

Stability of Adversarial Markov Chains, with an Application to Adaptive MCMC Algorithms

by (in alphabetical order)

Radu V. Craiu¹, Lawrence Gray², Krzysztof Łatuszyński³, Neal Madras⁴,

Gareth O. Roberts³, and Jeffrey S. Rosenthal^{1,*}

(March, 2014)

Abstract. We consider whether ergodic Markov chains with bounded step size remain bounded in probability when their transitions are modified by an adversary on a bounded subset. We provide counterexamples to show that the answer is no in general, and prove theorems to show that the answer is yes under various additional assumptions. We then use our results to prove convergence of various adaptive Markov chain Monte Carlo algorithms.

1. Introduction.

This paper considers whether bounded modifications of stable Markov chains remain bounded in probability. Specifically, we let P be a fixed time-homogeneous ergodic Markov chain kernel with bounded step size, and let $\{X_n\}$ be a stochastic process which follows the transitions P except on a bounded subset K where an “adversary” can make arbitrary bounded jumps. Under what conditions must such a process $\{X_n\}$ be bounded in probability?

One might think that such boundedness would follow easily, at least under mild regularity and continuity assumptions, i.e. that modifying a stable continuous Markov chain inside a bounded set K couldn’t possibly lead to unstable behaviour out in the tails. In fact the situation is rather more subtle, as we explore herein. We will provide counterexamples to show that boundedness may fail even for well-behaved continuous chains. We will then show

¹Dept. of Statistics, University of Toronto, 100 St. George Street, Toronto, Ontario, M5S 3G3, Canada.

²Dept. of Mathematics, University of Minnesota, 206 Church St. SE, Minneapolis, MN 55455-0488, U.S.A.

³Dept. of Statistics, University of Warwick, Coventry, CV4 7AL, United Kingdom

⁴Dept. of Mathematics & Statistics, York University, 4700 Keele St., Toronto, Ontario, M3J 1P3, Canada.

*Corresponding author. E-mail: jeff@math.toronto.edu

that under various additional conditions, including bounds on transition probabilities and/or small set assumptions and/or geometric ergodicity, such boundedness does hold.

The specific question considered here appears to be new, though it is somewhat reminiscent of previous bounds on non-Markovian stochastic processes such as those related to *adversarial queueing theory* [11, 19, 6]. We present our formal setup in Section 2, our main results in Section 3, and some counterexamples in Section 4. Our results are then proven in Sections 5 through 10. In Section 11 we describe how our results can be applied to adaptive MCMC algorithms, to verify their convergence more easily than previous approaches.

2. Formal setup and assumptions.

Let $(\mathcal{X}, \mathcal{F})$ be a metricised measure space, and let P be the transition probability kernel for a fixed time-homogeneous Markov chain on \mathcal{X} . Assume that P is Harris ergodic with stationary probability distribution π , so that

$$\lim_{n \rightarrow \infty} \|P^n(x, \cdot) - \pi\| := \lim_{n \rightarrow \infty} \sup_{A \in \mathcal{F}} |P^n(x, A) - \pi(A)| = 0, \quad x \in \mathcal{X}. \quad (1)$$

We assume, to relate the Markov chain to the geometry of \mathcal{X} , that there is a constant $D < \infty$ such that P never moves more than a distance D , i.e. such that

$$P(x, \{y \in \mathcal{X} : |y - x| \leq D\}) = 1, \quad x \in \mathcal{X}. \quad (2)$$

Let $K \in \mathcal{F}$ be a fixed *bounded* non-empty subset of \mathcal{X} , and for $r > 0$ let K_r be the set of all states within a distance r of K (so each K_r is also bounded).

In terms of these ingredients, we define our “adversarial Markov chain” process $\{X_n\}$ as follows. It begins with $X_0 = x_0$ for some specific initial state $x_0 \in K$. Whenever it is outside of K , it moves according to the Markov transition probabilities P , i.e.

$$\mathbf{P}(X_{n+1} \in A \mid X_0, X_1, \dots, X_n) = P(X_n, A), \quad n \geq 0, \quad A \in \mathcal{F}, \quad X_n \notin K, \quad (3)$$

When the process is inside of K , it can move *arbitrarily*, according to an adversary’s wishes, perhaps depending on the time n and/or the chain’s history in a non-anticipatory manner

(i.e., adapted to $\{X_n\}$; see also Example #3 below), subject only to measurability (i.e., $\mathbf{P}(X_{n+1} \in A \mid X_0, X_1, \dots, X_n)$ must be well-defined for all $n \geq 0$ and $A \in \mathcal{F}$), and to the restriction that it can't move more than a distance D at each iteration – or more specifically that from K it can only move to points within K_D . In summary, $\{X_n\}$ is a stochastic process which is “mostly” a Markov chain following the transition probabilities P , except that it is modified by an adversary when it is within the bounded subset K .

We are interested in conditions guaranteeing that this process $\{X_n\}$ will be bounded in probability, i.e. will be tight, i.e. will satisfy that

$$\lim_{L \rightarrow \infty} \sup_{n \in \mathbf{N}} \mathbf{P}(|X_n| > L \mid X_0 = x_0) = 0. \quad (4)$$

3. Main Results.

We now consider various conditions under which (4) will or will not hold. For application of our results to the verification of adaptive MCMC algorithms, see Section 11 below.

3.1. First results.

We first note that such boundedness is guaranteed in the *absence* of an adversary:

Proposition 1. *In the setup of Section 2, suppose $\{X_n\}$ always follows the transitions P (including when it is within K , i.e. there is no adversary). Then (4) holds.*

Proposition 1 is proved in Section 3.5 below.

Boundedness also holds for a lattice like \mathbf{Z}^d , or more generally if the state space \mathcal{X} is topologically discrete (i.e. countable and such that each state x is topologically isolated and hence open in \mathcal{X}). In this case, bounded subsets like K_{2D} must be *finite*, and the result holds without any further assumptions:

Proposition 2. *In the setup of Section 2, suppose P is an irreducible positive-recurrent Markov chain with stationary probability distribution π on a countable state space \mathcal{X} such*

that K_{2D} is finite. Then (4) holds.

Proposition 2 is proved in Section 5 below.

However, (4) does not hold in general, not even under a strong continuity assumption:

Proposition 3. *There exist adversarial Markov chain examples following the setup of Section 2, on state spaces which are countable subsets of \mathbf{R}^2 , which fail to satisfy (4), even under the strong continuity condition that \mathcal{X} is closed and*

$$\forall x \in \mathcal{X}, \forall \epsilon > 0, \exists \delta > 0 \text{ s.t. } \|P(y, \cdot) - P(x, \cdot)\| < \epsilon \text{ whenever } |y - x| < \delta. \quad (5)$$

Proposition 3 is proved in Section 4 below, using two different counterexamples.

Proposition 3 says that the adversarial process $\{X_n\}$ may not be bounded in probability, even if we assume a strong continuity condition on P . Hence, additional assumptions are required, as we consider next.

Remark. The counterexamples in Proposition 3 are discrete Markov chains in the sense that their state spaces are countable. However, their state spaces \mathcal{X} are not *topologically* discrete, since they contain accumulation points, and in particular sets like K_{2D} are not finite there, so there is no contradiction with Proposition 2.

3.2. A result using expected hitting times.

We now consider two new assumptions. The first provides an upper bound on the Markov chain transitions out of K_D :

- (A1) There is $M < \infty$, and a probability measure μ_* concentrated on $K_{2D} \setminus K_D$, such that $P(x, dz) \leq M \mu_*(dz)$ for all $x \in K_D \setminus K$ and $z \in K_{2D} \setminus K_D$.

Note that in (A1) we always have $z \neq x$, which is helpful when considering e.g. Metropolis algorithms which have positive probability of not moving. Choices of μ_* in (A1) might include $\text{Uniform}(K_{2D} \setminus K_D)$, or $\pi|_{K_{2D} \setminus K_D}$. The second assumption bounds an expected hitting time:

(A2) The expected time for a Markov chain following the transitions P to reach the subset K_D , when started from the distribution μ_* in (A1), is finite.

In terms of these two assumptions, we have:

Theorem 4. *In the setup of Section 2, if (A1) and (A2) hold for the same μ_* , then (4) holds, i.e. $\{X_n\}$ is bounded in probability.*

Theorem 4 is proved in Section 5 below.

3.3. A result assuming a small set condition.

The condition (A2), that the hitting time of K_D has finite expectation, may be difficult to verify directly. As an alternative, we consider a different assumption:

(A3) The set $K_{2D} \setminus K_D$ is small for P , i.e. there is some probability measure ν_* on \mathcal{X} , and some $\epsilon > 0$, and some $n_0 \in \mathbf{N}$, such that $P^{n_0}(x, A) \geq \epsilon \nu_*(A)$ for all states $x \in K_{2D} \setminus K_D$ and all subsets $A \in \mathcal{F}$.

We then have:

Theorem 5. *In the setup of Section 2, if (A1) and (A3) hold where either (a) $\nu_* = \mu_*$, or (b) P is reversible and $\mu_* = \pi|_{K_{2D} \setminus K_D}$, then (4) holds, i.e. $\{X_n\}$ is bounded in probability.*

Theorem 5 is proved in Section 7 below.

Assumption (A3) is often straightforward to verify. For example:

Proposition 6. *Suppose \mathcal{X} is an open subset of \mathbf{R}^d which contains a rectangle J which contains $K_{2D} \setminus K_D$. Suppose there are $\delta > 0$ and $\epsilon > 0$ such that*

$$P(x, dy) \geq \epsilon \text{Leb}(dy) \quad \text{whenever } x, y \in J \quad \text{with } |y - x| < \delta, \quad (6)$$

where Leb is Lebesgue measure on \mathbf{R}^d . Then (A3) holds with $\nu_* = \text{Uniform}(K_{2D} \setminus K_D)$.

Proposition 6 is proved in Section 8 below.

3.4. A result assuming geometrically ergodicity.

Assumption (A3) can be verified for various Markov chains, as we will see below. However, its verification will sometimes be difficult. An alternative approach is to consider *geometric ergodicity*, as follows (see e.g. [15] for context):

- (A4) The Markov chain transition kernel P is geometrically ergodic, i.e. there is $\rho < 1$ and $c < \infty$ and a π -a.e. finite function $V : \mathcal{X} \rightarrow [1, \infty]$ such that $\|P^n(x, \cdot) - \pi\| \leq c V(x) \rho^n$ for $n \in \mathbf{N}$ and $x \in \mathcal{X}$.

