ON THE POISSON EQUATION FOR METROPOLIS-HASTINGS CHAINS

ALEKSANDAR MIJATOVIĆ AND JURE VOGRINC

Abstract. This paper defines an approximation scheme for a solution of the Poisson equation of a geometrically ergodic Metropolis-Hastings chain Φ. The approximations give 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 result of the paper shows that the sequence of the asymptotic variances (in the CLTs for the control-variate estimators) converges to zero. We apply the algorithm to geometrically and non-geometrically ergodic chains and present numerical evidence for a significant variance reduction in both cases.

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*

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

with a *force function* $F\colon \mathcal{X} \to \mathbb{R}$ (see [MT09, Sec.17.4]), where 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 [MT09, Sec.17.7, p.459]). 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 [Hen97]).

This paper develops a novel approximation scheme for the solution $\hat{F}$ of (PE($\mathcal{P}$, F)) for the class of Metropolis-Hastings chains and (possibly discontinuous) force functions F satisfying certain general assumptions. This class of Markov chains is of great importance in statistics and other areas of science, see e.g. review papers [RR04, Tie94] and the references therein. In this context, the main motivation for building approximations to $\hat{F}$ is to reduce the asymptotic variance in the CLT of the Markov Chain Monte Carlo (MCMC) estimators based on the Metropolis-Hastings algorithm. The remainder of the introduction is structured as follows: Section 1.1 states our approximation algorithm, Section 1.2 describes the convergence criterion, based on the asymptotic variance in the CLT, and states our main result and Section 1.3 relates our result to the relevant literature and describes the structure of the paper.

2000 *Mathematics Subject Classification.* 60J10, 60J22.

Key words and phrases. Poisson equation for Markov chains, variance reduction, Metropolis-Hastings algorithm, Central Limit Theorem, asymptotic variance, Markov Chain Monte Carlo, weak approximation.

We thank Petros Dellaportas for suggesting the problem and useful conversations. The work was carried out while AM was at Imperial College.

1.1. **Algorithm.** A potential direct approximation approach for computing $\hat{F}$, based on the Poisson equation (PE($\mathcal{P}$, F)) itself, would suffer from at least two problems: (1) in discrete time, the transition kernel $\mathcal{P}$ is typically non-local, implying that the value of the solution $\hat{F}$ at any point in the state space $\mathcal{X}$ may depend on the values of $\hat{F}$ at all other points of the state space (rather than only on the points near by), and (2) the value of the constant $\pi(F)$ is *a priori* unknown. Our proposed algorithm circumvents these issues by exploiting the probabilistic structure underpinning the Poisson equation in (PE($\mathcal{P}$, F)). More precisely, the approximation of $\hat{F}$ is based on the *weak approximation* of the chain Ψ by a sequence of "simpler" Markov chains (converging in law to Ψ), such that the solutions of the Poisson equations for the approximating chains can be characterised algebraically. Our approximation of $\hat{F}$ is expressed in terms of the numerical solution of these linear-algebraic equations. The finite state Markov chain underpinning our algorithm 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 an element in J_i into the set J_j.

Algorithm

Data: Transition kernel $\mathcal{P}$, function F, partition $\{J_0, J_1, \ldots, J_m\}$ of $\mathcal{X}$ and a representative $a_j \in J_j$ for each $j \in \{0, 1, \ldots, m\}$.

Result: Approximate solution $\tilde{F}$ to the Poisson equation in (PE($\mathcal{P}$, F)).

(I) Define a matrix $A \in \mathbb{R}^{(m+1)\times(m+1)}$ with entries A_{ij}, where $i, j \in \{0, 1, \ldots, m\}$ and

$$A_{ij} := \begin{cases} \mathcal{P}(a_i, J_j), & \text{if } i \neq j; \\ -\sum_{k\in\{0,\ldots,m\}\setminus\{i\}} \mathcal{P}(a_i, J_k), & \text{if } i = j; \end{cases}$$

(II) Replace the first column of A by a column of ones: $A_{i0} := 1$, $i = 0, \ldots, m$;

(III) Define a vector $f \in \mathbb{R}^{m+1}$ with entries $f_j := F(a_j)$, $j = 0, \ldots, m$;

(IV) Solve $A\hat{f} = -f$ to find $\hat{f} \in \mathbb{R}^{m+1}$;

(V) Define $\tilde{F} := \sum_{j=1}^{m} \hat{f}_j 1_{J_j}$;

This approximation algorithm is naturally phrased for a general transition kernel $\mathcal{P}$, with $\hat{f}$ in step (IV) being the solution of the Poisson equation for the approximating chain. The convergence analysis under the precise assumptions on the partition of the state space, stated in Section 2, will be carried out for the Metropolis-Hastings kernel P (see (MH(q, π)) below). Numerical examples and the implementation of the algorithm are discussed in Section 5 below.

1.2. **Convergence.** In order to specify the convergence criterion for the successive approximations of $\hat{F}$ produced by the Algorithm, assume that the random sequence $(S_k(F))_{k\in\mathbb{N}}$ satisfies the law of large numbers (LLN), $\lim_{k\to\infty} S_k(F) = \pi(F)$ a.s., and the CLT

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

where $N(0, 1)$ is a standard normal distribution and σ_F^2 is a positive constant known as the *asymptotic variance*. Put differently, the variance of the estimator $S_k(F)$ 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 e.g. [RR04, Sec.5] and the references therein), the variance of the estimator $S_k(F)$ 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 a deterministic starting point π-a.e. and a π-integrable F in (PE($\mathcal{P}$, F)). As mentioned above, solving (PE($\mathcal{P}$, F)) is not feasible, but a good approximate solution $\tilde{F}$ to (PE($\mathcal{P}$, F)) could lead to a significantly reduced asymptotic variance ($\sigma^2_{F+U} \ll \sigma^2_F$) in the CLT($\Psi$, $F+U$), where $U = \mathcal{P}\tilde{F} - \tilde{F}$ (and hence $\pi(F+U) = \pi(F)$). This would reduce the error of the estimator $S_k(F+U)$ (cf. figures in the examples of Section 5).

This method of variance reduction is well-known and has been developed in various Markovian settings [AHO93, Hen97, HG02, HMT03]. Its applications in stochastic networks theory are described in [Mey08, Ch. 11], while applications in statistics for the random scan Gibbs sampler were developed in [DK12]. However, to the best of our knowledge, no systematic approach capable of (at least theoretically) reducing the asymptotic variance arbitrarily for a general class of discrete-time (e.g. Metropolis-Hastings) Markov chains has been developed so far. For example, [DK12] guesses the function G that solves (PE($\mathcal{P}$, F)) in the special case of the random scan Gibbs sampler with a multivariate normal target distribution and the force function $F(x) = x$. It then constructs control variates of the form $\mathcal{P}\tilde{G} - \tilde{G}$, where $\tilde{G}$ is releated to G, for other target distributions and the same F without specifying a procedure to arbitrarily reduce, algorithmically improve or otherwise analyse the achieved variance reduction.

The main contribution of the present paper is to prove that successive applications of the Algorithm can produce approximations $\tilde{F}$ to the solution of the Poisson equation for a Metropolis-Hastings chain Φ on $\mathbb{R}^d$ (with an invariant measure π and a transition kernel P in (MH(q, π))), such that the asymptotic variance in the CLT(Φ, $F + P\tilde{F} - \tilde{F}$) is arbitrarily small for a large class of π-a.e. continuous functions $F \colon \mathbb{R}^d \to \mathbb{R}$. By [GM96, Thm 2.3], the Poisson equation in (PE(P,F)) possesses a solution if we assume that P satisfies a geometric drift condition with a drift function V (see Assumptions A1-A3 below for the precise formulation) and that the force function F is globally bounded by a positive multiple of V. A drift function V in this context is by definition strictly positive and typically “bowl” shaped, i.e. it takes “uniformly” large values on the complements of large compact sets. In order to improve arbitrarily the quality of the approximate solution produced by the Algorithm, assume we have a sequence $\big(\mathbb{J}_n = \{J_0^n, \ldots, J_{m_n}^n\}\big)_{n\in\mathbb{N}}$ of partitions of $\mathbb{R}^d$, such that the set $\mathbb{R}^d \setminus J_0^n = \cup_{j=1}^{m_n} J_j^n$ is bounded for every $n \in \mathbb{N}$. Moreover, assume that the diameter of J_j^n (for any $1 \le j \le m_n$) tends to zero and $\inf_{x\in J_0^n} V(x)$ tends to infinity, as $n \to \infty$. The following theorem gives the main result of the paper (for the precise formulation of the assumptions see Theorem 2.4 in Section 2):

Theorem 1.1. *For each $n \in \mathbb{N}$, let $\tilde{F}_n \colon \mathbb{R}^d \to \mathbb{R}$ be the function obtained by the Algorithm with input: P, F, $\mathbb{J}_n$ and representatives $(a_j^n)_{0\le j\le m_n}$ (i.e. $a_j^n \in J_j^n$) chosen appropriately. Let σ_n^2 denote the asymptotic variance in CLT(Φ, $F + P\tilde{F}_n - \tilde{F}_n$). Then it holds that* $\lim_{n\to\infty} \sigma_n^2 = 0$.

It is clear that the Algorithm does not require the chain Φ to be reversible. We stress here that, likewise, the proof of the main result relies only on the weak approximation of the kernel P in (MH(q, π)) and does not use its reversibility. The precise condition on the choice of representatives in Theorem 1.1 is given in Section 2.2 (see Definition 2.2) and is, by Remark 2.4, a mild technical requirement easily satisfied in applications. The proof of our main result depends crucially on two ingredients: (i) the uniform (in $n \in \mathbb{N}$) convergence to stationarity of the approximating Markov chains, a fact based on the deep results [MT94, Th. 2.3] and [Bax05, Th. 1.1] in the theory of general Markov chains, and (ii) an *a priori* bound of the solution of (PE($\mathcal{P}$, F)) given in [GM96, Thm 2.3]. For an overview of the proof see Section 3.1.

A natural question arising from Theorem 1.1 is about the rate of decay of the sequence of asymptotic variances $\sigma_n^2 \to 0$. Theorem 4.1 below gives an upper bound on this rate. It transpires that, under certain general integrability conditions, the decay is governed by the greater of the two quantities: the mesh of the partition $\{J_1^n, \ldots, J_{m_n}^n\}$ of the bounded set $\mathbb{R}^d \setminus J_0^n$ and the integral $\pi(V^2 1_{J_0^n})$. This result, which we hope is of independent interest, can be used in applications as a guide for balancing the size of the bounded set $\mathbb{R}^d \setminus J_0^n$ and the mesh of its partition $\{J_1^n, \ldots, J_{m_n}^n\}$. Furthermore, in the case of random walk Metropolis chains studied in [RT96a, JH00], Theorem 4.1 yields the rate of decay expressed in terms of the target density π under easier to check sufficient conditions involving π only (see Proposition 4.2).

1.3. **Literature overview.** The construction of control variates using the function $\mathcal{P}G - G$, where G is an approximation of the solution to the Poisson equation, goes back to the PhD thesis [Hen97]. This approach, extended in [HG02, HMT03] and applied in the context of stochastic networks [Mey08], is in spirit close to ours as the construction of G depends on solving Poisson's equation for a related Markov process. The definition of the related Markov process in these contexts relies on the particular model under consideration and it is not immediately clear how to transfer the construction to a more general setting. In contrast, the Algorithm, based on the weak approximation by simple Markov chains, can be applied at least in principle to any discrete time Markov chain with very little modification. Analogous weak approximation ideas have been applied in continuous time in the context of Brownian motion [Mij07], Lévy [MVJ14] and Feller [MP13] processes.

When approximately solving Poisson's equation, one typically chooses basis functions, either by picking a commonly used basis (e.g. [DK12]) or using insight into the structure of the underlying problem (e.g. [Hen97]), and then seeks to represent the solution to Poisson's equation as a linear combination of these basis functions (see [DK12] and the relevant references for a description of this general methodology). In all cases known to us, thus obtained control variates depend in some way on the quantities they are supposed to estimate (recall that $\pi(F)$ features in (PE($\mathcal{P}$, F))), therefore having to rely on their estimation by more basic methods using the sampled path of the chain. Hence, even though they are typically consistent, the resulting estimators introduce a bias even if the chain is started from stationarity. Our method is based purely on the weak approximation of the Metropolis-Hastings chain and does not require an *a* *priori* estimate of $\pi(F)$ in order to construct the control variate, as all the necessary information is contained in the transition matrix approximating the kernel P in (MH(q, π)) (cf. Remark 5.2 in Section 5). As a consequence, when the chain is started in stationarity, the method based on the Algorithm remains unbiased.

The reminder of the paper is organised as follows. Section 2 describes our assumptions, gives examples of widely used Metropolis-Hastings chains satisfying these assumptions and formulates the full version of our main result (Theorem 2.4). In Section 3 we prove Theorem 2.4. The structure of the proof is given in Section 3.1, while Sections 3.2, 3.3, 3.4 and 3.5 carry out the steps. Section 4 formulates and establishes an upper bound on the rate of decay of the asymptotic variance for the Algorithm. Section 5 applies the Algorithm to specific geometrically and non-geometrically ergodic chains and quantifies numerically the variance reduction. It also discusses the construction of the matrix A in the Algorithm. Section 6 concludes the paper.

2. Assumptions and the main result

2.1. **Setting.** 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, i.e. 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

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

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

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

where δ_x is Dirac's measure centred at x, and the Markov chain $(\Phi_k)_{k\in\mathbb{N}}$ generated by it is known as the *Metropolis-Hastings chain* (see [MRR$^+$53, Has70]). 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 $\pi(x)dxP(x,dy) = \pi(y)dyP(y,dx)$) 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 MH(q,π) satisfies the following assumptions:

- A1: There exists a compact set $C_V \subset \mathbb{R}^d$, positive constants $\lambda_V < 1, \kappa_V$ and a π-integrable function $V\colon \mathbb{R}^d \to [1,\infty)$ mapping bounded sets to bounded sets, having bounded sub-level sets (i.e. $V^{-1}([0,c])$ is bounded $\forall c \in \mathbb{R}$) and satisfying

$$PV(x) \le \lambda_V V(x) + \kappa_V 1_{C_V}(x), \quad \forall x \in \mathbb{R}^d.$$

- 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.

Remark 2.1. (i) The inequality in A1 is known as the *geometric drift condition* and is our most important assumption. Under A1, we may without loss of generality assume that the *drift function* V satisfies $\pi(V^2) < \infty$, as we may work with $\sqrt{V}$ instead of V. Indeed, the geometric drift condition and Jensen's inequality imply the inequalities

$$P(\sqrt{V}) \leq \sqrt{PV} \leq \sqrt{\lambda_V V + \kappa_V 1_{C_V}} \leq \sqrt{\lambda_V}\sqrt{V} + \sqrt{\kappa_V} 1_{C_V}.$$

(ii) Assumptions A2 and A3 guarantee that Metropolis-Hastings chain Φ driven by P in MH(q, π) is π-irreducible (i.e μ^{Leb}-irreducible), strongly aperiodic and positive Harris recurrent (see [MT96, Lemmas 1.1, 1.2] and [Tie94, Thm 1, Cor. 2]). We refer the reader to the monograph [MT09] for the definitions of these notions and the theory of Markov Chains based on them. In our setting, their main relevance lies in the fact that, since π is invariant for Φ, they imply the LLN for any π-integrable F [MT09, Thm 17.1.7] and the CLT for any F with modulus bounded by a positive multiple of V (assuming $\pi(V^2) < \infty$) [MT09, Thm 17.4.4].

