*J. Appl. Probab.* 1–27 (2025)
*doi:*10.1017/jpr.2025.10025

# ON THE VARIANCE REDUCTION OF HAMILTONIAN MONTE CARLO VIA AN APPROXIMATION SCHEME

ZHONGGEN SU,* ** AND
ZEYU YAO,* *Zhejiang University*

**Abstract**

Let $\pi$ be a probability distribution in $\mathbb{R}^d$ and $f$ a test function, and consider the problem of variance reduction in estimating $\mathbb{E}_\pi(f)$. We first construct a sequence of estimators for $\mathbb{E}_\pi(f)$, say $(1/k)\sum_{i=0}^{k-1} g_n(X_i)$, where the $X_i$ are samples from $\pi$ generated by the Metropolized Hamiltonian Monte Carlo algorithm and $g_n$ is the approximate solution of the Poisson equation through the weak approximate scheme recently invented by Mijatović and Vogrinc (2018). Then we prove under some regularity assumptions that the estimation error variance $\sigma_\pi^2(g_n)$ can be as arbitrarily small as the approximation order parameter $n \to \infty$. To illustrate, we confirm that the assumptions are satisfied by two typical concrete models, a Bayesian linear inverse problem and a two-component mixture of Gaussian distributions.

*Keywords:* Central limit theorem; Markov chain Monte Carlo; Poisson equation; weak approximation

2010 Mathematics Subject Classification: Primary 60J05
Secondary 60J22; 65C05

## 1. Introduction and main results

Let $\pi$ be a target distribution in $\mathbb{R}^d$ and $f\colon \mathbb{R}^d \mapsto \mathbb{R}$ a test function. We are mainly concerned with the issue of variance reduction in estimating $\mathbb{E}_\pi(f)$ based on the samples generated by the Hamiltonian Monte Carlo algorithm. The variance reduction is a widely used method in statistical inferences. In fact, if $\pi$ is comparatively simple so that independent samples are easily drawn, say $(X_i, i \geq 1)$, then we can use $(1/k)\sum_{i=1}^k f(X_i)$ as an estimator of $\mathbb{E}_\pi(f)$ by the well-known law of large numbers. The estimation error is asymptotically given by $(1/k)\mathbb{E}_\pi(f - \mathbb{E}_\pi(f))^2$. However, if $\mathbb{E}_\pi(f - \mathbb{E}_\pi(f))^2$ is itself not negligible, then sufficiently many samples must be drawn in order to make the error as small as possible. Thus, in order to upgrade the efficiency of estimation we need to resort to the variance reduction method. That is, to find a function $g$ with known $\mathbb{E}_\pi(g)$, say $\mathbb{E}_\pi(g) = 0$ and $\mathbb{E}_\pi(f + g - \mathbb{E}_\pi(f))^2 < \mathbb{E}_\pi(f - \mathbb{E}_\pi(f))^2$, so that we might prefer to use $(1/k)\sum_{i=1}^k (f(X_i) + g(X_i))$ as an alternative estimator of $\mathbb{E}_\pi(f)$ by the law of large numbers again. A natural problem now arises: how do we construct such a $g(x)$? This is not an easy task in general.

Received 10 April 2024; accepted 19 May 2025.
* Postal address: School of Mathematical Science, Zhejiang University, 866 Yuhangtang Road, Hangzhou, 310058, P.R. China.
** Email address: suzhonggen@zju.edu.cn

On the other hand, we are also required to solve the issue of simulating the target distribution $\pi$. In fact, this latter issue is more challenging and often appears in diverse research fields like high-dimensional data analysis, Bayesian statistical inference, high-performance numeric computation, and machine learning.

### 1.1. Metropolis–Hastings Monte Carlo

Markov chain theory turns out to offer a powerful tool for managing an issue like simulation, which resulted in the Markov chain Monte Carlo (MCMC) method. Assume that $(X_i,\ i \geq 0)$ is an ergodic Markov chain with $\pi$ as its stationary distribution. Then it follows by Birkhoff's theorem that $\lim_{k\to\infty}\left[(1/k)\sum_{i=0}^{k-1} f(X_i)\right] = \mathbb{E}_\pi(f)$ $\pi$-almost everywhere (a.e.). Therefore, the time mean $(1/k)\sum_{i=0}^{k-1} f(X_i)$ may be used as a good estimator for $\mathbb{E}_\pi(f)$ when the Markov chain is run for a sufficiently long time.

Constructing such a Markov chain that targets the desired distribution, however, is itself a nontrivial problem. Fortunately, various procedures have been outlined in the literature for automatically constructing appropriate transitions for any given target distribution, with the foremost among these the Metropolis–Hastings algorithm.

Let $\pi$ be a fixed probability measure, and assume it possesses a density function, denoted as $\pi(x)$ (with a minor abuse of notation). Start with a proposal transition kernel $Q(x, \cdot)$ with density function $q(x, y)$ and run the Metropolis–Hastings algorithm to generate a chain $(X_i,\ i \geq 0)$ as follows. Given the current state $X_i = x$, sample a candidate $Y = y$ from the proposal kernel $Q(x, \cdot)$, and then the calculate the acceptance rate:

$$\alpha(x, y) := \min\left\{1, \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}\right\}. \tag{1.1}$$

Then independently draw a uniform random variable $Z \sim U[0, 1]$. If $Z \leq \alpha(x, y)$ then set $X_{i+1} = y$; otherwise, set $X_{i+1} = x$. The associated Markov kernel can be written as follows. For any set $A \in \mathcal{B}(\mathbb{R}^d)$,

$$P(x, A) = \int_A \alpha(x, y)q(x, y)\,\mathrm{d}y + \delta_x(A)\int_{\mathbb{R}^d}(1 - \alpha(x, y))q(x, y)\,\mathrm{d}y.$$

The Markov kernel $P(x, \mathrm{d}y)$ is remarkably reversible with respect to $\pi$, i.e. $\pi(x)P(x, \mathrm{d}y) = \pi(y)P(y, \mathrm{d}x)$, and so the stationary distribution of the chain $(X_i,\ i \geq 0)$ is exactly the target distribution $\pi$ [5, Proposition 2.3.1]. It is worth noting that only the ratio $\pi(y)/\pi(x)$ is used in the acceptance/rejection probability (1.1), and thus it easily extends to a larger class of distributions, particularly like the Bayesian posterior distributions whose normalization constants are hardly computable.

As the reader may realize, the Metropolis–Hastings framework is quite flexible by instantiating it with different choices for the proposal transition kernel $Q(x, \mathrm{d}y)$, which in turn have a significant influence upon sampling. A simple and widely used choice is the so-called Metropolized random walk (MRW). This corresponds to simply taking a random walk with transition kernel $q(x, \cdot) \sim \mathcal{N}(x, hI_d)$ around the state space, where some steps are occasionally rejected. Note that the proposal is independent of the target $\pi$, and so uses only a zeroth-order oracle for $\pi$.

If $\pi$ has a smooth enough density function, particularly $\pi(x) \propto \mathrm{e}^{-U(x)}$, where $U(x) \in \mathcal{C}^2(\mathbb{R}^d)$, we can exploit the information from the gradient. Consider the stochastic Langevin diffusion process

$$\mathrm{d}X_t = -\nabla U(X_t)\,\mathrm{d}t + \sqrt{2}\,\mathrm{d}B_t, \tag{1.2}$$

https://doi.org/10.1017/jpr.2025.10025 Published online by Cambridge University Press

where $B_t$ is the standard Brownian motion on $\mathbb{R}^d$. Under certain smoothness conditions, the distribution of $X_t$ converges in probability to $\pi$ as $t \to \infty$ regardless of the initial value $X_0$ [22]. In practice, numerical machine computation requires finite precision and updates to $X$. We adapt the Euler–Maruyama discretization of (1.2) to offer a proper choice for transition,

$$X_{k+1} = X_k - h\nabla U(X_k) + \sqrt{2h}\xi_{k+1}, \tag{1.3}$$

where $h > 0$ is the step size and $(\xi_k,\ k \geq 1)$ is a sequence of independent and identically distributed (i.i.d.) $\mathcal{N}(0, 1)$ random variables. In other words, the proposal transition kernel $q(x, \cdot)$ is just $\mathcal{N}(x - h\nabla U(x), 2hI_d)$. The algorithm corresponding to (1.3) is often referred to as the unadjusted Langevin algorithm. Note that the invariant measure, say $\pi_h$, induced by (1.3) is not identical to the target $\pi$, so a bias must be corrected. The corresponding Metropolis–Hastings algorithm is called the Metropolized adjusted Langevin algorithm (MALA).

There have been plenty of works investigating the efficiency of MRW and MALA; see [10, 18–20, 22, 24] and references therein for more details. MRW is still popular in many applications because of its conceptual simplicity and the ease of implementation. Unfortunately, it is typically the slowest in terms of the total number of iterations, and performs poorly with increasing dimension and complexity of the target distribution.

To avoid the slow exploration of the state space that results from the diffusive behavior of simple random walk proposals, we resort to the Hamiltonian Monte Carlo (HMC), which automatically generate distant and coherent exploration for sufficiently well-behaved target distributions by carefully exploiting the differential structure of the target probability density.

### 1.2. Metropolized Hamiltonian Monte Carlo

Metropolized Hamiltonian Monte Carlo, abbreviated to MHMC, was introduced in [6] in computational physics, and came to the statistics community two decades later, quickly gaining popularity. The reader is referred to [2–4, 23] for nice introductory reviews.

The basic idea behind the MHMC method is as follows. Let $H(x, v) = U(x) + \|v\|^2/2$, $(x, v) \in \mathbb{R}^d \times \mathbb{R}^d$. We augment the target distribution $\pi \propto \mathrm{e}^{-U(x)}$ to add a momentum variable $v$. Specifically, define the Boltzman–Gibbs probability measure $\mu_{\mathrm{BG}}(x, v) \propto \mathrm{e}^{-H(x,v)}$. Obviously, the first marginal of $\mu_{\mathrm{BG}}$ is $\pi$, so if we obtain a sample from $\mu_{\mathrm{BG}}$, then upon projecting to the first coordinate we obtain a sample from $\pi$.

Simulating $\mu_{\mathrm{BG}}$ is in turn done using the Hamiltonian dynamics formulation, which is a reformulation of classical dynamics. Denote by $(x_t, v_t)$ the state at time $t$ of a physical system, where $x_t$ is the position vector and $v_t$ the momentum vector. The evolution of the system through time is then given by the Hamiltonian equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \nabla_v H(x_t, v_t) = v_t, \qquad \frac{\mathrm{d}v}{\mathrm{d}t} = -\nabla_x H(x_t, v_t) = -\nabla U(x_t).$$

Denote by $\Phi_t \colon (x, v) \mapsto (x_t, v_t)$ the Hamiltonian flow, where $(x, v)$ and $(x_t, v_t)$ are starting states and the states after time $t$, respectively. The key advantage of the Hamiltonian dynamics over other formulations of classical dynamics is that the Hamiltonian flow possesses the following fundamental properties.

*Conservation of energy*: Along the Hamiltonian dynamics, $(x_t, v_t)_{t\geq 0}$ satisfies $H(x_t, v_t) = H\big(x_0, v_0\big)$, $t \geq 0$. In fact, it is easy to check that $\partial H(x_t, v_t)/\partial t \equiv 0$.

https://doi.org/10.1017/jpr.2025.10025 Published online by Cambridge University Press

*Conservation of volume*: $\Phi_t$ is a volume-preserving map since $\det(\nabla\Phi_t(x, v)) = 1$ for all $t \geq 0$ and $x, v \in \mathbb{R}^d$.

*Time reversibility*: Assume $(x_t, v_t)_{0\leq t\leq T}$ solve Hamilton's equations, then $(x_{T-t}, -v_{T-t})_{0\leq t\leq T}$ also solve Hamilton's equations.

*Preservation of Boltzman–Gibbs measure*: $\Phi_t$ leaves the augmented target $\mu_{\mathrm{BG}}$ invariant, $\mu_{\mathrm{BG}} \circ \Phi_t^{-1} = \mu_{\mathrm{BG}}$.

However, simply running Hamilton's equations does not yield a convergent sampling algorithm. To get around this issue we need to refresh the momentum periodically. The ideal HMC algorithm is as follows:

> Pick an integration time $T > 0$ and draw $(X_0, V_0) \sim \mu_0$, where $\mu_0$ is a fixed but arbitrary probability distribution. For each iteration $k = 0, 1, 2, \ldots$
>
> **Step 1:** Refresh the momentum by drawing $\xi_{kT} \sim N(0, I_d)$.
>
> **Step 2:** Integrate Hamilton's equations, and set $(X_{(k+1)T}, V_{(k+1)T}) = \Phi_T(X_{kT}, \xi_{kT})$.

Since both steps of each iteration preserve $\mu_{\mathrm{BG}}$, the entire algorithm preserves $\mu_{\mathrm{BG}}$. At this stage, though, this algorithm is still idealized because it assumes the ability to exactly integrate Hamilton's equations. This is generally not possible outside a few special cases. We approximately implement Hamilton's equations through the use of a numerical integrator. The simplest and most well-known such integrator is the so-called leapfrog integrator.