We also require a slightly different version of (A1):

- (A5) There is $M < \infty$ such that $P(x, dz) \leq M \pi(dz)$ for all $x \in K_D$ and $z \in K_{2D}$.

(Of course, (A5) holds trivially for $z \notin K_{2D}$, since then $P(x, dz) = 0$.) We then have:

Theorem 7. *In the setup of Section 2, if (A4) and (A5) hold, then (4) holds, i.e. $\{X_n\}$ is bounded in probability.*

Theorem 7 is proved in Section 10 below.

3.5. A short proof of Proposition 1.

We end this section with the (easy) proof of Proposition 1.

Let $x_0 \in K$ be the initial state, let $B_L = \{x \in \mathcal{X} : |x| \geq L\}$ and let $\epsilon > 0$. Then by equation (1), we can find N large enough that $\|P^n(x_0, \cdot) - \pi\| < \epsilon/2$ for all $n \geq N$. Then choose L large enough that $\pi(B_L) < \epsilon/2$ and also $L > |x_0| + ND$. For $n < N$, it follows from equation (2) that $P^n(x_0, B_L) = 0$. For $n \geq N$, we have $P^n(x_0, B_L) \leq \pi(B_L) + \|P^n(x_0, \cdot) - \pi\| < \epsilon/2 + \epsilon/2 = \epsilon$. Hence, $\sup_n P^n(x_0, B_L) \leq \epsilon$, i.e. $\sup_n \mathbf{P}(X_n \in B_L | X_0 = x_0) \leq \epsilon$. Since $\epsilon > 0$ was arbitrary, (4) follows.

(Alternatively, if P is Harris ergodic as in (1), then it converges in distribution, so it must be tight and hence satisfy (4). In fact, even if P is just assumed to be ϕ -irreducible with period $d \geq 1$ and stationary probability distribution π , then this argument can apply

separately to each of the sequences $\{X_{dn+j}\}_{n=0}^{\infty}$ for $j = 0, 1, \dots, d-1$ to again conclude (4.) ■

4. Counterexamples to prove Proposition 3.

We next present two counterexamples to illustrate that with the setup and assumptions of Section 2, the bounded in probability property (4) might fail.

Example #1. Let $\mathcal{X} = \{(1/i, j) : i \in \mathbf{N}, j = 0, 1, \dots\}$ be the state space. That is, $\mathcal{X} = \bigcup_{i \in \mathbf{N}} \mathcal{X}_i$ where each $\mathcal{X}_i \equiv \{(1/i, j)\}_{j=0,1,\dots}$ is a different “column”. Let $\pi(1/i, j) = 2^{-i} (1/i) (1 - 1/i)^j$, so that π restricted to each \mathcal{X}_i is a geometric distribution with mean i . Let $K = \{(1/i, 0)\}$ consist of the bottom element of each column (see Figure 1).

Let the Markov chain P proceed, outside of K , by doing a simple ± 1 Metropolis algorithm up and down its current column \mathcal{X}_i in a manner which is reversible with respect to π . That is, for $j \geq 1$, $P((1/i, j), (1/i, j-1)) = 1/2$, and $P((1/i, j), (1/i, j+1)) = (1 - 1/i)/2$, and $P((1/i, j), (1/i, j)) = 1 - P((1/i, j), (1/i, j-1)) - P((1/i, j), (1/i, j+1))$. Intuitively, the larger the column number i , the higher is the conditional mean of π on \mathcal{X}_i , so the higher the chain will tend to move within \mathcal{X}_i , and the longer it will take to return to K .

Inside of K , choose any appropriate transitions to make the chain irreducible and reversible with respect to π , e.g. choose $P((1/i, 0), (1/(i+1), 0)) = i/8(i+1)$, $P((1/i, 0), (1/(i-1), 0)) = 1/4$, $P((1/i, 0), (1/i, 1)) = 1/2i$, and $P((1/i, 0), (1/i, 0)) = 1 - P((1/i, 0), (1/(i+1), 0)) - P((1/i, 0), (1/(i-1), 0)) - P((1/i, 0), (1/i, 1))$.

Let the adversary proceed within K as follows. If $X_n \in K$, then $X_{n+1} = (1/n, 1)$. That is, the chain moves from K to higher and higher column numbers as time goes on.

With these specifications, K is bounded, and the process $\{X_n\}$ never moves more than a distance $D = 1$, to the setup of Section 2 is satisfied. However, the process $\{X_n\}$ will, over time, move to higher and higher column numbers, and will then tend to climb higher and higher up the columns. So, it will not be bounded in probability, i.e. (4) will not hold.

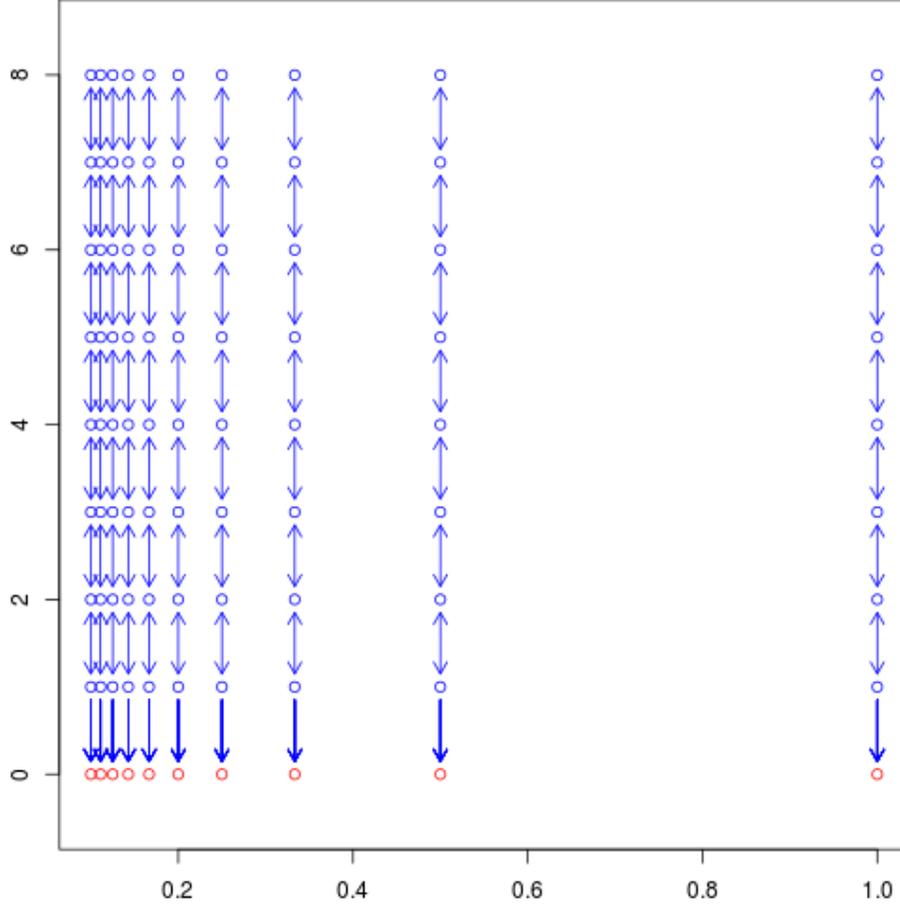


Figure 1. Part of the state space in Example #1.

(Alternatively, the adversary could proceed within K by moving from $(1/i, 0)$ to either $(1/i, 1)$ with probability $(1 - 1/i)/2$, or to $(1/(i - 1), 0)$ or $(1/(i + 1), 0)$ with equal probability $(1 + 1/i)/4$, to make the process $\{X_n\}$ be time-homogeneous Markov and reversible with respect to the infinite measure $\bar{\pi}$ defined by $\bar{\pi}(1/i, j) = (1/i)(1 - 1/i)^j$. Then $\{X_n\}$ will be null recurrent, and hence again (4) will not hold.) ■

Now, in the above Example, the state space \mathcal{X} is not closed. One could easily “extend” the example to include $\{(0, j) : j \in \mathbf{N}\}$ and thus make \mathcal{X} closed. However, this cannot be done in a continuous way, i.e. there is no way to satisfy (5) in this example. This might lead one to suspect that a continuity condition such as (5) suffices to guarantee (4). However,

this is not the case, as the following example shows:

Example #2. Our state space \mathcal{X} will be another countable subset of \mathbf{R}^2 , defined as follows. Let $O = (0, 0)$ be the origin. Let $S_0 = \{(i, 0) : i \in \mathbf{N}\}$. Let $\{\beta_k\}_{k=1}^\infty$ be an increasing sequence of integers with $\beta_k > k$ to be specified later. For $k \in \mathbf{N}$, let S_k consist of the k points $(i, i/k)$ for $i = 1, 2, \dots, k$, together with $\beta_k - 1$ additional points equally spaced on the line segment from $(k, 1)$ to the y -axis point $(0, \beta_k)$. Finally, let $\mathcal{Y} = \{(0, i) : i \in \mathbf{N}\}$ be the positive integer y -axis. Then $\mathcal{X} = O \cup \mathcal{Y} \cup \bigcup_{k=0}^\infty S_k$ (see Figure 2).

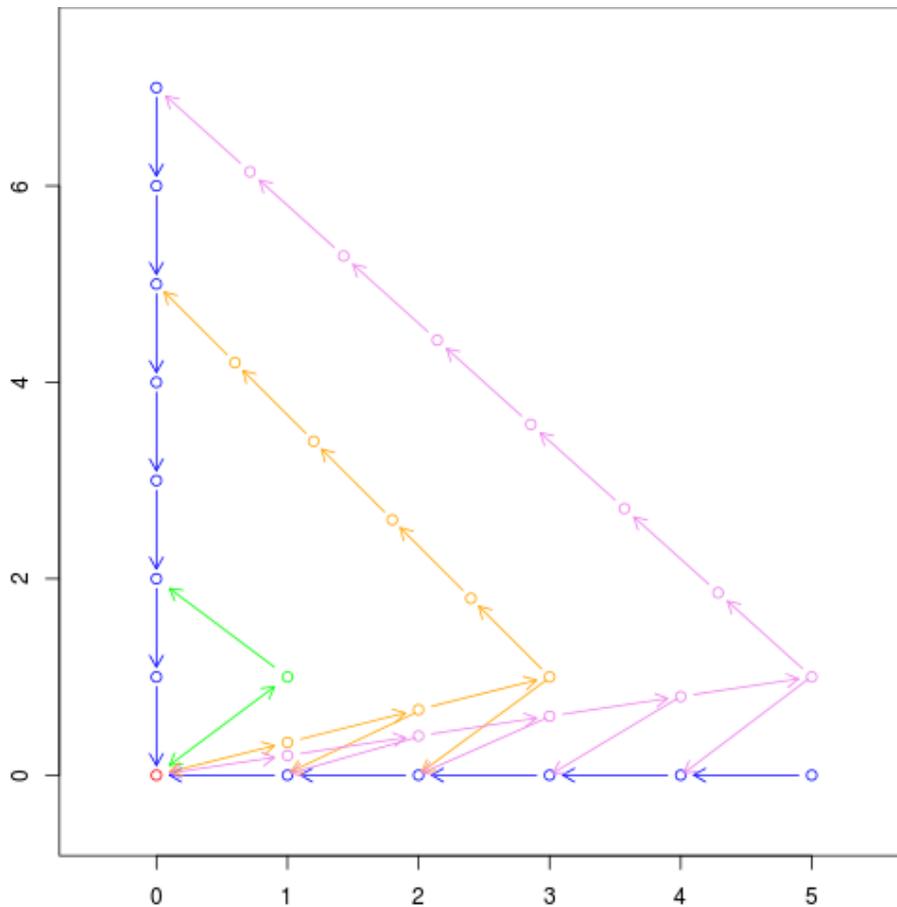


Figure 2. Part of the state space in Example #2.

