Recurrent and Ergodic Properties of Adaptive MCMC

By

CHAO YANG

Department of Mathematics,

University of Toronto,

M5S 2E4, Toronto ON, Canada

November 30, 2007

Abstract

We will discuss the recurrence on the state space of the adaptive MCMC algorithm using some examples. We present the ergodicity properties of adaptive MCMC algorithms under the minimal recurrent assumptions, and show the Weak Law of Large Numbers under the same conditions. We will analyze the relationship between the recurrence on the product space of state space and parameter space and the ergodicity, give a counter-example to open problem 21 in Roberts and Rosenthal's paper, and try to give the positive results under some stronger conditions.

1 Introduction

Markov chain Monte Carlo (MCMC) algorithms are widely used to generate samples from any probability distribution π on the state space $\mathcal{X}$. However it is generally acknowledged that the choice of an effective transition kernel is essential to obtain reasonable results by simulation in a limited amount of time. And such kernels are often very difficult to be well chosen (see Gelman et al.1996 [4]; Gilks et al 1996 [5]; Haario et al 1991 [7]; Roberts et al 1997 [11]). A possible solution so-called adaptive MCMC has been proposed recently. The adaptive MCMC algorithm will tune the transition kernel at each step using the past simulations and try to "learn" the best parameter values while the chain runs. See Gilks et al (1998) [6], Haario et al. (1999)[8]; (2001) [9], Andrieu and Moulines (2005) [2], Andrieu and Robert (2001)[3], Roberts and Rosenthal(2005) [14] [15], Atchade and Rosenthal (2005) [17], and Andieu and Achade (2005) [1] for example.

An important paper about the ergodicity of AMCMC was written by Roberts and Rosenthal [14] (2007). They present some simpler conditions, which still ensure the ergodicity of the specified target distribution. They also mentioned some research directions. We will continue to study the ergodicity of AMCMC along these directions, try to find some weaker conditions to ensure the ergodicity and discuss the relationship between the recurrence on the product space (of the state space and the parameter space) and the ergodicity.

The paper is organized as follows. Section 2 gives some introductions to the notations and definitions. In section 3, we will introduce our main results: the ergodic theorem of AMCMC under the weakest drift conditions such that each kernel is positive recurrence and the weak law of large numbers (WLLN) under the same conditions. Further we will discuss the uniformly recurrent conditions in the same section after constructing some simple examples to show that usually AMCMC does not have good recurrence property. In section 4 and section 5 we will give the proof of the ergodic theorem and the WLLN. Finally, we consider the recurrent property on the product space of the state space and the parameter one in section 6. We will give the negative answer to the open problem 21 in Roberts and Rosenthal (2005) [14] using a counter example, and present some positive results under stronger conditions.

2 Preliminaries

Before describing the procedure under study, it is necessary to introduce some notation and definitions.

2.1 Adaptive MCMC

Suppose $\pi(\cdot)$ is a fixed "target" probability distribution, on a state space $\mathcal{X}$ with $\sigma-$ algebra $\mathcal{F}$. The common MCMC algorithm is to construct Markov chain kernel P which has $\pi(\cdot)$ as its stationary distribution such that:

$$\|P^n(x,\cdot) - \pi(\cdot)\| \to_{n\to\infty} 0$$

for any $x \in \mathcal{X}$, where $\|\mu(\cdot) - \nu(\cdot)\| = \sup_{B\in\mathcal{F}} |\mu(B) - \nu(B)|$ is the usual total variation distance. However in an AMCMC we will try to select an "optimal" kernel at each step using the information from the historical simulation, like what Haario et al (2001) [9] did in their well-known AMCMC algorithm. Atchade and Rosenthal (2003) [17], Andrieu and Moulines (2003) [2] generalize their results with proving convergence of more general adaptive MCMC algorithms. Here we will formalize the AMCMC as what Roberts and Roenthal [14](2007) did.

We let $\{P_\gamma\}_{\gamma\in\mathcal{Y}}$ be a collection of Markov chain kernels on $\mathcal{X}$, each of which is $\phi-$irreducible and aperiodic(which it usually will be) and has $\pi(\cdot)$ as a stationary distribution: $(\pi P_\gamma)(x,\cdot) = \pi(\cdot)$, and we call the set $\mathcal{Y}$ parameter space. Let Γ_n be $\mathcal{Y}-$valued random variables which are updated according to specific rules. Consider a discrete time series $\{X_n\}$ on χ as below:

$$P[X_{n+1} \in A | X_n = x, \Gamma_n = \gamma, \mathcal{G}_n] = P_\gamma(x, A) \tag{2.1}$$

where $\mathcal{G}_n = \sigma(X_0, \cdots, X_n, \Gamma_0, \cdots, \Gamma_n)$. Then we call $\{X_n\}$ an adaptive MCMC with adaptive scheme Γ_n. Let

$$A^{(n)}((x,\gamma), B) = P[X_n \in B | X_0 = x, \Gamma_0 = \gamma], \qquad B \in \mathcal{F}$$

and

$$T((x,\gamma), n) = \|A^{(n)}((x,\gamma),\cdot) - \pi(\cdot)\|$$

According to the definition in Roberts, Rosenthal, and Schwartz [16] (1998), we say a family $\{P_\gamma\}_{\gamma\in\mathcal{Y}}$ of Markov chain kernels is simultaneously strongly aperiodically geometrically ergodic if there is $C \in \mathcal{F}$, $V : \mathcal{X} \to [1,\infty)$, $\delta > 0$, $\lambda < 1$, and $b < \infty$, such that $\sup_C V = v < \infty$, and

(i) for each $\gamma \in \mathcal{Y}$, there exists a probability measure $\nu_\gamma(\cdot)$ on C with $P_\gamma(x,\cdot) \geq \delta\nu_\gamma(\cdot)$ for all $x \in C$; and

(ii) $(P_\gamma V)(x) \leq \lambda V(x) + b\mathbb{I}_C(x)$

In Roberts and Rosenthal [14] (2007), they proved the following ergodic theorems:

Theorem 2.1. *Consider an adaptive MCMC algorithm on a state space* χ*, with adaptation index* $\mathcal{Y}$ *and the adaptive scheme is* Γ_n*.* $\pi(\cdot)$ *is stationary for each kernel* P_γ *for*

$\gamma \in \mathcal{Y}$. Suppose also that $\{P_\gamma\}_{\gamma\in\mathcal{Y}}$ is simultaneously strongly aperiodically geometrically ergodic and the Adaptive scheme satisfies the following condition:

[Diminishing Adaption] *$lim_{n\to\infty} D_n = 0$ in probability, where $D_n = sup_{x\in\mathcal{X}} \|P_{\Gamma_{n+1}} - P_{\Gamma_n}\|$ is a $\mathcal{G}_{n+1}$-measurable random variable.*

Then $lim_{n\to\infty} T(x,\gamma,n) = 0$ for all $x \in \mathcal{X}$ and $\gamma \in \mathcal{Y}$.

2.2 Recurrence Properties

In this part we will recall the definition of recurrence of general Markov chain and some related results. The recurrence property describes the behavior of the occupation time random variable $\eta_A = \sum_{n=1}^{\infty} \mathbb{I}\{X_n \in A\}$ which counts the number of visits to a set A. Therefor we have the following definition (see Chapter 8 in Meyn and Tweedie (1993) [11]):

The set A is called recurrent if $E_x[\eta_A] = \infty$ for all $x \in A$. If every A is recurrent, we say that the chain is recurrent.

3 The Ergodic Property And The Weak Law Of Large Numbers

3.1 The main results

First let us think about how to compare two elements γ_1 and γ_2 in the parameter space $\mathcal{Y}$. Actually what we need to describe is the difference between the respective kernels P_{γ_1} and P_{γ_2}, i.e. $\sup_{x\in\mathcal{X}} \|P_{\gamma_1}(x,\cdot) - P_{\gamma_2}(x,\cdot)\|$. Therefore we will define the metric $d(\gamma_1,\gamma_2)$ on $\mathcal{Y} \subset R^q$ as:

$$d(\gamma_1,\gamma_2) = \sup_{x\in\mathcal{X}} \|P_{\gamma_1}(x,\cdot) - P_{\gamma_2}(x,\cdot)\|$$

We suppose there exists a transition kernel P_γ corresponding to each $\gamma \in R^q$, and consider the following set:

$$\Delta = \{\gamma \in R^q |\ P_\gamma V \leq V - 1 + b1_C\}$$

Now we can state our main results as below:

Theorem 3.1. ***(Ergodicity Theorem)*** *Consider an adaptive MCMC algorithm with Diminishing Adaption, such that there is $C \in \mathcal{F}$, $V : \mathcal{X} \to [1, \infty)$ such that $\pi(V) < \infty$, $\delta > 0$, and $b < \infty$, with $\sup_C V = \nu < \infty$, and:*

(i) for each $\gamma \in \mathcal{Y}$, there exists a probability measure $\nu_\gamma(\cdot)$ on C with $P_\gamma(x, \cdot) \geq \delta \nu_\gamma(\cdot)$ for all $x \in C$; and

(ii)$P_\gamma V \leq V - 1 + b\mathbb{I}_C$ for each γ;

(iii)the set Δ is compact with respect to the metric d.

Suppose further that the sequence $\{V(X_n)\}_{n=0}^{\infty}$ is bounded in probability, given $X_0 = x_*$ and $\Gamma_0 = \gamma_*$. Then $\lim_{n\to\infty} T(x_*, \gamma_*, n) = 0$.

Usually we also want to estimate the integral $\pi(g) = \int_{\mathcal{X}} g(x)\pi(dx)$ of various functions $g : \mathcal{X} \to R$ using the laws of large numbers for ergodic averages of the form:

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i) \to_{n\to\infty} \pi(g) \text{ in probability or almost surely}$$

There are many references e.g Rosenthal and Tierney (1994) [18], Meyn and Tweedie (1993) [11] which give the proof and applications of the LLN of general Markov Chains. Regarding the LLN of AMCMC, there are also many papers e.g. Andieu and Achade (2005) [1], Andrieu and Moulines (2005) [2], Andrieu and Robert (2001)[3], Roberts and Rosenthal(2005) [14], Atchade and Rosenthal (2005) [15], C. Yang (2007) [19] giving the proof under various conditions. Especially, an counterexample was constructed in C. Yang (2007) [19] to show that the WLLN of AMCMC may NOT hold for unbounded measurable function even if the AMCMC is ergodic with respect to the target distribution. Here we will prove the WLLN of AMCMC for bounded function under the conditions of theorem 3.1.

Theorem 3.2. ***(WLLN)*** *Consider an adaptive MCMC algorithm. Suppose that the conditions of Theorem 3.1 hold. Let $g : \mathcal{X} \to R$ be a bounded measurable function. Then for any starting values $x \in \mathcal{X}$ and $\gamma \in \mathcal{Y}$, conditional on $X_0 = x$ and $\Gamma_0 = \gamma$ we have*

$$\frac{\sum_{i=1}^{n} g(X_i)}{n} \to \pi(g)$$

in probability as $n \to \infty$.

