Citation for published version (APA):

Citing this paper
Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights
Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the Research Portal

Take down policy
If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
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 scheme is based on the idea of weak approximation and gives rise to a natural sequence of control variates for the ergodic average $S_k(F) = (1/k) \sum_{i=1}^{k} F(\phi_i)$, where $F$ is the force function in the Poisson equation. The main results show that the sequence of the asymptotic variances (in the CLTs for the control-variate estimators) converges to zero and give a rate of this convergence. Numerical examples in the case of a double-well potential are discussed.

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 $P$ on a general state space $\mathcal{X}$, is well-known to be intimately linked with the solution $\hat{F}$ of the Poisson equation $(\text{PE}(P, F))$ $\hat{F} - P\hat{F} = F - \pi(F)$ with a force function $F : \mathcal{X} \to \mathbb{R}$ (see [MT09, Sec.17.4]). Here $\pi$ is the invariant probability measure of $\psi$ on $\mathcal{X}$, $\pi(F) = \int_{\mathcal{X}} F(x)\pi(dx)$ and $PG(x) = \mathbb{E}_x[G(\psi_1)]$ for any $G : \mathcal{X} \to \mathbb{R}$. In fact, the Poisson equation in $(\text{PE}(P, F))$ is of fundamental importance in many areas of probability, statistics and engineering (see [MT09, Sec.17.7, p.459]). In this context one of the main motivations for constructing approximations to $\hat{F}$ is to reduce the asymptotic variance in $(\text{CLT}(\psi, F))$ for the Markov Chain Monte Carlo (MCMC) estimators, thus speeding up the MCMC algorithms.

Assume that the random sequence $(S_k(F))_{k \in \mathbb{N}}$ satisfies the strong law of large numbers (SLLN), $\lim_{k \to \infty} S_k(F) = \pi(F)$ a.s., and the CLT $(\text{CLT}(\psi, F))$ $\sqrt{k} (S_k(F) - \pi(F)) \xrightarrow{d} \sigma_F \cdot N(0,1)$ as $k \to \infty$,

where $N(0,1)$ is a standard normal distribution and the constant $\sigma^2_F$ is the asymptotic variance. Put differently, the variance of $S_k(F)$ is approximately equal to $\sigma^2_F/k$. It is hence intuitively clear that if $\sigma^2_F$ is large, which occurs in applications particularly when $F$ has super-linear growth (as $\sigma^2_F \propto \text{Var}_{\pi}(F)$, see e.g. [RR04, Sec.5] and the references therein), the variance of $S_k(F)$ will also be big, requiring a large number of steps $k$ for convergence. In contrast, imagine we knew the solution $\hat{F}$ of the Poisson equation $(\text{PE}(P, F))$ and could evaluate the function $P\hat{F} - \hat{F}$. Then the estimator given by the ergodic average $S_k(F + P\hat{F} - \hat{F})$ (for any $k \in \mathbb{N}$) would be equal to the constant $\pi(F)$
for any (not necessarily stationary) path of the chain \( \Psi \), i.e. its variance vanishes for \( \pi \)-a.e. starting point. However, solving Poisson’s equation for the chains arising in most applications, even for very simple functions \( F \), is for all practical purposes impossible (see e.g. relevant comments in [Hen97]). Nevertheless, this line of reasoning suggests the following heuristic:

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

This heuristic is well known and strongly substantiated with numerical evidence. As a method of variance reduction it has been developed in various Markovian settings [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, schemes for constructing \( \tilde{F} \) found in the literature (a) depend strongly on the structure of the underlying model and, to the best of our knowledge, (b) there are no theoretical results quantifying a priori the amount of reduction in the asymptotic variance of \( (\text{CLT}(\Psi, F + \mathcal{P}F - \tilde{F})) \). This paper aims to address both (a) and (b) by introducing a general \textbf{Scheme} (see below) for constructing an approximate solution \( \tilde{F} \) to \( (\text{PE}(\mathcal{P}, F)) \), applicable to any discrete time Markov chain, and analysing the corresponding asymptotic variance in the setting of Metropolis-Hastings chains.

\textbf{Scheme}

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

\begin{align*}
\text{(I) Define } & p_X \in \mathbb{R}^{(m+1) \times (m+1)} \text{ and } f_X : \{a_0, a_1, \ldots, a_m\} \rightarrow \mathbb{R} \text{ respectively by } \quad (p_X)_{ij} := \mathcal{P}(a_i, J_j) \quad \text{and} \quad f_X(a_j) := F(a_j), \quad \text{where } i, j \in \{0, 1, \ldots, m\}. \\
\text{(II) Find a solution } & \hat{f}_X \text{ of Poisson’s equation \( (\text{PE}(p_X, f_X)) \).} \\
\text{(III) Define } & \tilde{F}_X := \sum_{j=1}^{m} \hat{f}_X(a_j)1_{J_j}. 
\end{align*}

\textbf{Output}: Approximate solution \( \tilde{F}_X : X \rightarrow \mathbb{R} \) to Poisson’s equation in \( (\text{PE}(\mathcal{P}, F)) \).

Our main result (Theorem 2.6 below) states that, for an appropriately chosen allotment \( X \), the function \( \tilde{F}_X \) can theoretically achieve an arbitrary reduction of the asymptotic variance for a class of Metropolis-Hastings chains and force functions \( F \) that satisfy natural growth conditions. To the best of our knowledge, this is the first systematic approach capable of reducing the asymptotic variance arbitrarily for a general class of discrete-time Markov chains. The proof hinges on the uniform convergence to stationarity of a sequence of approximating Markov chains, which in turn crucially depends on the results in [MT94, Bax05] (see Section 3.1 below for details). Step (II) in the Scheme amounts to solving a linear system and can be carried out provided that the stochastic matrix \( p_X \) is irreducible. Moreover, Poisson’s equation \( (\text{PE}(p_X, f_X)) \) has a solution that is unique up to the addition of a constant function (see [MS02, Theorem 9.3]). Furthermore, the asymptotic variance in \( (\text{CLT}(\Psi, F + \mathcal{P}F - \tilde{F}_X)) \) does not depend on the choice of \( \hat{f}_X \) in step (II) of the Scheme.

The approximation Scheme exploits the stochastic evolution implicitly present in \( (\text{PE}(\mathcal{P}, F)) \). As in [HG02, HMT03, Mey08], we are using the true solution of the Poisson equation for a related...
Markov process to construct \( \tilde{F} \). In our context, the approximation of \( \hat{F} \) is based on the weak approximation of the chain \( \Psi \) by a sequence of “simpler” finite state Markov chains (converging in law to \( \Psi \)), such that the solutions of the Poisson equations for the approximating chains can be characterised algebraically. The approximating Markov chain underpinning the Scheme mimics the behaviour of \( \Psi \) as follows: its state space is a partition \( \{ J_0, J_1, \ldots, J_m \} \) of the state space \( X \) and its transition matrix consists of the probabilities of \( \Psi \) jumping from a chosen element in \( J_i \) into the set \( J_j \). Analogous weak approximation ideas have been applied in continuous time to Brownian motion [MJJ07], Lévy [MVJ14] and Feller [MP13] processes. A recent interesting approach for approximating the solution of Poisson’s equation in discrete time has been proposed in [DM16]. The idea is to solve the equation obtained by differentiating both sides of (PE(\( \mathcal{P}, F \)) in the state variable. This leads to a new approximation method for \( \hat{F} \) but appears to require smoothness properties of the transition kernel, not afforded by the class of Metropolis-Hastings chains.

The approximation of a given Markov chain with a finite-state chain given by the Scheme is akin to others previously mentioned in the literature that are also based on a partition or a covering of the state space, see for instance [RS62, HSS4, Ros92] and [MR02]. These papers relate the spectral gap and the speed of convergence to equilibrium of the initial and of the approximating Markov chains. They do not however address potential similarity of Poisson’s equations.

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

A natural question arising from Theorem 2.6 is about the rate of the decay of the sequence of asymptotic variances \( \sigma^2_n \to 0 \). Theorem 4.1 shows that the decay is governed by the greater of the two quantities: the mesh of the partition of the bounded set \( \mathbb{R}^d \setminus J_0 \) and the \( \pi \)-average of the squared drift function of the chain over \( J_0 \) (see Section 2 for definitions). Furthermore, for the chains studied in [RT96a, JH00], Theorem 4.1 implies a bound on the rate of decay in terms of the target density \( \pi \) alone (see Proposition 4.3 below). We hope this result is both of some practical value (cf. Section 5.2.1) and independent interest.

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

## 2. Assumptions and the main result

Let \( \pi \) be a density function of a probability measure on \( \mathbb{R}^d \) with respect to the Lebesgue measure \( \mu^{\text{Leb}} \) and let \( q: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} \) 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

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

where $\delta_x$ is Dirac’s measure centred at $x$, and the Markov chain $(\Phi_k)_{k \in \mathbb{N}}$ generated by $P$ is known as the Metropolis-Hastings chain (see [MRR+53, Has70]). In this context, $\pi$ is termed a target density and $q$ a proposal density. It is easy to see that the chain $\Phi$ 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 $\pi$. The measure $\pi(x)dx$ is also known as the invariant probability measure for the chain $\Phi$. Throughout the paper we assume that the kernel $P$ in $\text{MH}(q, \pi)$ satisfies the following assumptions:

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

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

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

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

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

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

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

Note that $L^\infty_V$ equipped with the norm $|| \cdot ||_V$ is a Banach space (see [HL99, Proposition 7.2.1]).

Remark 2.1. (i) Assumptions $A1A3$ are standard. Widely used classes of Random walk Metropolis chains (i.e. $q(x, y) = q^*(y - x)$) satisfying $A1A3$ are given in [MT96, RT96a, JH00]. See also [RT96b] for examples of Metropolis Adjusted Langevin chains satisfying $A1A3$.

(ii) For Metropolis kernel $P$ satisfying $A1A3$ and $F \in L^\infty_V$ there exists a solution $\hat{F}$ to $\text{PE}(P, F)$ that is an element of $L^\infty_V$. The solution $\hat{F}$ is unique up to the addition of a constant function (see [GM99, Prop 1.1 and Thm 2.3]).

(iii) Assumptions $A2$ and $A3$ imply that Metropolis-Hastings chain $\Phi$ driven by $P$ is $\pi$-irreducible (i.e $\mu^{\text{Leb}}$-irreducible), strongly aperiodic and positive Harris recurrent (see [MT96, Lem 1.1&1.2], [Tie94, Thm 1, Cor 2] and monograph [MT09] as a general reference). In particular, the SLLN [MT09, Thm 17.1.7] and the CLT [MT09, Thm 17.4.4] hold for $F \in L^\infty_V$.

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

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


Remark 2.1(v) motivates the following definition.