Pick a step size $h > 0$ and a total number of iterations $K$, corresponding to the total integration time via $T = Kh$. Let $(x_0, v_0)$ be the initial point. For $i = 0, 1, 2, \ldots, K-1$,

$$\begin{cases} v_{(i+1/2)h} = v_{ih} - \dfrac{h}{2}\nabla U(x_{ih}), \\ x_{(i+1)h} = x_{ih} + hv_{(i+1/2)h}, \\ v_{(i+1)h} = v_{(i+1/2)h} - \dfrac{h}{2}\nabla U(x_{(i+1)h}). \end{cases} \tag{1.4}$$

Let $\Phi_{h,T}(x, v)$ be the output of the leapfrog integrator with $K$ steps started at $(x, v)$. Once we apply the leapfrog integrator to HMC, we obtain a discrete-time sampling algorithm that is once again biased. We correct the bias through the use of the Metropolis–Hastings acceptance/rejection dynamic. The MHMC algorithm is now summarized as follows:

> Initialize at $X_0 \sim \pi_0$, where $\pi_0$ is a fixed but arbitrary probability distribution. For iterations $i = 0, 1, 2, \ldots$
>
> **Step 1:** Refresh the momentum: draw $V_i \sim \mathcal{N}(0, I_d)$.
>
> **Step 2:** Propose a trajectory: let $(X_i', V_i') = \Phi_{h,T}(X_i, V_i)$.
>
> **Step 3:** Accept the trajectory with probability
>
> $$\alpha((X_i, V_i), (X_i', V_i')) = \min\left\{1, \frac{\exp(-H(X_i', V_i'))}{\exp(-H(X_i, V_i))}\right\}.$$

If the trajectory is accepted, set $X_{i+1} = X_i'$; otherwise, we set $X_{i+1} = X_i$. Iteratively, we can obtain a Markov chain $(X_i,\ i \geq 0)$ with kernel $P_{h,T}$ as follows:

https://doi.org/10.1017/jpr.2025.10025 Published online by Cambridge University Press

$$P_{h,T}(x, A) = \int_{\mathbb{R}^d} \mathbf{1}_{A\times\mathbb{R}^d}(\Phi_{h,T}(x, v))\alpha(x, v)\eta(v)\,\mathrm{d}v + \delta_x(A)\left(1 - \int_{\mathbb{R}^d} \alpha(x, v)\eta(v)\,\mathrm{d}v\right), \quad (1.5)$$

with $\alpha(x, v) := \alpha((x, v), \Phi_{h,T}(x, v)) = \min\{1, \exp(-H(\Phi_{h,T}(x, v)) + H(x, v))\}$ and $\eta(v) = (2\pi)^{-d/2}\exp\left(-\frac{1}{2}\|v\|^2\right)$. It is easy to see that $\pi$ is an invariant probability measure with respect to kernel $P_{h,T}$ [3]. More properties of the HMC algorithm are detailed in Section 2.

The main purpose of this article is to construct an efficient estimator whose error variances are asymptotically negligible based on the samples generated by the MHMC algorithm.

### 1.3. Main results

Besides the invariance of $\pi$, [9] showed that $P_{h,T}$ is ergodic under regular conditions, and hence, by Birkhoff's ergodic theorem, for $\pi$-almost every $x \in \mathbb{R}^d$, the following limit holds for $\pi$-integrable $f$: $\lim_{k\to\infty}(1/k)\sum_{i=0}^{k-1} f(X_i) = \mathbb{E}_\pi(f)$, $P_x$-a.e., where $P_x$ stands for the law with the initial state $X_0 = x$. Therefore, it is once again natural to take the mean value $(1/k)\sum_{i=0}^{k-1} f(X_i)$ as an approximation for $\mathbb{E}_\pi(f)$.

In order to establish the central limit theorem (CLT) and variance reduction, we need to make some additional assumptions on the target density $\pi(x)$, equivalently $U(x)$, and the test function $f$. Motivated by [9], we introduce the following assumptions.

**Assumption 1.1.** *For some fixed $l \in (1, 2]$, there exist $L_1 > 0$ and $A \in \mathbb{R}$ such that, for every $x \in \mathbb{R}^d$, $\langle \nabla U(x), x\rangle \geq L_1\|x\|^l - A$.*

**Assumption 1.2.** *For some fixed number $l \in (1, 2]$,*

(i) *$U \in C^3(\mathbb{R}^d)$ and there exists a constant $L_2 > 0$ such that $\|\nabla^k U(x)\| \leq L_2(1 + \|x\|^{l-k})$ for all $x \in \mathbb{R}^d$ and $k = 2, 3$.*

(ii) *There exist $L_3 > 0$ and $R_0 \geq 0$ such that $\langle \nabla^2 U(x)\nabla U(x), \nabla U(x)\rangle \geq L_3\|x\|^{3l-4}$ for all $x \in \mathbb{R}^d$, $\|x\| \geq R_0$.*

**Assumption 1.3.** *$U$ is a perturbation of a quadratic function. In particular, $U(x) = \frac{1}{2}x^\top \Sigma x + \zeta(x)$, where $\Sigma$ is a positive definite matrix, $\zeta : \mathbb{R}^d \to \mathbb{R}$ is a differentiable function, and there exist $L_4 \geq 0$ and $\gamma \in [1, 2)$ such that, for all $x, y \in \mathbb{R}^d$,*

$$|\zeta(x)| \leq L_4(1 + \|x\|)^\gamma, \quad (1.6a)$$

$$\|\nabla\zeta(x)\| \leq L_4(1 + \|x\|)^{\gamma-1}, \quad (1.6b)$$

$$\|\nabla\zeta(x) - \nabla\zeta(y)\| \leq L_4\|x - y\|. \quad (1.6c)$$

**Assumption 1.4.** *There exists an $r > 0$ such that $|f(x)| \leq V_r(x)$, where $V_r(x) = \mathrm{e}^{r\|x\|}$, $x \in \mathbb{R}^d$.*

**Remark 1.1.** If Assumptions 1.1 and 1.2 hold for some $l \in (1, 2]$, then for all $T \geq 1$ and $h > 0$ (we require further that $h < M/T^{3/2}$ for a certain constant $M > 0$ when $l = 2$), $P_{h,T}$ satisfies the drift condition $(V_r, \lambda, b, C)$, i.e. there exists a set $C > 0$, $\lambda \in [0, 1)$, and $b \in (0, +\infty)$ such that

$$P_{h,T}V_r(x) \leq \lambda V_r(x) + b\mathbf{1}_{x\in C} \quad \text{for all } x \in \mathbb{R}^d, \quad (1.7)$$

where $C = \{x\colon V_r(x) \leq L_5\}$ for a constant $L_5$. Consequently, $P_{h,T}$ is $V_r$-uniformly geometrically ergodic.

https://doi.org/10.1017/jpr.2025.10025 Published online by Cambridge University Press

Assumption 1.3 is stronger than Assumption 1.2, as mentioned in [9]. In fact, $U$ in Assumption 1.3 has a special form, namely a quadratic function with a perturbation. If Assumption 1.3 holds, then (1.7) holds for all $T \geq 1$ and $0 < h < M/T$, where $M$ is a constant. It is worth noting that the constant $M$ can be taken small enough that the bound (2.1) holds as well.

We are now ready to state our main results.

**Theorem 1.1.** (Central limit theorem.)

(i) *Fix $l \in (1, 2)$. Under Assumptions 1.1, 1.2 and 1.4, $f$ is $\pi$-integrable and, for each $T \geq 1$, with $h > 0$,*

$$\frac{1}{\sqrt{k}} \sum_{i=0}^{k-1} (f(X_i) - \mathbb{E}_\pi(f)) \xrightarrow{\mathrm{d}} \sigma_\pi(f)\mathcal{N}(0, 1), \quad k \to \infty, \tag{1.8}$$

*where $\sigma_\pi^2(f) = \mathrm{Var}_\pi(f) + 2\sum_{k=1}^{\infty} \mathbb{E}_\pi\left[(f - \mathbb{E}_\pi(f))P_{h,T}^k(f - \mathbb{E}_\pi(f))\right] < \infty$.*

(ii) *Fix $l = 2$. Under Assumptions 1.1, 1.2, and 1.4, $f$ is $\pi$-integrable and (1.8) holds for each $T \geq 1$, $0 < h < M/T^{3/2}$ with a constant $M > 0$.*

(iii) *Under Assumptions 1.1, 1.3, and 1.4, $f$ is $\pi$-integrable and (1.8) holds for each $T \geq 1$, $0 < h < M/T$ with a constant $M > 0$.*

As a direct consequence, we can estimate the error variance $\mathbb{E}_\pi\left[(1/k)\sum_{i=0}^{k-1}(f(X_i) - \mathbb{E}_\pi(f))\right]^2$ by $k^{-1}\sigma_\pi^2(f)$ provided $k$ is large enough, which characterizes the convergence rate of $(1/k)\sum_{i=0}^{k-1} f(X_i)$. If $\sigma_\pi^2(f)$ is comparatively large, however, more iterations of MHMC are required to decrease the error of the estimation, which may reduce the efficiency of sampling. To solve such an issue, we introduce another function $g(x)$ such that $\mathbb{E}_\pi(g)$ is known and $\mathbb{E}_\pi\left[(1/k)\sum_{i=0}^{k-1}((f(X_i) + g(X_i) - \mathbb{E}_\pi(g)) - \mathbb{E}_\pi(f))\right]^2$ can be substantially smaller. For example, let $\hat{f}$ be the solution to the Poisson equation

$$\hat{f} - P_{h,T}\hat{f} = f - \mathbb{E}_\pi(f); \tag{1.9}$$

then it is easy to see that $(1/k)\sum_{i=0}^{k-1}(f + P_{h,T}\hat{f} - \hat{f})(X_i)$ is an unbiased estimator for $\mathbb{E}_\pi(f)$ with variance zero, which induces high accuracy. In practice, it is almost impossible to obtain a precise solution to (1.9), however. We need a proper approximation $\tilde{f}$ of $\hat{f}$ and to calculate $\pi\left[(1/k)\sum_{i=0}^{k-1}(f + P_{h,T}\tilde{f} - \tilde{f})(X_i) - \mathbb{E}_\pi(f)\right]^2$ in order to determine the asymptotic variance.

Recently, [17] presented an approximation scheme for a solution of the Poisson equation of a geometrically ergodic Metropolis–Hastings chain, and further proved that the sequence of the asymptotic variance in the CLT for the control-variate estimators converged to zero. The major contribution of this article is to adapt that approximation scheme for a solution of the Poisson equation of MHMC chains to construct a sequence of control-variate estimators and to further prove the asymptotic variances converge to zero, and so realize the variance reduction. Specifically, split the whole space $\mathbb{R}^d$ into a number of subdomains, say $G_0^n, G_1^n, \ldots, G_{m_n}^n$, and choose a suitable point $a_i^n$ inside each subdomain $G_i^n$. Then, based on these points, construct a Markov chain with a finite number of states as a good approximation of the continuous-state Markov chain. In turn, there must exist a solution denoted by $\hat{f}_n$ to the Poisson equation induced by such a finite Markov chain. Finally, define $\tilde{f}_n(x) = \hat{f}_n(a_i^n)$ for $x \in G_i^n$ as in (3.4)

https://doi.org/10.1017/jpr.2025.10025 Published online by Cambridge University Press

and set $g_n = f + P_{h,T}\tilde{f}_n - \tilde{f}_n$, which is a perturbation of the original test function $f$. Note that $E_\pi g_n = E_\pi f$, and we have the following variance reduction theorem.

**Theorem 1.2** (Variance reduction.) *In the above setting and with the assumptions in Theorem 1.1, the central limit theorem holds for every $n \geq 1$:*

$$\frac{1}{\sqrt{k}} \sum_{i=0}^{k-1} (g_n(X_i) - \mathbb{E}_\pi(f)) \xrightarrow{\mathrm{d}} \sigma_\pi(g_n)\mathcal{N}(0, 1), \quad k \to \infty,$$

*where* $\sigma_\pi^2(g_n) = \mathrm{Var}_\pi(g_n) + 2\sum_{k=1}^{\infty} \mathbb{E}_\pi\big[(g_n - \mathbb{E}_\pi(f))P_{h,T}^k(g_n - \mathbb{E}_\pi(f))\big] < \infty$. *Moreover,* $\lim_{n\to\infty} \sigma_\pi^2(g_n) = 0$.

The rest of the paper is organized as follows. In Section 2, we first review some basic concepts and notions about Markov chains with general state spaces, and then give some more technical results about the MHMC Markov chains. In Section 3 we formulate the approximation scheme based on the idea of weak approximation following [17], and construct in a specific way $\tilde{f}_n$, which is used in the variance reduction. Section 4 is devoted to the detailed proofs of the main results. The proof of Theorem 1.1 is actually standard through the use of Lyapunov's drift conditions. However, Theorem 1.2 cannot be obtained by a simple combination of [9, 17]. In fact, there is an additional term $|\det \mathbf{J}_{\Phi_x}(\bar{x})|$ in the HMC kernel, which is so complex that it does not satisfy the assumption in [17]. Our proof makes delicate use of the specific approximate construction and moment controls. Section 5 is devoted to two concrete examples, a Bayesian linear inverse problem and a two-component mixture of Gaussians, which satisfy the conditions for variance reduction. In Section 6, we provide some numerical experiments to support our theoretical guarantees. The paper concludes with a discussion in Section 7.

### 1.4. Notation

Denote by $\mathbb{R}^+$ the set of positive real numbers. Denote by $\mathbb{N}^+$ the set of positive integers. Denote by $\|\cdot\|$ the Euclidean norm on $\mathbb{R}^d$. Denote by $\mathcal{B}(\mathbb{R}^d)$ the Borel $\sigma$-field of $\mathbb{R}^d$ and $\mathsf{F}(\mathbb{R}^d)$ the set of all Borel-measurable functions on $\mathbb{R}^d$. Denote by $\mu(\,\cdot\,)$ the Lebesgue measure on $\mathbb{R}^d$. For a Markov kernel $P$, denote by $(X_i)_{i\in\mathbb{N}}$ the corresponding canonical Markov chain, and denote by $(\Omega, \mathcal{F})$ the canonical space. For any probability measure $\nu$ on $\mathbb{R}^d$ and a Markov kernel $P$, denote by $\mathbb{P}_\nu$ the unique probability measure on the canonical space $(\Omega, \mathcal{F})$ such that the canonical process $X_i$ is a Markov chain with kernel $P$ and initial distribution $\nu$. For $f \in \mathsf{F}(\mathbb{R}^d)$, set $\|f\|_\infty = \sup_{x\in\mathbb{R}^d} |f(x)|$. For $f \in \mathsf{F}(\mathbb{R}^d)$ and $V\colon \mathbb{R}^d \to [1, \infty)$, the $V$-norm of $f$ is given by $\|f\|_V = \|f/V\|_\infty$. Suppose $M$ is a $p \times q$ matrix; denote by $M^\top$ and $\det(M)$ the transpose and the determinant of M, respectively. Denote by $C^k(\mathbb{R}^d)$ the set of all $k$-times continuously differentiable functions. Let $f \in C^k(\mathbb{R}^p, \mathbb{R}^q)$ and denote by $\mathbf{J}_f$ the Jacobian matrix of $f$. Let $g \in C^k(\mathbb{R}^p, \mathbb{R})$ and denote by $\nabla g$, $\nabla^2 g$, and $\nabla^3 g$ the first-, second-, and third-order derivatives of $g$, respectively.

