Abstract. The “drift-and-minorization” method, introduced and popularized in [Ros95; MT94; MT12], remains the most popular approach for bounding the convergence rates of Markov chains used in statistical computation. This approach requires estimates of two quantities: the rate at which a single copy of the Markov chain “drifts” towards a fixed “small set”, and a “minorization condition” which gives the worst-case time for two Markov chains started within the small set to couple with moderately large probability. In this paper, we build on [Oli12; PS15] and our work [ADS19a; ADS19b] to replace the “minorization condition” with an alternative “hitting condition” that is stated in terms of only one Markov chain.

1. Introduction

Since the seminal article by Gelfand and Smith [GS90], Markov chain Monte Carlo methods (MCMC) have become ubiquitous in statistical computing. These methods suffer from a well-known problem: although it is easy to construct many MCMC algorithms that are guaranteed to converge eventually, it is typically very difficult to tell how long convergence will take for any given MCMC algorithm (see e.g. [JH01; DKSC08]). Even worse, it is usually difficult to tell if an MCMC has converged yet even after running it (see e.g. [GR92] for an important early paper in the “diagnostics” literature, [BBM11] for a proof that the problem is intractable in general, and [Hsu+19; Hub16] for cases where the problem becomes tractable).

In the decades since the publication of Gelfand and Smith [GS90], many techniques have been developed to try to solve this problem for classes of MCMC algorithms that are important in statistics (see e.g. the surveys by Jones and Hobert [JH01] and Diaconis [Dia09]). The most popular of these techniques is the “drift-and-minorization” method, introduced and popularized in [Ros95; MT94; MT12]. The purpose of this note is to replace the minorization condition (see Inequality (1.2)) by a hitting condition (see Definition 2.5). It is straightforward to check that the usual drift-and-minorization condition immediately implies our drift-and-hitting condition, and in this sense our assumptions are weaker (and thus our result stronger) than typical drift-and-minorization conditions; see Section 4 for further discussion of the relationship between these conditions.

We now set some notation and recall an important special case of the drift-and-minorization bound. Throughout this paper, unless otherwise mentioned, we use $X$ to denote the state space of the underlying Markov process. Denote by \( \{g(x, 1, \cdot)\}_{x \in X} \) the transition kernel of some Markov chain on state space $X$ with (unique) stationary measure $\pi$ on a Borel $\sigma$-field $\mathcal{B}[X]$. For any $t \in \mathbb{N}$, we write $g(x, t, \cdot)$ to denote the $t$-step transition kernel generated from $\{g(x, 1, \cdot)\}_{x \in X}$. 

We say that \( g \) exhibits a “drift” or “Lyapunov” condition if there exists a function \( V : X \mapsto [0, \infty) \) and constants \( 0 \leq \lambda < 1, 0 \leq b < \infty \) so that
\[
(gV)(x) = \int V(y)g(x, 1, dy) \leq \lambda V(x) + b, \tag{1.1}
\]
for all \( x \in X \). We say that \( g \) exhibits a “minorization” condition if there exists a set \( C \in \mathcal{B}[X] \), constants \( 0 < \epsilon \leq 1, t_0 \in \mathbb{N} \), and measure \( \mu \) on \( X \) so that
\[
g(x, t_0, \cdot) \geq \epsilon \mu(\cdot) \tag{1.2}
\]
for all \( x \in C \). We say \( g \) satisfies conditions (1.1), (1.2) compatibly if it satisfies both and there exists \( d > \frac{2b}{1 - \lambda} \) such that
\[
C \supset \{ x \in X : V(x) \leq d \}. \tag{1.3}
\]
[Ros95, Thm. 12] says that any kernel \( g \) satisfying these conditions compatibly is geometrically ergodic, and also gives quantitative bounds on the convergence rate:

**Theorem 1.1** (Paraphrase of [Ros95, Thm. 12]). Let \( g \) satisfy conditions (1.1), (1.2) compatibly. Then there exists \( M : X \mapsto [0, \infty) \) and \( 0 < \rho \leq 1 \) such that
\[
\|g(x, t, \cdot) - \pi(\cdot)\|_{TV} \leq M(x)(1 - \rho)^t \tag{1.4}
\]
for all \( x \in X \) and \( t \in \mathbb{N} \). Moreover, there is an explicit formula for \( M \) and \( \rho \) in terms of the constants \( \lambda, b, t_0, \epsilon, d \).

There are more sophisticated versions of this drift-and-minorization bound, including results in the same paper [Ros95], but most are based on two conditions that are similar to (1.1) and (1.2).

The goal of the present paper is to present results similar to Theorem 1.1, in which the minorization condition (1.2) has been replaced by related hitting conditions. Recall that the hitting time of a set \( A \in \mathcal{B}[X] \) for a Markov chain \( \{X_t\}_{t \in \mathbb{N}} \) is:
\[
\tau(A) = \min\{t \in \mathbb{N} : X_t \in A\}. \tag{1.5}
\]

We now introduce the maximum hitting time (of sets with large measure):

**Definition 1.2.** Let \( \alpha \in \mathbb{R}_{>0} \). The maximum hitting time (with parameter \( \alpha \)) is
\[
t_H(g, \alpha) = \sup\{\mathbb{E}_x[\tau(A)] : x \in X, A \in \mathcal{B}[X] \text{ such that } \pi(A) \geq \alpha\}, \tag{1.6}
\]
where \( \mathbb{E}_x \) is the expectation of a measure in the space which generates the underlying Markov process and the subscript \( x \) is the starting point of the Markov process.