3.2 The Uniform Minimal Drift Condition

Intuitively, we hope the AMCMC is recurrent whenever each kernel is positive recurrent with respect to the target distribution π. However following the example below, we get the negative conclusion. Consider the following adaptive MCMC: suppose the state space $\mathcal{X} = \{1, 2\}$, the parameter space $\mathcal{Y} = \mathbb{N} \times \{1, 2\}$ with each kernel $P_{n,1} = \begin{pmatrix} 1 - \frac{1}{2^n} & \frac{1}{2^n} \\ \frac{1}{2^n} & 1 - \frac{1}{2^n} \end{pmatrix}$ and $P_{n,2} = \begin{pmatrix} \frac{1}{2^n} & 1 - \frac{1}{2^n} \\ 1 - \frac{1}{2^n} & \frac{1}{2^n} \end{pmatrix}$, and the stationary distribution $\pi(1) = \pi(2) = \frac{1}{2}$. We design an adaptive algorithm as:

$$\Gamma_n = \begin{cases} (n, 1) & \text{if } X_n = 1 \\ (n, 2) & \text{if } X_n = 2 \end{cases}$$

Lemma 3.1. *The above adaptive MCMC is NOT recurrent, although each kernel is positive recurrent with respect to the distribution* $\pi(\cdot)$. *Actually we have* $\mathbb{E}_2[\eta_2] < \infty$, *which means that the chain will NOT come back to* $\{2\}$ *after a long run when it starts from* $\{2\}$. *Therefore* $\lim_{n\to\infty} P(X_n = 2 | X_0 = i) = 0$ *for* $i = 1,\ 2$, *which is not equal to* $\pi(2)$.

Proof. Suppose $\eta_2 = \sum_{n=1}^{\infty} \mathbb{I}\{X_n = 2\}$. Then according to the adaptive algorithm, we have:

$$\begin{aligned} P_2(\eta_2 = n) &= \sum_{1 \le i_1 < i_2 \cdots < i_n < \infty} \frac{\Pi_{i=1}^{\infty}(1 - \frac{1}{2^i})}{\Pi_{j=1}^{n}(1 - \frac{1}{2^{i_j}})} \Pi_{j=1}^{n} \frac{1}{2^{i_j}} \\ &\le \sum_{1 \le i_1 < i_2 \cdots < i_n < \infty} \Pi_{j=1}^{n} \frac{1}{2^{i_j}} \\ &= \sum_{1 \le i_1 < i_2 \cdots < i_n < \infty} \frac{1}{2^{\sum_{j=1}^{n} i_j}} \\ &\le \sum_{m=\frac{n(n+1)}{2}}^{\infty} C_m^n \frac{1}{2^m} \\ &= \frac{1}{n!} \sum_{m=\frac{n(n+1)}{2}}^{\infty} m(m-1)\cdots(m-n+1)\frac{1}{2^m} \end{aligned}$$

Consider the functional series $S_n(x) = \sum_{m=\frac{n(n+1)}{2}}^{\infty} m(m-1)\cdots(m-n+1)x^m$ for

$0 < x < 1$, then we have:

$$\begin{aligned} S_n(x) &= x^n[\sum_{m=\frac{n(n+1)}{2}}^{\infty} x^m]^{(n)} \\ &= x^n[\frac{x^{\frac{n(n+1)}{2}}}{1-x}]^{(n)} \\ &= x^n \sum_{i=0}^{n} C_n^i \frac{(\frac{n(n+1)}{2})!}{(\frac{n(n+1)}{2}-i)!} x^{\frac{n(n+1)}{2}-i} i!(1-x)^{-i} \\ &\leq x^{\frac{n(n+1)}{2}} \times \sum_{i=0}^{n} C_n^i x^{n-i}(x-1)^{-i} \frac{(\frac{n(n+1)}{2})!}{(\frac{n(n+1)}{2}-n)!} n! \\ &\leq x^{\frac{n(n+1)}{2}} \times (x+\frac{1}{1-x})^n (\frac{n(n+1)}{2})^n n! \end{aligned}$$

Therefore we have:

$$\begin{aligned} P_2(\eta_2 = n) &\leq (\frac{1}{2})^{\frac{n(n+1)}{2}} \times (\frac{5}{2})^n \times (\frac{n(n+1)}{2})^n \\ &= \left[(\frac{1}{2})^{\frac{(n+1)}{2}} \times (\frac{5}{2}) \times (\frac{n(n+1)}{2})\right]^n \end{aligned}$$

We know that $\lim_{n\to\infty}(\frac{1}{2})^{\frac{(n+1)}{2}} \times (\frac{5}{2}) \times (\frac{n(n+1)}{2}) = 0$, i.e. there exists $N > 0$ such that for any $n > N$ we have $(\frac{1}{2})^{\frac{(n+1)}{2}} \times (\frac{5}{2}) \times (\frac{n(n+1)}{2}) < \frac{1}{2}$. So

$$\begin{aligned} E_2[\eta_2] &= \sum_{n=1}^{\infty} P_2(\eta_2 = n)n \\ &< \sum_{i=1}^{N} i + \sum_{i=N+1}^{\infty} i \times [\frac{1}{2}]^i \\ &< \infty \end{aligned}$$

Therefore the set $\{2\}$ is a transient set. Furthermore following that $\sum_{n=1}^{\infty} P_2(\eta_2 = n)n < \infty$, we know that $\lim_{n\to\infty} P(\eta_2 = n) = 0$, which is NOT equal to $\pi(2)$. $\square$

In the above example, we can ascribe the transience of the AMCMC to increasing of probability to $\{2\}$ as $n \to \infty$. Therefore we need the "uniform" recurrence property with respect to the parameter γ. Following the theorem 11.0.1 in Meyn and Tweedie [11], we know that an irreducible Markov chain is positive recurrent if and only if there exists some petite set C and some extend valued, non-negative test function V, which is finite for at least one state in the state space $\mathcal{X}$, satisfying:

$$PV(x) \leq V(x) - 1 + b\mathbb{I}_C(x), \quad x \in \mathcal{X}$$

Therefore we will suppose all the $\gamma \in \mathcal{Y}$ satisfy:

$$P_\gamma V(x) \leq V(x) - 1 + b\mathbb{I}_C(x), \quad x \in \mathcal{X}$$

4 The Proof of Ergodicity Theorem

Before we prove the theorem 3.1, let us think about the following lemma:

Lemma 4.1. *Consider an adaptive MCMC algorithm with Diminishing Adaptation, with a regular stationary measure π and an accessible atom $\alpha \in \mathcal{F}$ such that $P_\gamma(x, B) = \nu_\gamma(B)$ for any $x \in \alpha$ and $B \in \mathcal{B}(\mathcal{X})$, where $\nu_\gamma(\cdot)$ is a regular probability measure, let measurable function $W : \mathcal{X} \to [0, \infty)$, $0 < K < \infty$*
(i) $E_{\alpha,\gamma}[\tau_\alpha] \leq K$ and $E_{x,\gamma}[\tau_\alpha] \leq W(x)$ for any $x \in \alpha^c$ and $\gamma \in \mathcal{Y}$.
(ii) The parameter space $\mathcal{Y}$ is compact with respect to the metric d of the set Δ.
Suppose further that the sequence $\{W(X_n)\}_{n=0}^\infty$ is bounded in probability, given $X_0 = x_*$ and $\Gamma_0 = \gamma_*$. Then we have:

$$\lim_{n\to\infty} T(x_*, y_*, n) = 0$$

We will prove the lemma in section 4.3 after some technical preparations.

4.1 The Splitting Chain

To prove the above lemma we need a useful technique-splitting the chain, see Chapter 5 in Meyn and Tweedie (1993) [11]. Before we construct a splitting chain, we need to introduce the definition of atom and the Minorizaiton condition first:

Atoms: A set $\alpha \in \mathcal{B}(\mathcal{X})$ is called an atom for the Markov chain $\{X_n\}$ if there exists a measure ν on $\mathcal{B}(\mathcal{X})$ such that:

$$P(x, A) = \nu(A), \quad x \in \alpha.$$

If the chain $\{X_n\}$ is $\psi-$irreducible and $\psi(\alpha) > 0$, then α is called an accessible atom.

Minorization Condition: For some $\delta > 0$, some $C \in \mathcal{B}(\mathcal{X})$ and some probability measure ν with $\nu(C^c) = 0$ and $\nu(C) = 1$, $P(x, A) \geq \delta\mathbb{I}_C(x)\nu(A)$.

Consider a Markov chain with minorization condition, we can split chain. We first split the space $\mathcal{X}$ itself by writing $\check{\mathcal{X}} = \mathcal{X} \times \{0,1\}$, where $\mathcal{X}_0 = \mathcal{X} \times \{0\}$ and $\mathcal{X}_1 = \mathcal{X} \times \{1\}$ are thought of as copies $\mathcal{X}$ equipped with copies $\mathcal{B}(\mathcal{X}_0)$, $\mathcal{B}(\mathcal{X}_1)$ of the $\sigma-$ field $\mathcal{B}(\mathcal{X})$. We also let $\mathcal{B}(\check{\mathcal{X}})$ be the $\sigma-$field of $\check{\mathcal{X}}$ generated by $\mathcal{B}(\mathcal{X}_0)$, $\mathcal{B}(\mathcal{X}_1)$: that is $\mathcal{B}(\check{\mathcal{X}})$ is the smallest $\sigma-$field containing sets of the form $A_0 := A \times \{0\}$, $A_1 := A \times \{1\}$, $A \in \mathcal{B}(\mathcal{X})$.

We will write x_i, $i = 0, 1$ for elements of $\check{\mathcal{X}}$, with x_0 denoting members of the upper level $\mathcal{X}_0$ and x_1 denoting members of the lower level $\mathcal{X}_1$.

If λ is any measure on $\mathcal{B}(\mathcal{X})$, then the next step in the construction is to split the measure λ into two measures on each of $\mathcal{X}_0$ and $\mathcal{X}_1$ by defining the measure λ^* on $\mathcal{B}(\check{\mathcal{X}})$ through:

$$\begin{aligned} \lambda^*(A_0) &= \lambda(A \cap C)[1-\delta] + \lambda(A \cap C^c) \\ \lambda^*(A_1) &= \lambda(A \cap C)\delta \end{aligned}$$

Now we can step in the construction to the split the chain $\{X_n\}$ to the form a chain $\{\check{X}_n\}$ which lives on $(\check{\mathcal{X}}, \mathcal{B}(\check{X})$. Define the split kernel $\check{P}(x_i, A)$ for $x_i \in \check{\mathcal{X}}$ and $A \in \mathcal{B}(\check{X})$ by:

$$\begin{aligned} \check{P}(x_0, \cdot) &= P(x, \cdot)^*, \quad x_0 \in \mathcal{X}_0 - C_0; \\ \check{P}(x_0, \cdot) &= [1-\delta]^{-1}[P(x, \cdot)^* - \delta\nu^*(\cdot)], \quad x_0 \in C_0; \\ \check{P}(x_1, \cdot) &= \nu^*(\cdot), \quad x_1 \in \mathcal{X}_1 \end{aligned}$$

where C, δ and ν are the set, the constant and the measure in the Minorization Condition. We can see that outside C the chain $\{\check{X}_n\}$ behaves like $\{X_n\}$, moving on the "top" half $\mathcal{X}_0$ of the split space. Each time it arrives in C, it is "split"; with probability $1-\delta$ it remains in C_0, with probability δ it drops to C_1.