## 2. Properties of HMC

In this section we introduce important properties of HMC that will be used later. The reader is also referred to [2, 3, 23] for more details. For the reader's convenience, we briefly review some basic concepts and notions about Markov chains on $\mathbb{R}^d$ in Appendix A.

Suppose that the target density is $\pi(x) \propto \mathrm{e}^{-U(x)}$, with $U$ satisfying Assumptions 1.1–1.3. Thanks to the leapfrog algorithm and Metropolis–Hastings step in the MHMC, the resulting

https://doi.org/10.1017/jpr.2025.10025 Published online by Cambridge University Press

Markov chain $(X_i,\ i \geq 0)$ with kernel $P_{h,T}$ is reversible with respect to $\pi(x)$, and thus $\pi$ is an invariant probability measure [14].

As for the irreducibility and ergodicity of MHMC Markov chains, additional conditions are required. Recently, [9] provided some sufficient conditions for the irreducibility of MHMC Markov chains. Indeed, suppose that Assumption 1.2 holds and that both the step size $h$ and step number $T$ satisfy the inequality

$$\left[1 + \vartheta\left(hL_2^{1/2}\right)\right]^T < 2, \tag{2.1}$$

where $\vartheta(x) = x\left(1 + \frac{1}{2}x + \frac{1}{4}x^2\right)$; then, for all $x \in \mathbb{R}^d$ there exists a $C^1(\mathbb{R}^d, \mathbb{R}^d)$ diffeomorphism $\Psi_x \colon \bar{x} \mapsto \Psi_x(\bar{x})$ such that $x_T = \mathrm{proj}_1 \circ \Phi_{h,T}(x, \Psi_x(x_T))$, where $\mathrm{proj}_1 \colon (x, y) \mapsto x$ denotes coordinate projection and $\Phi_{h,T}$ is defined by (1.4).

Let $\mathbf{J}_{\Psi_x}(\bar{x})$ be the Jacobian matrix of $\Psi_x$ at the point $\bar{x}$. The following lemma is important when proving the main theorems.

**Lemma 2.1.** *If Assumption 1.2 and condition (2.1) hold, then there exists a constant $\kappa(h, T)$ such that, for any $x, \bar{x} \in \mathbb{R}^d$, $D(x, \bar{x}) := |\det \mathbf{J}_{\Psi_x}(\bar{x})| \leq \kappa(h, T)$.*

*Proof.* By the structure of the leapfrog integrator in (1.4), for all $(x_0, v_0) \in \mathbb{R}^d \times \mathbb{R}^d$ and $t \in \{1, 2, \ldots, T\}$, the $t$th iteration $(x_t, v_t) = \Phi_{h,t}(x_0, v_0)$ takes the form

$$x_t = x_0 + thv_0 - \frac{1}{2}th^2\nabla U(x_0) - h^2\sum_{k=1}^{t-1}(t-k)\nabla U(x_k), \tag{2.2}$$

$$v_t = v_0 - \frac{1}{2}h(\nabla U(x_0) + \nabla U(x_t)) - h\sum_{k=1}^{t-1}\nabla U(x_k).$$

Let $\Gamma_{h,t}(x, v) = \sum_{k=1}^{t-1}(t-k)\nabla U(x_k)$. Then, for any $t \geq 1$ and $h > 0$, Assumption 1.2 implies

$$\sup_{x,v,w\in\mathbb{R}^d} \frac{\|\Gamma_{h,t}(x, v) - \Gamma_{h,t}(x, w)\|}{\|v - w\|} \leq \frac{t}{h}(\kappa_{h,t} - 1),$$

where $\kappa_{h,t} = \left(1 + hL_2^{1/2}\vartheta\left(hL_2^{1/2}\right)\right)^t < 2$ [9]. Therefore, by (2.1), we have

$$\sup_{x,v\in\mathbb{R}^d} \frac{h}{t}\left\|\mathbf{J}_{\Gamma_{h,t}}(x, v)\right\| < 1, \tag{2.3}$$

where $\mathbf{J}_{\Gamma_{h,t}}(x, v)$ is the Jacobian matrix of the function $v \mapsto \Gamma_{h,t}(x, v)$.

As a direct consequence, the function $v_0 \mapsto x_t$ defined in (2.2) is a diffeomorphism and has a continuously differentiable inverse $\bar{x} \mapsto \Psi_x(\bar{x})$ for each $x \in \mathbb{R}^d$. In addition, by (2.2) and (2.3), there exists a constant $\kappa_1 \in (0, 1)$ such that, for any $x, v, w \in \mathbb{R}^d$,

$$\begin{aligned}&\|\mathrm{proj}_1 \circ \Phi_{h,T}(x, v) - \mathrm{proj}_1 \circ \Phi_{h,T}(x, w)\| \\ &\qquad = hT\left\|\left(v - \frac{h}{T}\Gamma_{h,T}(x, v)\right) - \left(w - \frac{h}{T}\Gamma_{h,T}(x, w)\right)\right\| \geq hT\kappa_1\|v - w\|,\end{aligned}$$

which in turn implies that there exists a constant $\kappa_2 > 0$ such that, for all $x, v, w \in \mathbb{R}^d$,

$$\|\Psi_x(v) - \Psi_x(w)\| \leq \kappa_2\|v - w\|.$$

https://doi.org/10.1017/jpr.2025.10025 Published online by Cambridge University Press

Therefore, for all $x, \bar{x} \in \mathbb{R}^d$, it follows from Hadamard's inequality that

$$|\det \mathbf{J}_{\Psi_x}(\bar{x})| \le \|\mathbf{J}_{\Psi_x}(\bar{x})\|^d \le \sup_{x,\bar{x}\in\mathbb{R}^d} \|\mathbf{J}_{\Psi_x}(\bar{x})\|^d \le \kappa_2^d =: \kappa(h, T) > 0,$$

where the last inequality is due to [7, Exercise 2.24]. □

**Lemma 2.2.** *If Assumption 1.2 and the condition (2.1) hold, then for any $T \ge 1$, $h \ge 0$, and $(x, v) \in \mathbb{R}^d \times \mathbb{R}^d$, $(x, v) \mapsto (\Phi_{h,T}(x, v), \Psi_x(v), D_{\Psi_x(\cdot)}(v))$ is continuous on $\mathbb{R}^d \times \mathbb{R}^d$.*

*Proof.* Continuity for the mapping $(x, v) \mapsto (\Phi_{h,T}(x, v), \Psi_x(v), D_{\Psi_x(\cdot)}(v))$ was presented in the proof of [9, Theorem 1]. □

**Remark 2.1.** The assumptions of Lemma 2.2 seem slightly different from those in Theorem 1.2. However, by adjusting the value of $M$ to make $h$ small enough, condition (2.1) can be satisfied and thus Lemma 2.2 remains valid.

It now follows from (1.5) that

$$P_{h,T}(x, \mathrm{d}\bar{x}) = \alpha_{\mathrm{H}}(x, \bar{x}) \frac{\exp\{-\|\Psi_x(\bar{x})\|^2/2\}}{(2\pi)^{d/2}} D(x, \bar{x})\, \mathrm{d}\bar{x} + \delta_x(\mathrm{d}\bar{x}) \int_{\mathbb{R}^d} (1 - \alpha(x, v)) \frac{\exp\{-\|v\|^2/2\}}{(2\pi)^{\frac{d}{2}}}\, \mathrm{d}v, \tag{2.4}$$

where $\alpha_{\mathrm{H}}(x, \bar{x}) = \alpha(x, \Psi_x(\bar{x}))$, $D(x, \bar{x})$ was defined in Lemma 2.1.

Furthermore, the Markov kernel $P_{h,T}$ is $\mu$-irreducible, aperiodic, and Harris recurrent, and each compact set is 1-small. As a consequence, for all $x \in \mathbb{R}^d$, $\lim_{k\to\infty} \left\| P_{h,T}^k(x, \cdot) - \pi(\cdot) \right\|_{\mathrm{TV}} = 0$. Note also that if a Markov chain has a unique invariant probability measure, then ergodicity is valid and the ergodic theorem can be obtained by [5, Theorem 5.2.6]. Therefore, by the irreducibility and [5, Theorem 9.2.15], $\pi$ is the unique invariant probability measure with respect to $P_{h,T}$ and hence, for all $f \in L^1(\pi)$ and $\pi$-almost every $x \in \mathbb{R}^d$, we have $\lim_{k\to\infty} (1/k) \sum_{i=0}^{k-1} f(X_i) = \mathbb{E}_\pi(f)$, $P_x$-a.e. This forms the starting point of our further study in this work.

## 3. Approximation scheme for variance reduction

**Definition 3.1.**

(i) Assume there exists a partition $\mathbb{G}$ of $(\mathbb{R}^d, \mu)$ into measurable subsets $G_0, G_1, \ldots, G_m$ such that $\mu(G_i) > 0$ holds for $0 \le i \le m$ and $\bigcup_{i=1}^m G_i$ is bounded. Choose $a_i \in G_i$ for all $0 \le i \le m$ and let $Y = \{a_0, \ldots, a_m\}$. We say that the pair $\mathbb{Y} = (\mathbb{G}, Y)$ is an allotment, with m being the size of $\mathbb{Y}$.

(ii) Let $W\colon \mathbb{R}^d \to [1, \infty)$ be a measurable function. The W-radius and W-mesh of an allotment $\mathbb{Y}$ are respectively defined by $\mathrm{rad}(\mathbb{Y}, W) = \inf_{y\in G_0} W(y)$ and

$$\delta(\mathbb{Y}, W) = \max\left\{ \max_{1\le i\le m} \sup_{y\in G_i} |y - a_i|,\ \max_{0\le i\le m} \sup_{y\in G_i} \left( \frac{W(a_i)}{W(y)} - 1 \right) \right\}.$$

(iii) If there exists a sequence of allotments $(\mathbb{Y}_n, n \ge 0)$ satisfying $\lim_{n\to\infty} \mathrm{rad}(\mathbb{Y}_n, W) = \infty$ and $\lim_{n\to\infty} \delta(\mathbb{Y}_n, W) = 0$, then we say that $\mathbb{Y}_n$ is exhaustive with respect to W.

https://doi.org/10.1017/jpr.2025.10025 Published online by Cambridge University Press

Consider $V_r(x) = \mathrm{e}^{r\|x\|}$, $x \in \mathbb{R}^d$. Obviously, $V_r\colon \mathbb{R}^d \to [1, \infty)$ is a continuous function with bounded sublevel sets, i.e. for every $c \in \mathbb{R}$, the pre-image $V_r^{-1}((-\infty, c])$ is a bounded ball. Proposition A.1 in [17] establishes the existence of an exhaustive sequence with respect to $V_r$. For the sake of completeness, we briefly review the construction of such an allotment here.

Let $r_1 > 1$, and let $(r_n, n \geq 1)$ be an increasing unbounded sequence of positive numbers. For each $n \geq 1$, define sets $L_n = V_r^{-1}((-\infty, r_n])$, $\tilde{L}_n = \{x \in \mathbb{R}^d\colon$ there exists $y \in L_n$ such that $\|x - y\| < \sqrt{d}\}$. The set $\tilde{L}_n$ is trivially a bounded and non-empty closed set. $V_r$ is uniformly continuous on $\tilde{L}_n$. So there exists a decreasing sequence $\varepsilon_n < 1$, vanishing as $n$ tends to infinity, such that $\|x - y\| < \varepsilon_n\sqrt{d}$ implies $|V_r(x) - V_r(y)| < 1/n$ for each $n$ and all $x, y \in \tilde{L}_n$.

Fix $n \geq 1$. For $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ write $K_x^n := \prod_{i=1}^d [x_i, x_i + \varepsilon_n)$. Pick points $x^{(1)}, x^{(2)}, \ldots, x^{(m_n)}$ such that the sets $G_j^n := K_{x^{(j)}}^n$ are disjoint and cover $L_n$. Let $G_0^n$ be the closure of $\mathbb{R}^d \setminus \bigcup_{j=1}^{m_n} K_{x^{(j)}}^n$. Finally, choose $a_0^n \in G_0^n$ such that $V_r(a_0^n) = \inf_{x \in G_0^n} V_r(x)$, and pick $a_j^n \in G_j^n$ arbitrarily. It is now easy to verify that the allotment defined above is exhaustive with respect to $V_r$.

Let $\mathbb{Y}_n = (\mathbb{G}^n, Y^n)$, where $\mathbb{G}^n = (G_i^n,\ 0 \leq i \leq m_n)$ and $Y^n = (a_i^n,\ 0 \leq i \leq m_n)$. We will construct a sequence of approximating solutions $\tilde{f}_n$ based on the exhaustive allotments $\mathbb{Y}_n$. Given the kernel $P_{h,T}$ with $h, T > 0$, define $Q_{ij}^{(n)} = P_{h,T}\big(a_i^n, G_j^n\big)$ and $Q^{(n)} = \big(Q_{ij}^{(n)}\big)_{(m_n+1)\times(m_n+1)}$. Since each part $G_i^n$ has a positive measure the transition matrix $Q^{(n)}$ is well defined, and for every $i, j \in \{0, 1, \ldots, m_n\}$, $Q_{ij}^{(n)} > 0$ by the $\mu$-irreducibility of $P_{h,T}$. Furthermore, $Q^{(n)}$ is irreducible, aperiodic, and recurrent, so there exists a unique invariant probability measure by Markov chain theory on a finite state space. Let $a_n(x) = \sum_{i=0}^{m_n} a_i^n \mathbf{1}_{x \in G_i^n}$, $x \in \mathbb{R}^d$. Then, by the definitions of $V_r$-mesh and exhaustive allotment, for every $x \in \mathbb{R}^d$ we have $\lim_{n\to\infty} a_n(x) = x$. Observe that $V_r(a_n(x))$ can be controlled by $V_r(x)$. Indeed, for all $n \geq 1$ and $x \in \mathbb{R}^d$, we have the useful inequality

$$\begin{aligned} V_r(a_n(x)) &= V_r(x)\left(1 + \frac{V_r(a_n(x)) - V_r(x)}{V_r(x)}\right) \\ &\leq V_r(x)\left(1 + \max_{0\leq i\leq m} \sup_{y \in G_i} \left(\frac{V_r(a_n(x))}{V_r(y)} - 1\right)\right) \leq V_r(x)(1 + \delta_n), \end{aligned} \tag{3.1}$$

where $\delta_n = \delta(\mathbb{Y}_n, V_r)$ denotes the $V_r$-mesh of the allotment $\mathbb{Y}_n$, the first inequality following by the exhaustivity property.