When the kernel is clear from the context, we write \( t_H(\alpha) = t_H(g, \alpha) \).

### 1.1. Guide to Paper

We give our main drift-and-hit theorem in Section 2, show that its conditions are satisfied for various common MCMC chains in Section 3, and describe why it can (sometimes) give better results than the usual drift-and-minorization theorem in Section 4.
1.2. Previous Work. This paper is based on the asymptotic equivalence of mixing and hitting times as established by in the sequence of papers [Oli12; PS15; BHP15; ADS19a; ADS19b]. This relationship allows us to show that our hitting condition (see Definition 2.5) implies something very close to the minorization condition (1.2), and thus to use the framework of Rosenthal [Ros95]. We note that our recent works [ADS19a; ADS19b] are based on the work of Peres and Sousi [PS15] and Oliveira [Oli12], which initially established this equivalence of mixing and hitting times in the special case that $X$ is finite. [ADS19a] and [ADS19b] make heavy use of hyperfinite representation of Markov processes, which is established by Duanmu, Rosenthal, and Weiss [DRW18].

2. Bounded Mixing Times for Dominated Chains

We recall the definition of mixing time:

**Definition 2.1.** Fix $\epsilon \geq 0$ and $A \in \mathcal{B}[X]$. The mixing time in $A$ with respect to $\epsilon$ is defined as

$$t_m(\epsilon, A) = \min \{ t \geq 0 : \sup_{x \in A} \| g(x, t, \cdot) - \pi(\cdot) \|_{TV} \leq \epsilon \}. \quad (2.1)$$

The lazy mixing time in $A$ with respect to $\epsilon$ is

$$t_L(\epsilon, A) = \min \{ t \in \mathbb{N} : \sup_{x \in A} \| g_L(x, t, \cdot) - \pi(\cdot) \|_{TV} \leq \epsilon \}, \quad (2.2)$$

where $g_L(x, 1, \cdot) = \frac{1}{2} g(x, 1, \cdot) + \frac{1}{2} \delta_x(\cdot)$.

For notational convenience, we write $t_m(A)$ or $t_L(A)$ to mean $t_m(\frac{1}{4}, A)$ or $t_L(\frac{1}{4}, A)$, write $t_m(\epsilon)$ or $t_L(\epsilon)$ to mean $t_m(\epsilon, X)$ or $t_L(\epsilon, X)$, and write $t_m$ or $t_L$ to mean $t_m(\frac{1}{4}, X)$ or $t_L(\frac{1}{4}, X)$.

In practice, it is rare for transition kernels $g$ corresponding to MCMC algorithms on unbounded state spaces to have finite mixing times (this would require jumps of unbounded size, and these are often difficult to construct). For this reason, the mixing and maximal hitting times cannot be directly compared in a meaningful way. We will fix this problem by constructing related Markov chains that both (i) have finite mixing times and (ii) have dynamics that are “slightly faster” than those of $g$.

We introduce the following definition to formalize this idea:

**Definition 2.2 (Dominated Chain).** Let $g$ be a transition kernel on state space $X$ with associated $\sigma$-field $\mathcal{B}[X]$. We say that a transition kernel $g^{(S)}$ with support contained in $S$ is $S$-dominated by $g$ if it satisfies the following two properties:

1. The stationary measure $\pi^{(S)}$ of $g^{(S)}$ is given by

$$\pi^{(S)}(A) = \frac{\pi(A \cap S)}{\pi(S)} \quad (2.3)$$

for all $A \in \mathcal{B}[X]$.

2. The transitions of $g^{(S)}$ satisfy

$$g^{(S)}(x, 1, A) \geq g(x, 1, A) \quad (2.4)$$

for all $x \in S$ and $A \subset S$ with $A \in \mathcal{B}[X]$.

There are many constructions in the literature that satisfy this $S$-domination property, such as the usual trace of a chain on $S$. In this paper, we focus on restrictions of Gibbs samplers and Metropolis–Hastings chains.
Definition 2.3 (Restricted Sampler). Recall that there is a unique Metropolis–Hastings kernel associated with any reversible proposal kernel $g$ and target distribution $\pi$. Let $g$ be any reversible transition kernel with stationary measure $\pi$ and let $S \in \mathcal{B}[X]$ have $\pi(S) > 0$. We define the restriction $g^{(S)}$ of $g$ to $S$ to be the Metropolis–Hastings kernel with proposal $g$ and target $\pi^{(S)}$ as in Eq. (2.3).

In the special case of the Gibbs sampler, another restriction is sometimes useful:

Definition 2.4 (Restricted Gibbs Samplers). Recall that the Gibbs sampler is uniquely defined by the target distribution. If $g$ is a Gibbs sampler targeting $\pi$ and $S \in \mathcal{B}[X]$ has $\pi(S) > 0$, we define the Gibbs restriction $g^{(S)}$ of $g$ to $S$ to be the Gibbs sampler $g^{(S)}$ targeting the measure $\pi^{(S)}$ defined in Eq. (2.3).