It is critical to note that the bottom level $\mathcal{X}_1$ is an atom with $\psi^*(X_1) == \delta\psi(C) > 0$ whenever the original chain is $\psi-$irreducible. We also have $\check{P}^n(x_i, \mathcal{X}_\infty - C_1) = 0$ for all $n \geq 1$ and all $x_i \in \check{\mathcal{X}}$, so that the atom $C_1 \subseteq \mathcal{X}_1$ is the only part of the bottom level which is reached with positive probability. We will use the notation $\check{\alpha} := C_1$ when we wish to emphasize the fact that all transitions out of C_1 are identical, so that C_1 is an atom in $\check{\mathcal{X}}$.

4.2 The Proof Of Lemma 4.1

For any initial value $x \in \mathcal{X}$ and measurable function $|f| \leq 1$, denote: $a_{x,\gamma}(n) = P_{x,\gamma}(\tau_\alpha = n)$, that is the first hitting time of α is n when the kernel is P_γ and the start value is x; similarly denote $u_\gamma(n) = (P_\gamma)_\alpha(\Phi_n \in \alpha)$ and define:

$$t_{f,\gamma}(n) = \int_\alpha P_\gamma^n(\alpha, dy)f(y) = (E_\gamma)_\alpha[f(\Phi_n)1\{\tau_{\alpha \geq n}\}]$$

Then following the first-entrance last-exit decomposition we have:

$$P_\gamma^n(x,B) =_\alpha P_\gamma^n(x,B) + \sum_{j=1}^{n-1}[\sum_{k=1}^{j} {}_\alpha P_\gamma^k(x,\alpha)P^{j-k}(\alpha,\alpha)]_\alpha P_\gamma^{n-j}(\alpha,B)$$

where ${}_\alpha P_\gamma^{n-j}(\alpha,B)$ is the taboo probability given by

$${}_\alpha P_\gamma^{n-j}(\alpha,B) = P_\gamma(X_{n-i} \in B, \tau_\alpha \geq n-j|X_0 \in \alpha)$$

Therefore for any $x \in \mathcal{X}$ and f, we have:

$$\int P_\gamma^n(x,d\omega)f(\omega) = \int {}_\alpha P_\gamma^n(x,d\omega)f(\omega) + a_{x,\gamma} * u_\gamma * t_{f,\gamma}(n)$$

then we will get:

$$\begin{aligned}
|E_{x,\gamma}[f(\Phi_n)] - E_\pi[f(\Phi_n)]| &\leq E_{x,\gamma}[f(\Phi_n)\mathbb{I}\{\tau_{\alpha \geq n}\}] \\
&+ |a_{x,\gamma} * u_\gamma - \pi(\alpha)| * t_{f,\gamma}(n) \\
&+ \pi(\alpha)\sum_{j=n+1}^{\infty} t_{f,\gamma}(j) \\
&\leq E_{x,\gamma}[f(\Phi_n)\mathbb{I}\{\tau_{\alpha \geq n}\}] + \sum_{j=1}^{n}|\sum_{i=1}^{j} a_x(j)u(j-i) - \pi(\alpha)|t_1(n-j) \\
&+ \pi(\alpha)\sum_{j=n+1}^{\infty} t_{f,\gamma}(j) \\
&\leq E_{x,\gamma}[f(\Phi_n)\mathbb{I}\{\tau_{\alpha \geq n}\}] + \sum_{j=1}^{n}\sum_{i=1}^{j} a_x|(j)u(j-i) - \pi(\alpha)|t_1(n-j) \\
&+ \pi(\alpha)\sum_{j=n+1}^{\infty} t_{f,\gamma}(j) \\
&\leq E_{x,\gamma}[f(\Phi_n)\mathbb{I}\{\tau_{\alpha \geq n}\}] + \sum_{j=1}^{n}\sum_{i=1}^{j} a_x(i)|u(j-i) - \pi(\alpha)|t_1(n-j) \\
&+ \pi(\alpha)\sum_{j=1}^{n}\sum_{i=j+1}^{\infty} a_x(i)t_1(n-j) + \pi(\alpha)\sum_{j=n+1}^{\infty} t_{1,\gamma}(j)
\end{aligned}$$

Now we can denote the first term as I, the second as II, the third as III and the fourth term as IV. And we have the following estimations.

4.2.1 The Estimation Of I and III

Lemma 4.2. *$I \leq \frac{W(x)}{n}$ for any $x \in \mathcal{X}$.*

Proof.

$$\begin{aligned} I &\leq E_{x,\gamma}[1_{\tau_\alpha \geq n}] \\ &= P_{x,\gamma}(\tau_\alpha \geq n) \\ &\leq \frac{E_{x,\gamma}(\tau_\alpha)}{n} \\ &\leq \frac{W(x)}{n} \end{aligned}$$

□

Lemma 4.3. *Let $a_n = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{i}$, then $III \leq 2a_n KW(x)$ for any $x \in \mathcal{X}$.*

Proof.

$$\begin{aligned} III &\leq \sum_{j=1}^{n} P_x(\tau_\alpha \geq j) P_\alpha(\tau_\alpha \geq n-j) \\ &\leq \sum_{j=1}^{n} \frac{W(x)}{j} \times \frac{K}{n-j} \\ &= KW(x)\frac{2}{n}\sum_{i=1}^{n}\frac{1}{i} \\ &= 2Ka_nW(x) \end{aligned}$$

And we know that $\lim_{n\to\infty} a_n = 0$. □

4.2.2 The Estimation Of Term IV

Following the structure of stationary distribution π, we know that:

$$\sum_{j=1}^{\infty} P_{\alpha,\gamma}(\tau_\alpha > j) = \frac{1}{\pi(\alpha)} = M$$

so for any $\epsilon > o$, there exists N_γ, such that for any $n_\gamma > N_\gamma$:

$$\sum_{j=1}^{n_\gamma} P_{\alpha,\gamma}(\tau_\alpha) > M - \epsilon$$

We define $n_\epsilon(\gamma) = \inf\{n : \sum_{j=1}^{n} P_{\alpha,\gamma}(\tau_\alpha > j) > M - \epsilon\}$, and prove that:

Lemma 4.4. *For any fixed γ_0, there exists $\delta > 0$ such that for any $d(\gamma, \gamma_0) < \delta$, we have $n_\epsilon(\gamma) = n_\epsilon(\gamma_0)$.*

Proof. Denote $\eta_1 = \sum_{j=1}^{n_{\gamma_0}} P_{\alpha,\gamma_0}(\tau_\alpha > j) - (M-\epsilon)$ and $\eta_2 = M-\epsilon-\sum_{j=1}^{n_{\gamma_0}-1} P_{\alpha,\gamma_0}(\tau_\alpha > j)$. Set $\delta = \frac{2\min\{\eta_1,\eta_2\}}{n_\epsilon(\gamma_0)(n_\epsilon(\gamma_0)+1)}$, then consider two Markov chain $\{X_i\}$ with kernel P_{γ_0} and $\{X_i^{'}\}$ with kernel P_{γ_1} such that $d(\gamma_0, \gamma_1) < \delta$. Then

$$\begin{aligned}
\mathbb{P}_x(X_i \neq X_i^{'} | X_{i-1} = X_{i-1}^{'}) &= \mathbb{E}(P_x(X_i \neq X_i^{'} | X_{i-1} = X_{i-1}^{'}, X_{i-1} = y)) \\
&\le \mathbb{E}(P(X_i \neq X_i^{'} | X_{i-1} = X_{i-1}^{'} = y)) \\
&= \mathbb{E}(\|P_{\gamma_0}(y,\cdot) - P_{\gamma_1}(y,\cdot)\|) \\
&\le \mathbb{E}(d(\gamma_0, \gamma_1)) \\
&< \delta.
\end{aligned}$$

The third equation $P(X_i \neq X_i^{'} | X_{i-1} = X_{i-1}^{'} = y) = \|P_{\gamma_0}(y,\cdot) - P_{\gamma_1}(y,\cdot)\|$ is following the Proposition 3(g) in [**?**]. Then we have

$$P_x(X_i \neq X_i^{'}, X_{i-1} = X_{i-1}^{'}) = P_x(X_i \neq X_i^{'} | X_{i-1} = X_{i-1}^{'}) P_x(X_{i-1} = X_{i-1}^{'}) \le \delta$$

With the same start value $x \in \alpha$, then we have:

$$\begin{aligned}
\mathbb{P}(X_i \neq X_i^{'} | X_0 = X_0^{'} = x) &= P_x(X_i \neq X_i^{'}, X_{i-1} \neq X_{i-1}^{'}) + P_x(X_i \neq X_i^{'}, X_{i-1} = X_{i-1}^{'}) \\
&\le P_x(X_{i-1} \neq X_{i-1}^{'}) + \delta \\
&\le P_x(X_{i-1} \neq X_{i-1}^{'}, X_{i-2} \neq X_{i-2}^{'}) + P_x(X_{i-1} \neq X_{i-1}^{'}, X_{i-2} = X_{i-2}^{'}) + \delta \\
&\le P_x(X_{i-2} \neq X_{i-2}^{'}) + 2\delta \\
&\le \cdots \\
&\le i\delta.
\end{aligned}$$

Therefore:

$$\sum_{i=1}^{n_{\gamma_0}(\epsilon)} P_x(X_i \neq X_i^{'}) \le \sum_{i=1}^{n_{\gamma_0}(\epsilon)} i\delta \le \min\{\eta_1, \eta_2\}.$$

So we still have $n_\epsilon(\gamma) = n_\epsilon(\gamma_0)$. □

Lemma 4.5. *For any $\epsilon > 0$, there exits $N > 0$ which is independent with γ, such that for any $n > N$, we have: $IV < \epsilon$*

Proof. Since

$$\begin{aligned} IV &\leq \pi(\alpha) \sum_{j=n+1}^{\infty} t_{1,\gamma}(j) \\ &= \pi(\alpha) \sum_{j=n+1}^{\infty} E_{\alpha,\gamma}[1_{\tau_\alpha \geq j}] \\ &= \pi(\alpha) \sum_{j=n+1}^{\infty} P_{\alpha,\gamma}(\tau_\alpha > j) \end{aligned}$$

following lemma 4.4, we know that for any $\epsilon > 0$, there exits $N > 0$ which is independent with γ, such that for any $n > N$, we have: $\sum_{j=n+1}^{\infty} P_{\alpha,\gamma}(\tau_\alpha > j) < \frac{\epsilon}{\pi(\alpha)}$. That is $IV \leq \epsilon$ for any $n > N$. □