(iii) Assumptions A2 and A3 can be relaxed somewhat (their current form is chosen to simplify the arguments in Sections 3 and 4). If the state space is an open subset of $\mathbb{R}^d$, q, π continuous μ^{Leb}-a.e., $\exists \epsilon_q > 0$ such that $q(x, y) > \epsilon_q$ for all x, y in a neighbourhood of C_V (C_V is the set in the drift condition in A1) and π bounded away from zero on compact sets, then the main result, Theorem 2.4 below, remains valid. The only difference may be that $\tilde{F}_n$ are well-defined only for all sufficiently large integers n.

(iv) Even if π is known only up to a normalising constant, the kernel P in (MH(q, π)) (and hence the chain Φ) is uniquely defined, as it depends only on the ratio $\pi(y)/\pi(x)$. As a consequence the Algorithm may be applied even if only an unnormalised version of π is known. Furthermore, Theorem 2.4 remains valid in this case.

2.2. **Main result.** Let the drift function V be as in A1 and define the function space

$$\mathrm{L}_V^\infty := \left\{ G\colon \mathbb{R}^d \to \mathbb{R};\, G \text{ measurable and } ||G||_V < \infty \right\}, \quad \text{where} \quad ||G||_V := \sup_{x \in \mathbb{R}^d} \frac{|G(x)|}{V(x)}. \tag{1}$$

Note that L_V^∞ equipped with the norm $||\cdot||_V$ is a Banach space (see [HLL99, Proposition 7.2.1]).

Remark 2.2. Since we are assuming $\pi(V^2) < \infty$ (cf. Remark 2.1), Assumption A1 implies that every $G \in \mathrm{L}_V^\infty$ satisfies the following: $\pi(G^2) < \infty$, $PG(x)$ is well defined for any $x \in \mathbb{R}^d$ (where the transition kernel P is given in MH(q, π)), $PG \in \mathrm{L}_V^\infty$ and $\pi(PG - G) = 0$. In particular, for any $F \in \mathrm{L}_V^\infty$ the LLN (for the chain Φ driven by P) and the CLT(Φ,F) hold (see [MT09, Thms 17.1.7 and 17.4.4] respectively). Hence, for an arbitrary $G \in \mathrm{L}_V^\infty$, the CLT$(\Phi, F + PG - G)$ holds with the same mean $\pi(F)$ as in CLT(Φ,F), but a possibly (substantially) different asymptotic variance σ^2_{F+PG-G}. This motivates the following general definition.

Definition 2.1. Let Ψ be a Markov chain with a transition kernel $\mathcal{P}$ and F a measurable function on its state space $\mathcal{X}$. Let $(G_n)_{n\in\mathbb{N}}$ be a sequence of measurable functions on $\mathcal{X}$, such that the CLT$(\Psi,\ F + \mathcal{P}G_n - G_n)$ holds with the asymptotic variance σ_n^2. We call $(G_n)_{n\in\mathbb{N}}$ a *sequence of approximate solutions of Poisson's equation* PE$(\mathcal{P},\ F)$ if $\lim_{n\to\infty} \sigma_n^2 = 0$.

Remark 2.3. The function $\mathcal{P}G_n - G_n$ does not change, if we shift G_n by a constant. Hence, if $(G_n)_{n\in\mathbb{N}}$ is a sequence of approximate solutions of Poisson's equation PE($\mathcal{P}$, F), then so is $(G_n + c_n)_{n\in\mathbb{N}}$ for any sequence $(c_n)_{n\in\mathbb{N}}$ of real numbers. This observation will be useful in the proof of Theorem 2.4, as solutions of Poisson's equation are unique only up to a constant shift.

The remainder of the section is devoted to defining a sequence of approximate solutions for the Metropolis-Hastings chain generated by P in MH(q,π). This requires the following concepts.

Definition 2.2. **(a)** Let $\mathbb{J}$ be a partition of $\mathbb{R}^d$ into measurable sets $J_0, J_1, \ldots, J_m$, such that $\cup_{j=1}^m J_j$ is bounded and $\mu^{\text{Leb}}(J_j) > 0$ holds for all $0 \le j \le m$. Let $X = \{a_0, a_1, \ldots, a_m\}$ be a set of *represntatives*: $a_j \in J_j$ for all $0 \le j \le m$. The pair $\mathbb{X} := (\mathbb{J}, X)$ is called an *allotment* and let m be the *size* of the allotment $\mathbb{X}$.

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

$$\text{rad}(\mathbb{X}, W) := \inf_{y\in J_0} W(y), \tag{2}$$

$$\delta(\mathbb{X}, W) := \max\left(\max_{1\le j\le m}\sup_{y\in J_j}|y-a_j|, \max_{0\le j\le m}\sup_{y\in J_j}(W(a_j)/W(y)-1)\right), \tag{3}$$

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 if the following limits hold: $\lim_{n\to\infty} \text{rad}(\mathbb{X}_n, W) = \infty$ and $\lim_{n\to\infty} \delta(\mathbb{X}_n, W) = 0$.

Remark 2.4. An allotment $\mathbb{X}$ is a partition of $\mathbb{R}^d$, together with representative points (one in each set), and J_0 is the only unbounded set in the partition. For the W-radius of $\mathbb{X}$ to be large, the union $\cup_{j=1}^m J_j$ of all the bounded sets in the partition has to cover the part of $\mathbb{R}^d$ where W is small. The W-mesh is a maximum of two quantities: the first is a standard mesh of the partition $\{J_1, \ldots, J_m\}$ of the bounded set $\mathbb{R} \setminus J_0 = \cup_{j=1}^m J_j$. The second quantity in (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 (3) is roughly equal to

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

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

Proposition 2.3. *Let $W\colon \mathbb{R}^d \to [1,\infty)$ be a continuous function with bounded sublevel sets. Then an exhaustive sequence of allotments with respect to W exists.*

Remark 2.5. The idea behind the proof of Proposition 2.3 is to use the uniform continuity of W on the set $W^{-1}\big((-\infty, r_n)\big)$ (for a sequence $r_n \uparrow \infty$) to define the n-th partition and its representatives. For a detailed proof see Appendix A below.

Given the transition kernel P in $\mathrm{MH}(q,\pi)$ and an allotment $\mathbb{X}=(\mathbb{J},X)$, we define a stochastic matrix $p_{\mathbb{X}}$ with entries $(0\le i,j\le m)$:

$$(p_{\mathbb{X}})_{ij} := P(a_i,J_j) = \begin{cases} \int_{J_j} \alpha(a_i,y)q(a_i,y)dy & \text{if} \quad i\neq j \\ 1-\int_{\mathbb{R}^d\setminus J_i} \alpha(a_i,y)q(a_i,y)dy & \text{if} \quad i=j. \end{cases} \tag{4}$$

Remark 2.6. Assumptions A2, A3 and Definition 2.2**(a)** ($\mu^{\mathrm{Leb}}(J_j)>0$ for all $0\le j\le m$) imply that all entries of $p_{\mathbb{X}}$ are strictly positive. Hence the chain on the state space X, driven by $p_{\mathbb{X}}$, is irreducible, recurrent, aperiodic and admits a unique invariant probability measure. Moreover, Poisson's equation for $p_{\mathbb{X}}$ and any force function on X possesses a solution, unique up to the addition of a constant function (see [MS02, Theorem 9.3]).

We can now state our main result.

Theorem 2.4. *Let the transition kernel P in MH(q,π) satisfy A1-A3 for a drift function V with $\pi(V^2)<\infty$. Let $F\in L^\infty_V$ 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. For each $n\in\mathbb{N}$ define $p_n:=p_{\mathbb{X}_n}$ and let $f_n\colon X_n\to\mathbb{R}$ be the restriction of F to X_n. Take $\hat{f}_n$ to be the unique solution of Poisson's equation PE(p_n, f_n), which satisfies $\hat{f}_n(a^n_0)=0$. For each $n\in\mathbb{N}$ define a function $\tilde{F}_n\colon\mathbb{R}^d\to\mathbb{R}$ by the formula*

$$\tilde{F}_n(x) := \sum_{j=1}^{m_n} \hat{f}_n(a^n_j)1_{J^n_j}(x) \quad \forall x\in\mathbb{R}^d.$$

Then, $(\tilde{F}_n)_{n\in\mathbb{N}}$ is a sequence of approximate solutions of Poisson's equation PE(P,F) and, for every $n\in\mathbb{N}$, the output of the Algorithm with input P, F and $\mathbb{X}_n$ is well defined and equals $\tilde{F}_n$.

We conclude this section by recalling well-known classes of examples of Metropolis-Hastings chains that satisfy Assumptions A1–A3 of Theorem 2.4.

Example 2.1. *Random walk Metropolis in* $\mathbb{R}$: the proposal density takes the form $q(x,y)=q^*(y-x)$ for some density $q^*\colon\mathbb{R}^d\to\mathbb{R}$. In [MT96] it is shown that geometric ergodicity of Φ (see [RR04, Sec. 3.4] for definition and properties) is essentially equivalent to the tails of the target π being exponential or lighter. More precisely, in [MT96] the following class of target densities on $\mathbb{R}$ was introduced: π is *log-concave in tails* if it is positive everywhere and there exist positive constants β and c such that $\frac{\pi(y)}{\pi(x)}\le e^{-\beta|y-x|}$ for all $y>x>c$ or $y<x<-c$. If π_1 is log-concave in tails and $q_1^*\colon\mathbb{R}\to(0,\infty)$ a positive, continuous, symmetric (i.e. $q_1^*(x)=q_1^*(-x)$) density satisfying $q_1^*(x)\le be^{-\beta x}$ (for some constant $b>0$ and all $x\in\mathbb{R}$), then [MT96, Thm 3.2] implies that for any $0<s<\beta$, the transition kernel P_1 in $\mathrm{MH}(q_1,\pi_1)$ (with the proposal $q_1(x,y)=q_1^*(y-x)$) satisfies A1 with the drift function $V(x)=e^{s|x|}$. Hence, P_1 satisfies A1-A3.

Example 2.2. *Random walk Metropolis in* $\mathbb{R}^d$: this example is based on [RT96a] and [JH00]. If the proposal density $q_2^*\colon\mathbb{R}^d\to\mathbb{R}$ is bounded away from zero in some neighbourhood of the origin and the target π_2 is positive, continuously differentiable and

$$\lim_{|x|\to\infty}\frac{x}{|x|}\cdot\nabla(\log\pi_2)(x)=-\infty \qquad \text{and} \qquad \limsup_{|x|\to\infty}\frac{x}{|x|}\cdot\frac{\nabla\pi_2(x)}{|\nabla\pi_2(x)|}<0 \tag{5}$$

hold, where ∇f denotes the gradient of a differentiable function $f : \mathbb{R}^d \to \mathbb{R}$, then the transition kernel P_2 in $\mathrm{MH}(q_2, \pi_2)$, with $q_2(x, y) = q_2^*(y - x)$, satisfies a geometric drift condition with the drift function $V(x) = c\pi_2^{-1/2}(x)$ (for some constant c that ensures $V \geq 1$), see [JH00, Thms 4.1 and 4.3]. Assuming further that q_2^* is continuous, strictly positive and bounded, the transition kernel P_2 satisfies A1-A3. Intuitively, the target densities satisfying (5) decay uniformly at a sub-exponential rate along any ray from the origin and the radial projection from the level sets $\{\pi = \varepsilon\} \subset \mathbb{R}^d$ to the unit sphere in $\mathbb{R}^d$ is one-to-one for all sufficiently small $\varepsilon > 0$. In particular a density proportional to $e^{-p(x)}$, where $p = p_k + p_{k-1}$ is a polynomial of order k (p_{k-1} is a polynomial of degree at most $k-1$ and p_k consists of the k-th order terms in p) and $p_k(x) \to \infty$ as $|x| \to \infty$, satisfies (5) (see [JH00, Theorem 4.6]).

Example 2.3. *Metropolis adjusted Langevin chain (MALA)*: the proposal process is a Markov chain on $\mathbb{R}^d$ with transition density

$$q_3(x, y) = (2\pi h)^{-d/2} \exp\Big(-|y - x - \frac{h}{2}\nabla(\log \pi_3)(x)|^2/2h\Big),$$

which incorporates information about the target π_3. Put differently, starting from x the proposal follows a normal distribution with mean $x + \frac{1}{2}h\nabla(\log \pi_3)(x)$ and covariance hI_d for some constant $h > 0$. In [RT96b], sufficient conditions for π_3 are given, such that A1 is satisfied with the drift function $V(x) = e^{s|x|}$ (for any sufficiently small $s > 0$). These conditions involve the acceptance region, where moves are accepted a.s., and the behaviour of the mean of the proposal $x + \frac{1}{2}h\nabla(\log \pi_3)(x)$ as $|x| \to \infty$ (see [RT96b, Thm 4.1]). Hence $(\mathrm{MH}(q_3, \pi_3))$ satisfies A1-A3.

Example 2.4. The Metropolis-Hastings chain in this example satisfies the generalised assumptions mentioned in Remark 2.1(iii), but not A1-A3. Consider an exponential target $\pi_4(x) = e^{-x}$ on $(0, \infty)$ and a proposal $q_4(x, y) = \frac{1}{x+1-\max(0,x-1)} 1_{[\max(0,x-1),x+1]}(y)$. Pick $s \in (0, 1)$, define $V(x) := e^{sx}$, and note that the kernel P_4 in $\mathrm{MH}(q_4, \pi_4)$ satisfies

$$P_4V(x) = \lambda V(x) \qquad \text{for any } x > 1, \text{ where } \quad \lambda := 1 - \frac{1}{2}\int_0^1 \left(1 - e^{-sz}\right)\left(1 - e^{-(1-s)z}\right) dz.$$

It is hence clear that $P_4V(x) \leq \lambda V(x) + \kappa 1_{[0,1]}(x)$ holds $\forall x > 0$, where $\kappa := e^{2s} > 0$ and $\lambda \in (0, 1)$. Furthermore, q_4 is μ^{Leb}-a.e. continuous and $q_4(x, y) \geq \frac{1}{2}$ for all $x, y \in [0, 1]$.

3. Proof of Theorem 2.4

3.1. **Overview of the proof.**

The proof of Theorem 2.4 is in two parts. In the first part we establish sufficient conditions for a sequence of functions to form a sequence of approximate solutions to Poisson's equation in the sense of Definition 2.1. This part of the proof, given in Section 3.2 below, relies on an *a priori* bound of the solution of the Poisson equation given in the main result of [GM96, Thm 2.3].

