Bernoulli **24**(3), 2018, 2401–2428
https://doi.org/10.3150/17-BEJ932

On the Poisson equation for Metropolis–Hastings chains

ALEKSANDAR MIJATOVIĆ[1] and JURE VOGRINC[2]

[1]*Department of Mathematics, King's College London, London WC2R 2LS, United Kingdom.*
E-mail: aleksandar.mijatovic@kcl.ac.uk
[2]*Department of Mathematics, Imperial College London, London SW7 2AZ, United Kingdom.*
E-mail: j.vogrinc13@imerial.ac.uk

This paper defines an approximation scheme for a solution of the Poisson equation of a geometrically ergodic Metropolis–Hastings chain Φ. The scheme is based on the idea of weak approximation and gives rise to a natural sequence of control variates for the ergodic average $S_k(F) = (1/k)\sum_{i=1}^{k} F(\Phi_i)$, where F is the force function in the Poisson equation. The main results show that the sequence of the asymptotic variances (in the CLTs for the control-variate estimators) converges to zero and give a rate of this convergence. Numerical examples in the case of a double-well potential are discussed.

Keywords: asymptotic variance; central limit theorem; Markov chain Monte Carlo; Metropolis–Hastings algorithm; Poisson equation for Markov chains; variance reduction; weak approximation

1. Introduction

A Central Limit Theorem (CLT) for an ergodic average $S_k(F) = \frac{1}{k}\sum_{i=1}^{k} F(\Psi_i)$ of a Markov chain $(\Psi_k)_{k\in\mathbb{N}}$, evolving according to a transition kernel $\mathcal{P}$ on a general state space $\mathcal{X}$, is well known to be intimately linked with the solution $\hat{F}$ of the *Poisson equation*

$$\hat{F} - \mathcal{P}\hat{F} = F - \pi(F) \qquad (\mathrm{PE}(\mathcal{P}, F))$$

with a *force function* $F\colon \mathcal{X} \to \mathbb{R}$ (see [20], Section 17.4). Here π is the invariant probability measure of Ψ on $\mathcal{X}$, $\pi(F) = \int_{\mathcal{X}} F(x)\pi(dx)$ and $\mathcal{P}G(x) = \mathbb{E}_x[G(\Psi_1)]$ for any $G : \mathcal{X} \to \mathbb{R}$. In fact, the Poisson equation in (PE($\mathcal{P}$, F)) is of fundamental importance in many areas of probability, statistics and engineering (see [20], Section 17.7, p. 459). In this context, one of the main motivations for constructing approximations to $\hat{F}$ is to reduce the asymptotic variance in (CLT(Ψ, F)) for the Markov Chain Monte Carlo (MCMC) estimators, thus speeding up the MCMC algorithms.

Assume that the random sequence $(S_k(F))_{k\in\mathbb{N}}$ satisfies the strong law of large numbers (SLLN), $\lim_{k\to\infty} S_k(F) = \pi(F)$ a.s., and the CLT

$$\sqrt{k}\big(S_k(F) - \pi(F)\big) \xrightarrow{d} \sigma_F \cdot N(0,1) \qquad \text{as } k \to \infty, \qquad (\mathrm{CLT}(\Psi, F))$$

where $N(0, 1)$ is a standard normal distribution and the constant σ_F^2 is the *asymptotic variance*. Put differently, the variance of $S_k(F)$ is approximately equal to σ_F^2/k. It is hence intuitively clear that if σ_F^2 is large, which occurs in applications particularly when F has super-linear growth (as $\sigma_F^2 \propto \mathrm{Var}_\pi(F)$, see, for example, [26], Section 5, and the references therein), the variance of $S_k(F)$ will also be big, requiring a large number of steps k for convergence. In contrast, imagine we knew the solution $\hat{F}$ of the Poisson equation (PE($\mathcal{P}, F$)) and could evaluate the function $\mathcal{P}\hat{F} - \hat{F}$. Then the estimator given by the ergodic average $S_k(F + \mathcal{P}\hat{F} - \hat{F})$ (for any $k \in \mathbb{N}$) would be equal to the constant $\pi(F)$ for any (not necessarily stationary) path of the chain Ψ, i.e. its variance vanishes for π-a.e. starting point. However, solving Poisson's equation for the chains arising in most applications, even for very simple functions F, is for all practical purposes impossible (see, e.g., relevant comments in [8]). Nevertheless, this line of reasoning suggests the following heuristic:

a good approximation $\tilde{F}$ to a solution of (PE($\mathcal{P}, F$)) *significantly reduces the asymptotic variance in the* (CLT($\Psi, F + \mathcal{P}\tilde{F} - \tilde{F}$)), *i.e.* $\sigma^2_{F+\mathcal{P}\tilde{F}-\tilde{F}} \ll \sigma_F^2$.

This heuristic is well known and strongly substantiated with numerical evidence. As a method of variance reduction, it has been developed in various Markovian settings [1,8–10]. Its applications in stochastic networks theory are described in [19], Chapter 11, while applications in statistics for the random scan Gibbs sampler were developed in [3]. However, schemes for constructing $\tilde{F}$ found in the literature (a) depend strongly on the structure of the underlying model and, to the best of our knowledge, (b) there are no theoretical results quantifying *a priori* the amount of reduction in the asymptotic variance of CLT($\Psi, F + \mathcal{P}\tilde{F} - \tilde{F}$). This paper aims to address both (a) and (b) by introducing a general *Scheme* (see below) for constructing an approximate solution $\tilde{F}$ to (PE($\mathcal{P}, F$)), applicable to any discrete time Markov chain, and analysing the corresponding asymptotic variance in the setting of Metropolis–Hastings chains.

Our main result (Theorem 2.6 below) states that, for an appropriately chosen allotment $\mathbb{X}$, the function $\tilde{F}_{\mathbb{X}}$ can theoretically achieve an arbitrary reduction of the asymptotic variance for a class of Metropolis–Hastings chains and force functions F that satisfy natural growth conditions. To the best of our knowledge, this is the first systematic approach capable of reducing the asymptotic variance arbitrarily for a general class of discrete-time Markov chains. The proof hinges on the uniform convergence to stationarity of a sequence of approximating Markov chains, which in turn crucially depends on the results in [2,21] (see Section 3.1 below for details). Step (II) in the

Scheme

Input: Transition kernel $\mathcal{P}$, function F, allotment $\mathbb{X} = (\mathbb{J}, X)$ consisting of a partition $\mathbb{J} = \{J_0, J_1, \dots, J_m\}$ of $\mathcal{X}$ and representatives $X = \{a_j \in J_j : j = 0, 1, \dots, m\}$.

begin

(I) Define $p_{\mathbb{X}} \in \mathbb{R}^{(m+1)\times(m+1)}$ and $f_{\mathbb{X}}: \{a_0, a_1, \dots a_m\} \to \mathbb{R}$ respectively by

$$(p_{\mathbb{X}})_{ij} := \mathcal{P}(a_i, J_j) \quad \text{and} \quad f_{\mathbb{X}}(a_j) := F(a_j), \qquad \text{where } i, j \in \{0, 1, \dots, m\}.$$

(II) Find a solution $\hat{f}_{\mathbb{X}}$ of Poisson's equation (PE($p_{\mathbb{X}}, f_{\mathbb{X}}$)).

(III) Define $\tilde{F}_{\mathbb{X}} := \sum_{j=1}^m \hat{f}_{\mathbb{X}}(a_j) 1_{J_j}$.

end

Output: Approximate solution $\tilde{F}_{\mathbb{X}} : \mathcal{X} \to \mathbb{R}$ to Poisson's equation in (PE($\mathcal{P}, F$)).

Scheme amounts to solving a linear system and can be carried out provided that the stochastic matrix $p_{\mathbb{X}}$ is irreducible. Moreover, Poisson's equation (PE($p_{\mathbb{X}}$, $f_{\mathbb{X}}$)) has a solution that is unique up to the addition of a constant function (see [16], Theorem 9.3). Furthermore, the asymptotic variance in CLT(Ψ, $F + \mathcal{P}\tilde{F}_{\mathbb{X}} - \tilde{F}_{\mathbb{X}}$) does not depend on the choice of $\hat{f}_{\mathbb{X}}$ in step (II) of the *Scheme*.

The approximation *Scheme* exploits the stochastic evolution implicitly present in (PE($\mathcal{P}$, F)). As in [9,10,19], we are using the true solution of the Poisson equation for a related Markov process to construct $\tilde{F}$. In our context, the approximation of $\hat{F}$ is based on the *weak approximation* of the chain Ψ by a sequence of "simpler" finite state Markov chains (converging in law to Ψ), such that the solutions of the Poisson equations for the approximating chains can be characterised algebraically. The approximating Markov chain underpinning the *Scheme* mimics the behaviour of Ψ as follows: its state space is a partition $\{J_0, J_1, \ldots, J_m\}$ of the state space $\mathcal{X}$ and its transition matrix consists of the probabilities of Ψ jumping from a chosen element in J_i into the set J_j. Analogous weak approximation ideas have been applied in continuous time to Brownian motion [22], Lévy [24] and Feller [23] processes. A recent interesting approach for approximating the solution of Poisson's equation in discrete time has been proposed in [4]. The idea is to solve the equation obtained by differentiating both sides of (PE($\mathcal{P}$, F)) in the state variable. This leads to a new approximation method for $\hat{F}$ but appears to require smoothness properties of the transition kernel, not afforded by the class of Metropolis–Hastings chains.

The approximation of a given Markov chain with a finite-state chain given by the *Scheme* is akin to others previously mentioned in the literature that are also based on a partition or a covering of the state space, see for instance [12,29,30] and [15]. These papers relate the speed of convergence to equilibrium of the initial and of the approximating Markov chains. They do not however address potential similarity of Poisson's equations.

Theorem 2.6 is theoretical in nature as the partition in $\mathbb{X}$ that provably reduces the variance below a prescribed level typically requires a large number of approximating states m. However, Example 5.2.2 in Section 5.2 below demonstrates numerically that in the case of a Random walk Metropolis chain converging to a double-well potential, the *Scheme* applied with only $m = 6$ points reduces the variance by approximately 10% (see Section 5 below for details).

A natural question arising from Theorem 2.6 is about the rate of the decay of the sequence of asymptotic variances $\sigma_n^2 \to 0$. Theorem 4.1 shows that the decay is governed by the greater of the two quantities: the mesh of the partition of the bounded set $\mathbb{R}^d \setminus J_0^n$ and the π-average of the squared drift function of the chain over J_0^n (see Section 2 for definitions). Furthermore, for the chains studied in [13,27], Theorem 4.1 implies a bound on the rate of decay in terms of the target density π alone (see Proposition 4.3 below). We hope this result is both of some practical value (cf. Section 5.2.1) and independent interest.

The reminder of the paper is organised as follows. Section 2 formulates our main result (Theorem 2.6). In Section 3, we prove Theorem 2.6. The structure of the proof is given in Section 3.1, while Sections 3.2, 3.3 and 3.4 carry out the steps. In Section 4, we state and prove Theorem 4.1 and Proposition 4.3, bounding the rate of convergence to zero of the asymptotic variances. Section 5 describes the implementation of the *Scheme* (Section 5.1) and gives numerical examples (Section 5.2).

2. Assumptions and the main result

Let π be a density function of a probability measure on $\mathbb{R}^d$ with respect to the Lebesgue measure μ^{Leb} and let $q\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a transition density function, that is, for every $x \in \mathbb{R}^d$, the function $y \mapsto q(x, y)$ is a density on $\mathbb{R}^d$. The idea behind the dynamics of a Metropolis–Hastings chain is to propose a move from a density $q(x, \cdot)$ to a new location, say y, and accept it with probability

$$\alpha(x, y) := \begin{cases} \min\left(1, \dfrac{\pi(y)q(y, x)}{\pi(x)q(x, y)}\right), & \pi(x)q(x, y) > 0, \\ 1, & \pi(x)q(x, y) = 0. \end{cases}$$

The Markov transition kernel $P(x, dy)$ for this dynamics is given by the formula

$$P(x, dy) := \alpha(x, y)q(x, y)\,dy + \left(1 - \int_{\mathbb{R}^d} \alpha(x, z)q(x, z)\,dz\right)\delta_x(dy), \qquad (\mathrm{MH}(q, \pi))$$

where δ_x is Dirac's measure centred at x, and the Markov chain $(\Phi_k)_{k\in\mathbb{N}}$ generated by P is known as the *Metropolis–Hastings chain* (see [7,18]). In this context, π is termed a *target density* and q a *proposal density*. It is easy to see that the chain Φ is reversible (i.e. it satisfies $\pi(x)\,dx\,P(x, dy) = \pi(y)\,dy\,P(y, dx)$) and hence stationary (i.e., $\int_{\mathbb{R}^d} P(x, dy)\pi(x)\,dx = \pi(y)\,dy$) with respect to π. The measure $\pi(x)\,dx$ is also known as the *invariant probability measure* for the chain Φ. Throughout the paper, we assume that the kernel P in $\mathrm{MH}(q, \pi)$ satisfies the following assumptions:

A1. *Geometric drift condition*: there exists a continuous function $V\colon \mathbb{R}^d \to [1, \infty)$, such that $\pi(V^2) < \infty$, V has bounded sublevel sets (i.e., $V^{-1}([1, c])$ is bounded for every $c \geq 1$) and

$$PV(x) \leq \lambda_V V(x) + \kappa_V 1_{C_V}(x), \qquad \text{for all } x \in \mathbb{R}^d,$$

for constants $\lambda_V \in (0, 1)$, $\kappa_V > 0$ and a compact set $C_V \subset \mathbb{R}^d$.

A2. The target density $\pi\colon \mathbb{R}^d \to (0, \infty)$ is continuous and strictly positive.

A3. The proposal density $q\colon \mathbb{R}^d \times \mathbb{R}^d \to (0, \infty)$ is continuous, strictly positive and bounded.

Associated with the *drift function* V is the function space

$$\begin{aligned} L_V^\infty &:= \left\{G\colon \mathbb{R}^d \to \mathbb{R};\ G \text{ measurable and } \|G\|_V < \infty\right\}, \\ &\text{where } \|G\|_V := \sup_{x\in\mathbb{R}^d} \frac{|G(x)|}{V(x)}. \end{aligned} \qquad (2.1)$$

Note that L_V^∞ equipped with the norm $\|\cdot\|_V$ is a Banach space (see [11], Proposition 7.2.1).