4.2.3 The Estimation On Term II

Lemma 4.6. *For any $\epsilon > 0$, there exists $N > 0$ which is independent with γ such that $II \leq \epsilon W(x)$.*

$$\begin{aligned} II &\leq \sum_{j=1}^{n} t_{1,\gamma}(n-j) \sum_{i=1}^{j} a_{x,\gamma}(i) i \frac{|u_\gamma(j-i) - \pi(\alpha)|}{i} \\ &\leq \sum_{j=1}^{n} t_{1,\gamma}(n-j) [\sum_{i=1}^{\infty} a_{x,\gamma}(i) i] \sum_{i=1}^{j} \frac{|u(j-i) - \pi(\alpha)|}{i} \\ &\leq \sum_{j=1}^{n} t_{1,\gamma}(n-j) E_{x,\gamma}(\tau_\alpha) \sum_{i=1}^{j} \frac{|u(j-i) - \pi(\alpha)|}{i} \\ &\leq W(x) \sum_{j=1}^{n} t_{1,\gamma}(n-j) \sum_{i=1}^{j} \frac{|u_\gamma(j-i) - \pi(\alpha)|}{i} \end{aligned}$$

Lemma 4.7. $\sum_{i=1}^{\infty} |u_\gamma(i) - \pi(\alpha)| < \infty$ *for each γ.*

Proof. Since $\sup_{\tilde{C}} V(x) = v$ and ν_γ is probability measure on α, $\int_{\mathcal{X}} V(x)\nu_\gamma(dx) < \infty$ and $\pi(V) < \infty$, following Theorem 11.3.12 of Meyn and Tweedie's book, we know that

ν_γ and $\pi(\cdot)$ are both regular measure. Then following Theorem 13.4.5 in Meyn and Tweedie's book, we know that:

$$\sum_{n=1}^{\infty} \|\nu_\gamma P_\gamma^n - \pi\| < \infty$$

Therefore we have $\sum_{n=1}^{\infty} \|P_\gamma^n(\alpha,\alpha) - \pi(\alpha)\| < \infty$ □

Lemma 4.8. $\lim_{n\to\infty} \sum_{j=1}^{n} t_{1,\gamma}(n-j) \sum_{i=1}^{j} \frac{|u_\gamma(j-i)-\pi(\alpha)|}{i} = 0$ *for any* $\gamma \in \mathcal{Y}$.

Proof. Let $s_j(\gamma) = \sum_{i=1}^{j} \frac{|u_\gamma(j-i)-\pi(\alpha)|}{i}$, following bounded convergence theorem and lemma 4.7, we have $s_j(\gamma) \to_{j\to\infty} 0$. Similarly following $\sum_{j=1}^{\infty} t_{1,\gamma}(j) = E_{\gamma,\alpha}(\tau_\alpha) \leq v < \infty$, we have $\lim_{n\to\infty} \sum_{j=1}^{n} t_{1,\gamma}(n-j) \sum_{i=1}^{j} \frac{|u_\gamma(j-i)-\pi(\alpha)|}{i} = 0$ □

Lemma 4.9. *For any* $\epsilon > 0$ *there exists* N *which is independent with* γ*, such that for any* $n > N$*, we have* $\sum_{j=1}^{n} t_{1,\gamma}(n-j) \sum_{i=1}^{j} \frac{|u_\gamma(j-i)-\pi(\alpha)|}{i} < \epsilon$.

Proof. Suppose there exist $\epsilon > 0$, and strictly increasing $\{n_i\}_{i=1}^{\infty}$ and $\gamma_{n_i} \in \mathcal{Y}$ such that $\sum_{j=1}^{n_i} t_{1,\gamma_{n_i}}(n-j) \sum_{i=1}^{j} \frac{|u_{\gamma_{n_i}}(j-i)-\pi(\alpha)|}{i} > \epsilon$. Then there exists γ_0 such that $\gamma_{n_i} \to \gamma_0$. Therefore we have:

$$\sum_{j=1}^{\infty} t_{1,\gamma_0}(n-j) \sum_{i=1}^{j} \frac{|u_{\gamma_0}(j-i)-\pi(\alpha)|}{i} > \epsilon$$

Contradiction!! □

From all above we have the following lemma:

Lemma 4.10. *For any* $\epsilon > 0$*, there exists* $N > 0$ *which is independent with the choice of* γ*, such that for any* $n > N$*, we have:*

$$\|P_\gamma^n(x,\cdot) - \pi(\cdot)\| \quad \leq \quad \frac{W(x)}{n} + \epsilon W(x) + \epsilon$$

4.2.4 The Proof Of Lemma 4.1

Proof. Let $M_\epsilon(x,\gamma) = \inf\{n \geq 1 : \|P_\gamma^n(x,\cdot) - \pi(\cdot)\| \leq \epsilon\}$. Then following the theorem 13 in Roberts and Rosenthal's paper [14] (2007), it suffices to prove that $\{M_\epsilon(X_n,\Gamma_n)\}_{n=0}^{\infty}$ is bounded in probability given $X_0 = x_*$ and $\Gamma_0 = \gamma_*$, i.e. for all $\delta > 0$, there is $N \in \mathbb{N}$ such that:

$$P[M_\epsilon(X_n,\Gamma_n) \leq N | X_0 = x_*, \Gamma_0 = \gamma_*] \geq 1 - \delta$$

Since for any $\epsilon > 0$, there exists $N > 0$ which is independent with the choice of γ, such that for any $n > N$, we have:

$$\|P_\gamma^n(x,\cdot) - \pi(\cdot)\| \leq \epsilon W(x) + \epsilon$$

and $W(X_n)$ is bounded in probability, we have the conclusion hold. □

4.3 The Proof Of Theorem 3.1

Proof. Consider the splitting chain $\{\check{X}_n^\gamma\}$, we know that the subset $\alpha = C_1 \in \check{\mathcal{X}}$ is an accessible atom of any chain $\{X_n^\gamma\}$. Before we prove the above inequalities, let us recall what the splitting chain is. Actually outside C the chain $\{\check{X}_n^\gamma\}$ behaves just like $\{X_n^\gamma\}$, moving on the "top" half $\mathcal{X}_0$ of the split space. Each time it arrives in C, it is "split"; with probability $1-\delta$ it remain in C_0, with probability δ it drops to C_1. We can prove the Theorem as following steps:

Step 1: Prove that there exists $K > 0$ such that

$$E_{\alpha,\gamma}(\tau_\alpha) \leq K;$$

Step 2: Prove that there exists a measurable function $W : \check{\mathcal{X}} \to [0,\infty)$ such that:

$$E_{x,\gamma}(\tau_\alpha) \leq W(x);$$

Step 3: Check the regularity of ν_γ and π.

Suppose $\check{\tau}_{A,\gamma}^{(m)}(B)$ is the $m-$th hitting time of B from A and with the kernel $\check{P}_\gamma$. Consider the random variable $\check{\tau}_{\alpha,\gamma}(\alpha)$, then $\check{\tau}_{\alpha,\gamma}(\alpha) = \check{\tau}_{\alpha,\gamma}(\check{C}) + \check{\tau}_{\check{C},\gamma}^{(k-1)}(\check{C})$with probability $(1-\delta)^{k-1}\delta$. If we denote the random variable $T =$ the number of $\{n \leq \check{\tau}_{\alpha,\gamma}(\alpha) | \check{X}_n \in \check{C}\}$, where $\check{C} = C_0 \cup C_1$, we have:

$$\begin{aligned} E_{\alpha,\gamma}(\tau_\alpha) &= E[E(\check{\tau}_{\alpha,\gamma}(\alpha)|T)] \\ &= \sum_{k=1}^{\infty} \left(E(\check{\tau}_{\alpha,\gamma}(\check{C})) + (k-1)E_{\check{C},\gamma}(\tau_{\check{C}}) \right) (1-\delta)^{k-1}\delta \\ &= E(\check{\tau}_{\alpha,\gamma}(\check{C})) + \frac{1-\delta}{\delta} E_{\check{C},\gamma}(\tau_{\check{C}}) \end{aligned}$$

and we also know that for any $x \in \check{C}, \gamma \in \mathcal{Y}$ $E[\check{\tau}_{x,\gamma}(\check{C})] = E_{x,\gamma}(\tau_C) \leq V(x) + b \leq v + b = K$. Therefore $E_{\alpha,\gamma}(\tau_\alpha) \leq K + \frac{1-\delta}{\delta}K = \frac{K}{\delta}$.

Similarly for any $x \notin \alpha$, we know that $\check{\tau}_{x,\gamma}(\alpha) = \check{\tau}_x(\check{C}) + \check{\tau}^{(k-1)}_{\check{C},\gamma}(\check{C})$with probability $(1-\delta)^{k-1}\delta$. Therefore we have:

$$\begin{aligned}
E_{x,\gamma}(\tau_\alpha) &= E[E(\check{\tau}_{x,\gamma}(\alpha)|T)] \\
&= \sum_{k=1}^{\infty}\Big(E(\check{\tau}_{x,\gamma}(\check{C})) + (k-1)E_{\check{C},\gamma}(\tau_{\check{C}})\Big)(1-\delta)^{k-1}\delta \\
&= E(\check{\tau}_{x,\gamma}(\check{C})) + \frac{1-\delta}{\delta}E_{\check{C},\gamma}(\tau_{\check{C}})
\end{aligned}$$

and we also have for any $x,\gamma \in \mathcal{Y}$, $E[\check{\tau}_{x,\gamma}(\check{C})] = E_{x,\gamma}(\tau_C) \le V(x) + b = W(x)$. Since $V(X_n)$ is bounded in probability, $W(X_n)$ is also bounded in probability.

Finally since $\int_{\mathcal{X}} V(y)\nu_\gamma(dy) < v$ and $\pi(V) < \infty$,the probability measures ν_γ and π are both regular. Then we can prove the theorem 3.1 following the lemma 4.1 □

5 The Proof Of The WLLN

Similar to the proof of theorem 3.1, it suffices to prove the following lemma before we prove the theorem 3.2:

Lemma 5.1. *Under the conditions of lemma 4.1. Let $g : \mathcal{X} \to R$ be a bounded measurable function. Then for any starting values $x \in \mathcal{X}$ and $\gamma \in \mathcal{Y}$, conditional on $X_0 = x$ and $\Gamma_0 = \gamma$ we have*

$$\frac{\sum_{i=1}^n g(X_i)}{n} \to \pi(g)$$

in probability as $n \to \infty$.