It is clear that both of these constructions give $S$-dominated chains, as does the usual trace process (see e.g. Section 6 of [BL10] for the last result); in general, all three are different. For the remainder of this section, we will always fix a kernel $g$ and denote by $g^{(C)}$ any specific kernel that is $C$-dominated by $g$.

The following hitting condition plays an essential role throughout the paper.

Definition 2.5. Fix $C \in \mathcal{B}[X]$. For $\alpha \in \mathbb{R}_{>0}$, denote by $t^{(C)}_H(\alpha)$ the maximum hitting time of large sets for $g^{(C)}$. We say that $g$ satisfies the $C$-hitting condition if there exists $0 < \alpha < 0.5$ such that $t^{(C)}_H(\alpha) < \infty$.

For a fixed collection of constants $D = \{d_\alpha\}_{0<\alpha<0.5}$, define $\mathcal{G}(D)$ to be the collection of kernels $g$ satisfying $t_m \leq d_\alpha t_H(\alpha)$.

We recall that the main results in [Oli12; PS15; BHP15; ADS19a; ADS19b] show that there exist universal constants $D$ so that this is satisfied for all $g$ in certain large classes.

Definition 2.6. We say that $g$ satisfies the drift condition (1.1) and the $C$-hitting condition (Definition 2.5) compatibly for $\alpha, C, D$ if:

1. $g$ satisfies both the drift condition (1.1) and the $C$-hitting condition.
2. $g^{(C)}$ satisfies the drift condition (1.1) with the same $V, \lambda, b$ as $g$.
3. $g^{(C)} \in \mathcal{G}(D)$.
4. $C \supset \{x \in X : V(x) \leq d\}$ for some $d > \frac{2b}{1-\lambda}$.

It is straightforward to check that these conditions imply that $t^{(C)}_H(\alpha) < \infty$ for all $0 < \alpha < 1$, but for convenience we focus on a single such value.

Finally, let $t^*_m$ denote the mixing time for $g^{(C)}$. We now show that we can obtain a minorization condition similar to Eq. (1.2) by combining the hitting condition in Definition 2.5 and the drift condition in Eq. (1.1).

Theorem 2.7. Let $g$ be a transition kernel that satisfies Eq. (1.1) and Definition 2.5 compatibly for some $C \in \mathcal{B}[X]$ with $\pi(C) > 0$ and some $D, \alpha$. Define

$$C' = \{x \in X : V(x) \leq d'\}$$

for some $d' > \frac{2b}{1-\lambda}$. Suppose $C \supset \{x \in X : V(x) \leq d\}$ for some $d > \frac{2b+24d'}{1-\lambda}$. Then, for $t > t^*_m(C')$, we have

$$\sup_{x,y \in C'} \|g(x, t, \cdot) - g(y, t, \cdot)\|_{TV} \leq \frac{2}{3}.$$
Proof. Fix \( x \in C' \) and pick \( 0 < \alpha < 0.5 \) as in Definition 2.5. Let \( T = t_m^{(C)}(C') \leq d\alpha t_H^{(C)}(\alpha) < \infty \). As \( x \in C' \), by assumption, we have

\[
\|g^{(C)}(x, T, \cdot) - \pi^{(C)}(\cdot)\|_{TV} \leq \frac{1}{4}. \tag{2.7}
\]

Next, we denote by \( \{X_t\}_{t \in \mathbb{N}} \) a Markov chain with transition kernel \( g \) starting at point \( X_0 = x \in C' \), and we denote by \( \{Y_t\}_{t \in \mathbb{N}} \) a Markov chain with transition kernel \( g^{(C)} \) starting at the same point \( x \). By Eq. (1.1), we have for all \( t \in \mathbb{N} \)

\[
\mathbb{E}[V(X_t)|X_{t-1}] \leq \lambda V(X_{t-1}) + b. \tag{2.8}
\]

Iterating this bound, we have

\[
\mathbb{E}[V(X_t)] \leq \lambda \mathbb{E}[V(X_{t-1})] + b \leq \lambda (\lambda \mathbb{E}[V(X_{t-2})] + b) + b \\
\leq \ldots \leq \lambda^t V(X_0) + \frac{b}{1 - \lambda}.
\]

By assumption, \( \max\{V(X_0), V(Y_0)\} \leq d' \), so we have

\[
\max\{\mathbb{E}[V(X_t)], \mathbb{E}[V(Y_t)]\} \leq \lambda^t d' + \frac{b}{1 - \lambda} \tag{2.9}
\]

for all \( t \in \mathbb{N} \). By a union bound and Markov’s inequality, we have

\[
\mathbb{P}\left[ \max_{0 \leq t \leq T} \{V(X_t), V(Y_t)\} > d \right] \leq \mathbb{P}\left[ \max_{0 \leq t \leq T} V(X_t) > d \right] + \mathbb{P}\left[ \max_{0 \leq t \leq T} V(Y_t) > d \right]
\]

\[
\leq \frac{2}{d - \frac{b}{1 - \lambda} \left( \sum_{t=0}^{T} \lambda^t d' \right) \leq \frac{2}{d - \frac{b}{1 - \lambda} \left( \frac{d'}{1 - \lambda} \right) \leq \frac{2d'}{d(1 - \lambda) - b}}
\]

\[
= \frac{2d'}{b + 24d'} \leq \frac{1}{12}.
\]

By the coupling inequality and Eq. (2.7), we have \( \|g(x, T, \cdot) - \pi^{(C)}(\cdot)\|_{TV} \leq \frac{1}{4} \). By triangle inequality again, we have \( \sup_{x, y \in C'} \|g(x, t, \cdot) - g(y, t, \cdot)\|_{TV} \leq \frac{2}{3} \) for all \( t > T = t_m^{(C)}(C') \).