Remark 2.1. (i) Assumptions A1–A3 are standard. Widely used classes of Random walk Metropolis chains (i.e., $q(x, y) = q^*(y - x)$) satisfying A1–A3 are given in [13,17,27]. See also [28] for examples of Metropolis Adjusted Langevin chains satisfying A1–A3.

(ii) For Metropolis kernel P satisfying A1–A3 and $F \in L_V^\infty$ there exists a solution $\hat{F}$ to PE(P, F) that is an element of L_V^∞. The solution $\hat{F}$ is unique up to the addition of a constant function (see [6], Proposition 1.1 and Theorem 2.3).

(iii) Assumptions A2 and A3 imply that Metropolis–Hastings chain Φ driven by P is π-irreducible (i.e. μ^{Leb}-irreducible), strongly aperiodic and positive Harris recurrent (see [17], Lemmas 1.1 and 1.2, [31], Theorem 1, Corollary 2, and monograph [20] as a general reference). In particular, the SLLN [20], Theorem 17.1.7, and the CLT [20], Theorem 17.4.4, hold for $F \in L_V^\infty$.

(iv) If $\pi(V) < \infty$ but $\pi(V^2) = \infty$, we may work with $\sqrt{V}$ instead of V, as Jensen's inequality implies $P(\sqrt{V}) \le \sqrt{\lambda_V}\sqrt{V} + \sqrt{\kappa_V} 1_{C_V}$, thus restricting our results to force functions $F \in L_{\sqrt{V}}^\infty$.

(v) Geometric drift condition A1 implies that for $G \in L_V^\infty$ we have $\pi(G^2) < \infty$, $PG(x)$ is well defined for any $x \in \mathbb{R}^d$, $PG \in L_V^\infty$ and $\pi(PG - G) = 0$. In particular, CLT(Φ, $F + PG - G$) holds with mean $\pi(F)$ and (possibly substantially reduced) asymptotic variance σ_{F+PG-G}^2.

Remark 2.1(v) motivates the following definition.

Definition 2.2. Let Φ be a Metropolis–Hastings chain driven by kernel P in MH(q, π). Let $(G_n)_{n\in\mathbb{N}}$ be a sequence in L_V^∞ with the asymptotic variance σ_n^2 in the CLT(Φ, $F + PG_n - G_n$). We say that $(G_n)_{n\in\mathbb{N}}$ *asymptotically solves Poisson's equation* PE(P, F) if $\lim_{n\to\infty} \sigma_n^2 = 0$.

Remark 2.3. (a) If $(G_n)_{n\in\mathbb{N}}$ asymptotically solves Poisson's equation PE(P, F), so does $(G_n + c_n)_{n\in\mathbb{N}}$ for any sequence $(c_n)_{n\in\mathbb{N}}$ of real numbers.

(b) Definition 2.2 does not require the Metropolis–Hastings structure on $\mathbb{R}^d$ and can be extended trivially to Markov chains on general state spaces satisfying an appropriate CLT.

We now define a sequence of functions that asymptotically solves Poisson's equation PE(P, F).

Definition 2.4. (a) Let $\mathbb{J}$ be a partition of $\mathbb{R}^d$ into measurable sets $J_0, J_1, \dots, J_m$, such that $\bigcup_{j=1}^m J_j$ is bounded and $\mu^{\mathrm{Leb}}(J_j) > 0$ holds for all $0 \le j \le m$. Let $X = \{a_0, a_1, \dots, a_m\}$ be a set of *representatives*: $a_j \in J_j$ for all $0 \le j \le m$. The pair $\mathbb{X} := (\mathbb{J}, X)$ is called an *allotment* and m be the *size* of the allotment $\mathbb{X}$.

(b) Let $W : \mathbb{R}^d \to [1, \infty)$ be a measurable function and $\mathbb{X}$ an allotment. *W-radius* and *W-mesh* of the allotment $\mathbb{X}$ are defined by

$$\mathrm{rad}(\mathbb{X}, W) := \inf_{y\in J_0} W(y), \tag{2.2}$$

$$\delta(\mathbb{X}, W) := \max\left(\max_{1\le j\le m} \sup_{y\in J_j} |y - a_j|, \max_{0\le j\le m} \sup_{y\in J_j} \left(\frac{W(a_j)}{W(y)} - 1\right)\right), \tag{2.3}$$

respectively, where $|x|$ denotes the Euclidean norm of any $x \in \mathbb{R}^d$.

(c) A sequence of allotments $(\mathbb{X}_n)_{n\in\mathbb{N}}$ is *exhaustive* with respect to the function W in (b) if the following holds: $\lim_{n\to\infty} \mathrm{rad}(\mathbb{X}_n, W) = \infty$ and $\lim_{n\to\infty} \delta(\mathbb{X}_n, W) = 0$.

Remark 2.5. (i) For any continuous function $W\colon \mathbb{R}^d \to [1,\infty)$ with bounded sublevel sets, there exists an exhaustive sequence of allotments (see the Appendix below).

(ii) Note that J_0 is the only unbounded set in the partition of an allotment $\mathbb{X}$. For the W-radius of $\mathbb{X}$ to be large, the union $\bigcup_{j=1}^m J_j$ of all the bounded sets in the partition has to cover the part of $\mathbb{R}^d$ where W is small.

(iii) The W-mesh is a maximum of two quantities: the first is a standard mesh of the partition $\{J_1,\dots,J_m\}$ of the bounded set $\mathbb{R}\setminus J_0=\bigcup_{j=1}^m J_j$. The second quantity in (2.3) implies that for the W-mesh to be small, representatives a_j have to be chosen so that $W(a_j)$ and $\inf_{y\in J_j} W(y)$ are close to each other, relative to size of W on J_j. Intuitively, if $W(a_0)$ is close to $\inf_{y\in J_0} W(y)$ and W is continuously differentiable, then the second term in (2.3) is approximately equal to

$$\max_{1\le j\le m}\sup_{y\in J_j}\big(\big(\nabla \log W(y)\big)^\top (y-a_j)\big).$$

Thus, if W does not exhibit super-exponential growth, the representatives $a_1,\dots,a_m$ can be chosen arbitrarily.

We can now state our main result.

Theorem 2.6. *Let the transition kernel P in $\mathrm{MH}(q,\pi)$ of a Metropolis–Hastings chain Φ satisfy A1–A3 for a drift function V. Let $F\in L_V^\infty$ be continuous π-a.e. and let $(\mathbb{X}_n=(\mathbb{J}_n,X_n))_{n\in\mathbb{N}}$ be an exhaustive sequence of allotments with respect to V, where $\mathbb{J}_n=\{J_0^n,\dots,J_{m_n}^n\}$ and $X_n=\{a_j^n\in J_j^n : j=0,1,\dots,m_n\}$. For each $n\in\mathbb{N}$, let $\tilde F_n$ be the output of the Scheme with input P, F and $\mathbb{X}_n$. Then the sequence $(\tilde F_n)_{n\in\mathbb{N}}$ asymptotically solves Poisson's equation $\mathrm{PE}(P,F)$, that is, the asymptotic variance σ_n^2 in $\mathrm{CLT}(\Phi, F+P\tilde F_n-\tilde F_n)$ converges to zero as $n\to\infty$.*

Remark 2.7. Functions $\tilde F_n$ from Theorem 2.6 are well defined. This is because all the entries

$$(p_n)_{ij}:=(p_{\mathbb{X}_n})_{ij}=P\big(a_i^n,J_j^n\big)=\begin{cases}\displaystyle\int_{J_j^n}\alpha\big(a_i^n,y\big)q\big(a_i^n,y\big)\,dy, & \text{if } i\neq j,\\ \displaystyle 1-\int_{\mathbb{R}^d\setminus J_i^n}\alpha\big(a_i^n,y\big)q\big(a_i^n,y\big)\,dy, & \text{if } i=j\end{cases}\tag{2.4}$$

of stochastic matrices p_n, constructed by the *Scheme* with input P, F and $\mathbb{X}_n$, are strictly positive by Assumptions A2, A3 and Definition 2.4(a) ($\mu^{\mathrm{Leb}}(J_j^n)>0$ for all $0\le j\le m_n$, where m_n is the size of allotment $\mathbb{X}_n$). Hence the chain on X_n, driven by p_n, is irreducible, recurrent, aperiodic and admits a unique invariant probability measure π_n. Moreover, Poisson's equation for p_n and any force function on X_n has a solution, unique up to addition of a constant (see [16], Theorem 9.3).

Remark 2.8. The proof of Theorem 2.6 does not rely heavily on the structure of Metropolis–Hastings kernels. Emulating the proof appears feasible at least for other specific T-chains (see [20], Chapter 6, for the definition). More specifically, reversibility is needed in Proposition 3.2, but an analogous result can be obtained without it. In the proof of Proposition 3.3(b), we use the fact that the non-Dirac component $T(x, dy)$ of $P(x, dy)$ has positive and continuous density with respect to the Lebesgue measure. Finally, in proofs of Proposition 3.7 and Theorem 2.6 we require $T(x, dy)$ to exhibit the following form of continuity, $\lim_{n\to\infty} \|T(x,\cdot) - T(a^n(x),\cdot)\|_V = 0$ for π-a.e. x (here $\|\cdot\|_V$ is the V total variation norm and $a^n(x) = \sum_{j=0}^{m_n} a_j^n 1_{J_j^n}(x)$).

3. Proof of Theorem 2.6

3.1. Overview of the proof

The central object in the proof of Theorem 2.6 is the function

$$\Delta(G) := PG - G + F - \pi(F), \tag{3.1}$$

which measures the failure of a function G to be a solution of the Poisson equation PE(P, F). Intuitively, the closer $\Delta(G)$ is to zero the better.

The proof is in two parts. In the first part (Section 3.2 below), we show that a sequence of functions $(G_n)_{n\in\mathbb{N}}$ in L_V^∞ asymptotically solves Poisson's equation PE(P, F) if $\lim_{n\to\infty} \pi(\Delta(G_n)^2) = 0$. This is a simple consequence of the representation of the asymptotic variance in terms of the spectral measure [14], equation (1.1), and the existence of a spectral gap for geometrically ergodic Markov chains established [25], Proposition 1.1.

The second part of the proof is more involved. It consists of verifying that functions $(\tilde{F}_n)_{n\in\mathbb{N}}$, defined in Theorem 2.6, indeed satisfy $\lim_{n\to\infty} \pi(\Delta(\tilde{F}_n)^2) = 0$. The key underlying fact needed for this purpose is that the family of the approximating finite state Markov chains driven by the stochastic matrices $(p_n)_{n\in\mathbb{N}}$ converge to their respective stationary distributions $(\pi_n)_{n\in\mathbb{N}}$ *uniformly* in $n \in \mathbb{N}$. This step is facilitated by the results in [21], Theorem 2.3 and [2], Theorem 1.1, which show that the constants appearing in the geometric ergodicity estimate depend only and explicitly on the constants in the drift, minorisation and strong aperiodicity conditions for that chain. In Section 3.3, we show that these constants can be chosen independently of $n \in \mathbb{N}$ (Proposition 3.3 below) and establish the uniform convergence to stationarity (Proposition 3.4 below).

In Section 3.4, we establish convergence in $L^2(\pi)$ of the sequence $(\Delta(\tilde{F}_n))_{n\in\mathbb{N}}$. In addition to the uniform convergence to stationarity, the proof requires a further weak approximation by a family of finite state Markov chains with stationary distributions that are explicit in the target density π (see (3.11) below). Note that the stationary laws π_n of the chains generated by the stochastic matrices p_n, defined in (2.4), cannot be expressed explicitly in terms of π.

Remark 3.1 (Auxiliary notation). In addition to the notation used in the statement of Theorem 2.6 and Remark 2.7, throughout the remainder of the section we will use the following objects:

- $\hat{F}$: solution of PE(P, F) in L_V^∞ (*cf.* Remark 2.1(ii)).
- f_n and v_n: restrictions of F and V to the set X_n, respectively.
- $\hat{f}_n$: solution of PE(p_n, f_n) constructed within the *Scheme* (*cf.* Remark 2.7).
- $\delta_n := \delta(\mathbb{X}_n, V)$: the V-mesh of the allotment $\mathbb{X}_n$ defined in (2.3).

3.2. Controlling the asymptotic variance

The following proposition gives a sufficient conditions for a sequence of functions $(G_n)_{n\in\mathbb{N}}$ to solve asymptotically the Poisson equation.

Proposition 3.2. *Let the sequence* $(G_n)_{n\in\mathbb{N}}$ *in* L_V^∞ *satisfy* $\lim_{n\to\infty}\pi(\Delta(G_n)^2)=0$. *Then* $(G_n)_{n\in\mathbb{N}}$ *asymptotically solves* PE(P,F) *in the sense of Definition 2.2.*

Proof. The kernel P is reversible and hence a bounded self-adjoint operator on the Hilbert space $L^2(\pi)$. Furthermore, the Hilbert subspace $\mathcal{H}:=\{G\in L^2(\pi):\pi(G)=0\}$ is invariant for P (i.e. $\pi(PG)=0$ for any $G\in\mathcal{H}$). By (3.1) and Remark 2.1(v) it follows that $\Delta(G_n)\in\mathcal{H}$ for all $n\in\mathbb{N}$. The asymptotic variance σ_n^2 in the CLT$(\Phi, F+PG_n-G_n)$ can be represented in terms of a positive (spectral) measure $E_{\Delta(G_n)}(d\lambda)$ on the spectrum $\sigma(P|_{\mathcal{H}})\subset\mathbb{R}$ associated with the function $\Delta(G_n)$, as follows (see [14] and [5], Theorem 2.1, for details):

$$\sigma_n^2=\int_{\sigma(P|_{\mathcal{H}})}\frac{1+\lambda}{1-\lambda}E_{\Delta(G_n)}(d\lambda). \tag{3.2}$$

Since the chain generated by P is geometrically ergodic by A1, [25], Proposition 1, implies that the spectral radius ρ of $P|_{\mathcal{H}}$ satisfies $\rho<1$. Hence, the inclusion $\sigma(P|_{\mathcal{H}})\subseteq[-\rho,\rho]$, the equality $E_{\Delta(G_n)}(\sigma(P|_{\mathcal{H}}))=\pi(\Delta(G_n)^2)$ (see e.g. [5], equation (2.2)) and the formula in (3.2) imply

$$\sigma_n^2\le\frac{1+\rho}{1-\rho}\cdot\int_{\sigma(P|_{\mathcal{H}})}E_{\Delta(G_n)}(d\lambda)=\frac{1+\rho}{1-\rho}\cdot\pi\big(\Delta(G_n)^2\big)\longrightarrow 0 \qquad \text{as } n\to\infty.$$

This proves the proposition. □

3.3. Uniform convergence to stationarity

Fix an exhaustive sequence of allotments $(\mathbb{X}_n)_{n\in\mathbb{N}}$ and stochastic matrices p_n, $n\in\mathbb{N}$, as in Theorem 2.6. The main aim of this section is to prove that the corresponding chains are geometrically ergodic uniformly in $n\in\mathbb{N}$. This is achieved as follows: first, the *uniform drift, minorisation* and *strong aperiodicity conditions* in (3.6), (3.7) and (3.8), respectively, are established. Then, the uniform convergence to stationarity follows from [2], Theorem 1.1 (cf. [21], Theorem 2.3).