Define transitions P on \mathcal{X} as follows. On S_0 , we have $P((i, 0), (i - 1, 0)) = 1$, i.e. it always moves towards the origin. Similarly, on \mathcal{Y} , we have $P((0, i), (0, i - 1)) = 1$, i.e. it again always

moves towards the origin. On the first $k - 1$ points of S_k , we have $P((i, i/k), (i + 1, (i + 1)/k)) = i/k$, and $P((i, i/k), (i - 1, 0)) = 1 - i/k$, i.e. it either continues upwards on S_k , or moves towards the origin on S_0 . On the remaining points of S_k , with probability 1 it moves one additional point along S_k 's path towards $(0, \beta_k)$.

Note that these transition probabilities are continuous in a very strong sense: if $x_n \rightarrow x$ (which can only happen for $x \in S_0$), then $P(x_n, y) \rightarrow P(x, y)$ for all $y \in \mathcal{X}$, and in particular $\|P(x_n, \cdot) - P(x, \cdot)\| \rightarrow 0$.

Note also that if this chain is started at $(1, 1/k)$, then it has probability $\prod_{i=1}^k (i/k) > 0$ of continuing along S_k all the way to $(k, 1)$, in which case it will take a total of $k + 2\beta_k$ iterations to return to O . Otherwise, for $1 \leq j \leq k - 1$, it takes $2j - 1$ iterations with probability $(\prod_{i=1}^{j-1} (i/k))(1 - (j/k))$. Thus, if $r_k = \mathbf{E}(\tau_O | X_0 = (1, 1/k))$ is the expected return time to O from $(1, 1/k)$, then

$$r_k = (k + 2\beta_k) \left(\prod_{i=1}^k (i/k) \right) + \sum_{j=1}^{k-1} (2j - 1) \left(\prod_{i=1}^{j-1} (i/k) \right) (1 - (j/k)).$$

In particular, by letting β_k grow sufficiently quickly, we can make r_k grow as quickly as desired.

Finally, we specify that from O , for $k \in \mathbf{N}$ the Markov chain moves to $(1, 1/k)$ with probability a_k , for some positive numbers a_k summing to 1 to be specified later.

Meanwhile, the adversary's compact set is given by the single state $K = \{O\}$. From O , the adversary proceeds simply by moving to each $(1, 1/k)$ with probability b_k , where the b_k are non-negative and sum to 1, and will be specified later. (Thus, the adversary's actions are chosen to still be time-homogeneous Markov.)

To complete the construction, we choose $\{\beta_k\}$ and $\{a_k\}$ and $\{b_k\}$ so that $\sum_k a_k r_k < \infty$ but $\sum_k b_k r_k = \infty$. For example, we can do this by first choosing β_k so that $r_k k^{-k} \rightarrow 1$, and then letting $a_k \propto (2k)^{-k}$ and $b_k \propto (k/2)^{-k}$.

It then follows that for the Markov chain P (governed by the $\{a_k\}$) the expected return time to O from O is finite, and hence the chain has a unique stationary probability measure

π . On the other hand, for the adversarial process $\{X_n\}$ (governed by the $\{b_k\}$) the expected return time to O from O is infinite. Hence, the adversarial process is null recurrent, so it will move to larger and larger S_k as time progresses. In particular, the adversarial process will *not* be bounded in probability, even though the transition probabilities P are continuous. ■

Remark. Example #2 is only defined on a countable state space \mathcal{X} , but if desired it could be “extended” to a counterexample on all of \mathbf{R}^2 , by e.g. convoluting $\pi(y)$ and $P(x, y)$ with a sufficiently tiny bivariate normal distribution at each $y \in \mathcal{X}$, and continuously interpolating the transition probabilities $P(x, A)$ at other $x \in \mathbf{R}^2$. In this way, (5) would still be satisfied, and (4) would still fail, even on the continuous state space \mathbf{R}^2 .

Finally, in a rather different direction, we consider what happens if the adversary is allowed to be anticipatory, i.e. to make moves based on future randomness, with (3) replaced by the weaker condition that $\mathbf{P}(X_{n+1} \in A | X_n = x) = P(x, A)$ but without conditioning on the previous history X_0, \dots, X_{n-1} . (Intuitively, the adversary is allowed to be “clairvoyant”.) It turns out that under this subtle change our theorems no longer hold:

Example #3. Let $\mathcal{X} = [0, \infty) \subseteq \mathbf{R}$. Define Markov chain transitions P as follows. For $x \leq 1$, $P(x, \cdot) = \text{Uniform}[0, 2]$. For $1 < x \leq 3$, $P(x, \cdot) = \text{Uniform}[x - 1, x + 1]$. For $3 < x \leq 4$, $P(x, \cdot) = \text{Uniform}[4, 5]$. For $x > 4$, $P(x, \cdot) = \frac{1}{2} \delta_{x+1}(\cdot) + \frac{1}{2} \text{Uniform}[x - 2, x - 1]$, where δ_{x+1} is a point-mass at $x + 1$. Then P is ϕ -irreducible, with negative drift for $x > 4$, so P must be positive recurrent with some stationary probability distribution π to which it converges as in (1). Also, P never moves more than a distance $D = 2$ as in (2).

We next define the adversarial chain $\{X_n\}$. Let $K = [0, 2]$, so $K_D = [0, 4]$ and $K_{2D} = [0, 6]$. Let $\{B_i\}_{i=0}^\infty$ be iid with $\mathbf{P}(B_i = 0) = \mathbf{P}(B_i = 1) = 1/2$, and let $\{U_i\}_{i=0}^\infty$ be iid $\sim \text{Uniform}[0, 1]$, and let $a_* = 4 + \sum_{i=1}^\infty B_i 2^{-i}$. For any $r \in \mathcal{X}$, let $r[i]$ be the coefficient of 2^i in the non-terminating binary expansion of r , so that $r = \sum_{i \in \mathbf{Z}} r[i] 2^i$. Conditional on X_n , we construct X_{n+1} by: (a) if $X_n \leq 1$ then $X_{n+1} = 2U_n$; (b) if $1 < X_n \leq 3$

then $X_{n+1} = X_n - 1 + 2U_n$; (c) if $3 < X_n \leq 4$ then $X_{n+1} = a_*$; (d) if $X_n > 4$ then $X_{n+1} = I_n(X_n + 1) + (1 - I_n)(X_n - 1 - U_n)$, where $I_n = \mathbf{1}_{X_n[-n]=B_n}$ is the indicator function of whether the coefficient of 2^{-n} in the binary expansion of X_n is equal to B_n .

Then it is easily checked that $\{X_n\}$ follows the one-step transitions P for all $x \in \mathcal{X}$ (including $x \in K$), in the sense that $\mathbf{P}(X_{n+1} \in A | X_n = x) = P(x, A)$ for all A (but without also conditioning on X_0, \dots, X_{n-1}). Furthermore, (A1) holds with $M = 1$ and $\mu_* = \text{Uniform}[4, 5]$. Also, (A2) holds for the same μ_* due to P 's negative drift for $x > 4$.

On the other hand, by construction a_* has the property that $a_*[-n] = B_n$ for all $n \in \mathbf{N}$. Hence, once the chain hits the interval $(3, 4]$, then it will move to a_* , and from there it will always add 1 with probability 1. Therefore, $X_n \rightarrow \infty$ with probability 1, so $\{X_n\}$ is not bounded in probability, so (4) does not hold. This process thus provides a counterexample to Theorem 4 if we assume only that $\mathbf{P}(X_{n+1} \in A | X_n = x) = P(x, A)$, without also conditioning on the previous history X_0, \dots, X_{n-1} as in (3). ■

5. Proof of Theorem 4 and Proposition 2.

We begin by letting $\{Y_n\}$ be a ‘‘cemetery process’’ which begins in the distribution μ_* at time 0, and then follows the fixed transition kernel P , and then *dies* as soon as it hits K_D . Assumption (A2) then says that this cemetery process $\{Y_n\}$ has finite expected lifetime. For $L > \ell_0 := \sup\{|x| : x \in K_D\}$, let $B_L = \{x \in \mathcal{X} : |x| \geq L\}$, and let N_L denote the cemetery process’s occupation time of B_L (i.e., the number of iterations that $\{Y_n\}$ spends in B_L before it dies). We then have:

Lemma 8. *Let $\{X_n\}$ be the adversarial chain as defined previously. Then assuming (A1), for any $n \in \mathbf{N}$, and any $L > \ell_0$, and any $x \in K$, we have*

$$\mathbf{P}(X_n \in B_L | X_0 = x) \leq M \mathbf{E}(N_L),$$

where N_L is the occupation time of B_L for the cemetery process $\{Y_n\}$ defined above.