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

**Remark 2.3.**
(a) If \((G_n)_{n \in \mathbb{N}}\) asymptotically solves Poisson’s equation \( \text{PE}(P, F) \), so does \((G_n + c_n)_{n \in \mathbb{N}}\) for any sequence \((c_n)_{n \in \mathbb{N}}\) of real numbers.
(b) Definition 2.2 does not require the Metropolis-Hastings structure on \( \mathbb{R}^d \) and can be extended trivially to Markov chains on general state spaces satisfying an appropriate CLT.

We now define a sequence of functions that asymptotically solves Poisson’s equation \( \text{PE}(P, F) \).

**Definition 2.4.**
(a) Let \( \mathbb{J} \) be a partition of \( \mathbb{R}^d \) into measurable sets \( J_0, J_1, \ldots, J_m \), such that \( \bigcup_{j=1}^m J_j \) is bounded and \( \mu^{\text{Leb}}(J_j) > 0 \) holds for all \( 0 \leq j \leq m \). Let \( X = \{a_0, a_1, \ldots, a_m\} \) be a set of representatives: \( a_j \in J_j \) for all \( 0 \leq j \leq m \). The pair \( X := (\mathbb{J}, X) \) is called an allotment and \( m \) is the size of the allotment \( X \).
(b) Let \( W : \mathbb{R}^d \to [1, \infty) \) be a measurable function and \( X \) an allotment. \( W \)-radius and \( W \)-mesh of the allotment \( X \) are defined by

\[
(2) \quad \text{rad}(X, W) := \inf_{y \in J_0} W(y),
\]

\[
(3) \quad \delta(X, W) := \max \left( \max_{1 \leq j \leq m} \sup_{y \in J_j} |y - a_j|, \max_{0 \leq j \leq m} \sup_{y \in J_j} (W(a_j)/W(y) - 1) \right),
\]

respectively, where \( |x| \) denotes the Euclidean norm of any \( x \in \mathbb{R}^d \).
(c) A sequence of allotments \((X_n)_{n \in \mathbb{N}}\) is exhaustive with respect to the function \( W \) in (b) if the following holds: \( \lim_{n \to \infty} \text{rad}(X_n, W) = \infty \) and \( \lim_{n \to \infty} \delta(X_n, W) = 0 \).

**Remark 2.5.**
(i) For any continuous function \( W : \mathbb{R}^d \to [1, \infty) \) with bounded sublevel sets, there exists an exhaustive sequence of allotments (see Appendix A below).
(ii) Note that \( J_0 \) is the only unbounded set in the partition of an allotment \( X \). For the \( W \)-radius of \( X \) to be large, the union \( \bigcup_{j=1}^m J_j \) of all the bounded sets in the partition has to cover the part of \( \mathbb{R}^d \) where \( W \) is small.
(iii) The \( W \)-mesh is a maximum of two quantities: the first is a standard mesh of the partition \( \{J_1, \ldots, J_m\} \) of the bounded set \( \mathbb{R} \setminus J_0 = \bigcup_{j=1}^m J_j \). 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 approximately equal to

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

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

We can now state our main result.
Theorem 2.6. Let the transition kernel $P$ in $[\text{MH}(q,\pi)]$ of a Metropolis-Hastings chain $\Phi$ satisfy $\text{A1-A3}$ for a drift function $V$. Let $F \in L_{\pi}^\infty$ be continuous $\pi$-a.e. and let $(\mathcal{X}_n = (J_n, X_n))_{n \in \mathbb{N}}$ be an exhaustive sequence of allotments with respect to $V$, where $J_n = \{J^n_0, \ldots, J^n_{m_n}\}$ and $X_n = \{a^n_i \in J^n_j : j = 0, 1, \ldots, m_n\}$. For each $n \in \mathbb{N}$, let $\tilde{F}_n$ be the output of the Scheme with input $P$, $F$ and $\mathcal{X}_n$. Then the sequence $(\tilde{F}_n)_{n \in \mathbb{N}}$ asymptotically solves Poisson’s equation $\text{PE}(P,F)$, i.e. the asymptotic variance $\sigma^2_n$ in $\text{CLT}(\Phi, F + PF_n - F_n)$ converges to zero as $n \rightarrow \infty$.

Remark 2.7. Functions $\tilde{F}_n$ from Theorem 2.6 are well defined. This is because all the entries
\begin{equation}
(p_n)_{ij} := (p_{\mathcal{X}_n})_{ij} = P(a^n_i, J^n_j) = \begin{cases}
\int_{J^n_j} \alpha(a^n_i, y)q(a^n_i, y)dy & \text{if } i \neq j \\
1 - \int_{\mathbb{R} \setminus J^n_j} \alpha(a^n_i, y)q(a^n_i, y)dy & \text{if } i = j
\end{cases}
\end{equation}
of stochastic matrices $p_n$, constructed by the Scheme with input $P$, $F$ and $\mathcal{X}_n$, are strictly positive by assumptions $\text{A2-A3}$ and Definition 2.4(a) ($\mu^{\text{Leb}}(J^n_j) > 0$ for all $0 \leq j \leq m_n$, where $m_n$ is the size of allotment $\mathcal{X}_n$). Hence the chain on $\mathcal{X}_n$, driven by $p_n$, is irreducible, recurrent, aperiodic and admits a unique invariant probability measure $\pi_n$. Moreover, Poisson’s equation for $p_n$ and any force function on $\mathcal{X}_n$ has a solution, unique up to addition of a constant (see [MS02, Theorem 9.3]).

Remark 2.8. The proof of Theorem 2.6 does not rely heavily on the structure of Metropolis-Hastings kernels. Emulating the proof appears feasible at least for other specific $T$-chains (see MT09, Chapte 6] for the definition). More specifically, reversibility is needed in Proposition 3.2 but an analogous result can be obtained without it. In the proof of Proposition 3.3(b), we use the fact that the non-Dirac component $T(x, dy)$ of $P(x, dy)$ has positive and continuous density with respect to the Lebesgue measure. Finally, in proofs of Proposition 3.4 and Theorem 2.6 we require $T(x, dy)$ to exhibit the following form of continuity, $\lim_{n \rightarrow \infty} ||T(x, \cdot)-T(a^n(x), \cdot)||_V = 0$ for $\pi$-a.e. $x$ (here $|| \cdot ||_V$ is the $V$ total variation norm and $a^n(x) = \sum_{j=0}^{m_n} a^n_j 1_{J^n_j}(x)$).

3. Proof of Theorem 2.6

3.1. Overview of the proof. The central object in the proof of Theorem 2.6 is the function
\begin{equation}
\Delta(G) := PG - G + F - \pi(F),
\end{equation}
which measures the failure of a function $G$ to be a solution of the Poisson equation $\text{PE}(P,F)$. Intuitively, the closer $\Delta(G)$ is to zero the better.

The proof is in two parts. In the first part (Section 3.2 below) we show that a sequence of functions $(G_n)_{n \in \mathbb{N}}$ in $L_{\pi}^\infty$ asymptotically solves Poisson’s equation $\text{PE}(P,F)$ if $\lim_{n \rightarrow \infty} \pi(\Delta(G_n)^2) = 0$. This is a simple consequence of the representation of the asymptotic variance in terms of the spectral measure [KV86, Eq 1.1] and the existence of a spectral gap for geometrically ergodic Markov chains established [RR97, Prop 1.1].

The second part of the proof is more involved. It consists of verifying that functions $(\tilde{F}_n)_{n \in \mathbb{N}}$, defined in Theorem 2.6 indeed satisfy $\lim_{n \rightarrow \infty} \pi(\Delta(\tilde{F}_n)^2) = 0$. The key underlying fact needed for this purpose is that the family of the approximating finite state Markov chains driven by the
stochastic matrices \((p_n)_{n \in \mathbb{N}}\) converge to their respective stationary distributions \((\pi_n)_{n \in \mathbb{N}}\) uniformly in \(n \in \mathbb{N}\). This step is facilitated by the results in \cite{MT94} Thm 2.3 and \cite{Bax05} Thm 1.1, which show that the constants appearing in the geometric ergodicity estimate depend only and explicitly on the constants in the drift, minorisation and strong aperiodicity conditions for that chain. In Section 3.3 we show that these constants can be chosen independently of \(n \in \mathbb{N}\) (Proposition 3.3 below) and establish the uniform convergence to stationarity (Proposition 3.4 below).

In Section 3.4 we establish convergence in \(L^2(\pi)\) of the sequence \((\Delta(\tilde{F}_n))_{n \in \mathbb{N}}\). In addition to the uniform convergence to stationarity, the proof requires a further weak approximation by a family of finite state Markov chains with stationary distributions that are explicit in the target density \(\pi\) (see \(\text{\cite{KPR96} Thm 1.1}\) below). Note that the stationary laws \(\pi_n\) of the chains generated by the stochastic matrices \(p_n\), defined in \(\text{\cite{KPR96}}\), cannot be expressed explicitly in terms of \(\pi\).

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

- \(\tilde{F}\): solution of \(\text{PE}(P,F)\) in \(L^\infty_V\) (cf. Remark 2.1(ii)).
- \(f_n\) and \(v_n\): restrictions of \(F\) and \(V\) to the set \(X_n\), respectively.
- \(\hat{f}_n\): solution of \(\text{PE}(p_n,f_n)\) constructed within the Scheme (cf. Remark 2.7).
- \(\delta_n := \delta(\mathcal{X}_n,V)\): the \(V\)-mesh of the allotment \(X_n\) defined in \(\text{\cite{Gey92}}\).

### 3.2. Controlling the asymptotic variance

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

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

**Proof.** The kernel \(P\) is reversible and hence a bounded self-adjoint operator on the Hilbert space \(L^2(\pi)\). Furthermore, the Hilbert subspace \(\mathcal{H} := \{G \in L^2(\pi) : \pi(G) = 0\}\) is invariant for \(P\) (i.e. \(\pi(PG) = 0\) for any \(G \in \mathcal{H}\)). By \(\text{\cite{Gey92}}\) and Remark 2.1(v) it follows that \(\Delta(G_n) \in \mathcal{H}\) for all \(n \in \mathbb{N}\). The asymptotic variance \(\sigma^2_n\) in the \(\text{CLT}(\Phi, F + PG_n - G_n)\) can be represented in terms of a positive (spectral) measure \(E_{\Delta(G_n)}(d\lambda)\) on the spectrum \(\sigma(P|\mathcal{H}) \subset \mathbb{R}\) associated with the function \(\Delta(G_n)\), as follows (see \(\text{\cite{KPR96}}\) and \(\text{\cite{Gey92}}\) Thm 2.1 for details):

\[
\sigma^2_n = \int_{\sigma(P|\mathcal{H})} \frac{1 + \lambda}{1 - \lambda} E_{\Delta(G_n)}(d\lambda).
\]

Since the chain generated by \(P\) is geometrically ergodic by \(\text{\cite{A1} RR97}\) Prop 1] implies that the spectral radius \(\rho\) of \(P|\mathcal{H}\) satisfies \(\rho < 1\). Hence the inclusion \(\sigma(P|\mathcal{H}) \subseteq [-\rho, \rho]\), the equality \(E_{\Delta(G_n)}(\sigma(P|\mathcal{H})) = \pi \left(\Delta(G_n)^2\right)\) (see e.g. \(\text{\cite{Gey92}}\) Eq (2.2)) and the formula in \(\text{(6)}\) imply

\[
\sigma^2_n \leq \frac{1 + \rho}{1 - \rho} \cdot \int_{\sigma(P|\mathcal{H})} E_{\Delta(G_n)}(d\lambda) = \frac{1 + \rho}{1 - \rho} \cdot \pi \left(\Delta(G_n)^2\right) \rightarrow 0 \quad \text{as} \quad n \to \infty.
\]

This proves the proposition. \(\square\)
3.3. Uniform convergence to stationarity. Fix an exhaustive sequence of allotments \((X_n)_{n \in \mathbb{N}}\) and stochastic matrices \(p_n, n \in \mathbb{N}\), as in Theorem 2.6. The main aim of this section is to prove that the corresponding chains are geometrically ergodic uniformly in \(n \in \mathbb{N}\). This is achieved as follows: first, the uniform drift, minorisation and strong aperiodicity conditions in \((\text{10}), (\text{11})\) and \((\text{12})\), respectively, are established. Then, the uniform convergence to stationarity follows from \([\text{Bax05 Thm 1.1}]\) (cf. \([\text{MT94 Thm 2.3}]\)).