For each $n\in\mathbb{N}$, let $a^n:\mathbb{R}^d\to\mathbb{R}^d$ map $x\in\mathbb{R}^d$ to its representative in $\mathbb{X}_n$. More precisely, let

$$a^n(x):=\sum_{j=0}^{m_n}a_j^n 1_{J_j^n}(x) \qquad \text{for every } x\in\mathbb{R}^d, \tag{3.3}$$

where $\{J_0^n,\dots,J_{m_n}^n\}$ is the partition and $X_n=\{a_0^n,\dots,a_{m_n}^n\}$ are the representatives in the allotment $\mathbb{X}_n$. Since the sequence of allotments is exhaustive, the following limit holds:

$$\lim_{n\to\infty}a^n(x)=x \qquad \text{for every } x\in\mathbb{R}^d. \tag{3.4}$$

Note that the definition of a V-mesh (see (2.3) in Definition 2.4) implies the inequality

$$V\big(a^n(x)\big) = V\big(a^n(x)\big) - V(x) + V(x) \le (1+\delta_n)V(x) \qquad \text{for all } n \in \mathbb{N} \text{ and } x \in \mathbb{R}^d. \quad (3.5)$$

Proposition 3.3 (Uniform drift, minorisation and strong aperiodicity conditions). *There exists a compact set $C \subset \mathbb{R}^d$ such that the following statements hold.*

(a) *There exist positive constants $\lambda < 1$, κ, such that the uniform drift condition holds*:

$$p_n v_n\big(a_j^n\big) \le \lambda v_n\big(a_j^n\big) + \kappa 1_C\big(a_j^n\big) \qquad \text{for all } n \in \mathbb{N}, \text{ and } a_j^n \in X_n. \quad (3.6)$$

(b) *Define $C_n := X_n \cap C$, for each $n \in \mathbb{N}$. There exist constants $\gamma, \tilde{\gamma} \in (0,\infty)$ and a measure ν_n, concentrated on X_n, such that the uniform minorisation condition,*

$$(p_n)_{ij} \ge \gamma \nu_n\big(\{a_j^n\}\big) \qquad \text{for all } n \in \mathbb{N}, \text{ and } i,j \in \{0,1,\dots,m_n\} \text{ satisfying } a_i^n \in C_n, \quad (3.7)$$

and the uniform strong aperiodicity condition,

$$\gamma \nu_n(C_n) \ge \tilde{\gamma} \qquad \text{for all } n \in \mathbb{N}, \quad (3.8)$$

hold.

Proof. (a) Fix an arbitrary $n \in \mathbb{N}$ and $j \in \{0,\dots,m_n\}$. By definition of the function $a^n(\cdot)$ in (3.3), we find

$$p_n v_n\big(a_j^n\big) - v_n\big(a_j^n\big) = \int_{\mathbb{R}^d} \big(V\big(a^n(y)\big) - V\big(a_j^n\big)\big)\alpha\big(a_j^n, y\big) q\big(a_j^n, y\big)\, dy.$$

By (3.5) we get $V(a^n(y)) - V(a_j^n) \le V(y) - V(a_j^n) + \delta_n V(y)$ for every $y \in \mathbb{R}^d$. The form of kernel P in (MH(q,π)) and this inequality imply

$$\begin{aligned} p_n v_n\big(a_j^n\big) - v_n\big(a_j^n\big) &\le PV\big(a_j^n\big) - V\big(a_j^n\big) + \delta_n \int_{\mathbb{R}^d} V(y)\alpha\big(a_j^n, y\big) q\big(a_j^n, y\big)\, dy \\ &\le PV\big(a_j^n\big) - V\big(a_j^n\big) + \delta_n PV\big(a_j^n\big) = (1+\delta_n)PV\big(a_j^n\big) - V\big(a_j^n\big). \end{aligned}$$

Since by definition $V(a_j^n) = v_n(a_j^n)$, the geometric drift condition in A1 implies

$$p_n v_n\big(a_j^n\big) \le (1+\delta_n)\lambda_V v_n\big(a_j^n\big) + (1+\delta_n)\kappa_V 1_{C_V}\big(a_j^n\big).$$

Since $\lim_{n\to\infty} \delta_n = 0$, if we define $C := C_V$, $\lambda := \frac{1+\lambda_V}{2}$ and $\kappa := \kappa_V(1+\sup_{n\in\mathbb{N}} \delta_n)$, there exists $N_0 \in \mathbb{N}$ such that the drift condition in (3.6) holds for all $n \ge N_0$. Note that if we enlarge C and increase κ, the uniform drift condition in (3.6) remains valid for all n it was valid for before the modification. Finally, if $N_0 > 1$, we enlarge C by all the representatives of the allotments $\mathbb{X}_1, \dots, \mathbb{X}_{N_0}$ (finitely many points) and increase κ sufficiently, so that (3.6) also holds for all $n \in \{1,\dots,N_0 - 1\}$.

(b) Recall that by Definition 2.4(c), the sequence $(r_n := \mathrm{rad}(\mathbb{X}_n, V))_{n\in\mathbb{N}}$ tends to infinity, though perhaps not monotonically. Let D be an open ball of radius $r_D > 2\sup_{n\in\mathbb{N}} \delta_n$ in $\mathbb{R}^d$.

Since D is a bounded set, by the definition of V-radius (see (2.2)) and Assumption A1, there exists $n_0 \in \mathbb{N}$ such that $D \subseteq \bigcap_{n \geq n_0} V^{-1}([1, r_n))$. We now enlarge the compact set C, constructed in part (a) of this proof, to contain the bounded set

$$\left(\bigcup_{n<n_0} \mathbb{R}^d \setminus J_0^n\right) \cup \bigcap_{n \geq n_0} V^{-1}\bigl([1, r_n)\bigr). \tag{3.9}$$

We may assume the set C is still compact, since the set in (3.9) is bounded, and hence the uniform drift condition in (3.6) still holds.

Define a measure ν on the Borel σ-algebra of $\mathbb{R}^d$ by $\nu(B) := \frac{\mu^{\mathrm{Leb}}(B \cap C)}{\mu^{\mathrm{Leb}}(C)}$ for any measurable set B. For each $n \in \mathbb{N}$, define a measure on the set of representatives X_n by $\nu_n(\{a_j^n\}) := \nu(J_j^n)$. Define the constant $\gamma := \mu^{\mathrm{Leb}}(C) \inf_{y,x \in C \times C} \alpha(x, y) q(x, y)$ and note that it is strictly positive by Assumptions A2 and A3 and Definition 2.4(a). For every $n \in \mathbb{N}$ and every $0 \leq i, j \leq m_n$, such that $a_i^n \in C_n$, the form of the kernel P in (MH(q, π)) implies the minorisation condition in (3.7):

$$(p_n)_{ij} = P\bigl(a_i^n, J_j^n\bigr) \geq \int_{J_j^n \cap C} \alpha\bigl(a_i^n, y\bigr) q\bigl(a_i^n, y\bigr)\, dy \geq \gamma \nu\bigl(J_j^n\bigr) = \gamma \nu_n\bigl(\bigl\{a_j^n\bigr\}\bigr).$$

We now establish the strong aperiodicity condition in (3.8). First assume that $n \geq n_0$, let D' be an open ball of radius $\frac{r_D}{2}$, with the same centre as D, and pick $y \in D'$. The definition of the V-radius $r_n = \operatorname{rad}(\mathbb{X}_n, V)$ in (2.2) implies $D \cap J_0^n \subseteq V^{-1}([1, r_n)) \cap V^{-1}([r_n, \infty))$ and hence $D \cap J_0^n = \varnothing$. Since the radius r_D of the ball D is strictly greater than $2 \sup_{n \in \mathbb{N}} \delta_n$ and the inequality $|y - a^n(y)| \leq \sup_{n \in \mathbb{N}} \delta_n$ holds, it follows that $a^n(y) \in D \subseteq C$. Hence, by definition (3.3), it holds that $D' \subseteq \bigcup_{\{j; a_j^n \in C\}} J_j^n$ and

$$\nu_n(C_n) = \nu_n(X_n \cap C) = \nu\Bigl(\bigcup_{\{j; a_j^n \in C\}} J_j^n\Bigr) \geq \nu\bigl(D'\bigr) = \frac{\mu^{\mathrm{Leb}}(D')}{\mu^{\mathrm{Leb}}(C)} > 0.$$

If $n < n_0$, then it holds that $C_n = X_n \cap C \supset \{a_j^n : j = 1, \ldots, m_n\}$, since C contains the set in (3.9) and hence $\mathbb{R}^d \setminus J_0^n$. Therefore, we find $\nu_n(C_n) \geq \frac{\mu^{\mathrm{Leb}}(\mathbb{R} \setminus J_0^n)}{\mu^{\mathrm{Leb}}(C)} > 0$. Hence, (3.8) holds for the positive constant

$$\tilde{\gamma} := \frac{1}{\gamma} \min\left\{\frac{\mu^{\mathrm{Leb}}(D')}{\mu^{\mathrm{Leb}}(C)}, \min_{n<n_0} \frac{\mu^{\mathrm{Leb}}(\mathbb{R} \setminus J_0^n)}{\mu^{\mathrm{Leb}}(C)}\right\}.$$

This concludes the proof of the proposition. □

Proposition 3.3 allows us to control the convergence to stationarity of the approximating chains uniformly in $n \in \mathbb{N}$. In the notation of Theorem 2.6 and Remarks 2.7 and 3.1, the following statement holds.

Proposition 3.4. *There exist positive constants ζ and $\theta < 1$, such that the inequality*

$$\sup_{\|g\|_{v_n} \le 1} \left|\left(p_n^k g\right)(b) - \pi_n(g)\right| \le \zeta\theta^k v_n(b) \qquad \textit{holds for all } b \in X_n,\ k \in \mathbb{N} \cup \{0\} \textit{ and } n \in \mathbb{N},$$

where the v_n-norm of a function $g : X_n \to \mathbb{R}$ is $\|g\|_{v_n} := \sup_{b \in X_n} |g(b)|/v_n(b)$ and $\pi_n(g)$ denotes the integral (*i.e.*, *weighted sum*) *of g with respect to π_n.*

Proof. Pick an arbitrary $n \in \mathbb{N}$. According to Proposition 3.3, the transition matrix p_n satisfies the drift condition in (3.6), the minorisation condition in (3.7) and the strong aperiodicity condition (3.8) with the constants $\kappa, \lambda, \gamma, \tilde{\gamma}$, which are independent of the choice of n. Hence, [2], Theorem 1.1 (see also [21], Theorem 2.3) applied to the transition kernel p_n on the state space X_n, yields

$$\sup_{\|g\|_{v_n} \le 1} \left|\left(p_n^k g\right)\left(a_j^n\right) - \pi_n(g)\right| \le \zeta(n) v_n\left(a_j^n\right)\theta(n)^k$$

for every $k \in \mathbb{N} \cup \{0\}$, $a_j^n \in X_n$ and constants $\zeta(n) \in (0, \infty)$ and $\theta(n) \in (0, 1)$. Furthermore, [2], Theorem 1.1, implies that the constants $\zeta(n), \theta(n)$ are only a (chain independent) function of $\kappa, \lambda, \gamma, \tilde{\gamma}$ in Proposition 3.3 and hence do not depend on n. This concludes the proof. □

3.4. Functions that asymptotically solve Poisson's equation PE(P, F)

In this section, we complete the proof of Theorem 2.6. By the Dominated Convergence Theorem (DCT), Proposition 3.2 implies that $(\tilde{F}_n)_{n\in\mathbb{N}}$ asymptotically solves PE(P, F) if the following conditions hold:

$$\sup_{n\in\mathbb{N}} \left\|\Delta(\tilde{F}_n)\right\|_V < \infty \quad \text{and} \quad \lim_{n\to\infty} \Delta(\tilde{F}_n)(x) = 0 \qquad \text{for } \pi\text{-a.e. } x \in \mathbb{R}^d. \tag{3.10}$$

The inequality in (3.10) follows from (3.1) and Proposition 3.5 below, which states that the V-norm $\tilde{F}_n$, shifted by an appropriate constant, is bounded uniformly in $n \in \mathbb{N}$. The existence of these constants rests on the uniform convergence to stationarity in Proposition 3.4 above.

The limit in (3.10) is established by bounding $|\Delta(\tilde{F}_n)|$ by a sum of three non-negative terms (see Lemma 3.8 below) and controlling each one separately. The first, given by $|F(x) - F(a^n(x))|$, tends to zero by (3.4) since the force function F is assumed to be continuous π-a.e. The second term $|U(x) - U(a^n(x))|$, where $U := P\tilde{F}_n - \tilde{F}_n$, is controlled by Proposition 3.5 and the DCT. Controlling the third term $|\pi_n(f_n) - \pi(F)|$ is more involved. It requires constructing a further approximating chain (based on the transition kernel P) with state space X_n and a transient matrix p_n^*, whose invariant distribution can be described analytically in terms of the density π (see (3.11) below). Proposition 3.7, whose proof also depends on the uniform convergence to stationarity in Proposition 3.4, establishes the desired limit. We now give the details of the outlined proof.