5.1 Some Technical Results

Following the usual laws of large numbers for Markov chain (see e.g. Meyn and Tweedie) imply that for each fixed $x \in \mathcal{X}$ and $\gamma \in \mathcal{Y}$, $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n g(X_i^\gamma) \to \pi(g)$ in probability, where $\{X_n^\gamma\}$ is the usual Markow chain with kernel P_γ. Actually we will prove that under the conditions in lemma 5.1 the above convergence is uniformly with respect to the parameter γ. Before we start the proof, let us define some symbols, let

$$s_i^\gamma(g) = \sum_{j=\tau_{\check{\alpha}}(i)+1}^{\tau_{\check{\alpha}}(i+1)} g(X_j^\gamma)$$

and

$$l_n^\gamma = \max\{i \geq 0 : \tau_{\check{\alpha}}(i) \leq n\}$$

Lemma 5.2. *Under the conditions of lemma 5.1, for any $\epsilon > 0$ and fixed start value x, there exists N which is independent with the choice of γ such that for any $n > N$ we have:*

$$P_x(|\frac{\sum_{i=1}^n g(X_i^\gamma)}{n} - \pi(g)| > \epsilon) < \epsilon W(x) + \epsilon$$

Proof. Without losing generalities, we suppose $\pi(g) = 0$ and $|g(x)| \leq M$ then

$$\begin{aligned} & P_x(|\frac{\sum_{i=1}^n g(X_i^\gamma)}{n}| > 3\epsilon) \\ = \quad & P_x(|\frac{\sum_{i=1}^{\tau_{\check{\alpha}}} g(X_i^\gamma)}{n} + \frac{\sum_{i=0}^{l_n} s_i(g)}{n} + \frac{\sum_{i=\tau_{\check{\alpha}}(l_n)+1}^n g(X_i^\gamma)}{n}| > 3\epsilon) \\ \leq \quad & P_x(|\frac{\sum_{i=1}^{\tau_{\check{\alpha}}} g(X_i^\gamma)}{n}| > \epsilon) + P_x(|\frac{\sum_{i=0}^{l_n} s_i^\gamma(g)}{n}| > \epsilon) + P_x(|\frac{\sum_{i=\tau_{\check{\alpha}}(l_n)+1}^n g(X_i^\gamma)}{n}| > \epsilon) \end{aligned}$$

Regarding the first term we have:

$$\begin{aligned} P_x(|\frac{\sum_{i=1}^{\tau_{\check{\alpha}}} g(X_i^\gamma)}{n}| > \epsilon) \quad & \leq \quad \frac{E_x[|\sum_{i=1}^{\tau_{\check{\alpha}}} g(X_i^\gamma)|]}{n\epsilon} \\ & \leq \quad \frac{E_x[\tau_{\check{\alpha}}]M}{n\epsilon} \\ & \leq \quad \frac{W(x)M}{n\epsilon} \end{aligned}$$

Regarding the third term we have:

$$\begin{aligned} P_x(|\frac{\sum_{i=\tau_{\check{\alpha}}(l_n)+1}^n g(X_i^\gamma)}{n}| > \epsilon) \quad & \leq \quad \frac{E_{\check{\alpha}}[|\sum_{i=\tau_{\check{\alpha}}(l_n)+1}^n g(X_i^\gamma)|]}{n\epsilon} \\ & \leq \quad \frac{E_{\check{\alpha}}[\tau_{\check{\alpha}}]M}{n\epsilon} \\ & \leq \quad \frac{KM}{n\epsilon} \end{aligned}$$

Actually the second term is independent with the choice of start value x,i.e.

$$P_x(|\frac{\sum_{i=0}^{l_n} s_i^\gamma(g)}{n}| > \epsilon) = P_{\check{\alpha}}(|\frac{\sum_{i=0}^{l_n} s_i^\gamma(g)}{n}| > \epsilon)$$

Suppose for any $n \in \mathbb{N}$, there exists γ_n such that $P_{\check{\alpha}}(|\frac{\sum_{i=0}^{l_n} s_i^{\gamma_n}(g)}{n}| > \epsilon) > \frac{\epsilon}{2}$, same as the proof of lemma 4.4, we can find some $\gamma_0 \in \Delta$ such that:

$$\lim_{n\to\infty} P_{\check{\alpha}}(|\frac{\sum_{i=0}^{l_n} s_i^{\gamma_n}(g)}{n}| > \epsilon) > \frac{\epsilon}{2}$$

Which is conflicting with the fact that for any $\gamma \in \Delta$ and $\epsilon > 0$, we have:

$$\lim_{n\to\infty} P_{\check{\alpha}}(|\frac{\sum_{i=0}^{l_n} s_i^{\gamma_n}(g)}{n}| > \epsilon) = \pi(g) = 0$$

Therefore there exists N_1, such that for any $n > N_1$ and γ, we have:

$$P_x(|\frac{\sum_{i=0}^{l_n} s_i^{\gamma}(g)}{n}| > \epsilon) < \frac{\epsilon}{2}$$

We also can find N_2 such that for any $n > N_2$ we have $\frac{M}{n} < \epsilon^2$ and $\frac{KM}{n} < \frac{\epsilon^2}{2}$. Then let $N = \max\{N_1, N_2\}$ we can get the conclusion. □

Lemma 5.3. *Given $\epsilon > 0$, we can find $N > 0$ such that when $n > N$ we have:*

$$E_{\gamma,x}[|\frac{\sum_{i=1}^{N} g(X_i)}{N}|] \leq \epsilon W(x) + \epsilon$$

Proof. Following lemma 5.2, we know that for any $\epsilon > 0$, there exists N such that:

$$P_x(|\frac{\sum_{i=1}^{n} g(X_i^{\gamma})}{n}| > \epsilon) < \frac{\epsilon}{M} W(x) + \frac{\epsilon}{2M}$$

We also have $|\frac{\sum_{i=1}^{n} g(X_i^{\gamma})}{n}| \leq M$. If we denote $\Lambda = \{\omega \in \Omega || \frac{\sum_{i=1}^{n} g(X_i^{\gamma})}{n}| > \frac{\epsilon}{2} \text{ given } X_0 = x\}$. Then we have:

$$\begin{aligned} E_{\gamma,x}[|\frac{\sum_{i=1}^{N} g(X_i)}{N}|] &= E_{\gamma,x}[|\frac{\sum_{i=1}^{N} g(X_i)}{N}| \times \mathbb{I}_\omega(\Lambda)] + E_{\gamma,x}[|\frac{\sum_{i=1}^{N} g(X_i)}{N}| \times \mathbb{I}_\omega(\Lambda^c)] \\ &\leq M[W(x)\frac{\epsilon}{M} + \frac{\epsilon}{2M}] + \frac{\epsilon}{2} \\ &\leq \epsilon W(x) + \epsilon \end{aligned}$$

□

5.2 The Proof Of Theorem 3.2

First we can prove the Lemma 5.1:

Proof. Given starting value $X_0 = x$, $\Gamma_0 = \gamma$ and $\epsilon > 0$, $W(X_n)$ is bounded in probability, i.e. for any $\epsilon > 0$, there exists $a > 0$ such that:

$$P(W(X_n) > a) < \frac{\epsilon}{4M} \text{ for all } n \in \mathbb{N}$$

Following lemma 5.3, we know that there exists $N = N(\epsilon)$, such that for any x and γ we have:

$$E_{\gamma,x}[|\frac{\sum_{i=1}^{N} g(X_i)}{N}|] \leq \frac{\epsilon W(x)}{4a} + \frac{\epsilon}{4}$$

Then let $D_n = \sup_{x\in\mathcal{X}} \|P_{\Gamma_{n+1}}(x,\cdot) - P_{\Gamma_n}(x,\cdot)\|$ and $H_n = D_n \geq \frac{\epsilon}{4MN^2}$. Using the Diminishing Adaptation condition to choose $n^* = n^*(\epsilon) \in \mathbb{N}$ large enough so that

$$P(H_n) \leq \frac{\epsilon}{4NM}, \quad n \leq n^*$$

To continue, fix a "target time" $K \geq n^*+N$. We shall construct a coupling which depends on the target time K (cf. Roberts and Rosenthal, 2002), to prove that $\mathcal{L}(X_k) \approx \pi(\cdot)$. Define the event $E = \cap_{i=n+1}^{n+N} H_i^c$, we have $P(E) \geq 1 - \frac{\epsilon}{4M}$. Now, it follows from the triangle inequality and induction that on the event E, we have:

$$\sup_{x\in\mathcal{X}} \|P_{\Gamma_{n+k}}(x,\cdot) - P_{\Gamma_n}(x,\cdot)\| < \frac{\epsilon}{4MN}, \quad k \leq N.$$

In particular, on E we have $\|P_{\Gamma_{L-N}}(x,\cdot) - P_{\Gamma_m}(x,\cdot)\| < \frac{\epsilon}{4MN}$ for all $x \in \mathcal{X}$ and $L-N \leq m \leq L$, so by induction again,

$$\|P^N_{\Gamma_{L-N}}(x,\cdot) - P_{\Gamma_n}(X_k \in \cdot|X_{L-N} = x, \mathcal{G}_{L-N})\| < \frac{\epsilon}{4M} \text{ on E, for } x \in \mathcal{X}.$$

To construct the coupling, first construct the original adaptive chain $\{X_n\}$ together with its adaption sequence $\{\Gamma_n\}$, starting with $X_0 = x$ and $\Gamma_0 = \gamma$.
We now claim that on E, we can construct a second chain $\{X_n^{'}\}_{n=L-N}^{L}$ such that $X_{L-N}^{'} = X_{L-N}$ and $X_n^{'}\tilde{P}_{\Gamma_{L-N}}(X_{n-1}^{'},\cdot)$ for $L-N+1 \leq n \leq L$, and such that $P(X_L^{'} \neq X_L) < \epsilon$. Indeed, conditional on $\mathcal{G}_{L-N}$, we have $X_L^{'}\tilde{P}^N_{\Gamma_{L-N}}(X_{L-N},\cdot)$. Then we have:

$$\|\mathcal{L}(X_k^{'}) - \mathcal{L}(X_k)\| < \frac{\epsilon}{4M}$$

The claim then follows from e.g. Roberts and Rosenthal (2004, Proposition 3(g)).
Since $|g| \leq M$, we have:

$$\begin{aligned} E\left(\frac{1}{N}|\sum_{i=n+1}^{n+N} g(X_i)|\mathcal{G}_n\right) &\leq E_{\Gamma_n,X_n}\left(\frac{1}{N}|\sum_{i=1}^{N} g(X_i)|\right) + M\frac{\epsilon}{4M} + MP(E^c) \\ &\leq \frac{\epsilon W(X_n)}{4a} + \frac{\epsilon}{2} \end{aligned}$$

and we also have:

$$E\left(\frac{1}{N}|\sum_{i=n+1}^{n+N} g(X_i)|\mathcal{G}_n\right) \quad \leq \quad M$$

Therefore,

$$\begin{aligned}
& E((\frac{1}{N}|\sum_{i=n+1}^{n+N} g(X_i)) \\
= \; & E\left(E(\frac{1}{N}|\sum_{i=n+1}^{n+N} g(X_i)|\mathcal{G}_n)\right) \\
= \; & E\left(E(\frac{1}{N}|\sum_{i=n+1}^{n+N} g(X_i)|\mathcal{G}_n, W(X_n) \leq a)\right) + E\left(E(\frac{1}{N}|\sum_{i=n+1}^{n+N} g(X_i)|\mathcal{G}_n, W(X_n) > a)\right) \\
\leq \; & \frac{\epsilon}{2} + \frac{\epsilon}{4} + M\frac{\epsilon}{4M} \\
= \; & \epsilon
\end{aligned}$$

Now consider any integer T sufficiently large that:

$$\max[\frac{Mn^*}{T}, \frac{MN}{T}] \leq \epsilon$$

Then we have:

$$\begin{aligned}
& E(|\frac{\sum_{i=1}^{T} g(X_i)}{T}||X_0 = x, \Gamma_0 = \gamma) \\
\leq \; & E(|\frac{\sum_{i=1}^{n^*} g(X_i)}{T}||X_0 = x, \Gamma_0 = \gamma) \\
+ \; & E(\frac{1}{\lfloor\frac{T-n^*}{N}\rfloor}\sum_{j=1}^{\lfloor\frac{T-n^*}{N}\rfloor}\frac{1}{N}\sum_{k=1}^{N} g(X_{N_1+(j-1)N+k}|X_0 = x, \Gamma_0 = \gamma)) \\
+ \; & E(|\frac{\sum_{n^*+\lfloor\frac{T-n^*}{N}\rfloor N+1}^{T} g(X_i)}{T}||X_0 = x_*, \Gamma_0 = \gamma) \\
\leq \; & \epsilon + \epsilon + \epsilon \\
= \; & 3\epsilon
\end{aligned}$$

Markove's inequality then gives that:

$$P(|\frac{\sum_{i=1}^{T} g(X_i)}{T}| \geq \epsilon^{\frac{1}{2}}|X_0 = x, \Gamma_0 = \gamma)) \leq 3\epsilon^{\frac{1}{2}}$$

Since this holds for all sufficiently large T and since $\epsilon > 0$ was arbitrary, the results follows. □

Secondly we can prove the theorem 3.2 easily using the lemma 5.1.

