(b) If H 2(2) and H 3(2) hold, then there exists  $\bar{S} > 0$  such that for all probability distributions  $\boldsymbol{\omega} = (\omega_i)_{i \in \mathbb{N}^*}$  on  $\mathbb{N}^*$ , all sequences  $\mathbf{h} = (h_i)_{i \in \mathbb{N}^*}$  satisfying  $\max_{i \in \text{supp}(\boldsymbol{\omega})} i^{3/2} h_i \leq \bar{S}$ , and  $a \in \mathbb{R}_+^*$ ,  $\bar{P}_{\mathbf{h}, \boldsymbol{\omega}}$  is  $V_a$ -uniformly geometrically ergodic.

(c) If H 4 holds, then there exists  $\bar{S} > 0$  (depending only on  $\Pi$  and  $A_5$ ) such that for all probability distributions  $\boldsymbol{\omega} = (\omega_i)_{i \in \mathbb{N}^*}$  on  $\mathbb{N}^*$ , all sequences  $\mathbf{h} = (h_i)_{i \in \mathbb{N}^*}$  satisfying  $\max_{i \in \text{supp}(\boldsymbol{\omega})} i h_i \leq \bar{S}$ , and  $a \in \mathbb{R}_+^*$ ,  $\bar{P}_{\mathbf{h}, \boldsymbol{\omega}}$  is  $V_a$ -uniformly geometrically ergodic.

PROOF. It is enough to consider (a) as the proofs of (b) and (c) are along the same lines. Set  $a \in \mathbb{R}_+^*$ . It is established in the proof of Theorem 10 that for all  $i \in \mathbb{N}^*$   $P_{i,h_i}$  satisfies a Foster–Lyapunov drift condition: there exists  $\check{\lambda}_i \in [0, 1)$  and  $\check{b}_i < \infty$  such that  $P_{i,h_i} V_a \leq \lambda_i V_a + b_i$ . By Corollary 3,  $\bar{P}_{\mathbf{h}, \boldsymbol{\omega}}$  is irreducible and aperiodic and all the compact sets are small. We conclude by applying [8], Theorem 15.2.4.  $\square$

3.2. *Comparison with the literature.* [2], Theorem 2.1, establishes geometric ergodicity of the HMC kernel but under an implicit assumption on the behavior of the acceptance rate (see [2], assumption (A3)). Our conditions are directly verifiable on the potential  $U$ .

[18, 19] mainly study different versions of the *unadjusted* versions of the HMC (based on the Verlet integrator but omitting the Metropolis–Hastings step). These articles provide quantitative results (as opposed to our results which are qualitative) with an explicit dependence on the dimension  $d$ , but under stringent assumptions on the target distribution, which is assumed to be strongly log-concave.

Our conditions are different from those given by [4] to establish the geometric ergodicity of the idealized randomized HMC, for which the Hamiltonian flow (1) is assumed to be known. In such cases, the proposals are always accepted, which considerably simplifies the proof: by far the most difficult part of the evidence is indeed to show that the acceleration rate tends towards 1. The conditions in [4] are as follows:

$$\begin{aligned}
 & \text{(i) } \int_{\mathbb{R}^d} \|q\|^2 d\pi(q) < +\infty, \\
 & \text{(ii) there exist } C_1 \in (0, 1) \text{ and } C_2 > 0 \text{ such that for all } q \in \mathbb{R}^d \\
 (28) \quad & (1/2)\langle \nabla U(q), q \rangle \geq C_1 U(q) + \frac{(\tau^{-1}C_1/4)^2 + \tau^{-2}C_1(1 - C_1)/4}{2(1 - C_1)} \|q\|^2 - C_2,
 \end{aligned}$$

where  $\tau > 0$  is the duration parameter of the RHMC algorithm.

Note that these conditions assumed that the tails of the target density are lighter than those of a Gaussian. In comparison, our results can be applied to subquadratic potentials. In addition, it can be shown that HMC is not geometrically ergodic under (28): a counterexample is given below.

The main difference with the setting of [4] is that HMC has a acceptance/rejection step and the integrated acceptance ratio

$$q \mapsto \int_{\mathbb{R}^d} \alpha_H\{(q, p), \Phi_h^{\circ(T)}(q, p)\} e^{-\|p\|^2/2} (2\pi)^{-d/2} dp$$

must not go to 0 as  $\|q\|$  goes to  $+\infty$ . Assumptions H 3 and H 4 are required to control the integrated acceptance ratio. Indeed, [26], Theorem 5.1, shows that a an irreducible Markov kernel  $P$  on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  is not geometrically ergodic with respect to an invariant measure  $\mu$  if

$$(29) \quad \inf\{\delta > 0 : \mu(\{q \in \mathbb{R}^d : P(q, \{q\}) \geq \delta\}) = 0\} = 1.$$