The following result is a minor modification of [Ros95, Thm. 12].

**Lemma 2.8.** Let \( g \) satisfy Eq. (1.1). Let \( S \) be a measurable subset of \( X \) such that

1. \( S \supseteq \{x \in X : V(x) \leq d\} \) for some \( d > \frac{2b}{1 - \lambda} \).
2. \( \sup_{x, y \in S} \|g(x, 1, \cdot) - g(y, 1, \cdot)\|_{TV} \leq 1 - \epsilon \) for some \( \epsilon > 0 \).

Then for every \( 0 < r < 1 \) and every \( x \in X \) we have

\[
\|g(x, t, \cdot) - \pi(\cdot)\|_{TV} \leq (1 - \epsilon)^r t + (1 + \frac{b}{1 - \lambda} + V(x))(1 + \frac{2b + \lambda d}{1 + d})^{1-r}(1 + 2(\lambda d + b))^r. \tag{2.10}
\]

**Proof.** We first show the condition \( \sup_{x, y \in S} \|g(x, 1, \cdot) - g(y, 1, \cdot)\|_{TV} \leq 1 - \epsilon \) implies a pseudo-minoration condition from [RR01], which is weaker than the minorization condition. By the definition of total variation distance, for any \( x, y \in S \), there exists \( C_{xy} \in \mathcal{B}[X] \) such that

\[
\|g(x, 1, \cdot) - g(y, 1, \cdot)\|_{TV} = g(x, 1, C_{xy}) - g(y, 1, C_{xy}) \leq 1 - \epsilon. \tag{2.11}
\]
Then we have \( g(x, 1, C_{xy}) + g(y, 1, C_{xy}) \geq \varepsilon \). For any \( A \in \mathcal{B}[X] \), define
\[
\mu_{xy}(A) := \frac{g(y, 1, A \cap C_{xy}) + g(x, 1, A \cap C_{xy})}{g(y, 1, C_{xy}) + g(x, 1, C_{xy})}.
\] (2.12)

Then it can be verified that \( \mu_{xy}(\cdot) \) is a valid probability measure and
\[
g(x, 1, \cdot) \geq \varepsilon \mu_{xy}(\cdot), \quad g(y, 1, \cdot) \geq \varepsilon \mu_{xy}(\cdot).
\] (2.13)

Therefore, \( S \) is a \((1, \varepsilon, \{\mu_{xy}\})\)-pseudo-small set in the sense of \([\text{RR01}]\). The result then follows directly from \([\text{RR01}, \text{Proposition 2}]\). \(\Box\)

We now present the main result of this section.

**Theorem 2.9.** Let \( g \) be a transition kernel that satisfies Eq. (1.1) with function \( V \) and constants \( \lambda \) and \( b \) and Definition 2.5 compatibly for some \( C \in \mathcal{B}[X] \) with \( \pi(C) > 0 \) and some \( D, \alpha \). Define
\[
C' = \{x \in X : V(x) \leq d'\}
\] (2.14)
for some \( d' > \frac{2b}{1-\lambda} \). Suppose \( C \supset \{x \in X : V(x) \leq d\} \) for some \( d > \frac{2b+24d'}{1-\lambda} \). Then there exists \( M : X \mapsto [0, \infty) \) and \( 0 < \rho \leq 1 \) such that
\[
\|g(x, t, \cdot) - \pi(\cdot)\|_{\text{TV}} \leq M(x)(1 - \rho)^t
\] (2.15)
for all \( x \in X \) and \( t \in \mathbb{N} \). Moreover, there is an explicit formula for \( M \) and \( \rho \) in terms of the constants \( \lambda, b, d, D, \alpha \).

**Proof.** It follows from Theorem 2.7 and Lemma 2.8. \(\Box\)

3. Application to Gibbs and Metropolis–Hastings Samplers

3.1. Application to Gibbs Samplers. We begin by studying a large class of Gibbs samplers defined in our companion paper \([\text{ADS19b}]\). We consider a class of Gibbs samplers targeting a measure \( \lambda \) supported on a compact subset of Euclidean space; without loss of generality we assume that the support of \( \lambda \) is inside of \([0, 1]^d\). For \( n \leq d \), let \( \pi_n \) denote the projection to the \( n \)-th coordinate. For \( x \in X \) and \( j \in \{1, 2, \ldots, d\} \), let
\[
A_{x,j} = \{a \in [0, 1]^d : \forall i \neq j, \pi_i(a) = \pi_i(x)\},
\] (3.1)
the line in \([0, 1]^d\) that passes through \( x \) and is parallel to the \( j \)-th coordinate axis. Let \( B_{x,j} \) be the connected component (the largest connected subset) of \( A_{x,j} \) that contains \( x \). For the remainder of this section, we consider fixed \( H \in \{A, B\} \), so that e.g. either \( H_{x,j} = A_{x,j} \) or \( H_{x,j} = B_{x,j} \). We make the following assumption on \( \lambda \):

**Assumption 1.** \( \lambda \) has a continuous density function \( \rho \) with respect to the Lebesgue measure. Furthermore, \( \int_{H_{x,j}} \rho(y)dy > 0 \) for every \( x \in X \) and \( j \in \{1, 2, \ldots, d\} \).