For each \(n \in \mathbb{N}\), let \(a^n : \mathbb{R}^d \to \mathbb{R}^d\) map \(x \in \mathbb{R}^d\) to its representative in \(X_n\). More precisely, let
\[
(7) \quad a^n(x) := \sum_{j=0}^{m_n} a^n_j 1_{j^n}(x) \quad \text{for every } x \in \mathbb{R}^d,
\]
where \(\{J^n_0, \ldots, J^n_m\}\) is the partition and \(X_n = \{a^n_0, \ldots, a^n_m\}\) are the representatives in the allotment \(X_n\). Since the sequence of allotments is exhaustive, the following limit holds:
\[
(8) \quad \lim_{n \to \infty} a^n(x) = x \quad \text{for every } x \in \mathbb{R}^d.
\]
Note that the definition of a \(V\)-mesh (see \([\text{3}]\) in Definition 2.4) implies the inequality
\[
(9) \quad V(a^n(x)) - V(a^n(x)) - V(x) + V(x) \leq (1 + \delta_n)V(x) \quad \text{for all } n \in \mathbb{N} \text{ and } x \in \mathbb{R}^d.
\]

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

(a) There exist positive constants \(\lambda < 1, \kappa\), such that the uniform drift condition holds:
\[
(10) \quad p_n \nu_n(a^n_j) \leq \lambda \nu_n(a^n_j) + \kappa 1_C(a^n_j) \quad \text{for all } n \in \mathbb{N}, \text{ and } a^n_j \in X_n.
\]

(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 \(\nu_n\), concentrated on \(X_n\), such that the uniform minorisation condition,
\[
(11) \quad (p_n)_{ij} \geq \gamma \nu_n(\{a^n_j\}) \quad \text{for all } n \in \mathbb{N}, \text{ and } i, j \in \{0, 1, \ldots, m_n\} \text{ satisfying } a^n_i \in C_n,
\]
and the uniform strong aperiodicity condition,
\[
(12) \quad \gamma \nu_n(C_n) \geq \tilde{\gamma} \quad \text{for all } n \in \mathbb{N},
\]
hold.

**Proof.** (a) Fix an arbitrary \(n \in \mathbb{N}\) and \(j \in \{0, \ldots, m_n\}\). By definition of the function \(a^n(\cdot)\) in \((7)\), we find
\[
p_n \nu_n(a^n_j) - \nu_n(a^n_j) = \int_{\mathbb{R}^d} \left( V(a^n(y)) - V(a^n_j) \right) \alpha(a^n_j, y) q(a^n_j, y) dy.
\]
By \((9)\) we get \(V(a^n(y)) - V(a^n_j) \leq V(y) - V(a^n_j) + \delta_n V(y)\) for every \(y \in \mathbb{R}^d\). The form of kernel \(P\) in \([\text{MH}(q, \pi)\]) and this inequality imply
\[
p_n \nu_n(a^n_j) - \nu_n(a^n_j) \leq PV(a^n_j) - V(a^n_j) + \delta_n \int_{\mathbb{R}^d} V(y) \alpha(a^n_j, y) q(a^n_j, y) dy
\]
\[
\leq PV(a^n_j) - V(a^n_j) + \delta_n PV(a^n_j) = (1 + \delta_n)PV(a^n_j) - V(a^n_j).
\]
Since by definition \(V(a^n_j) = \nu_n(a^n_j)\), the geometric drift condition in \([\text{A1}]\) implies
\[
p_n \nu_n(a^n_j) \leq (1 + \delta_n) \lambda V \nu_n(a^n_j) + (1 + \delta_n) \kappa 1_C \nu_n(a^n_j).
\]
Since $\lim_{n \to \infty} \delta_n = 0$, if we define $C := C_V$, $\lambda := \frac{1 + cV}{2}$ 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 (10) holds for all $n \geq N_0$. Note that if we enlarge $C$ and increase $\kappa$, the uniform drift condition in (10) 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 $X_1, \ldots, X_{N_0}$ (finitely many points) and increase $\kappa$ sufficiently, so that (10) also holds for all $n \in \{1, \ldots, N_0 - 1\}$.

(b) Recall that by Definition 2.4(c), the sequence $(r_n := \text{rad}(X_n, V))_{n \in \mathbb{N}}$ 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 $D \subseteq \bigcap_{n \geq n_0} V^{-1}([1, r_n])$. We now enlarge the compact set $C$, constructed in part (a) of this proof, to contain the bounded set

$$
(\bigcup_{n < n_0} \mathbb{R}^d \setminus J_n^0) \cup \bigcap_{n \geq n_0} V^{-1}([1, r_n]).
$$

We may assume the set $C$ is still compact, since the set in (13) is bounded, and hence the uniform drift condition in (10) still holds.

Define a measure $\nu$ on the Borel $\sigma$-algebra of $\mathbb{R}^d$ by $\nu(B) := \frac{\mu^{\text{Leb}}(B \cap C)}{\mu^{\text{Leb}}(C)}$ for any measurable set $B$. For each $n \in \mathbb{N}$, define a measure on the set of representatives $X_n$ by $\nu_n(\{a^n_j\}) := \nu(J^n_j)$. Define the constant $\gamma := \mu^{\text{Leb}}(C) \inf_{x, y \in C} \alpha(x, y)q(x, y)$ and note that it is strictly positive by Assumptions A2 and A3 and Definition 2.4(a). For every $n \in \mathbb{N}$ and every $0 \leq i, j \leq m_n$, such that $a^n_i \in C_n$, the form of the kernel $P$ in $\text{MH}(q, \pi)$ implies the minorisation condition in (11):

$$(p_n)_{ij} = P(a^n_i, J^n_j) \geq \int_{J^n_j \cap C} \alpha(a^n_i, y)q(a^n_i, y)dy \geq \gamma \nu(J^n_j) = \gamma \nu_n(\{a^n_j\}).$$

We now establish the strong aperiodicity condition in (12). 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 $r_n := \text{rad}(X_n, V)$ in (2) implies $D \cap J^n_0 \subseteq V^{-1}([1, r_n]) \cap V^{-1}(\{r_n, \infty\})$ and hence $D \cap J^n_0 = \emptyset$. Since the radius 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 (7), it holds that $D' \subseteq \bigcup_{\{i : a^n_j \in C\}} J^n_j$ and

$$
\nu_n(C_n) = \nu_n(X_n \cap C) = \nu \left( \bigcup_{\{j : a^n_j \in C\}} J^n_j \right) \geq \nu(D') = \frac{\mu^{\text{Leb}}(D')}{\mu^{\text{Leb}}(C)} > 0.
$$

If $n < n_0$, then it holds that $C_n = X_n \cap C \supset \{a^n_j : j = 1, \ldots, m_n\}$, since $C$ contains the set in (13) and hence $\mathbb{R}^d \setminus J^n_0$. Therefore we find $\nu_n(C_n) \geq \frac{\mu^{\text{Leb}}(\mathbb{R} \setminus J^n_0)}{\mu^{\text{Leb}}(C)} > 0$. Hence (12) holds for the positive constant

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

This concludes the proof of the proposition. \(\square\)

Proposition 3.3 allows us to control the convergence to stationarity of the approximating chains uniformly in $n \in \mathbb{N}$. In the notation of Theorem 2.6 and Remarks 2.7 and 3.1 the following statement holds.
Proposition 3.4. There exist positive constants $\zeta$ and $\theta < 1$, such that the inequality
\[ \sup_{\|g\|_{v_n} \leq 1} \left| (p^k_n g)(b) - \pi_n(g) \right| \leq \zeta \theta^k v_n(b) \]
holds for all $b \in X_n$, $k \in \mathbb{N} \cup \{0\}$ and $n \in \mathbb{N}$, where the $v_n$-norm of a function $g : X_n \to \mathbb{R}$ is $\|g\|_{v_n} := \sup_{b \in X_n} |g(b)|/v_n(b)$ and $\pi_n(g)$ denotes the integral (i.e. weighted sum) of $g$ with respect to $\pi_n$.

Proof. Pick an arbitrary $n \in \mathbb{N}$. According to Proposition 3.3, the transition matrix $p_n$ satisfies the drift condition in (10), the minorisation condition in (11) and the strong aperiodicity condition (12) with the constants $\kappa, \lambda, \gamma, \tilde{\gamma}$, which are independent of the choice of $n$. Hence, [Bax05, Theorem 1.1] (see also [MT94, Theorem 2.3]) applied to the transition kernel $p_n$ on the state space $X_n$, yields
\[ \sup_{\|g\|_{v_n} \leq 1} \left| (p^k_n g)(a^n_j) - \pi_n(g) \right| \leq \zeta(n)v_n(a^n_j)\theta(n)^k \]
for every $k \in \mathbb{N} \cup \{0\}$, $a^n_j \in X_n$ and constants $\zeta(n) \in (0, \infty)$ and $\theta(n) \in (0, 1)$. Furthermore, [Bax05, Theorem 1.1] implies that the constants $\zeta(n), \theta(n)$ are only a (chain independent) function of $\kappa, \lambda, \gamma, \tilde{\gamma}$ in Proposition 3.3 and hence do not depend on $n$. This concludes the proof. \hfill \Box

3.4. Functions that asymptotically solve Poisson’s equation PE($P, F$). In this section we complete the proof of Theorem 2.6. By the Dominated Convergence Theorem (DCT), Proposition 3.2 implies that $(\tilde{F}_n)_{n \in \mathbb{N}}$ asymptotically solves PE($P, F$) if the following conditions hold:
\[ (14) \quad \sup_{n \in \mathbb{N}} \|\Delta(\tilde{F}_n)\|_V < \infty \quad \text{and} \quad \lim_{n \to \infty} \Delta(\tilde{F}_n)(x) = 0 \quad \text{for } \pi\text{-a.e. } x \in \mathbb{R}^d. \]

The inequality in (14) follows from (5) and Proposition 3.5 below, which states that the $V$-norm $\tilde{F}_n$, shifted by an appropriate constant, is bounded uniformly in $n \in \mathbb{N}$. The existence of these constants rests on the uniform convergence to stationarity in Proposition 3.4 above.