Proof. Let σ be the last return time of $\{X_n\}$ to K_D by time n (which must exist since $X_0 \in K_D$), and let μ_k be the (complicated) law of X_k when starting from $X_0 = x_0$. Then letting $I = K_D \setminus K$ (“inside”) and $O = K_{2D} \setminus K_D$ (“outside”), we have

$$\begin{aligned}
\mathbf{P}(X_n \in B_L | X_0 = x_0) &= \sum_{k=0}^{n-1} \mathbf{P}(X_n \in B_L, \sigma = k | X_0 = x_0) \\
&= \sum_{k=0}^{n-1} \int_{y \in I} \int_{z \in O} \mathbf{P}(X_k \in dy, X_{k+1} \in dz, X_n \in B_L, \sigma = k | X_0 = x_0) \\
&= \sum_{k=0}^{n-1} \int_{y \in I} \int_{z \in O} \mu_k(dy) P(y, dz) \mathbf{P}(X_n \in B_L, \sigma = k | X_0 = x_0, X_k = y, X_{k+1} = z) \\
&\leq \sum_{k=0}^{n-1} \int_{y \in I} \int_{z \in O} \mu_k(dy) M \mu_*(dz) \mathbf{P}(X_n \in B_L, \sigma = k | X_0 = x_0, X_k = y, X_{k+1} = z) \\
&\leq \sum_{k=0}^{n-1} \int_{y \in I} \int_{z \in O} \mu_k(dy) M \mu_*(dz) \mathbf{P}(Y_{n-k-1} \in B_L | Y_0 = z)
\end{aligned}$$

[by letting $Y_n = X_{n+k+1}$, and noting that if $\sigma = k$ then the process did not return to K_D by time n so it behaved like the cemetery process between times $n - k - 1$ and n]

$$\begin{aligned}
&\leq M \sum_{k=0}^{n-1} \int_{z \in O} \mathbf{P}(Y_{n-k-1} \in B_L | Y_0 = z) \mu_*(dz) \\
&\leq M \sum_{j=0}^{\infty} \int_{z \in O} \mathbf{P}(Y_j \in B_L | Y_0 = z) \mu_*(dz).
\end{aligned}$$

But this last sum is precisely the expected total number of times that the cemetery process $\{Y_n\}$ is in B_L when started from the distribution μ_* . ■

Proof of Theorem 4. For each $A \in \mathcal{F}$, let $\nu(A)$ be its expected occupation measure, i.e. the expected number of times the cemetery process $\{Y_n\}$ is in the subset A . Then the total measure $\nu(\mathcal{X})$ equals the expected lifetime of the cemetery process, and is thus finite by (A2). Hence, by the usual Continuity of Measures,

$$\lim_{L \rightarrow \infty} \nu(B_L) = \nu\left(\bigcap_L B_L\right) = \nu(\emptyset) = 0.$$

This shows that $\mathbf{E}(N_L) \rightarrow 0$ as $L \rightarrow \infty$. Hence, by Lemma 8, $\lim_{L \rightarrow \infty} \mathbf{P}(X_n \in B_L | X_0 = x_0) \leq M \lim_{L \rightarrow \infty} \mathbf{E}(N_L) = 0$, so $\{X_n\}$ is bounded in probability. ■

We now turn our attention to discrete chains as in Proposition 2. We begin with a lemma.

Lemma 9. *For an irreducible Markov chain on a discrete state space with stationary probability distribution π , for any two states x and y , we have $\mathbf{E}_x(\tau_y) < \infty$, i.e. the chain will move from x to y in finite expected time.*

Proof. If this were not the case, then it would be possible from y to travel to x and then take infinite expected time to return to y . This would imply that $\mathbf{E}_y(\tau_y) = \infty$, contradicting the fact that we must have $\mathbf{E}_y(\tau_y) = 1/\pi(y) < \infty$ by positive recurrence. ■

Proof of Proposition 2. Since \mathcal{X} is countable and P is irreducible, $\pi(x) > 0$ for all $x \in \mathcal{X}$. Let $O = K_{2D} \setminus K_D$, and assume that $\pi(O) > 0$ (otherwise increase D to make this so, which can be done unless $\pi(K_D) = 1$ in which case the statement is trivial).

Since K_{2D} is finite, assumption (A1) with $\mu_* = \pi|_{K_{2D} \setminus K_D}$ follows immediately with e.g. $M = (\max_{x,z \in K_{2D}} P(x,y)) / (\min_{z \in K_{2D}} \pi(z)) < \infty$.

Next, note that it follows from Lemma 9 with any $y \in K_D$ (which must exist since we assume K is non-empty) that $\mathbf{E}_x(\tau_{K_D}) < \infty$ for each individual $x \in O$. But then $\mathbf{E}_{\mu_*}(\tau_{K_D}) = \sum_{x \in O} \mu_*(x) \mathbf{E}_x(\tau_{K_D})$, which must also be finite since O is finite. Hence, (A2) also holds. The result thus follows from Theorem 4. ■

6. Two additional probability lemmas.

In this section, we prove two probability results which we will use in the following section.

We first consider expected hitting times. Lemma 9 above shows that *discrete* ergodic Markov chains always have $\mathbf{E}_x(\tau_y) < \infty$. On a general state space, one might think by analogy that for any positive-recurrent ϕ -irreducible Markov chain with stationary distribution π , if $\pi(A) > 0$ and $\pi(B) > 0$, then we must have $\mathbf{E}_{\pi|A}(\tau_B) < \infty$, where τ_B is the hitting time of B . However, this is false. For example, consider a birth-death chain on the positive integers having stationary distribution $\pi(j) \propto j^{-2}$. Then if $B = \{1\}$ and $A = \{J, J+1, J+2, \dots\}$ for any $J > 1$, then $\mathbf{E}_{\pi|A}(\tau_B) \geq \sum_{j=J}^{\infty} \pi(j) (j-1) \propto \sum_{j=J}^{\infty} j^{-2} (j-1) = \infty$.

On the other hand, this result *is* true in the case $A = B$. Indeed, we have:

Lemma 10. *Consider a Markov chain with stationary probability distribution π , and let A be any measurable subset with $\pi(A) > 0$. Then*

- (i) $\mathbf{E}_{\pi|A}(\tau_A) = 1/\pi(A) < \infty$, where τ_A is the first return time to A .
- (ii) For all $k \in \mathbf{N}$, $\mathbf{E}_{\pi|A}(\tau_A^{(k)}) = k/\pi(A) < \infty$, where $\tau_A^{(k)}$ is the k^{th} return time to A .

Proof. Part (i) is essentially the formula of Kac [12]. Indeed, using Theorem 10.0.1 of [15] with $B = \mathcal{X}$, we obtain

$$1 = \pi(\mathcal{X}) = \int_{x \in A} \pi(dx) \mathbf{E}_x \left[\sum_{n=1}^{\tau_A} \mathbf{1}_{X_n \in \mathcal{X}} \right] = \int_{x \in A} \pi(dx) \mathbf{E}_x[\tau_A] = \pi(A) \mathbf{E}_{\pi|A}[\tau_A],$$

giving the result.

For part (ii), we expand the original Markov chain to a new Markov chain on $\mathcal{X} \times \{0, 1, \dots, k-1\}$, where the first variable is the original chain, and the second variable is the count (mod k) of the number of times the chain has returned to A . That is, each time the original chain visits A , the second variable increases by 1 (mod k). Then the expanded chain has stationary distribution $\pi \times \text{Uniform}\{0, 1, \dots, k-1\}$. Hence, by part (i), if we begin in $(\pi|A) \times \delta_0$, then the expected return time of the expanded chain to $A \times \{0\}$ equals $1 / [\pi(A) \times (1/k)] = k/\pi(A)$. But the first return time of the expanded chain to $A \times \{0\}$ corresponds precisely to the k^{th} return time of the original chain to A . ■

We also require the following generalisation of Wald's Equation.

Lemma 11. *Let $\{W_n\}$ be a sequence of non-negative random variables each with finite mean $m < \infty$, and let $\{I_n\}$ be a sequence of indicator variables each with $\mathbf{P}(I_n = 1) = p > 0$. Assume that the pairs sequence $\{(W_n, I_n)\}$ is iid (i.e., the sequence $\{Z_n\}$ is iid where $Z_n = (W_n, I_n)$). Let $\tau = \inf\{n : I_n = 1\}$, and let $S = \sum_{i=1}^{\tau} W_i$. Then $\mathbf{E}(S) < \infty$.*

Proof. If $p = 1$ then $\tau = 1$ and the statement is trivial, so assume $p \in (0, 1)$. We have

$$m = \mathbf{E}[W_i] = p \mathbf{E}[W_i | I_i = 1] + (1 - p) \mathbf{E}[W_i | I_i = 0],$$

so $\mathbf{E}[W_i | I_i = 1] \leq m/p$ and $\mathbf{E}[W_i | I_i = 0] \leq m/(1 - p)$. We then have that

$$\begin{aligned} \mathbf{E}(S) &= \sum_{n=1}^{\infty} \mathbf{P}[\tau = n] \mathbf{E}\left[\sum_{i=1}^n W_i \mid \tau = n\right] \\ &= \sum_{n=1}^{\infty} (1 - p)^{n-1} p \mathbf{E}\left[\sum_{i=1}^n W_i \mid I_1 = \dots = I_{n-1} = 0, I_n = 1\right] \\ &= \sum_{n=1}^{\infty} (1 - p)^{n-1} p \left[\left(\sum_{i=1}^{n-1} \mathbf{E}[W_i | I_i = 0] \right) + \mathbf{E}[W_n | I_n = 1] \right] \\ &\leq \sum_{n=1}^{\infty} (1 - p)^{n-1} p \left[(n - 1)m/(1 - p) + m/p \right] \\ &\leq \left[m / \min(p, 1 - p) \right] \sum_{n=1}^{\infty} n(1 - p)^{n-1} p. \end{aligned}$$

This last sum is a finite expected value, thus giving the result. ■

7. Proof of Theorem 5.

The key to the proof is the following fact about Markov chain hitting times.

Lemma 12. Consider a ϕ -irreducible Markov chain on a state space $(\mathcal{X}, \mathcal{F})$ with transition kernel P and stationary probability distribution π . Let $B, C \in \mathcal{F}$ with $\pi(B) > 0$ and $\pi(C) > 0$, and let μ be any probability measure on $(\mathcal{X}, \mathcal{F})$. Suppose C is a small set for P with minorising measure μ , i.e. there is $\epsilon > 0$ and $n_0 \in \mathbf{N}$ such that $P^{n_0}(x, A) \geq \epsilon \mu(A)$ for all states $x \in C$ and all subsets $A \in \mathcal{F}$. Let τ_B be the first hitting time of B . Then $\mathbf{E}_\mu(\tau_B) < \infty$.

Proof. It suffices to consider the case where $n_0 = 1$, since if not we can replace P by P^{n_0} and note that the hitting time of B by P is at most n_0 times the hitting time of B by P^{n_0} .

We use the Nummelin splitting technique [17, 15]. Specifically, we expand the state space to $\mathcal{X} \times \{0, 1\}$, where the second variable is an indicator of whether or not we are currently regenerating according to μ .