We now set some further notation. Set \( X = [0, 1]^d \) and let \( \lambda_{x,j} \) be the usual conditional distribution of \( \lambda \) on \( H_{x,j} \) (that is, the measure with density \( \rho \) and support \( H_{x,j} \) Define the typical Gibbs sampler with target \( \lambda \) by
\[
g(x, 1, A) = \frac{1}{d} \sum_{j=1}^{d} \lambda_{x,j}(A)
\] (3.2)
for every \( x \in X \) and \( A \in \mathcal{B}[X] \). In the context of the Gibbs sampler only, the \( C \)-dominated chain \( g(C) \) will always refer to the chain from Definition 2.4.
Theorem 3.1 ([ADS19b, Corollary. 4.5]). Let $0 < \alpha < \frac{1}{2}$. Suppose the target distribution $\lambda$ is supported on a compact subset of Euclidean space. Then there exist universal constants $d_\alpha, d'_\alpha$ with the following property: for every $\{g(x, 1, \cdot)\}_{x \in X}$ of the form (3.2) with $\lambda$ satisfying Assumption 1, we have

$$d'_\alpha t_H(\alpha) \leq t_L \leq d_\alpha t_H(\alpha).$$

We now state the following elementary facts about drift conditions.

Lemma 3.2. Suppose $g$ is a transition kernel satisfying Eq. (1.1) with function $V$ and constants $0 \leq \lambda < 1$ and $0 \leq b < \infty$. Then we have

$$(g_L V)(x) = \int V(y) g_L(x, 1, dy) \leq \frac{1 + \lambda}{2} V(x) + \frac{b}{2}$$

If $g$ is furthermore a Gibbs sampler, then

$$(g^{(C)}_L V)(x) \leq (g^{(C)}_V V)(x), \quad (g^{(C)}_L V)(x) \leq (g_L V)(x)$$

for any set $C$ of the form $C = \{y : V(y) \leq d\}$, any $d > \inf V(y)$ and any $x \in C$.

We now present the main result in this section:

Lemma 3.3. There exist a universal set of constants $D = \{d_\alpha\}_{0 < \alpha < 0.5}$ with the following property: Any Gibbs sampler $g$ that satisfies Assumption 1 and conditions (1), (4) of Definition 2.6 for some set $C$ of the form $C = \{x : V(x) \leq A\}$ in fact satisfies all four conditions of Definition 2.6 with this choice of $D$.

Proof. Condition (2) is immediate from Lemma 3.2. Condition (3) is exactly the content of Theorem 3.1.

Putting together Lemma 3.3 and Lemma 2.8 immediately gives:

Theorem 3.4. Let $g$ be the transition kernel of a Gibbs sampler with targeting distribution $\pi$. Suppose $g$ satisfies Eq. (1.1) and Definition 2.5 with $\alpha$ compatibly for some compact $C \in \mathcal{B}[X]$ with $\pi(C) > 0$. Suppose $\pi^{(C)}$ satisfies Assumption 1. Suppose

$$C \supset \{x \in X : V(x) \leq d\}$$

for some $d' > \frac{2b}{1 - \lambda_1}$, $d > \frac{2b + 24\lambda'}{1 - \lambda_1}$ where $\lambda_1 = \frac{1 + \lambda}{2}$ and $b_1 = \frac{b}{2}$. Then there exists $M : X \mapsto [0, \infty)$ and $0 < \rho \leq 1$ such that

$$|g_L(x, t, \cdot) - \pi(\cdot)||_{TV} \leq M(x)(1 - \rho)^t$$

for all $x \in X$ and $t \in \mathbb{N}$. Moreover, $M$ and $\rho$ depend only on the parameters $\lambda, b, \alpha, C$.

3.2. Application to Metropolis–Hastings Chain. In this section, we give results analogous to Lemma 3.3 for the following class of Metropolis–Hastings chain defined in [ADS19b]:

Definition 3.5 (Metropolis–Hastings Chain). Fix a distribution $\pi$ with continuous density function $\rho$ supported on a compact subset of Euclidean space; without loss of generality we assume that the support of $\pi$ is a subset of $[0, 1]^d$. We fix a reversible kernel $q$ with unique stationary measure $\nu$ whose support contains that of $\pi$. We assume that $\nu$ has continuous density $\phi$ and that, for all $x$ for which it is defined,
q(x, 1, \cdot) has density q_x. Finally, we assume that the mapping (x, y) → q_x(y) is continuous. We define the acceptance function by the formula
\[ \beta(x, y) = \min\{1, \frac{\rho(y)q_y(x)}{\rho(x)q_x(y)}\}. \]  
(3.8)
Define g to be the transition kernel given by the formula
\[ g(x, 1, A) = \int_{y \in A} q_x(y)\beta(x, y)dy + \delta(x, A) \int_{[0,1]^d} q_x(y)(1 - \beta(x, y))dy. \]  
(3.9)
Note that g is defined uniquely by its target \( \rho \) and proposal \( q \).