The limit in (14) is established by bounding $|\Delta(\tilde{F}_n)|$ by a sum of three non-negative terms (see Lemma 3.8 below) and controlling each one separately. The first, given by $|F(x) - F(a^n(x))|$, tends to zero by (8) since the force function $F$ is assumed to be continuous $\pi$-a.e. The second term $|U(x) - U(a^n(x))|$, where $U := P\tilde{F}_n - \tilde{F}_n$, is controlled by Proposition 3.5 and the DCT. Controlling the third term $|\pi_n(f_n) - \pi(F)|$ is more involved. It requires constructing a further approximating chain (based on the transition kernel $P$) with state space $X_n$ and a transient matrix $p^n_n$, whose invariant distribution can be described analytically in terms of the density $\pi$ (see (15) below). Proposition 3.7, whose proof also depends on the uniform convergence to stationarity in Proposition 3.4, establishes the desired limit. We now give the details of the outlined proof.

Proposition 3.5. There exists a constant $\xi > 0$ and a sequence of real numbers $(c_n)_{n \in \mathbb{N}}$, such that the following inequality holds for all $n \in \mathbb{N}$:
\[ \|\tilde{F}_n + c_n\|_V \leq \xi. \]

Proof. Pick an arbitrary $n \in \mathbb{N}$. Since $F \in L^\infty_V$ by assumption, its restriction $f_n : X_n \to \mathbb{R}$ satisfies $\|f_n\|_{v_n} \leq \|F\|_V$ (see Proposition 3.4 for definition of $v_n$-norm). By Proposition 3.4 the function
\( \tilde{f}_n : X_n \to \mathbb{R} \), given by
\[
\tilde{f}_n := \sum_{k=0}^{\infty} (p^k_n f_n - \pi_n(f_n)),
\]
is well defined and satisfies the inequality \( \| \tilde{f}_n \|_{v_n} \leq \frac{\zeta}{1-\theta} \| f_n \|_{v_n} \leq \frac{\zeta}{1-\theta} \| F \|_V \). Furthermore, by [MT09 Thm. 17.4.2], the function \( \tilde{f}_n \) solves Poisson’s equation \( \text{PE}(p_n, f_n) \). Since \( \tilde{f}_n : X_n \to \mathbb{R} \), in the definition of \( \hat{F}_n \), also solves \( \text{PE}(p_n, f_n) \) by Remark 2.7 there exists a constant \( c_n \in \mathbb{R} \) such that \( \tilde{f}_n + c_n = \hat{f}_n \).

Recall that \( \hat{F}_n = \sum_{j=0}^{m_n} \hat{f}_n(a^n_j) 1_{j_n} \), pick an arbitrary \( x \in \mathbb{R}^d \) and note that definition (7) implies \( \hat{F}_n(x) = \hat{f}_n(a^n(x)) \). Hence, we obtain
\[
\left| \hat{F}_n(x) + c_n \right| &= \left| f_n(a^n(x)) \right| \leq \frac{\zeta}{1-\theta} \| F \|_V v_n(a^n(x)) \leq \frac{\zeta}{1-\theta} \| F \|_V \pi(a^n(x)) \leq \xi V(x),
\]
and the last inequality follows from (9). Since both \( x \in \mathbb{R}^d \) and \( n \in \mathbb{N} \) were arbitrary, this implies the proposition. \( \square \)

In order to analyse the behaviour of the limit in (14), we need to define a further approximating Markov chain on \( X_n \) with the transition matrix \( p^*_n \) and the invariant measure \( \pi^*_n \), given by
\[
(p^*_n)_{ij} := \int_{J^n_i} \frac{\pi(x)}{\pi(J^n_j)} P(x, J^n_j) dx \quad \text{and} \quad \pi^*_n \{(a^n_j)\} := \pi(J^n_j), \quad \text{for } i, j \in \{0, \ldots, m_n\},
\]
respectively. Note that \( (p^*_n)_{ij} = \mathbb{P}_x[\Phi_1 \in J^n_i | \Phi_0 \in J^n_j] \), where \( \Phi \) is the Metropolis-Hastings chain we are analysing. It is clear from the definition in (15) that the equality \( \pi^*_n p^*_n = \pi^*_n \) holds. Furthermore, if we define a function \( h_n : X_n \to \mathbb{R} \) by
\[
h_n(a^n_j) := \int_{J^n_i} \frac{\pi(x)}{\pi(J^n_j)} F(x) dx \quad \text{for } a^n_j \in X_n \), it holds that \( \pi^*_n(h_n) = \pi(F) \).

Remark 3.6. \quad (i) Let \( \mu \) be a signed measure on \( X_n \) and \( \| \mu \|_{v_n} := \sup_{\| g \|_{v_n} \leq 1} | \mu(g) | \) its \( v_n \)-norm, where the norm \( \| g \|_{v_n} \) was defined in Proposition 3.4 and \( \mu(g) \) denotes the integral (i.e. weighted sum) of \( g : X_n \to \mathbb{R} \) with respect to \( \mu \). Furthermore, it is natural to define the dual normed vector spaces \( (L^\infty_{v_n}, \| \cdot \|_{v_n}) \) (analogous to \( L^\infty_V \) in (1)) and \( (M^\infty_{v_n}, \| \cdot \|_{v_n}) \) of functions on \( X_n \) and signed measures on \( X_n \), respectively. Since \( X_n \) is finite, the vector spaces \( L^\infty_{v_n} \) and \( M^\infty_{v_n} \) are isomorphic to \( \mathbb{R}^{1+m_n} \). Furthermore, any linear function \( B : L^\infty_{v_n} \to L^\infty_{v_n} \), mapping \( g \mapsto Bg \), induces a linear map on the dual \( B^* : M^\infty_{v_n} \to M^\infty_{v_n} \), given by \( \mu \mapsto B^* \mu := \mu B \) (in this definition we interpret \( \mu \) as a row vector and \( B \) as a matrix). It is well known that the operator norms coincide \( \| B \|_{v_n} = \| B^* \|_{v_n} \). This fact, which holds in a much more general setting (see [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 \( n \in \mathbb{N}, y \in \mathbb{R}^d \):
\[
\alpha(a^n(x), y) q(a^n(x), y) \leq \frac{q(y, a^n(x))}{\pi(a^n(x))} \leq \eta_x \pi(y), \quad \text{where } \eta_x := \frac{\sup_{z \in \mathbb{R}^d} q(z, y)}{\inf_{y \in \mathbb{N}} \pi(a^n(x))}.
\]
By (8) and \( \text{(A2)} \) we have \( 0 < \inf \{ \pi(z) : |z-x| \leq \sup_{k \in \mathbb{N}} \delta_k \} \leq \pi(a^n(x)), \) where \( \delta_k = \delta(X_k, V) \) (see Definition 2.4), for all sufficiently large \( n \in \mathbb{N} \). Thus, by \( \text{(A2)} \) and \( \text{(A3)} \) we have \( \eta_x \in \mathbb{R} \).
(0, ∞) and the inequalities in (17), which will be used in the proofs of Proposition 3.7 and Theorem 2.6 hold.

Proposition 3.7. The following inequalities hold for the measure π_n^* defined in (15):

\begin{equation}
|\pi_n^* - \pi_n|(f_n) \leq \frac{\zeta ||F||_V}{1 - \theta} ||\pi_n^* - \pi_n^* p_n||_{v_n},
\end{equation}

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

\begin{equation}
||\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} (V(y) + V(x)) Z_n(x, y) dy \pi(x) dx,
\end{equation}

where \( Z_n(x, y) := |\alpha(a^n(x), y) q(a^n(x), y) - \alpha(x, y) q(x, y)| \) for any \( x, y \in \mathbb{R}^d \) and the function \( a^n(\cdot) \) is given in (7). Furthermore, the following limit holds: \( \lim_{n \to \infty} |\pi_n(f_n) - \pi(F)| = 0. \)

Proof. We estimate the difference |\pi_n(f_n) - \pi(F)| using the invariant distribution \( \pi_n^* \) of the chain driven by \( p_n^* \) and the function \( h_n \), defined in (15) and (16) respectively, as follows

\begin{equation}
|\pi_n(f_n) - \pi(F)| = |\pi_n(f_n) - \pi_n^*(f_n) + \pi_n^*(f_n) - \pi_n^*(h_n)|
\end{equation}

\begin{equation}
\leq |(\pi_n - \pi_n^*)(f_n)| + |\pi_n^*(f_n - h_n)|.
\end{equation}

We will prove that both terms on the right-hand side converge to zero as \( n \to \infty \). The definitions of \( \pi_n^* \) and \( h_n \) (in (15) and (16) above) and the function \( a^n(\cdot) \) (see (7)) imply that the second term on the right-hand side of (20) takes the form

\[ \pi_n^*(f_n - h_n) = \sum_{j=0}^{m_n} \pi(J_n^j) \left( F(a^n_j) - \int_{J_n^j} \frac{\pi(x)}{\pi(J_n^j)} F(x) dx \right) = \int_{\mathbb{R}^d} (F(a^n(x)) - F(x)) \pi(x) dx. \]

Since \( F \) is continuous \( \pi \text{-a.e.} \), the integrand converges to zero \( \pi \text{-a.e.} \) by (8). Furthermore, for any \( x \in \mathbb{R}^d \) it holds that

\[ |F(a^n(x)) - F(x)| \leq |F(a^n(x))| + |F(x)| \leq ||F||_V (V(a^n(x)) + V(x)) \]

\[ \leq ||F||_V (2 + \sup_{k \in \mathbb{N}} \delta_k) V(x), \]

where the last inequality follows from (9). Therefore, by the DCT (recall that by the assumption in A1 we have \( \pi(V) < \infty \)), the second term in (20) indeed converges to zero.