Proposition 3.5. *There exists a constant $\xi > 0$ and a sequence of real numbers $(c_n)_{n\in\mathbb{N}}$, such that the following inequality holds for all $n \in \mathbb{N}$:*

$$\|\tilde{F}_n + c_n\|_V \le \xi.$$

Proof. Pick an arbitrary $n \in \mathbb{N}$. Since $F \in L_V^\infty$ by assumption, its restriction $f_n : X_n \to \mathbb{R}$ satisfies $\|f_n\|_{v_n} \le \|F\|_V$ (see Proposition 3.4 for definition of v_n-norm). By Proposition 3.4, the function $\bar{f}_n : X_n \to \mathbb{R}$, given by

$$\bar{f}_n := \sum_{k=0}^{\infty} \bigl(p_n^k f_n - \pi_n(f_n)\bigr),$$

is well defined and satisfies the inequality $\|\bar{f}_n\|_{v_n} \le \frac{\zeta}{1-\theta}\|f_n\|_{v_n} \le \frac{\zeta}{1-\theta}\|F\|_V$. Furthermore, by [20], Theorem. 17.4.2, the function $\bar{f}_n$ solves Poisson's equation $\mathrm{PE}(p_n, f_n)$. Since $\hat{f}_n : X_n \to \mathbb{R}$, in the definition of $\tilde{F}_n$, also solves $\mathrm{PE}(p_n, f_n)$, by Remark 2.7 there exists a constant $c_n \in \mathbb{R}$ such that $\hat{f}_n + c_n = \bar{f}_n$.

Recall that $\tilde{F}_n = \sum_{j=0}^{m_n} \hat{f}_n(a_j^n) 1_{J_j^n}$, pick an arbitrary $x \in \mathbb{R}^d$ and note that definition (3.3) implies $\tilde{F}_n(x) = \hat{f}_n(a^n(x))$. Hence, we obtain

$$\begin{aligned}\bigl|\tilde{F}_n(x) + c_n\bigr| &= \bigl|\bar{f}_n\bigl(a^n(x)\bigr)\bigr| \le \frac{\zeta}{1-\theta}\|F\|_V v_n\bigl(a^n(x)\bigr) = \frac{\zeta}{1-\theta}\|F\|_V V\bigl(a^n(x)\bigr) \\ &\le \xi V(x), \qquad \text{where } \xi := \frac{\zeta}{1-\theta}\Bigl(1 + \sup_{k\in\mathbb{N}} \delta_k\Bigr)\|F\|_V\end{aligned}$$

and the last inequality follows from (3.5). Since both $x \in \mathbb{R}^d$ and $n \in \mathbb{N}$ were arbitrary, this implies the proposition. □

In order to analyse the behaviour of the limit in (3.10), we need to define a further approximating Markov chain on X_n with the transition matrix p_n^* and the invariant measure π_n^*, given by

$$\begin{aligned}\bigl(p_n^*\bigr)_{ij} &:= \int_{J_i^n} \frac{\pi(x)}{\pi(J_i^n)} P\bigl(x, J_j^n\bigr)\,dx \quad \text{and} \\ \pi_n^*\bigl(\bigl\{a_j^n\bigr\}\bigr) &:= \pi\bigl(J_j^n\bigr), \qquad \text{for } i, j \in \{0, \dots, m_n\},\end{aligned} \tag{3.11}$$

respectively. Note that $(p_n^*)_{ij} = \mathbb{P}_\pi[\Phi_1 \in J_i^n | \Phi_0 \in J_j^n]$, where Φ is the Metropolis–Hastings chain we are analysing. It is clear from the definition in (3.11) that the equality $\pi_n^* p_n^* = \pi_n^*$ holds. Furthermore, if we define a function $h_n\colon X_n \to \mathbb{R}$ by

$$h_n\bigl(a_j^n\bigr) := \int_{J_j^n} \frac{\pi(x)}{\pi(J_j^n)} F(x)\,dx \qquad \text{for } a_j^n \in X_n, \text{ it holds that} \qquad \pi_n^*(h_n) = \pi(F). \tag{3.12}$$

Remark 3.6. (i) Let μ be a signed measure on X_n and $\|\mu\|_{v_n} := \sup_{\|g\|_{v_n} \le 1} |\mu(g)|$ its v_n-norm, where the norm $\|g\|_{v_n}$ was defined in Proposition 3.4 and $\mu(g)$ denotes the integral (i.e. weighted sum) of $g : X_n \to \mathbb{R}$ with respect to μ. Furthermore, it is natural to define the *dual* normed vector spaces $(L_{v_n}^\infty, \|\cdot\|_{v_n})$ (analogous to L_V^∞ in (2.1)) and $(M_{v_n}^\infty, \|\cdot\|_{v_n})$ of functions on X_n and signed measures on X_n, respectively. Since X_n is finite, the vector spaces $L_{v_n}^\infty$ and $M_{v_n}^\infty$ are isomorphic to

$\mathbb{R}^{1+m_n}$. Furthermore, any linear function $B : L^\infty_{v_n} \to L^\infty_{v_n}$, mapping $g \mapsto Bg$, induces a linear map on the dual $B^* : M^\infty_{v_n} \to M^\infty_{v_n}$, given by $\mu \mapsto B^*\mu := \mu B$ (in this definition we interpret μ as a row vector and B as a matrix). It is well known that the operator norms coincide $\|B\|_{v_n} = \|B^*\|_{v_n}$. This fact, which holds in a much more general setting (see [11], Section 7), plays an important role in the proof of Proposition 3.7.

(ii) The following estimate holds for any point $x \in \mathbb{R}^d$ and all $n \in \mathbb{N}$, $y \in \mathbb{R}^d$:

$$\begin{aligned}\alpha\big(a^n(x), y\big)q\big(a^n(x), y\big) &\le \frac{q(y, a^n(x))}{\pi(a^n(x))}\pi(y) \\ &\le \eta_x \pi(y), \qquad \text{where } \eta_x := \frac{\sup_{z,y\in\mathbb{R}^d} q(z, y)}{\inf_{n\in\mathbb{N}} \pi(a^n(x))}.\end{aligned} \tag{3.13}$$

By (3.4) and A2 we have $0 < \inf\{\pi(z) : |z - x| \le \sup_{k\in\mathbb{N}} \delta_k\} \le \pi(a^n(x))$, where $\delta_k = \delta(\mathbb{X}_k, V)$ (see Definition 2.4), for all sufficiently large $n \in \mathbb{N}$. Thus, by A2 and A3, we have $\eta_x \in (0,\infty)$ and the inequalities in (3.13), which will be used in the proofs of Proposition 3.7 and Theorem 2.6, hold.

Proposition 3.7. *The following inequalities hold for the measure π^*_n defined in (3.11):*

$$\big|\big(\pi^*_n - \pi_n\big)(f_n)\big| \le \frac{\zeta\|F\|_V}{1-\theta}\big\|\pi^*_n - \pi^*_n p_n\big\|_{v_n}, \tag{3.14}$$

where the constants $\theta \in (0, 1)$ and $\zeta > 0$ are as in Proposition 3.4, and

$$\big\|\pi^*_n - \pi^*_n p_n\big\|_{v_n} \le \Big(1 + \sup_{k\in\mathbb{N}} \delta_k\Big) \int_{\mathbb{R}^d\times\mathbb{R}^d} \big(V(y) + V(x)\big)Z_n(x, y)\,dy\,\pi(x)\,dx, \tag{3.15}$$

where $Z_n(x, y) := |\alpha(a^n(x), y)q(a^n(x), y) - \alpha(x, y)q(x, y)|$ for any $x, y \in \mathbb{R}^d$ and the function $a^n(\cdot)$ is given in (3.3). Furthermore, the following limit holds: $\lim_{n\to\infty} |\pi_n(f_n) - \pi(F)| = 0$.

Proof. We estimate the difference $|\pi_n(f_n) - \pi(F)|$ using the invariant distribution π^*_n of the chain driven by p^*_n and the function h_n, defined in (3.11) and (3.12) respectively, as follows

$$\begin{aligned}\big|\pi_n(f_n) - \pi(F)\big| &= \big|\pi_n(f_n) - \pi^*_n(f_n) + \pi^*_n(f_n) - \pi^*_n(h_n)\big| \\ &\le \big|\big(\pi_n - \pi^*_n\big)(f_n)\big| + \big|\pi^*_n(f_n - h_n)\big|.\end{aligned} \tag{3.16}$$

We will prove that both terms on the right-hand side converge to zero as $n \to \infty$. The definitions of π^*_n and h_n (in (3.11) and (3.12) above) and the function $a^n(\cdot)$ (see (3.3)) imply that the second term on the right-hand side of (3.16) takes the form

$$\begin{aligned}\pi^*_n(f_n - h_n) &= \sum_{j=0}^{m_n} \pi\big(J^n_j\big)\bigg(F\big(a^n_j\big) - \int_{J^n_j} \frac{\pi(x)}{\pi(J^n_j)} F(x)\,dx\bigg) \\ &= \int_{\mathbb{R}^d} \big(F\big(a^n(x)\big) - F(x)\big)\pi(x)\,dx.\end{aligned}$$

Since F is continuous π-a.e., the integrand converges to zero π-a.e. by (3.4). Furthermore, for any $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}\left|F\left(a^n(x)\right) - F(x)\right| &\le \left|F\left(a^n(x)\right)\right| + \left|F(x)\right| \le \|F\|_V\left(V\left(a^n(x)\right) + V(x)\right)\\ &\le \|F\|_V\left(2 + \sup_{k\in\mathbb{N}} \delta_k\right)V(x),\end{aligned}$$

where the last inequality follows from (3.5). Therefore, by the DCT (recall that by the assumption in A1 we have $\pi(V) < \infty$), the second term in (3.16) indeed converges to zero.

Establishing the convergence of the first term on the right-hand side in (3.16) is more involved. We start by establishing the following representation of the signed measure $\pi_n^* - \pi_n$.

Claim. There exists a linear map $B_n : L^\infty_{v_n} \to L^\infty_{v_n}$, with the dual $B_n^* : M^\infty_{v_n} \to M^\infty_{v_n}$, satisfying $\pi_n^* - \pi_n = B_n^*(\pi_n^* - \pi_n^* p_n) = (\pi_n^* - \pi_n^* p_n)B_n$ and $\|B_n^*\|_{v_n} = \|B_n\|_{v_n} \le \zeta/(1-\theta)$, where the constants $\theta \in (0,1)$ and $\zeta > 0$ are as in Proposition 3.4 (see Remark 3.6(I) for the definition of $L^\infty_{v_n}$ and $M^\infty_{v_n}$).

Define a transition matrix $1 \otimes \pi_n$ on the state space X_n by $(1 \otimes \pi_n)_{ij} := \pi_n(a_j^n)$. The corresponding chain is a sequence of independent r.v.s. with the law given by π_n (independently of the starting distribution). The inequality in Proposition 3.4 can therefore be expressed as $\|p_n^k - 1 \otimes \pi_n\|_{v_n} \le \zeta\theta^k$, for all $k \in \mathbb{N} \cup \{0\}$, implying that $B_n := \sum_{k=0}^\infty (p_n^k - 1 \otimes \pi_n)$ is a well defined linear map on the normed space $L^\infty_{v_n}$, such that $\|B_n\|_{v_n} \le \zeta/(1-\theta)$. In order to establish the first equality in the claim above, note that $\mu(1 \otimes \pi_n) = \pi_n$ for any probability measure $\mu \in M^\infty_{v_n}$ and, by Remark 3.6(I) and Proposition 3.4, the $\|\cdot\|_{v_n}$-norm of the linear operator $\mu \mapsto \mu(p_n^k - 1 \otimes \pi_n)$ on $M^\infty_{v_n}$ is bounded above by $\zeta\theta^k$ for all $k \in \mathbb{N}$. In particular, $\lim_{k\to\infty} \pi_n^* p_n^k = \pi_n$ in v_n-norm since $\|\pi_n^* p_n^k - \pi_n\|_{v_n} = \|\pi_n^*(p_n^k - 1 \otimes \pi_n)\|_{v_n} \le \zeta\theta^k \|\pi_n^*\|_{v_n}$ for all $k \in \mathbb{N}$. Consider the identity

$$\left(\pi_n^* - \pi_n^* p_n\right)\sum_{k=0}^{\ell}\left(p_n^k - 1 \otimes \pi_n\right) = \pi_n^* - \pi_n^* p_n^{\ell+1} \qquad \text{for all } \ell \in \mathbb{N},$$

and note that both sides converge in the appropriate $\|\cdot\|_{v_n}$-norms as $\ell \to \infty$. In the limit, the left-hand side equals $(\pi_n^* - \pi_n^* p_n)B_n$ and the right-hand side is $\pi_n^* - \pi_n$. This concludes the proof of the claim.

In order to establish the inequality in (3.14), note that $\|f_n\|_{v_n} \le \|F\|_V$ and Remark 3.6(I) imply $|(\pi_n^* - \pi_n)(f_n)| \le \|F\|_V(\pi_n^* - \pi_n)(f_n/\|f_n\|_{v_n}) \le \|F\|_V\|\pi_n^* - \pi_n\|_{v_n}$. This inequality and the claim imply (3.14).