Consider the target density  $\pi$  with potential  $U$  given for all  $q = (q_1, q_2) \in \mathbb{R}^2$  by

$$(30) \quad U(q) = -\log(e^{-q_1^2 - 5q_2^2} + e^{-5q_1^2 - q_2^2}).$$

Note that  $U$  satisfies the condition (28). On the contrary, we may show that (29) holds and, therefore, HMC is not geometrically ergodic for such a potential  $U$ . However, the detailed calculations are very technical and not particularly informative and we prefer to present a numerical evidence that (29) holds. Indeed, Figure 2 displays numerical computations of the mean acceptance ratio,  $\int_{\mathbb{R}^2} \alpha_H\{(q, p), \Phi_h^{\circ(T)}(q, p)\} e^{-\|p\|^2/2} (2\pi)^{-1} dp = 1 - P_{h,T}(q, \{q\})$  for  $q_1 \in \{200 + j50, j = 0, \dots, 6\}$ ,  $q_2 \in [q_1 + 10^{-4}, q_1 + 2 \cdot 10^{-4}]$  and  $T = 1$  which corresponds to MALA. We can observe that the larger  $q_1$ , the smaller  $1 - P_{h,T}(q, \{q\})$ , which illustrates that (29) holds for the HMC kernel.

However, our result can be applied to  $d$ -dimensional Gaussian mixtures with a dominating precision matrix. Consider potentials given for any  $q \in \mathbb{R}^d$  by

$$U(q) = -\log \left\{ \sum_{i=1}^N w_i \exp(-\langle \Pi_i(q - q_i), q - q_i \rangle / 2) \right\},$$

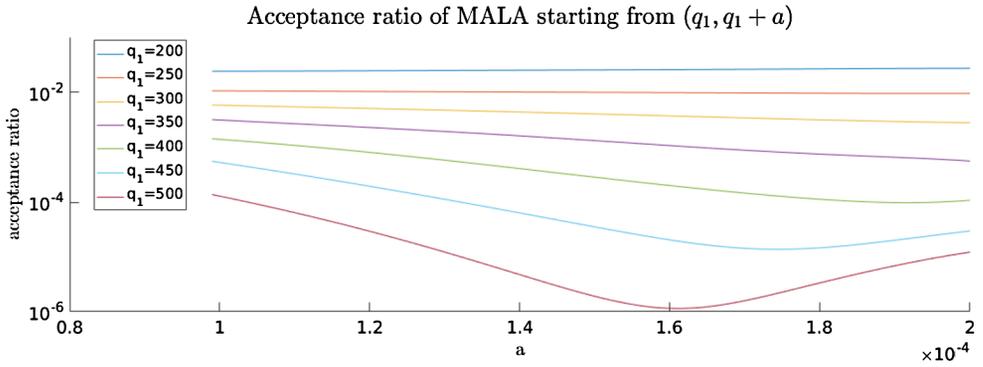


FIG. 2. Acceptance ratio for MALA applied to (30).

where for any  $i \in \{1, \dots, N\}$ ,  $w_i > 0$  and  $\sum_{i=1}^N w_i = 1$ ,  $q_i \in \mathbb{R}^d$ ,  $\Pi_i$  is a positive definite matrix. Assume that  $\Pi_i - \Pi_1$  is positive definite for  $i \in \{2, \dots, N\}$ . Indeed  $q \mapsto \langle \Pi_1(q - q_1), q - q_1 \rangle / 2$  and its first- and second-order partial derivatives are bounded functions and, therefore, H 4 is satisfied which shows that the conclusions of Theorem 10c hold. Note that the Gaussian mixture we consider in (30) has no dominating precision matrix.

Bou-Rabee et al. [3] is more closely related to our work and studies the same (metropolized) HMC with a Verlet integrator. Bou-Rabee et al. [3] evaluates the mixing time (with respect to a carefully designed Kantorovitch distance) for a target distributions which are log-concave outside a compact set. Two types of results are established. Reference [3], Theorems 2.4, 2.7, establish a contraction for the HMC kernel for smooth potential function  $U$  ( $U$  should be 4 times continuously differentiable with bounded second, third and fourth differential) and additional conditions on the stepsize and the number of integration steps. Moreover, this result holds only if the initial points are in a compact set. The second result [3], Theorem 2.12, establishes explicit complexity bounds in a specially crafted Kantorovitch distance of order 1, that is, given a precision parameter  $\epsilon > 0$ , [3], Theorem 2.12, gives a number of iterations  $n$  which is sufficient to ensure that the Kantorovitch distance of order 1 between the  $n$ th iterate of HMC and the target distribution  $\pi$  is smaller than  $\epsilon$ . Note that the discretization step and the number of iterates are functions of the total number of samples: therefore, the nature of this result is different from ours. Furthermore, compared to [3] we establish convergence in total variation distance or  $V$ -norm and not in a Kantorovitch distance.