**Lemma 3.1**

(i) *Drift condition. There exist constants $\lambda_v \in (0, 1)$, $b_v > 0$ satisfying*

$$Q^{(n)} V_r(a_i^n) \leq \lambda_v V_r(a_i^n) + b_v \mathbf{1}_{a_i^n \in C} \tag{3.2}$$

*for all $n \geq 1$ and $a_i^n \in Y_n$.*

(ii) *Minorization condition. There exist a compact set $C \subset \mathbb{R}^d$, a constant $\varepsilon \in \mathbb{R}^+$, and a probability measure $\nu_n$ on $(Y_n, \mathcal{P}(Y_n))$ such that*

$$Q_{ij}^{(n)} \geq \varepsilon \nu_n(\{a_j^n\}) \tag{3.3}$$

*for all $n \geq 1$ and $i, j \in \{0, 1, \ldots, m_n\}$ satisfying $a_i^n \in Y_n \cap C$.*

https://doi.org/10.1017/jpr.2025.10025 Published online by Cambridge University Press

(iii) *Strong aperiodicity condition. There exists a constant* $\bar{\varepsilon} \in (0, \infty)$ *such that* $\varepsilon \nu_n(C \cap Y_n) \geq \bar{\varepsilon}$.

The proof of Lemma 3.1 can be found in Appendix B.

Let $\hat{\pi}_n$ be an invariant probability measure of $Q^{(n)}$.

**Proposition 3.1.** *The approximate Markov kernel* $Q^{(n)}$ *is* $V_r$*-uniformly geometrically ergodic, i.e. there exist constants* $M > 0$ *and* $\rho \in (0, 1)$ *such that, for* $k, n \geq 1$,

$$\left\| \left(Q^{(n)}\right)^k (y, \cdot) - \hat{\pi}_n(\cdot) \right\|_{V_r} \leq M V_r(y) \rho^k \quad \text{for all } y \in Y_n.$$

*Proof.* By Lemma 3.1, the existence of $M$ and $\rho$ is a direct consequence of [1]. It is also implied that $M$ and $\rho$ only depend on the parameters $\varepsilon$, $\lambda_v$, $b_v$, and $\bar{\varepsilon}$ in Lemma 3.1. □

We are now in a position to construct the approximation solution $\tilde{f}_n$. To this end, we denote by $\hat{f}_n$ a solution to the Poisson equation $\hat{f}_n(y) - Q^{(n)}\hat{f}_n(y) = f(y) - \mathbb{E}_{\hat{\pi}_n}(f)$, $y \in Y_n$. Having $\hat{f}_n$, the approximating solution $\tilde{f}_n$ is defined by

$$\tilde{f}_n(x) = \sum_{i=0}^{m} \hat{f}_n\left(a_i^n\right) \mathbf{1}_{x \in G_i^n}, \quad x \in \mathbb{R}^d. \tag{3.4}$$

Obviously, the more precisely $Q^{(n)}$ approximates $P_{h,T}$, the more closely $\tilde{f}_n$ approaches the exact solution of the Poisson equation.

## 4. Proofs of main results

Throughout this section, the parameter $r > 0$ is fixed. We start with the proof of Theorem 1.1.

*Proof of Theorem* 1.1. We only prove (i) with $l \in (1, 2)$, since the other cases are similar. Recall that $P_{h,T}$ satisfies the drift condition (1.7) with $V_r(x) = \mathrm{e}^{r\|x\|}$. By [5, Theorem 15.2.4], $P_{h,T}$ is both $V_r$ and $V_{2r}$-uniformly geometrically ergodic, and so it follows by [21, Fact 10] that $V_r$ and $V_{2r}$ are $\pi$-integrable.

In addition, we obviously have the following equivalence relations:

$$\begin{aligned} & P_{h,T} V_r \leq \lambda V_r + b \mathbf{1}_C \\ \Leftrightarrow & P_{h,T} V_r + (1 - \lambda) V_r \leq V_r + b \mathbf{1}_C \\ \Leftrightarrow \quad & P_{h,T} \frac{V_r}{1 - \lambda} + V_r \leq \frac{V_r}{1 - \lambda} + \frac{b}{1 - \lambda} \mathbf{1}_C, \end{aligned}$$

where $C = \{x \colon V_r(x) \leq L_2\}$. Therefore, for every $f$ such that $\|f\|_{V_r} < \infty$, $f$ is $\pi$-integrable, and further the CLT (1.8) holds according to [5, Theorem 21.2.11]. □

Next, we turn to the proof of Theorem 1.2. We first give an upper bound for error variance in terms of the spectral radius of $P_{h,T}$ on $L_0^2(\pi)$.

**Proposition 4.1.** *Denote by* $\rho$ *the spectral radius of* $P_{h,T}|_{L_0^2(\pi)}$. *For any* $g \in L_0^2(\pi)$,

$$\sigma_\pi^2(g) \leq \frac{1 + \rho}{1 - \rho} \mathbb{E}_\pi(g^2).$$

The proof of Proposition 4.1 can be found in Appendix D.

https://doi.org/10.1017/jpr.2025.10025 Published online by Cambridge University Press

For notational simplicity, set $g_n = f + P_{h,T}\tilde{f}_n - \tilde{f}_n$, where $\tilde{f}_n$ is given by (3.4). Trivially, $\mathbb{E}_\pi(g_n) = \mathbb{E}_\pi(f)$, and by Proposition 4.1,

$$\sigma_\pi^2(g_n) \le \frac{1+\rho}{1-\rho}\mathbb{E}_\pi(g_n - \mathbb{E}_\pi(f))^2.$$

Therefore, it suffices to prove that $\mathbb{E}_\pi(g_n - \mathbb{E}_\pi(f))^2 \to 0$ as $n \to \infty$. In turn, it is sufficient to verify the following statements due to the dominated convergence theorem:

- For $\pi$-almost every $x \in \mathbb{R}^d$, $\lim_{n\to\infty} g_n(x) = \mathbb{E}_\pi(f)$.
- The $V_r$-norm is uniformly bounded: $\sup_{n\ge1} \|g_n - \mathbb{E}_\pi(f)\|_{V_r} < \infty$.

For clarity, we restate these statements as Propositions 4.2 and 4.3.

**Proposition 4.2** *For $\pi$-almost every $x \in \mathbb{R}^d$, $\lim_{n\to\infty} g_n(x) = \mathbb{E}_\pi(f)$.*

Observe that for any $x \in \mathbb{R}^d$, $\tilde{f}_n(a_n(x)) - P_{h,T}\tilde{f}_n(a_n(x)) = f(a_n(x)) - \mathbb{E}_{\hat{\pi}_n}(f)$, and so

$$\begin{aligned}|g_n(x) - \mathbb{E}_\pi f| &\le |f(x) - f(a_n(x))| + |\mathbb{E}_{\hat{\pi}_n}(f) - \mathbb{E}_\pi(f)| \\ &\quad + |(P_{h,T}\tilde{f}_n - \tilde{f}_n)(x) - (P_{h,T}\tilde{f}_n - \tilde{f}_n)(a_n(x))| \\ &=: M_n^{(1)}(x) + M_n^{(2)} + M_n^{(3)}(x).\end{aligned}$$

Then it reduces to verifying that $M_n^{(1)}$, $M_n^{(2)}$, and $M_n^{(3)}$ converge $\pi$-a.e. to zero as $n \to \infty$, hence the following three lemmas.

**Lemma 4.1.** $\lim_{n\to\infty} M_n^{(1)}(x) = 0$ *for $\pi$-almost every $x$.*

*Proof.* Note that $f$ is $\pi$-a.e. continuous and $a_n(x) \to x$ as $n \to \infty$. This concludes the proof. □

**Lemma 4.2.** $\lim_{n\to\infty} M_n^{(2)} = 0$.

The proof of Lemma 4.2 can be found in Appendix C. The following lemma offers a uniform bound for the $V_r$-norm of $\tilde{f}_n$.

**Lemma 4.3.** *There exist a constant $\beta > 0$ and a sequence of real numbers $\{b_n\}_{n\in\mathbb{N}}$ such that* $\sup_{n\ge1} \|\tilde{f}_n + b_n\|_{V_r} \le \beta$.

*Proof.* This is a direct consequence of Proposition 3.1 combined with [17, Proposition 3.5]. □

**Lemma 4.4.** $\lim_{n\to\infty} M_n^{(3)}(x) = 0$ *for $\pi$-almost every $x$.*

*Proof.* The form (2.4) and the structure of $P_{h,T}$ together yield

$$\begin{aligned}M_n^{(3)} &= |P_{h,T}\tilde{f}_n(x) - P_{h,T}\tilde{f}_n(a_n(x))| \\ &= \left| \int_{\mathbb{R}^d} (\tilde{f}_n(\bar{x}) - \tilde{f}_n(x))\big(\alpha_{\mathrm{H}}(x,\bar{x})D(x,\bar{x})\mathrm{e}^{-\|\Psi_x(\bar{x})\|^2/2} \right. \\ &\qquad \left. - \alpha_{\mathrm{H}}(a_n(x),\bar{x})D(a_n(x),\bar{x})\mathrm{e}^{-\|\Psi_{a_n(x)}(\bar{x})\|^2/2}\big)\,\mathrm{d}\bar{x} \right|,\end{aligned}$$

where in the first equation we used the fact that $\tilde{f}_n(a_n(x)) = \tilde{f}_n(x)$.

https://doi.org/10.1017/jpr.2025.10025 Published online by Cambridge University Press

Since $(x, \bar{x}) \mapsto (D(x, \bar{x}), \alpha_{\mathrm{H}}(x, \bar{x}))$ is continuous by Lemma 2.2 and, for each $\bar{x} \in \mathbb{R}^d$,

$$\left|\tilde{f}_n(\bar{x}) - \tilde{f}_n(x)\right| \le \left|\tilde{f}_n(\bar{x}) + b_n - (\tilde{f}_n(x) + b_n)\right| \le \beta(V_r(\bar{x}) + V_r(x)), \tag{4.1}$$

where $b_n$ and $\beta$ were given in Lemma 4.3, the integrand above converges to zero.

Analogous to the proof of Lemma 4.2, we obtain, for each $\bar{x} \in \mathbb{R}^d$,

$$\begin{aligned}\left|\alpha_{\mathrm{H}}(x, \bar{x})D(x, \bar{x})\mathrm{e}^{-\|\Psi_x(\bar{x})\|^2/2} - \alpha_{\mathrm{H}}(a_n(x), \bar{x})D(a_n(x), \bar{x})\mathrm{e}^{-\|\Psi_{a_n(x)}(\bar{x})\|^2/2}\right| \\ \le \delta_x \pi(\bar{x}) + \alpha_{\mathrm{H}}(x, \bar{x})k(x, \bar{x}).\end{aligned} \tag{4.2}$$

Obviously, the product of the right-hand sides in (4.1) and (4.2) is integrable with respect to $\mu$ over $\mathbb{R}^d$. Therefore, $M_n^{(3)}$ tends to zero by the dominated convergence theorem. □

**Proposition 4.3.** $\sup_{n\ge1} \|g_n - \mathbb{E}_\pi(f)\|_{V_r} < \infty$.

*Proof.* By the drift condition of $P_{h,T}$, the solution $\hat{f}$ to the Poisson equation (1.9) belongs to $L^\infty_{V_r} = \{f\colon \|f\|_{V_r} < \infty\}$, and it is unique up to a constant [5]. Therefore, by Lemma 4.3, there exist a constant $\beta'$ and a sequence of real numbers $\{b_n\}_{n\ge1}$ such that, for all $n \ge 1$ and $x \in \mathbb{R}^d$, $\left|\tilde{f}_n(x) + b_n - \hat{f}(x)\right| \le \beta' V_r(x)$. Observe that

$$g_n - \mathbb{E}_\pi(f) = P_{h,T}\tilde{f}_n - \tilde{f}_n + f - \pi(f) = P_{h,T}\big(\tilde{f}_n + b_n - \hat{f}\big) - \big(\tilde{f}_n + b_n - \hat{f}\big).$$

By the definition of $P_{h,T}$, it easily follows that, for all $n \ge 1$ and $x \in \mathbb{R}^d$,

$$\begin{aligned}|g_n - \mathbb{E}_\pi(f)| &= \left| \int_{\mathbb{R}^d} \big(\tilde{f}_n + b_n - \hat{f}\big)(\Phi_{h,T}(x, v))\alpha(x, v)\eta(v)\,\mathrm{d}v \right. \\ &\qquad \left. + \big(\tilde{f}_n + b_n - \hat{f}\big)(x) \cdot \left[1 - \int_{\mathbb{R}^d} \alpha(x, v)\,\eta(\mathrm{d}v)\right] - \big(\tilde{f}_n + b_n - \hat{f}\big)(x)\right| \\ &\le \int_{\mathbb{R}^d} \left[\left|\big(\tilde{f}_n + b_n - \hat{f}\big)(\Phi(x, v))\right| + \left|\big(\tilde{f}_n + b_n - \hat{f}\big)(x)\right|\right]\alpha(x, v)\eta(v)\,\mathrm{d}v \\ &\le \int_{\mathbb{R}^d} \beta' V_r(\Phi_{h,T}(x, v))\alpha(x, v)\,\eta(\mathrm{d}v) + \beta' V_r \int_{\mathbb{R}^d} \alpha(x, v)\eta(v)\,\mathrm{d}v \\ &\le \beta'(P_{h,T}V_r(x) + V_r(x)) \le (1 + \lambda + b)\beta' V_r(x),\end{aligned}$$

which together with the definition of the $V_r$-norm implies the desired result. □

*Proof of Theorem* 1.2. Combining Propositions 4.1, 4.2, and 4.3 concludes the proof. □

## 5. Applications

In this section we aim to show how to verify the assumptions on the target distributions given in Section 1.3 through two concrete models.