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 2.4, indeed satisfy the sufficient conditions from Section 3.2 (see Sections 3.3 and 3.4 below). The key underlying fact needed for this purpose is that the family of the approximating finite state Markov chains driven by the stochastic matrices $(p_n)_{n\in\mathbb{N}}$, defined in

Theorem 2.4 (cf. (4)), converge to their respective stationary distributions $(\pi_n)_{n\in\mathbb{N}}$ **uniformly** in $n \in \mathbb{N}$. This step is made possible by the deep results in [MT94, Theorem 2.3] and [Bax05, Theorem 1.1] for a general state space Markov chain, 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 (see Theorem 3.4 below for the precise statement of this result). In Section 3.3 we establish the uniform convergence to stationarity of the sequence of our approximating chains (see Corollary 3.5 below) by proving that they satisfy the drift, minorisation and strong aperiodicity conditions uniformly in $n \in \mathbb{N}$ (i.e. the constants in these inequalities do not depend on n, see Proposition 3.3) and applying [Bax05, Theorem 1.1].

In Section 3.4 we show that the sequence $(\tilde{F}_n)_{n\in\mathbb{N}}$ from Theorem 2.4 satisfies Definition 2.1. This proof relies heavily on the uniform convergence to stationarity mentioned above. However, in order to control the asymptotic variance in the CLT(Φ, $F + P\tilde{F}_n - \tilde{F}_n$), 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 (19) for the definition of these Markov chains and their invariant distributions), which is not the case for the stationary laws π_n of the chains generated by the stochastic matrices p_n. The introductory paragraphs of Section 3.4 describe how this new family of chains is used in the proof of Theorem 2.4.

The final step in the proof of Theorem 2.4 consists of showing that the solution to Poisson's equation PE(p_n,f_n) coincides with the corresponding output of the Algorithm. This argument is given in Section 3.5. We conclude this section with a remark on notation.

Remark 3.1. Throughout Section 3 we assume that the transition kernel P in MH(q, π) satisfies Assumptions A1-A3 for a drift function V with $\pi(V^2) < \infty$ (cf. Remark 2.1(i)). In addition to the notation used in Theorem 2.4, throughout the remainder of the section we will use the following objects: a solution $\hat{F} \in \mathrm{L}^\infty_V$ of PE(P,F) (which exists by Theorem 3.1), the restrictions v_n of the drift function V to the state space X_n and the unique invariant probability measure π_n of the stochastic matrix p_n on X_n (see Remark 2.6).

3.2. **Controlling the asymptotic variance.** Let $(G_n : \mathbb{R}^d \to \mathbb{R})_{n\in\mathbb{N}}$ be a sequence of functions in L^∞_V. In this section we give sufficient conditions, in terms of the functions

$$\Delta_n := PG_n - G_n + F - \pi(F), \tag{6}$$

for the asymptotic variance σ_n in the CLT(Φ, $F + PG_n - G_n$) to converge to zero as $n \uparrow \infty$. The key tool we deploy is the deep result in [GM96] on the existence and uniqueness of solutions to Poisson's equation for general Markov chains. For ease of reference we recall [GM96, Prop. 1.1 and Thm 2.3] stated in our setting.

Theorem 3.1. *Let $\mathcal{P}$ be a Markov kernel on $\mathcal{X}$ with unique invariant probability measure π. Let $V : \mathcal{X} \to [1, \infty)$ be a measurable function, $C \subseteq \mathcal{X}$ a measurable set and $\lambda < 1$, κ positive constants such that $\mathcal{P}V(x) \le \lambda V(x) + \kappa 1_C(x)$ holds for all $x \in \mathcal{X}$. Then there exists a positive constant c_V, such that for any force function $F \in L^\infty_V$ (defined as in* (1)*), Poisson's equation*

PE($\mathcal{P}$, F) admits a solution $\hat{F} \in L_V^\infty$ satisfying $||\hat{F}||_V \leq c_V ||F||_V$. If $\hat{F}_1$ is any other π-integrable solution of Poisson's equation, then $\hat{F} - \hat{F}_1$ is a constant π-a.e.

The next result gives sufficient conditions for the functions $(G_n)_{n\in\mathbb{N}}$ to form a sequence of approximate solutions to Poisson's equation. It is stated in the setting of Metropolis-Hastings chains, but Proposition 3.2 holds for any Markov chain satisfying the assumptions of Theorem 3.1 with a virtually identical proof.

Proposition 3.2. *Let $(G_n)_{n\in\mathbb{N}}$ and F be elements of L_V^∞ and, for Δ_n in (6), assume that*

$$\lim_{n\to\infty} \pi\left(\Delta_n^2\right) = 0 \quad \text{and} \quad \sup_{n\in\mathbb{N}} ||\Delta_n||_V < \infty.$$

Then $(G_n)_{n\in\mathbb{N}}$ is a sequence of approximate solutions to PE(P, F). Furthermore, there exists a constant $C_0 > 0$ such that $\sigma_n^2 \leq C_0\sqrt{\pi(\Delta_n^2)}$ for all $n \in \mathbb{N}$.

Proof. For any $n \in \mathbb{N}$, Remark 2.2 and the definition in (6) imply $\Delta_n \in \mathrm{L}_V^\infty$ and $\pi(\Delta_n) = 0$. Theorem 3.1, applied to Poisson's equation in PE(P,$-\Delta_n$), yields a function $H \in \mathrm{L}_V^\infty$, such that

$$PH - H = \Delta_n \qquad \text{and} \qquad ||H||_V \leq c_V ||\Delta_n||_V,$$

for a constant c_V, which is independent of n. Note that $G_n - \hat{F}$ also solves PE(P,$-\Delta_n$) by the definition of Δ_n in (6) and the fact that $\hat{F}$ is a solution of PE(P,F). Since $G_n - \hat{F} \in \mathrm{L}_V^\infty$, Theorem 3.1 implies that there exists $c_n \in \mathbb{R}$ such that the equality $G_n - \hat{F} + c_n = H$ holds π-a.e. Note further that substituting G_n by $G_n + c_n$ in definition (6) does not alter the function Δ_n. Hence, by Remark 2.3, we may assume that $c_n = 0$. This implies the inequality

$$\left|(G_n - \hat{F})(x)\right| \leq c_V \sup_{n\in\mathbb{N}} ||\Delta_n||_V V(x) \quad \text{for all } x \in \mathbb{R}^d.$$

Squaring this inequality, integrating with respect to π, taking a supremum in $n \in \mathbb{N}$ and applying the assumption in the proposition yields

$$\sup_{n\in\mathbb{N}} \pi\left((G_n - \hat{F})^2\right) \leq c_V^2 \pi\left(V^2\right) \sup_{n\in\mathbb{N}} ||\Delta_n||_V^2 < \infty. \tag{7}$$

The asymptotic variance σ_G^2 in CLT(Φ,G) can by [MT09, Theorem 17.4.4] be expressed in terms of any solution $\hat{G} \in \mathrm{L}_V^\infty$ of Poisson's equation PE(P,G) as

$$\sigma_G^2 = \pi\left(\hat{G}^2 - (P\hat{G})^2\right).$$

Trivially, $G_n - \hat{F} \in \mathrm{L}_V^\infty$ is a solution of PE(P, $F + PG_n - G_n$). Hence it holds that

$$\sigma_n^2 = \sigma_{F+PG_n-G_n}^2 = \pi\left((G_n - \hat{F})^2 - (P(G_n - \hat{F}))^2\right). \tag{8}$$

Jensen's inequality and the invariance of π imply

$$\pi\left(\left(P(G_n - \hat{F})\right)^2\right) \leq \pi\left(P\left((G_n - \hat{F})^2\right)\right) = \pi\left((G_n - \hat{F})^2\right). \tag{9}$$

Let $K := G_n - \hat{F}$, $L := P(G_n - \hat{F})$ and note that $K - L = -\Delta_n$ (by (6)), $\pi(L^2) \leq \pi(K^2) < \infty$ (by (7) and (9)) and $\pi(K^2 - L^2) = \sigma_n^2$ (by (8)). Furthermore, Cauchy's inequality yields

$$\pi(K^2 - L^2) = \pi\left((K-L)^2\right) + \pi(2L(K-L)) \leq \pi\left((K-L)^2\right) + 2\left[\pi\left(L^2\right)\pi\left((K-L)^2\right)\right]^{\frac{1}{2}},$$

which implies

$$\sigma_n^2 \leq \pi(\Delta_n^2) + 2\left[\pi(\Delta_n^2)\cdot \sup_{k\in\mathbb{N}} \pi\left((G_k - \hat{F})^2\right)\right]^{\frac{1}{2}}.$$

This, together with (7) and the assumption $\lim_{n\uparrow\infty}\pi(\Delta_n^2) = 0$, concludes the proof. □

3.3. **Uniform convergence to stationarity.** A family of finite state Markov chains, corresponding to the exhaustive sequence of allotments $(\mathbb{X}_n)_{n\in\mathbb{N}}$ (see Definition 2.2) and driven by generator matrices $p_n = p_{\mathbb{X}_n}$ (see (4) for definition), was introduced in the statement of Theorem 2.4. The main aim of this section is to prove that these chains are geometrically ergodic uniformly in $n \in \mathbb{N}$. This constitutes a key step in the proof of Theorem 2.4 and is achieved as follows: first, the *uniform drift, minorisation* and *strong aperiodicity conditions* in (13), (14) and (15), respectively, are established. Second, the uniform convergence to stationarity is deduced from the general result on the convergence of Markov chains in [MT94, Thm 2.3] (see also [Bax05, Thm 1.1]).

Before proving the uniform drift, minorisation and strong aperiodicity conditions, we introduce a function $a^n(\cdot)$, mapping $x \in \mathbb{R}^d$ to the representative (in $\mathbb{X}_n$) of its partition set, and record its basic properties. For each $n \in \mathbb{N}$, let $a^n\colon \mathbb{R}^d \to \mathbb{R}^d$ be defined by

$$a^n(x) := \sum_{j=0}^{m_n} a_j^n 1_{J_j^n}(x) \qquad \text{for every } x \in \mathbb{R}^d, \tag{10}$$

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 of the allotment $\mathbb{X}_n$. Since the sequence of allotments is exhaustive, the following limit holds:

$$\lim_{n\to\infty} a^n(x) = x \qquad \text{for every } x \in \mathbb{R}^d. \tag{11}$$

The definition of a V-mesh (see (3) in Definition 2.2) implies the inequality

$$V(a^n(x)) = V(a^n(x)) - V(x) + V(x) \leq (1+\delta_n)V(x) \qquad \text{for all } n \in \mathbb{N} \text{ and } x \in \mathbb{R}^d, \tag{12}$$

where we denote $\delta_n := \delta(\mathbb{X}_n, V)$.

We are now ready to establish a family of geometric drift conditions, analogous to A1, and a family of related minorisation and strong aperiodicity conditions for transition matrices $p_n = p_{\mathbb{X}_n}$ (given by (4)).

Proposition 3.3. ***Uniform drift, minorisation and strong aperiodicity conditions.*** *Let $(\mathbb{X}_n)_{n\in\mathbb{N}}$ be an exhaustive sequence of allotments with respect to the drift function V from Assumption A1. For each $n \in \mathbb{N}$, define a function $v_n : X_n \to \mathbb{R}$ on the set of representatives of the allotment $\mathbb{X}_n$ by $v_n(a_j^n) := V(a_j^n)$, $j \in \{0, \dots, m_n\}$. Then there exists a compact set $C \subset \mathbb{R}^d$ such that the following statements hold.*

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

$$p_n v_n(a_j^n) \leq \lambda v_n(a_j^n) + \kappa 1_C(a_j^n) \quad \forall n \in \mathbb{N},\ \forall a_j^n \in X_n. \tag{13}$$

(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 for each $n \in \mathbb{N}$, such that the uniform minorisation condition,*

$$(p_n)_{ij} \geq \gamma \nu_n\big(\{a_j^n\}\big) \quad \forall n \in \mathbb{N}, \forall i, j \in \{0, 1, \dots, m_n\} \text{ satisfying } a_i^n \in C_n, \tag{14}$$

and the uniform strong aperiodicity condition,

$$\gamma \nu_n(C_n) \geq \tilde{\gamma} \quad \forall n \in \mathbb{N}, \tag{15}$$

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 (10), we find

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

By (12) we get $V(a^n(y)) - V(a_j^n) \leq V(y) - V(a_j^n) + \delta_n V(y)$ for every $y \in \mathbb{R}^d$. The form of kernel P in (MH(q, π)) and this inequality imply

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

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

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

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 (13) holds for all $n \geq N_0$. Note that if we enlarge C and increase κ, the uniform drift condition in (13) 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 (13) also holds for all $n \in \{1, \dots, N_0 - 1\}$.

(b) Recall that by Definition 2.2**(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)) and Assumption A1, there exists $n_0 \in \mathbb{N}$ such that $D \subseteq \bigcap_{n\geq n_0} V^{-1}\big((-\infty, r_n)\big)$. We now enlarge the compact set C, constructed in part **(a)** of this proof, to contain the bounded set

$$\Big(\bigcup_{n<n_0} \mathbb{R}^d \setminus J_0^n\Big) \cup \bigcap_{n\geq n_0} V^{-1}\big([0, r_n)\big). \tag{16}$$

We may assume the set C is still compact, since the set in (16) is bounded, and hence the uniform drift condition in (13) 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 2.2**(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 (14):

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

We now establish the strong aperiodicity condition in (15). 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 $r_n = \mathrm{rad}(\mathbb{X}_n, V)$ in (2) implies $D \cap J_0^n \subseteq V^{-1}([0, r_n)) \cap V^{-1}([r_n, \infty))$ and hence $D \cap J_0^n = \emptyset$. 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 (10), it holds that $D' \subseteq \cup_{\{j; a_j^n \in C\}} J_j^n$ and

$$\nu_n(C_n) = \nu_n(X_n \cap C) = \nu\left(\cup_{\{j; a_j^n \in C\}} J_j^n\right) \geq \nu(D') = \frac{\mu^{\mathrm{Leb}}(D')}{\mu^{\mathrm{Leb}}(C)} > 0.$$

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 (16) 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 (15) 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. □

The following result for general state space Markov chains (see [MT94, Theorem 2.3] and [Bax05, Theorem 1.1]) is essential for establishing the uniform geometric ergodicity of the Markov chains generated by the transition matrices $p_n = p_{\mathbb{X}_n}$ (see (4)). We state this result because it plays a key role in the proof of Theorem 2.4.

Theorem 3.4. *Let a transition kernel $\mathcal{P}$ on a general state space $\mathcal{X}$ satisfy the assumptions of Theorem 3.1. In particular, assume there exist $\kappa > 0$ and $\lambda \in (0,1)$ such that $\forall x \in \mathcal{X}$ and a measurable $C \subseteq \mathcal{X}$ we have $\mathcal{P}V(x) \leq \lambda V(x) + \kappa 1_C(x)$. Assume further $\exists \gamma, \tilde{\gamma} \in (0, \infty)$ such that the inequalities $\mathcal{P}(x, B) \geq \gamma \nu(B)$, $\forall x \in C$ and all measurable $B \subseteq \mathcal{X}$, and $\gamma \nu(C) \geq \tilde{\gamma}$ hold. Then there exist constants $\zeta > 0, \theta \in (0,1)$,* ***depending only on*** *$\kappa, \lambda, \gamma, \tilde{\gamma}$, such that (recall the definitinon of the V-norm $\|\cdot\|_V$ in (1))*

$$\sup_{\|G\|_V \leq 1} \left|\mathcal{P}^k G(x) - \pi(G)\right| \leq \zeta V(x) \theta^k \qquad \textit{for all } x \in \mathcal{X} \textit{ and } k \in \mathbb{N} \cup \{0\}.$$

Proposition 3.3 and Theorem 3.4 allow us to control the convergence to stationarity of the approximating chains, with transition matrices $p_n = p_{\mathbb{X}_n}$ (defined in (4)), uniformly in $n \in \mathbb{N}$.