**4. Proofs of Section 2.** Note that a simple induction (see [2], Proposition 4.2) implies that for all  $(q_0, p_0) \in \mathbb{R}^d \times \mathbb{R}^d$  and  $k \in \{1, \dots, T\}$ , the  $k$ th iteration of the leapfrog integration,  $(q_k, p_k) = \Phi_h^{\circ(k)}(q, p)$ , where  $\Phi_h^{\circ(k)}$  is defined by (3), takes the form

$$\begin{aligned}
 (31) \quad & q_k = q_0 + kh p_0 - \frac{kh^2}{2} \nabla U(q_0) - h^2 \mathfrak{E}_{h,k}(q_0, p_0) \\
 & p_k = p_0 - \frac{h}{2} \{ \nabla U(q_0) + \nabla U \circ \tilde{\Phi}_h^{\circ(k)}(q_0, p_0) \} - h \sum_{i=1}^{k-1} \nabla U \circ \tilde{\Phi}_h^{\circ(i)}(q_0, p_0),
 \end{aligned}$$

where  $\mathfrak{E}_{h,k} : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  is given for all  $(q, p) \in \mathbb{R}^d \times \mathbb{R}^d$  by

$$(32) \quad \mathfrak{E}_{h,k}(q, p) = \sum_{i=1}^{k-1} (k-i) \nabla U \circ \tilde{\Phi}_h^{\circ(i)}(q, p).$$

4.1. *Proof of Theorem 1.* We first prove (9). Under the assumption that  $U$  is twice continuously differentiable, it follows by a straightforward induction, that for all  $h > 0$  and  $q \in \mathbb{R}^d$ ,  $p \mapsto \tilde{\Phi}_h^{\circ(k)}(q, p)$ , defined by (4), and  $p \mapsto \Xi_{h,k}(q, p)$ , defined by (32), are continuously differentiable and for all  $(q, p) \in \mathbb{R}^d \times \mathbb{R}^d$ ,

$$\mathbf{J}_{p, \Xi_{h,T}}(q, p) = \sum_{i=1}^{T-1} (T-i) \{ \nabla^2 U \circ \tilde{\Phi}_h^{\circ(i)}(q, p) \} \mathbf{J}_{p, \tilde{\Phi}_h^{\circ(i)}}(q, p),$$

where for all  $q \in \mathbb{R}^d$ ,  $\mathbf{J}_{p, \Xi_{h,k}}(q, p)$  ( $\mathbf{J}_{p, \tilde{\Phi}_i^{\circ(h)}}(q, p)$ , resp.) is the Jacobian of the function  $\tilde{p} \mapsto \Xi_{h,k}(q, \tilde{p})$  ( $\tilde{p} \mapsto \tilde{\Phi}_i^{\circ(h)}(q, \tilde{p})$ , resp.) at  $p \in \mathbb{R}^d$ .

Under H 1,  $\sup_{x \in \mathbb{R}^d} \|\nabla^2 U(x)\| \leq L_1$ , therefore by Lemma S3, we have that for any  $T \in \mathbb{N}^*$  and  $h > 0$ ,

$$(33) \quad \sup_{(q,p) \in \mathbb{R}^d \times \mathbb{R}^d} \|\mathbf{J}_{p, \Xi_{h,T}}(q, p)\| \leq T(\{1 + hL_1^{1/2} \vartheta_1(hL_1^{1/2})\}^T - 1)/h.$$

For any  $q \in \mathbb{R}^d$ ,  $T \in \mathbb{N}^*$  and  $h > 0$ , define  $\phi_{q,T,h}(p)$  for all  $p \in \mathbb{R}^d$  by

$$\phi_{q,T,h}(p) = p - (h/T)\Xi_{h,T}(q, p).$$

It is a well-known fact (see, e.g., [10], Exercise 3.26) that if

$$(34) \quad \sup_{(q,p) \in \mathbb{R}^d \times \mathbb{R}^d} (h/T)\|\mathbf{J}_{p, \Xi_{h,T}}(q, p)\| < 1,$$

then for any  $q \in \mathbb{R}^d$ ,  $\phi_{q,T,h}$  is a diffeomorphism and, therefore, by (31), the same conclusion holds for  $p \mapsto \tilde{\Phi}_h^{\circ(T)}(q, p)$ . Using (33), if  $T \in \mathbb{N}^*$  and  $h > 0$  satisfies (8), then the condition (34) is verified and (9) follows.

Denoting for any  $q \in \mathbb{R}^d$  by  $\tilde{\Psi}_h^{(T)}(q, \cdot) : \mathbb{R}^d \rightarrow \mathbb{R}$  the continuously differentiable inverse of  $p \mapsto \tilde{\Phi}_h^{\circ(T)}(q, p)$  and using a change of variable with  $\tilde{\Psi}_h^{(T)}(q, \cdot)$  in (6) concludes the proof of (10).

We now show that  $T_{h,T}$  satisfies the condition which implies that  $P_{h,T}$  is a T-kernel. We first establish some estimates on the function  $(q, p) \mapsto \tilde{\Psi}_h^{(T)}(q, p)$ . By (34) and (31), for any  $q, p, v \in \mathbb{R}^d$ , there exists  $\varepsilon \in (0, 1)$  such that  $\|\tilde{\Phi}_h^{\circ(T)}(q, p) - \tilde{\Phi}_h^{\circ(T)}(q, v)\| \geq (hT)\|\phi_{q,T,h}(p) - \phi_{q,T,h}(v)\| \geq (hT)(1 - \varepsilon)\|p - v\|$  which implies that there exists  $C \geq 0$  satisfying

$$(35) \quad \begin{aligned} \|\tilde{\Psi}_h^{(T)}(q, p) - \tilde{\Psi}_h^{(T)}(q, v)\| &\leq (1 - \varepsilon)^{-1}\|v - p\|, \\ \|\tilde{\Psi}_h^{(T)}(q, p)\| &\leq C\{\|p\| + \|\tilde{\Phi}_h^{\circ(T)}(q, 0)\|\}. \end{aligned}$$