Establishing the convergence of the first term on the right-hand side in (20) is more involved. We start by establishing the following representation of the signed measure \( \pi_n^* - \pi_n. \)

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

Define a transition matrix \( 1 \otimes \pi_n \) on the state space \( X_n \) by \( (1 \otimes \pi_n)_{ij} := \pi_n(a^n_i) \). The corresponding chain is a sequence of independent r.v.s. with the law given by \( \pi_n \) (independently of the starting distribution). The inequality in Proposition 3.4 can therefore be expressed as \( \|p_n^k - 1 \otimes \pi_n\|_{v_n} \leq \zeta \theta^k \), for all \( k \in \mathbb{N} \), implying that \( B_n := \sum_{k=0}^\infty (p_n^k - 1 \otimes \pi_n) \) is a well defined linear map on the normed space \( L_{v_n}^\infty \), 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 \( \mu \in M_{v_n}^\infty \) and, by
Remark 3.6(I) and Proposition 3.4, the \( \| \cdot \|_{v_n} \)-norm of the linear operator \( \mu : M_{v_n}^\infty \rightarrow \frac{\mu(p_n^k - 1 \otimes \pi_n)}{v_n} \) 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 

\[
\| \pi_n^* p_n^k - \pi_n \|_{v_n} = \| \pi_n^* (p_n^k - 1 \otimes \pi_n) \|_{v_n} \leq \zeta \theta^k \| \pi_n^* \|_{v_n} \text{ for all } k \in \mathbb{N}.
\]

Consider the identity

\[
(\pi_n^* - \pi_n^* p_n) \sum_{k=0}^\ell (p_n^k - 1 \otimes \pi_n) = \pi_n^* - \pi_n^* p_n^{\ell + 1} \quad \text{for all } \ell \in \mathbb{N},
\]

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

In order to establish the inequality in (18), note that \( \| f_n \|_{v_n} \leq \| F \|_V \) and Remark 3.6(I) imply 

\[
\| (\pi_n^* - \pi_n^*)(f_n) \| \leq \| F \|_V (\pi_n^* - \pi_n)(f_n) / \| f_n \|_{v_n} \leq \| F \|_V \| \pi_n^* - \pi_n \|_{v_n}.
\]

This inequality and the Claim imply (18).

The next task is to prove (19). Let \( g : X_n \to \mathbb{R} \) 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 (7). We apply the definition of the stochastic matrix \( p_n^* \) and its stationary law \( \pi_n^* \), given in (15), to obtain

\[
(\pi_n^* - \pi_n^* p_n) g = \pi_n^* (p_n^* - p_n) g = \sum_{j=0}^{m_n} \sum_{i=0}^{m_n} \left[ \pi(J^n) ((p_n^*)_{ij} - (p_n)_{ij}) \right] g(a^n_j)
\]

\[
= \sum_{j=0}^{m_n} \left[ \int_{\mathbb{R}^d} (P(x, J^n_j) - P(a^n(x), J^n_j)) \pi(x) dx \right] g(a^n_j)
\]

\[
= \int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^d} g(a^n(y)) \left[ \alpha(x, y) q(x, y) - \alpha(a^n(x), y) q(a^n(x), y) \right] dy \right) \pi(x) dx
\]

\[
+ \int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^d} g(a^n(x)) \left[ \alpha(a^n(x), y) q(a^n(x), y) - \alpha(x, y) q(a^n(x), y) \right] dy \right) \pi(x) dx,
\]

where the identity \( \delta_x(J^n_j) g(a^n_j) = \delta_{a^n(x)}(J^n_j) g(a^n_j) = \delta_{a^n(x)}(J^n_j) g(a^n(x)) \), for any \( x \in \mathbb{R}^d \) and \( j \in \{0, \ldots, m_n + 1\} \), implies the final equality. Since the function \( g \in L_{v_n}^\infty \), with \( \| g \|_{v_n} \leq 1 \), in the calculation above was arbitrary and satisfies \( \| g(a^n(x)) \| \leq V(a^n(x)) \) for all \( x \in \mathbb{R}^d \), we find

\[
\| \pi_n^* - \pi_n^* p_n \|_{v_n} = \sup_{\| g \|_{v_n} \leq 1} \| (\pi_n^* - \pi_n^* p_n) g \| \leq \int_{\mathbb{R}^d} \left( V(a^n(y)) + V(a^n(x)) \right) Z_n(x, y) \pi(x) dy dx,
\]

which, together with (9), implies (19).

We now apply the DCT to deduce that the right-hand side in (19) 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 (9) imply the estimates

\[
\int_{\mathbb{R}^d} \left( V(y) + V(x) \right) Z_n(x, y) dy \leq PV(x) + PV(a^n(x)) + 2V(x)
\]

\[
\leq (2 + \sup_{k \in \mathbb{N}} \delta_k \left( \lambda_V + \kappa_V \right) + 2) V(x)
\]

for all \( x \in \mathbb{R}^d \). Since, by Assumption A1, we have \( \pi(V) < \infty \), by the DCT the right-hand side in (19) tends to zero (as \( n \to \infty \)) if

\[
(21) \quad \lim_{n \to \infty} \int_{\mathbb{R}^d} \left( V(y) + V(x) \right) Z_n(x, y) dy = 0 \quad \text{for all } x \in \mathbb{R}.
\]
To establish the limit in (21), 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 (8) and the assumptions in A2 and A3. Hence the integrand in (21) converges to zero point-wise. By the estimate in (17), the integrand in (21) is bounded above by that \( \lim \)

\[
\text{lim}_{n \to \infty} Z_n(x, y) = 0
\]

which does not depend on \( n \) and is \( \mu^{\text{Leb}} \)-integrable in \( y \in \mathbb{R}^d \). Hence the limit in (21) holds by the DTC and, consequently, the right-hand side in (19) converges to zero as \( n \to \infty \). This fact and the estimates in (18) and (19) imply that the first term on right-hand side of (20) tends to zero as \( n \to \infty \) and the proposition follows.

In order to prove that the limit \( \lim_{n \to \infty} \Delta(\hat{F}_n) = 0 \) holds \( \pi \text{-a.e.} \) (i.e. the second condition in (14)), we need the following elementary estimate.

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

\[
\left| \Delta(\hat{F}_n)(x) \right| \leq |F(x) - F(a^n(x))| + |\pi_n(f_n) - \pi(F)| \\
+ \left| \left( P\hat{F}_n - \hat{F}_n \right)(x) - \left( P\hat{F}_n - \hat{F}_n \right)(a^n(x)) \right| \quad \text{for all } x \in \mathbb{R}^d.
\]

**Proof.** Recall that \( \hat{F}_n(x) = \sum_{j=0}^{m_n} \hat{f}_n(a^{n}_j, J^{n}_j)(x) \). Hence also \( P\hat{F}_n(x) = \sum_{j=0}^{m_n} \hat{f}_n(a^{n}_j, P(x, J^{n}_j)) \). The following equalities hold

\[
\Delta(\hat{F}_n)(b) = P(\hat{F}_n - \hat{F})(b) - (\hat{F}_n - \hat{F})(b) = \pi_n(f_n) - \pi(F) \quad \text{for any } b \in X_n,
\]

since \( \hat{F} \) (resp. \( \hat{f}_n \)) solves the Poisson equation in \( \text{PE}(P,F) \) (resp. \( \text{PE}(p_n, f_n) \)). Recall that the function \( a^n(\cdot) \) is defined in (7). Using the definition of \( \Delta(\hat{F}_n) \), the equalities in (22) and the fact that \( \hat{F} \) solves \( \text{PE}(P,F) \) yields

\[
\Delta(\hat{F}_n)(x) = \left( \hat{F} - P\hat{F} \right)(x) - \left( \hat{F} - P\hat{F} \right)(a^n(x)) + \left( \hat{F} - P\hat{F} \right)(a^n(x)) \\
- \left( \hat{F}_n - P\hat{F}_n \right)(a^n(x)) + \left( \hat{F}_n - P\hat{F}_n \right)(a^n(x)) - \left( \hat{F}_n - P\hat{F}_n \right)(x)
\]

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

**Proof of Theorem 2.6:** By Proposition 3.2, it is sufficient to verify that the conditions in (14) hold for the sequence of functions \( (\Delta(\hat{F}_n))_{n \in \mathbb{N}} \). By Proposition 3.5 there exists a constant \( \xi' \) and a sequence \( (c_n)_{n \in \mathbb{N}} \) such that the following estimate holds

\[
\left| \hat{F}_n(x) + c_n - \hat{F}(x) \right| \leq \xi' V(x) \quad \text{for all } n \in \mathbb{N} \text{ and } x \in \mathbb{R}^d.
\]
Note that we have \( \Delta(\tilde{F}_n) = P(\tilde{F}_n + c_n - \tilde{F}) - (\tilde{F}_n + c_n - \tilde{F}) \). The structure of the transition kernel \( P \) in \((\text{MH}(q, \pi))\) implies the following bounds for all \( n \in \mathbb{N} \) and \( x \in \mathbb{R}^d \):

\[
\left| \Delta(\tilde{F}_n)(x) \right| \leq \int_{\mathbb{R}^d} \left( \left| \tilde{F}_n(y) + c_n - \tilde{F}(y) \right| + \left| \tilde{F}_n(x) + c_n - \tilde{F}(x) \right| \right) \alpha(x, y)q(x, y)dy \\
\leq \int_{\mathbb{R}^d} \xi'V(y)\alpha(x, y)q(y, x)dy + \xi'V(x)\int_{\mathbb{R}^d} \alpha(x, y)q(x, y)dy \\
\leq \xi'(PV(x) + V(x)) \leq (\xi' + \xi'\lambda_V + \xi\kappa_V)V(x),
\]

where the last inequality is a consequence of the drift condition in A1. This inequality and the definition of the \( V \)-norm in (1) imply that the first condition in (14) is satisfied.