The next task is to prove (3.15). Let $g\colon X_n \to \mathbb{R}$ be a function satisfying $\|g\|_{v_n} \le 1$. Recall that $m_n + 1$ is the cardinality of X_n and that the function $a^n(\cdot)$ is defined in (3.3). We apply the definitions of the stochastic matrix p_n^* and its stationary law π_n^*, given in (3.11), to obtain

$$\begin{aligned}&\left(\pi_n^* - \pi_n^* p_n\right)g\\ &\quad = \pi_n^*\left(p_n^* - p_n\right)g = \sum_{j=0}^{m_n}\sum_{i=0}^{m_n}\left[\pi\left(J_i^n\right)\left(\left(p_n^*\right)_{ij} - (p_n)_{ij}\right)\right]g\left(a_j^n\right)\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=0}^{m_n} \left[\int_{\mathbb{R}^d} \left(P\left(x, J_j^n\right) - P\left(a^n(x), J_j^n\right) \right) \pi(x)\, dx \right] g\left(a_j^n\right) \\
&= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} g\left(a^n(y)\right) \left[\alpha(x,y) q(x,y) - \alpha\left(a^n(x), y\right) q\left(a^n(x), y\right) \right] dy \right) \pi(x)\, dx \\
&\quad + \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} g\left(a^n(x)\right) \left[\alpha\left(a^n(x), y\right) q\left(a^n(x), y\right) - \alpha(x,y) q(x,y) \right] dy \right) \pi(x)\, dx,
\end{aligned}
$$

where the identity $\delta_x(J_j^n) g(a_j^n) = \delta_{a^n(x)}(J_j^n) g(a_j^n) = \delta_{a^n(x)}(J_j^n) g(a^n(x))$, for any $x \in \mathbb{R}^d$ and $j \in \{0, \dots, m_n + 1\}$, implies the final equality. Since the function $g \in L^\infty_{v_n}$, with $\|g\|_{v_n} \le 1$, in the calculation above was arbitrary and satisfies $|g(a^n(x))| \le V(a^n(x))$ for all $x \in \mathbb{R}^d$, we find

$$
\begin{aligned}
\left\| \pi_n^* - \pi_n^* p_n \right\|_{v_n} &= \sup_{\|g\|_{v_n} \le 1} \left| \left(\pi_n^* - \pi_n^* p_n \right) g \right| \\
&\le \int_{\mathbb{R}^d \times \mathbb{R}^d} \left(V\left(a^n(y)\right) + V\left(a^n(x)\right) \right) Z_n(x,y) \pi(x)\, dy\, dx,
\end{aligned}
$$

which, together with (3.5), implies (3.15).

We now apply the DCT to deduce that the right-hand side in (3.15) converges to zero as $n \to \infty$. The definition of $Z_n(x,y)$ in the proposition, the form of the transition kernel P in (MH(q, π)), the drift condition in A1 and the inequality in (3.5) imply the estimates

$$
\begin{aligned}
\int_{\mathbb{R}^d} \left(V(y) + V(x) \right) Z_n(x,y)\, dy &\le PV(x) + PV\left(a^n(x)\right) + 2V(x) \\
&\le \left(\left(2 + \sup_{k \in \mathbb{N}} \delta_k \right) (\lambda_V + \kappa_V) + 2 \right) V(x)
\end{aligned}
$$

for all $x \in \mathbb{R}^d$. Since, by Assumption A1, we have $\pi(V) < \infty$, by the DCT the right-hand side in (3.15) tends to zero (as $n \to \infty$) if

$$
\lim_{n \to \infty} \int_{\mathbb{R}^d} \left(V(y) + V(x) \right) Z_n(x,y)\, dy = 0 \qquad \text{for all } x \in \mathbb{R}. \tag{3.17}
$$

To establish the limit in (3.17), pick an arbitrary $x \in \mathbb{R}^d$ and note that for every $y \in \mathbb{R}$ it holds that $\lim_{n \to \infty} Z_n(x,y) = 0$ by (3.4) and the assumptions in A2 and A3. Hence the integrand in (3.17) converges to zero point-wise. By the estimate in (3.13), the integrand in (3.17) is bounded above by the function

$$
y \mapsto \left(V(y) + V(x) \right) \left(\eta_x \pi(y) + \alpha(x,y) q(x,y) \right)
$$

which does not depend on n and is μ^{Leb}-integrable in $y \in \mathbb{R}^d$. Hence, the limit in (3.17) holds by the DTC and, consequently, the right-hand side in (3.15) converges to zero as $n \to \infty$. This fact and the estimates in (3.14) and (3.15) imply that the first term on right-hand side of (3.16) tends to zero as $n \to \infty$ and the proposition follows. □

In order to prove that the limit $\lim_{n\to\infty}\Delta(\tilde{F}_n)=0$ holds π-a.e. (i.e., the second condition in (3.10)), we need the following elementary estimate.

Lemma 3.8. *The function $\Delta(\tilde{F}_n):\mathbb{R}^d\to\mathbb{R}$, can be bounded above as follows*:

$$\begin{aligned}\left|\Delta(\tilde{F}_n)(x)\right| \le{}& \left|F(x)-F\big(a^n(x)\big)\right| + \left|\pi_n(f_n)-\pi(F)\right| \\ &+ \left|(P\tilde{F}_n-\tilde{F}_n)(x)-(P\tilde{F}_n-\tilde{F}_n)\big(a^n(x)\big)\right| \qquad \textit{for all } x\in\mathbb{R}^d.\end{aligned}$$

Proof. The form $\tilde{F}_n(x)=\sum_{j=0}^{m_n}\hat{f}_n(a_j^n)1_{J_j^n}(x)$ implies $P\tilde{F}_n(x)=\sum_{j=0}^{m_n}\hat{f}_n(a_j^n)P(x,J_j^n)$. The following equalities hold

$$\Delta(\tilde{F}_n)(b)=P(\tilde{F}_n-\hat{F})(b)-(\tilde{F}_n-\hat{F})(b)=\pi_n(f_n)-\pi(F) \qquad \text{for any } b\in X_n, \tag{3.18}$$

since $\hat{F}$ (resp. $\hat{f}_n$) solves the Poisson equation in PE(P, F) (resp. PE(p_n, f_n)). Recall that the function $a^n(\cdot)$ is defined in (3.3). Using the definition of $\Delta(\tilde{F}_n)$, the equalities in (3.18) and the fact that $\hat{F}$ solves PE(P, F) yields

$$\begin{aligned}\Delta(\tilde{F}_n)(x)={}&(\hat{F}-P\hat{F})(x)-(\hat{F}-P\hat{F})\big(a^n(x)\big)+(\hat{F}-P\hat{F})\big(a^n(x)\big)\\&-(\tilde{F}_n-P\tilde{F}_n)\big(a^n(x)\big)+(\tilde{F}_n-P\tilde{F}_n)\big(a^n(x)\big)-(\tilde{F}_n-P\tilde{F}_n)(x)\\={}&F(x)-F\big(a^n(x)\big)+\pi_n(f_n)-\pi(F)+(P\tilde{F}_n-\tilde{F}_n)(x)-(P\tilde{F}_n-\tilde{F}_n)\big(a^n(x)\big)\end{aligned}$$

for all $x\in\mathbb{R}^d$. The triangle inequality implies the lemma. □

Proof of Theorem 2.6. By Proposition 3.2, it is sufficient to verify that the conditions in (3.10) hold for the sequence of functions $(\Delta(\tilde{F}_n))_{n\in\mathbb{N}}$. By Proposition 3.5 there exists a constant ξ' and a sequence $(c_n)_{n\in\mathbb{N}}$ such that the following estimate holds

$$\left|\tilde{F}_n(x)+c_n-\hat{F}(x)\right|\le\xi' V(x) \qquad \text{for all } n\in\mathbb{N} \text{ and } x\in\mathbb{R}^d.$$

Note that we have $\Delta(\tilde{F}_n)=P(\tilde{F}_n+c_n-\hat{F})-(\tilde{F}_n+c_n-\hat{F})$. The structure of the transition kernel P in (MH(q, π)) implies the following bounds for all $n\in\mathbb{N}$ and $x\in\mathbb{R}^d$:

$$\begin{aligned}\left|\Delta(\tilde{F}_n)(x)\right|&\le\int_{\mathbb{R}^d}\left(\left|\tilde{F}_n(y)+c_n-\hat{F}(y)\right|+\left|\tilde{F}_n(x)+c_n-\hat{F}(x)\right|\right)\alpha(x,y)q(x,y)\,dy\\&\le\int_{\mathbb{R}^d}\xi' V(y)\alpha(x,y)q(x,y)\,dy+\xi' V(x)\int_{\mathbb{R}^d}\alpha(x,y)q(x,y)\,dy\\&\le\xi'\big(PV(x)+V(x)\big)\le\big(\xi'+\xi'\lambda_V+\xi'\kappa_V\big)V(x),\end{aligned}$$

where the last inequality is a consequence of the drift condition in A1. This inequality and the definition of the V-norm in (2.1) imply that the first condition in (3.10) is satisfied.

We now establish the limit in (3.10). Fix an arbitrary $x\in\mathbb{R}^d$, such that F is continuous at x. The first term on the right-hand side of the inequality in Lemma 3.8 therefore converges to zero by (3.4). The second term, which is independent of x, tends to zero by Proposition 3.7. In order to deal with the third term on the right-hand side of the inequality in Lemma 3.8, note that, by the definition of $\tilde{F}_n$ in Theorem 2.6, it holds that $\tilde{F}_n(a^n(x)) = \tilde{F}_n(x)$ for all $n \in \mathbb{N}$. Consequently, the structure of the transition kernel P in (MH(q, π)) implies that this term equals $|\int_{\mathbb{R}^d}(\tilde{F}_n(y) - \tilde{F}_n(x))[\alpha(x, y)q(x, y) - \alpha(a^n(x), y)q(a^n(x), y)]\,dy|$. The integrand converges to zero for every $y \in \mathbb{R}^d$ by (3.4) and Assumptions A2–A3. Furthermore, by Proposition 3.5, we obtain the inequality

$$\left|\tilde{F}_n(y) - \tilde{F}_n(x)\right| = \left|\tilde{F}_n(y) + c_n - \tilde{F}_n(x) - c_n\right| \le \xi\big(V(y) + V(x)\big) \qquad \text{for every } y \in \mathbb{R}^d. \tag{3.19}$$

The inequality in (3.13) yields an upper bound

$$\left|\alpha(x, y)q(x, y) - \alpha\big(a^n(x), y\big)q\big(a^n(x), y\big)\right| \le \eta_x\pi(y) + \alpha(x, y)q(x, y) \quad \text{for all } y \in \mathbb{R}^d. \tag{3.20}$$

The product of the right-hand sides in the inequalities (3.19) and (3.20) is integrable over $\mathbb{R}^d$ with respect to $\mu^{\mathrm{Leb}}(dy)$. Hence, the DCT implies that the third term on the right-hand side of the inequality in Lemma 3.8 converges to zero. Therefore, $\lim_{n\to\infty}\Delta(\tilde{F}_n)(x) = 0$ holds for all $x \in \mathbb{R}^d$ at which F is continuous. It only remains to note that, by the assumption on F in Theorem 2.6, this limit holds π-a.e. □

4. The rate of decay of asymptotic variances

Theorem 2.6 states that, under A1-A3, the asymptotic variance σ_n^2 in CLT($\Phi, F + P\tilde{F}_n - \tilde{F}_n$) converges to zero as $n \to \infty$. This section investigates the speed of this convergence. We show that, under suitable Lipschitz and integrability conditions, the rate of decay is bounded above by the slower of the decay rates of the sequences $\pi(V^2 1_{J_0^n})$ and $\delta_n^2 = \delta(\mathbb{X}_n, V)^2$ (see Remark 2.1(i) and equation (2.3), respectively). This result suggests that, when constructing an exhaustive sequence of allotments (see Definition 2.4 above) with respect to the drift function V, we can guarantee fastest rate of decay of the asymptotic variance σ_n^2 when the growth of the bounded set $\mathbb{R}^d \setminus J_0^n$ and the decay of the V-mesh of the partition of $\mathbb{R}^d \setminus J_0^n$ are balanced appropriately (δ_n^2 and $\pi(V^2 1_{J_0^n})$ must be comparable in size as $n \to \infty$).

Theorem 4.1. *Let the assumptions of Theorem 2.6 be satisfied and assume that the conditions*

$$\limsup_{n\to\infty} \delta_n^{-2} \int_{\mathbb{R}^d\setminus J_0^n} \left(\int_{\mathbb{R}^d} \big(V(x) + V(y)\big) Z_n(x, y)\,dy\right)^2 \pi(x)\,dx < \infty, \tag{4.1}$$

$$\limsup_{n\to\infty} \delta_n^{-2} \int_{\mathbb{R}^d\setminus J_0^n} \left|F(x) - F\big(a^n(x)\big)\right|^2 \pi(x)\,dx < \infty \tag{4.2}$$

hold, where $Z_n(x, y)$, for $x, y \in \mathbb{R}^d$, is defined in Proposition 3.7 and the function $a^n(\cdot)$ is given in (3.3). Then there exists a constant $C_0 > 0$ such that

$$\sigma_n^2 \le C_0 \max\left\{\pi\big(V^2 1_{J_0^n}\big), \delta_n^2\right\} \qquad \textit{for all } n \in \mathbb{N}.$$

Theorem 4.1, proved in Section 4.1 below, holds under general conditions that may be hard to verify in specific examples as the functions in (4.1)–(4.2) depend on the drift function V, often not available in closed form. With this in mind we study a broad class of Metropolis–Hastings chains with the property that V can be described in terms of the target density π and conditions (4.1)–(4.2) can be deduced from certain geometric properties of the level sets of π near infinity. Our approach builds on the results in [13,27].

Consider the class of Random walk Metropolis chains in $\mathbb{R}^d$. Put differently, the proposal density takes the form $q(x,y) = q^*(y-x)$ for some density $q^*: \mathbb{R}^d \to \mathbb{R}$. Assume q^* is continuous, strictly positive and bounded. Assume also that the target π is continuously differentiable, positive and satisfies:

$$\lim_{|x|\to\infty} \frac{x}{|x|} \cdot \nabla(\log \pi)(x) = -\infty \quad \text{and} \quad \limsup_{|x|\to\infty} \frac{x}{|x|} \cdot \frac{\nabla\pi(x)}{|\nabla\pi(x)|} < 0. \tag{4.3}$$

Under these assumptions the kernel P in (MH(q,π)) satisfies A1–A3 with a drift function $V_\gamma := c_\gamma \pi^{-\gamma}$ (where c_γ is a constant that ensures $V_\gamma > 1$) for any $0 < \gamma < \frac{1}{2}$ (see [13], Theorems 4.1 and 4.3, and Remark 2.1(iv)). Then the V_γ-radius (see (2.2)) equals $\mathrm{rad}(\mathbb{X}_n, V_\gamma) = \inf_{y\in J_0^n} c_\gamma \pi^{-\gamma}(y)$ and the V_γ-mesh $\delta_{\gamma,n} = \delta(\mathbb{X}_n, V_\gamma)$, defined in (2.3), takes the form

$$\delta_{\gamma,n} = \max\Bigl(\sup_{x\notin J_0^n} \bigl|x - a^n(x)\bigr|, \sup_{x\in\mathbb{R}^d} \bigl(\pi(x)/\pi\bigl(a^n(x)\bigr)\bigr)^\gamma - 1\Bigr). \tag{4.4}$$

The main assumptions in Proposition 4.3 below are:
(i) there exists a function $K_q : \mathbb{R}^d \to \mathbb{R}$ and $\varepsilon_q > 0$ such that

$$\begin{aligned} &\int_{\mathbb{R}^d} K_q(z)\,dz < \infty \quad \text{and} \\ &\bigl|q^*(z) - q^*(\tilde z)\bigr| \le |z-\tilde z| K_q(z) \qquad \text{for all } z, \tilde z \in \mathbb{R}^d \text{ with } |z-\tilde z| < \varepsilon_q; \end{aligned} \tag{4.5}$$

(ii) there exist constants $\beta \in (\frac{1}{2}, 1)$, $c_\beta > 0$ and $\varepsilon_\pi > 0$ such that

$$\bigl|\nabla\pi(\tilde x)\bigr| < c_\beta \pi(x)^\beta \qquad \text{for all } x, \tilde x \in \mathbb{R}^d \text{ with } |x - \tilde x| < \varepsilon_\pi. \tag{4.6}$$

Remark 4.2. Assumption (4.5) is a version of a local Lipschitz condition and holds for many proposals q^* used in practice, for example, normal densities. Assumption (4.6) and condition (4.3) hold for instance, when target density π is proportional to $e^{-p(x)}$, for a polynomial p of degree k with leading order terms p_k satisfying $p_k(x) \to \infty$ as $|x| \to \infty$.