Let $\alpha = \mathcal{X} \times \{1\}$. Then α is a Markov chain atom (i.e. the chain has identical transition probabilities from every state in α), and it has stationary measure $\pi(\alpha) = \epsilon \pi(C) > 0$. So, by Lemma 10(i) above, if the expanded chain is started in α (corresponding to the original chain starting in μ), then it will return to α in finite expected time $1/\pi(\alpha) < \infty$.

We now let W_n be the number of iterations between the $(n - 1)^{\text{st}}$ and n^{th} returns to α , and let $I_n = 1$ if this n^{th} tour visits B , otherwise $I_n = 0$. Then $\mathbf{P}[I_n = 1] > 0$ by the ϕ -irreducibility of P . Hence, $\{(W_n, I_n)\}$ satisfies the conditions of Lemma 11.

Therefore, by Lemma 11, the expected number of iterations until we complete a tour which includes a visit to B is finite. Hence, the expected hitting time of B is finite. ■

Corollary 13. (A3) with $\nu_* = \mu_*$ implies (A2).

Proof. This follows immediately by applying Lemma 12 with $C = K_{2D} \setminus K_D$, and $B = K_D$, and $\mu = \mu_* = \nu_*$. ■

Proof of Theorem 5. Under the assumption (a) that $\nu_* = \mu_*$, the result (4) follows by combining Corollary 13 with Theorem 4. Under the assumption (b) that P is reversible and $\mu_* = \pi_{K_{2D} \setminus K_D}$, it follows from the Appendix (Section 12 below) that (A3) also holds with $\nu_* = \pi|_{K_{2D} \setminus K_D} = \mu_*$. Hence, assumption (a) still applies, so (4) again follows. ■

Remark. One might wonder if it suffices in Theorem 5 to assume (A1) with *any* distribution μ_* , and (A3) with *any* distribution ν_* , without requiring that either $\nu_* = \mu_*$ or $\mu_* = \pi|_{K_{2D} \setminus K_D}$. Under these assumptions, it would still follow from Lemma 10(ii) that the return times to K_{2D} all have finite expectation. And it would still be true that *if* we regenerate from ν_* in finite expected time, then we will eventually hit K_D in finite expected time. The problem is that the expected time to *first* regenerate from ν_* might be infinite. Indeed, conditional upon visiting K_{2D} but repeatedly *failing* to regenerate, the chain could perhaps move to worse and worse states from which it would then take longer and longer to return to K_{2D} . (It is tempting to apply Lemma 11 here where W_n is the time between consecutive visits to K_{2D} and $I_n = 1$ if we regenerate otherwise 0, but unfortunately in that case $\{(W_n, I_n)\}$ are not iid, and conditional on non-regeneration the values of $\mathbf{E}[W_n | I_1 = \dots = I_n = 0]$ could grow unboundedly.)

8. Proof of Proposition 6.

Let $A \subseteq \mathbf{R}^d$ be the ball centered at the origin of radius 1, and let $B \subseteq \mathbf{R}^d$ be the ball centered at the point $(3/2, 0, 0, \dots, 0)$ of radius 1. Then $A \cap B$ has non-empty interior, so $v_d := \text{Leb}(A \cap B) > 0$. In terms of this, we have:

Lemma 14. *Let $A, B \subseteq \mathbf{R}^d$ be two balls with radii $r \leq R$, such that their centres are a distance $w \leq 3r/2 + (R - r)$ apart. Then $\text{Leb}(A \cap B) \geq r^d v_d$.*

Proof. If $r = R = 1$ then this is just the definition of v_d . If one of the balls is stretched by a factor $R > 1$ while moving its center a distance $R - r$ further away, then the new ball contains the old ball, so $\text{Leb}(A \cap B)$ can only increase. Finally, if each of w and r and R are multiplied by the same constant $a > 0$, then the entire geometry is scaled by a factor of a , so $\text{Leb}(A \cap B)$ is multiplied by a^d . Combining these facts, the result follows. \blacksquare

Lemma 15. *Let P be a Markov chain on an open subset $\mathcal{X} \subseteq \mathbf{R}^d$. Let J be a rectangular subset of \mathcal{X} , of the form $J = (a_1, b_1) \times \dots \times (a_d, b_d) \subseteq \mathcal{X}$, where $a_i < b_i$ are extended real numbers (i.e. we might have $a_i = -\infty$ and/or $b_i = \infty$ for some of the i). Suppose there are $\delta > 0$ and $\epsilon > 0$ satisfying the condition (6) that $P(x, dy) \geq \epsilon \text{Leb}(dy)$ whenever $x, y \in J$ with $|y - x| < \delta$. Then for each $n \in \mathbf{N}$, there is $\beta_n > 0$ such that $P^n(x, dy) \geq \beta_n \text{Leb}(dy)$ whenever $x, y \in J$ with $|y - x| < \delta(n + 1)/2$.*

Proof. We first consider the case where $a_i = -\infty$ and $b_i = \infty$ for all i . The result for $n = 1$ follows by assumption. Suppose the result is true for some $n \geq 1$. Let $|y - x| < \delta(n + 1)/2$, let A be the ball centered at x of radius $\delta(n + 1)/2$. and let B be the ball centered at y of radius δ . Then applying Lemma 14 with $r = \delta$ and $R = \delta(n + 1)/2$ and $w = \delta(n + 2)/2$, we see that $\text{Leb}(A \cap B) \geq \delta^d v_d$. The result now follows from the calculation

$$\begin{aligned} P^{n+1}(x, dy) &= \int_{z \in \mathcal{X}} P^n(x, dz) P(z, y) \geq \int_{z \in A \cap B} P^n(x, dz) P(z, y) \\ &\geq \int_{z \in A \cap B} \beta_n \text{Leb}(dz) \epsilon \text{Leb}(dy) \geq \text{Leb}(A \cap B) \beta_n \epsilon \text{Leb}(dy) \\ &\geq \delta^d v_d \beta_n \epsilon \text{Leb}(dy) =: \beta_{n+1} \text{Leb}(dy). \end{aligned}$$

For the general case, by shrinking δ as necessary, we can assume that $\delta < \frac{1}{2} \min_i (b_i - a_i)$. Then in the above calculation we can only use those parts of $A \cap B$ which are still inside J . But here J must contain at least half of $A \cap B$ in each coordinate, i.e. at least $1/2^d$ of $A \cap B$ overall. Hence, $\text{Leb}(A \cap B \cap J) \geq (1/2^d) \text{Leb}(A \cap B)$. So, the above calculation still goes

through, except now with β_{n+1} multiplied by an extra factor of $1/2^d$. ■

Proof of Proposition 6. Let $z = \text{Diam}(J)$. Find $n_0 \in \mathbf{N}$ such that $\delta(n_0 + 1)/2 > z$. Then it follows from Lemma 15 that there is $\epsilon_{n_0} > 0$ such that $P^{n_0}(x, dy) \geq \epsilon_{n_0} \text{Leb}(dy)$ for all $x, y \in J \supseteq K_{2D} \setminus K_D$. Hence, (A3) holds for this n_0 with $\nu_* = \text{Uniform}(K_{2D} \setminus K_D)$ and $\epsilon = \epsilon_{n_0} \text{Leb}(K_{2D} \setminus K_D)$. ■

9. Some facts about geometric ergodicity.

To prove Theorem 7, we need to understand the implications of the geometric ergodicity assumption (A4). The following proposition shows that we can always find a geometric drift function of the form $V(x) = \mathbf{E}_x(\kappa^{\tau_C})$ for some $\kappa > 1$.

Proposition 16. *If P is geometrically ergodic, then there is a small set $C \subseteq \mathcal{X}$ with $\pi(C) > 0$, and a real number $\kappa > 1$, such that if $V(x) = \mathbf{E}_x(\kappa^{\tau_C})$, then V is π -a.e. finite, and $r := \sup_{x \in C} V(x) < \infty$, and V is a geometric drift function in the sense that $PV(x) \leq \lambda V(x) + b \mathbf{1}_C(x) \leq \lambda V(x) + b$ for some $b < \infty$, where $\lambda = \kappa^{-1} < 1$, and $\|P^n(x, \cdot) - \pi\| \leq c V(x) \rho^n$ for $n \in \mathbf{N}$ and $x \in \mathcal{X}$ for some $\rho < 1$ and $c < \infty$, (Here $PV(x) = \int_{y \in \mathcal{X}} V(y) P(x, dy)$, and $\tau_C = \inf\{n \geq 1 : X_n \in C\}$ is the first hitting time of C .)*

Proof. The existence of *some* small set C with $\pi(C) > 0$ follows from e.g. [18] (where they are called “ C -sets”) or [17] or Theorem 5.2.2 of [15]. Geometric ergodicity then implies condition (15.1) of [15] for this C . The existence a (possibly different) small set C and $\kappa > 1$ with $\pi(C) > 0$ and $r < \infty$ then follows from Theorem 15.0.1(ii) of [15].

We then compute directly that if $\{W_n\}$ follows P , then for $x \notin C$,

$$V(x) = \mathbf{E}(\kappa^{\tau_C} | W_0 = x) = \mathbf{E}\left[\mathbf{E}(\kappa^{\tau_C} | W_1) \mid W_0 = x\right]$$

$$\begin{aligned}
&= \int_{y \in \mathcal{X}} \mathbf{E}(\kappa^{\tau_C} | W_1 = y) P(x, dy) = \int_{y \in \mathcal{X}} \mathbf{E}(\kappa^{\tau_C+1} | W_0 = y) P(x, dy) \\
&= \int_{y \in \mathcal{X}} \kappa \mathbf{E}(\kappa^{\tau_C} | W_0 = y) P(x, dy) = \kappa \int_{y \in \mathcal{X}} V(y) P(x, dy) = \kappa PV(x),
\end{aligned}$$

which shows that $PV(x) = \kappa^{-1}V(x)$ for $x \notin C$.