We now establish the limit in (14). Fix an arbitrary \( x \in \mathbb{R}^d \), such that \( F \) is continuous at \( x \). The first term on the right-hand side of the inequality in Lemma 3.8 therefore converges to zero by (8). The second term, which is independent of \( x \), tends to zero by Proposition 3.7. In order to deal with the third term on the right-hand side of the inequality in Lemma 3.8, note that, by the definition of \( \tilde{F}_n \) in Theorem 2.6, it holds that \( \tilde{F}_n(a^n(x)) = \tilde{F}_n(x) \) for all \( n \in \mathbb{N} \). Consequently, the structure of the transition kernel \( P \) in \((\text{MH}(q, \pi))\) implies that this term equals \( \int_{\mathbb{R}^d}(\tilde{F}_n(y) - \tilde{F}_n(x))[\alpha(x, y)q(x, y) - \alpha(a^n(x), y)q(a^n(x), y)]dy \). The integrand converges to zero for every \( y \in \mathbb{R}^d \) by (8) and Assumptions A2-A3. Furthermore, by Proposition 3.5 we obtain the inequality

\[
(23) \quad \left| \tilde{F}_n(y) - \tilde{F}_n(x) \right| = \left| \tilde{F}_n(y) + c_n - \tilde{F}_n(x) - c_n \right| \leq \xi(V(y) + V(x)) \quad \text{for every} \quad y \in \mathbb{R}^d.
\]

The inequality in (17) yields an upper bound

\[
(24) \quad |\alpha(x, y)q(x, y) - \alpha(a^n(x), y)q(a^n(x), y)| \leq \eta_\pi(y) + \alpha(x, y)q(x, y) \quad \text{for all} \quad y \in \mathbb{R}^d.
\]

The product of the right-hand sides in the inequalities (23) and (24) is integrable over \( \mathbb{R}^d \) with respect to \( \mu^{-\text{Leb}}(dy) \). Hence, the DCT implies that the third term on the right-hand side of the inequality in Lemma 3.8 converges to zero. Therefore, \( \lim_{n \to \infty} \Delta(\tilde{F}_n)(x) = 0 \) holds for all \( x \in \mathbb{R}^d \) at which \( F \) is continuous. It only remains to note that, by the assumption on \( F \) in Theorem 2.6, this limit holds \( \pi \)-a.e. \( \square \)

4. The rate of decay of asymptotic variances

Theorem 2.6 states that, under A1-A3, the asymptotic variance \( \sigma_n^2 \) in \( \text{CLT}(\Phi, F + P\hat{F}_n - \hat{F}_n) \) converges to zero as \( n \to \infty \). This section investigates the speed of this convergence. We show that, under suitable Lipschitz and integrability conditions, the rate of decay is bounded above by the slower of the decay rates of the sequences \( \pi(V^21_{J_0^n}) \) and \( \delta_n^2 = \delta(X_n, V)^2 \) (see Remark 2.1(i) and Equation (3) respectively). This result suggests that, when constructing an exhaustive sequence of allotments (see Definition 2.4 above) with respect to the drift function \( V \), we can guarantee fastest rate of decay of the asymptotic variance \( \sigma_n^2 \) when the growth of the bounded set \( \mathbb{R}^d \setminus J_0^n \) and the decay of the V-mesh of the partition of \( \mathbb{R}^d \setminus J_0^n \) are balanced appropriately (\( \delta_n^2 \) and \( \pi(V^21_{J_0^n}) \) must be comparable in size as \( n \to \infty \)).
Theorem 4.1. Let the assumptions of Theorem 2.6 be satisfied and assume that the conditions
\begin{align}
\limsup_{n \to \infty} \delta_n^{-2} \int_{\mathbb{R}^d \setminus J_0^c} \left( \int_{\mathbb{R}^d} (V(x) + V(y))Z_n(x, y) dy \right)^2 \pi(x) dx < \infty, \tag{25} \\
\limsup_{n \to \infty} \delta_n^{-2} \int_{\mathbb{R}^d \setminus J_0^c} |F(x) - F(a^n(x))|^2 \pi(x) dx < \infty \tag{26}
\end{align}
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 (\ref{eq:an}). Then there exists a constant \( C_0 > 0 \) such that
\[ \sigma_n^2 \leq C_0 \max\{\pi(V^21_{J_0^c}), \delta_n^2\} \quad \text{for all } n \in \mathbb{N}. \]

Theorem 4.1 proved in Section 4.1 below, holds under general conditions that may be hard to verify in specific examples as the functions in (25)–(26) depend on the drift function \(|\nabla V|\) and Remark 2.1(iv)). Then the kernel \( P \) builds on the results in [RT96a, JH00].

Consider the class of Random walk Metropolis chains in \( \mathbb{R}^d \). Put differently, the proposal density takes the form \( q(x, y) = q^*(y - x) \) for some density \( q^*: \mathbb{R}^d \to \mathbb{R} \). Assume \( q^* \) is continuous, strictly positive and bounded. Assume also that the target \( \pi \) is continuously differentiable, positive and satisfies:
\[ \lim_{|x| \to \infty} \frac{x}{|x|} \cdot \nabla \log \pi(x) = -\infty \quad \text{and} \quad \limsup_{|x| \to \infty} \frac{x}{|x|} \cdot \frac{\nabla \pi(x)}{|\nabla \pi(x)|} < 0. \tag{27} \]

Under these assumptions the kernel \( P \) in (\ref{eq:MH}) satisfies A1–A3 with a drift function \( V_{\gamma} := c_{\gamma} \pi^{-\gamma} \) (where \( c_{\gamma} \) is a constant that ensures \( V_{\gamma} > 1 \) for any \( 0 < \gamma < \frac{1}{2} \) (see [JH00] Thms 4.1 and 4.3) and Remark 2.1(iv)). Then the \( V_{\gamma} \)-radius (see (\ref{eq:radius})) equals \( \text{rad}(X_n, V_{\gamma}) = \inf_{y \in J_0^c} c_{\gamma} \pi^{-\gamma}(y) \) and the \( V_{\gamma} \)-mesh \( \delta_{\gamma,n} = \delta(X_n, V_{\gamma}) \), defined in (\ref{eq:delta}), takes the form
\[ \delta_{\gamma,n} = \max \left( \sup_{x \in J_0^c} |x - a^n(x)|, \sup_{x \in \mathbb{R}^d} (\pi(x)/\pi(a^n(x)))^{\gamma} - 1 \right). \tag{28} \]

The main assumptions in Proposition 4.3 below are:

(i) there exists a function \( K_q: \mathbb{R}^d \to \mathbb{R} \) and \( \gamma_q > 0 \) such that
\[ \int_{\mathbb{R}^d} K_q(z) dz < \infty \quad \text{and} \quad |q^*(z) - q^*(\tilde{z})| \leq |z - \tilde{z}| K_q(z) \quad \text{for all } z, \tilde{z} \in \mathbb{R}^d \text{ with } |z - \tilde{z}| < \epsilon_q; \tag{29} \]

(ii) there exist constants \( \beta \in (\frac{1}{2}, 1), c_{\beta} > 0 \) and \( \epsilon_\pi > 0 \) such that
\[ |\nabla \pi(\tilde{x})| < c_{\beta} \pi(x)^{\beta} \quad \text{for all } x, \tilde{x} \in \mathbb{R}^d \text{ with } |x - \tilde{x}| < \epsilon_\pi. \tag{30} \]

Remark 4.2. Assumption (29) is a version of a local Lipschitz condition and holds for many proposals \( q^* \) used in practice, e.g. normal densities. Assumption (30) and condition (27) hold for instance when target density \( \pi \) is proportional to \( e^{-p(x)} \), for a polynomial \( p \) of degree \( k \) with leading order terms \( p_k \) satisfying \( p_k(x) \to \infty \) as \( |x| \to \infty \).

An application of Theorem 4.1 in this setting yields the following result.
Proposition 4.3. Assume that (29) hold and fix $\gamma \in (0, \beta - \frac{1}{2})$. Let $(\mathcal{X}_n)_{n \in \mathbb{N}}$ be an exhaustive sequence of allotments with respect to $V_\gamma$ defined above. Let $F \in L^\infty_{\mathcal{V}_\gamma}$ be a continuously differentiable function satisfying the inequality $|\nabla F(x)| < c_F \pi^{\gamma - \frac{1}{2}}(x)$ for all $x, \bar{x} \in \mathbb{R}^d$ with $|x - \bar{x}| < \epsilon_F$ (for some constants $c_F, \epsilon_F > 0$). Then there exists a constant $C_{\gamma} > 0$ such that the asymptotic variance $\sigma^2_n \in CLT(\Phi, F + PF_n - F_n)$ where $F_n$ is constructed by the Scheme with input $P$, $F$ and $\mathcal{X}_n$, satisfies

$$\sigma^2_n \leq C_{\gamma} \max \left( \delta^2_{\gamma,n}, \int_{J^*_0} \pi^{1-2\gamma}(x)dx \right)$$

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

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

4.1. Proofs.

Proof of Theorem 4.1. Proposition 3.2 implies that there exists a constant $C_1 > 1$ such that $\sigma^2_n \leq C_1 \cdot \pi(\Delta(F_n)^2)$ for every $n \in \mathbb{N}$. Thus, $\limsup_{n \uparrow \infty} \sigma^2_n / \pi(\Delta(F_n)^2) < \infty$. Furthermore, the inequality in (14) implies that $\limsup_{n \uparrow \infty} \pi(\Delta(F_n)^21_{J_0^n}) / \pi(V^21_{J_0^n}) < \infty$.

Lemma 3.8 yields $\pi(\Delta(F_n)^21_{\mathbb{R}^d \setminus J_0^n}) \leq 3(T_1(n) + T_2(n) + T_3(n)$), where

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

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

Assumption (26) implies $\limsup_{n \uparrow \infty} T_2(n)/\delta^2_n < \infty$. The form of the kernel $P$ in MH($q, \pi$) and the fact that $\tilde{F}_n(x) = \tilde{F}_n(a^n(x))$ for all $x \in \mathbb{R}^d$ yield

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

The inequality in (14) therefore yields

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

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

Note that $T_3(n) = |\pi_n(f_n) - \pi(F)| \leq 2(|\pi_n - \pi_n^*|)(f_n)|^2 + 2|\pi_n^*(f_n - h_n)|^2$ (recall (15)–(16)). Since $\pi_n^*(f_n - h_n) = \int_{\mathbb{R}^d} (F(x) - F(a^n(x))\pi(x)dx$, the inequality $F \leq \|F\|_{V'V}$ and (9) hold, we find

$$|\pi_n^*(f_n - h_n)|^2 \leq \int_{\mathbb{R}^d} |F(x) - F(a^n(x))|^2 \pi(x)dx \leq \|F\|^2_{V'V}(2 + \sup_{n \in \mathbb{N}} \sigma_n^2)\pi(V^21_{J_0^n}) + \int_{\mathbb{R}^d \setminus J_0^n} |F(x) - F(a^n(x))|^2 \pi(x)dx.$$

Therefore (26) yields $\limsup_{n \uparrow \infty} |\pi_n^*(f_n - h_n)|^2 / \max(\pi(V^21_{J_0^n}), \delta^2_n) < \infty$. Similarly, inequalities (18) and (19) in Proposition 3.7 imply

$$\limsup_{n \uparrow \infty} \frac{|(\pi_n - \pi_n^*)(f_n)|^2}{\int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^d} (V(y) + V(x))Z_n(x, y)dy \right)^2 \pi(x)dx} < \infty.$$
Again, splitting the integral with respect to $x$ into the parts over $J^n_1$ and $\mathbb{R}^d \setminus J^n_1$ and applying (25), A1 and (9) yields $\limsup_{n \to \infty} |(\pi_n - \pi^*_n)(f_n)|^2 / \max(\pi(V^2 | J^n_1), \delta_n^2) < \infty$. Hence $\limsup_{n \to \infty} T_3(n) / \max(\pi(V^2 | J^n_1), \delta_n^2) < \infty$. This concludes the proof of the theorem.

**Proof of Proposition 4.3** Since $P$, $F$ and $\chi_n$ in Proposition 4.3 satisfy the assumptions of Theorem 2.6, we need only to establish that conditions (25) and (26) in Theorem 4.1 hold for $V = V_\gamma$ and $\delta_n = \delta_{\gamma,n}$, defined just before Proposition 4.3 above. Then, since $\pi(V^2 | J^n_1) = c^2 \int_{\mathbb{R}^d} \pi^{1 - 2\gamma}(x)dx,$ the proposition will follow by Theorem 4.1.