### 5.1. Bayesian linear inverse problem

Let $\beta \in \left(\frac{1}{2}, 1\right)$, $\lambda_1, \lambda_2, \delta > 0$, and $\mathbf{A}$ a $p \times d$ matrix. Consider the linear model $b = \mathbf{A}X + \varepsilon$, where $X$ has a prior distribution $\pi_X(x)$ given by

https://doi.org/10.1017/jpr.2025.10025 Published online by Cambridge University Press

$$\pi_X(x) \propto \exp\left(-\lambda_1\left(x^\top x+\delta\right)^\beta - \frac{\lambda_2}{2}x^\top x\right), \quad x\in\mathbb{R}^d,$$

and $\varepsilon \sim N(0, I_d)$.

The so-called Bayesian linear problem is to infer $X$ from the observations $b$. In fact, the density of the posterior distribution of interest us given by $\pi(x) \propto \exp(-U(x))$, where

$$U(x) = -\frac{1}{2}x^\top\left(\mathbf{A}^\top\mathbf{A}+\lambda_2 I_d\right)x - \lambda_1\left(x^\top x+\delta\right)^\beta + \langle b, \mathbf{A}x\rangle.$$

The exponential integrator version of MALA (EI-MALA) was devised to generate a Markov chain with invariant distribution $\pi$. The Wasserstein convergence rate of a class of EI-MALA algorithms was studied at length in [8], and the CLT for the corresponding Markov chains established in [13].

We will confirm that the potential $U$ satisfies Assumptions 1.1 and 1.2, and so both the CLT and the variance reduction theorem are applicable for the MHMC Markov chains as well. It is apparent that $U(x)$ can be decomposed as

$$U(x) = U_1(x) + U_2(x) = \left[\frac{1}{2}x^\top\left(\mathbf{A}^\top\mathbf{A}+\lambda_2 I_d\right)x\right] + \left[\lambda_1\left(x^\top x+\delta\right)^\beta - \langle b, \mathbf{A}x\rangle\right]. \tag{5.1}$$

[9] offers a sufficient condition for Assumptions 1.1 and 1.2. The following is a slightly stronger version.

**Lemma 5.1.** *Fix $l\in(1, 2]$. Suppose that the potential $U(x)$ can be decomposed as $U(x) = U_1(x) + U_2(x)$, where $U_1(x)$ and $U_2(x)$ satisfy the following conditions:*

(i) *$U_1$ and $U_2$ belong to $C^3(\mathbb{R}^d)$.*

(ii) *For all $k\geq 1$ and $x\in\mathbb{R}^d$, $U_1(kx) = k^l U_1(x)$ and $\{y\in\mathbb{R}^d\colon U_1(y)\leq U_1(x)\}$ is a convex set.*

(iii) *$\lim_{\|x\|\to\infty} U_1(x) = \infty$.*

(iv) *For $k = 2, 3$,*

$$\lim_{\|x\|\to\infty}\frac{\|\nabla^k U_2(x)\|}{\|x\|^{l-k}} = 0. \tag{5.2}$$

*Then the potential $U$ satisfies Assumptions 1.1 and 1.2.*

**Proposition 5.1.** *The potential $U$ given by (5.1) satisfies Assumptions 1.1 and 1.2.*

The proof of Proposition 5.1 can be found in Appendix E.

### 5.2. Two-component mixture of Gaussians

Consider the sampling from a mixture of Gaussians with two components

$$\pi \sim p\mathcal{N}(\mu, I_d) + (1-p)\mathcal{N}(\nu, I_d), \tag{5.3}$$

where $\mathcal{N}(\cdot, I_d)$ is a $d$-dimensional normal distribution with identity variance.

Without loss of generality, assume $p = \mathrm{e}^c/(\mathrm{e}^c+\mathrm{e}^{-c})$ for some $c\in\mathbb{R}$, and let $U_1(x) = \frac{1}{2}\|x\|^2$, $U_2(x) = -\log\left(\mathrm{e}^{c-\|\mu\|^2/2}\mathrm{e}^{-\langle x,\mu\rangle} + \mathrm{e}^{-c-\|\nu\|^2/2}\mathrm{e}^{-\langle x,\nu\rangle}\right)$. Then

https://doi.org/10.1017/jpr.2025.10025 Published online by Cambridge University Press

$$\begin{aligned}\pi(x) &= \frac{1}{(2\pi)^{d/2}}\left[p\exp\left\{-\frac{1}{2}\|x-\mu\|^2\right\} + (1-p)\exp\left\{-\frac{1}{2}\|x-\nu\|^2\right\}\right] \\ &= \frac{1}{(2\pi)^{d/2}(\mathrm{e}^c+\mathrm{e}^{-c})}\mathrm{e}^{-\|x\|^2/2}\left[\mathrm{e}^{c-\|\mu\|^2/2}\mathrm{e}^{-\langle x,\mu\rangle} + \mathrm{e}^{-c-\|\nu\|^2/2}\mathrm{e}^{-\langle x,\nu\rangle}\right] \\ &\propto \mathrm{e}^{-U(x)} = \mathrm{e}^{-(U_1(x)+U_2(x))}.\end{aligned}$$

While there is a huge literature about this model, we mention only the recent work [15] related to ours. Note that the potential function $U(x)$ is Lipschitz smooth but not strongly convex in $\mathbb{R}^d$. In fact, it is double-well and strongly convex only when $|x| \geq r$ for a sufficiently large $r$. MALA and the classical expectation-maximization (EM) optimization algorithm were compared in [15], which claimed that MALA sampling can be faster than the EM algorithm. We will examine Assumption 1.3 and verify the conditions (1.6a)–(1.6c), so that we can apply MHMC sampling to the double-well potential case.

**Proposition 5.2.** *Consider a two-component Gaussian mixture model as in (5.3). Then Assumptions 1.3 hold.*

The proof of Proposition 5.2 can be found in Appendix F.

**Remark 5.1.** In fact, this result can be extended to the case where $\pi \sim p\mathcal{N}(\mu, \Sigma) + (1-p)\mathcal{N}(\nu, \Sigma)$ for some symmetric positive definite matrix $\Sigma$. We can find that the linear terms $\langle \mu, x\rangle$, $\langle \nu, x\rangle$ in $U_2$ are replaced by $\langle \Sigma^{-1}\mu, x\rangle$, $\langle \Sigma^{-1}\nu, x\rangle$, respectively, and the bounds of $\|\nabla U_2\|$ and $\|\nabla^2 U_2\|$ can be obtained similarly.

**Remark 5.2.** By the approximation scheme, the allotment only needs to be exhaustive with respect to the function $V_r(x) = \exp(r\|x\|)$ for some $r$ such that $f(x) \leq V_r(x)$ for all $x \in \mathbb{R}^d$. Consequently, the specific allotment construction depends solely on the test function $f$. Suppose a constant $r$ is such that $f(x) \leq V_r(x)$ for all $x$; then we could construct an allotment according to Section 3. In particular, let $r_n = \mathrm{e}^n$ and $\varepsilon_n = (n^2\sqrt{d} + nrd)^{-1}$. This gives $L_n = \{x \in \mathbb{R}^d : \|x\| \leq n/r\}$ and $\tilde{L}_n = \{x \in \mathbb{R}^d : \|x\| \leq \sqrt{d} + n/r\}$. Following the construction of $G_j^n$ in Section 3, we can generate an allotment such that $\|x-y\| < \varepsilon_n\sqrt{d}$ implies $|V_r(x) - V_r(y)| < n^{-1}$ for each $n$ and all $x, y \in \tilde{L}_n$, by Lagrange's theorem. By [17, Proposition A.1], such an allotment forms an exhaustive sequence with respect to $V_r$.

## 6. Simulations

In this section we explain how to implement the MHMC algorithm for variance reduction through two simple examples. In fact, we have to tackle the following two issues in practice:

- The stochastic matrix $Q = (Q_{ij})_{(m+1)\times(m+1)}$ cannot be computed analytically.
- Once the approximate solution $\tilde{f}$ has been computed, the function $P_{h,T}\tilde{f}$, and thus the control variate $P_{h,T}\tilde{f} - \tilde{f}$, are not accessible in closed form.

Construct a partition $\{G_0, G_1, \ldots, G_m\}$ of $\mathbb{R}^d$ in such a way that the probability of $\pi(G_0)$ is small. We assume that it is easy to sample $d$-dimensional normal random points. Let $a_i \in G_i$ for $i > 0$ be arbitrary and choose $a_0$ on the boundary of $G_0$.

Let $Y = \{a_0, a_1, \ldots, a_m\}$, $\mathbb{G} = \{G_0, G_1, \ldots, G_m\}$. This induces an allotment $(Y, \mathbb{G})$ in $\mathbb{R}^d$. Recall that $Q_{ij} = P_{h,T}(a_i, G_j)$, where the transition kernel $P_{h,T}$ was given by (1.5). As the

https://doi.org/10.1017/jpr.2025.10025 Published online by Cambridge University Press

precise computation of each entry is not feasible, we construct an estimate of $Q_{ij}$ via another i.i.d. Monte Carlo. With this in mind, let $Z_1, Z_2, \ldots, Z_N$ be i.i.d. samples from $\mathcal{N}(0, I_d)$, where $N$ is sufficiently large. Define

$$\hat{Q}(a_i, a_j) := \begin{cases} \frac{1}{N}\sum_{k=1}^{N} \mathbf{1}_{Z_k \in G_j} \alpha(a_i, Z_k) & \text{if } j \neq i, \\ 1 - \sum_{k \in \{0,1,\ldots,m\}\setminus i} \hat{Q}(a_i, a_k) & \text{if } j = i. \end{cases} \tag{6.1}$$

The approximating transition matrix $Q = (Q_{ij})$ is now well estimated by $\hat{Q} = (\hat{Q}_{ij})$ by the law of large numbers, so we use $\hat{Q}$ in the MHMC algorithm instead of $Q$.

Turning to the second issue, given a $\pi$-integrable function $f$, we aim to estimate $\mathbb{E}_\pi(f)$ by $(1/k)\sum_{i=0}^{k-1}(f + P_{h,T}\tilde{f} - \tilde{f})(X_i)$, where $(X_i,\ i \geq 0)$ is a Markov chain generated by the MHMC algorithm. However, $P_{h,T}\tilde{f}$ is not accessible in a closed form once again. Similarly to (6.1), define, for any $x \in \mathbb{R}^d$,

$$\hat{Q}(x, a_j) := \begin{cases} \frac{1}{N}\sum_{k=1}^{N} \mathbf{1}_{Z_k \in G_j} \alpha(a_i, Z_k) & \text{if } j \neq i(x), \\ 1 - \sum_{k \in \{0,1,\ldots,m\}\setminus i} \hat{Q}(x, a_k) & \text{if } j = i(x), \end{cases} \tag{6.2}$$

where $i(x)$ is the unique index $i \in \{0, 1, \ldots, m\}$ such that $x \in G_{i(x)}$; then define, for any $x \in \mathbb{R}^d$, $\hat{Q}\tilde{f}(x) = \sum_{j=0}^{m} \hat{f}(a_j)\hat{Q}(x, a_j)$. Finally, we use $\hat{Q}\tilde{f}$ in place of $P_{h,T}\tilde{f}$. We remark that the term $(1/k)\sum_{i=0}^{k-1}(f + \hat{Q}\tilde{f} - \tilde{f})(X_i)$ is unbiased to some extent; the interested reader is referred to [17, Remark 5.1] for details.

In the next subsection we provide a simple example to illustrate the efficiency of variance reduction in MHMC through the weak approximation scheme.

> It is not obvious [how] to do so in a way which substantially decreases the variance of the simulation without substantially increasing the complexity of the simulations, and for which optimal values of the parameters can be approximated with a reasonable numerical cost. [12, p. 60].

### 6.1. Gaussian mixture distribution

Consider a one-dimensional Gaussian mixture distribution as mentioned in Section 5.2. Let $\mu = 3$, $\nu = -2$, and $p = 0.3$. Choose $f(x) = x^2$ as the force function. For any $m \geq 1$, $l > 0$, let $G_0^{m,l} = \mathbb{R}\setminus[-l, l]$ and $G_i^{m,l}$, $i = 1, 2, \ldots, m$, be intervals of length $2l/m$ partitioning $[-l, l]$. The step size and length are chosen to be $h = 0.1$ and $T = 20$. The construction of $\hat{Q}$ is derived from (6.1) using $N = 10^7$, and the estimator $(1/k)\sum_{i=0}^{k-1}(f + \hat{Q}\tilde{f} - \tilde{f})(X_i)$ is then derived from (6.2) using $N = 10^5$ and the approximating scheme. We study three allotments, where the parameters take values as shown in Table 1. We evaluate the associated quantities after 6000 iterations. The results are shown in Figure 1 for different iterations and different estimations from 50 replications. The red curve corresponds to conventional MHMC estimates, and the blue curve corresponds to the modified estimators $(1/k)\sum_{i=0}^{k-1}(f + \hat{Q}\tilde{f} - \tilde{f})(X_i)$.

We observe that the mean square errors (MSE) of the modified estimator are consistently lower than those of the original MHMC estimator. Moreover, as the allotments become

https://doi.org/10.1017/jpr.2025.10025 Published online by Cambridge University Press

TABLE 1. The parameter setup for different allotments.

| Allotment index | $l$ | $m$ |
|---|---|---|
| 1 | 2 | 18 |
| 2 | 4 | 15 |
| 3 | 6 | 30 |

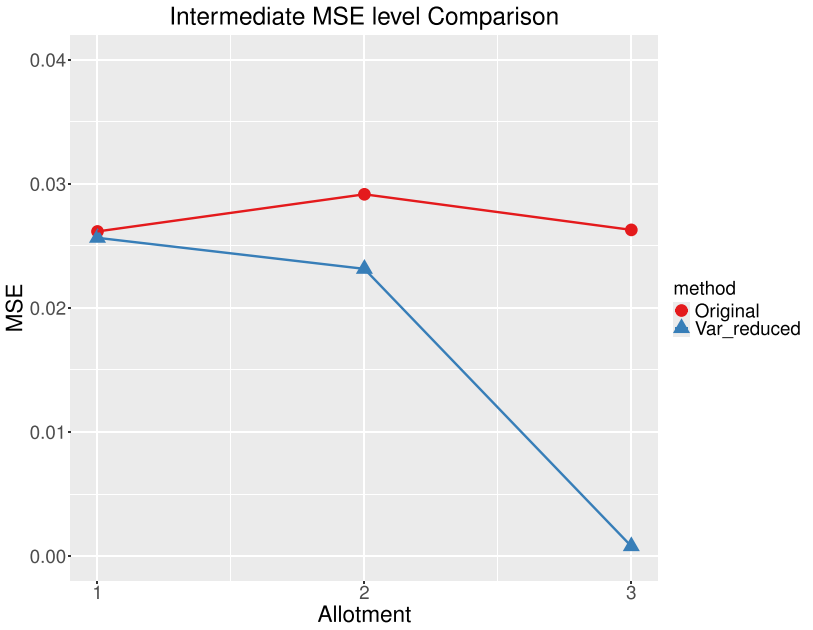


FIGURE 1. The estimates for $\mathbb{E}_\pi(f)$ using different methods with different allotments.

denser, the MSE decreases further. This result suggests that the variance reduction technique is effective and demonstrates that achieving improved convergence in sampling does not require excessively fine allotments. These findings are encouraging and point to the potential for applying this approach to a wider range of models, particularly those in high-dimensional settings.