In addition, for  $q, x, p \in \mathbb{R}^d$ , we have setting  $\tilde{q} = \tilde{\Psi}_h^{(T)}(q, p)$  that

$$\begin{aligned} \|\tilde{\Psi}_h^{(T)}(q, p) - \tilde{\Psi}_h^{(T)}(x, p)\| &= \|\tilde{q} - \tilde{\Psi}_h^{(T)}(x, \tilde{\Phi}_h^{\circ(T)}(q, \tilde{q}))\| \\ &= \|\tilde{\Psi}_h^{(T)}(x, \tilde{\Phi}_h^{\circ(T)}(x, \tilde{q})) - \tilde{\Psi}_h^{(T)}(x, \tilde{\Phi}_h^{\circ(T)}(q, \tilde{q}))\|, \end{aligned}$$

which implies by (35) and Lemma S1 that there exists  $C \geq 0$  satisfying

$$(36) \quad \|\tilde{\Psi}_h^{(T)}(q, p) - \tilde{\Psi}_h^{(T)}(x, p)\| \leq C\|q - x\|.$$

We now can prove that  $T_{h,T}$  is the continuous component of  $P_{h,T}$ . First by (11), for all  $B \in \mathcal{B}(\mathbb{R}^d)$ ,

$$T_{h,T}(q, B) \geq (2\pi)^{-d/2} \text{Leb}(B) \times \inf_{\tilde{q} \in B} \{\bar{\alpha}_H(q, \tilde{q})e^{-\|\tilde{\Psi}_q(\tilde{q})\|^2/2} D_{\tilde{\Psi}_h^{(T)}(q, \cdot)}(\tilde{q})\},$$

with the convention  $0 \times +\infty = 0$  and

$$\bar{\alpha}_H(q, \bar{q}) = \alpha_H\{(q, \bar{\Psi}_h^{(T)}(q, \bar{q})), \Phi_h^{\circ(T)}(q, \bar{\Psi}_h^{(T)}(q, \bar{q}))\}.$$

Since the function  $(q, p) \mapsto (\tilde{\Phi}_h^{\circ(T)}(q, p), \bar{\Psi}_h^{(T)}(q, p), D_{\bar{\Psi}_h^{(T)}(q, \cdot)}(p))$  is continuous on  $\mathbb{R}^d \times \mathbb{R}^d$  by Lemma S1, (35) and (36), and for any  $q, p \in \mathbb{R}^d, \mathbf{J}_{\tilde{\Phi}_h^{\circ(T)}(q, \cdot)}(\bar{\Psi}_h^{(T)}(q, p))\mathbf{J}_{\bar{\Psi}_h^{(T)}(q, \cdot)}(p) = \mathbf{I}_n$ , we get that  $T_{h,T}(q, \mathbf{B}) > 0$  for all  $q \in \mathbb{R}^d$  and all compact sets  $\mathbf{B}$  satisfying  $\text{Leb}(\mathbf{B}) > 0$ . Therefore, using that the Lebesgue measure is regular which implies that for any  $\mathbf{A} \in \mathcal{B}(\mathbb{R}^d)$  with  $\text{Leb}(\mathbf{A}) > 0$ , there exists a compact set  $\mathbf{B} \subset \mathbf{A}, \text{Leb}(\mathbf{B}) > 0$ , we can conclude that  $P_{h,T}$  is irreducible with respect to the Lebesgue measure. In addition, we get  $T_{h,T}(q, \mathbb{R}^d) > 0$  and, therefore, we obtain that  $P_{h,T}$  is aperiodic. Similarly, we get that any compact set is 1-small.

It remains to show that for any  $\mathbf{B} \in \mathcal{B}(\mathbb{R}^d), q \mapsto T_{h,T}(q, \mathbf{B})$  is lower semi-continuous which is a straightforward consequence of Fatou’s lemma and that for any  $p \in \mathbb{R}^d, q \mapsto (\Phi_h^{\circ(T)}(q, p), \bar{\Psi}_h^{(T)}(q, p), D_{\bar{\Psi}_h^{(T)}(q, \cdot)}(p))$  is continuous.

Finally, the last statements of (ii) follows from Proposition S11 in [11], Section S3, which implies that  $P_{h,T}$  is Harris recurrent and [21], Theorem 13.0.1, which implies (12).

4.2. *Proof of Theorem 2.* We use Corollary 14 of the Appendix. Indeed  $P_{h,T}$  is of form (37) and it is straightforward to check that it satisfies G 1 (note that Lemma S1 shows that  $\Phi_h^{\circ(T)}$  is a Lipschitz function on  $\mathbb{R}^{2d}$ ).