Corollary 3.5. *There exist positive constants ζ and $\theta < 1$, such that the inequality*

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

holds, where the notation is as in Proposition 3.3, π_n is the unique invariant probability measure for the Markov chain on X_n generated by the stochastic matirx p_n (cf. Remark 2.6), the supremum is taken over the functions $g : X_n \to \mathbb{R}$ with the v_n-norm, $||g||_{v_n} := \sup_{b \in X_n} |g(b)|/v_n(b)$, bounded above by one and $\pi_n(g)$ denotes the integral (i.e. weighted sum) of g with respect to π_n.

Proof. Pick an arbitrary $n \in \mathbb{N}$ and note that, by Remark 2.6, there exists a unique stationary measure π_n on the state space X_n. According to Proposition 3.3, the transition matrix p_n satisfies the drift condition in (13), the minoriation condition in (14) and the strong aperiodicity condition (15) with the constants $\kappa, \lambda, \gamma, \tilde{\gamma}$, which are independent of the choice of n. Hence, by Theorem 3.4 applied to the transition kernel p_n on the state space X_n, we have $\sup_{\|g\|_{v_n} \leq 1} \left|(p_n^k g)(a_j^n) - \pi_n(g)\right| \leq \zeta(n) v_n(a_j^n) \theta(n)^k$ for every $k \in \mathbb{N} \cup \{0\}$ and $a_j^n \in X_n$ and some positive constants $\zeta(n) \in (0, \infty)$ and $\theta(n) \in (0,1)$, which are in fact independent of n, as they are only a function of $\kappa, \lambda, \gamma, \tilde{\gamma}$ from Proposition 3.3. This concludes the proof. □

3.4. **A sequence of approximate solutions of Poisson's equation PE(P,F).** In this section we prove that the functions $(\tilde{F}_n)_{n \in \mathbb{N}}$, defined in Theorem 2.4 by $\tilde{F}_n = \sum_{j=0}^{m_n} \hat{f}_n(a_j^n) 1_{J_j^n}$, are a family of approximate solutions of PE(P, F) in the sense of Definition 2.1. The function Δ_n, defined in (6), that corresponds to the sequence $(\tilde{F}_n)_{n \in \mathbb{N}}$, can be expressed as

$$\Delta_n = P(\tilde{F}_n - \hat{F}) - (\tilde{F}_n - \hat{F}), \tag{17}$$

where $\hat{F}$ is a solution to PE(P, F) with a finite V-norm (its existence under A1-A3 is implied by Theorem 3.1). As we are assuming $\pi(V^2) < \infty$ (cf. Remark 2.1(i)), Proposition 3.2 implies that $(\tilde{F}_n)_{n \in \mathbb{N}}$ is a family of approximate solutions of PE(P, F) if the following conditions hold:

$$\sup_{n \in \mathbb{N}} ||\Delta_n||_V < \infty \qquad \text{and} \qquad \lim_{n \to \infty} \Delta_n = 0 \quad \pi\text{-a.e.} \tag{18}$$

The first condition in (18) is implied by Proposition 3.6 below, which shows that $\tilde{F}_n$, shifted by a constant, has its V-norm bounded uniformly in $n \in \mathbb{N}$. This result crucially depends on the uniform convergence to stationarity of the approximating chains (see Corollary 3.5).

The second condition in (18) requires bounding $|\Delta_n|$ in (17) by a sum of three non-negative terms (see Lemma 3.8 bellow) and controlling each of them separately. The first, given by $|F(x) - F(a^n(x))|$, tends to zero by (11) since the force function F is assumed to be continuous π-a.e. Proposition 3.6 and the Dominated Convergence Theorem (DCT) are applied to control the second term, which is of the form $|U(x) - U(a^n(x))|$ with $U := P\tilde{F}_n - \tilde{F}_n$. Controlling the difference $|\pi_n(f_n) - \pi(F)|$, which arises naturally as the third term in the bound on $|\Delta_n|$, is more involved (here π_n is the unique invariant measure on the state space X_n of the approximating chain generated by the stochastic matrix p_n and f_n is the restriction of the force function F to X_n). It requires constructing a further approximating chain (based on the transition kernel P) with state space X_n and a transiont matrix p_n^*, whose invariant distribution can be described analytically in terms of the density π (see equation (19) below). Proposition 3.7, whose proof depends on the uniform convergence to stationarity of the approximating chains driven by the transition matrices p_n (see Corollary 3.5), establishes the desired limit.

Proposition 3.6. *Let functions $\tilde{F}_n$, $n \in \mathbb{N}$, be as defined as in Theorem 2.4. Then there exists a positive constant ξ and a sequence of real numbers $(c_n)_{n \in \mathbb{N}}$, such that the following inequality*

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

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

Proof. Pick an arbitrary $n \in \mathbb{N}$. Since the force function F is in L_V^∞ by assumption, its restriction $f_n : X_n \to \mathbb{R}$ ($f_n(b) = F(b)$ for any $b \in X_n$) satisfies $||f_n||_{v_n} \leq ||F||_V$, where v_n is itself the restriction of the drift function V to X_n and the v_n-norm, $\|\cdot\|_{v_n}$, is defined in Corollary 3.5. By the same corollary, the function $\bar{f}_n : X_n \to \mathbb{R}$, given by

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

is well defined and satisfies the inequality $\left|\left|\bar{f}_n\right|\right|_{v_n} \leq \frac{\zeta}{1-\theta}||f_n||_{v_n} \leq \frac{\zeta}{1-\theta}||F||_V$. Furthermore, by [MT09, Thm. 17.4.2], the function $\bar{f}_n$ solves Poisson's equation PE(p_n,f_n). Since $\hat{f}_n : X_n \to \mathbb{R}$, in the definition of $\tilde{F}_n$, also solves PE(p_n,f_n), by Remark 2.6 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 (10) implies $\tilde{F}_n(x) = \hat{f}_n(a^n(x))$. Hence, we obtain

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

and the last inequality follows from (12). 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 (18), we need to define a furhter approximating Markov chain on X_n with the transition matrix p_n^* and the invariant measure π_n^*, given by

$$(19) \qquad (p_n^*)_{ij} := \int_{J_i^n} \frac{\pi(x)}{\pi(J_i^n)} P(x, J_j^n)dx \quad \text{and} \quad \pi_n^*(\{a_j^n\}) := \pi(J_j^n), \quad \text{for } i,j \in \{0,\dots,m_n\},$$

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 (19) 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

$$(20) \qquad h_n(a_j^n) := \int_{J_j^n} \frac{\pi(x)}{\pi(J_j^n)} F(x)dx \quad \text{for } a_j^n \in X_n, \text{ it holds that} \quad \pi_n^*(h_n) = \pi(F).$$

Proposition 3.7. *The following inequalities hold for the measure π_n^* defined in* (19) *and the transition matrix p_n of the approximating chain defined in Theorem 2.4 (cf.* (4)*):*

$$(21) \qquad \left|(\pi_n^* - \pi_n)(f_n)\right| \leq \frac{\zeta||F||_V}{1-\theta} \, ||\pi_n^* - \pi_n^* p_n||_{v_n},$$

where the constants $\theta \in (0,1)$ and $\zeta > 0$ and the measure π_n are as in Corollary 3.5, and

$$(22) \qquad \|\pi_n^* - \pi_n^* p_n\|_{v_n} \leq (1+\sup_{k\in\mathbb{N}}\delta_k) \int_{\mathbb{R}^d\times\mathbb{R}^d} \big(V(y)+V(x)\big) Z_n(x,y)dy\,\pi(x)dx,$$

where $Z_n(x,y) := \left|\alpha(a^n(x),y)q(a^n(x),y) - \alpha(x,y)q(x,y)\right|$ for any $x,y \in \mathbb{R}^d$ and the function $a^n(\cdot)$ is given in (10) *(see Remark 3.2(I) for the definition of $\|\pi_n^* - \pi_n^* p_n\|_{v_n}$). Furthermore, the following limit holds*

$$\lim_{n\to\infty} |\pi_n(f_n) - \pi(F)| = 0.$$

Remark 3.2. **(I)** For a signed measure μ on X_n, define its v_n-norm to be $\|\mu\|_{v_n} := \sup_{\|g\|_{v_n}\leq 1} |\mu(g)|$, where $\mu(g)$ denotes the weighted sum (i.e. the integral) of the values of a function $g : X_n \to \mathbb{R}$ with weights given by μ. This defines the left-hand side of the inequality in (22), which itself plays an important role in the proof of Proposition 3.7. Furthermore, it is natural to define the *dual* normed vector spaces $(\mathrm{L}^\infty_{v_n}, ||\cdot||_{v_n})$ (analogous to L^∞_V in (1)) and $(\mathrm{M}^\infty_{v_n}, ||\cdot||_{v_n})$ of functions on X_n (with the norm defined in Corollary 3.5) and signed measures on X_n (with the norm defined above). Note that since X_n is finite, the vector spaces $\mathrm{L}^\infty_{v_n}$ and $\mathrm{M}^\infty_{v_n}$ are isomorphic to the Euclidean space of dimension given by the cardinality of X_n. Furthermore, any linear map $B : \mathrm{L}^\infty_{v_n} \to \mathrm{L}^\infty_{v_n}$, $g \mapsto Bg$, induces a linear map on the dual $B^* : \mathrm{M}^\infty_{v_n} \to \mathrm{M}^\infty_{v_n}$, $\mu \mapsto B^*\mu := \mu B$ (in this definition we interpret μ as a row vector and B as a matrix).

It is well known that the oparator norms conicide $\|B\|_{v_n} = \|B^*\|_{v_n}$.[1] Indeed, for any $b \in X_n$, the measure $\mu_b(f) := |f(b)|/v_n(b)$ satisfies $\|\mu_b\|_{v_n} = 1$ and hence it holds that $\|Bg\|_{v_n} \leq \|B^*\|_{v_n}$, where $g \in \mathrm{L}^\infty_{v_n}$, $\|g\|_{v_n} \leq 1$, since $|Bg(b)|/v_n(b) = |\mu_b(Bg)| = |(B^*\mu_b)(g)| \leq \|B^*\mu_b\|_{v_n} \leq \|B^*\|_{v_n}$. Hence, $\|B\|_{v_n} \leq \|B^*\|_{v_n}$. To get the opposite inequality, note that $\|B^*\mu\|_{v_n} \leq \|B\|_{v_n}$ for any $\mu \in \mathrm{M}^\infty_{v_n}$ with $\|\mu\|_{v_n} \leq 1$, since $|B^*\mu(g)| = \|B\|_{v_n}|\mu(Bg/\|B\|_{v_n})| \leq \|B\|_{v_n}$ for all $g \in \mathrm{L}^\infty_{v_n}$ with $\|g\|_{v_n} \leq 1$. Hence it holds that $\|B\|_{v_n} \geq \|B^*\|_{v_n}$, implying the stated equality of the operator norms. This fact, which holds in a much more general Banch space setting (see e.g. [HLL99, 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 $n \in \mathbb{N}, y \in \mathbb{R}^d$:

$$\alpha(a^n(x), y)q(a^n(x), y) \leq \frac{q(y, a^n(x))}{\pi(a^n(x))}\pi(y) \leq \eta_x\pi(y), \qquad \text{where } \eta_x := \frac{\sup_{z,y\in\mathbb{R}^d} q(z, y)}{\inf_{n\in\mathbb{N}} \pi(a^n(x))}. \tag{23}$$

By (11) and A2 we have $0 < \inf\{\pi(z) : |z - x| \leq \sup_{k\in\mathbb{N}} \delta_k\} \leq \pi(a^n(x))$, where $\delta_k = \delta(\mathbb{X}_k, V)$ (see Definition 2.2), for all sufficiently large $n \in \mathbb{N}$. Thus, by A2 and A3, we have $\eta_x \in (0, \infty)$ and the inequalities in (23), which will be used in the proofs of Proposition 3.7 and Theorem 2.4, hold.

Proof of Proposition 3.7. We can estimate the difference $|\pi_n(f_n) - \pi(F)|$ using the invariant distribution π^*_n of the chain driven by the stochastic matrix p^*_n and the function h_n, defined in (19) and (20) respectively, as follows

$$\begin{aligned} |\pi_n(f_n) - \pi(F)| &= |\pi_n(f_n) - \pi^*_n(f_n) + \pi^*_n(f_n) - \pi^*_n(h_n)| \\ &\leq |(\pi_n - \pi^*_n)(f_n)| + |\pi^*_n(f_n - h_n)|. \end{aligned} \tag{24}$$

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 (19) and (20) above) and the function $a^n(\cdot)$ (see (10)) imply that the second term on the right-hand side of (24) takes the form

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

[1]Recall that $\|B\|_{v_n} = \sup\{\|Bg\|_{v_n} : g \in \mathrm{L}^\infty_{v_n}, \|g\|_{v_n} \leq 1\}$ and $\|B^*\|_{v_n} = \sup\{\|B^*\mu\|_{v_n} : \mu \in \mathrm{L}^\infty_{v_n}, \|\mu\|_{v_n} \leq 1\}$.

Since F is continuous π-a.e., the integrand converges to zero π-a.e. by (11). Furthermore, for any $x \in \mathbb{R}^d$ it holds that

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

where the last inequality follows from (12). Therefore, by the DCT (recall that by the assumption in A1 we have $\pi(V) < \infty$), the second term in (24) indeed converges to zero.

Establishing the convergence of the first term on the right-hand side in (24) is more involved. We start by establishing the following representation of the signed measure $\pi_n^* - \pi_n$.

Claim. There exists a linear map $B_n : \mathrm{L}^\infty_{v_n} \to \mathrm{L}^\infty_{v_n}$, with the dual $B_n^* : \mathrm{M}^\infty_{v_n} \to \mathrm{M}^\infty_{v_n}$, satisfying

$$\pi_n^* - \pi_n = B_n^* \left(\pi_n^* - \pi_n^* p_n\right) = \left(\pi_n^* - \pi_n^* p_n\right) B_n \quad \text{and} \quad \|B_n^*\|_{v_n} = \|B_n\|_{v_n} \leq \frac{\zeta}{1-\theta},$$

where the constants $\theta \in (0,1)$ and $\zeta > 0$ are as in Corollary 3.5. For the definition of the vector spaces $\mathrm{L}^\infty_{v_n}$ and $\mathrm{M}^\infty_{v_n}$ and the respective norms see Remark 3.2(I).