## 7. Discussion

There are two interesting recent works in the literature: [17], in which the authors develop an approximation scheme for a solution of the Poisson equations of a geometrically ergodic Metropolis–Hastings chain, and construct a sequence of control-variate estimators to decrease the error variance in the mean estimate; and [9], in which the authors discussed the irreducibility and geometric ergodicity of the popular Hamiltonian Monte Carlo algorithm in a rigorous mathematical sense. Motivated by these two works, we have established the CLT and variance reduction theorem under mild regular conditions for the MHMC algorithm, which implies that the MHMC algorithm can be applied in a wide range of fields.

https://doi.org/10.1017/jpr.2025.10025 Published online by Cambridge University Press

In addition, we offer several concrete examples satisfying the regular conditions, which suggests that our results can be applied to a broad class of potentials. The simulation results imply that the method works well in practice.

Some questions arise from our work. The simulations suggest that the approximation scheme takes effect even if $m$ is not so large, though it is desirable to obtain non-asymptotic results for the convergence of variance reduction, like in [17, Section 4]. It seems difficult to analyze the convergence rate when using the diffeomorphism mentioned in (2.4), and alternative ways are needed. On the other hand, when implementing the approximation scheme, we use standard Monte Carlo (6.1) to approximate the kernel $Q$, which may affect the performance of the scheme. Lastly, an interesting question is how to execute an efficient implementation for the approximation scheme. We leave these for future work.

## Appendix A. Markov chains on $\mathbb{R}^d$

For the reader's convenience, we briefly review some basic concepts and notions about Markov chains on $\mathbb{R}^d$ in this subsection. [5, 16] are good classic books in this field.

Let $(\mathbb{R}^d, \|\cdot\|)$ be a standard Euclidean space equipped with its Borel $\sigma$-field $\mathcal{B}$ and Lebesgue measure $\mu$. Consider a time-homogeneous Markov chain $(X_i,\ i \geq 0)$ with transition kernel $P$, i.e. $P(x, A) = \mathbb{P}(X_i \in A \mid X_{i-1} = x)$ for all $i \geq 1$, $x \in \mathbb{R}^d$, and $A \in \mathcal{B}$. For each $\sigma$-finite measure $\pi$ on $(\mathbb{R}^d, \mathcal{B})$, define, for a measurable set $A$, $\pi P(A) = \int_{x\in\mathbb{R}^d} \pi(\mathrm{d}x)\, P(x, A)$, and, for a measurable function $f\colon \mathbb{R}^d \to \mathbb{R}$, $Pf(x) = \int_{y\in\mathbb{R}^d} f(y)\, P(x, \mathrm{d}y)$. If $\pi P = \pi$ then we say $\pi$ is an invariant measure for the kernel $P$. If $\pi(\mathrm{d}x)\, P(x, \mathrm{d}y) = \pi(\mathrm{d}y)\, P(y, \mathrm{d}x)$, then $\pi$ is reversible with respect to the kernel $P$. It is well known that reversibility implies the existence of invariant probability.

The Markov kernel $P$ is $\pi$-irreducible if, for every $x \in \mathbb{R}^d$ and $A \in \mathcal{B}$ with $\pi(A) > 0$, there exists $k \geq 1$ such that $P^k(x, A) > 0$. A set $C \in \mathcal{B}$ is small if there exist $k \geq 1$ and a non-zero measure $\nu$ such that, for every $x \in C$, $P^k(x, \cdot) \geq \nu(\cdot)$. In particular, a set $C$ is usually referred to as 1-small if $k = 1$. With the concept of a small set, irreducibility admits another interpretation. A Markov chain is called irreducible if there exists a small set $C$ such that, for some $k_0 \geq 1$, $P^{k_0}(x, C) > 0$ for all $x \in \mathbb{R}^d$. Note that if $P$ is $\pi$-irreducible, then it is also irreducible since $\mathcal{B}$ is countably generated. A set satisfying the last inequality is sometimes called accessible.

For an accessible small set $C \in \mathcal{X}$, its period is defined by

$$d_C = \gcd\left\{k \geq 1;\ \inf_{x\in C} P^k(x, C) > 0\right\}.$$

Note that the set on the right-hand side is well defined, and all accessible small sets have a common period, say $\kappa$. A Markov kernel $P$ is called aperiodic if $\kappa = 1$. One of the remarkable properties of aperiodicity is that if $P$ is aperiodic then, for each $x \in \mathbb{R}^d$ and each accessible set $A \in \mathcal{B}$, there exists $N = N(x, A) \geq 1$ such that $P^k(x, A) > 0$ for all $k \geq N$.

A Markov chain with stationary distribution $\pi$ is called *Harris recurrent* if, for any $C \in \mathcal{B}$ with $\pi(C) > 0$ and any $x \in \mathbb{R}^d$, the chain eventually reaches $C$ from $x$ with probability 1, i.e. $\mathbb{P}(\text{there exists } n\colon X_n \in C \mid X_0 = x) = 1$.

Total variation distance is often used to measure the distance between distributions. Let $\nu_1$, $\nu_2$ be two probability measures; their total variation distance is defined by

$$\|\nu_1 - \nu_2\|_{\mathrm{TV}} = \sup_{A\in\mathcal{X}} |\nu_1(A) - \nu_2(A)| \equiv \frac{1}{2} \sup_{|f|\leq 1} \left| \int f\, \mathrm{d}\nu_1 - \int f\, \mathrm{d}\nu_2 \right|.$$

https://doi.org/10.1017/jpr.2025.10025 Published online by Cambridge University Press

Given a function $V\colon \mathbb{R}^d \to \mathbb{R}^+$, the $V$-norm for a function $f$ is defined by

$$\|f\|_V = \sup_{x\in\mathbb{R}^d} \frac{|f(x)|}{V(x)},$$

and the $V$-distance between $\nu_1$ and $\nu_2$ is given by

$$\|\nu_1 - \nu_2\|_V = \frac{1}{2} \sup_{\|f\|_V\le 1} \left| \int f \,\mathrm{d}\nu_1 - \int f \,\mathrm{d}\nu_2 \right|.$$

For an MCMC algorithm with the target distribution $\pi$, we are eager to know whether the distribution of $X_k$ is sufficiently close to $\pi$ as the iteration proceeds. The quantity commonly studied is $\|P^k(x,\cdot) - \pi(\cdot)\|_{\mathrm{TV}}$ and the like.

It is well known that if the kernel $P$ is irreducible and aperiodic then, for $\pi$-almost every $x \in \mathbb{R}^d$, $\lim_{k\to\infty} \|P^k(x,\cdot) - \pi(\cdot)\|_{\mathrm{TV}} = 0$. In addition, under some extra conditions the rate of convergence in this would be geometric. We say the kernel $P$ is geometrically ergodic if there is a positive constant $\rho < 1$ such that, for $\pi$-almost every $x \in \mathbb{R}^d$, there exists a finite number $\kappa_x$ with $\|P^n(x,\cdot) - \pi(\cdot)\|_{\mathrm{TV}} \le \kappa_x \rho^n$. The concept of *geometrically ergodic* has been intensively studied by much literature; see, e.g., [5, 16, 21]. If there exist a $\pi$-a.e. finite measurable function $V\colon \mathbb{R}^d \to [1,\infty]$, a small set $C \in \mathcal{B}$, and constants $\lambda \in (0,1)$, $b > 0$ such that $PV(x) \le \lambda V(x) + b\mathbf{1}_{x\in C}$, we say that $P$ satisfies the Lyapunov drift condition $(V, \lambda, b, C)$.

It is well known that the drift condition is equivalent to geometrical ergodicity when the kernel $P$ is irreducible, aperiodic, and admits an invariant probability. And the converse is also true [11].

Geometrical ergodicity plays an important role in the study of Markov chains. In particular, together with some regular conditions it guarantees that the central limit theorem holds, which is our concern in the present paper.

## Appendix B. Proof of Lemma 3.1

*Proof.* (i) We borrow some technical tricks from [17, Proposition 3.3]. For any fixed $n \ge 1$, we extend $a_n$ from $\mathbb{R}^d$ to $\mathbb{R}^d \times \mathbb{R}^d$, still denoted by $a_n$, where $a_n(x,v) = a_n(x)$. By (3.1), it follows that, for each $i = 0, 1, \ldots, m_n$,

$$\begin{aligned}
&Q^{(n)}V_r(a_i^n)\\
&= \int_{\mathbb{R}^d} V_r\big(a_n\big(\Phi_{h,T}\big(a_i^n, v\big)\big)\big)\eta(v)\alpha\big(a_i^n, v\big)\,\mathrm{d}v + V_r\big(a_i^n\big)\left(1 - \int_{\mathbb{R}^d} \eta(v)\alpha\big(a_i^n, v\big)\,\mathrm{d}v\right)\\
&\le (1+\delta_n)\int_{\mathbb{R}^d} V_r\big(\Phi_{h,T}\big(a_i^n, v\big)\big)\eta(v)\alpha\big(a_i^n, v\big)\,\mathrm{d}v + V_r\big(a_i^n\big)\left(1 - \int_{\mathbb{R}^d} \eta(v)\alpha\big(a_i^n, v\big)\,\mathrm{d}v\right)\\
&= P_{h,T}V_r\big(a_i^n\big) + \delta_n \int_{\mathbb{R}^d} V_r\big(\Phi_{h,T}\big(a_i^n, v\big)\big)\eta(v)\alpha\big(a_i^n, v\big)\,\mathrm{d}v \le (1+\delta_n)P_{h,T}V_r\big(a_i^n\big).
\end{aligned}$$

By the drift condition (1.7), there exist $\lambda_v \in [0,1)$, $b_v \in (0,+\infty)$, and a compact set $C$ such that $Q^{(n)}V_r\big(a_i^n\big) \le \big(1+\delta_n\big)P_{h,T}V_r\big(a_i^n\big) \le (1+\delta_n)\lambda_v V_r\big(a_i^n\big) + (1+\delta_n)b_v\mathbf{1}_{a_i^n\in C}$. Since $\lim_{n\to\infty}\delta_n = 0$, there exists $n_0 \ge 1$ such that, for all $n > n_0$, $(1+\delta_n)\lambda_v < \frac{1}{2}(1+\lambda_v) < 1$. Set

https://doi.org/10.1017/jpr.2025.10025 Published online by Cambridge University Press

$\lambda_v = \frac{1}{2}(1+\lambda_v)$ and $b_v = (1+\sup_n \delta_n)b_v$. If $N_0 > 1$ then we enlarge $C$, keeping the new set compact, and increase $b_v$, still denoted by $b_v$, such that (3.2) holds for $n \le N_0$. Therefore, we obtain a compact set $C$, $\lambda_v \in [0, 1)$, and $b_v > 0$ such that the drift condition (3.2) holds for all $n \ge 1$.

(ii) Let $C$ be the compact set constructed above; it suffices to establish (3.3). Since $C$ is compact, there exists $R_0 < \infty$ such that $C \subseteq B(0, R_0)$. It follows from [9, Theorem 12] that the ball $B(0, R_0)$ is 1-small for $P_{h,T}$. Precisely, there exist constants $(L_0, v_0, s_0) \in \mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^+$ such that, for each $i, j0, 1, \ldots, m_n\}$,

$$\begin{aligned}\big(Q^{(n)}\big)_{ij} &= P_{h,T}\big(a_i^n, G_j^n\big)\\ &\ge L_0 \min_{(x,v)\in B(0,R_n)\times B(v_0,s_0)} \{\alpha(x,v)\eta(v)\}\mu\big(G_j^n \cap B(0,R_0)\big)\\ &=: \varepsilon'\mu\big(G_j^n \cap B(0,R_0)\big)\\ &= \varepsilon'\mu(B(0,R_0))\frac{\mu\big(G_j^n \cap B(0,R_0)\big)}{\mu(B(0,R_0))} =: \varepsilon \nu_n\big(\{a_j^n\}\big),\end{aligned}$$

where $\varepsilon = \varepsilon'\mu(B(0, R_0))$ and $\nu_n\big(\{a_j^n\}\big) = \mu\big(G_j^n \cap B(0, R_0)\big)/\mu(B(0, R_0))$. The fact that $\varepsilon' > 0$ results from the continuity of $\alpha(x, v)$ and $\eta(v)$ by Lemma 2.2.

(iii) Set $\delta^* = \sup_{k\ge1} \delta_k$ and let $E \subseteq \mathbb{R}^d$ be an open set of radius $r_{\mathrm{E}} > \delta^*$. Recall that $V_r(x) = \mathrm{e}^{r\|x\|}$; then $r_n := \mathrm{rad}(\mathbb{Y}_n, V_r) \to \infty$ as $n \to \infty$, so there exists $n_0 \ge 1$ such that $E \subseteq \bigcap_{n\ge n_0} V_r^{-1}([1, r_n))$. Then we enlarge the set $C$ such that

$$C \supseteq \Big(\bigcup_{n<n_0} \mathbb{R}^d \backslash J_0^n\Big) \cup \bigcap_{n\ge n_0} V_r^{-1}([1, r_n)). \tag{B.1}$$

It is clear that the right-hand side of (B.1) is bounded, so we may, and do, assume that $C$ is compact, and thus the drift condition and minorization condition remain valid. Assume further that $R_0$ in $\nu_n$ is so large that $B(0, R_0)$ still contains $C$. It suffices to estimate $\nu_n\big(C \cap Y_n\big)$.