**Theorem 3.6** ([ADS19b, Cor. 6.5]). Let \( 0 < \alpha < \frac{1}{2} \). Then there exist universal constants \( d_\alpha, d'_\alpha \) with the following property: for every Metropolis–Hastings chain \( g \) of the form (3.9) satisfying conditions in Definition 3.5, we have
\[ d'_\alpha t_H(\alpha) \leq t_L \leq d_\alpha t_H(\alpha). \]  
(3.10)
In the context of the Metropolis–Hastings sampler only, the \( C \)-dominated chain \( g^{(C)} \) will always refer to the chain from Definition 2.3. We note that Lemma 3.2 remains true with “Gibbs” replaced by “Metropolis–Hastings” the one time it is used. We then have the main result of this section:

**Lemma 3.7.** There exist a universal set of constants \( D = \{d_\alpha\}_{0<\alpha<0.5} \) with the following property:
Any Metropolis–Hastings sampler \( g \) of the form (3.9) satisfying conditions in Definition 3.5 and conditions (1), (4) of Definition 2.6 for some set \( C \) of the form \( C = \{x : V(x) \leq A\} \) in fact satisfies all four conditions of Definition 2.6 with this choice of \( D \).

Proof. Condition (2) is immediate from Lemma 3.2\(^1\). Condition (3) is exactly the content of Theorem 3.1.
\[ \square \]
Combining Lemma 2.8 and Lemma 3.7 tells us that Theorem 3.4 holds for Metropolis–Hastings samplers as well as Gibbs samplers (we don’t rewrite the theorem to save space).

4. **Comparison to Minorization and Illustrative Example**

The main goal of this paper was to illustrate how the minorization condition (1.4) in Theorem 1.1 can be replaced by a hitting-time condition. We have not yet seriously addressed the obvious question: why would anyone wish to do this?

We would like to be able to say that using maximum hitting times gives better bounds. However, the equivalence results of [Oli12; PS15; BHP15; ADS19a; ADS19b] say that this nice response can’t be correct. After all, if mixing times and maximum hitting times are equal up to some universal constants, methods based on maximum hitting times must give bounds that are essentially equivalent to those obtained from the pseudo-minorization condition of [RR01].

Instead of obtaining better results, we claim that bounds on the maximum hitting time often have easier proofs. The “difficulty” of a proof is of course subjective, but we give evidence based on the following concrete claims:

\(^1\)As we have just noted, this applies with “Gibbs” replaced by “Metropolis–Hastings.”
Maximum hitting time bounds are never “harder” than pseudo-minorization bounds: Any pseudo-minorization bound can immediately be turned into a maximum hitting time bound (see (4.2)).

In some realistic examples, we can compute maximum hitting bounds but not pseudo-minorization bounds: We have a broad class of simple but realistic examples for which we can compute good bounds on the maximum hitting time, but have no realistic way to directly compute good bounds on the mixing time (see Theorem 4.2).

To see the first, consider any Markov chain with transition kernel \( g \), unique stationary measure \( \pi \) and mixing time \( t_m \). Then for any measurable \( A \) with \( \pi(A) \geq \frac{1}{3} \) and starting point \( x \),

\[
g(x, t_m, A) \geq \pi(A) - \|g(x, t_m, \cdot) - \pi(\cdot)\|_{TV} \geq \frac{1}{12},
\]

which immediately implies

\[
t_H(\frac{1}{3}) \leq 12t_m.
\] (4.2)

Next, we consider a class of examples with state space \([0, 1]\). We begin by defining the following class of nearly-unimodal measures. These are meant to mimic the typical “small set” used in a pseudo-minorization argument - we expect the target distribution to be reasonably close to unimodal, but might not have much detailed control over the size of the mode or fluctuations.

**Definition 4.1.** Fix \( 0 < \alpha < 0.5 \) and \( 1 \leq A < \infty \). For a density \( \rho \) and constant \( 0 < p < 1 \), denote by \( m(\rho, p) \) the \( p \)th quantile of \( \rho \). We say that a density \( \rho \) on \([0, 1]\) is **unimodal** if there exists some \( m \in [0, 1] \) such that

\[
\rho(x) < \rho(y)
\] (4.3)

for all \( 0 \leq x < y \leq m \) and

\[
\rho(x) > \rho(y)
\] (4.4)

for all \( m \leq x < y \leq 1 \). We say that a density \( \psi \) is \((\alpha, A, \rho)\)-nearly unimodal if

1. \( \rho \) is unimodal, and
2. \( A^{-1} \leq \frac{\rho(x)}{\psi(x)} \leq A \) for all \( x \in [0, 1] \), and
3. \( A^{-1} \leq \frac{\psi(x)}{\psi(y)} \leq A \) for all \( x, y \in [m(\psi, \alpha), m(\psi, 1 - \alpha)] \).