To prove the geometric drift condition, it remains only to prove that $b := \sup_{x \in C} PV(x)$ is finite. For this we use some addition results from [15]. We first compute that in the special case $f \equiv 1$, we have that

$$\begin{aligned}
\sup_{x \in C} \mathbf{E}_x \left(\sum_{k=0}^{\tau_C-1} f(W_k) \kappa^k \right) &= \sup_{x \in C} \mathbf{E}_x \left(\sum_{k=0}^{\tau_C-1} \kappa^k \right) \\
&= \sup_{x \in C} \mathbf{E}_x \left(\frac{\kappa^{\tau_C} - 1}{\kappa - 1} \right) = \frac{\sup_{x \in C} \mathbf{E}_x(\kappa^{\tau_C}) - 1}{\kappa - 1} = \frac{r - 1}{\kappa - 1} < \infty.
\end{aligned}$$

This means that C is an “ f -Kendall set” for $f \equiv 1$, as defined on p. 368 of [15]. Hence, by Theorem 15.2.4 of [15], the function $G(x) := G_C^{(\kappa)}(x, f)$ which equals 1 inside C and equals

$$\mathbf{E}_x \left(\sum_{k=0}^{\tau_C} f(W_k) \kappa^k \right) = \mathbf{E}_x \left(\sum_{k=0}^{\tau_C} \kappa^k \right) = \frac{\mathbf{E}_x(\kappa^{\tau_C+1}) - 1}{\kappa - 1} = \frac{\kappa V(x) - 1}{\kappa - 1} \quad (7)$$

outside of C , satisfies its own geometric drift condition, say $PG(x) \leq \lambda_G G(x) + b_G \mathbf{1}_C(x)$ where $\lambda_G < 1$ and $b_G < \infty$. In particular, since $G(x) = 1$ for $x \in C$, this means that $\sup_{x \in C} PG(x) \leq \lambda_G + b_G < \infty$. Now, by (7), for $x \notin C$ we have $V(x) = \frac{1}{\kappa} [1 + (\kappa - 1)G(x)] \leq 1 + G(x)$. Since $V(x) \leq r$ for $x \in C$, it follows that for all $x \in \mathcal{X}$, we have $V(x) \leq r + G(x)$. Therefore, $PV(x) \leq r + PG(x)$. This shows, finally, that

$$b := \sup_{x \in C} PV(x) \leq r + \sup_{x \in C} PG(x) \leq r + \lambda_G + b_G < \infty.$$

Putting these two facts together shows that $PV(x) \leq \kappa^{-1}V(x) + b$ with $b < \infty$.

The bound on $\|P^n(x, \cdot) - \pi\|$ then follows from Theorem 16.0.1 of [15]. ■

We next establish some bounds based on geometric-drift-type inequalities.

Lemma 17. *Let $\{Z_n\}$ be any stochastic process. Suppose there are $0 < \lambda < 1$ and $b < \infty$ such that for all $n \in \mathbf{N}$, we have $\mathbf{E}(Z_n | Z_0, \dots, Z_{n-1}) \leq \lambda Z_{n-1} + b$. Then for all $n \in \mathbf{N}$,*

$$\mathbf{E}(Z_n | Z_0) \leq \lambda^n Z_0 + \frac{b}{1-\lambda} \leq Z_0 + \frac{b}{1-\lambda}.$$

Proof. We claim that for all $n \geq 0$,

$$\mathbf{E}(Z_n | Z_0) \leq \lambda^n Z_0 + (1 + \lambda + \dots + \lambda^{n-1}) b. \quad (8)$$

Indeed, for $n = 0$ this is trivial, and for $n = 1$ this is equivalent to the hypothesis of the lemma. Suppose now that (8) holds for some value of n . Then

$$\begin{aligned} \mathbf{E}(Z_{n+1} | Z_0) &= \mathbf{E}\left(\mathbf{E}(Z_{n+1} | Z_0, \dots, Z_n) \mid Z_0\right) \leq \mathbf{E}\left(\lambda Z_n + b \mid Z_0\right) \\ &\leq \lambda\left(\lambda^n Z_0 + (1 + \lambda + \dots + \lambda^{n-1}) b\right) + b = \lambda^{n+1} Z_0 + (1 + \lambda + \dots + \lambda^{n-1} + \lambda^n) b, \end{aligned}$$

so (8) holds for $n + 1$. Hence, by induction, (8) holds for all $n \geq 0$.

The result now follows since $1 + \lambda + \dots + \lambda^{n-1} = \frac{1-\lambda^n}{1-\lambda} \leq \frac{1}{1-\lambda}$. ■

Proposition 18. *If P is geometrically ergodic with stationary probability distribution π and π -a.e. finite geometric drift function V satisfying $PV(x) \leq \lambda V(x) + b$ where $0 \leq \lambda < 1$ and $0 \leq b < \infty$, then $\mathbf{E}_\pi(V) \leq b/(1-\lambda) < \infty$.*

Proof. Choose any $x \in \mathcal{X}$ with $V(x) < \infty$ (which holds for π -a.e. $x \in \mathcal{X}$). Then applying Lemma 17 to $Z_n = P^n V(x)$ gives $P^n V(x) \leq V(x) + \frac{b}{1-\lambda}$, and in particular $P^n V(x) \not\rightarrow \infty$. But Theorem 14.3.3 of [15] with $f = V$ states that if $\pi(V) = \infty$, then $P^n V(x) \rightarrow \infty$ for all $x \in \mathcal{X}$. Hence, by contraposition, we must have $\pi(V) < \infty$.

Finally, we have by stationarity that $\pi(V) = \pi(PV)$. So, taking expectations with respect to π of both sides of the inequality $PV \leq \lambda V + b$ and using that $\pi(V) < \infty$, we obtain that $\pi(V) \leq \lambda \pi(V) + b$, whence $\pi(V) \leq b/(1-\lambda)$. ■

Remark. If P is *uniformly ergodic*, meaning that (A4) holds for a *constant* function $V < \infty$, then it follows from Theorem 16.0.2(vi) of [15] that $U := \sup_{x \in \mathcal{X}} \mathbf{E}_x(\tau_{K_D}) < \infty$, which implies that $\mathbf{E}_{\mu_*}(\tau_{K_D}) \leq U < \infty$, so (A2) holds.

10. Proof of Theorem 7.

The key to the proof is a uniform bound on certain powers of P :

Lemma 19. *Assuming (A4) and (A5), with V as in Proposition 16, $\sup_{x \in K_D} \sup_{n \geq 0} P^n V(x) < \infty$.*

Proof. For $x \in K_D$, $PV(x) = \mathbf{E}_{y \sim P(x, \cdot)} V(y) \leq M \mathbf{E}_{y \sim \pi} V(y) = M \pi(V) < \infty$ by (A5) and Proposition 18. Then applying Lemma 17 to $Z_n = P^n V(x)$ gives $P^n V(x) \leq M \pi(V) + \frac{b}{1-\lambda}$. In particular, $\sup_{x \in K_D} \sup_{n \geq 1} P^n V(x) < \infty$.

Furthermore, for $x \in K_D$, $V(x) = \mathbf{E}_x(\kappa^{\tau_C}) = \kappa \mathbf{E}_{P(x, \cdot)}(\kappa^{\tau_C}) \leq \kappa M \mathbf{E}_\pi(\kappa^{\tau_C}) = \kappa M \pi(V) < \infty$ by Proposition 18, so the above “sup” can be extended to include $n = 0$ too. ■

Remark. For Metropolis algorithms on continuous state spaces, usually $P(x, \{x\}) > 0$ for most $x \in \mathcal{X}$, so (A5) usually won’t hold (though (A1) often will; see Section 11). On the other hand, if $P(x, \cdot) = r(x) \delta_x(\cdot) + (1 - r(x)) R(x, \cdot)$ where δ_x is a point-mass at x and $0 \leq r(x) \leq 1$ and R satisfies (A5), then it is easily seen that if $\kappa r(x) \leq B < 1$ for all $x \in K_D$, then Lemma 19 still holds with $\sup_{x \in K_D} V(x) \leq \kappa M \pi(V)/(1 - B) < \infty$, and the rest of the proof of Theorem 7 then goes through without change.

Proposition 20. *Assuming (A4) and (A5), the random sequence $\{V(X_n)\}$ is bounded in probability.*

Proof. By Lemma 19, $U := \sup_{x \in K_D} V(x) < \infty$. Since the adversary can only adjust the values of $\{X_n\}$ within K_D , it follows that $\{X_n\}$ still satisfies a drift condition like (A4)

except with b replaced by $b + U < \infty$. Hence, Lemma 17 applied to $Z_n = V(X_n)$ says that $\{\mathbf{E}[V(X_n)]\}$ is bounded in probability, i.e. that $\zeta := \sup_{x \in K_D} \sup_{n \geq 0} \mathbf{E}[V(X_n) | X_0 = x] < \infty$. It follows by Markov's inequality that $\mathbf{P}[V(X_n) \geq R] \leq \zeta/R$ for all n and all $R > 0$. Hence, $\{V(X_n)\}$ is bounded in probability. ■

Remark. Proposition 20 immediately implies a bound on the ϵ convergence times [21] defined by $M_\epsilon(x) = \inf\{n \geq 1 : \|P^n(x, \cdot) - \pi(\cdot)\| \leq \epsilon\}$. Indeed, by Proposition 16 we have $\|P^n(x, \cdot) - \pi\| \leq cV(x)\rho^n$, whence $M_\epsilon(x) \leq \lceil \log(cV(x)/\epsilon) / \log(1/\rho) \rceil$. Since $\{V(X_n)\}$ is bounded in probability by Proposition 20, it follows that $\{M_\epsilon(X_n)\}$ is bounded in probability too. (See also the Containment condition (11) below.)

Proof of Theorem 7. The bounded-jumps condition (2) implies that the small set C must be bounded (in fact, of diameter $\leq 2Dn_0$). Let $r = \sup\{|x| : x \in C\} < \infty$. Then if $|x| > r$, then it takes at least $(|x| - r)/D$ steps to return to C from x . Hence, $V(x) \geq \kappa^{(|x|-r)/D}$. Therefore, $|x| \leq r + D \log(V(x))/\log(\kappa)$, so $|X_n| \leq r + D \log(V(X_n))/\log(\kappa)$. But $\{V(X_n)\}$ is bounded in probability by Proposition 20. Hence, so is $\{X_n\}$. ■

11. Application to adaptive MCMC algorithms.

Markov chain Monte Carlo (MCMC) algorithms proceed by running a Markov chain $\{X_n\}$ with stationary probability distribution π , in the hopes that $\{X_n\}$ converges in probability to π , i.e. that

$$\lim_{n \rightarrow \infty} \sup_{A \in \mathcal{F}} |\mathbf{P}(X_n \in A) - \pi(A)| = 0, \quad x \in \mathcal{X}, \quad A \in \mathcal{F}. \quad (9)$$

If so then for large n , the value of X_n is approximately a “sample” from π . Such algorithms are hugely popular in e.g. Bayesian statistical inference; for an overview see e.g. [7].