For $n < n_0$, note that $C \cap Y_n \supseteq \big\{a_i^n : i = 1, 2, \ldots, m_n\big\}$, since $C \supseteq \big(G_0^n\big)^{\mathrm{c}}$ by (B.1). Hence,

$$\nu_n(C \cap Y_n) \ge \nu_n\Big(\bigcup_{i=1}^{m_n} \{a_i^n\}\Big) = \frac{\mu\big(\big(G_0^n\big)^{\mathrm{c}} \cap B(0,R_0)\big)}{\mu(B(0,R_0))} = \frac{\mu\big(\big(G_0^n\big)^{\mathrm{c}}\big)}{\mu(B(0,R_0))} > 0, \tag{B.2}$$

where we used the fact that $B(0, R_0) \supseteq C \supseteq \big(G_0^n\big)^{\mathrm{c}}$.

For $n \ge n_0$, let $E$' be an open ball of radius $\frac{1}{2}r_{\mathrm{E}}$ centered at the center of $E$. Since $r_n = \mathrm{rad}(\mathbb{Y}_n, V_r) = \inf_{y\in G_0^n} V_r(y)$, we have $E \cap G_0^n \subseteq V_r^{-1}([1, r_n)) \cap V_r^{-1}([r_n, \infty)) = \varnothing$, which implies that $E \cap G_0^n = \varnothing$. Note that $|y - a_n(y)| \le \delta^* < \frac{1}{2}r_{\mathrm{E}}$ for any point $y \in E'$, so $a_n(y) \in E \subseteq C$. Therefore, $E' \subseteq \bigcup_{i:a_i^n\in C} G_i^n$ and

https://doi.org/10.1017/jpr.2025.10025 Published online by Cambridge University Press

$$\nu_n(C \cap Y_n) = \frac{\mu\big(\bigcup_{i:a_i^n \in C} G_i^n \cap B(0, R_0)\big)}{\mu(B(0, R_0))} \geq \frac{\mu(E' \cap B(0, R_0))}{\mu(B(0, R_0))} = \frac{\mu(E')}{\mu(B(0, R_0))} > 0, \tag{B.3}$$

where the last inequality results from $B(0, R_0) \supseteq C \supseteq E \supseteq E'$.

Therefore, if we set

$$\bar{\varepsilon} = \frac{\min\big\{\mu(E'),\ \min_{n<n_0} \mu\big((G_0^n)^{\mathrm{c}}\big)\big\}}{\varepsilon \mu(B(0, R_0))},$$

then $\bar{\varepsilon} > 0$ and $\varepsilon \nu_n(C \cap Y_n) \geq \bar{\varepsilon}$ by (B.2) and (B.3). The proof is complete. □

## Appendix C. Proof of Lemma 4.2

*Proof.* The proof basically follows the same line as [17, Proposition 3.7], with suitable modification. First, define a new approximate Markov chain on $Y_n$ whose transition matrix and invariant measure are denoted by $Q^{(n)*}$ and $\pi_n^*$, respectively. Precisely, define, for $i, j \in \{0, 1, \ldots, m_n\}$, $\pi_n^*\big(\{a_i^n\}\big) = \pi\big(G_i^n\big)$,

$$\big(Q^{(n)*}\big)_{ij} = \mathbb{P}_\pi\big(X_1 \in G_j^n \mid X_0 \in G_i^n\big) = \frac{1}{\pi\big(G_i^n\big)} \int_{G_i^n} \pi(x) P_{h,T}\big(x, G_j^n\big)\, \mathrm{d}x,$$

$$h_n\big(a_i^n\big) = \frac{1}{\pi\big(G_i^n\big)} \int_{G_i^n} \pi(x) f(x)\, \mathrm{d}x.$$

Then we estimate $|\mathbb{E}_{\hat{\pi}_n}(f) - \mathbb{E}_\pi(f)|$ as follows:

$$|\mathbb{E}_{\hat{\pi}_n}(f) - \mathbb{E}_\pi(f)| \leq |\mathbb{E}_{\hat{\pi}_n}(f) - \mathbb{E}_{\pi_n^*}(f)| + |\mathbb{E}_{\pi_n^*}(f) - \mathbb{E}_{\pi_n^*}(h_n)|, \tag{C.1}$$

where we used the fact that $\mathbb{E}_{\pi_n^*}(h_n) = \mathbb{E}_\pi(f)$.

Note that $|\mathbb{E}_{\hat{\pi}_n}(f) - \mathbb{E}_{\pi_n^*}(f)| \leq \|f\|_{V_r} \|\pi_n^* - \hat{\pi}_n\|_{V_r} \leq (M/(1-\rho)) \|f\|_{V_r} \|\pi_n^* - \pi_n^* Q^{(n)}\|_{V_r}$. We will show that $\|\pi_n^* - \pi_n^* Q^{(n)}\|_{V_r}$ converges to zero. Let $g\colon Y_n \to \mathbb{R}$ be a function such that $\|g\|_{V_r} \leq 1$; we obtain, by the definition of the $V$-norm,

$$\begin{aligned}
\big(\pi_n^* - \pi_n^* Q^{(n)}\big) g &= \pi_n^*\big(Q^{(n)*} - Q^{(n)}\big) g \\
&= \sum_{j=0}^{m_n} \left( \sum_{i=0}^{m_n} \pi_n^*\big(a_i^n\big)\big[Q_{ij}^{(n)*} - Q_{ij}^{(n)}\big] g\big(a_j^n\big) \right) \\
&= \sum_{j=0}^{m_n} \left( \sum_{i=0}^{m_n} \pi\big(G_i^n\big) \left[ \int_{G_i^n} \frac{\pi(x)}{\pi\big(G_i^n\big)} P_{h,T}\big(x, G_j^n\big)\, \mathrm{d}x - P_{h,T}\big(a_i^n, G_j^n\big) \right] \right) g\big(a_j^n\big) \\
&= \sum_{j=0}^{m_n} \left( \int_{\mathbb{R}^d} \pi(x) \big[P_{h,T}\big(x, G_j^n\big) - P_{h,T}\big(a_n(x), G_j^n\big)\big]\, \mathrm{d}x \right) g\big(a_j^n\big).
\end{aligned}$$

https://doi.org/10.1017/jpr.2025.10025 Published online by Cambridge University Press

By the equivalence relation in (2.4), it further follows that

$$\begin{aligned}
&\sum_{j=0}^{m_n}\Big(\int_{\mathbb{R}^d}\pi(x)\big[P_{h,T}\big(x,G_j^n\big)-P_{h,T}\big(a_n(x),G_j^n\big)\big]\,\mathrm{d}x\Big)g\big(a_j^n\big)\\
&=\sum_{j=0}^{m_n}\int_{\mathbb{R}^d}\pi(x)g(a_n(x))\,\mathrm{d}x\int_{G_j^n}[\alpha(a_n(x),v)-\alpha(x,v)]\eta(v)\,\mathrm{d}v\\
&\quad+\sum_{j=0}^{m_n}\int_{\mathbb{R}^d}\pi(x)\,\mathrm{d}x\int_{G_j^n}g(a_n(\bar{x}))\{\alpha_{\mathrm{H}}(x,\bar{x})\bar{\eta}(x,\bar{x})-\alpha_{\mathrm{H}}(a_n(x),\bar{x})\bar{\eta}(a_n(x),\bar{x})\}\,\mathrm{d}\bar{x}\\
&=\int_{\mathbb{R}^d}\pi(x)\,\mathrm{d}x\int_{\mathbb{R}^d}g(a_n(x))\{\alpha(a_n(x),v)-\alpha(x,v)\}\eta(v)\,\mathrm{d}v\\
&\quad+\int_{\mathbb{R}^d}\pi(x)\,\mathrm{d}x\int_{\mathbb{R}^d}g(a_n(\bar{x}))\{\alpha_{\mathrm{H}}(x,\bar{x})\bar{\eta}(x,\bar{x})-\alpha_{\mathrm{H}}(a_n(x),\bar{x})\bar{\eta}(a_n(x),\bar{x})\}\,\mathrm{d}\bar{x}=:I_n^{(1)}+I_n^{(2)},
\end{aligned}$$

where $\bar{\eta}(x,\bar{x})=(2\pi)^{-d/2}\exp\big\{-\frac{1}{2}\|\Psi_x(\bar{x})\|^2\big\}D(x,\bar{x})$. Then it is sufficient to show that both $I_n^{(1)}$ and $I_n^{(2)}$ tend to zero.

Note that the integrand of $I_n^{(1)}$ is dominated by a $\pi$-integrable entity. Indeed, for every $x\in\mathbb{R}^d$, we have

$$\begin{aligned}
&\Big|\int_{\mathbb{R}^d}g(a_n(x))\{\alpha(a_n(x),v)-\alpha(x,v)\}\eta(v)\,\mathrm{d}v\Big|\\
&\quad\le\int_{\mathbb{R}^d}V_r(a_n(x))|\alpha(a_n(x),v)-\alpha(x,v)|\eta(v)\,\mathrm{d}v\\
&\quad\le\big(1+\sup_{n\ge1}\delta_n\big)V_r(x)\int_{\mathbb{R}^d}|\alpha(a_n(x),v)-\alpha(x,v)|\eta(v)\,\mathrm{d}v,
\end{aligned}\tag{C.2}$$

which is $\pi$-integrable. Also, since the function $(x,v)\mapsto\alpha(x,v)$ is continuous on $\mathbb{R}^d$, the right-hand side of (C.2) converges to zero since $\lim_{n\to\infty}a_n(x)=x$. Thus, by the dominated convergence theorem, we have $\lim_{n\to\infty}I_n^{(1)}=0$.

Turning to $I_n^{(2)}$, set $Z_n(x,\bar{x})=\alpha_{\mathrm{H}}(x,\bar{x})\bar{\eta}(x,\bar{x})\alpha_{\mathrm{H}}(a_n(x),\bar{x})\bar{\eta}\big(a_n(x),\bar{x}\big)$. Using the structure of $P_{h,T}$ and (3.1), we get, for every $x\in\mathbb{R}^d$,

$$\begin{aligned}
&\Big|\int_{\mathbb{R}^d}g(a_n(\bar{x}))\{\alpha_{\mathrm{H}}(x,\bar{x})\bar{\eta}(x,\bar{x})-\alpha_{\mathrm{H}}(a_n(x),\bar{x})\bar{\eta}(a_n(x),\bar{x})\}\,\mathrm{d}\bar{x}\Big|\\
&\quad\le\int_{\mathbb{R}^d}V_r(a_n(\bar{x}))|Z_n(x,\bar{x})|\,\mathrm{d}\bar{x}\\
&\quad\le\big(1+\sup_{n\ge1}\delta_n\big)\int_{\mathbb{R}^d}V_r(\bar{x})|Z_n(x,\bar{x})|\,\mathrm{d}\bar{x}\\
&\quad\le\big(1+\sup_{n\ge1}\delta_n\big)(P_{h,T}V_r(x)+P_{h,T}V_r(a_n(x)))\\
&\quad\le\big(1+\sup_{n\ge1}\delta_n\big)\big(2+\sup_{n\ge1}\delta_n\big)(\lambda+b)V_r(x).
\end{aligned}\tag{C.3}$$

https://doi.org/10.1017/jpr.2025.10025 Published online by Cambridge University Press

Fix $x \in \mathbb{R}^d$. Since $Z_n(x, \bar{x})$ is continuous on $\mathbb{R}^d \times \mathbb{R}^d$ by Lemma 2.2, it obviously follows that

$$\lim_{n\to\infty} V_r(\bar{x})|Z_n(x, \bar{x})| = 0, \quad \mu\text{-a.e.} \tag{C.4}$$

On the other hand, we claim that there exists a constant $\kappa_x$ such that

$$\alpha_{\mathrm{H}}(a_n(x), \bar{x})\bar{\eta}(a_n(x), \bar{x}) \le \pi(\bar{x})\kappa_x \quad \text{for all } \bar{x} \in \mathbb{R}^d. \tag{C.5}$$

Indeed, by the structure of the Metropolis–Hastings algorithm, we have

$$\begin{aligned}\alpha_{\mathrm{H}}(a_n(x), \bar{x})\bar{\eta}(a_n(x), \bar{x}) &\le \frac{1}{(2\pi)^{d/2}} D(a_n(x), \bar{x}) \exp\{U(a_n(x)) - U(\bar{x})\} \\ &\quad \times \exp\left\{-\frac{1}{2}\|\mathrm{proj}_2 \circ \Phi_{h,T}(a_n(x), \Psi_{a_n(x)}(\bar{x}))\|^2\right\} \\ &\le \pi(\bar{x})\frac{C'}{\pi(a_n(x))} \le \pi(\bar{x})\kappa_x,\end{aligned}$$

where $\mathrm{proj}_2(x, v) := v$ and $\kappa_x := C'\big(\inf_{n\ge1} \pi(a_n(x))\big)^{-1}$ for all $x, v \in \mathbb{R}^d$.

Note that $D(a_n(x), \bar{x})$ is bounded by Lemma 2.1. Also, it follows from the continuity of $\pi$ and the definition of $\delta_n$ that $\pi(a_n(x)) \ge \inf\{\pi(y) : y \in B(x, \sup_{n\ge1} \delta_n)\} > 0$ for all sufficiently large $n$. Thus, we have $0 < \kappa_x < \infty$, and so (C.5) holds true.

Then we can show that, for the fixed $x$,

$$V_r(\bar{x})|Z_n(x, \bar{x})| \le V_r(\bar{x})[\pi(\bar{x})\delta_x + \alpha_{\mathrm{H}}(x, \bar{x})\bar{\eta}(x, \bar{x})] \in L^1(\mu). \tag{C.6}$$

Combining (C.3), (C.4), and (C.6) implies that $I_n^{(2)} \to 0$.