We have the following:

**Theorem 4.2.** Fix \( 1 \leq A < \infty \), \( 0 < \alpha < 0.5 \). There exists a universal constant \( L = L(\alpha, A) \) so that the following holds:

Let \( \psi \) be a continuous density that is \((\alpha, A, \rho)\)-nearly unimodal for some \( \rho \). For \( c > 0 \), let \( g_c \) be the transition kernel given in Definition 3.5 with proposal kernel \( q(x, 1, B) = \frac{|B \cap [x-c, x+c]|}{2c} \) and target \( \psi \). Then the maximum hitting time \( t_H(g_c, \alpha) \) of \( g_c \) satisfies

\[
t_H(g_c, \alpha) \leq Le^{-2}
\] (4.5)

for all \( c > C(\psi, \alpha) \) sufficiently large.
Proof: We check that the conclusion holds for a simple discretization of the process, then translate the bound back to our original kernel \( g_c \). For \( c^{-1} \in \mathbb{N} \), we define the transition kernel of a closely-related “birth-and-death” chain \( h_c \) on \([c]\) \(\equiv \{0,c,2c,\ldots,1-2c,1-c\}\) by the formula
\[
h_c(x,1,x+1) = \frac{1}{2} \min \left(1, \frac{\psi(x+1)}{\psi(x)} \right), \quad x < 1
\]
\[
h_c(x,1,x-1) = \frac{1}{2} \min \left(1, \frac{\psi(x-1)}{\psi(x)} \right), \quad x > 0
\]
and \( h_c(x,1,x) = 1 - h_c(x,1,x-1) - h_c(x,1,x+1) \). Since \( h_c \) is a birth-and-death chain, it is reversible with respect to some measure \( \psi_c \).

Next, fix a measurable set \( S \) with \( \int_{x \in S} \psi(x) dx > \alpha \); by the (uniform) continuity of \( \psi \), it is straightforward to check that \( \sum_{x \in S \cap \{c\}} \psi_c(x) > \alpha \) for all \( c > C(S,\psi,\alpha) \) sufficiently large. Since \( \int_{x \in S} \psi(x) dx > \alpha \), there exist a point \( p \in \{0,1,\ldots,\lfloor \frac{1}{c} m(\psi,1-\alpha) \rfloor \} \) so that
\[
\int_{x \in S \cap \{p \ell ,\ell +1\}} 1dx > \frac{c}{8}. \tag{4.6}
\]
Similarly, there exists \( p_r \in \{\lfloor \frac{1}{c} m(\psi,\alpha) \rfloor,\ldots,\lceil \frac{1}{c} \rceil \} \) satisfying (4.6). Denote by \( p'_r(c) \), \( p'_r(c) \) the closest points in \([c]\) to \( \frac{1}{c} p_r(c) \) and \( \frac{1}{c} p_r(c) \), respectively (break ties arbitrarily).

Denote by \( \tau(h,S) \) the hitting time of a set \( S \) for a transition kernel \( h \), extending the notion in Equation (1.5). Since birth-and-death chains can’t “skip” any values,
\[
\max_{x \in [c]} E_x[\tau(h_c,\{p'_r(c),p'_r(c)\})] \leq \max_{s \in \{\alpha,1-\alpha\}} E_x[\tau(h_c,\{m(\psi,s)\})] = t_H(h_c,\alpha). \tag{4.7}
\]

The expectations appearing in the middle expression have simple explicit formulas (see e.g. [PT96]). Comparing the explicit formula for \( h_c \) to the explicit formula for the simple random walk \( w_c \) on \([0,c,2c,\ldots,1-c,1]\), we see that all terms in the explicit formula differ by at most a multiplicative factor depending only on \( \alpha,c \). Thus, there exists a universal constant \( L_1 = L_1(\alpha, A) \) such that
\[
t_H(h_c,\alpha) \leq L_1 t_H(w_c,\alpha). \tag{4.8}
\]
Since \( t_H(w_c,\alpha) \leq c^{-2} \) (see e.g. Example 10.20 of [LPW09]), we have shown that
\[
\max_{x \in [c]} E_x[\tau(h_c,\{p'_r,c),p'_r(c)\})] \leq L_1 c^{-2}. \tag{4.9}
\]

Define \( I(c) = [p'_r(c) - \frac{1}{c} c, p'_r(c) + \frac{1}{c} c] \cup [p'_r(c) - \frac{1}{c} c, p'_r(c) + \frac{3}{c} c] \). By (4.6), there exists a universal constant \( L_2 = L_2(\alpha, A) \) such that
\[
\sup_{x} E_x[\tau(g_c,S)] \leq L_2 \sup_{x} E_x[\tau(g_c, I(c))]. \tag{4.10}
\]

We now relate this to \( h_c \). Fix some \( B > 0 \) and consider sample paths \( \{X_t^{(c)}\}_{t=0}^{Bc^{-2}} \sim g_c \), \( \{Y_t^{(c)}\}_{t=0}^{Bc^{-2}} \sim h_c \). Applying standard convergence results (e.g. Theorem 2.5 of [EK09]), one can quickly confirm that both of these processes converge to the same limiting continuous-time process in Skorohod’s topology on \( \mathcal{D}[0,B] \) as \( c \to 0 \) and time is rescaled in the usual way, as \( t \to c^{-2} t \). Since \( g_c \) can’t jump more than
Combining this with Inequalities (4.9) and (4.10), we have

\[
\sup_x \mathbb{E}_x [\tau(g_c, S)] \leq L_2 \sup_x \mathbb{E}_x [\tau(g_c, I(c))] \leq L_2 L_3 \max_{x \in [c]} \mathbb{E}_x [\tau(h_c, \{p'_e(c), p'_r(c)\})] \leq L_1 L_2 L_3 c^{-2}.
\]

Since this holds for all measurable \( S \), this completes the proof. \( \square \)

When \( \psi(x) \equiv 1 \), the kernel analyzed above is exactly the usual “ball” walk on \([0, 1]\), so this upper bound is in fact sharp. While our analysis is fairly simple, we know of no simple way to get comparable results by directly analyzing the mixing time. See further discussion in \( \textit{e.g.} \) [JS18; Yue00], where the methods studied cannot give bounds better than \( t_m(g_c) = O(c^{-3}) \) for any walk in this class.