Proof. Similar to proof of theorem 3.1, the splitting chain of $\{X_n^\gamma\}$ satisfies the conditions of lemma 5.1 for any $\gamma \in \mathcal{Y}$. Therefore we have the WLLN hold. □

6 Recurrence On The Product Space $\mathcal{X} \times \mathcal{Y}$

The adaptive MCMC induces sample paths on the product space $\mathcal{X} \times \mathcal{Y}$. We will study the recurrent property on the product space in this section. When each kernel P_γ has good ergodic property and the random variable sequence (X_n, Γ_n) is also recurrent on the $\mathcal{X} \times \mathcal{Y}$, we hope to get the ergodicity of AMCMC. But following the computation in section 6.1, we get the negative answer. Fortunately Roberts and Rosenthal's paper [14] (2007) offered us a proper condition–"Diminishing Adaptation conditions" and showed some positive results, however they mentioned an open problem as well. We will state the open problem in section 6.2 and give a counter-example to the open problem 21 in Roberts and Rosenthal's paper [14] (2007) in section 6.2.1. Finally we present some positive results about the relationship between ergoidicity and recurrence on the space $\mathcal{X} \times \mathcal{Y}$.

6.1 Recurrence Of Running Example

Even we take finite kernels with good ergodic property(uniformly ergodic) so that we can make the adaptive MCMC recurrent, we still can not guarantee the AMCMC is ergodic with respect to the target distribution π. A good counter example is one-two version running example which was presented in Roberts and Rosenthal(2005) [14] and simulated in the related Java applet. The example was also discussed in Atchade and Rosenthal (2005) [17]. Here we will consider the AMCMC algorithm as a general Markov chain on the product space $\mathcal{X} \times \mathcal{Y}$. We will give the explicit form of the transition matrix on the product space, and analysis the recurrent and ergodic property of such a Markov chain on the product space $\mathcal{X} \times \mathcal{Y}$.

Let $\mathcal{X} = \{1, 2, 3, 4\}$, $\pi(2) = b > 0$ be very small, and $\pi(1) = a$ and $\pi(2) = \pi(3) =$

$\frac{1-a-b}{2} > 0$. Let $\mathcal{Y} = \{1, 2\}$. For $\gamma \in \mathcal{Y}$, let P_γ be the kernel corresponding to a random-walk Metropolis algorithm for $\pi(\cdot)$, with proposal distribution:

$$Q_\gamma(x, \cdot) = Uniform\{x-\gamma, x-\gamma+1, \cdots, x-1, x+1, x+2, \cdots, x+\gamma\}$$

i.e.uniform on all the integers within γ of x, aside from x itself. The kernel P_γ then proceeds, given X_n and Γ_n, by first choosing a proposal state $Y_{n+1} \sim Q_{\Gamma_n}(X_n, \cdot)$. With probability $\min[1, \frac{\pi(Y_{n+1})}{\pi(X_n)}]$ it then accepts this proposal by setting $X_{n+1} = Y_{n+1}$. Otherwise, with probability $1 - \min[1, \frac{\pi(Y_{n+1})}{\pi(X_n)}]$, it rejects this proposal by setting $X_{n+1} = X_n$. (If $Y_{n+1} \notin \mathcal{X}$, then the proposal is always rejected; this corresponds to setting $\pi(y) = 0$ for $y \notin \mathcal{X}$.). We define the adaptive scheme such that $\Gamma_n = 2$ if the previous proposal was accepted, otherwise $\Gamma_n = 1$ if the previous proposal was rejected.

We can compute the kernels induced by the proposals Q_i, $i = 1, 2$:

$$P_1 = \begin{pmatrix} \frac{2a-b}{2a} & \frac{b}{2a} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{b}{1-a-b} & \frac{1}{2} - \frac{b}{1-a-b} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

$$P_2 = \begin{pmatrix} \frac{3}{4} - \frac{b}{4a} & \frac{b}{4a} & \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{a}{2(1-a-b)} & \frac{b}{2(1-a-b)} & \frac{3}{4} - \frac{a+b}{2(1-a-b)} & \frac{1}{4} \\ 0 & \frac{b}{2(1-a-b)} & \frac{1}{4} & \frac{3}{4} - \frac{b}{2(1-a-b)} \end{pmatrix}$$

In the above AMCMC, we can observe that the distribution of Γ_n given X_0 and Γ_0 does NOT depend on the value of $\{X_i | 0 \le i \le n-1\}$, therefore we call this kind of Markovian AMCMC. The $n-$th transition kernel $Q_{(n)}$ induced by Markovian adaptive algorithm is as below:

$$Q^{(n)}((x, \gamma), A \times B) \quad = \quad \int_A \int_B \Gamma_n(d\gamma_1 | x, y, \gamma) P_\gamma(x, dy)$$

Then in the one-two running example, if given the value of $X_{n-1} = x, X_n = y$ and $\Gamma_{n-1} = \gamma$, then Γ_n is a measurable function of x, y and γ. We have:

$$\Gamma_n(x, y, \gamma) = \delta(x = y) + 2\delta(x \neq y)$$

So we can compute the $n-th$ transition kernel on $(\mathcal{X}\times\mathcal{Y})$:

$$\begin{aligned} Q((x,\gamma), y\times\gamma_1) &= \int_A\int_B \Gamma_n(d\gamma_1|x,y,\gamma)P_\gamma(x,dy) \\ &= P_\gamma(x,y)\delta(x=y)\delta(\gamma_1=1)+P_\gamma(x,y)\delta(x\neq y)\delta(\gamma_1=2) \end{aligned}$$

Since the transition kernel is independent of n, the one-two version running example presents a general Markov Chain with transition kernle Q as:

$$Q=\begin{pmatrix} \frac{2a-b}{2a} & 0 & 0 & \frac{b}{2a} & 0 & 0 & 0 & 0 \\ \frac{3}{4}-\frac{b}{4a} & 0 & 0 & \frac{b}{4a} & 0 & \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & 0 & 0 & \frac{b}{1-a-b} & \frac{1}{2}-\frac{b}{1-a-b} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{a}{2(1-a-b)} & 0 & \frac{b}{2(1-a-b)} & \frac{3}{4}-\frac{a+b}{2(1-a-b)} & 0 & 0 & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & \frac{b}{2(1-a-b)} & 0 & \frac{1}{4} & \frac{3}{4}-\frac{b}{2(1-a-b)} & 0 \end{pmatrix}$$

Now we take the value $a=0.1$ and $b=0.01$, then $\pi(1)=0.1$; $\pi(2)=0.01$; $\pi(3)=\pi(4)=0.445$.

And we have the following lemma:

Lemma 6.1. *The above one-two version running example is recurrent, but for any starting value (x_*,γ_*), and $A\in\mathcal{B}\{\mathcal{X}\}$, we have:*

$$\lim_{n\to\infty}P_{(x_*,\gamma_*)}(X_n\in A)\neq\pi(A)$$

Proof. Let us calculate the eigenvalues of the above transition matrix, we have: $\lambda_1=1$; $\lambda_2=0.95445494$; $\lambda_3=0.12887658+0.4670861i$; $\lambda_4=0.12887658-0.4670861i$; $\lambda_5=-0.25615654$; $\lambda_6=0.03778642+0.1057364i$; $\lambda_7=0.03778642-0.1057364i$; $\lambda_8=-0.09286036$. Then compute the eigenvector of Q^T with respect to the eigenvalue $\lambda_0=1$, it is

$$(-0.48637045,-0.03354279,-0.00867102,-0.03468408,$$
$$-0.49208038,-0.36554543,-0.51525761,-0.34609757)$$

i.e the stationary distribution $\tilde{\pi}$ is: $\tilde{\pi}(1,1) = 0.213110130$, $\tilde{\pi}(1,2) = 0.014697250$, $\tilde{\pi}(2,1)0.003799331$, $\tilde{\pi}(2,2) = 0.015197323$, $\tilde{\pi}(3,1) = 0.215612017$, $\tilde{\pi}(3,2) = 0.160168927$, $\tilde{\pi}(4,1) = 0.225767451$, $\tilde{\pi}(4,2) = 0.151647571$. Therefore for any start value (x_*, γ_*), we have:

$$\begin{aligned} \lim_{n\to\infty} P_{(x_*,\gamma_*)}(X_n = 1) &= \lim_{n\to\infty} P_{(x_*,\gamma_*)}(X_n = 1, \Gamma_n = 1) + P_{(x_*,\gamma_*)}(X_n = 1, \Gamma_n = 1) \\ &= 0.21311 + 0.014697 = 0.227807 \end{aligned}$$

similarly

$$\lim_{n\to\infty} P_{(x_*,\gamma_*)}(X_n = 2) = 0.003799 + 0.015197 = 0.018996$$

$$\lim_{n\to\infty} P_{(x_*,\gamma_*)}(X_n = 3) = 0.215612 + 0.160168 = 0.37578$$

$$\lim_{n\to\infty} P_{(x_*,\gamma_*)}(X_n = 4) = 0.225767 + 0.151647 = 0.377414$$

Therefore for any $1 \le i, j \le 4$, we have:

$$E_i[\eta_j] = \infty$$

because $P_i(\eta_j = \infty) = 1$. But we can observe that $P_{(x_*,\gamma_*)}(X_n \in A) \to_{n\to\infty} \pi^{'}(A)$ which is the marginal distribution of $\tilde{\pi}$, however $\pi^{'}(\cdot) \neq \pi(\cdot)$. □

6.2 The Open Problem 21 In Roberts And Rosenthal's Paper

In the Theorem 13 of Roberts and Rosenthal [14] (2007), they present the following results that an adaptive MCMC algorithm with Diminishing Adaptation is ergodic provided that it is recurrent in probability in some sense. Before we state the Theorem 13, let us recall the definition "ϵ convergence time function" $M_\epsilon : \mathcal{X} \times \mathcal{Y} \to N$::

$$M_\epsilon(x, \gamma) = \inf\{n \ge 1 : \|P_\gamma^n(x, \cdot) - \pi(\cdot)\| \le \epsilon\}$$

Obviously if each individual P_γ is ergodic, then $M_\epsilon(x, \gamma) < \infty$.