An application of Theorem 4.1 in this setting yields the following result.

Proposition 4.3. *Assume that (4.5)–(4.6) hold and fix $\gamma \in (0, \beta - \frac{1}{2})$. Let $(\mathbb{X}_n)_{n\in\mathbb{N}}$ be an exhaustive sequence of allotments with respect to V_γ defined above. Let $F \in L^\infty_{V_\gamma}$ be a continuously differentiable function satisfying the inequality $|\nabla F(\tilde x)| < c_F \pi^{\gamma - \frac{1}{2}}(x)$ for all $x, \tilde x \in \mathbb{R}^d$ with $|x - \tilde x| < \varepsilon_F$ (for some constants $c_F, \varepsilon_F > 0$). Then there exists a constant $C_\gamma > 0$ such that the* *asymptotic variance* σ_n^2 *in the* $\mathrm{CLT}(\Phi, F + P\tilde{F}_n - \tilde{F}_n)$, *where* $\tilde{F}_n$ *is constructed by the Scheme with input* P, F *and* $\mathbb{X}_n$, *satisfies*

$$\sigma_n^2 \le C_\gamma \max\left(\delta_{\gamma,n}^2, \int_{J_0^n} \pi^{1-2\gamma}(x)\,dx\right) \qquad \textit{for all } n \in \mathbb{N}.$$

Remark 4.4. Any polynomial F, and in fact any function whose gradient grows no faster than a polynomial, satisfies assumptions of Proposition 4.3 for any $\gamma \in (0, \beta - \frac{1}{2})$.

4.1. Proofs

Proof of Theorem 4.1. Proposition 3.2 implies that there exists a constant $C_1 > 1$ such that $\sigma_n^2 \le C_1 \cdot \pi(\Delta(\tilde{F}_n)^2)$ for every $n \in \mathbb{N}$. Thus, $\limsup_{n\uparrow\infty} \sigma_n^2/\pi(\Delta(\tilde{F}_n)^2) < \infty$. Furthermore, the inequality in (3.10) implies that $\limsup_{n\uparrow\infty} \pi(\Delta(\tilde{F}_n)^2 1_{J_0^n})/\pi(V^2 1_{J_0^n}) < \infty$.

Lemma 3.8 yields $\pi(\Delta(\tilde{F}_n)^2 1_{\mathbb{R}^d\setminus J_0^n}) \le 3(T_1(n) + T_2(n) + T_3(n))$, where

$$T_1(n) := \int_{\mathbb{R}^d\setminus J_0^n} \left|(P\tilde{F}_n - \tilde{F}_n)(x) - (P\tilde{F}_n - \tilde{F}_n)\big(a^n(x)\big)\right|^2 \pi(x)\,dx,$$

$$T_2(n) := \int_{\mathbb{R}^d\setminus J_0^n} \left|F(x) - F\big(a^n(x)\big)\right|^2 \pi(x)\,dx \quad \text{and} \quad T_3(n) := \left|\pi_n(f_n) - \pi(F)\right|^2.$$

Assumption (4.2) implies $\limsup_{n\uparrow\infty} T_2(n)/\delta_n^2 < \infty$. The form of the kernel P in $\mathrm{MH}(q,\pi)$ and the fact that $\tilde{F}_n(x) = \tilde{F}_n(a^n(x))$ for all $x \in \mathbb{R}^d$ yield

$$T_1(n) = \int_{\mathbb{R}^d\setminus J_0^n} \left| \int_{\mathbb{R}^d} \big(\tilde{F}_n(y) - \tilde{F}_n(x)\big)\big[\alpha(x,y)q(x,y) - \alpha\big(a^n(x),y\big)q\big(a^n(x),y\big)\big]\,dy \right|^2 \pi(x)\,dx.$$

The inequality in (3.10) therefore yields

$$\limsup_{n\uparrow\infty} T_1(n) \Big/ \int_{\mathbb{R}^d\setminus J_0^n} \left(\int_{\mathbb{R}^d} \big(V(x) + V(y)\big) Z_n(x,y)\,dy\right)^2 \pi(x)\,dx < \infty.$$

Put differently we obtain $\limsup_{n\uparrow\infty} T_1(n)/\delta_n^2 < \infty$.

Note that $T_3(n) = |\pi_n(f_n) - \pi(F)| \le 2|(\pi_n - \pi_n^*)(f_n)|^2 + 2|\pi_n^*(f_n - h_n)|^2$ (recall (3.11)–(3.12)). Since $\pi_n^*(f_n - h_n) = \int_{\mathbb{R}^d}(F(x) - F(a^n(x)))\pi(x)\,dx$, the inequality $F \le \|F\|_V V$ and (3.5) hold, we find

$$\begin{aligned}
\left|\pi_n^*(f_n - h_n)\right|^2 &\le \int_{\mathbb{R}^d} \left|F(x) - F\big(a^n(x)\big)\right|^2 \pi(x)\,dx \\
&\le \|F\|_V^2 \Big(2 + \sup_{n\in\mathbb{N}} \delta_n\Big)^2 \pi\big(V^2 1_{J_0^n}\big) + \int_{\mathbb{R}^d\setminus J_0^n} \left|F(x) - F\big(a^n(x)\big)\right|^2 \pi(x)\,dx.
\end{aligned}$$

Therefore, (4.2) yields $\limsup_{n\uparrow\infty}|\pi_n^*(f_n-h_n)|^2/\max(\pi(V^2 1_{J_0^n}),\delta_n^2)<\infty$. Similarly, inequalities (3.14) and (3.15) in Proposition 3.7 imply

$$\limsup_{n\uparrow\infty}\bigl|(\pi_n-\pi_n^*)(f_n)\bigr|^2\Big/\int_{\mathbb{R}^d}\biggl(\int_{\mathbb{R}^d}\bigl(V(y)+V(x)\bigr)Z_n(x,y)\,dy\biggr)^2\pi(x)\,dx<\infty.$$

Again, splitting the integral with respect to x into the parts over J_0^n and $\mathbb{R}^d\setminus J_0^n$ and applying (4.1), A1 and (3.5) yields $\limsup_{n\uparrow\infty}|(\pi_n-\pi_n^*)(f_n)|^2/\max(\pi(V^2 1_{J_0^n}),\delta_n^2)<\infty$. Hence, $\limsup_{n\uparrow\infty}T_3(n)/\max(\pi(V^2 1_{J_0^n}),\delta_n^2)<\infty$. This concludes the proof of the theorem. $\square$

Proof of Proposition 4.3. Since P, F and $\mathbb{X}_n$ in Proposition 4.3 satisfy the assumptions of Theorem 2.6, we need only to establish that conditions (4.1) and (4.2) in Theorem 4.1 hold for $V=V_\gamma$ and $\delta_n=\delta_{\gamma,n}$, defined just before Proposition 4.3 above. Then, since $\pi(V_\gamma^2 1_{J_0^n})=c_\gamma^2\int_{\mathbb{R}^d}\pi^{1-2\gamma}(x)\,dx$, the proposition will follow by Theorem 4.1.

Start by establishing (4.2). We have $|x-a^n(x)|<\delta_{\gamma,n}$ for every $x\in\mathbb{R}^d\setminus J_0^n$ by (4.4). Consequently, Lagrange's theorem applied to F along a line segment connecting x and $a^n(x)$ yields a point $\tilde{x}^n$ on this segment such that

$$\begin{aligned}\delta_{\gamma,n}^{-2}\int_{\mathbb{R}^d\setminus J_0^n}\bigl|F(x)-F\bigl(a^n(x)\bigr)\bigr|^2\pi(x)\,dx&\le\int_{\mathbb{R}^d\setminus J_0^n}\biggl(\frac{|F(x)-F(a^n(x))|}{|x-a^n(x)|}\biggr)^2\pi(x)\,dx\\&=\int_{\mathbb{R}^d\setminus J_0^n}\bigl|\nabla F\bigl(\tilde{x}^n\bigr)\bigr|^2\pi(x)\,dx\\&\le c_F\int_{\mathbb{R}^d}\pi^{2\gamma-1}(x)\pi(x)\,dx=c_F\int_{\mathbb{R}^d}\pi^{2\gamma}(x)\,dx\end{aligned}$$

holds for a sufficiently large n by assumptions on F. Target π decays supper-exponentially along any ray from the origin and so does $\pi^{2\gamma}$. Thus, the integral $\int_{\mathbb{R}^d}\pi^{2\gamma}(x)\,dx$ is finite and (4.2) follows.

Next, we prove that (4.1) holds. In the setting of a symmetric Random walk Metropolis we have $\alpha(x,y)=\min(1,\pi(x)/\pi(y))$. Let $\mathcal{A}_x:=\{y\in\mathbb{R}^d;\pi(x)\le\pi(y)\}$ and note that $y\in\mathcal{A}_x$ if and only if $\alpha(x,y)=1$ and $V_\gamma(x)\ge V_\gamma(y)$. Recall $Z_n(x,y)=|\alpha(x,y)q^*(y-x)-\alpha(a^n(x),y)q^*(y-a^n(x))|$ and, for any $\mathcal{B}\subseteq\mathbb{R}^d$ and $x\in\mathbb{R}^d$, denote

$$\mathcal{I}_n(x,\mathcal{B}):=\delta_{\gamma,n}^{-2}\biggl(\int_{\mathcal{B}}\bigl(V_\gamma(x)+V_\gamma(y)\bigr)Z_n(x,y)\,dy\biggr)^2.$$

Condition (4.1) is equivalent to $\limsup_{n\to\infty}\int_{\mathbb{R}^d\setminus J_0^n}\mathcal{I}_n(x,\mathbb{R}^d)\pi(x)\,dx<\infty$. With this in mind, we split the integral in $\mathcal{I}_n(x,\mathbb{R}^d)$ into two integrals, depending on which of the disjoint sets $\mathcal{A}_x$ and $\mathcal{A}_x^c$ the point y belongs to (for any $A\subset\mathbb{R}^d$, A^c denotes $\mathbb{R}^d\setminus A$).

Note that it holds

$$\mathcal{I}_n\bigl(x,\mathbb{R}^d\bigr)\le 2\mathcal{I}_n(x,\mathcal{A}_x)+2\mathcal{I}_n\bigl(x,\mathcal{A}_x^c\bigr)\qquad\text{for all }x\in\mathbb{R}^d.$$

For all sufficiently large n, Lagrange's theorem, (4.4) and (4.6) imply that

$$\begin{aligned}\frac{|\pi(a^n(x))-\pi(x)|}{\delta_{\gamma,n}} &\le \frac{|\pi(a^n(x))-\pi(x)|}{|x-a^n(x)|}\\ &\le \left|\nabla\pi\left(\tilde{x}^n\right)\right| \le c_\beta \pi^\beta(x) \qquad \text{for all } x\in\mathbb{R}^d\setminus J_0^n.\end{aligned}\tag{4.7}$$

The following holds for all $x, y\in\mathbb{R}^d$:

$$Z_n(x,y)\le \alpha\big(a^n(x),y\big)\big|q^*\big(y-a^n(x)\big)-q^*(y-x)\big|+q^*(y-x)\big|\alpha(x,y)-\alpha\big(a^n(x),y\big)\big|.\tag{4.8}$$

If $y\in\mathcal{A}_x$ and n is large enough, then for every $x\in\mathbb{R}^d\setminus J_0^n$, using (4.5) and (4.7), the right hand side of (4.8) can be further bounded as follows (note that $\pi(a^n(x))\ge\pi(y)\ge\pi(x)$ is crucial in the analysis of the right term):

$$\begin{aligned}Z_n(x,y) &\le \delta_{\gamma,n}K_q^*(y-x)+q^*(y-x)\frac{|\pi(a^n(x))-\pi(y)|}{\pi(a^n(x))}1_{\{\pi(a^n(x))>\pi(y)\}}(x,y)\\ &\le \delta_{\gamma,n}K_q^*(y-x)+\delta_{\gamma,n}c_\beta q^*(y-x)\pi^{\beta-1}(x).\end{aligned}$$

Since the Lebesgue measure is translation invariant, there exists a constant $c_Z>0$ such that for sufficiently large $n\in\mathbb{N}$ we have

$$\delta_{\gamma,n}^{-1}\int_{\mathcal{A}_x} Z_n(x,y)\,dy < c_Z\pi^{\beta-1}(x) \qquad \text{for all } x\in\mathbb{R}^d\setminus J_0^n.\tag{4.9}$$

As $y\in\mathcal{A}_x$, we have $V_\gamma(x)\ge V_\gamma(y)$. Thus, (4.9) and $2\beta-2\gamma-1>0$ imply the following:

$$\begin{aligned}\int_{\mathbb{R}^d\setminus J_0^n}\mathcal{I}_n(x,\mathcal{A}_x)\pi(x)\,dx &\le \int_{\mathbb{R}^d\setminus J_0^n} 4V_\gamma(x)^2c_Z^2\pi^{2\beta-1}(x)\,dx\\ &= 4c_\gamma c_Z^2\int_{\mathbb{R}^d\setminus J_0^n}\pi^{2\beta-2\gamma-1}(x)\,dx<\infty.\end{aligned}\tag{4.10}$$