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 rvs with the law given by π_n (independently of the starting distribution). The inequality in Corollary 3.5 can therefore be expressed as

$$\|p_n^k - 1 \otimes \pi_n\|_{v_n} \leq \zeta\theta^k, \quad \text{for all } k \in \mathbb{N} \cup \{0\}, \text{ implying that} \quad B_n := \sum_{k=0}^{\infty} \left(p_n^k - 1 \otimes \pi_n\right)$$

is a well defined linear map on the normed space $\mathrm{L}^\infty_{v_n}$, such that $\|B_n\|_{v_n} \leq \zeta/(1-\theta)$. In order to establish the first equality in the Claim above, note that $\mu(1 \otimes \pi_n) = \pi_n$ for any probability measure $\mu \in \mathrm{M}^\infty_{v_n}$ and, by Remark 3.2(I) and Corollary 3.5, the $\|\cdot\|_{v_n}$-norm of the linear operator $\mu \mapsto \mu(p_n^k - 1 \otimes \pi_n)$ on $\mathrm{M}^\infty_{v_n}$ 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 $\|\pi_n^* p_n^k - \pi_n\|_{v_n} = \|\pi_n^*(p_n^k - 1 \otimes \pi_n)\|_{v_n} \leq \zeta\theta^k ||\pi_n^*||_{v_n}$ for all $k \in \mathbb{N}$. Consider the identitiy

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

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

In order to establish the inequality in (21), note that $\|f_n\|_{v_n} \leq ||F||_V$ and Remark 3.2(I) imply

$$\left|(\pi_n^* - \pi_n)(f_n)\right| \leq \|F\|_V (\pi_n^* - \pi_n)(f_n/\|f_n\|_{v_n}) \leq ||F||_V ||\pi_n^* - \pi_n||_{v_n}.$$

This inequality and the Claim imply (21).

The next task is to prove (22). Let $g \colon X_n \to \mathbb{R}$ be a function satisfying $\|g\|_{v_n} \leq 1$. Recall that $m_n + 1$ is the cardinality of X_n and that the function $a^n(\cdot)$ is defined in (10). We apply the definitinons of the stochastic matrix p^* and its stationary law π^*, given in (19), to obtain

$$
\begin{aligned}
&(\pi_n^* - \pi_n^* p_n)\, g = \pi_n^* \, (p_n^* - p_n)\, g = \sum_{j=0}^{m_n}\sum_{i=0}^{m_n} \big[\pi(J_i^n)\,((p_n^*)_{ij} - (p_n)_{ij})\,\big] g(a_j^n) \\
&= \sum_{j=0}^{m_n} \left[\int_{\mathbb{R}^d} \big(P(x, J_j^n) - P(a^n(x), J_j^n)\big)\, \pi(x) dx\right] g(a_j^n) \\
&= \int_{\mathbb{R}^d} \left(\sum_{j=0}^{m_n} \int_{J_j^n} \big[\alpha(x,y)q(x,y) - \alpha(a^n(x),y)q(a^n(x),y)\big] g(a_j^n) dy \right) \pi(x)dx \\
&\ + \int_{\mathbb{R}^d} \left[\sum_{j=0}^{m_n} \left[\left(1 - \int_{\mathbb{R}^d} \alpha(x,z)q(x,z)dz\right) \delta_x(J_j^n) \right.\right. \\
&\ \left.\left. - \left(1 - \int_{\mathbb{R}^d} \alpha(a^n(x),z)q(a^n(x),z)dz\right) \delta_{a^n(x)}(J_j^n)\right] g(a_j^n)\right] \pi(x)dx \\
&= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} g(a^n(y)) \big[\alpha(x,y)q(x,y) - \alpha(a^n(x),y)q(a^n(x),y)\big] dy \right) \pi(x)dx \\
&\ + \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} g(a^n(x)) \big[\alpha(a^n(x),y)q(a^n(x),y) - \alpha(x,y)q(x,y)\big] dy \right) \pi(x)dx,
\end{aligned}
$$

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

$$
||\pi_n^* - \pi_n^* p_n||_{v_n} = \sup_{\|g\|_{v_n}\le 1} |\,(\pi_n^* - \pi_n^* p_n)\, g| \le \int_{\mathbb{R}^d\times\mathbb{R}^d} \big(V(a^n(y)) + V(a^n(x))\big) Z_n(x,y)\pi(x)dydx,
$$

which, together with (12), implies (22).

We now apply the DCT to deduce that the right-hand side in (22) 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 (12) imply the estimates

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

for all $x \in \mathbb{R}^d$. Since, by Assumption A1, we have $\pi(V) < \infty$, by the DCT the right-hand side in (22) tends to zero (as $n \to \infty$) if

$$
\lim_{n\to\infty} \int_{\mathbb{R}^d} \big(V(y) + V(x)\big) Z_n(x,y)dy = 0 \quad \forall x \in \mathbb{R}^d. \tag{25}
$$

To establish the limit in (25), 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 (11) and the assumptions in A2 and A3. Hence the integrand in (25) converges to zero point-wise. By the estimate in (23), the integrand in (25) is bounded above by the function

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

which does not depend on n and is μ^{Leb}-integrable in $y \in \mathbb{R}^d$. Hence the limit in (25) holds by the DTC and, consequently, the right-hand side in (22) converges to zero as $n \to \infty$. This fact and the estimates in (21) and (22) imply that the first term on right-hand side of (24) tends to zero as $n \to \infty$ and the proposition follows. □

In order to prove that the limit $\lim_{n\to\infty} \Delta_n = 0$ holds π-a.e. (i.e. the second condition in (18)), where the function Δ_n is given in (17), we need the following elementary estimate.

Lemma 3.8. *The function $\Delta_n : \mathbb{R}^d \to \mathbb{R}$, given in* (17)*, can be bounded above as follows:*

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

Proof. Recall from the statement of Theorem 2.4 that $\tilde{F}_n(x) = \sum_{j=0}^{m_n} \hat{f}_n(a_j^n) 1_{J_j^n}(x)$. Hence it holds that $P\tilde{F}_n(x) = \sum_{j=0}^{m_n} \hat{f}_n(a_j^n) P(x, J_j^n)$. The following equalities hold

$$\Delta_n(b) = P(\tilde{F}_n - \hat{F})(b) - (\tilde{F}_n - \hat{F})(b) = \pi_n(f_n) - \pi(F) \qquad \text{for any } b \in X_n, \tag{26}$$

since $\hat{F}$ (resp. $\hat{f}_n$) solves the Poisson equation in PE(P,F) (resp. PE(p_n,f_n)), the transition matrix p_n takes the form in (4) for the state space X_n and, for any $a_j^n \in X_n$, by definitnion it holds $f_n(a_j^n) = F(a_j^n)$. Recall that the function $a^n(\cdot)$ is defined in (10). Applying the definition of Δ_n in (17), the equalities in (26) and the fact that $\hat{F}$ solves PE(P,F) yields

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

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

Proof of Theorem 2.4: $(\tilde{F}_n)_{n\in\mathbb{N}}$ ***are an approximate solution for PE(P,F).*** By Proposition 3.2, it is sufficient to verify that the conditions in (18) hold for the sequence of functions $(\Delta_n)_{n\in\mathbb{N}}$ in (17). By Theorem 3.1 the solution of the Poisson equation PE(P,F) satisfies $\hat{F} \in \mathrm{L}_V^\infty$. Hence, Proposition 3.6 implies the existance of a constant ξ' and a real sequence $(c_n)_{n\in\mathbb{N}}$ such that the following estimate holds

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

Note that by (17) we have $\Delta_n = P(\tilde{F}_n + c_n - \hat{F}) - (\tilde{F}_n + c_n - \hat{F})$. Hence the structure of the transition kernel P in (MH(q, π)) implies the following bounds for all $n \in \mathbb{N}$ and $x \in \mathbb{R}^d$:

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

where the last inequality is a consequence of the drift condition in A1. Hence, by the definition of the V-norm in (1), we find that $\sup_{n\in\mathbb{N}} ||\Delta_n||_V < \infty$, which is the first condition in (18).

The second condtion in (18) stipulates $\lim_{n\to\infty} \Delta_n = 0$ π-a.e. Fix an arbitrarty $x \in \mathbb{R}^d$, 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 (11). 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 2.4, 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 $|\int_{\mathbb{R}^d}(\tilde{F}_n(y) - \tilde{F}_n(x))\big[\alpha(x,y)q(x,y) - \alpha(a^n(x),y)q(a^n(x),y)\big]dy|$. The integrand converges to zero for every $y \in \mathbb{R}^d$ by (11) and Assumptions A2–A3. Furthermore, by Proposition 3.6, we obtain the inequality

$$\left|\tilde{F}_n(y) - \tilde{F}_n(x)\right| = \left|\tilde{F}_n(y) + c_n - \tilde{F}_n(x) - c_n\right| \le \xi\,(V(y) + V(x)) \quad \text{for every } y \in \mathbb{R}^d. \tag{27}$$

The inequality in (23) yields an upper bound

$$|\alpha(x,y)q(x,y) - \alpha(a^n(x),y)q(a^n(x),y)| \le \eta_x \pi(y) + \alpha(x,y)q(x,y) \quad \text{for all } y \in \mathbb{R}^d. \tag{28}$$

The product of the right-hand sides in the inequalities (27) and (28) is integrable over $\mathbb{R}^d$ with respect to $\mu^{\mathrm{Leb}}(dy)$, since $\pi(V) < \infty$ (see A1 and the definition of α in Section 2.1). Hence, the DCT implies that the third term on the right-hand side of the inequality in Lemma 3.8 converges to zero. Therefore, $\lim_{n\to\infty} \Delta_n(x) = 0$ 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 2.4, this limit holds π-a.s. □

3.5. **Compatibility with the Algorithm.** Let A_n be the matrix appearing in the Algorithm with the input P, F and $\mathbb{X}_n$. The first task is to show that A_n is non-singular.

Proposition 3.9. *Let p be a transition matrix for an irreducible Markov chain with a unique invariant probability measure μ on a state space with $\ell \in \mathbb{N}$ elements. Then the vector $\mathbf{1} \in \mathbb{R}^\ell$, with all the coordinates equal to one, is not in the image of $p - I$ (where I is the identity matrix of size ℓ) and any collection of $\ell - 1$ columns of $p - I$ is linearly independent.*

Remark 3.3. By Proposition 3.9 and Remark 2.6, under Assumptions A2 and A3, the Algorithm produces a well-defined output for each allotment in an exhaustive sequence $(\mathbb{X}_n)_{n\in\mathbb{N}}$ (see Definition 2.2**(c)**). Furthermore, Proposition 3.9 implies that the Algorithm will produce a well defined output for a much broader class of Markov Chains with a general transition kernel $\mathcal{P}$.

Proof. Interpret μ as a row vector with non-negative coordinates such that $\mu\mathbf{1} = 1$ and $\mu p = \mu$. It holds that μ is a left eigenvector of $p - I$ for the eigenvalue zero. If $\exists x \in \mathbb{R}^\ell$, such that $(p - I)x = \mathbf{1}$, we would get $0 = (\mu(p - I))x = \mu\mathbf{1} = 1$. Hence $\mathbf{1}$ is not in the image of $p - I$.

Since the chain is irreducible, all the entries of μ are strictly positive. If μ' is another left eigenvector of p, so is $\mu' + \beta\mu$ for any large $\beta > 0$. Since the invariant measure μ is unique, μ' is hence proportional to μ and the rank of $p - I$ is $\ell - 1$. Moreover $\ker(p - I) := \{x \in \mathbb{R}^\ell : px = x\}$ equals $\ker(p - I) = \{\lambda\mathbf{1} : \lambda \in \mathbb{R}\}$ and the proposition follows. □

***Proof of Theorem 2.4: Algorithm with input** P**, **F** and **$\mathbb{X}_n$** produces** $\tilde{F}_n$.* The matrix A_n in the Algorithm is equal to $p_n - I$ with the first column replaced by a column of ones, where p_n is the stochastic matrix defined in Theorem 2.4 (see also (4)) and I is the identity matrix of dimension $1+m_n$. By Remark 3.3, the Algorithm returns the unique solution vector $\hat{f}$ of the system $A_n\hat{f} = -f_n$, where the function f_n is identified with a column vector.

Since p_n is irreducible, the invariant measure π_n charges all the points in X_n. Hence, by Theorem 3.1, $\exists!\ \hat{f}_n : X_n \to \mathbb{R}$ satisfying PE(p_n, f_n) and $\hat{f}_n(a_0^n) = 0$ (recall $X_n = \{a_0^n, \dots, a_{m_n}^n\}$). We need to show that $\hat{f}_i = \hat{f}_n(a_i^n)$ for $1 \le i \le m_n$, where $\hat{f}_i$, $i \in \{0, \dots, m_n\}$, are the coordinates of $\hat{f}$ solving $A_n\hat{f} = -f_n$. Poisson's equation $(p_n - I)\hat{f}_n = \pi_n(f_n)\mathbf{1} - f_n$ can be viewed as a linear system of m_n+1 equations with m_n+1 unknowns $\hat{f}_n(a_i^n)$, $i \in \{1, \dots, m_n\}$, and $\pi_n(f_n)$:

$$-\pi_n(f_n) + \sum_{j=1}^{m_n} (p_n)_{ij}\hat{f}_n(a_j^n) - \hat{f}_n(a_i^n) = -f_n(a_i^n) \qquad i \in \{0, \cdots, m_n\}.$$

Hence $\hat{g} \in \mathbb{R}^{1+m_n}$, given by $\hat{g}_0 := -\pi_n(f_n)$ and $\hat{g}_i := \hat{f}_n(a_i^n)$, $1 \le i \le m_n$, satisfies $A_n\hat{g} = -f_n$ for A_n from the Algorithm. Since A_n is non-singular (by Proposition 3.9), the proof is complete. □

4. The rate of decay of asymptotic variances

Theorem 2.4 states that, under A1-A3, the asymptotic variance σ_n^2 in CLT$(\Phi, F + P\tilde{F}_n - \tilde{F}_n)$ tends to zero as $n \uparrow \infty$. 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 $\delta_n = \delta(\mathbb{X}_n, V)$ (see Remark 2.1(i) and Equation (3) respectively). This result suggests that, when constructing an exhaustive sequence of allotments (see Definition 2.2 above) with respect to the drift function V in A1, it is optimal to balance the growth of the size of the bounded set $\mathbb{R}^d \setminus J_0^n$ and the mesh of the partition of $\mathbb{R}^d \setminus J_0^n$ in such a way that δ_n and $\pi(V^2 1_{J_0^n})$ are comparable in size.

Theorem 4.1. *Let the assumptions of Theorem 2.4 be satisfied and assume that the conditions*

$$\limsup_{n\to\infty} \int_{\mathbb{R}^d \setminus J_0^n} \int_{\mathbb{R}^d} \left(V(x)^2 + V(x)V(y)\right) \frac{Z_n(x,y)}{\delta_n} \pi(x)dydx < \infty, \tag{29}$$

$$\limsup_{n\to\infty} \int_{\mathbb{R}^d \setminus J_0^n} V(x) \frac{|F(x) - F(a^n(x))|}{\delta_n} \pi(x)dx < \infty \tag{30}$$

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 (10)*. Then there exists a constant $C_0 > 0$ such that*

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

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 featuring in (29)–(30) depend on each other in a rather complicated way and an appropriate drift function V is 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 (29)–(30) can be deduced from certain geometric properties of the level sets of π near infinity. Our approach builds on the work in [RT96a] and [JH00], mentioned in Example 2.2 above.