We now check that  $P_{h,T}$  satisfies G 2( $R, 0, M$ ) for all  $R, M \in \mathbb{R}_+^*$  using Proposition 15. By (31), for all  $T \in \mathbb{N}^*, h > 0, q, p \in \mathbb{R}^d$ ,

$$\tilde{\Phi}_h^{\circ(T)}(q, p) = Thp + g_{q,T,h}(p),$$

where  $g_{q,T,h}(p) = q - (Th^2/2)\nabla U(q) - h^2\Xi_{h,T}(q, p)$  where  $\Xi_{h,T}$  is defined by (32). Lemma S3 shows that for any  $T \in \mathbb{N}^*$  and  $h > 0$ , it holds that

$$\sup_{p,v,q \in \mathbb{R}^d} \frac{\|g_{q,T,h}(p) - g_{q,T,h}(v)\|}{\|p - v\|} \leq Th[\{1 + hL_1^{1/2}\vartheta_1(hL_1^{1/2})\}^T - 1],$$

which implies that the condition Proposition 15(i) is satisfied. To check that condition Proposition 15(ii) holds, we consider separately the two cases:  $\beta < 1$  and  $\beta = 1$ .

- Consider first the case  $\beta < 1$ . By H 1(ii), for any  $T \in \mathbb{N}^*$  and  $h > 0$ , we get

$$\|\Xi_{h,T}(q, p)\| \leq T \sum_{i=1}^{T-1} \|\nabla U \circ \tilde{\Phi}_h^{\circ(i)}(q, p)\| \leq M_1 T \sum_{i=1}^{T-1} \{1 + \|\tilde{\Phi}_h^{\circ(i)}(q, p)\|^\beta\}.$$

Hence, by Lemma S2(i) there exists  $C \geq 0$  such that for all  $R \in \mathbb{R}_+^*$  and  $q, p \in \mathbb{R}^d, \|q\| \leq R,$

$$\|g_{q,T,h}(p)\| \leq C\{1 + R^\beta + \|p\|^\beta\},$$

which implies that condition (ii) of Proposition 15 holds for any  $T \in \mathbb{N}^*$  and  $h > 0$ .

- Consider now the case  $\beta = 1$ . For any  $T \in \mathbb{N}^*, h > 0, q, p \in \mathbb{R}^d$  we get using H 1(i)

$$\begin{aligned} \|g_{q,T,h}(p)\| &\leq \|q\| + Th^2L_1\|q\|/2 + Th^2\|\nabla U(0)\|/2 \\ &\quad + h^2\|\Xi_{h,T}(q, p) - \Xi_{h,T}(q, 0)\| + h^2\|\Xi_{h,T}(q, 0)\|. \end{aligned}$$

Therefore, using Lemma S3, for any  $q, p \in \mathbb{R}^d, \|q\| \leq R$  for  $R \geq 0$ , for any  $T \in \mathbb{N}^*$  and  $h > 0$  satisfying (8), there exists  $C \geq 0$  such that

$$\|g_{q,T,h}(p)\| \leq C + hT[\{1 + hL_1^{1/2}\vartheta_1(hL_1^{1/2})\}^T - 1]\|p\|,$$

showing that condition (ii) of Proposition 15 is satisfied.

Therefore, Proposition 15 can be applied and for any  $T \in \mathbb{N}^*$  and  $h > 0$  if  $\beta < 1$  and for any  $h > 0$  and  $T \in \mathbb{N}^*$  satisfying (8) if  $\beta = 1$ ,  $P_{h,T}$  satisfies G 2( $R, 0, M$ ) for all  $R, M \in \mathbb{R}_+^*$ . Corollary 14 concludes the proof of (a) and (b). The last statement then follows from [21], Theorem 14.0.1.

APPENDIX: IRREDUCIBILITY FOR A CLASS OF ITERATIVE MODELS

In this section, we establish the irreducibility of a Markov kernel associated to a random iterative model. These results are of independent interest. Let  $f : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  and  $\alpha : \mathbb{R}^d \times \mathbb{R}^d \rightarrow [0, 1]$  be Borel measurable functions and  $\phi : \mathbb{R}^d \rightarrow [0, +\infty]$  be a probability density with respect to the Lebesgue measure. Consider the Markov kernel  $K$  defined for all  $x \in \mathbb{R}^d$  and  $A \in \mathcal{B}(\mathbb{R}^d)$  by

$$(37) \quad K(x, A) = \int_{\mathbb{R}^d} \mathbb{1}_A(f(x, z))\alpha(x, z)\phi(z) dz + \bar{\alpha}(x)\delta_x(A),$$

where  $\bar{\alpha}(x) = \int_{\mathbb{R}^d} \alpha(x, z)\phi(z) dz$ . Define for all  $x \in \mathbb{R}^d$ ,  $f_x : \mathbb{R}^d \rightarrow \mathbb{R}^d$  by  $f_x = f(x, \cdot)$ .