Start by establishing (26). We have $|x - a^n(x)| < \delta_{\gamma,n}$ for every $x \in \mathbb{R}^d \setminus J^n_0$ by (28). Consequently, Lagrange’s theorem applied to $F$ along a line segment connecting $x$ and $a^n(x)$ yields a point $\tilde{x}^n$ on this segment such that

$$\delta_{\gamma,n}^{-2} \int_{\mathbb{R}^d \setminus J^n_0} |F(x) - F(a^n(x))|^2 \pi(x)dx \leq \int_{\mathbb{R}^d \setminus J^n_0} \left( |F(x) - F(a^n(x))| / |x - a^n(x)| \right)^2 \pi(x)dx$$

$$= \int_{\mathbb{R}^d \setminus J^n_0} |\nabla F(\tilde{x}^n)|^2 \pi(x)dx \leq c_F \int_{\mathbb{R}^d} \pi^{2\gamma - 1}(x)\pi(x)dx = c_F \int_{\mathbb{R}^d} \pi^{2\gamma}(x)dx$$

holds for a sufficiently large $n$ by assumptions on $F$. Target $\pi$ decays supper-exponentially along any ray from the origin and so does $\pi^{2\gamma}$. Thus, the integral $\int_{\mathbb{R}^d} \pi^{2\gamma}(x)dx$ is finite and (26) follows.

Next, we prove that (25) holds. In the setting of a symmetric Random walk Metropolis we have $\alpha(x,y) = \min(1, \pi(x) / \pi(y))$. Let $\mathcal{A}_x := \{ y \in \mathbb{R}^d; \pi(x) \leq \pi(y) \}$ and note that $y \in \mathcal{A}_x$ if and only if $\alpha(x,y) = 1$ and $V_\gamma(x) \geq V_\gamma(y)$. Recall $Z_n(x,y) = |\alpha(x,y)q^*(y - x) - \alpha(a^n(x), y)q^*(y - a^n(x))|$ and, for any $B \subseteq \mathbb{R}^d$ and $x \in \mathbb{R}^d$, denote

$$\mathcal{I}_n(x,B) := \delta^{-2}_{\gamma,n} \left( \int_B (V_\gamma(x) + V_\gamma(y)) Z_n(x,y)dy \right)^2.$$  

Condition (25) is equivalent to $\limsup_{n \to \infty} \int_{\mathbb{R}^d \setminus J^n_0} \mathcal{I}_n(x,\mathbb{R}^d)\pi(x)dx < \infty$. With this in mind, we split the integral in $\mathcal{I}_n(x,\mathbb{R}^d)$ into two integrals, depending on which of the disjoint sets $\mathcal{A}_x$ and $\mathcal{A}^c_x$ the point $y$ belongs to (for any $A \subset \mathbb{R}^d$, $A^c$ denotes $\mathbb{R}^d \setminus A$).

Note that it holds

$$\mathcal{I}_n(x,\mathbb{R}^d) \leq 2\mathcal{I}_n(x,\mathcal{A}_x) + 2\mathcal{I}_n(x,\mathcal{A}^c_x) \quad \text{for all } x \in \mathbb{R}^d.$$

For all sufficiently large $n$, Lagrange’s theorem, (28) and (30) imply that

$$\frac{|\pi(a^n(x)) - \pi(x)|}{\delta_{\gamma,n}} \leq \frac{|\pi(a^n(x)) - \pi(x)|}{|x - a^n(x)|} \leq |\nabla \pi(\tilde{x}^n)| \leq c_\beta \pi^\beta(x) \quad \text{for all } x \in \mathbb{R}^d \setminus J^n_0.$$  

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

$$Z_n(x,y) \leq \alpha(a^n(x), y)|q^*(y - a^n(x)) - q^*(y - x)| + q^*(y - x)|\alpha(x,y) - \alpha(a^n(x), y)|.$$  

If $y \in \mathcal{A}_x$ and $n$ is large enough, then for every $x \in \mathbb{R}^d \setminus J^n_0$, using (29) and (31), the right hand side of (32) can be further bounded as follows (note that $\pi(a^n(x)) \geq \pi(y) \geq \pi(x)$ is crucial in the
analysis of the right term):
\[ Z_n(x, y) \leq \delta_{\gamma,n} K_q^*(y - x) + q^*(y - x) \frac{|\pi(a^n(x)) - \pi(y)|}{\pi(a^n(x))} 1\{a^n(x) > \pi(y)\} (x, y) \leq \delta_{\gamma,n} K_q^*(y - x) + \delta_{\gamma,n} c_{\beta} q^*(y - x) \pi^{\beta - 1}(x). \]

Since the Lebesgue measure is translation invariant, there exists a constant \( c_Z > 0 \) such that for sufficiently large \( n \in \mathbb{N} \) we have
\[ \delta_{\gamma,n} \int_{A_z} Z_n(x, y) dy < c_Z \pi^{\beta - 1}(x) \quad \text{for all } x \in \mathbb{R}^d \setminus J_0^n. \]

As \( y \in A_z \), we have \( V_{\gamma}(x) \geq V_{\gamma}(y) \). Thus, (33) and \( 2\beta - 2\gamma - 1 > 0 \) imply the following:
\[ \int_{\mathbb{R}^d \setminus J_0^n} \mathcal{I}_n(x, A_z) \pi(x) dx \leq \int_{\mathbb{R}^d \setminus J_0^n} 4V_{\gamma}(x)^2 c_Z^2 \pi^{2\beta - 1}(x) dx \]
\[ = 4c_Z^2 \int_{\mathbb{R}^d \setminus J_0^n} \pi^{2\beta - 2\gamma - 1}(x) dx < \infty. \]

If \( y \in A^c_z \) and \( n \) is large enough, then for every \( x \in \mathbb{R}^d \setminus J_0^n \), using (29) and (31), we differently bound the right hand side of (32) as follows:
\[ Z_n(x, y) \leq \frac{\pi(y)}{\pi(a^n(x))} \delta_{\gamma,n} K_q^*(y - x) + q^*(y - x) \frac{\pi(y)}{\pi(a^n(x))} \frac{|\pi(a^n(x)) - \pi(x)|}{\pi(x)} \leq \delta_{\gamma,n} \frac{\pi(y)}{\pi(a^n(x))} \left( K_q^*(y - x) + c_{\beta} q^*(y - x) \pi^{\beta - 1}(x) \right) \]
\[ \leq \delta_{\gamma,n} c_{\pi} \frac{\pi(y)}{\pi(x)} \left( K_q^*(y - x) + c_{\beta} q^*(y - x) \pi^{\beta - 1}(x) \right), \]
where \( c_{\pi} := (1 + \sup_{n \in \mathbb{N}} \delta_{\gamma,n})^{1/\gamma} \) (note that \( \sup_{n \in \mathbb{N}} \sup_{x \in \mathbb{R}^d} \frac{\pi(x)}{\pi(a^n(x))} < c_{\pi} \) by (28)). Hence, similarly to (33) there exists a constant \( c_Z' > 0 \) such that
\[ \delta_{\gamma,n} \int_{A^c_z} Z_n(x, y) dy < c_Z' \pi^{\beta - 1}(x) \quad \text{for all } x \in \mathbb{R}^d \setminus J_0^n. \]

Recall that \( V_{\gamma}(y) \geq V_{\gamma}(x) \) for \( y \in A^c_z \) and apply the Cauchy-Schwarz inequality to obtain for each \( x \in \mathbb{R}^d \setminus J_0^n \) the bound:
\[ \mathcal{I}_n(x, A^c_z) \leq 4\delta_{\gamma,n}^2 \int_{A^c_z} Z_n(x, y) dy \cdot \int_{A^c_z} V_{\gamma}(y)^2 Z_n(x, y) dy \]
\[ \leq 4c_Z^2 c_{\pi} \pi^{\beta - 1}(x) \int_{A^c_z} V_{\gamma}(y)^2 \frac{\pi^\beta(y)}{\pi(x)} \left( c_{\beta} q^*(y - x) \pi^{\beta - 1}(y) + K_q(y - x) \right) dy. \]

The second inequality follows by (33)–(36) and the inequalities \( \pi(y)/\pi(x) < 1 \) and \( \pi(y)^{\beta - 1} \geq \pi(x)^{\beta - 1} \) for \( y \in A^c_z \) (recall that \( \beta \in (1/2, 1) \)). It is clear that if we substitute \( A^c_z \) with \( \mathbb{R}^d \) in (37), the inequality remains true. Hence the Fubini theorem implies
\[ \int_{\mathbb{R}^d \setminus J_0^n} \mathcal{I}_n(x, A^c_z) \pi(x) dx \]
\[ \leq 4c_Z^2 c_{\pi} \int_{\mathbb{R}^d} V_{\gamma}(y)^2 \left( c_{\beta} \pi(y)^{\beta - 1} \int_{\mathbb{R}^d} q^*(y - x) dx + \int_{\mathbb{R}^d} K_q(y - x) dx \right) \pi^\beta(y) dy \]
\[ \leq 4c_Z^2 c_{\pi} c_{\gamma}^2 \left( c_{\beta} \int_{\mathbb{R}^d} \pi^{2\beta - 2\gamma - 1}(y) dy \int_{\mathbb{R}^d} \pi^{\beta - 2\gamma}(y) dy \int_{\mathbb{R}^d} K_q(z) dz \right) < \infty. \]
Account, that \( q^* \) is a density and note that assumptions \( \gamma \in (0, \beta - 1/2) \) and \( \beta \in (1/2, 1) \) imply both \( \beta - 2\gamma, 2\beta - 2\gamma - 1 \in (0, 1) \) making the integrals in (38) finite. This together with (34) implies the inequality \( \limsup_{n \to \infty} \int_{\mathbb{R}^d \setminus \mathbb{R}_0^d} I_n(x, \mathbb{R}^d) \pi(x) dx < \infty \) and (25) follows. \( \square \)

5. Applications of the Scheme

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

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

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

Given the allotment \( (X, \{J_0, \ldots, J_m\}) \), where \( X = \{a_0, \ldots, a_m\} \), and the Metropolis-Hastings kernel \( \text{MH}(q, \pi) \), we have the input required to construct the matrix \( p_X \) (step I) of the Scheme. As the precise computation of its entries is not feasible in general, we construct an estimate \( \tilde{p}_X \) of \( p_X \) via IID Monte Carlo. 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

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

where \( n_1, n_2 \in \mathbb{N} \), random vectors \( Y_{j,l}^i, l_1 = 1, \ldots, n_1 \), are IID uniform in the set \( J_j \) for any \( j \in \{1, \ldots, m\} \) (subscript \( x \) indicates that \( Y_{j,l}^i \) are simulated at the point \( x \) but does not influence
the distribution) and \( Z_{x}^{ij}, l_{2} = 1, \ldots, n_{2}, \) are IID random vectors, independent of all \( Y_{j,x}^{i1} \) and distributed according to the proposal distribution \( q(x,z)dz \) in \( \mathcal{M}(\eta, \pi) \). We construct the matrix \( \hat{P}_{X} \) with entries \((\hat{P}_{X})_{ij} := \hat{P}(a_{i}, a_{j})\) and use it in the Scheme instead of \( P_{X} \).