More precisely, let π and q^* be as in Example 2.2, so that the kernel P in (MH(q,π)) (with $q(x,y) = q^*(y-x) = q^*(x-y)$) generates a symmetric random walk Metropolis chain in $\mathbb{R}^d$ and satisfies assumptions A2 and A3, while (5) holds for the differentiable target π. The kernel P in (MH(q,π)) hence satisfies A1-A3 with a drift function proportional to $\pi^{-1/2}(x)$ (see [JH00, Thms 4.1 and 4.3] and Example 2.2 above). By an argument analogous to the one in Remark 2.1(i), we can take the drift function to be $V_\gamma(x) := c_\gamma \pi^{-\gamma}(x)$ for any $0<\gamma<\beta/2$ and some $c_\gamma>0$, ensuring that $V_\gamma>1$. Assume further that the following two technical conditions hold: (i) there exists a function $K_q:\mathbb{R}^d\to\mathbb{R}$ and $\epsilon_q>0$ such that

$$\int_{\mathbb{R}^d} K_q(z)dz < \infty \quad \text{and} \quad |q^*(z)-q^*(\tilde z)| \le |z-\tilde z| K_q(z) \quad \forall z,\tilde z\in\mathbb{R}^d \text{ with } |z-\tilde z|<\epsilon_q; \tag{31}$$

(ii) there exist constants $\beta\in(0,1)$, $c_\beta>0$ and $\epsilon_\pi>0$ such that

$$|\nabla\pi(\tilde x)| < c_\beta \pi(x)^\beta \quad \forall x,\tilde x\in\mathbb{R}^d \text{ with } |x-\tilde x|<\epsilon_\pi. \tag{32}$$

Remark 4.1. Assumption (31) is a version of a local Lipschitz condition and holds for many proposals q^* used in practice, e.g. Gaussian proposals. Assumption (32) and condition (5) hold for the target densities π proportional to $e^{-p(x)}$, where p is a polynomial of degree k with leading order terms satisfying $p_k(x)\to\infty$ as $|x|\to\infty$ (see Example 2.2 for the precise definition of p_k).

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

Proposition 4.2. *Let P in MH(q,π) be the transition kernel of a random walk Metropolis chain described above, i.e. $q(x,y)=q^*(y-x)$, q^* is even and satisfies* (31)*, π satisfies* (5) *and* (32) *and A1-A3 hold. Fix $\gamma\in(0,\beta/2)$ and let $(\mathbb{X}_n)_{n\in\mathbb{N}}$ be an exhaustive sequence of allotments with respect to V_γ (cf. Definition 2.2 and paragraph above). Then the V_γ-radius in* (2) *equals 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* (3)*, takes the form*

$$\delta_{\gamma,n} = \max\left(\sup_{x\notin J_0^n}|x-a^n(x)|, \sup_{x\in\mathbb{R}^d}\left(\pi(x)/\pi(a^n(x))\right)^\gamma - 1\right), \tag{33}$$

where $a^n(\cdot)$ is defined in (10)*. Let $F\in L^\infty_{V_\gamma}$ be continuously differentiable function satisfying the inequality $|\nabla F(\tilde x)| < c_F\pi^{2\gamma-1}(x)$ for all $x,\tilde x\in\mathbb{R}^d$ with $|x-\tilde x|<\epsilon_F$ (for some constants $c_F,\epsilon_F>0$). Let σ_n^2 be the asymptotic variance in the CLT$(\Phi,\ F+P\tilde F_n-\tilde F_n)$, where $\tilde F_n$ is constructed by the Algorithm with input P, F and $\mathbb{X}_n$. Then there exists a constant $C_\gamma>0$ such that*

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

Remark 4.2. (i) Any polynomial F, and in fact any function whose gradient grows no faster than a polynomial, satisfies assumptions of Proposition 4.2 regardless of the chosen $\gamma\in(0,\beta/2)$. Such functions for example include the mean and the variance of any coordinate.

(ii) A natural question that arises in this context is the following: is it possible to take the limit as $\gamma \to 0$ in Proposition 4.2? Put differently, does the inequality $\sigma_n^2 \le C_0 \max(\delta_{\gamma,n}, \pi(J_0^n))$ hold for all $n \in \mathbb{N}$, some positive constant C_0 and a class of force functions (e.g. polynomials)? We conjecture that the answer is negative but were unable to find such an example.

4.1. **Proofs.** For any two sequences of real numbers $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$, we say that $(a_n)_{n\in\mathbb{N}}$ is of *lesser order* than $(b_n)_{n\in\mathbb{N}}$ if there exists a constant $C_0 > 0$ such that $a_n \le C_0 b_n$ holds for all $n \in \mathbb{N}$. We first establish Theorem 4.1 and then apply it to prove Proposition 4.2.

Proof of Theorem 4.1. The asymptotic variance appearing in CLT(P, G) can be expressed in terms of a solution to PE(P, G) (see [MT09, Theorem 17.4.4]). Apply this to Poisson's equation PE(P, $F + P\tilde{F}_n - \tilde{F}_n$) and its solution $\tilde{F}_n - \hat{F}$ to obtain $\sigma_n^2 = \pi\big((\tilde{F}_n - \hat{F})^2 - (P(\tilde{F}_n - \hat{F}))^2\big)$. Recall the definition of Δ_n from (17) and bound σ_n^2 as follows (constants c_n are those from Proposition 3.6):

$$\begin{aligned}\sigma_n^2 &= \pi\big((\tilde{F}_n + c_n - \hat{F})^2 - \big(P(\tilde{F}_n + c_n - \hat{F})\big)^2\big)\\ &= \pi\Big(\big((\tilde{F}_n + c_n - \hat{F}) - P(\tilde{F}_n + c_n - \hat{F})\big)\big((\tilde{F}_n + c_n - \hat{F}) + P(\tilde{F}_n + c_n - \hat{F})\big)\Big)\\ &= \pi\Big(V\big((\tilde{F}_n - \hat{F}) - P(\tilde{F}_n - \hat{F})\big)\frac{(\tilde{F}_n + c_n - \hat{F}) + P(\tilde{F}_n + c_n - \hat{F})}{V}\Big)\\ &\le \pi(V|\Delta_n|)(\xi + \|\hat{F}\|_V)(1 + \lambda_V + \kappa_V).\end{aligned}$$

The equalities hold, since neither σ_n^2 nor Δ_n change, if we perturb $\tilde{F}_n$ by a constant. The inequality is a consequence of Proposition 3.6 and A1.

Thus, the sequence $(\sigma_n^2)_{n\in\mathbb{N}}$ is of lesser order than $\big(\pi(V|\Delta_n|)\big)_{n\in\mathbb{N}}$. Express $\pi(V|\Delta_n|)$ as the sum $\pi(V|\Delta_n|1_{J_0^n}) + \pi(V|\Delta_n|1_{\mathbb{R}^d\setminus J_0^n})$. Since $\Delta_n \in \mathrm{L}_V^\infty$, sequence $\big(\pi(V|\Delta_n|1_{J_0^n})\big)_{n\in\mathbb{N}}$ is clearly of lesser order than $\big(\pi(V^2 1_{J_0^n})\big)_{n\in\mathbb{N}}$.

Now consider the other term, $\pi(V|\Delta_n|1_{\mathbb{R}^d\setminus J_0^n})$. By Lemma 3.8 it is bounded by the sum of the following three terms

$$(34)\quad T_1(n) := \int_{\mathbb{R}^d\setminus J_0^n} V(x)\left|\left(P\tilde{F}_n - \tilde{F}_n\right)(x) - \left(P\tilde{F}_n - \tilde{F}_n\right)(a^n(x))\right|\pi(x)dx,$$
$$T_2(n) := \int_{\mathbb{R}^d\setminus J_0^n} V(x)|F(x) - F(a^n(x)|\pi(x)dx \quad \text{and} \quad T_3(n) := |\pi_n(f_n) - \pi(F)|\pi(V).$$

Assumption (30) implies, that the sequence of second terms $(T_2(n))_{n\in\mathbb{N}}$ in (34) is of the order less than $(\delta_n)_{n\in\mathbb{N}}$. Using the form of kernel P =MH(q,π) and the fact, that $\tilde{F}_n(x) = \tilde{F}_n(a^n(x))$ ($\tilde{F}_n$ is piecewise constant), the first term can be transformed into:

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

This, in turn, can be bounded by a constant multiplier of

$$\int_{\mathbb{R}^d\setminus J_0^n}\int_{\mathbb{R}^d}\big(V(x)^2 + V(x)V(y)\big)Z_n(x,y)\pi(x)dydx$$

using $\tilde{F}_n \in \mathrm{L}_V^\infty$, definition of Z_n and triangle inequality. Hence, the sequence of first terms $(T_1(n))_{n\in\mathbb{N}}$ in (34) is also of lesser order than $(\delta_n)_{n\in\mathbb{N}}$, by assumption (29).

The sequence $(T_3(n))_{n\in\mathbb{N}}$ in (34) is obviously of the same order as $\big(|\pi_n(f_n) - \pi(F)|\big)_{n\in\mathbb{N}}$. But $|\pi_n(f_n) - \pi(F)|$ can be bounded above by $|(\pi_n - \pi_n^*)(f_n)| + |\pi_n^*(f_n - h_n)|$ (to recall definitions see (19) and (20)). It is straightforward to argue $\pi_n^*(f_n - h_n) = \int_{\mathbb{R}^d}(F(x) - F(a^n(x))\pi(x)dx$. Use this together with triangle inequality and bounds $V \geq 1$, $F \leq \|F\|_V V$ and (12) to conclude:

$$|\pi_n^*(f_n - h_n)| \leq \int_{\mathbb{R}^d} V(x)|F(x) - F(a^n(x))|\pi(x)dx$$
$$\leq ||F||_V(2 + \sup_{n\in\mathbb{N}} \delta_n)\pi(V^2 1_{J_0^n}) + \int_{\mathbb{R}^d\setminus J_0^n} V(x)|F(x) - F(a^n(x))|\pi(x)dx.$$

Hence, by (30), $\big(|\pi_n^*(f_n - h_n)|\big)_{n\in\mathbb{N}}$ is of lesser order than $\big(\max(\pi(V^2 1_{J_0^n}), \delta_n)\big)_{n\in\mathbb{N}}$.

Similarly, using inequalities (21) and (22) from Proposition 3.7, we can argue, that $|(\pi_n - \pi_n^*)(f_n)|$ is upper bounded by a constant multiplier (independent of $n \in \mathbb{N}$) of $\int_{\mathbb{R}^d} \big(\int_{\mathbb{R}^d} \big(V(y) + V(x)\big) Z_n(x, y)dy\big)\, \pi(x)dx$. Again we split integration with respect to x into parts integrating over J_0^n and $\mathbb{R}^d \setminus J_0^n$, and then use (29), A1 and (12) to conclude, that $\big(|(\pi_n - \pi_n^*)(f_n)|\big)_{n\in\mathbb{N}}$ is of lesser order than $\big(\max(\pi(V^2 1_{J_0^n}), \delta_n)\big)_{n\in\mathbb{N}}$.

Together this implies that the sequence of terms $(T_3(n))_{n\in\mathbb{N}}$ of (34) is of lesser order than $\big(\max(\pi(V^2 1_{J_0^n}), \delta_n)\big)_{n\in\mathbb{N}}$ as well, and the proof is finished. □

Proof of Proposition 4.2. Since P, F and $\mathbb{X}_n$ in Proposition 4.2 satisfy the assumptions of Theorem 2.4, we need only to establish that conditions (29) and (30) in Theorem 4.1 hold for $V = V_\gamma$ and $\delta_n = \delta_{\gamma,n}$. 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.

We will first establish (30) in this setting. By (33) we have $|x - a^n(x)| < \delta_{\gamma,n}$. Thus, by Lagrange's theorem and assumptions (32), we have, for all large enough n and some $\tilde{x}^n$ on a line between x and $a^n(x)$, the following:

$$\int_{\mathbb{R}^d\setminus J_0^n} V_\gamma(x)\frac{|F(x) - F(a^n(x))|}{\delta_{\gamma,n}}\pi(x)dx \leq \int_{\mathbb{R}^d\setminus J_0^n} V_\gamma(x)\frac{|F(x) - F(a^n(x))|}{|x - a^n(x)|}\pi(x)dx$$
$$= \int_{\mathbb{R}^d\setminus J_0^n} V_\gamma(x)|\nabla F(\tilde{x}^n)|\pi(x)dx \leq c_\gamma c_F \int_{\mathbb{R}^d} \pi^{-\gamma}(x)\pi^{2\gamma-1}(x)\pi(x)dx = c_\gamma c_F \int_{\mathbb{R}^d} \pi^\gamma(x)dx.$$

Target π decays supper-exponentially along any ray from the origin, hence so does π^γ. Thus, the integral $\int_{\mathbb{R}^d} \pi^\gamma(x)dx$ is finite (the same holds for any other positive exponent) and (30) follows.

Next, we show (29). As we are studying a symmetric random walk Metropolis, the acceptance probability equals $\alpha(x, y) = \min\Big(1, \frac{\pi(x)}{\pi(y)}\Big)$. Denote $\mathcal{A}_x := \{y \in \mathbb{R}^d;\ \pi(x) \leq \pi(y)\}$. Keep in mind, that if $y \in \mathcal{A}_x$, then $\alpha(x, y) = 1$ and $V_\gamma(x) \geq V_\gamma(y)$ and that this inequality is reversed otherwise.

For a set $\mathcal{B} \subseteq \mathbb{R}^d$ denote (recall $Z_n(x, y) = \big|\alpha(x, y)q^*(y - x) - \alpha(a^n(x), y)q^*(y - a^n(x))\big|$)

$$\mathcal{I}_n(\mathcal{B}) := \int_{\mathbb{R}^d\setminus J_0^n} \left(\int_{\mathcal{B}} \big(V_\gamma(x)^2 + V_\gamma(x)V_\gamma(y)\big)\frac{Z_n(x, y)}{\delta_{\gamma,n}}dy\right)\pi(x)dx.$$

Condition (29) will follow, if we manage to prove $\limsup_{n\to\infty} \mathcal{I}_n(\mathbb{R}^d)$ is finite. To do that, we split the integral with respect to y into four integrals, depending on whether or not $y \in \mathcal{A}_x$ and $y \in \mathcal{A}_{a^n(x)}$.