Consequently, there exists a constant $\kappa_0 > 0$ such that

$$|\mathbb{E}_{\hat{\pi}_n}(f) - \mathbb{E}_{\pi_n^*}(f)| \le \kappa_0\big(I_n^{(1)} + I_n^{(2)}\big) \to 0 \quad \text{as } n \to \infty. \tag{C.7}$$

Finally, we verify that the second term on the right-hand side of (C.1) tends to zero. By the definition of $h_n$ and $\pi_n^*$, we obtain

$$\begin{aligned}|\mathbb{E}_{\pi_n^*}(f) - \mathbb{E}_{\pi_n^*}(h_n)| &= \sum_{i=0}^{m_n} \pi\big(G_i^n\big)\big(f_n\big(a_i^n\big) - h_n\big(a_i^n\big)\big) \\ &= \sum_{i=0}^{m_n} \pi\big(G_i^n\big)\left[f\big(a_i^n\big) - \frac{1}{\pi\big(G_i^n\big)}\int_{G_i^n} \pi(x)f(x)\,\mathrm{d}x\right] \\ &= \sum_{i=0}^{m_n} \int_{G_i^n} \big[f\big(a_i^n\big) - f(x)\big]\pi(x)\,\mathrm{d}x = \int_{\mathbb{R}^d} (f(a_n(x)) - f(x))\pi(x)\,\mathrm{d}x.\end{aligned}$$

Note that $f$ is $\pi$-a.e. continuous and thus $f(a_n(x)) - f(x)$ $\pi$-a.e. converges to zero in $\mathbb{R}^d$. In addition, it is easy to see that

$$|f(a_n(x)) - f(x)| \le \|f\|_{V_r}\big(2 + \sup_{k\ge1} \delta_k\big)V_r(x) \in L^1(\pi).$$

Therefore, the dominated convergence theorem implies that

$$\lim_{n\to\infty} |\mathbb{E}_{\pi_n^*}(f) - \mathbb{E}_{\pi_n^*}(h_n)| = 0. \tag{C.8}$$

We conclude the proof by inserting (C.7) and (C.8) into (C.1). □

https://doi.org/10.1017/jpr.2025.10025 Published online by Cambridge University Press

## Appendix D. Proof of Proposition 4.1

*Proof.* Note that $\pi$ is reversible with respect to the Markov kernel $P_{h,T}$. By the spectral theory in [5, Chapter 22] we obtain

$$\sigma_\pi^2(g) = \int_S \frac{1+t}{1-t}\,\zeta_g(\mathrm{d}t),$$

where $S = \mathrm{Spec}(P_{h,T}|_{L_0^2(\pi)})$ and $\zeta_g$ is the spectral measure associated with $g$.

Since $P_{h,T}$ is $V_r$-uniformly geometrically ergodic, the spectral theory implies that $P_{h,T}$ has an absolute $L^2(\pi)$-spectral gap, i.e. the spectral radius $\rho < 1$. Therefore, we obtain

$$\sigma_\pi^2(g) \le \frac{1+\rho}{1-\rho}\int_S \zeta_g(\mathrm{d}t) = \frac{1+\rho}{1-\rho}\mathbb{E}_\pi(g^2),$$

where we used the fact that $\zeta_g(S) = \|g\|^2_{L_0^2(\pi)}$. The proof is complete. □

## Appendix E. Proof of Proposition 5.1

*Proof.* By direct calculation, we easily find that $U_1$ belongs to $C^3(\mathbb{R}^d)$ and satisfies the conditions in Lemma 5.1. It is now sufficient to show that $U_2$ belongs to $C^3(\mathbb{R}^d)$ and admits the property (5.2). Note that $U_2(x) = \lambda_1\big(x^\top x + \delta\big)^\beta + \langle b, \mathbf{A}x\rangle$; then some simple calculus yields

$$\begin{aligned}
\frac{\partial U_2}{\partial x_i} &= 2\lambda_1\beta\big(x^\top x+\delta\big)^{\beta-1}x_i - \big(\mathbf{A}^\top b\big)_i,\\
\frac{\partial^2 U_2}{\partial x_i\partial x_j} &= 2\lambda_1\beta\Big[\big(x^\top x+\delta\big)^{\beta-1}\delta_{ij} + 2x_ix_j(\beta-1)\big(x^\top x+\delta\big)^{\beta-2}\Big],\\
\frac{\partial^3 U_2}{\partial x_i\partial x_j\partial x_k} &= 2\lambda_1\beta\Big[2x_k(\beta-1)\big(x^\top x+\delta\big)^{\beta-2}\delta_{ij} + 2(\beta-1)\big(x^\top x+\delta\big)^{\beta-2}(x_j\delta_{ik}+x_i\delta_{jk})\\
&\qquad + 4x_ix_jx_k\beta_1\beta_2\big(x^\top x+\delta\big)^{\beta-3}\Big],
\end{aligned}$$

where $\delta_{ij}$ is the Kronecker delta function.

Obviously, all the partial derivatives are continuous and so $U_2 \in C^3(\mathbb{R}^d)$. Turning to study $\nabla^2 U_2(x)$, by the equivalence between norms we only consider the Hilbert–Schmidt (HS) norm of $\nabla^2 U_2(x)$ to obtain

$$\begin{aligned}
&\|\nabla^2 U_2(x)\|^2_{\mathrm{HS}}\\
&\le 4\lambda_1^2\beta^2\sum_{i,j}\big[(\|x\|^2+\delta)^{\beta-1}\delta_{ij} + 2x_ix_j(\beta-1)(\|x\|^2+\delta)^{\beta-2}\big]^2\\
&= C\big[d(\|x\|^2+\delta)^{2(\beta-1)} + 4(\beta-1)^2\|x\|^4(\|x\|^2+\delta)^{2(\beta-2)} + 4(\beta-1)\|x\|^2(\|x\|^2+\delta)^{2\beta-3}\big],
\end{aligned}$$

where $C$ is a numeric constant.

Since $\beta \in \big(\frac{1}{2}, 1\big)$ and $\delta > 0$, as $\|x\| \to \infty$ we have

$$\|\nabla^2 U_2(x)\| \le \big(\|\nabla^2 U_2(x)\|^2_{\mathrm{HS}}\big)^{1/2} \to 0. \tag{E.1}$$

https://doi.org/10.1017/jpr.2025.10025 Published online by Cambridge University Press

As for the term $\|\nabla^3 U_2(x)\|$, we obtain, by a similar argument,

$$\begin{aligned}\|\nabla^3 U_2(x)\|_{\mathrm{HS}}^2 &\le 4\lambda_1^2\beta^2 \sum_{i,j,k}\Big[2(\beta-1)\big(x^\top x+\delta\big)^{\beta-2}(x_k\delta_{ij}+x_j\delta_{ik}+x_i\delta_{jk})\\ &\quad + 4x_ix_jx_k(\beta-1)(\beta-2)\big(x^\top x+\delta\big)^{\beta-3}\Big]^2\\ &= C(\beta-1)^2\big[(6d+3d^2)(\|x\|^2+\delta)^{2\beta-4}\|x\|^2+4(\beta-2)^2(\|x\|^2+\delta)^{2\beta-6}\|x\|^6\\ &\quad + 12(\beta-2)(\|x\|^2+\delta)^{2\beta-5}\|x\|^4\big],\end{aligned}$$

which in turn implies, as $\|x\|\to\infty$,

$$\|\nabla^3 U_2(x)\|\,\|x\| \le \|\nabla^3 U_2(x)\|_{\mathrm{HS}}\|x\| \to 0. \tag{E.2}$$

Thus, combining (E.1) and (E.2) proves the condition (5.2). □

## Appendix F. Proof of Proposition 5.2

*Proof.* The function $|U_2(x)|$ is obviously continuous on $\mathbb{R}^d$. Also, note that

$$|U_2(x)| \le \log\big[\mathrm{e}^{c-\|\mu\|^2/2}\mathrm{e}^{-\langle x,\mu\rangle}+\mathrm{e}^{-c-\|\nu\|^2/2}\mathrm{e}^{-\langle x,\nu\rangle}\big].$$

Thus, for any $\gamma>1$,

$$\lim_{\|x\|\to\infty}\frac{|U_2(x)|}{(1+\|x\|)^\gamma}=0.$$

Hence, there exists $A_1>0$ such that

$$\sup_{x\in\mathbb{R}^d}\frac{|U_2(x)|}{(1+\|x\|)^\gamma}\le A_1. \tag{F.1}$$

Namely, condition (1.6a) is verified.

Next, we turn to $\|\nabla U_2(x)\|$. Note that

$$\nabla U_2(x)=\frac{\mathrm{e}^{c-\|\mu\|^2/2}\mathrm{e}^{-\langle x,\mu\rangle}\mu+\mathrm{e}^{-c-\|\nu\|^2/2}\mathrm{e}^{-\langle x,\nu\rangle}\nu}{\mathrm{e}^{c-\|\mu\|^2/2}\mathrm{e}^{-\langle x,\mu\rangle}+\mathrm{e}^{-c-\|\nu\|^2/2}\mathrm{e}^{-\langle x,\nu\rangle}}.$$

Then $\|\nabla U_2(x)\|\le\|\mu\|+\|\nu\|$. Therefore, for any $\gamma$ in (F.1), there exists $A_2>0$ such that

$$\sup_{x\in\mathbb{R}^d}\frac{\|\nabla U_2(x)\|}{(1+\|x\|)^{\gamma-1}}\le A_2,$$

which implies (1.6b).

It remains to prove (1.6c). Some simple calculations yield

$$\begin{aligned}\|\nabla^2 U_2(x)\|_{\mathrm{HS}} &= \frac{\exp\big\{\langle\mu+\nu,x\rangle-\frac{1}{2}(\|\mu\|^2+\|\nu\|^2)\big\}}{\big(\exp\big\{c-\frac{1}{2}\|\mu\|^2\big\}\exp\{-\langle x,\mu\rangle\}+\exp\big\{-c-\frac{1}{2}\|\nu\|^2\big\}\exp\{-\langle x,\nu\rangle\}\big)^2}\|\mu-\nu\|^2\\ &\le \frac{1}{2}\|\mu-\nu\|^2<\infty.\end{aligned}$$

Hence, by the mean value theorem there exists a constant $A_3>0$ such that, for any $x,y\in\mathbb{R}^d$,

$$\|\nabla U_2(x)-\nabla U_2(y)\|\le A_3\|x-y\|.$$

Set $L_4:=\max\{A_1,A_2,A_3\}$. Conditions (1.6a)–(1.6c) are now verified. □

https://doi.org/10.1017/jpr.2025.10025 Published online by Cambridge University Press

## Acknowledgements

The authors would like to express their gratitude to the anonymous referees for their careful reading and constructive comments.

## Funding information

This study is supported by National Natural Science Foundation of China grants (12271475, 11871425) and Fundamental Research Funds for Central Universities grant (2020XZZX002-03).

## Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

## References

[1] Baxendale, P. H. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Prob.* **15**, 700–738.
[2] Betancourt, M. (2018). A conceptual introduction to Hamiltonian Monte Carlo. Preprint, arXiv:1701.02434.
[3] Bou-Rabee, N. and Sanz-Serna, J. M. (2018). Geometric integrators and the Hamiltonian Monte Carlo method. *Acta Numerica* **27**, 113–206.
[4] Brooks, S., Gelman, A., Jones, G. and Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*, 1st edn. Chapman and Hall/CRC, New York.
[5] Douc, R., Moulines, E., Priouret, P. and Soulier, P. (2018). *Markov Chains*, 1st edn. Springer, Cham.
[6] Duane, S., Kennedy, A. D., Pendleton, B. J. and Roweth, D. (1987). Hybrid Monte Carlo. *Phys. Lett. B* **195**, 216–222.
[7] Duistermaat, J. J. and Kolk, J. A. (2004). *Multidimensional Real Analysis I: Differentiation*, 1st edn. Cambridge University Press.
[8] Durmus, A. and Moulines, E. (2015). Quantitative bounds of convergence for geometrically ergodic Markov chain in the Wasserstein distance with application to the Metropolis-adjusted Langevin algorithm. *Statist. Comput.* **25**, 5–19.
[9] Durmus, A., Moulines, E. and Saksman, E. (2020). Irreducibility and geometric ergodicity of Hamiltonian Monte Carlo. *Ann. Statist.* **48**, 3545–3564.
[10] Eberle, A. and Majka, M. B. (2019). Quantitative contraction rates for Markov chains on general state spaces. *Electron. J. Prob.* **24**, 1–36.
[11] Gallegos-Herrada, M. A., Ledvinka, D. and Rosenthal, J. S. (2024). Equivalences of geometric ergodicity of Markov chains. *J. Theoret. Prob.* **37**, 1230–1256.
[12] Graham, C. and Talay, D. (2013). *Stochastic Simulation and Monte Carlo Methods: Mathematical Foundations of Stochastic Simulation*, 1st edn. Springer, Berlin.
[13] Jin, R. and Tan, X. (2020). Central limit theorems for Markov chains based on their convergence rates in Wasserstein distance. Preprint, arXiv:2002.09427.
[14] Kamatani, K. and Song, X. (2021). Haar–Weave–Metropolis kernel. Preprint, arXiv:2111.06148.
[15] Ma, Y., Chen, Y., Jin, C., Flammarion, N. and Jordan, M. (2019). Sampling can be faster than optimization. *Proc. Nat. Acad. Sci. USA* **116**, 20881–20885.
[16] Meyn, S. and Tweedie, R. L. (2009) *Markov Chains and Stochastic Stability*, 2nd edn. Cambridge University Press.
[17] Mijatovic, A. and Vogrinc, J. (2018). On the Poisson equation for Metropolis–Hastings chains. *Bernoulli* **24**, 2401–2428.
[18] Pillai, N. S., Stuart, A. M. and Thiéry, A. H.(2012). Optimal scaling and diffusion limits for the Langevin algorithm in high dimensions. *Ann. Appl. Prob.* **22**, 2320–2356.
[19] Roberts, G. O. and Gilks, A. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Ann. Appl. Prob.* **7**, 110–120.
[20] Roberts, G. O. and Rosenthal, J. S. (1998). Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Statist. Soc. B* **60**, 255–268.
[21] Roberts, G. O. and Rosenthal, J. S. (2004). General state space Markov chains and MCMC algorithms. *Prob. Surv.* **1**, 20–71.

[22] ROBERTS, G. O. AND TWEEDIE, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2**, 341–363.
[23] VISHNOI, N. K. (2021). An introduction to Hamiltonian Monte Carlo method for sampling. Preprint, arXiv:2108.12107.
[24] XIFARA, T., SHERLOCK, C., LIVINGSTONE, S., BYRNE, S. AND GIROLAMI, M. (2014). Langevin diffusions and the Metropolis-adjusted Langevin algorithm. *Statist. Prob. Lett.* **91**, 14–19.

https://doi.org/10.1017/jpr.2025.10025 Published online by Cambridge University Press