Theorem 6.1. *Consider an adaptive MCMC algorithm with Diminishing Adaption (i.e., $\lim_{n\to\infty} \sup_{x\in\mathcal{X}} \|P_{\Gamma_{n+1}}(x, \cdot) - P_{\Gamma_n}(x, \cdot)\| = 0$ in probability). Let $x_* \in \mathcal{X}$ and $\gamma_* \in \mathcal{Y}$.*

Then $\lim_{n\to\infty} T(x_*, \gamma_*, n) = 0$ *provided that for all* $\epsilon > 0$*, the sequence* $\{M_\epsilon(X_n, \Gamma_n)\}_{n=0}^{\infty}$ *is bounded in probability given* $X_0 = x_*$ *and* $\Gamma_0 = \gamma_*$*,i.e. for all* $\delta > 0$*, there is* $N \in \mathbb{N}$ *such that* $P[M_\epsilon(X_n, \Gamma_n) \leq N | X_0 = x_*, \Gamma_0 = \gamma_*] \leq 1 - \delta$ *for all* $n \in \mathbb{N}$*.*

We can observe that in the above theorem the adaptive chain pair (X_n, Γ_n) has good "fast convergence" property in probability. Therefore this leads to the following open problem using recurrence concept.

Open Problem 21. Consider an adaptive MCMC algorithm with Diminishing Adaptation. Let $x_* \in \mathcal{X}$ and $\gamma_* \in \mathcal{Y}$. Suppose that for all $\epsilon > 0$, there is $m \in \mathbb{N}$ such that $P[M_\epsilon(X_n, \Gamma_n) < m \text{ i.o.} | X_0 = x_*, \Gamma_0 = \gamma_*] = 1$. Does this imply that $\lim_{n\to\infty} T(x_*, \gamma_*, n) = 0$?

The problem seems reasonable, however the following example gives us the negative answer.

6.2.1 The Counterexample To The Open Problem

Let us see the following example:

Consider $\mathcal{X} = R \bmod Z$ i.e. the state space is the real number mod the integers. Define $\mathcal{Y} = \mathbf{N} \cup \mathcal{X}$, and suppose $Z_{k,x}$ are random variable with distribution $Uniform[x - \frac{1}{2^{k+1}}, x + \frac{1}{2^{k+1}}]$ for any $(x, \gamma) \in \mathcal{X} \times \mathcal{Y}$. When $k \in N$, we define:

$$P_k(x, A) = \frac{1}{2^k} P(Z_{k,x} \in A) + (1 - \frac{1}{2^k}) \delta_x(A)$$

When $y \in \mathcal{X}$, suppose $\pi(\cdot)$ is the Lebesgue measure on $\mathcal{X}$.

we define:

$$P_y(x, A) = \begin{cases} \frac{2}{3}\pi(A) + \frac{1}{3}\delta_x(A) & x \neq y \\ \frac{2}{3}Uniform[0, \frac{3}{4}] + \frac{1}{3}\delta_0(A) & x = y \end{cases}$$

Lemma 6.2. *For each* $k \in \mathbb{N}$*,* P_k *is stationary with respect to* π*.*

Proof. It is suffice to prove that for any interval $A = [a, b] \subset [0, 1]$ we have:

$$\int_{\mathcal{X}} P_k(x, A)\pi(dx) = \pi(A)$$

Case 1:$|b-a| \geq \frac{1}{2^k}$

$$\begin{aligned}
\int_{\mathcal{X}} P_k(x,A)\pi(dx) &= \frac{1}{2^k} \times \int_0^1 P(Z_{x,k} \in A)dx + (1-\frac{1}{2^k})\pi(A) \\
&= \frac{1}{2^k} \times [2^k \int_{a-\frac{1}{2^{k+1}}}^{a+\frac{1}{2^{k+1}}} [x + \frac{1}{2^{k+1}} - a]dx + 2^k \int_{b-\frac{1}{2^{k+1}}}^{b+\frac{1}{2^{k+1}}} [-x + \frac{1}{2^{k+1}} + b]dx \\
&+ (b-a-\frac{1}{2^k})] + (1-\frac{1}{2^k})\pi(A) \\
&= \frac{1}{2^k} \times [2^{k+1} \int_0^{\frac{1}{2^k}} tdt + (b-a-\frac{1}{2^k})] + (1-\frac{1}{2^k})\pi(A) \\
&= b-a
\end{aligned}$$

similarly we can prove **Case 2:**$|b-a| < \frac{1}{2^k}$. □

Lemma 6.3. *For each $y \in \mathcal{X}$, P_y is stationary with respect to π.*

Proof.

$$\begin{aligned}
\int_{\mathcal{X}} P_y(x,A)\pi(dx) &= \int_{x\neq y} [\frac{2}{3}\pi(A) + \frac{1}{3}\delta_x(A)]\pi(dx) \\
&= \frac{2}{3}\pi(A) + \frac{1}{3}\pi(A) \\
&= \pi(A)
\end{aligned}$$

□

Define the independent random variable I_n as below:

$$I_n = \begin{cases} 1 & \text{w.p.} \frac{\sqrt{n}-1}{\sqrt{n}} \\ 0 & \text{w.p.} \frac{1}{\sqrt{n}} \end{cases}$$

And independent random variable Y_n as below: $Y_0 = Y_1 = 1$ and

$$Y_n = \begin{cases} n+1 & \text{with probability } \frac{1}{n} \\ n+2 & \text{with probability } \frac{1}{n} \\ . & \\ . & \\ . & \\ 2n & \text{with probability } \frac{1}{n} \end{cases}$$

Define the adaptive scheme as:

$$\Gamma_n = \begin{cases} Y_n & \text{if } I_n = 1 \\ X_n & \text{if } I_n = 0 \end{cases}$$

Lemma 6.4. *Such an adaptive scheme satisfies the diminishing condition.*

Proof. Actually $P_{Y_n}(x, A) = \frac{1}{n}\sum_{i=n+1}^{2n} P_i(x, A)$, so

$$\begin{aligned}
& |P_{\Gamma_{n+1}}(x, A) - P_{\Gamma_n}(x, A)| \\
\leq\ & |P_{Y_{n+1}}(x, A) - P_{Y_n}(x, A)| + P(I_n = 0 \text{ or } I_{n+1} = 0) \\
\leq\ & |\frac{1}{n+1}\sum_{i=n+2}^{2n+2} P_i(x, A) - \frac{1}{n}\sum_{i=n+1}^{2n} P_i(x, A)| + \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n+1}} \\
\leq\ & \frac{1}{n(n+1)}\sum_{i=n+2}^{2n} P_i(x, A) + \frac{1}{n(n+1)}|P_{n+2}(x, A) + P_{2n+2}(x, A) - P_{n+1}(x, A)| + \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n+1}} \\
\leq\ & \frac{1}{n} + \frac{3}{n(n+1)} + \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{n+1}} \\
\rightarrow\ & 0 \text{ as } n \rightarrow \infty
\end{aligned}$$

□

Lemma 6.5. *Given $x_* = 0$ and $\gamma_* = 0$. Then for any $\epsilon > 0$, there is $m \in N$ such that:*

$$P[(X_n, \Gamma_n) \in \mathcal{Z}_{m,\epsilon} \quad i.o. \quad |X_* = 0, \Gamma_0 = 0] = 1$$

Proof. We know P_0 is uniformly ergordic with respect to $\pi(\cdot)$, so for any $\epsilon > 0$ there exists m such that:

$$\|P_0^m(0, \cdot) - \pi(\cdot)\| < \epsilon \tag{6.2}$$

If we suppose

$$J = \begin{cases} 1 & \text{w.p.} \frac{2}{3} \\ 0 & \text{w.p.} \frac{1}{3} \end{cases}$$

Then we can consider $P_x(x, A)$ as the following: if $J = 0$, the chain will move to 0, otherwise select one point on the interval $[0, \frac{3}{4}]$ with uniform distribution.

And we have:

$$P[X_{n+1} = 0, \Gamma_{n+1} = 0 \text{ i.o.}] \geq P[I_n = 0, I_{n+1} = 0 \text{ and} J = 0 \text{ i.o}]$$

since $\sum_{i=1}^{\infty} P(I_{2i} = 0, I_{2i+1} = 0, J = 0) = \sum_{i=1}^{\infty} \frac{1}{3} \frac{1}{\sqrt{2i(2i+1)}} = \infty$. That is:

$$P[I_{2n} = 0, I_{2n+1} = 0 \text{ and} J = 0 \text{ i.o}] = 1$$

Therefore $P[(X_n, \Gamma_n) = (0, 0) \text{ i.o.}] = 1$. Following (6.2) we know that

$$\begin{aligned} 1 &\geq P[(X_n, \Gamma_n) \in \mathcal{Z}_{m,\epsilon} \;\; i.o. \;\; |X_* = 0, \Gamma_* = 0] \quad (6.3) \\ &\geq P[(X_n, \Gamma_n) = (0, 0) \;\; i.o. \;\; |X_* = 0, \Gamma_* = 0] = 1 \quad (6.4) \end{aligned}$$

□

Lemma 6.6. *Suppose $\{a_i\}_{i=1}^{\infty}$ is a decreasing positive sequence such that $0 < a_i < 1$, and if $\sum_{i=1}^{\infty} a_i < \infty$, then*

$$\lim_{N \to \infty} \prod_{i=N}^{\infty} (1 - a_i) = 1 \tag{6.5}$$

Proof. When $0 < a_i < 1$, we have:

$$ln(1 - a_i) \leq -a_i$$

Therefore

$$\begin{aligned} 1 &\geq \lim_{N \to \infty} \prod_{i=N}^{\infty} (1 - a_i) \\ &\geq \lim_{N \to \infty} e^{\sum_{i=N}^{\infty}(-a_i)} \\ &= 1 \end{aligned}$$

□

Lemma 6.7. *Given $X_* = 0$ and $\Gamma_* = 0$, we do NOT have* $\lim_{n \to \infty} T(x_*, \gamma_*, n) = 0$

Proof. Suppose $\lim_{n \to \infty} T(x_*, \gamma_*, n) = 0$, that is for any $\epsilon > 0$, there exists N_1 such that for any $n > N$ and $A \in \mathcal{B}(\mathcal{X})$,

$$|P[X_n \in A | X_* = 0, \Gamma_* = 0] - \pi(A)| < \epsilon \tag{6.6}$$

According to the above adaptive scheme, if $\Gamma_n \in [0, 1]$, then Γ_n must be equal to X_n, in other words the case of kernel $P_y(x, \cdot)$ but $y \neq x$ will NOT happen in this adaptive

Markov Chain. So if $X_n \in [0, \frac{3}{4}]$, there are four cases maybe happen at X_{n+1}