If $y \in \mathcal{B}_x^{n,1} := \mathcal{A}_x \cap \mathcal{A}_{a^n(x)}$, then we have

$$\frac{Z_n(x,y)}{\delta_{\gamma,n}} \le \frac{\left|q^*(y-x) - q^*(y-a^n(x))\right|}{|x-a^n(x)|} \le K_q(y-x) \tag{35}$$

for all large enough n, using (31) and $|x-a^n(x)| \le \delta_{\gamma,n}$. Because $V_\gamma(x) \ge V_\gamma(y)$ and $1-2\gamma > 0$, (31) and (35) imply the following:

$$\begin{aligned} \mathcal{I}_n(\mathcal{B}_x^{n,1}) &\le \textstyle\int_{\mathbb{R}^d \setminus J_0^n} 2V_\gamma(x)^2 \left(\int_{\mathcal{B}_x^{n,1}} K_q(y-x)dy \right) \pi(x)dx \\ &\le 2\pi(V_\gamma^2) \textstyle\int_{\mathbb{R}^d} K_q(z)dz \le 2 \int_{\mathbb{R}^d} \pi^{1-2\gamma}(x)dx \int_{\mathbb{R}^d} K_q(z)dz < \infty. \end{aligned} \tag{36}$$

If $y \in \mathcal{B}_x^{n,2} := (\mathbb{R}^d \setminus \mathcal{A}_x) \bigcap (\mathbb{R}^d \setminus \mathcal{A}_{a^n(x)})$, then bound as follows:

$$\begin{aligned} Z_n(x,y) &\le q^*(y-x)\pi(y)\left|\frac{1}{\pi(x)} - \frac{1}{\pi(a^n(x))}\right| + \frac{\pi(y)}{\pi(a^n(x))}\left|q^*(y-a^n(x)) - q^*(y-x)\right| \\ &\le q^*(y-x)\frac{\pi(y)}{\pi(a^n(x))}\frac{|\pi(a^n(x)) - \pi(x)|}{\pi(x)} + \frac{\pi(y)}{\pi(a^n(x))}\left|q^*(y-a^n(x)) - q^*(y-x)\right|. \end{aligned} \tag{37}$$

By (32) and Lagrange's theorem we have

$$\frac{|\pi(a^n(x)) - \pi(x)|}{\delta_{\gamma,n}} \le \frac{|\pi(a^n(x)) - \pi(x)|}{|x-a^n(x)|} \le |\nabla\pi(\tilde{x}^n)| \le c_\beta \pi^\beta(x)$$

for all sufficiently large n. Putting together the right inequality in (35), bound (37) and the above yields

$$\frac{Z_n(x,y)}{\delta_{\gamma,n}} \le \frac{\pi(y)}{\pi(a^n(x))}\left(c_\beta q^*(y-x)\pi^{\beta-1}(x) + K_q(y-x)dy\right). \tag{38}$$

By definition of $\delta_{\gamma,n}$ (see (33)), there exists $c_\pi > 0$, such that $\sup_{n\in\mathbb{N}} \sup_{x\in\mathbb{R}^d} \frac{\pi(x)}{\pi(a^n(x))} < c_\pi$. We can, for instance, take $c_\pi = (1 + \sup_{n\in\mathbb{N}} \delta_{\gamma,n})^{1/\gamma}$. We argue:

$$\begin{aligned} \mathcal{I}_n(\mathcal{B}_x^{n,2}) &\le \int_{\mathbb{R}^d\setminus J_0^n}\int_{\mathcal{B}_x^{n,2}} 2V_\gamma(y)^2 \frac{\pi(y)}{\pi(a^n(x))}\left(c_\beta q^*(y-x)\pi^{\beta-1}(x) + K_q(y-x)dy\right)\pi(x)dx \\ &\le \int_{\mathcal{B}_x^{n,2}} 2c_\pi V_\gamma(y)^2\left(c_\beta \pi(y)^{\beta-1}\int_{\mathbb{R}^d} q^*(y-x)dx + \int_{\mathbb{R}^d} K_q(y-x)dx\right)\pi(y)dy \\ &\le 2c_\pi c_\beta c_\gamma^2 \int_{\mathbb{R}^d} \pi^{\beta-2\gamma}(y)dy + 2c_\pi c_\gamma^2 \int_{\mathbb{R}^d}\pi^{1-2\gamma}(y)dy\int_{\mathbb{R}^d} K_q(z)dz < \infty. \end{aligned} \tag{39}$$

The first inequality holds by (38) and since $V_\gamma(y) \ge V_\gamma(x)$ for $y \in \mathcal{B}_x^{n,2}$. For the second we have used Fubini's theorem, bound $\frac{\pi(x)}{\pi(a^n(x))} < c_\pi$ and the fact, that $\pi(y)^{\beta-1} \ge \pi(x)^{\beta-1}$ (due to $y \in \mathcal{B}_x^{n,2}$ and $\beta < 1$). For the final one we have merely increased the integration domain and taken into the account, that q^* is a density, that $\beta > 2\gamma$ and the definition of V_γ.

Denote $\mathcal{B}_x^{n,3} := \mathcal{A}_x \bigcap (\mathbb{R}^d \setminus \mathcal{A}_{a^n(x)})$ and $\mathcal{B}_x^{n,4} := \mathcal{A}_{a^n(x)} \bigcap (\mathbb{R}^d \setminus \mathcal{A}_x)$. In a similar way as in (36) and (39), we can also find finite upper bounds on $\mathcal{I}_n(\mathcal{B}_x^{n,3})$ and $\mathcal{I}_n(\mathcal{B}_x^{n,4})$. Since $\mathbb{R}^d = \mathcal{B}_x^{n,1} \cup \mathcal{B}_x^{n,2} \cup \mathcal{B}_x^{n,3} \cup \mathcal{B}_x^{n,4}$, this implies $\limsup_{n\to\infty} \mathcal{I}_n(\mathbb{R}^d) < \infty$ and (29) follows.

□

5. Applications of the Algorithm

In this section we discuss the implementation of the Algorithm and describe numerical examples. Section 5.1 gives a simple IID Monte Carlo procedure to construct the matrix A defined in steps (I) and (II) of the Algorithm. We stress here that this procedure depends neither on the force function $F : \mathbb{R}^d \to \mathbb{R}$ nor the simulated path $(\Phi_i)_{i=1,\dots,k}$ of the Metropolis-Hastings chain Φ. It depends only on the characteristics of the underlying Markov process Φ. In particular, the same matrix A can be used over a family of functions F and any set of simulated paths of Φ.

In Section 5.2 we present numerical results for a variety of Metropolis-Hastings chains and force functions. We study numerically both geometrically ergodic (Example 5.2.1) and non-geometrically ergodic chains (Example 5.2.2), including an example where the CLT for ergodic averages is known not to hold (Example 5.2.3). Example 5.2.4 deals with the well known case of a slowly converging random walk Metropolis chain with a target distribution that has irregularly ("banana") shaped level contours. In these examples we use force functions that are not necessarily Lipschitz and may have super-linear growth and discontinuities.

In order to compare the variances of the ergodic averages $S_k(F)$ and $S_k(F+P\tilde{F}-\tilde{F})$ numerically, where $\tilde{F}$ and $P\tilde{F}$ are constructed in Section 5.1, in all the examples below we simulate 200 independent paths of the chain (started from stationarity) and report the quotient of estimated variances (see (43)). We find that the variance of $S_k(F+P\tilde{F}-\tilde{F})$ is between several hundred to several thousand times smaller than that of $S_k(F)$ (see Section 5.2 for precise figures), including Examples 5.2.2 and 5.2.3 where Theorem 2.4 is not applicable (A1 is violated).

5.1. **Implementation.** Given the Metropolis-Hastings kernel (MH(q, π)), define a partition $\{J_0,\dots,J_m\}$ of the state space such that the probability $\pi(J_0)$ is small, e.g. of order 10^{-6} or J_0 is a five standard deviations event under π. This is feasible in the low dimensional examples in this section. In practice J_0 could be chosen so that the simulated path is contained in the complement $\mathbb{R}^d \setminus J_0$. For example, for the specific chains in Sections 5.2.1–5.2.3 below, we set J_0 to be the complement of a large "box" and J_i, for $i = 1,\dots,m$, to be small boxes decomposing it. We pick $a_j \in J_j$, for $j > 0$, to be the centres of the boxes and choose a_0 to be close to the boundary of J_0.

Remark 5.1. The choice of "partition into boxes" in our examples below is made for illustrative purposes only as it is very simple to specify. But it is by no means optimal, particularly in dimensions greater than one, where it may lead to many states in the weak approximation being redundant as the Metropolis-Hastings chain spends no or very little time in most of the sets J_i, for $i = 1,\dots,m$. A systematic investigation of the efficient constructions of allotments for specific chains of interest in applications is left for future research. The choice of representatives $\{a_0,\dots,a_m\}$ is arbitrary and bears little influence on the variance reduction levels in Section 5.2.

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 A (steps (I)-(II) of the Algorithm). As the precise computation of its entries is not feasible in general, we construct an unbiased estimate $\hat{A}$ of A. With this in mind, let $i(x)$ be the unique index $i \in \{0, \ldots, m\}$, 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

$$\hat{P}(x, a_j) := \begin{cases} \frac{1}{n_1} \sum_{l=1}^{n_1} \mu^{\text{Leb}}(J_j) \alpha(x, Y_j^l) q(x, Y_j^l) & \text{if } j \notin \{0, i(x)\}, \\ \frac{1}{n_2} \sum_{l=1}^{n_2} 1_{J_0}(Z_{i,l}^x) \alpha(x, Z_{i,l}^x) & \text{if } j = 0 \neq i(x), \\ 1 - \sum_{k \in \{0,\ldots,m\} \setminus \{j\}} \hat{P}(x, a_k) & \text{if } i(x) = j, \end{cases} \tag{40}$$

where $n_1, n_2 \in \mathbb{N}$, random vectors Y_j^l, $l = 1, \ldots, n_1$, are IID uniform in the set J_j for any $j \in \{1, \ldots, m\}$ (recall that if $j \neq 0$, the set J_j is bounded) and $Z_{i,l}^x$, $l = 1, \ldots, n_2$, are IID random vectors distributed according to the proposal distribution $q(x, z)dz$ in (MH(q, π)). It is clear from this description that $\hat{P}(x, a_j)$ is an unbiased estimator of the transition probability $P(x, J_j)$. We define the estimator $\hat{A}$ for the matrix A in the Algorithm by the formula

$$\hat{A} := \hat{B} + (\mathbf{1} - \hat{B} e_0) e_0^\top, \quad \text{where} \quad \hat{B}_{ij} := \hat{P}(a_i, a_j) - \delta_{ij} \quad i, j \in \{0, \ldots, m\}, \tag{41}$$

δ_{ij} is the Kronecker delta, e_0 is the column vector in $\mathbb{R}^{1+m}$ with the first coordinate equal to one and the rest zero, $e_0^\top$ is its transpose and $\mathbf{1}$ is the column vector with all coordinates one.

Given a function $F : \mathbb{R}^d \to \mathbb{R}$, we can execute steps (III)-(V) in the Algorithm. Constructing the ergodic average estimator $S_k(F + P\tilde{F} - \tilde{F})$ requires the evaluation of the function $P\tilde{F}$ along the simulated path $(\Phi_i)_{i=1,\ldots,k}$ of the Metropolis-Hastings chain. We use the form of $\tilde{F}$ and the formula in (40) to find an unbiased estimate

$$\hat{P}\tilde{F}(x) := \sum_{j=0}^{m} \hat{f}_j \hat{P}(x, a_j) \tag{42}$$

for $P\tilde{F}(x)$ for any $x \in \mathbb{R}^d$, where $\hat{f}$ is the solution of the system in step (IV) of the Algorithm.

Remark 5.2. **(I)** There of course exist other unbiased estimators for $P(x, J_j)$, different from the one in (40). The choice made here works well with small random samples: in all the examples below we use $n_1 = 1$ and $n_2 = 10$. Note also that, in the construction of the ergodic average estimator $S_k(F + P\tilde{F} - \tilde{F})$, the function $\hat{P}\tilde{F}$ is used in the evaluation of $P\tilde{F}$ along the path of the chain. In this context it is important that the uniform vectors Y_j^l in the formula above do not depend on the value of the argument x and can be reused. In the case of random walk Metropolis, i.e. $q(x, y) = q^*(y - x)$, we have $Z_{i,l}^x = x + Z_{i,l}$ for any $x \in \mathbb{R}^d$ and $Z_{i,l}$ are also simulated only once. This is clearly more efficient than simulating IID random vectors from the distributions $P(x, dz)$ in (MH(q, π)), which would also lead to an unbiased estimate of $P\tilde{F}(x)$, as the random variates cannot be reused at distinct values x taken by the chain.
(II) Neither Theorem 2.4 nor the implementation of the Algorithm depend on the simulated path $(\Phi_i)_{i=1,\ldots,k}$. This should be contrasted with the 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. [DK12] for this approach applied to random scan Gibbs samplers). This produces a consistent but biased estimator $S_k(F + P\tilde{F} - \tilde{F})$, even if the chain is started in stationarity, as the control variate is a non-linear function of the estimate $S_k(F)$ of $\pi(F)$. The bias can be avoided by splitting

the path $(\Phi_i)_{i=1,\dots,k}$ into two parts, using the first part to construct the control variate and the second for the estimation. But this approach requires additional simulation and was not used in [DK12] (the level of variance reduction in the examples of [DK12] increases with k, likely due to the reduction in the bias). Our implementation of the Algorithm does not depend on the simulated trajectory. Hence using the entire path yields an unbiased estimator.

5.2. **Examples.** In order to analyse numerically the level of variance reduction produced by the implementation of the Algorithm in Section 5.1, let

$$r_{k,n} := \frac{\sum_{i=1}^{n}(S_k^i(F)-\pi(F))^2/n}{\sum_{i=1}^{n}(S_k^i(F+U)-\pi(F))^2/n}, \tag{43}$$

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 $(S_k^i(F), S_k^i(F+U))$, for $i=1,\dots,n$, are IID samples of the pair of ergodic average estimators $(S_k(F), S_k(F+U))$ evaluated on the simulated paths, where $U := \hat{P}\tilde{F} - \tilde{F}$ with $\hat{P}\tilde{F}$ and $\tilde{F}$ computed as in Section 5.1. Put differently, since the numerator (resp. denominator) of $r_{s,n}$ is an unbiased IID estimator of the variance of $S_k(F)$ (resp. $S_k(F+U)$), the quotient $r_{s,n}$ specifies the factor of the variance reduction achieved by the Algorithm. In the examples below we start the chain in stationarity, thus eliminating the bias of $(S_k(F), S_k(F+U))$ and allowing us to focus on the variance.

5.2.1. *Bimodal normal distribution.* Let the target law be $\pi := \rho N(\mu_1, \sigma_1^2) + (1-\rho)N(\mu_2, \sigma_2^2)$, where the parameters take the values $\mu_1 = -3$, $\sigma_1 = 1$, $\mu_2 = 4$, $\sigma_2 = 1/2$, $\rho = 2/5$. In this example the target density $\pi(\cdot)$ is a mixture of two normal densities with the modes at -3 and 4. Moreover, $\pi(\cdot)$ takes values close to zero in the neighbourhood of the origin. Let $F(x) := x^3$, $x \in \mathbb{R}$, be the force function and let the proposal density $q(x,\cdot)$ be $N(x,1)$. Construct $\hat{A}$, $\tilde{F}$ and $\hat{P}\tilde{F}$ by the formulae in (40)–(42) as in the previous example (decompose $\mathbb{R} \setminus J_0 := [-8, 7)$ into 700 subintervals of equal lengths and use $n_1 = 1$, $n_2 = 10$).

The assumptions of Theorem 2.4 are satisfied in this example and the chain is geometrically ergodic. 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 $\pi(F)$. The variance of the estimator $S_k(F+U)$ is thousands of times smaller than that of $S_k(F)$ as the function U takes into account the existence of both modes (see Figure 1 for the evolution of the estimators):

path length k of the chain ($n = 200$ stationary paths)	10^3	$5 \cdot 10^3$	$5 \cdot 10^4$
factor of variance reduction $r_{n,k}$ in Eq. (43)	1170.4	3281.7	5735.9

5.2.2. *Heavy tailed distribution.* Let $\pi(x) := (3/2\pi)(1+x^6)^{-1}$ and $F(x) = 1_{[0,1]}(x)$ $(x \in \mathbb{R})$. Since the target distribution π is heavy tailed, we take the proposal density $q(x,\cdot)$ to be the density of $N(x, 100)$. As in the previous two examples, we construct $\hat{A}$, $\tilde{F}$ and $\hat{P}\tilde{F}$ by the formulae in (40)–(42) using the decomposition of $\mathbb{R} \setminus J_0 := [-15, 15)$ into 1500 subintervals of equal lengths and $n_1 = 1$, $n_2 = 10$.