First, we give a result from geometric measure theory together with a proof for the reader’s convenience, which will be essential for the proof of the statements of this section. Let  $U \subset \mathbb{R}^d$  be an open set and  $\Theta : U \rightarrow \mathbb{R}^d$  be a measurable function such that there exist  $y_0, \tilde{y}_0 \in \mathbb{R}^d$  and  $M, \tilde{M} > 0$  satisfying  $B(\tilde{y}_0, \tilde{M}) \subset U$  and

$$(38) \quad B(y_0, M) \subset \Theta(B(\tilde{y}_0, \tilde{M})).$$

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

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

Note that  $\lambda_\Theta$  is a finite measure. Therefore, by the Lebesgue decomposition theorem (see [27], Section 6.10) there exist two measures  $\lambda_\Theta^{(a)}, \lambda_\Theta^{(s)}$  on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ , 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 12.** *Let  $U \subset \mathbb{R}^d$  be open and  $\Theta : U \rightarrow \mathbb{R}^d$  be a Lipschitz function satisfying (38). For any version  $\phi_\Theta$  of the density of  $\lambda_\Theta^{(a)}$  with respect to the Lebesgue measure on  $\mathbb{R}^d$ , it holds*

$$\phi_\Theta(y) \geq \mathbb{1}_{B(y_0, M)}(y) \|\Theta\|_{\text{Lip}}^{-d}, \quad \text{Leb-a.e.}$$

**PROOF.** Denote by  $L = \|\Theta\|_{\text{Lip}}$ . Let  $y \in B(y_0, M)$ . By (38), we may pick  $z \in B(\tilde{y}_0, \tilde{M})$  such that  $\Theta(z) = y$ . Let  $\delta_0 > 0$  be such that  $B(z, \delta_0/L) \subset B(\tilde{y}_0, \tilde{M})$ . 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 [27], Theorem 7.14. □

We can now state our main results. Let  $R, M \in \mathbb{R}_+^*$  and  $y_0 \in \mathbb{R}^d$ . Consider the following assumptions.

- G 1.  $\phi$  and  $\alpha$  are lower semicontinuous and positive on  $\mathbb{R}^d$  and  $\mathbb{R}^{2d}$ , respectively.
- G 2 ( $R, y_0, M$ ).

- (i) There exists  $L_f \in \mathbb{R}_+$  such that for all  $x \in B(0, R)$ ,  $f_x$  is  $L_f$ -Lipschitz, that is, 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 exist  $\tilde{y}_0 \in \mathbb{R}^d$  and  $\tilde{M} \in \mathbb{R}_+^*$ , such that for all  $x \in B(0, R)$ ,  $B(y_0, M) \subset f_x(B(\tilde{y}_0, \tilde{M}))$ .

**THEOREM 13.** *Assume G 1 and that there exist  $y_0 \in \mathbb{R}^d$ ,  $R > 0$  and  $M > 0$  such that G 2( $R, y_0, M$ ) is satisfied. Then  $B(0, R)$  is 1-small for  $K$ : for all  $x \in B(0, R)$  and  $A \in \mathcal{B}(\mathbb{R}^d)$ ,*

$$K(x, A) \geq L_f^{-d} \min_{(x,z) \in B(0,R) \times B(\tilde{y}_0, \tilde{M})} \{\alpha(x, z)\phi(z)\} \text{Leb}\{A \cap B(y_0, M)\},$$

where  $(\tilde{y}_0, \tilde{M}) \in \mathbb{R}^d \times \mathbb{R}_+^*$  is defined in G 2( $R, y_0, M$ ).

**PROOF.** For all  $x \in B(0, R)$  and  $A \in \mathcal{B}(\mathbb{R}^d)$ , we get

$$\begin{aligned} K(x, A) &= \int_{\mathbb{R}^d} \mathbb{1}_A(f(x, z))\alpha(x, z)\phi(z) dz = \int_{\mathbb{R}^d} \mathbb{1}_{f_x^{-1}(A)}(z)\alpha(x, z)\phi(z) dz \\ &\geq \min_{(x,z) \in B(0,R) \times B(\tilde{y}_0, \tilde{M})} \{\alpha(x, z)\phi(z)\} \text{Leb}\{f_x^{-1}(A) \cap B(\tilde{y}_0, \tilde{M})\}. \end{aligned}$$

The proof follows from Proposition 12 and G 2( $R, y_0, M$ )(i) which imply  $\text{Leb}\{f_x^{-1}(A) \cap B(\tilde{y}_0, \tilde{M})\} \geq L_f^{-d} \text{Leb}\{A \cap B(y_0, M)\}$ .  $\square$

The following corollary is a straightforward consequence of Theorem 13.

**COROLLARY 14.** *Assume G 1 and that there exists  $(y_0, M) \in \mathbb{R}^d \times \mathbb{R}_+^*$  such that for all  $R \in \mathbb{R}_+^*$  G 2( $R, y_0, M$ ). Then  $K$  is irreducible with irreducibility measure  $\text{Leb}\{\cdot \cap B(y_0, M)\}$ . In addition, all the compact sets are 1-small.*

In the next proposition, we give examples of functions  $f$  which satisfy G 2.