Case 1: $X_{n+1} = X_n$

Case 2: $X_{n+1} = 0$

Case 3: $X_{n+1} = Z_{x_n,n}$

Case 4: $X_{n+1} \sim Uniform[0, \frac{3}{4}]$

Only in the case 3, X_{n+1} maybe jump out of $[0, \frac{3}{4}]$, so $P(X_{n+1} \in [0, \frac{3}{4}] | X_n \in [0, \frac{3}{4}]) > 1 - \frac{1}{2^n}$. Since this is a Markovian adaptive MCMC,

$$\begin{aligned} & P(X_{n+2} \in [0, \frac{3}{4}] | X_n \in [0, \frac{3}{4}]) \\ \geq \; & P(X_{n+2} \in [0, \frac{3}{4}] | X_{n+1} \in [0, \frac{3}{4}]) P(X_{n+1} \in [0, \frac{3}{4}] | X_n \in [0, \frac{3}{4}]) \\ \geq \; & (1 - \frac{1}{2^n})(1 - \frac{1}{2^{n+1}}) \end{aligned}$$

Similarly for any $m > 0$, we have:

$$P(X_{n+m} \in [0, \frac{3}{4}] | X_n \in [0, \frac{3}{4}]) \geq \prod_{i=n}^{n+m-1} (1 - \frac{1}{2^i}) \tag{6.7}$$

Following lemma 6.6 we select $N_2 > 0$ such that $\prod_{i=N_2}^{\infty} (1 - \frac{1}{2^i}) > 1 - \frac{\epsilon}{2}$ Let $N = max\{N_1, N_2\}$, then following (6.4) there exist K large enough such that:

$$P[\exists N \leq n < N^K \text{ such that } (X_n, \Gamma_n) = (0,0)] > \frac{\frac{3}{4} + 2\epsilon}{1 - \frac{\epsilon}{2}} \tag{6.8}$$

whenever $(X_n, \Gamma_n) = (0,0)$, then X_{n+1} must be in $[0, \frac{3}{4}]$, so following (6.7) we have:

$$\begin{aligned} & P(X_{N^K+1} \in [0, \frac{3}{4}]) \\ = \; & P(X_{N^K+1} \in [0, \frac{3}{4}] | \exists N < n \leq N^K \text{ s.t. } X_n \in [0, \frac{3}{4}]) \cdot P(\exists N < n \leq N^K \text{ s.t. } X_n \in [0, \frac{3}{4}]) \\ \geq \; & \prod_{i=N}^{\infty} (1 - \frac{1}{2^i}) \cdot P[\exists N < n \leq N^K \text{ s.t. } (X_{n-1}, \Gamma_{n-1}) = (0,0)] \\ \geq \; & (1 - \frac{\epsilon}{2}) \times \frac{\frac{3}{4} + 2\epsilon}{1 - \frac{\epsilon}{2}} \\ = \; & \frac{3}{4} + 2\epsilon \end{aligned}$$

Which is conflicting with (6.6). □

6.3 Strengthen The Diminishing Adaption Condition

Following the counterexample in the section 6.2.1, we know that the Diminishing Adaption condition and the recurrence property to the "good convergence" set are not sufficient to get the ergodicity of the AMCMC. Therefore we can strengthen the Diminishing Adaption condition such that it can match with the recurrence condition, so that we can use the coupling methods to prove the ergodicity.

For any $m \in \mathbb{N}$ and $\epsilon > 0$, we can define the $i-$th hitting time $\tau_{x,\gamma}^{(i)}(m, \epsilon)$ as below:

$$\tau_{x,\gamma}^{(i)}(m, \epsilon) = \min\{n > \tau_{x,\gamma}^{(i-1)}(m, \epsilon) | M_\epsilon(X_n, \Gamma_n) \leq m \text{ given } X_0 = x, \Gamma_0 = \gamma\}$$

and the hitting number within n step

$$c_{x,\gamma}^{m,\epsilon}(n) = \text{ the number of } \{0 \leq j \leq n | M_\epsilon(X_j, \Gamma_j) \leq m \text{ given } X_0 = x, \Gamma_0 = \gamma\}$$

Furthermore we can define:

$$s_{x,\gamma}^{(i)}(m, \epsilon) = \sum_{j=\tau_{x,\gamma}^{(i)}(m,\epsilon)+1}^{\tau_{x,\gamma}^{(i+1)}(m,\epsilon)} D_j$$

Then we have the following theorem:

Theorem 6.2. *Consider an adaptive MCMC algorithm , let $x_* \in \mathcal{X}$ and $\gamma_* \in \mathcal{Y}$. Suppose that for all $\epsilon > 0$, there is $m \in \mathbb{N}$ such that $P[M_\epsilon(X_n, \Gamma_n) < m$ i.o.$|X_0 = x_*, \Gamma_0 = \gamma_*] = 1$ and $s_{x,\gamma}^{(i)}(m, \epsilon) \to_{i\to\infty} 0$ in probability. Then* $\lim_{n\to\infty} T(x_*, \gamma_*, n) = 0$.

Proof. For any $\epsilon > 0$, there is $m \in \mathbb{N}$ such that

$$P[M_\epsilon(X_n, \Gamma_n) < m \text{ i.o.} | X_0 = x_*, \Gamma_0 = \gamma_*] \quad = \quad 1$$

and there exists $N_1 > 0$ such that for any $n > N_1$ we have:

$$P\left[\sum_{j=n}^{n+m} s_{x,\gamma}^{(i)}(m, \epsilon) > \epsilon\right] \quad \leq \quad \epsilon$$

Following $P[M_\epsilon(X_n, \Gamma_n) < m$ i.o.$|X_0 = x_*, \Gamma_0 = \gamma_*] = 1$, we know that there is $N > 0$ such that

$$P[c_{x,\gamma}^{m,\epsilon}(N) > N_1 + m] > 1 - \epsilon \tag{6.9}$$

Consider any $n > N$, the above formula indicates that:

$$P[\exists k > N_1 + m \text{ such that } \tau_{x,\gamma}^{(k)}(m,\epsilon) \leq n < \tau_{x,\gamma}^{(k+1)}(m,\epsilon)] > 1-\epsilon$$

We set $l = \tau_{x,\gamma}^{(k-m)}(m,\epsilon)$. we can construct a second chain $\{X_i^{'}\}_{i=l}^{n}$ such that $X_l^{'} = X_l$ and $X_i^{'} \sim P_{\Gamma_l}(X_{i-1},\cdot)$ for $l \leq i \leq n$. If we denote the event $E = \{\sum_{i=l}^{n} P(X_i^{'} \neq X_i) < \epsilon\}$, then from (5.8) we have:

$$P[E] > 1-\epsilon$$

On the other hand we have:

$$\|P_{\Gamma_l}^{n-l}(X_l,\cdot) - \pi(\cdot)\| \quad \leq \quad \|P_{\Gamma_l}^{l+m}(X_l,\cdot) - \pi(\cdot) \leq \epsilon$$

we can construct $Z \sim \pi(\cdot)$, then

$$\begin{aligned}
& \|P(X_n \in \cdot|X_0 = x, \Gamma_0 = \gamma) - \pi(\cdot)\| \\
\leq \;& P(X_n \neq Z|X_0 = x, \Gamma_0 = \gamma) \\
\leq \;& P(X_n \neq X_n^{'}, E|X_0 = x, \Gamma_0 = \gamma) + P(X_n^{'} \neq Z, E|X_0 = x, \Gamma_0 = \gamma) + P(E^c|X_0 = x, \Gamma_0 = \gamma) \\
\leq \;& 3\epsilon
\end{aligned}$$

i.e. $T(x,\gamma,n) < 3\epsilon$. □

Following theorem 6.2, we can get the following corollary easily.

Corollay 6.8. *Consider an adaptive MCMC algorithm such that* $\sum_{i=1}^{\infty} D_i < \infty$ *in probability. Let* $x_* \in \mathcal{X}$ *and* $\gamma_* \in \mathcal{Y}$*. Suppose that for all* $\epsilon > 0$*, there is* $m \in \mathbb{N}$ *such that* $P[M_\epsilon(X_n,\Gamma_n) < m \text{ i.o.}|X_0 = x_*, \Gamma_0 = \gamma_*] = 1$*. Then* $\lim_{n\to\infty} T(x_*,\gamma_*,n) = 0$.

Proof. Since $\sum_{i=1}^{\infty} D_i < \infty$ in probability, we know that $s_{x,\gamma}^{(i)}(m,\epsilon) \to_{i\to\infty} 0$ in probability. Therefore following the theorem 6.2, we have the conclusion. □

Acknowledgements. We would like to thank Prof. Jeffrey Rosenthal for his help in writing this paper.

REFERENCES

[1] Adrieu, C. & Atchade, Y. F. (2005). On the efficiency of adaptive MCMC algorithms

[2] Andrieu, C. & Moulines, E. (2005). On the ergodicity properties of some Adaptive MCMC Algorithms. *To appear Ann. Appl. Probab.*.

[3] Andrieu, C. & Roberts, C. P. (2001). Conrolled MCMC for optimal samling. Technique report. University Paris Dauphine, Ceremade 0125.

[4] Gelman, A. G. & Roberts, G. O. & Gilks, W.R. (1996). Efficient Metropolis jumping rules. In J.M. Bernardo, J.O. Berger, A.F. David and A.F.M. Smith (eds), Bayesian Statistics V. pp. 599-608. Oxford: Oxford University Press.

[5] Gilks, W.R. Richardson, S. & Spiegelhalter, D. J. (1996). Markov chain Monte Carlo in practice. *Interdisciplinary Statistics*, Chapman & Hall, London.

[6] Gilks, W.R.Roberts, G. O. & Sahu,S.K. (1998). Adaptive Markov chain Monte Carlo through regeneration.*J. Amer. Statist. Assoc.*, **93**, 1045-1054.

[7] Haario, H. , Saksman, E. (1991). Simulated annealing process in general state space. *Adv. Appl. Probab.*, **23**, 866-893

[8] Haario, H. , Saksman, E. & Tamminen, J. (1999). Adaptive proposal distribution for random walk Metropolis algorithm. *Comput. Statist.* **14**, 375-395

[9] Haario, H. , Saksman, E. & Tamminen, J. (2001). An adaptive metropolis algorithm. *Bernoulli* **7**, 223-242

[10] Haario, H. , Saksman, E. & Tamminen, J. (2001). An adaptive metropolis algorithm. *Bernoulli* **7**, 223-242

[11] Meyn, S. P. & Tweedie, R. L. (1993). Markov Chains and Stochasitic Stability. *Springer-Verlag*

[12] Roberts, G. O. & Rosenthal, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probability Surveys* 1:20-71, 2004.

[13] Roberts, G. O. & Rosenthal, J. S. (2007). Coupling and ergodicity of adaptive MCMC (See also the related Java applet.)*J. Appl. Prob.*, **44** 458-475.

[14] Roberts, G. O. & Rosenthal, J. S. (2005). Example of adaptive MCMC.

[15] Roberts, G. O. & Rosenthal, J. S., and & Schwartz P.O. (1998). Convergence properties of perturbed Markov chains. *J. Appl. Prob.*, **35**, 1-11.

[16]Rosenthal, J. S. & Atchade, Y. F. (2005). On adaptive Markov chain Monte Carlo algorithm *Bernoulli*, **11** 815-828.

[17]Rosenthal, J. S. & Tierney, L. (1994). Markov chains for exploring posterior distribution (with discussion) *Ann. Stat.*, **22** 1701-1762.

[18]Yang, C. (2007). On the weak law of large numbers for unbounded functionals for adaptive MCMC *To be submitted*.