If $y\in\mathcal{A}_x^c$ and n is large enough, then for every $x\in\mathbb{R}^d\setminus J_0^n$, using (4.5) and (4.7), we differently bound the right hand side of (4.8) as follows:

$$\begin{aligned}Z_n(x,y) &\le \frac{\pi(y)}{\pi(a^n(x))}\delta_{\gamma,n}K_q^*(y-x)+q^*(y-x)\frac{\pi(y)}{\pi(a^n(x))}\frac{|\pi(a^n(x))-\pi(x)|}{\pi(x)}\\ &\le \delta_{\gamma,n}\frac{\pi(y)}{\pi(a^n(x))}\big(K_q^*(y-x)+c_\beta q^*(y-x)\pi^{\beta-1}(x)\big)\\ &\le \delta_{\gamma,n}c_\pi\frac{\pi(y)}{\pi(x)}\big(K_q^*(y-x)+c_\beta q^*(y-x)\pi^{\beta-1}(x)\big),\end{aligned}\tag{4.11}$$

where $c_\pi := (1+\sup_{n\in\mathbb{N}}\delta_{\gamma,n})^{1/\gamma}$ (note that $\sup_{n\in\mathbb{N}}\sup_{x\in\mathbb{R}^d}\frac{\pi(x)}{\pi(a^n(x))} < c_\pi$ by (4.4)). Hence, similarly to (4.9) there exists a constant $c'_Z > 0$ such that

$$\delta_{\gamma,n}^{-1}\int_{\mathcal{A}_x^c} Z_n(x,y)\,dy < c'_Z\pi^{\beta-1}(x) \qquad \text{for all } x\in\mathbb{R}^d\setminus J_0^n. \tag{4.12}$$

Recall that $V_\gamma(y)\ge V_\gamma(x)$ for $y\in\mathcal{A}_x^c$ and apply the Cauchy–Schwarz inequality to obtain for each $x\in\mathbb{R}^d\setminus J_0^n$ the bound:

$$\begin{aligned}\mathcal{I}_n\big(x,\mathcal{A}_x^c\big) &\le 4\delta_{\gamma,n}^{-2}\int_{\mathcal{A}_x^c} Z_n(x,y)\,dy\cdot\int_{\mathcal{A}_x^c} V_\gamma(y)^2 Z_n(x,y)\,dy \\ &\le 4c'_Z c_\pi\pi^{\beta-1}(x)\int_{\mathcal{A}_x^c} V_\gamma(y)^2\frac{\pi^\beta(y)}{\pi^\beta(x)}\big(c_\beta q^*(y-x)\pi^{\beta-1}(y)+K_q(y-x)\big)\,dy.\end{aligned} \tag{4.13}$$

The second inequality follows by (4.11)–(4.12) and the inequalities $\pi(y)/\pi(x) < 1$ and $\pi(y)^{\beta-1}\ge\pi(x)^{\beta-1}$ for $y\in\mathcal{A}_x^c$ (recall that $\beta\in(1/2,1)$). It is clear that if we substitute $\mathcal{A}_x^c$ with $\mathbb{R}^d$ in (4.13), the inequality remains true. Hence, the Fubini theorem implies

$$\begin{aligned}&\int_{\mathbb{R}^d\setminus J_0^n}\mathcal{I}_n\big(x,\mathcal{A}_x^c\big)\pi(x)\,dx \\ &\quad\le 4c'_Z c_\pi\int_{\mathbb{R}^d} V_\gamma(y)^2\bigg(c_\beta\pi(y)^{\beta-1}\int_{\mathbb{R}^d} q^*(y-x)\,dx \\ &\qquad+\int_{\mathbb{R}^d} K_q(y-x)\,dx\bigg)\pi^\beta(y)\,dy \\ &\quad\le 4c'_Z c_\pi c_\gamma^2\bigg(c_\beta\int_{\mathbb{R}^d}\pi^{2\beta-2\gamma-1}(y)\,dy+\int_{\mathbb{R}^d}\pi^{\beta-2\gamma}(y)\,dy\int_{\mathbb{R}^d} K_q(z)\,dz\bigg) < \infty.\end{aligned} \tag{4.14}$$

Account, that q^* is a density and note that assumptions $\gamma\in(0,\beta-1/2)$ and $\beta\in(1/2,1)$ imply both $\beta-2\gamma, 2\beta-2\gamma-1\in(0,1)$ making the integrals in (4.14) finite. This together with (4.10) implies the inequality $\limsup_{n\to\infty}\int_{\mathbb{R}^d\setminus J_0^n}\mathcal{I}_n(x,\mathbb{R}^d)\pi(x)\,dx < \infty$ and (4.1) follows. $\square$

5. Applications of the *Scheme*

Any implementation of the *Scheme* has to tackle the following two issues: (a) the stochastic matrix $p_\mathbb{X}$ in step (I) of the *Scheme* cannot be computed analytically; (b) once the approximate solution $\tilde{F}_\mathbb{X}$ has been computed, the function $P\tilde{F}_\mathbb{X}$, and thus the control variate $P\tilde{F}_\mathbb{X}-\tilde{F}_\mathbb{X}$, are again not accessible in closed form. In Section 5.1, we present an implementation of the *Scheme*, feasible for general Metropolis–Hastings chains that addresses these issues. In Section 5.2, we apply the method to the symmetric Random walk Metropolis chains with stationary distribution given by a double-well potential (i.e., a mixture of normals). The examples below, satisfying our assumptions, are chosen because they are well-known to converge very slowly in the case of the classical ergodic estimator.

Section 5.2 illustrates two points. First, Example 5.2.1 empirically confirms the arbitrary reduction of the asymptotic variance of the ergodic average in Theorem 2.6 as the partition of the state space is refined sufficiently. Furthermore, the numerical results indicate that the rate of convergence to zero of the asymptotic variance is of the order specified in Theorem 4.1. Second, and perhaps more importantly for future practical applications, Example 5.2.2 demonstrates that an asymptotic variance reduction can be achieved using a coarse partition with few states. This suggests that a similar approach of constructing control variates could be used for reducing the variance of MCMC algorithms in real-world applications and highlights the need for further research on how to efficiently construct weak approximations to the chains of interest in higher dimensions.

5.1. Implementation

Construct a partition $\{J_0, \dots, J_m\}$ with properties: (1) the probability $\pi(J_0)$ is small; (2) it is easy to sample uniform random points from sets J_j for $j \neq 0$. Let $a_j \in J_j$, for $j > 0$, be arbitrary and choose a_0 on the boundary of J_0. One may choose J_0 such that $\mathbb{R}^d \setminus J_0$ contains (most of) the simulated path of the chain. This works well in practice but does not guarantee (1) and makes the partition dependent on the random output.

Given the allotment $(X, \{J_0, \dots, J_m\})$, where $X = \{a_0, \dots, a_m\}$, and the Metropolis–Hastings kernel (MH(q, π)), we have the input required to construct the matrix $p_{\mathbb{X}}$ (step (I) of the *Scheme*). As the precise computation of its entries is not feasible in general, we construct an estimate $\hat{p}_{\mathbb{X}}$ of $p_{\mathbb{X}}$ via i.i.d. Monte Carlo. With this in mind, let $i(x)$ be the unique index $i \in \{0, \dots, m\}$, such that $x \in J_{i(x)}$, and define a random function $\hat{P} : \mathbb{R}^d \times X \to \mathbb{R}_+$ by the formula

$$\hat{P}(x, a_j) := \begin{cases} \frac{1}{n_1} \sum_{l_1=1}^{n_1} \mu^{\mathrm{Leb}}(J_j) \alpha\big(x, Y_{j,x}^{l_1}\big) q\big(x, Y_{j,x}^{l_1}\big), & \text{if } j \notin \{0, i(x)\}, \\ \frac{1}{n_2} \sum_{l_2=1}^{n_2} 1_{J_0}\big(Z_x^{l_2}\big) \alpha\big(x, Z_x^{l_2}\big), & \text{if } j = 0 \neq i(x), \\ 1 - \sum_{k \in \{0,\dots,m\} \setminus \{j\}} \hat{P}(x, a_k), & \text{if } i(x) = j, \end{cases} \tag{5.1}$$

where $n_1, n_2 \in \mathbb{N}$, random vectors $Y_{j,x}^{l_1}$, $l_1 = 1, \dots, n_1$, are i.i.d. uniform in the set J_j for any $j \in \{1, \dots, m\}$ (subscript x indicates that $Y_{j,x}^{l_1}$ are simulated at the point x but does not influence the distribution) and $Z_x^{l_2}$, $l_2 = 1, \dots, n_2$, are i.i.d. random vectors, independent of all $Y_{j,x}^{l_1}$ and distributed according to the proposal distribution $q(x, z)\,dz$ in (MH(q, π)). We construct the matrix $\hat{p}_{\mathbb{X}}$ with entries $(\hat{p}_{\mathbb{X}})_{ij} := \hat{P}(a_i, a_j)$ and use it in the *Scheme* instead of $p_{\mathbb{X}}$.

Given a function $F : \mathbb{R}^d \to \mathbb{R}$, we can execute steps (II)–(III) in the *Scheme*. Constructing the ergodic average estimator $S_k(F + P\tilde{F}_{\mathbb{X}} - \tilde{F}_{\mathbb{X}})$ requires the evaluation of the function $P\tilde{F}_{\mathbb{X}}$ along the simulated path $(\Phi_i)_{i=1,\dots,k}$ of the Metropolis–Hastings chain. We use the form of $\tilde{F}_{\mathbb{X}}$ and the formula in (5.1) to define

$$\hat{P}\tilde{F}_{\mathbb{X}}(x) := \sum_{j=0}^{m} (\hat{f}_{\mathbb{X}})_j \hat{P}(x, a_j) \tag{5.2}$$

for any $x \in \mathbb{R}^d$, where $\hat{f}_{\mathbb{X}}$ is the solution of the system in step (II) of the *Scheme* obtained by solving Poisson's equation PE($\hat{p}_{\mathbb{X}}$, $f_{\mathbb{X}}$). Moreover, the function $\hat{P}\tilde{F}_{\mathbb{X}}$ is used in place of $P\tilde{F}_{\mathbb{X}}$ along the entire path of the chain. Put differently, to estimate $\pi(F)$, we use a modified ergodic estimator $S_k(F + \hat{P}\tilde{F}_{\mathbb{X}} - \tilde{F}_{\mathbb{X}})$ instead of the original one $S_k(F + P\tilde{F}_{\mathbb{X}} - \tilde{F}_{\mathbb{X}})$.

This choice of estimator can be justified as follows: since $Y^{l_1}_{j,\Phi_i}$ and $Z^{l_2}_{\Phi_i}$, generated at each time step i, in the construction of $\hat{P}\tilde{F}_{\mathbb{X}}(\Phi_k)$ are independent of the past $(\Phi_j)_{j=1,\dots,i-1}$, we can construct a Markov chain $\hat{\Phi}$ with augmented state space $\mathbb{R}^d \times (J_1)^{n_1} \times \cdots \times (J_m)^{n_1} \times (\mathbb{R}^d)^{n_2}$, which keeps track of Φ_i and the auxiliary variables $Y^{l_1}_{j,\Phi_i}$ and $Z^{l_2}_{\Phi_i}$. It is not hard to see that the chain $\hat{\Phi}$ has a unique invariant measure $\hat{\pi}$ satisfying $\hat{\pi}(F + \hat{P}\tilde{F}_{\mathbb{X}} - \tilde{F}_{\mathbb{X}}) = \pi(F + P\tilde{F}_{\mathbb{X}} - \tilde{F}_{\mathbb{X}}) = \pi(F)$. Furthermore, $\hat{\Phi}$ is positive Harris recurrent and hence (by [20], Theorem 17.1.7) the SLLN $S_k(F + \hat{P}\tilde{F}_{\mathbb{X}} - \tilde{F}_{\mathbb{X}}) \xrightarrow{k\uparrow\infty} \pi(F)$ a.s. holds for any fixed $n_1, n_2 \in \mathbb{N}$.

Remark 5.1. The estimator $S_k(F + \hat{P}\tilde{F}_{\mathbb{X}} - \tilde{F}_{\mathbb{X}})$ is unbiased in the following sense: if the chain $\hat{\Phi}$ is started from stationarity (i.e., $\hat{\Phi}_0 \sim \hat{\pi}$) we have $\mathrm{E}_{\hat{\pi}}[S_k(F + \hat{P}\tilde{F}_{\mathbb{X}} - \tilde{F}_{\mathbb{X}})] = \pi(F)$ for any $k \in \mathbb{N}$. This should be contrasted with the general approach to variance reduction based on the Poisson equation (PE($\mathcal{P}$, F)), where the estimator $S_k(F)$ of $\pi(F)$ is essential in constructing a guess for the solution of (PE($\mathcal{P}$, F)) and hence the control variate itself (see, e.g., [3] for this approach applied to random scan Gibbs samplers and [4] for sufficiently smooth transition kernels). The latter approach produces a consistent but biased estimator even if the chain is started in stationarity.

In order to analyse numerically the level of improvement due to our implementation of the *Scheme*, denote

$$r_{k,n}(\mathbb{X}) := \frac{\sum_{i=1}^{n}(S^i_k(F) - \pi(F))^2/n}{\sum_{i=1}^{n}(S^i_k(F + \hat{P}\tilde{F}_{\mathbb{X}} - \tilde{F}_{\mathbb{X}}) - \pi(F))^2/n}, \tag{5.3}$$

where n is the number of simulated paths of the chain (started in stationarity at independent starting points) and k is the length of each path. The random vectors $(S^i_k(F), S^i_k(F + \hat{P}\tilde{F}_{\mathbb{X}} - \tilde{F}_{\mathbb{X}}))$, for $i = 1, \dots, n$, are i.i.d. samples of the pair of ergodic average estimators $(S_k(F), S_k(F + \hat{P}\tilde{F}_{\mathbb{X}} - \tilde{F}_{\mathbb{X}}))$ evaluated on the simulated paths. Put differently, $r_{k,n}$ is the ratio of mean square errors of estimators $S_k(F)$ and $S_k(F + \hat{P}\tilde{F}_{\mathbb{X}} - \tilde{F}_{\mathbb{X}})$, numerically evaluated on the same random collection of n independent simulated paths and will serve as an estimate of the improvement.