**PROPOSITION 15.** *Let  $g$  a function from  $\mathbb{R}^d \times \mathbb{R}^d$  to  $\mathbb{R}^d$  and  $R \in \mathbb{R}_+^*$ . Assume that:*

- (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 2( $R, 0, M$ ) for all  $M \in \mathbb{R}_+^*$  with  $\tilde{y}_0 = 0$  and

$$(39) \quad \tilde{M} = \{M + C_{R,0}\} / (\|b\| - C_{R,1}).$$

We preface the proof by recalling some basic notions of degree theory. Let  $D$  be a bounded open set of  $\mathbb{R}^d$ . Let  $f : \bar{D} \rightarrow \mathbb{R}^d$  be a continuous function on  $\bar{D}$  continuously differentiable on  $D$ . An element  $x \in D$  is said to be a *regular point* of  $f$  if the Jacobian matrix of  $f$  at  $x$ ,  $J_f(x)$ , 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  $f : \bar{D} \rightarrow \mathbb{R}^d$  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 [24], Proposition and Definition 1.1, that the set  $f^{-1}(\{y\})$  is finite. The degree of  $f$  at  $y$  is defined by

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

**PROPOSITION 16** ([24], Proposition and Definition 2.1). *Let  $f : \bar{D} \rightarrow \mathbb{R}^d$  be a continuous function and  $y \in \mathbb{R}^d \setminus f(\partial D)$ .*

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

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

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

Under the assumptions of Proposition 16, the degree of  $f$  at  $y$  is then defined for any  $g : \bar{D} \rightarrow \mathbb{R}^d$  satisfying (a) by

$$\text{deg}(f, D, y) = \text{deg}(g, D, y).$$

**PROPOSITION 17** ([24], Proposition 2.4). *Let  $f, g : \bar{D} \rightarrow \mathbb{R}^d$  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  $y \in \mathbb{R}^d \setminus H([0, 1] \times \partial D)$ . Then*

$$\text{deg}(f, D, y) = \text{deg}(g, D, y).$$

We have now all the necessary results to prove Proposition 15.

**PROOF OF PROPOSITION 15.** 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  $B(0, R)$ ,  $f_x^g$  is Lipschitz with bounded Lipschitz constant over this ball. Hence G 2(R, 0, M)(i) holds.

For all  $x \in \mathbb{R}^d$ , denote by  $f_x^g : z \mapsto f^g(x, z)$  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 (39), which is precisely G 2(R, 0, M)(ii).

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

$$H^g(t, z) = tbz + (1 - t)f_x^g(z) = bz + (1 - t)g(x, z).$$

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 (39),

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

In particular, we have  $H^g([0, 1] \times \partial B(0, \tilde{M})) \subset \mathbb{R}^d \setminus B(0, M)$ . Let  $z \in B(0, M)$ , then by Proposition 17 we have

$$\text{deg}(f_x^g, B(0, \tilde{M}), z) = \text{deg}(b \text{Id}, B(0, \tilde{M}), z) = 1.$$

Besides, by [24], Corollary 2.5, Chapter IV,  $\text{deg}(f_x^g, B(0, \tilde{M}), z) \neq 0$  implies that there exists  $y \in B(0, \tilde{M})$  such that  $f_x^g(y) = z$ . Finally G 2(R, 0, M)(ii) follows since this result holds for all  $z \in B(0, M)$ .  $\square$

**Acknowledgments.** The article was partly prepared within the framework of the HSE University Basic Research Program. AD acknowledges support from Polish National Science Center grant: NCN UMO-2018/31/B/ST1/00253

## SUPPLEMENTARY MATERIAL

**Supplement to “Irreducibility and geometric ergodicity of Hamiltonian Monte Carlo”** (DOI: [10.1214/19-AOS1941SUPP](https://doi.org/10.1214/19-AOS1941SUPP); .pdf). Supplementary information and postponed proofs.