**References**

[ADS19a] R. M. Anderson, H. Duanmu, and A. Smith. “Mixing Times and Hitting Times for General Markov Processes” (2019). Submitted.

[ADS19b] R. M. Anderson, H. Duanmu, and A. Smith. “Mixing Times and Hitting Times for Gibbs Samplers and Other Non-Feller Processes” (2019). Submitted.

[BBM11] N. Bhatnagar, A. Bogdanov, and E. Mossel. “The computational complexity of estimating MCMC convergence time”. In: Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques. Springer, 2011, pp. 424–435.

[BHP15] R. Basu, J. Hermon, and Y. Peres. “Characterization of cutoff for reversible Markov chains”. In: Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms. Society for Industrial and Applied Mathematics. 2015, pp. 1774–1791.

[BL10] J. Beltran and C. Landim. “Tunneling and metastability of continuous time Markov chains”. Journal of Statistical Physics 140.6 (2010), pp. 1065–1114.

[Dia09] P. Diaconis. “The Markov chain Monte Carlo revolution”. Bulletin American Math. Soc. 46.2 (2009), pp. 179–205.

[DKSC08] P. Diaconis, K. Khare, and L. Saloff-Coste. “Gibbs sampling, exponential families and orthogonal polynomials”. Stat. Science 23.2 (2008), pp. 151–178.

[DRW18] H. Duanmu, J. Rosenthal, and W. Weiss. “Ergodicity of Markov processes via non-standard analysis”. Memoirs of the American Mathematical Society, to appear. 2018.

[EK09] S. N. Ethier and T. G. Kurtz. Markov processes: characterization and convergence. Vol. 282. John Wiley & Sons, 2009.

[GR92] A. Gelman and D. B. Rubin. “Inference from iterative simulation using multiple sequences”. Stat. Science 7.4 (1992), pp. 457–472.
REFERENCES

[GS90] A. E. Gelfand and A. F. Smith. “Sampling-based approaches to calculating marginal densities”. *JASA* 85.410 (1990), pp. 398–409.

[Hsu+19] D. Hsu, A. Kontorovich, D. A. Levin, Y. Peres, C. Szepesvári, and G. Wolfer. “Mixing time estimation in reversible Markov chains from a single sample path”. *Annals of Applied Probability* 29.4 (2019), pp. 2439–2480.

[Hub16] M. L. Huber. *Perfect simulation*. Chapman and Hall/CRC, 2016.

[JH01] G. L. Jones and J. P. Hobert. “Honest exploration of intractable probability distributions via Markov chain Monte Carlo”. *Stat. Science* (2001), pp. 312–334.

[JS18] J. Johndrow and A. Smith. “Fast mixing of Metropolis-Hastings with unimodal targets”. *Electronic Communications in Probability* 23 (2018).

[LPW09] D. A. Levin, Y. Peres, and E. L. Wilmer. *Markov chains and mixing times*. With a chapter by James G. Propp and David B. Wilson. American Mathematical Society, Providence, RI, 2009, pp. xviii+371. ISBN: 978-0-8218-4739-8.

[MT12] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.

[MT94] S. P. Meyn and R. L. Tweedie. “Computable bounds for geometric convergence rates of Markov chains”. *Annals of Applied Probability* 4.4 (1994), pp. 981–1011.

[Oli12] R. Oliveira. “Mixing and hitting times for finite Markov chains”. *Electronic Journal of Probability* 17 (2012).

[PS15] Y. Peres and P. Sousi. “Mixing times are hitting times of large sets”. *Journal of Theoretical Probability* 28.2 (2015), pp. 488–519. ISSN: 0894-9840. DOI: 10.1007/s10959-013-0497-9.

[PT96] J. Palacios and P. Tetali. “A note on expected hitting times for birth and death chains”. *Statistics & Probability Letters* 30.2 (1996), pp. 119–125.

[Ros95] J. S. Rosenthal. “Minorization conditions and convergence rates for Markov chain Monte Carlo”. *JASA* 90.430 (1995), pp. 558–566.

[RR01] G. O. Roberts and J. S. Rosenthal. “Small and pseudo-small sets for Markov chains”. *Stochastic Models* 17.2 (2001), pp. 121–145.

[Yue00] W. K. Yuen. “Applications of geometric bounds to the convergence rate of Markov chains on $\mathbb{R}^n$”. *Stochastic processes and their applications* 87.1 (2000), pp. 1–23.

University of California, Berkeley, Department of Economics

University of California, Berkeley, Department of Economics

University of Ottawa, Department of Mathematics and Statistics

University of Toronto, Department of Statistical Sciences