The random walk Metropolis chain in the present example is known not to be geometrically ergodic [MT96, Theorem 3.3] and hence does not satisfy the main assumption A1 of Theorem 2.4.

Furthermore, the force function is not continuous, potentially leading to an increase in variance. However, the variance reduction achieved by the Algorithm is significant:

path length k of the chain ($n = 200$ stationary paths)	10^3	$5 \cdot 10^3$	$5 \cdot 10^4$
factor of variance reduction $r_{n,k}$ in Eq. (43)	52437	15662	1427

The right plot in Figure 1 shows typical paths of the estimators.

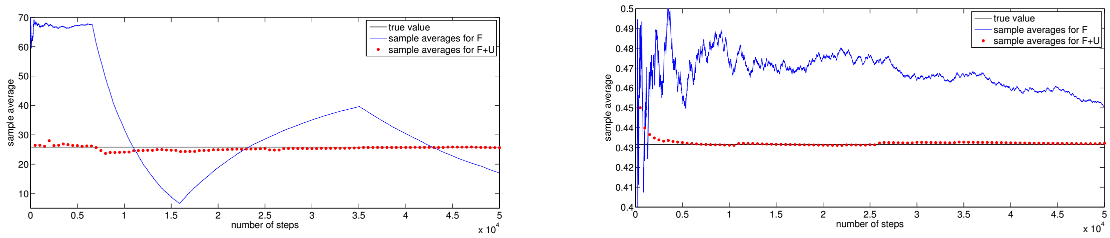

FIGURE 1. Evolution of the path averages $S_i(F)$ and $S_i(F+U)$, $i = 1, \ldots, k$, over $k = 10^5$ time steps in Examples 5.2.1 (left graph) and 5.2.2 (right graph).

5.2.3. *A non-geometrically ergodic chain without a CLT.* Consider the exponential target density $\pi(x) := e^{-x}$ on the positive reals and a proposal density $q(x,y) := 3e^{-3y}$, $x, y \in (0, \infty)$. The chain Φ, generated by the transition kernel (MH(q, π)), is the so-called independence sampler (the proposed value is independent of the current state). This chain is well known not to be geometrically ergodic and the CLT(Φ,F) fails for the force function $F(x) := x$ (see [Rob99, Sec. 4]). Furthermore, the slow convergence properties of the ergodic average $S_k(F)$ is well documented in the literature (e.g. the simulations in [RR98] indicate that the average of the path of such a chain over a million iterations returns a value of 0.8 (instead of $\pi(F) = 1$) and occasionally returns a very large value). In this example the chain tends to either spend a lot of time jumping around the level 1/3 or jump to a value much higher than 1 and stay there. This leads to a very unstable behaviour of the ergodic average $S_k(F)$.

In order to investigate numerically the level of improvement achieved by the Algorithm, let $(0, \infty) \setminus J_0 := (0, 13)$ and decompose it into 200 intervals of equal length. Using $n_1 = 1$, $n_2 = 10$ and the formulae in (40)–(42), compute $\hat{A}$, $\tilde{F}$ and $\hat{P}\tilde{F}$. Although this example is clearly outside the scope of Theorem 2.4, the variance of the estimator $S_k(F + U)$ is significantly reduced compared to that of $S_k(F)$:

path length k of the chain ($n = 200$ stationary paths)	10^3	$5 \cdot 10^3$	$5 \cdot 10^4$
factor of variance reduction $r_{n,k}$ in Eq. (43)	985.32	1063.8	2038.7

We obtain such a vast reduction in variance with longer paths, as the estimator $S_k(F)$ consistently and significantly underestimates the mean. See Figure 2 for a path of the estimators. Furthermore, note that [Rob99, Thm 3] implies that the CLT also fails for $S_k(F + U)$.

5.2.4. *A target density with an unbounded curvature of level contours.* Let $\phi_B : \mathbb{R}^2 \to \mathbb{R}^2$, $\phi(x,y) := (x, y + Bx^2 - 100B)$, be a diffeomorphism of $\mathbb{R}^2$ with a fixed "bananicity" constant $B > 0$. It is well known [HST01] that a random walk Metropolis chain with a normal proposal and the target distribution $\pi := f \circ \phi_B$, where f is the density of a bivariate normal distribution with independent components and zero mean $N(\mathbf{0}, \mathrm{diag}(100, 1))$, has very poor convergence properties. In fact, even if the adaptive random walk Metropolis algorithm is used (for $B = 0.1$) the mixing of the first component of the chain is still very slow after $5 \cdot 10^6$ iterations [RR09, Sec.2.1]. Since the part of $\mathbb{R}^2$, where most of the mass of π lies, is heavily curved (i.e. banana shaped), the chain struggles to traverse this set leading to poor mixing and slow convergence.

As in [RR09, Sec.2.1], we fix $B = 0.1$ and define $F(x, y) := x$. Hence it holds that

$$\pi(x, y) \propto \exp(-(x^2/100 + (y + 0.1x^2 - 10)^2)/2)$$

and $\pi(F) = 0$. This examples is based on a random walk Metropolis chain with the density of the proposed increments given by $N(\mathbf{0}, \mathrm{diag}(100, 100))$ and the target π. Let $\mathbb{R}^2 \setminus J_0$ be the image under ϕ_B of the rectangle $[-50, 50] \times [-5, 5]$. Subdivide the longer (resp. shorter) of the two sides of this rectangle into 300 (resp. 30) intervals of equal length, thus obtaining a partition into 9000 rectangles of the box $[-50, 50] \times [-5, 5]$. Define the partition sets J_j and the representatives a_j, $j = 1, \ldots, 9000$, to be the images of these rectangles and their centres under the diffeomorphism ϕ_B. Figure 2 shows typical paths of the two estimators. The achieved factor of variance reduction is several hundred:

path length k of the chain ($n = 200$ stationary paths)	10^3	$5 \cdot 10^3$	$5 \cdot 10^4$
factor of variance reduction $r_{n,k}$ in Eq. (43)	247.4	238.2	255.2

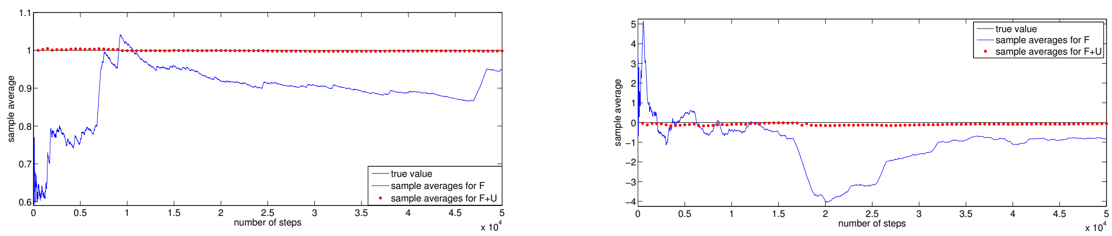

FIGURE 2. Evolution of the path averages $S_i(F)$ and $S_i(F + U)$, $i = 1, \ldots, k$, over $k = 10^5$ time steps in Examples 5.2.3 (left graph) and 5.2.4 (right graph).

6. Conclusion

In this paper we apply the idea of weak approximation of Markov processes to construct approximate solutions of Poisson's equation for discrete-time Markov chains. We show that, under general conditions, these approximations in the case of Metropolis-Hastings chains lead to ergodic averages with arbitrarily small asymptotic variance.

A number of questions of interest remain open. On the theoretical side, the key step in the proof presented here consists of establishing the uniform convergence to stationarity of a sequence of approximating chains (Section 3.3). Under suitable assumptions this fact is sufficient for the convergence of the Algorithm to the solution of Poisson's equation (measured by the size of the corresponding asymptotic variances, see Section 1.2 and Definition 2.1). It is feasible that the principle of uniform convergence to stationarity could be established in other contexts (e.g. queueing models and stochastic networks [Mey08]), both in discrete and continuous time, with the approximating Markov processes not necessarily having a finite state spaces. The main requirement for the approximating processes is that they should be sufficiently simple that their Poisson equations can be solved numerically. The key advantage of this approach is that the control variates do not require prior estimates of $\pi(F)$ or any other functional of the law π. They only depend on the characteristics of the underlying process (i.e. a transition kernel (resp. generator) in discrete (resp. continuous) time) converging to π.

A very simple application of the Algorithm, described in Section 5, shows numerically that the variance of ergodic estimators for the well-known slowly converging low-dimensional examples of the Metropolis-Hastings chains can be reduced arbitrarily. Developing the idea of weak approximation for the Poisson equation in the context of improving convergence of the estimators in Bayesian hierarchical models (see e.g. [Ros95] and [RR04, Sec. 2.4]) is a natural next step. For example, the chains based on the Metropolis-within-Gibbs [RR04] and the delayed Metropolis [CF05] samplers appear to lend themselves well to weak approximations using simpler Markov chains. These questions are left as a topic for future research.

Appendix A. Proof of Proposition 2.3

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 $L_n := W^{-1}\big((-\infty, r_n)\big)$,

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

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 $(\epsilon_n)_{n\in\mathbb{N}}$ (satisfying $\lim_{n\to\infty}\epsilon_n = 0$ and $\sup_{n\in\mathbb{N}}\epsilon_n < 1$) such that $|x-y| < \epsilon_n\sqrt{d}$ 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 $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ denote $K_x^n := [x_1, x_1 + \epsilon_n) \times \cdots \times [x_d, x_d + \epsilon_n)$. Clearly, it is possible to pick $x^1, x^2, \ldots x^{m_n} \in \mathbb{R}^d$ 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 $|x-y| < \epsilon_n\sqrt{d}$, for x, y from the same $\in J_j^n$. Since $\epsilon_n < 1$ 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,

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

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 (3)). By (2) and definition of L_n, $\mathrm{rad}(\mathbb{X}_n, W) \ge r_n$ for every $n \in \mathbb{N}$. So, $\lim_{n\to\infty}\mathrm{rad}(\mathbb{X}_n, W) = \infty$.

References

[AHO93] Sigrún Andradóttir, Daniel P Heyman, and Teunis J Ott. Variance reduction through smoothing and control variates for markov chain simulations. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 3(3):167–189, 1993.

[Bax05] Peter H. Baxendale. Renewal theory and computable convergence rates for geometrically ergodic Markov chains. *Ann. Appl. Probab.*, 15(1B):700–738, 2005.

[CF05] J. Andrés Christen and Colin Fox. Markov chain Monte Carlo using an approximation. *J. Comput. Graph. Statist.*, 14(4):795–810, 2005.

[DK12] Petros Dellaportas and Ioannis Kontoyiannis. Control variates for estimation based on reversible markov chain monte carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(1):133–161, 2012.

[GM96] Peter W. Glynn and Sean P. Meyn. A Liapounov bound for solutions of the Poisson equation. *Ann. Probab.*, 24(2):916–931, 1996.

[Has70] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[Hen97] Shane G Henderson. *Variance Reduction Via an Approximating Markov Process.* PhD thesis, Department of Operations Research, Stanford University, 1997. Available at `http://people.orie.cornell.edu/shane/pubs/thesis.pdf`.

[HG02] Shane G. Henderson and Peter W. Glynn. Approximating martingales for variance reduction in Markov process simulation. *Math. Oper. Res.*, 27(2):253–271, 2002.

[HLL99] Onésimo Hernández-Lerma and Jean Bernard Lasserre. *Further topics on discrete-time Markov control processes*, volume 42 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 1999.

[HMT03] Shane G. Henderson, Sean P. Meyn, and Vladislav B. Tadić. Performance evaluation and policy selection in multiclass networks. *Discrete Event Dyn. Syst.*, 13(1-2):149–189, 2003. Special issue on learning, optimization and decision making.

[HST01] Heikki Haario, Eero Saksman, and Johanna Tamminen. An adaptive metropolis algorithm. *Bernoulli*, 7(2):223–242, 2001.

[JH00] Søren Fiig Jarner and Ernst Hansen. Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.*, 85(2):341–361, 2000.

[Mey08] Sean Meyn. *Control techniques for complex networks.* Cambridge University Press, Cambridge, 2008.

[Mij07] Aleksandar Mijatović. Spectral properties of trinomial trees. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 463(2083):1681–1696, 2007.

[MP13] Aleksandar Mijatović and Martijn Pistorius. Continuously monitored barrier options under Markov processes. *Mathematical Finance*, 23(1):1–38, 2013.

[MRR$^+$53] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

[MS02] Armand M. Makowski and Adam Shwartz. The Poisson equation for countable Markov chains: probabilistic methods and interpretations. In *Handbook of Markov decision processes*, volume 40 of *Internat. Ser. Oper. Res. Management Sci.*, pages 269–303. Kluwer Acad. Publ., Boston, MA, 2002.

[MT94] Sean P. Meyn and R. L. Tweedie. Computable bounds for geometric convergence rates of Markov chains. *Ann. Appl. Probab.*, 4(4):981–1011, 1994.

[MT96] K. L. Mengersen and R. L. Tweedie. Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.*, 24(1):101–121, 1996.

[MT09] Sean Meyn and Richard L. Tweedie. *Markov chains and stochastic stability.* Cambridge University Press, Cambridge, second edition, 2009.

[MVJ14] Aleksandar Mijatović, Matija Vidmar, and Saul Jacka. Markov chain approximations for transition densities of Lévy processes. *Electron. J. Probab.*, 19:no. 7, 37, 2014.

[Rob99] G. O. Roberts. A note on acceptance rate criteria for CLTs for Metropolis-Hastings algorithms. *J. Appl. Probab.*, 36(4):1210–1217, 1999.

[Ros95] Jeffrey S. Rosenthal. Rates of convergence for Gibbs sampling for variance component models. *Ann. Statist.*, 23(3):740–761, 1995.

[RR98] Gareth O. Roberts and Jeffrey S. Rosenthal. Markov-chain Monte Carlo: some practical implications of theoretical results. *Canad. J. Statist.*, 26(1):5–31, 1998. With discussion by Hemant Ishwaran and Neal Madras and a rejoinder by the authors.

[RR04] Gareth O. Roberts and Jeffrey S. Rosenthal. General state space Markov chains and MCMC algorithms. *Probab. Surv.*, 1:20–71, 2004.

[RR09] Gareth O. Roberts and Jeffrey S. Rosenthal. Examples of adaptive MCMC. *J. Comput. Graph. Statist.*, 18(2):349–367, 2009.

[RT96a] G. O. Roberts and R. L. Tweedie. Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika*, 83(1):95–110, 1996.

[RT96b] Gareth O. Roberts and Richard L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.

[Tie94] Luke Tierney. Markov chains for exploring posterior distributions. *Ann. Statist.*, 22(4):1701–1762, 1994. With discussion and a rejoinder by the author.

Department of Mathematics, King's College London, UK
E-mail address: aleksandar.mijatovic@kcl.ac.uk

Department of Mathematics, Imperial College London, UK
E-mail address: j.vogrinc13@imperial.ac.uk