Adaptive MCMC algorithms [10] attempt to speed up the convergence (9) and thus make MCMC more efficient, by modifying the Markov chain transitions during the run (i.e. “on the fly”) in a search for a more optimal chain; for a brief introduction see e.g. [23]. Such algorithms often appear to work very well in practice (e.g. [22, 9, 8, 4]). However, they are no longer Markov chains (since the adaptations typically depend on the process’s entire history), making it extremely difficult to establish mathematically that the convergence (9) will even be preserved (much less improved). As a result, many papers either make the artificial assumption that the state space \mathcal{X} is compact (e.g. [10, 8, 4]), or prove the convergence (9) using complicated mathematical arguments requiring strong and/or uncheckable assumptions (e.g. [3, 1, 21, 9, 2, 25, 13]), or do not prove (9) at all and simply hope for the best. It is difficult to find simple, easily-checked conditions which provably guarantee the convergence (9) for adaptive MCMC algorithms.

One step in this direction is in [21], where it is proved that the convergence (9) is implied by two conditions. The first condition is *Diminishing Adaptation*, which says that the process adapts less and less as time goes on, or more formally that

$$\lim_{n \rightarrow \infty} \sup_{x \in \mathcal{X}} \sup_{A \in \mathcal{F}} |P_{\Gamma_{n+1}}(x, A) - P_{\Gamma_n}(A)| = 0 \quad \text{in probability.} \quad (10)$$

The second condition is *Containment*, which says that the process’s convergence times are bounded in probability, or more formally that

$$\{M_\epsilon(X_n, \Gamma_n)\}_{n=1}^\infty \quad \text{is bounded in probability,} \quad (11)$$

where $M_\epsilon(x, \gamma) = \inf\{n \geq 1 : \|P_\gamma^n(x, \cdot) - \pi(\cdot)\| \leq \epsilon\}$ is the ϵ *convergence time*. The first of these two conditions is usually easy to satisfy directly by wisely designing the algorithm, so it is not of great concern. However, the second condition is notoriously difficult to verify (see e.g. [5]) and thus a severe limitation (though an essential condition, c.f. [14]). On the other hand, the Containment condition (11) is reminiscent of the boundedness in probability of $\{X_n\}$ in (4), which is implied by our various theorems above. This suggests that the

theorems herein might be useful in establishing the Containment condition (11) for certain adaptive MCMC algorithms, as we now explore.

11.1. The adaptive MCMC setup.

Within the context of Section 2, we define an adaptive MCMC algorithm as follows. Let \mathcal{X} be an open subset of \mathbf{R}^d for some $d \in \mathbf{N}$, on which π is some probability distribution. Assume there is a collection $\{P_\gamma\}_{\gamma \in \mathcal{Y}}$ of Markov kernels on \mathcal{X} which are Harris-ergodic to π as in (1), for some compact index set \mathcal{Y} . The adversary proceeds by choosing, at each iteration n , an index $\Gamma_n \in \mathcal{Y}$ (possibly depending on n and/or the process's entire history). The process $\{X_n\}$ then moves at time n according to the transition kernel P_{Γ_n} , i.e.

$$\mathbf{P}(X_{n+1} \in A \mid X_n = x, \Gamma_n = \gamma, X_0, \dots, X_{n-1}) = P_\gamma(x, A).$$

To reflect the bounded jump condition (2), we assume

$$P_\gamma\left(x, \{y \in \mathcal{X} : |y - x| \leq D\}\right) = 1, \quad x \in \mathcal{X}, \quad \gamma \in \mathcal{Y}. \quad (12)$$

To reflect that the adversary can only adapt inside K , we assume that the P_γ kernels are all equal outside of K , i.e. that

$$P_\gamma(x, A) = P(x, A), \quad A \in \mathcal{F}, \quad x \in \mathcal{X} \setminus K, \quad (13)$$

for some fixed Markov chain kernel $P(x, dy)$. We further assume that

$$\exists M < \infty \quad \text{s.t.} \quad P(x, dy) \leq M \text{Leb}(dy), \quad x \in K_D \setminus K, \quad z \in K_{2D} \setminus K_D. \quad (14)$$

We also assume the ϵ - δ condition (6) that $P(x, dy) \geq \epsilon \text{Leb}(dy)$ whenever $x, y \in J$ with $|y - x| < \delta$, for some rectangle J with $K_{2D} \setminus K_D \subseteq J \subseteq \mathcal{X}$.

We shall particularly focus on the case where each P_γ is a Metropolis-Hastings algorithm. This means that P_γ proceeds, given X_n , by first choosing a proposal state $Y_{n+1} \sim Q_\gamma(X_n, \cdot)$ for some proposal kernel $Q_\gamma(x, \cdot)$ having a density $q_\gamma(x, y)$ with respect to Leb . Then, with

probability $\alpha_\gamma(X_n, Y_{n+1}) := \min[1, \frac{\pi(Y_{n+1}) q_\gamma(Y_{n+1}, X_n)}{\pi(X_n) q_\gamma(X_n, Y_{n+1})}]$ it *accepts* this proposal by setting $X_{n+1} = Y_{n+1}$. Otherwise, with probability $1 - \alpha_\gamma(X_n, Y_{n+1})$, it *rejects* this proposal by setting $X_{n+1} = X_n$. Thus,

$$P_\gamma(x, A) = r(x) \delta_x(A) + \int_{y \in A} Q_\gamma(x, dy) \alpha_\gamma(x, y)$$

where $\delta_x(\cdot)$ is a point-mass at x , and $r(x) = 1 - \int_{y \in \mathcal{X}} Q_\gamma(x, dy) \alpha_\gamma(x, y)$ is the overall probability of rejecting. Note that (12) and (13) and (14) are each automatically satisfied for P_γ and P if the same equations are satisfied for corresponding Q_γ and Q .

To state a clean theorem, we also assume continuous densities, as follows:

- (A6)** π has a continuous positive density function (with respect to Lebesgue), and the transition probabilities $P_\gamma(x, dy)$ either (i) have densities which are continuous functions of x and y and γ , or (ii) are Metropolis-Hastings algorithms whose proposal kernel densities $q_\gamma(x, dy)$ are continuous functions of x and y and γ .

11.2. An adaptive MCMC theorem.

In terms of the above setup, we have:

Theorem 21. *Consider an adaptive MCMC algorithm as in Section 11.1, on an open subset \mathcal{X} of \mathbf{R}^d , such that the kernels P_γ (or the proposal kernels Q_γ in the case of adaptive Metropolis-Hastings) have bounded jumps as in (12), and no adaption outside of K as in (13), with the fixed kernel P (or a corresponding fixed proposal kernel Q) bounded above as in (14). We further assume the ϵ - δ condition (6) on P , and the continuous densities condition (A6). Then the algorithm satisfies the Containment condition (11). Hence, assuming the Diminishing Adaptation condition (10), the algorithm converges in distribution to π as in (9).*

Theorem 21 is proved in Section 11.3 below. Clearly, similar reasoning also applies with alternative assumptions, and to other versions of adaptive MCMC including e.g. adaptive Metropolis-within-Gibbs algorithms (with P replaced by P^d for random-scan), c.f. [13].

Theorem 21 requires many conditions, but they are all conditions which are easy to ensure in practice, as illustrated by the following specific application example.

Example. Consider an adaptive MCMC algorithm of the following type. Let \mathcal{X} be an open subset of \mathbf{R}^d containing a large bounded rectangle K . Let \mathcal{Y} be a compact collection of d -dimensional positive-definite matrices. Define a process $\{X_n\}$ as follows: $X_0 = x_0$ for some fixed $x_0 \in K$. If $X_n \notin K$, it simulates a proposal $Y_{n+1} \sim N(X_n, \Sigma_0)$ for some fixed $\Sigma_0 \in \mathcal{Y}$. Then if $|Y_{n+1} - X_n| > D$, the proposal is rejected and $X_{n+1} = X_n$. Otherwise, with probability $\min[1, \frac{\pi(Y_{n+1})}{\pi(X_n)}]$ the proposal is accepted and $X_{n+1} = Y_{n+1}$, or with the remaining probability the proposal is rejected and $X_{n+1} = X_n$. (Here $q_\gamma(Y_{n+1}, X_n) = q_\gamma(X_n, Y_{n+1})$, so those factors cancel.) If $X_n \in K$, then some matrix $\Sigma_{n+1} \in \mathcal{Y}$ is chosen in some fashion, perhaps depending on X_n and on the chain's entire history, subject to the continuity condition that as a function of X_n the matrix $\Sigma_{n+1}(X_n)$ converges to Σ_0 as X_n converges to the boundary of K . The process then proceeds as above, except with Σ_{n+1} in place of Σ_0 .

This example is designed to satisfy all of the conditions of Theorem 21. Bounded jumps (12), and no adaption outside of K (13), are both immediate. Here the fixed kernel Q is bounded above (14) by the constant $M = (2\pi)^{-d/2} |\Sigma_0|^{-1/2}$, and the ϵ - δ condition (6) holds by the formula for Q together with the continuity of the density π (which guarantees that it is bounded above and below on the compact set K_{2D}). Furthermore the continuous densities condition (A6) holds by construction. Hence, by Theorem 21, this algorithm satisfies Containment (11). It follows that if the selection of the matrices Σ are done in a way which satisfies Diminishing Adaptation (10), then convergence (9) holds. ■

11.3. Proof of Theorem 21.

We begin with a result linking the boundedness property (4) for $\{X_n\}$ with the Containment condition (11) for $\{M_\epsilon(X_n, \Gamma_n)\}$, as follows:

Proposition 22. *Suppose (4) holds, and for each $n \in \mathbf{N}$ the mapping $(x, \gamma) \mapsto \Delta(x, \gamma, n) := \|P_\gamma^n(x, \cdot) - \pi(\cdot)\|$ is continuous. Then Containment (11) holds.*

Proof. Since each P_γ is Harris ergodic, $\lim_{n \rightarrow \infty} \Delta(x, \gamma, n) = 0$ for each fixed $x \in \mathcal{X}$ and $\gamma \in \mathcal{Y}$. If the mapping $(x, \gamma) \mapsto \Delta(x, \gamma, n)$ is continuous, then it follows by Dini's Theorem (e.g. [24], p. 150) that for any compact subset $C \subseteq \mathcal{X}$, since \mathcal{Y} is compact,

$$\lim_{n \rightarrow \infty} \sup_{x \in C} \sup_{\gamma \in \mathcal{Y}} \Delta(x, \gamma, n) = 0.$$

Hence, given C and $\epsilon > 0$, there is $n \in \mathbf{N}$ with $\sup_{x \in C} \sup_{\gamma \in \mathcal{Y}} \Delta(x, \gamma, n) < \epsilon$. It follows that $\sup_{x \in C} \sup_{\gamma \in \mathcal{Y}} M_\epsilon(x, \gamma) < \infty$ for any fixed $\epsilon > 0$.