5.2. Examples

In both examples, we use the target law $\pi := \rho N(\mu_1, \sigma_1^2) + (1 - \rho)N(\mu_2, \sigma_2^2)$, where $N(\cdot, \cdot)$ is a normal distribution of the appropriate dimension.

Table 1. The ratios of improvement $r_{k,n}(\mathbb{X}_m)$ with $n = 1000$ and varying path length k and partition size m

$m \setminus k$	$k = 5 \cdot 10^3$	$k = 2 \cdot 10^4$	$k = 5 \cdot 10^4$	$k = 2 \cdot 10^5$
$m = 30$	5.93	8.56	9.37	9.62
$m = 50$	18.0	32.1	34.2	34.7
$m = 70$	39.1	75.5	96.8	97.1
$m = 100$	76.9	$1.76 \cdot 10^2$	$2.22 \cdot 10^2$	$2.40 \cdot 10^2$
$m = 300$	$6.96 \cdot 10^2$	$1.75 \cdot 10^3$	$2.13 \cdot 10^3$	$2.36 \cdot 10^3$
$m = 500$	$2.14 \cdot 10^3$	$4.64 \cdot 10^3$	$6.05 \cdot 10^3$	$6.92 \cdot 10^3$
$m = 700$	$3.77 \cdot 10^3$	$8.90 \cdot 10^3$	$1.16 \cdot 10^4$	$1.32 \cdot 10^4$

5.2.1. *One dimensional double-well potential*

Let $\mu_1 = -3$, $\sigma_1 = 1$, $\mu_2 = 4$, $\sigma_2 = 1/2$, $\rho = 2/5$. The target density $\pi(\cdot)$ is a mixture of two normal densities with the modes at -3 and 4 which takes values close to zero in the neighbourhood of the origin. Let $F(x) := x^3$ be the force function and let the proposal density $q(x, \cdot)$ be $N(x, 1)$. The assumptions of Theorem 2.6 are satisfied in this example. However, the estimator $S_k(F)$ struggles to converge as the chain tends to get "stuck" under one of the modes for a long time, sampling values of F far away from $\pi(F)$.

Let the allotment $\mathbb{X}_m$ be defined so that $J_0^m := \mathbb{R} \setminus (-8, 7]$ and J_j^m for $j = 1, 2, \dots, m$ are intervals of equal length partitioning $(-8, 7]$. We take a_j^m for $j > 0$ to be the center of the interval J_j^m and we take $a_0^m = -8$. We construct $\hat{p}_{\mathbb{X}_m}$ by the formula in (5.1) (using $n_1 = n_2 = 1000$) and $\hat{P}\tilde{F}_{\mathbb{X}_m} - \tilde{F}_{\mathbb{X}_m}$ by the formulae in (5.1)–(5.2) (using $n_1 = 1, n_2 = 10$) and then use (5.3) to estimate the factor of improvement of the estimator $S_k(F + \hat{P}\tilde{F}_{\mathbb{X}_m} - \tilde{F}_{\mathbb{X}_m})$ in comparison to the estimator $S_k(F)$.

Table 1 shows the ratios of improvement $r_{k,n}(\mathbb{X}_m)$ as the length of the paths varies from $k = 5 \cdot 10^3$ to $2 \cdot 10^5$ and the number of intervals the set $(-8, 7]$ is partitioned into varies from $m = 30$ to $m = 700$. Each entry was computed using an independent sample of $n = 1000$ independent paths of the chain started in stationarity.

The numerical results support Theorem 2.6 as they demonstrate that the algorithm is capable of reducing the asymptotic variance arbitrarily. Note that the rate of the decay of the asymptotic variance (as the mesh of the allotment decreases) in Theorem 4.1 and Proposition 4.3 appears to coincide with the growth of the entries in the columns of the table (as m increases). This suggests that the bound in Theorem 4.1 (as a function of the mesh) is asymptotically sharp.

5.2.2. *Two dimensional double-well potential*

Let $\mu_1 = (-3, 0)$, $\sigma_1^2 = I$, $\mu_2 = (4, 0)$, $\sigma_2^2 = 1/4 \cdot I$, $\rho = 3/5$ (I is a two dimensional identity matrix). Let the force function be $F(x, y) := x$ and let the proposal density $q(x, \cdot)$ be $N(x, I)$. Again, the assumptions of Theorem 2.6 are satisfied.

To specify the allotment, decompose $B := (-7, 6] \times (-4, 4]$ into $6 = 3 \times 2$ equally sized rectangles and define them to be $J_1, J_2, \dots, J_6$. Take $J_0 := \mathbb{R}^2 \setminus B$, $a_0 := (-7, 0)$ and a_j to be the center of the box J_j for $j > 0$. Construct $\hat{p}_{\mathbb{X}_m}$ by the formula in (5.1) (using $n_1 = n_2 = 1000$) and

$\hat{P}\tilde{F}_{\mathbb{X}_m} - \tilde{F}_{\mathbb{X}_m}$ by the formulae in (5.1)–(5.2) (using $n_1 = 1$, $n_2 = 10$) and estimate the factor of improvement $r_{k,n}$ in (5.3). We obtain approximately a 10% reduction in variance. More precisely, we get

$$r_{k,n} = 1.09 \qquad \text{(resp. 1.08) for the path of length } k = 2 \cdot 10^5 \; \big(\text{resp. } k = 5 \cdot 10^4\big),$$

where $n = 1000$ sample paths were used. Moreover, $\pi_{\mathbb{X}}(f_{\mathbb{X}})$ is a poor estimator of $\pi(F)$ as $(\pi_{\mathbb{X}}(f_{\mathbb{X}}) - \pi(F))^2 = 1.52$, while the mean square error of $S_{2\cdot 10^5}(F + \hat{P}\tilde{F}_{\mathbb{X}} - \tilde{F}_{\mathbb{X}})$ is 0.85.

This indicates that a very fine discretisation need not be necessary to achieve variance reduction of MCMC estimators. Analogous implementations, using for example partitions of the state space based on F and π, might lead to variance reduction in higher dimensional models.

Appendix: Existence of exhaustive allotments

Proposition A.1. *Let $W : \mathbb{R}^d \to [1, \infty)$ be a continuous function with bounded sublevel sets, that is, for every $c \in \mathbb{R}$ the pre-image $W^{-1}((-\infty, c])$ is bounded. Then an exhaustive sequence of allotments with respect to W exists.*

Proof. Let $(r_n)_{n\in\mathbb{N}}$ be an increasing unbounded sequence of positive numbers, such that $r_1 > \inf_{x\in\mathbb{R}^d} W(x)$. For each $n \in \mathbb{N}$ define sets $L_n := W^{-1}((-\infty, r_n))$,

$$\tilde{L}_n := \big\{x \in \mathbb{R}^d; \exists y \in L_n, \text{ such that } |x - y| < \sqrt{d}\big\}.$$

Set $\tilde{L}_n$ is bounded and non-empty by definitions of W and r_n. So, W is uniformly continuous on $\tilde{L}_n$. There exists a positive sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ (satisfying $\lim_{n\to\infty} \varepsilon_n = 0$ and $\sup_{n\in\mathbb{N}} \varepsilon_n < 1$) such that $|x - y| < \varepsilon_n\sqrt{d}$ implies $|W(x) - W(y)| < \frac{1}{n}$ for each $n \in \mathbb{N}$ and all $x, y \in \tilde{L}_n$.

Fix $n \in \mathbb{N}$. For $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ denote $K_x^n := [x_1, x_1 + \varepsilon_n) \times \cdots \times [x_d, x_d + \varepsilon_n)$. Clearly, it is possible to pick $x^1, x^2, \dots, x^{m_n} \in \mathbb{R}^d$ so that sets $K_j^n := K_{x^j}^n$ (for $1 \le j \le m_n$) are disjoint and cover L_n (assume the cover is minimal). Finally, take J_0^n to be the closure of $\mathbb{R} \setminus \bigcup_{j=1}^{m_n} K_j^n$ and define $J_j^n := K_j^n \setminus J_0^n$. Note that $\mu^{\mathrm{Leb}}(J_j^n) > 0$ for all $0 \le j \le m_n$. For $1 \le j \le m_n$ pick arbitrary $a_j^n \in J_j^n$ and choose $a_0 \in J_0^n$, so that $W(a_0^n) = \inf_{x\in J_0^n} W(x)$ (possible since W has bounded sublevel sets and J_0^n is closed). Sets J_j^n together with representatives a_j^n define an allotment $\mathbb{X}_n$.

By Pythagoras theorem $|x - y| < \varepsilon_n\sqrt{d}$, for x, y from the same $\in J_j^n$. Since $\varepsilon_n < 1$ and $K_j^n \cap L_n \ne \emptyset$, we get $J_j^n \subset K_j^n \subset \tilde{L}_n$ for all $1 \le j \le m_n$. Hence,

$$\max_{1\le j\le m_n} \sup_{y\in J_j^n} \big|y - a_j^n\big| \le \varepsilon_n\sqrt{d}$$

and by uniform continuity (recall $W \ge 1$)

$$\max_{0\le j\le m_n} \sup_{y\in J_j^n} \frac{W(a_j^n) - W(y)}{W(y)} \le \frac{1}{n}.$$

Doing the above for every $n \in \mathbb{N}$ shows $\lim_{n\to\infty} \delta(\mathbb{X}_n, W) = 0$ (by (2.3)). By (2.2) and definition of L_n, $\mathrm{rad}(\mathbb{X}_n, W) \geq r_n$ for every $n \in \mathbb{N}$. So, $\lim_{n\to\infty} \mathrm{rad}(\mathbb{X}_n, W) = \infty$. $\square$

Acknowledgements

We thank Petros Dellaportas for suggesting the problem, Sean Meyn for helpful discussions and the referees for useful comments that improved the paper. AM was partially supported by the EPSRC grant EP/P003818/1.

References

[1] Andradóttir, S., Heyman, D.P. and Ott, T.J. (1993). Variance reduction through smoothing and control variates for Markov chain simulations. *ACM Trans. Model. Comput. Simul.* **3** 167–189.

[2] Baxendale, P.H. (2005). Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.* **15** 700–738. MR2114987

[3] Dellaportas, P. and Kontoyiannis, I. (2012). Control variates for estimation based on reversible Markov chain Monte Carlo samplers. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 133–161. MR2885843

[4] Devraj, A. and Meyn, S. (2016). Differential TD learning for value function approximation. Available at arXiv:1604.01828v1.

[5] Geyer, C.J. (1992). Practical Markov chain Monte Carlo. *Statist. Sci.* 473–483.

[6] Glynn, P.W. and Meyn, S.P. (1996). A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.* **24** 916–931.

[7] Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109. MR3363437

[8] Henderson, S.G. (1997). Variance Reduction Via an Approximating Markov Process. Ph.D. thesis, Department of Operations Research, Stanford University.

[9] Henderson, S.G. and Glynn, P.W. (2002). Approximating martingales for variance reduction in Markov process simulation. *Math. Oper. Res.* **27** 253–271.

[10] Henderson, S.G., Meyn, S.P. and Tadić, V.B. (2003). Performance evaluation and policy selection in multiclass networks. *Discrete Event Dyn. Syst.* **13** 149–189.

[11] Hernández-Lerma, O. and Bernard Lasserre, J. (1999). *Further Topics on Discrete-Time Markov Control Processes*. *Applications of Mathematics* (*New York*) **42**. New York: Springer.

[12] Hoekstra, A.H. and Steutel, F.W. (1984). Limit theorems for Markov chains of finite rank. *Linear Algebra Appl.* **60** 65–77. MR0749176

[13] Jarner, S.F. and Hansen, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.* **85** 341–361. MR1731030

[14] Kipnis, C. and Varadhan, S.R.S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.* **104** 1–19.

[15] Madras, N. and Randall, D. (2002). Markov chain decomposition for convergence rate analysis. *Ann. Appl. Probab.* **12** 581–606. MR1910641

[16] Makowski, A.M. and Shwartz, A. (2002). The Poisson equation for countable Markov chains: Probabilistic methods and interpretations. In *Handbook of Markov Decision Processes*. *Internat. Ser. Oper. Res. Management Sci.* **40** 269–303. Boston, MA: Kluwer Academic. MR1887206

[17] Mengersen, K.L. and Tweedie, R.L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24** 101–121. MR1389882

[18] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953). Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21** 1087–1092.

[19] Meyn, S. (2008). *Control Techniques for Complex Networks*. Cambridge: Cambridge Univ. Press. MR2372453

[20] Meyn, S. and Tweedie, R.L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge: Cambridge Univ. Press.

[21] Meyn, S.P. and Tweedie, R.L. (1994). Computable bounds for geometric convergence rates of Markov chains. *Ann. Appl. Probab.* **4** 981–1011.

[22] Mijatović, A. (2007). Spectral properties of trinomial trees. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **463** 1681–1696. MR2331531

[23] Mijatović, A. and Pistorius, M. (2013). Continuously monitored barrier options under Markov processes. *Math. Finance* **23** 1–38. MR3015232

[24] Mijatović, A., Vidmar, M. and Jacka, S. (2014). Markov chain approximations for transition densities of Lévy processes. *Electron. J. Probab.* **19** no. 7, 37. MR3164760

[25] Roberts, G.O. and Rosenthal, J.S. (1997). Geometric ergodicity and hybrid Markov chains. *Electron. Commun. Probab.* **2** 13–25. MR1448322

[26] Roberts, G.O. and Rosenthal, J.S. (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv.* **1** 20–71.

[27] 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

[28] Roberts, G.O. and Tweedie, R.L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2** 341–363. MR1440273

[29] Rosenthal, J.S. (1992). Convergence of pseudo-finite markov chains. Unpublished manuscript.

[30] Runnenburg, J.T. and Steutel, F.W. (1962). On Markov chains, the transition function of which is a finite sum of products of functions on one variable: Preliminary report. *Stichting Mathematisch Centrum. Statistische Afdeling* **S304** 1–22.

[31] Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist.* **22** 1701–1762.

Received August 2016 and revised October 2016