Given a function \( F: \mathbb{R}^{d} \to \mathbb{R} \), we can execute steps (II)-(III) in the Scheme. Constructing the ergodic average estimator \( S_{k}(F + \hat{P}_{X} - \hat{F}_{X}) \) requires the evaluation of the function \( P_{X} \) along the simulated path \((\Phi_{i})_{i=1,\ldots,k}\) of the Metropolis-Hastings chain. We use the form of \( \hat{F}_{X} \) and the formula in [39] to define

\[
\hat{P}_{X}(x) := \sum_{j=0}^{m} (\hat{f}_{X})_{j} \hat{P}(x, a_{j})
\]

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

This choice of estimator can be justified as follows: since \( Y_{j,\Phi_{i}}^{i1} \) and \( Z_{\Phi_{i}}^{ij} \), generated at each time step \( i \), in the construction of \( \hat{P}_{X}(\Phi_{k}) \) are independent of the past \((\Phi_{j})_{j=1,\ldots,i-1}\), we can construct a Markov chain \( \hat{\Phi} \) with augmented state space \( \mathbb{R}^{d} \times (J_{1})^{n_{1}} \times \cdots \times (J_{m})^{n_{1}} \times (\mathbb{R}^{d})^{n_{2}} \), which keeps track of \( \Phi_{i} \) and the auxiliary variables \( Y_{j,\Phi_{i}}^{i1} \) and \( Z_{\Phi_{i}}^{ij} \). It is not hard to see that the chain \( \Phi \) has a unique invariant measure \( \hat{\pi} \) satisfying \( \hat{\pi}(F + \hat{P}_{X} - \hat{F}_{X}) = \pi(F + \hat{P}_{X} - \hat{F}_{X}) = \pi(F) \).

Furthermore, \( \hat{\Phi} \) is positive Harris recurrent and hence (by [MT09, Theorem 17.1.7]) the SLLN \( S_{k}(F + \hat{P}_{X} - \hat{F}_{X}) \xrightarrow{k \to \infty} \pi(F) \) a.s. holds for any fixed \( n_{1}, n_{2} \in \mathbb{N} \).

**Remark 5.1.** The estimator \( S_{k}(F + \hat{P}_{X} - \hat{F}_{X}) \) is unbiased in the following sense: if the chain \( \Phi \) is started from stationarity (i.e. \( \Phi_{0} \sim \hat{\pi} \)) we have \( \mathbb{E}_{\hat{\pi}} \left[ S_{k}(F + \hat{P}_{X} - \hat{F}_{X}) \right] = \pi(F) \) for any \( k \in \mathbb{N} \). This should be contrasted with the general approach to variance reduction based on the Poisson equation \( \text{PE}(\hat{P}, F) \), where the estimator \( S_{k}(F) \) of \( \pi(F) \) is essential in constructing a guess for the solution of \( \text{PE}(\hat{P}, F) \) and hence the control variate itself (see e.g. [DK12] for this approach applied to random scan Gibbs samplers and [DM16] for sufficiently smooth transition kernels). The latter approach produces a consistent but biased estimator even if the chain is started in stationarity.

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

\[
r_{k,n}(\mathcal{X}) := \frac{\sum_{i=1}^{n}(S_{k}^{i}(F) - \pi(F))^{2}/n}{\sum_{i=1}^{n}(S_{k}^{i}(F + \hat{P}_{X} - \hat{F}_{X}) - \pi(F))^{2}/n},
\]

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 \((S_{k}^{i}(F), S_{k}^{i}(F + \hat{P}_{X} - \hat{F}_{X}))\), for \( i = 1, \ldots, n, \) are IID samples of the pair of ergodic average estimators \((S_{k}(F), S_{k}(F + \hat{P}_{X} - \hat{F}_{X}))\) evaluated on the simulated paths. Put differently, \( r_{k,n} \) is the ratio of mean square errors of estimators \( S_{k}(F) \) and \( S_{k}(F + \hat{P}_{X} - \hat{F}_{X}) \), numerically evaluated on the same random collection of \( n \) independent simulated paths and will serve as an estimate of the improvement.
5.2. Examples. In both examples we use the target law \( \pi := \rho N(\mu_1, \sigma_1^2) + (1 - \rho) N(\mu_2, \sigma_2^2) \), where \( N(\cdot, \cdot) \) is a normal distribution of the appropriate dimension.

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

Let the allotment \( \mathcal{X}_m \) be defined so that \( J_0^m := \mathbb{R} \setminus (-8, 7] \) and \( J_j^m \) for \( j = 1, 2, \ldots, m \) are intervals of equal length partitioning \((-8, 7]\). We take \( a_j^m \) for \( j > 0 \) to be the center of the interval \( J_j^m \) and we take \( a_0^m = -8 \). We construct \( \hat{p}_{\mathcal{X}_m} \) by the formula in (39) (using \( n_1 = n_2 = 1000 \)) and \( \hat{P}\hat{F}_{\mathcal{X}_m} - \hat{F}_{\mathcal{X}_m} \) by the formulae in (39)–(40) (using \( n_1 = 1, n_2 = 10 \)) and then use (41) to estimate the factor of improvement of the estimator \( S_k(F + \hat{P}\hat{F}_{\mathcal{X}_m} - \hat{F}_{\mathcal{X}_m}) \) in comparison to the estimator \( S_k(F) \).

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

<table>
<thead>
<tr>
<th>( m \backslash k )</th>
<th>( k = 5 \cdot 10^3 )</th>
<th>( k = 2 \cdot 10^4 )</th>
<th>( k = 5 \cdot 10^4 )</th>
<th>( k = 2 \cdot 10^5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m = 30 )</td>
<td>5.93</td>
<td>8.56</td>
<td>9.37</td>
<td>9.62</td>
</tr>
<tr>
<td>( m = 50 )</td>
<td>18.0</td>
<td>32.1</td>
<td>34.2</td>
<td>34.7</td>
</tr>
<tr>
<td>( m = 70 )</td>
<td>39.1</td>
<td>75.5</td>
<td>96.8</td>
<td>97.1</td>
</tr>
<tr>
<td>( m = 100 )</td>
<td>76.9</td>
<td>1.76 \cdot 10^2</td>
<td>2.22 \cdot 10^2</td>
<td>2.40 \cdot 10^2</td>
</tr>
<tr>
<td>( m = 300 )</td>
<td>6.96 \cdot 10^2</td>
<td>1.75 \cdot 10^3</td>
<td>2.13 \cdot 10^3</td>
<td>2.36 \cdot 10^3</td>
</tr>
<tr>
<td>( m = 500 )</td>
<td>2.14 \cdot 10^3</td>
<td>4.64 \cdot 10^3</td>
<td>6.05 \cdot 10^3</td>
<td>6.92 \cdot 10^3</td>
</tr>
<tr>
<td>( m = 700 )</td>
<td>3.77 \cdot 10^3</td>
<td>8.90 \cdot 10^3</td>
<td>1.16 \cdot 10^4</td>
<td>1.32 \cdot 10^4</td>
</tr>
</tbody>
</table>

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

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

To specify the allotment, decompose \( B := (-7, 6] \times (-4, 4] \) into \( 6 = 3 \times 2 \) equally sized rectangles and define them to be \( J_1, J_2 \ldots J_6 \). Take \( J_0 := \mathbb{R}^2 \setminus B, a_0 := (-7, 0) \) and \( a_j \) to be the center of the box \( J_j \) for \( j > 0 \). Construct \( \hat{p}_{\mathcal{X}_m} \) by the formula in (39) (using \( n_1 = n_2 = 1000 \)) and \( \hat{P}\hat{F}_{\mathcal{X}_m} - \hat{F}_{\mathcal{X}_m} \) by the formulae in (39)–(40) (using \( n_1 = 1, n_2 = 10 \)) and estimate the factor of improvement \( r_{k,n} \)
in (41). We obtain approximately a 10% reduction in variance. More precisely we get

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

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

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

Appendix A. Existence of exhaustive allotments

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

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

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

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 \( K^n_x := \{ x_1, x_1 + \epsilon_n \} \times \cdots \times \{ x_d, x_d + \epsilon_n \} \). Clearly, it is possible to pick \( x^1, x^2, \ldots, x^m \in \mathbb{R}^d \) so that sets \( K_j^n := \tilde{L}_n \cap \bigcup_{j=1}^{m_n} K^n_{x_j} \) (for \( 1 \leq j \leq m_n \)) are disjoint and cover \( L_n \) (assume the cover is minimal). Finally, take \( J^n_0 \) to be the closure of \( \mathbb{R} \setminus \bigcup_{j=1}^{m_n} K^n_j \) and define \( J^n_j := K^n_j \setminus J^n_0 \). Note that \( \mu^{\text{Leb}}(J^n_j) > 0 \) for all \( 0 \leq j \leq m_n \). For \( 1 \leq j \leq m_n \) pick arbitrary \( a^n_j \in J^n_j \) and choose \( a_0 \in J^n_0 \), so that \( W(a^n_0) = \inf_{x \in \tilde{L}_n} W(x) \) (possible since \( W \) has bounded sublevel sets and \( J^n_0 \) is closed). Sets \( J^n_j \) together with representatives \( a^n_j \) define an allotment \( \mathcal{X}_n \).

By Pythagoras theorem \( |x - y| < \epsilon_n \sqrt{d} \), for \( x, y \) from the same \( \in J^n_j \). Since \( \epsilon_n < 1 \) and \( K^n_j \cap L_n \neq \emptyset \), we get \( J^n_j \subset \bigcup_{j=1}^{m_n} K^n_j \cap L_n \) for all \( 1 \leq j \leq m_n \). Hence,

\[
\max_{1 \leq j \leq m_n} \sup_{y \in J^n_j} |y - a^n_j| \leq \epsilon_n \sqrt{d}
\]

and by uniform continuity (recall \( W \geq 1 \))

\[
\max_{0 \leq j \leq m_n} \sup_{y \in J^n_j} \frac{W(a^n_j) - W(y)}{W(y)} \leq \frac{1}{n}.
\]

Doing the above for every \( n \in \mathbb{N} \) shows \( \lim_{n \to \infty} \delta(\mathcal{X}_n, W) = 0 \) (by (3)). By (2) and definition of \( L_n, \) \( \text{rad}(\mathcal{X}_n, W) \geq r_n \) for every \( n \in \mathbb{N} \). So, \( \lim_{n \to \infty} \text{rad}(\mathcal{X}_n, W) = \infty. \)

\[ \square \]
References


**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*