Now, if $\{X_n\}$ is bounded in probability as in (4), then for any $\delta > 0$, we can find a large enough compact subset C such that $P(X_n \notin C) \leq \delta$ for all n . Then given $\epsilon > 0$, if $L := \sup_{x \in C} \sup_{\gamma \in \mathcal{Y}} M_\epsilon(x, \gamma)$, then $L < \infty$, and $P(M_\epsilon(X_n, \Gamma_n) > L) \leq \delta$ for all n as well. Since δ was arbitrary, it follows that $\{M_\epsilon(X_n, \Gamma_n)\}_{n=0}^\infty$ is bounded in probability. ■

We then need a lemma guaranteeing continuity of $\Delta(x, \gamma, n)$:

Lemma 23. *Under the continuous density assumptions (A6), for each $n \in \mathbf{N}$, the mapping $(x, \gamma) \mapsto \Delta(x, \gamma, n)$ is continuous.*

Proof. Assuming (A6)(ii), this fact is contained in the proof of Corollary 11 of [21]. The corresponding result assuming (A6)(i) is similar but easier. ■

Proof of Theorem 21. The bounded jumps condition (12), together with no adaption outside of K (13), ensure that the algorithm $\{X_n\}$ fits within the setup of Section 2. Since the densities of $P(x, dy)$ are bounded above by (14), it follows that (A1) holds with $\mu_* = \text{Uniform}(K_{2D} \setminus K_D)$. Also, using the ϵ - δ condition (6), it follows from Proposition 6

that (A3) holds for $\nu_* = \mu_*$. Hence, by Theorem 5(a), $\{X_n\}$ is bounded in probability, i.e. (4) holds. In addition, using the continuity assumption (A6), it follows from Lemma 23 that $\Delta(x, \gamma, n)$ is a continuous function. The result then follows from Proposition 22. ■

12. Appendix: Replacing the minorising measure by π .

Recall that Assumption (A3) requires that the set $K_{2D} \setminus K_D$ be small for P , with some minorising measure ν_* . It turns out that if Assumption (A3) holds for any ν_* , and if P is reversible, then Assumption (A3) also holds for the specific choice $\nu_* = \pi|_{K_{2D} \setminus K_D}$, i.e. where $\nu_*(A) = \pi(A \cap (K_{2D} \setminus K_D)) / \pi(K_{2D} \setminus K_D)$, except with the step size n_0 replaced by $2n_0$. Under the additional assumption of uniform ergodicity, this fact is Proposition 1 of [20]. For arbitrary reversible chains, this fact follows from Lemma 5.9 of the Polish doctoral thesis [16], which for completeness is reproduced as Lemma 24 below.

Lemma 24. *(Lemma 5.9 of [16]) Let P be a Markov chain transition kernel on $(\mathcal{X}, \mathcal{F})$. Let $C \in \mathcal{F}$ such that $\pi(C) > 0$. Assume that C is a small set for P , i.e. for some $n_0 \in \mathbb{N}$ and $\beta > 0$ and probability measure ν ,*

$$P^{n_0}(x, A) \geq \beta \mathbf{1}_C(x) \nu(A), \quad A \in \mathcal{F}. \quad (15)$$

Then

$$P^{n_0}(P^*)^{n_0}(x, A) \geq \frac{1}{4} \beta^2 \mathbf{1}_C(x) \pi(A \cap C), \quad A \in \mathcal{F}, \quad (16)$$

where P^* is the adjoint of P . In particular, if P is reversible, then $P^* = P$, and

$$P^{2n_0}(x, A) \geq \frac{1}{4} \beta^2 \mathbf{1}_C(x) \pi(A \cap C), \quad A \in \mathcal{F}.$$

Hence if $K_{2D} \setminus K_D$ is an n_0 -small set with minorising measure ν , and P is reversible, then $K_{2D} \setminus K_D$ is a $2n_0$ -small set with minorising measure $\pi|_{K_{2D} \setminus K_D}$.

Proof. By replacing P by P^{n_0} and P^* by $(P^*)^{n_0}$, it suffices to assume that $n_0 = 1$. We wish to bound the Radon-Nikodym derivative of ν with respect to π . Integrating both sides of (15) with respect to π yields that

$$\frac{d\nu}{d\pi} \leq \beta\pi(C) \quad \text{and} \quad \int_{\mathcal{X}} \frac{d\nu}{d\pi} \pi(dx) = 1.$$

Hence, for every $\varepsilon \in [0, 1]$, the set

$$D(\varepsilon) := \{x \in \mathcal{X} : \frac{d\nu}{d\pi} \geq \varepsilon\} \tag{17}$$

has $\pi(D(\varepsilon)) > 0$. We then compute that

$$\nu(D(\varepsilon)^c) = \int_{D(\varepsilon)^c} \frac{d\nu}{d\pi}(x) \pi(dx) \leq \varepsilon \int_{\mathcal{X}} \pi(dx) = \varepsilon$$

and hence

$$\nu(D(\varepsilon)) \geq 1 - \varepsilon. \tag{18}$$

Recall also that the adjoint P^* satisfies

$$\pi(dx) P(x, dy) = \pi(dy) P^*(y, dx). \tag{19}$$

Now let $x \in C$, and $A \in \mathcal{F}$ with $A \cap C \neq \emptyset$. Using first (15) and then (17),

$$\begin{aligned} PP^*(x, A) &= \int_{z \in \mathcal{X}} P^*(z, A) P(x, dz) \geq \int_{z \in \mathcal{X}} P^*(z, A \cap C) \nu(dz) \\ &\geq \beta \int_{z \in D(\varepsilon)} \int_{y \in A \cap C} P^*(z, dy) \varepsilon \pi(dz). \end{aligned}$$

To continue, use (19), and then (15) again, and finally (18), to obtain

$$\begin{aligned} PP^*(x, A) &\geq \beta \varepsilon \int_{z \in D(\varepsilon)} \int_{y \in A \cap C} \pi(dy) P(y, dz) \\ &\geq \beta^2 \varepsilon \nu(D(\varepsilon)) \pi(A \cap C) \geq \beta^2 \varepsilon (1 - \varepsilon) \pi(A \cap C). \end{aligned}$$

Setting $\varepsilon = 1/2$ yields (16). ■

References

- [1] C. Andrieu and E. Moulines (2003), On the ergodicity properties of some adaptive Markov Chain Monte Carlo algorithms. *Ann. Appl. Prob.* **16**, 1462–1505.
- [2] Y. Atchadé and G. Fort (2010), Limit theorems for some adaptive MCMC algorithms with subgeometric kernels. *Bernoulli* **16(1)**, 116–154.
- [3] Y. Atchadé and J.S. Rosenthal (2005), On Adaptive Markov Chain Monte Carlo Algorithms. *Bernoulli* **11(5)**, 815–828.
- [4] Y. Bai, R.V. Craiu, and A.F. Di Narzo (2011), Divide and conquer: a mixture-based approach to regional adaptation for MCMC. *J. Comp. Graph. Stat.* **20(1)**, 63–79.
- [5] Y. Bai, G.O. Roberts, and J.S. Rosenthal (2011), On the Containment condition for adaptive Markov chain Monte Carlo algorithms. *Adv. Appl. Stat.* **21**, 1–54.
- [6] A. Borodin, J. Kleinberg, P. Raghavan, M. Sudan, and D.P. Williamson (2001), Adversarial queuing theory. *J. ACM* **48(1)**, 13–38.
- [7] S. Brooks, A. Gelman, G.L. Jones, and X.-L. Meng, eds. (2011), *Handbook of Markov chain Monte Carlo*. Chapman & Hall / CRC Press.
- [8] R.V. Craiu, J.S. Rosenthal, and C. Yang (2009), Learn from thy neighbor: parallel-chain adaptive MCMC. *J. Amer. Stat. Assoc.* **488**, 1454–1466.
- [9] P. Giordani and R. Kohn (2010), Adaptive independent Metropolis-Hastings by fast estimation of mixtures of normals. *J. Comp. Graph. Stat.* **19(2)**, 243–259.
- [10] H. Haario, E. Saksman, and J. Tamminen (2001), An adaptive Metropolis algorithm. *Bernoulli* **7**, 223–242.
- [11] B. Hajek (1982). Hitting-time and occupation-time bounds implied by drift analysis with applications. *Adv. Appl. Prob.* **14**, 502–525.

- [12] M. Kac (1947), On the notion of recurrence in discrete stochastic processes. *Bull. Amer. Math. Soc.* **53**, 1002–1010.
- [13] K. Latuszynski, G.O. Roberts, and J.S. Rosenthal (2013), Adaptive Gibbs samplers and related MCMC methods. *Ann. Appl. Prob.* **23(1)**, 66–98.
- [14] K. Latuszynski and J.S. Rosenthal (2014), The Containment Condition and AdapFail algorithms. *J. Appl. Prob.* **51(4)**, to appear.
- [15] S.P. Meyn and R.L. Tweedie (1993), *Markov chains and stochastic stability*. Springer-Verlag, London. Available at: <http://probability.ca/MT/>
- [16] B. Miasojedow (2011), *Oszacowania błędów estymatorów stosowanych w markowowskich metodach Monte Carlo*. Ph.D. thesis, University of Warsaw. (In Polish.)
- [17] E. Nummelin (1984), *General irreducible Markov chains and non-negative operators*. Cambridge University Press.
- [18] S. Orey (1971), *Lecture notes on limit theorems for Markov chain transition probabilities*. Van Nostrand Reinhold, London.
- [19] R. Pemantle and J.S. Rosenthal (1999), Moment conditions for a sequence with negative drift to be uniformly bounded in L^r . *Stoch. Proc. Appl.* **82**, 143–155.
- [20] G.O. Roberts and J.S. Rosenthal (1998), Two convergence properties of hybrid samplers. *Ann. Appl. Prob.* **8(2)**, 397–407.
- [21] G.O. Roberts and J.S. Rosenthal (2007), Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *J. Appl. Prob.* **44**, 458–475.
- [22] G.O. Roberts and J.S. Rosenthal (2009), Examples of adaptive MCMC. *J. Comp. Graph. Stat.* **18(2)**, 349–367.

- [23] J.S. Rosenthal (2011), Optimal Proposal Distributions and Adaptive MCMC. Chapter 4 of the book [7].
- [24] W.R. Rudin (1976), Principles of mathematical analysis, 3rd ed. McGraw-Hill.
- [25] E. Saksman and M. Vihola (2010), On the ergodicity of the adaptive Metropolis algorithm on unbounded domains. *Ann. Appl. Prob.* **20(6)**, 1967–2388.