## REFERENCES

- [1] BETANCOURT, M., BYRNE, S., LIVINGSTONE, S. and GIROLAMI, M. (2017). The geometric foundations of Hamiltonian Monte Carlo. *Bernoulli* **23** 2257–2298. MR3648031 <https://doi.org/10.3150/16-BEJ810>
- [2] LIVINGSTONE, S., BETANCOURT, M., BYRNE, S. and GIROLAMI, M. (2016). On the geometric ergodicity of Hamiltonian Monte Carlo. *Bernoulli*. Preprint. Available at [arXiv:1601.08057v2](https://arxiv.org/abs/1601.08057v2).
- [3] BOU-RABEE, N., EBERLE, A. and ZIMMER, R. (2018). Coupling and convergence for Hamiltonian Monte Carlo. Preprint. Available at [arXiv:1805.00452](https://arxiv.org/abs/1805.00452).
- [4] BOU-RABEE, N. and SANZ-SERNA, J. M. (2017). Randomized Hamiltonian Monte Carlo. *Ann. Appl. Probab.* **27** 2159–2194. MR3693523 <https://doi.org/10.1214/16-AAP1255>
- [5] BOU-RABEE, N. and SANZ-SERNA, J. M. (2018). Geometric integrators and the Hamiltonian Monte Carlo method. *Acta Numer.* **27** 113–206. MR3826507 <https://doi.org/10.1017/s0962492917000101>
- [6] BYRNE, S. and GIROLAMI, M. (2013). Geodesic Monte Carlo on embedded manifolds. *Scand. J. Stat.* **40** 825–845. MR3145120 <https://doi.org/10.1111/sjos.12036>
- [7] CANCÈS, E., LEGOLL, F. and STOLTZ, G. (2007). Theoretical and numerical comparison of some sampling methods for molecular dynamics. *ESAIM Math. Model. Numer. Anal.* **41** 351–389. MR2339633 <https://doi.org/10.1051/m2an:2007014>
- [8] DOUC, R., MOULINES, E., PRIOURET, P. and SOULIER, P. (2018). *Markov Chains. Springer Series in Operations Research and Financial Engineering*. Springer, Cham. MR3889011 <https://doi.org/10.1007/978-3-319-97704-1>
- [9] DUANE, S., KENNEDY, A. D., PENDLETON, B. J. and ROWETH, D. (1987). Hybrid Monte Carlo. *Phys. Lett. B* **195** 216–222. MR3960671 [https://doi.org/10.1016/0370-2693\(87\)91197-x](https://doi.org/10.1016/0370-2693(87)91197-x)
- [10] DUISTERMAAT, J. J. and KOLK, J. A. C. (2004). *Multidimensional Real Analysis. I. Differentiation. Cambridge Studies in Advanced Mathematics* **86**. Cambridge Univ. Press, Cambridge. Translated from the Dutch by J. P. van Braam Houckgeest. MR2121976 <https://doi.org/10.1017/CBO9780511616716>
- [11] DURMUS, A., MOULINES, É. and SAKSMAN, E. (2020). Supplement to “Irreducibility and geometric ergodicity of Hamiltonian Monte Carlo.” <https://doi.org/10.1214/19-AOS1941SUPP>
- [12] FANG, Y., SANZ-SERNA, J. M. and SKEEL, R. D. (2016). Compressible generalized hybrid Monte Carlo (vol 140, 174108, 2014). *J. Chem. Phys.* **144**.
- [13] GIROLAMI, M. and CALDERHEAD, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** 123–214. MR2814492 <https://doi.org/10.1111/j.1467-9868.2010.00765.x>
- [14] LUBICH, C., HAIRER, E. and WANNER, G. (2002). *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, 2nd ed. *Springer Series in Computational Mathematics* **31**. Springer, Heidelberg.
- [15] LIU, J. S. (2008). *Monte Carlo Strategies in Scientific Computing. Springer Series in Statistics*. Springer, New York. MR2401592
- [16] LIVINGSTONE, S., FAULKNER, M. F. and ROBERTS, G. O. (2019). Kinetic energy choice in Hamiltonian/hybrid Monte Carlo. *Biometrika* **106** 303–319. MR3949305 <https://doi.org/10.1093/biomet/asz013>
- [17] LU, X., PERRONE, V., HASENCLEVER, L., TEH, Y. W. and VOLLMER, S. (2017). Relativistic Monte Carlo. In *Proceedings of Machine Learning Research* **54**. Artificial Intelligence and Statistics, AISTAT.
- [18] MANGOUBI, O. and SMITH, A. (2017). Rapid mixing of Hamiltonian Monte Carlo on strongly log-concave distributions. Preprint. Available at [arXiv:1708.07114](https://arxiv.org/abs/1708.07114).
- [19] MANGOUBI, O. and VISHNOI, N. (2018). Dimensionally tight bounds for second-order Hamiltonian Monte Carlo. In *Advances in Neural Information Processing Systems* 6027–6037.
- [20] MENGERSEN, K. L. and TWEEDIE, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24** 101–121. MR1389882 <https://doi.org/10.1214/aos/1033066201>
- [21] MEYN, S. and TWEEDIE, R. L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge Univ. Press, Cambridge. MR2509253 <https://doi.org/10.1017/CBO9780511626630>
- [22] NEAL, R. M. (1993). Bayesian learning via stochastic dynamics. *Adv. Neural Inf. Process. Syst.* 475–475.

- [23] NEAL, R. M. (2011). MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC Handb. Mod. Stat. Methods 113–162. CRC Press, Boca Raton, FL. [MR2858447](#)
- [24] OUTERELO, E. and RUIZ, J. M. (2009). *Mapping Degree Theory. Graduate Studies in Mathematics* **108**. Amer. Math. Soc., Providence, RI. [MR2566906](#) <https://doi.org/10.1090/gsm/108>
- [25] ROBERTS, G. O. and TWEEDIE, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2** 341–363. [MR1440273](#) <https://doi.org/10.2307/3318418>
- [26] ROBERTS, G. O. and TWEEDIE, R. L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83** 95–110. [MR1399158](#) <https://doi.org/10.1093/biomet/83.1.95>
- [27] RUDIN, W. (1987). *Real and Complex Analysis*, 3rd ed. McGraw-Hill, New York. [MR0924157](#)
- [28] SCHOFIELD, M. R., BARKER, R. J., GELMAN, A., COOK, E. R. and BRIFFA, K. R. (2016). A model-based approach to climate reconstruction using tree-ring data. *J. Amer. Statist. Assoc.* **111** 93–106. [MR3494640](#) <https://doi.org/10.1080/01621459.2015.1110524>
- [29] TANG, Y., SRIVASTAVA, N. and SALAKHUTDINOV, R. R. (2014). Learning generative models with visual attention. In *Advances in Neural Information Processing Systems* 1808–1816.